Marius IOSIFESCU, Costache MOINEAGU, Vladimir TREBICI
Emiliana URSIANU

MICĂ ENCICLOPEDIE DE STATISTICĂ

Coordonator Vladimir Trebici

Colegului și prietenului dr. Dumitru SANDU
cu caldă apreciere pentru
ceea ce realizează.
Omagiu
V. Trebici
15 aprilie 1985.

EDITURA ȘTIINȚIFICA ȘI ENCICLOPEDICA
București, 1985

V.TREBICI

Cap. IV. *Prelucrarea datelor statistice* **a fost elaborat de dr. ing. Nicolae Costake**

Redactor coordonator: Maria Blaga

Coperta și supracoperta: Simona DUMITRESCU

V.TREBICI

CUVÎNT ÎNAINTE

Dezvoltarea impresionantă a științei și explozia informațională — tendințe atît de caracteristice epocii contemporane — obligă obiectiv fiecare știință să recurgă la formula enciclopedică de prezentare a noțiunilor sale, ca una care să înlesnească specialiștilor și cercurilor largi de cititori o informare rapidă și economicoasă, la nivelul actual al științei respective. Nevoia tot mai accentuată de cunoaștere nu are numai un caracter abstract, teoretic; știința s-a transformat, sub ochii noștri, în ultimele decenii, într-o autentică forță de producție. Importanța ei a fost statornic subliniată în documentele de partid și de stat, ea ocupă un loc central în gîndirea Președintelui țării, tovarășul Nicolae Ceaușescu. „Partidul pornește — se spune în Programul P.C.R. — de la considerentul că știința constituie factorul primordial al progresului contemporan, că societatea socialistă multilateral dezvoltată și comunismul nu pot fi edificate decît pe baza celor mai înalte cuceriri ale științei și tehnicii"[1]. Lucrarea „Mică enciclopedie de statistică" pe care o oferim cititorilor din țara noastră încearcă să răspundă îndemnului partidului de a da lucrări științifice originale, în spiritul tradiției științei românești și care să fie în același timp utile în activitatea practică.

Progresele statisticii, pe plan național și internațional, sînt cu adevărat spectaculoase, după cum atestă literatura din ultimele trei-patru decenii. Dezvoltarea teoretică a acestei științe multicentenare este însoțită de extinderea considerabilă a aplicabilității sale practice. Numeroasele discipline care au apărut în ultimul timp, cu pronunțat caracter cantitativ, nu numai că nu au redus importanța statisticii, dar dimpotrivă, într-un grad sau altul, o implică. „Statisticizarea" este un proces care a cuprins în orbita sa mai toate științele, inclusiv pe cele sociale.

O lucrare enciclopedică despre statistică a devenit așadar o necesitate obiectivă. Literatura internațională consemnează cîteva lucrări; în țara noastră au existat de asemenea încercări de acest gen, printre care se cer menționate cele două ediții (1962 și 1969) ale „Dicționarului statistic-economic" al Direcției Centrale de Statistică, la elaborarea căruia au participat numeroși specialiști, printre care și autorii prezentei lucrări.

Folosind experiența națională și cea internațională, autorii „Micii enciclopedii de statistică" au recurs, după o îndelungată chibzuință, la o formulă originală, ținînd seama de necesitățile specifice de informare din țara noastră, dar, în același timp, de nivelul actual de dezvoltare a statisticii pe plan internațional. Din cele două modalități de expunere a materiei — logică și alfabetică — a fost aleasă prima. Ordinea alfabetică, oricît de avantajoasă ar părea la prima

---

1 Programul Partidului Comunist Român de făurire a societății socialiste multilateral dezvoltate și înaintare a României spre comunism, Ed. politică, București, 1975, p. 95.

vedere, ar restrînge atît scopul unei lucrări enciclopedice cît și cercul utilizatorilor ei : într-o asemenea formă, lucrarea s-ar adresa specialiștilor cărora ea ar putea să le procure lesnicios o informație sau alta, fiind așadar mai curînd un „memento" sau un „formular". Expunerea logică își dovedește superioritatea prin aceea că îmbină avantajul informativ cu cel formativ. Cititorul care parcurge sistematic „Mica enciclopedie de statistică" își însușește progresiv noțiunile și metodele ajungînd, la capătul lecturii, să aibă o imagine coerentă despre statistică.

În această concepție, lucrarea noastră se deosebește de cele similare editate în străinătate. Lucrarea lui M. G. Kendall și W. R. Buckland („A Dictionary of Statistical Terms", 1966) este o prezentare succintă, în ordine alfabetică, a principalelor noțiuni de teoria probabilităților și de statistică matematică. Lucrarea lui I. Paenson („Systematic Glossary of the Terminology of Statistical Methods", 1970, în limbile engleză, franceză, rusă, spaniolă), expusă tematic, este destinată în special traducătorilor. În schimb, lucrarea lui W. H. Kruskal și J. M. Tanur („International Encyclopedia of Statistics", 2 volume, 1979), prezintă 75 de articole despre statistică, scrise de un mare număr de specialiști, însoțite de 42 articole de științe sociale, cu referire specială la statistică, precum și 57 biografii de personalități care au ilustrat statistica. Desigur, fiecare din aceste lucrări, în funcție de concepția de bază, are o finalitate circumscrisă. Adoptarea ordinii logice își găsește expresia în structura „Micii enciclopedii de statistică". Părțile întîi și a doua oferă noțiuni generale și sistemul informațional statistic ; părțile a treia, a patra și a cincea expun materia propriu-zisă după demersul gnoseologic : statistica descriptivă — teoria probabilităților — statistica matematică. Într-o formă simplificată, demersul ar însemna trecerea de la statistica elementară la statistica matematică, legătura fiind asigurată de noțiunea de probabilitate. De aceea, în cuprinsul lucrării, se întîlnesc situații — dealtfel puține la număr — cînd unul și același termen este prezentat de două ori : prima dată, la nivelul statisticii descriptive, într-o formă mai curînd intuitivă, și a doua oară, în cadrul statisticii matematice, în formulare riguros matematică.

S-a socotit util ca „Mica enciclopedie de statistică" să dea și unele informații istorice, atît sub formă de istoric general al statisticii și de istorie a statisticii românești, cît și sub forma unor notițe istorice pentru fiecare noțiune sau metodă mai importante. De asemenea au fost incluse 84 biografii ale unor personalități care au dat contribuții importante la progresul statisticii. Un sistem de trimiteri asigură legăturile necesare dintre noțiuni, iar indicele alfabetic al termenilor înlesnește un acces rapid la informația căutată de cititor.

Desigur, ca orice lucrare de tip enciclopedic, lucrarea este susceptibilă de îmbunătățire. Autorii vor primi cu recunoștință observațiile și sugestiile cititorilor. Gîndul care i-a călăuzit a fost acela de a pune la dispoziția celor dornici să se instruiască și să se informeze în problemele statisticii o lucrare utilă, rod al experienței lor didactice și științifice de cîteva decenii. Dacă acest scop va fi atins aceasta o vor hotărî cititorii. Din parte-le, autorii însoțesc modesta lor lucrare cu clasicul :

„Habent sua fata libelli".

V.TREBICI

Lista de abrevieri

C.O. = caracteristică operativă
d.p. = densitate de probabilitate
f.f. = funcție de frecvență
f.r. = funcție de repartiție
g.l. = grade de libertate
i. = indici
i.a. = ipoteză alternativă
i.n. = ipoteză nulă
m.l. = m.l.-estimator
p.s. = prag de semnificație
s.c. = serie cronologică
t.s. = test statistic
v.a. = variabilă aleatoare

SUMAR

CUVÎNT ÎNAINTE 5
LISTA DE ABREVIERI 7

PARTEA ÎNTÎI. **GENERALITĂȚI** 17

I. *CONCEPTUL DE STATISTICA* 17
Statistică 17
Teoria probabilităților 31
Statistică matematică 34
Statistică descriptivă 36
Statistică informațională Onicescu 37
Statistică economică 39
Statistică analitică (inductivă sau inferențială) 40
Statistică teoretică (teorie a statisticii) 41
Statistică de ramură 41
Statistică actuarială 42
Statistică oficială 42

II. *CATEGORIILE ȘTIINȚEI STATISTICE* 43
Fenomen de masă 43
Lege statistică (stohastică) 43
Relații cauzale statistice (legături statistice) 44
Legea numerelor mari 45
Model (statistic) 46
Populație (statistică) 48
Eșantion (statistic) 49
Unitate statistică (individ) 49
Caracteristică statistică 50
Variabilă statistică 50
Variabilă aleatoare 51
Repartiție statistică 52
Parametru statistic 52
Indicator statistic 53

PARTEA A DOUA. **SISTEMUL INFORMAȚIONAL STATISTIC** 54

III. *CULEGEREA DATELOR STATISTICE* 54
Observare statistică (înregistrare statistică) 54
Eroare statistică 55
Recensământ 56
Recensământ al populației 56
Recensământ economic 57
Recensământ experimental (de probă) 57
Microrecensământ 57
Anchetă statistică 57
Anchetă prin sondaj (statistic) 57
Anchetă prin sondaj aleator 58
Anchetă prin sondaj nealeator 58
Anchetă (sondaj) de opinie 58
Panel 58
Chstionar 58
Experiment 58

IV. *PRELUCRAREA DATELOR STATISTICE* 61
Prelucrarea datelor statistice 61
Grupare statistică 61
Prelucrarea datelor statistice cu ajutorul mijloacelor tehnicii de calcul. 62
Organizarea prelucrării datelor statistice 63
Unitate de informatică statistică 65
Sistem informațional statistic 65
Flux informațional statistic 66
Nomenclator statistic unitar. 66
Registru statistic automat 68
Sistem informatic statistic 68
Calculator electronic (utilizat în prelucrarea datelor statistice) . . 71
Programe și sisteme de programe standard pentru prelucrarea datelor statistice 72
Bază de metadate statistice 73
Bază de date statistice 73
Bancă de date statistice 74

Prelucrare automată a datelor statistice operative 74
Prelucrare automată a datelor recensământului populației și al locuințelor 74
Prelucrare integrată a datelor demografice 75
Integrarea sistemului informatic statistic în sistemul informatic național 76

V. *PREZENTAREA DATELOR STATISTICE* 76
Tabel statistic 76
Tabel cu dublă intrare 78
Tabel de asociație 80
Tabelă (statistică) 81
Reprezentare grafică 81
Diagramă în batoane (bare) 85
Diagramă cumulativă 86
Diagramă prin coloane (diagramă prin tuburi de orgă) 86
Diagramă circulară (în sectoare) 88
Diagramă prin suprafețe 89
Historiogramă (cronodiagramă) 89
Diagramă semilogaritmică 91
Diagramă logaritmică (diagramă log—log) 98
Histogramă 99
Poligon al frecvențelor 102
Curbă a frecvențelor (curbă de repartiție) 105
Diagramă polară (radială) 109
Diagramă triunghiulară 111
Diagramă în z 112
Grafic Gantt 113
Piramidă a vîrstelor 114
Graficul lui Lexis 116
Cartogramă 120
Cartodiagramă 120
Corelogramă 122
Stereogramă 122
Diagramă Barlow-Campo 122

PARTEA A TREIA. **STATISTICA DESCRIPTIVĂ** 123

VI. *REPARTIȚII EMPIRICE (DE FRECVENȚE)* 123
Serie statistică (repartiție statistică) 123
Clasă 124

Frecvență 125
Repartiție de frecvențe 127

VII. *TENDINȚA CENTRALĂ (MEDII UZUALE)* 133
Valoare medie (medie) 133
Medie aritmetică 135
Medie armonică 140
Medie geometrică 143
Medie pătratică 147
Mediană 148
Cuantile (de ordinul *n*) 152
Mod 154
Medie cronologică 158
Condițiile lui Yule 159
Momente 160

VIII. *DISPERSIA (IMPRĂȘTIEREA)* 163
Parametrii dispersiei 163
Amplitudine (interval de variație, ω) 163
Abatere cuartilă (interval intercuartil, intercuartilă) 164
Abatere medie (abatere absolută medie) 165
Abatere mediană (abatere probabilă, abatere echiprobabilă) . . . 167
Abatere pătratică medie (abatere standard, abatere tip, σ) . . . 167
Diferență medie a lui Gini (*Dm*) 172
Parametrii formei (repartiției) 174

IX. *CORELAȚIA ȘI REGRESIA* 181
Corelație (statistică) 181
Regresie 182
Tabel de corelație 183
Corelație aparentă (corelație falsă) 183
Ecuație de regresie (ecuație de estimare) 187
Dreaptă de regresie (dreaptă de estimare) 189
Coeficient de regresie 193
Repartiție marginală 194
Repartiție condiționată (repartiție legată) 195
Covarianță 196
Coeficient de corelație (coeficient de corelație liniară, coeficient de corelație Bravais-Pearson) 197
Raport de corelație (indice de corelație) 202

Coeficient de corelaţie biserial 206
Coeficient de corelaţie tetrachoric 206
Corelaţie multiplă 206
Corelaţie a rangurilor 208

PARTEA A PATRA. **INDICII ŞI SERIILE CRONOLOGICE** . . . 213

X. *INDICII* 213
Indice 213
Indice elementar (individual) 221
Indice sintetic (total, general) 221
Pondere 222
Indice simplu 222
Indice ponderat 222
Indice general 222
Indice parţial (de grup) 222
Indice agregat 222
Perioadă de bază (de referinţă) 223
Perioadă curentă 223
Indice mediu 223
Serie de indici 223
Indice Laspeyres 224
Indice Paasche 224
Indice Edgeworth (Marshall-Edgeworth) 224
Indice „ideal" Fisher 224
Indice Divisia 224
Formula Bortkiewicz 226
Indice aritmetic mediu ponderat 227
Indice armonic mediu ponderat 227
Teste ale indicilor 227
Indice analitic 230
Indice al valorii medii 230
Indice cu structură variabilă 231
Indice cu structură fixă 231
Indice de structură (al schimbărilor structurale) 231
Indice standardizat 232
Ponderi standard 233
Indice teritorial 233
Determinarea contribuţiei factorilor la modificarea unui indicator complex 233

Indice al preţurilor 234
Indice al costului vieţii 235
Indice al preţurilor la consumator 235
Coeficient bugetar 235
Indice cu utilitate constantă 236
Indice al volumului fizic al producţiei 236
Foarfecele preţurilor 237
Indice al puterii de cumpărare a monedei 237
Racordarea indicilor 237

XI. *SERIILE CRONOLOGICE* 239
Serie cronologică (serie în timp, serie temporală, cronică) . . 239
Descompunerea seriilor cronologice 245
Ajustarea seriei cronologice 247
Analiza tendinţei generale (seculare) a serici cronologice . . . 251
Analiza variaţiilor (fluctuaţiilor) sezoniere ale serici cronologice 281
Analiza variaţiilor (fluctuaţiilor) ciclice 292
Analiză spectrală a seriei cronologice 295
Extrapolarea seriei cronologice 296
Ciclu economic (ciclu de afaceri) 297
Conjunctură economică 297
Econometrie 297

PARTEA A CINCEA. **TEORIA PROBABILITĂŢILOR ŞI STATISTICA MATEMATICĂ** . . . . . .

XII. *ELEMENTE DE TEORIA PROBABILITĂŢILOR* . . . 300
Spaţiu de evenimente elementare 300
Eveniment (aleator). 300
Probabilitate 302
Cimp de probabilitate 305
Evenimente independente 305
Evenimente dependente 305
Probabilitate condiţionată 305
Variabilă aleatoare 306
Repartiţie (a unei variabile aleatoare) 308
Funcţie de repartiţie 308
Operaţii cu variabile aleatoare 309
Valori medii 309
Valori medii condiţionate 312

Variabile aleatoare independente 313
Proces aleator (stohastic) cu parametru discret 314
Lanț 314
Proces aleator (stohastic) cu parametru continuu 315

XIII. *LEGI DE REPARTIȚIE* 315
Curbă de frecvență și curbă cumulativă 315
Cuantilă (α %-cuantilă) 316
Legi de repartiție discrete 318
Legi de repartiție continue 325
Legi de repartiție multidimensionale (n-dimensionale) 350
Legea Poisson multidimensională 361

XIV. *TEORIA SONDAJULUI STATISTIC (TEORIA SELECȚIEI)* 36.
Experiment 362
Eșantion (de volum n) 362
Spațiul eșantioanelor 362
Selecție 363
Sondaj 363
Statistică (m-dimensională) 364
Spațiul parametrilor 364
Funcție empirică de repartiție (funcție de repartiție de selecție) F_n 364
Moment empiric (moment de selecție) de ordin k 365
Moment empiric centrat (moment centrat de selecție) de ordin k . 366

XV *TEORIA ESTIMAȚIEI* 366
Estimator 367
Estimație 372
Funcție parametrică 372
Funcție estimabilă 372
Principiul verosimilității maxime 372
Estimator de verosimilitate maximă 373
Funcție de verosimilitate 375
Metoda celor mai mici pătrate 376
Estimator MMP 377
Metoda momentelor 378
Metoda minimului lui χ^2 379

XVI. *VERIFICAREA IPOTEZELOR STATISTICE* 379
Verificarea (testarea) ipotezelor statistice 380
Ipoteză statistică 380

Prag de semnificație 380
Eroare de genul întîi (I) 380
Eroare de genul al doilea (II) 380
Riscul erorilor asociate unui test statistic 381
Regiune (interval) de acceptare 381
Regiune (interval) de respingere (regiune critică) 381
Puterea unui test 381
Caracteristică operativă (funcția C.O.) 382
Forța unui test 382
Statistică 382
Test statistic 382
Test χ^2 387
Test z 387
Test t 393
Test statistic privind egalitatea a 2 medii 394
Test statistic privind egalitatea a k medii 409
Test de semnificație pentru dispersie 409
Test statistic privind egalitatea a două dispersii 412
Test statistic privind egalitatea a k dispersii ($k > 2$) 414
Teste de omogenitate generală 420
Teste de concordanță 426
Teste pentru verificarea normalității 441
Teste pentru verificarea caracterului aleator 459
Teste privind coeficientul de corelație 466
Teste privind repartiția Poisson 472
Teste privind repartiția binomială 477
Teste privind parametrii legii multinomiale 493
Control statistic al calității 506
Funcție de fiabilitate 510

XVII. *ANALIZA DE VARIANȚĂ (ANOVA). PLANIFICAREA EXPERIMENTELOR* 511
Analiză de varianță (ANOVA) 511
Tratament 511
Factor 511
Model liniar general 512
Formă canonică a ipotezelor fundamentale 515
Ipoteză nulă în modelul liniar 516
Sumă totală a pătratelor (Sumă totală de pătrate) 517
Sumă a pătratelor erorilor 517

Sumă a pătratelor datorată (pentru verificarea) ipotezei nule H_0 . 517
Testul ipotezei nule 518
Tabel ANOVA 519
Model I unifactorial ANOVA 519
Model I bifactorial ANOVA (clasificare încrucișată) 524
Model I bifactorial ANOVA (clasificare ierarhică (cuibărită)) . . 527
Model I ANOVA multifactorial formal (clasificare încrucișată) . . . 529
Model I ANOVA multifactorial formal (clasificare ierarhică) . . . 532
Model II unifactorial ANOVA 534
Model II bifactorial ANOVA (clasificare încrucișată) 535
Model II bifactorial ANOVA (clasificare ierarhică) 536
Model de analiză de covarianță 536
Metode de comparație multiplă (inferență statistică simultană) . . 539
Planificarea experimentelor 542
Plan experimental 542
Pătrat latin 544
Plan în blocuri 545
Randomizare 546
Confounding. 547

PARTEA A ȘASEA. **VARIA** 548

XVIII. *ISTORICUL STATISTICII DIN ROMÂNIA* 548
XIX. *PERSONALITĂȚI* 563
XX. *INSTITUȚII ȘI PUBLICAȚII DE STATISTICĂ* 590
INDICE . . . . . .

PARTEA ÎNTÎI

GENERALITĂȚI

Statutul actual al statisticii pune în evidență o știință complexă, cu un corp diversificat de metode și cu o vastă arie de aplicații, în cele mai diferite domenii ale vieții materiale și spirituale, ca rezultat al unei îndelungate evoluții. În această secțiune au fost reținute conceptele și noțiunile cele mai generale, valabile, în principiu, pentru orice domeniu de aplicație. Alegerea acestora are un caracter selectiv și, ca atare, în oarecare măsură, arbitrar, ceea ce era inevitabil. În felul acesta, se oferă o introducere generală în statistică, înlesnită de folosirea criteriului logic combinat cu cel genetic-istoric. Urmărirea istorică a statisticii redă în fapt constituirea acestei științe, în etape succesive, cu relicfarea aportului fiecărei etape.

I. CONCEPTUL DE STATISTICĂ

STATISTICĂ. Știința care se ocupă cu descrierea și analiza numerică a *fenomenelor de masă* *, dezvăluind particularitățile lor de volum, structură, dinamică, conexiune, precum și regularitățile sau legile ce le guvernează. Caracterul precumpănitor metodologic al **s.** o situează în metodologia și filozofia științei. Demersul complet gnoseologic al **s.** se desfășoară în următoarele faze [33]: 1. Observarea — culegerea informației despre procesul sau obiectul luat în studiu; 2. Formularea ipotezelor; 3. Predicția — deducțiile ce se fac pe baza ipotezelor formulate; 4. Validarea — culegerea de noi date pentru a confrunta predicțiile, efectuate pe baza teoriei formulate anterior. Există un mare număr de definiții ale **s.**, fapt care are mai multe explicații; *istorice* — **s.** s-a constituit la confluența mai multor orientări și discipline, dar avînd la origine un caracter social, proces care încă nu este încheiat; *metodologice* — **s.** incorporează o serie de metode în care predomină cele cantitative, cu fundamentarea epistemologică corespunzătoare; *aplicative* — prin generalitatea metodelor sale, **s.** se aplică în cele mai variate domenii, de la astronomie la societate, de la microcosm la macrocosm în natură și societate, de la fizica statistică pînă la **s.** socială. O definiție sau alta a **s.** stăruie fie asupra unei faze din demersul metodologic, fie asupra unui domeniu restrîns, fie asupra unui grup de metode, fie, în sfîrșit, asupra teoriei sau, dimpotrivă, a aplicațiilor. Ca oricare altă știință, **s.** s-a născut din necesități practice; procesul de teoretizare s-a desfășurat treptat putînd fi identificate, cu aproximație, următoarele etape: descriptivă, aritmetică, probabilistă, informațională, sistemică [28]. În prezent, **s.** dispune de un corp de metode a căror putere de cunoaștere este verificată în variate domenii și de o teorie generală în continuă dezvoltare. Cultura **s.** devine tot mai mult o componentă din cultura generală a omenirii contemporane, iar gîndirea **s.** — o modalitate științifică indispensabilă pentru analiza și interpretarea unor clase foarte largi de fenomene. **S.** a apărut ca o generalizare

sui-generis a unor date despre fenomene sociale. Istoricește ea s-a situat în mediul științelor sociale. Încă pe la mijlocul secolului trecut, sfera **s.** era considerată numai viața socială : treptat, a început să se „infiltreze" în domenii ale științelor naturii. Astfel, alături de **s.** socială s-au format și s-au dezvoltat, în decursul timpului, noi discipline științifice : **s.** matematică, fizica **s.**, mecanica **s.**, **s.** biologică etc. Astăzi, **s.** are numeroase ramuri care au ca obiect fenomene sociale și nesociale. Accepțiile curente ale noțiunii de **s.** sînt și ele multiple.

1. *Date* **s.** — date cantitative afectate în mare măsură de o mulțime de cauze.
2. *Metode* **s.** — metode special adaptate la studierea datelor statistice.
3. *Metodă* **s.** — metodă științifică de tip special care se aplică în procesul de cunoaștere a fenomenelor de masă.
4. *Teoria* **s.** sau **s.** — disciplină care se ocupă cu expunerea sistematică a conceptelor, principiilor și metodelor de cercetare **s.** [35].

Concepută ca știință sau ca metodă, **s.** actuală prezintă — după consensul general — două aspecte diferite și, dealtminteri, complementare : *descrierea* **s.** — fixarea informației, așa cum a rezultat din prelucrarea datelor empirice, și *inferența* **s.** — tratarea teoretică a datelor pentru a trage concluziile logice, asociate observațiilor efectuate. Primul aspect, mai puțin ambițios, evocă partea clasică a **s.** și este în general cunoscut sub numele de **s.** *descriptivă**, iar al doilea aspect reprezintă **s.** *inductivă* *. În demersul s., cele două aspecte trebuie să fie privite ca două etape ale unuia și aceluiași proces de cunoaștere.

Istoric. Istoria consemnează operații de numărare a populației, de măsurare a terenurilor sau a altor resurse la popoarele antice. Chinezii, bunăoară, dispuneau de date asupra populației, pămînturilor și recoltelor încă de pe timpul legendarului împărat Yu (pe la 2300 î.e.n.), inserate în „Shu-King" (Cartea istoriei) — o culegere de cele mai vechi documente scrise ale Chinei. Egiptenii cunoșteau cadastrări și numărători de populație, cu caracter regulat, din timpuri străvechi. Este atestată o numărătoare pe vremea faraonului Amenophis III (1408—1372 î.e.n.), care surprinde prin ingeniozitate. Evreii, care au stat mulți ani în Egipt, au preluat, se pare, de la egipteni ideea numărării populației. În textele biblice sînt menționate diferite numărători ; unele din ele au cuprins numai bărbații „buni de oaste", ca numărătoarea întreprinsă de Moise în deșertul peninsulei Sinai (1491 î.e.n.), și care a stabilit un număr de 600 000 de bărbați, altele — întreg poporul, cum a fost numărătoarea regelui David (1017 î.e.n.). Pentru statele antice (Lagash, Ur, Nippur, Akkad) din ținutul aflat între Tigru și Eufrat, tăblițe de lut („cărțile de lut") au conservat date cu privire la agricultură, creșterea animalelor, construcții de corăbii, navigație etc. (mileniul al 3-lea î.e.n.).

Numărătorile de populație și de bunuri, indispensabile pentru scopuri militare, așezarea impozitelor și evaluarea bogățiilor, cunoscute bine și în *polis*-urile Greciei antice, au culminat cu recensămintele (*census*) și anchetele romane. Tradiția antică atribuie instituirea census-ului regelui Servius Tullius (sec. 6 î.e.n.). Cu ocazia census-ului, fiecare cetățean trebuia să declare numele și prenumele, vîrsta, familia sa, numărul sclavilor și bunurile de orice natură. După cum ne informează Titus Livius, census-ul a permis stabilirea obligațiilor civile și militare ale cetățenilor nu numai pe cap de om, ca mai înainte, ci și după avere (*tributum ex censu*). O investigație de acest fel, dar mai evoluată și mai complexă, a avut loc pe vremea împăratului August (anii 27 î.e.n. și 14 e.n.). P. C. Tacitus pomenește de o vastă anchetă, care a inclus cetățenii și alții aflați sub arme, navele, resursele de toate felurile, veniturile publice. Bilanțul bogățiilor imperiului, scris cu mîna lui August, a fost prezentat senatului de

succesorul său. Înregistrări de tipul census-ului s-au practicat și în Dacia romană. În imperiul roman s-au organizat și înregistrări ale mișcării naturale a populației. Sub Antonius Pius (138—161), declararea nașterilor a devenit obligatorie și se făcea în Roma la prefectul tezaurului, iar în provincie, la *tabularium* (birouri speciale în care se păstrau registrele fiscale, vamale, cadastrale etc.). Investigații analoage au continuat în evul mediu. Caracteristice sînt inventarierile de bunuri ale feudelor, mănăstirilor, corporațiilor sau breslelor și însemnările curente ale nașterilor, căsătoriilor, deceselor la parohii, premergătoare actelor de stare civilă. N-au lipsit lucrări de evidență din inițiativa statului. Citām, de exemplu, lucrările : „Breviarium Fiscalium" din epoca lui Carol cel Mare (771—814), în care sînt minuțios descrise locuințele, uneltele de casă, proviziile, animalele de curte, speciile de culturi și „Doomesday Book" sau „Liber iudiciarius seu censusalis Willelmi I regis Angliae", care, în afară de enumerarea locuitorilor, include descrierea statistică a patrimoniilor, veniturilor, pămînturilor, servituțiilor prediale (funciare). Cartea lui William Cuceritorul, care datează din anii 1083—1086, este considerată prima însemnare censuară de mare proporție după census-ul roman.

Contabilizarea resurselor umane și materiale în lumea antică și medievală a izvorît dintr-o gîndire practică, utilitară și s-a executat, desigur, fără criterii riguroase, în mod empiric. Informațiile se culegeau mai cu seamă în scopuri fiscale, militare ori administrative. Tipurile de colectivități observate și modalitatea de descriere fac parte, incontestabil, din domeniul atribuit s. ; nu s-a conturat însă în acele vremuri vreo concepție despre s. Depășirea stadiului de s. practică sau materială în exclusivitate s-a produs în momentul în care, peste interesele practice imediate, s-au evidențiat interese de ordin științific, de cunoaștere critică a fenomenelor consemnate. Înregistrările empirice au folosit disciplinei s., la începutul ei, ca instrument de vehiculare a unui nou domeniu de cunoaștere și ca sursă documentară pozitivă, oferindu-i material de analiză și interpretare.

Faza descriptivă. Într-o primă accepție, s. echivala cu *descrierea statului* — expunerea situației geografice, economice, politice. Acest gen de s. a fost cultivat mai întîi de italieni. Încă din sec. al XIII-lea și al XIV-lea, celebrele *relazioni* (rapoarte) ale republicii venețiene concentrau, grație negustorilor și apoi agenților diplomatici, multiple informații despre condițiile partenerilor comerciali. Lucrările de descriere a statelor se înmulțesc în epoca Renașterii. Printre cele mai reprezentative amintim „Del governo e amministrazione di diversi regni e republiche" (F. Sansovino) și „Le relazioni universali" (G. Botero). Lucrarea lui *F. Sansovino* (1521—1583) cuprinde descrierile a 22 state, începînd cu Sparta, Atena, Roma etc. și terminînd cu statul imaginar Utopia ; lui i se atribuie meritul de a fi luat în considerare, în cartea sa, descrierea numerică a statelor : bogățiile, populația, comerțul, industria și finanțele lor. În chip analog, *G. Botero* (1533—1617) se referea la o serie de state printre care și la Moldova ; la el apar rudimentele unei s. comparate. Cei doi autori au avut numeroși continuatori în alte țări și îndeosebi în Germania. Dezvoltarea etatismului mercantilist european antrenează o intensificare a preocupărilor pentru cunoașterea situației cît mai multe țări. În întîmpinarea cererii sporite de informații a apărut faimoasa colecție „Respublicae Elzevirianae", un fel de enciclopedie a statelor. Colecția — intitulată astfel după numele ilustrei familii de tipografi Elzevir din Olanda — a însumat cîteva zeci de volume, cele mai multe publicate între anii 1625 și 1652 (și a avut ca principal editor pe Jan de Laet). Curentul „descrierea statului" a atins apogeul în sec. al XVII-lea și al XVIII-lea cînd, în Germania, s-a constituit o adevărată școală,

cunoscută sub denumirea de *școala descriptivă germană*. Fondatorul acestei școli a fost *Hermann Conring* * (1606—1681), profesor la Universitatea din Helmstedt. Prin originea sa olandez, Conring a cunoscut pe Jan de Laet și a contribuit la realizarea bibliotecii Elzevir, care a format dealtfel și baza cursului său universitar despre știința descriptivă a statului. Prelegerile lui Conring au fost, fără voia sa, publicate de un discipol, în anul 1671, sub titlul „Notitia rerum publicarum" sau „Descriptio singularis rerum publicarum aevi nostri". Ca cel dintîi manual de știință a statului se citează însă lucrarea „Teutscher Fürstenstaat" (1656), elaborată de *Veit Ludwig von Seckendorff* (1626—1692), profesor la Universitatea din Halle. *Martin Schmeitzel* (1679—1757) a introdus la Universitatea din Halle, cursuri despre *știința cameralâ* și „Notitia rerum publicarum", apoi a ținut un curs intitulat „Collegium politico-statisticum". *Gottfried Achenwall* * (1719—1772) continuă învățămîntul denumit *Staatskunde* la Universitatea din Göttingen. Prin învățămîntul pe care-l asigurau, profesorii universitari erau în mod automat consilieri în gestiunea afacerilor statului și responsabili de formarea funcționarilor necesari administrației.

Descrierea statului, pînă atunci vulgară, a devenit disciplină de predare academică, încadrată într-un sistem construit după norme teoretice și practice, care s-au elaborat și dezvoltat în universitățile germane. Noua disciplină (*Staatskunde*), derivată din literatura în materie existentă și nu din cercetări originale ale autorilor ei, a primit ulterior numele de **s.** (*Statistik*). De obicei, se atribuie lui G. Achenwall prima întrebuințare a termenului de **s.** în lucrarea „Abriss der Staatswissenschaft der europäischen Reiche" (1749). Adjectivul *statisticus* apăruse mai devreme în lucrări scrise în limba latină, ca de exemplu „Microscopium statisticum quo status imperii Romano-Germanici repraesentatur" (1672) a lui *Helenius Politanus* ; „Itinerarium Germanice politicum" (1675) de *Phil. Andr. Oldenburg* (în care se vorbește de *rationes statisticae*). Adesea se acordă lui Achenwall nu numai paternitatea denumirii, ci și a disciplinei, onoare pe care el n-a revendicat-o niciodată și care pare exagerată. „A declara pe Achenwall părintele **s.** — scrie M. Block — înseamnă a face un al doilea Amerigo Vespucci, care dă numele său unui continent pe care nu l-a descoperit" [2]. Însăși denumirea de **s.** se crede că a fost popularizată de către Achenwall, însă inventată de profesorul său Schmeitzel. Etimologic, cuvîntul **s.** provine de la latinescul medieval *status*, „stat" sau „stare politică", prin intermediul substantivului italian *statista*, cu înțelesul „om politic, versat în treburile statului". **S.** avea să înfățișeze, după Achenwall, situația (status) a tot ceea ce era remarcabil, esențial, într-un stat. În „Cuvînt înainte" la ediția a șasea a lucrării sale citim : „Dacă privesc un stat oarecare văd în el o mulțime mare de lucruri, care se prezintă așa cum sînt ele în realitate. Printre acestea citeva se referă în măsură considerabilă la prosperitatea sa, fie că ele o împiedică, fie că o stimulează : se pot numi acestea lucruri remarcabile într-un stat (*Staatsmerkwürdigkeiten*)" [1] și mai departe : „Ansamblul lucrurilor remarcabile reale ale unui imperiu sau ale unei republici constituie organizarea de stat (*Staatsverfassung*) în sens mai larg și învățătura despre organizarea de stat a unuia sau mai multor state este **s.** — *Staatskunde* — sau descrierea statului" [1]. În aceeași concepție, folosul principal al **s.** consta în faptul că „prin ea sîntem în stare nu numai să judecăm corect și temeinic tot felul de probleme de stat, ci și să dobîndim iscusințe ca, la caz de nevoie, să putem fi folosiți, cu sfaturi sau cu fapta, la tratarea lor" [1]. *A. L. Schlözer* (1735—1809), unul dintre cei mai fideli discipoli ai lui Achenwall și succesorul

său la catedra universitară, a exprimat „esența statului" și, ca atare, obiectul s. în trei cuvinte : *Vires unitae agunt*. Prin *vires* a înțeles forțele naturale și dobîndite (oamenii, teritoriul și avantajele naturale), *unitae* — forma și organele de guvernămînt, iar *agunt* — guvernarea și administrația. Lui îi aparține aforismul : „S. este istorie în repaus, istoria — s. în mișcare" [34]. S. descriptivă s-a numit și s. *universitară*, deoarece reprezentanții săi erau învățați de la universități (mai ales germane). Termenul s., definit în spiritul școlii din Göttingen, se întîlnește la mulți autori de renume ai epocii. Unul din capitolele lucrării „The Elements of Universal Erudition", scrisă de baronul *J. F. Bielfed* (trad. de W. Hooper, Londra, 1770), se intitulează „Statistics" și cuprinde și o definiție a acesteia : „Știința despre organizarea politică a statelor moderne ale lumii" [35]. Cu o definiție ceva mai largă, termenul s. apare în prefața la un studiu politic al situației Europei : „A Political Survey of the Present State of Europe", publicat în 1787 de *E. A. Zimmermann*, profesor de filozofie a naturii la Braunschweig. „Sînt aproape patruzeci de ani — scrie el — de cînd o ramură a cunoașterii politice, care are ca obiect puterea actuală și relativă a statelor moderne... a devenit o disciplină distinctă... În forma convenabilă pe care a căpătat-o acum... această disciplină, desemnată prin noul nume de s., a devenit un obiect de studiu favorit în Germania" [35]. Ulterior, *John Sinclair*, editorul și organizatorul primului volum al lucrării, „Statistical Account of Scotland" arăta : „ ...mulți oameni au fost intii surprinși că utilizez cuvinte noi ca s. și *statistic*, presupunînd că un alt termen din limba noastră ar avea același sens... și deoarece am crezut că un cuvînt nou ar atrage mai mult atenția publicului, am hotărît să-l adopt și cred că, acum, acest termen este complet naturalizat și încorporat în limba noastră" [35].

În decursul jumătății de secol care a urmat nașterii noii discipline, structura studiilor de descriere a statului a început să se modifice în sensul consolidării *determinărilor numerice*. Un reputat geograf și statistician, *A. F. Büsching* (1724—1793), profesor de filozofie la universitatea din Göttingen, a creat s. comparativă (*vergleichende Staatenstatistik*), în cartea sa : Vorbereitung zur gründlichen und nützlichen Kenntnis der geographischen Beschaffenheit und Staatsverfassung der europäischen Reiche" (Hamburg, 1758). Este considerat drept primul, după Achenwall, care a preluat critic date s. în lucrările sale geografice. Se pare că A. F. Büsching a fost influențat de lucrarea istoricului și filozofului *J. P. Anchersen*, intitulată „Descriptio statuum cultiorum in tabulis" (1741). Inovația lui Büsching a fost proclamată drept excelentă de însuși Schlözer, „păstrător intratabil al tradițiilor maestrului său Achenwall". El a recunoscut, în cele din urmă, că o s. fără cifre nu înseamnă nimic : „Indicații generale ca acestea : țara produce un vin *delicios*, ea are *bogate* manufacturi, un comerț *înfloritor*, *puțin grîu*, ne sînt furnizate de toate geografiile ; dar aceste indicații, atît timp cît nu sînt precizate prin cifre, nu servesc la nimic" [2]. Să mai amintim că A. F. Büsching a tradus din latină în germană „Descriptio Moldaviae" a lui Dimitrie Cantemir. Pătrunderea mai puternică a elementului numeric, favorizată de evidențe oficiale mai abundente și mai ușor accesibile, nu a schimbat conținutul s., așa cum a fost concepută de autorii germani din sec. al XVIII-lea. S. a intrat în istoria disciplinelor științifice ca o disciplină descriptivă, orientată spre *informație* și nu spre căutarea de legi, cu un mod de expunere preponderent verbal.

Faza aritmeticii politice. Pe vremea constituirii s. ca disciplină descriptivă a statului, se năștea în Anglia, în afara universităților, o s. cu totul deosebită, sub numele de *aritmetică politică*. Denumirea însăși este revelatoare pentru

semnificația atașată conceptului de s.: analiza datelor de observație prin procedee matematice, desprinderea regularităților în fenomenele social-economice și chiar formularea de previziuni. Folosirea metodei analitice, preconizată de filozofii materialiști metafizicieni ai timpului, recurgerea la instrumentul matematic și căutarea legităților marcau un substanțial progres, prefigurînd s. modernă.

Reprezentanții școlii aritmeticii politice din Anglia au pus accentul mai ales pe cercetarea fenomenelor demografice, deoarece vedeau în populație izvorul bogăției și puterii acestei țări. Ca material de explorare au servit rapoartele săptămînale asupra deceselor, introduse de administrația engleză încă din 1592, sub denumirea de „Bills of mortality". *John Graunt* * (1620—1674), negustor de stofe și comandant al miliției londoneze, cu modestă cultură, dar cu o intuiție genială, a prelucrat datele din rapoartele (listele) pe o perioadă de treizeci de ani ale Londrei și a pus în evidență legitățile creșterii populației, ale echilibrului numeric dintre sexe, ale fertilității, mortalității etc. El a atras atenția, pentru întîia oară, asupra predominanței nașterilor masculine în raport cu nașterile feminine (107 băieți la 100 fete), a calculat primele rate de mortalitate pe grupe de vîrstă și a alcătuit — în ipoteza unei populații staționare — o rudimentară tabelă de mortalitate. Rezumat în lucrarea „Natural and political observations upon the bills of mortality chiefly with reference to the government, religion, trade, growth, air, diseases etc. of the City of London" (Londra, 1662), studiul era un efort de pionier în analiza statistică. După aprecierea demografului și statisticianului american Walter Willcox, în prefața ediției americane a lucrării lui Graunt (1939), acesta : „ ...a descoperit uniformitatea și posibilitatea de predicție a unor fenomene demografice importante, luate în masă. Procedînd în acest fel, el a deschis calea pentru noi descoperiri de uniformități în numeroase fenomene sociale și comportamentale, precum și pentru studiul acestor uniformități, al naturii și al limitelor acestora..." [31]. În explorările și în popularizarea descoperirilor sale, Graunt a fost sprijinit de amicul său, *William Petty* * (1623—1687), un „cugetător plin de cutezanță" „întemeietorul economiei politice moderne, unul din cei mai geniali și mai originali cercetători ai economiei" și „într-o oarecare măsură inventatorul s." cum l-a caracterizat K. Marx. Petty, creator al denumirii *Political Arithmetick*, a invitat, într-o operă cu același titlu, la studiul fenomenelor social-economice în expresie cantitativă : „Metoda pe care o întrebuințez... nu este pînă acum prea comună, căci în loc de a mă servi de termeni la comparativ și la superlativ și de argumente pur raționale, am adoptat metoda (ca specimen al aritmeticii politice ce am avut-o demult în vedere) care constă în a mă exprima în termeni de *numere*, greutăți și *măsuri*..." [25]. În lucrarea „Political Arithmetick" (Londra, 1690) sînt folosite date și calcule numerice în compararea potențialului uman și economic al diferitelor state și, în particular, în evaluarea posibilităților comerciale și maritime ale Angliei în rivalitatea sa cu Franța.

Cercetări interesante în domeniul aritmeticii politice a întreprins astronomul englez Edmund Halley (1656—1742). De numele lui sînt legate estimări ale populației cîtorva state, întemeiate pe criterii mai raționale decît cele adoptate anterior, alcătuirea primei tabele complete de mortalitate, introducerea ideii de durată probabilă a vieții, cu aplicație la calcularea rentei viagere anuale în funcție de vîrstă. Alți doi reprezentanți de seamă ai școlii engleze, *Charles Davenant* (1656—1714) și *Gregory King** (1648—1712), au contribuit la promovarea teoriei și practicii aritmeticii politice în investigarea fenomenelor economice. Davenant—pentru care aritmetica politică era „arta de a judeca cu

ajutorul cifrelor, lucrurile care privesc guvernul" — s-a ocupat de aplicații la veniturile publice, elaborînd lucrarea „Discours on the public revenues" (1698). King s-a dedicat calculelor privind veniturile statului, pe calea impozitelor din agricultură, clădiri, bunuri mobile; de asemenea a încercat să stabilească mărimea capitalurilor de care dispunea Anglia în sec. al XVII-lea, pierderile suferite în urma războaielor, incendiilor, epidemiilor. În preocupările lui King a intrat și populația Angliei, acordînd atenție studiului veniturilor, cheltuielilor și acumulărilor realizate de diferite categorii de familii. Este considerat de asemenea precursor al econometriei.

Școala aritmeticii politice engleze și-a extins influența și în alte țări. Curînd a apărut o literatură aritmetică politică europeană, avînd repercusiuni profunde asupra curentului statistic universitar. În Germania, aritmetica politică a repurtat un succes decisiv în lupta aprigă cu școala descriptivă din Göttingen, prin aportul lui *Johann Peter Süssmilch* * (1707—1767). Beneficiar al literaturii adunate de Academia de Științe din Berlin, Süssmilch a reușit să realizeze o sinteză magistrală a aritmeticii politice. În lucrările sale se găsesc elemente ale unei concepții fecunde în rezultate științifice: căutarea explicației fenomenelor vieții umane și indicarea legilor ce le guvernează, corelarea fenomenelor demografice și economice, examinarea fenomenelor în dinamică, identificarea regularităților în masa de observații, scrupuloasa alegere și evaluarea critică a datelor. Opera principală a lui Süssmilch, „Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechts aus der Geburt, dem Tode un der Fortpflantzung erwiesen" (1741), este o carte de s. a populației și de demografie [29].

Aritmetica politică s-a dezvoltat tot mai mult în demografie, care se mulțumea în acea epocă să fie statistică. *Antoine Deparcieux* (1703—1768) a construit primele tabele franceze de mortalitate, iar *Pehr Wilhelm Wargentin* (1717—1783) a elaborat tabela de mortalitate pentru Suedia (cele dintîi tabele care se referă la populația unui întreg stat). Studiile demografice lăsau să se întrevadă că previziunea fenomenelor colective ar putea avea o însemnătate practică; dealtfel, tabelele de mortalitate, întocmite în sec. al XVIII-lea, au constituit punctul de plecare al înfloritoarei industrii de asigurări.

Faza probabilistă. În disputa care opunea cele două curente, a triumfat curentul de tendință modernă reprezentat de artimetica politică. S. descriptivă, agonizantă în pragul sec. al XIX-lea, a sfîrșit prin a se combina cu s. practică, deznodămînt pricinuit și de asocierea, pe măsura constituirii în discipline autonome, a celei mai mari părți din domeniile asupra cărora se întindea odinioară. Deși optica numerică prevalase față de optica verbală, denumirea de aritmetică politică a dispărut, substituindu-i-se denumirea de s., rezervată de școala germană „descrierii statului" (*Staatskunde*). În timp ce descrierile statistice devin obiect al instituțiilor publice, domeniul conceptual al aritmeticii politice capătă o nouă dimensiune o dată cu introducerea *calculului probabilităților* *. Primele încercări teoretice au fost făcute de *Jacob Bernoulli* * (1654—1705). În lucrarea postumă „Ars conjectandi" (1713), Bernoulli a formulat faimoasa „lege a numerelor mari", a deosebit probabilitatea apriori (matematică, logică sau deductivă) de probabilitatea aposteriori (statistică, empirică sau inductivă) și a schițat posibilitatea aplicării calculului probabilităților în economie (parte neterminată). *Pierre-Simon Laplace* * (1749—1827) a adunat în „Théorie analytique des probabilités" (1812), toate aplicațiile calculului probabilităților, a evidențiat avantajele în studiul fenomenelor dependente de cauze complexe și a demonstrat utilizarea posibilă a acestui calcul în exploatarea rezultatelor statistice privind fenomene sociale. După o apreciere necontes-

tată. Laplace „trebuie situat pe primul loc între acei savanți care, într-un fel, au aruncat o punte între aritmetica politică și calculul probabilităților" [23]. La aceeași epocă, *Karl Friedrich Gauss* * (1775—1855) definește legea normală de repartiție, pusă în lumină încă de *A. de Moivre* * (1667—1754), Jean Baptiste Joseph Fourier (1768—1830) promovează aplicațiile matematice în s. prin studii publicate în „Recherches statistiques sur la ville de Paris et le département Seine". iar *Simon Denis Poisson* * (1781—1840) descoperă legea de repartiție a evenimentelor rare și se ocupă de probleme de demografie, de statistică morală și judiciară, cu aplicații ale calculului probabilităților.

În sec. al XIX-lea, numeroși alți matematicieni și învățați au realizat lucrări valoroase, care au consolidat gîndirea probabilistă în s. și i-au asigurat progresul. Cel care a dominat însă ca statistician epoca sa a fost *Adolphe Quételet** (1796—1874). Cu el s-a produs o nouă cotitură în evoluția conceptului de s. De obicei, numele lui Quételet este asociat cu teoria „omului mediu", un *homo statisticus*, care venea să înlocuiască pe *homo oeconomicus* al clasicilor. Conform acestei teorii statistice, de la care se pleca în constituirea unui fel de fizică socială, diversele caractere fizice, intelectuale și morale ale oamenilor, observați în masă, s-ar echilibra într-o ființă fictivă, ideală, dorită de natură, dar neizbutită; oamenii „reali" s-ar distribui în jurul ei „la întîmplare", analog erorilor în măsurarea unei mărimi. În lucrarea „Sur l'homme et le développement de ses facultés ou essai de physique sociale" (1835), citim: „Noi trebuie, înainte de toate, să pierdem din vedere omul luat izolat și să nu-l considerăm decît o fracțiune a speciei. Despuindu-l de individualitatea sa, vom elimina tot ceea ce nu este decît accidental, iar particularitățile individuale, care nu exercită decît puțină sau nici o acțiune asupra masei, se vor șterge și vor permite sesizarea rezultatelor generale" [27]. Lăsînd la o parte latura sa fabuloasă, conceptul de „om mediu" se cuvine evocat pentru ideile implicate: repartiție, medie, dispersie, observare de masă, regularitate, esențiale în cercetările statistice. „Cînd Quételet — scrie Durkheim — semnala atenției filozofilor surprinzătoarea regularitate cu care anumite fenomene sociale se repetă, el ținea seama de teoria sa despre omul mediu, care a rămas, dealtfel, singura explicație sistematică a acestei remarcabile proprietăți" (în lucrările sale — *n.n.*) [9]. Cu prilejul centenarului morții lui Quételet (1974), au fost organizate sesiuni științifice și reeditări ale lucrărilor sale, vădindu-se un interes mare pentru ideile sale și reactualizîndu-se și concepția sa despre „omul mediu", într-o viziune modernă. Concepută ca o „fizică socială", s. quételetiană încorporează ideea că regularitățile descoperite prin observarea de masă rezultă din legi care se impun omului în ordinea fizică și de asemenea în ordinea morală. „Eu nu am încetat o clipă să repet — declara Quételet cu referire la criminalitate — că există un buget ce se plătește cu o regularitate nemaipomenită, acesta este bugetul închisorilor, ocnelor și eșafodurilor . . . ; anual numerele confirmă previziunile mele pînă într-un asemenea grad încît aș putea să spun cumva cu mai multă exactitate: există un tribut pe care omul îl plătește cu mai mare regularitate decît tributul datorat naturii sau vistieriei statului — este tributul pe care-l plătește crimei" [27]. Acest punct de vedere determinist-mecanicist era de nuanță moderată, deoarece admitea posibilitatea transformării ambianței sociale. Astfel, după ce afirmă că „societatea conține în sine germenii tuturor crimelor ce se vor comite" și că „orice stare socială presupune deci un anumit număr de crime și o anumită categorie de delicte, care apar drept consecință a organizării sale", Quételet continuă: „Această observație, care poate părea descurajatoare din primul moment, devine consolantă, dimpotrivă, cînd se examinează cu atenție, fiindcă ea arată posibilitatea de a îndrepta oamenii, modificînd insti-

tuţiile lor, obiceiurile lor, nivelul cunoştinţelor lor şi, în general, tot ceea ce influenţează felul lor de a fi" [27]. O altă lucrare a lui Quételet, cu semnificaţie aparte în dezvoltarea istorică a s., este manualul „La méthode statistique" (1848), scris în calitate de tutore, pentru prinţul Albert al Belgiei [10]. Elementele de s. formulate, fără preocupare de sistematizare, în „Essai de physique sociale" şi alte lucrări, au fost dezvoltate şi întrunite, cu prilejul manualului, într-o expunere coerentă. Avem de-a face, pentru prima oară, cu o teorie sintetică a s., cu un corp distinct de principii, tehnici şi procedee statistice. Elaborat după o perioadă de ample cercetări întemeiate pe practica s., manualul „La méthode statistique" corespunde unui stadiu mult mai evoluat al gîndirii autorului său. E adevărat, în primele sale lucrări, Quételet nu are despre s. o opinie prea diferită de cea curentă. S. este numită „anatomia societăţii" şi i se atribuie scopul „de a observa modificările pe care le suferă diferitele popoare în starea lor fizică şi morală şi de a căuta să pătrundă motivele". Treptat însă, conceptul de s. include elemente noi, îndreptînd s. pe căi noi : cercetarea variaţiei şi a relaţiilor cauzale în masa fenomenelor, dezvăluirea regularităţilor, investigarea cu predilecţie a fenomenelor vieţii umane etc. În final, Quételet ajunge să privească s. drept singura metodă utilizabilă pentru studiul fenomenelor de masă „în care nu sîntem în stare să facem să varieze circumstanţele" şi care sînt susceptibile „de a fi tratate prin calculul probabilităţilor". Interpretarea probabilistă a s., caracteristică lui Quételet, este comună şi altor oameni de ştiinţă. După *A. Cournot* (1801—1877), unul din fondatorii economici matematice, s. ar fi „ştiinţa care are ca obiect culegerea şi coordonarea faptelor numeroase de orice fel încît să se obţină raporturi numerice sensibil independente de anomaliile hazardului şi care denotă existenţa cauzelor regulate a căror acţiune s-a combinat cu cea a cauzelor întîmplătoare". Concepţia sa despre s., sintetizată în această definiţie, o găsim mai ales în lucrarea „Exposition de la théorie des chances et des probabilités" (1843).

Noua orientare n-a întîrziat să pătrundă în gîndirea s. a epocii, dar controversele în jurul conceptului de s. au continuat ; definiţiile în circulaţie s-au înmulţit. Gustav von Rümelin (1815—1889), statistician german, a numărat 63 de definiţii (inclusiv pe a sa) în 1863. Cel mai controversat era domeniul atribuit s. : unii admiteau, în sfera de cuprindere a s., toate fenomenele exprimabile numeric, indiferent de natura lor, alţii — numai fenomenele sociale, iar un alt grup restringe cîmpul de aplicare al s. la fenomenele vieţii umane („societatea civilă"). Multă vreme încă, autori de talent au păstrat, cu unele nuanţe, conceptul de statistică descriptivă al şcolii din Göttingen sau conceptul de aritmetică politică la faza iniţială şi au scris lucrări în concordanţă cu conceptul însuşit. O remarcabilă contribuţie la progresul s., al afirmării ei ca ştiinţă au adus şi alte şcoli naţionale, în decursul sec. al XIX-lea, denumit sugestiv „era entuziasmului statistic" (H. Westergaard). Înainte de constituirea şi afirmarea deplină a s. matematice, la sfîrşitul sec. al XIX-lea şi primele decenii ale sec. al XX-lea, sînt de consemnat cîteva contribuţii importante. În epoca post-quételetiană orientarea principală a fost aceea a unei s. sociale, cu o pondere crescîndă a metodelor numerice. De asemenea, în aceeaşi perioadă se remarcă o intensificare a s. populaţiei şi demografiei.

Şcoala germană a continuat, în linii generale, tradiţia unei s. concepute ca ştiinţă socială, cu numeroase elemente de s. descriptivă. Unul din reprezentanţii de seamă este *Wilhelm Lexis* * (1837—1914), după care domeniul s. îl constituie societatea, procesele desfăşurate în cadrul ei, ca procese de masă, supuse legilor statistice. Aproximativ pe aceeaşi linie se situează *Georg von Mayr* (1841—1925). Revendicînd energic pentru s. dreptul la demnitatea de ştiinţă, Mayr

îi atribuie ca obiect special „studiul maselor sociale de orice fel, ... prin metoda de analiză numerică" [13]. Originalitatea sa constă în a face din s. o știință socială generală, un fel de sociologie (manualul său este clasic și poartă titlul „Statistik und Gesellschaftslehre". I—1895 și II—1897). Remarcabile sînt contribuțiile unor statisticieni și demografi ca *G. P. Knapp* (1842—1926), *K. Becker* (1823—1896) și *R. Böckh* (1824—1907).

S. rusă a fost reprezentată în s. socială mai ales de *D. P. Juravski* (1810—1856), care consideră s. drept o știință a „categoriilor", a „comparației", punînd un mare accent pe metoda grupărilor, o știință materială avînd ca obiect studiul vieții sociale și, în același timp, știință metodologică. Lucrarea sa principală „Despre sursele de date statistice și folosirea lor" (1846) marchează un pas important spre s. cantitativă. De asemenea, contribuții importante se datorează lui *A. A. Ciuprov* * (1874—1926).

S. italiană, cu o veche tradiție, a primit un impuls viguros din partea lui *Luigi Bodio* (1840—1920), care a condus s. oficială a Italiei între 1872 și 1900. „Annali di statistica", inițiate în 1871, au colectat importante studii originale autohtone de teorie și metodologie s. (medii, reprezentări grafice, comparații s.) și relatări critice despre lucrări străine. Mențione merită și lucrarea „Teoria generale della statistica" (Milano, 1888), scrisă de *Antonio Gabaglio*. *R. Benini* (1862—1956), profesor de s. la diferite universități din Italia, a contribuit în măsură însemnată la progresul s. sociale și, mai ales, demografice. O figură aparte este *L. Amoroso* ale cărui contribuții de matematică a populației au marcat un progres în s. și demografie.

Școala franceză, în afară de contribuțiile sale remarcabile la teoria probabilităților, a dat, prin cîteva personalități, un impuls notoriu s. sociale. *A. Guillard* (1790—1876), *L. A. Bertillon* (1821—1883), *Jacques Bertillon* (1851—1922), *P. E. Levasseur* (1829—1911) a adus o contribuție importantă la progresul s. sociale și al s. demografice.

În s. engleză, o contribuție importantă a adus *W. Farr* * (1807—1887), organizator, împreună cu A. Quételet, al primului Congres internațional de s. (1853). Contribuții importante se datorează unor statisticieni ca norvegianul *A. N. Kiaer* * (1838—1919), ale cărui eforturi pentru afirmarea sondajului s. sînt foarte importante; belgianul *P. F. Verhulst* (1804—1849), descoperitorul curbei logistice; suedezul *G. Sundbărg*, cu importante studii de s. demografică. S. americană, prin *C. D. Wright* (1840—1909) și *W. F. Willcox** (1861—1964) se alătură progresului general al s. sociale.

În ansamblul curentelor s. ale timpului s-a încadrat organic s. românească ale cărei începuturi sînt marcate de două personalități proeminente: *Ion Ionescu de la Brad* * (1818—1891) și *Dionisie Pop Marțian* * (1829—1865). Primul și-a expus concepția despre s. în lucrarea „Povățuri pentru catagrafia Moldovei" (1859), un sintetic manual de teorie a s.

După Dionisie Pop Marțian, s. studiază, în expresie numerică, „fenomenele naturale și sociale" și „are de scop a conduce la descoperirea cauzelor și efectelor și de aici la legile care domnesc viața universală a naturii și cea particulară a popoarelor și a arăta prin coordonarea acestor legi unitatea și armonia lumii fizice și sociale în deosebitele lor fenomene" [18]. Rezultă în mod evident adeziunea sa la teza de largă circulație, potrivit căreia obiectul de cercetare al s. excede sfera fenomenelor sociale. Organizarea s. de stat și introducerea s. ca disciplină în învățămînt au contribuit sensibil la progresul s. din România. Crearea școlii de teoria probabilităților, legată de numele acad. *O. Onicescu* * (1892—1983) și de acad. *Gh. Mihoc* * (1906—1981), marchează o nouă etapă în istoria s. românești (→ *XVIII, Istoricul statisticii din România.*)

V.TREBICI

La progresul s. cantitative și pregătirea premiselor care au dus la constituirea s. moderne, au contribuit înființarea de oficii naționale de s., organizarea congreselor internaționale de s., apariția primelor reviste de s., introducerea s. în învățămîntul secundar și universitar. Congresele internaționale de s., inițiate de A. Quételet și W. Farr, s-au desfășurat între anii 1853 și 1872, favorizînd mult comparabilitatea internațională. O etapă nouă este marcată de înființarea *Institutului Internațional de Statistică* * (1885). Se înființează, de asemenea, societăți naționale de s. Primele reviste de s. sînt : „Journal of the Royal Statistical Society" din Londra * (1838), „Journal de la Société de Statistique de Paris" * (1860). În România, prima revistă de s. „Analele statistice" (1860) se datorează lui D. P. Marțian.

Pe baza tuturor acumulărilor înregistrate în sec. al XIX-lea, grație contribuției școlilor naționale de s. a fost cu putință, spre sfîrșitul secolului, trecerea la o nouă etapă.

Statistica modernă. Către sfîrșitul secolului trecut se conturează o nouă etapă în evoluția procesului de conceptualizare a s. — etapa constituirii s. *moderne*. Depășirea stadiului descriptiv, al simplei reprezentări cantitative, a însemnat, istoric, deplasarea gîndirii s. spre *interpretarea analitică* a fenomenelor de masă și obținerea de *concluzii inductive*, pe baza observațiilor empirice. Această schimbare de conținut a făcut ca s. să se întrepătrundă în mod constant cu matematica, în căutarea de metode corespunzătoare obiectivelor sale. După *M. G. Kendall** [15] (1907—1983) acest proces are loc în jurul anului 1890. O mare parte din metodele care reprezintă astăzi instrumentul obișnuit al analizei s. a fost elaborată de școala anglo-saxonă de s. matematică, fondată de *Francis Galton* * (1822—1911) și *Karl Pearson* * (1857—1936) și continuată de *Ronald Aylmer Fisher* * (1890—1962) și alții. Semnificativ este faptul că noua etapă este legată de lărgirea aplicațiilor și mai ales de experimente agricole și biologice.

Dintru început, s. se îmbogățește cu ideea de „variație legată" a variabilelor observate în colectivități s., dezvăluită de Galton în cartea sa „Natural inheritance" (1889). Transpusă în limbaj matematic de către Pearson, ea s-a constituit într-o teorie coerentă — teoria corelației s. — ale cărei aplicații au permis identificarea unor „noi forme de legi de dependență, specific s. și adaptate la formele complexe și variate pe care le oferă viața în diferitele sale manifestări" [34]. Alături de Pearson, la primele dezvoltări ale calculelor de corelație au contribuit *F. Y. Edgeworth* *, *A. L. Bowley* *, *G. U. Yule* *, *C. E. Spearman* *. Lucrările lui Spearman despre comportamentul indivizilor, dezvoltate mai tîrziu în psihologia aplicată, au condus la metodele de analiză factorială, continuare logică a studiului corelațiilor.

R. A. Fisher, calificat „cel mai mare dintre toți" (Kendall) [15], preconizează extinderea preocupărilor statistice la trei categorii de probleme [26] : a) probleme de specificație — expresia matematică a legii de repartiție a populației ; b) probleme de estimație — depistarea metodelor convenabile pentru calculul coeficienților proprii estimației parametrilor necunoscuți ai populației ; c) probleme de repartiție — repartiția estimațiilor parametrice și a unor coeficienți aleși pentru a testa validitatea specificației. Probabiliștii și statisticienii sec. al XIX-lea au avut în vedere numai prima categorie de probleme. În demersul modern, conceput de Fisher, se pornește de la *ipoteze statistice* * asupra fenomenului sau procesului observat, consecințele lor logice deduse se compară apoi cu datele disponibile și, dacă sînt în concordanță, ipotezele sînt justificate, cel puțin pînă la noi observații mai riguroase.

În accepția modernă, caracteristică pentru s. este preocuparea de a trage, printr-o analiză complexă, concluzii de o valoare generală din informații culese

prin cercetări pe eşantion. Dacă mai înainte aproape întreaga activitate s. se sprijinea pe *observarea totală* *, acum *sondajul* * tinde să devină de aplicaţie curentă. Dar a conchide de la eşantion la populaţie presupune o anumită incertitudine, pe care cercetătorul caută să o reducă. Unul dintre obiectivele esenţiale ale s., care-i dă nota sa originală, se consideră tocmai măsurarea incertitudinii concluziilor inductive. Cîteva idei intuitive pe această temă au apărut încă din sec. al XVIII-lea, însă posibilitatea tratării ei cu precizie matematică nu a fost pe deplin apreciată decît în secolul nostru. În linii generale, s. modernă, în care partea centrală o constituie s. matematică, cu conceptele sale probabiliste, s-a desăvîrşit în anii 1890—1940. Contribuţiile din perioada 1940 nu au schimbat — după părerea lui M. G. Kendall — profilul şi tematica fundamentală a s.

Statutul actual. În cursul istoriei sale tricentenare ca disciplină ştiinţifică, s. a fost definită — cum s-a mai spus — foarte diferit. După Walter Willcox, numărul definiţiilor date s. de către diverşi autori, între 1749 şi 1934, a depăşit o sută. Şi lista definiţiilor nu s-a închis, deoarece „printre temele în legătură cu care statisticienii nu sînt de acord, se găseşte definiţia ştiinţei lor" (Kendall) [26]. În diversitatea definiţiilor, se disting două opinii : unii califică s. drept *ştiinţă*, alţii o admit numai ca *metodă*. Sub impulsul şcolii Pearson-Fisher, s. şi-a găsit fundamentări foarte precise ca disciplină ştiinţifică ; ea se dezvoltă azi ca o ştiinţă evasiindependentă, avînd însuşirea — comună ştiinţelor — de a fi, în acelaşi timp, ştiinţă şi metodă utilizabilă şi în alte ştiinţe.

Opticile noi în ştiinţă — informaţională şi sistemică — care au pătruns în toate ştiinţele sînt proprii şi s. Mai mult, formularea şi afirmarea lor datorează mult s. *Etapa informaţională* a gîndirii s. începe cu teoria matematică a comunicaţiei, a lui Cl. Shannon (1948). Teoria informaţiei este considerată drept ramură a matematicii, strîns înrudită cu calculul probabilităţilor şi cibernetica şi care se ocupă cu problemele matematice ale obţinerii, transmiterii, stocării şi transformării informaţiei. Modelul shannonian este statistico-probabilist. Conceptul central este cel al entropiei informaţionale *(H)* care măsoară informaţia medie (a unei surse), dar care poate fi înţeles ca un simplu indice de diversitate :

$$H = -\Sigma\, p_i \log p_i,$$

p_i fiind probabilităţile sau frecvenţele semnalelor sursei. Se cuvine menţionată s. *informaţională Onicescu* * cu ale sale : *energia informaţională* * şi *corelaţia informaţională* *.

Etapa sistemică a gîndirii s. reprezintă — după V. Săhleanu [28] — mai mult decît o explicitare a ceea ce s-a realizat de-a lungul timpului sau decît o transcriere într-o nouă terminologie. Teoria generală a sistemelor (L. von Bertalanffy) şi cibernetica (N. Wiener) impun o reconsiderare a s.. Pe de altă parte, o serie de discipline noi — teoria jocurilor, teoria învăţării — fac apel la conceptele şi metodele s..

Rezultat al unui îndelungat proces, s. se prezintă cu un corp comun de noţiuni, organizat în jurul conceptului de probabilitate, de metode ştiinţifice, elaborate de s. matematică, cu fundamentare epistemologică explicită şi cu o varietate de domenii de aplicare, acoperind practic natura şi societatea.

S. constituie în prezent un sistem cu o deplină coerenţă logică, apţă pentru explorare în activitatea ştiinţifică ca şi activitatea de conducere a societăţii contemporane. În *domeniul ştiinţific*, s. a dobîndit o aplicabilitate generală. În prezent, se cuvine recunoscut că disciplinele, care dezvoltă direcţii cantitative

de cunoaștere, sînt toate tributare s. Așa cum Ronald Fisher a remarcat : „Către statisticieni ne îndreptăm pentru a răspunde la problemele cele mai serioase puse de marile aventuri ale epocii noastre. De fiecare dată cînd se întreprinde ceva important, statisticienii sînt în culise. În toate cercetările științifice de bază, ei sînt cei care elaborează planurile de experiență sau observații și tot la ei se recurge pentru a analiza rezultatele, pentru a evalua constatările și pentru a separa faptele clar demonstrate de cele ce mai cer confirmarea" [36]. Cercetările științifice simt necesitatea unei „escorte statistice" în două momente fundamentale ale contactului cu obiectul lor : momentul empiric — culegerea și prelucrarea faptelor brute, și momentul aplicativ, verificarea teoriilor pe calea confuntării cu realitatea. S. pune la dispoziție tehnici și procedee adecvate operațiilor întreprinse în fiecare din cele două momente metodologice ale cunoașterii științifice. Ajutorul solicitat s. de alte științe este facilitat prin aptitudinea ei de acomodare la caracterul procesului investigat, prin eficacitatea și economicitatea procedeelor sale.

Pe terenul ori în vecinătatea s. au apărut numeroase discipline și teorii noi, cum sînt „metriile" (biometria, econometria, psihometria, socioemetria, tehnometria, jurimetria etc.), cercetarea operațională, cibernetica, teoria informației, teoria deciziilor ș.a.

Această „explozie" a s. a făcut să se răspundă la întrebarea ce este s. astăzi, cu următoarea formulare : „s., în sensul cel mai larg, este matricea oricărei științe experimentale", (M. G. Kendall), adică o disciplină metodologică care ar trebui să fie familiară cercetătorilor, de exemplu, ca și logica. Prezența s., ca furnizor principal de informații, s-a socotit întotdeauna indispensabilă în *activitatea de conducere*. Dealtfel, cronologic, necesitatea s. s-a validat în practica conducerii. Astăzi este îndeobște cunoscut că nu este posibil a conduce o politică economică și socială modernă fără s. Cu aparatul subtil de investigare, calcul și previziune la îndemînă, s. sprijină cunoașterea realităților, fundamentează alternativele și documentează opțiunile în dezvoltarea societății contemporane. S. apare ca un instrument de studiu privilegiat la elaborarea deciziilor în mai bună cunoștință de cauză, indicînd incertitudinile care subzistă în condițiile unor informații prin natura lor incomplete. De aici și sublinierea rolului s. în procesele decizionale, ce ne conduce la Abraham Wald : „s. este o colecție de metode care ne permit să luăm decizii juste optime în caz de incertitudine". Pentru dezvoltarea s. aplicative sînt caracteristice și schimbările în tehnicile de prelucrare a datelor. Introducerea metodelor informatice de prelucrare eliberează pe statisticieni de calcule îndelungate „neinteresante" și le îngăduie să pună mai mult în valoare informația statistică și să mărească potențialul ei de cunoaștere. Ordinatoarele au creat posibilitatea să se abordeze noi tipuri de analiză, care nu erau de conceput cu tehnicile de calcul anterioare ; ele ajută să se opereze pe un plan mai larg și într-un mod mai aprofundat în analiza s. a fenomenelor. Astfel, tehnica electronică de calcul sporește forța s. ; conceptele și metodele sale continuă să pătrundă cuceritoare și înnoitoare în cele mai variate cîmpuri ale cunoașterii, pretutindeni unde fenomenele sînt de natură stohastică și se supun legilor s.

Menționăm că în ciuda importanței istoriei s., ca mod de gîndire, nu există pînă în prezent decît trei lucrări consacrate acestei probleme : a lui *V. John* (1858—1900), fost profesor de s. la Universitatea din Cernăuți (1888) [11], lucrarea similară a danezului *H. Westergaard* * (1853—1936), apărută în 1932 [34], precum și lucrarea lui *John Koren* (ed.) „The History of Statistics" (MacMillan, New York, 1918).

BIBLIOGRAFIE

1. ACHENWALL, G., *Staatsverfassung der heutigen vornehmsten europäischen Reiche und Völker im Grundrisse*, Göttingen, 1781, p. 4—5, reproduse în (30).
2. BLOCK, M., *Traité théorique et pratique de statistique*, Guillaumin et Cie, Libraires, Paris, 1878.
3. CROXTON, F. E., COWDEN, D. J., KLEIN, S., *Applied General Statistics*, ed. 3, Prentice-Hall, New York, 1967.
4. *Dicționar de filozofie*, Ed. politică, București, 1978.
5. *Dicționar statistic-economic*, Direcția Centrală de Statistică, București, 1969.
6. *Din istoria statisticii românești*, culegere de articole, Direcția Centrală de Statistică, București, 1969.
7. *Filozofie, materialism dialectic și istoric*, (coord. I. Tudosescu, M. Florea, C. Popa). Ed. științifică și enciclopedică, București, 1975.
8. GALVANI, L., *Statistica*, în *Enciclopedia italiana di scienze, lettere ed arti*, Istituto della enciclopedia italiana fondata da Giovanni Treccani, Roma, 1950.
9. HALBWACHS, M., *La théorie de l'homme moyen. Essai sur Quételet et la statistique morale*, Librairie Félix Alcan, Paris, 1913.
10. HORVATH, A. R., *Le concept de statistique internationale et son évolution historique*, în „International Statistical Review", nr. 3/1972.
11. JOHN, V., *Geschichte der Statistik*, I. Teil. *Von dem Ursprung bis auf Quételet, 1835*. Stuttgart, 1884.
12. IONESCU DE LA BRAD, I., *Povățuiri pentru catagrafia Moldovei precedate de oarecare elemente de statistică și publicate de Direcția Centrală de Statistică din Ministerul Interior al Moldovei*, Tipografia Buciumul Român. Iași, 1859.
13. LACQUART, C., *Statistique et science sociale*, Bruxelles, 1907.
14. JULIN, A., *Principes de statistique théorique et appliquée*, Bruxelles,1921.
15. KENDALL, M. G., *The History of Statistical Method*, în „*International Encyclopedia of the Social Sciences*", vol. 15, The Macmillan Company & The Free Press, 1968.
16. KENDAL, M. G., BUCKLAND, W. R., *A Dictionary of Statistical Terms*, Oliver & Boyd, London, 1960.
17. LHESSE, A., *La statistique — ses procédés — ses résultats*, Paris, 1905.
18. MARȚIAN, D. POP., *Articol program*, în „Anale statistice" nr. 1/1860.
19. MARȚIAN, D. POP, *Statistica justițială*, în „Anale statistice" nr. 17—20/1864.
20. MOINEAGU, C., NEGURĂ, I., URSEANU, V., *Statistica*, (Coordonator și prefață V. Trebici). Ed. științifică și enciclopedică, București, 1976.
21. MOOD, A. M., GRAYBILL, Fr. A., *Introduction à la statistique théorique*. Traduit de l'anglais, Dunod, Paris, 1973.
22. MOREAU de JONNÈS, A., *Elèments de statistique*, Deuxième édition. Paris, 1856.
23. MORICE, E., CHARTIER, F., *Méthode statistique*, Première partie,Elaboration des statistiques. Imprimerie Nationale, Paris, 1954.
24. ONICESCU, O., *Statistica și sociologia*, în „Arhiva pentru știință și reformă socială", Organ al Institutului Social Român, anul IX, nr. 4/1931.
25. PETTY, W., *Arithmétique politique*, în *Les oeuvres économiques de Sir William Petty*, traduit de l'anglais par Henry Dussauze-Maurice Pasquier, Tome premier, Paris, 1905.

V.TREBICI

26. PIATIER, A., *Statistique et observation économique*, Tome premier, *Méthodologie statistique*, Presses Universitaires de France, Paris, 1961.
27. QUÉTELET, A., *Sur l'homme et le développement de ses facultés ou Essai de physique sociale*, Tome premier, Paris, 1835.
28. SĂHLEANU, V., *Evoluția gîndirii statistice*, în „Revista de Filozofie" nr. 5, 1979.
29. SÜSSMILCH, J. P., *„L'ordre divin" aux origines de la démographie*. Traduction originale par Jacqueline Hecht, vol. I și II, Institut National d'Etudes démographiques, Paris, 1979.
30. SWOBODA, H., *Knaurs Buch der modernen Statistik*, Droemersche Verlagsanstalt Th. Knaur Nachf., München/Zürich, 1971.
31. TREBICI, V., *Mică enciclopedie demografică*, Ed. științifică și enciclopedică, București, 1975.
32. VODĂ, V., Gh., *Gîndirea statistică — un mod de gîndire a viitorului*, Ed. Albatros, București, 1977.
33. WALLIS, W. A., ROBERTS, H. V., *The Nature of Statistics*, Collier Books, New York, 1962.
34. WESTERGAARD, H., *Contributions to the History of Statistics*, London, 1932.
35. YULE, G. U., KENDALL, M. G., *Introducere în teoria statisticii*, Ed. științifică, București, 1969.
36. (...), *Les sciences sociales dans l'enseignement supérieur. La statistique*, U.N.E.S.C.O., Paris, 1957.

TEORIA PROBABILITĂȚILOR. Teorie matematică ce se ocupă de studiul fenomenelor întîmplătoare sau aleatoare utilizînd în acest scop metodele deductive ale matematicii. Fenomenelor aleatoare de masă le este caracteristică proprietatea de stabilitate a frecvențelor. Conceptul fundamental de *probabilitate* * corespunde intuitiv constantei în jurul căreia oscilează frecvențele și către care șirul acestor frecvențe tinde cînd numărul probelor crește nelimitat. Există păreri potrivit cărora t.p. nu ar fi matematică ci ar reprezenta o ramură a logicii; după alții, t.p. ar face parte din studiul naturii. În interpretarea probabilității, se distinge — cu oarecare convenționalitate — o definiție clasică și o alta modernă, numită axiomatică.

Istoric. Preocupările teoretice legate de probabilități au fost generate de jocurile de noroc și de asigurările de persoane. Cavalerul de Méré, om de spirit și amator de jocuri de noroc, s-a adresat lui *B. Pascal* (1623—1662) cu unele probleme legate de jocurile cu zaruri și cu șansele de cîștig în diferite situații. Curînd, de aceste probleme se interesează matematicianul *P. Fermat* * (1601—1665) și matematicianul actuar *Chr. Huygens* * (1629—1695). Începuturile t.p. datorate acestor oameni de știință, la care se adaugă studiile lui Cardano, cu a sa „Liber de Ludo Aleae" (Cartea despre jocul de noroc), apărută în 1663, și al lui Galileo privind erorile de măsurare, ca și studiile desfășurate de specialiștii în teoria asigurărilor, se continuă rapid, culminînd cu lucrarea lui *J. Bernoulli* * (1654—1705), intitulată „Ars Conjectandi" și apărută postum (1713), în care se dă prima formă a legii numerelor mari. *A. de Moivre* * (1667—1754) face primele observații asupra legii normale, care va fi dezvoltată de *C. F. Gauss** (1777—1855). Epoca clasică a t.p. este ilustrată de *P. S. Laplace* * (1749—1827), a cărui concepție este expusă mai cu seamă în „Théorie analytique des probabilités" (1812) și „Essai ..." (1814). Remarcabil este faptul că Laplace este printre primii care se preocupă de aplicațiile practice ale t.p., ceea ce prefigu-

rează constituirea statisticii matematice. Contribuții importante se datorează lui *S. D. Poisson* * (1781—1840), care a pus în evidență legea evenimentelor rare, ca și școlii ruse de **t.p.**, reprezentată de *P. L. Cebîșev* * (1821—1894), *A. M. Leapunov* * (1857—1918) și *A. A. Markov* * (1856—1922), care au dezvoltat *teorema limită centrală** și studiul variabilelor aleatoare dependente. Aceste școli naționale contribuie la progresul **t.p.** Școala italiană este reprezentată de Cantelli și Bruno de Finetti; cea franceză este legată de numele lui *E. Borel* * (1871—1956), P. Levy (1886—1972), *M. Fréchet* * (1878—1973); cea germană a dat în special pe *R. v. Mises* * (1883—1953). Trebuie consemnate totodată lărgirea aplicațiilor **t.p.** în diferite domenii și progresul considerabil al *statisticii matematice* *.

Faza modernă a **t.p.** este marcată în special de *A. N. Kolmogorov* * (n.1903), cu a sa lucrare fundamentală „Grundbegriffe der Wahrscheinlichkeiten" (1933), în care este expusă axiomatizarea probabilităților. Alte contribuții ale probabiliștilor sovietici dezvoltă această teorie: *S. N. Bernstein* * (1880—1968), care a sintetizat realizările școliilor franceză și rusă. *A. A. Hincin* * (1894—1959), care a dezvoltat teoria proceselor aleatoare staționare, *N. S. Smirnov* * (1900—1966), care a pus bazele teoriei metodelor neparametrice. Școala română de **t.p.**, legată de numele Acad. *O. Onicescu* * (1892—1983) și Acad. *Gh. Mihoc** (1906—1981), are contribuții originale la **t.p.**, ale căror începuturi datează încă din deceniul al patrulea. Cercetătorii români sînt printre primii care au studiat teoria lanțurilor Markov, au introdus noțiunea de probabilitate inversă, au rezolvat problema limită centrală pentru lanțuri Markov finite omogene, au introdus un nou tip de dependență a variabilelor aleatoare (dependența cu legături complete). **T.p.** și-a lărgit aplicabilitatea în ultimele decenii, putînd fi amintită contribuția sa la teoria informației, teoria fiabilității, teoria așteptării, controlul statistic al calității, teoria sistemelor.

Stadiul actual. Problematica **t.p.** este deosebit de vastă. Ca orice teorie matematică **t.p.** operează cu concepte abstracte, relațiile dintre acestea fiind de natură logică. Ea se construiește plecînd de la anumite noțiuni de bază definite cu ajutorul obiectelor fundamentale ale matematicii (mulțimi, relații, aplicații etc.) și satisfăcînd un sistem determinat de axiome. Alegerea noțiunilor și a axiomelor este călăuzită de preocuparea de a adapta la maximum conceptele noi, astfel elaborate, la conținutul „intuitiv" al situațiilor naturale ce se cer analizate. Atît istoria cît și dezvoltările recente ale **t.p.** atestă interesul pentru cunoașterea și explicarea anumitor legități ale existenței reale. În 1923, Borel scria: „... strict vorbind, s-ar putea considera o teorie (**t.p.** — *n.n.*) pur logică, fără a se preocupa de existența posibilă a aplicațiilor sale, dar o astfel de teorie ar fi un curat joc de gîndire fără interes, care nu ar merita numele de știință". Lucrările moderne, în special ale lui A. N. Kolmogorov, au făcut din **t.p.** o ramură a științelor matematice logic construită, asemenea geometriei, plecînd de la axiome (*teoria axiomatică* a probabilității întemeiată pe teoria mulțimilor). O abordare a **t.p.** pe această linie modernă se datorează lui O. Onicescu și este prezentată în „Principiile teoriei probabilităților" (1969). Cît privește domeniul actual al **t.p.**, în general se disting două părți: partea clasică, în care se studiază spațiul de probabilitate și variabilele aleatoare definite pe el, și partea modernă, consacrată proceselor stohastice. Partea din urmă se numește și dinamica **t.p.** față de cea dintîi care apare ca o știință a structurii obiectelor acestei teorii.

Probabilitate. Caracterizează gradul de posibilitate a producerii unui eveniment în condiții bine determinate. Procesul de determinare a p. unui eveniment este

tot atît de complex ca și natura fenomenelor. În primă abordare — *abordare clasică* —, *p.* unui eveniment este raportul dintre numărul rezultatelor (cazurilor) favorabile producerii acestui eveniment și numărul total al rezultatelor (cazurilor) egal posibile (egal probabile). Astfel, *p.* căderii pe „față" a monedei la jocul „față sau revers" este 1/2; cînd se trage o carte într-un joc de 52 de cărți, *p.* de a obține un caro este 13/52, o carte roșie 1/2, o carte roșie sau o carte neagră 1 (certitudine), o carte care să nu fie nici roșie, nici neagră 0 (imposibilitate). Pentru ca definiția să fie valabilă, trebuie satisfăcută condiția de „echiprobabilitate a evenimentelor elementare"(de unde, în exemplele considerate, monedă omogenă, pachet de cărți de joc standard) și condiția de„spațiu finit de evenimente".

P. clasică este uneori numită *p. apriori* (alteori *p. teoretică*), deoarece poate fi enunțată dinainte, fără experimentare prealabilă. De exemplu, știind că în jocul de cărți sînt 4 valeți printre cele 52 de cărți ce se pot trage succesiv, la întîmplare, atunci $P_{(valet)} = 4/52 = 1/13 = 0{,}077$. În fapt, există numeroase fenomene aleatoare în care evenimentele elementare nu se pot socoti și deci evaluarea *apriori* a unei *p.* apare imposibilă (și din incapacitatea de a verifica ipoteza echiprobabilității). Regula clasică devine, bunăoară, inaplicabilă pentru calculul *p.* ca o persoană, astăzi în viață, să supraviețuiască un an, doi ani, trei ani, ... sau *p.* cu care peste o regiune geografică dată va cădea ploaie într-un interval de timp dat. Ca urmare, s-a recurs utilizînd un număr mare de observații, la o evaluare (statistică) a frecvenței relative a evenimentului și s-a convenit să se adopte această frecvență drept *p.*

Tabela de mortalitate, construită pe bază de date statistice, indică, de pildă, *p.* ca o persoană de vîrsta x să ajungă la vîrsta $x + 1$.

Abordarea *p.* ca un *concept experimental* își are temeiul în una din însușirile fundamentale ale fenomenelor aleatoare, anume *stabilitatea frecvențelor relative*: la capătul unei lungi serii de repetări ale probei, este practic sigur că frecvența relativă a evenimentului va fi aproximativ egală cu *p.*

Exemplu. Dacă se aruncă un zar (perfect simetric din punct de vedere geometric și omogen în repartiția materiei din care este construit) de repetate ori și se notează frecvența lui 6, se constată că frecvența de apariție a acestui număr de puncte tinde — pe măsură ce numărul de aruncări crește — să se stabilizeze în jurul valorii 1/6, adică a valorii *p.*

Probabilitate subiectivă. Este o codificare de informații (sau, pentru a relua o formulă sugestivă, „traducerea bunului simț în cifre"), emanată de la o persoană interesată în evaluarea probabilității. Un conducător, spre exemplu, care trebuie să ia o decizie privind o anumită comportare în activitatea pe care o conduce, intuiește, pe baza informației *apriori*, mai multe posibilități de acțiune despre al căror rezultat nu este sigur. În această situație, el evaluează șansele de reușită ale unei acțiuni sau alteia. Firește, interpretarea probabilității ca limită de frecvență relativă nu are sens. Ca urmare, subiectul deciziei va fi îndemnat să codifice informația de la 0 la 1 și, astfel, să ierarhizeze opțiunile respective. Probabilitățile atribuite se interpretează ca un *pariu echitabil*. Presupunînd probabilitatea 0,6 pentru opțiunea întîia, înseamnă că, dacă conducătorul ar trebui să parieze un leu, el pariază 60 de bani pe un rezultat pozitiv și 40 bani pe un rezultat contar: o asemenea diviziune a riscurilor ar fi coerentă cu întreaga informație rn posesia sa. Abordarea subiectivă a probabilității a fost introdusă în 1926i de Frank Ramsey în lucrarea sa „The Foundation of Mathematics and Other Logical Essays". Conceptul a fost dezvoltat de Bernard Koopman, Richard Good, Bruno de Finetti și Leonard

Savage. Precursor al conceptului de probabilitate statistică trebuie considerat *Thomas Bayes"* * (1702—1761) (inferența bayesiană).

BIBLIOGRAFIE

1. BOREL, Emile, *Probabilities and Life* (trad. versiunii franceze, 1943), New York, 1962.
2. GNEDENKO, B. V., KHINCIN, A., *An Elementary Introduction to the Theory of Probability* (trad. versiunii ruse, ed. V), New York, 1962.
3. IOSIFESCU, M., *Lanțuri Markov finite și aplicații*, Ed. tehnică, București, 1977.
4. KOLMOGOROV, A. N., *Foundations of the Theory of Probability*, New York, 1956 (apărută prima dată în limba germană, 1933; ediția rusă 1936).
5. KOLMOGOROV, A. N., *Matematica: conținutul, metodele și importanța ei* (cap. XI. *Teoria probabilităților*), Ed. științifică, București, 1960.
6. METVIER, M., *Notions fondamentales de la théorie des probabilités*, Dunod, Paris, 1972.
7. MIHOC, Gh., BERGTHALLER, C., URSEANU, V., *Procese stohastice, Elemente de teorie și aplicații*, Ed. științifică și enciclopedică, București, 1978.
8. MIHOC, Gh., IOSIFESCU, M., URSEANU, V., *Elemente de teoria probabilităților și aplicațiile ei*, Ed. științifică, București, 1966.
9. MIHOC, Gh. ș.a., *Teoria probabilităților și statistica matematică*, Ed. didactică și pedagogică, București, 1970.
10. MIHOC, Gh., MICU, N., *Teoria probabilităților și statistică matematică*, Ed. didactică și pedagogică, București, 1980.
11. ONICESCU, O., MIHOC, Gh., IONESCU-TULCEA, I., *Calculul probabilităților și aplicații*, Ed. Academiei R.S.R., București, 1956.
12. ONICESCU, O., *Principiile teoriei probabilităților*, Ed. Academiei R.S.R., București, 1969.
13. WEAVER, W., *Doamna Șansă*, Ed. tehnică, București, 1970.

STATISTICĂ MATEMATICĂ. Disciplină matematică consacrată elaborării noțiunilor și metodelor specifice studiului statistic al fenomenelor de masă. Bazată pe *teoria probabilităților* *, din care s-a desprins, **s.m.** oferă metodele generale decurgînd din caracterul abstract al relațiilor tratate. Într-o formulare mai generală, **s.m.** se ocupă cu gruparea, analiza și interpretarea datelor referitoare la anumite fenomene, precum și cu unele previziuni privind producerea lor în viitor. Spre deosebire de teoria probabilităților al cărei demers este deductiv, **s.m.** urmează calea inductivă: pornind de la cunoașterea modului de repartizare a frecvențelor statistice, **s.m.** își propune ca prin metode inductive să obțină informații referitoare la legile de probabilitate ale fenomenului care a produs frecvențele. În domeniul **s.m.** se includ: *teoria sondajului* * — formularea principiilor și procedeelor de prelevare a eșantioanelor și de prelucrare a datelor de observație potrivit cu regulile calculului probabilităților, în scopul de a dobîndi cantitatea de informație necesară estimării parametrilor populației reprezentate: *teoria estimației* * — construirea formulelor care să permită evaluarea parametrilor unei populații pe baza informației rezultate din cercetarea pe eșantion; *verificarea ipotezelor statistice* * — elaborarea metodelor cu ajutorul cărora să se judece valabilitatea ipotezelor privind parametrii sau forma de repartiție a unei populații (plecînd de la eșantion); *teoria*

deciziilor statistice — stabilirea metodelor de inferență statistică aplicabile la luarea deciziilor în condiții de incertitudine (alegerea între alternative cu includerea în calcul a repercusiunilor acțiunii).

Adesea, termenul **s.m.** este folosit pentru a denumi partea „inductivă" a statisticii, adică partea de analiză a informației obținute pornind de la un mecanism aleator. Distincția între statistica analitică și statistica descriptivă ar corespunde aproximativ împărțirii în **s.m.** și statistică elementară. **S.m.** s-a dezvoltat însă ca o disciplină pur formală, în care numărul este studiat în afara oricărei referințe la lucrul numărat. Ca disciplină matematică specializată, **s.m.** se ocupă de aparatul matematic utilizat în statistică ca mijloc de a trage pe deplin folosae din cantitatea de informație conținută în datele statistice. *Istoric.* Începuturile **s.m.** sînt prefigurate în special în lucrările lui Laplace. Constituirea ei — spre sfîrșitul sec. al XIX-lea — a fost pregătită de discipline ca : teoria probabilităților, statistica demografică, teoria erorilor, teoria seriilor dinamice, unele contribuții ale psihologiei și sociologiei. Crearea ei este legată de școala anglo-saxonă de **s.m.**, legată în special de numele lui *Fr. Galton* * (1828—1911) și *K. Pearson* * (1857—1936), a cărui carte fundamentală „The Grammar of Science" (1892) marchează un nou mod de gîndire statistică. Studiul relațiilor dintre variabile de către K. Pearson, în jurul anului 1890, a dus la formularea unor concepte și metode care constituie și astăzi corpul **s.m.** : corelația și regresia (indicate de F. Galton), sondajul, criteriul χ^2, formulat de F. Helmert (1875) și independent de K. Pearson, ca și alte metode. Ele sînt dezvoltate de o întreagă pleiadă cuprinzînd pe : *G. U. Yule* * (1871—1951), *W. S. Gosset* * (1876—1937), cunoscut sub pseudonimul de Student, și mai ales *R. A. Fisher* * (1890—1962). Progrese importante se realizează în teoria eșantioanelor.

Începutul a fost marcat de cercetările lui Student și Fisher. Student a prezentat în 1908 una din cele mai importante repartiții statistice — așa-numita repartiție *t*, provenind dintr-o problemă de eșantioane mici. Într-o succesiune rapidă, Fisher a descoperit repartiția, în eșantioane dintr-o populație normală a coeficientului de corelație, a coeficienților de regresie, a coeficienților de corelație multiplă și a raportului de varianțe, cunoscut ca raportul *F*. Între 1925 și 1935, s-au elaborat : teoria estimației, sub influența lui Fisher și teoria testării ipotezelor, inspirată de *Egon Pearson* * (1895—1980) și *Jerzy Neyman* * (1894—1981). Paralel cu dezvoltările teoriei sondajelor, se află în progres remarcabil teoria planificării experimentelor, bazată pe lucrările lui Fisher asupra experimentării în agrobiologie. Au apărut apoi lucrările lui Williams Shewhart (1891—1967), care au fost larg extinse la metodele utilizate de industrie pentru controlul statistic al calității și fiabilității. În perioada celui de-al doilea război mondial, impulsuri noi au venit de la *Abraham Wald* * (1902—1950), autorul teoriei testului secvențial. Wald a avut în vedere o abordare mai generală, care s-a unit cu ideile lui Neyman despre testarea ipotezelor și s-a dezvoltat într-o teorie de mare atracție intelectuală — teoria funcțiilor de decizie. Alți autori de prestigiu. John Wishart și *M. G. Kendall* * (1907—1983) în Anglia, *Harold Hotelling* * (1890—1973) și Samuel Wilks în S.U.A., *Corrado Gini* * (1884—1965) în Italia, S. N. Roy și R. C. Bose în India, dar mai ales *P. C. Mahalanobis* * (1893—1972), *Ragnar Frisch* * (1894—1973) în Norvegia, *V. S. Nemcinov* * (1894—1964) în U.R.S.S., adaugă un mare număr de noi rezultate, mai ales în domeniul analizei multidimensionale. Expansiunea cercetării științifice din ultimele decenii, pe căi originale, a dus la descoperirea de procedee de investigație din ce în ce mai elaborate, mai eficiente și la cucerirea unor noi sfere de aplicație statistică.

V.TREBICI

Planificarea experimentelor * (*experimental designs*) a fost domeniul predilect al celebrei „Rothamsted Experimental Station" din Anglia, unde între anii 1919—1930 au lucrat R. A. Fisher și o serie de alți savanți. Școala a fost continuată în S.U.A., la Iowa State College, Laboratorul de statistică, cu reprezentanți remarcabili ca *G. W. Snedecor* * (1881—1974), *W. G. Cochran* (1909—1980). G. Cox și alții. Analiza factorială inițiată mai ales de psihologi a fost reprezentată de *C. E. Spearman* * (1883—1945), L. C. Thurstone, C. Burt, G. H. Thompson și *H. Hotelling* * (1890—1973). În ultimele decenii interesul s-a îndreptat spre teoria funcțiilor de decizie statistică, teoria metodelor neparametrice, analiza multidimensională, inferența statistică pentru procese stohastice.

Cu totul remarcabilă este lărgirea cîmpului de aplicare a metodelor **s.m.** în biologie, agricultură, controlul calității. Contribuții deosebite la dezvoltarea controlului statistic al calității are *W. A. Shewhart* (1891—1967), care și-a expus principiile în 1931, în lucrarea „Controlul economic al calității produselor industriale" și în „Metoda statistică din punctul de vedere al controlului calității", în colaborare cu W. E. Deming (1939). La dezvoltarea **s.m.** o contribuție importantă are revista „Biometrika", înființată de F. Galton și K. Pearson în 1901 și în care s-au publicat numeroase studii ale școlii anglo-saxone. Treptat, au fost înființate și alte reviste de **s.m.** în numeroase țări.

BIBLIOGRAFIE

1. CEAPOIU, N., *Metode statistice aplicate în experiențele agricole și biologice*, Ed. agro-silvică, București, 1968.
2. CIUCU, G. ș.a., *Statistică matematică și cercetări operaționale*, Ed. didactică și pedagogică, București, 1974.
3. GEORGESCU-ROEGEN, N. St., *Metoda statistică. Elemente de statistică matematică*, Imprimeria Națională, București, 1933.
4. IOSIFESCU, M., TĂUTU, P., *Procese stohastice și aplicații în biologie și medicină*, Ed. Academiei R.S.R., București, 1968.
5. KENDALL, M. G., STUART, A., *The advanced theory of statistics*, tome 1, Hafner, New York, 1958.
6. McGEE, V., *Principles of statistics. Traditional and Bayesian*, New York, 1971.
7. MORICE, E., *Dictionnaire de statistique*, Dunod, Paris, 1968.
8. ONICESCU, O., MIHOC, Gh., *Lecții de statistică matematică*, Ed. tehnică, București, 1958.
9. RANCU, N., TÖVISSI, L., *Statistica matematică cu aplicații în producție*, Ed. Academici R.S.R., București, 1963.
10. SĂHLEANU, V., *Metode matematice în cercetarea medico-biologică*, Ed. medicală, București, 1957.
11. SNEDECOR, G. W., *Metode statistice aplicate în cercetările de agricultură și biologie*, Ed. didactică și pedagogică, București, 1968.
12. TÖVISSI, L., VODĂ, V. Gh., *Metode statistice. Aplicații în producție*, Ed. științifică și enciclopedică. București, 1982.

STATISTICĂ DESCRIPTIVĂ. 1. (ist.) Fază a statisticii în sec. XVII—XIX. 2. Parte a statisticii avînd ca obiect culegerea și clasificarea datelor statistice și, pe această bază structurală, descrierea fenomenelor investigate. După G. Calot, **s.d.** încearcă să rezume cantitativ informația culeasă asupra unui univers concret cu ajutorul unei cercetări exhaustive: scopul ei nu este de a

explica ci de a descrie, de a pune în evidență esențialul, de a realiza sinteze cu ajutorul unui limbaj numeric. În demersul său descriptiv, **s.d.** pornește de la observarea masei de fenomene singulare, materializată în strîngerea datelor de bază, originale. Datele observării, neavînd o semnificație simplă și imediată, sînt supuse apoi unor operații de prelucrare pentru a le aduce la o formă clară și utilizabilă. În prelucrarea inițială, operația de prim ordin este *gruparea*; repartițiile statistice, rezultate din grupări se prezintă în *tabele* și, dacă imprejurările o cer, se transpun în *grafice*.

Odată ce materialul statistic a fost „turnat în forme" convenabile, se caută parametrii care pot descrie cel mai bine *cîmpurile de variabilitate* după diferitele lor caractere. Obiectivul este atins prin calcule de valori relative, valori medii, dispersii, abateri și alte valori tipice. Un alt grup de preocupări vizează o formă de realitate superioară în care se integrează populația statistică observată și se concretizează în prelucrarea aprofundată a *legăturilor* dintre elementele populației. Procedeele ce se folosesc sînt multiple; de ele se ocupă teoria asocierii, regresiei și corelației. Între **s.d.** și statistica matematică există o simetrie: populație — spațiu de evenimente elementare; frecvență — probabilitate, repartiții empirice — repartiții teoretice, variabilă statistică — variabilă aleatoare, medie statistică — speranță matematică (media unei variabile aleatoare). Legătura dintre **s.d.** și statistică matematică se realizează prin conceptul de probabilitate. Așadar, **s.d.** trebuie considerată o etapă a demersului statistic, o fază introductivă a acestuia. În general, domeniul ei cuprinde: repartiții empirice, tabele statistice, reprezentări grafice, indicatori de poziție, dispersie și concentrare, repartiții bidimensionale, regresie și corelație, la care se adaugă seriile cronologice și indicii care se aplică la fenomenele economice.

BIBLIOGRAFIE

GÉRARD CALOT, *Cours de statistique descriptive*, Deuxième édition, Dunod, Paris, 1973.

STATISTICĂ INFORMAȚIONALĂ ONICESCU. Are drept domeniu proprietățile globale de sistem ale uneia sau mai multor populații (sau repartiții). A fost formulată în 1966, ca dezvoltare a teoriei matematice a informației a lui Ch. Shannon, prin introducerea conceptului de *energie informațională* *. **Energie informațională.** Fie S un sistem cu un număr finit de caracteristici $Y_1, Y_2, \ldots, Y_n$ cu probabilitățile (sau frecvențele sau ponderile) respective $p_1, p_2, \ldots, p_n$. Expresia $E = \sum_{i=1}^{n} p_i^2$ este numită (după forma ei, fără altă semnificație mecanică) *e.i.*

Valoarea maximă este realizată cînd sistemul are o *stare de determinare* statistică, adică una dintre probabilități are valoarea 1. Valoarea minimă este obținută pentru sistemele cu *stare de nedeterminare* (statistică) completă, adică toate stările au aceeași probabilitate. Deci

$$\frac{1}{n} \leqslant E \leqslant 1.$$

Corelație informațională. Dacă $p_1, p_2, \ldots, p_n$ și respectiv $q_1, q_2, \ldots, q_n$ sînt probabilitățile (frecvențele sau ponderile) a două sisteme S_1 și S_2 relativ la

aceleași caracteristici comune Y_1, Y_2, ..., Y_n, expresia

$$c_{12} = \sum_{i=1}^{n} p_i q_i$$

reprezintă *c.i.* a celor două sisteme relativ la sistemul de caracteristici Y_1, Y_2, ..., Y_n.

Corelația unui sistem cu el însuși este identică cu energia informațională. În general

$$0 \leqslant c_{12} \leqslant 1.$$

Dacă $C_{12} = 0$, cele două sisteme sînt „străine" între ele; n-au nici o caracteristică comună (din punct de vedere statistic, bineînțeles).

Coeficient de corelație informațională între S_1 și S_2. Este definit de expresia

$$r_{12} = \frac{c_{12}}{\sqrt{E_1 E_2}} = \frac{\sum p_i q_i}{\sqrt{(\sum p_i^2)(\sum q_i^2)}}.$$

În general,

$$0 \leqslant r \leqslant 1.$$

Teoremă importantă. $r_{12} = 1$ dacă și numai dacă cele două sisteme au repartiții identice (sînt identice din punct de vedere statistic). *c.c.i.* constituie instrumentul principal în studiile și aplicațiile statisticii informaționale.

Cazul $n > 2$. Definițiile și proprietățile corelației informaționale se generalizează direct pentru orice număr finit de sisteme.

Dacă $n \to \infty$ și seriile care definesc energia informațională și coeficientul respectiv sînt convergente, definițiile și proprietățile enunțate rămîn valabile.

Dacă repartițiile sînt continue și $\rho_i(x)$ sînt densitățile respective, generalizările noțiunilor informaționale se obțin fără vreo modificare esențială. De pildă, coeficientul de corelație pentru două repartiții continue cu densitățile $\rho_i(x)$. $\rho_2(x)$ este dat de expresia

$$r = \frac{\int \rho_1(x)\, \rho_2(x)}{\left(\int \rho_1^2(x)\ dx\right)\left(\int \rho_2^2(x)\ dx\right)},$$

fără a face ipoteze sau a introduce condiții noi. Conceptele și metodele statisticii informaționale Onicescu cunosc o largă aplicare în economie, demografie, marketing [6], sociologie [2], pedagogie [3] etc. O aplicație recentă este econometria informațională. O dezvoltare a energiei informaționale se datorează lui Cresin [1]; o încercare de generalizare a făcut Gh. Mihoc [2].

BIBLIOGRAFIE

1. CRESIN, R., *Aplicarea energiei informaționale în studiul longevității*, în „Revista de Statistică" nr. 12, 1966.
2. MIHOC, Gh., URSEANU, V., URSIANU, E., *Modele de analiză statistică*, Ed. științifică și enciclopedică, București, 1982 (cap. 4. Statistica informațională. Energia și corelația informațională).
3. NOVAK, Andrei, *Metode statistice în pedagogie și psihologie*, Ed. didactică și pedagogică, București, 1977 (cap. 7 Elemente de teoria informației).
4. ONICESCU, O., *Energia informațională, componentă a unui barometru statistic al sistemelor*, în „Revista de Statistică" nr. 11, 1966.
5. ONICESCU, O., *Energie informationelle*, C.R. Acad. Sci. Paris, Serie A, 26 nov. 1966.
6. ONICESCU, O., *Energia informațională și aplicațiile ei*, în *Marketing intern și internațional, concepte, metode și instrumente analitice pentru decizie*, Ed. politică, București, 1976.
7. ONICESCU, O., ȘTEFĂNESCU, V., *Elemente de statistică informațională cu aplicații*, Ed. tehnică, București, 1979.
8. STERIAN, P., *Aplicarea energiei informaționale în studii sociologice*, în „Revista de Statistică" nr. 2, 1967.
9. ȘTEFĂNESCU, V., *Aplicațiile energiei și corelației informaționale*, Ed. Academiei R. S. România, București, 1979.

STATISTICĂ ECONOMICĂ. Disciplină care are ca obiect studiul statistic al fenomenelor și proceselor economice de masă. Sub termenul de **s.e.** se reunește întregul studiu statistic al vieții economice în ansamblul său și diferențiată pe ramuri economice. În mod corespunzator se disting: **s.e.** *generală* și *statistica de ramură* *. Capitolele importante ale **s.e.** generale sînt: statistica avuției naționale, statistica produsului social, statistica venitului național, sistemul de balanțe al economiei naționale (balanța reproducției produsului social și a venitului național, balanța legăturilor dintre ramuri, balanța avuției naționale, balanța muncii). Uneori, prin **s.e.** se înțelege, în esență, *statistica social-economică*, care cuprinde studiul statistic nu numai al fenomenelor economice, dar și al tuturor celorlalte fenomene sociale. În acest context se mai includ: statistica populației (statistica demografică — cu cea mai veche tradiție), statistica învățămîntului, statistica sănătății și ocrotirii sănătății, statistica judiciară etc. Procesul de diversificare a statisticii pe ramuri este în continuă desfășurare.

Domeniul predilect al **s.e.**: metoda balanțelor, seriile cronologice, indicii, analiza tendințelor; contribuții importante: *E. Laspeyres* * (1834—1913); *Fr. Y. Edgeworth* * (1845—1926); *E. Engel* * (1821—1896), *I. Fisher* * (1867—1947). *H. Paasche* * (1851—1925), W. M. Persons și alții.

BIBLIOGRAFIE

1. ARAGON, Y., TRINQUIER-ALCOUFFE, C., *Introduction à la statistique en sciences sociales*, Col. Societas, Toulouse, 1979.
2. BĂRBAT, Al., *Teoria statisticii sociale*, Ed. didactică și pedagogică, București, 1972.
3. CAPANU, I., *Statistica economică*, ed. a II-a, Ed. didactică și pedagogică, București, 1978.
4. CAPĂTĂ, M., *Despre rolul și locul statisticii în cadrul sistemului informațional economico-statistic*, în „Revista de Statistică" nr. 11, 1973.

V.TREBICI

5. *Dicționar statistic-economic*, ed. a II-a, Direcția Centrală de Statistică, București, 1969.
6. DUMAS, R., *L'entreprise et la statistique*, Tome I. Dunod, Paris, 1967.
7. LEVIN, I., RICHARD, *Statistics for Management*, Prentice-Hall Inc., London, 1978.
8. PINTY J.-J., GAULTIER, C., *Dictionnaire de mathématiques et statistiques en sciences humaines*, Encyclopédie Universitaire, Paris, 1971.
9. YEOMANS, K. A., *Statistics for the social scientist*, Vol. I și II, Penguin Books, Middlesex, 1968.

STATISTICĂ ANALITICĂ (*inductivă* sau *inferențială*). Parte a statisticii care cuprinde metodele de apreciere critică a valabilității parametrilor empirici; parte centrală a statisticii matematice. Parametrii calculați sînt definiți și precizați (cu condiția aplicării unor procedee corecte) numai în ce privește colecția de unități observate; ei pot fi luați drept reflectări mai mult sau mai puțin deformate ale unor parametri într-o populație mai întinsă, reală sau ipotetică, cunoscută sau necunoscută. Pentru estimații dincolo de datele disponibile, se aplică metodele de inferență statistică, intrînd în acțiune, ca un instrument complementar, conceptul de probabilitate. Mai întîi, prin inducții dirijate și controlate, se dezvăluie modelele teoretice (probabiliste) aplicabile observațiilor: apoi, reprezentările concrete sînt confruntate cu modelele induse și se formulează — în termeni de probabilitate — gradul de încredere ce se poate acorda concluziilor. Este evident că un astfel de demers comportă un risc de eroare, imposibil de anulat. Se recurge la calculul probabilităților pentru a măsura acest risc și a selecționa procedurile care să-l minimizeze pe cît este posibil.

Cea mai mare parte a cercetărilor statistice — chiar cînd par exhaustive — au ca obiect „eșantioane" dintr-o populație infinită și impun parcurgerea completă a drumului metodologic schițat. Trebuie să admitem că, dacă nu s-ar realiza saltul spre utilizarea probabilităților, statistica nu s-ar elibera niciodată de cazurile particulare și vederile sale ar fi extrem de limitate. Probabilitatea „imprimă marca de adevărat obiect al științei statistice. Cei ce nu vor să se uite la această marcă pot, totuși, să facă știință bună dacă marca există. Dar ei nu o știu sau, mai curînd, nu vor să o știe. Dacă, însă, această marcă nu există, simpla elaborare a datelor statistice constituie o muncă zadarnică, o moară care se învîrte în gol. Acesta este riscul empiriștilor puri" (Onicescu). Noțiunile de statistică și probabilități formează aceeași linie de gîndire, care oferă cadrul conceptual capabil să depisteze structurile ascunse sub mulțimea faptelor observate. În sensul modern al termenului, statistica este privită tot mai mult ca „știința interpretării faptelor prin raportare la modele probabiliste".

BIBLIOGRAFIE

1. CEAPOIU, N., *Metode statistice aplicate în experiențele agricole și biologice*, Ed. agro-silvică, București, 1968.
2. CIUCU, G., ș.a., *Statistică matematică și cercetări operaționale*, Ed. didactică și pedagogică, București, 1974.
3. GEORGESCU-ROEGEN, N. St., *Metoda statistică. Elemente de statistică matematică*, Imprimeria Națională, București, 1933.
4. IOSIFESCU, M., TĂUTU, P., *Procese stohastice și aplicații în biologie și medicină*, Ed. Academiei R.S.R., București, 1968.

V.TREBICI

5. KENDALL, M. G., STUART, A., *The advanced theory of statistics*, tome 1 Hafner, New York, 1958.
6. McGEE V., *Principles of statistics. Traditional and Bayesian*, New York, 1971.
7. MORICE, E., *Dictionnaire de statistique*, Dunod, Paris, 1968.
8. ONICESCU, O., MIHOC, Gh., *Lecții de statistică matematică*, Ed. tehnică, București, 1958.
9. RANCU, N., TÖVISSI, L., *Statistica matematică cu aplicații în producție*, Ed. Academiei R.S.R., București, 1963.
10. SĂHLEANU, V., *Metode matematice în cercetarea medico-biologică*, Ed. medicală, București, 1957.
11. SNEDECOR, G. W., *Metode statistice aplicate în cercetările de agricultură și biologie*, Ed. didactică și pedagogică, București, 1968.
12. TÖVISSI, L., VODĂ, V. Gh., *Metode statistice. Aplicații în producție*, Ed. științifică și enciclopedică, București, 1982.

STATISTICĂ TEORETICĂ (*teorie a statisticii*). 1. Sinonimă cu statistica matematică [3]; 2. termen mai general cuprinzînd statistica matematică [4], unele metode și concepte cum ar fi seriile cronologice, indicii, metode elaborate de psihologie și alte discipline, avînd un caracter de maximă generalitate; 3. teorie a unui domeniu limitat în care se aplică metodele statistice: teorie a statisticii sociale, teoria statisticii economice; expune obișnuit metodele statisticii descriptive adaptate la domeniul respectiv; 4. termen folosit pentru a marca diferența față de statistica aplicată.

BIBLIOGRAFIE

1. BĂRBAT, Al., *Teoria statisticii sociale*, Ed. didactică și pedagogică, București, 1972.
2. BIJI, M., BIJI E., *Statistica teoretică*, Ed. didactică și pedagogică, București, 1979.
3. MOOD, A. M., GRAYBILL, Fr. A, *Introduction à la statistique théorique*, Dunod, Paris, 1973.
4. YULE, G. U., KENDALL, M. G., *Introducere în teoria statisticii*. Ed. științifică, București, 1969.

STATISTICĂ DE RAMURĂ. Parte a statisticii social-economice, reprezentînd aplicarea metodelor statistice la studiul unei ramuri a economiei naționale: statistica industriei, statistica agriculturii, statistica comerțului etc. Fiecare **s.r.** se bazează, pe de o parte, pe principiile unitare ale teoriei statisticii, iar pe de altă parte, pe tezele științelor speciale care studiază esența fenomenului supus cercetării. În cadrul **s.r.** se aplică și se diferențiază procedeele statistice și indicatorii statistici în funcție de specificul obiectului de studiu. Pentru a pune în evidență procesele, interdependențele unei ramuri, **s.r.** poate folosi și alte metode cantitative decît cele strict statistice.

BIBLIOGRAFIE

1. BARON, T., KORKA, M., PECICAN, E., STĂNESCU, M., *Statistica*, Ed. didactică și pedagogică, București, 1981.
2. TÖVISSI, L., ISAIC-MANIU, Al., ANTONESCU, C., *Statistica*, Curs pentru Facultatea de Economie a industriei, construcțiilor, transporturilor (text xerox), A.S.E., București, 1981.

V.TREBICI

STATISTICĂ ACTUARIALĂ. Are ca obiect fenomenele ce pot constitui riscuri asigurate : incendiul, cutremurul, grindina, inundația, epizootiile (asigurări de bunuri) sau moartea asiguratului, accidentele (asigurări de persoane) urmînd să stabilească frecvența, regularitatea și întinderea pagubelor ce le pot produce în vederea analizării posibilităților de cuprindere a acestor fenomene în sfera asigurărilor, ca riscuri asigurate, a stabilirii cuantumului primelor necesare acoperirii acestor riscuri și determinării mărimii sumelor asigurate.

În **s.a.** se folosesc teoria probabilităților și statistica matematică, se efectuează calcule actuariale, se întocmesc tabele actuariale de mortalitate etc. **S.a.** cuprinde : funcțiile biometrice, construirea tabelelor de mortalitate, studiul rentelor viagere, al rezervei matematice, al asigurărilor de diferite tipuri etc. Din punct de vedere istoric, **s.a.** și, mai general, matematica actuarială a evoluat în strînsă legătură cu teoria probabilităților și demografia.

În România, **s.a.** s-a dezvoltat grație lucrărilor unor matematicieni și statisticieni ca : M. Saniclevici, M. Praporgescu, *Gh. Mihoc* * (1906—1981), legate de nevoile asigurărilor de viață.

BIBLIOGRAFIE

1. MIHOC, Gh., *Tratat de matematici actuariale*, Biblioteca Institutului de statistică, actuariat și calcul al Universității din București, București, 1943.
2. MIHOC Gh., IONESCU, H., *Curs de matematici pentru statisticieni și economiști* (§ 188. Funcții biometrice continue), București, 1956.
3. MIHOC, Gh., CRAIU, I., IONESCU, H., *Teoria matematică în operațiunile financiare*, vol. I și II, Ed. științifică, București, 1960.

STATISTICĂ OFICIALĂ. Instituții de statistică ale statului, împreună cu activitatea desfășurată de acestea în domeniul culegerii, prelucrării și publicării datelor statistice. Instituțiile respective poartă denumirea de oficii centrale de statistică, birouri naționale de statistică, institute centrale de statistică, direcții centrale de statistică, existînd astăzi în majoritatea covîrșitoare a țărilor. Printre atribuțiile acestor instituții figurează efectuarea de recensăminte și anchete statistice, înregistrarea prețurilor, întocmirea conturilor naționale, publicarea de anuare statistice.

În cadrul cooperării internaționale, instituțiile naționale de statistică furnizează date Organizației Națiunilor Unite (Biroul de Statistică, Divizia de Populație etc.) și organismelor din cadrul O.N.U. (F.A.O., O.M.S., UNESCO, B.I.T. etc.) care publică anuare internaționale, studii etc. În țările dezvoltate, oficiile naționale de statistică au fost organizate în cursul sec. al XIX-lea. Un rol important în această privință l-au avut Congresele internaționale de statistică (primul — 1853) și *Institutul Internațional de Statistică* *. În perioada postbelică, rolul hotărîtor îl au Comisia de Statistică a O.N.U. și Biroul de Statistică O.N.U., precum și Comisiile economice regionale. În cadrul Comisiei Economice O.N.U. pentru Europa funcționează Conferința statisticienilor europeni.

În unele țări se face distincția între *statistica generală a statului*, reprezentată de Oficiul național de statistică și *statistica departamentală* (→ și *Sistem informațional statistic*).

V.TREBICI

II. CATEGORIILE ŞIINŢEI STATISTICE

FENOMEN DE MASĂ (gr. *phainomenon* — ceea ce apare). În raportul dialectic dintre esenţă şi fenomen, acesta din urmă exprimă trăsăturile individuale, exterioare, întîmplătoare şi vremelnice ale obiectelor, trăsături prin care domeniile realităţii se manifestă nemijlocit şi multiplu. Ca noţiune generală, fenomenul reprezintă orice proces din natură şi societate (procese fizice, chimice, biologice, economice, sociale). Elementul particular al **f.m.** este conferit de faptul că el se constituie sub acţiunea colectivă şi repetată a unui număr mare de factori întîmplători şi că deci pentru punerea în evidenţă a esenţei este nevoie de un număr mare de cazuri. Formarea preţurilor pe piaţa neorganizată presupune un număr mare de producători care îşi confruntă cantităţile de muncă individuală; formarea unui comportament demografic în sînul unei populaţii presupune, pe de o parte, un număr mare de persoane, iar pe de altă parte, un număr mare de factori sociali, culturali, educaţionali, psihologici, din a căror influenţă şi combinaţii se constituie o anumită atitudine faţă de familie, faţă de numărul de copii etc. Denumiri sinonime: „fenomen statistic", „fenomen colectiv". Un termen mai vechi „fenomen atipic", mai puţin uzitat, exprimă sugestiv trăsăturile unui **f.m.**, fenomen care nu se poate reproduce în mod identic aproape niciodată.

F.m. ne duce la problema raportului dintre necesitate şi întîmplare, dintre *legea statistică* * şi legea dinamică, dintre *modelul stohastic* * şi cel determinist, cu largi interpretări şi implicaţii atît pe plan filozofic cît şi în domeniul statistic. **F.m.** se întîlnesc aproape în toate domeniile ştiinţei şi ale activităţii sociale. Sub aspectele cele mai complexe ele apar însă în domeniul social. În cadrul fenomenelor sociale se ivesc forme de relaţii cauzale deosebit de complexe: afară de cauzele esenţiale şi neesenţiale intervin lanţurile de evenimente cauzale, în care se integrează, ca o verigă, scopurile umane — personale şi sociale. Fenomenalitatea neobişnuit de bogată a vieţii sociale ridică cele mai mari dificultăţi în stabilirea schemelor fenomene — cauze supuse analizei statistice.

LEGE STATISTICĂ (stohastică). Lege care guvernează comportamentul fenomenelor de masă, descriind *relaţii cauzale statistice* *. Particularitatea esenţială a **l.s.** constă în faptul că ea este stabilită şi nu valorează decît statistic, adică aplicată numai unei mase de evenimente cu structură completă şi nicidecum cazurilor particulare. O astfel de lege se realizează ca o tendinţă predominantă, ca o necesitate care îşi croieşte drum printr-un număr foarte mare de întîmplări şi care se manifestă în aceste întîmplări ca medie a unui număr mare de abateri întîmplătoare.

L.s. este una din formele de manifestare a conexiunii universale a fenomenelor din natură şi societate. Noţiunea opusă şi complementară a **l.s.** este *legea dinamică* (deterministă) a cărei definiţie este următoarea: valorilor strict determinate ale unor factori le corespund întotdeauna valori strict determinate ale mărimilor dependente de aceşti factori.

În cazul legii dinamice, raporturile cantitative dintre mărimi rămîn adevărate pentru fiecare caz individual, pentru fiecare element al colectivităţii supus acţiunii legii respective. Un exemplu este legea atracţiei universale. În timp ce legea dinamică descrie comportamentul riguros al unor sisteme sumative şi închise şi deci prevederea riguroasă a stărilor ulterioare ale unor asemenea sisteme, **l.s.** exprimă comportamentul sistemelor complexe, integrative şi deschise.

Introducerea **l.s.** în metodologia științei a fost stimulată în mod deosebit de cercetările în domeniile termodinamicii, mecanicii fluidelor și a gazelor, domenii în care nu se mai putea opera cu legi dinamice, specifice mecanicii clasice. Este relevantă, în acest sens, „mișcarea browniană", în care, pentru a se putea caracteriza comportamentul unui gaz închis într-un recipient, la o anumită presiune și temperatură, nu se poate proceda la studierea mișcării fiecărei molecule în parte, ci se procedează la o evaluare a mediei de comportament al tuturor moleculelor gazului.

Fizica contemporană se folosește aproape exclusiv de metodele statistice, formulînd legile ei pentru un număr mare de sisteme. Media sau rezultanta statistică a comportamentului unui ansamblu de elemente are caracter de lege, deoarece este stabilă și repetabilă. Astfel, dacă se consultă statisticile demografice, se constată „stabilitatea statistică" a raportului de masculinitate la naștere (în țara noastră în jur de 106 băieți la 100 fete); talia adulților, în diferite populații, se distribuie după „curbe în clopot" mai mult sau mai puțin aplatizate (boltite) dar de aceeași alură generală; studiul unor fenomene economice relevă cicluri de variație adesea foarte regulate etc. **L.s.** constituie modelul *teoretic* aplicabil la observații; pentru a o surprinde, demersul statistic operează probabilist, de unde și denumirea de *lege de probabilitate* dată **l.s.**

Legile statistice sînt legi cu statut propriu, tot așa de obiective ca și legile dinamice; ele sînt mai cuprinzătoare, corespunzînd înțelegerii actuale și statutului contemporan al științei.

RELAȚII CAUZALE STATISTICE (*legături statistice*). Relații prin care se realizează mecanismul procesului de determinare și dezvoltare a fenomenelor de masă. Spre deosebire de relațiile cauzale mecanice (descrise de legi dinamice), **r.c.s.** (descrise de legi statistice) exprimă relații suple, neunivoce, relații în care cauzele interacționează cu factorii aleatori, cu un cerc larg de condiții, astfel încît trecerea de la cauză la efect este mediată de jocul întîmplărilor, cuprins într-un *model statistic* * (stohastic).

Fiind vorba de raportul dintre cauză și efect, de succesiunea genetică a fenomenelor în funcție de ansamblul condițiilor care însoțesc dinamica lor obiectivă, legea cauzalității se manifestă diferit în funcție de varietatea formelor de mișcare și a nivelurilor de organizare a materiei. Dacă mecanica clasică se concretizează prin relații cauzale directe și univoce, știința contemporană a pus în evidență relații cauzale indirecte și neunivoce de tipul unor înlănțuiri neliniare în care trecerea de la cauză la efect se realizează statistic. În structurarea acestor relații cauzale neliniare care au loc în *procesele aleatoare stohastice* *, precum și în cercetarea mecanismului lor de funcționare intervine *probabilitatea** . (→ și *Corelație* *.)

Istoric. Legătura de tip statistic sau stohastic a fost dezvăluită de școala biometrică Galton-Pearson, la sfîrșitul sec. al XIX-lea, sub forma „variației legate" a variabilelor observate în colectivități statistice. Utilizarea conceptului de cauzalitate statistică (în formularea „. . . *case of statistical causation* . . .") se întîlnește în anul 1909, în memoriul lui G. U. Yule privind aplicațiile metodei corelației în statistica socială și economică, prezentat la sesiunea a XII-a a Institutului Internațional de Statistică. De atunci, acest concept s-a înscris în sfera disputelor teoretice purtate în jurul problemei determinismului și cauzalității în știință.

Elucidarea esenței cauzalității statistice prezintă un interes teoretic și metodologic major pentru cercetarea statistică a legăturilor dintre fenomenele de masă. Într-o reprezentare dialectică a cauzalității, între legătura statistică și legătura

funcțională (rigidă) există doar o deosebire relativă, care apare în modalitatea de manifestare a cauzalității — neunivoc sau univoc. În fond, legăturile rigide sînt „cazuri limită" ale legăturilor statistice, un „model ideal, abstract al cauzalității concrete" și se pot identifica numai în sistemele în care practic (și nu principial) factorii aleatori nu influențează calitativ starea lor viitoare. (IX. *Regresie și Corelație.)*

LEGEA NUMERELOR MARI. Teoremă a statisticii matematice potrivit căreia frecvența unui eveniment converge în probabilitate către probabilitatea acelui eveniment.

Într-o formă mai generală, **l.n.m.** caracterizează legătura dintre numărul evenimentelor și gradul de acțiune a legii generale, cu respectarea cîtorva condiții simple.

Pentru ilustrarea **l.n.m.** se obișnuiește a se utiliza experiența constînd din aruncarea unui ban sau a unui zar, din scoaterea unei bile dintr-o urnă etc. Așadar, dacă într-o probă (extragerea unei bile dintr-o urnă care conține bile albe și bile negre) probabilitatea unui eveniment (apariția unei bile albe) este p și dacă se repetă proba de un număr mare de ori în condiții identice (repunerea bilei în urnă după ce s-a notat culoarea sa), raportul între numărul de ori în care se produce evenimentul și numărul total de probe — adică frecvența relativă f a evenimentului — tinde să se apropie din ce în ce mai mult de probabilitatea p. Mai precis, dacă numărul de probe este suficient de mare, devine absolut improbabil ca abaterea între f și p să depășească o valoare oarecare, oricît de mică ar fi ea, dată cu anticipație.

Primul care a enunțat **l.n.m.** a fost J. Bernoulli în opera sa „Ars conjectandi", apărută în 1713 :

Dacă notăm cu ν_n numărul apariților unui eveniment A în n experiențe independente și cu $p = P(A)$ — probabilitatea evenimentului A, atunci

$$\lim_{n\to\infty} P\left(\left|\frac{\nu_n}{n} - p\right| < \varepsilon\right) = 1,$$

oricare ar fi $\varepsilon > 0$.

Există și alte variante ; ele stabilesc condițiile în care un șir de variabile aleatoare $X_1, X_2, \ldots, X_n$ independente satisfac relația

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \ldots + X_n}{n} - \frac{M[X_1] + \ldots + M[X_n]}{n}\right| < \varepsilon\right) = 1$$

oricare ar fi $\varepsilon > 0$.

L.n.m. are o utilizare curentă, întrucît ea permite să se aprecieze probabilitatea unui eveniment doar prin referire la numărul de cazuri în care s-a produs, o atare apreciere fiind cu atît mai bună cu cît numărul observațiilor este mai mare și cu atît mai utilă cu cît nu există în general nici un alt mijloc de a calcula această probabilitate.

Cei dintîi statisticieni, care au constatat regularități în fenomenele sociale cercetate, s-au referit statornic la principiul numerelor mari. J. P. Süssmilch, de exemplu, a scris că *ordinea* în frecvența deceselor, nașterilor, căsătoriilor este vizibilă în numărul mare de cazuri : „o proprietate a acestei ordini constă în aceea că se ascunde în dezordinea numărului mic de cazuri și poate fi desco-

perită numai dintr-un mare număr de registre despre multe localităţi şi pentru mulţi ani". A. Quételet, căruia după expresia plastică a unui critic „îi tremura mîna de emoţia bucuriei" consemnînd rezultatele investigaţiilor sale statistice, a condiţionat „cunoaşterea legilor generale" de reunirea observaţiilor „în destul de mare număr", pentru a elimina din ele „tot ceea ce nu este decît întîmplător şi individual".

În virtutea **l.n.m.**, acţiunea combinată a unui mare număr de factori întîmplători duce, în anumite condiţii foarte generale, la rezultate aproape independente de întîmplare, dînd deci posibilitate necesităţii să se realizeze. În numărul mare de observaţii, abaterile individuale (pozitive şi negative) de la *tendinţa centrală* * se pot compensa. Aplicarea unei legi statistice unei populaţii ce nu prezintă toate variaţiile de cazuri în proporţia lor normală este tot aşa de absurdă ca şi tentativa de a prevedea momentul morţii unui individ cu ajutorul unei tabele de mortalitate. Ca exemplu, în acest sens, să facem apel la una din primele legi statistice descoperite : la naştere, numărul băieţilor întrece pe cel al fetelor ; raportul iniţial admis era de B 107 : F 100. Căutăm să probăm păstrarea raportului de masculinitate la naştere în zilele noastre. Este în afară de orice îndoială că, dacă ne-am opri asupra distribuţiei născuţilor pe sexe în sînul fiecărei familii în parte, ar fi imposibil să indicăm ce fel de raport tipic există, date fiind fluctuaţiile sale destul de numeroase şi importante pentru diferite familii. O cercetare, singură posibilă şi singură valabilă într-o asemenea situaţie, trebuie întreprinsă pe mase îndeajuns de mari de naşteri. Într-adevăr, observarea cazurilor de naşteri, bunăoară pe cuprinsul ţării noastre, constituie mărturia faptului că an de an se nasc ceva mai mulţi băieţi decît fete ; raportul general este de B 106 : F 100.

Cînd mecanismul aleator, care permite să se atribuie *apriori* o valoare evenimentului, este cunoscut — ca în cazul de mai sus — frecvenţa constatată empiric pe un mare număr de cazuri va fi luată drept măsură a probabilităţii evenimentului.

l..n.m. are o mare însemnătate nu numai ca „teză abstractă a ştiinţei" : ea joacă un rol de seamă şi în activitatea practică a oamenilor. J. Bernoulli a ţinut să menţioneze că procedeul sporirii numărului observaţiilor în scopul de a ridica gradul de siguranţă al judecăţii „nu este nou şi nu este neobişnuit" şi că „chiar omul simplu, după un instinct natural oarecare, el singur şi fără nici un fel de instruire prealabilă (lucru de mirare) ştie că, cu cît mai multe observaţii sînt luate în considerare, cu atît este mai mic pericolul de a nu se ajunge la ţintă". Foarte multe dintre fenomenele care se petrec în jurul nostru par să se producă cu totul întîmplător, dacă ne mărginim la a considera doar cîteva cazuri izolate. Dar, odată considerate la scară foarte mare, aceste fenomene întîmplătoare „se cuminţesc" şi prezintă un comportament din ce în ce mai regulat, cu cît numărul cazurilor sau al experimentelor creşte.

MODEL (STATISTIC). Termenul „model" este lipsit de univocitate şi se utilizează pentru a desemna concepte diferite. Există numeroase definiţii ale m. şi ale condiţiilor pe care trebuie să le îndeplinească acesta. Reproducem definiţia lui S. Marcus [1] : „Fiind dat un obiect A, un alt obiect B este un m. al lui A, dacă următoarele condiţii sînt îndeplinite :

a) există o analogie între B şi A ; cu alte cuvinte, B îndeplineşte o funcţie iconică în raport cu A ;

b) obiectul B poate fi investigat prin cel puţin o metodă care nu e compatibilă cu natura lui A, fapt care impune o anumită eterogenitate a lui B în raport cu A ;

c) există cel puțin o metodă de tipul afirmat la punctul b), în raport cu care investigarea lui B conduce la concluzii netriviale ; cu alte cuvinte, există o deosebire netrivială între elementele de intrare și cele de ieșire în investigarea lui B prin metoda considerată ;
d) concluziile relative la B au o anumită relevanță în raport cu A ;
e) obiectul B are o capacitate de autonomizare și deschidere ; el poate fi conceput independent, poate părăsi funcțiile a), b), c) și d) relativ la A, pentru a dobîndi funcții similare în raport cu un alt obiect A. ...
... Cea mai bună șansă de a satisface cererile c) și d) se obține prin constituirea obiectului B în *contextul unei teorii* (*s.n.*) cit mai avansate, bogată în concepte, metode și rezultate aflate într-un stadiu avansat de elaborare".
Prin urmare, **m.** este o verigă într-un proces epistemologic mai larg care integrează : ipoteza, reprezentarea schematică a unui proces, prin **m.**, testarea acestuia (de obicei statistică) și reluarea procesului, în cadrul unei teorii generale. Toate aceste principii sînt valabile și pentru **m.** (**s.**).
Pentru domeniul care ne preocupă, **m.** reprezintă, în general, transcrierea matematică a unor ipoteze simplificate despre starea sau evoluția unui fenomen sub influența anumitor variabile. Din punctul de vedere al variabilelor care definesc structura **m.**, se disting două categorii.

Model determinist. M. care nu conține elemente aleatoare și care determină în mod cert starea sau evoluția fenomenului studiat în funcție de variabilele luate în considerare. **M.d.** descrie o relație pură între fenomene (o unică și bine determinată valoare a funcției pentru o valoare dată a argumentului).
Model statistic (stohastic). M. în care intervin, pe lîngă variabilele măsurabile sau observabile, una sau mai multe variabile aleatoare, corespunzător efectului posibil al factorilor necontrolați prin variabilele specificate în **m. M.s.** descrie o relație complexă între fenomene (un ansamblu de valori ale funcției pentru o valoare dată a argumentului). Exemplu : **m.** consumului de energie electrică de forma :

$$E_t = a + bP_t + CF_t + U_t.$$

Variabila endogenă E (consumul de energie electrică) depinde de variabilele exogene P (producția industrială) și F (consumul unui produs substituibil), precum și de o perturbare aleatoare U (*eroare la ecuație* în opoziție cu eroarea de observație — *eroare la variabile*). Termenul U însumează, oarecum sub o formă aleatoare, efectul tuturor factorilor nespecificați ; el este considerat aleator, deoarece ignoranța noastră asupra modalităților sale de determinare îi conferă natura unei variabile aleatoare. Ca urmare a relației formulate, mărimea E este aleatoare, ca și variabila U.

Variabilă endogenă. Variabilă ale cărei valori sînt deduse din rezolvarea **m.**

Variabilă exogenă. Variabilă introdusă „din afară", ale cărei valori sînt determinate independent de **m.**
În **m.** statistic de corelație, denumirile clasice ale variabilelor corelate sînt : *variabilă dependentă* și *variabilă independentă*. Variabila cunoscută se numește variabilă *independentă*, iar variabila care poate fi cunoscută cînd se cunoaște variabila independentă (sau variabilele independente) se numește variabilă *dependentă*. Se spune că variabila dependentă este *funcție* de variabila independentă. Se consideră că termenii de *variabilă explicată* și de variabile *explicative* (sau *descriptive*) sînt mai potriviți, acestea din urmă nefiind întotdeauna independente.

Dacă prima se suprapune cu noțiunea de variabilă endogenă, cea de-a doua excede sfera noțiunii de *v.e.*, cuprinzînd și variabila endogenă „întîrziată" sau „avansată" în timp, utilizată pentru explicarea variabilei endogene. De exemplu, pentru a atenua variațiile conjuncturale ale consumului calculat de energie electrică prin **m.** precedent, care ar apărea eventual prea puternice în comparație cu variațiile consumului observat, se va putea introduce în acest scop o variabilă stabilizatoare E_{t-1}, adică :

$$E_t = a + bP_t + cF_t + dE_{t-1} + U_t.$$

E_{t-1} reprezintă consumul de energie electrică din perioada precedentă, traducînd regularitatea obiceiurilor de consum. E_{t-1} este o variabilă endogenă, dar inclusă în **m.** imaginat ca o variabilă explicativă.

În vocabularul modern, se desemnează uneori o ipoteză explicativă prin denumirea de **m.** statistic (**m.** χ^2, model *F*, model *t* etc.). Orice **m.** statistic implică o selecție a variabilelor ce-l constituie și deci o simplificare, de unde preocuparea atentă de a se obține o reprezentare simplificată, dar nu deformată. La construirea **m.** statistic mai ales, cercetătorul este amenințat de două pericole foarte sugestiv caracterizate de cunoscutul matematician R. Bellman : „capcana suprasimplificării și mlaștina supracomplicării".

În fond, **m.** traduce modul în care gîndirea omului concepe fenomenul modelat. Această concepere depinde nu numai de fenomenul în sine — a cărui existență este independentă de felul cum l-a înțeles cercetătorul, dar și de starea generală a cunoștințelor umane. Un același fenomen poate să dea naștere la mai multe **m.**, fiecare din ele făcînd loc altuia, mai conform cu ansamblul cunoștințelor dobîndite. Progresul științei și introducerea tehnicii electronice de calcul conduc la **m.** mai elaborate, care pot să se substituie **m.** pînă acum familiare.

Caracteristic pentru stadiul actual al modelării este preocuparea de a elabora o teorie a coerenței și coeziunii care să împletească punctul de vedere determinist cu cel probabilist.

Modelarea este un proces iterativ care presupune utilizarea unor date experimentale în vederea alegerii unei anumite repartiții statistice și în același timp examinarea validității acestei repartiții. Secvențele ar fi :

1. formularea unei ipoteze ; 2. colectarea materialului experimental ; 3. verificarea ipotezei. După L. Tövissi și V. Vodă [2], secvențele sînt : a) formularea **m.** inițial (clase de repartiții) ; b) culegerea materialului experimental (datele) ; c) alegerea unei repartiții particulare ; d) verificarea gradului de concordanță a repartiției alese cu situația reală ; e) formularea ipotezelor ce explică mecanismele aleatoare ce au generat datele.

BIBLIOGRAFIE

1. MARCUS, S., *De la propoziție la text*, în volumul *Semantică și semiotică* (sub red. acad. I. Coteanu și prof. dr. Lucia Wald), Ed. științifică și enciclopedică, București, 1981.
2. TÖVISSI, L., VODĂ, V., *Metode statistice. Aplicații în producție*, Ed. științifică și enciclopedică, București, 1982.

POPULAȚIE (STATISTICĂ). Mulțime de elemente care aparțin de o aceeași familie și care fac obiectul cercetării statistice. **P.(s)** (sau simplu *populația*) poate fi alcătuită dintr-un ansamblu de persoane (populația României la recensămîntul din 5 ianuarie 1977), obiecte (parcul de tractoare și mașini în agricultura românească la o anumită dată), evenimente (căsătoriile în cursul unei

perioade), idei sau opinii (opiniile consumatorilor despre calitatea unor noi tipuri de bunuri de uz îndelungat puse în circulație) etc. O populație este *finită*, dacă include un număr determinat de elemente, dar ea poate fi considerată drept reprezentativă a unei populații teoretic *infinite*. care acoperă universul conceptual statistic (un lot de piese fabricate în condițiile date este o populație finită ce se poate considera ca reprezentînd populația infinită constituită din mulțimea tuturor pieselor care ar putea să fie produse în aceleași condiții). Este fundamental să se fixeze frontierele **p.(s.)** supuse efectiv observării. De exemplu, populația, obiect al recensămîntului din 1977, a fost delimitată în :

— *conținut* : s-au cuprins cetățenii români domiciliați în țară sau aflați temporar în străinătate la data recensămîntului, precum și cetățenii străini sau persoanele fără cetățenie cu domiciliul permanent în România ;
— *spațiu* : s-au recenzat locuitorii de pe întreg teritoriul țării ;
— *timp* : declarațiile înscrise în formularele de recensămînt au trebuit să se raporteze la starea populației existentă la ora „0" din noaptea de 4 spre 5 ianuarie 1977 (deși perioada de completare a formularelor a fost 5—12 ianuarie 1977).

În general, fără o circumscriere riguroasă a obiectului investigației statistice — a **p.(s.)** — în funcție de scopul investigației, apare pericolul de a se înregistra fenomene din afara populației studiate sau, dimpotrivă, de a se omite unele fenomene ținînd de populația respectivă. Se înțelege că, în ambele situații, s-ar ajunge la o reflectare deformată a mediului observat, iar ulterior, în stadiile următoare ale cercetării, la caracterizări și concluzii eronate.

Termenul de **p.(s.)** este preluat din studiul concret al populației umane, devenind însă un concept abstract, de maximă generalitate. Se precizează că obiectele care intră în populație dispun de unele caracteristici generale și se deosebesc prin alte caracteristici. Populația omogenă este compusă din obiecte în care una sau mai multe caracteristici esențiale sînt comune.

Termeni sinonimi pentru **p.(s.)** : colectivitate statistică, ansamblu statistic, colectiv statistic, a căror folosire nu este recomandabilă.

Colectivitate statistică. Sinonim al *populației statistice* *.

EȘANTION (statistic). Submulțime sau parte dintr-o *populație* (*statistică*) *, constituită după criterii bine stabilite, pentru a putea servi la determinarea caracteristicilor *populației (statistice)**. (→ III. *Anchetă prin sondaj* și XIV. *Teoria sondajului statistic*.)

UNITATE STATISTICĂ (*Individ*). Element al unei populații statistice, asupra căruia se va efectua nemijlocit observarea. În populațiile statistice citate (→ *Populație*), **u.s.** erau respectiv : persoana, gospodăria, tractorul agricol, mașina agricolă, căsătoria, opinia.

U.s. este purtătorul originar de informație sau, altfel spus, subiectul logic al informației statistice. Datorită varietății aspectelor sub care se poate prezenta în fapt, **u.s.** comportă o definiție precisă, care exclude orice posibilitate de interpretare diferită de către observatori și astfel orice eroare ce poate prejudicia valoarea investigației. Definirea unității întrebuințate în statistică se impune chiar și atunci cînd ea ar părea clară pentru oricine. Să considerăm, de pildă, „născutul viu" unitate principală în statistica demografică. La prima vedere s-ar părea că nu mai este nevoie să-l definim clar și distinct ; totuși în literatura domeniului, precum și în limbajul comun identificăm existența unor înțelesuri divergente. De aceea, pentru o recunoaștere fără echivoc, a fost adoptată următoarea definiție a noțiunii de născut viu : „produs al concepției expulzat

V.TREBICI

sau extras complet din corpul mamei, independent de durata gestației, și care, după această separație, prezintă un semn de viață precum : respirația, activitatea cardiacă, pulsațiile cordonului ombilical sau contracția musculară dependentă de voință".

U.s. sînt uneori constituite din mai multe unități simple (**u.s.** *complexe*) ; definirea lor poate să ridice chestiuni din cele mai complicate. Un caz tipic de acest gen îl reprezintă definiția gospodăriei, unitate de observare obișnuită la recensămîntul populației. În definiția recensămîntului din 5 ianuarie 1977, prin gospodărie s-a înțeles : „grupul de două sau mai multe persoane care locuiesc împreună în mod obișnuit, avînd în general legături de rudenie, care participă în totalitate sau parțial la formarea veniturilor și la cheltuirea lor". Definiția a fost însoțită de o serie de alte precizări. În investigațiile aceleiași populații se pot alege **u.s.** diferite și o aceeași **u.s.** poate să fie definită în chip diferit. De exemplu, recensămintele populației, microrecensămintele veniturilor populației, sondajul statistic al bugetelor de familie au drept **u.s.** fie familia, fie gospodăria sau menajul, existînd diferențe în definirea lor de la o țară la alta, uneori de la o înregistrare la alta. Cercetătorul ce-și propune să analizeze manifestările unui fenomen de masă, oglindite în seriile de date statistice, are datoria ca, înainte de toate, să cunoască unitatea întrebuințată la culegerea datelor și modul cum a fost ea definită. Această cunoaștere îl apără de compararea unor date eventual necomparabile, însoțită de aprecieri înșelătoare.

CARACTERISTICĂ STATISTICĂ. Însușire comună unităților unei populații, reținută în tematica investigației statistice și care capătă accepții sau valori diferite de la o unitate la alta sau de la un grup de unități la altul.

VARIABILĂ STATISTICĂ. Caracteristică statistică care are proprietatea de a-și schimba în timp sau spațiu nivelul de dezvoltare sub influența diferiților factori ce acționează asupra populației statistice. Variația unei caracteristici poate fi continuă și discontinuă (discretă).

Variabilă calitativă. Caracteristică ale cărei modalități sînt consemnate, pentru fiecare unitate, sub formă „literală" și nu numerică. Așa s-ar exprima, într-o populație de muncitori, meseria (strungar, frezor, lăcătuș...) ori forma de calificare (școală, cursuri de calificare, ucenicie la locul de muncă...), sexul (masculin, feminin). Caracterul cantitativ al observațiilor rezultă doar din numărarea unităților corespunzătoare fiecărei modalități a caracteristicii respective.

Variabilă cantitativă. Caracteristică măsurabilă ; fiecărei unități statistice îi corespunde un număr care este măsura (valoarea) caracteristicii. Observațiile sînt susceptibile de exprimare *ab initio* prin numere, care se pot ordona și ierarhiza (proprietăți ordinale) și asupra cărora se pot executa operații (proprietăți cardinale). Observatorul va nota, de exemplu, cu referire la muncitori, vechimea în meserie în ani, productivitatea muncii în unități valorice (lei) sau în unități naturale (bucăți, kilograme etc.), retribuția în lei.

V.c. poate fi :

Variabilă discretă. Variabilă care poate lua numai anumite valori și în general valori întregi pe scara lor de variație firească (numărul copiilor pe familie, numărul personalului pe întreprindere, numărul de piese rebutate dintr-un lot de piese etc.).

Variabilă continuă. Variabilă care poate lua orice valori într-un interval finit sau infinit (întinderea unei ferme agricole vegetale, randamentul unei culturi, beneficiul unei întreprinderi, puterea instalată a unui grup electrogen etc.).

Distincția din urmă privește natura intrinsecă a variabilelor numerice ; în fapt, și o variabilă continuă măsurată în mod empiric, pe un anumit număr de unități statistice, apare ca variabilă discontinuă. Intervalul care separă valorile exprimate în ordine discretă poate fi ales arbitrar de observator, în funcție de precizia dorită și de metoda de măsurare utilizată. De exemplu, producția anuală de cereale a diferitelor cooperative agricole de producție, alcătuind o populație statistică, se consideră satisfăcător evaluată în tone, deși precizia de măsurare poate să treacă dincolo de această limită. Dacă totuși „producția" este tratată ca *v.c.*, se are în vedere faptul că apriori orice valoare (pozitivă) poate să reprezinte — într-un anumit interval — o producție. Prin extensiunea noțiunii de *v.c.*, cînd valorile posibile ale unei cantități de natură discontinuă sînt foarte numeroase, aceasta se asimilează unei variabile continue, pentru a ușura calculele.

Observație. Adesea, noțiunea de variabilă este utilizată pentru a numi o caracteristică numerică, rezervîndu-se, în principiu, noțiunea de „caracteristică" (uneori „atribut") observațiilor de ordin calitativ. În controlul statistic al calității produselor, de exemplu, sînt utilizate expresiile „control prin atribute" (corespunzător sau necorespunzător) și „control prin variabile" (variațiile parametrilor supuși măsurătorii).

VARIABILĂ ALEATOARE. Noțiune de bază în teoria probabilităților, desemnînd variabila ale cărei valori, apriori necunoscute, apar în împrejurări datorate întîmplării, cu probabilități determinate. **V.a.** poate fi considerată drept o variabilă „specială" în sensul că ea comportă nu numai un ansamblu de valori posibile, ci și o funcție bine definită care indică probabilitatea fiecăreia dintre valorile posibile.

Să considerăm, de exemplu, numărul X de puncte obținute la aruncarea simultană a două zaruri. X nu poate lua decît una sau alta din unsprezece valori, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 și nu se va cunoaște valoarea lui X înainte de aruncarea efectivă a zarurilor. Dar, cunoscînd compoziția jocului, se poate dinainte calcula probabilitatea ca să se obțină, aruncînd simultan două zaruri, fie două puncte ($x = 2$), fie trei puncte ($x = 3$), fie patru puncte ($x = 4$) etc. ; fie maximul de puncte ($x = 12$).

Astfel, în ipoteza echiprobabilității evenimentelor elementare :

Valori posibile (x_i)	2	3	4	5	6	7	8	9	10	11	12
Probabilități corespunzătoare (p_i)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

(suma probabilităților este egală cu 1).

Tabelul pune în evidență corespondența între valorile posibile ale **v.a.** și probabilitățile lor.

Exemplul prezentat ilustrează o **v.a.** de tip discret, a cărei repartiție se prezintă sub forma generală :

$$X\begin{pmatrix} x_1, & x_2, & \ldots, & x_n \\ f(x_1), & f(x_2), & \ldots, & f(x_n), \end{pmatrix}$$

în care

$$f(x_i) = P(X = x_i) ; \sum_{i=1}^{n} f(x_i) = f(x_1) + f(x_2) + \ldots + f(x_n) = 1.$$

În numeroase cazuri, **v.a.** este susceptibilă de a lua oricare din valorile situate într-un interval (a, b), a sau b sau împreună putînd fi infiniți. De exemplu,

talia indivizilor dintr-un grup de populație (mai multe sute de mii) ilustrează foarte bine noțiunea de v.a. continuă. A ști care este probabilitatea de a întîlni indivizi măsurînd între 1,65 m și 1,70 m pune problema construirii intervalului. În general, se va adopta valoarea $X = x_i$ pentru a reprezenta toate valorile care aparțin intervalului mic :

$$x_i - \frac{1}{2}\Delta x_i \leqslant x \leqslant x_i + \frac{1}{2}\Delta x_i$$

a cărui amplitudine este aleasă arbitrar.
Se asociază atunci fiecărei valori x_i a lui X probabilitatea p_i cu care această valoare poate fi atinsă, probabilitate care se va raporta la amplitudinea Δx_i a intervalului ce cuprinde toate valorile lui X reprezentate prin $X = x_i$.
Raportul

$$\frac{p_i}{\Delta x_i} = \varphi(x_i)$$

este *densitatea de probabilitate* a variabilei aleatoare X.
În calculele practice, cu utilizarea tabelelor numerice, se va spune că probabilitatea de a obține valoarea x este

$$p = \varphi(x) \cdot \Delta x.$$

REPARTIȚIE STATISTICĂ. Fie o variabilă statistică X care ia valorile $x_1, x_2, \ldots, x_n$, cu frecvențele $f_1, f_2, \ldots, f_n$. Mulțimea ale cărei elemente sînt perechile ordonate (x_i, f_i) constituie repartiția variabilei statistice X :

$$X : \begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ f_1 & f_2 & \ldots & f_n \end{pmatrix}.$$

PARAMETRU STATISTIC. Valoare reprezentativă dedusă dintr-o operație numerică (definită fără ambiguitate) aplicată la o repartiție statistică. O astfel de valoare este rezultatul celei mai mari condensări posibile a informațiilor furnizate de o masă de observații. Se mai numește valoare tipică.
Nivel (al repartiției). Ordin de mărime caracterizat de valoarea centrală a variabilei în jurul căreia gravitează celelalte valori.
Dispersie (a repartiției). Amplitudinea împrăștierii valorilor observate în jurul valorii centrale.
Formă (a repartiției). Îndepărtarea repartiției statistice de la o repartiție teoretică.
P.s. cel mai des utilizați :
a) parametrii de nivel : media, mediana, modul ; b) parametrii de dispersie : varianța, abaterea medie pătratică, coeficientul de variație ; c) parametrii de asimetrie și de boltire (aplatizare) : coeficienții lui Pearson și ai lui Fisher. În analiza statistică sînt desemnate sub numele generic de *parametri* și alte valori sintetice, ca : indicii de concentrare, coeficienții de corelație, coeficienții de contingență etc. Se obișnuiește de asemenea (îndeosebi în literatura anglo-saxonă) ca termenul „parametru" să fie folosit pentru a se indica o valoare

reprezentativă teoretică, în care caz valoarea reprezentativă corespunzătoare, măsurată pe eşantion, este denumită „statistică". Astfel, media de sondaj este o statistică (engl. *statistics*) ce estimează media populaţiei, care este deci parametrul (în engl.: *parameter*).

INDICATOR STATISTIC. Mărime statistică cu ajutorul căreia se caracterizează un fenomen — în cazul cel mai frecvent un fenomen social-economic — sub raportul structurii, interdependenţelor, al modificării sale în timp sau în spaţiu. Expresia numerică este legată de calitatea fenomenului; ca atare, i.s. este expresia numerică a unei categorii riguros definite. De exemplu, indicatorul „venitul naţional al R. S. României, pe anul 1980, de 508,7 miliarde lei" este expresia cantitativă a categoriei economice de venit naţional. Din punctul de vedere al expresiei numerice, i.s. poate fi o mărime absolută (volum, efectiv, total) şi o mărime derivată (medie, indice, coeficient etc.). I.s. este rezultatul unei observări şi prelucrări statistice.

PARTEA A DOUA

SISTEMUL INFORMAȚIONAL STATISTIC

Cercetarea statistică, ca oricare altă cercetare științifică se realizează în trei etape distincte, legate între ele și intercondiționate. Prima etapă, cea a *observării statistice* *, asigură culegerea informației primare de la unitățile statistice ce compun populația; a doua etapă este a *prelucrării datelor statistice** ; ultima este a analizei și interpretării rezultatelor obținute, în vederea formulării concluziilor referitoare la populația investigată. Acest demers este clasic în cercetarea statistică a fenomenelor social-economice, avînd o îndelungată tradiție. În linii generale, acest demers este valabil și pentru alte fenomene sau domenii, cum ar fi *planificarea experimentelor* *, cu particularități determinate de natura acestui procedeu. Volumul mare al populației, numărul de caracteristici, complexitatea acestora, cheltuielile importante angajate impun o organizare corespunzătoare a cercetării statistice, realizată astăzi cu ajutorul unor metode moderne ca : teoria grafurilor, metoda drumului critic, metoda analizei cost-beneficiu etc.

Partea a doua întrunește conceptele și metodele legate de culegerea datelor statistice, prelucrarea lor și prezentarea acestora în formă convenabilă pentru analiză statistică și cea de conținut.

III. CULEGEREA DATELOR STATISTICE

OBSERVARE STATISTICĂ *(înregistrare statistică)*. Acțiunea de culegere a informațiilor referitoare la unitățile statistice și caracteristicilor acestora, după criterii riguros definite. **O. s.** se realizează pe bază de plan în cadrul căruia sînt specificate problemele legate de obiectul observării, scopul acesteia, problemele organizatorice, programul, timpul și locul observării etc. Datele înregistrate în cadrul **o.s.** trebuie să respecte două condiții : să fie în volum complet și să aibă calitatea corespunzătoare. Nerespectarea primei condiții are ca rezultat omisiunea unităților (subînregistrarea), iar a celei de a doua condiții se materializează în *erori de observare* *. Verificarea datelor observării reprezintă o fază de cea mai mare importanță a **o.s.**

Observare totală *(exhaustivă)*. Înregistrarea tuturor unităților, fără excepție, care constituie populația finită, supusă cercetării. Exemplul clasic este *recensămîntul populației* * sau sistemul dărilor de seamă statistice, în industrie, de pildă, care cuprinde toate întreprinderile.

Observare parțială. Înregistrarea unui număr de unități din populația finită, alese după criterii bine definite. Exemplul tipic îl reprezintă *ancheta prin sondaj* *.

Observare curentă. Înregistrarea sistematică a unităților (și a caracteristicilor acestora), în mod permanent, pe măsura producerii fenomenelor respective.

Exemplu: înregistrarea evenimentelor demografice de naşteri, decese, căsătorii.
Observare periodică. Înregistrare efectuată la intervale de timp, bine stabilite, a unităţilor supuse observării (recensămîntul populaţiei şi al locuinţelor, recensămîntul animalelor).
Observare directă. Înregistrare nemijlocită ce se efectuează prin contactul direct al cercetătorului cu unităţile respective.
Observare indirectă. Înregistrare a datelor din surse care au consemnat anterior fenomenul, ca de pildă, înregistrare pe bază de documente.

EROARE STATISTICA. În investigaţia statistică se produc inevitabil erori de înregistrare şi de calcul, care pot fi clasificate convenţional în: a) erori de observare **statistică**; b) erori de prelucrare; c) erori de reprezentativitate şi d) erori de modelare. Cele de observare sînt:
Eroare sistematică. Rezultă din jocul unui factor a cărui acţiune face să devieze variabila studiată întotdeauna în acelaşi sens. Ca surse potenţiale de astfel de erori citărm: încălcarea regulilor de circumscriere a cîmpului de investigaţie, de definire a unităţilor de observare, de alcătuire a chestionarelor, prelevări de eşantioane nereprezentative, procedee de prelucrare inadecvate, nerespectarea premiselor de calcul statistic. Adesea distorsiunile sînt introduse de răspunsurile denaturate ale persoanelor anchetate din motive psihologice, economice sau de altă natură. Bunăoară, persoana anchetată dă alt răspuns decît cel corect, uneori din dorinţa de a face impresie (efect de „vanitate"), alteori socotind că răspunsul dat va face plăcere anchetatorului (efect de „simpatie"). Întrebările care au incidenţă fiscală ori implică interese economice, sociale etc. prezintă propriile lor riscuri de *e.s.* Există încă riscul adiţionării de erori generate de prelucrările succesive ale observaţiilor statistice.
În calculele statistice, erorile se propagă pas cu pas fără nici o garanţie că ele se vor anula. Dacă *e.s.* este cunoscută, atunci este uşor de corectat rezultatul final; dacă ea nu este cunoscută, rezultatul poate fi iremediabil compromis.
Eroare accidentală. Imputabilă micilor cauze fortuite şi se produce în ambele sensuri: atît în sensul exagerării cît şi în sensul diminuării fenomenului. Aceste erori se numesc şi *erori compensatoare*, denumire ce subliniază faptul că cu cît observaţiile sînt mai numeroase, cu atît erorile comise într-un sens au ocazia să compenseze pe cele comise în celălalt sens şi, în consecinţă, valoarea finală (medie, proporţie, dispersie) nu va fi sensibil afectată.
E.a. sînt inerente şi inevitabile. Nu există estimaţie numerică a unei mărimi (care nu rezultă dintr-o simplă numărătoare directă) fără eroare. Să presupunem, de exemplu, că vrem să măsurăm cu precizie — în cadrul controlului statistic al calităţii — lungimea unui produs. Efectuînd măsurători repetate, vom ajunge la o diversitate de rezultate, apropiate între ele, dar nu identice. Atunci avem obiceiul să considerăm valoarea medie ca „adevărată" lungime a produsului şi divergenţele valorilor observate ca eroare accidentală. Asupra rezultatului operaţiei de măsurare avute în vedere acţionează un complex de factori, care în mare parte scapă posibilităţii de a fi stăpîniţi şi provoacă diferenţele constate între măsurători. Măsurarea este afectată de perturbaţii datorate fie agenţilor exteriori, fie instrumentelor folosite, fie însuşi observatorului. Desigur, dacă — în cazul considerat — am măsura lungimea produsului cu ajutorul unui instrument defect, eroarea s-ar repeta de fiecare dată în acelaşi sens, adică ne-am afla în prezenţa unei erori sistematice.
Aşadar, o operaţie statistică antrenează erori, cum se întîmplă în orice măsurare. Ceea ce interesează este ordinul de mărime al erorii, ca să se ştie între ce

limite de acceptabilitate se găseşte afirmaţia unei investigaţii statistice. Stabilirea erorii de estimaţie probabilă are valoare de regulă deontologică pentru statistică şi statistician. „O eroare de 30% este mai puţin gravă, mai ales că nu e o raritate în unele domenii, decît neprecizarea erorii, care aduce cu ea neîncrederea totală ... După cum medicul trebuie să distingă eventualele erori ale termometrului sau ale unui tratament de abaterile de la normalitate pe care le prezintă o maladie, tot aşa statisticianul trebuie să distingă, estimînd cît mai exact erorile de înregistrare şi de calcul statistic, de normalitatea sau anormalitatea proceselor studiate" (Onicescu).
Teoria erorilor, elaborată de Gauss şi dezvoltată considerabil ulterior, reprezintă un capitol important al statisticii matematice.

RECENSĂMÎNT. Observare statistică totală, cu caracter periodic, de obicei la nivel naţional, după criterii bine stabilite.
Pentru efectuarea sa, este necesară elaborarea unui plan cuprinzînd toate etapele, de la pregătirea r. pînă la publicarea rezultatelor. Principiile generale ale r. sînt : a) pregătirea minuţioasă a lucrărilor în cadrul cărora o atenţie deosebită se acordă pregătirii cadrelor; b) testarea r. proiectat; c) efectuarea r. simultan pe întregul teritoriu al ţării; d) alegerea perioadei optime de efectuare a r.; e) repetarea r. la intervale egale de timp, în vederea asigurării comparabilităţii. Tipuri de r. : al populaţiei, al locuinţelor, al agriculturii, al animalelor. Cel mai frecvent practicat în prezent este r. *populaţiei* *; celelalte r. se efectuează mult mai rar.

RECENSĂMÎNT AL POPULAŢIEI. Observare statistică totală a populaţiei de pe un teritoriu delimitat (naţional), cu respectarea unor principii şi criterii. După O.N.U., „un r.p. se defineşte ca un proces de culegere, prelucrare şi publicare a datelor demografice, economice şi sociale la un timp specificat şi valabile pentru toate persoanele din ţara respectivă sau de pe un teritoriu delimitat" [1].
Caracteristicile unui r.p. sînt : [2] a) este o operaţie iniţiată de stat; b) se efectuează pe un teritoriu bine determinat; c) se înregistrează toate persoanele, fără excepţie (universalitate); d) înregistrarea are loc la acelaşi moment de timp (simultaneitate); e) unitatea de înregistrare este persoana ale cărei caracteristici se consemnează amănunţit. Caracteristicile ce se înregistrează sînt demografice, sociale, economice, culturale. La recensămîntul populaţiei şi locuinţelor din 5 ianuarie 1977 s-au înregistrat : 1. numele şi prenumele; 2. situaţia faţă de capul gospodăriei; 3. situaţia la recensămînt (prezentă, temporar absentă, plecată pentru o perioadă îndelungată; 4. locul naşterii; 5. anul stabilirii în localitate; 6. ultima reşedinţă avută; 7. domiciliul legal; 8. sexul; 9. data naşterii; 10. starea civilă; 11. pentru femeile născute înainte de 1962 : numărul copiilor născuţi şi numărul de copii în viaţă; 12. cetăţenia; 13. naţionalitatea; 14. felul şcolii absolvite; 15. profesia; 16. felul şcolii pe care o urmează; 17. limba maternă; 18. sursa principală de existenţă; 19. locul de muncă; 20. sectorul social-economic al locului de muncă; 21. ocupaţia la locul de muncă. Prelucrarea se face în mod detaliat, asigurîndu-se la echipamente electronice. *(→ Prelucrarea automată a datelor recensămîntului populaţiei şi al locuinţelor.)*
Datele unui r.p. sînt folosite şi ca *bază de sondaj* * pentru diferite anchete. În ultimele 3—4 decenii, un r.p. este combinat cu metoda sondajului : a) pentru *recensămîntul de probă* *; b) pentru înregistrarea unor caracteristici numai în cadrul unui eşantion; c) pentru verificarea datelor obţinute prin înregistrare; d) pentru controlul datelor; e) pentru obţinerea operativă a datelor. Se practică combi-

narea unui r.p. cu cel al locuințelor, al agriculturii sau al întreprinderilor economice.

Istoric. Odată cu apariția statului au început să se practice, într-o formă rudimentară, numărători ale populației. În antichitate se cunosc recensămintele romane *(census)*, efectuate cu regularitate. În Europa, era recensămintelor moderne începe cu cel din Suedia (1749), din S.U.A. (1790), Anglia (1801), Franța (1801). Problemele metodologice și organizatorice ale r.p. au făcut obiectul congreselor internaționale de statistică (1853—1872); după cel de-al doilea război mondial, Biroul de Statistică al O.N.U. și Divizia de Populație a O.N.U. s-au preocupat de perfecționarea metodologică și organizatorică a r.p. În prezent, practic nu există țară în care să nu se fi efectuat cel puțin un r.p.

În România, r.p., în mod autonom sau combinate cu alte recensăminte, s-au efectuat în anii : 1838, 1859, 1899, 1912, 1930, 1941, 1948, 1956, 1966 și 1977.

BIBLIOGRAFIE

1. O.N.U. *Principes et recommandations concernant les recensements de population de 1970* („Etudes statistiques", Série M. nr. 44, 1967).
2. V. TREBICI, *Demografia*, Ed. științifică și enciclopedică, București, 1979.

RECENSĂMÎNT ECONOMIC. Observare statistică totală, de tipul recensământului, avînd ca obiect întreprinderile industriale, comerciale, de transport, agricultură, pe un teritoriu bine delimitat și cu respectarea unor principii și criterii.

În România, ultimul recensămînt general al întreprinderilor industriale, comerciale, de transport etc. s-a efectuat la 15 noiembrie 1948. În prezent, se mai practică recensămîntul animalelor.

RECENSĂMÎNT EXPERIMENTAL *(de probă)*. Aplicație a metodei sondajului pentru testarea unui recensămînt al populației în ceea ce privește metodologia și organizarea. În acest scop, se constituie un eșantion redus, compus din orașe și comune, în care se procedează la efectuarea recensămîntului; constatările servesc la definitivarea metodologiei și organizării recensămîntului populației.

MICRORECENSĂMÎNT. Înregistrare prin sondaj a populației, în vederea culegerii datelor privind populația, unele caracteristici demografice, veniturile și cheltuielile, locuințele etc. **M.** a fost inițiat și se practică în R. F. Germania.

ANCHETĂ STATISTICĂ. Observare statistică parțială, efectuată în domenii ca : nivelul de trai, cererea de mărfuri, opiniile populației etc. **A.s.** cunoaște o mare dezvoltare în ultimele decenii. În majoritatea cazurilor prin a.s. se înțelege *ancheta prin sondaj* *.

ANCHETĂ PRIN SONDAJ *(statistic)*. Investigație întreprinsă pe o fracțiune reprezentativă a populației statistice, numită *eșantion* *. În virtutea caracterului reprezentativ al eșantionului, informațiile obținute prin sondaj sînt extinse la întreaga populație. De exemplu, pentru cunoașterea nivelului de trai al populației unei țări, se recurge obișnuit la ancheta bugetelor de familie. Două soluții pot fi considerate : fie să se realizeze o anchetă exhaustivă și deci să se supună investigației toate familiile din țara respectivă, fie să se procedeze prin sondaj, anchetîndu-se un număr limitat de familii. Rațiuni lesne de înțeles fac ca, în practica investigațiilor statistice, să se adopte cea de-a doua soluție. În țara noastră, eșantionul actual include 11 000 de familii din cele aproximativ 6,8 milioane de familii. În genere, se socotește că un sondaj cu o rată mai mică,

anume o familie din o mie, permite să se estimeze, cu suficientă precizie, caracteristicile nivelului de trai al *populației de referință* *.
Circumscris numai la o parte din populație, sondajul apare infinit mai suplu, mai economic, mai operativ și astfel mai avantajos decît investigația exhaustivă. Grație dimensiunii sale, **a.s.** permite adesea ridicarea calității informației statistice, deoarece este posibilă organizarea mai temeinică a observării, folosirea unui personal de anchetă mai bine pregătit, aplicarea mai multor verificări și teste de exactitate și implicit reducerea la minimum a erorilor de observare. În plus, un eșantion îngăduie observarea simultană a unui număr mare de caracteristici și deci aprofundarea cunoașterii masei de fenomene. În sfîrșit, metoda sondajelor se impune nu numai pentru rațiuni de cost, rapiditate, suplețe, ci și de fiecare dată cînd măsurarea alterează unitatea statistică (control de fabricație). Dacă se cercetează, de exemplu, punctul de rezistență la șoc a unui lot de becuri electrice, evident nu se va putea avea informația completă decît cu prețul distrugerii întregului lot.
Informația culeasă pe un eșantion este o informație parțială în raport cu ceea ce o conține populația din care s-a extras eșantionul. Prin aplicarea sondajului se renunță la inducția completă și se face apel la inducția incompletă, adică se caută să se tragă concluzii general valabile dintr-un număr restrîns de cazuri. Inducția incompletă, folosită dintotdeauna în viața cotidiană ca și în știință, stă la baza „metodei reprezentative", care și-a cucerit dreptul de cetate în statistică la începutul secolului nostru.
ANCHETĂ PRIN SONDAJ ALEATOR. Anchetă al cărei eșantion este prelevat și tratat după principiile *sondajului aleator (probabilist)**.
ANCHETĂ PRIN SONDAJ NEALEATOR. Anchetă efectuată cu metoda *sondajului prin alegere rațională* * și a *sondajului prin metoda cotelor* *.
ANCHETĂ (SONDAJ) DE OPINIE. Anchetă parțială, efectuată obișnuit cu ajutorul metodei sondajului, avînd drept scop cunoașterea părerilor persoanelor în diferite probleme. Opiniile individuale, subiective și atitudinile izolate prezintă o semnificație ce poate fi înfățișată obiectiv, cuantificată și măsurată. Folosită pe scară largă în anchetele sociologice, demografice, psihosociale, de marketing. Acest tip de anchetă presupune mai multe etape: a) delimitarea obiectului; b) preancheta; c) determinarea obiectivelor și formularea explicită a ipotezelor cercetării; d) universul anchetei; e) alcătuirea eșantionului; f) alegerea tehnicilor de cercetare; g) pretestarea instrumentelor de cercetare; h) definitivarea acestora; i) aplicarea în teren a instrumentelor de cercetare; j) prelucrarea datelor; k) analiza rezultatelor; l) redactarea raportului de cercetare. Organizarea se face uneori cu tehnica grafurilor sau al metodei PERT.

PANEL. Eșantion cu caracter stabil (eșantion fix) compus dintr-un număr de persoane de la care se obțin informații în legătură cu opiniile lor; cercetarea se face după metoda longitudinală, persoanele din eșantion fiind chestionate la diferite momente de timp.

CHESTIONAR. Instrument și tehnică de investigare a opiniilor, motivațiilor, intereselor, atitudinilor, reprezentînd o succesiune logică și psihologică de întrebări scrise sau de imagini grafice cu funcție de stimuli, în raport cu ipotezele cercetării. Întrebările sînt: a) introductive; b) de trecere; c) filtru; d) bifurcate; e) de control.

EXPERIMENT. Observarea și măsurarea efectelor manipulării unor variabile independente asupra variabilelor dependente într-o situație în care acțiunea

altor factori — prezenți efectivi, dar străini studiului — este redusă la minimum. După G. Cuvier: „observatorul ascultă natura, experimentatorul o întreabă și o silește să i se dezvăluie". E. se practică în cele mai variate domenii, inclusiv în cel psihosocial.

(→ XVII. *Planificarea experimentelor.*)

Istoric. Ancheta prin sondaj s-a practicat cu mult înaintea elaborării matematice a sondajului. Începuturile cercetărilor de acest gen se plasează în Anglia, spre sfîrșitul sec. al XVIII-lea, obiectul lor fiind condițiile de viață și, mai particular, „paupertatea". Prima cercetare de acest gen poate fi socotită aceea a lui sir Frederick Morton Eden, publicată în 1797, la Londra, sub titlul: „Starea săracilor sau istoria claselor muncitoare din Anglia ..."; materialul a fost cules de către parohii sătești și cuprindea descrierea a 73 familii muncitorești. În sec. al XIX-lea se continuă asemenea cercetări. Ch. Booth, considerat părintele anchetei sociale științifice, publică 17 volume (1889—1902) din ancheta sa: „Labour and life of the people of London". Un alt pionier este Frédéric Le Play, cu clasicile sale cercetări în mediul muncitoresc: „Les ouvriers européens", 6 vol. Paris, 1855. În 1890, A. Toque și E. Cheyson au prezentat la Sesiunea Institutului Internațional de Statistică comunicarea „Les budgets comparés de cent monographies de famille". Din aceeași perioadă datează și lucrările lui *Ernst Engel* * (1821—1896) consacrate bugetelor de familie. Comunicarea sa, prezentată Institutului Internațional de Statistică (Bull. de l'Institut International de Statistique, Tome VI, Rome, 1892) face mențiunea că a folosit 149 bugete pentru anul 1800, 240 pentru perioada 1801—1850 și 2889 bugete pentru perioada 1851—1890.

Istoria propriu-zisă a metodei sondajului statistic începe cu ultimul deceniu al sec. al XIX-lea, fiind legată de numele lui A. N. Kiaer, statistician norvegian. Acesta a efectuat pe teren cercetări după principiile metodei sondajului, ale cărui rezultate au fost prezentate la sesiunea din Berna (1895) a Institutului Internațional de Statistică. Ideile sale au fost primite cu ostilitate, printre adversari numărîndu-se Georg von Mayr și Luigi Bodio. Urmează o lungă perioadă de discuții, în cadrul Institutului Internațional de Statistică (sesiunile din 1895, 1897, 1901). Sesiunea din 1903 recomandă această metodă, denumită reprezentativă, în următoarea specificare: „cu obligația ca în expunerea rezultatelor să se arate complet în ce condiții s-a făcut alegerea unităților observate". Totuși rezervele au persistat; abia la sesiunea din 1925 a Institutului Internațional de Statistică (Roma), la avizul comisiei special instituite („Commission sur l'application des méthodes représentatives"), compusă din *A. L. Bowley**, *Corrado Gini* *, *A. Jensen*, Lucien March și Fr. Zizek, metoda a fost acceptată și recomandată oficial. Primele teoretizări matematice se datorează lui *A. L. Bowley* *, în lucrarea sa „Elements of statistics" (2 vol., ed. London, 1902, p. 308—315), care a efectuat și o anchetă asupra condițiilor de viață din Anglia. Cercetările prin sondaj s-au diversificat și, paralel, s-a dezvoltat teoria matematică a sondajului.

Aplicațiile au cuprins condițiile de viață, opiniile persoanelor în legătură cu diferite probleme, controlul produselor etc. În 1927, L. L. Thurstone folosește metoda pentru măsurarea atitudinilor (L. L. Thurstone și E. J. Chave, „The Measurement of Attitudes", Chicago, 1929).

O contribuție importantă au avut anchetele de opinie, cele sociologice ca și de psihologie socială, iar ceva mai tîrziu, cele de marketing. În 1935, G. Gallup înființează Institutul American de Opinie Publică, cunoscut sub denumirea

Gallup Poll; în aceeași perioadă, E. Roper efectuează sondaje în cadrul universității Princeton; în 1940 se înființează Biroul de Cercetări ale Opiniei Publice de la Universitatea Princeton (H. Cantril); în 1936 ia ființă Insitutul Britanic de Opinie Publică, în Franța, în 1938, se întemeiază Institutul Francez de Opinie Publică (IFOP), cu revista „Sondages". În 1947, are loc, la Paris, Congresul institutelor europene de opinie publică. Iau ființă sute de institute și de centre de studiere a opiniei publice, grupate în ESOMAR (Consiliul european pentru studierea opiniei publice și a pieței), și WAPOR (Asociația Mondială pentru Studiul Opiniei Publice). În 1949, S. A. Stouffer publică „The American Soldier" (Princeton); anchetele sociologice și psihosociale se dezvoltă considerabil împreună cu metodele statistico-matematice respective (C. P. Lazarsfeld, K. Merton, Katona). Un domeniu predilect de aplicație al sondajului devine și demografia, al cărui început este marcat de ancheta „Indianopolis Study" (1938), continuată apoi cu numeroase alte anchete. Sute de anchete prin sondaj cu caracter demografic se efectuează în întreaga lume; cea mai mare este Ancheta Mondială a Fertilității (W.F.S.), efectuată în ultimii ani și la care s-a afiliat și România.

În România, prima mare anchetă prin sondaj este cercetarea bugetelor de familie, inițiată în 1948, și care se continuă pînă astăzi, cu un eșantion de 11 000 bugete de familie (6500 personal muncitor, 4 000 țărani și 500 pensionari), ancheta prin sondaj a prețurilor de pe piața țărănească, în 95 orașe, cercetarea unui număr de 800 magazine, ancheta unui panel de consumatori compus din 7 000 familii, în 20 orașe și 30 localități rurale. Se efectuează de asemenea diferite anchete de marketing, de sociologie și psihologie socială etc.

BIBLIOGRAFIE

1. U.N.O. *A Short Manual on Sampling*, Volume I, *Elements of Sample Survey Theory* (Studies in Methods, Series F., nr. 9), New York, 1960.
2. O.N.U. *Enquêtes par sondage d'intérêt actuel* (Serie C), Onzième Rapport, nr. 12, 1968.
3. BOUDON., R. LAZARSFELD, P. (editori), *L'analyse empirique de la causalité*. Ed. Mouton, Paris, La Haye, 1969.
4. CHELCEA, S., *Chestionarul în investigația sociologică*, Ed. științifică și enciclopedică, București, 1975.
5. COCHRAN, W. G., *Sampling Techniques*, John Wiley and Sons, New York, 1953.
6. DEMING, W. E., *Some Theory of Sampling*, John Wiley and Sons, New York, 1950.
7. DESABIE, J., *Théorie et pratique des sondages*, Ed. Dunod, 1966.
8. HANSEN, M. H., HURWITZ, W. N., MADOW, W. G., *Sample Survey Methods and Theory*, Vol. I and II, John Wiley and Sons, New York, 1953.
9. KIAER, A., *Die representative Untersuchungsmethode*, Allgemeine Arhiv, Bd. V. Tübingen, 1898.
10. KIAER, A., *Sur les méthodes représentatives typologiques appliquées à la statistique*, Bull. de l'Institut International de Statistique, 11, 1915.
11. MIHOC, Gh., URSEANU, V., *Sondaje și estimații statistice. Teorie și aplicații*. Ed. tehnică, București, 1977.
12. MOSER, C. A., *Metodele de anchetă în investigarea fenomenelor sociale* (trad. rom. după ed. 1963), Ed. științifică, București, 1967.

V.TREBICI

13. MURTHY, M. N., *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta, 1967.
14. STOETZEL, J., GIRARD, A., *Sondajele de opinie publică*, Ed. ştiinţifică şi enciclopedică, Bucureşti, 1975.
15. SUKHATME, P. V., *Sampling Theory of Surveys with applications*, Indian Society of Agricultural Statistics, New Delhi and Jowa State College Press, Ames, Jowa, 1953.
16. TREBICI, V., *Metoda sondajului statistic: istoric, concepte, noţiuni*, în „Revista de Statistică", anul XVIII, nr. 12, 1969.
17. YATES, F., *Sampling Methods for Censuses and Surveys*, Charles Griffin et Co., London, 1949.

IV. PRELUCRAREA DATELOR STATISTICE

Informaţiile obţinute în cadrul *observării statistice* * se referă la *unităţi statistice*. Pentru a deveni utile analizei ele sînt supuse concentrării şi sistematizării. Aceasta reprezintă a doua fază a cercetării statistice — prelucrarea datelor statistice. Trecerea de la unităţi la grupe, apoi la întreaga populaţie sau colectivitate statistică, este însoţită de o pierdere inevitabilă a informaţiei.

Sub raport gnoseologic, prelucrarea datelor statistice presupune o operaţie de proiectare a grupărilor statistice care să ofere un cadru logic în care se sistematizează informaţii după criterii ce ţin de specificul fenomenelor investigate. Din punctul de vedere al tehnicilor şi metodelor de prelucrare, în decursul timpului prelucrarea datelor statistice a cunoscut faza manuală, faza mecanografică, iniţiată în jurul anilor 1890 şi faza electronică sau automatizată, începînd cu circa trei decenii în urmă şi astăzi generalizată în activitatea oficiilor naţionale de statistică din întreaga lume.

Unii termeni cu veche tradiţie în statistică, dar figurînd şi astăzi în manualele de statistică, şi-au pierdut actualitatea; în schimb s-au extins pe scară largă termenii legaţi de prelucrarea automatizată. Din numărul acestora nu au fost reţinuţi decît cei care sînt legaţi nemijlocit de prelucrarea datelor statistice, evitîndu-se cei ce se referă în general la calculatoare electronice, programe etc. şi care figurează în lucrările de specialitate.

PRELUCRAREA DATELOR STATISTICE. Ansamblu de operaţii cuprinzînd verificarea, sistematizarea şi prelucrarea datelor obţinute prin observarea statistică.

Prelucrare primară. Prelucrarea informaţiei statistice primare.

Prelucrare secundară. Prelucrarea informaţiei obţinute în cadrul prelucrării primare.

Prelucrare centralizată. Prelucrare efectuată de obicei la nivel central — de exemplu, în cadrul oficiului central de statistică.

Prelucrare descentralizată. Prelucrare efectuată la nivel local (judeţ, municipiu, întreprindere) sau departamental (minister).

GRUPARE STATISTICĂ. Operaţie de descompunere a unei populaţii statistice sau colectivităţii în subpopulaţii denumite generic grupe, pe baza unor caracteristici (cantitative sau calitative), a căror variaţie justifică asemenea grupe.

Grupare statistică simplă. Grupare după o singură caracteristică (Exemplu : gruparea întreprinderilor industriale după numărul muncitorilor).
Grupare statistică combinată. Grupare după mai multe caracteristici.

Grupare statistică tipologică. Grupare după caracteristici ce pun în evidență tipuri de fenomene *(Clasificări și nomenclatoare)* având un caracter, de cele mai multe ori, oficial. De exemplu, clasificarea ramurilor economiei naționale, clasificarea populației pe categorii sociale etc.

Grupare statistică analitică. Grupare în care se pune în evidență relația reciprocă dintre două sau mai multe caracteristici, din care unele au rol de factor, iar altele sînt caracteristice rezultative.
Există o strînsă legătură între grupări, tabele și serii statistice.

PRELUCRAREA DATELOR STATISTICE CU AJUTORUL MIJLOACELOR TEHNICII DE CALCUL. Aplicarea tehnicilor și metodelor informaticii la prelucrarea datelor statistice, ținînd seama de specificul statisticii. Datorită volumului mare de date ce necesită să fie prelucrate într-un interval de timp cît mai scurt, precum și complexității calculelor statistico-matematice, prelucrarea datelor statistice constituie un domeniu tipic al prelucrării informației cu ajutorul tehnicii de calcul.

În prelucrarea datelor statistice, se utilizează o varietate de mijloace ale tehnicii de calcul, în funcție de cerințele specifice activității statistice luate în considerare. Astfel, la nivelul personalului ocupat cu corectarea și verificarea datelor statistice poate fi suficientă utilizarea de *microcalculatoare electronice de buzunar* sau de *mașini de calculat de birou* (tendința modernă constă în elaborarea de microcalculatoare specializate pentru culegerea și verificarea primară a datelor statistice inclusiv acumularea lor pe o casetă magnetică); la nivelul organelor teritoriale de statistică, a căror activitate este orientată cu prioritate pe prelucrarea datelor statistice operative și curente, pot fi suficiente *mașini electronice de facturat-contabilizat programabile, terminale* conectate cu un calculator electronic central, *minicalculatoare electronice* sau chiar *calculatoare electronice**, în funcție de volumul activității statistice; la nivelul organelor centrale de statistică, este necesară asigurarea capacității de prelucrare necesare asigurării și a prelucrării *anchetelor statistice * și recensămintelor statistice **, precum și a *analizei informației statistice*, ceea ce implică, de regulă, utilizarea de *calculatoare electronice de mare capacitate **. Soluțiile de dotare a sistemului statistic cu tehnică de calcul sînt funcție și de concepția *organizării prelucrării datelor statistice **.

Istoric. Tehnica de calcul a apărut sub impulsul necesității obținerii într-un timp util a datelor recensămîntului populației Statelor Unite ale Americii din anul 1890, prin inventarea și utilizarea echipamentelor de convertire și centralizare a datelor bazate pe utilizarea cartelei perforate, de către dr. Hermann Hollerith, de la U. S. Bureau of the Census. Ele s-au dezvoltat ulterior pentru prelucrarea datelor de gestiune în general sub denumirea de *seturi mecanografice*, principalul producător al acestor echipamente, firma americană IBM (International Business Machines) a devenit în continuare principalul producător de calculatoare pe plan mondial. Organizațiile de statistică naționale au fost printre primele utilizatoare ale *calculatorului electronic* (din anii '50) și au contribuit la introducerea unor metode tehnice specializate (de exemplu, lectura optică a documentelor cu performanțe foarte ridicate — cititorul de documente FOSDIC dezvoltată de U. S. Bureau of the Census pentru recensămîntul popu-

lației din anul 1960). În România, prelucrarea datelor statistice cu ajutorul echipamentelor mecanografice a început în anii '30 și s-a extins considerabil în anii '50. În 1960 a intrat în exploatare la Direcția Centrală de Statistică primul echipament electronic de prelucrare a datelor convertite pe cartele perforate din România, iar în 1967 a intrat în exploatare la Direcția Centrală de Statistică primul calculator electronic cu multiprogramare și memorie externă de mare capacitate din țara noastră. În anii '70, în cadrul sistemului Direcției Centrale de Statistică, au intrat în exploatare teleprelucrarea și teleinterogarea datelor statistice cu ajutorul terminalelor cu tub catodic conectate prin linii telefonice normale (comutabile) cu un calculator electronic al Centrului de calcul al Direcției Centrale de Statistică.

În cadrul Comisiei Permanente pentru Statistică din cadrul C.A.E.R. *funcționează grupa permanentă de lucru pentru automatizarea prelucrării informației statistice*; în cadrul Conferinței Statisticienilor Europeni a Comisiei Economice pentru Europa — O.N.U. funcționează o *grupă de lucru pentru automatizarea prelucrării informației statistice* și un *seminar internațional* (seminarul ISIS — Integrated Statistical Information System), organizat periodic de către Centrul de calcul de cercetări de la Bratislava; sub auspiciile Organizației Națiunilor Unite și ale guvernelor țărilor gazdă au fost înființate *Centrul de calcul de cercetări* (CRC-Bratislava) și *Centrul internațional de instruire* (SZAMOK) în Ungaria; în sfîrșit a fost lansat și un *proiect internațional — CEE PNUD* în domeniul automatizării prelucrării automate a informației statistice SCP — Statistical Computing Project.

BIBLIOGRAFIE

1. COSTAKE, N., *Seminarul ISIS 1982*, în „Revista de Statistică", nr. 9, 1982.
2. CRISTEA, V., DUMITRU, P., *Dicționar de informatică*, Ed. științifică și enciclopedică, București. 1981.
3. MOINEAGU, C., NEGURĂ, I., URSEANU, V., *Statistica*, Ed. științifică și enciclopedică, București, 1976.

ORGANIZAREA PRELUCRĂRII DATELOR STATISTICE. Structurarea resurselor, mijloacelor și procedurilor utilizate în prelucrarea datelor statistice și a relațiilor dintre ele sub forma *fluxurilor informaționale* *.

În cadrul unei economii naționale, prelucrarea datelor statistice are, în general, un caracter difuz în sensul că la nivelul fiecărei unități economico-sociale este necesară cunoașterea și analiza rezultatelor propriei activități, elaborarea unor calcule statistice matematice în scopul organizării și conducerii unor activități ceea ce implică realizarea unui *sistem informațional statistic* *.

Din punctul de vedere al obiectivelor și al sferei de cuprindere, la nivelul unei economii naționale prelucrarea datelor statistice este organizată de regulă astfel:

a) culegerea datelor primare, prelucrarea și elaborarea de rapoarte și formulare statistice la nivelul *unităților economico-sociale de bază*, conform metodologiei elaborate cu colaborarea organelor centrale interesate și difuzate de către organul național de statistică;

b) culegerea și validarea rapoartelor și formularelor statistice eventual și prelucrarea lor primară de către *organele teritoriale de statistică*, transmiterea lor la *organul central de statistică* în vederea validării și prelucrării la nivelul întregii economii, elaborării de rapoarte și analize destinate conducerii superioare, administrației de stat și/sau publicului larg;

c) culegerea, validarea și prelucrarea rapoartelor și formularelor statistice *în cadrul departamentelor*, mai ales referitor la informația statistică specifică departamentului, raportîndu-se după metodologia elaborată și difuzată de către organul central de statistică.

În România, activitatea de culegere, transmitere, prelucrare, memorie și analiză a datelor statistice este organizată la nivelul unităților economico-sociale și la nivelul centralelor, ministerelor și celorlalte organe centrale și la nivelul sistemului Direcției Centrale de Statistică (direcțiile județene de statistică, direcția de statistică a municipiului București, Direcția Centrală de Statistică și Centrul de calcul al Direcției Centrale de Statistică).

În linii mari, Direcția Centrală de Statistică elaborează *sistemul metainformațional statistic* * în colaborare cu Comitetul de Stat al Planificării, Ministerul Finanțelor și celelalte organe centrale și teritoriale interesate, efectuează verificarea informației statistice și asigură realizarea rapoartelor și analizelor cu privire la dezvoltarea economico-socială a economiei naționale, iar organele teritoriale de statistică realizează culegerea și controlul datelor statistice și elaborează rapoartele și analizele dezvoltării economico-sociale pe plan local.

Organizare centralizată *(a prelucrării datelor statistice)*. Prin aceasta se înțelege că prelucrarea automată are loc în principal la nivel central, elaborîndu-se atît rezultatele necesare la nivel teritorial cît și la nivelul ansamblului economiei naționale. În aceste cazuri există o *unitate de informatică statistică* * centrală puternică.

Organizare descentralizată *(a prelucrării datelor statistice)*. Prelucrarea automată are loc în principal la nivel teritorial, la nivel central efectuîndu-se în special centralizări și analize. În aceste cazuri prelucrarea datelor statistice se bazează pe un ansamblu de unități de informatică.

Organizare mixtă (a prelucrării datelor statistice). Combinare a celor două tipuri : centralizată și descentralizată.

În cadrul sistemului Direcției Centrale de Statistică este reprezentată atît prelucrarea automată centralizată — direcțiile județene de statistică dotate cu terminale conectate cu Centrul de Calcul al Direcției Centrale de Statistică — cît și cea mixtă — direcțiile județene prelucrează datele statistice la centrele de calcul teritoriale, din subordinea Institutului Central pentru Conducere și Informatică.

În afară de organizarea prelucrării datelor statistice la nivel național este organizată și prelucrarea datelor statistice la nivel internațional, pe baza unor acorduri internaționale conform cărora organul central de statistică sau un organ departamental autorizat, cu acordul Direcției Centrale de Statistică (pentru datele statistice referitoare la un domeniu limitat), transmit periodic datele convenite. Exemple de prelucrări de date statistice la nivel internațional sînt cele ale C.A.E.R., O.N.U. (inclusiv ale agențiilor specializate), E.M.I., Comunității Economice Europene, care elaborează și publicații statistice internaționale.

BIBLIOGRAFIE

1. COSTAKE, N., *Centrul de calcul electronic și prelucrare mecanografică a informației statistice*, în „Revista de statistică" nr. 4, 1971.
2. COSTAKE, N., *Rolul și locul statisticii în cadrul sistemului informațional economico-social*, în „Revista de Statistică" nr. 12, 1973.

V.TREBICI

3. TREBICI, V., *Demografia*, Ed. ştiinţifică şi enciclopedică, Bucureşti, 1979.
4. *Dicţionar statistic-economic*, Direcţia Centrală de Statistică, Bucureşti, 1969.

UNITATE DE INFORMATICĂ STATISTICĂ. Organizaţie avînd ca obiect al activităţii prelucrarea informaţiei statistice, eventual şi culegerea şi/sau verificarea şi/sau transmiterea informaţiei statistice, eventual şi proiectarea, implementarea şi întreţinerea sistemului informatic statistic sau subsistemelor acestuia.

De regulă, **u.i.s.** dotate cu calculator electronic, care dispun de personal numeros şi realizează şi proiectarea de sisteme informatice, sînt denumite *centre de calcul electronic*; unităţile dotate cu calculator electronic sau minicalculator electronic cu personal puţin numeros cuprinzînd şi analişti sau programatori pentru proiectarea şi/sau implementarea de sisteme informatice sînt denumite *oficii de calcul*; celelalte unităţi de informatică sînt numite *staţii de calcul*. **U.i.s.** se caracterizează prin : ponderc relativ mare a personalului din activitatea de prelucrare automată a datelor (operatori şi controlori de date), ca urmare a volumului mare de date şi de lucrări şi a necesităţii asigurării calităţii rezultatelor, precum şi număr relativ mare al personalului din activitatea de cercetare — proiectare (analişti, programatori, ingineri de sistem), ca urmare a posibilităţii reduse de utilizare a soluţiilor standard de firmă sau curente pentru alte domenii de activitate, ceea ce implică necesitatea elaborării şi dezvoltarea soluţiilor de prelucrare a datelor statistice de către unităţile de informatică. În ţara noastră **u.i.s.** sînt organizate conform decretului 499/1973 care stabileşte normele legale de organizare a unităţilor de informatică din R. S. România. Există două unităţi de informatică statistică : Centrul de calcul al Direcţiei Centrale de Statistică şi staţia de calcul din cadrul Centrului de calcul.

Dată fiind importanţa sistemului informaţional statistic şi tradiţia îndelungată de utilizare a echipamentelor pentru prelucrarea informaţiei, în majoritatea ţărilor membre ale C.A.E.R., organului central de statistică i s-au subordonat şi unităţile de informatică teritoriale (de exemplu, ansamblul centrelor de calcul teritoriale), deşi capacitatea acestora nu este încărcată decît într-o măsură relativ mică pentru prelucrarea datelor statistice.

BIBLIOGRAFIE

Decret cu privire la organizarea unitară a activităţii de informatică (nr. 499, 1973)
COSTAKE, N., *Alegerea sistemelor electronice numerice de prelucrare a informaţiei*, Ed. tehnică, Bucureşti, 1970.

SISTEM INFORMAŢIONAL STATISTIC. *Lato sensu*, ansamblul mijloacelor, metodelor şi procedeelor necesare realizării culegerii, validării, transmiterii, prelucrării, memorării, regăsirii şi analizei informaţiei statistice. *Stricto sensu*, **s.i.s.** se compune din :

1. *ansamblul datelor statistice* (*indicatorilor economico-sociali statistici*) înscrise, respectiv păstrate pe suporturi de informaţie de tip document şi/sau magnetice (dări de seamă statistice, formulare statistice, arhive şi serii cronologice de date statistice, rapoarte, analize şi publicaţii statistice). Acestea, la rîndul lor, sînt : a) *date statistice de intrare* : furnizate de unităţile economico-sociale raportoare de date statistice sau culese de personalul specializat sau alocat temporar al organului central de statistică, şi anume : dări de seamă statistice ale unităţilor economico-sociale ; rapoarte statistice ocazionale (înregistrări statistice

speciale) ale unităților economico-sociale; formulare statistice de evidență primară (de exemplu, ale populației, vămilor etc.); formulare de anchete statistice; formulare de recensăminte statistice; date statistice memorate; date statistice de ieșire destinate organelor de conducere supreme, teritoriale și departamentale ale statului; acestea pot fi rapoarte, analize și studii cu caracter de uz intern sau secret;

2. *informația de descriere a datelor statistice și fluxurilor informaționale statistice* (clasificări statistice; nomenclatoare statistice; norme metodologice statistice; norme tehnice statistice; documentație de prelucrare statistică) prin care se descriu procedurile de elaborare a ieșirilor statistice și de memorare a datelor statistice).

Ansamblul informației de descriere a datelor statistice și a fluxurilor informaționale statistice este denumit și *sistem metainformațional statistic.*
Ansamblul datelor statistice arhivate și/sau memorate sau păstrate ca serii de date statistice este denumit și *fond de date statistice.* Organului central de statistică din fiecare țară îi și revine atribuția organizării fondului național de date statistice.

BIBLIOGRAFIE

1. COSTAKE, N., *Cu privire la noțiunea de sistem informațional statistic integrat*, în „Revista de Statistică" nr. 1, 2, 1973.
2. SUNDGREN, B., *An infological approach to data bases S.C.B.*, Statistika Centralbyrån, Stockholm, 1973.

FLUX INFORMAȚIONAL STATISTIC. Circuitul informației statistice de la *unitatea de raportare* până la *realizarea lucrării statistice* care iese din sistemul informațional statistic sau actualizarea *fondului de date statistice* destinat accesului mai multor utilizatori. De regulă, în cadrul unei economii naționale există numeroase **f.i.s.** paralele.

Flux informațional statistic cu canale multiple. Datele statistice culese circulă pe două sau mai multe circuite paralele, rezultatele centralizării fiind în principiu cel puțin în parte identice. **F.i.s.c.m.** prezintă avantajul posibilității verificării rezultatelor prin confruntare și dezavantajul dublicării efortului. Avantajul poate fi important în cazul prelucrării informației statistice fără suportul calculatoarelor electronice.

Flux informațional statistic cu canal unic. Fluxuri în care aceleași date statistice circulă pe un singur circuit informațional.
În România, există ambele tipuri de **f.i.s.**

BIBLIOGRAFIE

1. COSTAKE, N., *Sistemul informațional economic*, în „Revista de Statistică" nr. 1, 2, 3, 1967.

NOMENCLATOR STATISTIC UNITAR. Convenție de reprezentare codificată a unei clasificări statistice astfel încît codurile atribuite să fie invariante. O clasificare statistică realizează o partiție pe mai multe niveluri ierarhice, definind un graf de tip arbore care constituie suportul centralizării indicatorilor economico-sociali însumabili. Pentru realizarea invarianței codurilor:
a) fiecărei entități a clasificării statistice și, respectiv, fiecărui nod al grafului

i se asociază două coduri : codul de identificare propriu și codul de identificare a unității de nivel ierarhic imediat superior ;
b) codurile atribuite sînt nesemnificative (de exemplu, numere întregi pozitive oarecare) ;
c) o entitate este definită într-o singură clasificare ;
d) actualizarea nomenclatorului se face de către o singură organizație, care difuzează în sistemul informațional economico-social modificările operate și/sau ediții actualizate ale nomenclatorului.
Pentru siguranța codificării sînt utilizate coduri cu caracter (cifră) de control.
Există cel puțin trei clasificări care au caracterul unor partiții generale ale sistemului economic : *clasificarea administrativă* — structura ierarhică a unităților economico-sociale ; *clasificarea administrativ-teritorială* — structura ierarhică a zonelor teritoriale și localităților ; *clasificarea ramurilor, produselor și serviciilor.*
N.s.u. sînt create, întreținute și difuzate cu ajutorul calculatorului electronic și sînt definite legal.
În general, o poziție dintr-un nomenclator unitar conține : cele două coduri asociate entității — codul de identificare propriu și cel al entității ierarhic imediat superioară ; textul descriptiv ; codurile asociate din alte nomenclatoare unitare : codurile asociate din alte nomenclatoare. Codurile asociate creează deci legături cu alte nomenclatoare.
În țara noastră, primele nomenclatoare unitare, și anume SIRUES (nomenclatorul unitar al unităților economico-sociale) și SIRUTA (nomenclatorul unitar al unităților teritorial-administrative), au fost definite prin H.C.M. 74/1976. Sînt în curs de elaborare sau finalizare nomenclatorul unitar al ramurilor, produselor și serviciilor, nomenclatorul unitar al indicatorilor dezvoltării economico-sociale și nomenclatorul unitar al clasificării unice a produselor și serviciilor.
Existența unor subordonări multiple are drept consecință ca relațiile de subordonare efective să nu poată fi reprezentate printr-un graf de tip arbore, ci printr-un graf de tip rețea. Necesitatea regulilor unice de centralizare a datelor statistice impune însă adoptarea unor convenții necesare pentru realizarea nomenclatoarelor unitare sub formă de graf arbore. Utilizarea exclusivă de nomenclatoare unitare reprezintă condiția de bază a integrării sistemului informațional economico-social.

În afară de nomenclatoarele unitare, în prelucrarea statistică se utilizează și *nomenclatoare specifice*, elaborate pentru necesitățile unei singure prelucrări statistice (de exemplu, un recensămînt sau o singură dare de seamă), într-o anumită perioadă (de exemplu un an). Nomenclatoarele specifice sînt liste simple utilizîndu-se codificarea semnificativă, ceea ce poate ușura programarea prelucrării datelor. Nomenclatoarele specifice prezintă însă următoarele dezavantaje, care sînt în mod evident precumpănitoare : a) codificarea semnificativă este sensibilă la modificări organizatorice ; b) analiza statistică este orientată pe domenii specializate și nu pe ansamblul sistemului economic, constituind un factor de perturbare a deciziei macroeconomice.
Pentru asigurarea suportului pregătirii deciziilor macroeconomice este necesară realizarea unui sortiment minim de nomenclatoare unitare, de interes general, care să cuprindă : nomenclatorul unitar al unităților teritoriale (coordonate geografice) ; nomenclatorul unitar al profesiilor ; nomenclatorul unitar al ocupațiilor ; nomenclatorul unitar al cetățenilor și naționalităților.

Din considerente de compatibilitate cu nomenclatoarele internaţionale se recomandă ca fiecărei poziţii dintr-un nomenclator unitar să i se asocieze şi eventualele coduri internaţionale asociate.

REGISTRU STATISTIC AUTOMAT. Extindere a noţiunii de nomenclator statistic unitar. Reprezintă un nomenclator statistic creat, întreţinut şi difuzat cu ajutorul calculatorului electronic. La poziţiile acestui nomenclator sînt asociate şi valorile curente ale unor indicatori caracteristici (de exemplu, fiecărei unităţi economico-sociale i se pot asocia numărul mediu de personal şi valoarea fondurilor fixe). Primele **r.s.a.** intrate în practica sistemelor informaţionale ale statisticii moderne au fost registrele automate ale populaţiei (în Suedia, Norvegia). Realizarea registrelor automate ale populaţiei implică şi legiferarea codului de identificare personal (adoptat şi în ţara noastră).

În România a fost realizat registrul automat al localităţilor care conţine datele demografice totale aferente fiecărei localităţi.

O dezvoltare a **r.s.a.** o constituie asocierea nu numai a valorilor curente ale unor indicatori, ci a *seriilor cronologice* *. În România a fost realizat **r.s.a.** al seriilor cronologice ale judeţelor, cuprinzînd cîteva mii de serii asociate fiecărui judeţ. În cazul în care seriile sînt asociate unor nomenclatoare unitare se poate vorbi despre existenţa unor *baze de date statistice* * în sensul restrîns al termenului.

BIBLIOGRAFIE

1. *Ordinul Directorului General al Direcţiei Centrale de Statistică,* B. Of., partea 1, nr. 35 (5.5.1980).
2. COSTAKE, N., CUCINSCHI, E., ZAMFIRESCU, R., *Registrul unităţilor economico-sociale* (SIRUES), în „Revista de Statistică" nr. 5, p. 1976.

SISTEM INFORMATIC STATISTIC. Ansamblul echipamentelor şi programelor care asigură automatizarea prelucrării, memorării, regăsirii, eventual şi a culegerii şi publicării informaţiei statistice.

S.i.s. se poate defini prin:

a) *structura funcţională* constituită din următoarele subsisteme:

— *subsistemul metainformatic statistic,* care cuprinde baza de metadate statistice sau cel puţin ansamblul nomenclatoarelor unitare statistice şi/sau al registrelor automate statistice;

— *subsistemul prelucrării datelor recensămintelor şi anchetelor statistice* de mare volum, subsistem elaborat şi utilizat de regulă, de către organul naţional al statisticii de stat sau organe naţionale specializate pentru cercetări statistice;

— *subsistemul prelucrării datelor statistice operative şi curente,* realizat şi utilizat, de regulă, de toate unităţile informatice coordonatoare ale prelucrării automate a datelor statistice din economie;

— *subsistemul automatizării pregătirii publicării informaţiei statistice,* subsistem realizat şi utilizat de regulă de organele naţionale ale statisticii de stat;

— *subsistemul bazelor de date statistice,* subsistem realizat şi utilizat de regulă de organele naţionale ale statisticii de stat, care asigură accesul şi al altor unităţi interesate;

— *subsistemul modelelor matematice macroeconomice statistice,* subsistem realizat şi utilizat de regulă de organele naţionale ale statisticii de stat, care asigură şi accesul altor organe ale administraţiei de stat interesate;

b) *structura de echipamente de prelucrare a informației statistice* constă din :
— *echipamentul electronic central*, în principal unul sau mai multe calculatoare pentru prelucrarea datelor statistice ; *echipamentul pentru introducerea și verificarea datelor statistice la nivel central*.
Tendința mondială actuală constă în renunțarea la mașinile de perforare/verificare cartele în favoarea următoarelor : pentru volume mari de date statistice — cititoare optice automate ale formularelor statistice și corectarea prin terminale cu tub catodic a formularelor completate într-un mod care nu a putut fi recunoscut de cititorul optic automat ; pentru volumele uzuale de date statistice operative și curente — echipamente de tip mini/microcalculator care realizează introducerea/verificarea datelor cu ajutorul terminalelor cu tub catodic și conversia lor pe suport magnetic (disc sau bandă magnetică), eventual conectate la echipamentul electronic central ;
— *echipamentul pentru interogarea/utilizarea datelor statistice la nivel central*. Pentru interogarea/analiza datelor statistice și metadatelor statistice memorate, echipamentul curent este în prezent terminalul cu tub catodic echipat cu imprimantă terminal pentru reținerea rezultatelor apreciate ca imprimante :
— *echipamentul pentru realizarea publicațiilor statistice*. Tendința mondială actuală constă în utilizarea echipamentelor automate pentru trasarea graficelor (inclusiv în culori), precum și a unor tehnici avansate de realizare automată a publicațiilor pe baza datelor statistice și textelor rezultate din prelucrarea automată și furnizate pe suport magnetic (de exemplu, metoda fotocompoziției) ;
— *echipamentul pentru teletransmisia automată a datelor statistice*. Realizarea rețelelor de teleprelucrare a datelor statistice constituie una din tendințele actuale. Rețelele de teleprelucrare a datelor statistice pot necesita : minicalculatoare de telecomunicații (pentru descărcarea echipamentului central și/sau pentru concentrarea mesajelor de la/la mai multe terminale) ; modemuri (pentru realizarea legăturii dintre echipamentul electronic de calcul și liniile de telecomunicații) ; secretizoare (de exemplu, pentru realizarea confidențialității datelor transmise pe liniile de telecomunicații publice) ;
— *echipamentul electronic pentru prelucrarea datelor statistice la nivel teritorial* depinde de mărimea și specificul teritoriului pe care îl servește în relațiile cu volumul datelor statistice culese, dar și cu complexitatea problemelor de analiză statistică a dezvoltării economico-sociale a teritoriului. În cazul unor teritorii mari, organele teritoriale de statistică sînt dotate cu calculatoare electronice (de regulă, sub forma centrelor de calcul teritoriale, care servesc și unitățile economice mici și celelalte unități economico-sociale din teritoriu). Există și situația dotării cu minicalculatoare, acestea putînd fi interconectate într-o *rețea de teleprelucrare a datelor statistice* *. În cazul unităților teritoriale relativ mici, este suficientă dotarea cu echipamente mai simple (de genul mașinilor de facturat-contabilizat electronice programabile și/sau cu terminale cu tub catodic (cu cel puțin o imprimantă terminal).
În România, direcțiile județene de statistică sînt dotate cu mașini electronice de facturat-contabilizat programabile FC 64 și FC 128 și/sau cu terminale cu tub catodic, conectate la calculatoarele Centrului de calcul al Direcției Centrale de Statistică ;
— *echipamentul electronic pentru prelucrarea datelor statistice la nivelul centralelor, departamentelor și ministerelor* nu are caracter specific, o parte din capacitatea echipamentului existent fiind utilizată pentru prelucrarea datelor statistice.

— *canalul de transmisie a datelor statistice între organele centrale și organele teritoriale de statistică respectiv organele departamentale*: linia telefonică închiriată sau comutabilă pentru teletransmiterea automată a informației; acest canal este utilizat în cadrul rețelelor de teleprelucrare automată a datelor statistice; legătura telex, pentru transmiterea volumelor de informație reduse; transmiterea automată a documentelor; transportul unor suporturi materiale de informație: bandă magnetică, bandă perforată (dacă rezultă ca produs secundar al unei prelucrări) sau documente (dările de seamă și/sau formularele statistice, centralizatoare de date statistice).

Tendința modernă constă în utilizarea liniei telefonice pentru transmisia automată a datelor statistice operative și curente, bandă magnetică în cazul dărilor de seamă statistice cu volum mare de date și transportul documentelor (în vederea citirii optice automate) în cazul volumelor foarte mari de date statistice. Conform acestei tendințe, introducerea și controlul datelor statistice, sînt pe cît posibil, descentralizate, deci cît mai aproape de locul generării informației statistice.

c) *Structura programelor de prelucrare automată a datelor statistice,* care cuprinde:
— *sistemul programelor pentru crearea/întreținerea, livrarea, copierea pe suport magnetic a nomenclatoarelor unitare statistice*;
— *sistemul programelor pentru prelucrarea automată a datelor statistice în loturi* (*în mod lot*): programe și generatoare de programe de introducere și control logic automat ale datelor statistice; generatoare de tabele statistice; programe pentru crearea/întreținerea textelor care nu sînt preluate automat din nomenclatoarele unitare; programe de elaborare a tabelelor statistice; programe pentru automatizarea pregătirii publicațiilor statistice în sensul asigurării confidențialității, precum și în sensul elaborării suporturilor pentru comanda echipamentelor de trasare grafic și/sau imprimare/multiplicare;
— *sistemul programelor pentru teleprelucrarea datelor statistice operative și curente*: programe pentru teleintroducerea/validarea/centralizarea datelor statistice; programe pentru apelul la distanță a tabelelor statistice; programe pentru teleprelucrarea datelor statistice în timp real (realizînd practic simultan introducerea/validarea și elaborarea datelor statistice);
— *sistemul programelor de gestiune a bazelor de date/metadate statistice* asigurînd: crearea/actualizarea bazelor de date statistice; telcinterogarea bazelor de date statistice: interfața de utilizare a bazei de date statistice cu sistemul programelor de prelucrare a datelor statistice în mod lot;
— *sistemul programelor de analiză a datelor statistice*: programe generale pentru calcule statistico-matematice; programe pentru calcule de analiză statistică specializate; programe pentru modelarea dezvoltării proceselor economico-sociale.

Tendințele actuale în domeniul programelor de prelucrare a datelor statistice constau în: standardizarea soluțiilor; mărirea performanțelor, prin adaptarea cît mai bună la necesitățile prelucrării datelor statistice și utilizarea tehnicilor avansate; creșterea ponderii programelor de teleprelucrare a datelor statistice; realizarea unei „interfețe prietenoase" între calculator și utilizatorul statistician. Caracteristicile aproximative ale unei asemenea interfețe pot fi următoarele: universalitate (permite accesul utilizatorului la toate datele statistice și procedurile necesare); transparență (utilizarea programelor poate fi realizată de personal statistic fără pregătire specială informatică); cooperativitate (de exemplu, posibilitatea utilizatorului de a întrerupe și relua execuția cu o even-

tuală modificare a parametrilor, furnizarea de mesaje cu înțeles clar, de natură să ghideze acțiunile utilizatorului; personalizare (de exemplu, utilizarea de mesaje ierarhice, permițînd distincția dintre utilizatorul neinstruit sau infrecvent și utilizatorul experimentat — această caracteristică fiind importantă în special în cazul programelor de teleprelucrare).

S.i.s. este format din subsisteme orientate pe domenii: subsistemul informatic al statisticii industriei, subsistemul informatic statistic al agriculturii și silviculturii, subsistemul informatic statistic al muncii și retribuirii, subsistemul informatic statistic al populației etc.

În România a fost elaborat un sistem de programe de prelucrare standard generale pentru prelucrarea dărilor de seamă statistice operative și curente, utilizat direct de către economiștii statisticieni: sistemele de programe ESOP.

Sistem informatic statistic integrat. Sistem caracterizat prin următoarele: a) utilizarea exclusivă a nomenclatoarelor unitare, ceea ce are drept consecințe descrierea unitară a proceselor macroeconomice, complexe și posibilitatea memorării datelor statistice în baze de date;

b) un anumit indicator economico-social (respectiv o anumită metadată) este cules, la un moment dat, o singură dată, ceea ce are drept consecințe, printre altele, și organizarea fluxului informațional pe principiul canalului unic și actualizarea centralizată a nomenclatoarelor unitare, precum și a fișierelor și bazelor de date statistice.

Integrarea sistemului informatic statistic constituie o cerință evidentă pentru pregătirea deciziilor de conducere macroeconomică, întrucît permite elaborarea modelelor matematice macroeconomice statistice. Lipsa integrării sistemului informatic statistic prezintă un număr de dezavantaje evidente, printre care imposibilitatea corelării datelor statistice și imposibilitatea prevederii consecințelor probabile ale deciziilor posibile.

BIBLIOGRAEIE

1. COSTAKE, N. *Despre sistemul informațional statistic,* în „Revista de Statistică" nr. 7, 1979.
2. FEDORENKO, N. și colab., *Dicționar de matematică și cibernetică în economie,* Ed. științifică și enciclopedică, București, 1979.
3. HUGHES, T. (ed.), *Statistics beyond 1980,* UNO-ECE, 1979.
4. SUNDGREN, B., *Statistical data processing systems — architecture and design methodology,* Comunicare la sesiunea Institutului Internațional de Statistică, 1981.

CALCULATOR ELECTRONIC *(utilizat în prelucrarea datelor statistice).* Calculator electronic destinat unui organ central de statistică avînd următoarele caracteristici: memorie internă 2...4 MB; memorie externă cu acces direct (de exemplu unități cu discuri magnetice), 1... 10 GB; în cazul prelucrării automate a datelor statistice există o regulă empirică potrivit căreia raportul dintre capacitatea memoriei externe și capacitatea memoriei interne este de ordinul a 10^3; 4...6 unități cu bandă magnetică; memorie virtuală; viteză de prelucrare 1...2 mil. operații/s; sistem de operare permițînd teleprelucrarea interactivă, la nivelul unei rețele de ordinul zecilor, sutelor de terminale cu tub catodic (mod mesaj), inclusiv prin intermediul unor minicalculatoare; sisteme de programe pentru aplicații statistice care permit: prelucrarea automată a formularelor statistice în loturi; teleprelucrarea automată a dărilor de

seamă statistice ; crearea/întreținerea/utilizarea bazelor de metadate statistice ; crearea/întreținerea/utilizarea bazelor de date statistice ; analiza statistico-matematică ; elaborarea/utilizarea modelelor macroeconomice dinamice cu mai multe ramuri ; elaborarea interactivă de programe, inclusiv în limbaje de programare de nivel înalt. (FORTRAN, COBOL, PL/1, PASCAL, BASIC, APL). În etapa actuală, caracteristicile tehnice minime ale *minicalculatoarelor* utilizate în prelucrarea descentralizată a datelor statistice se pot considera următoarele : memorie internă 128...256 KB ; memorie externă 40...100 MB în acces direct pe discuri magnetice +2 unități cu bandă magnetică ; unități periferice ; 4...6 terminale cu tub catodic, cititor cartele, imprimante ; posibilitatea (echipament/sistem de operare) de conectare la rețele de teleprelucrare ; sisteme de programe standard pentru prelucrarea formularelor statistice în loturi ; crearea/întreținerea unei baze de serii de date statistice ; analiza statistico-matematică cu sistem de programare interactivă, la nivelul cel puțin a unui limbaj de nivel înalt.

BIBLIOGRAFIE

1. *Dicționar de informatică*, Ed. științifică și enciclopedică, București, 1981.
2. JOVMIR, S., MITRAN, T., *Mic dicționar de informatică*, în „Revista de Statistică" nr. 1, 1980

PROGRAME ȘI SISTEME DE PROGRAME STANDARD PENTRU PRELUCRAREA DATELOR STATISTICE. Programe și sisteme de programe specifice informaticii statistice elaborate ca soluții standard cu caracter general. Standardizarea se realizează la următoarele niveluri : programe cu caracter specializat (execută, de exemplu o singură funcțiune) ; ele constituie o restricție în proiectarea prelucrării, dar pot asigura performanțe înalte ; programe parametrizate (pot fi utilizate în principiu de către utilizatorul nespecializat și adaptate necesităților aplicației în limite relativ largi) ; *programe generate automat* pe baza descrierii efectuate într-un limbaj orientat pe aplicație, eventual direct de către utilizatorul statistician.

Programe și sisteme de programe standard pentru prelucrarea dărilor de seamă și formularelor statistice în mod lot care asigură :

— crearea, întreținerea, copierea pe bandă magnetică a nomenclatoarelor utilizate în prelucrare ;

— introducerea/validarea informației cu semnalarea erorilor și eventual cu opțiunea de corecție automată și calculul de indicatori suplimentari (programe de control logic). Controlul logic al datelor statistice este un control de plauzibilitate și se referă în mod curent la : verificarea unor relații aritmetice între indicatorii raportați (cheie de tip „total"=„suma elementelor componente") sau între indicatorii raportați în perioada curentă și cea precedentă (cheie de tip „cumulat" : valoare cumulată raportată = valoarea cumulată raportată anterior + valoare din perioada curentă) ; verificarea unor relații de ordin (de exemplu : „din care" mai mic decît „total") ; verificarea unor relații de plauzibilitate între valorile unor indicatori (de exemplu : gradul de general nu este plauzibil la o vîrstă sub 30 ani) ; acest tip de control logic utilizează nomenclatoare ; verificarea primirii datelor statistice de la toate unitățile raportoare implicate (controlul de integritate).

V.TREBICI

Automatizarea controlului logic al informației statistice are o importanță deosebită în asigurarea calității rezultatelor. În vederea creșterii acestei calități are sens și minimizarea informației introduse, de exemplu preluarea caracteristicilor unităților de raportare dintr-un nomenclator memorat și nu din formularul raportat, în vederea evitării erorilor de testare ;
— elaborarea tabelelor statistice. Din punctul de vedere al prelucrării datelor statistice, tabelele pot fi de două tipuri de bază ; tabele cu structură fixă, ale căror dimensiuni sînt sau pot fi predeterminate (de exemplu : populația, pe grupe de vîrstă și sexe) ; tabele liniare, ale căror dimensiuni nu pot fi predeterminate sau nu este sigur că pot fi predeterminate (de exemplu : valorile indicatorilor producției industriale pe feluri de unități raportoare, pe unități raportoare). Un element de tabel liniar poate fi un subtabel fix (de exemplu, pentru fiecare nivel din tabelul liniar anterior se pot imprima 2 rînduri : valori în prețuri constante, valori în prețuri curente).
Există și alte tipuri, a căror utilizare nu se recomandă.
Elaborarea ambelor tipuri de tabele statistice implică citirea întregului fișier de date statistice validate în urma controlului logic automat și cumularea în diferitele niveluri de totaluri ale valorilor indicatorilor specificați prin denumirea tabelului sau +1 (în cazul calculului de frecvențe).
Operația de calcul al tabelelor statistice în memoria internă a calculatorului se numește și ventilare, caracterizată prin definirea elementului de tabel care este ventilat, definirea desfășurării (nivelurilor de totalizare) și a criteriului de apartenență la tabel a datelor din arhiva supusă prelucrării.
Poate avea sens și calculul de coloane suplimentare (după efectuarea centralizării), iar în cazul tabelelor cu structură fixă sau a subtabelelor cu structură fixă ale tabelelor liniare — și calculul de rînduri suplimentare ;
— imprimarea tabelelor care constă în imprimarea datelor în forma fișierelor tabel elaborate cu textele conform nomenclatoarelor, eventual cu calcule de coloane suplimentare și/sau diferite editări).

Programe standard pentru prelucrarea dărilor de seamă/formularelor statistice în mod interactiv. Programe care realizează de regulă parțial funcțiile citate mai sus, utilizînd ca unitate periferică de intrare și de ieșire terminale. În soluțiile de tip timp real, elaborarea tabelelor are loc simultan cu introducerea și validarea datelor.

Sisteme de gestiune a bazelor de date statistice. Sisteme al căror specific, în raport cu alte sisteme de gestiune de baze de date, constă în : asigurarea memorării de serii cronologice de date statistice, cu posibilități de actualizare a valorilor din trecut în funcție de modificările organizatorice ; existența, în general, a două niveluri de agregare : nivelul datelor individuale la nivelul unităților de raportare : nivelul datelor agregate pe nivelurile standard generale ; prevederea modului de lucru interactiv (pentru realizarea răspunsului rapid la solicitări), precum și lot (pentru realizarea actualizărilor bazelor de date) utilizînd fișierele create în prelucrarea datelor statistice operative și curente.
BAZĂ DE METADATE STATISTICE : Colecție de metadate statistice unitare memorate și accedate sub controlul unui sistem de gestiune a bazelor de date. O bază de metadate statistice nu cuprinde în mod obligatoriu serii de metadate, ci versiunea curentă (eventualele versiuni anterioare putînd fi copiate în fișiere de date și reîncărcate în caz de necesitate în bază de metadate).
BAZĂ DE DATE STATISTICE. Colecție structurată de serii cronologice de date statistice, memorate conform unei baze de metadate statistice, al cărei

conținut reflectă concepția unui model macroeconomic sau a unui submodel al acestuia și este actualizat sistematic pe baza datelor statistice operative și curente, care permit accesul rapid al utilizatorilor pentru interogare și analiză, pe baza unui limbaj orientat pe aplicație și/sau unor convenții de utilizare simple și al cărui conținut este relativ independent de modificările organizatorice din economia națională. Pentru a realiza independența maximă posibilă față de modificările organizatorice, datele statistice se memorează la nivelul minim. Pentru a realiza viteza de răspuns maximă, sistemul de programe pentru gestiunea **b.d.s.** prevede atît agregarea la cerere (de-a lungul unui subgraf al unui nomenclator unitar), precum și memorarea datelor agregate periodic în mod lot (baza de date statistice primare și **b.d.s.**)
Conținutul **b.d.s.** poate fi perturbat (în sensul compatibilității termenilor seriilor de date statistice memorate) de modificări metodologice și/sau de prețuri. Exploatarea unei **b.d.s.** implică desemnarea administratorului bazei de date statistice.

BANCĂ DE DATE STATISTICE. Ansamblul constituit dintr-o bază de metadate statistice și o bază de date statistice utilizate cu ajutorul unui sistem de gestiune specific. Reprezintă elementul caracteristic sistemelor informatice statistice moderne integrate.

BIBLIOGRAFIE

1. ARVAS, C., *Joint use of data banks*, în „Revista de Statistică", Anexă, 1968.
2. MARTIN, M. J., *Computer Data Base Organisation*, Prentice-Hall. 1977.
3. PESCARU, V., CATONA, I., DUNA, D., POPESCU, C., ȘATRAN, I., *Fișiere, baze și bănci de date*, Ed. tehnică, București, 1976.

PRELUCRARE AUTOMATĂ A DATELOR STATISTICE OPERATIVE. Prelucrare automată periodic — zilnică, decadală, lunară, trimestrială — relativ simplă (orientată pe validare și centralizare pe niveluri standard (de regulă minister/departament/centrală/unitate sau județ/municipii + orașe, respectiv comune) avînd drept obiectiv esențial obținerea rezultatelor într-un timp cît mai scurt (ciclul total de prelucrare administrat pe ansamblul organizației statistice de ordinul 24—72 ore).
Teleprelucrarea datelor statistice operative cu utilizarea nomenclatoarelor unitare constituie o soluție eficientă pentru sistemul informatic statistic.

PRELUCRARE AUTOMATĂ A DATELOR RECENSĂMÎNTULUI POPULAȚIEI ȘI AL LOCUINȚELOR. Prelucrarea automată a datelor statistice individuale înregistrate pe bază de declarație, cu privire la situația la un moment dat a fiecărui locuitor al unei țări. Datorită înregistrării la nivel individual, prelucrarea datelor recensămintelor populației și ale locuințelor este posibilă într-un număr foarte mare de profiluri, putînd constitui suportul unor analize complexe macroeconomice. Prelucrarea automată a datelor recensămîntului populației și locuințelor este caracterizată în afară de volumul mare de date, prin următoarele :

a) culegere de date individuale exhaustive, ceea ce impune *proiectarea formularului de înregistrare* astfel încît să se simplifice cît mai mult operația de înregistrare (în unele țări autoînregistrare), precum și de *introducere în calculator* (preferabil prin citirea automată a formularului) ;

b) prevederea unui control logic automat al datelor înregistrate : ținînd seama de volumul foarte mare de date este permisă și utilizarea de corecții automate de tip determinist (de exemplu, *sex feminin*, și *vîrsta 12 ani*, și profesia *elev*, și situația față de capul gospodăriei *fiică*, și sursa de existență *întreținută* de părinți arată că numărul de copii născuți vii în viață este nul, indiferent de numărul înregistrat) sau corecții de tip statistic (de exemplu, stare civilă *căsătorit, sex bărbătesc, tehnician ocupat în industrie*, cu *vîrsta nedeclarată*, poate primi vîrsta individului anterior cu aceleași caracteristici de sex, profesie, stare civilă, ramură). Corecția automată nu poate fi utilizată în toate cazurile, fiind necesară semnalarea ca eroare a cazurilor nerezolvabile (de exemplu, sexul și vîrsta necunoscute).

În urma controlului logic automat se creează *arhiva datelor recensămîntului* populației și al locuințelor, de regulă pe bandă magnetică (există și cazuri în care arhiva datelor recensămîntului populației s-a păstrat pe discuri magnetice, pentru a permite elaborarea rapidă de tabele la cerere). Se recomandă și elaborarea statisticii corecțiilor automate efectuate, pentru a evalua plauzibilitatea deciziilor metodologice (de exemplu, concentrarea corecțiilor automate pe o singură procedură și în același timp o frecvență ridicată a acestora sugerează posibilitatea unei imperfecțiuni metodologice).

c) Elaborarea și imprimarea tabelelor recensămîntului, validarea acestora (la nivelul unei unități teritoriale) și transmiterea lor spre publicare. Avînd în vedere volumul foarte mare de date, precum și complexitatea metodologiei recensămîntului, organizarea prelucrării nu se face pe etape funcționale (introducere, constituire arhivă, elaborare tabele, imprimare tabele) la nivelul întregului volum de date ci pe unități teritoriale în cazul țării noastre, pe județe. Tendința modernă constă în atribuirea de coduri de identificare personale cu ocazia efectuării unui recensămînt al populației, care devine astfel și sursa de încărcare a unui registru permanent al populației.

BIBLIOGRAFIE

1. COSTAKE, N., FILOTTI, I., ZAMFIRESCU, R., *Unele observații privind programarea prelucrării datelor recensămîntului populației și locuințelor din 15 martie 1966*, în „Revista de Statistică" nr. 2, 1968.

PRELUCRARE INTEGRATĂ A DATELOR DEMOGRAFICE. Prelucrare automată a datelor statistice demografice proiectată și organizată astfel încît toate prelucrările statistice de date demografice să constituie un sistem informatic statistic integrat. Exemplu : încărcarea unui registru automat al populației cu ocazia unui recensămînt, actualizarea sa permanentă prin prelucrarea buletinelor demografice și elaborarea de tabele statistice demografice exclusiv pe baza datelor memorate în registrul automat al populației.

În România, prelucrarea integrată a datelor demografice are loc la un nivel agregat (nivelul minim fiind cel al populației totale pe sexe și ani de vîrstă la nivel de comună). Sistemul informatic statistic de prelucrare a datelor demografice, care utilizează nomenclatorul unitar al unităților teritorial-administrative, cuprinde :

— prelucrarea automată a datelor buletinelor demografice, din care rezultă datele demografice curente ;

— actualizarea datelor demografice ale localităților obținute la ultimul recensămînt al populației prin rezultatele prelucrării datelor demografice curente ;

— actualizarea registrului permanent demografic al localităților :
— calculul ratelor de fertilitate și mortalitate pe medii ;
— subsistemul prognozei demografice ;
— subsistemul analizei matematice a migrației populației.

BIBLIOGRAFIE

1. COSTAKE, N., GOREAC, I., *Program conversațional pentru modelarea creșterii populației și calculul unor indicatori asociați,* în „Studii și cercetări de calcul economic și cibernetică economică" nr. 2, 1981.

INTEGRAREA SISTEMULUI INFORMATIC STATISTIC ÎN SISTEMUL INFORMATIC NAȚIONAL. Constituirea sistemelor informatice naționale care constau în esență din următoarele :

a) elaborarea, întreținerea și difuzarea nomenclatoarelor unitare de către sistemul informatic statistic către celelalte subsisteme informatice din economie ;

b) elaborarea, întreținerea și difuzarea de către sistemul informatic statistic a sistemelor informatice standard de prelucrare a datelor statistice ;

c) asigurarea accesului la băncile de date statistice ;

d) conectarea calculatoarelor electronice din cadrul sistemului informatic statistic la rețeaua națională de calculatoare electronice ;

e) definirea interfețelor dintre sistemul informatic statistic și celelalte componente ale sistemului informatic național.

BIBLIOGRAFIE

1. COSTAKE, N., *Cu privire la proiectarea și realizarea sistemului național de informatică al R. S. România,* în *Cibernetica aplicată în economie,* Ed. Academiei R.S.R., București, 1982.
2. COSTAKE, N., PETREANU, N., *Unele probleme ale sistemului informatic național,* în „Revista Economică", nr. 34, 1982.

V. PREZENTAREA DATELOR STATISTICE

Rezultatele prelucrării primare sau secundare a datelor statistice sînt prezentate în *tabele statistice* * sau iau forma *reprezentărilor grafice* *. Acestea sînt forme comode pentru înțelegerea populațiilor și caracteristicilor acestora despre care s-au obținut informații cu prilejul observării statistice. Funcția tabelelor statistice și a reprezentărilor grafice nu se reduce la aceea de prezentare a rezultatelor prelucrării ; în numeroase cazuri, ea este și analitică. Aceasta explică reluarea unor tabele și reprezentări în capitolele următoare (*Regresie și corelație* ; *Analiza seriilor cronologice* ; *Control statistic al calității*).

TABEL STATISTIC. Formă de prezentare a rezultatelor prelucrării primare sau secundare, a unei sistematizări logice de date statistice primare sau derivate (medii, indici, coeficienți etc.), cu ajutorul căruia se caracterizează populația

(colectivitatea) investigată, părțile sale componente (subpopulațiile), legăturile dintre ele. În ultima analiză, **t.s.** este un ansamblu de reprezentări despre populație, care constituie *subiectul* * **t.s.** și despre *predicat* *.

Redus la esențial, **t.s.** se prezintă sub forma a două coloane : una conține valorile variabilei (x_i) care delimitează clasele și alta—efectivele (f_i) observate pentru fiecare clasă.

Subiectul **t.s.** Populația sau subpopulațiile statistice despre care se fac propoziții logice în **t.s.**

Predicatul **t.s.** Elementele statistice ce caracterizează subiectul **t.s.** : numărul unităților, valoarea însumată a caracteristicilor etc., adică ceea ce se spune despre subiectul **t.s.**

Introducerea **t.s.** marchează un moment important în istoria statisticii (→ *Statistică*.)

Între *grupările statistice* și **t.s.** există o simetrie.

Tabel statistic simplu. (*Tabel cu simplă intrare*). Tabel prezentînd repartiția unei colectivități după o singură caracteristică (atributivă, de timp sau spațiu). Macheta tabelului este formată din două coloane : prima destinată înscrierii modalităților caracteristicii, a doua — înscrierii frecvenței sau nivelului fiecărei modalități. Un tabel statistic care descrie, de exemplu, colectivitatea după o variabilă discretă are forma generală :

Valorile variabilei x_i	Frecvențele f_i
x_1	x_1
x_2	f_2
⋮	⋮
x_i	f_i
⋮	⋮
x_n	f_n
Total	$\sum_{i=1}^{n} f_i$

În măsura în care se consideră util, se adaugă coloane referitoare la frecvențe relative, frecvențe cumulate, valori globale, abateri etc.

Exemplu de **t.s.s.** în care datele sînt ordonate cronologic. Populația României la recensămintele din 1930, 1948, 1956, 1966, 1977

Tabelul 1

Data recensămîntului	Numărul populației totale
29 decembrie 1930	14 280 729
25 ianuarie 1948	15 872 623
21 februarie 1956	17 489 450
15 martie 1966	19 103 163
5 ianuarie 1977	21 559 910

Sursa : Anuarul statistic al R. S. România, 1981, p. 45.

Tabel statistic pe grupe. T.s. în care populația (colectivitatea) este prezentată pe grupe sau *clase* după o singură caracteristică ; aceasta din urmă poate fi cantitativă sau calitativă.

Numărul întreprinderilor și al personalului muncitor după forma de proprietate, în 1980

Tabelul 2

Forma de proprietate	Numărul întreprinderilor	Numărul personalului muncitor	Numărul personalului muncitor pe o întreprindere
INDUSTRIE	1752	3 198 100	1825
Industria de stat	1334	2 896 600	2171
Industria cooperatistă	418	301 500	721

Sursa : Anuarul statistic al R. S. România, 1981, p. 135.

Tabel statistic combinat. T.s. în care populația apare sub formă de grupări combinate, cu predicatul prelucrat sau neprelucrat ; populația este grupată după două sau mai multe caracteristici.

TABEL CU DUBLĂ INTRARE. Tabel înfățișînd repartiții combinate după două caracteristici. Cînd există o dependență între cele două caracteristici, tabelul capătă denumirea de *tabel de corelație*. Să considerăm, de exemplu, o colectivitate descrisă simultan după două variabile X și Y. Notînd prin x_1, $x_2, \ldots, x_i, \ldots, x_p$ valorile luate de variabila X prin $y_1, y_2, \ldots y_j, \ldots, y_q$ valorile

Populația României pe provincii istorice la recensământul din 5 ianuarie 1977 și ratele de natalitate, mortalitate și ale excedentului natural în perioda 1966—1980

Tabelul 3

Provincii istorice	Numărul populației	Rate (la 1000 locuitori)		
		Natalitate	Mortalitate	Excedent natural
ROMÂNIA	21 559 910	20,2	9,5	10,7
Oltenia	2 354 265	19,7	10,3	9,4
Muntenia	4 505 385	20,6	9,5	11,1
Dobrogea	863 348	20,6	8,1	12,5
Moldova	4 402 798	24,2	8,1	16,1
Banat	1 082 461	15,4	12,1	3,3
Transilvania	4 385 954	19,9	9,6	10,3
Crișana-Maramureș	2 031 814	19,0	11,1	7,9
Municipiul București	1 933 885	14,3	9,0	5,3

Sursa : V. Trebici, *Les provinces historiques de la Roumanie : aspects démographiques*, în vol. *Papers in Honour of Octav Onicescu* on His 90th Birthday, Nagard Publisher, Roma, 1983, p. 470.

luate de variabila Y, iar prin f_{ij} numărul unităților colectivității care au valoarea x_i a variabilei X și totodată valoarea y_j a variabilei Y, tabelul este de forma :

X \ Y	y_1	y_2	...	y_j	...	y_q	Total
x_1	f_{11}	f_{12}	...	f_{1j}	...	f_{1q}	$f_{1\cdot}$
x_2	f_{21}	f_{22}	...	f_{2j}	...	f_{2q}	$f_{2\cdot}$
⋮	⋮	⋮	...	⋮	...	⋮	⋮
x_i	f_{i1}	f_{i2}	...	f_{ij}	...	f_{iq}	$f_{i\cdot}$
⋮	⋮	⋮	...	⋮	⋮	⋮	⋮
x_p	f_{p1}	f_{p2}	...	f_{pj}	...	f_{pq}	$f_{p\cdot}$
Total	$f_{\cdot 1}$	$f_{\cdot 2}$	...	$f_{\cdot j}$	...	$f_{\cdot q}$	$f_{\cdot\cdot}$

Pe rîndurile tabelului figurează valorile variabilei X și pe coloane, valorile variabilei Y. Tabelul are p rînduri și q coloane. Pentru scurtarea scrierii simbolice, se indică printr-un punct o totalizare după indicele i sau indicele j.

$f_{i\cdot}$ este totalul frecvențelor rîndului i, însumarea fiind efectuată după indicele j :

$$f_{i\cdot} = f_{i1} + f_{i2} + \ldots + f_{ij} + \ldots + f_{iq} = \sum_{j=1}^{q} f_{ij};$$

$f_{\cdot j}$ este totalul frecvențelor coloanei j, însumarea fiind efectuată după indicele i :

$$f_{\cdot j} = f_{1j} + f_{2j} + \ldots + f_{ij} + \ldots + f_{pj} = \sum_{i=1}^{p} f_{ij}.$$

$f_{\cdot\cdot}$ reprezintă volumul colectivității, obținut prin însumarea efectuată fie rînd cu rînd, fie coloană cu coloană.
Însumarea rînd cu rînd

$$f_{\cdot\cdot} = \begin{cases} f_{11} + f_{12} + \ldots + f_{1j} + \ldots + f_{1q} \\ + f_{21} + f_{22} + \ldots + f_{2j} + \ldots + f_{2q} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ + f_{i1} + f_{i2} + \ldots + f_{ij} + \ldots + f_{iq} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ + f_{p1} + f_{p2} + \ldots + f_{pj} + \ldots + f_{pq} \end{cases} = \begin{cases} f_{1\cdot} \\ + f_{2\cdot} \\ \ldots \\ + f_{i\cdot} \\ \ldots \\ + f_{p\cdot} \end{cases}.$$

ceea ce se scrie

$$f_{\cdot\cdot} = \sum_{i=1}^{p} \sum_{j=1}^{q} f_{ij} = \sum_{i=1}^{p} f_{i\cdot}.$$

Însumarea coloană cu coloană se scrie

$$f_{\cdot\cdot} = \sum_{j=1}^{q} \sum_{i=1}^{p} f_{ij} = \sum_{j=1}^{q} f_{\cdot j}.$$

În definitiv

$$f_{\cdot\cdot} = \sum_{i=1}^{p} \sum_{j=1}^{q} f_{ij} = \sum_{i=1}^{p} f_{i\cdot} = \sum_{j=1}^{q} f_{\cdot j}.$$

Uneori se reunesc datele într-o repartiție de frecvențe cu trei dimensiuni (X, Y, Z) și în acest caz se construiește un tabel statistic *cu trei intrări*.

Examinat în raport cu forma sa, tabelul statistic are următoarele elemente :

a) un corp constituit dintr-un număr variabil de rînduri și coloane ;

b) indicații de identificare : număr, titlu general, titluri de rînduri și coloane, unități de măsură ;

c) indicații complementare : surse, note explicative.

Un tabel statistic bine alcătuit dispensează de pagini de comentarii și interpretări, fiind un instrument de lucru deosebit de util pentru specialiști.

TABEL DE ASOCIAȚIE. Tabel statistic care prezintă repartiția elementelor unei populații după două caracteristici corelate logic și care alternează între două posibilități.

Dacă X_1, avînd două posibilități, X_{11} și X_{12}, și X_2, avînd două posibilități, X_{21} și X_{22}, prezintă frecvențe comune, acestea sînt înfățișate în tabel de forma :

Valorile caracteristicii X_1	Valorile caracteristicii X_2		TOTAL
	X_{21}	X_{22}	
X_{11}	n_{11}	n_{12}	$n_{11} + n_{12}$
X_{12}	n_{21}	n_{22}	$n_{21} + n_{22}$
TOTAL	$n_{11}+n_{21}$	$n_{12}+n_{22}$	n

T.a. servește la calculul *coeficientului de asociație* *.

TABELĂ (statistică). Termen folosit în exclusivitate pentru un tip special de tabel statistic cuprinzînd funcțiile biometrice, cum ar fi *tabela de mortalitate* * sau *tabela de viață* * cu largă aplicabilitate în demografie, actuariat, controlul statistic al calității.

REPREZENTARE GRAFICĂ. Reprezentare prin figuri geometrice, naturale sau simbolice a datelor cuprinse într-un *tabel statistic* *. Transpunerea în grafice a datelor statistice constituie, în multe cazuri, un mijloc de studiu avantajos, deoarece oferă o vedere sintetică a fenomenelor studiate, înlesnește comparații între fenomene, pune în lumină caracteristicile lor (structuri, relații, tendințe, regularități) și eventual conduce la descoperirea de anomalii, erori sau omisiuni în datele originale. Unele grafice îndeplinesc și funcții de calcul statistic, cu ajutorul lor determinîndu-se mai simplu anumite valori statistice, pentru care nu se cere o precizie prea mare. În activitatea curentă de planificare și de previziune, metoda grafică se dovedește de asemenea un auxiliar prețios.

R.g. nu sînt însă destinate să înlocuiască tabelele statistice, ci să le completeze, făcîndu-le mai asimilabile. Folosul graficelor, deși însemnat, este limitat prin posibilitățile de reprezentare într-un singur plan. Nu se poate înfățișa, în general, decît un singur fenomen sau fenomene juxtapuse ; reprezentarea evoluției simultane a mai multor variabile legate se arată dificilă și adesea imposibilă. Chiar cu ajutorul reliefului, cu greu se pot reprezenta imagini grafice cu mai mult de trei dimensiuni.

Reprezentarea în coordonate rectangulare (ortogonale) Reprezentare în raport și proporțional cu două axe perpendiculare între ele (fig. 1).

Poziția unui punct P oarecare din plan se determină în raport cu *punctul de intersecție* sau *originea* celor două drepte, numite respectiv *axa x-ilor* și *axa y-ilor* (sau *axa absciselor* și *axa ordonatelor*). Distanța a socotită pe axa x-ilor se numește *abscisa* punctului P, distanța b pe axa y-ilor, *ordonata* sa ; împreună, cele două distanțe constituie *coordonantele* punctului P. Abscisele duse la dreapta și ordonatele duse deasupra originii sînt *pozitive*, celelalte coordonate sînt *negative*. În cea mai mare parte a graficelor statistice și abscisele și ordonatele sînt pozitive. Pentru a localiza un punct $P(x_i, y_i)$ este necesară gradarea axelor după o anumită scară. Se alege în mod arbitrar o unitate de lungime și se stabilește o corespondență între această unitate de lungime și unitățile relative

la x și y. Transportată pe o axă, o scară numerică se transformă în scară grafică. Scările grafice uzuale sînt : *scara aritmetică* și *scara logaritmică*.

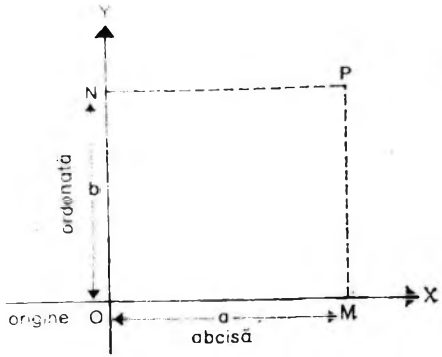


Fig. 1 Reprezentare în coordonate rectangulare

Scara aritmetică (*naturală*) traduce proporționalitatea între numerele (x_1, y_1) și lungimile absciselor și ordonatelor, iar *scara logaritmică*, proporționalitatea între logaritmii numerelor (x_1, y_1) și lungimile absciselor și ordonatelor.
Exemplu Scara verticală aritmetică (a) și logaritmică (b) (fig. 2).

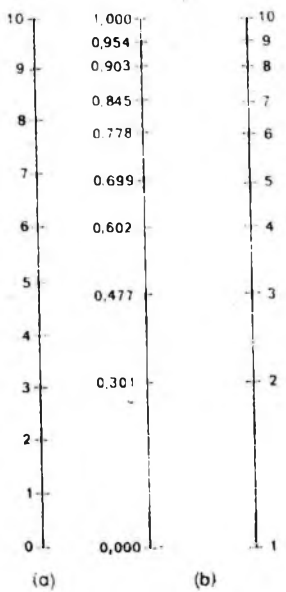


Fig. 2 Scară aritmetică și logaritmică

În cazul scării aritmetice, intervalul corespunzător unei unități rămîne același pe întreaga lungime a scării. La scara logaritmică, intervalele grafice, adică segmentele dintre punctele cotate, sînt inegale. Ca atare, scara aritmetică este o *scară uniformă* sau *scară liniară*, iar scara logaritmică, *scară neuniformă* sau *scară neliniară*. Scara logaritmică se stabilește mai întîi pentru numerele cuprinse între 1 și 10, adică, fiind aleasă unitatea de lungime, se înscriu punctele de diviziune la o distanță de origine măsurată succesiv prin log 2, log 3, log 4 ... etc., log 9, numere citite în tabela de logaritmi. Fiecare dintre diviziuni poate fi gradată în același mod : de exemplu, segmentul ale cărui extremități poartă gradările 2 și 3 va putea fi împărțit în 10 părți numerotate 2,1 ; 2,2 ; etc., punctele de diviziune găsindu-se la distanțe de origine respectiv egale cu log 2,1 ; log 2,2 ; ... ; log 2,9. Subdiviziunile se extind atît cît se dorește și atît cît ele rămîn ușor de descifrat. Prelungindu-se segmentul astfel gradat printr-un segment identic, se realizează scara logaritmică a numerelor de la

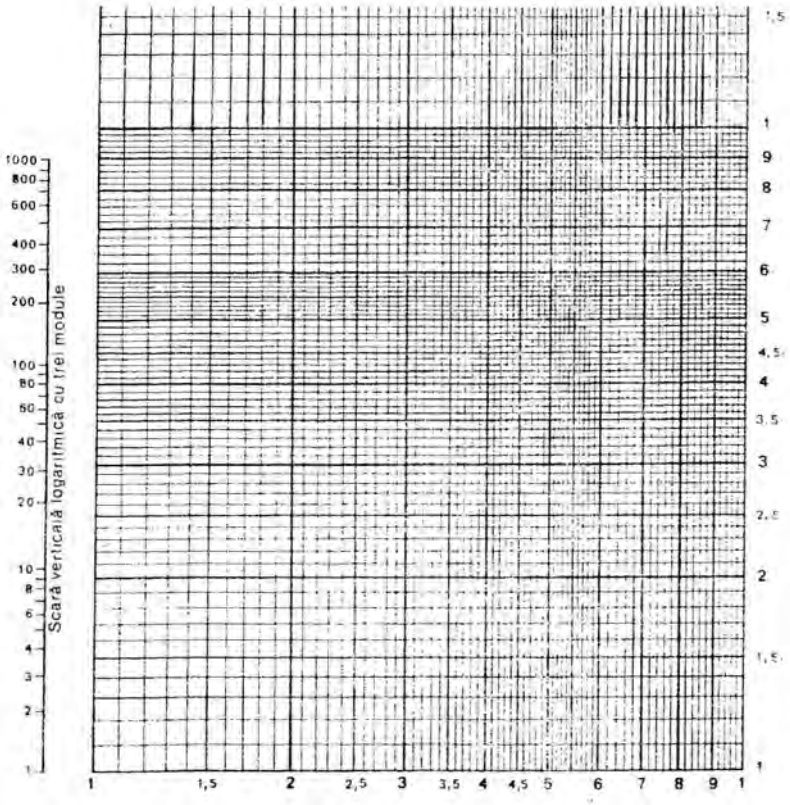


Fig. 3 Hîrtie logaritmică

10 la 100, cu gradarea 20, 30, 40 . . . , 90 și eventual cu gradări intermediare. O a doua translație va realiza scara logaritmică a numerelor de la 100 la 1 000 și așa mai departe. Amplitudinea translației se numește *modul logaritmic* (scară grafică a unei hîrtii logaritmice în care gradarea este limitată la o clasă de numerație zecimală, fie de la 1 la 10 sau de la 10 la 100 sau de la 100 la 1 000 etc.). O scară logaritmică include, după caz, 2, 3, 4 etc. module.

Diagrame logaritmice se pot construi fie pe hîrtie cadrilată obișnuită (în general *hîrtie milimetrică*), transpunînd logaritmii în coordonate aritmetice ca și numerele naturale, fie pe *hîrtie logaritmică (hîrtie log-log)* — hîrtie gradată în scări logaritmice, dar cu numerele naturale indicate pe scări. Mai frecvent se utilizează *hîrtia semilogaritmică*, hîrtie în care axa absciselor este gradată în scară aritmetică, iar axa ordonatelor, în scară logaritmică. Hîrtia logaritmică sau semilogaritmică preimprimată permite să se plaseze direct datele originale pe grafic, fără a fi nevoie să se determine în prealabil logaritmii lor (fig. 3 și 4).

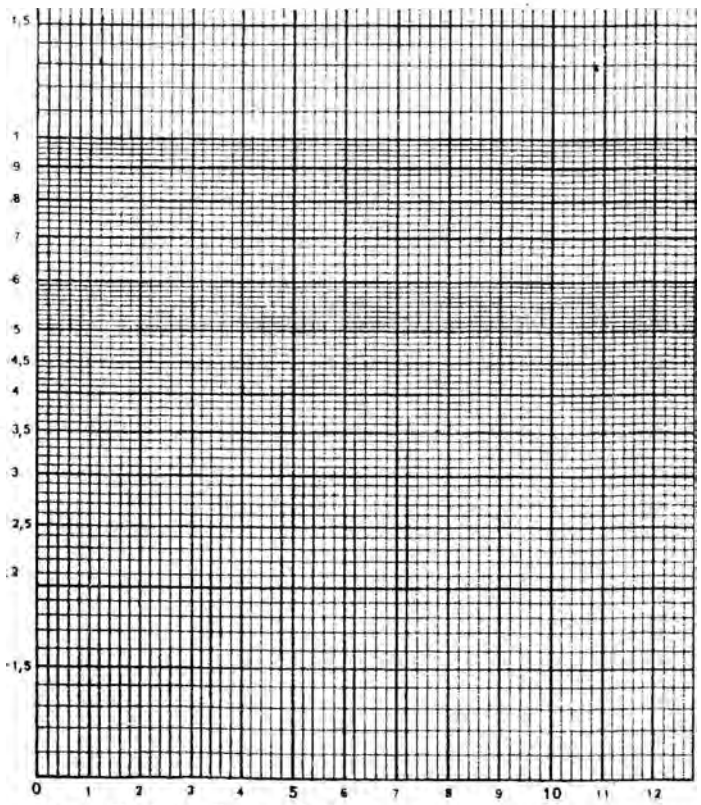


Fig. 4 Hîrtie semilogaritmică

Rețea probabilistă (hîrtie probabilistă). Hîrtie special gradată (simplu sau dublu logaritmică) pe care se reprezintă perechile de puncte (t_i, F_t, n), avînd ca rezultat liniarizarea funcției de repartiție. Se recunoaște astfel repartiția respectivă și se estimează parametrii ei fără să se apeleze la metode analitice.

Reprezentare în coordonate polare Reprezentare în coordonate raportate la o axă x și un pol o situat pe x. Coordonatele polare ale unui punct P, într-un plan, sînt : distanța $\rho = OP$ (raza vectoare) și θ = unghiul xOP (unghi polar). (fig. 5).

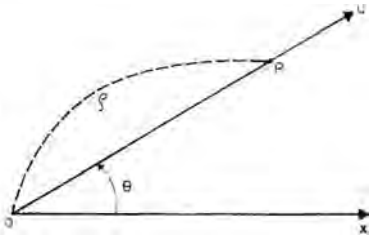


Fig. 5 Reprezentare în coordonate polare

Cu condiția ca să se fixeze un sens pozitiv de rotație pornind de la axa Ox (axă polară), corespondența astfel stabilită este biunivocă, adică oricărui punct din plan îi corespunde un cuplu unic (ρ, θ) și, reciproc, dacă se dă un cuplu (ρ, θ), acestuia îi corespunde un punct și unul singur în plan. Unghiul θ se exprimă în grade sau în radiani, iar raza vectoare, într-o unitate de lungime oarecare. Coordonatele polare sînt folosite în special pentru reprezentarea seriilor cronologice (date săptămînale, lunare, trimestriale sau anuale). $\overline{OP} = \rho$ este o lungime proporțională cu importanța fenomenului la un moment dat, θ intervalul de timp care separă două măsuri succesive, 2π reprezentînd unitatea de timp. Dacă 2π reprezintă anul, θ este egal cu $\pi/26$ pentru variațiile săptămînale, $\pi/6$ pentru variațiile lunare, $\pi/2$ pentru variațiile trimestriale.

DIAGRAMĂ ÎN BATOANE *(bare)*. Diagramă în care fiecărei valori a variabilei, înscrisă în abscisă, i se asociază o ordonată („baton", „bară") de lungime proporțională cu frecvența valorii. Lungimile batoanelor însumate fac deci Σf_t, dacă se consideră frecvențele absolute sau 1 (resp. 100), dacă se consideră frecvențele relative. Acest tip de diagramă este propriu variabilelor discrete (fig. 6).

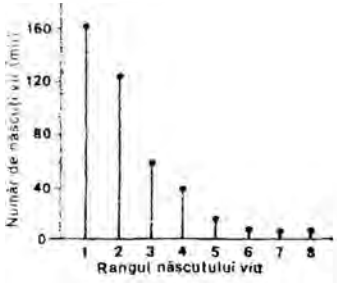


Fig. 6 Repartiția născuților vii după rangul născutului viu în R.S. România, 1979

Pentru compararea a două sau mai multe repartiții cu frecvențe absolute inegale se recurge la reprezentarea frecvențelor relative, modificîndu-se în consecință scara de-a lungul axei verticale.

DIAGRAMĂ CUMULATIVĂ. Pentru o repartiție discretă, diagramă construită „în scară" : se trasează, ca și în cazul *diagramei în batoane*, segmente de drepte de lungimi proporționale cu frecvențele, dar decalîndu-se progresiv

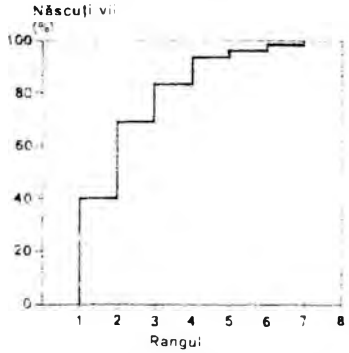


Fig. 7 Diagramă cumulativă a născuților vii după rangul născutului viu în R. S. România, 1979

în sus, astfel încît originea fiecărui segment să se afle la nivelul extremității segmentului precedent. Segmentele verticale se unesc prin segmente orizontale (fig. 7). Palierele orizontale ale diagramei au drept ordonate :

$$F_i = f_1 + f_2 + \ldots + f_i, \text{ respectiv } \tilde{F}_i = \tilde{f}_1 + \tilde{f}_2 + \ldots + \tilde{f}_i.$$

Înălțimea fiecărei trepte este proporțională cu o frecvență simplă (f_i, resp. $\tilde{f}_i$). Diagrama (fig. 7) prezintă frecvențele cumulate ascendente și arată, pentru fiecare rang al născutului viu, proporția născuților vii avînd un rang cel mult egal cu cel considerat. Reprezentarea frecvențelor cumulate în sens invers, după aceeași metodă, ar arăta, pentru fiecare rang al născutului viu, proporția născuților vii avînd un rang cel puțin egal cu cel considerat.

DIAGRAMĂ PRIN COLOANE *(Diagramă prin tuburi de orgă).* Reprezen-

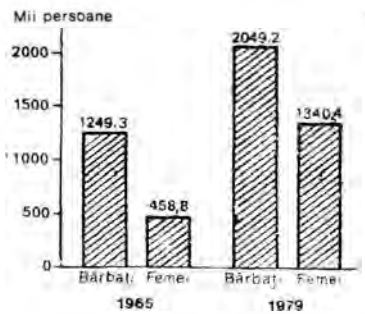


Fig. 8 Personalul ocupat în industria României, pe sexe

tare a unei serii sau repartiții simple prin dreptunghiuri avînd bazele egale și înălțimile (prin urmare și suprafețele) proporționale cu frecvențele (fig. 8).

Dreptunghiurile succesive pot fi alăturate sau separate (printr-un interval constant).

Acest tip de diagramă permite să i se alăture o scară care să pună și mai bine în evidență raporturile între frecvențe.

Uneori, dreptunghiurile sînt construite pe o *linie de bază* verticală și în acest caz diagrama este numită *diagramă prin benzi* (fig. 9).

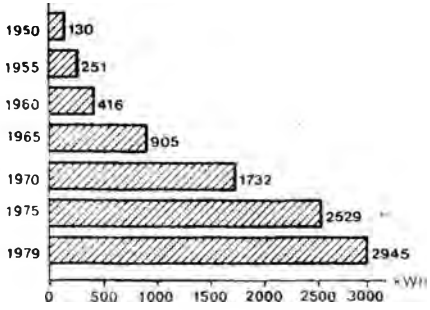


Fig. 9 Producția de energie electrică pe locuitor în R. S. România

Se poate reprezenta, prin dreptunghiuri, evoluția mai multor fenomene juxtapuse, haşurîndu-se în chip diferit (fig. 10).

Întrucît la diagramele descrise intră în calcul numai înălțimile dreptunghiurilor, ele sînt *diagrame cu o dimensiune* (*liniare*). Dreptunghiurile nu sînt decît „ordonate lărgite" pentru a fi mai vizibile. Ordonatele sînt reprezentate prin ariile dreptunghiurilor, dar se păstrează proprietatea proporționalității lungimilor (lungimea dreptunghiurilor), dîndu-se fiecărui dreptunghi aceeași lățime.

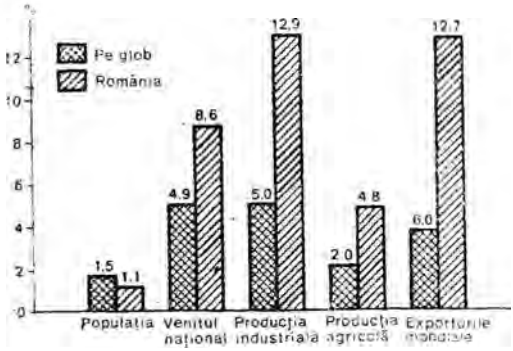


Fig. 10 Ritmurile medii mondiale și ale R. S. România în perioada 1950—1975

Dreptunghiurile sînt adoptate de asemenea pentru reprezentarea grafică a structurii populației sau eșantionului în raport cu variabila observată (fig. 11). Un dreptunghi se împarte în mai multe altele, înălțimea dreptunghiurilor divizionare variază proporțional cu mărimea ponderii.

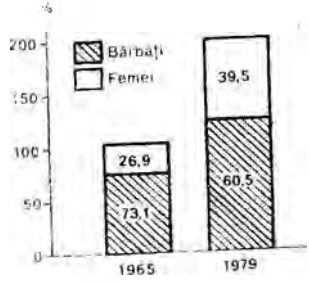
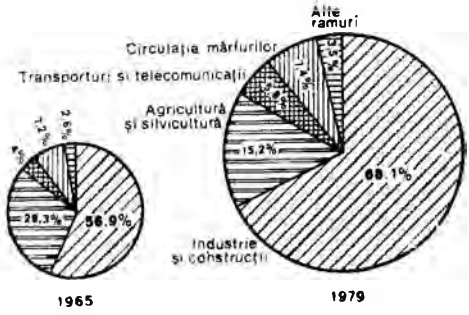


Fig. 11 Structura pe sexe a personalului ocupat în industria României

DIAGRAMĂ CIRCULARĂ *(în sectoare)*. Diagramă în care repartiția se reprezintă sub forma unui cerc divizat în sectoare avînd unghiurile la centru (deci și suprafețele) proporționale cu frecvențele (fig. 12).

Figl 12 Structura venitului național al R. S. România



În aplicație :

— pentru fiecare din cei doi ani s-a construit un cerc a cărui arie este proporțională cu mărimea venitului național (146 mld. lei în 1965 și 499 mld. lei în 1979) ; razele celor două cercuri (r_1 pentru 1965 și r_2 pentru 1979) sînt în raportul :

$$\frac{r_2}{r_1} = \sqrt{\frac{\pi r_2^2}{\pi r_1^2}} = \sqrt{\frac{499}{146}} = \sqrt{3,41} = 1,85 ;$$

— împărțind 360° proporțional cu ponderile deținute de ramurile economice considerate s-au aflat unghiurile la centru ale sectoarelor de cerc (3,6 × 56,9 ≈ ≈205° ; 3,6 × 29,3 ≈106° etc.).

V.TREBICI

O variantă a acestei tehnici constă în înlocuirea cercului printr-un semicerc, unghiul unui sector care reprezintă $p(\%)$ fiind egal cu $(p \times 180)/100$. Diagrama astfel construită se numeşte *diagramă semicirculară* (fig. 13).

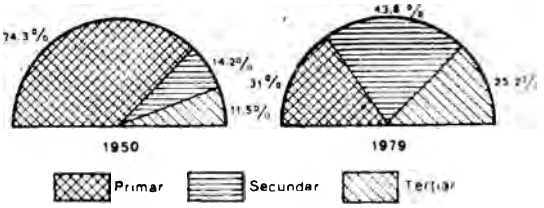


Fig. 13 Structura populaţiei ocupate a României pe sectoare de activitate

Pentru comparaţii în timp sau în spaţiu, diagrama circulară scoate în evidenţă deodată diferenţele în valoare absolută şi în valoare relativă. Diagrama semicirculară nu evidenţiază decît raportul părţilor cu întregul, importanţa lor relativă.

DIAGRAMĂ PRIN SUPRAFEŢE. Diagramă formată de obicei din pătrate sau cercuri avînd suprafeţele proporţionale cu valorile reprezentate (fig. 14).

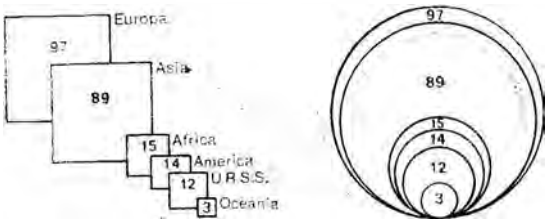


Fig. 14 Densitatea populaţiei pe continente (locuitor/kmp). 1978

Pentru construirea unei astfel de diagrame trebuie extrase rădăcinile pătrate ale valorilor seriei. Laturile pătratelor, razele cercurilor se iau proporţionale cu aceste rădăcini.

HISTORIOGRAMĂ *(cronodiagramă).* Reprezentare grafică a unei serii cronologice bazată pe sistemul coordonatelor rectangulare. Valorile succesive ale seriei sînt reprezentate de linia frîntă trasată prin punctele de coordonate y, t (fig. 15).

Pe axa absciselor s-au reprezentat „anii" (t), iar pe axa ordonatelor „valorile variabilei observate" (y). Fiecărei perechi de valori a seriei îi corespunde în planul axelor un punct: intersecţia dreptelor ridicate perpendicular pe axele de coordonate. Caracterul de „serie de stocuri" s-a marcat grafic prin plasarea punctului reprezentativ în dreptul datei de referinţă („sfîrşitul anului"). În caz de „serie de fluxuri", punctul reprezentativ va fi fixat în mijlocul segmentului corespunzător intervalului de timp de un an (fig. 16).

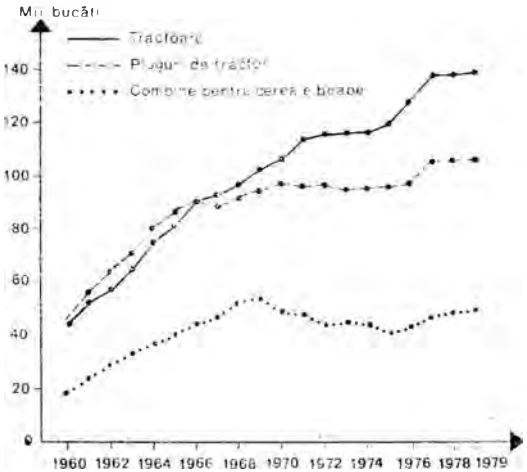


Fig. 15 Parcul de tractoare și mașini agricole principale (la sfîrșitul anului)

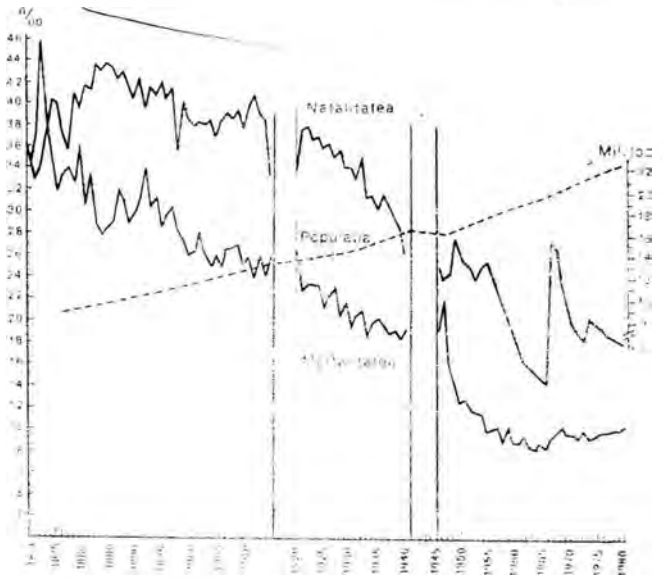


Fig. 16 Evoluția numărului populației României a natalității și mortalității, 1871—1980

H. sînt, mai mult decît alte grafice, instrumente de analiză și de cercetare. Ele înlesnesc studiul componentelor seriilor cronologice, mai ales a *trendului* și a *fluctuațiilor sezoniere* și permit compararea mai multor serii cronologice. De asemenea, o **h.** face posibilă urmărirea cu ușurință a evoluției diferenței între două fenomene, cînd aceasta are o semnificație concretă (diferența între costurile totale și costurile fixe, între venituri și cheltuieli, între importuri și exporturi etc.). În exemplul luat (fig. 16), diferențele între natalitate și mortalitate marchează sporul natural al populației, clar sesizat, în evoluția sa, la simpla lectură a graficului. În legătură cu construirea **h.**, problema cea mai importantă este aceea a raportului dintre unitatea de lungime adoptată pentru măsura timpului și cea care servește pentru scara ordonatelor. Fenomenele reflectate în seriile cronologice se caracterizează prin numeroase variații a căror înfățișare grafică depinde de alegerea scărilor. Prin lărgirea sau comprimarea scărilor, într-o anumită măsură, se pot atenua sau, dimpotrivă, amplifica variațiile. Dacă scările nu au fost judicios alese, există riscul unor interpretări false. Se reține principiul general după care o variație recunoscută ca interesantă trebuie să fie dezvăluită prin **h.** Fig. 17 arată cum alegerea scărilor poate influența judecata cititorului asupra amplitudinii variației. Institutul Internațional de Statistică a stabilit cu prilejul celei de a VIII-a sesiuni, pe baza raportului prezentat de Jacques Bertillon, că unitățile trebuie astfel alese încît alura tangențială a liniei din **h.** să fie reprezentată de o dreaptă înclinată la 45° față de axe (I.I.S., VIII e session, J. Bertillon : *Propositions relatives à l'uniformité à apporter dans l'établissement des graphiques*, p. 313, apud N. Georgescu-Roegen, *op. cit.*, p. 57).

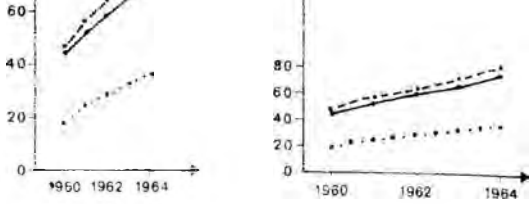


Fig. 17 Historiograme cu scări diferite

Tot pentru a nu avea o impresie falsă despre dimensiunile variațiilor fenomenului reprezentat, este necesar ca întotdeauna scara verticală să pornească de la zero. Cînd ordinul de mărime al datelor din serie depășește cu mult nivelul zero, se întrerupe scara, așa cum se vede în fig. 18.

Prin întreruperea operată se evită un gol prea mare pe grafic și se permite alegerea unei scări care să pună mai bine în evidență variațiile fenomenului descris.

DIAGRAMĂ SEMILOGARITMICĂ. Diagramă care comportă o scară logaritmică și o scară aritmetică.

În general, **d.s.** se aplică la seriile cronologice și se construiește cu scara logaritmică pe axa ordonatelor (fig. 19).

Utilizarea axei ordonatelor ca scară logaritmică înseamnă să se substituie funcția $Y = \log y$ funcției y. Valorile y_i sînt ordonatele nominale, iar Y_i (logaritmii zecimali ai lui y_i), ordonatele reale. *(→ Scară logaritmică)*. Pe diagramă se înscriu ordonatele nominale y_i (valorile exporturilor), ceea ce simplifică lectura.

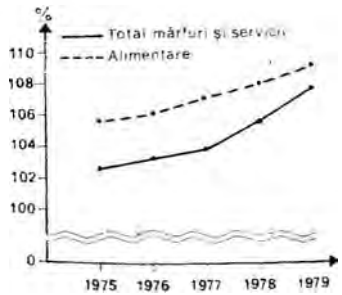


Fig. 18 Indicii prețurilor (1970 = 100)

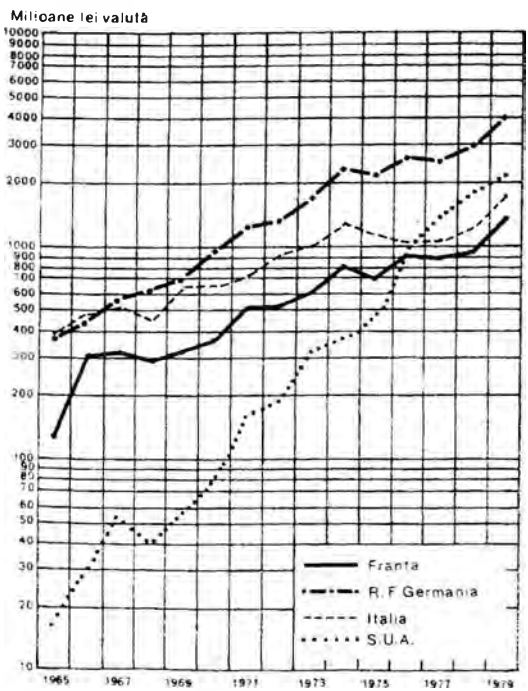


Fig. 19 Exporturile României în țări capitaliste dezvoltate, 1965—1979

V.TREBICI

Axa absciselor, gradată aritmetic, poartă scara timpului (anii). Paralele la axe, duse prin punctele gradărilor, determină o grilă care ușurează plasarea punctelor și citirea valorilor.

D.s. are citeva *proprietăți* de utilitate practică.

1. *Ordonatelor nominale în progresie geometrică le corespund (pe diagramă) ordonate reale în progresie aritmetică.*

Astfel, notăm corespondența :

$y \quad a \quad aq \quad aq^2 \quad aq^3 \ldots$

$Y \quad \log a \ (\log a + \log q) \ (\log a + 2 \log q) \ (\log a + 3 \log q)$

2. *Dacă raportul dintre două valori ale lui y este egal cu raportul dintre două alte valori ale sale, diferența ordonatelor reale ale primelor două este egală cu aceea a ordonatelor reale ale ultimelor două.*

De exemplu, egalitatea

$$\frac{y_2}{y_1} = \frac{y_4}{y_3}$$

se traduce prin

$$\log \frac{y_2}{y_1} = \log \frac{y_4}{y_3}$$

sau

$$\log y_2 - \log y_1 = \log y_4 - \log y_3,$$

adică

$$Y_2 - Y_1 = Y_4 - Y_3.$$

Pe axa ordonatelor, lungimile $(Y_2 - Y_1)$ și $(Y_4 - Y_3)$ sînt egale.

3. *Pentru o aceeași variație a lui x, două segmente sînt paralele, dacă rapoartele ordonatelor nominale ale punctelor pe care le unesc sînt egale.*

Într-o diagramă cu scări aritmetice, dreapta care unește două puncte $A(x_1y_1)$ și $B(x_2y_2)$ are drept coeficient director :

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

și două drepte cu același coeficient director sînt paralele.

Pe **d.s.**, coeficientul director real este :

$$m' = \frac{Y_2 - Y_1}{x_2 - x_1} = \frac{\log \dfrac{y_2}{y_1}}{x_2 - x_1}$$

De aceea se spune că **d.s.** păstrează rapoartele.

D.s. devine deci utilă cînd se studiază *variațiile relative*, adică rapoartele de mărimi. Prin reprezentarea grafică, de exemplu, a unei serii cronologice se ur-

mărește mai ales *tendința generală* de evoluție a fenomenului și *oscilațiile în jurul acestei tendințe*. Importanța unei schimbări absolute date depinde de mărimea bazei la care se raportează schimbarea. Ca atare, pentru reprezentarea seriei cronologice, **d.s.** este potrivită, întrucît, într-o asemenea diagramă, distanțele verticale egale pe axa ordonatelor reprezintă variații procentuale egale, în timp ce pe o diagramă aritmetică de tip obișnuit ele reprezintă variații absolute egale.

D.s. permite determinarea raportului între două valori ale variabilei și implicit rata de variație, precum și compararea creșterilor (sau descreșterilor) relative a mai multor variabile.

Pe grafic, lungimea ΔY este egală cu $\log \frac{y_2}{y_1}$. Prin urmare, este suficient să se ducă lungimea ΔY pe scară pornind de la origine și să se citească indexul corespunzător pentru a se obține valoarea raportului $\frac{y_2}{y_1}$ (fig. 20).

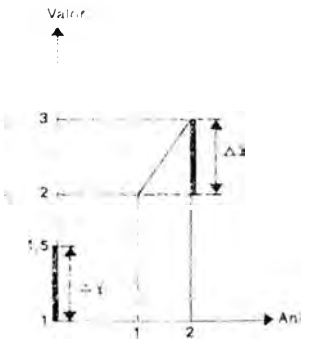


Fig. 29 Reprezentarea raportului $\frac{y_2}{y_1}$

Presupunînd

$y_2 = 3$; $y_1 = 2$

și reportînd pe scară lungimea ΔY, cu începere de la origine, rezultă

$$\frac{y_2}{y_1} = 1{,}5,$$

de unde, rata de creștere egală cu 0,5 sau 50 %.

Avantajul rețelei semilogaritmice în compararea creșterilor relative a două sau mai multe variabile apare clar din confruntarea graficelor 21 a și 21 b, care reprezintă aceleași serii de date.

Panta lui A este mai mare decît cea a lui B și a lui C; panta lui C este cea mai mică. A are, așadar, o creștere relativă mai intensă decît B și C. Graficul pe rețea aritmetică (fig. 21 a) nu permite formularea unei astfel de judecăți (remarcăm îndeosebi că variațiile lui A, pe graficul cu scară aritmetică, sînt greu de citit, datorită scării utilizate).

O modalitate de a măsura procentul de creştere sau de descreştere, reprezentat în d.s. printr-o distanţă verticală dată, ca şi raportul procentual între o serie cronologică şi o alta constă în construirea scărilor corespunzătoare pentru un modul (fig. 22).

Diviziunile pe scară (a), citită de jos în sus, marchează creşterile procentuale, care se întind de la 0 la 900% (începutul şi sfîrşitul modulului; distanţa de la 1

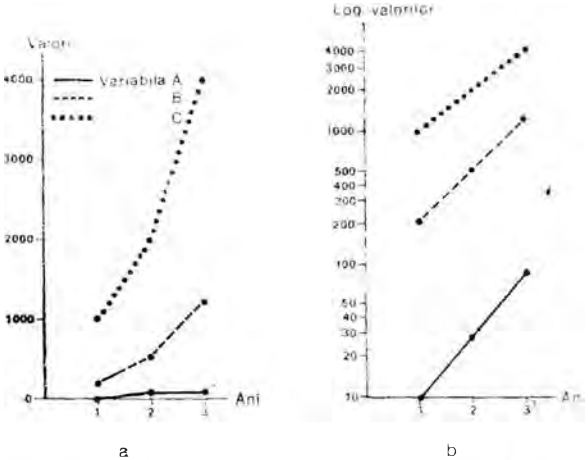


Fig. 21 Diagrama aritmetică (a) şi semilogaritmică (b) a aceleiaşi serii statistice

la 2 reprezintă o creştere de 100%, de la 1 la 3 o creştere de 200%, ..., de la 1 la 10 o creştere de 900%; aceleaşi creşteri sînt şi în cazul celorlalte module, adică de la 10 la 100, de la 100 la 1000, de la 1000 la 10 000 etc.).

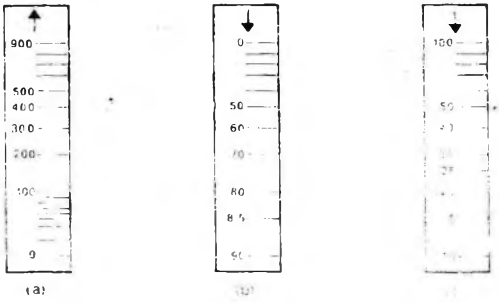


Fig. 22 Scări în procente: a) scara creşterilor; b) scara descreşterilor; c) scara comparaţiilor

Pentru a determina cu ajutorul scării (a), de exemplu, procentul de creştere a exportului ţării noastre în R. F. Germania, între anii 1977 şi 1979, (fig. 19) măsurăm pe o verticală distanţele între punctele care înfăţişează aceşti ani;

așezînd această distanță pe scară, aflăm o creștere de aproape 60%. Pe scara descreșterilor (b), citită de sus în jos, distanța de la o diviziune la alta reprezintă 10%. Astfel, de la diviziunea notată cu 100 la diviziunea notată cu 90, de la diviziunea notată cu 90 la acea notată cu 80 etc. sînt cîte 10%; descreșterea de la punctul notat cu 0 pînă la ultimul punct al scării este de 90% (la fel în celelalte două module). Scara (b) se folosește ca și scara (a): se măsoară pe verticală distanța între două puncte și segmentul determină descreșterea procentuală pe scara (b). De exemplu, pentru anii 1967 și 1968, se găsește o descreștere a exporturilor în S.U.A. egală cu 25%.

Cu ajutorul scării de comparație (c) se poate determina, în exemplul dat, raportul procentual între exportul în o țară și exportul în o altă țară. Dacă vrem să cunoaștem, de exemplu, raportul procentual între exportul în Franța și exportul în R. F. Germania în anul 1979, măsurăm distanța între punctele corespunzătoare ale diagramei și apoi o transpunem pe scara de comparație, socotind de sus în jos; găsim că exportul în Franța a reprezentat 30% din exportul în R. F. Germania.

Pentru aprecierea ratei de variație se poate de asemenea utiliza un *indicator de pantă* — un desen construit pe hîrtie transparentă, comportînd un fascicul de drepte ale căror pante sînt indicate în procente (fig. 23).

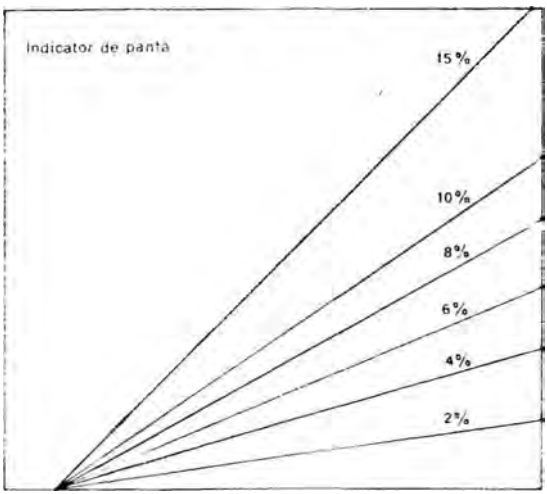


Fig. 23 Indicator de pantă

Este știut că panta crește sau descrește cu rata de variație. Examinarea diverselor pante ale aceleiași diagrame informează deci despre variațiile ratei. În construirea indicatorului de pantă, s-ar putea, teoretic, determina creșterea ordonatei referitoare la pantă pentru un interval Δx egal cu unitatea, dar o astfel de construcție ar fi adesea imprecisă. Este preferabil să se ia un interval Δx cît mai mare și care să conțină un număr comod de unități. În cazul tratat

(fig. 24) s-a luat $\Delta x = 30$ ani. Propunîndu-ne să construim, de exemplu, ordonate referitoare la panta de 2%, avînd :
$r = 0.02$ și $1 + r = 1,02$
(r este rata de creștere anuală),

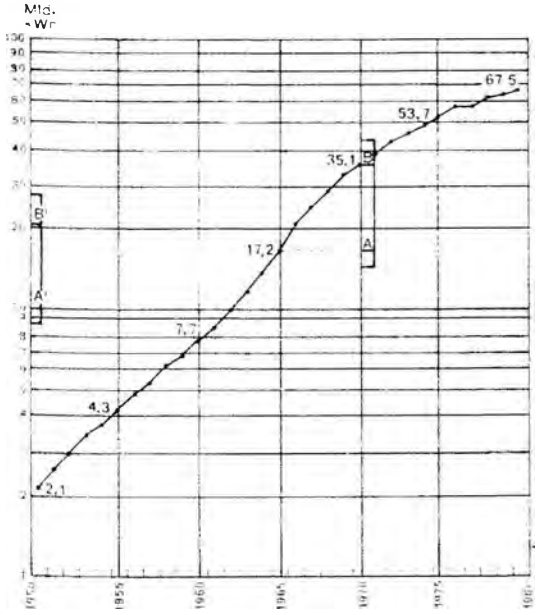


Fig. 24 Producția de energie electrică în România

atunci

$$\frac{\Delta y}{\Delta x} = \log\ (1 + r) = \log\ 1,02.$$

În tabela de logaritmi se găsește :
log 1,02 = 0.00860.
de unde

$$Y = 30 \times 0,00860 = 0,258.$$

Această valoare a lui ΔY corespunde modulului 1 al scării logaritmice, dar modulul fiind 5 cm (fig. 24), ordonata $\Delta Y = 5 \times 30 \times 0,00860 = 1,29$ cm.
Ordonata relativă la panta de 4% (log 1,04 = 0,01703) va fi :
$5 \times 30 \times 0,01703 = 2,55$ cm ș.a.m.d.
Aplicînd indicatorul de pantă pe diagrama evoluției producției de energie electrică a României în decursul a 30 de ani, se identifică pentru perioada 1951—1970 un ritm mediu anual de creștere egal cu 15%.

În cincinalul 1971—1975, ritmul mediu anual a fost de circa 9%, iar în ultimul cincinal 4,7%.

D.s. este utilizată nu numai pentru a evidenția proporțiile, ci și pentru a reproduce *gama întinsă a valorilor unei variabile.* Un același interval ΔY pe o scară funcțională logaritmică reprezintă un domeniu de variație cu atît mai mare a funcției y studiate cu cît acest interval este mai îndepărtat de origine. Considerînd, de exemplu, evoluția lui y de la 1 la 10000 și modulul scării logaritmice 5 cm, $\Delta Y = 5$ cm reprezintă:

9	unități	de la	1	la	10
90	,,	,,	10	,,	100
900	,,	,,	100	,,	1000
9000	,,	,,	1000	,,	10000.

Cu scara 0,5 cm = 1 unitate, ar trebui o lungime de 5000 cm sau 50 m pe scara aritmetică pentru a reprezenta variația lui y de la 1 la 10000, în timp ce pe scara logaritmică trebuie numai 20 cm (4 module × 5 cm) pentru a reprezenta această variație.

Grație distorsiunii pe care o introduce (dilatarea relativă a zonelor de valori mari în raport cu zonele de valori mici), **d.s.** permite un studiu fin al fenomenului reprezentat. La **d.s.** se recurge în cazul *variațiilor exponențiale.* Dacă reprezentarea grafică pe scară aritmetică evidențiază o curbă care sugerează o funcție exponențială, expresia sa va fi de forma

$$y = ab^x$$

(cu a și b constante și x variabilă).
Scriem egalitatea logaritmilor celor doi membri:

$$\log y = \log (a \cdot b^x) = \log a + x \log b$$

Întrucît $\log a$ și $\log b$ sînt constante ca și a și b rezultă că $\log y$ sau Y este o funcție liniară de x.

Se obține o dreaptă pe graficul semilogaritmic. Așadar, un grafic semilogaritmic permite o *anamorfoză* (transformare geometrică a unei curbe) utilă: *curbele exponențiale sînt transformate în drepte.*

DIAGRAMĂ LOGARITMICĂ (diagramă log-log). Diagramă în coordonate dublu logaritmice.
Utilizarea a două axe gradate în scări logaritmice prezintă interes în materie de *ajustare* *. Rețeaua logaritmică transformă în dreaptă, prin anamorfoză, o funcție putere de forma
$y = ax^b$,
dar, în această transformare, trebuie să se treacă nu numai de la y la $\log y$, ci și de la x la $\log x$. Se obține, într-adevăr:
$\log y = \log a + b \log x$.

Exemplu. Relația între producția globală industrială (y) și fondurile fixe industriale (x), considerată de forma $y = ax^b$ (fig. 25).

Punctele definite de cuplurile ($\log x$, $\log y$) — corespunzătoare datelor înregistrate în 35 județe — se situează în jurul unei drepte care poate fi trasată vizual sau printr-o altă metodă. Dreapta pare să ajusteze cel mai bine *norul statistic** figurat.

O aplicație clasică a graficului logaritmic este aceea întreprinsă de economistul italian V. Pareto la studiul repartiției veniturilor. Pareto a remarcat că „ajustarea" cea mai satisfăcătoare a unei serii de puncte de coordonate reprezentînd

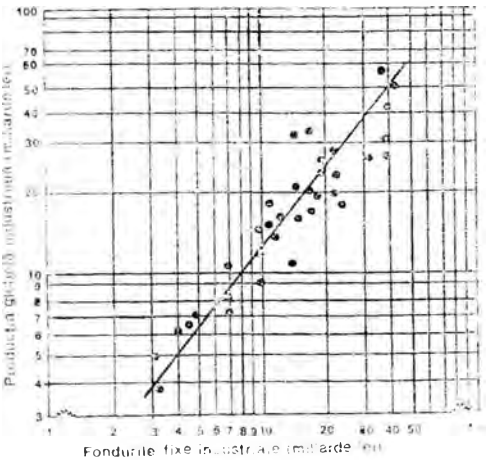


Fig. 25 Producția globală industrială și fondurile fixe industriale, 1979

respectiv valorile veniturilor x și numărul y al veniturilor superioare lui x (frecvențe cumulate) pare să fie de forma hiperbolei. Adoptînd coordonate dublu logaritmice, el a obținut o ajustare liniară. Autorul dă atunci curbei de frecvențe cumulate o expresie analitică :

$\log y = \log a - b \log x$ (pantă negativă)

și pune astfel în evidență repartiția asimetrică a veniturilor. Fig. 26 ilustrează constatarea lui Pareto.

Punctele M_1, M_2, ..., M_8 sînt foarte apropiate de o dreaptă, care se poate trasa, fără eroare sensibilă, prin punctele M_2 și M_7. Dacă se asimilează seria cu această dreaptă și se calculează valorile constantelor din ecuația $\log y = \log a - b \log x$, rezultă : $\log y = 16{,}1786 - 2{,}4099$.

Această relație care leagă y și x, caracteristică aproape tuturor repartițiilor de venituri, a primit numele de „legea lui Pareto".

D.I. se aplică frecvent la studiul *elasticităților* (de exemplu, elasticitatea diferitelor categorii de cheltuieli din bugetul de familie : alimentație, îmbrăcăminte, locuință, igienă și sănătate etc. în funcție de cheltuielile totale).

Natural, toate proprietățile specifice scărilor funcționale logaritmice sînt aici valabile. În particular, **d.l.** va fi indicată atunci cînd y și x prezintă amplitudini de variație cu atît mai mari cu cît se depărtează de origine (proprietate de „tasare" dînd o precizie relativ constantă).

HISTOGRAMĂ. Grafic al repartiției pe clase de valori format dintr-o succesiune de dreptunghiuri alăturate, avînd suprafețele proporționale cu frecvențele în clase. Termen introdus de K. Pearson (gr. *histos* — țesut ; *grama* — scriere, desen).

H. se construiește în coordonate rectangulare (fig. 27); pe axa absciselor se trec valorile variabilei care delimitează clasele, iar în ordonate valorile frecven-

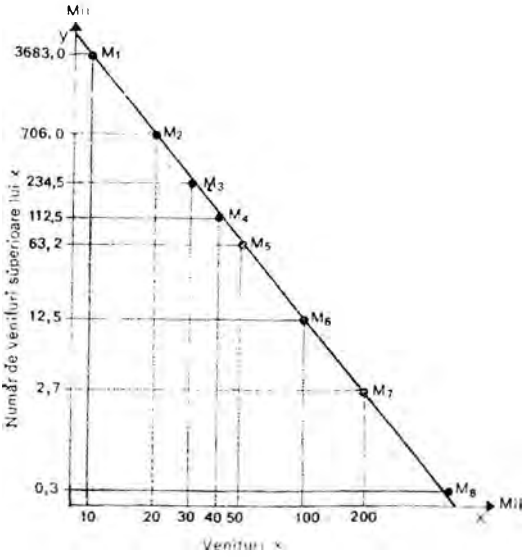


Fig. 26 Curba lui Pareto privind repartiția veniturilor

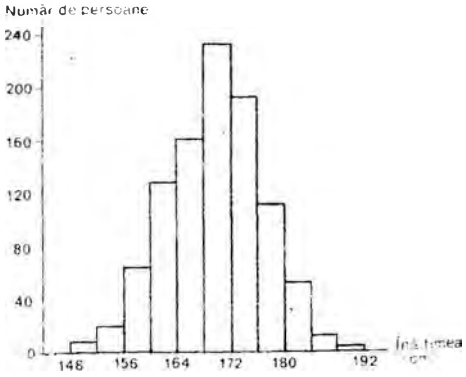


Fig. 27 Histograma repartiției a 988 de tineri în vîrstă de 19 ani după înălțime

telor de clasă; punctele reprezentînd limitele fiecărui interval de clasă se unesc prin linii orizontale. Suprafața închisă de linia în scară ce mărginește **h.** și de axa absciselor este proporțională cu suma frecvențelor.

În aplicație, s-a reprezentat prin h. repartiția populației observate făcută pe clase de talii cu amplitudine constantă (148—152 cm, 152—156 cm, ..., 188—192 cm) și atunci s-au construit, pe intervalele scării egal spațiate, dreptunghiuri de înălțimi proporționale cu frecvențele repartiției (7 persoane, 20 persoane, ..., 5 persoane). În cazul contrar — al repartiției pe clase de amplitudine variabilă — relația de proporționalitate trebuie să fie aplicată la suprafețele dreptunghiurilor, nu la înălțimile lor. De aceea, se operează o anumită transformare a repartiției înainte de a fi reprezentată grafic.

Exemplu. Repartiția personalului ocupat în economia României după retribuția netă lunară, martie 1979.

Tabelul 4

Retribuția (lei)	Numărul personalului (%)	Frecvențe reduse la intervalul de 100 lei
Pînă la 1 300	4,0	4,0
1 301—1 500	9,2	4,6
1 501—1 700	14,8	7,4
1 701—2 000	26,1	8,7
2 001—2 500	27,5	5,5
Peste 2 500	18,4	3,7
Total	100,0	—

Sursa: *Anuarul statistic al R. S. România 1980*, p. 124.

Clasele de retribuție nu sînt toate egale între ele, iar prima și ultima clasă sînt nedeterminate. Pentru a înlătura mai întîi nedeterminarea, convenim să închidem primul interval la 1 200 lei, nivelul minim al retribuției la data observării și ultimul interval la 3 000 lei, adica egal cu intervalul precedent. În ceea ce

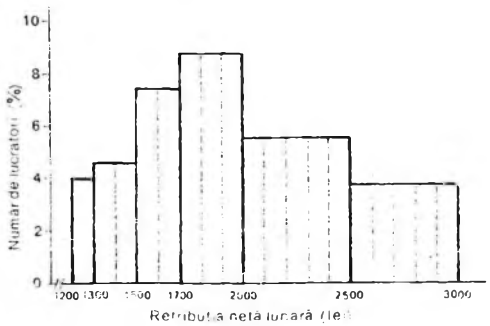


Fig. 28 Histograma repartiției personalului din economia națională după retribuția netă lunară realizată, martie 1979

privește inegalitatea claselor, soluția constă în a alege un interval de clasă unitar (de exemplu, primul interval = 100 lei), a cărui amplitudine să fie divizor comun tuturor celorlalte intervale.

Construim **h**. (fig. 28) ridicînd dreptunghiuri care au bazele proporționale cu frecvențele pe interval unitar (reduse la intervalul de 100 lei).

Uneori, poate fi utilă trasarea **h**, pe hîrtie semilogaritmică, cu scară logaritmică în abscise și liniară în ordonate, ca în fig. 29. Tabelul 5 conține elementele necesare pentru a construi h. repartiției log x. S-a presupus că valoarea inferioară a primului interval de clasă este 50, iar valoarea superioară a ultimului interval 20 000, acesta divizîndu-se în intervalele : 5 000—10 000 și 10 000—20 000.

Frecvența unitară (col. 5) — calculată ca raport între frecvență (în speță, frecvență relativă, col. 2) și mărimea intervalului de clasă în logaritmi zecimali (col. 4) — constituie, la scara aleasă, înălțimea dreptunghiului **h**. pentru clasa corespunzătoare. În acest mod se asigură proporționalitatea suprafeței dreptunghiului cu frecvența de clasă.

Tabelul 5

Număr de muncitori	Număr de întreprinderi (%)	Logaritmi zecimali ai limitelor de clasă	Intervale de clasă în logaritmi zecimali	Frecvențe unitare
(1)	(2)	(3)	(4)	(5)
50		1,69897		
	8,8		0,60206	14,6
200		2,30103		
	19,0		0,39794	47,8
500		2,69897		
	23,7		0,30103	78,7
1000		3,00000		
	22,6		0,30103	75,1
2000		3,30103		
	12,0		0,17609	68,2
3000		3,47712		
	8,3		0,22185	37,4
5000		3,69897		
	4,2		0,30103	14,0
10000		4,00000		
	1,4		0,30103	4,7
20000		4,30103		

Diagramă diferențială. Reprezentare în scară logaritmică în abscise, în care unele repartiții asimetrice de stînga devin aproximativ simetrice. Histograma reprezentînd în ordonate frecvențele $f(x)$ în funcție de valorile variabilei x este uneori numită **d.d**. (pentru că face să apară diferențele între ordonate).

POLIGON AL FRECVENȚELOR. Grafic al repartiției pe clase de valori definit de o linie frîntă care unește perpendicularele, proporționale cu frecvențele, ridicate din centrele claselor înscrise pe axa absciselor (fig. 30 a). Dacă s-a construit *histograma* repartiției, **p.f.** se obține unind succesiv, prin segmente

de dreaptă, mijloacele bazelor superioare ale dreptunghiurilor (fig. 30 b) Suprafața cuprinsă între linia poligonală și axa absciselor este echivalentă cu suprafața histogramei.

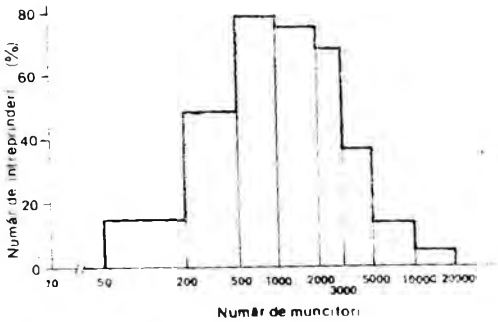


Fig. 29 Histograma repartiției întreprinderilor industriale după numărul muncitorilor

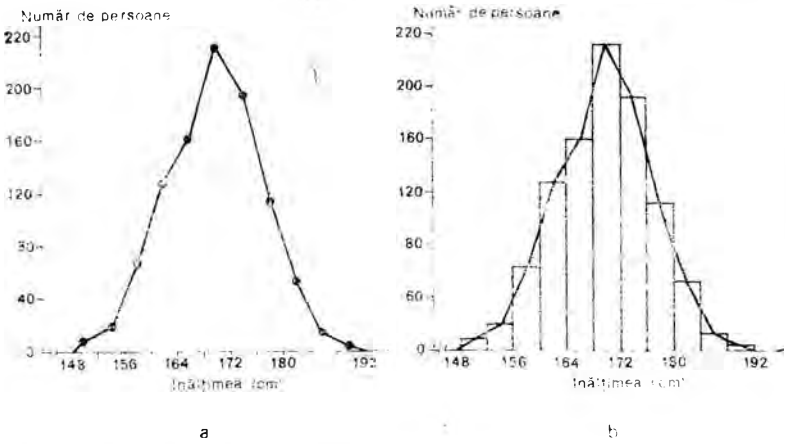


Fig. 30 a) Poligonul frecvențelor; b) Histograma și poligonul frecvențelor.

Exemplu: Repartiția a 988 tineri în vîrstă de 19 ani, după înălțime. (→*Histogramă*).

Abscise:	150	154	158	162	166	170	174	178	182	186	190
Ordonate:	7	20	65	129	161	233	191	113	53	11	5

V.TREBICI

În cazul repartiției pe intervale neegale, ordonatele punctelor înscrise în plan se iau proporționale cu frecvențele reduse la intervalul ales ca unitate *(→Histogramă)*.

P.f. se utilizează frecvent pentru reprezentarea simultană a două sau mai multe repartiții (fig. 31).

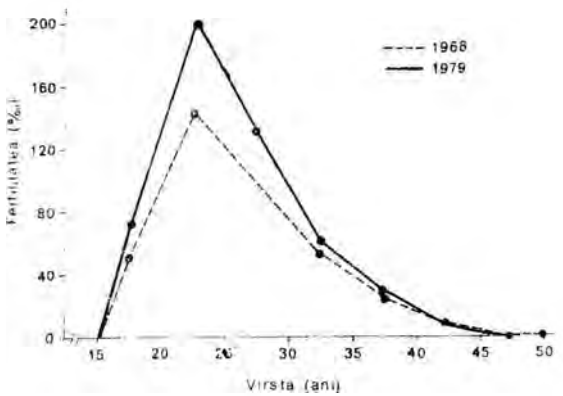


Fig. 31 Fertilitatea populației feminine pe grupe cincinale de vîrstă, 1966 și 1979

Dacă frecvențele totale ale repartițiilor de reprezentat nu sînt identice, trebuie să se ducă în ordonate frecvențele relative și dacă intervalul de clasă nu este același pentru diferitele repartiții, se vor utiliza frecvențele unitare. Prin **p.f.** poate fi reprezentată grafic și repartiția frecvențelor cumulate. Se iau în abscise *limitele superioare* ale claselor și în ordonate *frecvențele cumulate*, absolute sau relative.

Diagramă integrală (cumulativă). Linia poligonală obținută prin unirea punctelor $[x, F(x)]$ corespunzătoare extremităților claselor succesive. Se mai numește *poligonul frecvențelor cumulate* (fig. 32).

Pentru repartiția luată drept exemplu s-au construit : A) poligonul frecvențelor cumulate ascendente și B) poligonul frecvențelor cumulate descendente. Cum repartiția este continuă, diagrama cumulativă ia forma unei diagrame de puncte (și nu de *diagramă cumulativă în scară*, ca pentru o repartiție de variabilă discretă). Fiecărui punct de pe axa abscielor îi corespunde pe linia poligonală un punct a cărui ordonată indică, în valoare absolută sau relativă :
— în primul caz (A), numărul de observații inferioare abscisei considerate (numărul de persoane avînd înălțimea sub x cm) ;

— în al doilea caz (B), numărul de observații egale sau superioare abscisei considerate (numărul de persoane de înălțime egală sau mai mare decît x cm). De observat că ordonatele reprezentînd frecvențele retrocumulate s-au trasat în dreptul limitei inferioare a claselor.

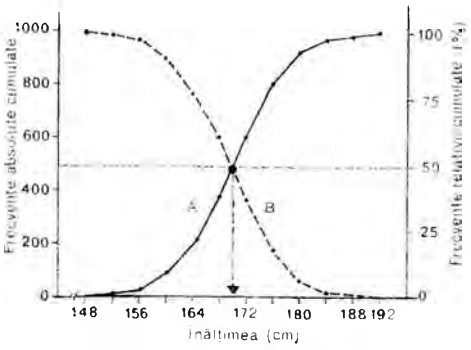


Fig. 32 Repartiția cumulativă a 988 tineri în vîrstă de 19 ani, după înălțime

Cele două diagrame sînt simetrice în raport cu dreapta paralelă la axa absciselor de ordonată $\frac{1}{2}$ din frecvența totală și, ca atare, se construiește numai una din ele, obișnuit diagrama de tipul A.

Cu ajutorul poligonului frecvențelor cumulate se pot obține informații interesante despre repartiția studiată. De exemplu, 25% din tinerii înregistrați au înălțimea peste 1,75 m. Sînt tot atîția tineri cu înălțimea sub 1,70 m cîți tineri sînt cu înălțimea de peste 1,70 m (proporție care corespunde ordonatei 50% determinată prin interpolare din ochi pe fig. 32).

Diagrama cumulativă, reprezentînd în ordonate sumele succesive ale frecvențelor simple este uneori numită pentru acest motiv *diagrama integrală*.

Dacă se asociază *diagrama diferențială* diagramei integrale, cum se arată în fig. 33, se realizează corespondența:

$x_0 \rightarrow f(x_0) \rightarrow F(x_0)$,

unde

x_0 reprezintă o valoare a variabilei;

$f(x_0)$ — frecvența relativă a valorii x_0;

$F(x_0)$ — frecvența relativă cumulată pînă la x_0.

$F(x_0)$ ordonata diagramei integrale pentru valoarea x_0 este egală cu aria diagramei diferențiale care corespunde părții situate la stînga ordonatei în x_0. Această corespondență se folosește adesea pentru a preciza caracteristicile repartițiilor studiate.

CURBĂ A FRECVENȚELOR *(curbă de repartiție)*. Curbă continuă către care tinde *histograma* (sau *poligonul frecvențelor*) cînd intervalul de clasă este suficient de mic și numărul de observații relativ mare (fig. 34).

Se înlocuiește conturul histogramei cu o curbă lină care limitează o arie totală egală cu cea a histogramei. Părțile din aria dreptunghiurilor histogramei lăsate de **c.f.** în afara ariei ei se compensează cu părțile adăugate, care nu fac parte din aria dreptunghiurilor. Cu cît este mai mare numărul de clase și acel al observațiilor (frecvențelor) în fiecare clasă cu atît mai regulat și lin devine conturul

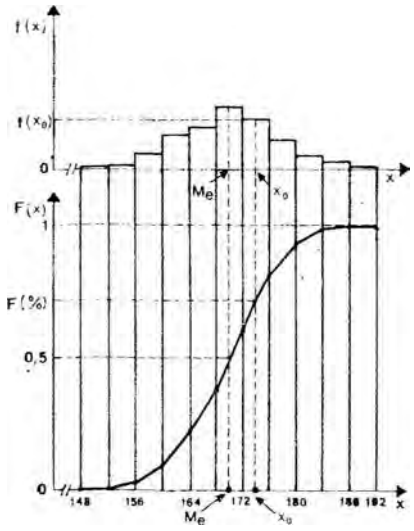


Fig. 33 Corespondența dintre diagrama diferențială și diagrama integrală

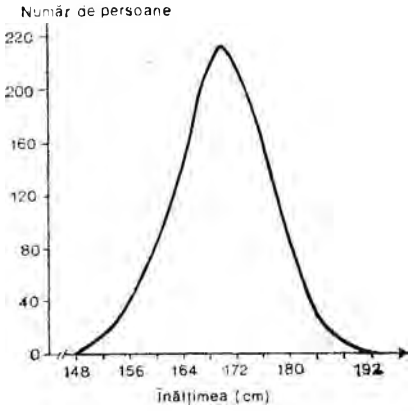


Fig. 34 Curba frecvențelor : repartiția a 988 tineri în vîrstă de 19 ani, după înălțime

histogramei, deoarece iregularitățile accidentale cauzate de numărul mic de observații sînt eliminate. „Netezirea" curbei se bazează, din punct de vedere logic, pe presupunerea că variabila pe care o reprezintă este *continuă*. Fig. 35 ilustrează convergența histogramei către e.f. cînd volumul observațiilor crește indefinit.

Histogramele succesive se referă la eșantioane de piese fabricate de o mașină, înregistrate după un anumit parametru, al căror volum sporește treptat. Se

ajunge la constatarea : **c.f.** unei populații infinit de mare este limita către care tinde histograma (sau poligonul frecvențelor) privind un eșantion din acea populație, cînd volumul eșantionului crește indefinit.
Semnificația **c.f.** este capitală : se consideră că ea reprezintă repartiția ideală

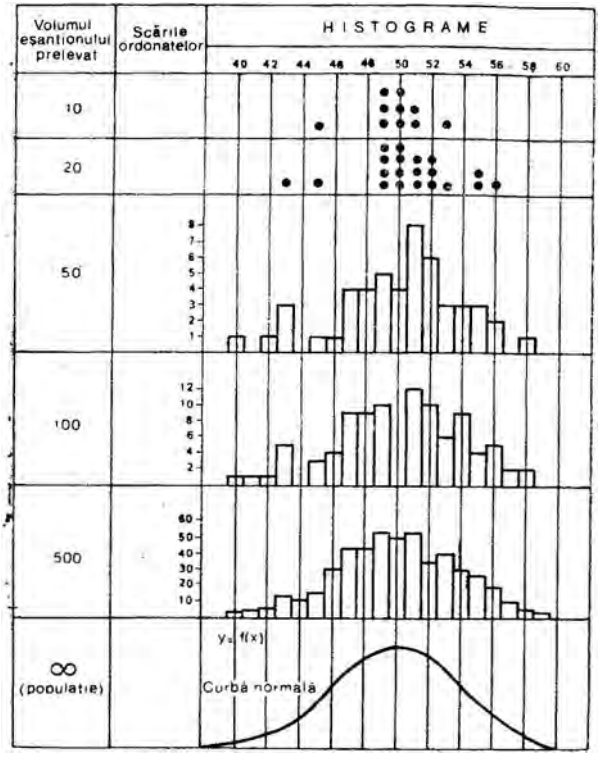


Fig. 35 Histograme și curba frecvențelor

sau teoretică, repartiție care nu poate fi obținută prin înregistrarea statistică din cauza jocului întîmplării. De aceea, calcularea **c.f.** pornind de la datele statistice, constituie una din problemele de bază ale statisticii matematice.
C.f. este o curbă reprezentativă a unei funcții $y = f(x)$, numită *funcția de frecvență** sau *lege de frecvență* *. Pentru studiul teoretic al repartiției statistice, este comod să se substituie frecvențelor f_i frecvențele relative $\tilde{f}_i = \frac{f_i}{\Sigma f_i}$, ceea ce echivalează cu reducerea ordonatelor în raportul $\frac{1}{\Sigma f_i}$ sau, ceea ce înseamnă

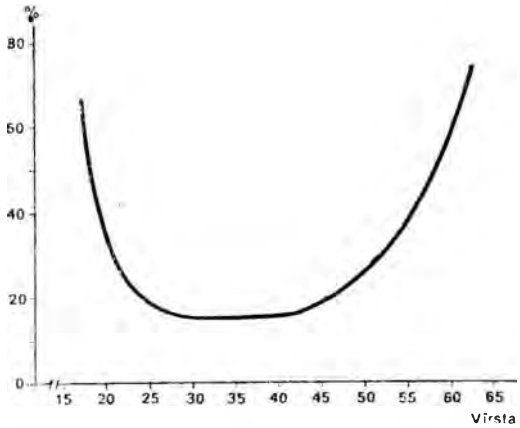


Fig. 36 Proporția (%) a femeilor inactive în populația feminină pe grupe de vîrstă, la recensămîntul din 5 ianuarie 1977

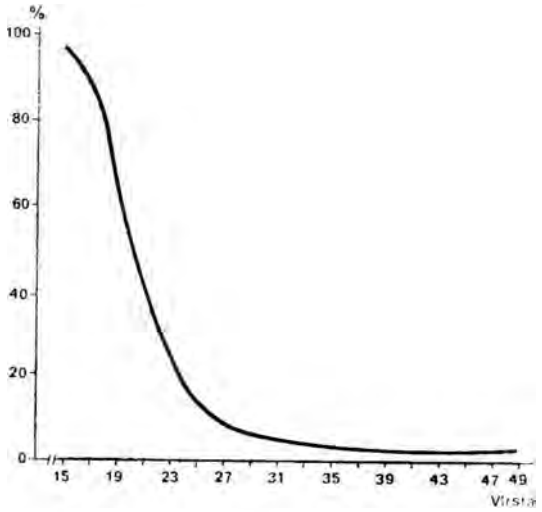


Fig. 37 Proporția (%) a femeilor nemăritate în populația feminină, pe ani de vîrstă, la recensămîntul din 5 ianuarie 1977

același lucru, cu modificarea scării pe axa ordonatelor. În limbajul probabilist, funcția de frecvență (relativă) se numește *densitate de probabilitate.** Construirea curbei permite să se determine prin *interpolare* (deducerea valorilor unei variabile între două valori date) frecvențele pentru valorile intermediare ale abscișelor.

Prin **c.f.** se pune în evidență *forma* repartiției. Astfel, repartiția taliilor într-o populație omogenă se prezintă sub formă simetrică (profil de clopot) : la mijlocul scării de repartiție se găsește o zonă de frecvență maximă, încadrată de două zone în care frecvențele scad în mod progresiv. Alteori se întîlnesc repartiții în formă de ,,u'' (fig. 36) sau de ,,i'' (fig. 37).

Repartiția în ,,u'' se caracterizează printr-un maximum de frecvențe la fiecare extremitate și, în partea centrală, o zonă de frecvențe minime ; în cazul repartiției în ,,i'', frecvențele merg constant în descreștere de la stînga la dreapta. Dacă numărul de observații crește indefinit, *poligonul frecvențelor cumulate* tinde către o curbă numită *curba frecvențelor cumulate* sau *curba integrală a frecvențelor* sau *ogivă*. Pentru repartiția în clopot, această curbă îmbracă forma unui ,,S'' alungit, formă numită sigmoidă (fig. 33).

Ogiva lui Galton. Formă specială a ogivei ; inversarea ordinii uzuale de a trasa graficul, și anume reprezentarea valorilor variabilei pe axa y și a frecvențelor pe axa x.

Față de curba frecvențelor simple, ogiva prezintă avantajul că forma sa nu depinde de mărimea intervalelor de clasă — aceste intervale pot fi inegale. Prin urmare, este posibil să se compare ogivele, chiar dacă intervalele de clasă nu sînt de aceeași mărime.

DIAGRAMĂ POLARĂ *(radială)*. Diagramă construită în *coordonate polare* și întrebuințată de obicei pentru reprezentarea variațiilor sezoniere ale unui fenomen. Se disting două tipuri clasice de **d.p.(r)**, ilustrate de fig. 38 și fig. 39.
a) S-a împărțit cercul în 12 sectoare egale și s-au determinat razele sectoarelor proporționale cu valorile seriei :

Tabelul 6

Luna	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Căsătorii (medii 1976—1980)	6291	7019	4760	5872	6958	6064	7027	8063	7938	9051	6416	4781

Raza cercului este proporțională cu numărul mediu lunar al căsătoriilor pe întreaga perioadă 1976—1980, adică 6 687 căsătorii.
b) Pe 12 raze ce fac unghiuri egale între ele, — și deci egale cu 30° — s-au luat lungimi proporționale cu datele tab. 7.

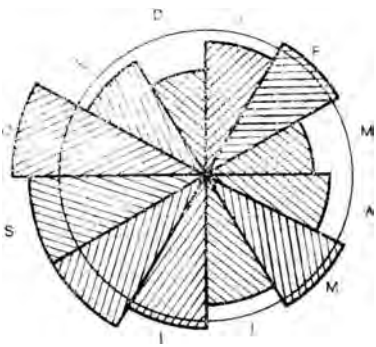


Fig. 38 Variația lunară a căsătoriilor în mediul rural din România (media pe cinci ani 1976—1980).

Tabelul 7

Anul	Luna I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1978	6410	7354	5509	5380	7212	6364	7728	8425	8505	9195	6602	4883
1979	6138	7240	5017	5867	7284	6600	7202	8249	8342	9017	6208	4684
1980	5329	5925	4268	6604	6724	5470	6623	8021	7173	7806	5922	4228

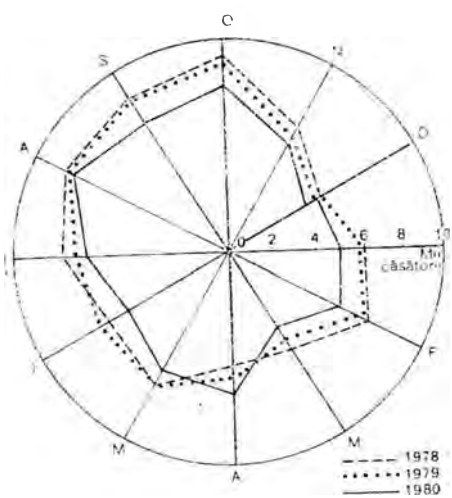


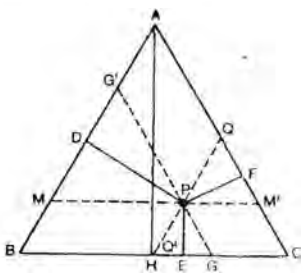
Fig. 39 Variația lunară a căsătoriilor în mediul rural din România, în anii 1978-1980

Această diagramă permite reprezentarea variațiilor în decursul mai multor ani. Ea devine sugestivă îndeosebi cînd sensul variației variabilei este constant crescător sau descrescător. Diagrama prezintă atunci aspectul unei spirale, de unde și denumirea ce i se dă de *diagramă în spirală*.

DIAGRAMĂ TRIUNGHIULARĂ. Diagramă construită cu ajutorul triunghiului echilateral, reprezentînd un ansamblu de trei variabile a căror sumă este constantă. **D.t.** are la bază una din proprietățile următoare ale triunghiului echilateral :

1) pentru orice punct, situat în interiorul triunghiului, suma distanțelor sale la cele trei laturi este constantă și egală cu înălțimea triunghiului ;

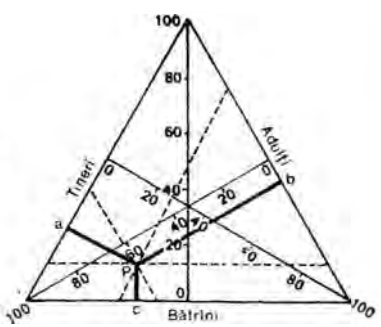
Fig. 40 Diagramă triunghiulară



2) paralelele la laturi duse printr-un punct oarecare din interiorul triunghiului determină cu laturile două grupe de trei segmente a căror sumă este constantă și egală cu latura triunghiului (fig. 40).

Exemplu : În populația României din 1979, populația tînără deține 26,3%, populația adultă 60,5% și populația vîrstnică 13,2% (*Anuarul statistic al R. S. România*, 1980, p. 55—57).

Fig. 41 Structura populației României pe grupe mari de vîrstă, 1979



Conform primei proprietăți, cele trei componente sînt reprezentate pe cele tre înălțimi, gradate de la 0 la 100% și orientate în direcția vîrfurilor triunghiului. Deci, înălțimile sînt scările (fig. 41).

Dreptele duse perpendicular la scări prin diviziunile 26,3%, 60,5% și 13,2% se intersectează în punctul *P*. Acest punct se unește cu laturile triunghiului prin perpendicularele paralele cu înălțimile triunghiului (pe care sînt fixate scările), reprezentînd — prin lungimea lor — proporțiile celor trei populații.

În baza celei de-a doua proprietăți, laturile triunghiului constituie scările. Laturile sînt gradate succesiv, fiecare vîrf corespunzînd unei origini și unei extremități a axei (fig. 42).

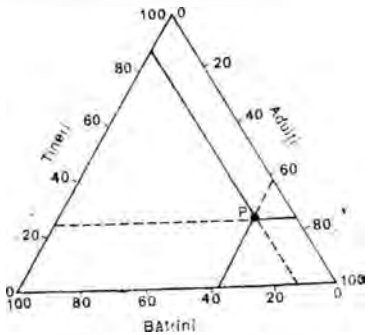


Fig. 42 Structura populației României pe grupe mari de vîrstă, 1979

Punctul reprezentativ (P) se găsește la intersecția paralelelor la laturile triunghiului duse prin punctele de coordonate respective. Utilizarea scărilor construite pe laturile triunghiului ușurează citirea valorilor de structură reprezentate.

DIAGRAMĂ ÎN Z. Grafic al unei serii cronologice lunare, reprezentînd simultan: datele simple, datele cumulate, sumele mobile pe douăsprezece luni. *Exemplu.* Vînzările lunare înregistrate la un magazin universal în anii 1978, 1979 și 1980 (mil. lei).

Tabelul 8

Luna	Vînzările lunare			Vînzările cumulate		Totalurile mobile (12 luni)	
	1978	1979	1980	1979	1980	1979	1980
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ianuarie	21	27	34	27	34	253	311
Februarie	16	20	23	47	57	257	314
Martie	12	23	33	70	90	268	324
Aprilie	13	19	27	89	117	274	332
Mai	24	26	35	115	152	276	341
Iunie	32	35	40	150	192	279	346
Iulie	46	42	39	192	231	275	343
August	11	14	25	206	256	278	354
Septembrie	10	18	31	224	287	286	367
Octombrie	19	26	42	250	329	293	383
Noiembrie	14	22	31	272	360	301	392
Decembrie	29	32	43	304	403	304	403
Total	247	304	403	—	—	—	—

Un procedeu de calcul rapid al totalurilor mobile (col. 6 și 7):
— total mobil ian. 1979 = total 1978 + ian. 1979 — ian. 1978 = 247 + 27 — −21 = 253 mil. lei:
— total mobil febr. 1979 = total mobil ian. 1979 + febr. 1979 — febr. 1978 = = 253 + 20 — 16 = 257
etc.
Cu ajutorul tab. 8 s-a stabilit graficul *Z* pentru anii 1979 și 1980 (fig. 43).

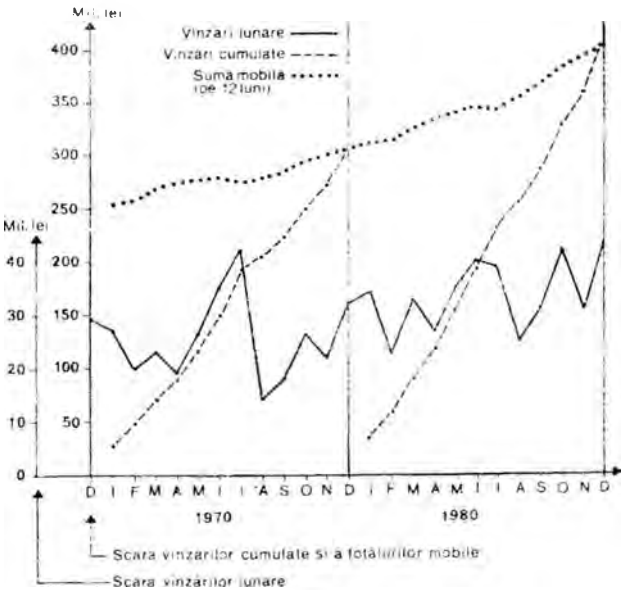


Fig. 43 Diagramă în Z a vînzărilor unui magazin cu amănuntul

Pentru a da graficului dimensiuni normale, se utilizează două scări distincte în ordonate: o scară a datelor simple și o altă scară a datelor cumulate și a totalurilor mobile (care ar putea fi de asemenea gradată logaritmic). Denumirea de **d.** în **Z** se datorează aspectului său.

GRAFIC GANTT. Grafic întrebuințat pentru a confrunta imediat programul stabilit pentru un lucrător, o echipă, o mașină sau un atelier și munca efectiv realizată în decursul unei perioade date. **G.G.** este constituit dintr-o serie de coloane, fiecare corespunzînd unei unități de timp lucrătoare (oră, zi, săptămînă) și avînd o lățime proporțională cu cantitatea lucrului prevăzut; o linie orizontală în interiorul coloanei arată raportul între această cantitate și cantitatea realizată.

Exemplu: Producția unui muncitor, pentru fiecare zi a săptămînii, trebuie să fie — după normă — de 200 piese și s-au executat: luni 160 piese, marți 220 piese, miercuri 210 piese, joi 170 piese, vineri 200 piese și simbătă 230 piese.

G.G. este cel mai răspîndit și cel mai vechi grafic de planificare și de previziune. Avantajul practic al unui astfel de grafic rezidă mai ales în faptul că se ține continuu la zi fără vreo dificultate.

PIRAMIDĂ A VÎRSTELOR. Reprezentarea grafică a repartiției populației după sex și vîrstă, cu ajutorul a două histograme, a căror așezare în grafic are o particularitate. Fie ${}_1P_x$ — efectivul populației de vîrstă x, în care x ia valorile

		Luni	Marți	Miercuri	Joi	Vineri	Sîmbătă
Planificat		200	200	200	200	200	200
Realizat	zilnic	150			170	200	
	cumulat		220	210			230
		160	380	590	760	960	1190

Fig. 44 Graficul Gantt

0, 1, 2, ..., 100 ani sau ${}_5P_x$ — efectivul populației pe clase cincinale de vîrstă: 0—4, 5—9, 10—14, ..., 95—99 ani, 100 ani și peste (interval deschis). Se construiesc două histograme; una pentru populația masculină, cealaltă pentru populația feminină, cu două scări în abscisă, figurate de semiaxe, pe care se trec efectivele populației; pe axa ordonatelor, la scară convenabilă, se trec valorile vîrstei, cu intervale de 1 an sau cinci ani. Cele două histograme se așază în așa fel încît sugerează o piramidă. Efectivele populației sînt reprezentate prin suprafețe care sînt date de intervalul caracteristicii „vîrstă" și de frecvențele reprezentînd populația. Întrucît numărul populației și repartiția ei după vîrstă și sex se determină cu prilejul recensămîntului populației, piramida vîrstelor va reprezenta o asemenea repartiție la momentul critic al recensămîntului, de exemplu 15 martie 1966 sau 5 ianuarie 1977 — ultimele recensăminte din România. Pentru sporirea proprietăților analitice ale **p.v.** este util ca efectivele populației să fie recalculate pentru data de 1 ianuarie al anului respectiv, în această situație efectivele populației pe clase de vîrstă corespund cu generațiile respective, permițînd aprecierea evoluției generației pînă la data de 1 ianuarie. În fig. 45, piramida vîrstelor populației României a fost construită pentru 1 ianuarie 1980, trecîndu-se în dreptul valorilor vîrstelor anii de naștere (generațiile).

De pildă, generația 1979 corespunde vîrstei de 0 ani la 1 ianuarie 1980, generația 1879 corespunde vîrstei de 100 ani. Avînd deci 100 clase de vîrstă corespunzînd unui număr de 100 generații, **p.v.** este o cronică a unei populații pe o perioadă de 100 ani. În absența migrației și a unor evenimente excepționale cum ar fi războaiele și în cazul în care generațiile succesive nu ar fi afectate de oscilații în ce privește regimul de natalitate, atunci o asemenea reprezentare grafică ar avea înfățișarea unei piramide deoarece singurul factor care modifică efectivele populației este mortalitatea, ca funcție de vîrstă. Asemenea populații, de tip închis și neafectate de evenimente particulare, sînt relativ puține și sînt caracteristice de obicei țărilor în curs de dezvoltare.

Populația României, reprezentată în piramida vîrstelor la 1 ianuarie 1980, are cîteva trăsături caracteristice. Unele clase de vîrstă sînt disproporționate față de cele anterioare sau posterioare: cele notate pe grafic cu 1 încercuit,

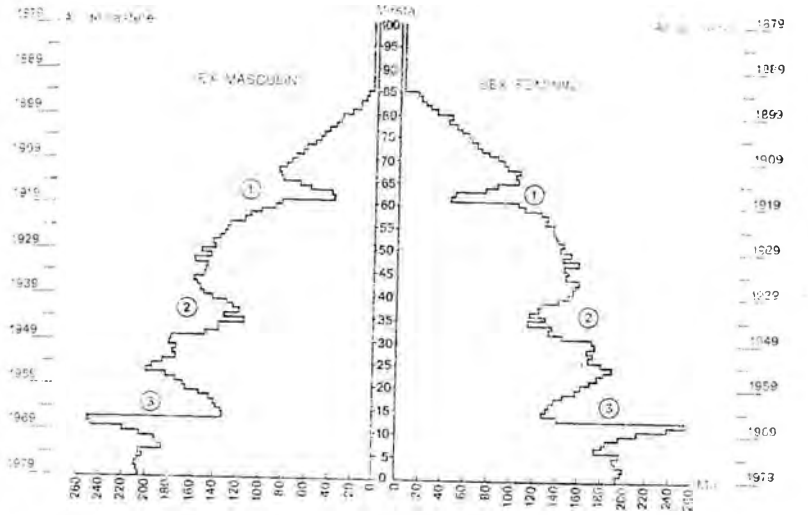
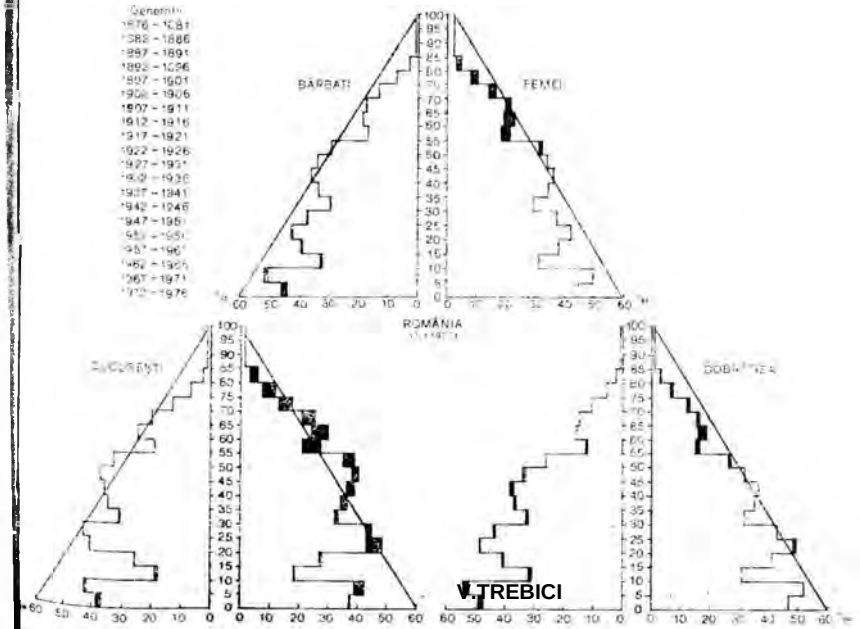


Fig. 45 Piramida vîrstelor populației României la 1 ianuarie 1980



V.TREBICI

Fig. 46 Piramida vîrstelor populației României, municipiului București și pro-

reprezentînd clasele 61—64 ani, și respectiv generațiile 1916—1919, sînt marcate de ceea ce se numește ,,deficitul de naștere'' din timpul primului război mondial. Cele notate cu 2 încercuit sînt clasele de 35—39 ani (generațiile 1941—1945), afectate de cel de-al doilea război mondial; în sfîrșit, cele notate cu 3 încercuit sînt generațiile 1958—1966, apărute în perioada scăderii puternice a natalității. Asemenea clase de vîrstă se numesc ,,întrinduri'' (în fr., *classes creuses*). Forma **p.v.** este o indicație intuitivă a proporțiilor dintre sexe, precum și a proporțiilor dintre grupele mari de vîrstă: populația ,,tînără'' (0—14 ani), populația ,,adultă'' (15—59 ani) și populația ,,bătrînă'' (60 ani și peste) și ca atare permite aprecierea dacă populația respectivă se află în curs de îmbătrînire demografică sau, dimpotrivă, este o populație cu caracteristici tinere.

P.v. se construiește pentru populația națională, populația urbană și rurală, populația județelor și provinciilor istorice; de asemenea este indicat să se construiască **p.v.** pentru diferite subpopulații (corpul didactic, corpul medical, populația activă și inactivă etc.).

În scopuri de comparabilitate pentru mai multe populații în timp și în spațiu este recomandabilă construirea **p.v.** cu frecvențe relative: populația totală (ambele sexe) este considerată egală cu 1 000 sau 10 000 și fiecare efectiv de vîrstă reprezintă o pondere în așa fel încît suma frecvențelor relative este egală cu 1 000 sau 10 000. Fig. 46 ne înfățișează populația României la 1 ianuarie 1977 și pe cea a populației municipiului București și provinciei istorice Dobrogea. Se citesc cu ușurință particularitățile celor trei piramide ale vîrstelor ca și influența unor factori ca: natalitatea, mortalitatea și migrația. De asemenea apare cu claritate excedentul de bărbați sau de femei la diferitele vîrste. Fiind o repartiție de frecvențe, pe baza **p.v.** se calculează valori ca: mediana, cuartile etc.

P.v. poate fi folosită de asemenea pentru reprezentarea unor modele de populații cum ar fi populația staționară sau populația stabilă: în acest caz, efectivele populației sînt date de numărul supraviețuitorilor dintr-o tabelă de mortalitate, în care se consideră o cohortă fictivă de 100 000 persoane la naștere și care se reduce sistematic sub efectul mortalității, pînă la dispariția ei.

Deși **p.v.** este un grafic elaborat de demografi în scopuri demografice, ca poate fi aplicată în oricare alt domeniu în care sînt situații similare.

BIBLIOGRAFIE

R. PRESSAT, *Dictionnaire de démographie*, Presses Universitaires de France, Paris, 1979, p. 173—174.

V. TREBICI, *Demografia*, Editura științifică și enciclopedică, București, 1979, p. 77—80.

GRAFICUL LUI LEXIS. Reprezentare grafică constînd dintr-o rețea de paralele la un sistem obișnuit de axe rectangulare, în așa fel încît înfățișarea generală este de rețea. Scara absciselor este aceeași cu scara ordonatelor: pe prima se trec datele observării, iar pe scara ordonatelor — valori reprezentînd durate de timp (vîrsta, vechimea unei cohorte etc.). Aplicația cea mai frecventă a **g.L.** este observarea unei cohorte demografice — generație, promoție de căsătorii — în decursul unei perioade de timp, cu care prilej se schimbă efectivul inițial al cohortei sub efectul evenimentelor demografice. Astfel, numărul inițial al unei generații se modifică sub efectul deceselor care se înregistrează în fiecare an, numărul unei promoții de căsătorii se diminuează sub efectul divorțului și al decesului, efectivul unei cohorte de persoane necăsătorite, odată cu intrarea

în vîrsta nubilă (16, respectiv 18 ani), se diminuează prin căsătorii etc. Fig. 47 ne dă înfățișarea generală a **g.L.**
Culoarul haşurat, ales ca exemplu, este drumul afectat unei cohorte pe măsura îmbătrînirii ei (durata în ani) și a scurgerii timpului (perioade). După parcurgerea a șapte perioade, cohorta respectivă ajunge la durata de 6 ani (vîrstă

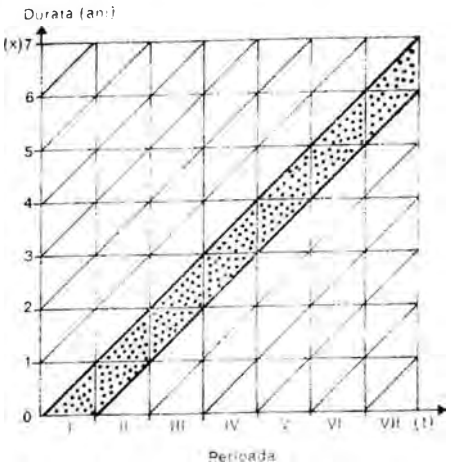


Fig. 47 Graficul lui Lexis

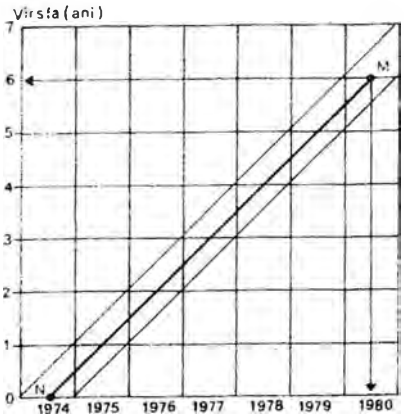


Fig. 48 Linia vieții în graficul lui Lexis

sau vechime); cu acest prilej se poate face contabilizarea evenimentelor demografice care au afectat efectivul cohortei, precum și determinarea efectivului la fiecare dată calendaristică (1 ianuarie), fie la fiecare aniversare.
Pentru înțelegerea proprietăților descriptive și analitice ale **g.L.** luăm cazul cel mai simplu al unei persoane.

O persoană născută la 1 iulie 1974 parcurge culoarul determinat de cele două paralele, iar la 1 iulie 1980 survine decesul, la vîrsta exactă de 6 ani. Linia NM este linia vieții; prin extindere, aplicăm linia vieții la o căsătorie contractată în momentul C și desfăcută în momentul D. Se vede clar din grafic că se corelează trei variabile: data nașterii (t), data observării decesului (z) și vîrsta persoanei decedate (x). Demografia, ca și statistica, nu se ocupă de cazuri, ci de mase de evenimente și de efective ale cohortelor. De pildă, efectivul generației feminine 1970, de 207 551 s-a diminuat în anul 1971 cu 9547 persoane (decese infantile), ajungînd la 198 004; în 1972 se diminuează cu 502 persoane (decese

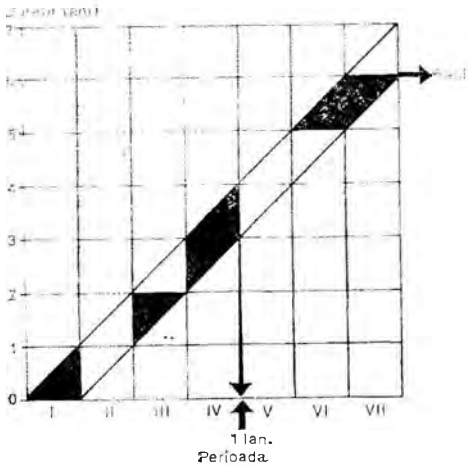


Fig. 49 Mase de evenimente demografice în graficul lui Lexis

între vîrsta de 1 an și 2 ani), ajungînd deci la 197 502 persoane, în așa fel încît efectivul generației la 1 ianuarie 1977 ajunge la 196 857, diferență de 207 551 — 196 857 = 10 694 fiind numărul celor decedați de la naștere (1970) pînă la data de 1 ianuarie 1977.

Masele de evenimente, în cazul nostru decesele, constituie mase de puncte și care formează triunghiuri în **g.L.** (fig. 49); ele reprezintă dubla grupare a evenimentelor demografice.

De pildă, triunghiul din perioada I la 0 ani, reprezintă numărul deceselor ce au avut loc pînă la 1 ianuarie perioada II, la vîrsta de 0 ani; triunghiul din perioada III reprezintă decesele survenite în perioada III și care corespund vîrstei de 1 an (pînă la doi ani); paralelogramul din perioada IV cuprinde decesele generației survenită în perioada IV, produse între 2 și 3 ani, între două date calendaristice, iar paralelogramul din perioada VII cuprinde decesele generației survenite între două aniversări exacte, 5 și 6 ani, dar care s-au produs în două perioade: VI și VII. Masele de evenimente demografice modifică efectivele cohortei; or, efectivul se determină fie la o dată calendaristică (1 ianuarie perioada V, în fig. 49), fie la o aniversare exactă (6 ani, fig. 49). Cu alte cuvinte, efectivele cohortei sînt fluxurile liniilor vieții reprezentate de segmente corespunzînd datelor sau duratelor (aniversărilor).

V. TREBICI

Fig. 50 ne oferă alt exemplu de **g.L.** Promoția primelor căsătorii încheiate în România în anul 1975 a fost de 170 825. Pînă la 1 ianuarie 1976 promoția respectivă a avut 20 339 născuți-vii la o durată de 0 ani;

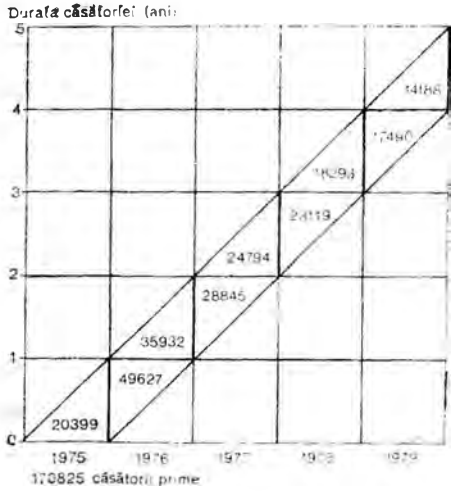


Fig. 50 Evoluția primelor căsătorii încheiate în 1979, în perioada 1975—1979

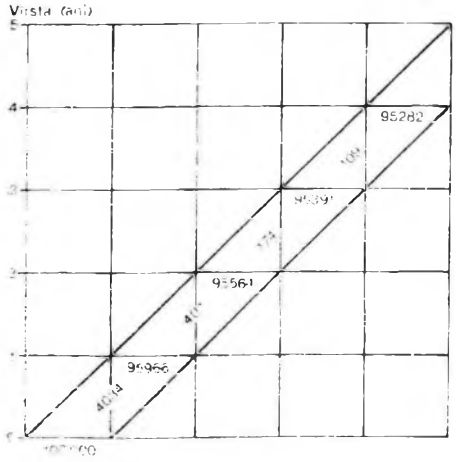


Fig. 51 Tabela de mortalitate feminină 1970—1972 (fragment)

în anul 1976 numărul de născuți-vii pînă la prima aniversare a căsătoriei a fost de 49 629; așadar, pînă la împlinirea duratei de 1 an, numărul total al născuților a fost de 70 026; în 1976, un număr de

născuți-vii s-a realizat între vîrsta de 1 an și 2 ani; în total, promoția de căsătorii 1975 a avut în anul 1976 un număr de 85 559 născuți-vii ș.a.m.d. De la formarea promoției de căsătorii în 1975 și pînă la 1 ianuarie 1980 s-au realizat 232 687 născuți care pot fi contabilizați pe fiecare an calendaristic (între 1 ianuarie al unui an și al celui imediat următor) și pe fiecare durată (între două aniversări exacte). Rezultă limpede marile avantaje oferite de **g.L.** pentru calculul diferitelor valori și, în general, pentru modelarea matematică.

Cel mai vechi și mai important model în demografie este *tabela de mortalitate* *, ale cărei funcții se determină pe baza efectivului inițial al cohortei și a numărului deceselor survenite între două aniversări. În acest scop, se pleacă de la o cohortă fictivă, cu un efectiv inițial de 10 000 sau 100 000, urmărindu-se scăderea efectivelor pe măsura creșterii vîrstei, ca efect al mortalității.

Fig. 51 ne arată un fragment din tabela de mortalitate a populației feminine, pe anii 1970—1972. Pînă la prima aniversare a generației aceasta se diminuează cu 4 034 decese, deci efectivul supraviețuitorilor la 1 an este de 95 966, pînă la a doua aniversare, supraviețuitorii se diminuează cu 401 persoane etc., în așa fel încît la vîrsta de 100 ani dispare ultimul supraviețuitor.

G.L. se utilizează în demografie pentru orice cohortă — grup de persoane care în decursul unei perioade a avut același eveniment — în vederea urmăririi schimbării efectivelor și a maselor de evenimente. El se poate folosi și în alte domenii care au caracteristici similare. Un exemplu este observarea unui grup de mașini în timp în raport cu unele caracteristici, pentru determinarea unei legi de mortalitate, a unei repartiții de tip Weibull, a unui efectiv de animale etc.

G.L. este de fapt un grafic la a cărui realizare au contribuit și alți statisticieni, afară de *W Lexis* * (1837—1914), cum ar fi Zeuner, Becker, Perozzo. **G.L.** mia este cunoscut și sub denumirea de „rețea demografică" și este un mijloc foarte util în observarea și analiza longitudinală (diacronică).

BIBLIOGRAFIE

R. PRESSAT, *Dictionnaire de démographie*, Presses Universitaires de France, Paris, 1979.

V. TREBICI, *Mică enciclopedie de demografie*, Ed. științifică și enciclopedică, București, 1975.

CARTOGRAMĂ. Reprezentare grafică a repartițiilor statistice teritoriale (pe țări, continente, județe etc.), fără să comporte elemente matematice deosebite. Se disting **c.** în culori, **c.** hașurate, **c.** în bandă, **c.** cu izolinii sau diferite combinații.

Fig. 52 ne prezintă harta României pe județe, acestea din urmă fiind colorate diferit în raport cu caracteristica „densitatea populației" la recensămîntul din 1977.

CARTODIAGRAMĂ. Reprezentare grafică constînd în desemnarea unei hărți împărțite din punct de vedere teritorial, în interiorul căreia sînt înscrise diferite diagrame care reprezintă grafic indicatorii fenomenului cercetat. Diagramele pot fi : coloane, cercuri, pătrate, sfere, puncte etc. Regulile formulate pentru întocmirea diagramelor sînt în întregime valabile la cartodiagramă, cu un element în plus : plasarea lor în hartă.

Un caz particular este combinarea unei hărți cu o reprezentare de tipul stereogramei.

Suprafețci unei provincii i se adaugă o înălțime, care este proporțională cu mărimea fenomenului reprezentat la scară convenabilă. Se remarcă cu ușurință

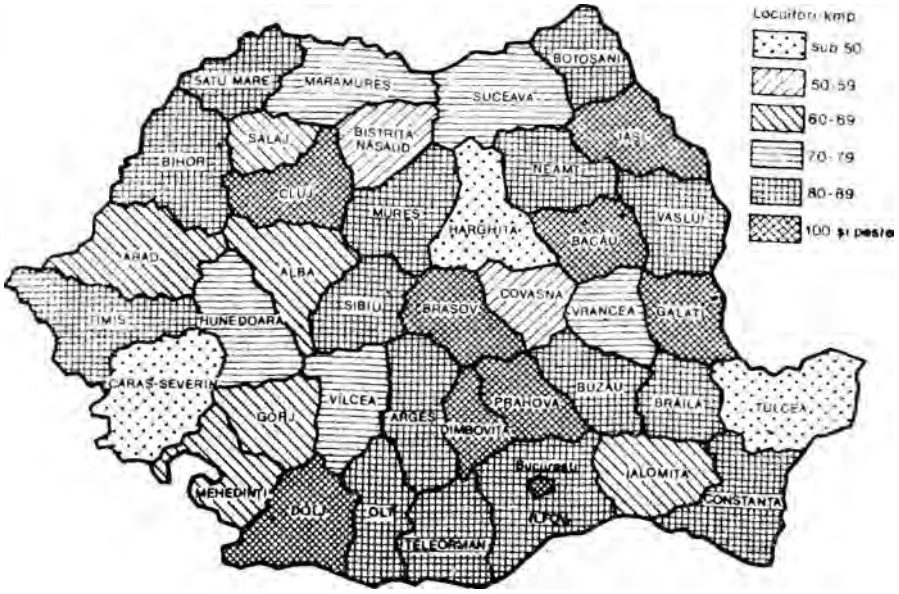


Fig. 52 Densitatea populației României în 1977, pe județe



V.TREBICI

Fig. 53 Dinamica producției globale industriale, pe județe, 1979

că excedentul cel mai înalt se înregistrează în Moldova, iar cel mai scăzut, în provincia istorică Banat. O asemenea reprezentare grafică sporește posibilitățile de analiză.

BIBLIOGRAFIE

D. HAȘIGAN, I. MARINESCU, *Grafice și elemente de calcul grafic*, Ed. științifică. București, 1968.

CORELOGRAMĂ. Reprezentare grafică a variației coeficientului de autocorelație de ordinul K în funcție de K. (→ *Analiza seriilor cronologice, Autocorelație).*

STEREOGRAMĂ. Reprezentare grafică în spațiu a repartiției empirice bidimensionale, adică a corespondenței care există între valorile vectorului (x,y) în cazul cînd x, y sînt variabile aleatoare continue, pe de o parte, și frecvențele perechilor x_i, y_i corespund simultan celor două intervale de clasă ale lui X și respectiv Y. Este ansamblul format din paralelipipede dreptunghiulare avînd ca bază dreptunghiul cu laturile egale, respectiv, cu intervale de clasă ale lui X și Y, și volumele proporționale cu frecvențele n_{ij} ale perechilor $x_i y_j$.

DIAGRAMĂ BARLOW—CAMPO. Reprezentare grafică simplă pentru determinarea formei funcției de repartiție și specificarea parametrilor ei. Spre deosebire de rețelele probabiliste, **d.B.—C.** are avantajul că nu necesită hîrtie specială gradată. Se bazează pe statistica *timpului total de testare* (→ XVI. *Verificarea ipotezelor statistice, Teste pentru verificarea normalității* și *Control statistic al calității).*

V.TREBICI

PARTEA A TREIA

STATISTICA DESCRIPTIVĂ

Ca primă fază a demersului statistic, statistica descriptivă are ca noțiune centrală *repartiția de frecvențe* (empirică), împreună cu indicatorii ce caracterizează forma, tendința centrală și împrăștierea, precum și regresia și corelația.

VI. REPARTIȚII EMPIRICE (DE FRECVENȚE)

SERIE STATISTICĂ *(repartiție statistică)*. Șir de valori numerice ale unei caracteristici, ordonate după un anumit principiu, în funcție de șirul valorilor unei alte caracteristici. S.s. este deci corespondența dintre două șiruri de date, din care primul are în vedere caracteristica folosită la sistematizarea dublă, iar cel de-al doilea — mărimea caracteristicii, repartiția unităților populației etc. S.s. poate fi formată atît din mărimile absolute rezultate din observare sau măsurare, cît și din mărimi statistice obținute prin diferite prelucrări. S.s. permite comparații atît din punctul de vedere al timpului, cît și al diferitelor tipuri ale procesului.
Notă. Termenul de *serie* este obișnuit rezervat *seriei cronologice* *.

Serie statistică spațială. Șir de valori care arată variația volumului populației sau a nivelului caracteristicii în raport cu variația unei caracteristici teritoriale : populația pe provincii istorice, pe județe etc.
Serie statistică cronologică (de timp). Șir de valori care arată variația volumului populației sau a nivelului caracteristicii în raport cu timpul (o dată sau o perioadă). (→ XI. *Seriile cronologice*).

Serie statistică de repartiție. Șir de valori care arată schimbarea frecvențelor unei caracteristici în raport cu variația unei caracteristici (→ *Repartiția de frecvențe*). Denumirea cea mai frecventă este de *repartiție de frecvențe*.
Pentru a fi eficiente și comode, **s.s.r.** se stabilesc după o serie de reguli, unele generale, cu aplicație la toate tipurile de repartiții, altele specifice fiecărui tip în parte. Mai întîi, orice repartiție se construiește potrivit unei idei directoare sau unui interes de cunoaștere, practic ori teoretic. Aceasta implică operații de selecționare în optica interesului reținut : în unele cazuri, repartiția se formează în raport cu o singură variabilă, iar în alte cazuri, în raport cu două sau mai multe variabile ; uneori se adoptă o gradație naturală între diferitele clase ale variabilei, alteori, una conceptuală sau convențională. Numărul claselor unei variabile poate de asemenea să varieze conform informației dorite. Clasele aceleiași variabile trebuie să fie însă totdeauna *incompatibile* — să se excludă reciproc și totodată *exhaustive* — să acopere toate cazurile particulare. Respectarea regulii din urmă apără de erori în clasarea unităților și interpretarea repartiției. Procesul de construire a **s.s.r.** este un proces de sintetizare a observărilor primare, care atrage după sine sacrificarea unei anumite cantități de

informație. Există astfel riscul ca structurile reale să fie deformate și, implicit, repartițiile construite să devină inutilizabile. Regula constă în păstrarea la maximum a integrității informațiilor cuprinse în documentația de bază. Formarea de clase omogene — o altă regulă elementară — presupune o condensare a datelor cu măsură și, în mod corespunzător, minimizarea pierderilor de informație.

Regulile de construire a **s.s.r.** diferă apoi după cum variabilele cercetate sînt *calitative* sau *cantitative*.

CLASĂ. Subdiviziune a domeniului de variație al variabilei observate.

C. sînt definite verbal în cazul *variabilei calitative* * (**c.** de stare civilă : necăsătorit, căsătorit, văduv, divorțat) și numeric în cazul *variabilei cantitative* * (**c.** de vîrstă anuale : 0 ani, 1 an, 2 ani, 3 ani, ... sau **c.** de vîrstă cincinale : 0—4 ani, 5—9 ani, 10—14 ani, ...).

Termeni sinonimi : *categorie, grupă, modalitate.*

Pentru **c.** constituite din grupe de valori trebuie să se determine : amplitudinea **c.**, limitele **c.** și centrul acesteia.

Amplitudinea clasei. Mărimea intervalului de clasă sau de grupare.

Limitele clasei. Valoarea cea mai mică *(limita inferioară)* și valoarea cea mai mare *(limita superioară)* a **c.**

Centrul (mijlocul) clasei (sau al intervalului). Valoarea variabilei socotită că reprezintă valorile tuturor unităților cuprinse în clasă (media aritmetică a limitelor).

Exemplul 1. Valorile unei variabile, ordonate după mărime . . . . . .
. . . .25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35 . . . . . .etc.
sînt întrunite în clase de cîte 5 valori.

. . . . .
25—29
30—34
. . .

Amplitudinea **c.** (intervalul de clasă) este $30-25=5$. Limitele **c.**: limita inferioară 25 (resp. 30) și limita superioară 29 (resp. 34).

Centrul **c.** $(25+29)/2=27$ [resp. $(30+34)/2=32$].

Observație. În demografie, **c.** de vîrstă x include toate persoanele a căror vîrstă este cuprinsă între x și $x+1$ ani. De unde, considerînd **c.** de valori, luate ca exemplu, drept **c.** de vîrstă, centrele **c.** vor fi :

C. *de vîrstă*	*Centrul de* **c.**
. . . . .	. . . . .
25—29 ani (=29 ani+364 zile)	27,5
30—34 ani	32,5
. . . . .	. . . . .

Variabila „vîrstă" este tratată ca *variabilă implicit continuă.*

Clasă deschisă. C. de amplitudine nedeterminată. Prima și ultima **c.** pot fi „închise" sau „deschise", după cum seria comportă sau nu un număr de valori „aberante".

V.TREBICI

Exemplul 2. Cînd populația totală se repartizează în **c.** de vîrstă, se convine ca ultima **c.** să rămînă *deschisă:*
0—4 ani
5—9 ani
...
85—89 ani
90 ani și peste

Exemplul 3. Într-o clasificare a căsătoriilor după vîrsta soților, cu un interval de **c.** de 5 ani, ambele **c.** extreme sînt lăsate de obicei *deschise:*
sub 20 ani
20—24 ani
...
55—59 ani
60 ani și peste.

FRECVENȚĂ. Număr de observații corespunzător fiecărei *clase* * din *repartiția statistică* *

F. astfel definită este numită și **f.** *absolută* sau *efectiv* (notată f_i).

Frecvență de clasă. F. cu care se înscrie fiecare clasă în repartiția statistică.

Frecvență totală. Sumă a tuturor **f.** unei repartiții statistice.

Frecvență relativă. F. individuală (de clasă) exprimată ca o proporție a **f.** totale

Exemplu. Dacă într-o repartiție a unei *variabile discrete* *, X, rezultată din N observații (f. *totală*) s-a găsit de f_i ori valoarea X_i (**f.** *de clasă*), f_i reprezintă, prin definiție, **f.** absolută a valorii x_i, iar raportul $\frac{f_i}{N}$ **f.** sa *relativă* *).

Considerînd că variabila X poate lua n valori, $x_1, x_2, \ldots, x_n$, evident

$$f_1 + f_2 + \ldots + f_n = \sum_{i=1}^{n} f_i = N$$

și

$$\frac{f_1}{N} + \frac{f_2}{N} + \ldots + \frac{f_n}{N} = \sum_{i=1}^{n} \frac{f_i}{N} = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} f_i} = 1.$$

F. relativă, notată $\tilde{f}_i$, este în mod necesar cuprinsă între 0 și 1 :

$$0 \leqslant \tilde{f}_i \leqslant 1.$$

$\tilde{f}_i = 0$ dacă nu există nici un element în clasa de rangul i.
$\tilde{f}_i = 1$ dacă toate elementele populației aparțin aceleiași clase.
Aceste cazuri extreme nu prezintă nici un interes.

---

*) În terminologia unor autori, o f. (fără altă precizare) desemnează o f. relativă.

Ca să se evite prea multe cifre zecimale, **f.r.** se exprimă obișnuit în procente și suma lor devine egală cu 100. **F.r.** servesc la compararea directă a structurilor după variabila studiată, cînd populațiile diferă ca volum.

Frecvență cumulată. F. dedusă din adunarea succesivă a **f.** ordinare (simple) Cumularea **f.** poate să înceapă de la prima clasă sau de la ultima clasă a repartiției statistice (în sensul citirii normale a repartiției).

În prima variantă se obțin **f.c.** ascendente :

$$F_1 = f_1,$$

$$F_2 = f_1 + f_2,$$

$$F_3 = f_1 + f_2 + f_3,$$

. . . . . .

$$F_i = f_1 + f_2 + f_3 + \ldots + f_i.$$

Cumulul de rangul n este egal cu **f.** totală :

$$F_n = f_1 + f_2 + f_3 + \ldots + f_n = \sum_{i=1}^{n} f_i.$$

În a doua variantă se obțin *f.c.* descendente :

$$F'_1 = f_n,$$

$$F'_2 = f_n + f_{n-1},$$

$$F'_3 = f_n + f_{n-1} + f_{n-2},$$

. . . . . .

$$F'_i = f_n + f_{n-1} + f_{n-2} + \ldots + f_{n-i+1}.$$

Cumulul de rangul n reproduce **f.** totală

$$F'_n = f_n + f_{n-1} + f_{n-2} + \ldots + f_1 = \sum_{i=1}^{n} f_i.$$

și totdeauna

$$F_i + F'_{n-i} = \sum_{i=1}^{n} f_i.$$

La fel se definesc **f.** *relative cumulate*. Cumulul — direct sau invers — de rangul n este egal cu 1 (respectiv 100%). Adică

$$\bar{F}_n = \bar{f}_1 + \bar{f}_2 + \bar{f}_3 + \ldots + \bar{f}_n = 1.$$

și

$$\tilde{F}'_n = \tilde{f}_n + \tilde{f}_{n-1} + \tilde{f}_{n-2} + \ldots + \tilde{f}_1 = 1.$$

De asemenea

$$\tilde{F}_i + \tilde{F}'_{n-i} = \sum_{i=1}^{n} \tilde{f}_i = 1.$$

F.c. sînt necesare pentru anumite calcule statistice și, uneori, pentru o caracterizare mai sugestivă a repartițiilor statistice.

Frecvență unitară (frecvență pe interval de clasă unitar sau densitate de frecvență) — **f.** definită de raportul :

$$\frac{f_i}{\Delta x_i}, \text{ respectiv } \frac{\tilde{f}_i}{\Delta x_i},$$

în care Δx_i este *intervalul de clasă.**

În cazul repartiției pe o scară de *intervale neegale**, **f.** (absolute sau relative) nu sînt comparabile nemijlocit între ele. În scopul asigurării comparabilității, se obișnuiește a le împărți la intervalele de clasă corespunzătoare, obținîndu-se noi **f.** numite *f.u.*

REPARTIȚIE DE FRECVENȚE. Repartiție a unităților populației în raport cu o variabilă calitativă sau cantitativă.

*Pentru o variabilă calitativă,** repartiția are forma generală :

Clasa	a	b	...	i	...	n
Frecvența	$f(a)$	$f(b)$	...	$f(i)$	...	$f(n)$

Trăsătura distinctivă a acestei repartiții este *simetria claselor*, în sensul că toate clasele se consideră egale între ele ($a = b = \ldots = i = \ldots = n$). Pentru a se arăta că nu există nici o deosebire de grad între clasele componente, repartiția a fost denumită *homogradă*.

Cînd variabila cercetată are clasele clar definite (populația : masculin — feminin ; transportul de călători : feroviar — auto — fluvial — maritim — aerian), repartiția se construiește fără nici o dificultate. Dar, adesea clasele scării de repartiție nu pot fi circumscrise decît grație admiterii anumitor convenții, avînd deci un caracter mai mult sau mai puțin subiectiv (categorii socioprofesionale, ramuri ale economiei naționale). Într-o atare circumstanță este necesar să se întocmească un *nomenclator* — o listă a claselor — pentru a se preîntîmpina posibilitățile de interpretare personală. Dacă s-au fixat clasele variabilei, nu rămîne decît să se socotească, pentru fiecare din ele, frecvența cu care s-a înregistrat. De obicei, repartiția se prezintă în ordinea descrescătoare a frecvențelor.

Exemplul 1. Populația României pe naționalități, la recensămîntul din 5 ianuarie 1977 (→tab. 9).

Exigența exhaustivității repartiției conduce destul de des la închiderea sa cu o clasă suplimentară : „alte cazuri", „alte categorii", „nedeclarați" etc. (în exemplu : „alte naționalități"). Desigur, numărul de unități înglobate la

„alte...“ va trebui să fie relativ restrins, pentru ca repartiția obținută să păstreze un interes complet.

Tabelul 9

Naționalitatea	Numărul populației	%
Total	21559910	100,0
Români	19003544	88,1
Maghiari	1706874	7,9
Germani	358732	1,7
Alte naționalități	490760	2,3

Sursa: „Revista de statistică“ nr. 6, 1977.

Elementul prin care repartiția unei variabile de tip calitativ intră în judecățile statistice este *frecvența* *. Frecvențele caracterizează importanța părților componente ale populației și asocierile variabilelor calitative sînt precizate prin comparații între frecvențe.

O variabilă cantitativă poate lua un număr restrins de valori și atunci, asemănător cazului variabilei calitative, repartiția decurge ușor din gruparea unităților statistice pe fiecare din valorile observate. Dacă $x_1, x_2, \ldots, x_i, \ldots x_n$ sînt valorile observate pentru o variabilă X și $f_1, f_2, \ldots f_i, \ldots, f_n$ frecvențele corespunzătoare, repartiția se scrie :

$$X\begin{pmatrix} x_1 & x_2 & \ldots & x_i & \ldots & x_n \\ f_1 & f_2 & \ldots & f_i & \ldots & f_n \end{pmatrix}.$$

Valorile variabilei se înșiruie în ordinea mărimii, crescătoare și citeodată descrescătoare. Unitățile înregistrate se diferențiază după mărimea variabilei ($x_1 < x_2 < x_3 \ldots$) și de aceea repartiția rezultată s-a numit *heterogradă*.*)

Exemplul 2. Născuții vii, după rang, în România, 1979.

Tabelul 10

Rangul născutului viu x_i	Numărul născuților vii f_i	% $\tilde{f}_i$
Total	410603	100,0
1	161940	39,5
2	123498	30,0
3	57957	14,1
4	35264	8,6
5	14487	3,5
6	6917	1,7
7	4018	1,0
8 și peste	6522	1,6

Sursa: *Anuarul statistic al R. S. România 1980*, p. 72.

*) Termenii „homograd“ și „heterograd“ — termeni vechi, rar folosiți — au fost introduși de statisticianul suedez Charlier.

Repartiția este formată din *clase discrete* (entități separate) : rangul născutului viu poate fi 1 sau 2 sau 3 etc. Seria claselor se termină cu o *clasă deschisă* (8 și peste), deoarece născuții vii de ranguri superioare rangului al 7-lea nu s-au înscris cu frecvențe semnificative, care să justifice evidențierea lor în clase deosebite. De regulă, variabilele observate apar cu o gamă largă de valori și, în acest caz, se aplică *repartiția pe clase* (*grupe*) *de valori*. Domeniul de variație al variabilei se împarte într-un număr „convenabil" de clase consecutive de amplitudine egală sau nu și se indică pentru fiecare din ele frecvența înregistrată. Repartiția ia forma :

$$X\begin{pmatrix} x_0 \leqslant x < x_1 & x_1 \leqslant x < x_2 \ldots x_{i-1} \leqslant x < x_i \ldots x_{n-1} \leqslant x < x_n \\ f_1 & f_2 \quad \ldots \quad f_i \quad \ldots \quad f_n \end{pmatrix}.$$

O clasă este delimitată prin două valori ale variabilei — *limitele* (*extremitățile*) *clasei* (x_{i-1} și x_i), a căror diferență constituie *intervalul de clasă* ($x_i - x_{i-1}$). Se atribuie unităților clasei aceeași valoare a variabilei — în general, *mijlocul intervalului*, egal cu semisuma valorilor extreme :

$$\frac{x_{i-1} + x_i}{2}.$$

Exemplul 3. Repartiția cooperativelor agricole *) din județul Ialomița după producția medie de porumb la hectar, în 1978.

Tabelul 11

Producția medie la hectar (chintale)	Număr de cooperative	%
Total	99	100,0
25—30	4	4,0
30—35	3	3,0
35—40	6	6,1
40—45	11	11,1
45—50	10	10,1
50—55	17	17,2
55—60	12	12,1
60—65	19	19,2
65—70	8	8,1
Peste 70	9	9,1

*) Cooperative cu recolte superioare la porumb.

În spațiul colecției de unități tratate „producția medie la hectar" variază de o manieră cvasicontinuă și, dacă s-ar opera pe o scară de valori distincte, ar rezulta o repartiție întinsă, complicată și neexpresivă. Ca urmare, intervalul de variație s-a împărțit în 10 clase egale ca dimensiune (5 chintale), cooperativele grupîndu-se în funcție de clasele stabilite. Problema nouă și de primă importanță, generată de o asemenea repartiție, privește deci determinarea

grupelor de valori ce formează scara clasificării unităților observate. Trebuie avut în vedere faptul că fixarea intervalelor de grupare include decuparea mai mult sau mai puțin arbitrară a cîmpului de variație al variabilei, cu implicațiile ei. Astfel, o scară greșit aleasă poate schimba complet aspectul repartiției. Și aici, experiența cercetătorului și buna cunoaștere a domeniului investigat au menirea să joace rolul esențial. Multă grijă comportă decizia în alegerea, printre un număr mare de soluții posibile, a mărimii intervalelor de clasă (grupare). Firește, nu se poate da o indicație universală în ceea ce privește mărimea optimă a intervalelor; numai o analiză prealabilă, pentru fiecare caz în parte, este în stare să ofere un răspuns. Se recomandă doar ca intervalele să nu fie prea largi, întrucît ar produce o pierdere prea mare de informație, ar disimula anumite particularități ale repartiției și, implicit, ar micșora precizia rezultatelor și nici prea înguste, căci, deși ar conserva o informație mai deplină, n-ar elimina iregularitățile accidentale, în schimb ar complica fără real folos calculele. Un echilibru trebuie să fie găsit între „o precizie amăgitoare și o informație suficientă" *).

Intervalul de clasă depinde de numărul de clase și de întinderea repartiției. O regulă empirică — *regula lui Sturges* — definește numărul minim de clase (k) în funcție de numărul de observații (N):

$$k = 1 + \frac{10}{3} \log_{10} N. \tag{1}$$

Amplitudinea intervalului (d) — pe o scară cu intervale egale — poate fi dedusă din împărțirea diferenței între valoarea cea mai mare (x_{max}) și valoarea cea mai mică (x_{min}) a serici de valori observate la numărul de clase, calculat după formula (1).

Așadar:

$$d = \frac{x_{max} - x_{min}}{1 + \frac{10}{3} \log_{10} N}. \tag{2}$$

O altă relație pentru stabilirea intervalului de grupare (d) este următoarea:

$$d = \frac{1}{100} \cdot 8(x_{max} - x_{min}), \tag{3}$$

în care 8 este un factor empiric.

De la o clasă la alta, intervalele pot avea o amplitudine constantă sau variabilă, adică pot fi *egale* sau *neegale*. Este de preferat repartiția pe o scară cu intervale egale, fiindcă numai așa frecvențele diferitelor clase sînt comparabile între ele și potrivite pentru calcule ulterioare. Nu întotdeauna însă divizarea pe intervale egale este cea adecvată.

Exemplul 4. Repartiția întreprinderilor industriale din România după numărul muncitorilor în 1979 * .(→ tab.12).

*) Numărul „optim" de clase admis de unii autori se situează între 10 și 15, de alții, între 8 și 20, iar de alții, între 15—25. În practică, rareori se folosesc mai puțin de 8 sau mai mult de 15 clase.

Intervalele marginale sînt *deschise*; intervalele intermediare au amplitudinile: 300, 500, 1000, 1000, 2000,... muncitori.

În principiu, o scară cu intervale de mărime egală presupune o populație perfect omogenă în ceea ce privește compoziția; în caz contrar, intervalele se iau de mărimi diferite, pentru ca diferențierile calitative în structura

Tabelul 12

Numărul muncitorilor	Numărul întreprinderilor	%
Total	1701	100,0
Pînă la 200	149	8,8
201—500	323	19,0
501—1000	404	23,7
1001—2000	385	22,6
2001—3000	204	12,0
3001—5000	141	8,3
Peste 5000	95	5,6

* la sfîrșitul anului.
Sursa: *Anuarul statistic al R. S. România 1980*, p. 136.

populației să nu fie eclipsate. Fiecare clasă apare definită de o *limită inferioară* și o *limită superioară*. Ca să se suprime orice risc de eroare, limitele claselor trebuie să fie explicit indicate. Formele utilizate sînt următoarele (cu aplicație la clasele de retribuții lunare):

(a)	(b)	(c)
1600—1799	1601—1800	1600—1800
1800—1999	1801—2000	1800—2000
2000—2199	2001—2200	2000—2200
. . .	. . .	. . .

Formele (a) și (b) sînt explicite, dar forma (c), în care valoarea limitei superioare a unei clase se confundă cu valoarea limitei inferioare a clasei următoare, poate da naștere la ambiguități. Într-adevăr, valorile observate ce s-ar găsi la frontiera claselor ar putea fi repartizate în chip diferit (cum ar fi: persoana cu retribuția de 1800 lei pe lună ar putea fi repartizată fie în prima clasă, fie în cea de-a doua clasă). Neajunsul se înlătură printr-o convenție prealabilă, care, în general, admite cuprinderea în clasă a limitei sale inferioare și excluderea limitei superioare.

Adică:

1600 la mai puțin de 1800 lei,
1800 „ „ „ „ 2000 „ ,
2000 „ „ „ „ 2200 „ ,
. . .

Termenul *limite de clasă* se folosește pentru a desemna cea mai mică și cea mai mare valoare într-o clasă de valori dată. Dar, în cazul unei variabile

V.TREBICI

continue, operîndu-se rotunjiri, se face distincție între limitele de clasă „enunțate" și „reale". Să considerăm, de exemplu, clasele de retribuție medie orară alcătuite în modul următor :

8 — 8,9 lei,
9 — 9,9 „ ,
10 — 10,9 „ ,
. . .

Se presupune că retribuțiile au fost aproximate la prima zecimală, așa încît clasele nu încep în realitate de la valorile indicate în scară. Astfel, clasa indicată prin limitele 8—8,9 lei conține de fapt retribuțiile mai mari sau egale cu 7,95 lei și mai mici decît 8,95 lei. Valorile 7,95 și 8,95 sînt numite *limite reale de clasă*. În această situație, particulară unor variabile continue, centrul (mijlocul) clasei rămîne același, oricare ar fi considerate limitele de clasă — cele „enunțate" sau cele „reale" :

$$(8 + 8{,}9)/2 = (7{,}95 + 8{,}95)/2 = 8{,}45 \text{ lei.}$$

Repartiția cumulativă de frecvențe este formată prin adunarea succesivă a frecvențelor ordinare (absolute sau relative → *Frecvență*).

O repartiție cumulativă de frecvențe ne permite să determinăm numărul total de observații inferioare sau superioare unei valori date.

Exemplul 5. Repartiția țărilor după produsul național brut pe locuitor în 1977.

Tabelul 13

PNB*) pe locuitor (dolari) $x_{i-1} \leqslant x < x_i$	Număr de țări**) f_i	Frecvențe cumulate : Număr de țări cu	
		PNB/loc. $< x_i$ F_i	PNB/loc $\geqslant x_{i-1}$ F_i'
(1)	(2)	(3)	(4)
Pînă la 200	21	21	178
200 — 500	41	62	157
500 — 2000	56	118	116
2000 — 5000	31	149	60
5000 și peste	29	178	29

*) La prețurile pieței.
**) Nu sînt incluse : Iran, Kampuchia Democrată, Liban, Uganda și R. S. Vietnam.
Sursa : *World Bank Atlas*, Population, per Capita Product and Growth Rates, 1979.

Frecvențele absolute au fost adunate succesiv în sensul valorilor crescătoare (col. 3) și în sensul valorilor descrescătoare (col. 4) ale variabilei. Numerele obținute în col. 4 sînt numite și *frecvențe retrocumulate*.

În mod analog se formează *repartiția cumulativă de frecvențe relative*.

VII. TENDINȚA CENTRALĂ (MEDII UZUALE)

VALOARE MEDIE *(medie)*. Număr susceptibil de a rezuma ansamblul valorilor observate ale unei variabile statistice, reprezentat de o funcție de aceste valori.

Fie $\varphi(x)$ o funcție *crescătoare* sau *descrescătoare*.

Dacă observațiile dau ca rezultat numerele tone):

$$x_1, x_2, \ldots, x_n$$

numărul $\bar{x}$ (citit „x barat"), definit prin relația

$$\varphi(\bar{x}) = \frac{1}{n}\,[\varphi(x_1) + \varphi(x_2) + \ldots + \varphi(x_n)],$$

este φ — medie, adică **m.** definită plecînd de la funcția $\varphi(x)$.

Conform relației (1), o **m.** exprimă *nivelul* atins de o variabilă *numeronc.* ea este *uniform* repartizată pe unitățile care compun o populație *det.*

Conceptul de **m.** este familiar. De obicei, cuvîntul **m.**, fără nici o alt x_i și de care, se folosește în accepția de *medie aritmetică* — **m.** fundamentală [illegible] și comodă în practică. Alte **m.** uzuale sînt : *media armonică* *, *media geometrică* *, *media pătratică* *.

După expresia lui $\varphi(x)$, se obține drept caracteristică $\bar{x}$:

a) pentru $\varphi(x) = x$, **m.** aritmetică $\bar{x}_a$;

b) pentru $\varphi(x) = \frac{1}{x}$ (sau x^{-1}), **m.** armonică $\bar{x}_h$;

c) pentru $\varphi(x) = \log x$, **m.** geometrică $\bar{x}_g$;

d) pentru $\varphi(x) = x^2$, **m.** pătratică $\bar{x}_q$.

Diferitele forme de medie se înrudesc : formulele lor provin dintr-o aceeași proprietate :

$$\varphi(x_1) + \varphi(x_2) + \ldots + \varphi(x_n) = \varphi(\bar{x}) + \varphi(\bar{x}) + \ldots + \varphi(\bar{x}). \qquad (2)$$

Sub formă generalizată, **m.** este dată de asemenea de formula :

$$\bar{x}^k = \frac{1}{n}\sum_{i=1}^{n} x_i^k, \qquad (3)$$

m. *de ordinul* k a valorilor lui X (presupuse pozitive) *). Făcînd să varieze k, se obține o infinitate de medii. În particular, dacă mărimea lui k variază între -1 și $+2$, se obțin :

a) $k = -1$, **m.** armonică $\bar{x}_h$;

b) $k = 0$, **m.** geometrică $\bar{x}_g$;

c) $k = 1$, **m.** aritmetică $\bar{x}_a$;

d) $k = 2$, **m.** pătratică $\bar{x}_q$.

Relația între aceste **m.** este următoarea :

$$\bar{x}_h < \bar{x}_g < \bar{x}_a < \bar{x}_q.$$

*) Numită și *formula generalizată a mediilor*.

Verificată, de exemplu, pe un șir de cinci valori :

$$12,\quad 14,\quad 15,\quad 18,\quad 20$$

se află

$$\bar{x}_h = 15{,}29,\ \bar{x}_g = 15{,}54,\ \bar{x}_a = 15{,}80,\ \text{și}\ \bar{x}_q = 16{,}05.$$

se numește :

simplă, cînd se aplică la o serie de valori individuale (n valori x_i, Se pr; 2, ..., n) și

clasei *ponderată*, în cazul unei repartiții de frecvențe (n valori x_i ; cu ponderile p_i, prin li 2, ..., n).

7,95 le a din urmă, formulele (1) (2) și (3) devin (scrise prescurtat) :

reale de c

(mijlocul)

cele „enu

$$\varphi(\bar{x}) = \frac{1}{\sum_{i=1}^{n} p_i} \sum_{i=1}^{n} p_i \varphi(x_i) \tag{4}$$

R

$$\sum_{i=1}^{n} p_i \varphi(x_i) = \varphi(\bar{x}) \sum_{i=1}^{n} p_i \tag{5}$$

$$\bar{x}^k = \frac{1}{\sum_{i=1}^{n} p_i} \sum_{i=1}^{n} p_i x_i^k. \tag{6}$$

Ponderea unei valori x_i este numărul p_i asociat acelei valori, de natură să-i dea o anumită importanță relativă (în raport cu importanța atribuită celorlalte valori ale variabilei).

Într-o *repartiție de frecvențe* * constituie și sînt numite de obicei :

— *ponderi*, frecvențele absolute f_i a căror sumă Σf_i este indiferentă și

— *coeficienți de ponderare*, frecvențele relative $\tilde{f}_i = \dfrac{f_i}{\Sigma f_i}$ a căror sumă $\Sigma \tilde{f}_i$ este 1.

Notă

M. ponderată se reduce la **m.** simplă atunci cînd frecvențele valorilor repartiției sînt egale între ele $f_i = f$.

Relativ la **m.**, ca noțiune generică, sînt de remarcat următoarele particularități :

a) **m.** este cuprinsă între cea mai mică și cea mai mare dintre valorile luate de variabilă ;

b) **m.** este egală cu valorile date, dacă acestea sînt egale între ele :

c) în raport cu valorile concrete pe care le sintetizează, **m.** apare ca o mărime abstractă ;

d) substituită fiecăreia dintre valorile seriei, **m.** păstrează o anumită proprietate a ansamblului de valori (care conduce la formula de calcul a fiecărei forme de **m.**).

V.TREBICI

MEDIE ARITMETICĂ ($\bar{x}_a$). Media obținută ca raport între suma valorilor observate ale unei variabile și numărul lor.
Medie aritmetică simplă (a unei serii de n valori x_i, $i = 1, 2, \ldots, n$):

$$\bar{x}_a = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\Sigma x}{n}. \qquad (1)$$

Exemplul 1. Producția mondială de grîu în anii 1974—1978 (milioane tone):

1974	1975	1976	1977	1978
360	355	417	387	441

Sursa: *Anuarul statistic al R. S. România, 1976, ..., 1980.*
Producția de grîu în medie pe an:

$$\bar{x}_a = \frac{360 + 355 + 417 + 387 + 441}{5} = \frac{1960}{5} = 392 \text{ mil. tone.}$$

Media aritmetică ponderată (a unei repartiții de frecvențe, de valori x_i și de frecvențe f_i, $i = 1, 2, \ldots, n$):

$$\bar{x}_a = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_n f_n}{f_1 + f_2 + \ldots + f_n} = \frac{\Sigma xf}{\Sigma f}, \qquad (2)$$

care se mai poate scrie

$$\bar{x}_a = x_1 \frac{f_1}{\Sigma f} + x_2 \frac{f_2}{\Sigma f} + \ldots + x_n \frac{f_n}{\Sigma f} = \Sigma x \frac{f}{\Sigma f}, \qquad (3)$$

$\frac{f_i}{\Sigma f_i}$ fiind frecvența relativă (coeficient de ponderare).

Exemplul 2. Numărul mediu de rebuturi într-un lot conform repartiției din tab. 18.

$$\bar{x}_a = \frac{(0 \times 13) + (1 \times 21) + (2 \times 36) + (3 \times 19) + (4 \times 7) + (5 \times 3) + (6 \times 1)}{13 + 21 + 36 + 19 + 7 + 3 + 1} =$$

$$= 1{,}99 \approx 2 \text{ rebuturi.}$$

Expresia (3) arată că **m.a.** poate fi calculată și cu ajutorul frecvențelor relative, ajungîndu-se la același rezultat.
Într-adevăr.

$$\bar{x}_a = 0 \times \frac{13}{100} + 1 \times \frac{21}{100} + 2 \times \frac{36}{100} + 3 \times \frac{19}{100} + 4 \times \frac{7}{100} +$$

$$+ 5 \times \frac{3}{100} + 6 \times \frac{1}{100} = 1{,}99 \approx 2 \text{ rebuturi.}$$

Modificarea proporțională a frecvențelor nu modifică valoarea medie.
În cazul unei *repartiții de frecvențe pe clase de valori* * (pe o scară de intervale), formula de definiție a *m.a.p.* nu poate fi aplicată direct, întrucît frecvențele privesc grupe de valori și nu valori distincte ale variabilei. Calculul mediei include o ipoteză suplimentară despre valoarea x-ilor: se presupune distribuirea uniformă a frecvențelor în interiorul intervalului dat și, în consecință, se admite că valorile x sînt reprezentate prin centrele intervalelor. Odată determinat centrul fiecărui interval, se ajunge la calcul **m.a.** în cazul unei repartiții pe valori distincte, ca în aplicația precedentă.

Exemplul 3. Retribuția medie lunară a personalului întreprinderii, potrivit repartiției din tab. 19.

$$\bar{x}_a = \frac{\frac{1600+1800}{2} \times 5 + \frac{1800+2000}{2} \times 13 + \ldots + \frac{3600+3800}{2} \times 11}{800} =$$

$$= \frac{2177600}{800} = 2722 \text{ lei}$$

sau ceea ce este totuna

$$\bar{x}_a = \frac{\frac{1600+1800}{2} \times 1{,}1 + \frac{1800+2000}{2} \times 1{,}6 + \ldots + \frac{3600+3800}{2} \times 1{,}4}{100} =$$

$$= 2722{,}6 \text{ lei.}$$

Cum intervalele au aceeași amplitudine, calculele se pot prescurta, efectuînd o *schimbare de variabilă.*
Se definește o nouă variabilă X', numită *variabilă auxiliară*, prin transformare liniară *)

$$x'_i = \frac{x_i - x_0}{d} \text{ (în care } x_0 \text{ și } d \text{ sînt constante)}$$

sau de formă echivalentă

$$x_i = dx'_i + x_0. \qquad (4)$$

Fiecărei valori a centrului de interval x_i i se substituie o valoare a variabilei auxiliare x'_i. **M.a.** a lui X se deduce ușor din cea a lui X'.
Cu frecvențele introduse în formula (4)

$$x_i f_i = dx'_i f_i + x_0 f_i$$

*) Se numește *transformare liniară* o transformare în care ecuația ce permite să se treacă de la variabila X la variabila X' este o ecuație liniară. În cazul dat:

$$x'_i = \frac{1}{d} x_i - \frac{x_0}{d}.$$

adunînd membru cu membru diversele egalităţi

$$\begin{aligned} x_1 f_1 &= d x'_1 f_1 + x_0 f_1 \\ x_2 f_2 &= d x'_2 f_2 + x_0 f_2 \\ &\cdots\cdots\cdots \\ x_n f_n &= d x'_n f_n + x_0 f_n \end{aligned}$$

$$\Sigma xf = d\,\Sigma' xf + x_0 \Sigma f \qquad (5)$$

şi împărţind la Σf, rezultă

$$\frac{\Sigma xf}{\Sigma f} = d\frac{\Sigma x'f}{\Sigma f} + x_0 \frac{\Sigma f}{\Sigma f} \qquad (6)$$

sau

$$\bar{x}_a = d\bar{x}'_a + x_0.$$

Altfel formulat, procedeul constă în aplicarea combinată a proprietăţilor **m.a**, care permit ca fiecare valoare a variabilei să fie micşorată o dată prin scăderea unei constante ($x_i - x_0$) şi a doua oară prin împărţirea restului obţinut la o altă constantă $[(x_i - x_0)/d]$. Cantitatea cu care se scad termenii repartiţiei (x_0), numită şi *origine*, poate fi oricare dintre termenii repartiţiei şi chiar o mărime din afara ei. În general, se reţine ca nouă origine o valoare apropiată de valoarea centrală a repartiţiei: centrul *intervalului modal* * sau centrul *intervalului median* *. Drept nouă unitate de măsură (d) se alege mărimea intervalului de clasă, care dă, pentru noua variabilă X', numerele cele mai simple.

Mersul operaţiilor este arătat în tab. 14.

Tabelul 14

Centrul intervalului x_i	Frecvenţa f_i	$x'_i = \frac{x_i - 2700}{200}$	$x'_i f_i = \frac{x_i - 2700}{200} f_i$
(1)	(2)	(3)	(4)
1700	9	−5	− 45
1900	13	−4	− 52
2100	27	−3	− 81
2300	64	−2	−128
2500	148	−1	−148
2700	261	0	0
2900	123	1	123
3100	89	2	178
3300	34	3	102
3500	21	4	84
3700	11	5	55
Total	800	—	88

$$\bar{x}_a = 2700 + 200\frac{88}{200} = 2700 + 22 = 2722 \text{ lei.}$$

Evident, media calculată pentru o repartiție cu intervale exprimă, cu oarecare aproximație, mărimea medie reală a variabilei, deoarece nu poate fi vorba de o corespondență perfectă între ipoteza de calcul admisă și realitate. Gradul de aproximație este în funcție de două elemente : uniformitatea repartiției unităților din grupe în interiorul intervalelor și dimensiunea intervalelor de clasă. Media va fi cu atît mai precisă cu cît repartiția unităților cuprinse în fiecare interval este mai uniformă și cu cît mărimea intervalelor este mai redusă. Repartițiile statistice cu o scară de intervale prea mari nu permit o evaluare satisfăcătoare a mediei.

Proprietăți.

1. *Suma algebrică a abaterilor valorilor observate de la media lor aritmetică este nulă.*

Demonstrată în variantele (a) serie de valori individuale și (b) repartiție de frecvențe :

(a)	(b)
$x_1 - \bar{x}$	$(x_1 - \bar{x}) f_1$
$x_2 - \bar{x}$	$(x_2 - \bar{x}) f_2$
⋮ ⋮	⋮ ⋮ ⋮
$x_n - \bar{x}$	$(x_n - \bar{x}) f_n$
$S = \Sigma x_i - n\bar{x}$	$S = \Sigma(x_i - \bar{x}) f_i = \Sigma x_i f_i - \bar{x} \Sigma f_i$

or, prin definiție,

$$\bar{x} = \frac{\Sigma x_i}{n} \qquad \bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i},$$

deci

$$\Sigma x_i = n\bar{x} \qquad \Sigma x_i f_i = \bar{x} \Sigma f_i$$

și

$$S = n\bar{x} - n\bar{x} = 0 \quad S = \bar{x} \Sigma f_i - x \Sigma f_i = 0.$$

O verificare numerică a proprietății enunțate :

(a) $\Sigma(x_i - \bar{x}) = \Sigma x_i - n\bar{x} = 1960 - (5 \times 392) = 0$ (→ exemplul 1).

(b) $\Sigma(x_i - \bar{x}) f_i = \Sigma x_i f_i - \bar{x} \Sigma f_i = 2177600 - (2722 \times 800) = 0$

(→ exemplul 3).

2. *Suma pătratelor abaterilor valorilor observate de la media lor aritmetică este inferioară sumei pătratelor abaterilor în raport cu oricare altă valoare.*

Fie o serie de n valori x_i ($i = 1, 2, \ldots, n$) și x_0 o valoare alta decît m.a. $\bar{x}$ a acestor valori.

Fie $(x_i - x_0)^2 - (x_i - \bar{x})^2 = a_i$.

Problema constă în a demonstra că $\Sigma a_i > 0$.

Avem :

$$a_i = (x_i - x_0)^2 - (x_i - \bar{x})^2 = -2x_i(x_0 - \bar{x}) + x_0^2 - \bar{x}^2,$$

deci

$$\Sigma a_i = -2\, x_i(x_0 - \bar{x}) + n(x_0^2 - \bar{x}^2).$$

Şi cum

$$\Sigma x_i = n\bar{x},$$

rezultă

$$\Sigma a_i = n(-2\bar{x}x_0 + 2\bar{x}^2 + x_0^2 - \bar{x}^2) = n(x_0 - \bar{x})^2,$$

de unde

$$\Sigma a_i > 0.$$

Mai simplu, proprietatea poate fi enunţată astfel : suma pătratelor abaterilor în raport cu o valoare este minimală, cînd această valoare este **m.a.**

3. **M.a.** *a unui ansamblu de valori compus din două sau mai multe subansambluri poate fi exprimată direct în funcţie de m.a. ale subansamblurilor, aplicînd formula* :

$$\bar{x}_a = \frac{\bar{x}_1 f_1 + \bar{x}_2 f_2 + \ldots + \bar{x}_n f_n}{f_1 + f_2 + \ldots + f_n} = \frac{\Sigma\, \bar{x}_i f_i}{\Sigma f_i}, \qquad (7)$$

în care $\bar{x}_1, f_1$; $\bar{x}_2, f_2$; ... ; x_n, f_n sînt **m.a.** parţiale cu frecvenţele respective.

Exemplul 4. Suprafaţa cultivată, producţia totală şi producţia medie la hectar de grîu şi secară (medii anuale)

Tabelul 15

	1971—1975	1976—1979
1. Suprafaţa cultivată (mii hectare)	2463,9	2300,8
2. Producţia totală (mii tone)	5449,7	6075,5
3. Producţia medie la hectar (chintale)	22,1	26,4

Sursa : *Anuarul statistic al R. S. România 1980*, p. 257—259.

Mediile anuale pe întreaga perioadă 1971—1979 vor fi:

(1) $$\frac{(2463{,}9 \times 5) + (2300{,}8 \times 4)}{5 + 4} = 2391{,}4 \text{ mii hectare.}$$

(2) $$\frac{(5449{,}7 \times 5) + (6075{,}5 \times 4)}{5 + 4} = 5727{,}8 \text{ mii tone.}$$

(3) $$\frac{57278}{2391{,}8} = 24 \text{ chintale.}$$

Formula (7) devine

$$\bar{x}_a = \frac{\Sigma \bar{x}_i}{n}.$$

în cazul cînd subansamblurile componente conţin acelaşi număr de unităţi (frecvenţe egale).
Se spune că **m.a.** admite o *teoremă de adiţiune.*
4. *Dacă se înlocuieşte fiecare valoare observată a variabilei cu* **m.a.**, *suma valorilor nu se schimbă.*
Din această proprietate se deduce însăşi formula de definiţie a **m.a.** Astfel, pentru :
a) seria de valori individuale (n valori x_i, $i = 1, 2, \ldots, n$)

$$x_1 + x_2 + \ldots + x_n = \bar{x} + \bar{x} + \ldots + \bar{x}$$

sau

$$x_1 + x_2 + \ldots + x_n = n\bar{x},$$

de unde

$$\bar{x}_a = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

scrisă prescurtat

$$\bar{x}_a = \frac{\Sigma x_i}{n};$$

b) repartiţia de frecvenţe (n valori x_i, cu frecvenţele f_i, $i = 1, 2, \ldots, n$)

$$x_1 f_1 + x_2 f_2 + \ldots + x_n f_n = \bar{x} f_1 + \bar{x} f_2 + \ldots + \bar{x} f_n$$

sau

$$x_1 f_1 + x_2 f_2 + \ldots + x_n f_n = \bar{x}(f_1 + f_2 + \ldots + f_n),$$

de unde

$$\bar{x}_a = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_n f_n}{f_1 + f_2 + \ldots + f_n},$$

scrisă prescurtat

$$\bar{x}_a = \frac{\Sigma x_i f_i}{\Sigma f_i}.$$

MEDIE ARMONICĂ ($\bar{x}_h$). Mărimea inversă a mediei aritmetice calculată din mărimile inverse ale valorilor variabilei.
Medie armonică simplă (a unei serii de n valori x_i, $i = 1, 2, \ldots, n$).

$$\bar{x}_h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}} = \frac{n}{\Sigma \frac{1}{x}}. \qquad (1)$$

Exemplul 1. Un muncitor cheltuiește pentru executarea unei operații 20 minute, iar altul 30 minute. Care este timpul mediu cheltuit?
Alegerea formei de medie este, în acest caz, determinată de relația :

$$\frac{T}{t} = Q, \tag{2}$$

în care
T înseamnă timpul lucrat ;
t — timpul cheltuit pentru o operație ;
Q — numărul de operații.
Într-o oră s-au executat :

$$t_1 = 20 \text{ minute} \ldots \frac{60}{20} = 3 \text{ operații},$$

$$t_2 = 30 \quad \text{„} \quad \ldots \frac{60}{30} = 2 \quad \text{„} \quad ,$$

$$\text{total} \ldots \left(\frac{60}{20} + \frac{60}{30} \right) = 5 \quad \text{„} \quad ,$$

de unde

$$\bar{t} = \frac{\text{Timpul total lucrat}}{\text{Numărul total de operații}},$$

adică

$$\bar{t} = \frac{60 + 60}{\frac{60}{20} + \frac{60}{30}},$$

$$\bar{t} = \frac{1 + 1}{\frac{1}{20} + \frac{1}{30}} = \frac{2}{\frac{3 + 2}{60}} = \frac{120}{5} = 24 \text{ minute}.$$

În ipoteza unui consum de timp de muncă ($\bar{t}_i$) egal cu timpul mediu ($\bar{t} = 24$ minute) pentru fiecare din cei doi muncitori, într-o oră aceștia ar fi executat

$$\frac{60}{24} + \frac{60}{24} = \frac{120}{24} = 5 \text{ operații},$$

cantitate egală cu cea efectiv realizată.

V.TREBICI

Medie armonică ponderată (a unei repartiții de frecvențe, de valori x_i și de frecvențe f_i, $i = 1, 2, \ldots, n$).

$$\bar{x}_h = \frac{f_1 + f_2 + \ldots + f_n}{\frac{1}{x_1} f_1 + \frac{1}{x_2} f_2 + \ldots + \frac{1}{x_n} f_n} = \frac{\Sigma f}{\Sigma \frac{1}{x} f}. \qquad (3)$$

Exemplul 2. Se consideră datele înregistrate, pentru aceeași variabilă (exemplul 1), în cursul mai multor cronometrări succesive, repartizate în clase de valori egale :

Tabelul 16

Consumul specific de muncă (om-ore/operație) t_i	număr de om-ore lucrate T_i	$\frac{1}{t_i} T_i$
0,30	27	90
0,35	42	120
0,40	84	210
0,45	162	360
0,50	65	130
0,55	33	60
0,60	18	30
Total	431	1000

Potrivit relației (2)

$$\frac{T_1}{t_1} + \frac{T_2}{t_2} + \ldots + \frac{T_n}{t_n} = \frac{T_1}{\bar{t}} + \frac{T_2}{\bar{t}} + \ldots + \frac{T_n}{\bar{t}} \qquad (4)$$

sau

$$\frac{T_1}{t_1} + \frac{T_2}{t_2} + \ldots + \frac{T_n}{t_n} = \frac{1}{\bar{t}} (T_1 + T_2 + \ldots + T_n),$$

de unde ecuația mediei

$$\bar{t} = \frac{T_1 + T_2 + \ldots + T_n}{\frac{1}{t_1} T_1 + \frac{1}{t_2} T_2 + \ldots + \frac{1}{t_n} T_n} = \frac{\Sigma T}{\Sigma \frac{1}{t} T},$$

deci

$$\bar{t} = \frac{431}{1000} = 0{,}431 \text{ om-ore, aproximativ 23 minute.}$$

Utilizarea **m.a.** devine obligatorie în cazul *relațiilor de proporționalitate inversă* *) ca în exemplul evocat. Într-adevăr, dacă privim variabila t o derivată din T Q, ne aflăm în prezența unei relații inverse : consumul specific de muncă (t), față de aceeași cantitate de cheltuieli de muncă (T joacă rol de pondere), crește pe măsura descreșterii volumului de operații (Q) și invers. În varianta în care ponderea lui t s-ar face cu Q, existența unei proporționalități directe (t față de aceeași mărime a lui Q crește pe măsura creșterii lui T și invers) justifică aplicarea mediei aritmetice. Adoptînd $T = tQ$ și notațiile din formula generală (x în loc de t și f în loc de Q) se vede limpede că în funcție de datele disponibile se calculează fie media aritmetică, fie **m.a.** și se ajunge la aceeași valoare medie :

$$\bar{x} = \frac{\Sigma\, xf}{\Sigma\, \frac{1}{x}\, xf} = \frac{\Sigma\, xf}{\Sigma f}.$$

M.a. apare astfel ca o formă particulară a mediei aritmetice.

MEDIE GEOMETRICĂ ($\bar{x}_g$). Medie calculată prin extragerea rădăcinii de un ordin egal cu numărul valorilor observate ale variabilei din produsul valorilor Medie geometrică simplă (a unei serii de n valori x_i, $i = 1, 2, \ldots, n$).

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = (\Pi\, x_i)^{\frac{1}{n}} \qquad (1)$$

sau exprimată în logaritmi

$$\log \bar{x}_g = \frac{\log x_1 + \log x_2 + \ldots + \log x_n}{n} = \frac{\Sigma \log x_i}{n},$$

de unde

$$\bar{x}_g = \text{antilog} \frac{\Sigma \log x_i}{n}. \qquad (2)$$

M.g. este deci egală cu antilogaritmul mediei aritmetice a logaritmilor valorilor seriei. De aici și denumirea de *medie logaritmică* sub care se întîlnește uneori **m.g.** Se aplică în mod curent la calculul ritmului mediu de variație a fenomenelor.

Exemplul 1. Dinamica venitului național al României, 1976—1979.

Tabelul 17

Anul	Indici (anul precedent = 100)
1976	110,5
1977	107,9
1978	107,6
1979	107,5

Sursa : *Anuarul statistic al R. S. România 1980*, p. 91.

*) Se spune că **m.a.** a fost introdusă în teoria muzicii pentru compararea lungimii coardelor vibrante cu înălțimea sunetului lor. Cum „numărul vibrațiilor este *invers proporțional* cu lungimea coardei, nota medie, care corespunde numerelor de vibrații date de două coarde de lungimi diferite, este dată de o coardă a cărei lungime este **m.a.** a lungimilor (A. Piatier, *Statistique et observation économique*, t. I, Presses Universitaires de France, Paris, 1961, p 217—218).

Indicele mediu anual de creștere :

$$\bar{x}_g = \sqrt[4]{1{,}105 \times 1{,}079 \times 1{,}076 \times 1{,}075}$$

logaritmînd

$$\log \bar{x}_g = \frac{\log 1{,}105 + \log 1{,}079 + \log 1{,}076 + \log 1{,}075}{4} =$$

$$= \frac{0{,}04336 + 0{,}03302 + 0{,}03181 + 0{,}03141}{4} = \frac{0{,}13960}{4} = 0{,}03490$$

$$\bar{x}_g = \text{antilog} \cdot 0{,}03490 = 1{,}084 \text{ sau } 108{,}4\%.$$

Medie geometrică ponderată (a unei repartiții de frecvențe, de valori x_i și de frecvențe f_i, $i = 1, 2, \ldots, n$).

$$\bar{x}_g = \sqrt[\Sigma f_i]{x_1^{f_1},\ x_2^{f_2},\ \ldots,\ x_n^{f_n}} = (\Pi\, x_i^{f_i})^{\frac{1}{\Sigma f_i}} \tag{3}$$

sau exprimată în logaritmi

$$\log \bar{x}_g = \frac{f_1 \log x_1 + f_2 \log x_2 + \ldots + f_n \log x_n}{f_1 + f_2 + \ldots + f_n} = \frac{\Sigma f_i \log x_i}{\Sigma f_i},$$

de unde

$$\bar{x}_g = \text{antilog} \frac{\Sigma f_i \log x_i}{\Sigma f_i}. \tag{4}$$

Exemplul 2. Venitul național al țării noastre a crescut, în medie pe an, cu 11,3% în anii 1971—1975 și cu 8,4% în anii 1976—1979. Care a fost creșterea medie anuală în cei nouă ani?

$$\bar{x}_g = \sqrt[9]{1{,}113^5 \times 1{,}084^4}$$

$$\log \bar{x}_g = \frac{(5 \times \log 1{,}113) + (4 \times \log 1{,}084)}{5 + 4} =$$

$$= \frac{(5 \times 0{,}04650) + (4 \times 0{,}03503)}{9} = \frac{0{,}37262}{9} = 0{,}04140$$

$$\bar{x}_g = \text{antilog} \cdot 0{,}04140 = 1{,}1 \text{ sau } 110\%.$$

În limbajul curent, se spune că *ritmul* (*rata*) *mediu anual* de creștere a venitului național a fost de 110—100 = 10% în perioada 1971—1979.

Proprietăți

1. **M.g.** *a unui ansamblu de valori compus din două sau mai multe subansambluri poate fi exprimată direct în funcție de* **m.g.** *ale subansamblurilor*, aplicînd formula :

$$\bar{x}_g = \text{antilog} \frac{\Sigma f_i \log \bar{x}_i}{\Sigma f_i}, \tag{5}$$

în care :
$\bar{x}_i$ este **m.g.** parțială ;
f_i — numărul de observații în subansamblu.
Formula (5) devine

$$\bar{x}_g = \text{antilog} \frac{\Sigma \bar{x}_i}{n} \tag{6}$$

în cazul în care subansamblurile componente conțin același număr de observații (frecvențele sînt egale între ele). Așadar, ca și media aritmetică, **m.g.** admite o teoremă de adițiune.
Exemplul 2 ilustrează proprietatea enunțată, cu aplicarea formulei (5).
2. **M.g.** *a raportului a două variabile este dată de raportul* **m.g.** *ale celor două variabile.*
Fie seriile de valori ale variabilelor X, Y, Z, cu $Z = X/Y$:

$$x_1, x_2, \ldots, x_n$$

$$y_1, y_2, \ldots, y_n$$

$$z_1, z_2, \ldots, z_n,$$

unde

$$z_i = \frac{x_i}{y_i}. \tag{7}$$

Atunci

$$\bar{x}_{g(z_i)} = \sqrt[n]{z_1 \cdot z_2 \cdot \ldots \cdot z_n} = \frac{\sqrt[n]{x_1 \cdot x_2 \cdot \ldots x_n}}{\sqrt[n]{y_1 \cdot y_2 \cdot \ldots y_n}} = \frac{\bar{x}_{g(x_i)}}{\bar{x}_{g(y_i)}}.$$

Exemplul 3. În intervalul 1976—1979, producția globală industrială (X) a României a crescut, în medie pe an, cu 10,3%, iar forța de muncă (Y) ocupată în industrie, cu 2,7%. Cum raportul între prima și cea de-a doua variabilă exprimă productivitatea muncii (Z), ritmul mediu anual de creștere a acesteia a fost :

$$\frac{110{,}3}{102{,}7} \times 100 - 100 = 7{,}4\%.$$

3. **M.g.** *a unui produs de variabile este dată de produsul* **m.g.** *ale variabilelor respective.*

Considerînd acum (→ relaţia 7)

$$x_i = z_i \cdot y_i \tag{8}$$

avem

$$\bar{x}_{g(x_i)} = \sqrt[n]{x_1 x_2 \dots x_n} = \sqrt[n]{z_1 z_2 \dots z_n} \cdot \sqrt[n]{y_1 y_2 \dots y_n} = \bar{x}_{g(z_i)} \cdot \bar{x}_{g(y_i)}.$$

Altfel spus : m.g. a produsului este produsul m.g.

Exemplul 4. Se presupun cunoscute ritmurile medii anuale de creştere a productivităţii muncii (7,4%) şi forţei de muncă (2,%) în perioada 1976—1979 (ex. 3). Ritmul mediu anual de creştere a producţiei globale va fi :

$$\frac{107{,}4 \times 102{,}7}{100} \times 100 - 100 = 10{,}3\%.$$

4. *Dacă se înlocuieşte fiecare valoare observată a variabilei cu* **m.g.**, *produsul valorilor nu se schimbă.*

Această proprietate defineşte **m.g.**, care se bazează pe proporţia geometrică. Astfel, pentru o serie de n valori x_i $(i = 1, 2, \dots, n)$:

$$x_1 \cdot x_2 \dots x_n = \bar{x} \cdot \bar{x} \dots \bar{x}$$

sau

$$x_1 \cdot x_2 \dots x_n = \bar{x}^n,$$

de unde

$$\bar{x} = \sqrt[n]{x_1 x_2 \dots x_n}.$$

Ca aplicaţie la seria de indici din exemplul 1 :

$$1{,}105 \times 1{,}079 \times 1{,}076 \times 1{,}075 = \bar{I}^4$$

şi

$$\bar{I}_{76-79} = \sqrt[4]{1{,}105 \times 1{,}079 \times 1{,}076 \times 1{,}075} = \sqrt[4]{1{,}38}.$$

Venitul naţional a fost de 361,9 mild. lei în 1975 şi de 49,93 mld. lei în 1979. În funcţie de nivelul anului 1975, venitul naţional în 1979 se scrie :

$$361{,}9 \times 1{,}105 \times 1{,}079 \times 1{,}076 \times 1{,}075 = 499{,}3 \text{ mld. lei.}$$

Substituit indicilor concreţi, indicele mediu păstrează egalitatea scrisă. Adică :

$$361{,}9 \times \bar{I}^4 = 499{,}3 \text{ mld. lei.}$$

Ca urmare

$$\bar{I}_{76-79} = \sqrt[4]{\frac{499{,}3}{361{,}9}}$$

$$\log I_{76-79} = \frac{\log 499{,}3 - \log 361{,}9}{4} = \frac{2{,}69836 - 2{,}55859}{4} =$$

$$= \frac{0{,}13977}{4} = 0{,}03494$$

$$I_{76-79} = \text{antilog } 0{,}03494 = 1{,}084 \text{ sau } 108{,}4\,\%.$$

Deci, în general:

$$I = \sqrt[n-1]{\frac{x_n}{x_0}}, \tag{9}$$

în care:

x_n este termenul final al seriei de mărimi absolute,

x_0 — termenul inițial al seriei de mărimi absolute,

n — numărul de intervale (momente) de timp.

Este ușor de observat că, dispunînd de *indicii cu bază constantă**, atunci:

$$I = \sqrt[n-1]{\frac{I_{n/0}}{I_{0/0}}}. \tag{10}$$

Cu datele exemplului

$$I_{75/75} = 100\,\%\,;\ I_{79/75} = 138\,\%$$

$$I_{76-79} = \sqrt[4]{\frac{138}{100}}\,;\ \log I_{76-79} = \frac{\log 138 - \log 100}{4} = \frac{2{,}13988 - 2}{4} =$$

$$= \frac{0{,}13988}{4} = 0{,}03497$$

$$I_{76-79} = \text{antilog } 0{,}03497 = 1{,}084 \text{ sau } 108.4\,\%.$$

MEDIE PĂTRATICĂ ($\bar{x}_q$). Medie egală cu rădăcina pătrată a mediei aritmetice a pătratelor valorilor variabilei.

Medie pătratică simplă (a unei serii de n valori x_i, $i = 1, 2, \ldots, n$).

$$\bar{x}_q = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n}} = \sqrt{\frac{\Sigma x_i^2}{n}}. \tag{1}$$

Medie pătratică ponderată (a unei repartiții de frecvențe, de valori x_i și de frecvențe f_i, $i = 1, 2 \ldots n$)

$$\bar{x}_q = \sqrt{\frac{x_1^2 f_1 + x_2^2 f_2 + \ldots + x_n^2 f_n}{1 + f_2 + \ldots + f_n}} = \sqrt{\frac{\Sigma x_i^2 f_i}{\Sigma f_i}}$$

Deşi ar putea fi calculată pentru orice valori (pozitive, nule sau negative), **m.p.** nu are sens ca valoare centrală decît dacă toate valorile x_i sînt pozitive. **M.p.** se aplică la calcularea *abaterii medii pătratice**.

MEDIANĂ (*Me*). Valoare situată la mijlocul seriei sau repartiţiei statistice avînd valorile aranjate în ordine de mărime crescătoare sau descrescătoare. **Me.** împarte colecţia de unităţi observate în două grupe egale ca număr.

1. Pentru o serie cu un număr impar de valori, $2k + 1$, **me.** este valoarea de rangul $k + 1$.

Exemplul 1. **Me.** seriei :

4 7 12 26 32 38 59.
↑
Me

Dacă seria cuprinde un număr par de valori, 2 k, **me.** se găseşte între valorile de rangul k şi $k + 1$ şi, în general, se convine să fie reprezentată de media aritmetică a celor două valori.

Exemplul 2. **Me.** seriei

4 7 12 26 32 38.
Interval median

Se ia deci ca **me.** $(12 + 26)/2 = 19$.

2. În cazul unei repartiţii pe valori distincte, semnificaţia **me.** apare frecvent nesigură.

Exemplul 3. Un control de calitate efectuat asupra 100 de loturi dintr-un produs a dat rezultatele următoare :

Tabelul 18

Număr de rebuturi într-un lot x_i	Număr de loturi f_i	Număr cumulat de loturi F_i
0	13	13
1	21	34
2	36	70
3	19	89
4	7	96
5	3	99
6	1	100
Total	100	

Numărul median de rebuturi s-ar considera 2. Dar valoarea reţinută nu răspunde definiţiei **me.** în sensul de valoare cu proprietatea ca valorile mai mari sau mai mici decît ea să apară cu frecvenţe egale. Nu sînt decît 30% de loturi cu un număr de rebuturi superior numărului median, în loc de 50%. Alteori

proporția ar putea să fie mult mai mică (20, 10 sau chiar 5%) și atunci trebuie să se recurgă la alte valori tipice pentru caracterizarea tendinței centrale a repartiției.

3. Cînd datele sînt grupate pe intervale și pierd deci individualitatea lor, me. se calculează prin procedeul interpolării liniare, bazat pe ipoteza repartiției uniforme a frecvențelor în intervalul median (clasa mediană).

Exemplul 4. Repartiția personalului unei întreprinderi pe grupe de retribuții lunare nete realizate.*)

Tabelul 19

Retribuția lunară (lei)	Numărul personalului	Numărul cumulat al personalului	
		ascendent	descendent
1600—1800	9	9	800
1800—2000	13	22	791
2000—2200	27	49	778
2200—2400	64	113	751
2400—2600	148	261	687
2600—2800	261	522	539
2800—3000	123	645	278
3000—3200	89	734	155
3200—3400	34	768	66
3400—3600	21	789	32
3600—3800	11	800	11
Total	800		

*) Fără personalul de conducere.

Lectura frecvențelor cumulate arată că valorile de rangul $k = 400$ și $k + 1 = 401$ sînt cuprinse în intervalul „2600—2800 lei". Admițînd ipoteza echirepartiției, se împarte acest interval cu frecvența :

$$\frac{2800 - 2600}{261} = \frac{200}{261} \text{ lei.}$$

Pînă la valoarea de rangul $(400 + 401)/2 = 400{,}5$ revin intervalului median :

$$400{,}5 - 261 = 139{,}5 \text{ frecvențe.}$$

Partea din mărimea intervalului median adunată la limita sa inferioară dă valoarea mediană. Astfel :

$$Me = 2600 + \frac{200}{261}(400{,}5 - 261) = 2600 + 106{,}9 = 2706{,}9 \approx 2707 \text{ lei.}$$

Formula de calcul a me. este :

$$Me = x_k + d_k \frac{1/2(F_n + 1) - F_{k-1}}{f_k},$$

unde

x_k este limita inferioară a intervalului median;

d_k — mărimea intervalului median;

F_n — Σf;

F_{k-1} — frecvența cumulată a intervalului imediat inferior intervalului median;

f_k — frecvența simplă a intervalului median.

Curbele frecvențelor cumulate ascendente și descendente permit o determinare grafică rapidă a **me.**; ea este determinată prin intersecția acestor curbe (fig. 54)

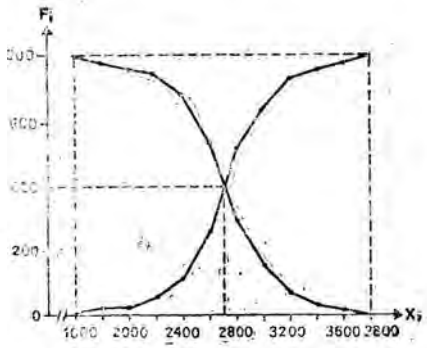
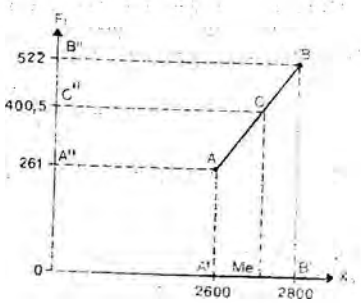


Fig. 54 Determinarea grafică a medianei

Pornind de la graficul stabilit se poate determina și modul de calcul al **me.** Reproducem partea din curba cumulativă ascendentă care corespunde intervalului median (fig. 55).

Fig. 55 Determinarea grafică a medianei (fragment)



Într-adevăr:

$$\overline{OM_e} = \overline{OA'} + \overline{A'Me}.$$

Dar, în virtutea teoremei lui Tales:

$$\overline{AMe}/\overline{A'B'} = \overline{AC}/\overline{AB} = \overline{A''C''}/\overline{A''E''},$$

de unde

$$\overline{A'Me} = \overline{A'B'} \times (\overline{A''C''}/\overline{A''B''})$$

și, prin urmare:

$$\overline{OMe} = \overline{OA'} + \overline{A'B'} \times (\overline{A''C''}/\overline{A''B''}).$$

Înlocuind simbolurile prin valorile lor, se obține:

$$\overline{OMe} = 2600 + (2800 - 2600) \times \frac{490.5 - 261}{522 - 261} = 2706,9$$

Proprietate. Suma abaterilor, în valoare absolută, a tuturor valorilor variabilei în raport cu o valoare arbitrară este minimă cînd aceasta se confundă cu **me.** Adică:

$$\sum_i |x_i - a| \text{ este minimum pentru } a = \mathbf{Me}$$

proprietate utilizată pentru a caracteriza o dispersie prin *medie**.

Medială (*Ml*). **Me.** aplicată la repartiția valorilor globale ale variabilei derate. **Ml.** separă în două părți egale suma produselor dintre valorile variabilei și frecvențele corespunzătoare:

$$\sum_{i=1}^{n} x_i f_i.$$

De exemplu, în repartiția referitoare la retribuțiile lunare (tabelul de mai jos) medială va fi aceea care împarte în două părți egale suma retribuțiilor primite de lucrători:

Retribuția lunară (lei) x_i	Numărul personalului f_i	$x_i f_i$
⋮	⋮	⋮
2400—2600	261	370000
2600—2800	522	704700
⋮	⋮	⋮
Total	—	2177600

x_i este mijlocul intervalului i

$$Ml = 2600 + 200 \frac{1/2(2177600 + 1) - 613900}{704700} = 273$$

V.TREBICI

Ml. se aplică la repartițiile în care ansamblul valorilor $x_i\ f_i$, constituit de valorile variabilei ponderate cu frecvențele, are o semnificație concretă. Confruntarea valorilor caracteristice **Me.** și **Ml.** pune în evidență o eventuală repartiție inegală a variabilei reprezentate.

CUANTILE (de ordinul n). Valori ale variabilei care separă repartiția ordonată în n părți cuprinzînd același efectiv egal cu $1/n$ din efectivul total.

Dacă $n = 4$ cuantilele se numesc *cuartile*;

$n = 10$ „ „ *decile*;

$n = 100$ „ „ *centile*.

C. corespunzătoare lui $n = 2$ este *mediana* *.

Cuartile. În număr de trei, ele împart domeniul valorilor observate (așezate în ordine de mărime) în patru părți de frecvențe egale cu 1/4. Ca mărime, prima *c.* (Q_1) este inferioară medianei, a doua *c.* (Q_2) se confundă cu mediana, iar a treia *c.* (Q_3) este superioară medianei. Analog medianei, *c. inferioară* se calculează după formula:

$$Q_1 = x_k + d_k \frac{1/4(F_n + 1) - F_{k-1}}{f_k} \tag{1}$$

și *c. superioară*

$$Q_3 = x_k + d_k \frac{3/4(F_n+1) - F_{k-1}}{f_k}, \tag{2}$$

unde, în fiecare caz, k este indicele clasei care conține *c.* considerată, x_k limita inferioară a acestei clase și d_k intervalul său. Se substituie valorii $1/2(F_n + 1)$ valoarea $1/4(F_n + 1)$ pentru Q_1 și $3/4(F_n + 1)$ pentru Q_3. Aplicate la repartiția din tab. 19:

$$Q_1 = 2400 + 200 \frac{1/4(800 + 1) - 113}{148} = 2517{,}9 \text{ lei}$$

și

$$Q_3 = 2800 + 200 \frac{3/4(800 + 1) - 522}{123} = 2928 \text{ lei}.$$

Ca și mediana, *c.* pot fi determinate, cu o oarecare aproximație, grafic (fig. 56 a și 56 b).

Rezultă că:

1) în *diagrama diferențială* * dreptele având drept abscise *c.* împart diagrama în 4 arii egale;

2) în *diagrama integrală* *, *c.* sînt abscisele punctelor ale căror ordonate sînt respectiv: 25% (*c.* Q_1), 50% (*c.* Q_2 sau mediana), 75% (*c.* Q_3).

Decile. În număr de nouă, *d.* împart cîmpul de variabilitate în zece intervale, conținînd fiecare 10% din observații. A cincea *d.* se confundă deci cu mediana.

V.TREBICI

În calculul *d.*, se substituie valorii $1/2(F_n+1)$ a frecvențelor cumulate, succesiv valorile $m/10\ (F_n + 1)$, pentru $m = 1, 2, 3, \ldots, 9$.
Cu aplicație la aceeași repartiție (tab. 19) :

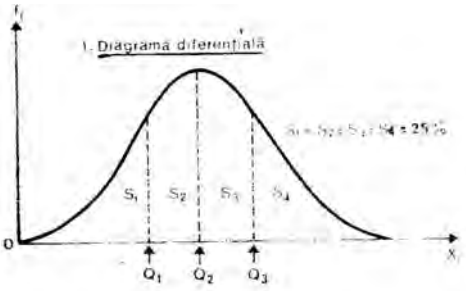


Fig. 56a Determinarea cuartilelor (diagrama diferențială)

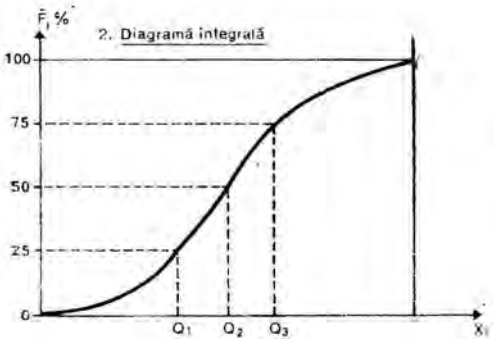


Fig. 56 b Determinarea cuartilelor (diagrama integrală)

$$D_1 = 2200 + 200\,\frac{1/10(800 + 1) - 49}{64} = 2297{,}19 \text{ lei}$$

$$D_2 = 2400 + 200\,\frac{2/10(800 + 1) - 113}{148} = 2463{,}78 \text{ lei}$$

. . .

$$D_5 = 2600 + 200\,\frac{5/10(800 + 1) - 261}{261} = 2706{,}90 \text{ lei}$$

. . .

$$D_9 = 3000 + 200\,\frac{9/10(800 + 1) - 645}{89} = 3170{,}56 \text{ lei.}$$

D. se notează de obicei cu litera D afectată de indici :

$$D_1,\ D_2,\ D_3, \ldots, D_5 = Me, \ldots, D_9.$$

Fig. 57 ilustrează principiul reprezentării grafice a d.

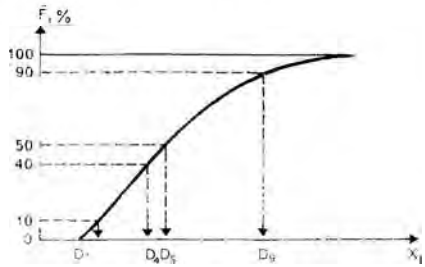


Fig. 57 Reprezentarea grafică a decilelor

Centile. În număr de 99, c. separă repartiția în o sută de intervale egale. Cea de-a cincizecea c. se identifică cu mediana, c. de rangurile 25 și 75 se identifică cu prima și a treia cuartilă, iar c. de rangurile 10, 20, 30 etc., cu decilele. C. se reprezintă de obicei prin litera C afectată de indici : $C_{01}, C_{02}, \ldots, C_{50}, \ldots, C_{99}$. Calculele se desfășoară la fel ca în cazul medianei, substituind valorii $1/2(F_n + 1)$ a frecvențelor cumulate, valorile $m/100(F_n + 1)$ pentru $m = 1, 2, 3, \ldots, 98, 99$ și considerînd clasele de valori corespunzătoare.
C repartiției din tab. 19 sînt :

$$C_{01} = 1600 + 200\ \frac{1/100(800 + 1) - 0}{9} = 1778{,}00 \text{ lei}$$

$$C_{02} = 1800 + 200\ \frac{2/100(800 + 1) - 9}{13} = 1908{,}00 \text{ lei}$$

. . .

$$C_{50} = 2600 + 200\ \frac{50/100(800 + 1) - 261}{261} = 2706{,}90 \text{ lei}$$

. . .

$$C_{99} = 3600 + 200\ \frac{99/100(800 + 1) - 789}{11} = 3672{,}55 \text{ lei.}$$

Calculul decilelor și c. se aplică la seriile care comportă suficient de numeroase elemente, în prealabil așezate în ordine de mărime.
MOD (Mo). Valoare cu cea mai mare frecvență în repartiția statistică (maximul *curbei de frecvențe*) *. De vreme ce se repetă de cele mai multe ori, valoarea modală reprezintă și valoarea cea mai „probabilă".
Termen sinonim : *dominantă* *).

*) Termenul englezesc ***mode***, introdus de K. Pearson, a fost tradus în franțuzește prin *dominante*.

În cazul unei repartiții discrete, **mo.** se identifică imediat, pe calea simplei examinări a șirului de frecvențe. Astfel, în tab. 18 numărul de 2 produse-rebut s-a înregistrat la 36 loturi; nici o altă mărime a rebuturilor nu revine atît de des: numărul „2 produse-rebut" este așadar **mo.** repartiției.

Cînd repartiția îmbracă forma grupării pe clase de valori, se va defini un *interval modal* (*clasă modală*) — interval căruia îi aparține **mo.** Mai precis, *intervalul modal* este intervalul cu *densitatea de frecvență* maximă. De exemplu, intervalul modal în repartiția din tab. 19 este intervalul „2600—2800 lei" cu frecvența 261. O estimare mai precisă a **mo.** necesită anumite ipoteze și calcule.

În condiția unei repartiții simetrice sau atunci cînd intervalele (clasele) adiacente celui modal s-ar deosebi puțin unul de altul după frecvență sau greutate specifică, se poate lua, cu o aproximație satisfăcătoare, mijlocul intervalului modal ca valoare modală. Dacă repartițiile sînt de alt tip, **mo.** se va găsi pe scara de variație a intervalului modal mai aproape sau mai departe de centrul acestui interval, în funcție de forma repartiției. Poziția **mo.** în intervalul modal se determină în raport cu abaterea frecvenței intervalului premodal și a celui postmodal de la frecvența intervalului modal. Procedeul de calcul este un procedeu de interpolare, constînd în aplicarea formulei următoare:

$$Mo = x_k + d_k \frac{f_k - f_{k-1}}{(f_k - f_{k-1}) + (f_k - f_{k+1})}, \qquad (1)$$

în care

x_k este limita inferioară a intervalului modal;
d_k — mărimea intervalului modal;
f_k — frecvența intervalului modal;
f_{k-1} — frecvența intervalului premodal;
f_{k+1} — frecvența intervalului postmodal.

Exemplu. Retribuția modală a repartiției din tab. 19.

	x_i	f_i	
	⋮	⋮	
	2400—2600	148	← f_{k-1}
x_k →	2600—2800	261	← f_k
	2800 ↑ —3000	123	← f_{k+1}
	⋮	⋮	
	d_k		

$$\mathbf{Mo} = 2600 + 200 \frac{261 - 148}{(261 - 148) + (261 - 123)} = 2600 + 90 = 2690 \text{ lei.}$$

V.TREBICI

Valoarea modală se estimează grafic cu ajutorul *histogramei* *.
În intervalul modal se duce o primă diagonală între limita inferioară, la nivelul indicat de intervalul precedent, și punctul opus al dreptunghiului care reprezintă intervalul modal. Se trasează o a doua diagonală, pornind de la limita superioară la nivelul indicat de frecvența intervalului următor. Proiectînd punctul P — intersecția celor două diagonale — pe axa orizontală se obține, prin citire directă, valoarea modală.

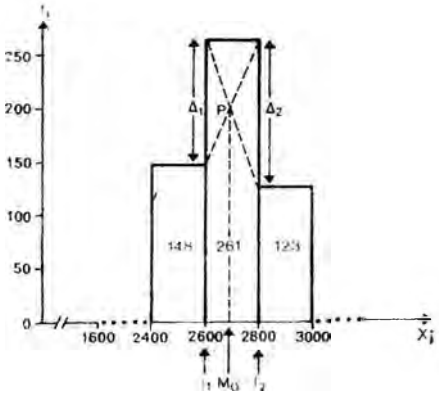


Fig. 58 Determinarea grafică a modului

Diagrama (fig. 58) ilustrează metoda de interpolare a valorilor pentru determinarea **mo.** Folosind notațiile pe grafic, scriem proporția:

$$\frac{Mo - l_1}{l_2 - Mo} = \frac{\Delta_1}{\Delta_2},$$

de unde

$$\Delta_2 Mo - \Delta_2 l_1 = \Delta_1 l_2 - \Delta_1 Mo,$$

$$\Delta_1 Mo + \Delta_2 Mo = \Delta_1 l_2 + \Delta_2 l_1,$$

$$Mo(\Delta_1 + \Delta_2) = \Delta_1 l_2 + \Delta_2 l_1.$$

Dar

$$l_2 = l_1 + d,$$

deci

$$Mo = \frac{\Delta_1 l_1 + \Delta_1 d + \Delta_2 l_1}{\Delta_1 + \Delta_2} = \frac{\Delta_1 l_1 + \Delta_2 l_1}{\Delta_1 + \Delta_2} + \frac{\Delta_1 d}{\Delta_1 + \Delta_2} =$$

$$= l_1 + d \frac{\Delta_1}{\Delta_1 + \Delta_2},$$

în care (compară cu 1)

$$l_1 \text{ este } x_k$$
$$\Delta_1 = f_k - f_{k-1}$$
$$\Delta_2 = f_k - f_{k+1}.$$

Un alt procedeu pentru aflarea **mo.**, propus de K. Pearson, se sprijină pe relația care există între **mo.**, mediană și medie în repartițiile moderat asimetrice. În orice repartiție simetrică unimodală (un singur maximum), media aritmetică, mediana și modul sînt identice ; într-o repartiție ușor asimetrică și unimodală, se constată că mediana se găsește între medie și **mo.**, distanța sa față de **mo.** fiind, aproximativ, dublul distanței sale de la media aritmetică (fig. 59).

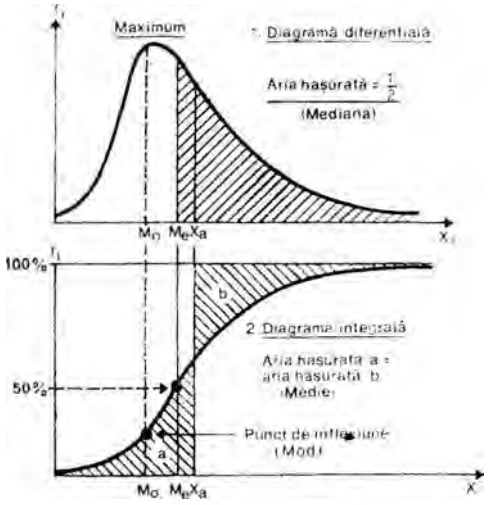


Fig. 59 Raportul dintre mod, mediană și media aritmetică

Prin urmare, cunoscîndu-se media aritmetică și mediana, **mo.** se deduce simplu din formula :

$$Mo = 3Mc - 2\,\bar{x}_a. \qquad (2)$$

Retribuția modală, calculată după formula (1), este de 2690 lei, iar după formula (2)

$$Mc = 2707 \text{ lei}$$

$$\bar{x}_a = 2722 \text{ lei}$$

$$Mo = (3 \times 2707) - (2 \times 2722) = 2677 \text{ lei}.$$

Diferența dintre valorile obținute este neimportantă.
În practica statistică se întîlnesc :
Repartiții unimodale. Repartiții cu o singură valoare sau clasă de valori modală.
Repartiții plurimodale (multimodale). Repartiții cu mai multe valori sau clase de valori modale.

S-a considerat, în exemplul tratat, o repartiție unimodală (repartiție cu o singură „cocoașă" sau repartiție „în clopot"). Prezența a două sau a mai multor moduri (un mod principal și alte moduri secundare) arată, de regulă, caracterul eterogen al repartiției: repartiția plurimodală corespunde unei contopiri de mai multe repartiții („amestec" de populații: de exemplu, într-o repartiție

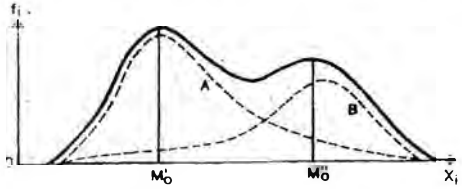


Fig. 60 Reprezentarea grafică a unei repartiții bimodale

a cererilor de pantofi după numărul pantofului, apar două moduri: numărul 42 — repartiția pentru bărbați și numărul 37 — repartiția pentru femei).
Fig. 60 reprezintă o repartiție bimodală compusă din două repartiții clasice (A și B).
Pentru repartițiile în formă de *U*, K. Pearson a introdus noțiunea de *antimod*, a cărei valoare corespunde frecvenței minime. Formula (1) se aplică și la calcularea antimodului.

MEDIE CRONOLOGICĂ ($\bar{x}_c$). Formă particulară a mediei aritmetice aplicată la determinarea nivelului mediu al unei *serii cronologice de momente.** În funcție de egalitatea sau neegalitatea intervalelor de timp ale seriei, se calculează **m.c.** sub formă simplă sau ponderată.

Medie cronologică simplă:

$$\bar{x}_c = \frac{\frac{x_1 + x_2}{2} + \frac{x_2 + x_3}{2} + \ldots + \frac{x_{n-1} + x_n}{2}}{n - 1}$$

care, prin transformare, devine

$$\bar{x}_c = \frac{1/2\, x_1 + x_2 + \ldots + x_{n-1} + 1/2 x_n}{n - 1}. \qquad (1)$$

Exemplul 1. Parcul de tractoare din agricultura românească, 1975—1980.

Tabelul 20

Anul	Număr de tractoare (mii bucăți, la sfîrșitul anului)
1975	119,5
1976	128,0
1977	138,5
1978	138,8
1979	139,8
1980	146,5

Sursa: *Anuarul statistic al R. S. România, 1980*, p. 235.

Numărul mediu de tractoare de care a dispus agricultura în cincinalul 1976—1980 a fost:

$$\bar{x}_c = \frac{(0{,}5 \times 119{,}5) + 128 + 138{,}5 + 138{,}8 + 139{,}8 + (0{,}5 \times 146{,}5)}{5} =$$

$$= 135{,}62 \text{ mii tractoare.}$$

Medie cronologică ponderată:

$$\bar{x}_c = \frac{\frac{x_1 + x_2}{2} t_1 + \frac{x_2 + x_3}{2} t_2 + \ldots + \frac{x_{n-1} + x_n}{2} t_n}{t_1 + t_2 + \ldots + t_n}, \qquad (2)$$

în care t este intervalul de timp.

Exemplul 2. Stocul de mărfuri inventariat într-o unitate comercială (la 1 ale lunii, în mii lei):

ianuarie	*martie*	*iunie*	*iulie*
694	742	810	770

Stocul mediu de mărfuri:

$$\bar{x}_c = \frac{\left(\frac{694 + 742}{2} \times 2\right) + \left(\frac{742 + 810}{2} \times 3\right) + \left(\frac{810 + 770}{2} \times 1\right)}{2 + 3 + 1} = 759 \text{ mii lei.}$$

Calculul *m.c.p.* dă un rezultat aproximativ. Pentru a fi cît mai aproape de realitate se cer următoarele condiții: a) perioadele de timp care separă termenii seriei să fie cît mai scurte și b) diferența între termenii seriei să fie cît mai mică.

CONDIȚIILE LUI YULE. Condiții precizate de statisticianul englez *Yule* *) care definesc un model ideal de medie*), și anume:

a) să fie definită în mod obiectiv, independent de aprecierea subiectivă a cercetătorului;

b) să fie expresia tuturor termenilor repartiției;

c) să posede proprietăți simple și evidente, făcînd posibilă înțelegerea sensului ei general;

d) să poată fi calculată cu ușurință și rapiditate;

e) să fie afectată cît mai puțin cu putință de fluctuațiile de selecție;

f) să se preteze ușor la calcule algebrice ulterioare.

Dintre parametrii cunoscuți, *media aritmetică* se apropie cel mai mult de modelul imaginat, ceea ce și explică utilizarea ei curentă. Ea este o funcție de toate valorile observate (ca de altfel orice formă de medie: armonică, geometrică etc.), admite o teoremă de adițiune, se bucură de proprietăți matematice simple, permite calcule algebrice, nu e decît puțin sensibilă la fluctuațiile de eșantionare și este ușor de manipulat.

---

*) Prin *medie* sînt desemnați toți parametrii de nivel: medie, mediană, mod. (→G. Udny Yule, M. G. Kendall, *Introducere în teoria statisticii*, Editura științifică, București, 1969, p. 124—125.)

Firește, dedusă din toți termenii repartiției, media aritmetică își pierde stabilitatea în prezența unor valori „aberante", anormal de mari sau anormal de mici. Acest „inconvenient" obligă la asigurarea premisei de omogenitate a datelor pentru care se calculează media. Nu trebuie să se uite că o medie reflectă cu atît mai precis proprietățile tipice ale populației cu cît mai omogenă este populația. Relativ la calculul mediilor („mijlociilor"), socotit „una din combinările cele mai obișnuite de statisticieni" — scria cu peste un secol în urmă I. Ionescu de la Brad — „... este foarte îmbelșugat în iluzii mai ales cînd extremele, cînd minimile și maximile sînt foarte depărtate ...*)

Mediana, prin definiție, nu ține seama de mărimea tuturor termenilor repartiției și ea este deci mai puțin afectată de mărimea termenilor extremi. O unitate statistică cu valoare excepțională ca mărime influențează totalul valorilor din care se calculează media aritmetică, aceeași unitate nu influențează mediana decît prin rangul ei.**) Calculul medianei este simplu, dar ea nu se pretează la operații algebrice. Mediana unei repartiții globale nu poate fi dedusă din medianele repartițiilor componente. Un avantaj practic al medianei, deloc neglijabil, constă în faptul că poate fi calculată chiar dacă nu se cunosc în mod precis limitele extreme ale primului și ultimului interval al repartiției considerate (repartiție deschisă). Semnificația sa devine însă incertă în cazul repartițiilor discrete cu „salturi" importante sau în cazul repartițiilor continue de volum redus.

Modul răspunde și mai puțin condițiilor lui Yule : nu se determină în funcție de toate observațiile, nu permite calcule algebrice, este foarte instabil.

Calculat pentru repartiții cu un număr mare de termeni și cu o tendință centrală netă, modul are totuși o valoare descriptivă importantă : exprimă mărimea cea mai comună a variabilei, mărimea a cărei cunoaștere este adesea deosebit de utilă. Se citează un exemplu, întîlnit în multe aplicații statistice : dacă o fabrică de confecții studiază, printre altele, taliile cumpărătorilor, ea nu caută să fabrice „pantaloni medii", care nu se vor potrivi nimănui, ci „pantaloni pentru clasa modală", adică pentru „marea majoritate" ; celelalte talii le va repartiza în raport cu dispersia în jurul valorilor centrale.***) Ca și în cazul medianei, intervalele marginale deschise ale unei repartiții nu împiedică calculul modului.

Exceptînd situațiile care motivează clar alegerea unui singur parametru, utilizarea împreună a celor trei parametri oferă o mai completă reprezentare a realității. Repartițiile asimetrice impun oricum calculul modului, în afară de cel al mediei aritmetice și al medianei.

MOMENTE**). Valori care sintetizează o repartiție, permițînd precizarea anumitor caracteristici ale repartiției.

Moment de ordinul k al variabilei X în raport cu o valoare arbitrară x_0.

Media aritmetică a diferențelor $x_i - x_0$, ridicate la puterea k :

$$x_0 m_k = \frac{\Sigma f_i (x_i - x_0)^k}{\Sigma f} . \qquad (1)$$

---

*) *I. Ionescu de la Brad, Povățuiri pentru catagrafia Moldovei precedate de oarecare elemente de statistică și publicată de Direcția Centrală de Statistică din Ministerul Interior al Moldovei*, Tip. „Buciumul Român", Iași, 1859, p. 7.

**) Mediana este numită și *medie de poziție*, deoarece se reține doar locul, și nu mărimea termenilor seriei.

***) A. Piatier, *Statistique et observation économique*, t. I., Presses Universitaires de France, Paris, 1961, p. 221 – 222.

****) Termenul **m.** a fost împrumutat din mecanică unde este folosit pentru a descrie distribuția de mase

V.TREBICI

În practică, se utilizează aproape exclusiv **m.** în raport cu originea ($x_0 = 0$) și **m.** în raport cu media ($x_0 = \bar{x}_a$).

Moment simplu (m_k) **M.** calculat în raport cu originea măsurătorilor ($x_0 = 0$);

$$m_k = \frac{\Sigma f_i x_i^k}{\Sigma f}. \tag{2}$$

Moment centrat (μ_k) **M.** calculat în raport cu media aritmetică a repartiției ($x_0 - \bar{x}_a$):

$$\mu_k = \frac{\Sigma f_i (x_i - \bar{x}_a)^2}{\Sigma f}. \tag{3}$$

Momente uzuale

a) *simple*

$$m_1 = \frac{\Sigma f_i x_i}{\Sigma f} = \bar{x}_a \qquad m_3 = \frac{\Sigma f_i x_i^3}{\Sigma f}$$

$$m_2 = \frac{\Sigma f_i x_i^2}{\Sigma f} \qquad m_4 = \frac{\Sigma f_i x_i^4}{\Sigma f}$$

b) *centrate*

$$\mu_1 = \frac{\Sigma f_i (x_i - \bar{x}_a)}{\Sigma f} = 0 \qquad \mu_3 = \frac{\Sigma f_i (x_i - \bar{x}_a)^3}{\Sigma f}$$

$$\mu_2 = \frac{\Sigma f_i (x_i - \bar{x}_a)^2}{\Sigma f} = \sigma^2 \qquad \mu_4 = \frac{\Sigma f_i (x_i - \bar{x}_a)^4}{\Sigma f}.$$

În particular, **m.** simplu de ordinul 1 se confundă cu media aritmetică, **m.** centrat de ordinul 1 este nul [proprietatea mediei $\Sigma f_i (x_i - \bar{x}_a) = 0$] și **m.** centrat de ordinul 2 se confundă cu *varianța.**

Primele patru **m.** ale unei repartiții de frecvențe sînt, în cea mai mare parte a cazurilor, suficiente pentru a descrie caracteristicile principale. Din aceste momente sînt derivați *coeficienții de asimetrie* * și *coeficienții de aplatizare* *. *Relații între* **m.** *simple și cele centrate uzuale*:

$$\mu_2 = m_2 - m_1^2 \tag{4}$$

$$\mu_3 = m_3 - 3m_1 m_2 + 2m_1^3 \tag{5}$$

$$\mu_4 = m_4 - 4m_1 m_3 + 6m_1^2 m_2 - 3m_1^4. \tag{6}$$

Corecția lui Sheppard. Cantitate care trebuie scăzută din **m.** *calculat* al unei repartiții pe clase de valori, pentru a corecta eroarea sistematică introdusă prin substituirea centrelor de clasă valorilor adevărate ale variabilei observate.

V.TREBICI

*Sheppard**) a propus următoarele formule de calcul a **m.** *corectate* μ_2' și μ_4' din **m.** *brute* (adică necorectate):

$$\mu_2' = \mu_2 - \frac{1}{12} d^2 \tag{7}$$

$$\mu_4' = \mu_4 = \frac{1}{2} \mu_2 d^2 + \frac{7}{240} d^4; \tag{8}$$

în care d reprezintă amplitudinea, presupusă comună, a claselor. Formulele (7) și (8) pot fi aplicate în cazul unei repartiții unimodale („în clopot") cu interval de clasă constant, frecvența tinzînd către zero în ambele direcții („aripile" curbei de frecvențe tind să se apropie progresiv de axa x-ilor).
Se sugerează ca „de regulă, să nu se facă corecții decît atunci cînd frecvența este cel puțin 1000 sau cînd gruparea este mai largă decît cea în care intervalul de grupare constituie a douăzecea parte din intervalul de variație al variabilei**).
Corecția lui Sheppard nu este aplicabilă:

1) repartițiilor pe valori distincte, căci atunci dispare motivul care a condus la corecție;
2) repartițiilor în formă de „J" sau „U" sau chiar formelor moderat asimetrice (oblice);
3) momentelor centrate de ordin impar, deoarece alternările de semne provoacă compensarea erorilor; în particular, în repartițiile perfect simetrice, **m.** impare sînt nule.

M. centrate de ordin par sînt, în general, supraestimate cînd se calculează din repartiții pe clase de valori, de unde și necesitatea corecției în sensul arătat.
Cumulant de ordinul n. Numărul $M_{[n]}$ determinat de egalitatea

$$i^n M_{[n]} = \left[\frac{d^{n-1}}{dt^{n-1}} \left(\frac{\varphi'(t)}{\varphi(t)} \right) \right]_{t=0} = \left[\frac{d^n}{dt^n} (\ln \varphi(t) \right]_{t=0} \tag{10}$$

în ipoteza că există derivatele pînă la ordinul n inclusiv ale *funcției caracteristice****) $\varphi(t)$, deci că există momentele pînă la acest ordin inclusiv; i este simbolul unității imaginare. Primii patru cumulanți pot fi exprimați ca funcții simple de **m.** de primele patru ordine:

$$M_1 = m_1$$

$$M_2 = m_2 - m_1^2$$

$$M_3 = m_3 - 3m_1m_2 + 2m_1^3$$

$$M_4 = m_4 - 4m_1m_3 - 3m_2^2 + 12m_1^2m_2 - 6m_1^4.$$

---

*) *W. F. Sheppard, On the Calculation of the Most Probable Values of Frequency Constants for Data Arranged According to Equidistant Divisions of a Scale*, Proc. Lon. Math. Soc. vol. XXIX, p. 353.
**) G. Udny Yule, M. G. Kendall, *op. cit.*, p. 153.
***) Cumulanții au fost introduși în statistică de *T. N. Thiele**, sub numele de „semiinvarianți". Cuvîntul „cumulant" este preferabil și de uz general, întrucît există și alte tipuri de mărimi care posedă proprietatea de semiinvarianță cu sens algebric (G. Udny Yule, M. G. Kendall, *op. cit.*, p. 182).

în particular, în raport cu media aritmetică,

$$M_1 = 0$$

$$M_2 = \mu_2$$

$$M_3 = \mu_3$$

$$M_4 = \mu_4 - 3\,\mu_2^2.$$

Cumulanții, numiți uneori *semiinvarianți*, au două proprietăți remarcabile :
a) cu excepția primului, toți cumulanții sînt independenți de originea de calcul ;
b) dacă scala de măsurare a variabilei este modificată prin înmulțirea tuturor valorilor cu o constantă a, al n-lea cumulant este înmulțit cu a^n.

VIII. DISPERSIA (ÎMPRĂȘTIEREA)

PARAMETRII DISPERSIEI. Parametri care servesc la măsurarea împrăștierii valorilor unei variabile în jurul valorii centrale. Media, mediana, modul ne informează, fiecare în felul său, despre ordinul de mărime mijlociu sau valoarea centrală a unei repartiții statistice. Orice repartiție are însă o organizare interioară definită de modul în care valorile variabilei gravitează în jurul valorii centrale și care poate diferi de la o repartiție la alta. (→ *Parametru statistic.*) Ca atare, *parametrii de nivel* trebuie neapărat completați cu parametrii de dispersie *). Termen sinonim : *varianță* (→XVII. *Analiza de varianță ANOVA.*) Ca și în cazul tendinței sau valorii centrale, statisticienii au imaginat mai mulți parametri pentru măsurarea dispersiei, întemeiați, în general, pe noțiunea de *abatere*. Aceștia caracterizează fie dispersia brută, fie dispersia relativă.
Dispersie brută. Dispersie exprimată în unități de măsură ale datelor originale.
Dispersie relativă. Dispersie exprimată în numere pure, independente de unitățile de măsură ale datelor originale și mai bine adaptate la compararea dispersiilor unor populații de natură diferite.
AMPLITUDINE *(interval de variație, ω).* Diferența între valoarea cea mai mare, x_{max}, și valoarea cea mai mică, x_{min}, a unei variabile

$$\omega = x_{max} - x_{min}.$$

În cazul repartiției pe intervale se consideră diferența între centrele de clasă extreme. Repartiția din tab. 19 are **a.** :

$$\omega = 3700 - 1700 = 2000 \text{ lei.}$$

A. are un sens clar și se determină extrem de repede. În anumite cazuri, ca atunci cînd singurele valori extreme ale variabilei condiționează o decizie,

*) Parametrii de nivel corespund punctelor celor mai remarcabile ale curbei de frecvențe, iar parametrii de dispersie — etalării curbei. Alura intrinseca a curbei de frecvență și poziția sa în raport cu axele de referință se reduc la un mic număr de mărimi, analog definirii unui cerc prin mărimile care fixează poziția centrului și raza.

ea constituie măsura dispersiei cea mai utilă (de exemplu, în controlul statistic al calității unor produse). Alteori, se utilizează doar ca parametru indicativ, deoarece prezintă inconveniente serioase : nu depinde decît de termenii extremi, care sînt adesea excepționali ; este supusă fluctuațiilor mari de la eșantion la eșantion ; ignoră forma repartiției între extreme (o aceeași valoare a **a.** poate să rezulte dintr-o repartiție simetrică sau dintr-una total asimetrică, deși se știe că în al doilea caz dispersia este mai mare decît în primul caz). De altfel, în general, **a.** apare definită cu aproximație pentru repartițiile pe scară de intervale, valorile extreme nemaifiind cunoscute cu exactitate în urma grupării în clase de valori.

ABATERE CUARTILĂ (*interval intercuartil, intercuartilă*). Abatere între prima și a treia cuartilă (→ *Cuartile*) :

$$Q_3 - Q_1. \tag{1}$$

Se obține o indicație despre dispersie comparîndu-se *amplitudinea* totală cu **a.c.**

Dacă $Q_3 - Q_1 < \frac{\omega}{2}$ repartiția este intens concentrată.

$Q_3 - Q_1 > \frac{\omega}{2}$ repartiția este intens dispersată.

Uneori, se utilizează *semiintercuartila* :

$$\frac{1}{2}(Q_3 - Q_1). \tag{2}$$

De asemenea, pentru anumite comparații (de exemplu, repartiții exprimate în unități diferite), se recurge la **a.c.** *relativă* (*intercuartila relativă*), definită de relația :

$$\frac{Q_3 - Q_1}{Q_2}, \tag{3}$$

adică de raportul dintre **a.c.** în valoare absolută și mărimea *mediană* ($Q_2 = Me$). Cu aplicație la repartiția din tab. 19 :

$$Q_3 - Q_1 = 2928 - 2518 = 410 \text{ lei}$$

$$Q_3 - Q_1 < \frac{\omega}{2} = 410 < \frac{2000}{2} \text{ lei}$$

$$\frac{1}{2}(Q_3 - Q_1) = \frac{1}{2}\,410 = 205 \text{ lei}$$

$$\frac{Q_3 - Q_1}{Q_2} = \frac{410}{2707} = 0{,}1515 \text{ sau } 15{,}2\%.$$

Evident, cu cît **a.c.**, în valoare absolută sau în valoare relativă, este mai mare, cu atît valorile variabilei sînt mai dispersate și invers.

Cu toate că **a.c.** prezintă avantajul posibilității de aplicare deopotrivă la repartiții închise ca și la repartiții deschise și, în al doilea rînd, nu este prea mult influențată de valorile extreme, ea rămîne totuși un parametru insuficient pentru măsurarea dispersiei, întrucît nu este expresia întregii repartiții. Un bun parametru de dispersie trebuie să integreze într-o valoare de sinteză abaterile tuturor valorilor observate, între ele sau față de medie.

Observație. Intervalul intercuantil conține 50% din observații. Se pot defini de asemenea :

1) *interval interdecil*

$D_9 - D_1$ (conține 80% din observații) ;

2) *interval intercentil*

$C_{99} - C_{01}$ (conține 98% din observații).

Kelley a găsit că riscul de eroare este minim dacă se utilizează parametrul :

$$C_{93} - C_{07}.$$

Intervalele intercentile nu pot fi calculate decît pe un număr suficient de mare de observații.

ABATERE MEDIE (*abatere absolută medie*). Medie aritmetică a abaterilor „absolute" ale valorilor observate ale variabilei față de media lor aritmetică. Expresia teoretică **a.m.** ($e_{\bar{x}_a}$) este :

$$e_{\bar{x}_a} = \frac{\Sigma | x - \bar{x}_a |}{n}, \text{ pentru o serie de valori individuale} \qquad (1)$$

și

$$e_{\bar{x}_a} = \frac{\Sigma | x - \bar{x}_a | f}{\Sigma f}, \text{ pentru o repartiție de frecvențe.} \qquad (2)$$

A.m. poate fi calculată și în raport cu *mediana*, înlocuindu-se $\bar{x}_a$ prin *Me* în formulele (1) și (2).

Exemplul 1. Producția de energie electrică a României în cincinalul 1976—1980.

Tabelul 21

Anul	Energie electrică (mld. kWh)	$\lvert x - 62{,}98 \rvert$	$\lvert x - 64{,}3 \rvert$
(1)	(2)	(3)	(4)
1976	58,3	4,68	6,0
1977	59,9	3,08	4,4
1978	64,3	1,32	0
1979	64,9	1,92	0,6
1980	67,5	4,52	3,2
Total	314,9	15,52	14,2

Sursa : *Anuarul statistic al R. S. România 1980*, p. 179.

$$\bar{x}_a = \frac{314{,}9}{5} = 62{,}98 \text{ mld. kWh}$$

Me = 64,3 mld. kWh

$$e_{\bar{x}_a} = \frac{15,52}{5} = 3,10 \text{ mld. kWh}$$

$$e_{Me} = \frac{14,2}{5} = 2,84 \text{ mld. kWh.}$$

Exemplul 2. Calculul a.m. pentru repartiția de frecvențe din tab. 19.

Tabelul 22

x	f	$\lvert x - 2722\rvert$	$\lvert x - 2722\rvert f$	$\lvert x - 2707\rvert$	$\lvert x - 2707\rvert f$
(1)	(2)	(3)	(4)	(5)	(6)
1700	9	1022	9198	1007	9063
1900	13	822	10686	807	10491
2100	27	622	16794	607	16389
2300	64	422	27008	407	26048
2500	148	222	32856	207	30636
2700	261	22	5742	7	1827
2900	123	178	21894	193	23739
3100	89	378	33642	393	34977
3300	34	578	19652	593	20162
3500	21	778	16338	793	16653
3700	11	978	10758	993	10923
Total	800	—	204568	—	200908

$$\bar{x}_a = 2722 \text{ lei}$$

$$Mc = 2706,\ 9 \approx 2707 \text{ lei.}$$

$$e_{\bar{x}_a} = \frac{204568}{800} = 255,71 \text{ lei}$$

$$e_{Me} = \frac{200908}{800} = 251,14 \text{ lei.}$$

În calculul practic al parametrului $e_{\bar{x}_a}$ ne putem mărgini la valorile lui x superioare lui $\bar{x}_a$, deoarece suma abaterilor pozitive este egală cu suma valorilor absolute ale abaterilor negative.
Vom avea atunci

$$ne_{\bar{x}_a} = 2 \sum_{x > \bar{x}_a} (x - \bar{x}_a), \text{ pentru cazul (1)}$$

$$5e_{\bar{x}_a} = 2(1,68 + 3,08) = 15,52$$

și,

$$(\Sigma f)\, e_{\bar{x}_a} = 2 \sum_{x > \bar{x}_a} (x - \bar{x}_a) f, \text{ pentru cazul (2)}$$

$$800\ e_{\bar{x}_a} = 2(9198 + 10686 + 16794 + 27008 + 32856 + 5742) = 201568.$$

Tendința centrală a unei repartiții se exprimă în mod curent prin media aritmetică și, ca atare, este logic să se definească dispersia pe baza abaterilor valorilor observate în raport cu media lor aritmetică. Cum însă, prin construcție, media abaterilor (în valoare algebrică) de la media aritmetică este nulă, se ia media abaterilor în valoare absolută. Procedeul de calcul se justifică și prin faptul că, în noțiunea de dispersie sau variabilitate, prezintă interes mărimea abaterilor și nu sensul lor (pozitiv sau negativ).

A.m. în raport cu mediana are valoarea cea mai mică (→ *Mediană*).

A.m. fie în raport cu media aritmetică, fie în raport cu mediana se folosește mai ales cînd numărul de observații este redus.

ABATERE MEDIANĂ (*abatere probabilă, abatere echiprobabilă*). Mediana valorilor absolute ale abaterilor calculate fie în raport cu media aritmetică, fie în raport cu mediana seriei de valori observate.

Pentru seria temporală din ex. 1 :

— mediana valorilor $| x_i - \bar{x}_a | = 3{,}08$ mld. kwh,

— „ „ $| x_i - Me | = 3{,}2$ „ „ ,

iar pentru repartiția de frecvențe din ex. 2

— mediana valorilor $| x_i - \bar{x}_a | = 222$ lei,

— „ „ $| x_i - Me | = 207$ „ .

Se numește și abatere probabilă sau echiprobabilă, fiindcă are o șansă din două de a fi depășită. Sînt cuprinse deci 50% din observații în intervalul celor două abateri probabile, centrat pe mediană.

Cînd repartiția este simetrică, abaterea mediană se confundă cu *semiintercuartila*. În consecință, într-o repartiție ușor asimetrică, semiintercuartila poate fi o estimație acceptabilă a abaterii mediane. Această abatere este puțin utilizată.

ABATERE PĂTRATICĂ MEDIE (*abatere standard, abatere tip*, σ). Rădăcină pătrată din media aritmetică a pătratelor abaterilor valorilor observate în raport cu media lor aritmetică :

$$\sigma = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n}}, \text{ pentru o serie de valori individuale} \qquad (1)$$

și

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2 f}{\Sigma f}}, \text{ pentru o repartiție de frecvențe.} \qquad (2)$$

A.p.m. se calculează pornind de la pătratele abaterilor, valori într-un fel mai semnificative decît abaterile simple în valoare absolută, căci accentuează efectul abaterilor mari, responsabile de dispersii mari.

Exemplul 1. Calculul **a.p.m.** pentru seria din tab. 21

Tabelul 23

	x	$(x - 62{,}98)^2$	x^2
	(1)	(2)	(3)
1	58,3	21,9024	3398,89
2	59,9	9,4864	3588,01
3	64,3	1,7424	4134,49
4	64,9	3,6864	4212,01
5	67,5	20,4304	4556,25
Σ	314,9	57,2480	19889,65

$\bar{x}_a = 62{,}98$ mld. kwh

$(x - 62{,}98) \rightarrow$ tab. 21, col. 3.

$$\sigma = \sqrt{\frac{57{,}2480}{5}} = 3{,}38 \text{ mld. kwh.}$$

Prin dezvoltarea formulei de definiție (1) se ajunge la o formulă convenabilă în calculul practic :

$$\sigma = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}. \qquad (3)$$

Cu datele aplicației :

$$\sigma = \sqrt{\frac{19\,889{,}65}{5} - \left(\frac{314{,}9}{5}\right)^2} = 3{,}38 \text{ mld. kwh.}$$

Formula (3) evită operația $(x - \bar{x})$, care poate provoca erori cînd media $(\bar{x}_a)$ este complicată și comportă îndeosebi mai multe zecimale. **A.p.m.** (σ) se exprimă în unitatea de măsură a variabilei (X) pentru care s-a calculat. Evident, cu cît valoarea sa este mai mare, cu atît mai intensă este variația variabilei cercetate și invers.

V.TREBICI

Exemplul 2. Calculul **a.p.m.** în cazul datelor grupate (cf. tab. 19):

Tabelul 24

x	f	$(x - 27{,}22)^2 f$	xf	x^2f
(1)	(2)	(3)	(4)	(5)
17	9	940.0356	153	2601
19	13	878,3892	247	4693
21	27	1044,5868	567	11907
23	64	1139,7376	1472	33856
25	148	729.4032	3700	92500
27	261	12,6324	7047	190269
29	123	389,7132	3567	103443
31	89	1271,6676	2759	85529
33	34	1135,8856	1122	37026
35	21	1271,0964	735	25725
37	11	1052,1324	407	15059
Σ	800	9865,2800	21776	602608

Notă. Valorile x împărțite la 100.

$$\bar{x}_a = 2722 \text{ lei.}$$

Conform formulei (2)

$$\sigma = \sqrt{\frac{98652800}{800}} = 351{,}16 \text{ lei,}$$

sau aplicînd formula

$$\sigma = \sqrt{\frac{\Sigma x^2 f}{\Sigma f} - \left(\frac{\Sigma xf}{\Sigma f}\right)^2}, \tag{4}$$

dedusă din formula de definiție (2),

$$\sigma = \sqrt{\frac{6026080000}{800} - \left(\frac{2177600}{800}\right)^2} = 351{,}16 \text{ lei.}$$

Calculul **a.p.m.**, pentru o repartiție cu intervale egale, se ușurează mult dacă se aplică formula:

$$\sigma = \sqrt{\frac{\Sigma\left(\frac{x - \bar{x}}{d}\right)^2 f}{\Sigma f} - \left[\frac{\Sigma \frac{x - \bar{x}}{d} f}{\Sigma f}\right]^2} \cdot d. \tag{5}$$

Se efectuează o schimbare de variabilă, procedeu al cărui principiu a fost expus în legătură cu calculul mediei aritmetice. (→*Medie aritmetică.*)

V.TREBICI

Mecanismul aplicării formulei (5) este arătat în tab. 25.

Tabelul 25

x	f	$\frac{x-2700}{200}$	$\left(\frac{x-2700}{200}\right)f$	$\left(\frac{x-2700}{200}\right)^2 f$
(1)	(2)	(3)	(4)	(5)
1700	9	−5	− 45	225
1900	13	−4	− 52	208
2100	27	−3	− 81	243
2300	64	−2	−128	256
2500	148	−1	−148	148
2700	261	0	0	0
2900	123	+1	+123	123
3100	89	+2	+178	356
3300	34	+3	+102	306
3500	21	+4	+ 84	336
3700	11	+5	+ 55	275
Σ	800	—	+ 88	2476

$$\sigma = \sqrt{\frac{2476}{800} - \left(\frac{88}{800}\right)^2} \cdot 200 = 351{,}16 \text{ lei.}$$

Calculele sînt simplificate ca urmare a reducerii apreciabile a valorilor x ca mărime. Procedeul poate servi dintr-odată la calculul mediei și a **a.p.m.**

De reținut faptul că procedeul de calcul rapid aplicat nu este un procedeu de aproximare. Formula (5) nu reprezintă decît o transformare algebrică a formulei de definiție (2), cînd toate unitățile dintr-o aceeași clasă primesc valoarea centrului clasei.

La baza metodelor practice de calcul al **a.p.m.** stă formula lui König, scrisă pentru *varianță* * :

$$V' = V + d^2, \tag{6}$$

unde

$$V = \frac{\Sigma(x - \bar{x}_a)^2}{n} \text{ (pentru o serie de valori individuale).}$$

adică pătratul **a.p.m.** (σ^2).

Abaterile $(x - \bar{x}_a)$ se exprimă frecvent în numere zecimale, cu mai multe zecimale, ceea ce face anevoios calculul pătratelor lor $\Sigma(x - \bar{x}_a)^2$. Așa fiind, se alege o altă valoare centrală C (adesea $C = 0$), astfel ca abaterile $x - C$ să fie numere simple, pe cît posibil întregi.

Se obține atunci ușor o varianță V', superioară lui V și legată cu V prin relația lui König :

$$V' = V + (C - \bar{x}_a)^2 \text{ sau } V' = V + d^2,$$

unde $d = |C - \bar{x}_a|$.

De aici rezultă :

$$V = V' - d^2, \tag{7}$$

unde

$$V' = \frac{\Sigma(x - C)^2}{n}.$$

Din varianță se deduce **a.p.m.** : $V = \sigma^2$ și $\sigma = \sqrt{V}$.

Varianța (*Dispersie* sau *fluctuație*). Este egală cu media aritmetică a pătratelor abaterilor. Definită astfel, ea se exprimă în unități bizare : acelea ale valorilor ridicate la pătrat. Ca să se rămînă la unități omogene, se reține, drept valoare tipică a dispersiei, **rădăcina pătrată a varianței** (σ).

Cînd se utilizează *centrele de clase* * pentru calculul varianței, ca în exemplul tratat, se admite că valorile în interiorul unei clase date sînt repartizate în mod uniform. Totuși, într-o *repartiție unimodală* * cu o *variabilă continuă* *, media reală a valorilor în interiorul fiecărei clase se va abate de la centrul clasei în direcția *modului*. Ca urmare, trebuie corectată *eroarea sistematică* în calculul varianței, scăzîndu-se 1/12 din pătratul *intervalului de clasă* (→ *Corecția lui Sheppard*). Astfel :

$$V_{cor} = \sigma^2_{cor} = \sigma^2 - \frac{1}{12} d^2 = \frac{98652800}{800} - \frac{1}{12} 200^2 = 123316 -$$

$$- \frac{1}{12} 40000 = 119983$$

și **a.p.m.** corectată :

$$\sigma_{cor} = \sqrt{119983} = 346{,}40 \text{ lei (în loc de } 351{,}16 \text{ lei).}$$

Noțiunile de *varianță* și *abatere pătratică medie* sînt fundamentale : se regăsesc explicit sau implicit formulate în aproape toate dezvoltările metodei statistice.

Coeficient de variație (v). Raport între **a.p.m.** și media aritmetică :

$$v = \frac{\sigma}{\bar{x}_a}. \tag{8}$$

În exemplele precedente :

(1) $\bar{x}_a = 62{,}98$ mld. kwh
$\sigma = 3{,}38$ " "

$$v = \frac{3{,}38}{62{,}98} = 0{,}0536 \text{ sau } 5{,}4\%.$$

(2) $\bar{x}_a = 2722$ lei
$\sigma = 351{,}16$ „

$$v = \frac{351{,}16}{2\,722} = 0{,}1281 \text{ sau } 12{,}8\%.$$

V.TREBICI

C.v. are aceeași semnificație ca și **a.p.m.** : cu cît valoarea lui este mai mare cu atît mai eterogenă este populația pe care o caracterizează media, cu atît mai intensă este variația variabilei și invers. Acest coeficient, exprimat de regulă în procente, constituie măsura cea mai răspîndită a *dispersiei relative* *.
Cu ajutorul *c.v.* se pot ierarhiza dispersiile valorilor variabilelor, măsurate cu unități diferite, în două sau mai multe populații, dar valoarea sa rămîne dependentă de nivelul mediei.
Notă
Alte măsuri ale dispersiei relative (rar folosite) :
— **a.p.m.** împărțită la mediană

$$\frac{\sigma}{Mc}; \tag{9}$$

— abaterea absolută medie împărțită la media aritmetică

$$\frac{\sigma_{\bar{x}_a}}{\bar{x}}; \tag{10}$$

— abaterea cuartilă împărțită la mediană

$$\frac{Q_3 - Q_1}{Mc}. \tag{11}$$

DIFERENȚĂ MEDIE A LUI GINI (*Dm*). Media aritmetică a diferențelor în valoare absolută, formate prin asocierea valorilor variabilei două cîte două în toate felurile posibile, inclusiv cu ele însele.
1. Fie seria simplă :
103, 115, 119, 128, 137, 138, 143.
Comodă este construirea unui tabel triunghiular : pe rîndul 1 se scriu valorile seriei, iar pe rîndurile următoare, diferențele între fiecare valoare și celelalte valori, începînd cu prima valoare.

Valorile seriei	103	115	119	128	137	138	143	Suma diferențelor
	0	12	16	25	34	35	40	162
		0	4	13	22	23	28	90
			0	9	18	19	24	70
				0	9	10	15	34
					0	1	6	7
						0	5	5
							0	0
								368

Dacă se socotesc diferențele într-un singur sens, ca în tabelul adoptat, atunci **d.m.** are valoarea :

$$Dm = \frac{368}{21} = 17{,}52.$$

Se observă că numărul total al diferențelor este egal cu $1 + 2 + 3 + \ldots + (n-1)$ pentru o serie de n elemente, adică expresia sumei unei progresii aritmetice de n termeni, cu primul termen 1 și cu rația 1.
Suma va fi :

$$\Sigma(n-1) = \frac{n(n-1)}{2} = \frac{7(7-1)}{2} = 21.$$

Definiția **d.m.** pretinde să se socotească diferențele termenilor seriei în ambele sensuri (de exemplu, 103—115, dar și 115—103) și de asemenea diferențele nule (103—103 ; 115—115 ; etc.). Astfel *) :

$$Dm = \frac{2 \times 368}{(7)^2} = \frac{736}{49} = 15{,}02.$$

C. Gini a conceput o formulă echivalentă :

$$Dm = 4\,\frac{\Sigma|x - Me||y_x - y_{Me}|}{n^2}, \qquad (1)$$

în care

x reprezintă termenii seriei ;
Me — mediana ;
y_x — rangul fiecărui termen al seriei ;
y_{Me} — rangul median ;
n — numărul termenilor seriei.

Cu datele seriei luate ca exemplu, se formează tabelul de calcul :

Tabelul 26

x	y_x	$\lvert x - 128\rvert$	$\lvert y_x - 4\rvert$	$\lvert x - 128\rvert \lvert y_x - 4\rvert$
103	1	25	3	75
115	2	13	2	26
119	3	9	1	9
128	4	0	0	0
137	5	9	1	9
138	6	10	2	20
143	7	15	3	45
Total	—	—	—	184

$$Dm = 4\,\frac{184}{(7)^2} = \frac{736}{49} = 15{,}02.$$

*) În loc de un tabel triunghiular se obține un tabel pătrat cu 7 coloane și 7 rînduri (49 diferențe) și suma diferențelor evident dublă (2×368).

2. În cazul unei repartiții de frecvențe cu intervale egale, formula lui Gini devine :

$$Dm = 2d \frac{\Sigma F_i(F_i - \Sigma f_i)}{(\Sigma f_i)^2}, \qquad (2)$$

în care :

d este mărimea intervalului ;
f — frecvențe simple ;
F — frecvențe cumulate.

De exemplu, **d.m.** a repartiției din tab. 19

Tabelul 27

f_i	F_i	$F_i - \Sigma f_i$	$F_i(F_i - \Sigma f_i)$
9	9	791	7119
13	22	778	17116
27	49	751	36799
64	113	687	77631
148	261	539	140679
261	522	278	145116
123	645	155	59975
89	734	66	48444
34	768	32	24576
21	789	11	8679
11	800	0	0
Σ 800	—	—	606134

$$Dm = 2 \times 200 \frac{606134}{(800)^2} = 378{,}83 \text{ lei.}$$

D.m. a lui Gini are proprietatea de a depinde de abaterile valorilor variabilei însăși, și nu de ,,răspîndirea valorilor în jurul unui punct arbitrar"*), cum ar fi o valoare centrală. Într-un anumit sens, ea măsoară ,,răspîndirea intrinsecă a populației independent de originea de calcul sau de nivelul repartiției". Această proprietate o recomandă ca parametru de dispersie pentru repartiții statistice care nu au o valoare centrală semnificativă (de exemplu, în cazul repartițiilor hiperbolice).

PARAMETRII FORMEI (REPARTIȚIEI). Parametrii care definesc *asimetria* și *aplatizarea* repartiției de frecvențe.

I. O repartiție se numește *simetrică* sau *asimetrică* după cum valorile variabilei sînt egal sau neegal dispersate de o parte și de alta a valorii centrale (fig. 61). Două sau mai multe repartiții asimetrice se pot deosebi atît prin *gradul* de asimetrie, cît și prin *elul* asimetriei :

Curba *B* este mai puțin asimetrică decît curba *C*. Curba *B* este *pozitiv asimetrică*, iar curba *C*, *negativ asimetrică*, dacă se convine să se considere asimetria

*) G. Udny Yule, M. G. Kendall, *op. cit.*, p. 165.

V.TREBICI

pozitivă sau negativă în sensul etalării repartiției spre partea valorilor mici sau spre partea valorilor mari.
Cînd repartiția este perfect simetrică :

$$\bar{x}_a = Me = Mo.$$

Q_1 și Q_3 — echidistante de mediană.
D_1 și D_9 — echidistante de mediană.
Poziția respectivă a acestor elemente permite calculul *coeficienților de asimetrie* (*disimetrie, oblicitate*), care sînt nuli în caz de simetrie și se îndepărtează

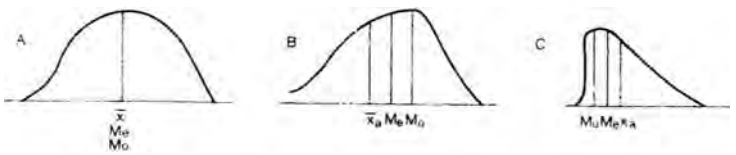


Fig. 61 Tipuri de repartiții (A-Simetrică; B-Asimetrică (oblică) la dreapta C-Asimetrică (oblică) la stînga

cu atît mai mult de valoarea 0 cu cît repartiția este mai asimetrică. Coeficienții concepuți sînt numere abstracte.
Coeficientul lui Yule. Raportul între diferența abaterilor cuartile și suma lor :

$$\frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)}. \qquad (1)$$

Coeficientul variază între -1 și $+1$. Dacă $Q_3 - Q_2 > Q_2 - Q_1$, valoarea sa este pozitivă, iar în caz contrar, negativă. Numitorul $(Q_3 - Q_2) + (Q_2 - Q_1) = Q_3 - Q_1$ este întotdeauna pozitiv.
Coeficienți pearsonieni

a) $$sk = \frac{\bar{x}_a - Mo}{\sigma}, \qquad (2)$$

diferența între medie și mod împărțită la abaterea pătratică medie (*sk*, notație legată de termenul *skewness*, oblicitate, întrebuințat de Pearson pentru denumirea asimetriei).

b) $$sk = \frac{3(\bar{x}_a - Me)}{\sigma}, \qquad (3)$$

coeficient empiric de asimetrie bazat pe relația între mod, mediană și medie într-o repartiție moderat asimetrică (→ *Mod*), prin care se înlătură inconvenientul determinării poziției modului ce afectează coeficientul calculat după formula (2). Măsura (3) se află între limitele -3 și $+3$.

c) $$sk = \frac{Q_3 + Q_1 - 2Me}{Q}, \qquad (4)$$

coeficient de asimetrie intercuartil (Q reprezintă *semiintercuartila*).

d) $$\beta_1 = \frac{(\mu_3)^2}{(\mu_2)^3}, \tag{5}$$

unde μ_2 și μ_3 sînt respectiv momentele de ordinul 2 și 3 în raport cu media aritmetică (*momente centrate**).

e) $$\chi = \frac{\sqrt{\beta_1}(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)}, \tag{6}$$

coeficient derivat din β_1 și β_2 care nu este decît o evaluare mai exactă a coeficientului (2).

Coeficientul lui Fisher

$$\gamma_1 = \sqrt{\beta_1} = \frac{\mu_3}{\sigma_3}. \tag{7}$$

Față de coeficienții (1), (2), (3) și (4), construiți cu ajutorul unor puncte de reper izolate — modul, mediana, media, cuartila, coeficienții (5), (6) și (7) prezintă avantajul că fac să intre în calcul ansamblul abaterilor față de media aritmetică, singurul care reflectă într-adevăr asimetria repartiției.

II. O repartiție este mai mult sau mai puțin *aplatizată* (sau *boltită*) după cum valorile corespunzătoare unor abateri mici de la valoarea centrală dețin o proporție mai mult sau mai puțin însemnată. Această proprietate a repartiției unimodale, cunoscută sub numele de *exces* (κυρτός — cocoșat), se măsoară prin coeficientul lui Pearson β_2 :

$$\beta_2 = \frac{\mu_4}{(\mu_2)^2} \tag{8}$$

sau derivatul său — coeficientul lui Fisher γ_2 :

$$\gamma_2 = \beta_2 - 3 = \frac{\mu_4}{\sigma^4} - 3. \tag{9}$$

Comparația se face cu boltirea *curbei normale* *, cînd $\beta_2 = 3$. Dacă $\beta_2 > 3$, curba este mai ascuțită spre vîrf decît curba normală și se numește *leptocurtică* (λεπτός — îngust + κυρτός); dacă $\beta_2 < 3$ curba este mai turtită și se numește *platicurtică* (πλατύς, larg + κυρτός)*).

Exemplul 1. Pentru repartiția personalului după retribuție din tab. 19 s-au aflat :

$\bar{x}_a$	= 2722 lei	*Coeficienții de asimetrie*
$Me(Q_2)$	= 2706,9 lei	(1) 0,0783
Mo	= 2690 ,,	(2) 0,0911
Q_1	= 2517,9 ,,	(3) 0,1290
Q_3	= 2928 ,,	
σ	= 351,16 ,,	(4) 0,1565

*) ,,*Student* (W. Gosset — *n.n.*) dă o amuzantă metodă de memorare pentru aceste nume : curbele platicurtice ca și *platypus* (ornitorinc) sînt turtite și au coada scurtă. Curbele leptocurtice sînt înalte, cu coada lungă ca și la cangurii propriu-ziși''. (După G. Udny Yule și M. Kendall, *op. cit.*, p. 181.)

V.TREBICI

Coeficienții obținuți sînt diferiți, după cum era de așteptat, dar toți indică o repartiție ușor asimetrică pozitivă.

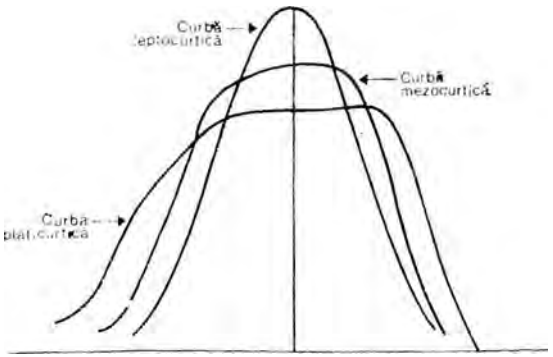


Fig. 62 Curba platicurtică, leptocurtică și mezocurtică

Exemplul 2. Căsătorii după grupa de vîrstă a soțului, 1979.

Tabelul 28

Grupa de vîrstă (ani) x	Număr de căsătorii f	x'	$x'f$	$(x')^2f$	$(x')^3f$	$(x')^4f$	$(x'-1)^4f$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sub 20	7132	−1	−7132	7132	−7132	7132	114112
20−24	97453	0	0	0	0	0	97453
25−29	54676	1	54676	54676	54676	54676	0
30−34	15344	2	30688	61376	122752	245504	15344
35−39	7366	3	22098	66294	198882	596646	117856
40−44	5216	4	20864	83456	333824	1335296	422496
45−49	3565	5	17825	89125	445625	2228125	912640
50−54	2653	6	15918	95508	573048	3438288	1658125
55−59	1894	7	13258	92806	649642	4547494	2454624
60+	2840	8	22720	181760	1454080	11632540	6818840
Total	198139	—	190915	732133	3825397	24085801	12611490

Sursa: *Anuarul statistic al R. S. România 1980*, p. 80.

Proba (testul, verificarea) lui Charlier. Ca mijloc de control al operațiilor cuprinse în tabel s-a recurs la un calcul auxiliar și anume: calculul momentului de ordinul 4 în raport cu o valoare care să difere de originea aleasă cu mărimea unui interval (col. 8, $x'_i - 1$ fiind $-2, -1, 0, 1, 2$ etc.). Totalurile obținute în col. 2, 4, 5, 6, 7 trebuie să verifice relația:

$$\Sigma(x'-1)^4f = \Sigma(x')^4f - 4\Sigma(x')^3f + 6\Sigma(x')^2f - 4\Sigma(x')f + \Sigma f. \qquad (10)$$

V.TREBICI

Într-adevăr,

$$12611490 = 24085801 - (4 \times 3825397) + (6 \times 732133) - (4 \times 190915) + 198139.$$

Paternitatea acestui procedeu de control este atribuită statisticianului suedez C. Charlier. În formulare generală: exactitatea calculului unui moment poate fi controlată prin calculul unei serii suplimentare de abateri, măsurate nu de la *originea arbitrară* întrebuințată pentru a determina momentul respectiv, ci de la o origine care se află la distanța de un interval de clasă mai jos. Pentru momentele de ordin inferior se va întrebuința după caz:

$$\Sigma(x' - 1)^3 f = \Sigma(x')^3 f - 3\Sigma(x')^2 f + 3\Sigma(x')f - \Sigma f \qquad (11)$$

$$\Sigma(x' - 1)^2 f = \Sigma(x')^2 f - 2\Sigma(x')f + \Sigma f. \qquad (12)$$

Fiecare din valorile înscrise în col. 5, 6, 7 se pot obține prin înmulțirea cu x'_i a valorilor coloanei ce o precede, evitîndu-se în acest fel calcularea în prealabil a puterilor $(x')^2$, $(x')^3$, $(x')^4$.

Coeficienții β și γ

a) În raport cu media de lucru (22,5 ani) ca origine, se obțin momentele simple:

$$m_1 = \frac{190915}{198138} = 0{,}963540; \qquad m_3 = \frac{3825397}{198139} = 19{,}306633;$$

$$m_2 = \frac{732133}{198139} = 3{,}695047; \qquad m_4 = \frac{24085801}{198139} = 121{,}560121.$$

b) Din momentele simple se deduc momentele centrate (→ *Momente*):

$$\mu_2 = (3{,}695047 - 0{,}96354^2) \times 5^2 = 69{,}165950;$$

$$\mu_3 = (19{,}306633 - 3 \times 0{,}96354 \times 3{,}695047 + 2 \times 0{,}96354^3) \times 5^3 = 1301{,}846875;$$

$$\mu_4 = (121{,}560121 - 4 \times 0{,}96354 \times 19{,}306633 + 6 \times 0{,}96354^2 \times 3{,}695047 - 3 \times \times 0{,}96354^4) \times 5^4 = 40716{,}580625.$$

c) Cu corecțiile aduse (→ *Corecția lui Sheppard*), momentele de ordinul 2 și 4 sînt:

$$\mu'_2 = 69{,}16595 - \frac{1}{12} 5^2 = 67{,}082617;$$

$$\mu'_4 = 40716{,}580625 - \frac{1}{2} 69{,}16595 \times 5^2 + \frac{7}{240} 5^4 = 40590{,}714062,$$

de unde (formulele (5), (6), (7)):

$$\beta_1 = \frac{1301{,}846875^2}{67{,}082617^3} = 5{,}614;$$

$$\chi = \frac{\sqrt{5{,}614}(9{,}02 + 3)}{2(5 \times 9{,}02 - 6 \times 5{,}614 - 9} = 5{,}897;$$

$$\gamma_1 = \sqrt{5{,}614} = 2{,}369.$$

Curba este pronunțat asimetrică. Formulele (8) și (9):

$$\beta_2 = \frac{40590.714062}{67{,}082617^2} = 9{,}02\,;$$

$$\gamma_2 = 9{,}02 - 3 = 6{,}02.$$

Curba este puternic leptocurtică. De remarcat este faptul că, întrucît coeficienții sînt numere abstracte, nu are importanță dacă se lucrează — în exemplul tratat — în unități de cinci ani sau în unități de un an.

Curbele lui Pearson. Curbe reprezentînd douăsprezece repartiții de frecvențe elaborate de K. Pearson și numite repartiție de tip I, repartiție de tip II, repartiție de tip III etc.

Ecuația diferențială a familiei de curbe Pearson este

$$\frac{df}{dx} = \frac{x - a}{b_0 + b_1 x + b_2 x^2 + }\,, \tag{13}$$

unde df înseamnă densitatea de repartiție. Constantele a, b_0, b_1, b_2 pot fi exprimate în funcție de primele patru momente, dacă acestea există. Prin transformări convenabile, multe din repartițiile importante din statistică pot fi deduse din ecuația (13).

Un caz special al repartiției de tip I este *repartiția beta* (B). Repartiția de tip III se numește de asemenea *repartiția gama* (Γ).

Curbă de concentrare (curbă Lorenz—Gini). Tip de *ogivă* adaptat pentru reprezentarea disparității între părțile relative atribuite claselor în repartiția valorilor globale ale variabilei observate. Curba se trasează prin punctele M de coordonate (p, q) definite de:

p_i, frecvența relativă cumulată pînă la $x = x_i$;

q_i, citul prin $\bar{x}_a \Sigma f$ a sumei tuturor valorilor observate pînă la $x = x_i$;

(x_i — limită superioară de clasă).

Valoarea lui q_i este în general calculată, cu aproximație, sub forma

$$q_i = \frac{\sum_1^i x_i f_i}{\sum_1^k x_i f_i}\,, \tag{14}$$

x_i *) fiind centrul clasei de frecvență f_i și k numărul de clase.

Exemplu. Tab. 29 cuprinde datele pentru construirea curbei de concentrare a industriei noastre republicane.

Cele două serii de date puse în corespondență sînt:

p_i, procentul cumulat al numărului de întreprinderi (col. 3);

q_i, procentul cumulat al numărului de muncitori (col. 5).

Examinarea comparativă a seriilor de procente oferă o idee despre gradul de concentrare, dar o idee vagă. Pentru a o preciza, se stabilește un grafic astfel conceput:

valorile p % — domeniu de la 0 la 100%, înscrise în abscise;

valorile q % — domeniu de la 0 la 100%, înscrise în ordonate.

*) Valorile q_i sînt calculate pe baza datelor înregistrate, și nu ca estimații conform formulei (14).

Tabelul 29

Grupe de întreprinderi după numărul muncitorilor	Întreprinderi		Muncitori	
	Frecvențe simple	Frecvențe cumulate	Frecvențe simple	Frecvențe cumulate
(1)	(2)	(3)	(4)	(5)
Pînă la 200	9,9	9,9	0,7	0,7
201— 500	13,0	22,9	2,3	3,0
501—1000	18,3	41,2	7,0	10,0
1001—2000	25,3	66,5	18,9	28,9
2001—3000	15,1	81,6	18,7	47,6
3001—5000	11,0	92,6	21,3	68,9
Peste 5000	7,4	100,0	31,1	100,0

Sursa : *Anuarul statistic al R. S. România 1980*, p. 136.

Curba de concentrare este deci reprezentarea grafică a lui q în funcție de p (fig. 63).

Suprafața mărginită de curba de concentrare și diagonala OB poartă numele de *arie de concentrare* : cu cît această arie este mai importantă cu atît concen-

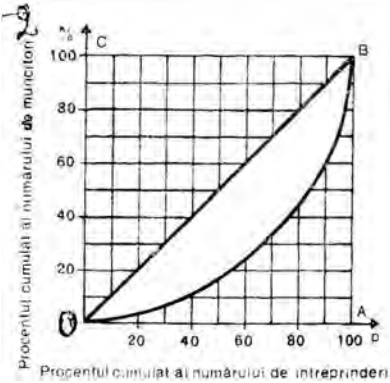


Fig. 63 Curba de concentrare a întreprinderilor industriale

trarea este mai puternică. La limită (în ipoteza teoretică de concentrare maximă), curba de concentrare s-ar confunda cu linia frîntă OAB. Și dacă repartiția ar fi perfect egalitară (9,9% de întreprinderi ar dispune de 9,9% din numărul total de muncitori, 22,9% ar dispune de 22,9% etc.), curba de concentrare s-ar reduce la diagonala OB, numită uneori *dreaptă de echilibru* sau *dreaptă de echirepartiție.**)

Coeficient de concentrare (G. indicele lui Gini). Măsură a concentrării definită de relația :

$$G = \frac{\text{Aria de concentrare}}{\text{Aria triunghiului } ABC}. \qquad (15)$$

*) Reprezentarea grafică a concentrării a fost introdusă de punct de M. O. Lorenz și C. Gini.

VASILE

Calcul practic : se transpun punctele reprezentative pe o hîrtie milimetrică și se socotește numărul de pătrățele (mm^2) cuprins între diagonală și curbă ; dublul acestui număr se împarte la numărul de pătrățele (mm^2) cuprins în pătrat. Acest calcul nu este decît aproximativ, dar, în materie de coeficient de concentrare, nu se caută decît un ordin de mărime convenabil. Aplicat la exemplul precedent (fig. 63), se găsește un coeficient $G = 0{,}5$. Coeficientul de concentrare variază între 0 (diagonală de echirepartiție, concentrare nulă) și 1 (laturile pătratului, concentrare maximă).

O altă măsură a concentrării propusă de Gini (1912) este egală cu raportul între diferența medie (→ *Diferența medie a lui Gini*) și dublul mediei aritmetice :

$$G = \frac{Dm}{2\bar{X}_a} = \frac{1}{2}\frac{Dm}{\bar{x}_a}. \qquad (16)$$

În cercetările statistice, noțiunea de concentrare se înrudește cu cea de dispersie, întrucît ea se raportează la intensitatea grupării datelor în jurul valorii centrale.

IX. CORELAȚIA ȘI REGRESIA

CORELAȚIE (STATISTICĂ). 1) Termen general folosit pentru a defini interdependența sau legătura între variabilele observate în populații statistice (și deci sinonim cu *legătura statistică*).

2) Metodă de descriere și analiză a legăturilor de tip statistic între două sau mai multe variabile.

În terminologia statistică, conceptul de c. apare, uneori, cu un înțeles foarte larg, acoperind orice legătură statistică — fie între variabile cantitative, fie între variabile calitative. Alteori, noțiunea de c. semnifică numai relațiile dintre variabilele cantitative. Într-un sens și mai restrîns, se numește c. numai măsura gradului de legătură între variabilele cantitative, respectiv *coeficientul de corelație* * sau *raportul de corelație* *.

Istoric. Ideea unei dependențe între variabilele observate în colectivități statistice este atribuită lui *F. Galton**. În prima sa lucrare asupra eredității, comunicată în anul 1877, se arăta — pe exemplul plantei *Lathyrus odoratus* (sîngele voinicului) — că talia medie a descendenților este legată liniar de talia ascendenților. Ulterior, în studiile despre ereditate la om, adunate în cartea *Natural inheritance* (1889), Galton folosește noțiunea de „variație legată" și imaginează, cel dintîi, o metodă statistică practică de exprimare a legăturii, precum și o unitate de măsură a gradului de legătură, cunoscută sub denumirea „funcția lui Galton".

Stimulat de lucrările lui Galton, K. Pearson cercetează repartițiile bidimensionale, mai ales din domeniul biologiei, unde dispunea de datele cele mai bogate, indică ecuația liniei drepte adecvată problemei de *regresie** a lui Galton și introduce celebrul său coeficient de corelație. Alături de Pearson, numeroși alți cercetători de prestigiu au contribuit la dezvoltarea teoriei corelației și la promovarea aplicațiilor ei în variate domenii : *F. Y. Edgeworth**, *A. L. Bowley**, *G. U. Yule**, iar mai tîrziu, *R. A. Fisher**, *A. A. Ciuprov* *, *C. Spearman,** M. Ezekiel, *N. S.* Cetverikov, *M. G. Kendall** ș.a.

V.TREBICI

Chiar de la începuturile ei, metoda cunoscută obișnuit sub numele de *metoda corelației* era socotită drept „... o metodă generală de o mare forță și de un cîmp de aplicație întins, pentru a investiga natura relațiilor existente între variabilele statistice". Matematic, ea se traduce prin „ ... formarea ecuație ce dă valoarea unei variabile în funcție de o alta sau de mai multe altele" și prin determinarea valorii coeficientului de corelație, care „... măsoară [redacted] ecuația de regresie se apropie" (Yule). Așadar, păstrînd concepția originară, metoda corelației întrunește — la dezvoltarea ei actuală — întregul complex de investigație a conexiunilor statistice, conducînd la expresii numerice sintetice : ecuații de regresie, coeficienți de regresie, coeficienți de corelație, coeficienți de determinare, erori de estimare etc. Cu toate că metoda corelației, în sensul restrîns al cuvîntului, și metoda regresiei diferă prin premisele lor, totuși ele se împletesc atît de strîns în analiza statistică de ansamblu a legăturii încît valoarea lor de cunoaștere poate fi apreciată numai judecîndu-le împreună, într-o sinteză metodologică. (→ *Regresie.*)

REGRESIE. Metodă de cercetare a unei relații predeterminate exprimînd legătura care există între o variabilă y, numită *variabilă dependentă** (*variabilă explicată* * sau *variabilă endogenă* *) și una sau mai multe variabile $x_1, x_2, \ldots,$ numite *variabile independente* (*variabile explicative* * sau *variabile exogene* *). Fie y o variabilă dependentă de variabilele $x_1, x_2, \ldots$ Relația exactă sau *ecuația de regresie* * a lui y în funcție de $x_1, x_2, \ldots,$ pusă sub forma

$$y = f(x_1, x_2, \ldots)$$

definește o *curbă* sau o *suprafață de regresie* ; ea are menirea să permită, pentru valorile date $x_1, x_2 \ldots,$ calculul unei estimații a lui y, cu variațiile aleatoare, neexplicate prin variabilele $x_1, x_2, \ldots$

Termenul r[redacted]) a fost folosit, pentru prima dată, ca un concept statistic, în [redacted]n în memoriul [redacted] [redacted]e", inclus în cartea sa „Natural inheritance" (1889). Studiind relația dintre talia copiilor și talia părinților, Galton a descoperit că din părinți a căror talie este mai mică decît media colectivității provin copii cu o talie superioară lor și invers. Din prelucrarea observațiilor asupra a 928 cazuri, conform tabelului construit de Galton, rezultă „mediile legate" ($\bar{y}$) :

x	sub 64,5	64,5	65,5	66,5	67,5	68,5	69,5	70,5	71,5	72,5	peste 72,5
$\bar{y}$	65,0	65,3	66,7	67,1	68,1	68,0	68,7	68,2	68,1	67,2	72,9

în care x reprezintă talia părinților (toli), $\bar{y}$ — talia medie a copiilor corespunzătoare taliei părinților din fiecare grupă (toli). Talia medie a părinților în colectivitatea generală este egală cu [redacted].

Se constată că înălțimea copiilor născuți din părinți înalți tinde să „regreseze" spre înălțimea medie a populației. Galton a indicat termenul de **r.** ca denumire a procesului general de predicție a unei variabile (înălțimea copiilor) din o altă variabilă (înălțimea părinților).

Puține fenomene pot fi descrise cu suficientă exactitate prin relații sau modele exacte. Ținînd seama de natura întîmplătoare a fenomenelor, de legitățile, de repetabilitatea lor este mai adecvat să se folosească modele de *dependență stohastică**, în care elementul aleator să apară sub formă de erori aditive.

Așadar, ecuația de regresie se scrie sub forma :

$$y = f(x_1, x_2, \ldots,) + \varepsilon$$

unde ε este o eroare aleatoare.

V.TREBICI

TABEL DE CORELAȚIE *Tabel cu dublă intrare**, în cazul repartiției după variabile cantitative, utilizat în calculele de corelație

Exemplu : Corelația între înălțimea (x = cm) și greutatea (y = kg) corpului la copii de 10 ani de sex masculin (tab. 30).

În capetele de rînduri și coloane ale tabelului se trec clasele de valori sau valorile distincte ale variabilelor x și y, iar în rubrici — frecvențele perechilor de valori $x_i y_j$. Tabelul conține rînduri și coloane suplimentare cu ajutorul cărora se determină elementele necesare obținerii *ecuației de regresie*, calculului *coeficientului de corelație* și al altor valori sintetice în analiza de corelație. Pentru asemenea calcule se poate folosi și un tabel redus, care este de aceeași formă cu tabelul, 30, dar în care variabilele x și y se înlocuiesc cu variabilele x' și y', date de formulele :

$$x' = \frac{x - C_x}{\Delta x} \; ; \qquad y' = \frac{y - C_y}{\Delta y} \, ,$$

unde C_x și C_y sînt origini de calcul convenabil alese pentru x și y, și Δx și Δy sînt, respectiv, mărimile intervalelor de grupare (constante), corespunzătoare lui x și y.

Variabilele x' și y' ușurează calculele.

Se va putea adopta, de exemplu, tab. 31 sau oricare altul echivalent.

Valorile variabilelor corelate s-au micșorat astfel :

$$x' = \frac{x - 131}{1} \quad \text{și} \quad y' = \frac{y - 29}{1} \, .$$

CORELAȚIE APARENTĂ (*corelație falsă*). Corelație constatată între două variabile, dar care nu are nici o semnificație explicativă sau cauzală. Este adesea imprudent să se interpreteze corelația în termeni de cauzalitate. Corelația este doar o reflectare a legăturilor cu caracter complex existente între fenomenele de masă și, ca atare, să nu se uite că o reflectare poate ușor să fie o aparență înșelătoare. Statisticile nu pot oferi informații despre natura legăturilor între două variabile ; ele pot spune cel mult pînă la ce grad cele două variabile tind să varieze concomitent. Corespondența între variabile poate rezulta din diverse condiții : o legătură nemijlocită între fenomene, unilaterală sau cu reacții inverse (interdependență), covariație a fenomenelor datorată unor cauze comune, un simplu paralelism întîmplător în variația a două sau mai multe fenomene. În prima categorie intră, de exemplu, dependențele economice : productivitatea muncii — calificarea forței de muncă, productivitatea muncii — înzestrarea tehnică, cererea de mărfuri — veniturile bănești sau interdependențele : venit național — acumulare, producție — — cerere, productivitate — remunerare. În alte cazuri, două fenomene se modifică în același sens sau în sensuri opuse, cu o anumită regularitate, ca urmare a modificării unui al treilea fenomen, fără ca între primele două fenomene să existe vreo legătură. Astfel, examinarea datelor statistice relevă variația concomitentă a cererii de mărfuri și a economiilor bănești ale populației. Un calcul de corelație, să spunem, ar fi lipsit de sens, fiecare variabilă fiind legată de veniturile bănești ale populației.

Pentru a treia situație se citează deseori următorul exemplu tipic : un autor a compilat, pentru diferite anotimpuri, frecvența cuiburilor de barză reperate

Tabelul 30

x \ y	24	25	26	27	28	29	30
126	1	1	3	1			
127		3	4	1		1	
128	1	4	5	7	1	2	1
129		1	2	6	9	6	4
130			1	7	36	17	6
131			1	6	27	56	39
132				2	10	16	26
133					1	7	10
134						1	2
135							1
136							
137							
f_y	2	9	16	30	84	106	89
yf_y	48	225	416	810	2352	3074	2670
y^2f_y	1152	5625	10816	21870	65856	89146	80100
Σyxf_{xy}	6096	28700	53170	104841	306852	402520	350940

Sursa: P. Mureșan, *Manual de statistică sanitară*, Editura medicală, 1980, București, p. 101 (datele din rîndurile și coloanele suplimentare sînt calculate de C. Moineagu).

V.TREBICI

(*continuare*)

31	32	33	34	35	f_x	xf_x	x^2f_x	Σxyf_{xy}
					6	756	95256	19404
					9	1143	145161	29845
					21	2688	344064	71552
1					29	3741	482589	105264
2	1				70	9100	1183000	258700
18	4	1			152	19912	2608472	583867
7	2	1			61	8448	1115136	249744
11	6	2	1		38	5054	672182	154812
4	7	3	1		18	2412	323208	76380
2	3	1	1		8	1080	145800	34425
	1	2	2	1	6	816	110976	27336
		1	2	1	4	548	75076	18632
45	24	11	7	2	425	55698	7300920	1629961
1395	768	363	238	70	12429			
43215	21576	11979	8092	2450	364907			
184140	102240	48675	32232	9555	1629961			

Tabelul 31

x' \ y'	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	$f_{x'}$	$x'f_{x'}$	$x'^2f_{x'}$	$\Sigma x'y'f_{x'y'}$
−5	1	1	3	1									6	−30	150	100
−4		3	4	1		1							9	−36	144	104
−3	1	4	5	7	1	2	1						21	−63	189	150
−2		1	2	6	9	6	4	1					29	−58	116	50
−1			1	7	36	17	6	2	1				70	−70	70	40
0			1	6	27	56	39	18	4	1			152	0	0	0
1				2	10	16	26	7	2	1			64	64	64	36
2					1	7	10	11	6	2	1		38	76	152	121
3						1	2	4	7	3	1		18	54	162	141
4							1	2	3	1	1		8	32	128	92
5									1	2	2	1	6	30	150	135
6										1	2	1	4	24	144	120
fy'	2	9	16	30	84	106	89	45	24	11	7	2	425	23	1469	1095
$y'fy'$	−10	−36	−48	−60	−84	0	89	90	72	44	35	12	104			
y'^2fy'	50	144	144	120	84	0	89	180	216	176	175	71	1450			
$\Sigma x'y'$ fxy	40	124	153	94	45	0	39	90	153	136	155	66	1095			

într-o suburbie a Londrei și de asemenea frecvența nașterilor de copii în aceeași suburbie. Datele arătau că perioada de cea mai înaltă frecvență pentru o variabilă era și perioada de cea mai înaltă frecvență pentru cealaltă variabilă; de asemenea cînd nașterile erau mai puțin numeroase, cuiburile erau mai rare. Statistic vorbind, corelația între cele două fenomene era ridicată. Dar s-ar putea vedea aici o relație de la cauză la efect? *)

Asemenea cazuri, nu prea rare și nici atît de simple, atrag atenția că sînt necesare eforturi pentru a distinge, în complexitatea fenomenelor, *legăturile cauzale* de *legăturile formale*. Succesul inferenței spre o relație cauzală printr-o relație statistică depistată empiric poate fi asigurat de cunoașterea temeinică a domeniului abordat și de experiența personală.

ECUAȚIE DE REGRESIE (*ecuație de estimare*). Relație matematică prin care se exprimă dependența între două sau mai multe variabile.

Dacă s-a admis, fie și numai ipotetic, dependența unui fenomen de unul sau mai mulți factori, atunci apare întrebarea: ce model matematic trebuie adoptat, ca să fie propriu dependenței cercetate și apt pentru [redacted] și, eventual, prev[redacted]e? Din punct de vedere metodologic, problema se reduce la stabilirea formei e.r., destinată să descrie relația dintre *variabila dependentă* * și *variabilele independente* * incluse în model. O soluție convenabilă presupune, în primul rînd, înțelegerea esenței și mecanismului legăturii, reprezentarea legității schimbării fenomenului modelat sub acțiunea factorilor aleși, iar, după aceea, alegerea — din arsenalul funcțiilor matematice — a tipului de funcție adecvat legității sesizate.

Cazul elementar de aplicare a metodei regresiei îl constituie legătura dintre două variabile (fig. 61, 65, 66, 67 și 68):

$$y = f(x) + \varepsilon.$$

Numeroase legături de acest gen se înscriu pe curbe definite de următoarele ecuații de regresie:

$$y = a + bx + \varepsilon \tag{1}$$

$$y = ax^b + \varepsilon \tag{2}$$

$$y = ab^x + \varepsilon \tag{3}$$

$$y = a + bx + cx^2 + \varepsilon \tag{4}$$

$$y = a + b\frac{1}{x} + \varepsilon. \tag{5}$$

Modelul liniar (1) semnifică [redacted] variabilei dependente [redacted] influența schimbării variabilei independente (**). Oricare ar fi nivelul variabilei independente, creșterea sau descreșterea ei cu o unitate se consideră că produce, în fiecare caz, creșterea sau descreșterea variabilei dependente cu aceeași mărime (mărimea [redacted] — *coeficientul de regresie**). Ecuația liniară își găsește aplicarea și cînd variațiile variabilelor nu urmează [redacted] dar ele sînt mici și, ca atare, curba descrisă este foarte aproape de

*) Citat de Will Scarlet în: Emmet A. Betts, *Foundations of Reading Instruction*, New York, American Book Co. 1946, p. 123.

**) Pentru funcția liniară (1), proporționalitatea directă există între $y - a$ și x, și nu între y și x.

linia dreaptă. Funcția putere (2) implică un ritm de variație constant, de unde și valabilitatea ei pentru fenomene cu evoluție continuă, neadmitînd — cel puțin în cuprinsul perioadei de analiză sau/și de previziune — nivel de saturație.

Aproximarea exponențială (3) este potrivită cînd variabila independentă crește în progresie aritmetică, iar variabila independentă crește în progresie geometrică. Parabola (4) și hiperbola (5) sînt specifice exprimării dependențelor cu o alură ce tinde către un punct maxim (minim); dacă acest punct este depășit curbele sau capătă stabilitate sau descresc (cresc).

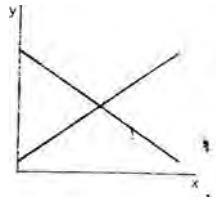


Fig. 64 Reprezentarea grafică a funcției $Y = a + bx$

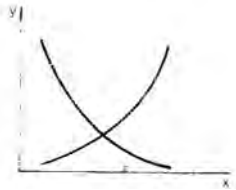


Fig. 65 Reprezentarea grafică a funcției $Y = ab^x$

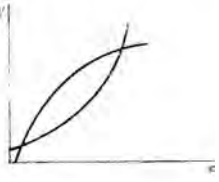


Fig. 66 Reprezentarea grafică a funcției $Y = ax^b$

Curbele corespunzătoare e.r. se numesc *curbe de regresie*. Dacă dependența lui y de x este liniară, *regresia* este numită *liniară* sau *dreaptă de regresie* *. Regresia definită de relația $y = f(x) + \varepsilon$, unde $f(x)$ nu este liniară, se numește *regresie curbilinie*.

Cele mai multe fenomene ale vieții reale — economice, sociale, biologice etc. — sînt afectate de variațiile mai multor factori; în acest caz se stabilește, prin extensiunea ecuației regresiei simple, *ecuația de regresie multiplă* *. De exemplu, ecuația de regresie liniară (1):

$$x_1 = a + b_{12.34\ldots n}x_2 + b_{13.24\ldots n}x_3 + \ldots + b_{1n.23\ldots(n-1)}x_n.$$

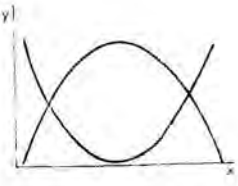


Fig. 67 Reprezentarea grafică a funcției $Y = a + bx + cx^2$

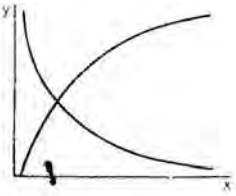


Fig. 68 Reprezentarea grafică a funcției $Y = a + b\frac{1}{x}$

Statisticienii au creat termenul de *regresie multiplă* pentru descrierea procesului prin care diverse variabile sînt folosite ca să se estimeze o altă variabilă. (→ *Dreaptă de regresie.*)

V.TREBICI

Dacă e.r. este de forma :

$$y(x) = a + bx + cx^2 + \ldots .$$

atunci se obține regresia ce se numește *polinomială* *. Coeficienții $a, b, c, \ldots$ se determină prin *metoda celor mai mici pătrate* *. În cazul în care valorile variabilei independente x sînt echidistante, coeficienții regresiei polinomiale se pot determina și prin metoda *polinoamelor ortogonale* *, care constă în determinarea regresiei polinomiale sub forma :

$$y(x) = b_0 P_0(x) + \ldots + b_k P_k(x),$$

unde $P_{k(x)}$ sînt polinoame de gradul k în x satisfăcînd relația de ortogonalitate :

$$\sum_{i=1}^{n} P_r(x_i) P_s(x_i) = 0, \quad r \neq s$$

pentru $x_1, \ldots, x_n$ observații asupra variabilei independente x. Pentru diferite valori ale lui n, s-au întocmit tabele care dau direct valorile polinoamelor $P_k(x)$ (Fisher, Jordan).
Cu ajutorul acestor valori se calculează coeficienții $\{b\}$ ai regresiei polinomiale :

$$b_k = \frac{\sum_{i=1}^{n} y P_k(x_i)}{\sum_{i=1}^{n} P_k^2(x_i)} .$$

Exactitatea ajustării valorilor calculate $y(x)$ față de valorile observate $\{y\}$ se îmbunătățește prin calcularea termenului complementar $b_{k+1} P_{k+1}(x)$, termenii precedenți rămînînd neschimbați.

DREAPTĂ DE REGRESIE (*dreaptă de estimare*). Denumirea rezervată, în general, dreptei definite de ecuația de regresie liniară $y = a + bx$.
Această dependență liniară este un model determinist și nu reflectă exact legătura dintre y și x. Valorile observate (x_i și y_i) nu se găsesc exact pe dreapta

$$y = a + bx$$

și, de fapt, unei valori x_i îi pot corespunde mai multe valori y_i.
Așadar, dependența liniară dintre y și x trebuie considerată ca o *dependență stohastică**, ținînd seama de componentele aleatoare sau nedeterminate ale modelului.
Comportarea lui y este descrisă de o anumită repartiție statistică, a cărei valoare medie depinde de x și este fixă cînd se specifică valoarea lui x

$$M(y/x) = a + bx.$$

Acest model presupune că variabila dependentă y_i este egală în medie cu o funcție liniară de x, depinzînd de parametrii necunoscuți a, b.
Abaterea unei valori observate y_i față de valoarea ei medie $M(y_i|x_i)$

$$e_i = y_i - M(y_i|x_i) = y_i - a - bx_i$$

reprezintă eroarea de observare (sau de măsurare). Valorile $\{e_i\}$ sînt considerate variabile aleatoare numite [redacted]e sau [redacted]* care satisfac condițiile :

$$M(e_i) = 0$$

$$\text{var } e_i = D^2(e_i) = M\,(e_i - M(e_i)^2 = \sigma^2, \quad \forall_i$$

$$\text{cov } (e_i,\ e_j) = M(e_i - M(e_i))(e_j - M(e_j) = M(e_i e_j) = 0, \quad i \neq j.$$

D.r. este numită [redacted]. Estimarea parametrilor a, b se face prin *metoda celor mai mici pătrate* * sau în *sensul celor mai mici pătrate*, adică în așa fel încît să se facă minimă suma pătratelor abaterilor între punctele observate și punctele corespunzătoare ale dreptei. Dispunînd de o serie de valori observate $(x_i,\ y_i)$, suma pătratelor abaterilor de minimizat este :

$$S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - y_j^*)^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2.$$

Valorile x_i și y_i fiind cunoscute, această sumă este funcție numai de parametrii a și b. Minimul poate fi determinat anulînd derivatele parțiale în raport cu a și în raport cu b :

$$\frac{\partial s}{\partial a} = \frac{\partial s}{\partial b} = 0.$$

Se obțin astfel *ecuațiile normale* * :

$$\begin{cases} \sum_{i=1}^{n} (y_i - a - bx_i) = 0 \\ \sum_{i=1}^{n} x_i(y_i - a - bx_i) = 0 \end{cases}$$

sau

$$\begin{cases} na + b \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_1 \\ a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i. \end{cases}$$

Prima ecuație arată că suma algebrică a abaterilor între valorile observate și ordonatele corespunzătoare ale **d.r.** este nulă : abaterile negative, relative la punctele situate sub dreaptă, compensează întocmai abaterile pozitive, relative la punctele situate deasupra dreptei. Se deduce astfel :

$$\bar{y} = a + b\bar{x} \text{ sau } a = \bar{y} - b\,\bar{x}.$$

adică **d.r.** trece prin *punctul mediu* $(\bar{x},\ \bar{y})$.

Pe de altă parte, înmulțind prima ecuație cu $\bar{x}$ și scăzînd-o din a doua, se obține :

$$b = \frac{\sum_{i=1}^{n} x_i y_i \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\text{cov}(x, y)}{s_x^2}.$$

Dezvoltînd produsele $(x_i - \bar{x})(y_i - \bar{y})$, se vede într-adevăr că :

$$\text{cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) =$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)\right].$$

D.r. a lui y în x are deci drept ecuație

$$y = \frac{\text{cov}(x, y)}{s_x^2}(x - \bar{x}) + \bar{y} \quad \text{sau} \quad y = b_{yx}(x - \bar{x}) + \bar{y}$$

coeficientul b_{yx} fiind [redacted].

Exemplu. **D.r.** a greutății corpului în funcție de înălțime la copii de 10 ani de sex masculin (tab. 30).

$$\bar{x} = \frac{55698}{425} = 131{,}05 \text{ cm}; \quad \bar{y} = \frac{12429}{425} = 29{,}24 \text{ kg}.$$

$$s_x^2 = \frac{7300920 - \frac{(55698)^2}{425}}{425} = 3{,}45;$$

$$\text{cov}(x, y) = \frac{1629961 - \frac{55698 \times 12429}{425}}{425} = 2{,}56;$$

Ecuația **d.r.** :

$$y^* = \frac{2.56}{3{,}45}(x - 131{,}05) + 29{,}24 = 0{,}742\,x - 67{,}99.$$

D.r. înlocuiește *norul statistic*, rezumînd legătura între y și x în sensul $y \to x$ Se notează obișnuit $\text{Dy}(x)$. Dacă legătura lui x cu y în sensul $x \to y$ prezintă interes, se calculează **d.r.** a lui x în raport cu y, notată $\text{D}x(y)$.
În exemplul considerat :

$$x^* = \frac{2{,}56}{3{,}35}(y - 29{,}24) + 131{,}05 = 0{,}764\,y + 108{,}71.$$

Pe grafic (fig. 69) s-au reprezentat și regresiile empirice (legătura între *mediile condiționate* * și valorile variabilei independente) : regresia lui y în x și regresia lui x în y. Forma liniei de regresie teoretică poate fi dedusă din aceea a liniei de regresie empirică (statistică). Cu referire la repartiția tratată apare pe deplin îndreptățită alegerea unei linii de regresie a cărei formă să fie linia dreaptă.

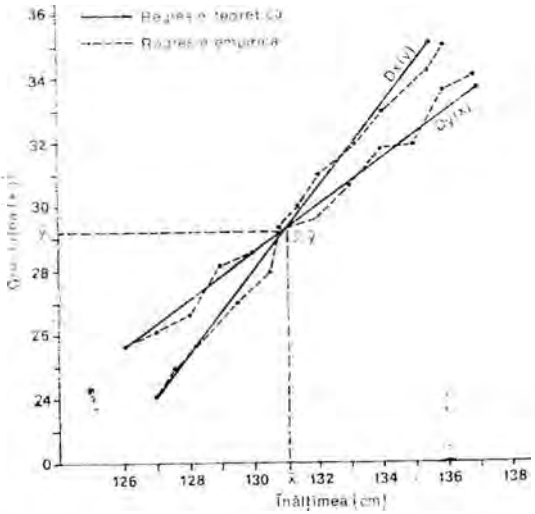


Fig. 69 Drepte de regresie ale repartiției copiilor, de 10 ani, de sex masculin, după greutate (y) și înălțime (x)

Regresia liniară * este cea mai des întîlnită, cea mai ușor de tratat datorită simplicității, dar în același timp și cea mai bogată în concluziile la care poate duce.

În cazul în care distanțele dintre valorile observate și punctele corespunzătoare ale **d.r.** sînt măsurate perpendicular pe dreaptă, **d.r.** se numește *regresie ortogonală* (denumire introdusă de R. Frisch, în 1934).

În cazul în care ecuația **d.r.** este de forma

$$\frac{y - \bar{y}}{\sigma_y} = \frac{x - \bar{x}}{\sigma_x}$$

regresia se numește *regresie diagonală* (termen introdus de asemenea de R. Frisch, în 1931).

Dacă $\sigma_y = \sigma_x$, atunci represia diagonală coincide cu regresia ortogonală.

Dreaptă de regresie a probiturilor. D.r. reprezentînd variația procentului de indivizi ce prezintă o reacție determinată la doze diferite ale unor stimuli în funcție de logaritmul dozei administrate. În ipoteza că procentele sînt repartizate aproximativ normal se spune că procentele sînt exprimate în *probituri*.

V.TREBICI

COEFICIENT DE REGRESIE. În ecuația de regresie

$$y = a + bx,$$

constanta **b**, definită de relația

$$b_{yx} = r \frac{s_y}{s_x}.$$

este [redacted] în raport cu variabila x.
În sens geometric, [redacted] reprezintă [redacted] de regresie.

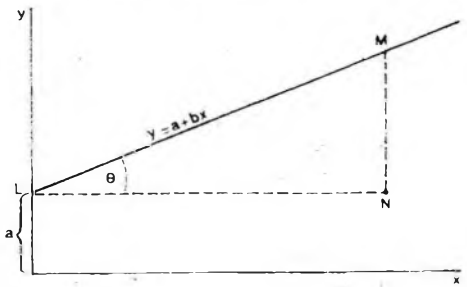


Fig. 70 Reprezentarea grafică a coeficientului de regresie

Constanta **a** este valoarea lui y cînd x este egal cu zero ; cu alte cuvinte, aceasta este distanța numerică între origine și punctul de intersecție al axei y și linia de regresie. Constanta **b** este valoarea tangentei pentru unghiul format între linie și o orizontală dusă din aceasta.

În fig. 70, $b = \text{tg } \theta = \frac{MN}{NL}$.

În calculul corelației [redacted] unei variabile y, cînd x [redacted], adică răspunde tocmai la problema regresiei.
În cazul *corelației directe* sau *pozitive* (variațiile celor două fenomene se produc în același sens), **c.r.** capătă valoare pozitivă, iar în cazul *corelației inverse* sau *negative* (variațiile sînt de sens contrar) — valoare negativă.
O formulă comodă pentru calculul **c.r.**
Ecuațiile normale sînt :

I. $na + b\Sigma x = \Sigma y$

II. $a\Sigma x + b\Sigma x^2 = \Sigma xy.$

Rezolvînd sistemul în raport cu a și b se obțin :

$$\hat{b} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

$$\hat{a} = \frac{\Sigma y - b\Sigma x}{n}$$

V.TREBICI

Exemplu. Cu elementele de calcul din tab. 30.

$$b_{y/x} = \frac{(425 \times 1629961) - (55698 \times 12429)}{(425 \times 7300820) - (55698)^2} = 0{,}74 \text{ kg.}$$

O modificare a înălțimii cu 1 cm este însoțită de o modificare a greutății cu 0,71 kg, în medie.

$$b_{x/y} = \frac{(425 \times 1629961) - (55698 \times 12429)}{(425 \times 364907) - (12429)^2} = 0{,}76 \text{ cm.}$$

Dacă s-ar considera înălțimea în funcție de greutate, atunci unei modificări cu 1 kg a greutății îi corespunde o modificare a înălțimii cu 0,75 cm, în medie.

Observație. În legătură cu analiza regresiei s-a pus problema [redacted] ecuației de regresie : cunoașterea formei analitice a legăturii lui y față de x nu ne permite să determinăm legătura lui x față de y? Dacă, de exemplu, în ecuația de regresie greutate (y)/înălțime (x)

$$y^* = -67{,}99 + 0{,}742\ x.$$

Coeficientul b arată o modificare a greutății la 1 cm înălțime, în medie, cu 0,74 kg, de aici n-ar reieși că la 1 kg greutate înălțimea va trebui să se modifice, în medie cu $1/0{,}742 = 1{,}34$ cm? [redacted] ție cunoscută în matematică sub numele de inversiunea funcțiilor este o imposibilitate" (*N. St. Georgescu-Roegen*). Această operație este posibilă numai în cazul legăturii rigide, caracterizată printr-o strictă proporționalitate în variațiile variabilelor corelate și, ca urmare, liniile de regresie y în x și x în y coincid, iar **c.r.** $b_{y/x}$ și $b_{x/y}$ sînt mărimi inverse.

Dependența statistică se definește tocmai prin aceea că variațiile variabilelor corelate nu se află în raport de strictă proporționalitate, coeficienții de corelație au valori subunitare, liniile de regresie y în x și x în y nu se suprapun, iar **c.r.** nu sînt mărimi inverse. Contradicția aparentă se poate ușor explica, avînd în vedere împrejurarea că ecuația de regresie y/x se construiește pe baza datelor statistice grupate după variabila x, iar ecuația de regresie x/y pe baza datelor grupate după variabilă y, ceea ce antrenează după sine și schimbarea în raportul dintre variațiile ambelor variabile.

REPARTIȚIE MARGINALĂ. Repartiție figurînd la marginea (dreaptă sau inferioară) a *tabelului cu dublă intrare.* *

Coloana marginală a tabelului cu dublă intrare, care conține frecvențele $f_{i\cdot}$, corespunzătoare fiecărei valori x_i a variabilei X, constituie **r.m.** a lui X. În mod analog, ultimul rînd al aceluiași tabel, care conține frecvențele $f_{\cdot j}$ constituie **r.m.** a lui Y.

Fiecăreia din aceste repartiții îi corespunde, dacă x și y sînt variabile cantitative, o *medie marginală* ($\bar{x}, \bar{y}$) și o *varianță marginală* (s_x^2, s_y^2).

Exemplu. Repartițiile marginale ale copiilor de 10 ani de sex masculin (tab. 30) : — după înălțime (coloana f_x) :

$$\text{media marginală } \bar{x} = \frac{\Sigma\, x f_x}{\Sigma f} = \frac{55698}{525} = 131{,}05 \text{ cm.}$$

varianța marginală $s_x^2 = \dfrac{\Sigma x^2 f_x - \dfrac{(\Sigma x f_x)^2}{\Sigma f}}{\Sigma f} = \dfrac{7300920 - \dfrac{(55698^2)}{425}}{425} = 3{,}15$

$s_x = 1{,}86$ cm.

— după greutate (rîndul f_y) :

media marginală $\bar{y} = \dfrac{\Sigma y f_y}{\Sigma f} = \dfrac{12429}{425} = 29{,}24$ kg

varianța marginală $s_y^2 = \dfrac{\Sigma y^2 f_y - \dfrac{(\Sigma y f_y)^2}{\Sigma f}}{\Sigma f} = \dfrac{364907 - \dfrac{(12429)^2}{425}}{425} = 3{,}35$

$s_y = 1{,}83$ kg.

REPARTIȚIE CONDIȚIONATĂ (*repartiție legată*). Într-un *tabel cu dublă intrare* *, repartiția care dă pentru o modalitate, o valoare sau o clasă de valori ale uneia din variabile repartiția relativă la cealaltă variabilă.

Repartiția după variabila X a $f_{\cdot j}$ unități prezentînd valoarea y_j a variabilei Y constituie r.c. *a lui* X *legat de* $Y = y_j$. În mod analog, linia i a tabelului descrie repartiția după variabila Y a $f_{i\cdot}$ unități prezentînd valoarea x_i a variabilei X și constitute r.c. a lui Y *legat de* $X = x_i$. În fiecare din aceste repartiții se va putea calcula — în cazul variabilelor cantitative — o medie numită *medie condiționată* și o varianță numită *varianță condiționată*. Astfel :

$$\hat{x}_j = \frac{\sum_{i=1}^{p} x_i f_{ij}}{f_{\cdot j}}$$

$$s^2(X) = \frac{\sum_{i=1}^{p} (\bar{x}_i - \bar{x}_j)^2 f_{ij}}{f_{\cdot j}}$$

și

$$\hat{y}_i = \frac{\sum_{j=1}^{q} y_j f_{ij}}{f_{i\cdot}}$$

$$s_i^2(Y) = \frac{\sum_{j=1}^{q} (y_j - \bar{y}_i)^2 f_{ij}}{f_{i\cdot}}$$

Exemplu : **R.c.** a copiilor de 10 ani de sex masculin după greutate legată de înălțime este constituită de rîndurile tab. 30. Caracteristicile condiționate corespunzătoare sînt prezentate în tab. 32.

Tabelul 32

Clasa de înălțime (cm)	Greutatea medie (kg) $(\hat{y}_i)$	Varianța și abaterea tip a greutății	
		$s^2_{i(Y)}$	$s_{i(Y)}$
126	25,67	0,72	0,85
127	26,11	1,49	1,22
128	26,62	2,09	1,45
129	28,14	1,73	1,32
130	28,43	0,96	0,98
131	29,32	1,50	1,22
132	29,54	2,67	1,63
133	30,63	1,80	1,34
134	31,67	1,23	1,11
135	31,88	1,04	1,02
136	33,50	0,92	0,96
137	34,00	0,50	0,71

Seria de medii condiționate (col. 2) o găsim în fig. 69.

Așadar, în cazul relațiilor care interesează două variabile, mediile condiționate, respectiv variantele condiționate ale uneia din variabile (în exempul nostru y) sînt valorile calculate plecînd de la valorile acestei variabile înregistrate pe unitățile pentru care valoarea celeilalte variabile (x) este situată într-un anumit interval (sau corespunde anumitei valori a lui x, dacă această variabilă este discretă).

COVARIANȚĂ. Măsură a variabilității legate a două variabile x și y în jurul mediilor lor, notată obișnuit prin simbolul cov (x, y) sau simplu *cov* :

$$\text{cov}\ (x, y) = \frac{1}{n} \sum_{i=1}^{n} (x - \bar{x})(y - \bar{y}),$$

respectiv

$$\text{cov}\ (x, y) = \frac{1}{f_{..}} \sum_{i} \sum_{j} f_{ij}(x_i - \bar{x})(y_j - \bar{y}),$$

f_{ij} fiind frecvența unităților pentru care se are simultan $X = x_i$, $Y = y_j$ într-o populație în volum total $f_{..}$

C este pozitivă sau negativă după cum relația între cele două serii de date este directă sau inversă. **C.** este nulă dacă două variabile sînt independente. Această mărime intervine în studiul legăturii între două variabile, mai ales în aceea a corelației liniare.

Pentru simplificarea calculului **c.**, se utilizează metode analoage celor întrebuințate la calculul *variației.**

COEFICIENT DE CORELAȚIE (*coeficient de corelație liniară, coeficient de corelație Bravais—Pearson*). Număr care măsoară gradul de legătură între variabile, avînd drept expresie (cazul a două variabile) :

$$r_{yx} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{nS_xS_y} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}} \quad (1)$$

Cei doi termeni ai acestei fracții se calculează ușor folosind formulele lui Koenig :

$$\Sigma(x-\bar{x})^2 = \Sigma x^2 - N\bar{x}^2 = \Sigma x^2 - \frac{1}{n}(\Sigma x)^2$$

$$\Sigma(y-\bar{y})^2 = \Sigma y^2 - N\bar{y}^2 = \Sigma y^2 - \frac{1}{n}(\Sigma y)^2$$

$$\Sigma(x-\bar{x})(y-\bar{y}) = \Sigma xy - N\bar{x}\bar{y} = \Sigma xy - \frac{1}{n}\Sigma x\,\Sigma y,$$

în care n este numărul total de observații cuplate.
Formula (1) prin dezvoltare devine

$$r_{yx} = \frac{n\Sigma xy - \Sigma x\,\Sigma y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2]\,[n\Sigma y^2 - (\Sigma y)^2]}} \quad (2)$$

sau

$$r = r_{xy} = r_{yx} = \frac{\operatorname{cov}(x, y)}{s_x s_y}$$

mai general :

$$r = r_{xy} = r_{yx} = \frac{\sum_{i=1}^{p}\sum_{i=j}^{q} f_{ij}(x_i-\bar{x})(y_j-\bar{y})}{\sqrt{\sum_{i=1}^{p} f_{i\cdot}(x_i-\bar{x})^2}\sqrt{\sum_{j=1}^{q} f_{\cdot j}(y_j-\bar{y})^2}}$$

C. c. fără altă specificare este folosit pentru a desemna coeficientul de corelație liniară.

TRANFORMATA LUI FISHER. Variabila Z' definită ca

$$Z' = \operatorname{arc\,tg} r = \frac{1}{2}\ln\frac{1+r}{1-r} = r + \frac{r^3}{3} + \frac{r^5}{5} + \dots$$

Dacă r este **c.c.** liniară calculat pe baza unui eşantion de n perechi (x, y) provenite dintr-o populaţie normală cu coeficientul de corelaţie ρ, atunci variabila Z' este repartizată aproximativ normal cu media arc tg ρ şi abaterea standard $\sigma_{z'} = \frac{1}{\sqrt{n-3}}$. Variabila liniară Z' se utilizează în statistica testelor relative la **c.c.** liniară.

C.c. este un număr abstract, independent de unităţile de exprimare a datelor originale şi deci comparabil cu oricare altul.

De fapt, **c.c.** s-ar putea socoti tot un *coeficient de regresie*, exprimat în abaterile normate pentru ambele variabile. Produsul mediu al abaterilor normate dă **c.c.**

$$\overline{t_x t_y} = r_{xy}, \tag{3}$$

în care

$$t_x = \frac{x - \bar{x}}{s_x} \text{ şi } t_y = \frac{y - \bar{y}}{s_y}. \tag{4}$$

Coeficientul de regresie a lui y în x şi **c.c.** sînt legaţi prin egalitatea

$$b_{y/x} = r_{yx} \frac{s_y}{s_x}, \tag{5}$$

de unde

$$r_{yx} = b_{y/x} \frac{s_x}{s_y}. \tag{6}$$

Într-adevăr, dacă scriem:

$$b_{y/x} = \frac{n\Sigma_{yx} - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}, \quad s_x^2 = \frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n} \text{ şi}$$

$$s_y^2 = \frac{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}{n},$$

atunci

$$\frac{n\Sigma xy - \Sigma xy}{n\Sigma x^2 - (\Sigma x)^2} \cdot \frac{\sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n}}}{\sqrt{\frac{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}{n}}} = \frac{n\Sigma xy - \Sigma xy}{n\Sigma x^2 - (\Sigma x)^2} \times$$

$$\times \frac{\sqrt{n\Sigma x^2 - (\Sigma x)^2}}{\sqrt{n\Sigma y^2 - (\Sigma y)^2}} = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2]\,[n\Sigma y^2 - (\Sigma y)^2]}} = r_{yx}.$$

În relația (6), c.c. apare drept un *coeficient de regresie standardizat* *. Ecuația de estimare scrisă sub forma :

$$y = \bar{y} = r \frac{s_y}{s_x} x - \bar{x}$$

arată că în sens geometric r este panta *dreptei de regresie* * (estimare) în unități ale abaterii standard.
O definiție alternativă a c.c. : r este rădăcina pătrată din produsul coeficienților de regresie :

$$b_{yx} = r \frac{s_y}{s_x} \; ; \; b_{xy} = r \frac{s_x}{s_y},$$

respectiv coeficienții de regresie ai lui y în x și ai lui x în y.
Semnificația c.c. se desprinde cu claritate dintr-o altă formulare a sa :

$$r_{yx} = \sqrt{\frac{\Sigma(y^* - \bar{y})^2}{\Sigma (y - \bar{y})^2}},$$

adică, r_{yx} nu este altceva decît [illegible] variabilei dependente, care poate fi explicată de ecuația regresiei liniare simple.
În calculele de corelație, *variația totală* $\Sigma(y - \bar{y})^2$ se descompune în *variația explicată* $\Sigma (y^* - \bar{y})^2$ și *variația neexplicată sau reziduală* $\Sigma (y - y^*)^2$, în care y reprezintă valorile empirice ale variabilei dependente, iar y^*, valorile calculate ale aceleași variabile. Între cele trei variabile există relația

$$\Sigma (y^* - \bar{y})^2 + \Sigma(y - y^*)^2 = \Sigma (y - \bar{y})^2.$$

C.c. r este cuprins între -1 și $+1$:

$$-1 < r < +1.$$

El este pozitiv ($r > 0$), dacă ambele variabile x și y variază în același sens și este negativ ($r < 0$), în caz contrar. Cînd valoarea absolută a c.c. este apropiată de unitate, se poate afirma că există o puternică dependență liniară între cele două variabile x și y. Dacă valoarea absolută a c.c. este slabă, nu trebuie în mod automat să se conchidă că nu există legătură între x și y ; legătura poate să existe, dar nu de formă liniară.
S-ar putea imagina, de exemplu, o relație perfectă între x și y cînd

$$y^2 = a^2 - x^2.$$

O serie de puncte determinate după această ecuație se vor plasa pe un cerc cu raza a. Dacă totuși s-ar reprezenta aceste puncte prin linie dreaptă, s-ar ajunge la o linie orizontală de regresie și la un coeficient de corelație egal exact cu zero. A conchide de aici că nu există nici o corelație ar fi absolut greșit. Trebuie reamintit postulatul lui Pearson : [illegible] în ce măsură linia dreaptă descrie relația între variabile. De asemenea, *testul de liniaritate* * se impune cînd c.c. este în vecinătatea lui zero.
În fond, dacă valoarea lui r între două variabile este zero, aceasta înseamnă că ori nu există vreo corelație între ele, ori că această corelație nu poate fi descrisă de o linie dreaptă.

V.TREBICI

Dependența constatată sau neconstatată constituie un fapt de natură pur statistică. Legătura (x_i, y_i) este o legătură între structurile de unități abstracte în mulțimile matematice. Dacă dependența poate fi prezumată plecînd de la rezultatul statistic, ea rămîne de stabilit pe planul cauzalității științifice.

În afară de cazul excepțional în care seria de cupluri (x_i, y_i) este exhaustivă, ne aflăm în prezența unui eșantion de n cupluri dintr-o populație N. Alte eșantioane de n cupluri din populația respectivă ar fi condus la c. c. r_1, r_2, r_3 ... etc. În aceste condiții c.c. r determinat pentru un eșantion n-are nici un sens prin el însuși. El permite numai să se afirme, cu un risc de eroare ales dinainte, că c.c. relativ la populația studiată aparține unui interval mergînd de la $r - \alpha$ la $r + \alpha$, α fiind cu atît mai slab (deci precizia mai mare) cu cît efectivul n al eșantionului este mai mare. Așadar, c.c. măsoară, cu un prag de probabilitate dat, corelația care există între variabilele cercetate.

Exemplul 1. (→ tab. 33.)

Cu elementele de calcul date în tabelul 33, se poate determina c.c. după formulele (1) și (2). Astfel, formula (1):

$$r_{yx} = \frac{5956{,}600}{\sqrt{32114{,}5}\ \sqrt{1119{,}2040}} = 0{,}993;$$

formula (2):

$$r_{yx} = \frac{(10 \times 87067) - (1785 \times 454{,}4)}{\sqrt{[(10 \times 350737) - (1785)^2]\ [(10 \times 21767{,}14) - (454{,}4)^2]}} = 0{,}993$$

Rezolvînd ecuațiile de regresie:

$$10a + 1785\ b = 454{,}4$$

$$1785\ a + 350737\ b = 87067$$

se obține:

$$y^* = 12{,}3312 + 0{,}18548\ x;$$

calculul se poate efectua și după formula:

$$r_{yx} = \sqrt{\frac{a \Sigma y + b \Sigma xy - \frac{(\Sigma y)^2}{n}}{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}} =$$

$$= \sqrt{\frac{(12{,}3318 \times 454{,}4) + (0{,}18548 \times 87067) - \frac{(454{,}4)^2}{10}}{21767{,}14 - \frac{(454.4)^2}{10}}} = 0{,}993.$$

Exemplul 2. C.c. între greutatea (y) și înălțimea (x) a copiilor de 10 ani de sex masculin, aplicînd formula (2), conform datelor tab. de corelație 30:

$$r_{yx} = \frac{(425 \times 1629961) - (55698 \times 12429)}{\sqrt{[(425 \times 730920) - (55698)^2]\ [(425 \times 364907) - (12429)^2]}} = 0{,}753.$$

V.TREBICI

Tabelul 33

Corelația dintre consumul de energie electrică (y) și indicele producției industriale (x) în economia României, 1970—1979.

Anii	x (%)	y (mlrd kw/h)	$x - \bar{x}$	$(x - \bar{x})^2$	$y - \bar{y}$	$(y - \bar{y})^2$	$(x - \bar{x}) \cdot (y - \bar{y})$	x^2	y^2	xy
0	1	2	3	4	5	6	7	8	9	10
1970	100	29,4	−78,5	6162,25	−16,04	257,2816	1259,140	10000	864,26	2940,0
1971	112	32,7	−66,5	4122,25	−12,74	162,3076	847,210	12544	1069,29	3662,4
1972	125	35,9	−53,5	2862,25	− 9,54	91,0116	510,390	15625	1288,81	4487,5
1973	143	39,2	−35,5	1260,25	− 6,24	38,9376	221,520	20449	1536,64	5605,6
1974	164	42,5	−14,5	210,25	− 2,91	8,6436	42,630	26896	1806,25	6970,0
1975	184	47,2	+5,5	30,25	+ 1,76	3,0976	9,680	33856	2227,84	8684,8
1976	205	52,7	+26,5	702,25	+ 7,26	52,7076	192,390	42025	2777,29	10803,5
1977	230	55,6	+51,5	2652,25	+10,16	103,2256	523,240	52900	3091,36	12788,0
1978	251	58,9	+72,5	5256,25	+13,46	181,1716	975,850	63001	3469,21	14783,9
1979	271	60,3	+92,5	8556,25	+14,86	220,8196	1374,550	73441	3636,09	16341,3
Total	1785	454,4	—	32114,50	—	1119,2040	5956,600	350737	21767,14	87067,0

Sursa: *Anuarul statistic al R. S. România 1976*, p. 122—123 și *1980*, p. 194—195.

Volumul operațiilor de calcul poate fi redus, dacă se consideră valorile micșorate ale variabilelor corelate (tab. 31):

$$r_{yx} = \frac{(425 \times 1095) - (23 \times 104)}{\sqrt{[(425 \times 1469) - (23)^2]\,[(425 \times 1450) - (104)^2]}} = 0{,}75.$$

(→ *Tabel de corelație*).

Tabelul lui Pearson arată că, dacă numai o sută de măsuri cuplate erau sub efectul hazardului, probabilitatea pentru valoarea absolută a lui r de a depăși 0,25 ar fi inferioară lui 0,01 (în aplicație, volumul eșantionului = 425). Dacă, deci, se infirmă că rezultatul obținut $r = 0{,}75$ nu se datorează hazardului, ci corelației statistice care există între mărimile observate, se riscă să se comită o eroare : probabilitatea acestei erori este cu mult inferioară lui 0,01. Se consideră deci că între cele două variabile există o corelație pozitivă (variațiile în același sens) sigură sau că valoarea $r = 0{,}75$ este foarte semnificativă la pragul de probabilitatea 0,01.

Postulatele corelației lui Pearson. Postulate subiacente aplicării c.c. imaginat de Pearson.

Principalele postulate sînt : *rectiliniaritatea* (liniaritatea) relației între variabilele cercetate și *omoscedasticitatea*. Satisfacerea primului postulat poate fi apreciată mai întîi prin simpla examinare a *liniei de regresie empirice* * : o linie aproximativ dreaptă. Micile divergențe cu rectiliniaritatea sînt adesea de puțină importanță. Dacă, dimpotrivă, se observă că linia de regresie are o tendință de a se curba într-un fel sau altul, se recomandă să se renunțe la c.c. în favoarea *raportului de corelație*, care este de întrebuințare mai generală și mai exact. Un al doilea mod de a examina liniaritatea se realizează cu ajutorul unui test precis — [illegible] *liniarității* *, al cărui principiu general constă în a judeca importanța abaterii între c.c. și raportul de corelație. Dacă această abatere se dovedește considerabilă, înseamnă că relația nu este liniară. Un al doilea postulat pe care se întemeiază folosirea c.c. — omoscedasticitatea — presupune ca dispersiile condiționate (în sînul coloanelor, respectiv în sînul rîndurilor, → *Tabel de corelație*) să fie aproximativ egale. În fond, această condiție este în mod satisfăcător îndeplinită cînd cele două repartiții sînt simetrice.

RAPORT DE CORELAȚIE (*Indice de corelație*). Număr care măsoară, sub aspectul intensității, corelația curbilinie, definit de formula (cazul a două variabile) :

$$\eta_{yx} = \sqrt{1 - \frac{\Sigma (y - y_x^*)^2}{\Sigma (y - \bar{y})^2}}, \quad \text{respectiv} \quad \sqrt{1 - \frac{\sum_i \sum_j f_{ij}(y_i - \bar{y})^2}{\sum_j f_{\cdot j}(y_j - \bar{y})^2}} \qquad (1)$$

sau, ceea ce este același lucru

$$\eta_{yx} = \sqrt{\frac{\Sigma (y_x^* - \bar{y})^2}{\Sigma (y - \bar{y})^2}}. \qquad (2)$$

R.c. joacă, în cazul corelației neliniare, un rol analog cu acel al coeficientului de corelație în cazul corelației liniare. El are la bază descompunerea dispersiei totale a variabilei dependente în dispersia valorilor empirice față de valorile

teoretice (estimate printr-o ecuație de regresie neliniară) și dispersia valorilor teoretice față de medie. Compararea dispersiei valorilor teoretice față de medie cu dispersia valorilor empirice de la medie dă **r.c.** Acest parametru este totdeauna cuprins între 0 și 1. Cantitatea η^2_{yx} fiind egală cu citul a două valori non-negative nu poate fi într-adevăr inferioară lui zero. Ea nu poate fi dealtminteri nulă decît dacă toate mediile condiționate sînt egale între ele, presupunînd, desigur cazul obișnuit, că există mai multe observații ale lui y pentru anumite valori ale lui x, ca în exemplul din tab 34., în care se găsesc datele necesare pentru rezolvarea sistemului de ecuații normale :

$$50\ a + 5{,}5203\ b = 241{,}2$$

$$5{,}5203\ a + 0{,}7265\ b = 28{,}3714.$$

Ecuația de regresie **hiperbolică este** :

$$y^* = 3{,}181 + 14{,}881 \cdot \frac{1}{x}.$$

Valorile y^* **sînt trecute în tabelul** 35, unde s-a calculat

$$\Sigma\,(y - y^*)^2 f = 7{,}8330.$$

Calculînd

$\Sigma\,(y - \bar{y})^2 f$ **cu ajutorul datelor tab.** 34.

$$1197{,}28 - \frac{(241{,}2)^2}{50} = 33{,}7312,$$

atunci, **conform formulei 1,**

$$\eta_{yx} = \sqrt{1 - \frac{7{,}8330}{33{,}7312}} = 0{,}876$$

sau, scriind

$\Sigma(y - \bar{y})^2 f - \Sigma\,(y - y^*)^2 f = 33{,}7312 - 7{,}8330 = 25{,}8982$, conform formulei (2)

$$\eta_{yx} = \sqrt{\frac{25{,}8982}{33{,}7312}} = 0{,}876.$$

Deoarece **s-a rezolvat ecuația de regresie,** calculul **r.c.** se poate efectua și după formula :

$$\eta_{yx} = \sqrt{\frac{a\Sigma y f_y + b\Sigma \frac{1}{x} y f_{yx} - \frac{(\Sigma y f_y)^2}{\Sigma f}}{\Sigma\ y^2 f_y - \frac{(\Sigma y f_y)^2}{\Sigma f}}}.$$

Tabelul 34

Corelația dintre nivelul relativ al cheltuielilor de circulație și volumul desfacerii mărfurilor

x \ y	3,4—3,8	3,8—4,2	4,2—4,6	4,6—5,0	5,0—5,4	5,4—5,8	5,8—6,2	6,2—6,6	f_x	$\frac{1}{x} f_x$	$\frac{1}{x^2} f_x$	yf_{xy}	$\Sigma \frac{1}{x} yf_{xy}$
18—20	3	1							4	0,2105	0,0111	14,8	0,7789
16—18	2	2							4	0,2353	0,0138	15,2	1,8941
14—16		3	2						5	0,3333	0,0222	20,8	1,3867
12—14		1	2	3					6	0,4615	0,0355	27,2	2,0923
10—12		1	4	3					8	0,7273	0,0661	36,0	3,2727
8—10			1	1	2	1	1		6	0,6667	0,0741	31,2	3,4667
6—8				3	3	2		1	9	1,2857	0,1837	47,6	6,8000
4—6						2	3	3	8	1,6000	0,3200	48,4	9,6800
f_y	5	8	9	10	5	5	4	4	50	5,5203	0,7265	241,2	28,3714
yf_y	18	32	39,6	48	26	28	24	25,6	241,2				
$y^2 f_y$	64,8	128	174,24	230,4	135,2	156,8	144	163,84	1197,28				
$\frac{1}{x} f_{xy}$	0,2755	0,5381	0,7619	1,0432	0,6508	0,7968	0,7111	0,7429	5,5203				
$\frac{1}{x^2} f_{xy}$	0,0152	0,0372	0,0661	0,1162	0,0859	0,1332	0,1323	0,1404	0,7265				
$\Sigma \frac{1}{x} yf_{xy}$	0,9918	2,1524	3,3523	5,0074	3,3842	4,4621	4,2666	4,7546	28,3714				

y reprezintă nivelul relativ al cheltuielilor de circulație (%);

Tabelul 35

y^* \ y	3,60	4,00	4,40	4,80	5,20	5,60	6,00	6,40	$\Sigma(y-y^*)^2 f$
3,96	0,2196 3 0,3888	0,0016 1 0,0016							0,3904
4,06	0,2116 2 0,4232	0,0036 2 0,0072							0,4304
4,17		0,0289 3 0,0867	0,0529 2 0,1058						0,1925
4,33		0,1089 1 0,1089	0,0049 2 0,0098	0,2209 3 0,6627					0,7814
4,53		0,2809 1 0,2809	0,0169 4 0,0676	0,0729 3 0,2187					0,5672
4,83			0,1849 1 0,1849	0,0009 1 0,0009	0,1369 2 0,2738	0,5929 1 0,5929	1,3689 1 1,3689		2,4214
5,31				0,2601 3 0,7803	0,0121 3 0,0363	0,0841 2 0,1682		1,881 1 1,1881	2,1729
6,16						0,3136 2 0,6272	0,256 3 0,0768	0,0576 3 0,1728	0,8768
$\Sigma(y-y^*)^2 f$	0,8120	0,4853	0,3681	1,6626	0,3101	1,3883	1,4457	1,3609	7,8330

cu datele calculate

$$\eta_{yx} = \sqrt{\frac{(3{,}181 \times 241{,}2) + 14{,}881 \times 28{,}3714) - \frac{(241{,}2)^2}{50}}{1197{,}28 - \frac{(241{,}2)^2}{50^2}}} = 0{,}876.$$

Notă. K. Pearson a propus r.c. care nu implică nici o ipoteză *apriori* asupra formei legăturii :

$$\eta_{yx} = \sqrt{1 - \frac{C^2_{yx}}{s^2_y}},$$

în care C^2_{yx} reprezintă media pătratelor abaterilor de la curba de regresie y în x (curba descrisă de mediile condiționale $\hat{y}_x$). Semnificația lui η_{yx} apare mai clară din calculul său sub formă transformată :

$$\eta_{yx} = \sqrt{\frac{\Sigma\,(\hat{y}_x - y)^2 f_i}{\Sigma\,(y - \bar{y})^2 f_j}}$$

(→ aplicația din vol. A. Piatier, *Statistique et observation économique*, t. I, p. 380—384).

COEFICIENT DE CORELAȚIE BISERIAL. Coeficient de corelație elaborat de K. Pearson pentru a măsura corelația între două caracteristici, dintre care una este o caracteristică măsurabilă (cantitativă) și alta o caracteristică calitativă (atributivă) cu două clase de valori posibile. Poate fi utilizat ca un estimator al coeficientului corelației liniare în cazul repartiției normale bidimensionale.

COEFICIENT DE CORELAȚIE TETRACHORIC. Estimator al coeficientului corelației liniare între două variabile, fiecare având cîte două clase sau categorii distincte în cazul repartiției normale bidimensionale (sau repartiții marginale normale).

CORELAȚIE MULTIPLĂ. Corelație între o variabilă dependentă și mai multe variabile independente.
Modelul are forma :

$$y = f(x_1, x_2, \ldots, x_n) + \varepsilon. \qquad (1)$$

Dacă se consideră c.m. liniară, ecuația de regresie se scrie :

$$y^* = a + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n. \qquad (2)$$

În sensul celor mai mici pătrate, coeficienții $a, b_1, \ldots, b_n$ trebuie să asigure, pentru o serie statistică, minimumul sumei :

$$\Sigma(y - y^*) = \Sigma\,[y - (a + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n)]^2. \qquad (3)$$

Anulînd cele n derivate parțiale ale acestei sume în raport cu $a, b_1, b_2, \ldots, b_n$, se obțin ecuațiile normale :

$$na + b_1 \Sigma x_1 + b_2 \Sigma x_2 + \ldots + b_n \Sigma x_n = \Sigma y$$

$$a \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2 + \ldots + b_n \Sigma x_1 x_n = \Sigma y x_1$$

$$a \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2 + \ldots + b_n \Sigma x_2 x_n = \Sigma y x_2 \qquad (4)$$

. . . . . .

. . . . . .

$$a \Sigma x_n + b_1 \Sigma x_1 x_n + b_2 \Sigma x_2 x_n + \ldots + b_n \Sigma x_n^2 = \Sigma y x_n.$$

Exemplu. O aplicație în care intervin trei variabile : rata de natalitate (y), rata generală de fertilitate (x_1) și ponderea populației feminine în populația totală (x_2).

Tabelul 36

Anii	Rata de natalitate (‰) y	Rata generală de fertilitate (‰) x_1	Ponderea populației feminine în populația totală (%) x_2
1971	19,5	75,3	25,9
1972	18,8	72,7	25,9
1973	18,2	70,4	25,8
1974	20,3	79,3	25,6
1975	19,7	77,5	25,4
1976	19,5	77,3	25,1
1977	19,6	78,8	24,8
1978	19,1	77,7	24,5
1979	18,6	76,6	24,3
1980	18,0	74,8	24,0

Sursa : Anuarul statistic al R. S. România 1981.

Calculele au condus la ecuațiile normale :

I. $10\ a + 760{,}4\ b_1 + 251{,}3\ b_2 = 191{,}30$

II. $760{,}4\ a + 57890{,}9\ b_1 + 19102{,}8\ b_2 = 14559{,}99$

III. $251{,}3\ a + 19102{,}8\ b_1 + 6319{,}57\ b_2 = 4809{,}05$

și ecuația de regresie

$$-18{,}8270 + 0{,}2564\ x_1 + 0{,}7346\ x_2.$$

Ca și în cazul *regresiei simple* *, coeficienții regresiei multiple arată cu cît se schimbă variabila explicată atunci cînd variabila explicativă respectivă se schimbă cu o unitate. Mărimea fiecărui coeficient al regresiei este o mărime „netă", curățată de influențele exercitate în mod indirect de celelalte variabile explicative specificate în model.

Principiile regresiei simple se aplică și în cazul regresiei multiple, dar calculele sînt mai laborioase.

Pentru a măsura gradul de legătură între o variabilă explicată y și mai multe variabile explicative x_1, x_2, ..., x_n luate împreună, se utilizează *raportul c.m.* (R), dat de formula :

$$R_{y \cdot x_1 x_2 \ldots x_n} = \sqrt{1 - \frac{\Sigma (y - y^*_{x_1 x_2 \ldots x_n})^2}{\Sigma (y - \bar{y})^2}}. \qquad (5)$$

Valoarea lui R depinde deci de raportul dintre dispersia valorilor determinate pe baza ecuației de regresie liniară și dispersia valorilor observate ale variabilei dependente. Cu cît valorile observate se abat mai puțin de la valorile calculate cu atît coeficientul de corelație are o valoare mai mare și ca atare corelația este mai intensă.

Dacă s-a stabilit ecuația de regresie, coeficientul de corelație se calculează cu ușurință după formula :

$$R_{y x_1 x_2 \ldots x_n} = \sqrt{\frac{a\Sigma y + b_1 \Sigma x_1 y + b_2 \Sigma x_2 y + \ldots + b_n \Sigma x_n y - \frac{(\Sigma y)^2}{n}}{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}}. \qquad (6)$$

Coeficientul **c.m.** măsoară influența simultană a variabilelor independente $x_1 x_2, \ldots, x_n$ asupra variabilei dependente y.

Exemplu. Dacă se estimează, conform ecuației de regresie multiplă (5), valorile $y^*_{x_1 x_2}$ — rata de natalitate și se aplică formula de calcul (5) al lui R, rezultă :

$$R_{y \cdot x_1 x_2} = \sqrt{1 - \frac{0{,}0017}{0{,}7242}} = 0{,}998.$$

Mărimea ratei de natalitate se află în dependență foarte strînsă față de rata generală de fertilitate (x_1) și de ponderea populației feminine în populația totală (x_2).

CORELAȚIE A RANGURILOR. Corelație între numerele care indică pozițiile în serii ale unităților după variabilele observate.

Cînd variabilele nu sînt susceptibile de măsurare sau cînd nu pot fi măsurate cu precizia necesară, dar se pretează la o clasificare prin ranguri, atunci acestea servesc ca bază pentru a stabili o eventuală corelație. Rangul ocupat de fiecare unitate se interpretează ca indicativul unei măsuri și se utilizează ca atare pentru calculul unui coeficient de corelație.

COEFICIENTUL LUI SPEARMAN. Coeficient de corelație liniară între rangurile diverselor unități în cele două clasificări, a cărui valoare este :

$$S = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)},$$

în care :

S este coeficientul corelației rangurilor (propus de Spearman) ;

d — diferența între ranguri pentru fiecare pereche de observații ;

n — numărul perechilor de observații.

V.TREBICI

Dacă se notează rangurile cu simbolurile folosite de obicei pentru valorile caracteristicilor corelate, formula de calcul (1) se deduce ușor din expresia :

$$S = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\sqrt{\Sigma(x - \bar{x})^2}\ \sqrt{\Sigma(y - \bar{y})^2}}.$$

Într-adevăr, cele două serii x și y, fiind constituite din primii n întregi, $1, 2, \ldots, n$, avem :

$$\Sigma x = \Sigma y = \frac{n(n+1)}{2} \quad \text{și} \quad \bar{x} = \bar{y} = \frac{n+1}{2}$$

apoi :

$$\Sigma(x - \bar{x})^2 = \Sigma(y - \bar{y})^2 = \sum_{i=1}^{n} \left(i - \frac{n+1}{2} \right)^2 = \sum_{i}^{n} i^2 - n \frac{(n+1)^2}{4} =$$

$$= \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4} = \frac{n(n+1)(n-1)}{12}.$$

Calculul lui $\Sigma(x - \bar{x})(y - \bar{y})$ se face comod prin relația :

$$2(x - \bar{x})(y - \bar{y}) = (x - \bar{x})^2 + (y - \bar{y})^2 - [(x - \bar{x}) - (y + \bar{y})]^2,$$

adică punînd $(x - \bar{x}) - (y - \bar{y}) = x - y = d$

$$2\Sigma(x - \bar{x})(y - \bar{y}) = \Sigma(x - \bar{x})^2 + \Sigma(y - \bar{y})^2 - \Sigma d^2 =$$

$$= \frac{n(n+1)(n-1)}{6} - \Sigma d^2,$$

de unde rezultatul :

$$S = \frac{\dfrac{n(n+1)(n-1)}{12} - \dfrac{1}{2}\Sigma d^2}{\dfrac{n(n+1)(n-1)}{12}},$$

adică :

$$S = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}.$$

Pentru două serii de ranguri x și y identice, toate diferențele $d = x - y$ sînt nule și $S = 1$; este ușor de verificat că dacă rangurile sînt exact inverse, atunci $S = -1$.

Între aceste două extreme, S este totdeauna cuprins între -1 și $+1$.

V. TREBICI

Coeficientul corelației rangurilor posedă deci proprietăți analoge celor ale coeficientului de corelație între variabile măsurabile.

Exemplu. Un număr de 10 orașe mari din lume se clasifică — să presupunem — după calitatea aerului (x) și îmbolnăvirile pulmonare (y), conform datelor tab. 37.

Tabelul 37

Orașul	Rangul orașului după variabila		d	d^2
	x	y		
A	3	5	−2	4
B	7	4	+3	9
C	9	7	+2	4
D	1	3	−2	4
E	2	1	+1	1
F	10	9	+1	1
G	4	2	+2	4
H	5	8	−3	9
I	6	10	−4	16
J	8	6	+2	4
Total	—	—	—	56

$$S = 1 - \frac{6 \times 56}{10(10^2 - 1)} = 1 - 0{,}339 = 0{,}661.$$

Un coeficient de corelație de 0,661 sugerează o legătură substanțială pozitivă între cele două variabile, cel puțin în orașele considerate.

Thornton *) a publicat un tabel pe care-l socotește drept cea mai bună bază pentru a aprecia valoarea probantă a lui S cînd numărul de cazuri (N) variază între 2 și 30 (→ tabelul 37 bis). $S = 0{,}661$ este semnificativ la nivelul de probabilitate 5% ($N = 10$, S va trebui să atingă sau să depășească mărimea lui 0,644).

Valorile indicate în tabelul Thornton au totuși o valoare limitată : ele au fost calculate în ipoteza că seriile nu cuprind cazuri *ex aequo*.

Observație. Cazuri de *ex aequo*. De exemplu, dacă orașele G, H, I ar fi obținut în prima serie același rang, 4, atunci fiecare din ele ar fi primit rangul 5, adică media aritmetică a rangurilor ocupate în fapt $(4 + 5 + 6) : 3 = 5$, Dacă, de asemenea, D și E ar fi obținut același rang, 1, ar fi fost clasate la rangul $(1 + 2) : 2 = 1{,}5$.

Coeficientul lui Kendall. Coeficient bazat pe toate diferențele rangurilor între cele două clasificări, definit de relația :

$$K = \frac{2T}{n(n-1)}. \qquad (2)$$

*) G. R. Thornton, *The Significance of Rank Difference Coefficients of Corelation*, in „Psychometriks" 8/1943, p. 212.

Tabelul 37 bis

Coeficienții corelației rangurilor, care sînt semnificative la nivelurile 1% și 5% de probabilitate, după *Thornton*

N	0,01	0,05	N	0,01	0,05	N	0,01	0,05
			11	0,764	0,626	21	0,571	0,450
2	—	—	12	0,736	0,599	22	0,558	0,440
3	—	—	13	0,711	0,575	23	0,547	0,430
4	—	—	14	0,687	0,553	24	0,536	0,420
5	—	0,933	15	0,667	0,535	25	0,525	0,412
6	0,960	0,843	16	0'648	0,517	26	0,516	0,404
7	0,906	0,781	17	0,630	0,501	27	0,506	0,396
8	0,869	0,725	18	0,614	0,487	28	0,497	0,389
9	0,825	0,681	19	0,598	0,474	29	0,488	0,382
10	0,788	0,644	20	0,583	0,462	30	0,479	0,375

Se poate observa că dacă clasificarea relativă la (y) ar fi aceeași cu cea care rezultă din clasificarea relativă la variabila (x), valorile rangurilor y s-ar succeda în ordinea naturală a n numere prime întregi. Orice modificare a acestei ordini se traduce deci prin faptul că există cel puțin un rang (y_j) mai mare decît unul (y_i) din cele care-l preced. Se caută cuplurile care prezintă această particularitate, asociindu-se lui y_1, succesiv, $y_2, y_3, \ldots, y_n$, apoi lui y_2, succesiv, $y_3, y_4, \ldots, y_n$ etc. pînă la y_{n-1} și y_n.

Cuplurile pentru care $y_j > y_i$ se notează prin valoarea $+ 1$, iar cuplurile pentru care $y_j < y_i$ cu $j > i$, prin $- 1$. Suma algebrică a valorilor astfel obținute se desemnează prin T sau S.
Numărul de comparații făcute fiind

$$(n - 1) + (n - 2) + (n - 3) + \ldots + 2 + 1 = \frac{n(n - 1)}{2},$$

raportul

$$K = \frac{T}{\frac{n(n - 1)}{2}} = \frac{2T}{n(n - 1)}$$

poate fi socotit ca un indicator de corelație.
Pentru clasificări identice, diferențele sînt toate egale cu $+ 1$ și numărul lor este $\frac{n(n - 1)}{2}$. Atunci $K = 1$. Pentru clasificări exact inverse, diferențele

sînt toate egale cu -1 și numărul lor este $\frac{n(n-1)}{2}$; atunci $K = 1$. În afară de aceste cazuri de legătură perfectă, coeficientul K rămîne cuprins între -1 și $+1$.

Exemplu. Reluînd datele din tab. 37, orașele sînt prezentate în tab. 38 în ordinea crescătoare a rangului după variabila X. T se determină după poziția rangurilor lui y (col. 2) în direcția corespunzătoare ordinei primei serii (col. 1).

Astfel, orașul D, care ocupă primul loc în raport cu variabila x, se situează pe locul al treilea în raport cu variabila y, avînd în ordine succesivă orașul E, de un rang inferior (-1), orașul A de un rang superior ($+1$), orașul G de un rang inferior (-1) ș.a.m.d. ($D \ldots -2+7$). $T = -11 + 34 = 23$. Înlocuind în formula (2) obținem

Tabelul 38

Orașul	Rangul orașului după variabile		Diferențe	
	x	y	$-$	$+$
0	1	2	3	4
D	1	3	2	7
E	2	1	0	8
A	3	5	2	5
G	4	2	0	6
H	5	8	3	2
I	6	10	4	0
B	7	4	0	3
J	8	6	0	2
C	9	7	0	1
F	10	9	0	0
Total :	—	—	11	34

$$K = \frac{2 \times 23}{10(10-1)} = \frac{46}{90} = 0{,}511.$$

PARTEA A PATRA

INDICII ȘI SERIILE CRONOLOGICE

Această secțiune este consacrată prezentării unor concepte statistice care descriu variația în timp a unor fenomene. Atît indicii cît și seriile cronologice sînt elaborate ale statisticii economice.

În lucrările de statistică descriptivă aceste probleme formează partea finală, datorită caracterului lor particular. Indicii, care au apărut cu aproximativ două secole în urmă (→ *Istoric*), în legătură cu preocupările de a măsura puterea de cumpărare a monedei, reclamă un aparat matematic modest; progresele recente au adăugat tratării indicilor aplicații ale teoriei sondajului și ale calculului diferențial și integral în ce privește mai ales elasticitatea și alte aplicații econometrice. În schimb, seriile cronologice, elaborate de economiști și de statisticieni in legătură cu măsurarea și analiza ciclurilor economice, s-au bucurat de o dezvoltare matematică considerabilă, împrejurare ce trebuie pusă în legătură cu progresele mari ale prospectivei, ale diferitelor metode de prognoză, facilitate și ele de dezvoltarea metodelor de simulare electronică. Partea esențială a seriilor cronologice, cunoscută sub denumirea de „analiză a seriilor cronologice", a fost dezvoltată, în ultimele 3—4 decenii, sub raport matematic, adăugindu-i-se o serie de metode noi.

Ca urmare, în timp ce domeniul predilect al indicilor rămîne cel economic, cel al seriilor cronologice a depășit sensibil acest cadru, cunoscînd aplicații în cele mai variate domenii, în afara celui economic.

Comun însă ambelor capitole este caracterul aplicativ foarte larg al acestor noțiuni în domeniul vieții social-economice. Zeci de variante de indici, ca și de serii cronologice, pe perioade mai scurte sau mai lungi, se întîlnesc în toate publicațiile statistice internaționale și naționale. În cele ce urmează ne vom limita la prezentarea noțiunilor fundamentale cu recomandarea adresată cititorului de a consulta—ca și pentru celelalte capitole—lucrările de specialitate.

X. INDICII

INDICE. Număr relativ rezultat din compararea valorilor unui indicator statistic, fie la diferite momente de timp, fie în spații diferite, fie, în sfîrșit, în diferite categorii, constituite în raport cu o caracteristică oarecare. Mai general, **i.** oferă o măsură a schimbării relative a unor variabile sau a unei grupe de variabile în raport cu caracteristicile: timp, spațiu sau categorii specificate. Fie seria de valori în timp ale mărimii M, dispusă astfel: M_0, M_1, ..., M_t la datele succesive $0, 1, \ldots, t$.

V.TREBICI

În aceste condiții, i. are formula generală :

$$I_{t/0}(M) = \frac{M_t}{M_0},$$

în care : M_t — mărimea din *perioada curentă* *, iar M_0 — mărimea din *perioada de bază* *. I. măsoară variația relativă a mărimii M, fiind un număr pur (fără dimensiuni).

După practica încetățenită, i. se exprimă în procente ; atunci cînd însă valorile sînt mari i. se exprimă la unitate. De exemplu, indicele producției industriale din România în 1981 (baza 1950 = 100) a fost de 3400% sau, altfel spus, producția industrială a crescut de 34 ori. În raport cu noțiunile folosite în prezenta lucrare, i. poate fi un număr relativ pentru mărimile : x, f, xf, Σf, Σxf, $\bar{x}$ etc., ceea ce presupune necesitatea unei definiri riguroase a mărimilor pentru care urmează să calculăm un i., precum și a unei clasificări a i.

Din punctul de vedere al funcției, un i. poate fi :

a) număr relativ al valorilor unei mărimi în diferite perioade (sau momente) de timp ; aceasta este accepția cea mai generală ;

b) număr relativ al valorilor aceluiași indicator în diferite spații, purtînd termenul comun de i. *teritorial* (*regional*) * ;

c) număr relativ al valorilor aceluiași indicator existent în diferite subcolectivități (categorii sociale, de exemplu).

Acestor domenii clasice ale i., statistica din țările socialiste, cu economie planificată, le anexează o funcție nouă a i.

d) număr relativ al valorilor aceluiași indicator din punctul de vedere al planului și al realizării.

Cum problemele metodologice au particularități numai pentru i. sub raport temporal și i. sub raport teritorial, nu ne vom ocupa, în continuare, decît de aceste două funcțiuni ale i. I. pentru cele două accepții mai poartă denumirea de : i. cronologic și i. regional.

Exemplul 1. Valoarea fondurilor fixe din unitățile socialiste din România (la sfîrșitul anului) a fost de 561 774 milioane în 1970 și de 1 786 443 milioane în anul 1981.

$$I_{1981/1970} = \frac{1\,786\,443}{561\,774} \cdot 100 = 318\%.$$

Valoarea fondurilor fixe, în 1981, a fost de circa 3,2 ori mai mare ca în 1970.

Exemplul 2. În anul 1981 producția medie de lapte pe o oaie a fost, la nivelul țării, de 34 litri ; în același an producția medie de lapte pe o oaie, în județul Suceava, a fost de 45 litri :

$$I_{\text{Suceava/România}} = \frac{45}{34} \cdot 100 = 132{,}4\%.$$

Exemplul 3. Retribuția medie netă lunară a personalului muncitor, în 1981, a fost de 2340 lei, cea a muncitorilor a fost de 2253 lei.

$$I_{\text{muncitori/personal muncitor}} = \frac{2253}{2340} \cdot 100 = 96{,}3\%.$$

Exemplul 4. Rata brută de mortalitate a populației României în 1981 a fost de 10 decese la 1000 locuitori ; aceeași rată în județul Iași a fost de 7,7 decese la 1000 locuitori

$$I \text{ Iași/România} = \frac{7{,}7}{10} \cdot 100 = 77\%.$$

Exemplul 5. Producția globală a industriei din România a fost, în anul 1970, de 307,2 miliarde lei, iar în anul 1980, de 936,1 miliarde lei (datele pe anul 1970 sînt exprimate în prețuri 1963 ; cele din 1980 — în prețurile de la începutul anului).

$$I \ 1980/1970 = \frac{936{,}1}{307{,}2} \cdot 100 = 307{,}4\%.$$

Volumul producției industriale a crescut de aproximativ 3,1 ori.
Exemplele de mai sus arată funcțiile i. pentru timp, spațiu și categorii. Ele atrag însă atenția asupra caracterului indicatorilor luați în calcul. În exemplul 1, avem de a face cu o masă foarte variată de fonduri fixe (mașini, construcții speciale etc.), întrunită într-o singură expresie, cu ajutorul valorilor, fără de care nu s-ar fi putut face omogenizarea. În exemplul 2, i. exprimă raportul dintre două valori care au fost calculate ca valori medii aritmetice ponderate, după formula :

$$\bar{x} = \frac{\Sigma xf}{\Sigma f},$$

cu alte cuvinte :

$$I = \frac{\bar{x}_1}{\bar{x}_0}.$$

Aceeași semnificație au exemplele 3 și 4, numai că în fiecare caz media aritmetică ponderată s-a calculat diferit. Dacă luăm exemplul 4, rata brută de mortalitate la nivel național s-a calculat după formula :

$$\bar{m}_R = \frac{\sum_{x=0}^{100} m_x P_x}{\sum_{x=0}^{100} P_x} \quad \text{și} \quad \bar{m}(\text{Iași}) = \frac{\sum_{x=0}^{100} m_x P_x(\text{Iași})}{\sum_{x=0}^{100} P_x(\text{Iași})},$$

deci i. este :

$$I = \frac{\bar{m}_R}{\bar{m}\ (\text{Iași})} = \frac{\sum_{x=0}^{100} m_x P_x^R}{\sum_{x=0}^{100} P_x^R} \Bigg/ \frac{\sum_{x=0}^{100} m_x P_x^{(I)}}{\sum_{x=0}^{100} P_x^{(I)}}.$$

Exemplul 5 este cel al unui i. *al volumului fizic al producției* *, care măsoară variația în timp a unui *agregat*, compus din zeci de mii de produse, omogenizat cu ajutorul prețurilor (de obicei constante sau comparabile).
De aici apare evident că i. ca număr relativ exprimă nu numai variația caracteristicii x ci și a frecvențelor, respectiv a structurii frecvențelor (Σf).
Se mai poate observa că mărimile care constituie obiectul măsurării cu ajutorul indicilor pot reprezenta un *nivel*, corespunzînd unui punct în timp (de exemplu, valoarea fondurilor fixe la sfîrșitul anului, prețul unui produs la o anumită dată etc.) și un *flux* (producția industrială dintr-un an calendaristic, numărul nașterilor într-o lună etc.)
Indicatorii de nivel se mai numesc *indicatori de moment*, iar cei de flux — *indicatori de interval*, noțiuni examinate la XI. „Seriile cronologice".
Prin urmare, mărimile care constituie obiectul indicilor sînt foarte variate ca domenii de aplicație și uneori foarte complexe. Aceste mărimi pot reprezenta elemente de natură *cantitativă* sau extensivă (cantități de produse, număr de persoane) care pot fi aditive sau nonaditive. Notînd cu Z un element cantitativ de aceeași natură, volumul totalizat va fi Σz (același produs fabricat în mai multe întreprinderi); în cazul cînd aceste elemente sînt nonaditive, omogenizarea lor se face cu ajutorul unor coeficienți de comensurare numiți *ponderi* *. În acest scop se folosește o caracteristică esențială x; volumul căpătă expresia: Σxz. Elementele pot fi și de natură calitativă (prețul unitar, retribuția individuală etc.). Însumarea lor este lipsită de sens economic; de aceea, nivelul lor totalizat va fi: Σxz, în care z îndeplinesc rolul de ponderi. Pentru aceeași caracteristică x, înregistrată la mai multe unități ale unei colectivități (prețuri diferite la același produs), vom determina mai întîi media:

$$\bar{x} = \frac{\Sigma xz}{\Sigma z}, \text{ de unde: i.} = \frac{\bar{x}_1}{\bar{x}_2} = \frac{\Sigma x_1 z_1}{\Sigma z_1} : \frac{\Sigma x_0 z_0}{\Sigma z_0}.$$

Diferența dintre aspectul cantitativ și cel calitativ, ca și caracterul concret al mărimilor de indexat reclamă o atenție particulară la constituirea i.
Considerațiile exemplificative de mai sus ne duc la ideea unei clasificări a i. după diferite criterii:
a) după sfera de cuprindere, i. poate fi: i. *elementar* * și i. *sintetic* *;
b) după perioada de comparare distingem:
i. *cu bază fixă* * și i. *în lanț* *;
c) după natura fenomenului, avem: i. *al prețurilor* *, i. *al producției* *, i. *al costului vieții* * etc.
În domeniul propriu-zis al teoriei și metodei i. intră de fapt numai i. *sintetic* *; acesta cunoaște următoarele modalități:
d) după procedeul de calcul: i. *agregat**, i. *mediu** (aritmetic, armonic, geometric);
e) după modul de evidențiere a influenței structurii (factor cantitativ): i. *cu structură variabilă**, i. *cu structură fixă**, i. *al variației structurii**.
Pe plan teoretic, există discuții în legătură cu cele două funcții ale unui i. sintetic:
a) *funcția sintetică*: i. măsoară schimbarea medie a unui fenomen complex oarecare;
b) *funcția analitică*: i. descompune și măsoară contribuția fiecărui factor la schimbarea medie a fenomenului complex respectiv, cum ar fi, de exemplu,

contribuția variației prețurilor și a cantității la volumul total al desfacerilor mărfurilor.

Termenul adoptat în limba română este cel de **i.**; încercările mai vechi (N. St. Georgescu-Roegen, 1933) de a pune în circulație expresia de „numere indice" (în engl.: *index numbers*) nu au avut succes, dealtfel ca și în alte limbi. În franceză, *nombres indices* a fost abandonat pentru *indices*.

Istoric. Apariția **i.** datează din sec. al XVIII-lea în legătură cu încercările de a măsura evoluția prețurilor și a valorii banilor. Primii **i.** au fost din categoria **i.** elementari sau simpli. Dutot (1738) a propus un indice al prețurilor de tipul:

$$\frac{p_1 + p_2 + \ldots + p_n}{P_1 + P_2 + \ldots + P_n} = \frac{\Sigma p}{\Sigma P},$$

în care: p — prețul din perioada curentă; P — prețul din perioada de bază. Economistul italian Gian-Rinaldo Carli (1764) a propus indicele de prețuri sub forma unei medii aritmetice simple:

$$\frac{1}{n}\left(\frac{p_1}{P_1} + \frac{p_2}{P_2} + \ldots + \frac{p_n}{P_n}\right) = \frac{1}{n}\sum\frac{p}{P}.$$

Carli a încercat să măsoare efectul pe care l-a avut asupra puterii de cumpărare a monedei afluxul de metale prețioase în Europa după descoperirea Americii. El a calculat rapoartele prețurilor pentru trei produse: grîu, vin și ulei pe perioada 1400—1750, făcînd o medie simplă a acestor rapoarte.

Primul indice ponderat îi aparține englezului Arthur Young (1812), propus în lucrarea sa: „An Enquiry into the Progressive Value of Money", care a introdus coeficienți pentru importanța relativă a diferitelor articole ce intră în componența indicelui:

$$\frac{p_1k_1 + p_2k_2 + \ldots + p_nk_n}{P_1k_1 + P_2k_2 + \ldots + P_nk_n} = \frac{\Sigma pk}{\Sigma Pk},$$

în care: k — coeficient al importanței relative a fiecărui preț (pondere). O altă formă, folosind ponderi constante, propusă de A. Young, a fost:

$$\frac{1}{\Sigma c'}\left(\frac{p_1}{P_1}c_1' + \frac{p_2}{P_2}c_2' + \ldots + \frac{p_n}{P_n}c_n'\right) = \frac{1}{\Sigma c'}\sum\frac{p}{P}c' = \sum\frac{p}{P}c',$$

în care:

$$c = \frac{c'}{\Sigma c'} \text{ și } \Sigma c = 1.$$

În Anglia, în revista „Economist", s-au publicat de către Thomas Tocke și William Newmarch, în vol. VI din „Istoria prețurilor" (1838—1857) **i.** de prețuri. Newmarch a publicat în „Journal of Royal Statistical Society" (1859) o serie de **i.** folosind, se pare, pentru prima dată expresia de *index numbers*. Un moment important în teoria **i.** este legat de numele statisticienilor și economiștilor germani *Etienne Laspeyres* * (1834—1913) și *Hermann Paasche* * (1851—1925).

V.TREBICI

Primul a propus în 1864 un i. de tip agregat, ce-i poartă numele, folosind drept ponderi cantităţile de produse din perioada de bază ; al doilea, a formulat un i. agregat pentru preţuri, folosind drept ponderi cantităţile din perioada curentă :

$$L = \frac{\Sigma pQ}{\Sigma PQ} \text{ şi } P = \frac{\Sigma pq}{\Sigma Pq},$$

în care : q şi Q — cantităţile de produse din perioada curentă şi perioada de bază.

Un impuls teoriei i. l-a dat economistul W. St. Jevons în studiile sale : „A Serious Fall in the Value of Gold ascertained and its Social Effects set forth" (1863) şi „The Variation of Prices and the Value of the Currency since 1782" (1865).

A. Marshall propune i. *in lanţ* * (1885), precum şi un sistem de ponderare care să concilieze i. *Laspeyres* şi i. *Paasche*. Această idee a fost dezvoltată de *F. Y. Edgeworth* * (1845—1926) ; i. este astăzi cunoscut sub denumirea de i. *Edgeworth* (1887 şi 1889). Tot Edgeworth este cel care a introdus noţiunea de „tip" în teoria i.

Sub raport practic este de consemnat formularea i. *Dow-Jones*, datorat ziaristului H. Ch. Dow care l-a lansat în 1889, ameliorat de asociatul său, financiarul Jones ; acesta este i. valorilor industriale americane.

O contribuţie importantă la teoria i. are *L. von Bortkiewicz* * (1868—1931) care, într-o serie de studii (1923, 1924 şi 1925) a propus o formulă pentru măsurarea diferenţei dintre i. Laspeyres şi i. Paasche, de tipul :

$$P(p) - L(p) = \frac{\Sigma x_1 y_1}{\Sigma x_0 y_1} : \frac{\Sigma x_1 y_0}{\Sigma x_0 y_0} = 1 + r_{i_x i_y} v_{i_x} v_{i_y},$$

în care : v_{i_x} — coeficientul de variaţie a i. individuali ai indicatorului calitativ ;

v_{i_y} — coeficientul de variaţie a i. individuali ai indicatorului cantitativ ;

$r_{i_x i_y}$ — coeficientul de corelaţie dintre i. individuali ai indicatorilor calitativ şi cantitativ.

O nouă fază în teoria i. este marcată de lucrările lui *Irving Fisher* * (1867—1946), în care acesta a examinat un mare număr de i. Dacă pe vremea lui Quetelet existau 66 i., I. Fisher inventariază 134. El propune o serie de condiţii — *teste ale* i. * — pe care trebuie să le îndeplinească un i. (reversibilitate, circularitate etc.), ajungînd la „formula ideală" a i. care este media geometrică. În lucrarea sa principală („The Making of Index Numbers", 1922) el stabileşte o clasificare a i. în : i. agregat, aritmetic, armonic, geometric, median şi modal. Şcoala de conjunctură economică de la Harvard a dat un impuls important teoriei i. prin lucrările lui W. M. Persons (1919, 1924 şi, mai ales 1928 : „The Construction of Index Numbers), ale lui W. C. Mitchell („The Making and Using of Index Numbers, 1915, reluată în 1938). Institutul din Harvard a publicat, din 1919, aşa-numitul barometru, care cuprinde trei i. : a) de pe piaţa efectelor (cursurilor de bază ale acţiunilor) ; b) de pe piaţa mărfurilor şi c) de pe piaţa financiară.

V. TREBICI

Statisticianul englez *A. L. Bowley* (1869—1957) a dat, printre altele, un studiu asupra i. („The Precision of Index Numbers", publicat în „Journal of the Royal Statistical Society", March, 1926).
Cu lucrările matematicianului francez F. Divisia (1925, 1926) se lărgește studiul i. în lanț (→ I. *Divisia* *). În perioada interbelică sînt de semnalat contribuțiile statisticianului italian *Corrado Gini* * (1884—1965), econometricianului norvegian *Ragnar Frisch* * (1894—1973). Un aport original are statisticianul sovietic A. A. Konius, cu i. ce-i poartă numele („The Problem of the True Index of the Cost of Living", în „Econometrica", nr. 7, 1939, publicat inițial în limba rusă, în 1924). Este de semnalat că dezvoltarea econometriei, după înființarea Societății de Econometrie (1930) ca și lărgirea preocupărilor studiilor în domeniul cererii și ofertei, al elasticității au stimulat preocupările pentru i. O lucrare de sinteză este cea a lui B. Mudgett.

În perioada postbelică studiul comparabilității internaționale a favorizat dezvoltarea i. M. Gilbert se ocupă de problema calității („Quality Changes and Index Numbers", 1961), problemă cercetată și de statisticianul indian *P. C. Mahalanobis* * (1893—1972). H. Theil (1960) se preocupă de găsirea celui mai bun i. liniar al prețurilor și cantităților.

Cîmpul de aplicare a i. s-a extins considerabil în perioada postbelică, așa cum se poate constata din publicațiile O.N.U. și publicațiile naționale. Un domeniu nou este abordat și dezvoltat, în special în statistica sovietică, acela al i. *teritorial* *; contribuții de seamă L. Kazineț, N. Peregudov, L. Satunovski. În evoluția teoriei i. s-au cristalizat două teorii în ce privește concepția asupra i.
a) *teoria statistică* a i., potrivit căreia i. se poate aplica la orice situație, făcînd abstracție de ceea ce reprezintă fenomenele respective. Începutul ei datează din sec. al XVIII-lea și a fost ilustrată mai ales de Laspeyres, Paasche, I. Fisher,

b) *teoria economică* a i., concentrată asupra problemei i. prețurilor la consumator; această teorie este preocupată în special de interpretarea economică a fenomenelor ce reprezintă conținutul unui i.
Sub raportul funcției i. s-au conturat de asemenea două tendințe:

a) *concepția sintetică a* i.— i. reprezintă o caracterizare comparativă a nivelurilor fenomenelor economice;

b) *concepția analitică a* i.— i. este un indicator al influenței factorilor.
În statistica sovietică (*L. Kazineț*, 1963; *I. Paenson*, 1970) predomină tendința analitică, dar sînt de semnalat în ultimii ani încercări de a concilia, pe plan teoretic, cele două tendințe.

În România, preocupări timpurii legate de teoria i. sînt legate de numele matematicianului și statisticianului *N. St. Georgescu-Roegen* (n. 1906), cu studii publicate între 1932—1940. V. Madgearu a folosit în lucrarea sa „Evoluția economiei românești după războiul mondial" (1940) așa-numitul „quantum index" al lui N. St. Georgescu-Roegen, cu baza 1929, pentru seria 1920—1938, la comerțul exterior.

O dezvoltare remarcabilă a cunoscut teoria i. în perioada postbelică. L. Tövissi (1964) s-a ocupat de relațiile dintre i. de tip Laspeyres și Paasche, stabilind o serie de relații importante din punct de vedere teoretic și practic. V. Urseanu (1962) și S. Falcan-Popescu (1963, 1967) s-au ocupat de i. teritorial; numeroși statisticieni și economiști au dezvoltat funcția i. ca instrument de analiză a factorilor. Un studiu de sinteză în domeniul se datorează lui V. V. Dumitrescu (teza de doctorat 1977).

V.TREBICI

Aplicațiile i. în România sînt de dată relativ recentă. Ziarul economic „Argus" publica un i. al costului vieții (1916 = 100), cu 39 articole : 21 alimentare, 9 nealimentare, 9 diverse. Institutul Central de Statistică publica în „Buletinul prețurilor" un i. general și altul parțial al prețurilor en gros și de detaliu. În „Anuarul statistic al României" 1937 și 1938 (p. 592—607) găsim seria statistică privind prețul mediu anual pe orașe al alimentelor de origine animală, în număr de 17 ; prețul mediu anual pe orașe al alimentelor de origine vegetală, în număr de 31 ; prețul mediu anual pe orașe al combustibilului și articolelor diverse (10 articole) ; prețul mediu lunar la București al articolelor de primă necesitate. Seria se referă la 70 de orașe și avea ca bază anul 1913. În „Anuarul Statistic al României" (1939 și 1940) s-au publicat date privind i. costului vieții. În perioada postbelică gama i. s-a lărgit, în prima perioadă s-au construit : un i. general al prețurilor și tarifelor ; un i. al prețurilor și tarifelor plătite de salariați (i. costului vieții), un i. al prețurilor și tarifelor calculate pe baza unei scheme fixe de consum („coșniță").

În evoluția i. prețurilor, se constată următoarele etape (V. V. Dumitrescu, 1977) :

a) etapa 1947—1951 : se calculează i. pentru un număr restrîns de localități, folosindu-se datele din bugetele familiilor de muncitori ;

b) etapa 1952—1954 : pe baza unificării prețurilor a devenit posibil un calcul al i. la nivelul țării, cu un nomenclator de circa 200 produse ;

c) etapa 1955—1974 : un i. cu sferă completă care reflectă totalitatea mișcărilor de prețuri cu amănuntul. I. se stabilesc pe grupe de produse, familii de produse și pe total comerț socialist ;

d) de la 1 ianuarie 1973, se introduce sistemul de urmărire a prețurilor de producție și de livrare din industrie, pe baza metodei excepțiilor. Se calculează un i. armonic pe total industrie, pe ministere, pe principalele destinații ;

e) la 1 ianuarie 1975 s-a introdus sistemul informațional de urmărire a prețurilor cu amănuntul.

În prezent se practică următorii i. (*M. Bulgaru*, 1982) :

1. I. prețurilor de producție și de livrare pentru industrie (i. armonic) ;

2. I. prețurilor produselor agricole care au fost vîndute către fondul de stat și pe piața țărănească, pe baza unui nomenclator de 42 produse principale (33 vegetale și 9 animale) ;

i. este un tip agregat ;

3. I. prețurilor cu amănuntul și al tarifelor care întrunește patru i. : i. prețurilor cu amănuntul din comerțul socialist ; i. prețurilor pe piața țărănească ; i. tarifelor pentru serviciile prestate de unitățile socialiste ; i. tarifelor pentru serviciile prestate de meseriașii necooperativizați ;

4. I. costului vieții ; ponderarea se face cu cheltuielile totale ale populației din bugete de familie.

Prețurile practicate pe piața țărănească se culeg, după metoda sondajului, de pe un număr de 270 piețe din 95 orașe, reprezentînd 80% din totalul vînzărilor. În ce privește cheltuielile din bugetele de familie, ele sînt clasificate în 79 grupe alimentare și 52 grupe nealimentare. Calculele se fac diferențiat : pe muncitori, țărani, intelectuali-funcționari și pensionari, apoi pe grupe de venituri, pe urban și rural.

În „Anuarul Statistic al Republicii Socialiste România" se publică numeroase serii de i., cu baza 1938, 1948, 1950 și cu alte baze.

V.TREBICI

INDICE ELEMENTAR *(individual).* **I.** ce măsoară variația unui element în timp, spațiu sau pe categorii, de tipul :

$$i_{t/0}(x) = \frac{x_t}{x_0} \quad \text{și} \quad i_{t/0}(f) = \frac{f_t}{f_0}$$

și

$$i_{t/0}(xf) = \frac{x_t f_t}{x_0 f_0}.$$

Se notează cu **i.** (mic). Cele mai frecvente aplicații :

— **i.e.(i.)** de preț ;

$$i_{t/0}(p) = \frac{p_t}{p_0};$$

— **i.e.(i.)** de cantități :

$$i_{t/0}(q) = \frac{q_t}{q_0};$$

— **i.e.(i.)** al valorii unui produs :

$$i_{t/0}(qp) = \frac{q_t p_t}{q_0 p_0},$$

în care : t — perioada curentă, 0 — perioada de bază.
Cum s-a mai spus, teoria **i.** nu se ocupă de **i.** individuali ; totuși, aceștia constituie obiectul **i.** mediu (aritmetic, geometric etc.)
I.e.(i.) răspunde la trei teste : *testul de reversibilitate* *, *testul de tranzitivitate** și *testul de circularitate* *, condiții pe care nu le au i. sintetici.
INDICE SINTETIC *(total, general).* Număr relativ ce măsoară variația unui fenomen complex, compus din elemente ce nu pot fi comensurate nemijlocit, adică din elemente nonaditive. Variația se înțelege în raport cu timpul, spațiul sau o categorie oarecare. Forma sa generală este de i. *agregat.* Întrucît variabila sau variabilele, în combinațiile lor, pot avea expresiile Σz, Σzx și Σxz numite și *agregate*, un **i.s.** (**t.g.**) capătă formele următoare : a) pentru caracteristica cantitativă z :

$$I_z = \frac{\Sigma z_1}{\Sigma z_0} \quad \text{(aditivă)}$$

$$I_z = \frac{\Sigma z_1 x_0}{\Sigma z_0 x_0} \quad \text{(nonaditivă)}$$

b) pentru caracteristica calitativă x :

$$I_{\bar{x}} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\Sigma x_1 z_1}{\Sigma z_1} : \frac{\Sigma x_0 z_0}{\Sigma z_0} \quad \text{(elemente cantitative aditive)}$$

$$I_x = \frac{\Sigma z x_1}{\Sigma z x_0} \quad \text{(elemente cantitative non aditive).}$$

PONDERE. Coeficient de comensurare folosit la calculul unui indice sintetic pentru o colectivitate de elemente necomensurabile nemijlocit, avînd drept funcție :
a) comensurarea elementelor necomensurabile ;
b) stabilirea importanței relative a fiecărui element în colectivitatea statistică. Atașarea prețului unitar la cantitățile fiecărui produs permite obținerea unui agregat :

$$q'p' + q''p'' + \ldots + q^n p^n = \Sigma q p \text{ (cazul a).}$$

Întrunirea **i.** individual într-un **i.** sintetic se face prin atașarea unor **p.** de tipul

$$\frac{p^i q^i}{\Sigma p^i q^i} \quad (\rightarrow \textit{coeficient bugetar}\ *) \quad \text{(cazul b).}$$

P. se notează cu w (în engl. : *weight*) ; în unele lucrări se folosește simbolul ω.
INDICE SIMPLU. Formă de **i.** în care nu intervin ponderile.
Exemplu. O medie simplă de i. individuali ai prețurilor. (→ *Indicele prețurilor**.)
INDICE PONDERAT. Formă de **i.** în care se folosesc ponderi. Teoretic orice **i.** sintetic este un i. ponderat.
INDICE GENERAL. I. sintetic construit pentru o colectivitate. *Exemplu. Indicele general al prețurilor cu amănuntul.* I se mai spune **i.** total.
INDICE PARȚIAL *(de grup).* I. sintetic construit pentru o subcolectivitate sau un grup. *Exemplu. Indice de grup al prețurilor produselor pe piața țărănească.* Metodologia unui **i.p.** (g.) nu diferă de cea a unui i. general ; diferență constă în numărul de elemente luate în calcul.
Deci perechile de i. sînt :
a) individual (elementar) — sintetic
b) simplu — ponderat
c) general — parțial (de grup).
De obicei, i. general sau total este considerat sinonim cu i. sintetic ; **i.p.** (g.) se întîlnește uneori cu denumirea de *subindice.*
INDICE AGREGAT. I. sintetic de tipul cel mai general ; forma fundamentală a i. sintetic, în care mărimile ce se compară au expresia unor agregate :

$$I_{t/0}(x) = \frac{\Sigma x_1 z_0}{\Sigma x_0 z_0} \quad \text{și} \quad I_{t/0}(z) = \frac{\Sigma z_1 x_0}{\Sigma z_0 x_0}.$$

Ponderile **i.a.** pot fi cele din perioada de bază sau din perioada curentă.
Un **i.a.** este un i. *mediu* * ; forma sa primară este de medie necxplicită ; în condițiile unui anumit sistem de ponderare **i.a.** devine o medie explicită a i. individuali.

V.TREBICI

Tipurile fundamentale de **i.a.** sînt : i. *prețurilor* * și i. *cantităților* sau *volumului fizic al producției*.

PERIOADĂ DE BAZĂ (*de referință*). Moment de timp sau perioadă care constituie baza față de care se compară o mărime în cadrul i.

Este egal cu 100, în mod convențional, iar perioada se notează cu 100. *Exemplu* : I. volumului fizic al producției industriale are drept bază anul 1938 (1938 = = 100), sau 1950 (1950 = 100), sau unul de sfîrșit al fiecărei perioade cincinale (1955 = 100, 1975 = 100, 1980 = 100). Se notează cu 0. *Exemplu*. $\Sigma q_0 p_0$.

PERIOADĂ CURENTĂ. Moment de timp sau perioadă a cărei mărime actuală se compară cu mărimea din perioada de bază. Se notează cu 1. *Exemplu*. $\Sigma q_1 p_1$. Cînd avem de a face cu *o serie de indici* *, mărimea din perioada curentă se notează cu t ($t = 1, 2, \ldots, n$ perioade).

INDICE MEDIU. Medie a i. individuali, ponderată sau neponderată.

Indice mediu aritmetic. Medie aritmetică, ponderată sau neponderată a i. individuali.

Indice mediu armonic. Medie armonică, ponderată sau neponderată, a i. individuali.

Indice mediu geometric. Medie geometrică, ponderată sau neponderată a i. individuali. (→ *Indice Fisher*).

Indice median. Valoare mediană a i. individuali dintr-o repartiție statistică a i. individuali, cu valori aranjate în ordine crescătoare. Se utilizează foarte rar.

Indice modal. Valoarea modală a i. dintr-o repartiție statistică a i. individuali. Se utilizează foarte rar.

SERIE DE INDICI. Șir de i. individuali sau sintetici în care numărul de perioade este de cel puțin trei ($t \geqslant 3$).

În aceste condiții se pun probleme precum : alegerea bazei, sistemul de ponderi constante sau variabile, i. cu bază fixă și în lanț (→ și *serii cronologice*).

Indice cu bază fixă. I. dintr-o seric de i. în care compararea se face în permanență cu mărimea din perioada de bază, care rămîne neschimbată.

În cazul unui i. individual :

$$i_{t/0} = \frac{z_1}{z_0}; \frac{z_2}{z_0}; \ldots; \frac{z_t}{z_0}.$$

În cazul unui i. sintetic :

$$I_{t/0} = \frac{\Sigma x_1 z_0}{\Sigma x_0 z_0}; \frac{\Sigma x_2 z_0}{\Sigma x_0 z_0}; \ldots; \frac{\Sigma x_t z_0}{\Sigma x_0 z_0}.$$

Indice în lanț. I. dintr-o serie de i. în care compararea se face în mod succesiv între mărimile din perioade alăturate.

În cazul unui i. individual :

$$i_{t/t-1} = \frac{z_1}{z_0}; \frac{z_2}{z_1} \ldots \frac{z_t}{z_{t-1}}.$$

În cazul unui i. sintetic :

$$I_{t/t-1} = \frac{\Sigma x_1 z_0}{\Sigma x_0 z_0}; \frac{\Sigma x_2 z_0}{\Sigma x_1 z_0} \ldots \frac{\Sigma x_t z_0}{\Sigma x_{t-1} z_0}.$$

Indice cu ponderi constante. I. sintetic dintr-o serie de i. în care ponderile rămîn neschimbate indiferent de numărul perioadelor.

$$I_{t/0} = \frac{\Sigma x_1 w_0}{\Sigma x_0 w_0}; \quad \frac{\Sigma x_2 w_0}{\Sigma x_0 w_0} \cdots \frac{\Sigma x_t w_0}{\Sigma x_0 w_0},$$

în care : w_0 — ponderile din perioada de bază.

Indice cu ponderi variabile. I. sintetic dintr-o serie de i. în care ponderile se schimbă cu fiecare perioadă.

$$I_{t/t-1} = \frac{\Sigma x_1 w_1}{\Sigma x_0 w_1}; \quad \frac{\Sigma x_2 w_2}{\Sigma x_1 w_2} \cdots \frac{\Sigma x_t w_t}{\Sigma x_{t-1} w_t}.$$

INDICE LASPEYRES. I. agregat în care ponderile i. sînt cele din perioada de bază. Notînd cu p — prețurile și cu q — cantitățile, vom avea :

$$L(p) = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}; \quad L(q) = \frac{\Sigma p_0 q_1}{\Sigma p_0 q_0}.$$

INDICE PAASCHE. I. agregat în care ponderile i. sînt cele din perioada curentă :

$$P(p) = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}; \quad P(p) = \frac{\Sigma p_1 q_1}{\Sigma p_1 q_0}.$$

INDICE EDGEWORTH *(Marshall-Edgeworth).* I. în care ponderile i. sînt cele din perioada de bază și perioada curentă. Pentru prețuri, I.E. este :

$$E(p) = \frac{\Sigma \{p_1 (q_0 + q_1)\}}{\Sigma \{p_0 (q_0 + q_1)\}}.$$

INDICE „IDEAL" FISHER. Media geometrică a i. Laspeyres și i. Paasche. Pentru prețuri

$$F = \sqrt{LP} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}.$$

INDICELE DIVISIA. I. în lanț instantaneu. În cazul indicelui prețurilor, considerînd că prețurile și cantitățile variază continuu :

$$\frac{D + \mathrm{d}D}{D} = L_t + \mathrm{d}t/t = \frac{\sum_i p^i(t) q^i(t) \dfrac{p'^i(t + \mathrm{d}t)}{p^i(t)}}{\sum_i p^i(t) q^i(t)} =$$

$$= 1 + \frac{\sum_i q^i(t)p'^i(t)}{\sum_i q^i(t)p^i(t)}\,dt = P_{t+dt/t}.$$

De aici :

$$D_{t/0}(p) = \exp\left[\int_0^t \frac{\sum_i q^i(t)p'^i(t)}{\sum_i q^i(t)p^i(t)}\,dt\right] = \exp\left[\int_0^t \varphi\ dt\right],$$

punînd $\varphi(t) = \dfrac{\sum_i q^i(t)p'^i(t)}{\sum_i q^i(t)p^i(t)}$.

Interesul teoretic al **I.D.** este de a prezenta proprietatea de circularitate, deci și de reversibilitate și de înlănțuire :

$$D_{t'/t}(p) = \exp\left[\int_t^{t'} \varphi(t)\,dt\right] = \frac{\exp\left[\int_0^{t'} \varphi(t)dt\right]}{\exp\left[\int_0^t \varphi(t)dt\right]} = \frac{D_{t'/0}(p)}{D_{t/0}(p)}.$$

I.D. propus în 1926 a fost dezvoltat de Törnquist (1937) și de H. Theil (1960). Forma obișnuită în care se întîlnește **I.D.** este :

$$P_{0t}\,Q_{0t} = \frac{\sum_{i=1}^n p_{it}\,q_{it}}{\sum_{i=1}^n p_{i0}q_{i0}}.$$

Logaritmînd ambele părți și diferențiind în raport cu t, se obține :

$$\frac{1}{P_{0t}} \cdot \frac{dP_{0t}}{dt} + \frac{1}{Q_{0t}} \cdot \frac{dQ_{0t}}{dt} =$$

$$= \frac{1}{\sum_{i=1}^n p_{it}q_{it}} \left(\sum_{i=1}^n q_{it} \frac{dp_{it}}{dt} + \sum_{i=1}^n p_{it} \frac{dq_{it}}{dt} \right).$$

Integrînd ecuațiile, obținem :

$$\log P_{01} = \sum_{i=1}^{n} \int_{t_0}^{t_1} c_i(t)\, \mathrm{d} \log p_{it},$$

în care :

$$c_i(t) = \frac{p_{it} q_{it}}{\sum_{i=1}^{n} p_{it} + q_{it}}.$$

Deci valoarea P_{01} depinde de evoluția lui $c_i(t)$ între t_0 și t_1 ; c este ponderea diferitelor produse în bugetul total (→ *indicele prețurilor* și *indicele costului vieții*).

Dacă valorile $c_i(t)$ sînt presupuse constante în timp (toate produsele au o elasticitate a prețurilor cererii egală cu 1), obținem indicele geometric :

$$P_0 = \prod_{i=1}^{n} \left(\frac{p_{i1}}{p_{i0}} \right)^{c_i}.$$

Dacă nu se face nici o ipoteză cu privire la c_i atunci :

$$\log P_{01} = \sum_{i=1}^{n} \bar{c}_i \left(\log p_{i1} - \log p_{i0} \right),$$

în care : $\bar{c}_i$ — media ponderată a valorilor c_i la momentele t_0 și t_1.

FORMULA BORTKIEWICZ. Formulă de estimare a diferenței dintre i. Laspeyres și i. Paasche. În cazul i. prețurilor :

$$L_{1/0}(p) = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}; \quad P_{1/0}(p) = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1},$$

de unde :

$$P_{1/0}(p) - L_{1/0}(p) = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} - \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} = \frac{\Sigma p_0 q_0}{\Sigma p_0 q_1} \left[\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} - \frac{\Sigma p_1 q_0 \, \Sigma p_0 q_1}{\Sigma p_0 q_0 \, \Sigma p_0 q_0} \right] =$$

$$= \frac{1}{L_{1/0}(p)} \left[\frac{\Sigma p_0 q_0 \frac{p_1 q_1}{p_0 q_0}}{\Sigma p_0 q_0} - \frac{\Sigma p_0 q_0 \frac{p_1}{p_0} \Sigma p_0 q_0 \frac{q_1}{q_0}}{\Sigma p_0 q_0 \, \Sigma p_0 q_0} \right].$$

Observăm că între parantezele mari avem : media ponderată a i. individuali cu w_0 (ponderi din perioada de bază), media ponderată a i. individuali ai prețurilor (ponderi w_0) și media i. individuali ai cantităților. Prin urmare, expresia

dintre parantezele mari este covarianța ponderată a i. individuali ai prețurilor și cantităților :

$$P_{1/0}(p) - L_{1/0}(p) = \frac{\operatorname{cov}[I_{1/0}(p) ; I_{1/0}(q)]}{L_{1/0}(q)}.$$

Pentru cantități :

$$P_{1/0}(q) - L_{1/0}(q) = \frac{\operatorname{cov}[I_{1/0}(p), I_{1/0}(q)]}{L_{1/0}(p)}.$$

În general :

a) $P < L$, cînd în medie, prețurile și cantitățile variază în sens opus ;
b) $P > L$, cînd, în medie, prețurile și cantitățile variază în același sens ;
c) $P = L$, cînd prețurile și cantitățile variază fără corelația liniară.

Prin urmare, sistemele de ponderare a i. sintetici au importanță în ce privește diferiți indici.

INDICE ARITMETIC MEDIU PONDERAT. I. sintetic reprezentînd o medie aritmetică ponderată, după schema :

$$I_{1/0}(i_x) = \frac{\Sigma x z i_x}{\Sigma x z} \quad \text{și} \quad I_{1/0}(i_z) = \frac{\Sigma x z i_z}{\Sigma x z}.$$

În cazul i. volumul fizic al producției, **i.a.m.p.** este :

$$I_{1/0}(i_q) = \frac{\Sigma q_0 p_0 i_q}{\Sigma q_0 p_0}.$$

În condiții specifice de ponderare, **i.a.m.p.** este egal cu i. agregat.

INDICE ARMONIC MEDIU PONDERAT. I. sintetic reprezentînd o medie armonică ponderată, după schema :

$$I_{1/0}(i_x) = \frac{\Sigma x z}{\Sigma \frac{1}{i_x} x z} \quad \text{și} \quad I_{1/0}(i_z) = \frac{\Sigma x z}{\Sigma \frac{1}{i_z} x z}.$$

În cazul *indicelui prețurilor* * **i.a.m.p.** este :

$$I_{1/0}(i_p) = \frac{\Sigma p_1 q_1}{\Sigma \frac{1}{i_p} p_1 q_1}.$$

TESTE ALE INDICILOR. Condiții pe care trebuie să le îndeplinească i. individual și i. sintetic. Majoritatea testelor a fost formulată de I. Fisher (1922), cu prilejul căutării unei formule a „indicelui ideal". Introducerea testelor în teoria i. reprezintă contribuția cea mai de seamă a orientării matematice.

V.TREBICI

Test de identitate a indicelui I. pentru perioada curentă trebuie să fie egal cu 1, dacă luăm perioada curentă drept perioadă de bază.

Test de reversibilitate în timp a indicelui. I. a două perioade din care prima este baza celeilalte, calculați cu aceeași metodă trebuie — în cazul cînd facem ca perioada a doua să fie baza primei perioade — să ne dea valorile inverse.

Dacă P_{01} este $\frac{1}{2}$, P_{10} trebuie să fie egal cu $2 = \frac{1}{0,5}$.

Test al schimbării bazei indicelui. Rezultatul împărțirii a doi indici din două perioade diferite, calculați cu aceeași metodă și avînd aceeași bază trebuie să fie independent de alegerea bazei.

Test circular al indicelui. Produsul înmulțirii *indicilor în lanț** succesivi, calculați prin schimbarea bazei, trebuie să fie egal cu *indicele cu bază fixă** corespunzător :

Test de reversibilitate a factorilor indicelui. Produsul a doi factori — de exemplu, i. prețurilor și i. cantităților — trebuie să fie egal cu i. produsul acestor factori. Astfel, i. *valorii* este produsul i. prețurilor și i. cantităților, deoarece schimbările valorii se datorează schimbărilor prețurilor sau cantităților.

Test de proporționalitate a indicelui. Dacă toți i. individuali sînt egali cu A, i. sintetic trebuie să fie egal și el cu A.

Test de determinare a indicelui. I. sintetic nu trebuie să devină nul, nedefinit sau nedeterminat dacă unul din i. parțiali devine nul, nedefinit sau nedeterminat.

Test de omogenitate (comensurabilitate) a indicelui. Valoarea i. trebuie să fie independentă de schimbările unităților de măsură.

NOTĂ. Cele mai importante **t. : t.** de reversibilitate în timp : **t.** circular și **t.** de reversibilitate a factorilor.

Teste ale indicelui elementar (individual). Condiții pe care le satisfac i. elementari :

a) **t.** de reversibilitate în timp ;

$$I_{1/0} = \frac{M_1}{M_0}, \quad \text{în care : } M - \text{mărimea ;}$$

de unde :

$$I_{0/1} = \frac{1}{I_{1/0}}, \quad \text{deoarece } \frac{M_0}{M_1} = \frac{1}{M_1/M_0}$$

b) **t.** circular :

$$I_{2/0} = I_{2/1} \cdot I_{1/0}, \text{ deoarece } \frac{M_2}{M_0} = \frac{M_2}{M_1} \cdot \frac{M_1}{M_0}.$$

Această proprietate a i. în lanț este folosită și pentru :

$$I_{2/1} = \frac{I_{2/0}}{I_{1/0}}.$$

c) t. de reversibilitate a factorilor

$$I_{2/1} \cdot I_{1/0} \cdot I_{0/2} = 1, \quad \text{deoarece}$$

$$I_{2/1} \cdot I_{1/0} = I_{2/0} = \frac{1}{I_{0/2}}.$$

Dacă M este produsul mărimilor h și k, i. elementar al mărimii produs este egală cu produsul i. mărimilor factori :

$$\text{Dacă } M_t = h_t \cdot k_t,$$

atunci

$$I_{1/0}(M) = I_{1/0}(h) \cdot I_{1/0}(k),$$

deoarece

$$\frac{M_1}{M_0} = \frac{h_1}{h_0} \cdot \frac{k_1}{k_0}.$$

Aplicație particulară : valoarea = preț × cantitate, de unde : i. elementar al valorii = i. elementar de prețuri × i. elementar de cantități.

Teste ale indicelui sintetic. Condiții pe care le satisfac i. sintetic.

Sistemul de ponderare, folosirea ponderilor constante sau variabile sint tot atîtea condiții care influențează t.i.s.

a) i. Laspeyres, Paasche și Fisher nu răspund testului circular ;

b) i. Laspeyres și Paasche nu răspund testului de reversibilitate, în timp ce i. Fisher are această proprietate :

$$F_{0/1}(M) = \sqrt{L_{0/1}(M) \cdot P_{0/1}(M)} =$$

$$= \frac{1}{\sqrt{P_{1/0}(M)\ L_{1/0}(M)}} = \frac{1}{F_{1/0}(M)}.$$

Este locul să facem comparația dintre i. Laspeyres, Paasche și Fisher. Dacă la momentul 0 ponderea unui component în mărimea complexă M o notăm cu w_0, iar în perioada curentă cu w_1 (coeficienții w joacă rolul frecvențelor f ale unei variabile statistice). Atunci :

a) i. Laspeyres este media aritmetică a i. elementari ponderată cu w_0 (perioada de bază sau de referință) ;

$$L_{1/0}(M) = \Sigma\, w_0 I_{1/0}(M) = \Sigma\, w_0 \frac{M_1}{M_0}.$$

b) i. Paasche este media armonică a i. elementari, ponderată cu w_1 (perioada curentă) :

$$\frac{1}{P_{1/0}(M)} = \sum \frac{w_1}{I_{1/0}(M)} = \sum w_1 \frac{M_0}{M_1}.$$

c) i. Fisher este media geometrică simplă a i. Laspeyres și Paasche :

$$F_{1/0}(M) = \sqrt{L_{1/0}(M) P_{1/0}(M)}$$

d) i. Fisher este cuprins între L și P, fiind o medie ;
e) i. Laspeyres și Paasche sînt cuprinși între i. elementari extremi, fiind medii ; proprietatea este valabilă și pentru i. Fisher :

$$\min I_{1/0}(M) \leqslant [P_{1/0}(M), F_{1/0}(M), L_{1/0}(M)] \max I_{1/0}(M) ;$$

f) adesea P este mai mic decît L, deoarece media armonică este mai mică decît media aritmetică ;
g) pentru considerente practice, cel mai avantajos este i. Laspeyres.
În cele două aplicații fundamentale, la prețuri și la cantități, ce i doi i. se prezintă astfel :

Indice	Prețuri (p)	Cantități (q)
Laspeyres	$L_{1/0}(p) = \dfrac{\Sigma p_0 q_0 \dfrac{p_1}{p_0}}{\Sigma p_0 q_0} = \dfrac{\Sigma p_1 q_0}{\Sigma p_0 q_0}$	$L_{1/0}(q) = \dfrac{\Sigma p_0 q_0 \dfrac{q_1}{q_0}}{\Sigma p_0 q_0} = \dfrac{\Sigma p_0 q_1}{\Sigma p_0 q_0}$
Paasche	$P_{1/0}(p) = \dfrac{\Sigma p_1 q_1}{\Sigma p_1 q_1 \dfrac{p_0}{p_1}} = \dfrac{\Sigma p_1 q_1}{\Sigma p_0 q_1}$	$P_{1/0}(q) = \dfrac{\Sigma p_1 q_1}{\Sigma p_1 q_1 \dfrac{q_0}{q_1}} = \dfrac{\Sigma p_1 q_1}{\Sigma p_1 q_0}$

Reguli pentru construirea i. sintetici :
a. Formula de bază a i. sintetic este aceea de i. agregat.
b. Un i. trebuie tratat sistemic : în cadrul unui sistem de i, la un fenomen complex, elementele sînt factori, iar i. devin i. factoriali (→ **I.** *analitic*).
c. Ponderile trebuie alese în funcție de natura elementelor componente ale mărimii complexe.
d. Perioada de bază trebuie aleasă în așa fel încît să nu fie afectată de accidente ;
e. Produsele (articolele) incluse în i. să fie reprezentative, să țină seama de calitate, de apariția unor produse noi și de dispariția celor vechi.
INDICE ANALITIC. I. care exprimă influența unui factor asupra unui fenomen complex. (→ *Concepția analitică a i.*). În aceste condiții, i. trebuie să formeze un sistem ; nivelurile unui factor se fixează ca mărimi constante pentru a se putea măsura variația celuilalt factor. Se aplică pe scară largă în analiza economică. (→ *Determinarea contribuției factorilor la modificarea unui indicator complex.*)
INDICE AL VALORII MEDII. I. sintetic al elementelor calitative dintr-un ansamblu de elemente aditive, avînd expresia unui raport între nivelul mediu comparat și cel de bază de comparație :

$$I_{\bar{x}} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\Sigma x_1 z_1}{\Sigma z_1} : \frac{\Sigma x_0 z_0}{\Sigma z_0} .$$

Să se observe că fiind vorba de medii ponderate, mărimea i. depinde nu numai de variația caracteristicii x, ci și de variația ponderilor.
Înlocuind ponderile z cu frecvența lor relativă :

$$z^* = \frac{z}{\Sigma z} \quad (\Sigma z^* = 1)$$

i.v.m. devine un i. de formă agregată :

$$I_{\bar{x}} = \frac{\Sigma x_1 z_1^*}{\Sigma x_0 z_0^*}.$$

De notat că nu frecvențele absolute, ci cele relative, adică raportul lor reciproc, influențează mărimea i.v.m.
Dacă în noul i. de formă agregată introducem i. individuali :

$$i_x = \frac{x_1}{x_0}$$

și i. individuali ai frecvențelor relative

$$i_z = \frac{z_1}{\Sigma z_1} : \frac{z_0}{\Sigma z_0} = \frac{z_1}{z_0} \text{ i } \frac{\Sigma z_1}{\Sigma z_0},$$

acesta devine :

$$I_{\bar{x}} = \frac{\Sigma i_x i_z x_0 z_0}{\Sigma x_0 z_0},$$

adică un i. mediu ponderat calculat din produsele i. individuali $i_x i_z$.
INDICE CU STRUCTURĂ VARIABILĂ. (→ *Indice al valorilor medii.*)
INDICE CU STRUCTURĂ FIXĂ. I. al valorii medii din care s-a eliminat efectul structurii :

$$I_{\bar{x}(x)} = \frac{\Sigma x_1 z_1^*}{\Sigma z_1^*} : \frac{\Sigma x_0 z_1^*}{\Sigma z_1^*} = \frac{\Sigma x_1 z_1^*}{\Sigma x_0 z_1^*}.$$

Acesta este indicele factorului calitativ.
INDICE DE STRUCTURĂ *(al schimbărilor structurale).* I. al factorului structural de tipul :

$$I_{\bar{x}(z)} = \frac{\Sigma x_0 z_1^*}{\Sigma z_1^*} : \frac{\Sigma x_0 z_0^*}{\Sigma z_0^*},$$

întrucît $\Sigma z_0^* = \Sigma z_1^* = 1,$

i. devine :

$$I_{\bar{x}}(z^*) = \frac{\Sigma x_0 z_1^*}{\Sigma x_0 z_0^*} .$$

Cei trei i., cu sistemul de ponderare adoptat, constituie un sistem, iar relația dintre ei este :
I cu structură variabilă = *I* cu structură fixă × *I* de structură

$$I_{\bar{x}}(x, z) = I_{\bar{x}}(x)\, I_{\bar{x}}(z^*) = \frac{\Sigma x_1 z_1^*}{\Sigma x_0 z_0^*} = \frac{\Sigma x_1 z_1^*}{\Sigma x_0 z_1^*} \times \frac{\Sigma x_0 z_1^*}{\Sigma x_0 z_0^*} .$$

Sistemul de i. prezentat mai sus are aplicație la analiza retribuției medii, a productivității medii, a prețului mediu, a diferiților i. demografici (de exemplu, rata de mortalitate).

INDICE STANDARDIZAT. I. ponderat cu *ponderi standard* *, reprezentînd proporțiile din colectivitatea generală, în vederea eliminării influenței structurii. Cea mai veche aplicare a **i.s.** se datorează demografiei, în particular, la analiza mortalității.
Fie m_x — ratele specifice de mortalitate după vîrstă și P_x — efectivele populației corespunzînd ratelor specifice, în care x — vîrsta (0, 1 ,... 100 ani). Rata generală de mortalitate, cu semnificație de medie genera lă, este :

$$\bar{m}_x = \frac{\sum_{x=0}^{100} m_x P_x}{\sum_{x=0}^{100} P_x} .$$

Putem adopta în locul ponderilor absolute pe cele relat i ve :

$$P_x^* = \frac{P_x}{\Sigma P_x}, \text{ de unde } \Sigma P_x^* = 1.$$

Presupunem că avem de comparat rata generală de mortalitate a două județe : *A* și *B*. I. va fi :

$$I_{A/B}(m) = \frac{\sum_{x=0}^{100} m_x^{(A)} P_x^{*(A)}}{\sum_{x=0}^{100} P_x^{*(A)}} : \frac{\sum_{x=0}^{100} m_x^{(B)} P_x^{*(B)}}{\sum_{x=0}^{100} P_x^{*(B)}} = \frac{\bar{m}_x(A)}{\bar{m}_x(B)} .$$

I. este un raport între două medii, fiecare avînd ponderile sale : mărimea obținută este influențată atît de variațiile caracteristicii cît și de variația ponderilor relative. Pentru a face comparabile ratele de mortalitate ale celor

două județe se introduc în ambele formule ponderile relative ale populației României :

$$I_{A/B}(m) = \frac{\Sigma m_x^{(A)} P_x^{*(R)}}{\Sigma m_x^{(B)} P_x^{*(R)}}.$$

Din i. cu structură variabilă, el a devenit *indice cu structură fixă* * și deci influența unor structuri diferite a fost eliminată. Procedeul se numește *metoda directă de standardizare* ; există și o metodă indirectă.

Noțiunea de **i.s.** și de standardizare, elaborate inițial de demografie, au căpătat o largă aplicare în cadrul comparațiilor internaționale și al comparațiilor teritoriale în aceeași țară.

(→ *Indice teritorial* *).

PONDERI STANDARD. Ponderi, sub forma frecvențelor relative, de tipul $z^* = \frac{z}{\Sigma z}$ sau, mai general, $f^* = \frac{f}{\Sigma f}$, stabilite pentru o colectivitate generală și care se aplică la calculul i. pentru două sau mai multe subcolectivități aparținînd colectivității date.

INDICE TERITORIAL. Raport între mărimile unui indicator oarecare în spațiu : pe țări, în cadrul țării — pe județe, orașe etc. Astfel, raportul dintre prețul mediu al unui produs vîndut pe piața țărănească dintr-un oraș, raportat la prețul mediu al aceluiași produs vîndut într-un alt oraș. Fiind vorba de compararea a două medii cu ponderi diferite, se folosesc *ponderi standard* *. **I.t.** se construiesc pentru comparațiile internaționale și naționale. Obiectul comparațiilor teritoriale naționale îl constituie prețul mediu, prețul de cost mediu, productivitatea medie a muncii, recolta medie la hectar, retribuția medie a muncii etc. Sistemul de ponderare are o importanță deosebită.

Se constată un interes deosebit pentru **i.t.** în perioada postbelică : în literatura și practica țărilor socialiste s-a dezvoltat considerabil teoria **i.t.** în jurul anilor 1960. (→ *Istoric.*)

DETERMINAREA CONTRIBUȚIEI FACTORILOR LA MODIFICAREA UNUI INDICATOR COMPLEX. Aplicație a *i.* analitic pentru măsurarea contribuției fiecărui factor la modificarea totală. Exemple : contribuția prețurilor și cantităților la modificarea valorii totale, a productivității muncii și a numărului muncitorilor la modificarea volumului total al producției etc. Schemele folosite pot fi de tip aditiv, multiplicativ și mixt (aditiv-multiplicativ).

Schema aditivă. Fie doi factori elementari x și z, cu i. complex (y):

$$y = x + z.$$

Rolul celor doi factori x și z :

— în formă relativă : $i_y = \frac{y_1}{y_0} = \frac{x_1+z_1}{x_0+z_0} = \frac{x_1}{x_0+z_0} + \frac{z_1}{x_0+z_0}$;

— în formă absolută (ca diferență) :

$$dy = y_1 - y_0 = (x_1+z_1) - (x_0+z_0) = (x_1 - x_0) + (z_1 - z_0).$$

Schema multiplicativă cu doi factori :

$$x = \frac{y}{z}, \qquad \text{de unde : } y = xz.$$

Rolul factorilor x și z la modificarea lui y :
— în formă relativă :

$$i_y = \frac{y_1}{y_0} = \frac{x_1 z_1}{x_0 z_0} = i_x \cdot i_z ;$$

— în formă absolută :

$$dy = y_1 - y_0 = x_1 z_1 - x_0 z_0.$$

În aceste condiții întîlnim așa-numitul spor nedescompus. Numeroase aplicații s-au realizat în ultimele două decenii în acest domeniu, ținînd seama în special de necesitățile practice. O expunere amplă a problemei aparține lui *L. Törnissi* (1982).

Metoda relativă a indicelui. Metodă constînd din construirea unui i. sintetic pe baza rapoartelor individuale. Termen folosit în statistica internațională (în engl. : *Relative method*).

Metoda agregată a indicelui. Metodă constînd din construirea unui i. sintetic prin raportarea *agregatelor**. Termen folosit în statistica internațională (în engl. : *Aggregative method*).

Raportul dintre cele două metode (pe exemplul retribuției), în care : r — retribuția (pe un muncitor) și T — numărul muncitorilor.

	Metoda „relativă"	Metoda „agregată"
Neponderat	$\Sigma \frac{r_1}{r_0}$	$\frac{\Sigma r_1}{\Sigma r_0}$
Ponderat	$\frac{\Sigma \frac{r_1}{r_0} T}{T}$	$\frac{\Sigma r_1 T}{\Sigma r_0 T}$

Expresia de „relativ" („mărime relativă") a fost folosită inițial pentru „prețul relativ", ca i. individual al prețurilor.

INDICE AL PREȚURILOR. I. care caracterizează variația prețurilor unitare. I. *individual* :

$$I_{1/0}(p) = \frac{p_1}{p_0} ;$$

I. *agregat* :

$$I_{1/0}(p) = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} \text{ și } \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0};$$

I. *armonic mediu* :

$$I_{1/0}(i_p) = \frac{\Sigma p_1 q_1}{\Sigma \frac{1}{i_p} p_1 q_1}.$$

Cu sistemul de ponderi adoptat i. armonic este egal cu i. agregat (ponderi din perioada curentă), adică i. Paasche.
În România i. prețurilor folosit oficial este i. Paasche.
Prețurile pot fi : de cost, de producție, cu ridicata, cu amănuntul, de unde o gamă largă de i. ai prețurilor.
Preț absolut. Prețul în expresia sa concretă într-o perioadă oarecare.
Preț relativ. Raport între valorile unui preț absolut în două perioade (π) :

$$\pi = \frac{p_1}{p_0}.$$

P.r. este un număr fără dimensiune. Pe această bază se calculează media aritmetică simplă :

$$I = \frac{1}{m} \Sigma \left(\frac{p_1}{p_0} \right).$$

În statistica din România nu se calculează asemenea tipuri de i. ; ei sînt în toate cazurile i. ponderați.

INDICE AL COSTULUI VIEȚII. I. al prețurilor cu amănuntul și al tarifelor serviciilor cumpărate de populația totală, de o categorie oarecare (personal muncitor, funcționari, pensionari). Este deci un i. particular al prețurilor. Întrucît i.c.v. servește la caracterizarea evoluției unui indicator important al nivelului de trai și este folosit practic pentru asigurarea unei corelații cu salariul (retribuția), se pun probleme deosebite în ce privește nomenclatura produselor, observarea prețurilor, perioada de bază și sistemul de ponderare. În general, drept ponderi se folosesc cheltuielile medii curente ale unei familii mijlocii pentru cumpărarea articolelor din fiecare grupă. **I.c.v.** este, în ultimă analiză, un i. general compus din mai mulți i. parțiali (de grup).

INDICE AL PREȚURILOR LA CONSUMATOR. (→ *Indice al costului vieții*).
COEFICIENT BUGETAR. Ponderea relativă a unui grup de articole și servicii în totalul cheltuielilor unei familii. Numărul articolelor, de ordinul sutelor, se întrunesc pe grupe mari.

În România, structura cheltuielilor totale de consum ale familiilor de personal muncitor și de țărani cooperatori a fost, în 1981, următoarea :

	Personal muncitor	Țărani cooperatori
Cheltuieli totale de consum	100,0	100,0
1. Alimentație și băuturi	46,8	62,9
2. Îmbrăcăminte și încălțăminte	17,6	14,2
3. Locuință	10,2	8,6
4. Înzestrare cu bunuri și cheltuieli pentru întreținere	7,5	4,6
5. Medicamente și îngrijire medicală	0,9	0,7
6. Transport și telecomunicații	8,1	4,0
7. Cultură și educație	5,4	2,8
8. Alte cheltuieli	3,5	2,2

Sursa: *Anuarul statistic al Republicii Socialiste România 1982*, p. 51.

Pentru fiecare grupă se construiește un i. parțial (de grup). (→ și *Istoric.*)

INDICE CU UTILITATE CONSTANTĂ. Formă specială a *indicelui costului vieții* *, care, în general, are drept obiect măsurarea modificărilor ce s-au produs în costul vieții unei persoane sau al unui grup de persoane cu gusturi identice pentru bunuri. O utilitate u este asociată cu orice combinație de cantități de bunuri care se ia în considerație. Persoana este presupusă a fi bine adaptată la situația dominantă a prețurilor, în așa fel încît, dat fiind venitul ei, persoana respectivă alege un set de cantități care-i dau nivelul cel mai înalt de utilitate. Pentru fiecare punct de timp t, definind un set de prețuri, există pentru fiecare nivel al utilității u o cheltuială $\mu(t, u)$ care este cheltuiala cea mai scăzută pentru a realiza pe u.

La punctul de timp t_0 se observă opțiunea persoanei pentru cantitățile și cheltuielile $\mu(t_0, u_0)$. Aceleași observații se fac la punctul t_1. Dacă se poate afi ma că pentru persoana respectivă $u_0 = u_1$, i. prețurilor pentru acest nivel de utilitate este :

$$P_{1/0}(u_0) = \frac{\mu(t_1\, u_0)}{\mu(t_0\, u_0)}.$$

I. se numește uneori „i. definit de indiferență" și a fost propus de statisticianul sovietic A. A. Konius (1924).

INDICE AL VOLUMULUI FIZIC AL PRODUCȚIEI (*Indice al cantităților*). I. care caracterizează variația cantităților de produse sau mărfuri.

I. individual :

$$I_{1/0}(q) = \frac{q_1}{q_0}$$

I. apregat :

$$I_{1/0}(q) = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}$$

I. aritmetic mediu :

$$I_{1/0} = \frac{\Sigma i_q\, q_0\, p_0}{\Sigma q_0\, p_0}.$$

În acest sistem de ponderare, i. aritmetic mediu este egal cu i. agregat. În România **I.v.f.p.** se calculează după formula Laspeyres.

FOARFECELE PREȚURILOR. Raport dintre i. produselor agricole și așa-numitul i. de paritate; acesta din urmă se calculează ca raport dintre i. prețurilor produselor agricole vîndute la punctele primare de desfacere și i. de paritate al prețurilor plătite de agricultori cumpărate de aceștia, plus dobînzile, impozitul funciar și impozitul pe salariile plătite muncitorilor agricoli. I. se calculează în numeroase țări pentru a caracteriza nivelul de trai al producătorilor agricoli.

INDICE AL PUTERII DE CUMPĂRARE A MONEDEI. I. ce caracterizează schimbarea masei de bunuri și servicii care poate fi achiziționată cu aceeași cantitate de bani; reprezintă mărimea inversă a i. prețurilor:

$$i_{p.c} = \frac{1}{I_p} = \frac{\Sigma q_1 p_0}{\Sigma q_1 p_1}.$$

RACORDAREA INDICILOR. Procedeu folosit în cazul unei *serii de indici** pe o perioadă mare de timp și care, în esență, constă din aplicarea proprietății i. în lanț. Operația se mai numește „înnodare" a i. cu bază fixă cu i. în lanț. De pildă, i. volumului fizic al producției, care a folosit pentru cîțiva ani anumite prețuri comparabile și trece la alte prețuri comparabile, va fi racordat, în cadrul seriei de i., astfel:

$$I_q^{T/0} \approx \frac{\Sigma qhp_I}{\Sigma q_1 p_I} \cdot \frac{\Sigma qhp_{II}}{\Sigma qhp_{II}} \cdot \ldots \frac{\Sigma q_T p_N}{\Sigma q_{T-N} p_N},$$

în care: p_I, $p_{II}, \ldots$, p_N sînt sisteme de prețuri comparabile anterioare $1, \ldots, h$; $h+1, \ldots, h+k$; $T-N, \ldots, T$.
Această operație are un caracter aproximativ.

BIBLIOGRAFIE

1. BĂRBAT, AL., *Teoria statisticii sociale*, Editura didactică și pedagogică, București, 1972.
2. BORTKIEWICZ, L. v., *Zweck un Struktur einer Preis-indexzahl*, in „Nordisk Statistik Tidskrift", B, 2, 1923; B. 3, 1924, B. 4, 1925.
3. BULGARU, MIRCEA, *Statistica ramurilor neindustriale. Metode și tehnici de calcul și analiză statistică*, Editura didactică și pedagogică, București, 1982.
4. CALOT, G., *Cours de statistique descriptive*, Deuxième édition, Dunod, Paris, 1973.
5. CETVERIKOV, N. S., *Statisticeskie i stohasticeskie issledovaniia*, Gostatizdat, Moscova, 1963. (capitolul: Metoda Index Numbers ca mijloc de studiere a modificării valorii banilor).
6. CROXTON, F. E., COWDEN, D. J., KLEIN, S., *Applied General Statistics*, Third Edition, Prentice-Hall, Inc. Englewood Cliffs, N. Y., 1967.
7. DUON, GASTON, *De la théorie à la pratique des indices statistiques*, Application aux problèmes économiques, sociaux et démographiques, Ed. Eyrolles Gauthier-Villars, Paris, 1955.

V.TREBICI

8. DUMITRESCU, V. V., *Preţurile şi indicii de preţuri pe principalele ramuri*, în, „Revista de statistică", nr. 2, 1972.
9. DUMITRESCU, VASILE V., *Elaborarea unui sistem integrat de indici pentru exprimarea venitului naţional în preţuri constante*, Teză de doctorat, A.S.E., Bucureşti, 1977.
10. EDGEWORTH, F. Y., *Defense of Index Numbers*, în Economic Journal, 1895.
11. FALCAN-POPESCU, SIMINA, *Probleme teoretice şi metodologice ale statisticii teritoriale*, în „Revista de statistică" nr. 9, 1964.
12. FISHER, IRVING, *The making of index numbers*, 1922.
13. FOURASTIÉ, JACQUELINE, *Les formules d'indices de prix*, Armand Colin, Paris, 1966.
14. FOURASTIÉ, JACQUELINE, *Les indices statistiques. Principes élémentaires*, Dunod, Paris, 1969.
15. GILBERT, M., KRAVIS, I., *Etude comparative des produits nationaux et au pouvoir d'achat des monnaies*, OECE, Paris, 1955.
16. GEORGESCU-ROEGEN, N. ST., *Metoda statistică. Elemente de statistică matematică*, Ed. Institutului Central de Statistică, Bucureşti, 1933.
17. INTERNATIONAL ENCYCLOPEDIA OF THE SOCIAL SCIENCES, David L. Sills (ed.), The MacMillan Company & The Free Press, Vol. 7, 1968.
18. JICA, P., *Consideraţii asupra noţiunii şi formelor productivităţii muncii*, în „Revista de Statistică", nr. 9 şi 10, 1970.
19. KAZINEŢ, L. S., *Teoria indeksov (Osnovnîie voprosi)*, Gostatizdat, Moscova, 1963.
20. KENDALL, M. G., BUCKLAND, W. R., *Dictionary of Statistical Terms*, Oliver & Boyd, London, 1957.
21. KONIUS, A. A., *The Problem of the True Index of the Cost of Living*, în „Econometrica" nr. 7, 1939 (iniţial în 1924 în limba rusă).
22. LÉVY, M. L., *Comprendre les statistiques*, Editions du Seuil, 1979.
23. MIHOC, GH., URSEANU, V., *Matematici aplicate în statistică*, Editura Academici Republicii Socialiste România, Bucureşti, 1962.
24. MILLS, F. C., *Statistical Methods*, Columbia University, New York, 1955 (trad. de uz intern D.C.S., 1959).
25. MITCHELL, W. C., *The Making and Using of Index Numbers*, 3d. ed., U. S. Bureau of Labor Statistics, Bulletin no. 656, Washington, 1938.
26. MOINEAGU, C., NEGURĂ, I., URSEANU, V., *Statistica*, Editura ştiinţifică şi enciclopedică, Bucureşti, 1976.
27. MORICE, F., CHARTIER, F., *Méthode statistique*, INSEE, Paris, 1954.
28. MOUCHEZ, PHILIPPE, *Les indices de prix*, Cujas, Paris, 1961.
29. MUDGETT, B. D., *Index Numbers*, New York, 1951.
30. PAENSON, I., *Glossaire systématique anglais-français-espagnol-russe de la terminologie des méthodes statistiques*, Pergamon Press, Oxford-New York-Toronto.Sydney-Braunschweig, 1970.
31. STATISTICESKII SLOVAR, A. I. Ejov (sub red.) Statistika, Moscova, 1965.
32. TÖVISSI, L., *Probleme teoretice ale calculului indicilor statistici*, Teză de doctorat, A.S.E., 1964.
33. TÖVISSI, L., ISAIC MANIU, AL., ANTONESCU C., *Statistica*, Curs pentru Facultatea Economia Industriei, Construcţiilor, Transporturilor. Pentru uzul studenţilor A.S.E., Bucureşti, 1982.

V.TREBICI

34. THEIL, H., *Best Linear Index Numbers of Prices and Quantities*, în *„Econometrica"* nr. 28, 1960.
35. YEOMANS, K. A., *Statistics for the social scientist*: Volume one, *Introducing Statistics*, Penguin Books, 1968.
36. YULE, G. U., KENDALL, M. G., *Introducere în teoria statisticii*, Editura ştiinţifică, Bucureşti. 1969.
37. *DICŢIONAR STATISTIC ECONOMIC* (sub red. C. Ionescu), D.C.S., Bucureşti, 1969.

XI. SERIILE CRONOLOGICE

SERIE CRONOLOGICĂ (*serie în timp, serie temporală, cronică*). Şir de valori pe care le înregistrează, la momente sau la intervale de timp succesive, o variabilă de definiţie constantă; prin urmare, un şir de valori care arată variaţia volumului populaţiei statistice sau a nivelului caracteristicii în raport cu timpul (→VI., p. 123). Termenul de serie dinamică, adoptat prin anii '50 de statistica românească şi întîlnit uneori şi astăzi în manuale de statistică ar trebui abandonat din cauza caracterului său imprecis. N. Georgescu-Roegen folosea, în 1933, termenul de „serie în timp", corespunzător engl. *Time series*. Expresia folosită în literatura franceză este de „serie cronologică" pe care o adoptăm în lucrarea de faţă.
Notînd cu x valorile variabilei şi cu t cele ale timpului, s.c., notată cu $\{x_t\}$, se prezintă astfel: $x_1, x_2, \ldots x_n$, la momentele sau intervalele de timp $t_1, t_2 \ldots \ldots, t_n$. În formă tabelară, întîlnită în publicaţiile statistice:

Timpul (t)	Variabila (x)
t_1	x_1
t_2	x_2
⋮	⋮
t_n	x_n

Se mai foloseşte notaţia x pentru valorile timpului şi y pentru variabila considerată, şi uneori u; vom avea deci y_t sau u_t. În prezenta lucrare se folosesc notaţiile y (variabila considerată) şi x (timpul). În unele cazuri particulare se va folosi y, deci y_t.
Unitatea de timp cel mai des folosită este anul; sînt cazuri cînd se adoptă ca unităţi de timp: ziua, săptămîna, luna, trimestrul, perioade cincinale. Pentru a pune în evidenţă particularităţile unor fenomene (demografice, meteorologice) se adoptă ca unităţi de timp ora şi chiar minutul.
Volumul populaţiei statistice poate fi: numărul populaţiei unei ţări sau unităţi teritorial-administrative, volumul producţiei industriale sau agricole, iar nivelul caracteristicii — preţul, productivitatea muncii, natalitatea, recolta medie la hectar.

Analiza seriilor cronologice. Operaţie complexă constînd din aplicarea metodelor statistico-matematice asupra s.c. empirice în vederea identificării *ten-*

dinței generale * de dezvoltare și a altor particularități ale acestora. Pregătirea **s.c.** pentru analiză cere calculul unor indicatori sintetici și aplicarea unor metode *de ajustare a s.c.**. *A. s.c.* reprezintă unul din capitolele cele mai dinamice ale statisticii contemporane, cu numeroase aplicații în alte domenii. (→ *Econometrie*; *Conjunctură economică*.)

Valorile variabilei ca funcție de timp pot fi fixate la un anumit moment de timp (la data recensământului, la sfîrșitul sau la începutul anului, la mijlocul anului etc.) sau să se refere la un interval de timp (lună, an etc.)

Serie cronologică de momente. Seric de observații ordonate în timp, exprimînd *stocuri*: efectivul unei populații, valoarea de inventar a fondurilor fixe la sfîrșitul anului etc. O asemenea **s.c.** mai este numită *discretă* sau discontinuă, deoarece variabila este considerată discretă, iar valorile sînt fixate la date echidistante.

Tabelul 39

Fondurile fixe în economia națională a României în anii 1970—1981 (în valoare completă de inventar, la sfîrșitul anului) — miliarde lei

Anul (x)	Valoarea fondurilor fixe (y)
1970	757
1971	825
1972	893
1973	977
1974	1075
1975	1203
1976	1325
1977	1454
1978	1521
1979	1719
1980	1864
1981	2019

Sursa: *Anuarul statistic al R. S. România 1976*, p. 56; *1981*, p. 97 și *1982*, p. 44.

O asemenea **s.c.** nu se poate cumula, deoarece volumul fondurilor fixe la sfîrșitul anului 1970 se regăsește în cea mai mare parte în volumul lor la sfîrșitul anului 1971 etc. Deci însumarea nu are sens din punctul de vedere al conținutului; ea este permisă pentru considerente de calcul, pentru ajustări etc. *S.c.m.*, cel mai adesea întîlnite, sînt cele ce reprezintă numărul populației, al personalului muncitor, al parcului de tractoare și mașini agricole etc., la o anumită dată.

Serie cronologică de intervale. Serie de observații ordonate în timp, exprimînd *fluxuri*: producția industrială sau agricolă pe un an, numărul de născuți vii într-o lună sau un an, cantitatea de precipitații într-o zi sau într-o lună etc. O asemenea **s.c.** mai este numită continuă, deoa-

rece variabila este continuă și poate lua un număr infinit de valori în limitele intervalului (lună, an etc.).

Tabelul 40

Producția de tractoare a României în anii 1971—1981
(în bucăți)

Anul (x)	Producția (y)	
	Anuală	Cumulată
1971	30400	30400
1972	34883	65283
1973	38800	104083
1974	44550	148633
1975	50003	198636
1976	53911	252547
1977	59306	311853
1978	65715	377568
1979	62494	440062
1980	70873	510935
1981	68093	579028

Sursa: *Anuarul statistic al R. S. România, 1976*, p. 117; *1980*, p. 185 *și 1982*, p. 94.

O asemenea s.c. se poate cumula: producția totală de tractoare (Σy) în anii 1971—1981 a fost de 579028, ceea ce revine în medie anuală la 52699 bucăți: $\frac{\Sigma y}{n} = \frac{579028}{11}$.

S.c. poate fi privită drept repartiția unei populații sau unui nivel după o singură caracteristică: timpul. De exemplu, seria anuală a producției de tractoare corespunde repartiției numărului de tractoare fabricate pe ani: populația statistică este mulțimea de tractoare, iar modalitățile caracteristicii sînt înlocuite prin unitățile de timp (ani). În general, se preferă totuși examinarea s.c. ca o repartiție după două caracteristici: nivelul fenomenului și timpul. Timpul devine atunci o veritabilă variabilă în funcție de care se analizează evoluția celeilalte variabile — nivelul fenomenului.

Un model care explică felul în care se dezvoltă seria y_t se numește *proces* și se notează cu $\{y_t\}$. Dacă procesul are anumite regulă probabiliste el este *proces aleator* sau *stohastic* (→ XII. p. 314).

Între repartițiile de frecvențe și s.c. există o serie de asemănări și de deosebiri. Pentru repartițiile de frecvențe ordinea cronologică nu are importanță: analiza acestora este o *analiză statică*. Pentru s.c., elementul principal este variația în timp, iar analiza devine o *analiză dinamică*. Din punctul de vedere al factorilor, se precizează că o repartiție de frecvențe este supusă integral influenței factorilor aleatori, în timp ce s.c. este influențată — uneori în proporție hotărîtoare — de factori nealeatori specifici. Majoritatea observațiilor ordonate după caracteristica timp nu sînt independente între ele și de aceea calculele probabiliste care se bazează pe ipoteza independenței reciproce nu sînt aplicabile s.c. Metodele inferenței statistice sînt mai puțin dezvoltate în domeniul s.c.; în schimb calculele sînt mult mai complicate.

Interdependența dintre observațiile succesive, caracteristice **s.c.**, ne duce la noțiunea de *autocorelație* *, pe care nu o întîlnim în analiza seriilor în care observațiile sînt independente. Într-o **s.c.** există corelație între y_0 și y_t.

Medie cronologică. Medie aritmetică particulară (→ *Medie aritmetică*), calculată pentru **s.c.** de moment după formula :

$$\bar{y} = \frac{\frac{1}{2}y_1 + y_2 + \ldots + \frac{1}{2}y_n}{n-1},$$

ce se mai poate scrie :

$$\bar{y} = \frac{\frac{y_1+y_2}{2} + \frac{y_2+y_3}{2} + \cdots + \frac{y_{n+1+n}}{2}}{n-1}.$$

Pentru **s.c.** de intervale media seriei se calculează după formula :

$$\bar{y} = \frac{\Sigma y}{n}.$$

Medie mobilă. Media aritmetică particulară, calculată din suma termenilor unei **s.c.**, în care numărul termenelor se deplasează sistematic.

Fie o funcție y de timp și y_t valoarea lui y corespunzătoare datei t ($t = 1,2 \ldots$). Se numesc *m.m.* pentru p luni ale funcției y mediile succesive ale funcției y calculate pentru p luni succesive și raportate la mijlocul perioadei :

$$\bar{y}_p = \left(t + \frac{p+1}{2} \right) = \frac{1}{p} \sum_{y=1}^{0} y_{t+k} = \frac{1}{p} \left(y_{t+1} + y_{t+2} + \ldots + y_{t+p} \right).$$

Data $t + \frac{p+1}{2}$ se află la mijlocul perioadei cuprinse între datele $t + 1$ și $t + p$.

M.m. este folosită în special la determinarea *variației sezoniere* *. Numărul de termeni luați în calcul diferă : 3, 5, 12 termeni. Există metode speciale în acest sens (de exemplu, metoda lui Spencer cu 15 și 21 termeni).

S.c. au ca domeniu principal fenomenele economice și demografice ; treptat, aplicațiile au cuprins și alte fenomene, ca cele meteorologice, psihologice etc. Cele mai multe metode și modele privind **s.c.** au fost elaborate de economiști, demografi și actuari ; în ultimul timp însă contribuția matematicienilor devine tot mai importantă, în legătură cu analiza proceselor stohastice. (→ *Istoric*).

Procesul complet de tratare a unei **s.c.** presupune mai multe faze și operații :

a) Premisa fundamentală este obținerea materialului statistic și sistematizarea acestuia sub formă de **s.c.** Perioada de timp poate fi : luna, anul, perioade cincinale. Pentru analiza unor fenomene este important să se asigure informația statistică pe o perioadă de timp cît mai mare : un secol și mai mult, în special pentru punerea în evidență a *tendinței seculare* * ;

b) O **s.c.** este supusă mai întîi observației de descriere, realizată cu mijloacele *statisticii descriptive* ; un loc important îl are reprezentarea grafică a **s.c.**, efectuată cu *historiograma* * sau *cronograma* * (fig. 15 și fig. 16 ne oferă două exem-

ple de reprezentare grafică a s.e. Există o deosebire între historiograma unei s.e. de momente și a uneia de intervale.
Prima este ilustrată de fig. 71 b, cea de intervale de fig. 71 a.

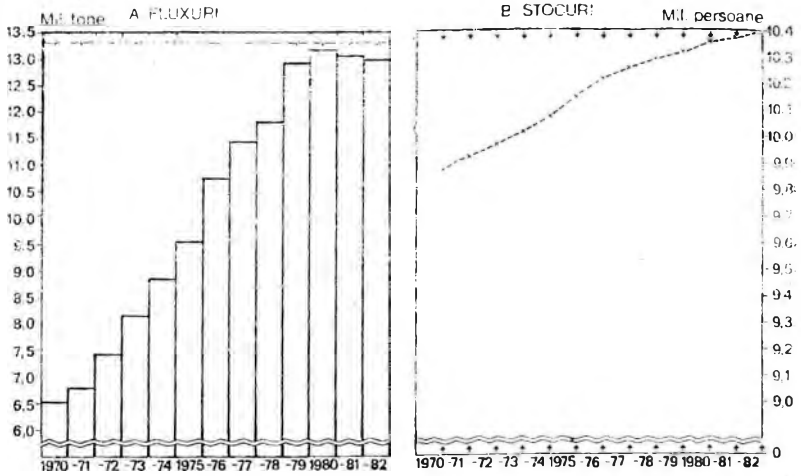


Fig. 71 a Producția de oțel (mil. tone) 1970—1982
Fig. 71 b Populația ocupată (mil. persoane, la sfîrșitul anului) 1970—1982

Se remarcă că s.e. de intervale are drept historiogramă o reprezentare asemănătoare cu histograma *, fiind vorba de intervale. Ea poate fi însă transformată într-o reprezentare de tipul *poligonului de frecvențe* *, așa cum se practică obișnuit :
c) S.e. este supusă apoi analizei, cu ajutorul diferitelor metode și modele statistico-matematice, cu care prilej se realizează *descompunerea seriei cronologice* *, identificarea *componentelor serici cronologice* *. operațiile cele mai importante fiind *ajustarea, interpolarea* și *netezirea*. Analiza s.e. este partea centrală a teoriei s.e. ;
d) ca urmare a analizei, s.e. devine aptă pentru prospectivă și prognoză, pe bază de *extrapolare a tendinței* *. Legătura dintre aceste faze și complexitatea analizei pot fi ilustrate cu exemplul oferit de fig. 16 în care sînt reprezentate evoluția natalității și mortalității populației României (1871—1980) și evoluția numărului populației (1877 — 1980). Dacă ne limităm la evoluția ratei brute de mortalitate (număr de decese la 1000 locuitori) pe o perioadă de 110 ani constatăm că *tendința generală** este cea de scădere ; aceasta este însă estompată de anumite variații anuale. De pildă, în anii imediat următori celor două războaie mondiale, mortalitatea a cunoscut o creștere, sub influența unor factori specifici. Mai mult, dacă dispunem și de valori lunare, constatăm o variație în cuprinsul anului care poate exprima o influență sezonieră sau periodică. Fig. 82 ne arată variația lunară a căsătoriilor în mediul rural 1957—1966, iar fig. 87 variația lunară a ratei totale de fertilitate 1971—1980. La alte s.e. se poate observa, pe fondul unei tendințe generale, o *mișcare*

ciclică *, alături de o *mișcare sezonieră* *. De asemenea, dacă urmărim o s.c. împreună cu o alta putem să identificăm o corelație, care mai poate fi exprimată sub formă de *corelație cu întîrziere*. De asemenea, s.c. le este propriu fenomenul de *autocorelație* *.

Istoric. Începuturile analizei moderne a s.c. datează din cea de-a doua jumătate a sec. al XIX-lea și primele două decenii ale sec. al XX-lea, fiind legate de numele economiștilor Clement Juglar (1819—1905), St. W. Jevons (1835 — 1883) și H. L. Moore care au încercat să evidențieze o regularitate ciclică a oscilațiilor fenomenelor economice. Dezvoltarea analizei s.c. a fost considerabil influențată de preocupările privind *ciclul economic**, studiul *conjuncturii economice* *, *teoria indicilor* *, *econometria* *, iar, mai recent, prospectiva și prognoza. Marile progrese recente sînt legate și de aplicarea metodelor de simulare electronică și de modelare economică. O contribuție de seamă se datorează Școlii economice de la Harvard și în special economiștilor *Warren M. Persons* * și Wesley Clair Mitchell (1874—1948). La Universitatea din Harvard ia ființă, în 1917, Comitetul de Cercetări Economice (Harvard Committee on Economic Research), care editează „Harvard Business Barometer" sau „Business Index". Studiul conjuncturii economice (pe termen scurt) solicita metode de analiză a s.c. W. C. Mitchell publică în 1913 „Business Cycles", iar W. M. Persons publică în 1919 „Indices of Business Conditions" și „Correlation of Time Series" (capitol în „Handbook of Mathematical Statistics", by H. L. Rietz, 1924). Potrivit teoriei lui W. M. Persons, puteau fi determinate patru tipuri de schimbări în s.c. : 1) schimbări care aveau loc în decursul fiecărui an ca rezultat al succesiunii anotimpurilor ; 2) schimbări datorate unei mișcări de durată sau tendința seculară care cuprinde un interval mare de ani ; 3) schimbări sau oscilații nesistematice ; 4) oscilații ciclice sau în formă de unde care puteau fi periodice sau neperiodice și care erau legate de perioadele de avînt sau de depresiune economică. Clasificarea lui W. M. Persons a rămas valabilă pînă în prezent. Alte contribuții se datorează lui A. Aftalion (1874 — 1956) ; Lucien March, în 1905, a propus termenul de covariație pentru s.c paralele și cel de covariație cu întîrziere ; K. Wiskell (1851—1926) a clasificat procesele cumulative ale ciclurilor economice prin analiza discrepanței dintre rata randamentului investițiilor și rata comercială a dobînzii. Odată cu constituirea *econometriei* * (în jurul anilor 1930), analiza s.c. intră într-o nouă fază. Sînt de menționat contribuțiile economiștilor R. Frisch, S. Kuznets, J. Tinbergen. În 1933, R. Frisch publică studiul „Propagation Problems and Impulse Problems in Dynamic Economics", J. Tinbergen dă la iveală, în 1935, studiul „Annual Survey : Quantitative Business Cycle Theory", iar în anii 1938—1939 publică „Statistical Testing of Business Cycle Theory" (2 vol. Geneva). Economistul polonez M. Kalecki (1899—1970) publică, în 1935, „A Macrodynamic Theory of Business Cycles". Economiști precum J. Schumpeter (1883—1956) și J. M. Keynes (1883—1946), prin analizele lor asupra ciclurilor economice, dau un impuls studiilor privind s.c. În istoria economică se consideră că cei ce au avut cele mai importante contribuții la teoria ciclurilor economice și la analiza lor statistică sînt W. M. Persons, S. Kuznets și J. Tinbergen. Li se adaugă, cu o contribuție însemnată, Kondratieff, cu ale sale „Cicluri mari", potrivit cărora din 1790 pînă în 1940 s-au observat trei cicluri mari : 1792—1850, 1850—1896 și 1896—1940. În perioada postbelică intervin elemente noi în analiza s.c. J. Tinbergen și L. R. Klein aplică metode noi în analiza s.c. („Economic Fluctuations in the United States", 1950) ; S. Kuznets publică, în 1961, „Capital in the American Economy : Its

Formation and Financing". Se dezvoltă *analiza spectrală* *; G. M. Jenkins publică, în 1957, „The Spectral Analysis of Time Series"; J. Johnston, în 1963, publică „Econometric Models". Autocorelației s.c. îi sînt consacrate studii printre care cele ale economiștilor Orcutt și James (1948), problemele de testare a tendinei generale sînt tratate în studiile lui H. Wold, T. W. Anderson, H. Theil, G. Tintner. Un moment important în istoria s.c. îl reprezintă studiile lui Bartlett (1935 și 1955) privind procesele stohastice.

O dezvoltare importantă cunoaște analiza s.c. în legătură cu preocupările de prognoză și planificare. Este de semnalat lucrarea fundamentală a lui G. E. P. Box și G. M. Jenkins, „Time Series Analysis, Forecasting and Control" (1970), ale căror metode au devenit clasice. L. H. Koopmans ne dă „The spectral analysis of time series" (1974), în timp ce G. Calot (1981) propune metode noi de analiză a variațiilor lunare ale i. demografici și de previziune a acestora pe termen scurt. Lui O. D. Anderson îi datorăm monografia „Time Series Analysis and forecasting — the Box — Jenkins approach" (1977). S-au organizat, în ultimii ani, sesiuni și seminarii internaționale consacrate analizei s.c. și s-au publicat numeroase monografii.

La dezvoltarea analizei s.c au contribuit actuari, biometricieni și demografi precum: P. Gompertz, W. M. Makeham, R. Pearl, L. Reed, ale căror curbe sînt folosite pe scară largă: curba *Gompertz* *, curba *Gompertz-Makeham* *, *curba logistică a lui Pearl-Reed* *; alți actuari și demografi au contribuții în domeniul metodelor de ajustare, interpolare și netezire: Sprague, Glover, Greville, King, Beers, J. Spencer, folosite dealtfel și în domeniul repartițiilor de frecvență.

În România, primele preocupări privind s.c. ca și i. se datorează lui N. St. *Georgescu Roegen* *, cu remarcabile studii publicate în țară și în străinătate, începînd cu anul 1930. Probleme speciale de ajustare, cu aplicații la s.c., au fost tratate de *Gh. Mihoc* * și *O. Onicescu* *. În perioada postbelică sînt de semnalat studii referitoare la caracterizarea ritmurilor medii datorate lui Al. Froda, I. Marinescu, *V. Urseanu* * (1914—1983), L. Tövissi. Numeroase aplicații privesc metodele de prognoză și de planificare, ca o continuare a analizei s.c.

DESCOMPUNEREA SERIILOR CRONOLOGICE. Operație statistico-matematică de identificare și exprimare cantitativă a componentelor unei s.c.: *tendința generală* * sau *tendința seculară* *, *variația ciclică* *, *variația sezonală* * și *variația aleatoare* * sau *variația reziduală* *. Înainte de a aplica metodele analizei s.c., seria statistică respectivă trebuie verificată dacă îndeplinește, pentru întreaga perioadă de timp, condițiile de comparabilitate. Pentru a elimina efectul variației prețurilor se calculează și se aplică *indicele prețurilor* *; în vederea eliminării efectului modificării numărului populației indicatorii s.c. se calculează ca valori pe un locuitor; se ajustează s.c. în funcție de numărul de zile din fiecare lună, de numărul zilelor lucrătoare și nelucrătoare.

Tendință generală (tendință seculară). Mișcare lină și regulată a unui fenomen, în cadrul unei s.c., în decursul întregii perioade, reprezentînd o creștere sau o descreștere; se notează cu *T*.

Variație ciclică (mișcare ciclică). Oscilații (fluctuații) în cadrul s.c. care exprimă particularitățile *ciclului economic* *; se notează cu *C*.

V.TREBICI

Variație sezonieră (mișcare sezonieră). Caz particular de *variație periodică* *, reprezentînd oscilații (fluctuații) ale s.c. în funcție de anotimpuri, luni sau zile ; se notează cu *S*.

Variație periodică (mișcare periodică). Oscilații ale s.c. în funcție de perioadă.

Variație reziduală (variație aleatoare sau accidentală). Oscilații (fluctuații) ale s.c. în funcție de factorii aleatori sau întîmplători. Se notează cu *R*.

Terminologie. 1) Se apreciază că folosirea expresiei „trend" în locul „tendinței generale" și a expresiei „trend secular" în locul „tendinței seculare" nu este justificată, termenul de „tendință" fiind suficient de clar pentru semnificația ce-o reprezintă în cadrul analizei s.c. 2) Propunem introducerea substantivului „sezonalitate" — termen ce lipsește din DELR — în locul termenului perifrastic de „caracter sezonier". În mod corespunzător vom avea „indice desezonalizat" (din care s-a eliminat influența sezonală).

Model aditiv al seriei cronologice. Reprezentare cantitativă a raportului dintre valorile s.c. și componentele sale, presupunînd o relație de independență a factorilor :

$$A = T + C + S + R.$$

în care : A sau y_t (sau u_t) reprezintă valorile observate (empirice) la un moment dat. Deci A este suma celor patru componente ale s.c. De aici rezultă :

a) Seria din care se elimină tendința generală (seria cu tendința eliminată) :

$$A - T = S + C + R.$$

b) Serie ajustată în funcție de influența factorilor sezonali :

$$A - S = T + C + R.$$

Se presupune că S este factorul predominant și că $C + R = 0$.

c) Seria din care se elimină influența factorilor sezonali și a tendinței generale (serie desezonalizată și fără tendință) :

$$A - T - S = C + R.$$

d) Seria din care se elimină tendința generală, influența factorilor ciclici și sezonali rămînînd numai influența factorilor reziduali :

$$A - T - S - C = R.$$

Model multiplicativ al seriei cronologice. Reprezentare cantitativă a raportului dintre valorile s.c. și componentele sale, presupunînd o relație de proporționalitate a factorilor cu T :

$$A = T \times C \times S \times R.$$

Valoarea observată la un moment dat este produsul factorilor (componentelor). De aici rezultă :

a) Seria din care se elimină tendința generală :

$$\frac{A}{T} = S \times C \times R.$$

b) Serie ajustată din punctul de vedere al influenței factorilor sezonali :

$$\frac{A}{S} = T \times C \times R.$$

Se presupune că S este predominant, iar $C \times R = 1$.
c) Seria din care se elimină influența factorilor sezonieri și a tendinței generale :

$$\frac{A}{T \times S} = C \times R.$$

d) Seria cu factori reziduali :

$$\frac{A}{T \times S \times C} = R.$$

Schema este generală și nu se aplică în toate cazurile. Dacă datele s.c. sînt anuale, atunci nu se poate vorbi de o componentă sezonieră ; în cazul unei s.c. pe o perioadă scurtă de timp, componenta ciclică poate fi omisă.
AJUSTAREA SERIEI CRONOLOGICE. Aplicarea unor metode statistico-matematice și grafice adecvate asupra s.c. empirice în vederea evidențierii a ceea ce este esențial și tipic în dezvoltarea fenomenelor reprezentate de s.c. și care are caracter de lege sau regularitate. Ca rezultat se obține o serie de valori calculate ale s.c. care înlocuiesc seria empirică. Există numeroase metode ; aplicarea lor trebuie precedată de o analiză multilaterală în care competența, intuiția și experiența cercetătorului au o importanță deosebită. Metodele de a.s.c. se folosesc de asemenea pentru interpolarea și extrapolarea unei s.c.
În linii generale, metodele de a.s.c. se grupează astfel :
Metode de ajustare grafică a seriei cronologice. Metode ce constau în trasarea liberă și aproximativă a unei curbe de către cercetător asupra unei s.c. empirice ; accasta a fost în prealabil trecută pe un grafic, cu scară verticală aritmetică sau logaritmică. (→ *Historiogramă*). I se mai spune ,,tendință vizuală" (Gh. Mihoc și V. Urseanu, 1962), corespunzătoare expresiei ,,courbe tracée à main levée" (în franceză) și ,,free-hand curve" și ,,by inspection" (în engleză). A mai existat și expresia de ,,french curve" (în engl.), subliniindu-se preferința statisticienilor francezi pentru ajustarea grafică.
O asemenea ajustare are un caracter orientativ și ne oferă primele informații privind *tendința generală* * și alte particularități ale s.c.
Un exemplu de ajustare grafică a unei s.c. este prezentat în fig. 80, România. Țiței extras (mii tone) 1950—1982, din care se vede că extracția de țiței se încetinește anunțînd o asimptotă superioară de circa 14 mil. tone.
Metode de ajustare mecanică a seriei cronologice. Metodele constau din aplicarea succesivă, în mod mecanic, a unor formule de calcul stabilite dinainte, pentru toți termenii seriei. Amintim : *metoda mediilor mobile* *, metoda sporului mediu, metoda ratei medii de creștere și altele.

Exemplu. Ajustarea unei s.c. cu ajutorul sporului mediu.

Tabelul 41

România. Populația ocupată 1970—1982
(mii persoane)

Anul (x)	Numărul populației (y)	Numărul populației ajustat cu sporul mediu (Y)	Anul (x)	Numărul populației (y)	Numărul populației ajustat cu sporul mediu (Y)
1970	9 875,0	9 875,00	1977	10 264,3	10 181,25
1971	9 938,9	9 918,75	1978	10 290,0	10 225,00
1972	9 970,9	9 962,50	1979	10 320,0	10 268,75
1973	10 020,6	10 006,25	1980	10 350,1	10 312,50
1974	10 070,1	10 050,00	1981	10 375,3	10 356,25
1975	10 150,8	10 093,75	1982	10 400,0	10 400,00
1976	10 227,0	10 137,50			

$n = 13$

Sporul mediu se determină după formula:

$$\overline{\Delta} = \frac{\Sigma\Delta}{n-1} = \frac{525\,000}{12} = 43\,750 \text{ persoane},$$

în care: Δ este sporul anual $y_1 - y_0$; $y_2 - y_1, \ldots, y_n - y_{n-1}$
sau $\Sigma\Delta = y_n - y_0 = 10\,400{,}0 - 9\,875{,}0 = 525{,}0$, de unde:

$$Y_{1971} = y_{1970} + \overline{\Delta}; \quad Y_{1972} = y_{1970} + 2\,\overline{\Delta} \ldots$$

$$Y_{1982} = y_{1970} + 12\,\overline{\Delta}.$$

Ajustarea cu sporul mediu este liniară, de tipul progresiei aritmetice:
$Y_n = y_0 + n\overline{\Delta}$.
Exemplu. Ajustarea unei s.c. cu ajutorul ratei medii de creștere.

Tabelul 42

România. Producția de oțel 1970—1980 (mii tone)

Anul (x) (1)	Producția de oțel (y) (2)	Producția ajustată cu rata medie de creștere r (Y) (3)	Anul (x) (1)	Producția de oțel (y) (2)	Producția ajustată cu rata medie de creștere r (Y) (3)
1970	6 517	6 517	1976	10 733	9 940
1971	6 803	6 992	1977	11 457	10 665
1972	7 401	7 502	1978	11 779	11 443
1973	8 161	8 049	1979	12 909	12 277
1974	8 848	8 635	1980	13 175	13 175
1975	9 549	9 265			

$n = 11$

$$(1 + r) = \sqrt[n]{\frac{y_n}{y_0}}; \quad \log (1 + r) = \frac{\log \left(\frac{y_n}{y_0}\right)}{n}, \text{ de unde:}$$

$$y_n = y_0 (1 + r)^n.$$

În cazul nostru:

$$(1 + r) = \sqrt[10]{\frac{13175}{6517}} = 2{,}0216$$

$$\log (1 + r) = \frac{\log 2{,}0216}{10} = \frac{0{,}30569}{10} = 0{,}030569$$

antilog 0,030569 = 1,0729, de unde

$r = 0{,}0729 \approx 0.073$ sau 7,3%

$$Y_n = Y_0(1{,}0729)^{10}.$$

Datele obținute sînt trecute în col. 3 a tab. 42.

Formula este a dobînzii compuse avînd la bază progresia geometrică. (→ *Media geometrică*.) Dacă se consideră timpul ca funcție continuă atunci se folosește formula exponențială:

$$\frac{y_n}{y_0} = e^{rn}, \text{ de unde: } y_n = y_0 . e^{rn}.$$

De unde deducem pe r:

$$r = \frac{\log \left(\frac{y_n}{y_0}\right)}{n \log e}.$$

În cazul nostru, $r = \frac{\log 2{,}0216}{10 \times 0{,}4342945} = 0{,}0803877$ sau 0,0704; în procente 7,04% ≈ 7%.

Să se observe că rata medie de creștere obținută cu funcția exponențială este întotdeauna mai mică decît cea obținută cu media geometrică. În cazul nostru, 7,0% < 7,3%. (→ *Ajustarea seriei cronologice cu ajutorul exponențialei.*)

Să se remarce că ajustarea cu ajutorul ratei medii de creștere se bazează pe o formulă care ia în considerare numai primul și ultimul termen al s.c., ceea ce a generat numeroase critici la adresa acestei metode, dealtfel larg folosită. În cazul nostru $\Sigma y > \Sigma Y$: 107333 > 104461. Diferența este de — 2,7%.

Metode de ajustare analitică a seriei cronologice. Aplicarea metodelor statistico-matematice cu ajutorul cărora aproximarea valorilor teoretice ale s.c.

se face pe baza unor funcții matematice alese în așa fel încît să descrie cel mai bine evoluția fenomenului. Ajustarea cu aceste metode ia în considerare toți termenii seriei ; metoda cea mai folosită este *metoda celor mai mici pătrate*. (→ *Ecuația de regresie*.) Valorile s.e. se notează cu y, iar cele obținute prin ajustare cu Y.

Clasificarea metodelor de ajustare în cele 3 grupe, așa cum a fost prezentată mai sus, este doar una din clasificări. După alți autori, metodele de ajustare sînt : a) ajustare grafică, b) semimediei, c) medii mobile simple ; d) medii mobile ponderate exponențial ; e) curbe matematice ale tendinței.

Netezire a seriei cronologice. Operație cu ajutorul metodelor statistico-matematice și grafice, efectuată asupra unei s.c., prin care se obțin noi valori. În acest caz curba are o înfățișare lină, „netezită". Valorile pentru completarea curbei și respectiv a s.c. se obțin prin interpolare. Operația de netezire a curbelor se bazează pe ipoteza că variabilele pe care curbele le reprezintă sînt *continue*. Seria lină sau netezită, care ia locul seriei empirice neregulate, este considerată ca reprezentînd legea ce guvernează comportamentul datelor observate. Pentru *n.s.c.* se folosesc : metoda diferențelor, interpolarea grafică, interpolarea polinomială. Metodele amintite se aplică atît repartițiilor de frecvențe cît și s.c. Amintim cîteva metode : Sprague, Shovelton, Jenkins, Karup, King, J. W. Glover, Beers, T. N. E. Greville, majoritatea autorilor fiind actuari, iar metodele propuse referindu-se la repartițiile populației după vîrstă, a deceselor după vîrstă etc. Contribuții românești se datorează în special lui *Gh. Mihoc** și *O. Onicescu** în legătură cu calculele actuariale și demografice.

Terminologie. În literatura românească există o oarecare imprecizie în legătură cu noțiunile de „ajustare" și „netezire". O analiză comparativă ne va permite să precizăm noțiunile. În engl. există patru expresii : *to fit (fitting, fitted)*, *to smooth (smoothing, smoothed)*, *to graduate (graduating, graduated)*, *to adjust (adjusting, adjusted)*.

Prin *curve fitting* (ajustarea curbei) se înțelege ajustarea la datele observate, prin aplicarea unei formule, în vederea reliefării legii fenomenului reprezentat. Este deci un tip de interpolare în care ajustarea valorilor empirice este aproximativă ; scopul ei este de a stabili tendința generală (*trend*) și tipul sau modelul (*pattern*). Prin *smoothing* (netezire) se înțelege o interpolare avînd drept scop să obțină serii sau curbe line (netezite). Operația se bazează pe ipoteza că variabila este continuă. Termenul de *graduation* trebuie considerat un sinonim al expresiei de *smoothing*. Prin *adjustement* se înțelege, în general, o operație de corectare ; în numeroase cazuri ea folosește operații de *fitting* și *smoothing*.

În alte limbi, se folosesc numai două sau trei expresii. În limba română, N. Georgescu Roegen (1933) folosea atît expresia de „ajustare" cît și cea de „netezire". Gh. Mihoc (1943) folosea expresia de „ajustare" (grafică, mecanică, analitică) și pe cea de „interpolare", în sensul de „netezire". În lucrările apărute în perioada postbelică, expresia cel mai des întîlnită este cea de „ajustare" înțelegîndu-se prin aceasta și pe cea de „netezire". Gh. Mihoc și V. Urseanu (1962) folosesc expresiile de „ajustare" și „interpolare", această din urmă pentru „netezire" dînd totodată explicații detaliate în legătură cu sensul celor două noțiuni. Cu prilejul traducerii unor lucrări din statistica anglo-saxonă s-a reintrodus expresia de „netezire" (Yule și Kendall, 1969).

V.TREBICI

Tabelul comparativ de mai jos ne ajută la clarificarea terminologiei.

Engleză	Franceză	Rusă	Română
1. to fit fitting fitted	ajuster ajustement ajusté (e)	virovniati viravnivanie viravnennaia	a ajusta ajustare ajustat(ă)
2. to smooth smoothing smoothed	lisser lissage lisse	sglajivati sglajivanie sglajennaia (viravnennaia)	a netezi netezire netezit, lin
smoothed curve	courbe lisse	krivaia sglajennaia (viravnennaia) și krivaia plavnaia	curbă netezită sau lină
3. to adjust adjusting adjusted (for trend)	corriger correction corrigé de la tendance générale	popravka popravki na dvijenie vekovogo urovnia	a ajusta, a corecta corectare corectat (ajustat) pentru a ține seama de tendința generală

Expresia *graduating* rămîne să fie considerată drept „netezire". Vom reține deci pentru limba română expresiile de „ajustare", „netezire" și „corectare", în înțelesul lor din tabelul comparativ de mai sus.

Interpolarea seriei cronologice. Operație de completare a termenilor lipsă din cadrul unei s.c. între valorile existente. Dacă valorile unei funcții de o singură variabilă x, de exemplu y_x — în care x este timpul —, au fost tabelate pentru valori echidistante ale variabilei x, $x + h$, $x + 2h$ etc. se pune problema de a afla valori ale funcției corespunzătoare unei valori intermediare a variabilei. Să presupunem că avem numărul populației la 2 sau 3 recensăminte efectuate la intervale decenale și vrem să determinăm numărul populației în așa-numitele perioade intercenzitare. Funcția matematică pe care o alegem poate fi liniară sau neliniară. Interpolarea se efectuează cu diferite metode : a diferențelor, a diferențelor unui polinom, cu formula lui Newton, Lagrange etc. Dar nici o s.c. empirică nu variază absolut lin ca o funcție matematică, în afară de cazul în care seria a fost „netezită". Interpolarea poate servi deci ca o metodă de estimare a datelor adevărate fără însă să se substituie acestora.

ANALIZA TENDINȚEI GENERALE (SECULARE) A SERIEI CRONOLOGICE. Operație complexă constînd din aplicarea metodelor statistico-matematice asupra s.c. empirice în vederea identificării *tendinței generale (seculare)**, principala componentă a unei s.c. Metodele cele mai frecvent folosite în **a.t.g.** (s) sînt cele de *ajustare analitică**, cu aplicarea metodei celor mai mici pătrate. În acest scop se alege o funcție matematică, adaptată la specificul curbei empirice și se estimează parametrii necunoscuți pe baza datelor statistice. Cele mai des intîlnite funcții și, respectiv, curbe sînt : dreapta și polinoamele de diferite grade, din care vor fi prezentate cîteva în cele ce urmează.

Ecuația dreptei : $Y = a + bx$, în care : Y este variabila dependentă ce urmează să fie ajustată, x este timpul, a și b — parametri ce urmează să fie estimați (constante).

Curbele de grad mai ridicat sînt :
— curba de gradul doi (parabola) :

$$Y = a + bx + cx^2;$$

— curbă de gradul trei (cubică) :

$$Y = a + bx + cx^2 + dx^3 \text{ etc.}$$

Ajustarea seriei cronologice cu ajutorul dreptei. Ajustare folosind ecuația dreptei :

$$Y = a + bx. \tag{1}$$

Estimarea parametrilor a și b se face cu ajutorul metodei celor mai mici pătrate. Această metodă satisface următoarele condiții : a) suma abaterilor verticale ale valorilor observate (empirice) de la dreapta ajustată este egală cu zero ; b) suma pătratelor tuturor abaterilor este mai mică decît suma abaterilor verticale pătrate de la oricare altă dreaptă.
Constantele a și b se determină pe baza datelor observate obținîndu-se și rezolvîndu-se simultan două ecuații normale :

$$\text{I } \Sigma y = na + b\Sigma x \tag{2}$$

$$\text{II } \Sigma xy = a\Sigma x + b\Sigma x^2$$

Ideea de bază este ca $\Sigma(y - Y)^2 =$ minimum.

Tabelul 43

Determinarea ecuațiilor normale și a sumelor pentru ajustarea cu dreapta folosind metoda celor mai mici pătrate
(date convenționale)

x	y	Ecuația datelor observate $y = a + bx$	Determinarea primei ecuații normale		Determinarea celei de-a doua ecuații normale		xy	x^2
			Coeficientul lui a	Ecuația de observație înmulțită cu coeficientul lui a	Coeficientul lui b	Ecuația de observație înmulțită cu coeficientul lui b		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0	2	$2=a$	1	$2=a$	0	...	0	0
1	1	$1=a+b$	1	$1=a+b$	1	$1=a+b$	1	1
2	3	$3=a+2b$	1	$3=a+2b$	2	$6=2a+4b$	6	4
3	2	$2=a+3b$	1	$2=a+3b$	3	$6=3a+9b$	6	9
4	4	$4=a+4b$	1	$4=a+4b$	4	$16=4a+16b$	16	16
5	3	$3=a+5b$	1	$3=a+5b$	5	$15=5a+25b$	15	25
6	5	$5=a+6b$	1	$5=a+6b$	6	$30=6a+36b$	30	36
21	20	...	...	$20=7a+21b$	...	$74=21a+91b$	74	91

Cele două ecuații normale sînt :

$$\text{I } 20 = 7a + 21\ b$$

$$\text{II } 74 = 21a + 91b.$$

Rezolvarea simultană a ecuațiilor se realizează cu ajutorul metodelor algebrei. De pildă, înmulțim ecuația normală I cu 3 și rezultatul îl scădem din ecuația normală II, eliminînd astfel pe a :

$$\begin{array}{r} \text{II } 74 = 21a + 91b \\ (\text{I} \times 3)\ 60 = 21a + 63b \\ \hline 14 = 28b \end{array}$$

$$b = \frac{14}{28} = 0{,}5.$$

Pentru a obține pe a introducem valoarea lui b în ecuația normală I sau II. Dacă folosim ecuația I vom avea :

$$20 = 7a + 21\ (0{,}5) = 7a + 10{,}5$$
$$7a = 20 - 10{,}5 = 9{,}5$$

$$a = \frac{9{,}5}{7} = 1{,}357 \approx 1{,}36.$$

Ecuația dreptei ajustate va fi :

$$Y = 1{,}36 + 0{,}5x$$

$$Y_1 = 1{,}36 + 0{,}5\ (0) = 1{,}36$$

$$Y_2 = 1{,}36 + 0{,}5\ (1) = 1{,}86$$

$$Y_3 = 1{,}36 + 0{,}5\ (2) = 2{,}36$$

$$Y_4 = 1{,}36 + 0.5\ (3) = 2.86$$

$$Y_5 = 1{,}36 + 0{,}5\ (4) = 3{,}36$$

$$Y_6 = 1{,}36 + 0.5\ (5) = 3.86$$

$$Y_7 = 1{,}36 + 0{,}5\ (6) = 4{,}36.$$

Creșterea este liniară și este dată de valoarea lui $b = 0{,}5$: reprezentarea grafică este aceea a dreptei.

Din punct de vedere practic se pune problema numărului par și impar al termenilor.

Exemplu. Ajustarea unei s.c. cu ajutorul dreptei.

Tabelul 44

Populația ocupată a României (milioane persoane) 1970—1982
— la sfîrșitul anului —

Anul	Populația (y)	Timpul (x)	x^2	xy	$Y = a + bx = 9.454{,}046 + 45{,}611x$
1	2	3	4	5	6
1970	9 875,0	1	1	9 875,0	9 899,7
1971	9 938,9	2	4	19 877,8	9 945,3
1972	9 970,9	3	9	29 912,7	9 990,9
1973	10 020,6	4	16	40 082,4	10 036,5
1974	10 070,1	5	25	50 350,5	10 082,1
1975	10 150,8	6	36	60 904,8	10 127,7
1976	10 227,0	7	49	71 589,0	10 173,3
1977	10 264,3	8	64	82 114,4	10 218,9
1978	10 290,0	9	81	92 610,0	10 264,5
1979	10 320,0	10	100	103 200,0	10 310,2
1980	10 350,1	11	121	113 851,0	10 355,8
1981	10 375,5	12	144	124 506,0	10 401,4
1982	10 400,0	13	169	135 200,0	10 447,0
Σ	132 253,2	91	819	934 073,6	132 253,3

Precizăm că originea este anul 1970, iar $n = 13$.

Sistemul de ecuații normale:

$$na + b\Sigma x = \Sigma y$$

$$a\Sigma x + b\Sigma x^2 = \Sigma xy$$

Introducînd valorile obținute avem:

$13\ a + 91\ b = 132\ 253{,}2$

$91\ a + 819\ b = 934\ 073{,}6.$

Rezolvarea simultană a ecuațiilor normale:

$$a = \frac{\begin{vmatrix} \Sigma y & \Sigma x \\ \Sigma xy & \Sigma x^2 \end{vmatrix}}{\begin{vmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{vmatrix}} = \frac{\Sigma y \cdot \Sigma x^2 - \Sigma xy \cdot \Sigma x}{n\Sigma x^2 - \Sigma x \cdot \Sigma x}.$$

$$a = \frac{(132\ 253{,}2 \times 819) - (934\ 073{,}6 \times 91)}{(13 \times 819) - (91 \times 91)} = 9\ 853{,}046 \text{ mii persoane}$$

$$b = \frac{\begin{vmatrix} n & \Sigma y \\ \Sigma x & \Sigma xy \end{vmatrix}}{\begin{vmatrix} n & \Sigma x \\ \Sigma x & \Sigma x^2 \end{vmatrix}} = \frac{n\Sigma xy - \Sigma x\Sigma y}{n\Sigma x^2 - \Sigma x\Sigma x}.$$

$$b = \frac{(13 \times 934\ 073{,}6) - (91 - 132\ 253{,}2)}{(13 \times 819) - (91 - 91)} = 45{,}611 \text{ mii persoane},$$

de unde: $Y = 9\ 454{,}06 + 45{,}611\ x$. Datele ajustate sînt trecute în col. 6 a tab. 44.

Calculele pot fi simplificate dacă schimbăm originea seriei în așa fel încît $\Sigma x = 0$. Întrucît seria are $n = 13$, vom lua ca origine anul $1976 = 0$.
În aceste condiții sistemul de ecuații normale devine:

$$na = \Sigma y$$

$$b\Sigma x^2 = \Sigma xy,$$

de unde:

$$a = \frac{\Sigma y}{n} \quad \text{și} \quad b = \frac{\Sigma xy}{\Sigma x^2}.$$

Exemplu. Ajustarea unei s.c. cu ajutorul dreptei ($\Sigma x = 0$).

Tabelul 45

Populația ocupată a României (mil. persoane) 1970—1982
(la sfîrșitul anului)

Anul	Populația (y)	Timpul (x)	x^2	xy	$Y = 10\,173{,}3 + 45{,}611\,x$
(1)	(2)	(3)	(4)	(5)	(6)
1970	9 875,0	—6	36	—59 250	9 899,7
1971	9 938,9	—5	25	—49 694,5	9 945,3
1972	9 970,9	—4	16	—39 883,6	9 990,9
1973	10 020,6	—3	9	—30 061,8	10 036,5
1974	10 070,1	—2	4	—20 140,2	10 082,1
1975	10 150, 8	—1	1	—10 150,8	10 127,7
1976	10 227,0	0	0	0	10 173,3
1977	10 264,3	1	1	10 264,3	10 218,9
1978	10 290,0	2	4	20 580,0	10 264,5
1979	10 320,0	3	9	30 960,0	10 310,1
1980	10 350,1	4	16	41 400,0	10 355,8
1981	10 375,5	5	25	51 577,5	10 401,4
1982	10 400,0	5	36	62 400,0	10 447,0
Σ	132 253,2	0	182	8 301,3	132 253,3
	Σy	Σx	Σx^2	Σxy	ΣY

1976 = 0 ; $n = 13$

Sistemul de ecuații normale :

$$13\,a = 132\,253{,}2$$

$$182\,b = 8\,301{,}3,$$

de unde : $a = \dfrac{132\,253{,}2}{13} = 10\,173{,}3$; $b = \dfrac{8\,301{,}3}{182} = 45{,}611.$

Să observăm că a este un termen al seriei (10 173 300 persoane) în timp ce b reprezintă creșterea anuală (45 611 persoane) ; în reprezentare grafică b este panta dreptei.

Este suficient să dispunem de două valori ale seriei ajustate pentru a putea trasa treapta. Menționăm că la ajustarea cu ajutorul dreptei $\Sigma Y = \Sigma y$.

Examinînd curba empirică și cea ajustată din fig. 72 avem impresia — nu certitudinea ! — că ajustarea cu dreapta este potrivită pentru seria noastră. Perioada este prea scurtă pentru aprecierea tendinței generale. Observăm că în subperioada 1970—1974 creșterea este liniară ; în subperioada 1974—1977 ea este de tip exponențial ; în ultimii ani, creșterea absolută și relativă este

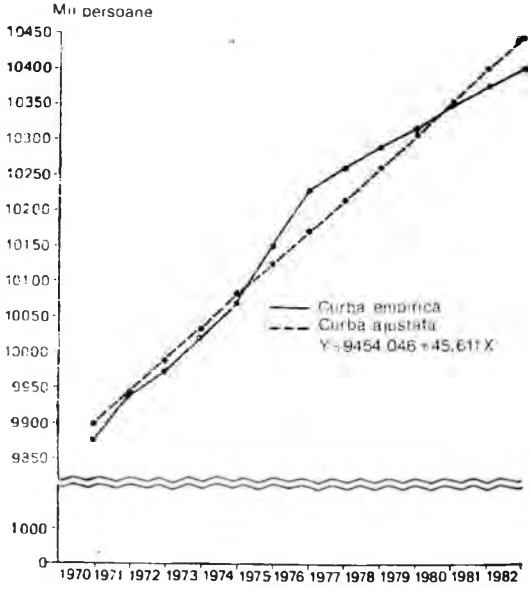


Fig. 72 România. Populația ocupată (mii persoane) 1970—1982 (date reale și date ajustate)

mai redusă. Asemenea constatări ridică două probleme : lungimea perioadei și alegerea curbei. Înainte de a le examina să determinăm pătratele abaterilor valorilor empirice de la valorile calculate.

Tabelul 46

Pătratele abaterilor valorilor y de la valorile Y (tab. 44 și tab. 45)

Anul	y	Y	$(y-Y)$	$(y-Y)^2$
1970	9 875,0	9 899,7	−24,7	610,09
1971	9 938,9	9 945,3	− 6,4	40,96
1972	9 970,9	9 990,9	−20,0	400,00
1973	10 020,6	10 036,5	−15,9	252,81
1974	10 070,1	10 082,1	−12,0	144,00
1975	10 150,8	10 127,7	23,1	533,61
1976	10 227,0	10 173,3	53,7	2 833,69
1977	10 264,3	10 218,9	45,4	2 061,16
1978	10 290,0	10 264,5	25,5	650,25
1979	10 320,0	10 310,1	9,9	98,01
1980	10 350,1	10 355,8	− 5,7	32,49
1981	10 375,5	10 401,5	−25,9	670,81
1982	10 400,0	10 447,0	−47,0	2 209,00
Σ	132 253,2	132 253,2	0	10 586,88

Deci $\Sigma(y - Y)^2 = 10\,586{,}88$, care este minimă în raport cu ecuația dreptei alese. Ar trebui încercate ajustări și cu alte funcții pentru a găsi pe cea mai corespunzătoare. Dealtfel, opt abateri sînt negative și cinci pozitive. Nu sîntem în situația privilegiată a repartițiilor de frecvențe unde *stabilitatea* este o caracteristică fundamentală și unde beneficiem de modele teoretice cu care se compară repartițiile de frecvențe (→ XIII. *Legi de repartiție.*)

Reamintind cît de importante sînt competența, intuiția și experiența cercetătorului la alegerea funcțiilor și curbelor. Statistica formulează următoarele recomandări generale, valabile nu numai pentru dreaptă ci și pentru alte curbe :

a) Pentru determinarea tendinței generale (seculare) este de dorit ca s.c. să se refere la o perioadă cît mai lungă de timp. Reprezentarea grafică a s.c. poate da o indicație în legătură cu caracterul tendinței generale.

b) Dacă din examinarea grafică a seriei rezultă anumite modificări semnificative, se impune folosirea a două tendințe generale.

c) La alegerea tipului de funcție pentru tendința generală se procedează astfel :

— datele empirice se trec pe un grafic și uneori tendința vizuală este suficientă ;

— dacă se observă că tendința nu este liniară, se alege una din curbele neliniare

d) Unele criterii formale există, cum se va vedea la ajustările după curbele. neliniare.

Un exemplu care ne arată importanța cunoașterii esenței proceselor poate fi dat în legătură cu tab. 44 și 45 și fig. 73. Populația ocupată a României în anii 1970—1982 a crescut în total cu 525 000 sau cu 43 750 în medie anuală.

Dar în timp ce în primii 6 ani creșterea medie anuală a fost de 58 634 de persoane, în următorii 6 ani creșterea medie anuală a fost de 28 834 sau aproximativ jumătate. În anii 1980—1982 creșterea a fost de circa 25 000 persoane anual. Or, pe de o parte, populația ocupată se alimentează din populația totală (populația în vîrstă de muncă) care impune așa-numita restricție demografică. Ca urmare a scăderii natalității, această resursă devine tot mai redusă.

Pe de altă parte, populația ocupată este funcție de crearea locurilor de muncă în economia națională care, la rîndu-i, este funcție de investiții. Trecerea la tipul de dezvoltare intensivă a economiei naționale va reduce necesarul de forță de muncă. Se poate deci aprecia că în aceste condiții creșterea absolută (incrementele absolute) și cea relativă vor fi tot mai mici. Ajustarea tendinței cu dreapta este nesatisfăcătoare ; probabil ajustarea cu curba Gompertz ar fi recomandabilă.

Să conchidem că pentru perioade scurte dreapta oferă o bună ajustare. Acesta este cazul cel mai simplu ; cele mai multe s.c. cer alte tipuri de curbe (polinoame de grad superior), adică curbe neliniare.

Ajustarea seriei cronologice cu ajutorul parabolei. Ajustare folosind un polinom :

$$Y = a + bx + cx^2 + dx^3 \ldots \qquad (3)$$

O asemenea ajustare are ca rezultat o tendință neliniară. Estimarea parametrilor se face cu metoda celor mai mici pătrate.

Din familia acestor curbe, cea mai des folosită este parabola de gradul 2 :

$$Y = a + bx + cx^2. \qquad (4)$$

Sistemul de ecuații normale este :

$$\begin{aligned} \text{I}\quad \Sigma y &= na + b\Sigma x + c\Sigma x^2 \\ \text{II}\quad \Sigma xy &= a\Sigma x + b\Sigma x^2 + c\Sigma x^3 \\ \text{III}\quad \Sigma x^2 y &= a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4. \end{aligned} \tag{5}$$

Dacă folosim procedeul simplificat ($\Sigma x = 0$), sistemul de ecuații normale devine :

$$\begin{aligned} \text{I}\quad \Sigma y &= na + c\Sigma x^2 \\ \text{II}\quad \Sigma xy &= b\Sigma x^2 \\ \text{III}\quad \Sigma x^2 y &= a\Sigma x^2 + c\Sigma x^4. \end{aligned} \tag{6}$$

Vom ilustra acest tip de ajustare cu două exemple : evoluția recoltei medii la hectar a porumbului (tab. 47) și evoluția mortalității infantile (tab. 48).
Recolta medie la hectar :

$n = 45$; originea 1960 = 0.

Folosim formula simplificată ($\Sigma x = 0$)

$$\begin{aligned} na + c\Sigma x^2 &= \Sigma y \\ b\Sigma x^2 &= \Sigma xy \\ a\Sigma x^2 + c\Sigma x^4 &= \Sigma x^2 y. \end{aligned}$$

Introducem valorile obținute în formula (6)

$$\begin{aligned} 45a + \quad 7\,590c &= \quad 818{,}6 \\ 7590\,b \quad\quad &= 5\,230{,}3 \\ 7590a + 2\,302\,806c &= 157\,087{,}5 \end{aligned}$$

Determinăm pe $b = \dfrac{\Sigma xy}{\Sigma x^2} = \dfrac{5\,230{,}3}{7\,590} = 0{,}6891$ chintale.

Pentru determinarea lui a și c pornim de la formula

$$\begin{aligned} na + c\Sigma x^2 &= \Sigma y \\ a\Sigma x^2 + c\Sigma x^2 &= \Sigma x^2 y \end{aligned}$$

$$\begin{aligned} 45a + \quad 7\,590\ c &= \quad 818{,}6 \\ 7590a + 230\,2806\,c &= 157\,087{,}5 \end{aligned}$$

Tabelul 47

România. Recolta medie la hectar la porumb 1938—1982 (în chintale)

Anul	Recolta la ha (y)	Timpul (x)	x^2	x^4	xy	x^2y	$Y = a+bx+cx^2=15{,}05+0{,}6891x+0{,}0186x^2$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1938	10,6	−22	484	234 256	−233,2	5 130,4	8,89
1939	13,3	−21	441	194 481	−279,3	5 865,3	8,78
1940	10,5	−20	400	160 000	−210,0	4 200,0	8,71
1941	10,3	−19	361	130 321	−195,7	3 718,3	8,67
1942	6,4	−18	324	104 976	−115,2	2 073,6	8,67
1943	9,1	−17	289	83 521	−154,7	2 629,9	8,71
1944	11,1	−16	256	65 536	−176,6	2 841,6	8,79
1945	3,2	−15	225	50 625	−48,0	720,0	8,90
1946	2,3	−14	196	38 416	−32,2	450,8	9,05
1947	11,3	−13	169	28 561	−146,9	1 909,7	9,24
1948	6,2	−12	144	20 736	−74,4	892,8	9,46
1949	5,4	−11	121	14 641	−59,4	653,4	9,72
1950	7,4	−10	100	10 000	−74,0	740,0	10,02
1951	10,8	−9	81	6 561	−97,2	874,8	10,35
1952	8,5	−8	64	4 096	−68,0	544,0	10,73
1953	11,2	−7	49	2 401	−68,4	548,8	11,14
1954	15,0	−6	36	12 096	−90,0	540,0	11,59
1955	18,0	−5	25	625	−90,0	450,0	12,07
1956	11,0	−4	16	256	−44,0	176,0	12,59
1957	17,0	−3	9	81	−51,0	153,0	13,5
1958	10,0	−2	4	16	−20,0	40,0	13,75
1959	16,0	−1	1	1	−16,0	16,0	14,38
1960	15,5	0	0	0	0	0	15,05
1961	16,7	1	1	1	16,7	16,7	15,76
1962	15,9	2	4	16	31,8	63,6	16,50
1963	17,8	3	9	81	53,4	160,2	17,28
1964	20,2	4	16	256	80,8	323,2	18,10
1965	17,8	5	25	625	89,0	445,0	18,96
1966	24,4	6	36	1 296	146,4	878,4	19,85
1967	21,3	7	49	2 401	149,1	1 043,7	20,79
1968	21,3	8	64	4 096	170,4	1 363,2	21,75
1969	23,3	9	81	6 561	209,7	1 887,3	22,76
1970	21,2	10	100	10 000	212,0	2 120,0	23,80
1971	25,1	11	121	14 641	276,1	3 037,1	24,88
1972	30,7	12	144	20 736	368,4	4 420,8	26,00
1973	25,0	13	169	28 561	325,0	4 225,0	27,15
1974	25,1	14	196	38 416	351,4	4 919,6	28,34
1975	27,8	15	225	50 625	417,0	6 255,0	29,57
1976	34,1	16	256	65 536	545,6	8 729,6	30,84
1977	30,4	17	289	83 521	516,8	8 785,6	32,14
1978	32,1	18	324	104 976	577,8	10 400,4	33,48
1979	37,2	19	361	130 321	706,8	13 429,2	34,86
1980	33,9	20	440	160 000	678,0	13 560,0	36,27
1981	35,1	21	441	194 481	737,1	15 479,1	37,72
1982	42,1	22	484	234 256	926,2	20 376,4	39,21
Σ	818,6	0	7 590	2 322 806	5 230,3	157087,5	818,6
	Σy	Σx	Σx^2	Σx^4	Σxy	Σx^2y	ΣY

Tabelul 47 (continuare)

de unde:

$$a = \frac{\begin{matrix}818{,}6 & 7590\\ 157\,087{,}5 & 2\,302\,806\end{matrix}}{\begin{matrix}45 & 7\,590\end{matrix}} = \frac{(818{,}6\times 2\,302\,806) - (157\,087{,}5\times(7590)}{(45\times 2\,302\,806) - (7590\times 7590)} =$$

$$= \frac{\begin{matrix}7590 & 2\,302\,806\\ 1885{,}0769 & 1192{,}2941\end{matrix}}{\begin{matrix}103{,}62627 & 57\,6081\end{matrix}} = \frac{692{,}7828}{46\,817} = 15{,}05 \text{ chintale.}$$

$$c = \frac{\begin{matrix}45 & 818{,}6\\ 7590 & 157087.5\end{matrix}}{\begin{matrix}45 & 7590\end{matrix}} = \frac{(45\times 157\,087.5) - (7590\times 818{,}6)}{(45\times 2\,302\,806) - (7590\times 7590)} =$$

$$= \frac{\begin{matrix}7590 & 2.302.806\\ 7{,}0689375 & 6{,}213174\end{matrix}}{46{,}01817} = \frac{0{,}8557635}{46{,}01817} = 0{,}0185962 \approx 0{,}0186 \text{ chintale.}$$

Tendința parabolică:

$$Y = 15{,}05 + 0{,}6891x + 0{,}0186x^2,$$

Valorile estimate sînt trecute în col. 8 a tab. 47.

Să se observe că $\Sigma y = \Sigma Y = 818{,}6$.

S.c. empirică și cea ajustată sînt reprezentate în fig. 73.

Curba empirică are oscilații anuale, uneori pronunțate, care exprimă, în general, efectul condițiilor meteorologice. Tendința generală a recoltei la ha este însă de creștere, ceea ce reflectă efectul factorilor agrotehnici, organizatorici etc.

Curba aleasă pare să estimeze corect această tendință. Desigur, cercetătorul va fi interesat să studieze și *oscilațiile sau variațiile periodice* *.

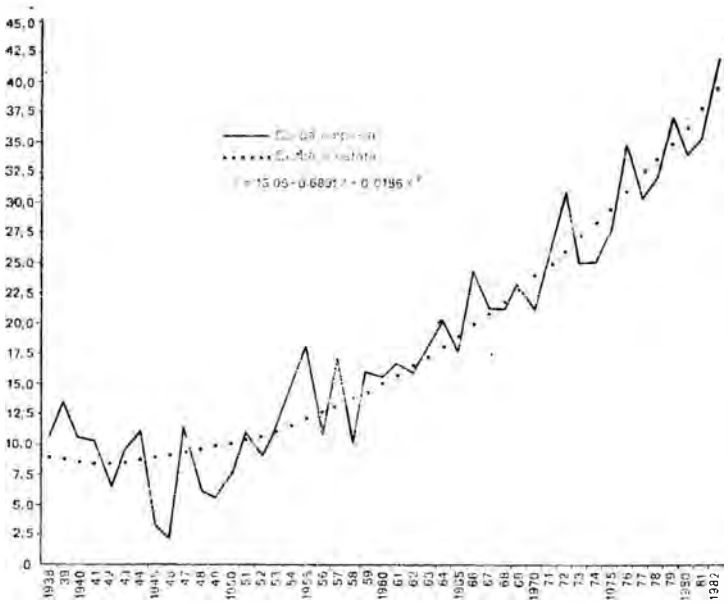


Fig. 73 România. Recolta medie la ha la porumb, 1938—1982 (date reale și date ajustate)

O altă ajustare parabolică este ajustarea s.c. a mortalității infantile (tab. 48).

În acest caz.

$$n = 33\,;\ \text{originea}\ 1966 = 0.$$

Introducem valorile obținute în formula (6) :

$$\begin{aligned} 33\,a + \quad 2992\,c &= \quad 1918{,}4 \\ 2992\,b \qquad\quad &= \quad 7788{,}3 \\ 2992\,a + 487696\,c &= 191385. \end{aligned}$$

De unde :

$$b = \frac{-\,7788{,}3}{2992} = -\,2{,}60 \ \text{promile}$$

$$a = \frac{191.385}{33} \ \frac{487.696}{2992} = 50{,}82 \ \text{promile}$$

$$c = \frac{2992}{33} \ \frac{191.385}{2992} = 0{,}08063.$$

Tabelul 48

România. Mortalitatea infantilă 1950—1982

(decese sub un an la 1000 născuți vii)

Anii	Mortalitatea infantilă (y)	Timpul (x)	x^2	x^4	xy	x^2y	$Y = a+bx+cx^2$ $Y = 50{,}82 - 2{,}60x + 0{,}08063x^2$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1950	116,7	—16	256	65 536	—1 867,2	29 875,2	113,1
1951	118,1	—15	225	50 625	—1 771,5	26 572,5	108,0
1952	104,7	—14	196	38 416	—1 465,8	20 522,2	103,0
1953	96,3	—13	169	28 561	—1 251,9	16 274,7	98,2
1954	88,8	—12	144	20 736	—1 065,6	12 787,2	93,6
1955	78,2	—11	121	14 641	—860,2	9 462,2	89,2
1956	81,5	—10	100	10 000	—815,0	8 150,0	84,9
1957	80,9	—9	81	6 561	—728,1	6 552,9	80,8
1958	69,4	—8	64	4 096	—555,2	4 441,6	76,8
1959	75,5	—7	49	2 401	—528,5	3 699,5	73,0
1960	74,6	—6	36	1 296	—447,6	2 685,6	69,3
1961	69,4	—5	25	625	—347,0	1 735,0	65,8
1962	58,8	—4	16	256	—235,2	940,8	62,5
1963	55,2	—3	9	81	—165,6	496,8	59,3
1964	48,6	—2	4	16	—97,2	194,4	56,3
1965	44,1	—1	1	1	—44,1	44,1	53,5
1966	46,6	0	0	0	0	0	50,8
1967	46,6	1	1	1	46,6	46,6	48,3
1968	59,5	2	4	16	119,0	238,0	45,9
1969	54,9	3	9	81	164,7	494,1	43,7
1970	49,4	4	16	256	197,6	790,4	41,7
1971	42,4	5	25	625	212,0	1 060,0	39,8
1972	40,0	6	36	1 296	240,0	1 440,0	38,1
1973	38,1	7	49	2 401	266,7	1 866,9	36,6
1974	35,0	8	64	4 096	280,0	2 240,0	35,2
1975	34,7	9	81	6 551	312,3	2 810,7	34,0
1976	31,4	10	100	10 000	314,0	3 140,0	32,9
1977	31,2	11	121	14 641	343,2	3 775,2	32,0
1978	30,3	12	144	20 736	363,6	4 363,2	31,2
1979	31,6	13	169	28 561	410,8	5 340,4	30,6
1980	29,3	14	196	38 416	410,2	5 742,8	30,2
1981	28,6	15	225	50 625	429,0	6 435,0	30,0
1982	28,0	16	256	65 536	448,0	7 168,0	29,9
Σ	1 918,4	0	2 992	487 696	—7788,3	191 385	1 918,4

Tendința parabolică: $Y = 50{,}82 - 2{,}60x + 0{,}08063x^2$.

Și aici $\Sigma y = \Sigma Y = 1918{,}4$.

Să precizăm că $a = 50{,}82$ promile este un termen mediu al seriei, în timp ce $b = -2{,}60$ promile este cantitatea cu care scade mortalitatea infantilă în fiecare an (fig. 74).

Mortalitatea infantilă la începutul perioadei era ridicată; în primii ani scăderea a fost rapidă; apoi, treptat, descreșterea a devenit mai lentă. Pentru comparație s-a luat nivelul mortalității infantile din Suedia din anul 1980 = 7,5‰, cel mai scăzut nivel din lume și care constituie un fel de asimptotă inferioară: nu poate fi imaginată o tendință de scădere avînd o asimptotă zero (în condițiile actuale).

Ajustarea seriei cronologice cu ajutorul exponențialei. Ajustare folosind funcția exponențială de tipul:

$$Y = ab^x, \tag{7}$$

care, prin logaritmare, devine:

$$\log Y = \log a + x \log b. \tag{8}$$

Estimarea parametrilor a și b se face cu ajutorul sistemului de ecuații normale:

$$n \log a + \log b \Sigma x = \Sigma' \log y$$

$$\log a \Sigma x + \log b \Sigma x^2 = \Sigma x \log y. \tag{9}$$

Procedeul simplificat ($\Sigma x = 0$):

$$n \log a = \Sigma \log y$$

$$\Sigma x^2 \log b = \Sigma x \log y, \tag{10}$$

de unde:

$$\log a = \frac{\Sigma \log y}{n}$$

$$\log b = \frac{\Sigma x \log y}{\Sigma x^2}.$$

Să reluăm seria cu producția de oțel din România 1970—1980 (tab. 42) pe care am ajustat-o cu rata medie de creștere și s-o ajustăm cu ajutorul funcției exponențiale, aplicînd metoda celor mai mici pătrate. Aplicația este strict ilustrativă, deoarece nu se poate vorbi de o tendință generală (seculară) la o serie în care $n = 11$ ani.

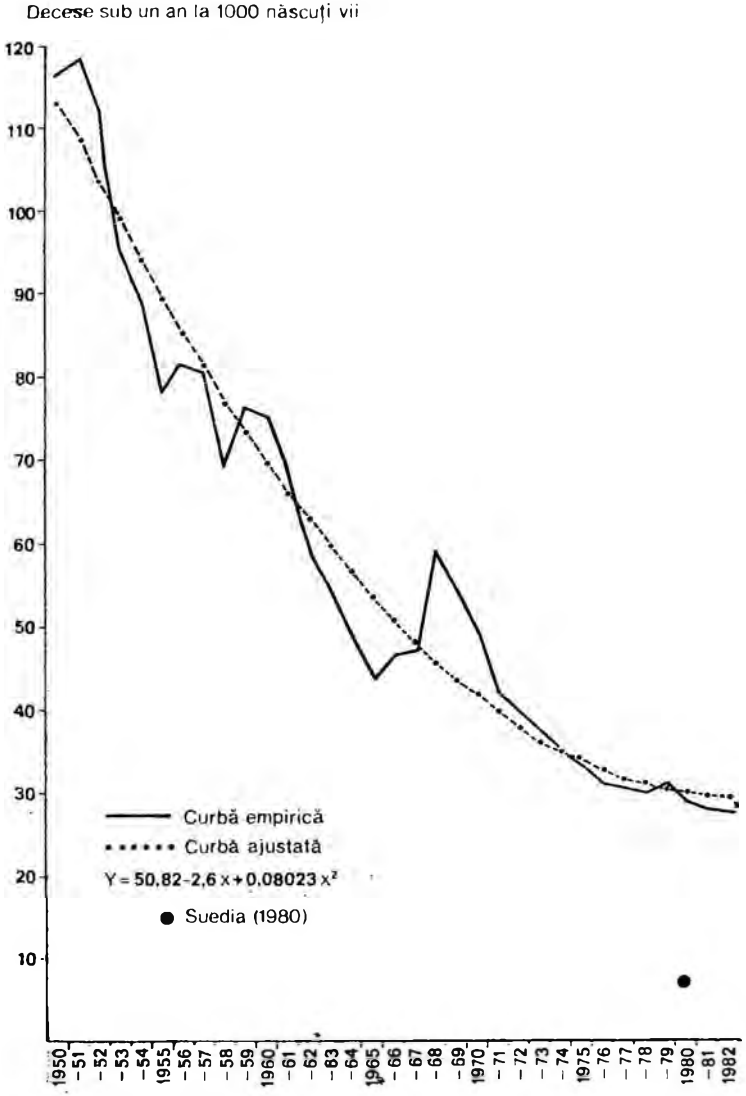


Fig. 74 Mortalitatea infantilă 1950—1982 (date reale și date ajustate)

Tabelul 49

România. *Producția de oțel 1970—1980 (mii tone)*. Ajustarea cu funcția exponențială.

Anul	Producția de oțel (y)	Timpul (x)	x^2	log y	$x \log y$	log Y = =3,779179+ 0.0329581x	Y
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1970	6 517	1	1	3,81405	3,81045	3,812151	6 488
1971	6 803	2	4	3,83270	7,66540	3,8451092	7 000
1972	7 401	3	9	3,86929	11,60787	3,8780673	7 552
1973	8 161	4	16	3,91174	15,64696	3,9110254	8 148
1974	8 848	5	25	3,94685	19,73425	3.9439835	8 790
1975	9 549	6	36	3,97996	23,87976	3,9769416	9 483
1976	10 773	7	49	4,03075	28,21525	4,0098997	10 230
1977	11 457	8	64	4,05907	32,47256	4,0428578	11 040
1978	11 779	9	81	4,07111	36,63990	4,0758159	11 910
1979	12 909	10	100	4,11093	41,10930	4,1090740	12 853
1980	13 175	11	121	4,1197	45,31736	4,1417321	13 860
Σ	107 332	66	506	43.74621	266,10266	—	107 354
	Σy	Σx	Σx^2	$\Sigma \log y$	$\Sigma x \log y$		ΣY

Originea : anul 1970 ; $n = 11$.

Introducem valorile obținute în formula (9) :

$$11 \log a + 66 \log b = 43{,}74621$$

$$66 \log a + 506 \log b = 266.10266$$

$$\log a = \frac{\begin{matrix} 266{,}10266 & 506 \\ 11 & 66 \end{matrix}}{\begin{matrix} 66 & 506 \end{matrix}} = \frac{(43{,}74621 \times 506) - (266{,}102666 \times 66)}{(11 \times 506) - (66 \times 66)} =$$

$$= \frac{4572{,}807}{1210} = 3{,}7791793.$$

$$\log b = \frac{\begin{matrix} 11 & 43{,}74621 \\ 66 & 266.10266 \end{matrix}}{\begin{matrix} 11 & 66 \end{matrix}} = \frac{(11 \times 266{,}10266) - (66 \times 43{,}74621)}{(11 \times 506) - (66 \times 66)} =$$

$$= \frac{39{,}8794}{1210} = 0{,}0329581.$$

Deci

$$\log Y = 3{,}7791793 + 0{,}0329581\, x.$$

Valorile lui log Y sînt trecute în col. 7 a tab. 49 ; antilogaritmii ne dau valorile lui Y (col. 8 tab. 49).

Să se observe că $\Sigma y = \Sigma Y$; diferența (0,02%) se datorează rotunjirii calculelor. Curba exponențială redă bine tendința producției de oțel, adică ajustează bine s.c. empirică (fig. 75). Dacă dreapta ne dă o creștere absolută constantă,

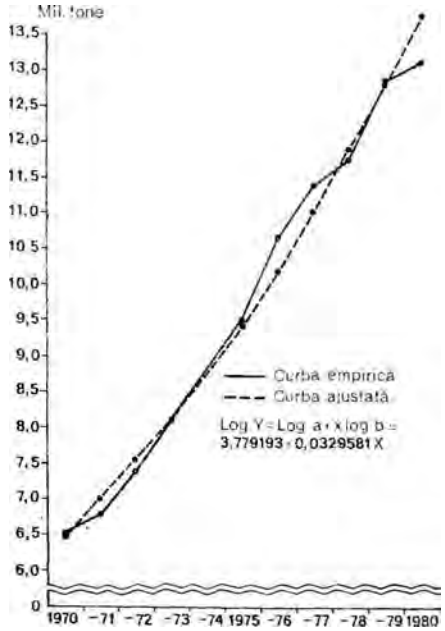


Fig. 75 România. Producția de oțel (mil. tone) 1970—1980 (date reale și date ajustate)

în schimb exponențiala se caracterizează printr-o creștere relativă constantă. În cazul nostru, rata medie de creștere este de 7,89 ≈ 7,9% anual. Întrucît dispunem de două ajustări pentru aceeași serie, să examinăm particularitățile lor.

Pentru ajustarea cu rata medie de creștere $\Sigma(y - Y)^2$ este 1 975 307 ; în timp ce în cazul ajustării cu funcția exponențială $\Sigma(y - Y)^2 = 1\ 028\ 600$. Prima este aproape dublă (192%). Conchidem că cea de-a doua ajustare este mai bună, adică mai apropiată de datele reale.

Curba exponențială și ajustările (interpolările) cu ajutorul ei, precum și extrapolările tendențiale se bucură de o largă aplicație pentru simplul motiv că în ultimele decenii numeroase fenomene economice, demografice, tehnologice se dezvoltă exponențial, uneori cu rate destul de ridicate. În numeroase țări, economia crește exponențial ; populația mondială a crescut în perioada 1950—1975 cu o rată medie anuală de 2% (perioada de dublare 35 ani), fenomen

Tabelul 50

Pătratele abaterilor valorilor empirice de la valorile ajustate

Anul	Seria empirică (y)	Seria ajustată cu rata medie de creștere ($r = 0{,}073$)	$(y - Y)^2$	Seria ajustată cu funcția exponențială (Y)	$(y - Y)^2$
(1)	(2)	(3)	(4)	(5)	(6)
1970	6 517	6 517	0	6 488	841
1971	6 803	6 993	36 100	7 000	38 809
1972	7 401	7 503	10 404	7 552	22 801
1973	8 161	8 051	12 100	8 148	169
1974	8 348	8 639	43 681	8 790	3 364
1975	9 549	9 269	78 400	9 483	4 356
1976	10 773	9 946	683 929	10 230	294 849
1977	11 451	10 672	616 226	11 040	173 889
1978	11 779	11 451	107 584	11 910	17 161
1979	12 909	12 287	386 884	12 853	3 136
1980	13 175	13 175	0	13 860	469 225
Σ	107 332	104 504	1 975 307	107 354	1 028 600

caracterizat drept „explozie demografică"; știința se dezvoltă exponențial (legea lui J. D. De Solla Price), dublîndu-se în circa 15 ani. Să mai adăugăm „explozia informațională", „explozia documentară", „explozia tehnologică", ca particularități ale actualei revoluții științifice și tehnice.

Dacă, deci, folosirea calculelor exponențiale este justificată, în schimb pericolele unei aplicări necritice sînt mari. Exemplul din tab. 51 ne permite să punem în evidență unele denaturări generate de folosirea exponențialei, neajunsuri de care nu sînt scutite nici ajustările cu alte funcții. Totodată, va reieși din acest exemplu importanța cunoașterii esenței fenomenului considerat pentru a evita asemenea situații.

Datele din tab. 51 sînt prezentate în fig. 76 și fig. 77.

Din ambele reprezentări grafice rezultă că producția de energie electrică a crescut exponențial în perioada 1950—1982 în strînsă legătură cu dezvoltarea economiei naționale, în special a industriei. Pentru întreaga perioadă 1950—1982 rata medie de creștere (r) a fost de 11,5% anual:

$$(1 + r)^{32} = \frac{y_{1982}}{y_{1950}}, \text{ de unde: } (1 + r) = \sqrt[32]{\frac{68900}{2113}} =$$

$$= \sqrt[32]{32{,}6077}; \log (1 + r) = \frac{\log 32{,}6077}{32}.$$

Făcînd operațiile necesare rezultă că $r = 0{,}115$ sau 11,5% anual. Ajustarea (interpolarea) s.c. cu această rată este trecută în col. 3a tab. 51. Din fig. 76 rezultă că între seria de date reale și cele ajustate există o diferență importantă.

Coincid, potrivit definiției ratei medii, numai termenii extremi : 1950 și 1982 ; pentru celelalte valori abaterile sînt foarte mari. Suma producției de energie electrică (este vorba de o s.c. de intervale) este de 937 732 mil.kWh ; suma

Tabelul 51

România. *Producția de energie electrică (mil. kWh) 1950—1982* date reale și date ajustate cu ajutorul ratelor medii anuale de creștere (r)

Anul	Producția de energie electrică (*y*)	Valori ajustate (*Y*) 1950—1982	Valori ajustate pe subperioade			
			1950—1960 (r = 0,137)	1960—1970 (r = 0,165)	1970—1980 (r=0,0677)	1980—1982 (r = 0,0105)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1950	2 113	2 113	2 113			
1951	2 473	2 356	2 402			
1952	2 886	2 627	2 732			
1953	3 411	2 929	3 106			
1954	3 697	3 266	3 531			
1955	4 340	3 641	4 015			
1956	4 930	4 060	4 565			
1957	5 440	4 527	5 191			
1958	6 184	5 048	5 902			
1959	6 824	5 628	6 710			
1960	7 650	6 275	7 650	7 650		
1961	8 657	6 997		8 912		
1962	10 087	7 802		10 383		
1963	11 682	8 699		12 096		
1964	13 851	9 699		14 092		
1965	17 215	10 815		16 417		
1966	20 806	12 059		19 126		
1967	24 769	13 445		22 281		
1968	27 828	14 992		25 958		
1969	31 509	16 716		30 241		
1970	35 088	18 638		35 088	35 088	
1971	39 454	20 781			37 464	
1972	43 439	23 171			40 000	
1973	46 779	25 836			42 708	
1974	49 063	28 807			45 599	
1975	53 721	32 120			48 686	
1976	58 266	35 813			51 982	
1977	59 858	39 932			55 501	
1978	64 255	44 524			59 259	
1979	64 933	49 644			63 271	
1980	67 486	55 354			67 486	67 486
1981	70 138	61 719				68 195
1982	68 900	68 817				68 900
Σ	937 732	848 850				

valorilor ajustate 648 850 mil.kWh, cu circa 30% mai mică. Explicația acestei anomalii rezidă în faptul că producția de energie electrică a crescut exponențial, dar cu *rate diferite* în cursul perioadei. Prin urmare, *tendința generală* * determinată în acest mod, se dovedește a fi departe de realitate.

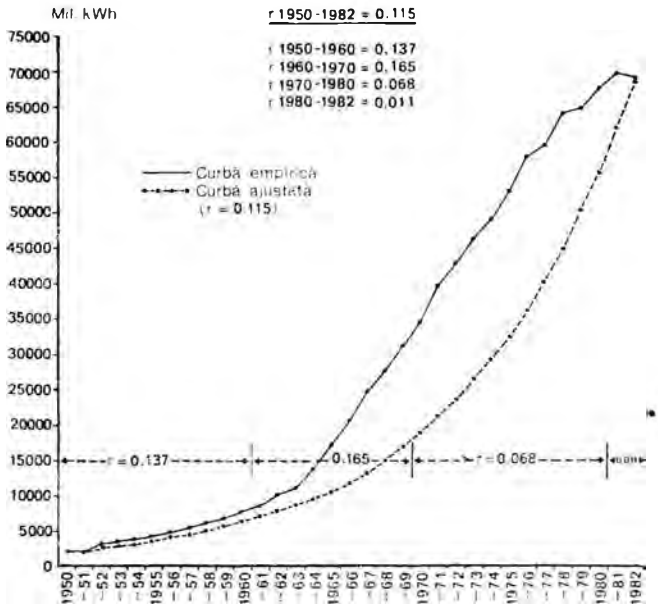


Fig. 76 România. Producția de energie electrică (mil. kWh) 1950—1982 (scară aritmetică)

Rate medii anuale

1950—1982.....0,115 sau 11,5% anual

din care :

1950—1960.....0,137 sau 13,7% anual
1960—1970.....0,165 sau 16,5% anual
1970—1980.....0,068 sau 6,8% anual
1980—1982.....0,0105 sau 1,1% anual

Ajustarea (interpolarea) pe subperioade este dată în coloanele 4, 5, 6 și 7 din tab. 51, iar în reprezentare grafică, în fig. 78.

Noua sumă a valorilor ajustate este de 891 562 mil.kWh, cu 5% mai mică decît suma valorilor reale.

Prin urmare la determinarea *tendinței generale* * trebuie observată modificarea ei, ceea ce impune ajustarea pe subperioade. Analiza de fond a s.c. a producției de energie electrică, în strînsă legătură cu procesele economice reale, ne arată că în prima perioadă a avut loc o *demarare*, relativ rapidă (13,7% anual); urmează apoi o *expansiune*, cu o rată înaltă de creștere (16,5%); se instalează apoi o *încetinire* a creșterii exponențiale (6,8%), și, în sfîrșit, o *stagnare* sau o

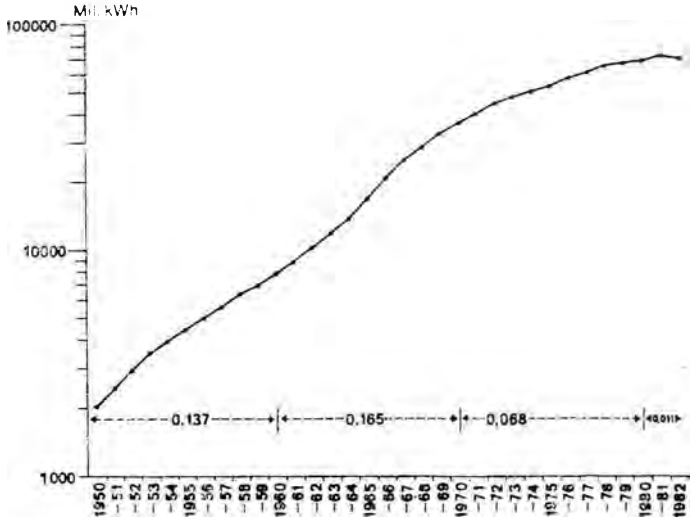


Fig. 77 România. Producția de energie electrică (mil. kWh) 1950—1982 (scară verticală logaritmică)

saturare (1,1%). În funcție de politica energetică și de industrializare (tip intensiv) s-ar putea ca producția energetică să fie în viitor plafonată de o asimptotă; este însă posibil ca după faza de saturare să reînceapă creșterea exponențială cu aceleași faze.

Metodologic, ajungem la necesitatea folosirii unor alte curbe: *exponențiala modificată* *, curbe de *creștere asimptotică* *, *curba logistică* * și la problema *analizei ciclului* *.

Totuși, cum s-a mai spus, rata medie de creștere, în forma de medie geometrică sau în cea a funcției exponențiale, se folosește pe scară largă. În publicațiile statistice O.N.U. ca și în cele naționale, după această metodă se calculează dinamica populației, a venitului național, a producției industriale etc. În România, aplicarea metodei ratei medii de creștere este cvasi-universală.

În legătură cu rata medie de creștere se pune o problemă practică: calculul perioadei de dublare (triplare etc.) a unui fenomen în condițiile unei anumite rate medii de creștere.

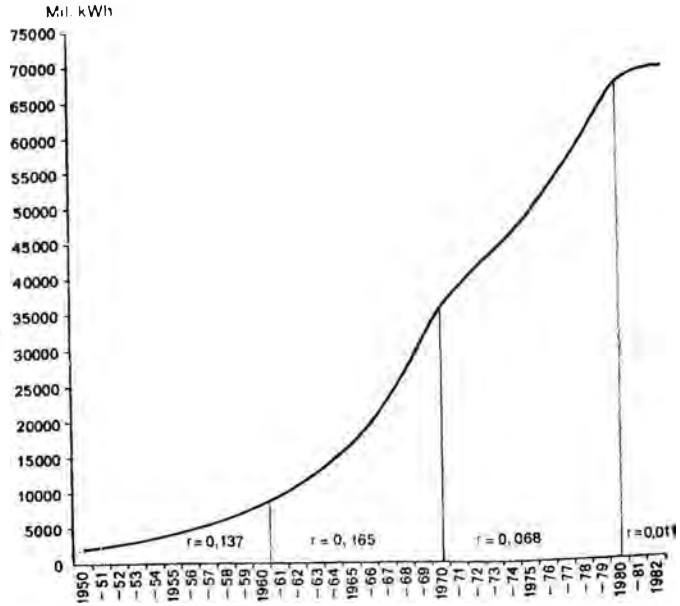


Fig. 78 România. Producția de energie electrică (mil. kWh) 1950—1982. Curbe ajustate pe subperioade

Formula este următoarea :

$$(1 + r)^n = 2,$$

în care : r — rata medie anuală de creștere ;
n — numărul de ani ;
2 — valoarea dublă a fenomenului.

Se rezolvă :

$$n = \frac{\log 2}{\log(1 + r)}.$$

Exemplu. Producția de energie electrică a crescut, în perioada 1970—1980, cu 6,8% anual (deci $1 + r = 1{,}068$).
Dublarea :

$$n = \frac{\log 2}{\log 1{,}068} = \frac{0{,}30103}{0{,}02857} = 10{,}5 \text{ ani.}$$

Dăm cîteva valori caracteristice pentru fenomenele demografice și economice (tab. 52).

Tabelul 52

Perioada de dublare (ani) a unui fenomen în condițiile unei anumite rate medii anuale de creștere (%)

Rata medie anuală (r%)	Perioada de dublare (ani)	Rata medie de creștere (r%)	Perioada de dublare (ani)
0,0	138,7	8,5	8,5
0,5	138,7	9,0	8,0
1,0	69,7	9,5	7,6
1,5	46,5	10,0	7,3
2,0	35,0	10,5	6,9
2,5	28,1	11,0	6,6
3,0	23,4	11,5	6,4
3,5	20,1	12,0	6,1
4,0	17,7	12,5	5,9
4,5	15,7	13,0	5,7
5,0	14,2	13,5	5,5
5,5	12,9	14,0	5,3
6,0	11,9	14,5	5,1
6,5	10,8	15,0	5,0
7,0	10,2	15,5	4,8
7,5	9,6	16,0	4,7
8,0	9,0	16,5	4,5

Cu o rată medie anuală înregistrată în anii 1950—1981 fenomenele de mai jos se pot dubla (aproximativ).
Exemple. Sursa: (Anuarul Statistic al R.S.R. 198, p. 42*).*

Fenomenul	Rata (%)	Perioada de dublare (ani)
1. Populația	1,0	70 ani
2. Populația ocupată	0,7	100 ani
3. Fondurile fixe	7,3	10 ani
4. Producția industrială	12,0	6 ani
5. Venitul național pe locuitor	8,0	9 ani
6. Retribuția reală total personal muncitor	4,8	15 ani

Ca un indicator să se dubleze în cursul unui cincinal, rata medie de creștere ar trebui să fie de 14,8% anual.

Ajustarea seriei cronologice cu ajutorul exponențialei modificate. — ajustare folosind formula :

$$Y = k + ab^x, \tag{11}$$

în care : ab^x — exponențiala ;

k — asimptotă

Exponențiala modificată face parte din familia *curbelor de creștere asimptotică*. Dacă dreapta ($Y = a + bx$) descrie o cantitate constantă (spor absolut) a creșterii sau descreșterii, iar curba exponențială ($Y = ab^x$) implică o rată constantă a schimbării, deci și o rată constantă a schimbării cantității ; în schimb, exponențiala modificată descrie o tendință potrivit căreia cantitatea creșterii scade cu o rată procentuală constantă, dar în același timp curba se apropie de o limită superioară (asimptotă).

Diferitele forme ale curbei exponențialei modificate se văd în fig. 79.

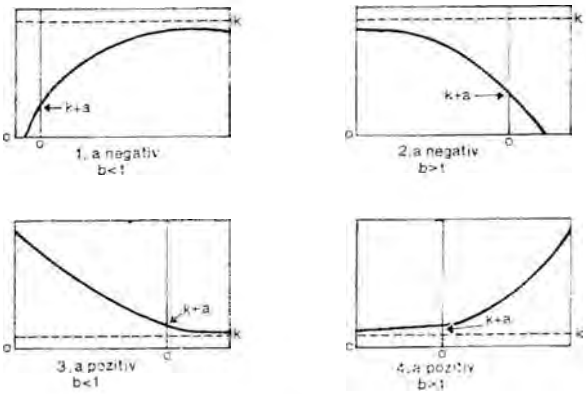


Fig. 79 Patru forme ale curbei exponențiale modificate $Y = k + ab$ *(după Croxton et al., p. 264)*

Un exemplu convențional (după Croxton, p. 263) ne edifică asupra semnificației exponențialei modificate și a modului de estimare a parametrilor (tab. 53). Primele două coloane din tab. 53 prezintă o serie care are o rată constantă de descreștere a cantității creșterii. Coloanele 4 și 5 reprezintă prima diferență care este de 75% față de prima diferență precedentă. Sporurile creșterii sînt : $\Delta_1, \Delta_2, \Delta_3, \Delta_4$ și Δ_5 și deci :

$$\frac{\Delta_2}{\Delta_1} = \frac{\Delta_3}{\Delta_2} = \frac{\Delta_4}{\Delta_3} = \frac{\Delta_5}{\Delta_4} = 0{,}75 \text{ sau } 75\%.$$

Tabelul 53

Date convenţiale privind curba exponenţială modificată ($k=114$)

x	y	Totaluri parţiale	Creşterea (sporul) lui y	Procentul sporului precedent
1	2	3	4	5
0	50		...	...
1	66	116.000	16	...
	....	....		
2	78		12	75
3	87	165.000	9	75
	....			
4	93,75	....	7,75	75
5	98,8125	192.5625	5,0625	75

Dacă linia tendinţei s-ar continua ea se va apropia tot mai mult de asimptota $k = 114$, fără însă să o atingă.

Procedura de calcul pentru estimarea parametrilor este :

$$b^n = \frac{\Sigma_{3y} - \Sigma_{2y}}{\Sigma_{2y} - \Sigma_{1y}} \; ; \; a = (\Sigma_{2y} - \Sigma_{1y}) \frac{b-1}{(b^n - 1)}$$

$$k = \frac{1}{n}\left[\Sigma_1 y - \left(\frac{b^n - 1}{b-1}\right) a\right],$$

în care : Σ_1, Σ_2 şi Σ_3 sînt totaluri parţiale ;

n — numărul de ani în fiecare din cele trei totaluri parţiale

Rezolvarea formulei cere să se determine mai întîi b, apoi a şi în sfîrşit k. Dacă a şi b sînt introduse în expresia dată pentru k obţinem :

$$k = \frac{1}{n}\left[\frac{(\Sigma_1 y)(\Sigma_3 y) - (\Sigma_2 y)^2}{\Sigma_1 y + \Sigma_3 y - 2\Sigma_2 y}\right].$$

S.c. reale nu se comportă în general după o curbă exponenţială modificată ; consideraţiile privind această curbă înlesnesc înţelegerea altor curbe de creştere asimptotică.

Ajustarea seriei cronologice cu ajutorul curbei Gompertz. Ajustarea folosind formula :

$$Y = k\, a^{b^x},$$

care, în formă logaritmică, devine

$$\log Y = \log k + (\log a)\, b^x.$$

Curba Gompertz descrie o tendință în care sporurile („incrementele") creșterii logaritmilor scad cu o rată constantă.
Estimarea parametrilor k, a și b se face cu formula de la pag. 275.
O serie la care se poate aplica curba Gompertz este cea din tab. 54.
Determinăm mai întîi pe k; $n = 11$.

$$\log k = \frac{1}{n}\left[\frac{(\Sigma_1 \log y)(\Sigma_3 \log y) - (\Sigma_2 \log y)^2}{\Sigma_1 \log y + \Sigma_3 \log y - 2\Sigma_2 \log y}\right] =$$

$$= \frac{1}{11}\left[\frac{(43{,}62864 \times 45{,}38676) - 45{,}11777^2}{43{,}62864 + 45{,}38676 - 2(45{,}11777)}\right] =$$

$$= \frac{1}{11}\left[\frac{1980{,}1626 - 2035{,}6131}{-1{,}22014}\right] = \frac{1}{11}\left(\frac{-55{,}4505}{-1{,}22014}\right) =$$

$$= \frac{45{,}446014}{11} = 4{,}1312526.$$

$$\log k = 4{,}1312526\,;\ \text{antilog} = 13\,530\,000.$$

Asimptota superioară este de 13 530 000 mii tone țiței extras (fig. 80).
Ajustarea — am mai spus — este de tip grafic („tendință vizuală"), cu o asimptotă de 14 mil. tone, stabilită arbitrar. Este posibil ca în urma programului special de creștere a extracției de țiței, aceasta să crească în anii următori. Ținînd seama de rezervele existente ca și de politica energetică, extracția de țiței s-ar putea plafona la 14 milioane tone.
Calculul celorlalți parametri se face după formulele deja citate:

$$b^n = \frac{\Sigma_3 \log y - \Sigma_2 \log y}{\Sigma_2 \log y - \Sigma_1 \log y}$$

$$\log a = (\Sigma_2 \log y - \Sigma_1 \log y)\,\frac{b-1}{(b^x-1)^2}.$$

Cu datele obținute:

$$b^{11} = \frac{45{,}38676 - 45{,}11777}{45{,}11777 - 43{,}62864} = \frac{0{,}26899}{1{,}48913} = 0{,}1806356.$$

Tabelul 54

România. Țiței extras 1950—1982 (mii tone)

Anul	Țiței extras (y)	log y	Anul	Țiței extras (y)	log y	Anul	Țiței extras (y)	log y
(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
1950	5 047	3,70303	1961	11 582	4,06371	1972	14 128	4,15008
1951	6 211	3,79316	1962	11 864	4,06421	1973	14 287	4,15494
1952	8 002	3,90320	1963	12 233	4,08754	1974	14 486	4,16095
1953	9 058	3,95703	1964	12 395	4,09325	1975	14 590	4,16406
1954	9 741	3,98860	1965	12 571	4,09934	1976	14 700	4,16732
1955	10 555	4,02345	1966	12 825	4,10805	1977	14 650	4,16584
1956	10 920	4,03822	1967	13 206	4,12079	1978	13 724	4,13748
1957	11 180	4,04844	1968	13 285	4,12336	1979	12 323	4,09072
1958	11 336	4,05446	1969	12 346	4,09149	1980	11 511	4,06108
1959	11 438	4,05835	1970	13 377	4,12636	1981	11 644	4,06610
1960	11 500	4,06070	1971	13 793	4,13966	1982	11 700	4,06819
Total parțial	104.988	43,62864	Total parțial	139.477	45,11777	Total parțial	147 743	45,38676
	$\Sigma_1 y$	$\Sigma_1 \log y$		$\Sigma_2 y$	$\Sigma_2 \log y$		$\Sigma_3 y$	$\Sigma_3 \log y$

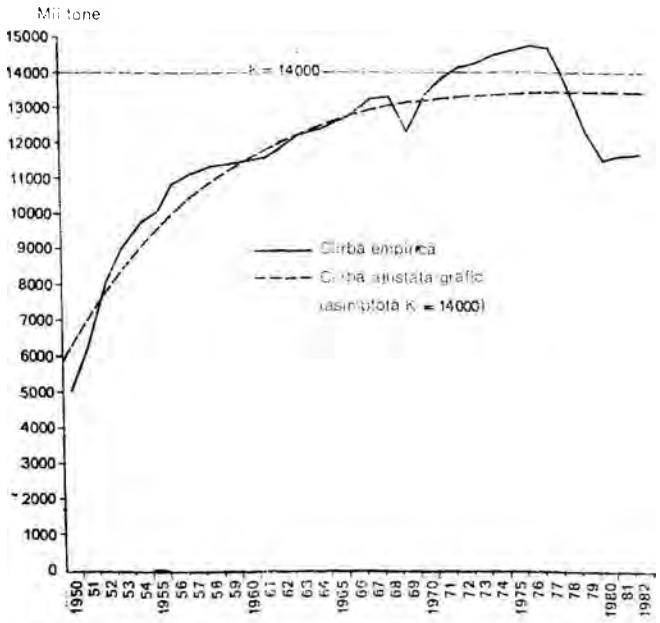


Fig. 80 România. Țiței extras (mii tone) 1950—1982

În mod succesiv se determină $\log b^{11}$, apoi $\log b$ și în sfîrșit b. Se trece apoi la determinarea lui a.

Schema de calcul al tendinței este următoarea :

Anul	x	Extracția (y)	$\log y$	Calculul valorilor tendinței			
				b^x	$(\log a)b^x$	$\log Y = \log k + (\log a)\, b^x$	Y
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)

Curba Gompertz, al cărei autor a fost actuar și demograf, a fost considerată în 1825 (Benjamin Gompertz, *On the Nature of the Function Expressive of the Law of Human Mortality*, în Philosophical Transanctions of the Royal Society of London) cu aplicații la probabilitatea instantanee de deces, exprimată ca funcție exponențială de vîrstă :

$q(x) = Ae^{k^x}$, în care : $q(x)$ — probabilitatea instantanee de deces. Formula a fost stabilită pornindu-se de la ideea că uzura progresivă a organismului se traduce printr-o creştere continuă, cu o rată constantă, a probabilităţii instantanee.

Ea a fost generalizată de actuarul şi demograful W. M. Makeham (*On the Law of Human Mortality and Construction of Annuity Tables*, în „Journal of the Institute of Actuaries", vol. VIII).

Formula Gompertz-Makeham, formulă privind probabilitatea de deces prin modificarea formulei Gompertz :

$$q(x) = Ae^{k^{x+k'}}.$$

Introducerea parametrului k' s-a făcut pentru a lua în considerare decesele accidentale care se adaugă deceselor prin uzura organismului. Termenul „accidental" este luat în sens larg, incluzînd şi decesele de origine infecţioasă. Forma sub care se întîlneşte această formulă cel mai adesea la calculul supravieţuitorilor este :

$$l_x = ks^x g^{c^x},$$

în care : l_x — numărul supravieţuitorilor.

O metodă de estimare a parametrilor din formula Gompertz-Makeham se găseşte în Mihoc şi Urseanu (1962, p. 330—331).

Curba Gompertz se foloseşte nu numai în demografie, biometrie şi actuariat ; ea se aplică pe scară largă la ajustarea s.c. în econometrie, marketing etc.

Formula Gompertz-Makeham se foloseşte numai în demografie şi actuariat la ajustarea probabilităţilor de deces şi, respectiv, a funcţiei de supravieţuire, în special la virstele înaintate. Aplicabilitatea curbei Gompertz la fenomenele economice şi în special la industrie se justifică prin faptul că dezvoltarea economică (sau a unui produs oarecare) are loc în patru faze : a) perioada de experimentare ; b) perioada de creştere (producţia de masă) ; c) atingerea unui punct cînd creşterea se continuă, dar cu o rată descrescătoare ; d) perioada de stabilitate.

Ajustarea seriei cronologice cu ajutorul curbei logistice. Ajustare folosind formula :

$$\frac{1}{Y} = k + ab^x,$$

care se mai scrie :

$$Y = \frac{k}{1 + 10^{a+bx}} \quad \text{şi} \quad Y = \frac{k}{1 + e^{a+bx}},$$

formulă pe care o întîlnim cel mai adesea în lucrările de specialitate.

Pentru a ajusta cu prima formulă cu ajutorul *metodei punctelor alese* avem nevoie să alegem trei ani : x_0, x_1 şi x_2, echidistanţi, primul la începutul perioadei, al doilea — la mijlocul ei şi ultimul — la sfîrşitul perioadei. Valorile lui y

se notează cu y_0, y_1 și y_2; numărul de ani (n) este cel de la x_0 la x_1 sau de la x_1 la x_2.
Cele trei constante se obțin astfel:

$$k = \frac{2y_0y_1y_2 - y_1^2(y_0 + y_2)}{y_0y_2 - y_1^2}$$

$$a = \log \frac{k - y_0}{y_0}$$

$$b = \frac{1}{n}\left[\log \frac{y_0(k - y_1)}{y_1(k - y_0)}\right].$$

Forma curbei logistice se vede din fig. 81.

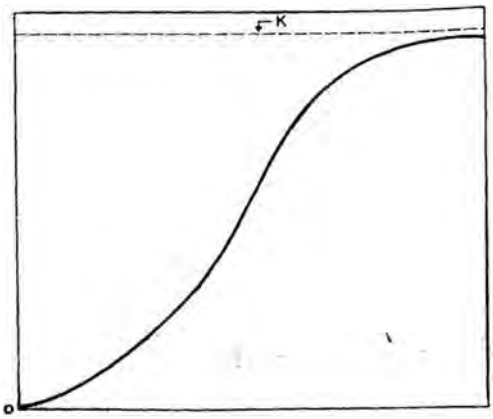


Fig. 81 Curba logistică $Y = \frac{k}{1 + 10^{a+bx}}$

Calculele sînt, în general, laborioase. Există în literatură procedee simplificate; unul din acestea este prezentat în lucrarea lui Shryock și Siegel (1971, vol. II, p. 382 și urm.).
Curba logistică a fost descoperită de matematicianul belgian P. F. Verhulst, în 1838 (*Notice sur la loi que la population suit dans son accroissement* în Correspondance mathématique et physique publiée par A. Quételet, X, 113—21, Bruxelles) și redescoperită de biologii și demografii americani R. Pearl și L. J. Reed, în 1920 (*On the Rate of Growth of the Population of the United States since 1790 and Its Mathemathical Representation*, în „Proceedings of the National Academy of Sciences VI", nr. 6, 275, 288). R. Pearl a efectuat și experimente cu musca *Drosophila Melanogaster* pentru a demonstra că orice specie biologică, inclusiv specia umană, se dezvoltă după o curbă logistică. Cele mai numeroase aplicații sînt în demografie; în ultimele decenii ele s-au extins și în economie,

ecologie etc. Considerațiile privind creșterea economică zero, populația staționară etc. au reactualizat curba logistică. Raționamentul principal este acela că trăim într-o lume finită (suprafață finită, resurse limitate) și deci dezvoltarea, multă vreme exponențială, tinde spre o stare de stabilitate, fiind plafonată de o asimptotă superioară.

În demografia românească sînt două aplicații cu creșterea logistică a populației: P. Râmneanțu (*Elemente de biometrie medicală și statistică vitală*, 1939, p. 274—280) pentru populația totală a României 1650—2200 și St. Știrbu, cu privire la populația Municipiului București (republicat în Vl. Trebici, *Statistica populației*, A.S.E., 1970, p. 622).

Criterii de alegere a curbelor de ajustare a tendinței generale (teste). Varietatea s.c. empirice, exprimînd multitudinea de fenomene sociale, economice și demografice cere ca la alegerea tipului de curbă de ajustare să se facă mai întîi o ajustare grafică în vederea evidențierii caracterului tendinței. Folosirea scării logaritmice este indicată în cazul în care tendința este neliniară. În toate cazurile judecata cercetătorului, bazată pe cunoașterea domeniului respectiv, are o importanță hotărîtoare. După ce s-au obținut curbele ajustate, cu ajutorul funcțiilor respective, se pot aplica următoarele teste (Croxton et al., 1967, p. 282—283):

1) Dacă primele diferențe (Δ^1) au tendința de a fi constante, se recomandă *dreapta.**

2) Dacă diferențele de ordinul doi (Δ^2) au tendința de a fi constante, se folosește o *curbă de gradul doi.**

3) Dacă primele diferențe au tendința de a descrește cu un procent constant, se recomandă *exponențiala modificată.**

4) Dacă tendința aproximată, reprezentată pe o rețea aritmetică, este o dreaptă, se folosește *dreapta.**

5) Dacă tendința aproximată, reprezentată pe o rețea semilogaritmică, este o dreaptă, se folosește *curba exponențială.* *

6) Dacă tendința aproximată, reprezentată pe o rețea semilogaritmică, seamănă cu o exponențială modificată, se recomandă *curba Gompertz.**

6) Dacă tendința aproximată, reprezentată pe o rețea cu scară verticală reciprocă și cu o scară orizontală, seamănă cu o exponențială modificată, se recomandă *curba logistică.**

8) Dacă primele diferențe seamănă cu o *curbă de frecvențe asimetrică*, se folosește *curba Gompertz* * sau o curbă logistică mai complexă.

9) Dacă primele diferențe seamănă cu o curbă normală de frecvențe, se folosește *curba logistică.**

10. Dacă primele diferențe ale logaritmilor sînt constante, se folosește o *curbă exponențială.**

11. Dacă diferențele de ordinul doi ale logaritmilor sînt constante, se aplică ajustarea cu o curbă de gradul doi la logaritmi.

Se întîmplă uneori ca două curbe să fie folosite pentru ajustarea aceluiași s.c. empirice. În acest caz va fi adoptată ajustarea cu acea curbă care ne dă cele mai mici abateri pătrate (a se vedea exemplul din tab. 50 cu două ajustări).

ANALIZA VARIAȚIILOR (FLUCTUAȚIILOR) SEZONIERE ALE SERIEI CRONOLOGICE — Operație complexă constînd din aplicarea metodelor statistico-matematice asupra s.c. empirice în vederea identificării sezonalității, ca formă a *variației periodice* *. Etimologic, este vorba de o variație determinată de succesiunea anotimpurilor („sezoanelor"); prin extindere, sezonalitatea poate avea însă și un alt caracter: repetarea unor particularități pe luni,

trimestre sau alte unități de timp. De aceea s.c. trebuie să cuprindă termenii aranjați pe luni, trimestre etc. Sezonalitatea poate să exprime un tip („pattern") de mișcare relativ regulată, care se menține de la an la an („tip constant"); ea poate însă să reflecte un tip ce se modifică de la o perioadă la alta („tip variabil"). Fenomenul de sezonalitate este caracteristic, în primul rînd, agriculturii (primăvara, vara, toamna și iarna); el se întîlnește și în alte ramuri ale economiei naționale, construcții, transporturi, industria alimentară etc. Un domeniu specific este cel al demografiei; nașterile, decesele, căsătoriile cunosc o anumită sezonalitate care, în special, în ultimul timp, face obiectul unor preocupări aprofundate.

Vom examina variația lunară a numărului de căsătorii din mediul rural în perioada 1957—1966 (tab. 55).

Din examinarea tabelului rezultă că numărul căsătoriilor variază de la o lună la alta în cuprinsul fiecărui an. Se mai constată însă că numărul căsătoriilor variază de la un an la altul; în perioada 1957—1966 *tendința generală* * este de scădere.

Fiind vorba de frecvențe absolute (numărul căsătoriilor), prima impresie este aceea că variația sezonieră se reduce, ceea ce este însă fals, deoarece tipul de variație sezonieră rămîne același, cum se va arăta în continuare, în condițiile diminuării numărului de căsătorii. Prin urmare, analiza trebuie să pună concomitent în evidență *tendința generală* * și *variația sezonieră* * și să țină seama totodată de *variația ciclică* * (dacă aceasta există) și de *variația reziduală* *. Numărul anual al căsătoriilor în mediul rural scade sistematic în perioada 1957—1966. Fig. 82 ne redă această diminuare prin linia punctată. De la

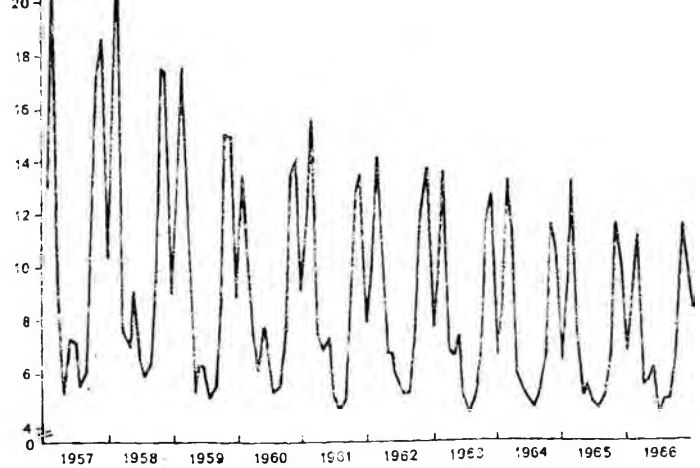


Fig. 82 Variația lunară a căsătoriilor în mediul rural în România 1957—1966 (Sursa: C. Moineagu et al., Statistica, *Ed. științifică și enciclopedică, 1976, p. 31)*

Tabelul 55

România. Variația lunară a căsătoriilor din mediul rural 1957—1966

Lunile	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	Total lunar
Ianuarie	13 583	16 266	13 071	13 569	11 358	9 589	9 707	8 810	8 461	8 760	113 174
Februarie	21 167	21 867	17 395	21 212	15 528	14 163	13 533	13 294	13 111	11 155	162 425
Martie	8 750	7 538	12 791	7 561	7 470	10 386	6 902	10 401	7 037	5 421	84 257
Aprilie	5 335	6 908	5 361	6 013	6 824	6 672	6 595	5 865	4 954	5 564	60 091
Mai	7 458	9 205	6 492	7 821	7 372	6 714	7 226	5 355	5 464	6 114	69 221
Iunie	6 705	6 612	7 238	6 085	5 169	5 953	5 137	5 029	4 827	4 387	56 142
Iulie	5 504	5 908	5 138	5 340	4 617	5 130	4 464	5 689	4 625	4 872	50 287
August	6 025	6 367	5 456	5 522	4 876	5 115	5 116	5 296	4 947	4 996	53 716
Septembrie	10 627	10 162	8 775	7 723	7 685	7 280	6 814	6 474	6 401	6 540	78 681
Octombrie	17 124	17 590	15 020	13 373	12 414	12 154	11 403	11 455	11 527	11 596	133 686
Noiembrie	18 842	17 519	14 976	14 117	13 479	13 716	12 685	10 613	11 043	10 197	137 187
Decembrie	10 442	9 126	8 932	9 096	7 803	7 625	6 771	6 313	6 844	8 286	81 238
Total anual	131 562	135 068	119 845	117 432	104 625	104 497	96 353	93 594	89 241	87 888	1 080 105
Medii lunare pe an	10963,5	11255,7	9987,1	9786,0	8718,8	8708,1	8029,4	7799,5	7436,8	7324,0	

Sursa: C. Moineagu, I. Negură, V. Urseanu, *Statistica*, Editura științifică și enciclopedică, București, 1976, p. 302.

131 562 căsătorii în 1957, se ajunge la 87 888, în anul 1966. O ajustare analitică, cu ajutorul unei parabole de gradul doi, de pildă, ar pune în evidență *tendința generală* *. Să precizăm că numărul căsătoriilor din mediul rural a scăzut, în perioada considerată, atît ca urmare a diminuării numărului absolut al populației rurale, cît mai ales datorită scăderii frecvenței căsătoriilor. Rata de nupțialitate (număr de căsătorii la 1 000 locuitori) a scăzut de la aproximativ 11 la 7,4‰. Cauzele sînt multiple și este de resortul analizei demografice explicarea lor. Dar pentru a preveni o eroare care poate fi sugerată chiar de forma graficului din fig. 82, variația lunară a căsătoriilor s-a modificat foarte puțin. Tipul de sezonalitate este următorul: o frecvență redusă a căsătoriilor în lunile decembrie, ianuarie, martie, aprilie, mai, iunie, iulie și august și frecvențe mari în lunile februarie, octombrie și decembrie. Influența factorilor economici, pe de o parte (activitatea agricolă), iar pe de altă parte, a tradiției (modelul cultural țărănesc) este evidentă. Fig. 83 și 84 pun bine în evidență aceste particularități.

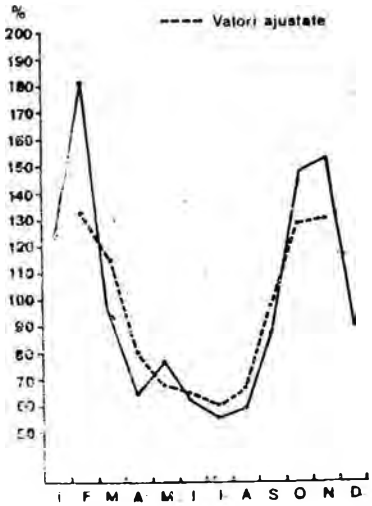


Fig. 83 Curba căsătoriilor lunare ajustate cu media de trei termeni, 1957—1966 (raportate în % față de media lunară)

Dacă „tipul" a rămas practic neschimbat — repartiția are forma de U — în schimb amplitudinea sezonalității s-a redus, după cum rezultă din fig. 84. Acest exemplu ne înlesnește prezentarea noțiunilor și metodelor de analiză a sezonalității s.c.

Indice sezonier (de sezonalitate) — Măsură a sezonalității, de obicei ca expresie relativă (în procente) a abaterii valorilor empirice ale s.c. de la anumite valori, calculate.

Acestea din urmă pot fi simple medii, *medii mobile* * sau valori ale tendinței, obținute printr-o ajustare analitică.

Indice sezonier simplu neajustat. I. obținut din raportarea valorilor empirice ale s.c. la media valorilor.

Din tab. 55 calculăm mediile fiecărei luni în perioada 1957—1966 și media lunară pentru toate lunile ale perioadei.

Seria i.sezonieri este trecută în fig. 79; ea exprimă tipul de sezonalitate. Fiind vorba de abateri de la media generală (media lunară a perioadei 1957—1966), este important să stabilim și abaterile valorilor lunare din cuprinsul

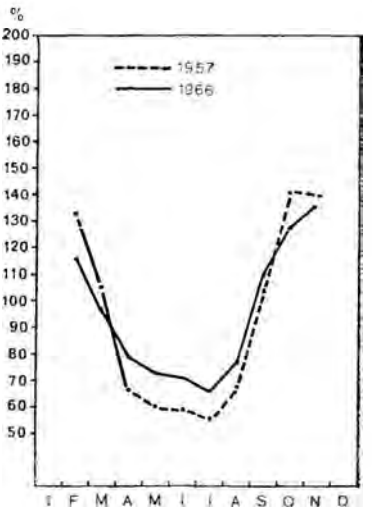


Fig. 84 Curba căsătoriilor lunare ajustate cu media mobilă de trei termeni, 1957 și 1966 (rapoarte în % față de media lunară

fiecărui an de la media lunară a anului respectiv. Procedînd astfel, obținem valorile din tab. 57.

Transpunerea valorilor i. sezonieri pe un grafic (fig. 85) ne arată dispersia valorilor individuale lunare de la media lunară pe 10 ani.

Tabelul 56

Mediile lunare și indicele sezonier 1957—1966

Luna	Numărul mediu lunar	Indice sezonier (%)	Luna	Numărul mediu lunar	Indice sezonier (%)
Ianuarie	11 317	125,7	Iulie	5 029	55,9
Februarie	16 243	180,5	August	5 372	59,7
Martie	8 426	93,6	Septembrie	7 868	87,4
Aprilie	6 009	66,8	Octombrie	13 369	148,9
Mai	6 922	76,9	Noiembrie	13 719	152,4
Iunie	5 614	62,4	Decembrie	8 124	90,3
			Media	9 001	100,0

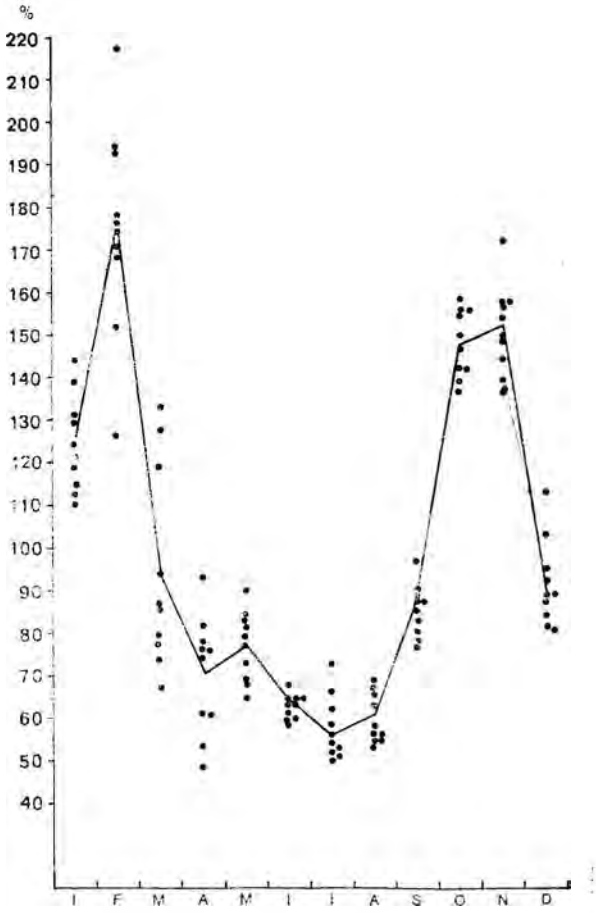


Fig. 85 Indicii sezonieri în perioada 1957—1966

Pentru unele luni, valorile arată o împrăștiere mare : afirmația este valabilă pentru lunile februarie, martie și aprilie, influențate de variația calendaristică a tradiției. Celelalte luni se caracterizează printr-o împrăștiere mai redusă. Pentru a pune mai bine în evidență tipul de sezonalitate, unii autori recomandă să se elimine valorile extreme și să se calculeze un i. sezonier luind un număr de valori care se situează în mijlocul seriei. De pildă, în cazul nostru media

Tabelul 57

Indicii sezonieri pentru fiecare an din perioada 1957—1966

Anul	Ianuarie	Februarie	Martie	Aprilie	Mai	Iunie	Iulie
1957	123,9	193,1	79,8	48,7	68,0	61,2	50,2
1958	144,5	194,3	67,0	61,4	81,8	58,7	52,3
1959	130,9	174,2	128,1	53,7	65,0	62,5	51,4
1960	138,7	217,0	77,3	61,4	69,9	62,2	54,6
1961	130,3	178,1	85,7	78,3	84,6	59,3	53,0
1962	110,1	126,6	119,3	76,7	77,1	68,4	58,9
1963	120,9	168,6	86,0	82,1	90,0	64,0	56,0
1964	112,9	170,4	133,3	75,2	68,7	64,5	72,9
1965	113,8	176,3	94,6	93,5	73,5	64,9	62,2
1966	119,6	152,3	74,0	76,0	83,5	60,0	66,5

Tabelul 57

—continuare—

Anul	August	Septembrie	Octombrie	Noiembrie	Decembrie	Media lunară
1957	55,0	96,9	156,2	171,9	95,2	100,0
1958	56,6	90,3	156,3	156,6	81,1	100,0
1959	54,6	87,9	150,4	150,0	89,4	100,0
1960	56,4	78,9	136,7	144,3	103,5	100,0
1961	55,9	88,1	142,7	154,7	89,5	100,0
1962	58,7	83,6	139,6	157,5	87,6	100,0
1963	63,7	84,9	142,0	158,0	84,3	100,0
1964	67,9	83,0	146,9	136,1	80,9	100,0
1965	66,5	86,1	155,0	148,5	92,0	100,0
1966	68,2	89,3	158,3	139,2	113,1	100,0

se va calcula din 8 termeni, renunțîndu-se la doi termeni extremi. Pentru aceasta este necesar însă să folosim un tabel cu rangurile valorilor.

Fiecare medie s-a calculat ca medie aritmetică simplă din valorile avînd rangurile 2—8. Noua serie de i. sezonieri este reprezentată în fig. 86 sub formă de linie frîntă. Se observă că ea trece în general prin mijlocul norului de puncte care reprezintă i. sezonieri individuali.

Indice sezonier ajustat. I. ale cărui valori reprezintă abateri de la o serie în prealabil ajustată fie cu metoda mediilor mobile, fie cu o metodă de ajustare analitică. Numărul de termeni luați în calculul mediilor mobile este diferit. Exemplul din tab. 59 se bazează pe medii calculate din 12 termeni (numărul lunilor). Fiind vorba de ilustrarea unei metode ne limităm la 3 ani (1957—1959).

Tabelul 58

Rangurile indicilor sezonieri ai căsătoriilor pe luni, perioada 1957—1966

Rangul	Ianuarie	Februarie	Martie	Aprilie	Mai	Iunie
1	144,5	217,0	133,3	93,5	90,0	68,4
2	138,7	194,3	128,1	82,1	84,6	64,9
3	130,9	193,1	119,3	78,3	83,5	64,5
4	130,3	178,1	94,6	76,6	81,8	64,0
5	123,9	176,3	86,0	76,0	79,9	62,5
6	120,9	174,2	85,7	75,2	77,1	62,2
7	119,6	170,4	79,8	61,4	73,5	61,2
8	113,8	168,6	77,3	61,4	68,7	60,0
9	112,9	152,3	74,0	53,7	68,0	59,3
10	110,1	126,6	67,0	48,7	65,0	58,7
Media din 8 termeni	123,9	175,9	93,1	70,6	77,1	62,3
Indice (%)	124,3	176,4	93,4	70,8	77,3	62,5

Tabelul 58

— continuare —

Rangul	Iulie	August	Septembrie	Octombrie	Noiembrie	Decembrie	Media
1	72,9	68,2	96,9	158,3	171,9	113,1	...
2	66,5	67,9	90,3	156,3	158,0	103,5	...
3	62,2	66,5	89,3	156,2	157,5	95,2	...
4	58,9	63,7	88,1	155,0	156,6	92,0	...
5	56,0	58,3	87,9	150,4	154,6	89,5	...
6	54,6	56,6	86,1	146,9	150,0	89,4	...
7	53,0	56,4	84,9	142,7	148,5	87,0	...
8	52,3	55,9	83,6	142,0	144,3	84,3	...
9	51,4	55,0	83,0	139,6	139,2	81,1	...
10	50,2	54,6	78,9	136,7	136,1	80,9	...
Media din 8 termeni	56,9	60,1	86,7	148,6	151,1	90,3	99,7
Indice (%)	57,1	60,3	87,0	149,0	151,6	90,6	100,0

Valorile reprezentînd totalurile anuale mobile sînt plasate la intervalul dintre două luni. De pildă, totalul pe 1957 de 131 562 căsătorii este trecut între luna iunie și iulie și a.ș.m.d. Întrucît valorile reale corespund fiecărei luni calendaristice, este necesar ca și valorile ajustate (col. 7) să respecte același principiu, drept care se aplică o metodă specială.

Tabelul 59

Numărul căsătoriilor pe luni și calculul numărului ajustat cu media mobilă pe 12 luni

Anul și luna	Numărul de căsătorii	Total mobil de 12 luni	Media mobilă de 12 luni (col. 3 : 12)	Total mobil de 2 luni	Media mobilă centrată de 12 luni (col. 5 : 2)	Indice sezonier (col. 2 : col. 6) %
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1957						
Ian.	13 583					
Febr.	21 167					
Mart.	8 750					
Apr.	5 335					
Mai	7 158					
Iunie	6 705	131 562	10 964			
Iulie	5 504	134 245	11 187	22 151	11 076	49,7
Aug.	6 025	134 945	11 245	22 432	11 216	53,7
Sept.	10 627	133 733	11 144	22 389	11 195	94,9
Oct.	17 124	135 306	11 276	22 420	11 210	152,8
Nov.	18 842	137 053	11 421	22 697	11 349	166,0
Dec.	10 442	136 960	11 413	22 834	11 417	91,5
1958						
Ian.	16 266	137 364	11 447	22 860	11 430	142,3
Febr.	21 867			22 923	11 462	190,8
Martie	7 538	137 706	11 476	22 913	11 457	65,8
Apr.	6 908	137 241	11 437	22 913	11 457	60,3
Mai	9 205	137 707	11 476	22 841	11 421	80,6
Iunie	6 612	136 384	11 365	22 621	11 311	58,5
Iulie	5 908	135 068	11 256	22 245	11 123	53,1
August	6 367	131 873	10 989	21 606	10 803	58,9
Sept.	10 162	127 401	10 617	21 672	10 836	93,8
Oct.	17 590	132 651	11 055	21 981	10 991	160,0
Nov.	17 519	131 107	10 926	21 726	10 813	162,0
Dec.	9 126	128 391	10 700	21 368	10 684	85,4
		128 020	10 668			

(continuarea tabelului 59

(1)	(2)	(3)	(4)	(5)	(6)	(7)
1959						
Ian.	13 071			21 272	10 636	122,9
Febr.	17 395	127 250	10 604	21 132	10 566	164,5
Martie	12 791	126 339	10 528	20 941	10 471	122,2
Apr.	5 361	124 952	10 413	20 612	10 306	52,0
Mai	6 492	122 382	10 199	20 186	10 093	64,3
Iunie	6 283	119 839	9 987	19 957	9 991	62,5
Iulie	5 138	119 645	9 970	19 982	9 991	51,4
August	5 456	120 143	10 012	20 342	10 171	53,6
Sept.	8 775	123 960	10 330	20 224	10 112	86,8
Oct.	15 020	118 730	9 891	19 843	9 922	151,4
Nov.	14 976	119 382	9 949	20 008	10 004	119,7
Dec.	8 932	120 711	10 059	20 106	10 053	88,8
1960		120 558	9 894			
Ian.	13 569					
Febr.	21 212					
Martie	7 561					
Apr.	6 013					
Mai	7 821					
Iunie	6 085					

Metoda centrării mediilor mobile. (→ *Medie mobilă.*) Metodă constînd în transformarea valorilor ajustate cu media mobilă și care se situează între două valori lunare alăturate în valori care să corespundă fiecărei luni calendaristice. Astfel în tab. 59 în col. 4 cifra de 10 964 arată numărul de căsătorii situat între iunie și iulie, 11 187 este cifra dintre lunile iulie și august etc. Pentru a le transforma în valori lunare, se calculează totalul mobil pe două luni (col. 5), iar prin împărțirea la 2 se obține valoarea ajustată ce corespunde lunii calendaristice, (col. 6). I. sezonier ajustat (col. 7) se calculează prin raportarea valorilor reale (y) din col. 2 la valorile ajustate (Y) din col. 6 și figurează în col. 7. Reprezentarea grafică (fig. 86) arată cum se situează valorile individuale în raport cu valorile ajustate.

I.s.a. se poate calcula și în raport cu valorile ajustate după alte metode. De pildă, pentru perioada 1957—1966, numărul căsătoriilor, în scădere, evoluează

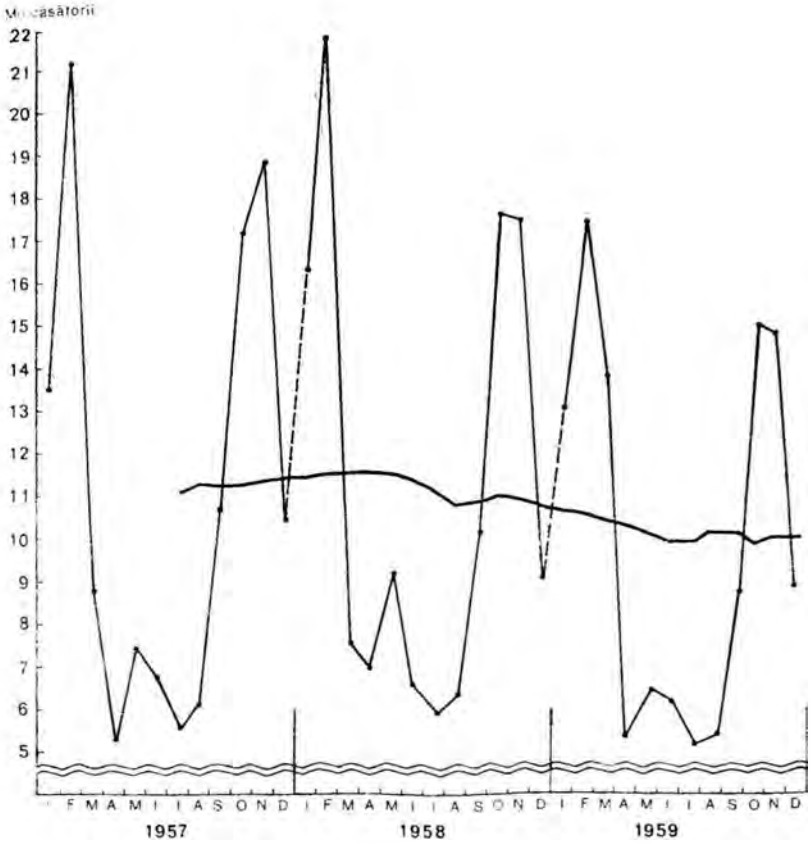


Fig. 56 Curba căsătoriilor din mediul rural 1957—1959 (reală și ajustată cu media centrată de 12 luni)

după o tendință liniară. Ajustarea cu dreapta ne-ar duce la obținerea valorilor ajustate Y cu care s-ar compara valorile empirice, adică la obținerea unei noi serii de *i.s.a.*

Întrucît avem de-a face cu medii și cu abateri de la medie, se impun diferite calcule privind *dispersia* *, *abaterea medie pătratică* *. (→ *Analiză de varianță.*)

Rapoarte în lanț. I. sezonieri calculați (în procente) prin raportarea fiecărui termen y la termenul imediat precedent. (→ *Indice în lanț.*) Pentru seria lunară medie 1957—1966, aceste rapoarte în lanț sînt :

Tabelul 60

Rapoarte în lanț pentru seria 1957—1966

Luna	Numărul mediu de căsătorii	Raport în lanț (%)
Ianuarie	11 317	—
Februarie	16 243	143,5
Martie	8 426	51,9
Aprilie	6 009	71,3
Mai	6 922	115,2
Iunie	5 614	81,1
Iulie	5 029	89,6
August	5 372	106,8
Septembrie	7 868	146,5
Octombrie	13 369	169,9
Noiembrie	13 719	102,6
Decembrie	8 124	59,2

Raportul cu bază fixă este 71,8%, iar media *n.l* 103,4%. Aceste rapoarte sînt și ele prelucrate cu diferite metode și permit calculul unor coeficienți sezonieri.

Rezultatele unei analize a sezonalității cu metode moderne (metoda G. Calot) referitoare la un fenomen demografic al populației României sînt prezentate în fig. 87.

Metoda Calot a fost expusă în 1973 și 1981. Meritul ei principal constă în aceea că ea este folosită și ca model prospectiv.

ANALIZA VARIAȚIILOR (FLUCTUAȚIILOR) CICLICE. Operație complexă constînd din aplicarea metodelor statistico-matematice asupra s.c. empirice în vederea identificării componentei ciclice. Domeniul predilect de aplicare a **a.v.c.** este *ciclul economic* *, proces caracteristic economiei capitaliste. Există însă cicluri în economia socialistă legate de formarea și reproducția fondurilor fixe, forței de muncă, resurselor energetice, de dezvoltarea cercetării științifice și inovarea tehnologică, de schimbarea structurii economiei naționale. Există cicluri abrupte și cicluri combinate.

Într-o s.c. poate surveni o schimbare de sens (crescător sau descrescător). Dacă un termen din s.c. depășește pe cel anterior și pe cel următor ($y_{t-1} < y_t > y_{t+1}$), el se numește vîrf (peak) ; în cazul contrar ($y_{t-1} > y_t < y_{t+1}$), el este o *adîncitură* (*trough*). Intervalul între două schimbări de sens succesive este o *fază*, iar intervalul dintre două vîrfuri (sau adînciturі) a două oscilații este *perioada*. O metodă curentă de **a.v.e.** este metoda reziduurilor, după formula :

$$\frac{D' - T}{T},$$

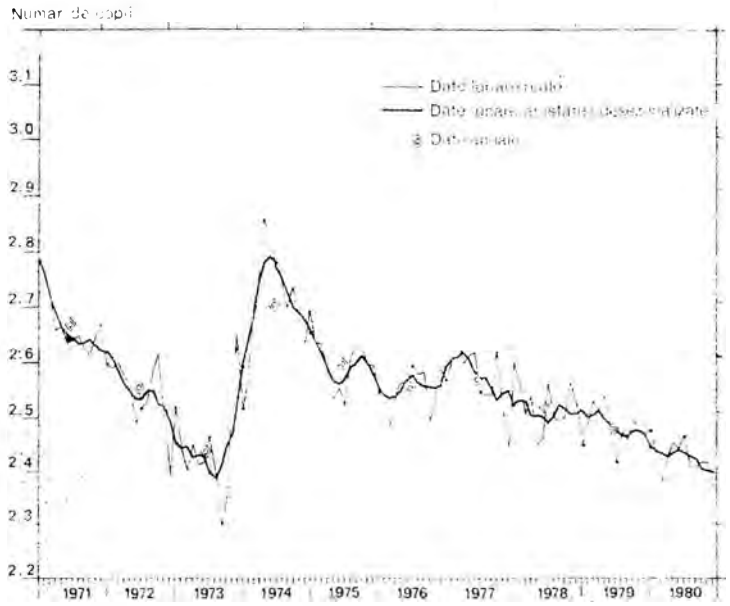


Fig. 87 România. Rata totală de fertilitate (număr de copii), 1971—1980

în care D' — date brute ale variațiilor ciclice, corectate din punctul de vedere al variațiilor sezoniere, iar T — tendința de lungă durată. Formula ne dă *mișcarea ciclică* *.

Analiză armonică a seriilor cronologice. Analiză bazată pe considerarea unei s.c ca sumă a funcțiilor periodice de forma :

$$x(t) = \sin \frac{2\pi}{T} t \text{ și } y(t) = \cos \frac{2\pi}{T} t,$$

în care : T_1, T_2, . . ., T_n sînt perioadele (necunoscute) ale acestor funcții. Se pot determina T_1, T_2, . . ., T_n alegîndu-se funcții periodice cunoscute și făcîndu-le să varieze, prin atribuirea succesivă a diferitelor valori, ajungîndu-se astfel la o sumă a acestor funcții care dă cea mai bună aproximare a variațiilor seriei observate.

Periodogramă. Reprezentarea grafică a *analizei armonice* *.

Analiză a periodogramei. Analiză a s.c. care se bazează pe ideea că o s.c. strict periodică poate fi exprimată ca sumă a numerelor valurilor (undelor), fiecare reprezentată de un sinus sau un cosinus.

Presupunem o variabilă y ca funcție de timp : $y_t = f(t)$. Dacă $f(t + T) = f(t)$ pentru toate valorile lui t, atunci y_t poate fi exprimată ca serie Fourier :

$$Y_t = \frac{1}{2} A_0 + \sum_{j=1}^{\infty} \left(A_j \sin \frac{360_{jt}}{T} + B_j \cos \frac{360_{jt}}{T} \right),$$

în care : T — perioada de oscilație ;

$$A_0,\ A_j \text{ și } B_j \text{ — constante.}$$

Se demonstrează că :

$$R_j^2 = A_j^2 + B_j^2,$$

în care : R_j — amplitudinea (valoarea maximă a lui y), corespunzătoare termenului armonic cu perioada $\frac{T}{j}$.

Atît periodograma cît și spectrul unei serii pot fi concepute ca descompuneri ale variației seriei la diferite frecvențe. Deoarece majoritatea s.c. economice nu sînt strict periodice, estimația statistică a spectrului este preferabilă *a.p.* În general, lipsurile *a.p.* pot fi înlăturate printr-o metodă mai generală care este *analiza spectrală* *.

Corelație serială a seriei cronologice. Corelație între termenii vecini ai unei s.c. care nu este determinată de schimbările variabilei independente (adică a mediei populației statistice) ; *c.s.s.c.* este corelația dintre abaterile succesive ca raport la mediile populației statistice. Se mai numește autocorelație și este exprimată de coeficientul de corelație al momentelor mixte (*Coeficient de corelație*). Este deci corelația între x_s și x_t într-o s.c. unde $s \neq t$.

Autocorelație de ordinul 1. Coeficient de corelație între doi termeni vecini ai s.c., x_t și x_{t+1}.

Autocorelație de ordinul k (p_k). Coeficient de corelație între doi termeni separați de $(k - 1)$ termeni :

$$p_k = \frac{\text{cov}\,(y_t,\ y_t + k)}{\sqrt{\{\text{var}\ y_t\ \text{var}\ y_{t+k}\}}}.$$

Valoarea teoretică a lui p_k pentru orice k (cu excepția lui zero). Graficul având pe t_k ca ordonată și pe k ca abscisă se numește *corelogramă* *.

Există diferite teste pentru autocorelație. Cel mai simplu test pentru x_t și x_{t-j} pentru o valoare dată a lui j este :

Test neparametric (Wald și Wolfowitz, 1943). Test circular statistic pe tipul :

$$R = \sum_{t=1}^{N-1} x_t\, x_{t+1} + x_n\, x_1,$$

care, pentru N mare, este repartizat normal. Ipoteza nulă testată înseamnă că valorile sînt independente. Mai există testul numit *raportul lui von Neumann*. Alte teste sînt mai rafinate.

Se menționează că modelul general al unei s.e. este un *proces stohastic.* (→ *Proces aleator* (*stohastic.*)

Variație concomitentă a seriei cronologice. Evoluție în timp a două (sau mai multe) s.c. reprezentînd fenomene total independente. Pentru a pune în evidență *v.c.s.c.* comparăm fie variațiile tendinței de lungă durată a s.c. (covariație tendențială), fie variațiile abaterilor s.c. de la linia de tendință (covariația diferențială). În acest scop formăm diferențele :

$$U_t = u_t - \bar{u} \text{ și } V_t = v_t - \bar{v},$$

unde : u_t și v_t sînt valorile observate ale variabilelor u și v la momentul t, iar $\bar{u}$ și $\bar{v}$ sînt mediile aritmetice ale acestora bazate pe n observații.

Indice simplu de covariație tendențială :

$$i = \frac{c - d}{n},$$

în care : c — numărul de cazuri de concordanță a semnelor (U_t și V_t au același semn) și

d — numărul de discordanțe (U_t și V_t sînt de semn contrar).

Indice simplu de covariație diferențială. I. similar cu cel de sus, cu particularitatea că diferențele se formează nu în raport cu $\bar{u}$ și $\bar{v}$, ci în raport cu valorile teoretice $u_{(t)}$ și $v_{(t)}$.

ANALIZĂ SPECTRALĂ A SERIEI CRONOLOGICE Aplicarea și adaptarea unei metode de analiză din științele naturii (fizică, astronomie, optică), avînd ca bază ideea de *spectru,* cu care este asimilat un fenomen reprezentat de s.c., cu toate componentele sale (tendință, sezonalitate etc.) Spectrul este definit ca o diagramă de descriere a fiecărei frecvențe a unei s.c. *A.s.c.* este o generalizare a teoriei descompunerii formulate de matematicianul J. Fourier (1768—1830) și are drept scop să studieze proprietățile intrinseci ale seriilor și amplitudinile componentelor periodice ; de asemenea, prin analiza încrucișată, ea pune în evidență legăturile dintre serii, măsurînd decalajele dintre ele. O primă aplicație a fost studiul valurilor mării la țărm.

Fie o s.c. x_t :

$$x_t = \Sigma A_j \sin(\omega_j t + \alpha_j) + \varepsilon_t = \Sigma B_j \cos(\omega_j t) + \Sigma C_j \sin(\omega_j t) + \varepsilon_t. \quad (17)$$

Seria este suma numărului de componente armonice și a unui reziduu necorelat (ε_t).

Dacă frecvențele ω_j (corespunzătoare variațiilor lunare, zilnice etc.) sînt cunoscute, în așa fel încît A_j și α_j urmează să fie estimate, atunci pe baza unei serii observate $x_1, x_2, \ldots, x_n$, estimatorii, pe baza metodei celor mai mici pătrate, sînt aproximativ :

$$\hat{B}_j = \frac{2}{n} \sum_{t=1}^{n} x_t \cos(\omega_j t)$$

$$\hat{C}_j = \frac{2}{n} \sum_{t=1}^{n} x_t \sin(\omega_j t).$$

Amplitudinea pătrată $A_j^2 = B_j^2 + C_j^2$ este estimată aproximativ cu

$$\hat{A}^2 = \frac{4}{n^2}\{[\Sigma x_t \cos(\omega_j t)]^2 + [\Sigma x_t \sin(\omega_j t)^2]\} = \frac{4}{n}\sum_{s=1}^{n}\sum_{t=1}^{n} x_s x_t \cos[\omega_j(s-t)]).$$

Se mai poate scrie :

$$\hat{A}_j^2 = \frac{4}{n^2}\left|\sum_{t=1}^{n} x_t \exp(-i\,\omega_j t)\right|^2 .$$

Importanța lui $\hat{A}_j^2$ constă în faptul că el măsoară descreșterea în suma reziduală a pătratelor. Cu cît este mai mare această cantitate cu atît mai mare este contribuția pe care componenta armonică a frecvenței ω_j aduce la variația lui x_t.
De aceea, dacă ω_j sînt necunoscute, pot fi căutate periodicitățile lor calculînd o cantitate analoagă cu formula $\hat{A}_j^2$ pentru variabila ω.
Periodograma va fi :

$$f_n(\omega) = \frac{1}{n}\left|\sum_{t=1}^{n} x_t \exp(-i\,\omega t)\right|^2 =$$

$$= \frac{1}{n}\sum_{s=1}^{n}\sum_{t=1}^{n} x_s x_t [\omega(s-t)] =$$

$$= \frac{1}{n}\{[\Sigma x_t \cos(\omega t)]^2 + [\Sigma x_t \sin(\omega t)]^2\}.$$

Se poate determina și o *funcție de densitate spectrală.*

Metoda **a.s.s.c.**, de dată relativ recentă în analiza economică, a cunoscut în ultimii ani dezvoltări promițătoare (Jenkins, O. Morgenstern, C. W. Granger). **EXTRAPOLAREA A SERIEI CRONOLOGICE.** Operație de stabilire a valorii unor termeni ai s.c. (de momente sau de intervale) care se situează în afara orizontului de analiză. **E.s.c.** presupune : a. adoptarea unui model de evoluție $y_t = f(t)$ și b. introducerea în model a valorii convențională a variabilei timp corespunzătoare momentului pentru care se efectuează extrapolarea. În general, extrapolarea devine posibilă după ce a fost determinată *tendința generală a s.c.**, cu unul din procedeele de ajustare; în acest caz, tendința este extrapolată pentru un orizont oarecare. O asemenea extrapolare se numește *exploativ-tendențială*. Se recomandă ca lungimea perioadei de extrapolare să nu depășească lungimea perioadei pentru care analiza s.c. a determinat tendința generală. Extrapolarea se folosește în prognoză sau previziune (în demografie termenul este de „proiectare"), în studiul *conjuncturii economice* *, în *econometrie* *, în demografie. În prognoza economică predomină metoda normativă. În funcție de modelul adoptat și de lungimea perioadei (scurtă, medie, lungă și foarte lungă), extrapolarea este însoțită de eroare. Fiind vorba de o *estimație punctuală*, se stabilește un interval al prognozei și se determină eroarea *ex ante* și *ex post*.

Exemplu. Populația României la 1 ianuarie 1983 a fost de 22 527 235. Între 1 ianuarie 1970 și 1 ianuarie 1981, rata medie anuală de creștere a fost de 0,9%; între 1 ianuarie 1978—1 ianuarie 1983, a fost de 0,7%. Extrapolarea acestor tendințe de tip exponențial ne dă următoarele valori:

Anul	Numărul populației la 1 ianuarie	
	$y_n = y_0(1{,}009)^n$	$y_n = y_0(1{,}007)^n$
1983	22 527 235	22 527 235
1984	22 729 980	22 684 925
1985	22 934 549	22 843 719
1986	23 140 959	23 003 625
1987	23 349 227	23 164 650
1988	23 559 342	23 326 802
1989	23 771 138	23 490 089
1990	23 985 318	23 654 519
1991	24 201 185	23 820 100

Populația la 1 iulie 1990 este de 24 093 251, (prima variantă) și de 23 737 250 (a doua variantă). Diferența între cele două variante este de — 1,5%. Extrapolarea se poate face cu tendință parabolică, logistică etc.

CICLU ECONOMIC (*Ciclu de afaceri*). Succesiune de faze în desfășurarea totalității elementelor care determină situația economică a unei țări (capitaliste). Principalele faze ale c.e. sînt expansiunea și contracția economică **C.e.** capitalist are patru faze: criza, depresiunea, înviorarea și avîntul (*boom*) **C.e.** poate fi: de durată mică, mare și de durată foarte mare („ciclurile Kondratieff"). Particularitățile **c.e.** exprimă influența factorilor dezvoltării economice și constituie un domeniu predilect al economiei politice. (→ *Conjunctură economică*; *Econometrie.*) Necesitatea de a pune în evidență **c.e.** a dus la dezvoltarea analizei s.c., la elaborarea diferitelor modele și metode (*variația ciclică* *: *mișcarea ciclică* *; → și *Istoric*).

CONJUNCTURĂ ECONOMICĂ (exact: studiu al conjuncturii economice). Domeniu de aplicare a metodelor statistico-matematice și, în special, a analizei **s.e.** în scopul caracterizării evoluției probabile a fenomenelor economice pe termen scurt și mediu. Întemeietorul **s.c.e.** este considerat *W. M. Persons* * de la Universitatea Harvard, cu al său studiu: „Introducerea barometrului economic pe baza datelor anuale" (1916), iar prima instituție consacrată **s.c.e** este considerat Comitetul de Cercetări Economice al Universității Harvard (1917). După acest prototip au fost înființate institute de conjunctură în Germania (1925), în Anglia („London and Cambridge Economic Service"), în același an. În anii 1930 a fost înființat Institutul de Conjunctură Economică la București, condus de Virgil Madgearu și la care au colaborat, printre alții, N. Georgescu-Roegen, P. Sterian, R. Moldovan, Gh. Rădulescu. În 1967 a fost înființat un Institut pentru Studiul Conjuncturii Economice pe lîngă Ministerul Comerțului Exterior și Cooperării Economice Internaționale.

ECONOMETRIE. Disciplină creată în anii '30 constînd din îmbinarea teoriei economice, a matematicii și statisticii. În 1930 a fost înființată societatea de econometrie (Econometric Society), datorită inițiativei unor economiști, statisticieni și matematicieni printre care *I. Fisher* *, *L. von Bortkiewicz* *,

F. Divisia *, *H. Hotelling* *: din 1933 apare revista „Econometrica". În primul număr din 1933, R. Frisch a dat următoarea definiție a e. „Experiența a arătat că fiecare din următoarele trei puncte de vedere, al statisticii, al teoriei economice și al matematicii, este o condiție necesară, dar nu suficientă pentru o înțelegere efectivă a relațiilor cantitative din economia modernă; unificarea lor este aceea care asigură eficiența. E. este tocmai această unificare". Inițial, e. își concentra atenția asupra studierii ciclurilor economice, elaborînd metode de studiere a conjuncturii economice, de determinare și de extrapolare a tendințelor, domeniu în care contribuția econometricienilor a fost foarte importantă. Treptat, e. și-a lărgit domeniul de studiu. O contribuție românească importantă este e. *informațională Onicescu*.

BIBLIOGRAFIE

1. ANDERSON, O. D., *Time series analysis and forecasting — the Box-Jenkins approaches*, Butterworts, London—Boston, 1977.
2. BIJI, ELENA, NEGURĂ, ION. *Aplicații statistice în conducerea activității din agricultură*, Ed. Ceres, București, 1980.
3. BOX, G. E. P., JENKINS, G. M., *Time series analysis, forecasting and control*, Holden Day, San Francisco, 1970.
4. CALOT, G., *Cours de statistique descriptive*, Deuxième édition, Dunod Paris, 1973, (cap. 7 și 8).
5. CALOT, G., *L'observation de la fécondité à court et moyen terme*, în Population (INED), 36e année, numéro 1, Janvier—Février, 1981.
6. CROXTON, F. E., COWDEN, I. J., KLEIN, S., *Applied General Statistics*, Third Edition, Prentice-Hall, Inc. Englewood Cliffs, N.Y. 1967 (cap. 11, 12, 13, 14, 15 și 16).
7. DEMETRESCU, M. C., *Elasticitatea cererii populației*, Ed. Academiei R.S.R., București, 1967.
8. DICȚIONAR STATISTIC-ECONOMIC (sub red. C. Ionescu). Direcția Centrală de Statistică, București, 1969.
9. GEORGESCU-ROEGEN, N. St., *Metoda statistică. Elemente de statistică matematică*, Ed. Institutului de Statistică generală a statului, București 1933 (cap. XIII).
10. GRANGER, C. W. J., *Analyse spectrale des séries temporelles en économie*, (prefață de Oskar Morgenstern), Dunod, Paris, 1969.
11. INTERNATIONAL ENCYCLOPEDIA OF THE SOCIAL SCIENCES, David L. Sills (Ed), The MacMillan Company & The Free Press, vol. 2, 16, 1968.
12. JENKINS, G. M., *A Survey of Spectral Analysis*, in Applied Statistics, 14, 1965.
13. KOOPMANS, L. M., *The spectral analysis of time series*, Academic Press, New York, 1974.
14. MARINESCU, I., *Metode de previziune pentru fenomene economice cu variații sezoniere*, în „Revista de Statistică" nr. 9, 1968.
15. MIHOC, GH. ș.a., *Teoria matematicii în operațiile financiare*, vol. I și II, Ed. științifică, București, 1960.
16. MIHOC, GH., URSEANU, V., *Matematici aplicate în statistică*, Ed. Academiei, București, 1962 (cap. XII, XIII și XIV).
17. MOINEAGU, C., NEGURĂ, I., URSEANU, V., *Statistica*, Ed. științifică și enciclopedică, București, 1976.

18. NEGURĂ, I., DINULESCU, S., GABURICI, A., *Cercetarea operațională în agricultură*, Ed. Ceres, București, 1975.
19. PAENSON, I., *Systematic glossary of the terminology of statistical terms*, Pergamon Press, Oxford — New York etc., 1970.
20. PECICAN, E., *Analiza spectrală și descompunerea seriilor dinamice*, în „Revista de statistică" nr. 8, 1973.
21. SHRYOCK, H. S., SIEGEL, J. S. and ass., *The Methods and Materials of Demography*, vol. 1 și 2, Bureau of the Census, Washington, 1971.
22. TEODORESCU, V., *Conceptul unui sistem de programe pentru crearea și administrarea unei baze de date a seriilor cronologice*, în „Revista de Statistică" nr. 12, 1979.
23. TÖVISSI, I., ISAIC-MANIU, AL., ANTONESCU, C., *Statistica*, Curs A.S.E., București, 1982.
24. VIȘINOIU, N., *Utilizarea analizei spectrale în studiul transporturilor de călători*, în „Studii și Cercetări de Calcul Economic și Cibernetică Economică" nr. 3, 1978.
25. WOLD, H. (ed.), *Bibliography on Time Series and Stochastic Processes*, International Statistical Institute, Oliver & Boyd, Edinburgh, 1965.
26. YEOMANS, K. A., *Introducing statistics. Statistics for the social scientist*, Volume one, cap. 6, Volume two, *Applied statistics*, cap. 5 Penguin Books, 1968.
27. YULE, G. U., KENDALL, M. G., *Introducere în teoria statisticii*, Ed. științifică, București, 1969.
28. NEWBOLD, P., *Some Recent Developments in Time Series Analysis*, în „International Statistical Review" 49 (1981), 53—66.

TEORIA PROBABILITĂȚILOR ȘI STATISTICA MATEMATICĂ

XII. ELEMENTE DE TEORIA PROBABILITĂȚILOR

SPAȚIU DE EVENIMENTE ELEMENTARE. Primul concept fundamental al teoriei probabilităților. Considerînd un experiment aleator, adică un experiment ale cărui rezultate nu pot fi prevăzute cu exactitate, cum ar fi aruncarea unei monede sau a unui zar etc., prin **s.e.e.** asociat acestuia se înțelege totalitatea rezultatelor sale posibile. (Evident, de aici rezultă că *eveniment elementar* este sinonim cu *rezultat posibil.*)
Notația tradițională pentru un **s.e.e.** este Ω. De exemplu, dacă o monedă este aruncată o singură dată, atunci notînd cu s și v, respectiv apariția stemei și apariția valorii, **s.e.e.** este $\Omega = \{s, v\}$. Dacă moneda este aruncată de două ori, **s.e.e.** corespunzător este $\Omega = \{ss, sv, vs, vv\}$. În cazul aruncării unui zar, o singură dată, $\Omega = \{1, 2, 3, 4, 5, 6\}$, iar dacă zarul este aruncat de două ori, atunci **s.e.e.** corespunzător constă din cele 36 de perechi de numere $(i, j,)$ $1 \leqslant i, j \leqslant 6$. Nu trebuie să se creadă că orice **s.e.e.** este finit. De exemplu, experimentul care constă în alegerea la întîmplare a unui număr natural are ca **s.e.e.** $\Omega = \{1, 2, 3, \ldots\}$, care este o mulțime infinită numărabilă, iar experimentul care constă în alegerea la întîmplare a unui număr real pozitiv are ca **s.e.e.** $\Omega = (0, \infty)$, care este o mulțime infinită nenumărabilă. În fapt, orice mulțime poate fi considerată ca **s.e.e.** asociat unui experiment aleator: (Punînd firește, în corespondență biunivocă elementele mulțimii cu rezultatele posibile ale unui experiment aleator convenabil construit.)

EVENIMENT (ALEATOR). Un **e.(a.)** (pe scurt **e.**) asociat unui experiment aleator, deci unui *spațiu de evenimente elementare* *, corespunde unui enunț legat de experiment, enunț care poate fi corect sau fals. Mulțimea evenimentelor elementare pentru care enunțul este corect constituie **e.** considerat. De exemplu, în cazul aruncării de două ori a unei monede, cînd $\Omega = \{ss, sv, vs, vv\}$, **e.** „rezultatul primei aruncări este identic cu rezultatul celei de a doua" este $\{ss, vv\}$. Astfel, un **e.** apare ca o *submulțime a spațiului de evenimente elementare.* Această afirmație trebuie precizată după cum urmează. În cazul în care *spațiul de evenimente elementare* * Ω este finit sau numărabil infinit se consideră ca **e.** asociate lui Ω *toate* submulțimile lui Ω, a căror clasă se notează $\mathscr{P}(\Omega)$. În cazul în care Ω este o mulțime infinită nenumărabilă nu este întotdeauna posibil să se considere ca **e.** asociate lui Ω toate submulțimile sale, întrucît se poate întîmpla să nu existe nici o *probabilitate* * pe $\mathscr{P}(\Omega)$. Ieșirea din această dificultate se bazează pe suprimarea calității de **e.** unor submulțimi ale lui Ω, considerînd ca **e.** *numai* elementele unei clase $\mathscr{K}$ de submulțimi ale lui Ω care are o structură de σ-*algebră* *

E. se notează cu litere mari $A, B, C, \ldots$. Întregul *spațiu de evenimente elementare* * Ω se numește **e.** *sigur*, iar **e.** care nu conține nici un eveniment elementar, deci mulțimea vidă $\emptyset$, se numește **e.** *imposibil.* Justificarea acestor denumiri este evidentă : e. sigur are loc în orice realizare a experimentului, iar cel imposibil nu are loc în nici o realizare a experimentului.

Se spune că **e.** A *implică* **e.** B, dacă $A \subset B$, adică A este o submulțime a lui B. Din nou, justificarea acestei terminologii este evidentă. Două **e.** A și B se numesc *echivalente* dacă fiecare implică pe celălalt, adică dacă $A = B$ în sensul teoriei mulțimilor.

Operații cu evenimente. Corespund construcției de enunțuri complexe pe baza unor enunțuri simple cu ajutorul operațiilor logice de disjuncție, conjuncție și negație.

Fie A și B **e.** în același *spațiu de evenimente elementare* Ω. Se definește *reuniunea* **e.** A și B, notată $A \cup B$, ca **e.** constînd din acele evenimente elementare care aparțin fie lui A, fie lui B, fie amîndurora. Se definește, de asemenea, *intersecția* **e.** A și B, notată $A \cap B$, ca e. constînd din acele evenimente elementare care aparțin atît lui A cît și lui B. În sfîrșit, se definește *complementarul* (sau *opusul*) **e.** A, notat A^c, ca mulțimea acelor evenimente elementare care nu aparțin lui A.

De exemplu, considerînd experimentul care constă în aruncarea unui zar o singură dată, cînd Ω constă din numerele întregi i, $1 \leqslant i \leqslant 6$, dacă

$$A = \{i \text{ este par}\}, \quad B = \{i \geqslant 4\},$$

atunci

$$A \cup B = \{i \text{ este par sau } i \geqslant 4\} = \{2, 4, 5, 6\},$$

$$A \cap B = \{i \text{ este par și } i \geqslant 4\} = \{4, 6\},$$

$$A^c = \{i \text{ este impar}\} = \{1, 3, 5\},$$

$$B^c = \{i < 4\} = \{1, 2, 3\}.$$

Mai general, reuniunea a n **e.** $A_1, \ldots, A_n$, notată $A_1 \cup \ldots \cup A_n$ sau $\bigcup_{i=1}^{n} A_i$, constă din acele evenimente elementare care aparțin cel puțin unuia din **e.** $A_1, \ldots, A_n$. Similar, reuniunea unui șir infinit $(A_i)_{i \geqslant 1}$ de **e.** notată $\bigcup_{i \geqslant 1} A_i$, constă din acele evenimente elementare care aparțin cel puțin unuia din **e.** A_i, $i \geqslant 1$.

De asemenea, intersecția a n **e.** $A_1, \ldots, A_n$, notată $A_1 \cap \ldots \cap A_n$ sau $\bigcap_{i=1}^{n} A_i$, constă din acele evenimente elementare care aparțin tuturor **e.** $A_1, \ldots, A_n$. Similar, intersecția unui șir infinit $(A_i)_{i \geqslant 1}$ de **e.**, notată $\bigcap_{i \geqslant 1} A_i$, constă din acele evenimente elementare care aparțin tuturor **e.** A_i, $i \geqslant 1$.

Două **e.** A și B se numesc *incompatibile* sau *disjuncte* dacă ele nu au nici un eveniment elementar comun, adică este imposibil ca atît A cît și B să apară simultan într-o aceeași realizare a experimentului. Cu alte cuvinte, A și B sînt incompatibile dacă $A \cap B = \emptyset$. Mai general, n **e.** $A_1, A_2, \ldots, A_n$ se numesc *disjuncte în ansamblul lor* dacă $A_i \cap A_j = \emptyset$, $i \neq j$, $1 \leqslant i, j \leqslant n$. Această înseamnă că nu mai mult decît unul din **e.** poate să apară într-o realizare a experimentului. Definiția se extinde, evident, și la cazul unei infinități numărabile de **e.** $A_1, A_2, \ldots$

Un sistem de **e.** disjuncte în ansamblul lor se numește *sistem complet* de **e.** dacă reuniunea lor este **e.** sigur. Cu alte cuvinte, un sistem complet de **e.** este o partiție a **e.** sigur.

Proprietățile cele mai importante ale operațiilor cu e. definite mai înainte sînt următoarele : (toate e. care apar fac parte din același *spațiu de evenimente elementare* *)

$$A \cup A = A,\ A \cap A = A,\ A \cap \Omega = A,\ A \cup \Omega = \Omega,$$

$$A \cup A^c = \Omega,\ A \cap A^c = \emptyset,\ A \cap \emptyset = \emptyset,$$

$$A \cup B = B \cup A,\ A \cup (B \cup C) = (A \cup B) \cup C,$$

$$A \cap B = B \cap A,\ A \cap (B \cap C) = (A \cap B) \cap C,$$

$$A \cap \left(\bigcup_{i \geqslant 1} B_i\right) = \bigcup_{i \geqslant 1} (A \cap B_i),$$

$$\left(\bigcup_{i \geqslant 1} A_i\right)^c = \bigcap_{i \geqslant 1} A_i^c,\ \left(\bigcap_{i \geqslant 1} A_i\right)^c = \bigcup_{i \geqslant 1} A_i^c$$

(ultimele două egalități sînt cunoscute sub denumirea de *regulile lui de Morgan*). Din ultimele trei egalități luînd $B_i = \emptyset$, $A_i = \emptyset$ și apoi $A_i = \Omega$ pentru $i > n$, se obține

$$A \cap \left(\bigcup_{i=1}^{n} B_i\right) = \bigcup_{i=1}^{n} (A \cap B_i),$$

$$\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c,\ \left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} A_i^c.$$

σ-algebră. Clasă $\mathscr{K}$ de submulțimi ale unui *spațiu de evenimente elementare** Ω cu următoarele proprietăți :

1. $\Omega \in \mathscr{K}$;

2. $A_1, A_2, \ldots \in \mathscr{K}$ implică $\bigcup_{i \geqslant 1} A_i \in \mathscr{K}$;

3. $A \in \mathscr{K}$ implică $A^c \in \mathscr{K}$.

În cuvinte, o σ-a. conține e. sigur Ω, odată cu un șir de elemente ale sale conține și reuniunea acestora și odată cu un element conține și complementarul său. Elementele unei σ-a. convenabil alese capătă în exclusivitate calitatea de e. asociate *spațiului de evenimente elementare* * Ω (→*Eveniment (aleator)*.)

PROBABILITATE. 1. Din punct de vedere filozofic, p. unui eveniment este o măsură a gradului de posibilitate a acestuia, grad de posibilitate care merge de la valoarea 0 (pentru evenimenul imposibil) pînă la valoarea 1 (pentru evenimentul sigur). Pentru un eveniment intermediar A, $0 \neq A \neq \Omega$, p. lui A trebuie să reflecte stabilitatea asimptotică constatată experimental (mai întîi în jocurile de noroc) a frecvenței lui A într-un număr arbitrar de mare de repetări independente ale experimentului aleator căruia A îi este asociat. În consecință, dacă A și B sînt evenimente incompatibile (→*Operații cu evenimente*), atunci numărul de apariții ale evenimentului $A \cup B$ într-un număr arbitrar de repetări ale experimentului fiind egal cu suma numărului de apariții ale lui A și a

numărului de apariții ale lui B, p. lui $A \cup B$ trebuie să fie egală cu suma p. lui A și a lui B. Prin urmare, p. trebuie să posede o proprietate de aditivitate pentru evenimente incompatibile.

2. Considerențele nematematice anterioare au influențat definirea conceptului matematic de probabilitate.

A. În cazul cel mai simplu în care *spațiul de evenimente elementare* * Ω este finit sau numărabil infinit, cînd ca evenimente asociate lui Ω se consideră toate submulțimile sale (clasa $\mathscr{P}(\Omega)$ a acestor submulțimi este în mod banal o *σ-algebră* *), prin p. pe $\mathscr{P}(\Omega)$ se înțelege o funcție P care asociază fiecărui *eveniment* * $A \in \mathscr{P}(\Omega)$ un număr nenegativ $\mathsf{P}(A)$, în particular $\mathsf{P}(\Omega) = 1$, și care este *complet aditivă*, adică

$$\mathsf{P}\left(\bigcup_{i \geqslant 1} A_i\right) = \sum_{i \geqslant 1} \mathsf{P}(A_i)$$

oricare ar fi evenimentele $A_1, A_2, \ldots$ disjuncte în ansamblul lor. (→ *Operații cu evenimente.*) Din proprietatea de aditivitate completă rezultă mai întîi (luînd $A_i = \emptyset$, $i \geqslant 1$) că $\mathsf{P}(\emptyset) = 0$ și apoi (luînd $A_i = \emptyset$, $i > n$) că

$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathsf{P}(A_i)$$

oricare ar fi evenimentele $A_1, \ldots, A_n$ disjuncte în ansamblul lor. Ultima relație arată că p. P este și *finit aditivă*. Nu este suficient să se postuleze ultima proprietate pentru a deduce aditivitatea completă. Postularea aditivității complete (intrucîtva nenaturală) asigură elaborarea unei teorii matematice desăvîrșite. Datorită faptului că $A \cup A^c = \Omega$, $A \cap A^c = \emptyset$, aditivitatea finită implică

$$\mathsf{P}(A^c) = 1 - \mathsf{P}(A)$$

și cum $\mathsf{P}(A^c) \geqslant 0$, rezultă că $\mathsf{P}(A) \leqslant 1$ pentru orice $A \in \mathscr{P}(\Omega)$. Mai general, dacă $A \subset B$, atunci $\mathsf{P}(A) \leqslant \mathsf{P}(B)$. Aceasta rezultă din faptul că $B = A \cup (B \cap A^c)$, iar evenimentele A și $B \cap A^c$ sînt disjuncte.

O consecință frecvent folosită a aditivității complete este faptul că pentru orice șir descendent de evenimente $A_1 \supset A_2 \supset \ldots$

$$\mathsf{P}\left(\bigcap_{i \geqslant 1} A_i\right) = \lim_{n \to \infty} \mathsf{P}(A_n).$$

(Se demonstrează că această proprietate este, de fapt, echivalentă cu aditivitatea completă a probabilității P.)

O p. P pe $\mathscr{P}(\Omega)$ poate fi construită după cum urmează. Fie $p(i)$ numere nenegative de sumă 1 puse în corespondență biunivocă cu evenimentele elementare ω_i din Ω.

Definind $\mathsf{P}(\omega_i) = p(i)$ și pentru un eveniment $A \in \mathscr{P}(\Omega)$ punînd

$$\mathsf{P}(A) = \sum_{\{i\,:\,\omega_i \in A\}} p(i),$$

se vede imediat că P este, într-adevăr, o p. în sensul definiției de mai înainte. (Reciproc, orice p. P pe $\mathscr{P}(\Omega)$ poate fi construită ca mai sus luînd $p(i) = \mathsf{P}(\omega_i)$.)

În cazul particular cînd $\Omega = \{\omega_1, \ldots, \omega_m)\}$ și $p(i) = 1/m$, $1 \leqslant i \leqslant m$, procedeul indicat conduce la relația

$$\mathsf{P}(A) = \frac{\text{numărul evenimentelor elementare din } A}{\text{numărul total de evenimente elementare}} =$$

$$= \frac{\text{numărul „cazurilor favorabile“}}{\text{numărul „cazurilor posibile“}}$$

Aceasta este așa-numita *definiție clasică* a p. care, după cum se vede, este aplicabilă numai în condiții deosebit de restrictive.
Cazul particular considerat poate fi materializat printr-o urnă conținînd m bile care diferă între ele numai prin culoare. Este natural, atunci, să se presupună că **p.** extragerii oricăreia dintre ele este $1/m$. Dacă în urnă există n_i bile de culoare i, $1 \leqslant i \leqslant r$, $\sum_{i=1}^{r} n_i = m$, atunci probabilitatea extragerii unei bile de culoare i va fi egală cu n_i/m. Atît experimentul aruncării monedei cît și cel al aruncării zarului pot fi considerate drept cazuri particulare ale experimentului extragerii unei bile dintr-o urnă de tipul celei descrise. În primul caz, $m = 2$, $n_1 = n_2 = 1$, culorile 1 și 2 corespunzînd stemei, respectiv, valorii. În cel de al doilea caz, $m = 6$, $n_1 = n_2 = n_3 = n_4 = n_5 = n_6 = 1$, culorile 1, 2, 3, 4, 5, 6, corespunzînd celor 6 fețe ale zarului.
În cazul în care *spațiul de evenimente elementare* *Ω este o mulțime infinită nenumărabilă, cînd ca evenimente asociate lui Ω se consideră *numai* elementele unei *σ-algebre* * $\mathscr{K}$ de submulțimi ale lui Ω, prin **p.** pe $\mathscr{K}$ se înțelege o funcție P care asociază fiecărui *eveniment** $A \in \mathscr{K}$ un număr nenegativ $\mathsf{P}(A)$, în particular $\mathsf{P}(\Omega) = 1$, și care este complet aditivă, adică

$$\mathsf{P}\Big(\bigcup_{i \geqslant 1} A_i\Big) = \sum_{i \geqslant 1} \mathsf{P}(A_i),$$

oricare ar fi *evenimentele* $A_1, A_2, \ldots, \in \mathscr{K}$ disjuncte în ansamblu lor.
Este de notat că $\Omega^c = \emptyset \in \mathscr{K}$ și că dacă $A_1, A_2, \ldots, \in \mathscr{K}$, atunci $A_1^c, A_2^c, \ldots \in \mathscr{K}$ deci $\bigcup_{i \geqslant 1} A_i^c \in \mathscr{K}$. Una din regulile lui de Morgan (→ *Operații cu evenimente*) arată atunci că $\bigcap_{i \geqslant 1} A_i = [\bigcup_{i \geqslant 1} A_i^c]^c \in \mathscr{K}$. Aceste observații erau, evident, banale în cazul $\mathscr{K} = \mathscr{P}(\Omega)$.
Toate proprietățile probabilității evidențiate în cazul $\mathscr{K} = \mathscr{P}(\Omega)$ se extind și la cazul unei *σ-algebre* * $\mathscr{K} \subset \mathscr{P}(\Omega)$. Alegerea σ-algebrei $\mathscr{K}$ este o problemă care se rezolvă de la caz la caz în funcție de specificul experimentului și de posibilitatea de a defini o p. pe $\mathscr{K}$. Astfel, în cazul $\Omega =$ mulțimea numerelor reale, drept $\mathscr{K}$ se ia cea mai mică σ-algebră de submulțimi ale lui Ω care conține toate intervalele. Elementele lui $\mathscr{K}$ se numesc, în acest caz, *mulțimi boreliene unidimensionale* (după numele lui *Emile Borel**) și se demonstrează că pe $\mathscr{K}$ se pot defini diverse probabilități.

Repartiție (de probabilitate). Termen în esență echivalent cu cel de probabilitate, folosit pentru a desemna o funcție de mulțime avînd toate proprietățile probabilității, funcție al cărui domeniu de definiție este o *σ-algebră** complet arbitrară (și nu una de evenimente).

Lege (de repartiție). Termen sinonim cu *repartiție (de probabilitate)**.
CÎMP DE PROBABILITATE. Triplet $(\Omega, \mathscr{K}, \mathsf{P})$ în care Ω este un *spațiu de evenimente elementare* *, $\mathscr{K}$ o *σ-algebră* * de submulțimi ale lui Ω, iar P o *probabilitate** pe $\mathscr{K}$. **C.p.** este temelia pe care se construiește teoria probabilităților.

EVENIMENTE INDEPENDENTE. Conceptul de independență a evenimentelor depinde esențial de *probabilitatea** P considerată pe *σ-algebra** $\mathscr{K}$ din *spațiul de evenimente elementare* Ω. Două evenimente $A, B \in \mathscr{K}$ se numesc *independente* dacă $\mathsf{P}(A \cap B) = \mathsf{P}(A)\,\mathsf{P}(B)$. Trei evenimente $A, B, C \in \mathscr{K}$ se numesc *independente* dacă $\mathsf{P}(A \cap B) = \mathsf{P}(A)\,\mathsf{P}(B)$, $\mathsf{P}(B \cap C) = \mathsf{P}(B)\,\mathsf{P}(C)$, $\mathsf{P}(C \cap A) = \mathsf{P}(C)\,\mathsf{P}(A)$, $\mathsf{P}(A \cap B \cap C) = \mathsf{P}(A)\,\mathsf{P}(B)\mathsf{P}(C)$. În general, dacă $(A_i)_{i \in I}$ este o familie finită sau infinită de evenimente, atunci evenimentele A_i, $i \in I$ se numesc *independente* dacă pentru orice mulțime finită de indici distincți $i_1, \ldots, i_k \in I$

$$\mathsf{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathsf{P}(A_{i_1}) \ldots \mathsf{P}(A_{i_k}).$$

Conceptul de **e.i.** se poate ilustra cel mai simplu pe cazul experimentului Bernoulli repetat. Prin *experiment Bernoulli* se înțelege un experiment aleator cu două rezultate posibile α (succes) și $\overline{\alpha}$ (eșec) avînd probabilitățile p, respectiv q, $p + q = 1$. Dacă un astfel de experiment este repetat de n ori, atunci *spațiul de evenimente elementare** asociat, Ω_n, va consta din cele 2^n șiruri ordonate de lungime n de componente α și $\overline{\alpha}$. Întrebarea „cum trebuie definite probabilitățile evenimentelor elementare pentru ca orice șir de rezultate ale experimentului Bernoulli repetat să fie construit din **e.i.**“? are un răspuns imediat: dacă $\omega \in \Omega_n$ conține de k ori componenta α și de $n-k$ ori componenta $\overline{\alpha}$ atunci pentru a avea independență trebuie ca $\mathsf{P}(\omega) = p^k q^{n-k}$. Întrucît există C_n^k evenimente elementare cu această componență, iar $\sum_{k=0}^{n} C_n^k p^n q^{n-k} = (p+q)^n = 1$, definiția probabilității P este perfect justificată. Notînd $B_i = \{$rezultatul celei de a i-a repetări a experimentului este $\alpha\}$ se vede imediat că $\mathsf{P}(B_i) = p$, de unde $\mathsf{P}(B_i^c) = 1 - p = q$, $1 \leqslant i \leqslant n$. De aici rezultă, într-adevăr, că orice sistem de evenimente $A_1, \ldots, A_n$, unde fiecare A_i este egal cu B_i sau cu B_i^c, este constituit din **e.i.**

EVENIMENTE DEPENDENTE. Dependența este proprietatea opusă independenței: două evenimente $A, B \in \mathscr{K}$ se numesc dependente dacă ele nu sînt independente. Pentru două **e.d.** A și B egalitatea $\mathsf{P}(A \cap B) = \mathsf{P}(A)\,\mathsf{P}(B)$ nu se mai verifică. Ea se înlocuiește cu egalitatea $\mathsf{P}(A \cap B) = \mathsf{P}(A)\,\mathsf{P}(B|A)$ unde $\mathsf{P}(B \mid A)$ este *probabilitatea condiționată* * de evenimentul A a evenimentului B.

PROBABILITATE CONDIȚIONATĂ. Probabilitatea condiționată de evenimentul A a evenimentului B, notată $\mathsf{P}(B|A)$, este raportul:

$$\frac{\mathsf{P}(A \cap B)}{\mathsf{P}(A)}.$$

Evident, dacă $\mathsf{P}(A) = 0$, raportul care definește $\mathsf{P}(B|A)$ capătă forma $0/0$ care nu are nici un sens. Ori de cîte ori se consideră o probabilitate condiționată $\mathsf{P}(B \mid A)$ se impune deci condiția ca $\mathsf{P}(A) > 0$, chiar dacă ea nu este menționată explicit.

Următoarele formule în care apar p.c. sînt fundamentale:
Formula condiționării repetate. Pentru evenimente arbitrare $A_1, \ldots, A_n \in \mathcal{K}$

$$\mathsf{P}(A_1 \cap A_2 \cap \ldots \cap A_n) =$$

$$= \mathsf{P}(A_1)\,\mathsf{P}(A_2|A_1)\mathsf{P}(A_3|A_1 \cap A_2) \ldots \mathsf{P}(A_n|A_1 \cap \ldots \cap A_{n-1})$$

ori de cîte ori $P(A_1 \cap \ldots \cap A_{n-1}) > 0$.
Formula probabilității totale. Oricare ar fi sistemul complet de evenimente (→ *Operații cu evenimente*) finit sau numărabil infinit $A_1, A_2, \ldots \in \mathcal{K}$ și evenimentul $B \in \mathcal{K}$

$$\mathsf{P}(B) = \sum_i \mathsf{P}(B \cap A_i),$$

formulă care se mai poate scrie

$$\mathsf{P}(B) = \Sigma' \mathsf{P}(B|A_i)\,\mathsf{P}(A_i),$$

unde ultima sumă se extinde la toate valorile i pentru care $\mathsf{P}(A_i) > 0$.
Formula lui Bayes (probabilitățile cauzelor). Oricare ar fi sistemul complet de evenimente finit sau numărabil infinit $A_1, A_2, \ldots \in \mathcal{K}$ și evenimentul $B \in \mathcal{K}$

$$\mathsf{P}(A_i|B) = \frac{\mathsf{P}(B|A_i)\,\mathsf{P}(A_i)}{\Sigma' \mathsf{P}(B|A_i)\,\mathsf{P}(A_i)}, \quad i = 1, 2, \ldots,$$

unde suma de la numitor se extinde la toate valorile i pentru care $\mathsf{P}(A_i) > 0$
VARIABILĂ ALEATOARE. Intuitiv, prin **v.a.** pe un *cîmp de probabilitate** $(\Omega, \mathcal{K}, \mathsf{P})$ se înțelege o cantitate care este măsurată în legătură cu un experiment aleator, pentru care Ω este *spațiul de evenimente elementare**. Dacă pentru un rezultat al experimentului $\omega \in \Omega$ o operație de măsurare furnizează un număr real $X(\omega)$, funcția reală X, care asociază fiecărui $\omega \in \Omega$ valoarea $X(\omega)$, se numește **v.a.** *reală* atunci cînd mulțimea $\{\omega : X(\omega) < x\}$ a evenimentelor elementare ω pentru care $X(\omega) < x$ este un eveniment ($\in \mathcal{K}$) oricare ar fi numărul real x. Această condiție restrictivă este introdusă pentru a putea vorbi de probabilitățile valorilor lui X (în cazul în care X are o mulțime finită sau numărabil infinită de valori) sau ale intervalelor de valori ale lui X (în cazul în care X are o mulțime infinită nenumărabilă de valori). (Pe de altă parte, este interesant de notat că definiția conceptului de variabilă aleatoare nu face să intervină *probabilitatea** P ci numai *spațiul de evenimente elementare** Ω și *σ-algebra** $\mathcal{K}$.)
În cazul în care Ω este finit sau numărabil infinit, caz în care orice submulțime a lui Ω este un *eveniment**, conceptul de **v.a.** reală coincide cu cel de funcție reală definită pe Ω.
Se demonstrează că dacă X este o **v.a.** reală, atunci și mulțimile $\{\omega : X(\omega) \in B\}$ sînt evenimente oricare ar fi mulțimea boreliană unidimensională B. (→ *Probabilitate*.) Acest fapt stă la baza definirii conceptului de **v.a.** *cu valori într-o mulțime arbitrară* M în care s-a introdus o *σ-algebră** de submulțimi $\mathcal{M}$: o **v.a.** cu valori în M pe un *cîmp de probabilitate** $(\Omega, \mathcal{K}, \mathsf{P})$ este o funcție X definită pe Ω cu valori în M astfel încît $\{\omega : X(\omega) \in B\} \in \mathcal{K}$ pentru orice $B \in \mathcal{M}$.

De obicei, pentru simplificarea scrierii, se suprimă în notație evenimentul elementar ω, adică se scrie $\{X < x\}$ și $\{X \in B\}$ în loc de $\{\omega : X(\omega) < x\}$, respectiv $\{\omega : X(\omega) \in B\}$.

Variabilă aleatoare discretă. V.a. ale cărei valori (reale sau nu) formează o mulțime M finită sau numărabil infinită. Dacă valorile **v.a.** discretă X sînt $x_1, x_2, \ldots$ atunci $\{X = x_i\}$ sînt evenimente și cunoașterea probabilităților $\mathsf{P}(X = x_i) = p_X(x_i)$, $i = 1, 2, \ldots$ este suficientă pentru determinarea probabilităților $\mathsf{P}(X \in B)$, $B \subset M$. Evident,

$$\mathsf{P}(X \in B) = \sum_{\{i;\, x_i \in B\}} p_X(x_i)$$

și, în particular,

$$1 = \mathsf{P}(X \in M) = \sum_i p_X(x_i) = 1.$$

Funcția p_X se numește *funcția de frecvență* sau *densitatea de probabilitate* a *v.a.d.* X.

Variabilă aleatoare continuă. V.a. (reală) X astfel încît $\mathsf{P}(X = x) = 0$ pentru orice x real. Dacă există o funcție măsurabilă nenegativă p_X definită pe dreapta reală astfel încît

$$\mathsf{P}(a < X < b) = \int_a^b p_X(x)\, dx,$$

oricare ar fi numerele reale a și b, se spune că p_X este *densitatea de probabilitate* a *v.a.c.* X. Evident, $\int_{-\infty}^{+\infty} p_X(x)\, dx = 1$. Se demonstrează că pentru orice mulțime boreliană unidimensională B (→ *Probabilitate*) are loc egalitatea $\mathsf{P}(X \in B) = \int_B p_X(x)\, dx$ (integrala fiind concepută ca o integrală Lebesgue).

Vector aleator (n-dimensional). V.a. $X = (X_1, \ldots, X_n)$ cu valori într-o submulțime M a spațiului euclidian n-dimensional R^n. $X_1, \ldots, X_n$ se numesc *componentele* vectorului X. Conceptul de *v.a. (n-dimensional) discret* este implicit în cel de *variabilă aleatoare discretă* *. Conceptul de *v.a. n-dimensional continuu* se definește similar celui de **v.a.** (reală) *continuă** : X este numit *v.a. n-dimensional-continuu* dacă $\mathsf{P}(X = (x_1, \ldots, x_n)) = 0$ pentru orice $(x_1, \ldots, x_n) \in R^n$; dacă există o funcție măsurabilă nenegativă p_X definită pe R^n astfel încît

$$\mathsf{P}((a_1 < X_1 < b_1) \cap \ldots \cap (a_n < X_n < b_n)) =$$

$$= \int_{a_1}^{b_1} \ldots \int_{x_n}^{b_n} p_X(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

pentru orice numere reale $a_1 < b_1, \ldots, a_n < b_n$, se spune că p_X este *densitatea de probabilitate* a *v.a. n-dimensional continuu* X (evident,

$$\int_R \ldots \int_R p_X(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = 1).$$

REPARTIȚIE (A UNEI VARIABILE ALEATOARE). *Repartiție (de probabilitate)** Q_X definită prin relația

$$Q_X(M) = P(X \in M), \quad M \in \mathcal{M},$$

unde X este o variabilă aleatoare cu valori în mulțimea M înzestrată cu σ-algebra de submulțimi $\mathcal{M}$. În cazul unei *variabile aleatoare discrete** repartiția (de probabilitate) este, în esență, densitatea sa de probabilitate. După cum X este discretă sau continuă repartiția se numește, de asemenea, *discretă* sau *continuă*. Dacă X este un *vector aleator n-dimensional** repartiția lui X se numește *n-dimensională*. Repartițiile componentelor *vectorului aleator* X se numesc repartiții *marginale*.

Prin *densitatea* repartiției Q_X se înțelege densitatea de probabilitate a lui X. (→ *Variabilă aleatoare discretă*, *Variabilă aleatoare continuă*, *Vector aleator*.)

FUNCȚIE DE REPARTIȚIE. 1. Pentru o *variabilă aleatoare** reală X funcția de repartiție F_X se definește pe dreapta reală prin egalitatea

$$F_X(x) = P(X < x) = Q_X((-\infty, x)).$$

Se vede imediat că F_X este nedescrescătoare (adică $F_X(x) \leq F_X(y)$, dacă $x < y$), $\lim_{x \to -\infty} F_X(x) = 0$, $\lim_{x \to \infty} F_X(x) = 1$, F_X este continuă la stînga, adică

$$\lim_{\substack{y \to x \\ y < x}} F_X(y) = F_X(x).$$

Este de notat, de asemenea, că

$$P(a \leq X < b) = F_X(b) - F_X(a)$$

oricare ar fi numerele reale $a < b$.

Dacă variabila X este discretă cu valorile reale $x_1, x_2, \ldots$, atunci F_X are un salt de mărime $p_X(x_i) = P(X = x_i)$ în punctul $x = x_i$, $i = 1, 2, \ldots$, ea rămînînd constantă între salturi. Cu alte cuvinte, F_X este o funcție etajată. Reciproc, cunoașterea funcției de repartiție etajate F_X permite să se construiască densitatea de probabilitate p_X, anume: $p_X(x_i) = \lim_{0 < h \to 0} F_X(x_i + h) - F_X(x_i)$, $i = 1, 2, \ldots$.

Dacă variabila X este continuă, cu densitatea de probabilitate p_X, atunci

$$F_X(x) = \int_{-\infty}^{x} p_X(u)\, du.$$

De aici rezultă că F_X este, în acest caz, o funcție absolut continuă. Reciproc, cunoașterea funcției de repartiție absolut continue F_X permite să se construiască densitatea de probabilitate p_X, care nu este altceva decît derivata lui F_X.

2. Pentru un *vector aleator n-dimensional** continuu $X=(X_1, \ldots, X_n)$ funcția de repartiție F_X se definește pe spațiul euclidian n-dimensional R^n prin egalitatea

$$F_X(x_1, \ldots, x_n) = P(X_1 < x_1, \ldots, X_n < x_n) =$$

$$= \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} p_X(x_1, \ldots, x_n)\, dx_1 \ldots dx_n,$$

unde p_X este densitatea de probabilitate a lui X.
De aici rezultă că F_X este absolut continuă și

$$p_X(x_1, \ldots, x_n) = \frac{\partial^n F(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n}.$$

OPERAȚII CU VARIABILE ALEATOARE. Plecînd de la variabile aleatoare date se pot obține alte variabile aleatoare prin aplicarea operațiilor aritmetice elementare. Se demonstrează că dacă X și Y sînt variabile aleatoare reale pe un același *cîmp de probabilitate* *, atunci $X + Y$, XY, X/Y ($Y \neq 0$) și deci, în particular, $X + a$ și aX, unde a este un număr real arbitrar, sînt, de asemenea, variabile aleatoare. Evident, $X + Y$ este variabila care în punctul ω ia valoarea $X(\omega) + Y(\omega)$ etc.

Mai general, dacă f este o funcție măsurabilă (de exemplu, continuă) de n variabile reale, iar $X_1, \ldots, X_n$ sînt variabile aleatoare reale, atunci $f(X_1, \ldots X_n)$, care este funcția care în punctul ω ia valoarea $f(X_1(\omega), \ldots, X_n(\omega))$, este, de asemenea, o variabilă aleatoare. Într-o exprimare lapidară, funcțiile de variabile aleatoare sînt tot variabile aleatoare. Proprietatea poate fi extinsă și la variabile aleatoare cu o mulțime arbitrară de valori.

VALORI MEDII. 1. X fiind o *variabilă aleatoare discretă* * cu funcția de frecvență $p_X(x_i) = p_i$, $i = 1, 2, \ldots$, iar f o funcție reală sau complexă arbitrară definită pe mulțimea valorilor lui X, se definește **v.m.** a lui $f(X)$, notată $\mathbf{E}(f(X))$, ca suma seriei $\sum_i f(x_i)p_i$ (presupunînd seria absolut convergentă, adică $\sum_i |f(x_i)|\, p_i < \infty$; această condiție este întotdeauna satisfăcută în cazul în care mulțimea valorilor lui X este finită.

2. X fiind o *variabilă aleatoare continuă* * cu densitatea p_X, iar f o funcție reală sau complexă, măsurabilă, de variabilă reală, se definește **v.m.** a lui $f(X)$, notată $\mathbf{E}(f(X))$, ca valoarea integralei $\int_{-\infty}^{+\infty} f(x)p_X(x)\, dx$ (presupunînd integrala absolut convergentă, adică $\int_{-\infty}^{+\infty} |f(x)|\, p_X(x)\, dx < \infty$: această condiție este întotdeauna satisfăcută în cazul în care f este o funcție mărginită).

Moment de ordinul m. V.m. corespunzătoare alegerii $f(x)=x^m$, m întreg pozitiv (pentru o variabilă aleatoare reală). Prin urmare, *m.o.* m se calculează prin formulele

$$\mathsf{E}\,X^m = \sum_i x_i^m p_i \quad \text{sau} \quad \mathsf{E}\,X^m = \int_{-\infty}^{+\infty} x^m p_X(x)\,dx.$$

Moment absolut de ordinul m. V.m. corespunzătoare alegerii $f(x) = |x|^m$, m real (pentru o variabilă aleatoare reală). Prin urmare, *m.a.o.* m se calculează prin formulele

$$\mathsf{E}|X|^m = \sum_i |x_i|^m p_i \quad \text{sau} \quad \mathsf{E}|X|^m = \int_{-\infty}^{+\infty} |x|^m p_X(x)\,dx.$$

Medie. 1. Termen similar celui de *moment de ordinul 1** (notat $\mathsf{E}\,X$) al variabilei aleatoare reale X. Prin urmare

$$\mathsf{E}X = \sum_i x_i p_i \quad \text{sau} \quad \mathsf{E}\,X = \int_{-\infty}^{+\infty} x\,p_X(x)\,dx.$$

Proprietățile mai importante ale mediei sînt următoarele:

1) Dacă $X = c = \text{const.}$, atunci $\mathsf{E}\,X = c$;

2) Dacă $X \leqslant Y$, atunci $\mathsf{E}\,X \leqslant \mathsf{E}\,Y$;

3) $\mathsf{E}\,(aX) = a\mathsf{E}\,X$ pentru orice număr real a;

4) $\mathsf{E}\,(X + Y) = \mathsf{E}\,X + \mathsf{E}\,Y$.

2. Vectorul valorilor medii $(\mathsf{E}X_1, \ldots, \mathsf{E}X_n)$ ale componentelor $X_1, \ldots, X_n$ ale vectorului aleator n-dimensional $X = (X_1, \ldots, X_n)$.

Moment centrat de ordinul m. V.m. corespunzătoare alegerii $f(x) = (x - \mathsf{E}\,X)^m$, m întreg pozitiv (pentru o variabilă aleatoare reală).

Moment centrat absolut de ordinul m. V.m. corespunzătoare alegerii $f(x) = |x - \mathsf{E}X|^m$, m real (pentru o variabilă aleatoare reală).

Dispersie (varianță). Termen similar celui de *moment centrat de ordinul 2** (notat $\mathsf{D}X$ sau $\mathrm{Var}\,X$). Prin urmare,

$$\mathsf{D}\,X = \sum_i (x_i - \mathsf{E}\,X)^2 p_i \text{ sau } \mathsf{D}\,X = \int_{-\infty}^{+\infty} (x - \mathsf{E}\,X)^2 p_X(x)\,dx$$

și are loc relația

$$\mathsf{D}X = \mathsf{E}X^2 - (\mathsf{E}X)^2.$$

Dispersia este un indicator important al împrăștierii valorilor variabilei aleatoare după cum rezultă din *inegalitatea lui Cebîșev*: $\mathsf{P}(|X - \mathsf{E}X| > a) \leq \mathsf{D}X/a^2$, $a > 0$. Sînt de notat următoarele proprietăți ale dispersiei:

1) $\mathsf{D}(aX) = a^2\mathsf{D}X$ pentru orice număr real a;

2) $\mathsf{D}(X + Y) = \mathsf{D}X + \mathsf{D}Y + 2\mathsf{E}(X - \mathsf{E}X)(Y - \mathsf{E}Y)$.

Abaterea medie pătratică. Rădăcina pătrată a *dispersiei**.

Corelație (covarianță). V.m. a produsului $(X - \mathsf{E}X)(Y - \mathsf{E}Y)$, unde X și Y sînt variabile aleatoare reale, dată de formula $\operatorname{cov}(X, Y) = \mathsf{E}(XY) - \mathsf{E}X\,\mathsf{E}Y$. Este de notat că $\operatorname{cov}(X, X) = \mathsf{D}X$.

Matrice de corelație (covarianță). Matricea cu elementele $\operatorname{cov}(X_i, X_j)$, $1 \leq i, j \leq n$, unde $X_1, \ldots, X_n$ sînt componentele vectorului aleator n-dimensional, $X = (X_1, \ldots, X_n)$.

Coeficient de corelație. Raportul

$$\rho(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sqrt{\mathsf{D}X\,\mathsf{D}Y}},$$

unde X, Y sînt variabile aleatoare reale. *C.c.* are valori cuprinse între -1 și $+1$. Valorile extreme corespund cazului în care între X și Y există o legătură funcțională liniară de forma

$$Y = \mathsf{E}Y \mp \left(\frac{\mathsf{D}Y}{\mathsf{D}X}\right)^{1/2}(X - \mathsf{E}X)$$

Valorile pozitive ale lui $\rho(X, Y)$ pot indica existența unei tendințe ca la valorile mari ale lui Y să corespundă valori mari ale lui X, iar la valorile mici ale lui Y, valori mici ale lui X. Valorile negative ale lui $\rho(X, Y)$ pot indica existența unei tendințe ca la valorile mici ale lui Y să corespundă valori mari ale lui X, iar la valorile mari ale lui Y, valori mici ale lui X. Valorile lui $\rho(X, Y)$ apropiate de zero pot indica faptul că aceste tendințe sînt slabe, în timp ce valorile lui ρ apropiate de ± 1 pot indica faptul că aceste tendințe sînt puternice.

Funcție caracteristică. 1. Pentru o variabilă aleatoare reală X funcția caracteristică $\varphi_X(t)$ se definește ca **v.m.** corespunzătoare alegerii $f(x) = \exp(\mathrm{i}tx) = \mathrm{e}^{\mathrm{i}tx}$ (i = unitatea imaginară $\sqrt{-1}$). Prin urmare

$$\varphi_X(t) = \sum_i \mathrm{e}^{\mathrm{i}tx_i}p_i \text{ sau } \varphi_X(t) = \int_{-\infty}^{+\infty} \mathrm{e}^{\mathrm{i}tx}p_X(x)\,\mathrm{d}x.$$

Întrucît $|\mathrm{e}^{\mathrm{i}tx}| = 1$, *f.c.* a unei variabile aleatoare reale există întotdeauna. Se demonstrează că existența *momentului de ordin m** al lui X implică existența derivatei de ordinul m a *f.c.* $\varphi_X(t)$ în punctul $t = 0$ și

$$\left.\frac{\mathrm{d}^m\varphi_X(t)}{\mathrm{d}t^m}\right|_{t=0} = \mathrm{i}^m\mathsf{E}X^m.$$

(Implicația inversă este adevărată numai pentru valorile pare ale lui m.)

2. Pentru un vector aleator n-dimensional $X = (X_1, \ldots, X_n)$ *funcția caracteristică* $\varphi_X(t_1, \ldots, t_n)$ se definește ca *media** variabilei aleatoare $\exp \mathrm{i}(t_1X_1 + \ldots + t_nX_n)$.

Entropie. V.m., notată H, corespunzătoare alegerii $f(x_i) = \log_2 \frac{1}{p_i}$, $1 \leqslant i \leqslant n$, pentru o *variabilă aleatoare discretă* * cu funcția de frecvență $p(x_i = p_i$, $1 \leqslant i \leqslant n$. Cu alte cuvinte, entropia

$$H = \sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i} = -\sum_{i=1}^{n} p_i \log_2 p_i$$

este media *informației locale* $\left(\log_2 \frac{1}{p_i}\right)$ furnizate de evenimentele $\{X = x_i\}$, $1 \leqslant i \leqslant n$.

E. apare, de asemenea, ca limita cînd $\alpha \to 1$ a *informației de ordin* α ($0 < \alpha \neq 1$) definită prin relația

$$I_\alpha = \frac{1}{1-\alpha} \log_2 \left(\sum_{i=1}^{n} p_i^\alpha\right).$$

Energie informațională. Mărimea

$$E = 2^{-I_2} = \sum_{i=1}^{n} p_i^2$$

(→*Entropie.*) Deși folosit de Corrado Gini încă din 1936 și ulterior și de alți cercetători (G. U. Yule, G. Herdan, A. Rényi, ș.a.) importanța deosebită a acestui indicator a fost pusă în evidență de Octav Onicescu în 1966 (→*Statistică informațională.*)

VALORI MEDII CONDIȚIONATE. 1. X și Y fiind variabile aleatoare discrete cu valorile $x_1, x_2, \ldots$, respectiv $y_1, y_2, \ldots$, iar f o funcție reală sau complexă arbitrară definită pe mulțimea valorilor lui X, se definește valoarea medie a lui $f(X)$ condiționată de $Y = y_j$, $j = 1, 2, \ldots$, notată $\mathsf{E}(f(X) \mid Y = y_j)$, ca *media** variabilei aleatoare discrete cu valorile $f(x_1)$, $f(x_2)$, ... și funcția de frecvență $p_X(x_i \mid y_j) = \mathsf{P}(X = x_i \mid Y = y_j)$, $i = 1, 2, \ldots$ (→*Probabilitate condiționată*). Prin urmare,

$$\mathsf{E}(f(X) \mid Y = y_j) = \sum_{i \geqslant 1} f(x_i)\, \mathsf{P}(X = x_i \mid Y = y_j).$$

2. (X, Y) fiind un *vector aleator** cu densitatea de probabilitate $p_{X,Y}(x, y)$, iar f o funcție reală sau complexă măsurabilă, de variabilă reală, valoarea medie $\mathsf{E}(f(X) \mid Y = y)$ a lui $f(X)$ condiționat de $Y = y$ se definește prin egalitatea

$$\mathsf{E}(f(X) \mid Y = y) = \frac{\displaystyle\int_{-\infty}^{+\infty} f(x)\, p_{X,Y}(x, y)\, \mathrm{d}x}{\displaystyle\int_{-\infty}^{+\infty} p_{X,Y}(x, y)\, \mathrm{d}x}.$$

VARIABILE ALEATOARE INDEPENDENTE. Extindere a conceptului de *evenimente independente* *. Variabilele aleatoare $X_1, \ldots, X_n$ cu valori în mulțimea arbitrară M cu σ-*algebra** de părți $\mathcal{M}$ se numesc *independente* dacă

$$\mathbf{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \mathbf{P}(X_1 \in B_1) \ldots \mathbf{P}(X_n \in B_n)$$

oricare ar fi mulțimile $B_1, \ldots, B_n \in \mathcal{M}$. (Pentru simplificarea scrierii, virgula înlocuiește semnul $\cap$. Astfel $(X_1 \in B_1, \ldots, X_n \in B_n)$ înseamnă $(X_1 \in B_1) \cap \ldots \ldots \cap (X_n \in B_n)$). În particular, dacă $X_1, \ldots, X_n$ sînt *variabile aleatoare discrete**, independența revine la satisfacerea relației

$$\mathbf{P}(X_1 = x_1, \ldots, X_n = x_n) = \mathbf{P}(X_1 = x_1) \ldots \mathbf{P}(X_n = x_n)$$

pentru valori arbitrare $x_1, \ldots, x_n$ ale variabilelor. Cu alte cuvinte, variabilele aleatoare $X_1, \ldots, X_n$ sînt independente dacă orice sistem de evenimente definite de aceste variabile în mod izolat una față de celelalte este constituit din *evenimente independente*.*

Se demonstrează că dacă X și Y sînt **v.a.i.** reale, atunci $\mathbf{E}(XY) = \mathbf{E}X\mathbf{E}Y$, de unde $\operatorname{cov}(X, Y) = 0$, egalitate care antrenează relația $\mathbf{D}(X + Y) = \mathbf{D}X + \mathbf{D}Y$.

O proprietate deosebit de importantă este aceea că *funcția caracteristică* * a unei sume de **v.a.i.** reale este egală cu produsul funcțiilor caracteristice ale termenilor sumei.

Teoria clasică a probabilităților culminează cu un studiu extensiv al șirurilor (infinite) de **v.a.i.** (reale). Un astfel de șir se caracterizează prin aceea că orice submulțime finită de elemente ale sale este constituită din **v.a.i.**

Următoarele rezultate (referitoare la un șir de **v.a.i.** reale având aceeași funcție de repartiție) sînt fundamentale :

Legea numerelor mari. Dacă media $\mathbf{E}X_1 (= \mathbf{E}X_2 = \ldots)$ există, atunci

$$\lim_{n \to \infty} \frac{X_1(\omega) + \ldots + X_n(\omega)}{n} = \mathbf{E}X_1$$

în toate punctele ω cu excepția eventuală a unei mulțimi de puncte constituind un eveniment de probabilitate nulă.

Legea numerelor mari este istoricește prima și cea mai fundamentală teoremă limită a teoriei probabilităților. Ea este asociată cu numele lui Jakob Bernoulli (1713). Forma dată aici se datorește în esență lui F. P. Cantelli (1917).

Teorema limită centrală. Dacă $\mathbf{E}X_1 (= \mathbf{E}X_2 = \ldots) = m$ și $\mathbf{D}X_1 (= \mathbf{D}X_2 = \ldots) = \sigma^2 \neq 0$, atunci

$$\lim_{n \to \infty} \mathbf{P}\left(\frac{X_1 + \ldots + X_n - nm}{\sigma\sqrt{n}} < x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du$$

pentru toate valorile reale x.

Istoria teoremei limite centrale începe cu de Moivre (1738). Laplace (1813) a evidențiat sfera largă de aplicare a teoremei, văzînd *repartiția normală** ca repartiția limită a unui mare număr de termeni independenți individual neglijabili, și, în particular, a dezvoltat această idee ca un argument fundamental în teoria erorilor.

Legea logaritmului iterat. În condițiile *teoremei limite centrale* *

$$\limsup_{n\to\infty} \frac{X_1(\omega) + \ldots + X_n(\omega) - nm}{\sigma\sqrt{2n\log\log n}} = 1,$$

$$\liminf_{n\to\infty} \frac{X_1(\omega) + \ldots + X_n(\omega) - nm}{\sigma\sqrt{2n\log\log n}} = -1$$

în toate punctele ω cu excepția eventuală a unei mulțimi de puncte constituind un eveniment de probabilitate nulă.

Legea logaritmului iterat se datorește în esență lui A. I. Hincin (1924).

PROCES ALEATOR (STOHASTIC) CU PARAMETRU DISCRET. Șir infinit de variabile aleatoare definite pe același *cîmp de probabilitate**.

LANȚ. Termen sinonim cu *proces aleator cu parametru discret**. Cel mai simplu exemplu de **l.** este un șir de variabile aleatoare independente. (→ *Variabile aleatoare independente.*)

Lanț Markov (finit). L. $(X_n)_{n\geq 1}$ cu valori într-o mulțime finită S, numită *spațiul stărilor*, cu proprietatea că probabilitatea condiționată

$$\mathsf{P}(X_{n+1} = i_{n+1} \mid X_n = i_n, \ldots, X_1 = i_1)$$

este egală cu probabilitatea condiționată

$$\mathsf{P}(X_{n+1} = i_{n+1} \mid X_n = i_n)$$

pentru toate *stările* $i_1, \ldots, i_n, i_{n+1} \in S$ și toate numerele naturale n.

Se poate spune că, spre deosebire de cazul unui lanț de variabile aleatoare independente, caracterizat de o lipsă completă de memorie asupra evoluției trecute, într-un *l.M.f.* se păstrează asupra evoluției trecute amintirea *cea mai recentă*.

Conceptul de *l.M.f.*, introdus de matematicianul rus A. A. Markov (1856—1922) în anul 1906, este unul dintre cele mai fecunde concepte ale teoriei probabilităților. Potențialitățile matematice ale acestuia nu pot fi comparate decît cu potențialitățile sale aplicative. O listă (incompletă) a domeniilor în care teoria *l.M.f.* se aplică cu succes include: controlul statistic al calității produselor, siguranța în funcționare a sistemelor tehnice complexe, biologie și medicină, demografie, geografie, geofizică, teoria mobilității sociale, sisteme de educație, economie, poluare, marketing, operații financiare, mobilitatea forței de muncă etc.

Teoria dependenței markoviene este un domeniu în care matematica românească a adus contribuții recunoscute, de valoare durabilă.

Lanț cu legături complete. L. $(X_n)_{n\geqslant 1}$ cu valori într-o mulțime finită S cu proprietatea că probabilitatea condiționată

$$\mathsf{P}(X_{n+1} = i_{n+1} \mid X_n = i_n, \ldots, X_1 = i_1)$$

este o funcție dată $\varphi^{(n)}_{i_n i_{n+1}}$ de probabilitățile condiționate

$$\mathsf{P}(X_n = i \mid X_{n-1} = i_{n-1} \ldots, X_1 = i_1), \quad i \in S.$$

pentru orice $i_1, \ldots, i_n, i_{n+1} \in S$, $n \geqslant 2$. Evident un *lanț Markov* * este un caz particular de *l.l.c.* (funcțiile $\varphi^{(n)}_{ij}$, $i, j \in S$ sînt constante).

Conceptul de *l.l.c.* a fost introdus în anul 1935 de Octav Onicescu și Gheorghe Mihoc. El a căpătat aplicații importante în teoria matematică a învățării și, în forme generalizate, se regăsește în diverse capitole ale matematicii. O lucrare recentă consacrată dependenței cu legături complete și aplicațiilor sale a apărut în 1982. *)

Lanț O-M (Onicescu-Mihoc). Termen sinonim cu *lanț cu legături complete.**

PROCES ALEATOR (STOHASTIC) CU PARAMETRU CONTINUU. Familie de variabile aleatoare $(X_t)_{t\in T}$ definite pe același *cîmp de probabilitate**, unde T este un interval al dreptei reale, întreaga dreaptă reală sau semiaxa reală nenegativă.

Teoria proceselor aleatoare cu parametru continuu este capitolul central al teoriei contemporane a probabilităților. Dezvoltarea sa a avut și are o influență decisivă asupra evoluției statisticii matematice în ultimele trei decenii.

XIII. LEGI DE REPARTIȚIE

Principalele repartiții ale v.a. caracterizate prin funcție de frecvență sau densitatea de probabilitate ; funcția de repartiție și unele caracteristici ale lor (media, dispersia, modul etc.).

CURBĂ DE FRECVENȚĂ și **CURBĂ CUMULATIVĂ** *(Curbă de repartiție* sau *curba frecvențelor cumulate)*. Graficul f.f. (d.p.) $p_X(x)$, respectiv f.r. $F_X(x)$ în funcție de x.

Exemplu. În cazul v.a. normale $Z \sim N(0, 1)$ **c.f.** este dată de fig. 88, iar **c.r.** este dată de fig. 89 (curba în formă de S sau *ogiva lui Galton*)*.

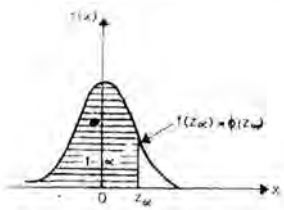


Fig. 88. Curba de frecvență a legii $N(0, 1)$

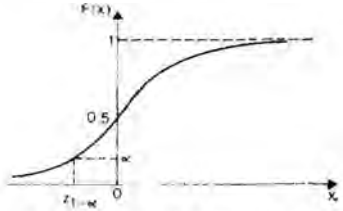


Fig. 89 Curba de repartiție a legii $N(0, 1)$ (Ogiva lui Galton)

*) Ș. Grigorescu, M. Iosifescu, *Dependență cu legături complete și aplicații*, Editura științifică și enciclopedică, București, 1982.

Din proprietățile f.r. se observă că aria totală mărginită de axa OX și de **c.f.** este egală cu 1.

Coada unei repartiții. Acea parte din funcția de repartiție a unei v.a. situată sub (peste) o valoare dată. Termenul se referă atît la valorile v.a., cît și la aria mărginită de axa OX, **c.f.** și dreapta $X=x_\alpha$ sau $X=x_{1-\alpha}$ (fig. 90 și fig. 91).

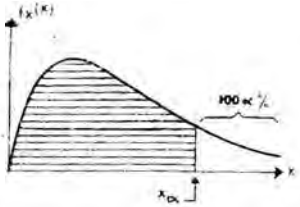


Fig. 90 α %-cuantila și coada unei repartiții

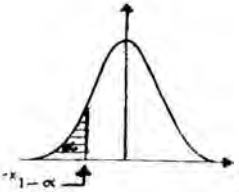


Fig. 91 α %-cuantila inferioară a legii $N(0, 1)$

Grad de libertate. Concept general desemnînd numărul de elemente independente (v.a., statistici etc.) necesar pentru a defini starea unui sistem sau a unui ansamblu (numărul de elemente considerate simultan minus numărul de relații independente care le leagă).

Exemplu. Dacă $x_1, x_2, \ldots, x_n$ este un *eșantion de volum* n* din v.a. X, atunci statistica :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

este o v.a. cu $(n-1)$ *g.l.*, deoarece în ansamblul constituit de eșantionul de mai sus (în care valorile x_i sînt independente) există o relație de legătură

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

(*media eșantionului**).

Se mai spune că valorile eșantionului $x_1, \ldots, x_n$ constituie un ansamblu cu $(n-1)$ *g.l.*

CUANTILĂ (*α%-cuantilă*) (VII).

α%-cuantilă superioară a v.a. X sau a f.r. F_X corespunzătoare. Valoarea x_α determinată astfel încît :

$$\alpha = P(X > x_\alpha), \quad 0 \leqslant \alpha \leqslant 1,$$

adică

$$F_X(x_\alpha) = P(X < x_\alpha) = 1 - \alpha.$$

Coada repartiției * este de arie egală cu α și se află în partea dreaptă față de abscisa x_α (fig. 90).
De exemplu, dacă v.a. $Z \sim N(0, 1)$ și $\alpha = 0{,}05$, atunci 0,05% — *c.s.* a legii $N(0, 1)$ este valoarea $z_{0,05} = 1{,}65$ determinată astfel încît:

$$0{,}05 = P(Z > z_{0,05} = 1{,}65) = 1 - \Phi(z_{0,05}) = 1 - 0{,}94950,$$

unde

$$\Phi(t) = 1/\sqrt{2\pi} \int_{-\infty}^{t} e^{-x^2/2}\, dx, \quad \text{f.r. a legii } N(0, 1).$$

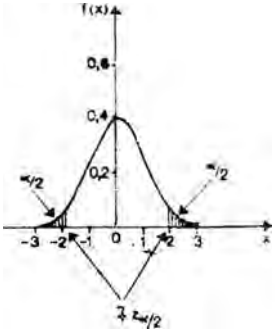


Fig. 92 α%-*cuantila bilaterală a legii* $N(0, 1)$

α%-cuantilă inferioară a v.a. X sau a f.r. F_X corespunzătoare. Valoarea $x_{1-\alpha}$ determinată astfel încît:

$$\alpha = P(X < x_{1-\alpha}),$$

adică

$$F_X(x_{1-\alpha}) = \alpha.$$

Coada repartiției * este de arie egală cu α și se află în partea stîngă față de abscisa $x_{1-\alpha}$ (fig. 91).
α%-cuantilă bilaterală (*α-cuantilă bilaterală*) a v.a. X. Valoarea $x_{\alpha/2}$ determinată astfel încît:

$$P(|X| > x_{\alpha/2}) = \alpha,$$

adică

$$P(X < -x_{\alpha/2}) + P(X > x_{\alpha/2}) = \alpha/2 + \alpha/2 = \alpha.$$

De exemplu, dacă v.a. $Z \sim N(0, 1)$ și $\alpha = 0{,}05$, atunci 0,05—*c.b.* este valoarea $z_{0,025} = 1{,}96$ determinată astfel încît (fig. 92):

$$P(|Z| > z_{\alpha/2} = 1{,}96) = \alpha = 0{,}05,$$

adică

$$P(Z < -z_{\alpha/2}) + P(Z > z_{\alpha/2}) = \Phi(-z_{\alpha/2}) + 1 - \Phi(z_{\alpha/2}) =$$

$$= \Phi(z_{1-\alpha/2}) + 1 - \Phi(z_{\alpha/2}) = 0.025 + 0{,}025 = 0{,}05.$$

Rezultă

$$\Phi(z_{0{,}025}) = 1 - 0{,}025 = 0{,}975,$$

de unde

$$z_{0{,}025} = 1.96$$

cu

$$\Phi(t) = 1/\sqrt{2\pi} \int_{-\infty}^{t} e^{-x^2/2}\,dx, \quad \text{f.r. a legii } N(0, 1)$$

$$1 - \Phi(-t) = \Phi(t)$$

α-*c.b.* este $\alpha/2$-cuantila superioară.

Observații.

1. α-cuantila inferioară și α-cuantila superioară se mai numesc *punctele critice* ale v.a. X.
2. Între α-cuantila superioară și α-cuantila inferioară există relația :

$$x_\alpha = -x_{1-\alpha}.$$

3. În cazul în care graficul d.p. a v.a. X este simetric, *cozile repartiției** (din stînga, respectiv din dreapta) sînt egale și de valoare $\alpha/2$.
4. Legătura dintre punctele critice și probablitățile corespunzătoare este ilustrată pe graficul d.p., respectiv al f.r. a v.a. în fig. 88, respectiv fig. 89 pentru v.a. $Z \sim N(0.1)$, precum și în fig. 92.

LEGI DE REPARTIȚIE DISCRETE. (→*Repartiția variabilei aleatoare discrete.*)

Legea (discretă) uniformă. *Repartiția unei v.a. discrete** X cu valorile $x = a + k,\ a + 2k, \ldots, a + nk$ (a și k constante date) și funcția de frecvență : (fig. 93)

$$P(X = x) = p_X(x) = \frac{1}{n}.$$

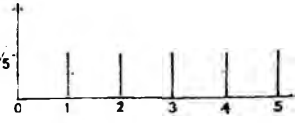


Fig. 93 Curba de frecvență a legii (discrete) uniforme pentru $n = 5$ ($a = 0$, $k = 1$)

Schematic *l. (d)u.* se poate reprezenta sub forma

$$X\begin{pmatrix} a+k & a+2k & \dots & a+nk \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix}.$$

Caracteristici principale

Media : $E(X) = a + k\dfrac{n+1}{2}$.

Dispersia : $\mathrm{Var}(X) = k^2\dfrac{n-1}{12}$.

Mediana : $Me(X) = a + k\dfrac{n+1}{2}$.

Legea binomială Bi(n, p). Repartiția unei v.a. X cu valorile 0, 1, 2, ..., și având funcția de frecvență :

$$P(X = x) = p_X(x) = C_n^x p^x q^{n-x} = \frac{n!}{x!(n-x)!}\, p^x q^{n-x},$$

unde $p + q = 1$, $p > 0$, $q > 0$, n-întreg pozitiv, $C_n^x = \dfrac{n!}{x!(n-x)!}$, adică :

$$X\begin{pmatrix} 0 & 1 & x & n-1 & n \\ C_n^0 p^0 q^{n-0} & C_n^1 p^1 q^{n-1} & \dots C_n^x p^x q^{n-x} \dots & C_n^{n-1} p^{n-1} q^1 & C_n^n p^n q^0 \end{pmatrix}$$

p_X este termenul general al dezvoltării binomului $(p + q)^n$:
X este v.a. reprezentînd numărul de apariții ale evenimentului A în n experimente independente (→ *extragere (selecție) bernoulliană sau sondaj cu revenire*)*, unde p este probabilitatea realizării lui A și q este probabilitatea realizării evenimentului complementar non A (→ *urna lui Bernoulli*).
Fiind dată o urnă — conținînd N bile dintre care Np bile albe și Nq bile negre — din care se face *un sondaj cu revenire** de n bile. p_X este probabilitatea obținerii unui minim de x bile albe și $n-x$ bile negre, indiferent de ordinea apariției lor.
Funcția de repartiție a v.a. $X \sim Bi(n, p)$:

$$F(x) = P(X < x) = \sum_{k=0}^{[x]} C_n^k p^k q^{n-k} = \begin{cases} 0 & \text{dacă } x < 0 \\ C_n^0 q^{n-0} & \text{dacă } 0 < x \leqslant 1 \\ \sum_{k=0}^{1} C_n^k p^k q^{n-k} & \text{dacă } 1 < x \leqslant 2 \\ \dots\dots\dots\dots\dots & \\ \sum_{k=0}^{n-1} C_n^k p^k q^{n-k} & \text{dacă } n-1 < x \leqslant n \\ 1 & \text{dacă } x > n \end{cases}$$

unde $[x]$ este cel mai mare întreg cuprins în numărul $x \in R$.

De exemplu, graficul f.r. a legii Bi (6.1/4) este dat de fig. 94.

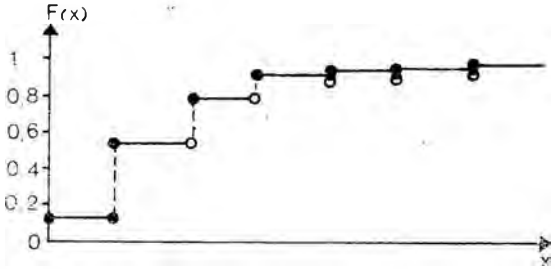


Fig. 94 Curba de repartiție. Graficul funcției de repartiție a legii $Bi(n = 6,\ p = 1{,}4)$

Caracteristici principale
Media : $E(X_{Bi(n,p)}) = np$.
Dispersia : $\mathrm{Var}\,(X_{Bi(n,p)}) = npq$.
Modul M_0 — valoarea cea mai probabilă — este numărul întreg care satisface relațiile : $np - q \leqslant M_0 \leqslant np + p$.
Funcția caracteristică :

$$\varphi_{X_{Bi(n,p)}}(t) = E(e^{itX}) = \sum_{x=0}^{n} e^{itx} C_n^x p^x q^{n-x} = (p\,e^{it} + q)^n.$$

Aplicînd *formula Moivre-Stirling* :

$$n! = 1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n = \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \cdot [1 + \theta(n)],$$

unde $\theta(n) \to 0$ pentru $n \to \infty$, se obține expresia asimptotică a probabilității p_X cînd $npq > 10$ și $n \to \infty$:

$$p_X(x) \approx \frac{1}{\sqrt{2\pi npq}}\, e^{-\frac{l_n^2}{2npq}},$$

adică *legea normală** de medie np și abatere medie pătratică $\sqrt{npq}$, unde $l_n = x - np$: $a\sqrt{n} < l_n < b\sqrt{n}$, a,b,n = constante.
i.b. a fost formulată de J. Bernoulli în „Ars Conjectandi" (1713).
Legea binomial negativă (repartiția lui Pascal sau Polya) $Pa\,(c, p)$. Repartiția unei v.a. X reprezentînd numărul de *extrageri (selecții) bernoulliene sau sondaje cu revenire** necesar obținerii unui număr dat c de bile albe, adică din n experi-

mente independente, numărul de probe independente în care se realizează evenimentul A de c ori.
F.f. a v.a. X este :

$$p_X = P(X = n) = C_{n-1}^{c-1} p^c q^{n-c}, \quad n > c \quad (p + q = 1).$$

adică termenul general din $p^c(1-q)^{-c}$ dacă $(1-q)^{-c}$ este dezvoltat în serie de puteri ale lui q, de unde justificarea denumirii *l.Pa(c, p)* de *l.b.n.*

Caracteristicile principale

Media : $E(X) = \dfrac{qc}{p}$.

Dispersia : $\operatorname{Var} X = \dfrac{qc}{p^2}$.

Modul M_0 : $\dfrac{c-1}{p} < M_0 < \dfrac{c-q}{p}$.

Funcția de repartiție :

$$F(c : x) = P(X \leqslant x) = 1 - \sum_{j=0}^{c-1} C_{[x]}^{j} p^j q^{[x]-j}.$$

Funcția caracteristică : $\varphi_{X_{Pa(c,p)}}(t) = (pe^{it})^c (1 - q^{it})^{-c}$.

Repartiția limită a *l.b.n.* pentru $c \to \infty$ este *repartiția normală* *.
Se aplică în medicină, biologie și anume teoria contagiunii, p fiind probabilitatea constantă ca un microb să fie atins în aplicarea repetată a unui tratament, care supune microbii la riscul de a fi atinși, presupunînd că microbul moare dacă este atins în c tratamente. (fig. 95).

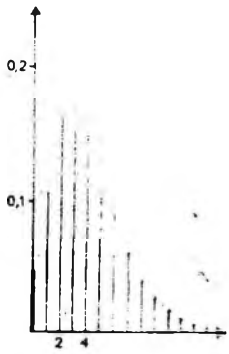


Fig. 95 Curba de frecvență a legii binomial negative pentru $c = 6$, $p = 0{,}6$

Legea geometrică. Repartiția unei v.a. X avînd f.f.

$$p_X(x) = P(X = k) = pq^{k-1}, \quad k = 1, 2, \ldots, n.$$

Schematic, *l.g.* se poate prezenta sub forma :

$$Q_X\begin{pmatrix}1 & 2 \ldots k \ldots & n\\ p & pq \ldots pq^{k-1} \ldots pq^{n-1}\end{pmatrix}.$$

X este o v. a. reprezentînd numărul de probe independente în care se realizează evenimentul A pentru prima dată, din n experimente independente.
L.g. se obține din *legea binomial negativă* * pentru $c = 1$.
Legea hipergeometrică. Repartiția v.a. X avînd funcția de frecvență

$$p_X = P_{N,n,k} = P(X = x) = \frac{C_k^x\, C_{N-k}^{n-x}}{C_N^n},\quad x = 0, 1, 2, \ldots, (n, k)$$

((n, k) înseamnă cel mai mic număr dintre n și k), adică repartiția v.a. X reprezentînd numărul de bile albe care apar într-un eșantion de n bile obținut printr-o operație de *extragere (selecție sau sondaj) fără revenire* * dintr-o urnă cu N bile, dintre care $k = Np$ bile albe și $N-k = Nq$ bile negre. Dacă în cele n extrageri succesive bila albă a ieșit de α ori ($0 \leqslant \alpha \leqslant k$) și bila neagră a ieșit de $n-\alpha$ ori ($0 \leqslant n-\alpha \leqslant k$) și bila neagră a ieșit de $n-\alpha$ ori ($0 < n-\alpha \leqslant N-k$), atunci repartiția v.a. X este :

$$X\begin{pmatrix}0 & 1 & \ldots & \alpha & \ldots & k\\ \dfrac{C_k^0\, C_{N-k}^n}{C_N^n} & \dfrac{C_k^1\, C_{N-k}^{n-1}}{C_N^n} & \ldots & \dfrac{C_k^\alpha C_{N-k}^{n-\alpha}}{C_N^n} & \ldots & \dfrac{C_k^k\, C_{N-k}^{n-k}}{C_N^n}\end{pmatrix}.$$

Caracteristicile principale
Media : $E(X) = np$.

Dispersia : $\mathrm{Var}\,(X) = npq\,\dfrac{N-n}{N-1}$.

Funcția de repartiție :

$$F(x) = P(X \leqslant x) = \sum_{z=\max(0,\, n-(N-k))}^{[x]} \frac{C_k^z\, C_{N-k}^{n-z}}{C_N^n},$$

unde $[x]$ reprezintă cel mai mare întreg conținut în x, astfel încît $[x] \leqslant \min(n, k)$.
Funcția generatoare este :

$$\psi(t) = \sum_{k=0}^{\infty} t^k P_{N,n,k} = \left(1 + \frac{a \cdot b}{1 \cdot c}\, t + \frac{a(a+1)b(b-1)}{1 \cdot 2 \cdot c^2}\, t^2 + \ldots\right) P_{N,n,0},$$

unde $a = -n$, $b = -k$, $c = N-k-n+1$.

Expresia din paranteză este seria hipergeometrică a lui Gauss, de unde și justificarea denumirii acestei legi.

Dacă N este suficient de mare, practic dacă $n/N \lhd 1/10$ atunci *l.h.* poate fi aproximată prin *legea binomială* Bi(n, p)*.

Legea Poisson. Repartiția unei v.a. X cu valorile $x = 0, 1, 2, \ldots$, având funcția de frecvență:

$$p_X = P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!},$$

unde λ este o constantă pozitivă și $e = 2{,}718\ldots$ este baza logaritmului neperian. (fig. 96).

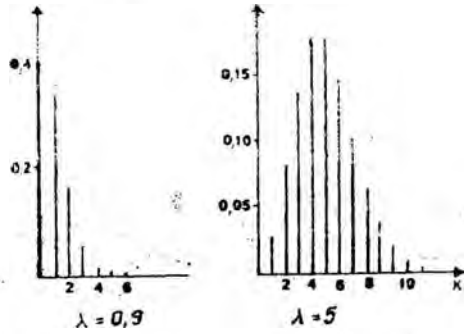


Fig. 96 Curba de frecvență a legii Poisson pentru $\lambda = 0{,}9$, *respectiv* $\lambda = 5$

Schematic, *l.P.* se poate reprezenta sub forma:

$$Q_X\begin{pmatrix} 0 & 1 & 2 & \ldots \\ e^{-\lambda} & e^{-\lambda}.\frac{\lambda}{1!} & e^{-\lambda}.\frac{\lambda^2}{2!} & \ldots \end{pmatrix}$$

și se notează:

$$X \sim Po(\lambda).$$

L.P. se obține ca un caz limită al *legii binomiale* Bi(n, p)* pentru n suficient de mare, p mic ($p < 0{,}1$), astfel încît să se poată considera produsul $\lambda = np$ de valoare constantă.

Cu alte cuvinte, dacă $\{X_n\}$ este un șir de v.a. $\sim Bi(n, p)$, $n = 1, 2; \ldots$, atunci șirul $\{X_n\}$ converge în repartiție către v.a. $X \sim Po(\lambda)$: $X_n \xrightarrow{r} X$, adică șirul funcțiilor de repartiție $\{F_n\}$ corespunzătoare v.a. $\{X_n\}$ converge către funcția de repartiție F a lui X, în fiecare punct de continuitate al acesteia.

Dacă $\lambda > 20$ sau chiar $\lambda > 10$, *l.P.* este aproximată de *legea normală* $N(\lambda, \sqrt{\lambda})$:

$$p_X \sim \frac{1}{\sqrt{2\pi}} \int_{u_1}^{u_2} e^{-u^2/2}\, du,$$

unde

$$u_1 = \frac{x - \lambda - 1/2}{\sqrt{\lambda}}, \qquad u_2 = \frac{x - \lambda + 1/2}{\sqrt{\lambda}}$$

(corecția de continuitate).

Caracteristici principale

Media: $E(X) = \lambda$.

Dispersia: $\operatorname{Var} X = \lambda$.

Funcția de repartiție: $F(x) = P(X \leqslant x) = \sum_{\alpha=0}^{[x]} \frac{\lambda^\alpha e^{-\lambda}}{\lambda!}$,

unde $[x]$ este cel mai mare întreg conținut în x.

Funcția caracteristică: $\varphi(t) = e^{\lambda(e^{it}-1)}$.

L.P. a fost considerată de Poisson în 1837 și a fost „redescoperită" de *L. V. Bortkiewicz* *, în 1898, în studiul mortalității prin accidente în cavalerie.

Bortkiewicz a denumit *l.P.* „legea evenimentelor rare" sau „legea probabilităților mici".

L.P. este folosită în aplicațiile statisticii matematice în biologie, telecomunicații, în controlul statistic al calității atunci cînd probabilitatea obținerii unei piese defecte este foarte mică etc.

Legea logaritmică. Repartiția unei v.a. X cu valorile $1, 2, \ldots$ avînd funcția de frecvență:

$$p_X = \frac{k\alpha^x}{x}, \quad 0 < \alpha < 1, \quad x = 1, 2, \ldots,$$

unde

$$k = \frac{-1}{\ln(1 - \alpha)}.$$

Caracteristici principale

Media: $E(X) = \frac{k\alpha}{1 - \alpha}$.

Dispersia: $\operatorname{Var} X = \frac{k\alpha(1 - k\alpha)}{(1 - \alpha)^2}$.

L.l. a fost considerată de R.A. Fisher în 1941.

LEGI DE REPARTIȚIE CONTINUE. (→ *Repartiția variabilei aleatoare continue.*) **Legea uniformă.** Repartiția unei v.a. X continue care ia valori în intervalul $[\alpha, \beta]$ și care are d.p.

$$f(x) = \frac{1}{\beta - \alpha}.$$

Se notează $X \sim U(\alpha, \beta)$ (fig. 97).

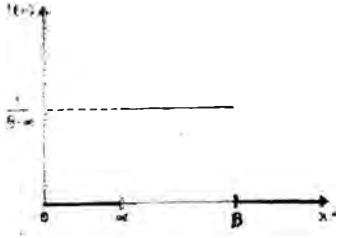


Fig. 97 Curba de frecvență a legii uniforme $U(\alpha, \beta)$

Caracteristici principale

Media : $E(X) = \dfrac{\alpha + \beta}{2}$.

Dispersia : $\operatorname{Var} X = \dfrac{(\beta - \alpha)^2}{12}$.

Funcția de **repartiție** $F(x) = \displaystyle\int_{\alpha}^{x} f(y)\,dy = \frac{x - \alpha}{\beta - \alpha}$.

Legea triunghiulară. Repartiția unei v.a. X a cărei d.p. este o curbă de formă triunghiulară. Prin alegerea convenabilă a originei, d.p. a *l.t.* se poate pune sub forma :

$$f(x) = \begin{cases} \dfrac{2}{a(a+b)}(a + x), & -a < x < 0 \\ \dfrac{2}{b(a+b)}(b - x), & 0 < x < b \\ 0, \text{ în rest.} \end{cases}$$

Prima lege a lui Laplace. Repartiția unei v.a. X continue, cu valori pe R având d.p. :

$$f(x) = \frac{h}{2} e^{-h|x|}, \quad -\infty < x < +\infty.$$

Prezintă interes istoric, fiind introdusă de Laplace în 1774 ca lege de repartiție a erorilor.

Legea normală. (*Legea Laplace-Gauss*). Repartiția v.a. X având d. p. :

$$f(x) = \left(1/\sigma\sqrt{2\pi}\right) e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

L.n. este o lege continuă, complet determinată de m și σ.
Constantele m, σ^2 se numesc parametrii *l.n.*

Media : $E(X) = \dfrac{1}{\sigma\sqrt{2\pi}} \displaystyle\int_{-\infty}^{+\infty} x\, e^{-\frac{(x-m)^2}{2\sigma^2}}\, dx = m.$

Dispersia : $\operatorname{Var} X = E(X - m)^2 = \sigma^2$.
Așadar, parametrii m, σ^2 reprezintă : m — parametrul de grupare este media v.a. $X \sim N(m, \sigma^2)$; σ^2 — parametrul de împrăștiere este dispersia v.a. $X \sim N(m, \sigma^2)$ (fig. 98).

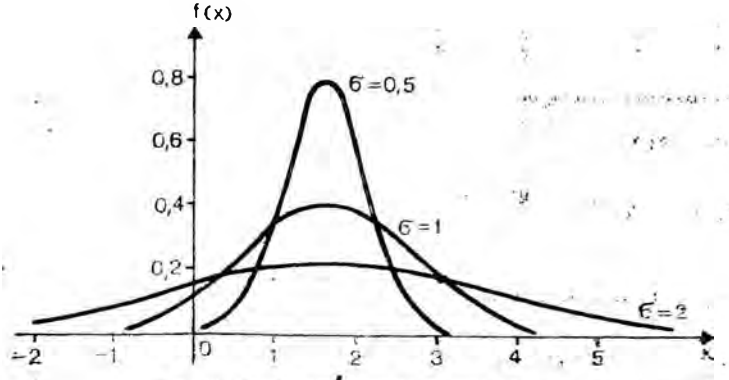


Fig. 98 Curba de frecvență a legii $N(m, \sigma)$

Legea normală normată este repartiția v.a. $U \sim N(0,1)$:

$$U = (X - m)/\sigma,$$

unde v.a. $X \sim N(m, \sigma^2)$.

V.a. $U \sim N(0, 1)$ are d.p. :

$\varphi(u) = 1/\sqrt{2\pi}\, e^{-u^2/2}$.
Media : $E(U) = 0$.
Dispersia : $\operatorname{Var}(U) = 1$.

Funcția de repartiție : $\Phi(u) = 1/\sqrt{2\pi} \displaystyle\int_{-\infty}^{u} e^{-x^2/2}\, dx$

se mai numește și *funcția lui Laplace* și este tabelată.

Observații

1. Schimbarea de variabilă $u = (x - m)/\sigma$ transformă orice *l.n.* $N(m, \sigma^2)$ în *l.n.* $N(0, 1)$.

2. Funcția de repartiție a v.a. $X \sim N(m, \sigma^2)$ este:

$$F(x) = P(X < x) = 1/\sigma\sqrt{2\pi} \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}} dt,$$

care poate fi exprimată prin f.r. $\Phi(u)$ a *l.n.* $N(0, 1)$:

$$F_X(x) = \Phi_U((x - m)/\sigma).$$

3. Se observă că

$$\Phi(-x) = 1 - \Phi(x), \quad x \geq 0.$$

4. Valorile α-**enantilelor** v.a. $U \sim N(0, 1)$ satisfac relația

$$u_{1-\alpha} = -u_\alpha.$$

5. Dacă $X \sim N(m, \sigma^2)$, atunci

$$P(a < X < b) = \Phi((b - m)/\sigma) - \Phi((a - m)/\sigma).$$

6. Dacă $(X_1, \ldots, X_n)$ este un eșantion de volum n dintr-o populație normală $N(m, \sigma^2)$, atunci *media de selecție**

$$\overline{X} = 1/n \sum_{i=1}^{n} X_i.$$

este o v.a. $\overline{X} \sim N(m, \sigma^2/n)$.

7. Graficul d.p. a *l.n.* $N(m, \sigma^2)$ este simetric față de $x = m$, are un maxim de coordonate $(m, 1/\sqrt{2\pi}\sigma)$ și două puncte de inflexiune pentru $x = m \pm \sigma$.

8. *L.n.* joacă un rol important în *teorema limită centrală* * și stă la baza modelării multor fenomene naturale.

Erorile întîmplătoare sau fluctuațiile rezultatelor unor experimente urmează, de regulă, o lege ce poate fi aproximată prin *l.n.* conform *teoremei limită centrală* *.

Legea lognormală. Repartiția v.a. X având d.p.:

$$f(x) = \begin{cases} \dfrac{1}{(x - \theta)\sigma\sqrt{2\pi}} e^{-\frac{[\ln(x-\theta)-m]^2}{2\sigma^2}}, & x > 0 \\ 0, & x \leq 0. \end{cases}$$

Adică, $X \sim l.l.$ de parametri θ, m, σ^2 dacă $Y = \ln(X - \theta) \sim N(m, \sigma^2)$.

Caracteristici principale

Media: $E(X) = \theta + e^{m+\sigma^2/2}$.

Dispersia: $\text{Var } X = e^{2m} e^{\sigma^2}(e^{\sigma^2} - 1)$.

Observații.

1. *L.l.* este o repartiție unimodală și cu asimetrie pozitivă.

2. Cînd σ^2 este mic, graficul d.p. *l.l.* este asemănător cu cel al legii normale.

3. Pentru diferite valori ale parametrilor θ, m, σ^2 în fig. 99 sînt date graficele d.p. a *l.l.*

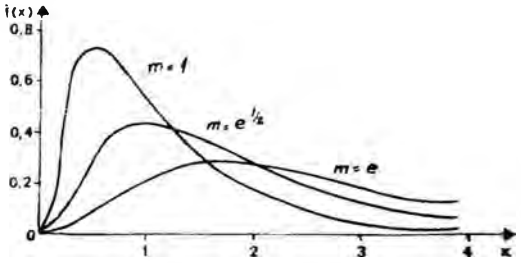


Fig. 99 Curba de frecvență a legii lognormale pentru diferite valori ale lui m *și* $\theta = 0$, $\sigma^2 = 0{,}5$

Legea formelor pătratice de variabile aleatoare normale. Repartiția formei pătratice

$$Q = \sum_{i=1}^{n} u_i^2 - \sum_{j=1}^{s} l_j^2,$$

unde u_i $(i = 1, 2, \ldots, n)$ sînt n v.a. independente $\sim N(0, 1)$, $l_j = \sum_{i=1}^{n} a_{j_i} u_i$ $(j = 1, 2, \ldots, s < n) \sim N(0, 1)$ sînt s funcții liniare de v.a. $\{u_i\}$, independente între ele cu:

$$\sum_{i=1}^{n} a_{j_i}^2 = 1, \qquad \sum_{i=1}^{n} a_{j_i} a_{k_i} = 0 \quad (j, k = 1, 2, \ldots, n).$$

Fie acum $l_k = \sum_{i=1}^{n} a_{k_i} u_i$ $(k = s+1, s+2, \ldots, n)$ alte $n-s$ funcții liniare de v.a. $\{u_i\}$, în care coeficienții a_{k_i} îndeplinesc condițiile de ortogonalitate

$$\sum_{i=1}^{n} a_{k_i}^2 = 1, \qquad \sum_{i=1}^{n} a_{j_i} a_{k_i} = 0, \quad (j, k = 1, \ldots, n).$$

Funcțiile $l_k \sim N(0, 1)$ sînt independente, deoarece $\operatorname{cov}(l_j, l_k) = 0$, $j \neq k$. Forma pătratică Q se scrie:

$$Q = \sum_{j=1}^{n} (u_j^2 - l_j^2) + \sum_{j=s+1}^{n} l_j^2$$

sau, aplicînd transformarea ortogonală $l_j = \sum_{i=1}^{n} a_{j_i} u_i$,

$$Q = \sum_{j=s+1}^{n} l_j^2,$$

unde

$$l_j = \sum_{i=1}^{n} a_{j_i} u_i, \quad j = s+1, \ldots, n$$

sînt funcţii liniare de v.a. $\{u_i\}$ cu coeficienţii a_{k_i} satisfăcînd condiţiile de ortogonalitate :

$$\sum_{i=1}^{n} a_{k_i}^2 = 1, \quad \sum_{i=1}^{n} a_{j_i} a_{k_i} = 0, \quad j, k = 1, 2, \ldots, n \quad j \neq k.$$

Repartiţia formei pătratice Q este legea χ^2 cu $(n-s)$ g.l., $Q \sim \chi^2_{n-s}$.

Observaţii

1. Forma pătratică $Q = \sum_{j=s+1}^{n} l_j^2$ este independentă de variabilele $l_1, l_2, \ldots l_s$, iar rangul formei pătratice Q coincide cu numărul g.l. ale legii sale.

2. Teorema de desfacere a legii χ^2 (elaborată de W. G. Cochran, în 1934). Fie

$$\chi^2 = \sum_{i=1}^{n} u_i^2 = Q_1 + Q_2 + \ldots + Q_s,$$

unde

v.a. independente $\{u_i\} \sim N(0, 1)$, $i = 1, 2, \ldots, n$;

forma pătratică Q_j de rang k_j, respectiv, $j = 1, \ldots, s$.

Condiţia necesară şi suficientă pentru ca formele pătratice $Q_1, Q_2, \ldots, Q_s$ să fie independente şi $Q_j \sim \chi^2_{k_j}$, $j = 1, \ldots, s$, respectiv este ca :

$$k_1 + k_2 + \ldots + k_s = n.$$

Legea Gamma. Repartiţia unei v.a. X cu valori nenegative avînd d.p.

$$f(x) = \begin{cases} 0, & x \leqslant 0 \\ \dfrac{a^p}{\Gamma(p)} x^{p-1} e^{-ax}, & x > 0, \quad (\beta = 1/a) \end{cases}$$

unde

$$\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} \, dx$$

(fig. 100).

Cazuri particulare :

1. Dacă $p = 1$, se obține

$$f(x) = \begin{cases} 0 & , x \leqslant 0 \\ ae^{-ax}, & x > 0. \end{cases}$$

adică d.p. a *legii exponențiale** de parametru a.

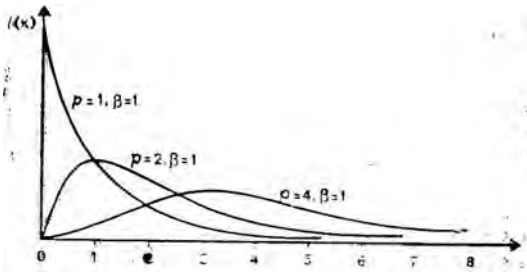


Fig. 100 Curba de frecvență a legii Gamma pentru diferite valori ale lui p și $\beta = 1/a = 1$

2. Dacă $p = k \in N$ și $a = \lambda k$, se obține

$$f(x) = \begin{cases} 0 & , x \leqslant 0 \\ \dfrac{(\lambda k)^k}{\Gamma(k)} x^{k-1} e^{k x}, & x > 0, \end{cases}$$

adică d.p. a *legii Erlang.*

3. Dacă $p = n/2$ și $a = \sigma^2/2$, $(\sigma > 0)$ se obține:

$$f(x) = \begin{cases} 0 & x \leqslant 0 \\ \dfrac{1}{2^{n/2}\, \sigma^n\, \Gamma\left(\dfrac{n}{2}\right)} x^{n/2-1} e^{-x/2\sigma^2}, & x > 0 \end{cases}$$

adică d.p. a *legii* χ^2*

Legea Beta. Repartiția v.a. X având d.p.:

$$f(x) = \begin{cases} \dfrac{\Gamma(\mu + \nu)}{\Gamma(\mu)\ \Gamma(\nu)} x^{\mu-1}(1-x)^{\nu-1}, & \text{dacă } x \in (0, 1) \\ 0 & , \text{dacă } x \notin (0, 1) \end{cases}$$

$\mu, \nu > 0$.

$X \sim Be(\mu, \nu)$, $\mu, \nu > 0$ (fig. 101).

Funcția de **repartiție** a lui *Be*:

$$F(x) = P(X < x) = \frac{\Gamma(\mu + \nu)}{\Gamma(\mu)\,\Gamma(\nu)} \int_0^x y^{\mu-1}(1-y)^{\nu-1}\,dy =$$

$$= \begin{cases} 1, & x \geqslant 1 \\ \dfrac{B_x(\mu, \nu)}{B(\mu, \nu)}, & x \in (0,1) \\ 0, & x \leqslant 0 \end{cases}$$

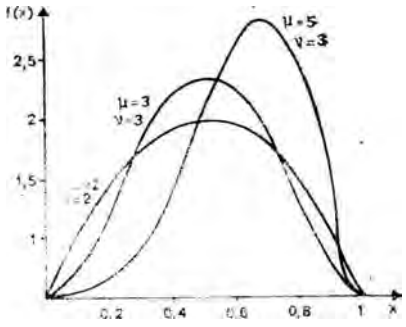


Fig. 101 Curba de **frecvență** *a legii Beta, Be(μ, ν) pentru diferite valori ale lui* **μ** *și* **ν**.

unde

$B_x(\mu, \nu) = \int_0^x y^{\mu-1}(1-y)^{\nu-1}\,dy$ este funcția Beta incompletă,

$B(\mu, \nu) = \int_0^1 y^{\mu-1}(1-y)^{\nu-1}\,dy = \dfrac{\Gamma(\mu)\,\Gamma(\nu)}{\Gamma(\mu+\nu)}$ este funcția Beta.

Caracteristice principale

Momentul de **ordinul** *r*

$$M_r = \frac{\Gamma(\mu+\nu)\,\Gamma(\mu+r)}{\Gamma(\mu)\,\Gamma(\mu+\nu+r)} = \frac{(\mu+r-1)(\mu+r-2)\ldots(\mu+1)\,\mu}{(\mu+\nu+r-1)(\mu+\nu+r-2)\ldots(\mu+\nu+1)(\mu+\nu)}.$$

de unde deducem :

Media : $E(X) = \dfrac{\mu}{\mu + \nu}$.

Dispersia : Var $X = \dfrac{\mu\nu}{(\mu + \nu)^2 (\mu + \nu + 1)}$.

Observații

1. Dacă X, Y sînt v.a. independente avînd *legea Gama** de parametrii μ, ν, respectiv, atunci v.a.

$$U = \frac{X}{X + Y}$$

are *l. Be* de parametrii μ, ν : $U \sim Be(\mu, \nu)$.

2. Pentru $\mu = \nu = 1$ se obține *legea uniformă**, deoarece $\Gamma(1) = \Gamma(2) = 1$.

3. Dacă v.a. independente X_1, $X_2 \sim$ *legea Gama* * de parametrii $(\alpha, 2)$, $(\beta, 2)$, respectiv, atunci v.a. $X_1/(X_1 + X_2) \sim Be(\alpha, \beta)$.

Legea Cauchy. Repartiția v.a. X avînd d.p. :

$$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2}, \quad -\infty < x < \infty$$

sau, mai general,

$$f(x; m, \sigma) = \frac{\sigma}{\pi} \frac{1}{\sigma^2 + (x - m)^2}, \quad x \in R, \ m \in R, \ \sigma > 0,$$

unde m, σ -parametrii *l.C.* (fig. 102).

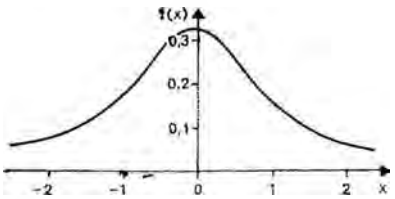


Fig. 102 Curba de frecvență a legii Cauchy cu $m = 0$ și $\sigma = 1$.

Caracteristici principale

Modul : $M_0 = 0$.

Funcția de repartiție :

$$F(x) = P(X < x) = \frac{1}{\pi} \int_{-\infty}^{x} \frac{dy}{1 + y^2} = \frac{1}{2} + \frac{1}{\pi} \operatorname{arc\,tg}\, x.$$

Funcția caracteristică : $\varphi(t) = e^{-|t|}$.

Observații

1. *L.C.* nu are momente de nici un ordin, deci nici dispersie.
2. Nu se aplică *teorema limită centrală** pentru *l.C.*
3. Este un caz particular al *legii t a lui Student**.

Legea arc sin. Este legea unei v.a. X a cărei d.p. este :

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}}.$$

Funcția de repartiție : $F(x) = 2/\pi \arcsin\sqrt{x}$, $0 \leqslant x \leqslant 1$. Această lege apare în teoria evenimentelor recurente.

Legea χ^2 (Helmert-Pearson). Repartiția v.a. X având d.p.

$$f(x) = \frac{1}{2^{\nu/2}\,\Gamma(\nu/2)}\, x^{\nu/2-1}\, e^{-x/2}, \quad \begin{matrix} 0 \leqslant x < \infty \\ \nu \geqslant 0, \text{ întreg.} \end{matrix}$$

Se spune că v.a. urmează *l.*χ^2 cu ν g.l. : $X \sim \chi^2_\nu$. *L.* χ^2 se numește *legea lui pătrat* cu ν g.l. sau *legea Helmert-Pearson* cu același număr de g.l. *L.* χ^2 a fost considerată prima oară în anul 1876 de Helmert și ulterior „redescoperită" de K. Pearson.

Caracteristici principale

Momentul de ordinul r :

$$M_r = 2^r\,(\nu/2) + (r-1)\,(\nu/2 + r - 2) \ldots (\nu/2 + 1)\,\nu/2,$$

de aici se obține :
Media : $E(X) = \nu$.
Dispersia : Var $X = 2\nu$.
Funcția caracteristică : $\varphi(t) = (1 - 2it)^{-\nu/2}$.
Graficele d.p. ale *l.*χ^2 pentru diferitele g.l. sînt date de fig. 103.

Observații

1. Dacă $\nu/2 = p$ și $a = \frac{1}{2}$, se obține d.p. a *legii Gama* *.

2. Dacă v.a. $X \sim \chi^2_\nu$, atunci v.a. $Z = \sqrt{2X} - \sqrt{2\nu - 1} \sim N(0,1)$.
3. $\chi^2_{\alpha;\,\nu}$ — α-cuantila superioară a legii χ^2 se determină astfel încît :

$$P(\chi^2_\nu > \chi^2_{\alpha;\nu}) = \alpha$$

și sînt tabelate.

4. Dacă v.a. independente $X_1 \sim \chi^2_{\nu_1}$, $X_2 \sim \chi^2_{\nu_2}$ atunci v.a. $X_1 + X_2 \sim \chi^2_{\nu_1+\nu_2}$ (proprietatea de stabilitate a legii χ^2).

5. Dacă v.a. independente $X_i \sim N(0,1)$, $i = 1, \ldots, s$, atunci v.a. $\sum_{i=1}^{s} X_i^2 \sim \chi^2_s$.

6. Dacă vectorul aleator $\mathbf{X} = (X_1, \ldots, X_s)'$ are *legea normală multidimensională* $N(\mathbf{m}, \mathbf{\Sigma})$, atunci forma pătratică pozitiv definită : $(\mathbf{x}-\mathbf{m})'\,\mathbf{\Sigma}^{-1}\,(\mathbf{x}-\mathbf{m})$ are legea χ^2_s.

Legea Helmert. Este legea v.a. abatere standard :

$$\sigma_e = \sqrt{\sum_i (x_i - \bar{x})^2/n}$$

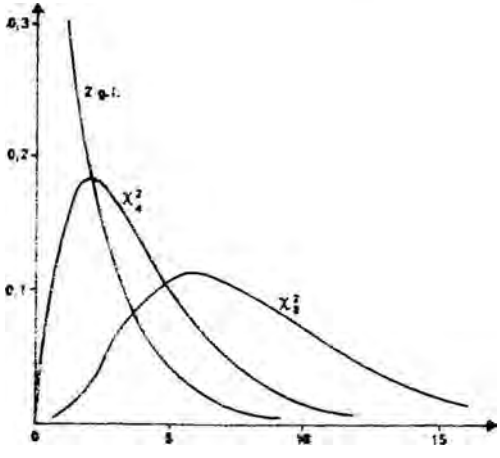


Fig. 103 Curba de frecvență a egii χ^2 cu diferite grade de libertate

a unui eșantion extras dintr-o populație normală cu dispersia σ^2. V.a. $n\,\sigma_e^2/\sigma^2$ este o v.a. cu *legea* χ^{2*} cu $\nu = n-1$ g.l.

Această lege a fost obținută de Helmert în 1876.

Legea t a lui Student. Repartiția v.a. X având d.p. :

$$f(t) = \frac{\Gamma\left(\frac{s+1}{2}\right)}{\sqrt{\pi\, s}\,\Gamma(s/2)} \left(1 + \frac{t^2}{s}\right)^{-(s+1)/2}$$

Se spune că X are *l.t S.* cu s g.l : $X \sim t_s$ **(fig. 104).**

Caracteristici principale

Funcția de repartiție :

$$F(x) = P(X \leqslant x) = \frac{\Gamma\left(\frac{s+1}{2}\right)}{\sqrt{\pi\, s}\,\Gamma(s/2)} \int_{-\infty}^{x} \left(1 + \frac{t^2}{s}\right)^{-(s+1)/2} dt.$$

Momente : momentele de ordin impar sînt nule ;
momente de ordin par :

$$M_{2p}(t) = s^p \frac{\Gamma\left(\frac{1}{2}+p\right)\Gamma\left(\frac{s}{2}-p\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma(s/2)}, \quad p < \frac{s}{2}.$$

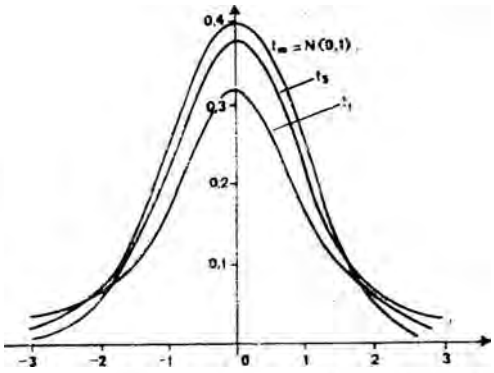


Fig. 104 Curba de frecvență a legii t Student cu diferite grade de libertate ($s = 1,5, \infty$)

De aici se obține :
Media : $E(X) = 0$.
Dispersia : $\text{Var } X = s/(s-2)$.
Observații
1. α-cuantila superioară a $t.S.$ cu s g.l., $t_{\alpha;s}$ se determină astfel încît :

$$P(t_s > t_{\alpha;s}) = \alpha$$

și sînt tabelate.
2. Dacă v.a. $X \sim N(0,1)$ și v.a. $Y \sim \chi^2_\nu$, X, Y v.a. independente, atunci v.a. $t = X/\sqrt{Y/\nu} \sim t_\nu$.
3. Legea limită a v.a. $X \sim t_s$, cînd $s \to \infty$, este legea $N(0,1)$.

Legea F Snedecor-Fisher. Repartiția v.a. X avînd d.p.

$$f(x) = \frac{\Gamma((s_1+s_2)/2)}{\Gamma(s_1/2)\,\Gamma(s_2/2)} (s_1/s_2)^{s_1/2}\, x^{s_1/2-1} \left(1 + \frac{s_1}{s_2}x\right)^{-\frac{s_1+s_2}{2}}, \quad 0 \leqslant x < \infty.$$

Se notează $X \sim F_{s_1, s_2}$, adică v.a. X are $t.F$ cu s_1, s_2 g.l.
Graficul d.p. al $t.F_{s_1, s_2}$ nu este simetric (fig. 105).

Caracteristici principale
Momentul de ordin r:

$$M_r = EX^r = \left(\frac{s_1}{s_2}\right)^r \frac{\Gamma\left(\frac{s_1}{2} + r\right)\Gamma\left(\frac{s_2}{2} - r\right)}{\Gamma\left(\frac{s_1}{2}\right)\Gamma\left(\frac{s_2}{2}\right)}, \quad r < \frac{s_2}{2}$$

unde $\Gamma(\nu)$ este *funcția Gamma**.

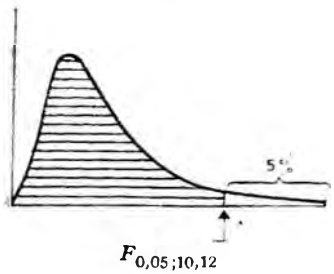


Fig. 105 Curba de frecvență a legii F cu $s_1 = 10$, $s_2 = 12$ grade de libertate

De aici se obține:
Media: $E(X) = s_2/(s_2-2)$
Momentul de ordin 2: $M_2(X) = s_2^2/s_1(s_2-2)(s_2-4)$

Dispersia: $\text{Var } X = \dfrac{2\,s_2^2\,(s_1 + s_2 - 2)}{s_1(s_2-2)^2(s_2-4)}$, $s_2 > 4$

Funcția de repartiție:

$$F(x) = P(X < x) = \frac{\Gamma\left(\frac{s_1 + s_2}{2}\right)}{\Gamma(s_1/2)\,\Gamma(s_2/2)} \left(\frac{s_1}{s_2}\right)^{\frac{s_1}{2}} \int_0^x y^{\frac{s_1}{2}-1} \left(1 + \frac{s_1}{s_2} y\right)^{-\frac{s_1+s_2}{2}} dy.$$

Cînd $s_2 \to \infty$, $E(X) = 1$, $\text{Var } X = 2/s_1$, adică media respectiv dispersia unei v.a. $\chi^2_{s_1}/s_1$. Deci v.a. $s_1 F_{s_1, \infty} \sim \chi^2_{s_1}$. (→Observația 2 de mai jos.)
$F_{\alpha; s_1, s_2}$ — α-cuantila v.a. $X \sim F_{s_1, s_2}$ se determină astfel încît:

$$P(F_{s_1, s_2} > F_{\alpha; s_1, s_2}) = \alpha.$$

Între v.a. F_{s_1, s_2} și v.a. F_{s_2, s_1} există relațiile:

$$F_{s_1, s_2} = \frac{1}{F_{s_2, s_1}}.$$

și, corespunzător, pentru f.r. respective :

$$F_{F_{s_1,\,s_2}}(x) = 1 - F_{F_{s_2,\,s_1}}(1/x).$$

Observații

1. Dacă v.a. X_1, X_2 independente, $X_1 \sim \chi^2_{s_1}$ și $X_2 \sim \chi^2_{s_2}$, atunci v.a.

$$F = \frac{X_1/s_1}{X_2/s_2} \sim F_{s_1,\,s_2}.$$

2. Cînd $s_2 \to \infty$, v.a. $s_1 F_{s_1,\,\infty} \sim \chi^2_{s_1}$.

3. *I..F.* are importanță deosebită cînd se compară *dispersiile de selecție** ale unor populații normale, în *analiza de regresie**, în *analiza de varianță**,

4. Fie v.a. $X \sim F_{s_1,\,s_2}$. Se arată ușor că v.a.

$$\frac{1}{1+\frac{s_1}{s_2}X}, \quad \frac{s_1/s_2\,X}{1+s_1/s_2\,X}$$

au, respectiv, *legea Beta**, $Be\left(\frac{s_2}{2}, \frac{s_1}{2}\right)$ și *legea Beta**, $Be\left(\frac{s_1}{2}, \frac{s_2}{2}\right)$.

5. Fie v.a. $X \sim F_{s_1,\,s_2}$, atunci :

$$F_{1,\,s_2} = \frac{U^2}{\chi^2_{s_2}/s_2} \sim t^2_{s_2},$$

unde v.a. $U \sim N(0,1)$.

De aici se deduc relațiile între α-cuantilele corespunzătoare ale legilor respective.

Legea Z (a lui Fisher). Repartiția v.a. $Z = \frac{1}{2}\ln X$, unde $X \sim F_{s_1,\,s_2}$ este *l. Z* cu s_1, s_2 g.l. A fost considerată de R. A. Fisher în 1924.

D.p. a v.a. Z este :

$$f(z) = 2\,\frac{\Gamma\left(\frac{s_1+s_2}{2}\right)}{\Gamma\left(\frac{s_1}{2}\right)\Gamma\left(\frac{s_2}{2}\right)}\left(\frac{s_1}{s_2}\right)^{\frac{s_1}{2}} e^{s_1 z}\left(1+\frac{s_1}{s_2}e^{2z}\right)^{-\frac{s_1+s_2}{2}},$$

$$-\infty < z < \infty.$$

Funcția de repartiție a v.a. Z :

$$F(x) = P(Z \leqslant x) = \int_{-\infty}^{x} f(z)\, dz = F_{F_{s_1, s_2}}(e^{2x}).$$

$Z_{\alpha; s_1, s_2}$ — α-cuantila superioară v.a. Z cu s_1, s_2 g.l. se determină astfel încît :

$$P(Z_{s_1, s_2} > Z_{\alpha; s_1, s_2}) = \alpha.$$

Între α-cuantilele v.a. Z_{s_1, s_2} și F_{s_1, s_2} are loc relația :

$$Z_{\alpha; s_1, s_2} = \frac{1}{2} \ln F_{\alpha; s_1, s_2}.$$

De asemenea se observă că :

$$Z_{\alpha; s_1, s_2} = - Z_{1-\alpha; s_1, s_2}.$$

Legea χ^2 necentrată. Repartiția v.a.

$$\chi^2_{\nu, \delta} = \sum_{i=1}^{\nu} U_i^2$$

unde $\{U_i\}$, $1 \leqslant i \leqslant \nu$ sînt v.a. independente, $U_i \sim N(m_i, 1)$.
Se notează $\chi^2_{\nu; \delta}$ și se numește v.a. χ^2 necentrată cu ν g.l. și cu parametrul de excentricitate (sau necentricitate) δ, unde :

$$\delta^2 = \sum_{i=1}^{\nu} m_i^2$$

D.p. a v.a. $\chi^2_{\nu; \delta}$ este :

$$h_{\nu; \delta}(x) = \frac{1}{2^{\nu/2} \sqrt{\pi}} e^{-\frac{1}{2}(\delta^2 + x)} x^{\nu/2 - 1} \sum_{\beta=0}^{\infty} \frac{(\delta^2 x)^\beta}{(2\beta)!} \frac{\Gamma(\beta + 1/2)}{\Gamma(\nu/2 + \beta)}$$

Funcția de repartiție :

$$H_{\nu, \delta}(x) = \int_0^x h_{\nu; \delta}(u)\, du.$$

Observații.

1. P. B. Patnaik în „Biometrika" 36/1949 a dat o aproximație a v.a. necentrate χ^2 cu ν g.l. și parametrul de excentricitate δ, printr-o v.a. χ^2 centrată :

$$\chi^2_{\nu;\,\delta} = C\,\chi^2_{\nu^*},$$

unde constantele C și ν^* se determină astfel încît :

$$E(\chi^2_{\nu;\,\delta}) = C\,E(\chi^2_{\nu^*})$$

$$\text{Var}\,(\chi^2_{\nu;\,\delta}) = C^2\,\text{Var}\,(\chi^2_{\nu^*}),$$

adică :

$$\nu + \delta^2 = C\ \nu^*$$

$$\nu + 2\delta^2 = C^2\,\nu^*$$

2. Funcția de repartiție a v.a. $\chi^2_{\nu;\,\delta}$ devine

$$H_{\nu;\,\delta}(x) = H_{\nu^*}\left(\frac{x}{C}\right),$$

unde $H_{\nu^*}(x/C)$ este funcția de repartiție a v.a. centrate $\chi^2_{\nu^*}$.

3. Valorile α-cuantilelor superioare ale *l.* χ^2 *n.* se determină astfel încît :

$$P(\chi^2_{\nu;\,\delta} < \chi^2_{\nu;\,\delta;\,\alpha}) = 1 - \alpha.$$

Legea t necentrată. Repartiția v.a.

$$t_\delta = U/\sqrt{V/s},$$

unde $U \sim N(0,1)$; $V \sim \chi^2_s$, v.a. independente.

L. t n. se obține dacă v.a. U sau V sînt fie pe rînd, fie amîndouă deodată v.a. necentrate.

Legea F necentrată. Repartiția v.a.

$$F_{\nu_1,\,\nu_2} = \frac{X_1/\nu_1}{X_2/\nu_2},$$

unde $X_1 \sim \chi^2_{\nu_1}$, $X_2 \sim \chi^2_{\nu_2}$ v.a. independente.

L. F n, se obține dacă v.a. X_1, X_2 sînt fie pe rînd, fie amîndouă deodată necentrate. Așadar, v.a.

$$F_{\nu_1,\,\nu_2;\,\delta_1} = \frac{\nu_2\,\chi^2_{\nu_1;\,\delta_1}}{\nu_1\,\chi^2_{\nu_2}}\ ;\ F_{\nu_1,\,\nu_2;\,\delta_2} = \frac{\nu_2\,\chi^2_{\nu_1}}{\nu_1\,\chi^2_{\nu_2;\,\delta_2}}$$

sînt v.a. simplu necentrate. V.a.

$$F_{\nu_1, \nu_2; \delta_1, \delta_2} = \frac{\nu_2 \chi^2_{\nu_1; \delta_1}}{\nu_1 \chi^2_{\nu_2; \delta_2}}$$

este o v.a. dublu necentrată.

Observații

1. Se observă că :

$$F_{\nu_1, \nu_2; \delta_1} = 1/F_{\nu_2, \nu_1; \delta_1}$$

și deci este suficient să fie cunoscută numai f.r. a v.a. $F_{\nu_1, \nu_2; \delta_1}$.

2. Între v.a. necentrate $t_{\nu; \delta}$ și $F_{1, \nu; \delta}$ există relația :

$$t^2_{\nu; \delta} = F_{1, \nu; \delta}.$$

3. D.p. a v.a. $F_{\nu_1, \nu_2; \delta_1}$ simplu necentrată este :

$$f_{\nu_1, \nu_2 | \delta_1}(x) = \frac{\nu_1}{\nu_2} \frac{e^{-\delta_1^2/2}}{\Gamma(\nu_2/2)} \sum_{\beta=0}^{\infty} \frac{\left(\frac{\delta_1^2}{2}\right)^{\beta} (\nu_1/\nu_2 \cdot x)^{\frac{\nu_1}{2}+\beta-1} \Gamma\left(\frac{\nu_1 + \nu_2}{2} + \beta\right)}{\beta ! \Gamma\left(\frac{\nu_1}{2} + \beta\right) \left(1 + \frac{\nu_1}{\nu_2} x\right)^{\frac{\nu_1+\nu_2}{2}+\beta}}.$$

Legea maximului studentizat al modulului (*legea maximului modulului studentizat*). Repartiția v.a.

$$m_{k, \nu} = \{\max_i | X_i |\} / s,$$

unde v.a. $\{X_i\}_{1 \leqslant i \leqslant k}$, independente, $X_i \sim N(0, \sigma^2)$, $\forall\ i$;
s^2 este un estimator pătratic independent cu ν g.l. al lui σ^2, adică $\nu s^2/\sigma^2 \sim \chi^2_\nu$;
s^2 independent de $\{X_i\}_{1 \leqslant i \leqslant k}$.

Valorile α-cuantilelor superioare ale *l.m.s.m.*, $m_{\alpha; k, \nu}$ cu k, ν g.l. se determină astfel încît :

$$P(m_{k, \nu} < m_{\alpha; k, \nu}) = 1 - \alpha.$$

Pentru diferite valori ale lui k, ν și $\alpha = 0{,}05$, respectiv 0,01, valorile α-cuantilelor superioare sînt tabelate.

Legea amplitudinii studentizate. Repartiția v.a.

$$q_{k, \nu} = (\max_i X_i - \min_i X_i)/s$$

cu k, ν g.l., unde v.a. X_i, $i = 1, \ldots, k$ sînt v.a. independente $X_i \sim N(m, \ldots$, $\nu s^2/\sigma^2 \sim \chi^2_\nu$ și s^2 este independent de $\{X_i\}$.

Valorile α-cuantilelor superioare $q_{\alpha;k,\nu}$ determinate astfel încît $P(q_{k,\nu} < q_{\alpha;k,\nu}) = 1 - \alpha$ sînt tabelate.

Legea amplitudinii studentizate mărite. Repartiția v.a. $q_{k,\nu}$ cu k, ν g.l.

$$q'_{k,\nu} = \max\{q_{k,\nu}, m_{k,\nu}\} = \max\left\{\frac{\max\limits_i X_i - \min\limits_i X_i}{s}, \frac{\max\limits_i |X_i|}{s}\right\},$$

unde v.a. $\{X_i\}$, $i = 1, \ldots, k$ independente $\sim N(0, \sigma^2)$; s^2 — un estim... pătratic independent cu ν g.l. al lui σ^2, adică $\nu s^2/\sigma^2 \sim \chi^2_\nu$.

Valorile α-cuantilelor superioare ale *l.a.s.m.* $q'_{\alpha;k,\nu}$ sînt determinate ...ncit

$$P(q'_{k,\nu} < q'_{\alpha;k,\nu}) = 1 - \alpha.$$

Dacă $\alpha \leqslant 0{,}05$ și $k > 2$, atunci

$$q'_{\alpha;k,\nu} \approx q_{\alpha;k,\nu}$$

unde $q_{\alpha;k,\nu}$ sînt α-cuantilele superioare ale *legii amplitudinii st...uden...*

Legea maximului studentizat al abaterii față de media de selec...ie. ...sideră v.a.

$$Z_i = (X_i - \overline{X})/\sigma, \quad i = 1, \ldots, k$$

unde $\{X_i\}$, $1 \leqslant i \leqslant k$ sînt v.a. independente $\sim N(m, \sigma^2)$; $\overline{X} = \dfrac{\sum_{i=1}^{n} \ldots}{n}$

a) *Cazul σ-cunoscut*

Se calculează:

$$E(Z_i) = 0$$

$$\text{Var}\ (Z_i) = (k-1)/k$$

$$\text{Cov}\ (Z_i, Z_j) = -1/k$$

$$\text{Cor}\ (Z_i, Z_j) = -1/(k-1).$$

V.a $H_{k,\nu} = \max\limits_{1 \leqslant i \leqslant k} |Z_i|$ se numește *maximul abaterii absolute față de ... de selecție* cu k și ν g.l. $H_{\alpha;k,\nu} = H_\alpha$ este α-cuantila superioară ...clete astfel încît

$$P(|Z_1| < H_\alpha, \ldots, |Z_k| < H_\alpha) = 1 - \alpha \tag{1}$$

sau, echivalent,

$$P(\max |Z_i| < H_\alpha) = 1 - \alpha.$$

V.a. $Z_1, \ldots, Z_k$ au repartiția comună normală k-dimensională de rang $(k-1)$, adică vectorul aleator $\mathbf{Z} = (Z_1, \ldots, Z_k)'$ urmează *legea normală k-dimensională** cu vectorul medie $(0, \ldots, 0)'$ și matricea de covarianță de rang $(k-1)$.
Cuantilele repartiției v.a. $\delta = \lim d$, unde $d = \{\max_{1 \leqslant i \leqslant k} | X_i - \overline{X} |\}/s$ se obțin ca soluție a ecuației (1). O soluție aproximativă se obține ținînd seamă de Cor (Z_i, Z_j) și aplicînd *inegalitățile lui Bonferroni* :

$$1 - \sum_{i=1}^{k} P(\bar{A}_i) \leqslant P(A_1, A_2 \ldots A_k) \leqslant 1 - \sum_{i=1}^{k} P(\bar{A}_i) + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} P(\bar{A}_i \bar{A}_j),$$

unde

$$P(\bar{A}_i) = P(|Z_i| > h_\alpha) = 2/\sqrt{2\pi} \int_{\sqrt{k/(k-1)} \cdot h_\alpha}^{\infty} e^{-x^2/2}\, dx, \quad i = 1, 2, \ldots, k$$

$$P(\bar{A}_i \bar{A}_j) = 2 \int_{\sqrt{k/(k-1)} h_\alpha}^{\infty} \int_{\sqrt{k/(k-1)} h_\alpha}^{\infty} p(x_1, x_2)\, dx_1 dx_2 +$$

$$+2 \int_{-\infty}^{-\sqrt{k/(k-1)}\, h_\alpha} \int_{\sqrt{k/(k-1)} h_\alpha}^{\infty} p(x_1, x_2)\, dx_1 dx_2 \quad i \neq j;\ i,\ j = 1, \ldots, k.$$

$p(x_1, x_2)$ este d.p. a *legii normale bidimensionale** a vectorului aleator $\mathbf{X} = (X_1, X_2)'$ cu componentele v.a. X_i cu: $E(X_i) = 0$, Var $X_i = 1$, $\rho = \text{Cor}\,(X_i, X_j)' = -1/(k-1)$.

b) *Cazul σ-necunoscut*

Fie s o *estimație nedeplasată* * a lui σ cu ν g.l. V.a.:

$$H_{k,\nu} = \frac{\max_i \{| X_i - \overline{X} |\}/\sigma}{s/\sigma}$$

se numește *m.s.a.a.m.s*, unde $\nu s^2/\sigma^2 \sim \chi^2_\nu$ cu ν g.l. și este independentă de $\{X_1, \ldots, X_k\}$.

Se reface raționamentul din cazul σ-cunoscut, înlocuind pe h_α cu $s/\sigma \cdot h_\alpha$ și ținînd seamă că:

$$P(\bar{A}_i) = \int_{\sqrt{k/(k-1)}h_\alpha}^{\infty} p(t)\,dt, \quad i = 1, \ldots, k$$

$$P(\bar{A}_i \bar{A}_j) = 2 \int_{\sqrt{k/(k-1)}h_\alpha}^{\infty} \int_{\sqrt{k/(k-1)}h_\alpha}^{\infty} p(t_1, t_2)\,dt_1 dt_2 +$$

$$+ 2 \int_{-\infty}^{-\sqrt{k/(k-1)}h_\alpha} \int_{\sqrt{k/(k-1)}h_\alpha}^{\infty} p(t_1, t_2)\,dt_1 dt_2,$$

unde $p(t)$ este d.p. *t a lui Student** cu ν g.l., iar $p(t_1, t_2)$ este d.p. *t Student bidimensională** cu $\varrho = -1/(k-1)$.

Valorile $H_{\alpha;\, k,\, \nu}$ au fost tabelate de M. Halperin, W. Greenhouse și J. Zalokar folosind o teoremă de acoperire simultană dată de T. W. Tukey, dedusă direct din *inegalitățile lui Bonferroni*.

Legea maximului studentizat al abaterii față de media teoretică. E. G. Schilling a tabelat valorile marginii superioare a maximului studentizat al abaterii absolute față de media populației sau media teoretică, notate cu h_α, adică

$$P\left(\max_i |X_i - m|/s_x \leqslant h_\alpha\right) = 1 - \alpha$$

cu $h_\alpha = h_{\alpha;\, k,\, \nu}$, unde $\{X_i\}$, $1 \leqslant i \leqslant k$ sînt v.a. considerate anterior.

Între α-cuantilele legii *m.s.a.f.m.t.* și cea a *maximului studentizat al abaterii față de media de selecție* există relația:

$$h_\alpha = H_\alpha \sqrt{k/(k-1)}.$$

O aproximație a intervalelor de încredere și teste bazate pe maximul studentizat al modulului pentru h au fost date de G. Enderlein, 1972.

Legea exponențială. Repartiția v.a. X având d.p.

$$f(x) = \lambda\, e^{-\lambda x};\ x > 0,\ \lambda > 0.$$

L.e. este caz particular al *legii Gamma** pentru $p = 1$, $a = 1/\lambda$ sau al *legii Weibull** pentru $\beta = 1$, $\gamma = 0$, cu notația $\lambda = 1/\eta$.

Caracteristici principale

Media: $E(X) = 1/\lambda$.

Dispersia: $\mathrm{Var}(X) = 1/\lambda^2$.

Rata de defectare: $r(x) = \lambda$.

Funcția de repartiție: $F(x) = 1 - e^{-\lambda x}$

Observații

1. *L.e.* joacă un rol important în *teoria fiabilității* *, aplicațiile statisticii matematice în tehnică. Prezintă însă o rigiditate la adecvarea sa unei clase largi de date experimentale care descriu timpul de funcționare fără defecțiuni.
2. V.a. având d.p.

$$f(x) = 1/2\ e^{-|x-\mu|}$$

urmează *legea dublu exponențială*.

Legea Weibull. Repartiția v.a. X având d.p.

$$f_x(x) = \frac{\beta}{\eta}\left(\frac{x-\gamma}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{x-\gamma}{\eta}\right)^{\beta}\right], \begin{matrix}\gamma \leqslant x < \infty \\ \gamma \geqslant 0\end{matrix}$$

β, γ, η reprezintă parametri pozitivi

β — parametrul de formă

η — parametrul de scală

γ — parametrul de locație.

Caracteristici principale

Media : $$E(X) = \gamma + \eta\,\Gamma\left(1 + \frac{1}{\beta}\right).$$

Dispersia : $\mathrm{Var}\ X = \eta^2\,[\Gamma(1 + 4\beta) - \Gamma^2(1 + 1/\beta)]$.

Modul : $M_0 = \gamma + \eta\ (1 - 1/\beta)^{1/\beta}$.

Mediana : $M_e = \gamma + \eta\ (\ln 2)^{1/\beta}$

unde $\Gamma(t)$ este *funcția Gamma* *

$$\Gamma(t) = \int_0^{\infty} a^{t-1}\ e^{-a}\ da.$$

Această lege joacă un rol important în *teoria fiabilității* * unui material, în testarea durabilității componentelor sistemelor electronice, în demografie etc. A fost considerată de W. Weibull în 1951.

În aplicațiile *l. W.* se folosesc frecvent conceptele de *rată de defectare** și *funcție de fiabilitate* *.

Rata de defectare $r(x)$ sau *rata de hazard* este probabilitatea de defectare într-un interval de timp foarte mic a unui sistem care a funcționat x ore, adică :

$r(x)\,dx = P\{$un sistem care a funcționat x ore să se defecteze în intervalul $(x,\ x + dx)\} = dF(x)/(1 - F(x)) = f(x)/(1 - F(x))$.

În cazul *l.W.* rata de defectare este :

$$r(x) = \frac{f(x)}{1 - F(x)} = \frac{\beta}{\eta}\left(\frac{x-\gamma}{\eta}\right)^{\beta-1}.$$

Funcția de fiabilitate este probabilitatea de supraviețuire la momentul x. În cazul *l.W.*, funcția de fiabilitate este :

$$R(x) = 1 - F(x) = P(X > x) = e^{-\left(\frac{x-\gamma}{\eta}\right)^{\beta}}.$$

Pentru $\beta = 1$ se obține *repartiția exponențială*.
Pentru $\beta = 2$ se obține *repartiția Rayleigh* *.
L.W. se mai poate prezenta astfel :

$$F(x) = 1 - e^{-\left(\frac{x}{a}\right)^{b}}, \quad x \geqslant 0 \quad a,\ b > 0$$

$$F(x) = 1 - e^{-x^k/\lambda}, \quad x \geqslant 0 \quad k,\ \lambda > 0$$

$$F(x) = 1 - e^{-\theta x^k} \quad , \quad x \geqslant 0 \quad \theta,\ k > 0.$$

O proprietate importantă este că dacă X este o v.a. Weibuli, atunci X^k este v.a. *de tip exponențial* *.

Legea Rayleigh. Caz particular al legii Weibull pentru $\beta = 2$.
Funcția de repartiție a *l.R.* se mai poate scrie astfel :

$$F(x;\theta) = 1 - e^{-\theta x^2}, \quad x \geqslant 0 \quad \theta > 0.$$

Rata de defectare :

$$r(x) = \frac{f(x,\theta)}{1 - F(x,\theta)} = 2\theta x, \quad \begin{matrix} x \geqslant 0 \\ \theta > 0. \end{matrix}$$

Această lege apare în aplicațiile statisticii matematice în tehnică fiind caracteristică proceselor de uzură, în fizica moleculară etc.

Legea Rayleigh generalizată. Repartiția unei v.a. X având d.p. :

$$f(t,\theta,k) = \frac{2\theta^{k+1}}{\Gamma(k+1)} t^{2k+1} e^{-\theta t^2}, \quad t \geqslant 0,\ \theta > 0,\ k > -1.$$

În particular, dacă $\theta = \frac{1}{\eta^2}$, $k = 0$, se obține d.p. a *legii Rayleigh*.

Pentru $\theta = \frac{1}{2\eta^2}$, $k = \frac{1}{2}$, se obține *legea Maxwell* având d.p.

$$f(t,\eta) = \frac{1}{\eta^3\sqrt{2\pi}} t^2 e^{-t^2/2\eta^2}, \quad \eta > 0.$$

Această lege își găsește aplicații în tehnică la procesele de uzură, în fizica moleculară etc.

Pentru $\theta = \frac{1}{2\tau^2}$, $k = \nu/2 - 1$, se obține *legea* χ^{2*} cu v.g.l. având d.p.:

$$f(t, \nu, \tau) = \frac{1}{2^{\nu/2-1}\tau^{\nu}\,\Gamma(\nu/2)}\; t^{\nu-1}\, e^{-t^2/2\tau^2}.$$

Caracteristici principale

Media: $E(X) = \Gamma(k - 3/2)/\Gamma(k + 1) \cdot 1/\sqrt{\theta}$

Dispersia: $\mathrm{Var}\ X = [(\Gamma(k + 1) - \Gamma^2(k - 3/2)/\Gamma^2(k + 1)] \cdot 1/\theta.$

Funcția de repartiție:

$$F(t, \theta, k) = \frac{2\theta^{k-1}}{\Gamma(k+1)} \int_0^t x^{2k-1} e^{-\theta x^2}\, dx = \frac{\Gamma_{\theta t^2}(k+1)}{\Gamma(k+1)},$$

unde $\Gamma_{t^2\theta}(k + 1)$ este funcția Gamma incompletă.

Legea Gram-Charlier. Legea unei v.a. Y prezentată sub forma unei serii propusă de Gram în 1883 și de Charlier în 1905 fie pentru a reprezenta legea unei v.a. discrete în funcție de diferențele succesive ale unei v.a. Poisson (*l.G.-C. – tip B*), fie pentru a obține o reprezentare aproximativă a legii unei v.a. continue în funcție de d.p. a legii normale $N(0,1)$ (*l.G.-C.–tip A*).

L. G.-C. –tip B. Se reprezintă f.f. $p_Y = P_j$ a v.a. discrete Y ca o serie liniară de *diferențe* succesive în raport cu j ale lui $p_x = \omega_j$, f.f. ale unei v.a. Poisson:

$$P_j = \sum_{i=0}^{\infty} a_i \nabla^i \omega_j,$$

unde ∇ este operatorul diferență succesivă:

$$\nabla \omega_j = \omega_j - \omega_{j-1}$$

cu $\omega_j = p_X$, f.f.a *legii Poisson* Po* (λ):

$$\omega_j = \begin{cases} e^{-\lambda}\ \lambda^j/j!, & j \geqslant 0 \\ 0\ , & j < 0 \end{cases}$$

$$\sum_{j=c}^{\infty} \omega_j = 1, \text{ toți } c \leqslant 0.$$

Funcția de repartiție a v.a. discrete Y:

$$F(c) = P(Y < c) = 1 - \sum_{j=c}^{\infty} P_j = 1 - \sum_{i=0}^{\infty} a_i \nabla^i \left(\sum_{j=c}^{\infty} \omega_j \right).$$

L. G.-C.–tip A. Se reprezintă aproximativ $g(x)$ – d.p. a v.a. continue Y sub forma unei serii în funcție de $f(x)$ – d.p. a *legii normale* N(0,1)*:

$$g(x) = f(x) - \varepsilon_1 Df(x) - 1/2(\varepsilon_1^2 + \varepsilon_2) D^2 f(x) -$$

$$- 1/6(\varepsilon_1^3 + 3\varepsilon_1\varepsilon_2 + \varepsilon_3) D^3 f(x) + \ldots,$$

unde D este operatorul de diferențiere :

$$D^j f(x) = \mathrm{d}^j f(x) / \mathrm{d}x^j$$

$$D^j f(x) = (-1)^j H_j(x) f(x)$$

cu $H_j(x)$ — polinomul Hermite de gradul j :

$\varepsilon_1 = \varepsilon_2 = 0$; $\varepsilon_3 = \sqrt{\beta_1}$, $\varepsilon_4 = \beta_2 - 3$, unde :
$\beta_1 = \mu_3^2/\mu_2^3$ — *coeficientul de asimetrie al lui K.Pearson**
$\beta_2 = \mu_4/\mu_2^2$ — *coeficientul de exces**.
Așadar, d.p. $g(x)$ a v.a. Y se scrie :

$$g(x) = [1 + 1/6\sqrt{\beta_1}\, H_3(x) + 1/24(\beta_2 - 3)H_4(x) + \ldots](\sqrt{2\pi})^{-1} e^{-x^2/2}$$

avînd funcția de repartiție ;

$$G(x) = P(Y \leqslant x) = \Phi(x) - 1/6\beta_1 \sqrt{\beta_1}(x^2 - 1)f(x) -$$
$$- 1/24(\beta_2 - 3)(x^3 - 3x) f(x) + \ldots,$$

unde $\Phi(x) = 1/\sqrt{2\pi} \int_{-\infty}^{x} e^{-t^2/2}\, \mathrm{d}t$, funcția de repartiție a v.a. $\sim N(0,1)$ cu d.p. $f(x)$. O altă dezvoltare este *dezvoltarea Edgeworth* prin care d.p. a v.a. Y, respectiv funcția de repartiție se scrie :

$$g(x) = [1 + 1/6\sqrt{\beta_1} H_3(x) + 1/24(\beta_2 - 3) H_4(x) +$$
$$+ 1/72\beta_1 H_6(x) + \ldots] f(x),$$

respectiv

$$G(x) = \Phi(x) - 1/6\sqrt{\beta_1}(x^2 - 1)f(x) - 1/24(\beta_2 - 3)(x^3 - 3x)f(x) -$$
$$- 1/72\beta_1(x^5 - 10x^3 + 15x)f(x) + \ldots$$

$f(x) = (\sqrt{2\pi})^{-1} e^{-x^2/2}$, d.p.a v.a. $X \sim N(0,1)$. $\Phi(x) = (\sqrt{2\pi})^{-1} \int_{-\infty}^{x} e^{-t^2/2}\, \mathrm{d}t$.

De cele mai multe ori, în aplicații este suficientă folosirea primelor 4 momente, așa încît d.p. în dezvoltarea Gram-Charlier are forma :

$$g(x) = [1 - 1/6\sqrt{\beta_1}(x^3 - 3x) + 1/24(\beta_2 - 3)(x^4 - 6x^2 + 3)] f(x)$$

și d.p. în dezvoltarea Edgeworth are forma :

$$g(x) = [1 - 1/6\sqrt{\beta_1}(x^3 - 3x) + 1/24(\beta_2 - 3)(x^4 - 6x^2 + 3) +$$
$$+ 1/72\beta_1(x^5 - 10x^3 + 15x)] \cdot f(x),$$

unde $f(x)$ este d.p. a v.a. $X \sim N(0,1)$.

Legi de tip Pearson. Legile unei familii importante de v.a. a căror d.p. verifică ecuația diferențială :

$$1/f(x)\cdot \mathrm{d}f(x)/\mathrm{d}x = (x - a)/(b_0 + b_1 x + b_2 x^2),$$

a, b_0, b_1, b_2 fiind parametri.
După valorile parametrilor se disting mai multe tipuri de *l.P.*
Legea de tip I a lui Pearson (*l.P.I.*) Legea unei v.a. a cărei d.p. este :

$$f(x) = k(1 + x/a_1)^{m_1}(1 - x/a_2)^{m_2}\ ;\ -a_1 \leqslant x \leqslant a_2,\quad m_1, m_2 > -1$$

cu

$$m_1/m_2 = a_1/a_2 \ (\rightarrow \text{fig. } 106).$$

Printr-o schimbare de variabile și de parametri :

$$x + a_1 = z(a_1 + a_2)\ ;\ p = m_1 + 1,\ q = m_2 + 1$$

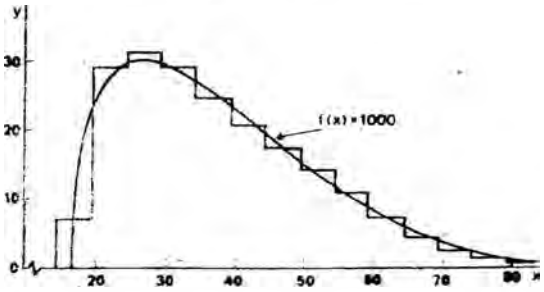


Fig. 106 Curba de frecvență a legii Pearson de tip I folosită pentru ajustarea unor date de frecvență

I.P.I. se reduce la *legea Beta* ** de parametri p și q (→fig. 101).
Legea de tip II a lui Pearson (*l.P. II*). Legea unei v.a. a cărei d.p. este :

$$f(x) = k(1 - x^2/a^2)^m\ ;\ -a \leqslant x \leqslant a,\ m > -1.$$

Este un caz particular al celui precedent care se poate pune sub forma d.p. a *legii Beta* * de parametri p și p.
Pentru $m = 0$ sau $p = 1$ se obține *legea uniformă* *.
Legea de tip III a lui Pearson (*l.P. III*). Legea unei v.a. a cărei d.p. este :

$$f(x) = k\,(1 + x/a)^{\gamma a}\mathrm{e}^{-\gamma x}\ ;\ -a \leqslant x \leqslant \infty$$

$$\gamma,\ a > 0.$$

Printr-o schimbare de variabilă și de parametri :

$$\gamma(x + a) = z,\quad \gamma a = p$$

se reduce la *legea Gamma* * de parametru p. (→fig. 100).
Conține drept caz particular *legea* χ^2.* (→fig. 103).

Legea de tip IV a lui Pearson (*l.P.* IV). Legea unei v.a. a cărei d.p. este :

$$f(x) = k(1+x^2/a^2)^{-m}e^{-\nu \operatorname{arctg}(x/a)};\ -\infty < x < \infty,\quad a, \nu > 0.$$

Legea de tip V a lui Pearson (*l.P.* V). Legea unei v.a. a cărei d.p. este :

$$f(x) = kx^{-p}e^{-\gamma/x};\ 0 \leqslant x < \infty,\ \gamma > 0,\ p > 1.$$

Printr-o schimbare de variabilă definită prin :

$$z = \gamma/x$$

se reduce la *l.P.* III.

Legea de tip VI a lui Pearson (*l.P.* VI). Legea unei v.a. a cărei d.p. este :

$$f(x) = kx^{-q_1}(x-a)^{q_2};\ a \leqslant x < \infty,\ q_1 > q_2 - 1$$

Printr-o schimbare de variabilă se reduce la *l.P.* I.

Legea de tip VII a lui Pearson (*l.P.* VII). Legea unei v.a. a cărei d.p. este :

$$f(x) = k(1 + x^2/a^2)^{-m};\ -\infty < x < \infty,\ m > 1/2$$

Caz particular : *legea t a lui Student**

Legea de tip VIII a lui Pearson (*l.P.* VIII). Legea unei v.a. a cărei d.p. este :

$$f(x) = k(1 + x/a)^{-m};\ -\infty < x < \infty,\ 0 \leqslant m \leqslant 1.$$

Legea de tip IX a lui Pearson (l.P. IX). Legea unei v.a. a cărei d.p. este :

$$f(x) = k(1+x/a)^{-m};\ -a \leqslant x \leqslant 0,\ m > -1.$$

Legea de tip X a lui Pearson (*l.P.* X). Legea unei v.a. a cărei d.p. este :

$$f(x) = 1/\sigma \exp\left[-\frac{x-m}{\sigma}\right],$$

adică este o lege de tip exponențial.

Legea de tip XI a lui Pearson (*l.P.* XI). Legea unei v.a. a cărei d.p. este :

$$f(x) = kx^{-m} = (m-1)a^{m-1}x^{-m};\ a \leqslant r < \infty,\ m > 1.$$

adică este o *lege Pareto.**

Legea de tip XII a lui Pearson (*l.P.* XII). Legea unei v.a. a cărei d.p. este :

$$f(x) = [(1 + x/a_1)/(1 - x/a_2)]^m;\ -a_1 \leqslant x \leqslant a_2,\ m > 1.$$

Legea Pareto. Legea unei v.a. a cărei d.p. este :

$$f(x) = \alpha x_0 (x_0/x)^{\alpha+1};\ 0 < x_0 < x < \infty,$$

unde $\alpha > 0$ — parametrul de formă numit constanta lui Pareto.

Caracteristici principale

Media : $E(X) = \alpha/(\alpha - 1)\ x_0,\ \alpha > 1.$

Dispersia : $\operatorname{Var} X = \alpha x_0^2/(\alpha - 2)(\alpha - 1), \quad \alpha > 2.$

Modul : $Mo = \alpha/x_0.$

Mediana : $Me = 2^{1/\alpha} x_0.$

Funcția de repartiție : $F_X(x) = P(X \leqslant x) = 1 - \left(\frac{x_0}{x}\right)^{\alpha}.$

L.P. a fost propusă inițial de *V. Pareto* (1848—1923) sub forma :

$$N = A/x^{\alpha}, \quad 0 < x_0 < x < \infty$$

pentru a descrie repartiția venitului unei populații, N fiind numărul de persoane avînd venit $\geqslant x_0$: A, α—parametri (fig. 26).

Observații

1. *L.P.* este un caz particular al *legii de tip VI Pearson* *.
Se mai întîlnește sub numele de *legea de tip I Pareto.*
2. Alte forme ale *l.P.* :
— *legea de tip II Pareto* (*l. Lomax*). Repartiția unei v.a. X avînd f.r.

$$F_X(x) = 1 - \frac{K_1}{(x + C)^{\alpha}}$$

(caz special al *legii de tip VI Pearson*)*.
— *legea de tip III Pareto.* Repartiția unei v.a. X avînd f.r.

$$F_X(x) = 1 - \frac{K_{2x} e^{-bx}}{(x + C)^{\alpha}}.$$

LEGI DE REPARTIȚIE MULTIDIMENSIONALE (*n-dimensionale*). Repartiția vectorului aleator $\mathbf{X} = (X_1, \ldots, X_n)'$ avînd componentele v.a. X_i, $i = 1, \ldots, n$.

A. *Legi de repartiție multidimensionale continue*

Legea normală multidimensională (*n*-dimensională). Repartiția vectorului aleator $\mathbf{X}$ cu d.p.

$$f(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})'\Sigma^{-1}(\mathbf{x}-\mathbf{m})}, \quad \mathbf{x} \in R^n,$$

unde $\mathbf{m}$ este un vector $n \times 1$, Σ este o matrice $n \times n$ pozitiv definită.
$\mathbf{X}$ se numește vector aleator normal n-dimensional și se notează
$\mathbf{X} \sim N_n(\mathbf{m}, \Sigma)$.
L.n.m. este extensia *legii normale* $N(m, \sigma^2)$ *Laplace-Gauss**, la cazul n-dimensional.

Caracteristici principale

Vectorul medie : $E(\mathbf{X}) = \mathbf{m} = \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix}.$

Matricea de covarianță: $\mathrm{Cov}\ (\mathbf{X}, \mathbf{X}') = \boldsymbol{\Sigma} = (\sigma_{ij})_{1 \leqslant i,\, j \leqslant n}$

cu

$$m_i = E(X_i)$$

$$\sigma_{ij} = \mathrm{Cov}\ (X_i, X_j) = \rho \sigma_i \sigma_j. \quad \rho = \mathrm{Cor}\ (X_i X_j),\ i \neq j$$

$$\sigma_{ii} = \sigma_i^2 = \mathrm{Var}\ X_i.$$

Dacă rang $\boldsymbol{\Sigma} < n$, atunci $\mathbf{X} \sim N_n(\mathbf{m}, \boldsymbol{\Sigma})$, degenerată.
Dacă rang $\boldsymbol{\Sigma} = n$, atunci $\mathbf{X} \sim N_n(\mathbf{m}, \boldsymbol{\Sigma})$ nedegenerată.
Rangul matricei de covarianță $\boldsymbol{\Sigma}$ se mai numește *rangul l.n.m.*
Observații
1. Dacă coeficientul de corelație $\rho = 0$, atunci $f(\mathbf{x})$ este produsul a n d.p. unidimensionale normale $N(m, \sigma^2)$ și cele n v.a. componente X_i sînt necorelate.
2. Repartițiile marginale ale v.a. componente X_i sînt legile normale $N(m_i, \sigma_i^2)$, $i = 1, \ldots, n$.
3. Dacă $\mathbf{X} \sim N_n(\mathbf{m}, \boldsymbol{\Sigma})$, $\mathbf{C}$ este o matrice $n \times n$ nesingulară, $\mathbf{d}$ este un vector $n \times 1$, atunci vectorul aleator $\mathbf{CX} + \mathbf{d} \sim N_n(\mathbf{Cm} + \mathbf{d}, \mathbf{C\Sigma C}')$.
4. Dacă vectorul aleator $\mathbf{X} = (X_1, \ldots, X_n)' \sim N_n(\mathbf{m}, \boldsymbol{\Sigma})$, atunci condiția necesară și suficientă ca vectorii aleatori $\mathbf{X}^{(1)} = (X_1, \ldots X_k)'$ și $\mathbf{X}^{(2)} = (X_{k+1} \ldots, \ldots, X_n)'$, $1 \leqslant k \leqslant n$ să fie independenți este ca $\boldsymbol{\Sigma}_{12} = \mathbf{0}$, adică matricea de covarianță să fie de forma:

$$\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{X}, \mathbf{X}') = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

unde $\boldsymbol{\Sigma}_{11} = \mathrm{Cov}(\mathbf{X}^{(1)}, \mathbf{X}^{(1)\prime})$; $\boldsymbol{\Sigma}_{22} = \mathrm{Cov}(\mathbf{X}^{(2)}, \mathbf{X}^{(2)\prime})$.

5. Repartiția condiționată a vectorului aleator $\mathbf{X}^{(1)}$ cu condiția ca componentele vectorului aleator $\mathbf{X}^{(2)}$ să fie fixe ($= \mathbf{x}^{(2)}$) este *l.n.m.* cu d.p.

$$f(\mathbf{x}^{(1)}/\mathbf{x}^{(2)}) = \frac{f(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})}{h(\mathbf{x}^{(2)})}$$

unde $f(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ este d.p. comună a vectorilor $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}$; $h(\mathbf{x}^{(2)})$ este d.p. a vectorului $\mathbf{X}^{(2)}$ în punctul $\mathbf{x}^{(2)}$ adică:

$$f(\mathbf{x}^{(1)}/\mathbf{x}^{(2)}) = \frac{1}{(2\pi)^{\frac{k}{2}} [\det(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})]^{1/2}} \times$$

$$\times e^{-\frac{1}{2}[(\mathbf{x}^{(1)} - \mathbf{m}_1) - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}^{(2)} - \mathbf{m}_2)]'(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})^{-1}[(\mathbf{x}^{(1)} - \mathbf{m}_1) - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}^{(2)} - \mathbf{m}_2)]}.$$

6. Dacă vectorii aleatori $\mathbf{X}_i \sim N_n\ (\mathbf{m}_i, \boldsymbol{\Sigma}_i)$ independenți, atunci vectorul aleator

$$\sum_{i=1}^{N} \beta_i \mathbf{X}_i \sim N_n \left(\sum_{i=1}^{N} \mathbf{m}_i, \sum_{i=1}^{N} \beta_i^2 \boldsymbol{\Sigma}_i \right),\ i = 1, \ldots, N.$$

7. Forma pătratică

$$(\mathbf{x} - \mathbf{m})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \mathbf{m}) = C = \mathrm{const.}$$

este ecuația unui elipsoid în spațiul n-dimensional. Cînd C variază, se obține o familie de elipsoizi cu centrul în punctul $\mathbf{m}$.

Legea normală bidimensională. Caz particular al *legii normale multidimensionale** pentru $n = 2$.

Caracteristici principale

$$m_1 = E(X_1)$$

$$m_2 = E(X_2)$$

$$\sigma_1^2 = \text{Var } X_1$$

$$\sigma_2^2 = \text{Var } X_2$$

$$\rho = \text{Cor}(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var } X_1 \cdot \text{Var } X_2}}, \quad -1 \leqslant \rho \leqslant 1.$$

Matricea de covarianță a vectorului aleator $\mathbf{X} = (X_1, X_2)'$ devine :

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

avînd $\det \mathbf{\Sigma} = |\mathbf{\Sigma}| = \sigma_1^2\sigma_2^2(1-\rho^2)$; $m_1, m_2, \sigma_1, \sigma_2, \rho$ sînt parametrii *l.n.b.* D.p. a *l.n.b.* este de forma (fig. 107) :

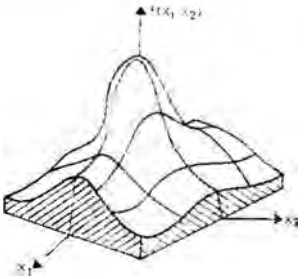


Fig. 107 Graficul densității de probabilitate a legii normale bidimensionale

$$f(\mathbf{x}) = f(x_1, x_2) = \frac{1}{2\pi \cdot \sigma_1\sigma_2 (1-\rho^2)^{1/2}} \cdot e^{-1/2 \cdot Q(x_1, x_2)}$$

unde forma pătratică pozitiv definită este :

$$Q(x_1, x_2) = \frac{1}{(1-\rho^2)} \left[\left(\frac{x_1 - m_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_1 - m_1}{\sigma_1} \right) \left(\frac{x_2 - m_2}{\sigma_2} \right) + \left(\frac{x_2 - m_2}{\sigma_2} \right)^2 \right].$$

Observații

1. Repartițiile marginale ale v.a. X_1 și X_2 sînt legile normale $N(m_1, \sigma_1^2)$, respectiv $N(m_2, \sigma_2^2)$ (fig. 107).

2. Repartițiile condiționate ale v.a. X_2 condiționată de valoarea x_1 a v.a. X_1 respectiv ale v.a. X_1 condiționată de valoarea x_2 a v.a. X_2, sînt *legi normale**

avînd mediile

$$m_{2/1} = E(X_2/x_1) = m_2 + \rho \frac{\sigma_2}{\sigma_1}(x_1 - m_1)$$

$$m_{1/2} = E(X_1/x_2) = m_1 + \rho \frac{\sigma_1}{\sigma_2}(x_2 - m_2)$$

și dispersiile:

$$\mathrm{Var}(X_2/x_1) = \sigma_2^2(1 - \rho^2)$$

$$\mathrm{Var}(X_1/x_2) = \sigma_1^2(1 - \rho^2).$$

3. Curbele $Q(x_1, x_2) = \text{const.}$ sînt elipsele de probabilitate, în fiecare punct al unei astfel de elipse d.p. fiind constantă.
Elipsa mediană numită și *elipsa probabilă* este acea elipsă pentru care probabilitatea totală asociată perechilor (x_1, x_2) situate în interior este egală cu 1/2.
Elipsa de concentrație este elipsa de ecuație $Q(x_1, x_2) = 4$, corespunzînd lui $\mathbf{X} = (X_1, X_2)'$ cu d.p. uniformă în toate punctele din R^2 și avînd aceiași parametri cu *l.n.b.*
4. Curbele de regresie a lui X_2 în x_1, respectiv a lui X_1 în x_2 sînt dreptele trecînd prin punctul $(x_1 = m_1, x_2 = m_2)$ și se confundă pentru $\rho = \pm 1$.
Dacă $\rho = \pm 1$, atunci punctele (x_1, x_2) sînt concentrate pe dreapta

$$\frac{x_2 - m_2}{\sigma_2} = \frac{x_1 - m_1}{\sigma_1} \quad \text{pentru } \rho = +1,$$

respectiv pe dreapta

$$\frac{x_2 - m_2}{\sigma_2} = -\frac{x_1 - m_1}{\sigma_1} \quad \text{pentru } \rho = -1.$$

5. *L.n.b. standardizată (normată)* este repartiția vectorului aleator (U, V) unde U, V sînt v.a. normale standardizate (normate):

$$U = \frac{X_1 - m_1}{\sigma_1}, \quad V = \frac{X_2 - m_2}{\sigma_2}.$$

D.p. a *l.n.b. standardizate:*

$$f(u, v) = \frac{1}{2\pi\sqrt{1 - \rho^2}} \cdot e^{-\frac{1}{2}\frac{1}{1-\rho^2}(u^2 - 2\rho uv + v^2)}.$$

Legea Beta multidimensională *(Repartiția Dirichlet).* Repartiția vectorului aleator $\mathbf{X} = (X_1, X_2, \ldots, X_s)'$ avînd d.p.:

$$f(x_1, x_2, \ldots, x_s) =$$

$$= \begin{cases} \dfrac{\Gamma(\mu_1 + \mu_2 + \ldots + \mu_{s+1})}{\Gamma(\mu_1)\Gamma(\mu_2)\ldots\Gamma(\mu_{s+1})} x_1^{\mu_1 - 1} \ldots x_s^{\mu_s - 1}(1 - x_1 \ldots x_s)^{\mu_{s+1} - 1} \\ \text{dacă} \\ \mathbf{x} \in V_s = \left\{(x_1, \ldots, x_s),\ x_k \geqslant 0,\ k = 1, 2, \ldots, s;\ \sum_{k=1}^{s} x_k \leqslant 1\right\} \\ 0 \qquad \text{dacă} \qquad \mathbf{x} \notin V_s \end{cases}$$

$$(\mu_1 > 0, \mu_2 > 0, \ldots, \mu_s > 0,\ \mu_{s+1} > 0).$$

Se spune că vectorul aleator $\mathbf{X} = (X_1, \ldots, X_s)'$ are *l.D.* cu s dimensiuni de parametri $\mu_1, \ldots, \mu_s : \mu_{s-1}$ și se notează
$\mathbf{X} \sim D(\mu_1, \ldots, \mu_s ; \mu_{s+1})$.
Caracteristici principale
Momentele *l.D.* $D(\mu_1, \ldots, \mu_s : \mu_{s+1})$:

$$M_{r_1 r_2 \ldots r_s} = \frac{\Gamma(\mu_1 + r_1)\Gamma(\mu_2 + r_2) \ldots \Gamma(\mu_s + r_s)\,\Gamma(\mu_1 + \mu_2 + \ldots + \mu_{s+1})}{\Gamma(\mu_1)\Gamma(\mu_2)\ldots\Gamma(\mu_s)\Gamma(\mu_1 + \mu_2 + \ldots + \mu_{s+1} + r_1 + r_2 + \ldots + r_s)}$$

ținînd seama de integrala Dirichlet :

$$\int_{V_s} x_1^{\mu_1 - 1} x_2^{\mu_2 - 1} \ldots x_s^{\mu_s - 1}(1 - x_1 - \ldots - x_s)^{\mu_{s+1} - 1}\, dx_1 \ldots dx_s =$$

$$= \frac{\Gamma(\mu_1)\Gamma(\mu_2) \ldots \Gamma(\mu_{s+1})}{\Gamma(\mu_1 + \mu_2 + \ldots + \mu_{s+1})}$$

De aici se deduc mediile și dispersiile v.a. componente X_i ale vectorului aleator $\mathbf{X} = (X_1, \ldots, X_s)'$. De exemplu :

$$M_{10\ldots0} = E(X_1) = \frac{\mu_1}{\mu_1 + \mu_2 + \ldots + \mu_{s+1}}$$

$$M_{010\ldots0} = E(X_2) = \frac{\mu_2}{\mu_1 + \mu_2 + \ldots + \mu_{s+1}}, \ldots$$

$$M_{20\ldots0} = E(X_1^2) = \frac{\mu_1(\mu_1 + 1)}{(\mu_1 + \ldots + \mu_{s-1})(\mu_1 + \ldots + \mu_{s+1} + 1)}$$

$$M_{020\ldots0} = E(X_2^2) = \frac{\mu_2(\mu_2 + 1)}{(\mu_1 + \ldots + \mu_{s+1})(\mu_1 + \ldots + \mu_{s-1} + 1)}, \ldots$$

$$\sigma_1^2 = \operatorname{Var} X_1 = \frac{\mu_1(\mu_1 + \ldots + \mu_{s+1} - \mu_1)}{(\mu_1 + \ldots + \mu_{s+1})^2 (\mu_1 + \ldots \mu_{s+1} + 1)}$$

$$\sigma_2^2 = \operatorname{Var} X_2 = \frac{\mu_2(\mu_1 + \ldots + \mu_{s-1} - \mu_2)}{(\mu_1 + \ldots + \mu_{s+1})^2(\mu_1 + \ldots + \mu_{s+1} + 1)}, \ldots$$

$$\operatorname{Cov}(X_1, X_2) = \frac{-\mu_1\mu_2}{(\mu_1 + \ldots + \mu_{s-1})^2(\mu_1 + \ldots + \mu_{s-1} + 1)}$$

$$\operatorname{Cov}(X_1, X_3) = \frac{-\mu_1\mu_3}{(\mu_1 + \ldots + \mu_{s-1})^2(\mu_1 + \ldots + \mu_{s-1} + 1)}, \ldots$$

Observații
1. Dacă componentele $X_1, X_2, \ldots, X_{s-1}$ ale vectorului aleator $\mathbf{X} = (X_1, X_2, \ldots, X_{s+1})'$ sînt v.a. independente și au, respectiv, *legea Gamma* * de parametri $\mu_1, \ldots, \mu_{s-2}$, atunci vectorul aleator $\mathbf{U} = (U_1, U_2, \ldots, U_s)'$ cu

$$U_k = X_k/(X_1 + \ldots + X_{s+1}), \qquad k = 1, \ldots, s + 1$$

are *l.D.* $\sim D(\mu_1, \mu_2, \ldots, \mu_s ; \mu_{s+1})$.

2. Dacă $\mathbf{X} = (X_1, \ldots, X_s)'$ un vector aleator $\sim$ *l.D.* și anume : $D(\mu_1, \mu_2, \ldots, \mu_s; \mu_{s+1})$, atunci v.a. $(X_1 + X_2 + \ldots + X_s) \sim$ *legea Beta* *. $Be(\mu_1 + \mu_2 + \ldots + \mu_s, \mu_{s+1})$.

Legea T^2 Hotelling. Generalizare a *legii t Student** pentru cazul multidimensional.

Repartiția formei pătratice :

$$T^2 = \mathbf{Y}'\mathbf{S}^{-1}\mathbf{Y}$$

$$(T^2 \sim T^2_n \text{ sau } T^2 \sim T^2_{p,n}),$$

unde vectorul aleator $\mathbf{Y} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$. $\mathbf{\Sigma}$ nesingulară. Matricea aleatoare $n\mathbf{S} \sim Wi\ (\mathbf{\Sigma}, n)$ independentă de $\mathbf{Y}$ și care se poate pune sub forma :

$$n\mathbf{S} = \sum_{i=1}^{n} \mathbf{z}_i \mathbf{z}_i'$$

cu $\mathbf{z}_i$, $1 \leqslant i \leqslant n$ vectori aleatori independenți $\sim N_p(\mathbf{0}, \mathbf{\Sigma})$.

Dacă $\mathbf{Y} \sim N_p(\mathbf{m}, \mathbf{\Sigma})$, atunci v.a. T^2 are *l.T² H. necentrată* cu n g.l. și parametrul de necentricitate :

$$\delta^2 = \mathbf{m}'\mathbf{\Sigma}^{-1}\mathbf{m}.$$

Observații

1. Dacă v.a. $\mathbf{T}^2 \sim$ *l.T² H.* necentrată cu n g.l., atunci v.a.

$$F = \frac{n - p + 1}{np} T^2 \sim F_{p, n-p+1; \delta}.$$

2. Între cuantilele *l. T². H.* și *legii F** centrate, respectiv necentrate, există relațiile :

$$\frac{n - p + 1}{np} T^2_{\alpha; p, n} = F_{\alpha; p, n-p+1},$$

respectiv

$$\frac{n - p + 1}{pn} T^2_{\alpha; p, n; \delta} = F_{\alpha; p, n-p+1; \delta}.$$

3. Fie un eșantion corespunzător vectorului aleator p-dimensional N_p $(\mathbf{m}, \mathbf{\Sigma})$ format din N vectori-observație independenți : $\mathbf{Y}_1, \ldots, \mathbf{Y}_N$.

Forma pătratică : $N(\overline{\mathbf{Y}} - \mathbf{m})' \mathbf{S}^{-1}(\overline{\mathbf{Y}} - \mathbf{m}) \sim T^2_{p, N-1}$ are repartiția *l. T² H.* cu $(N - 1)$ g.l, unde $\overline{\mathbf{Y}}$, $\mathbf{S}$ sînt *m. l.-estimatorii* * vectorului medie $\mathbf{m}$, respectiv matricei de covariantă $\mathbf{\Sigma}$.

4. *L. T² H.* se folosește pentru verificarea ipotezelor referitoare la media legii $N_p(\mathbf{m}, \mathbf{\Sigma})$: $H_0 : \mathbf{m} = \mathbf{m}_0$ cînd $\mathbf{\Sigma}$ este necunoscută.

Legea t multidimensională (O. J. Dunn). Repartiția vectorului aleator $\mathbf{t} = (t_1, \ldots, t_p)'$, unde

$$t_i = \frac{x_i - \mu_i}{s_i}, \quad i = 1, \ldots, p$$

sînt p v.a. avînd *legea t a lui Student* *, adică $\mathbf{x} = (x_1, \ldots, x_p)' \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ este un vector aleator avînd *legea normală p-dimensională** cu vectorul medie $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ și matricea de covarianță $\boldsymbol{\Sigma} = (\sigma_{ij})_{1 \leq i, j \leq p}$; s_i^2 este un estimator pătratic independent al lui σ_{ii} cu ν_i g.l. și care este independent de $x_1, \ldots, x_p$, adică v.a. $\nu_i s_i^2 / \sigma_{ii} \sim \chi^2_{\nu_i}$.

Repartițiile componentelor t_i ale vectorului aleator $\mathbf{t}$ sînt *repartițiile t a lui Student** cu ν_i g.l., respectiv.

Alte generalizări ale legii t la cazul multidimensional au fost date de C. W. Dunett și M. Sobel, în 1955, pentru cazul cînd v.a. $\{x_i\}$ au dispersii egale.

În cazul cînd matricea de covarianță este de forma $(\sigma^2 \rho_{ij})$, unde $\rho_{ii} = 1$, $i = 1, \ldots, p$ și v.a. $\nu s^2/\sigma^2 \sim \chi^2_\nu$ și independentă de $(x_1, \ldots, x_p)$, atunci v.a. $t_i = (x_i - \mu_i)/s \sim t$ *Student*, $i = 1, \ldots, p$, iar vectorul aleator $\mathbf{t} = (t_1, \ldots, t_p)'$ **are d.p.**

$$f(t_1, \ldots, t_p) = \frac{\Gamma\left(\frac{p+\nu}{2}\right) |\rho^{ij}|^{1/2}}{\nu^{p/2} \Gamma(\nu/2) \pi^{p/2}} \left[1 + \frac{1}{\nu} \sum_{i=1}^{p} \sum_{j=1}^{p} \rho^{ij} t_i t_j\right]^{-\frac{p+\nu}{2}},$$

unde (ρ^{ij}) este inversa matricei de covarianță $(\sigma^2 \rho_{ij})$, iar $|\rho^{ij}|$ este determinantul matricii (ρ^{ij}).

Cuantilele acestei legi au fost tabelate de O. J. Dunn în 1959.

Legea Wishart. Generalizarea *legii* χ^2 * Helmert - Pearson la cazul multidimensional. Repartiția unei matrici aleatoare simetrice, pozitiv definite $\mathbf{A}$ de forme pătratice și forme biliniare care poate fi transformată în suma :

$$\sum_{i=1}^{n} \mathbf{y}_i \mathbf{y}_i' = (y_{ij})_{1 \leq i, j \leq p}, \quad n = N - 1,$$

unde vectorii independenți p-dimensionali $\mathbf{y}_i \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$.

D.p. a matricei aleatoare $\mathbf{A}$ este :

$$w(\mathbf{A}; \boldsymbol{\Sigma}, n) = \begin{cases} \dfrac{|\mathbf{A}|^{\frac{n-p-1}{2}} e^{-\frac{1}{2} tr \mathbf{A}\boldsymbol{\Sigma}^{-1}}}{2^{np/2} \pi^{\frac{p(p-1)}{4}} |\boldsymbol{\Sigma}|^{n/2} \prod_{i=1}^{p} \Gamma\left(\frac{n-i+1}{2}\right)}, & \mathbf{A} \text{ pozitiv definită} \\ 0, \text{ în rest} \end{cases}$$

unde $tr \mathbf{A}\boldsymbol{\Sigma}^{-1} = \sum_{i=1}^{p} \sum_{j=1}^{p} y_{ij} \sigma^{ji}$, norma matricei $\mathbf{A}\boldsymbol{\Sigma}^{-1}$, cu

$$\boldsymbol{\Sigma} = (\sigma_{ij})_{1 \leq i, j \leq p} \text{ și } \boldsymbol{\Sigma}^{-1} = (\sigma^{ji})_{1 \leq i, j \leq p}.$$

Se spune că $w(\mathbf{A}; \boldsymbol{\Sigma}, n)$ este d.p. a *l. Wi.* de parametru $\boldsymbol{\Sigma}$ și cu n g.l.

L. Wi. a fost considerată de J. Wishart în 1928.

Observații

1. Pentru $p = 1$, $\boldsymbol{\Sigma} = 1$, se obține d.p. a *legii* χ^2 *.

2. Dacă $\mathbf{A}_1, \ldots, \mathbf{A}_k$ sînt matrici repartizate după *l. Wi.*, independente, cu același parametru — matricea $\mathbf{\Sigma}$ și cu $n_1, \ldots, n_k$ g.l., respectiv, atunci suma lor, $\mathbf{A}_1 + \ldots + \mathbf{A}_k$, este o matrice Wishart de parametru $\mathbf{\Sigma}$ și $n = n_1 + \ldots + n_k$ g.l. (proprietatea de stabilitate a *l.Wi.*).
3. Dacă se consideră matricea

$$\mathbf{A} = \sum_{h=1}^{N} (\mathbf{X}_h - \overline{\mathbf{X}})(\mathbf{X}_h - \overline{\mathbf{X}})' = \sum_{h=1}^{N} \mathbf{X}_h \mathbf{X}_h' - N\overline{\mathbf{X}}\overline{\mathbf{X}}'$$

(suma pătratelor și produsele mixte ale observațiilor), unde :
$\mathbf{X}_i$, $i = 1, \ldots, N$ — vectori aleatori independenți p-dimensionali $\mathbf{X}_i \sim N_p(\mathbf{m}, \mathbf{\Sigma})$.:
$\overline{\mathbf{X}} = \frac{1}{N} \sum_{h=1}^{N} \mathbf{X}_h$, vectorul mediilor de selecție ale observațiilor, atunci repartiția matricei $\mathbf{A}$ este legea $Wi\,(\mathbf{\Sigma}, n)$ cu $n = N-1$.
4. Fie observațiile $\{x\}$ extrase din k grupe independente sub forma a k vectori p-dimensionali corespunzători fiecărei grupe, avînd, respectiv, vectorul medie $\mathbf{m}_h$, $h = 1, \ldots, k$ și matricea de covariață $\mathbf{\Sigma}$, comună tuturor grupelor: Matricea aleatoare

$$\mathbf{S} = \frac{1}{N-k} \sum_{h=1}^{k} \mathbf{A}_h, \; N = N_1 + \ldots + N_k$$

este o estimație nedeplasată a lui $\mathbf{\Sigma}$, unde :

$$\mathbf{A}_h = \sum_{i=1}^{N_h} (\mathbf{X}_{ih} - \overline{\mathbf{X}}_h)(\mathbf{X}_{ih} - \overline{\mathbf{X}}_h)',$$

(matricea sumelor de pătrate și produselor mixte ale observațiilor din grupa h, $h = 1, \ldots, k$) ;
$\mathbf{X}_{ih}' = (x_{i1h}, \ldots, x_{ipk})$ — a i-a linie a matricei $\mathbf{X}_h$ a observațiilor din grupa h, $h = 1, \ldots, k$;
$\overline{\mathbf{X}}_h$ — vectorul medie de selecție din grupa h, adică *m.l.-estimatorul* * vectorului medie $\mathbf{m}_h$, $h = 1, \ldots, k$.
Repartiția matricei aleatoare $\mathbf{S}$ este *l. Wi* $(\mathbf{\Sigma}, n)$, $n = N-k$.
5. Matricea aleatoare $N\hat{\mathbf{\Sigma}} \sim Wi\,(\mathbf{\Sigma}, n)$ cu $n = N-1$, unde :
$\hat{\mathbf{\Sigma}} = \frac{1}{N} \mathbf{A}$ este *m.l. - estimația** matricei de covariață $\mathbf{\Sigma}$, și anume $\hat{\mathbf{\Sigma}}$ este *estimatorul deplasat** al lui $\mathbf{\Sigma}$.
6. Matricea aleatoare $(N-1)\,\mathbf{S} \sim Wi\,(\mathbf{\Sigma}, n)$, $n = N-1$, unde

$$\mathbf{S} = \frac{1}{N-1} \mathbf{A}$$

este *estimația nedeplasată** a lui $\mathbf{\Sigma}$ și se numește *matricea de covariață de selecție*.
Legea Gram-Charlier, Edgeworth generalizată. Dezvoltările Gram-Charlier sau Edgeworth generalizate se obțin, respectiv, ca extensii ale *legilor Gram-*

*Charlier** sau Edgeworth la cazul multidimensional și au apărut între 1910—1920. În 1930, S. J. Pretorius, în „Biometrika" 30 (1930), p. 109—223, a publicat o sinteză a rezultatelor obținute. În 1968, I. M. Mihăilă, în „Rev. Roum. Math. Pures et Appl." (1968), p. 803—813, a dat o dezvoltare Gram-Charlier a d.p. a unei v.a. tridimensionale în funcție de $f_{N_3(\mathbf{0},\mathbf{R})}$—d.p. standardizată a vectorului aleator tridimensional $N_3(\mathbf{0}, \mathbf{R})$ cu vectorul medie $\mathbf{0}$ și matrice de corelație $\mathbf{R}$, folosind polinoame Hermite tridimensionale:

$$p(x_1, x_2, x_3) = \sum_{j_1=0}^{\infty}\sum_{j_2=0}^{\infty}\sum_{j_3=0}^{\infty} C_{j_1j_2j_3}\frac{\partial^{j_1+j_2+j_3}}{\partial x_1^{j_1}\partial x_2^{j_2}\partial x_3^{j_3}} f_{N_3(\mathbf{0},\mathbf{R})}(\mathbf{x};\ \mathbf{0},\ \mathbf{R}) =$$

$$= \sum_{j_1}\sum_{j_2}\sum_{j_3} C_{j_1j_2j_3}(-1)^{j_1+j_2+j_3}H_{j_1j_2j_3}(\mathbf{x})f_{N_3(\mathbf{0},\mathbf{R})}(\mathbf{x};\ \mathbf{0},\mathbf{R}),$$

unde

$H_{000} = 1$; $H_{100} = x_1$; $H_{200} = x_1^2 - a_{11}$; $H_{110} = x_1x_2 - a_{12}$;

$H_{300} = x_1^3 - 3a_{11}x_1$; $H_{310} = x_1^2x_2 - 2a_{12}x_1 - a_{11}x_2$;

$H_{111} = x_1x_2x_3 - a_{23}x_2 - a_{12}x_3$;

$H_{400} = x_1^4 - 6a_{11}x_1^2 + 3a_{11}^2$; $H_{310} = x_1^3x_2 - 3a_{12}x_1^2 - 3a_{11}x_1x_2 + 3a_{11}a_{12}$;

$H_{220} = x_1^2x_2^2 - a_{22}x_1^2 - a_{11}x_2^2 - 4a_{12}x_1x_2 + a_{11}a_{22} + 2a_{12}^2$;

$H_{211} = x_1^2x_2x_3 - a_{23}x_1^2 - 2a_{13}x_1x_2 - 2a_{12}x_1x_3 - a_{11}x_2x_3 + a_{11}a_{23} + 2a_{12}a_{13}$ etc.

și coeficienții $C_{j_1j_2j_3}$ dați de formulele:

$$C_{000} = -\frac{1}{6}\mu_{300};\quad C_{210} = -1/2\mu_{210};\quad C_{111} = \mu_{111};$$

$$C_{400} = \frac{1}{24}(\mu_{400} - 3);\quad C_{310} = 1/6(\mu_{310} - 3\mu_{110});$$

$$C_{220} = \frac{1}{4}(\mu_{220} - \mu_{200} - \mu_{020} - 4\rho_{12}\mu_{110} - 1 + 2\rho_{12}^2);$$

$$C_{211} = \frac{1}{2}(\mu_{211} - \rho_{23}\mu_{200} - 2\rho_{13}\mu_{110} - 2\rho_{12}\mu_{101} - \mu_{011} + \rho_{23} + \rho_{12}\rho_{13})$$

cu $((a_{ij}))_{1\leq i,j\leq 3} = \mathbf{A} = \mathbf{R}^{-1}$

$$\mu_{r_1r_2r_3} = E\left(\prod_{j=1}^{3}(X_j - E(X_j))^{r_j}\right)$$

$$\rho_{r_ir_j} = \mathrm{Cor}(X_i, X_j),\quad 1 \leq i, j \leq 3.$$

Pentru cazul bidimensional, dezvoltarea Gram-Charlier se obține din cea anterioară luînd:

$$H_{r_1r_2} = H_{r_1,r_2,0};\quad C_{r_1r_2} = C_{r_1,r_2,0}\ \text{cu}\ \mu_{r_1r_20} = \mu_{r_1r_2};$$

$((a_{ij}))_{1\leq i,j\leq 2} = \mathbf{R}^{-1}$ cu $a_{11} = a_{22} = (1 - \rho_{12}^2)^{-1}$; $a_{12} = -\rho_{12}(1 - \rho_{12}^2)^{-1}$.

B. *Legile de repartiție multidimensionale discrete*

Legea multinomială. Repartiția vectorului aleator $\mathbf{X} = (X_1, X_2, \ldots, X_k)'$ cu componentele X_i ($i = 1, \ldots, k$) v.a. $Bi(n, p_i)$, avînd d.p. :

$$P(X_1 = x_1, \ldots, X_k = x_k) = p_{n;x_1,\ldots,x_k} = \frac{n!}{x_1!x_2!\ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k},$$

unde :

$$\sum_{i=1}^{k} p_i = 1; \quad x_1 + x_2 + \ldots + x_k = n, \quad x_i \geqslant 0, \ i = 1, \ldots, k.$$

V.a. X_i ($i = 1, \ldots, k$) reprezintă numărul de realizări „succese" ale evenimentelor A_i ($i = 1, \ldots, k$), respectiv, în n probe independente corespunzînd numărului de bile de culoare i ($i = 1, \ldots, k$) dintr-un eșantion de n bile obținut ca rezultat al unor *extrageri bernouillene (sondaje cu revenire* *)* dintr-o urnă conținînd N bile de k culori, dintre care $a_1 = Np_1$ bile de culoarea 1, $a_2 = Np_2$ bile de culoarea 2, ... $a_k = Np_k$ bile de culoarea k, unde

$$p_i = \frac{a_i}{\sum_{i=1}^{k} a_i} = \frac{a_i}{N}.$$

L.m. se folosește în multe aplicații ale statisticii matematice, în principiu în aceleași condiții în care se aplică *legea binomială* *, însă pentru situații mai generale cînd se consideră mai multe categorii de evenimente în locul unei simple dihotomii („da", „nu" : „realizat", „nerealizat" etc.)

L.m. își găsește aplicații în teoria cinetică a particulelor din fizică, în termodinamica statistică.

L.m. se aplică în probleme de estimare pentru populații, cînd datele de observație sînt grupate într-un număr finit de grupe, presupunînd observațiile independente.

L.m. se folosește în analiza *tabelelor de contingență* *, adică a tabelelor de clasificare după două, trei sau mai multe criterii a unor factori cu mai multe nivele.

Repartițiile componentelor X_i ($i = 1, \ldots, k$) sînt *legi binomiale* $Bi(n, p_i)$ avînd *caracteristicile* :

Media : $E(X_i) = np_i$.

Dispersia : $\mathrm{Var}\, X_i = np_i(1 - p_i)$, $i = 1, \ldots, k$.

Repartițiile condiționate ale lui X_j pentru $X_i = x_i$ sînt *legi binomiale** avînd :

$$E(X_j|x_i) = (n - x_i)\frac{p_j}{1 - p_i}$$

$$\mathrm{Var}\,(X_j|x_i) = (n - x_i)\frac{p_j}{1 - p_i}\left(1 - \frac{p_j}{1 - p_i}\right).$$

Covarianța a două v.a. X_i, X_j:

$$\operatorname{Cov}(X_i, X_j) = -np_ip_j, \quad i \neq j, \quad i, j = 1, \ldots, k.$$

Funcția de repartiție a vectorului aleator X:

$$F(n; u_1, \ldots, u_k) = P(X_1 \leqslant u_1, X_2 \leqslant u_2, \ldots, X_k \leqslant u_k) =$$

$$= \sum_{x_1=0}^{[u_1]} \sum_{x_2=0}^{[u_2]} \cdots \sum_{x_k=0}^{[u_k]} P_{n; x_1, x_2, \ldots, x_k}.$$

Funcția caracteristică a vectorului aleator X:

$$\varphi(t_1, \ldots, t_k; n) = \sum_{x_1+\ldots+x_k=n} p_{n; x_1, \ldots, x_k} e^{i(t_1x_1+\ldots+t_kx_k)} =$$

$$= (p_1 e^{it_1} + \ldots + p_k e^{it_k})^n.$$

Legea hipergeometrică generalizată. Repartiția vectorului aleator $\mathbf{X} = (X_1, \ldots, X_k)'$ având d.p.

$$p_{\mathbf{X}} = P(X_1 = x_1, \ldots, X_k = x_k) = \frac{C_{Np_1}^{x_1} \cdots C_{Np_k}^{x_k}}{C_N^n} p_1^{x_1} \cdots p_k^{x_k}$$

($N_1 = Np_1, \ldots, N_k = Np_k$, $x_1 + \ldots + x_k = n$; $N_1 + \ldots + N_k = N$), adică este repartiția vectorului aleator X având componentele v.a. X_i ($i = 1, \ldots, k$) care apar în experimentul descris la legea *multinomială**, în cazul de față efectuînd *extrageri (selecții) bernoulliene (sondaje fără revenire)**

Caracteristicile principale

Media: $E(X_i) = np_i$.

Dispersia: $\operatorname{Var} X_i = np_i(1 - p_i) \dfrac{N - n}{N - 1}$.

Covarianța: $\operatorname{Cov}(X_i, X_j) = np_i \cdot p_j \dfrac{N - n}{N - 1}$.

Funcția de repartiție:

$$F(x_1, \ldots, x_{k-1}) = P(X_1 \leqslant x_1, \ldots, X_k \leqslant x_k) =$$

$$= \sum_{\alpha_1=0}^{[x_1]} \sum_{\alpha_2=0}^{[x_2]} \cdots \sum_{\alpha_{k-1}=0}^{[x_{k-1}]} \frac{C_{N_1}^{x_1} C_{N_2}^{x_2} \cdots C_{N_k}^{x_k}}{C_N^n},$$

unde

$$\sum_{i=1}^{k-1} x_i = \max\left(0, n - \left(N - \sum_{i=1}^{k-1} N_i\right)\right); \quad [x_i],\ 1 \leqslant i \leqslant k-1$$

reprezintă cei mai mari întregi cuprinși în x_i, $1 \leqslant i \leqslant k-1$, îndeplinind condiția

$$\sum_{i=1}^{k-1} [x_i] = \min\left(n, \sum_{i=1}^{k-1} N_i\right).$$

Pentru $k = 2$ se obține *legea hipergeometrică**.

LEGEA POISSON MULTIDIMENSIONALĂ. Se obține ca un caz limită al *legii multinomiale** pentru $p_i \to 0$ $(i = 1, \ldots, k-1)$ $p_k \to 1$ și $n \to \infty$ astfel încît $np_i = \lambda_i = \text{const}$ $(i = 1, \ldots, k-1)$ și având d.p.:

$$p_{\mathbf{X}} = \frac{\lambda_1^{x_1} \lambda_2^{x_2} \ldots \lambda_{k-1}^{x_{k-1}}}{x_1!\ x_2! \ldots x_{k-1}!} e^{-\sum_{i=1}^{k-1} \lambda_i}$$

și f.r.

$$F(x_1, x_2, \ldots, x_{k-1}) = P(X_1 \leqslant x_1, \ldots, X_{k-1} \leqslant x_{k-1}) =$$

$$= \sum_{\alpha_1=0}^{[x_1]} \sum_{\alpha_2=0}^{[x_2]} \ldots \sum_{\alpha_{k-1}=0}^{[x_{k-1}]} \frac{\lambda_1^{\alpha_1} \lambda_2^{\alpha_2} \ldots \lambda_{k-1}^{\alpha_{k-1}}}{\alpha_1!\ \alpha_2! \ldots \alpha_{k-1}!} e^{-\sum_{i=1}^{k-1} \lambda_i}.$$

L.P.m. este repartiția unui vector aleator cu $(k-1)$ componente, fiecare fiind v.a. Poisson.

În particular, pentru $k = 2$ se obține legea Poisson bidimensională a cărei d.p. a fost obținută de J. T. Campbell (1934) și ulterior de P. Holgate (1964) sub forma:

$$p_{\mathbf{X}}(x_1, x_2) = e^{-(\xi + \lambda_1 + \lambda_2)} \sum_{j=0}^{\min(x_1, x_2)} \frac{\xi^j}{j!} \frac{\lambda_1^{x_1 - j}}{(x_1 - j)!} \frac{\lambda_2^{x_2 - j}}{(x_2 - j)!}$$

Repartițiile marginale ale v.a. X_1, X_2 sînt legi Poisson de medii $\xi + \lambda_1$, $\xi + \lambda_2$, respectiv.

L.P.m. a fost studiată de G. P. Patil și S. Bildikar (1967).

Expresiile momentelor *l.P.m.* au fost obținute de D. P. Banerjee (1959) și M. Sibuya ș.a. (1964).

XIV. TEORIA SONDAJULUI STATISTIC (TEORIA SELECȚIEI)

În acest capitol este inclus ansamblul de metode statistice care permit folosirea rezultatelor unor experiențe efectuate asupra unei părți a *populației statistice* considerate, pentru a trage concluzii asupra unor caracteristici ale întregii populații.

EXPERIMENT. Procedeu (procedură) organizat care poate fi repetat (ori de cite ori) în aceleași condiții și în urma căruia se obțin rezultate ce pot fi observate, măsurate și apoi interpretate.

Experiment determinist. E. ale cărui rezultate sînt complet determinate de condițiile în care acesta se desfășoară.

De exemplu, încălzirea la 100° C a unui recipient cu apă are drept rezultat fierberea apei.

Experiment aleator. E. ale cărui rezultate nu pot fi prevăzute cu exactitate, știindu-se doar o clasă de realizări (rezultate) posibile ale acestuia.

De exemplu, aruncarea unui zar este un *e.a.* avînd drept rezultat una din cele 6 realizări posibile : fața 1, fața 2, ..., fața 6 a zarului.

EȘANTION *) (DE VOLUM n). Fie v.a. $X \sim f_X(x)$, unde $f_X(x)$ este d.p. a legii ei (de repartiție). Un **e.(v.n.)** din $f_X(x)$ constă din vectorul aleator n-dimensional $(X_1, \ldots, X_n)'$, unde cele n componente sînt v.a. independente X_i, $1 \leqslant i \leqslant n$, avînd aceeași lege (de repartiție) ca și v.a. $X \sim f_X(x)$.

Valorile observate ale v.a. $X_1, \ldots, X_n$ sînt $x_1, \ldots, x_n$ numite *date (de observație)* sau *variabile de selecție (variabilele eșantionului)*.

Un mod de a construi un **e.(v.n.)** este următorul : se consideră v.a. X_i, $1 \leqslant i \leqslant n$, definite pe cîmpul de evenimente $(R^n, \mathscr{B}^n)$ prin relațiile :

$$X_i(\mathbf{x}^n) = X(x_i)$$

cu $\mathbf{x}^n = (x_1, \ldots, x_n)'$.

$(R^n, \mathscr{B}^n)$ se probabilizează considerînd probabilitatea P^n indusă de $F^n = \overset{n}{\underset{i=1}{*}} F_i$, $F_i = F_X$, $1 \leqslant i \leqslant n$, (unde $*$ este operația de convoluție a funcțiilor de repartiție F_i, toate aceleași cu F_X, f.r. a v.a. X) pe $\mathscr{B}^n$.

V.a. $X_1, \ldots, X_n$ definite pe cîmpul de probabilitate $(R^n, \mathscr{B}^n, P^n)$ sînt **elementele e.(v.n.).**

Caracteristicile principale ale v.a. X (media, dispersia, mediana etc.) se determină pe baza datelor **e.(v.n.).** $x_1, \ldots, x_n$.

SPAȚIUL EȘANTIOANELOR. Concept fundamental în teoria probabilistă a sondajului. **S.e.** este totalitatea realizărilor (rezultatelor) posibile ale unui experiment aleator. (→ *Spațiu de evenimente elementare*). **S.e.** se notează cu Ω. **S.e.** se organizează ca spațiu euclidian n-dimensional cu coordonate reale : *Produsul scalar* al vectorilor $\mathbf{x} = (x_1, \ldots, x_n)'$, $\mathbf{y} = (y_1, \ldots, y_n)' \in R^n$ este

$$(\mathbf{x}, \mathbf{y}) = \mathbf{x}'\mathbf{y} = \mathbf{x}\mathbf{y}' = \sum_{i=1}^{n} x_i y_i$$

---

*) **2)** Se atrage atenția asupra folosirii greșite a termenului de *selecție* în loc de e.(v.n.). Selecția este operația de extragere a unei părți din populația statistică, rezultatul acestei operații fiind un e.(v.n.).

2) E.(v.n.) este termen sinonim cu *eșantion bernoullian de volum* n din $f_X(x)$ sau asupra v.a. X.

Norma lui $y \in R^n$ este

$$\|y\| = \sqrt{(y, y)} = \sqrt{\sum_{i=1}^{n} y_i^2}$$

Rezultat posibil. *(→ Eveniment elementar).* Un element $\omega = (x_1, \ldots, x_n)$ al s.e. Ω.

SELECȚIE. Operația de extragere a unei părți dintr-o populație statistică.

Selecție bernoulliană (selecție (extragere) cu revenire) sau sondaj cu revenire. Operația de extragere a unui eșantion de volum n dintr-o populație statistică. Operația de *s.b. (s. (e) r.)* revine la a admite că după ce s-a extras la întîmplare un element oarecare din populația considerată, acest element este introdus din nou în populație înaintea unei noi extrageri. Rezultatul operației de *s.b. (s. (e.) r.)* este un *eșantion bernoullian de volum* n.*

De exemplu, în cazul în care se consideră o populație finită. operația de *s.b. (s.(e.) r.)* conduce la repartiția bernoulliană *(→ legea binomială, Bi(n, p)*.*

Selecție nebernoulliană (selecție fără revenire) sau sondaj* fără revenire. Operația de construire a unui vector aleator n-dimensional $(X_1, \ldots, X_n)'$, unde v.a. X_i, $1 \leqslant i \leqslant n$ au aceeași lege (de repartiție) ca și v.a. $X \sim f_X(x)$.

Operația *s.n. (s.f.r.)* revine la a admite că după ce s-a extras la întîmplare un element oarecare din populația considerată, acest element nu mai este introdus în populație înaintea unei noi extrageri.

S.n. (s.f.r.) prezintă interes în cazul unor populații finite. Pentru populații foarte largi, considerate populații infinite, deosebirea dintre operația de *s.n. (s.f.r.)* și operația de *selecție bernoulliană (selecție cu revenire)** dispare.

SONDAJ. Operația de prelevare a elementelor unui eșantion.

Sondaj aleator. Tip de s. în care unitățile populației statistice sînt prelevate la întîmplare. Fiecare element are aceeași probabilitate de a fi ales în eșantion.

Sondaj dirijat. Tip de s. în care unitățile populației statistice sînt prelevate după un anumit principiu prestabilit.

S.d. prezintă pericolul unor *erori sistematice**. Încrederea acordată acestui tip de s. depinde de circumstantele cazului dat, furnizînd citeodată eșantioane mai reprezentative decît sondajul aleator.

Sondaj mixt. Tip de *s.* care combină principiile *sondajului dirijat ** cu cele ale *sondajului aleator** și constă în divizarea întregii populații în straturi-grupe tipice în raport cu un principiu prestabilit și apoi extragerea a cîte unui *eșantion** aleator din fiecare strat. Formarea eșantionului se face în *faze.*

Sondaj cu revenire. *(→Selecție bernoulliană (cu revenire)).*

Sondaj fără revenire. *(→Selecție nebernoulliană (fără revenire)).*

Sondaj stratificat simplu fără revenire. *Sondaj mixt ** efectuat dintr-o populație finită, împărțită în k straturi de volum finit, după fiecare extragere unitatea extrasă nu mai revine în populația originară.

Sondaj stratificat (tipic) fără revenire în două faze. Sondaj stratificat fără revenire dintr-o populație finită care constă din alegerea la întîmplare a unui număr distinct de straturi din totalitatea straturilor (unități primare ale sondajului) — faza întîi a sondajului din care se fac apoi extrageri succesive fără revenire (unități secundare ale sondajului) — faza a doua a sondajului.

Sondaj stratificat fără revenire proporțional. Sondaj stratificat caracterizat prin faptul că din fiecare strat (grupă) tipic în care a fost divizată populația originară se extrag atîtea unități, încît raportul dintre volumul eșantionului

de strat și volumul stratului (grupei) să fie egal cu raportul dintre volumul eșantionului general și volumul total al populației. Termen sinonim cu *sondaj stratificat reprezentativ.*

Observații

1. Se pot efectua *sondaje mixte** sau *sondaje stratificate** din populații infinite : sondaj mixt simplu sau de o singură fază, sondaj mixt în două sau mai multe faze, sondaj mixt proporțional.
2. Sondajul stratificat proporțional este mai eficient decît sondajul stratificat simplu, întrucît dispersia *statisticii* * medie de selecție în cazul sondajului stratificat proporțional este mai mică decît dispersia mediei de selecție în cazul sondajului stratificat simplu.
3. Sondajul stratificat pentru care eficiența sondajului este maximă se numește *sondaj stratificat optim.* Se poate vorbi și de economicitatea sondajului stratificat optim în sensul obținerii maximului de informație statistică cu un cost minim.

Sondaj secvențial. Tip de s. în care volumul eșantionului nu este fixat de la începutul operației de prelevarea probelor, prelevarea continuînd pînă la un *moment (aleator) de oprire.** cînd se realizează un anumit eveniment specificat.

Moment aleator de oprire. Momentul de timp la care se termină operația de *sondaj secvențial* *.

M.a.o. definit pe șirul de v.a. $\{Z_n\}$ este o v.a. N cu valori întregi, pozitive, astfel încît evenimentul $(N = n)$ să aparțină σ-cîmpului indus de v.a. $Z_1, \ldots, Z_n$, $n \geqslant 1$, adică pentru orice n, evenimentul $(N = n)$ depinde numai de v.a. $Z_1, \ldots, Z_n$.

Operația de *sondaj secvențial* * se termină cu v.a. Z_N, definită astfel încît valoarea lui Z_N pentru $\omega \in \Omega$ să fie egală cu $Z_{N(\omega)}(\omega)$.

Sondaj din materiale în vrac. Tip de s. utilizat în cazul în care populația studiată nu poate fi considerată ca fiind formată din unități elementare fizic individualizate. Unitățile eșantionului se constituie a posteriori, ca urmare a unor prelevări elementare repartizate la întîmplare în masa populației studiate. Populația se consideră ca un eșantion global. Plecînd de la acest eșantion global, se prepară eșantionul destinat analizei sau controlului.

S.m.v. se aplică în cazul unor loturi formate din vagoane de cărbuni, minereuri, baloturi de bumbac etc.

STATISTICĂ *(m-dimensională).* Vectorul $\mathbf{T}$ cu m componente $T_1, \ldots, T_m$: $\mathbf{T}(X_1, \ldots, X_n) = (T_1(X_1, \ldots, X_n), \ldots, T_m(X_1, \ldots, X_n))'$, unde T_j, $1 \leqslant j \leqslant m$ sînt funcții măsurabile.

$T_j : R^n \to R$ depinde de *eșantionul* $X_1, \ldots, X_n$ *de volum** n : corespunzător v.a. $X \sim f_X(x)$.

Dacă $m = 1$, s. $\mathbf{T}(X_1, \ldots, X_n)$ se notează $T(X_1, \ldots, X_n)$.

SPAȚIUL PARAMETRILOR. Fie v.a. $X \sim f_X(., \boldsymbol{\theta})$ a cărei d.p. depinde de vectorul r-dimensional $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)'$ constant necunoscut numit *vectorul parametrilor*. S.p. este mulțimea $\Theta \leqslant R^r$, $r \geqslant 1$ a tuturor valorilor posibile ale lui $\boldsymbol{\theta}$.

FUNCȚIE EMPIRICĂ DE REPARTIȚIE *(Funcție de repartiție de selecție)* $\mathbf{F_n}$. Funcția $F_n(.)$ definită prin relația :

$$F_n(x) = \frac{n(x)}{n},$$

unde $n(x)$ este numărul v.a. X_i, $1 \leqslant i \leqslant n$ independente avînd aceeași lege (de repartiție) $F_X(.., \theta)$, adică numărul v.a. X_i ale *eșantionului* (de volum n)* ale căror valori sînt mai mici decît x, pentru $x \in R$.
F_n este o funcție în trepte (etajată), care este o *funcție de repartiție** pentru (*datele (de observație)** $x_1, \ldots, x_n$, ale v.a. $X_1, \ldots, X_n$. F_n este o v.a. ca funcție de v.a. $X_1, \ldots, X_n$, pentru orice $x \in R$, definită pe același cîmp de probabilitate pe care sînt definite v.a. X_i.
Legătura dintre **f.e.r. (f.r.s.)** și *funcția de repartiție** F_X (teoretică) a v.a. X este pusă în evidență de *Teorema Glivenko-Cantelli (→ Legea numerelor mari *)*: cu notațiile de mai sus, este adevărat că:

$$P(\sup \{|F_n(x) - F_X(x)|;\ x \in R\} \to 0) = 1,$$
$$n \to \infty$$

adică $F_n(x) \xrightarrow[n \to \infty]{\text{a.s.}} F_X(x)$, uniform în $x \in R$.

MOMENT EMPIRIC (MOMENT DE SELECȚIE) DE ORDIN k. Se consideră un *eșantion de volum* n* și fie $x_1, \ldots, x_n$ *variabilele de selecție (n observații)* corespunzătoare.
M.e.(m.s.).o.k este *statistica** $\overline{m}_k : R^n \to R$ astfel încît:

$$\overline{m}_k = \frac{x_1^k + x_2^k + \ldots + x_n^k}{n}.$$

Caracteristici principale
Media: $E(\overline{m}_k) = m_k$.
Dispersia: $\operatorname{Var}(\overline{m}_k) = \dfrac{m_{2k} - m_k^2}{n}$.

Observație. **M.e.(m.s.).o.k** *converge în probabilitate* către m_k, *momentul (teoretic) de ordin k**, cu condiția ca momentele teoretice de ordin k, respectiv $2k$, să existe.

Media de selecție. Pentru $k = 1$ se obține **m.e.o.** $k = 1$ sau *m.s.*:

$$\overline{m}_1 = \frac{x_1 + x_2 + \ldots + x_n}{n} = \overline{X}.$$

Caracteristici principale
Media: $E(\overline{X}) = m$ (media teoretică).
Dispersia: $\operatorname{Var}(\overline{X}) = \sigma^2/n$, unde σ^2 este dispersia teoretică.
Observație
1. *M. s. converge în probabilitate* către media teoretică m a v.a. X.
Exemple
1. Se consideră v.a. $X \sim N(m, \sigma^2)$ și un *eșantion de volum* n* corespunzător ei. *M.s.* $\overline{X}$ este o v.a. $\sim N(m, \sigma^2/n)$.
2. Dacă v.a. $X \sim$ *legea Cauchy **, atunci *m.s.* $\overline{X} \sim$ legea Cauchy, deși legea (de repartiție) Cauchy nu are momente de nici un ordin.
3. Dacă v.a. $X \sim$ *legea Poisson* Po(λ)*, atunci *m.s.* $\overline{X}$ este o v.a. a cărei lege

(de repartiție) este apropiată de legea Poisson $Po(\lambda)$ și are funcția caracteristică :

$$\varphi_X(t) = e^{n\lambda(e^{it/n}-1)}.$$

MOMENT EMPIRIC CENTRAT (MOMENT CENTRAT DE SELECȚIE) DE ORDIN k. Se consideră un *eșantion de volum* n* și fie $x_1, \ldots, x_n$ *variabilele de selecție (observațiile)* corespunzătoare. **M.e.c.o.** k este *statistica** $\overline{\mu}_k : R^n \to R$ definită prin relația :

$$\overline{\mu}_k = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{X})^k$$

unde $\overline{X}$ este *media de selecție** corespunzătoare eșantionului.
Caracteristici principale
Media : $E(\overline{\mu}_k) = \mu_k + \mathcal{O}(1/n)$.
Dispersia : $\operatorname{Var}(\overline{\mu}_k) = \dfrac{\mu_{2k} - \mu_k^2 - 2k\mu_{k-1}\mu_{k+1} + k^2\mu_2\mu_{k-1}^2}{n} + \mathcal{O}(1/n^2)$

cu condiția ca media v.a. X să fie nulă, adică $E(X) = 0$ și μ_k — *momentul central de ordin* k* al v.a. X.

Dispersia de selecție. Pentru $k = 2$ se obține **m.e.c.o.** 2 sau $d.s.$

$$\overline{\mu}_2 = 1/n \sum_{i=1}^{n} (x_i - \overline{X})^2 = s^2.$$

Exemplu. Se consideră v.a. $X \sim N(m, \sigma^2)$ și un eșantion de volum n corespunzător ei. V.a.

$$(n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$$

unde s^2 este $d.s.$ corespunzătoare eșantionului considerat.

XV. TEORIA ESTIMAȚIEI

În acest capitol este inclus ansamblul de metode ale statisticii matematice care permit determinarea valorilor unui parametru θ al legii (de repartiție) v.a. $X \sim f_X(x; \theta)$, $\theta \in \Theta \subset R^n$, pe baza rezultatelor experimentelor sau observațiilor unui *eșantion* din populația corespunzătoare v.a. X. Mai general, se estimează o funcție reală, măsurabilă $g(\theta)$, care depinde de parametrul θ al legii respective.

Aceste metode se aplică în probleme de estimare a parametrilor unor legi (de repartiție) discrete sau continue, pentru medii și combinații liniare de medii ale unor populații normale, pentru dispersii și raport de dispersii ale unor populații normale, pentru obținerea unor *intervale de încredere** referitoare la funcția de repartiție F_X a v.a. X.

ESTIMATOR. Se consideră v.a. $X \sim f_X(\cdot, \theta)$, $\theta \in \Theta \subset R^r$ și $g(\theta)$ o funcție reală, măsurabilă care depinde de parametrul necunoscut θ.

Fie $X_1, \ldots, X_n$ un *eșantion* * *de volum n* corespunzător v.a. X, adică n v.a. independente $X_1, \ldots, X_n$ având aceeași lege (de repartiție) ca și v.a. X.

Un **e.** al lui $g(\theta)$ este o *statistică** $U = U(X_1, \ldots, X_n)$ folosită pentru a determina funcția $g(\theta)$. **E.** fără altă specificare este termen sinonim cu *estimator punctual* (*point estimator*, în engl.). **E.** se mai numește și *funcție de estimație*. *Exemplu.* Fie un eșantion de volum n dintr-o populație normală $N(m, \sigma^2)$, pentru care m și (sau) σ^2 sînt parametri necunoscuți.

Pentru estimarea parametrului m — media teoretică, se pot folosi mai mulți estimatori: *media** (de selecție), *mediana** (de selecție) etc. calculați pe baza eșantionului considerat.

Estimator nedeplasat. Un **e.** $U = U(X_1, \ldots, X_n)$ este *e.n.* a lui $g(\theta)$ dacă:

$$E_\theta\, U(X_1, \ldots, X_n) = g(\theta), \text{ pentru orice } \theta \in \Theta.$$

Estimator absolut corect. **E.** *nedeplasat** pentru care:
$E_\theta U(X_1, \ldots, X_n) = g(\theta)$

$\mathrm{Var}_\theta U(X_1, \ldots, X_n) = \mathcal{O}(n)$,

unde $\mathcal{O}(n) \to 0$ pentru $n \to \infty$.
Se spune că **e.** U converge în medie pătratică către $g(\theta)$ cînd $n \to \infty$.
Nedeplasarea este unul din criteriile de alegere între 2 sau mai mulți **e.** ai lui $g(\theta)$ dintr-o clasă de **e.** avînd anumite proprietăți optimale.

Observație. Ținînd seama de legătura dintre convergența în medie pătratică și convergența în probabilitate, se spune că **e.** U este un *e.a.c.* al lui $g(\theta)$ dacă $U(X_1, \ldots, X_n)$ converge în probabilitate către $g(\theta)$ cînd $n \to \infty$, adică: pentru orice $\varepsilon > 0$, $P(|U(X_1, \ldots, X_n) - g(\theta)| < \varepsilon) \geqslant 1 - \alpha$.

Estimator consistent. Un **e.** U al lui $g(\theta)$ este *e.c.* dacă converge în probabilitate către U.

Estimator corect. Un **e.** $U = U(X_1, \ldots, X_n)$ este *e.c.* al lui $g(\theta)$ dacă:

$$E_\theta U(X_1, \ldots, X_n) = g(\theta) + f(g(\theta)\,;\, n),$$

unde $f(g(\theta)\,;\, n)$ este o funcție de $g(\theta)$ astfel încît $f(g(\theta)\,;\, n) \to 0$ cînd $n \to \infty$, adică:
$E_\theta U(X_1, \ldots, X_n) = g(\theta) + f(g(\theta)\,;\, n)$

$\mathrm{Var}_\theta\, U(X_1, \ldots, X_n) = \mathcal{O}(n)$,

unde $\mathcal{O}(n) \to 0$ cînd $n \to \infty$.
E.c. se mai numește și *e. consistent deplasat.*

Exemplu. Fie v.a. $X \sim Bi(n, p)$ reprezentînd numărul de apariții ale unui eveniment A în n experimente independente, unde p este probabilitatea realizării lui A în fiecare probă.

Se pune problema estimării parametrului necunoscut p al legii $Bi(n, p)$ pe baza observațiilor din cele n probe independente ale experimentului.

Fie $X = \nu$, numărul de realizări ale evenimentului A în cele n probe ale experimentului.

Statistica ν/n — frecvența relativă a lui A este un e. *absolut corect* al lui p, deoarece

$$E(\nu/n) = \frac{1}{n} E(\nu) = \frac{1}{n} np = p$$

$$\text{Var}\,(\nu/n) = 1/n^2 \text{Var}\,(\nu) = 1/n^2 \cdot npq = pq/n = \mathcal{C}(n),$$

unde $\mathcal{C}(n) = pq/n \to 0$ cînd $n \to \infty$.
Statistica $(\nu + 1)/(n + 1)$ este un e. *corect* al lui p, deoarece :

$$E((\nu + 1)/(n + 1)) = p + \frac{1 - p}{n + 1}$$

$$\text{Var}\left(\frac{\nu + 1}{n + 1}\right) = \frac{npq}{(n + 1)^2} = \mathcal{C}(n),$$

unde $\mathcal{C}(n) \to 0$, cînd $n \to \infty$.
Pentru n suficient de mare, diferența dintre e. absolut corect ν/n și e. corect $(\nu + 1)/(n+1)$ este foarte mică.

Estimator eficient (estimator uniform nedeplasat de dispersie minimă — — *UMVU estimator* în engl.).

Un *estimator nedeplasat** $U(X_1, \ldots, X_n)$ al lui $g(\boldsymbol{\theta})$, care are proprietatea de a avea cea mai mică dispersie în clasa tuturor estimatorilor nedeplasați ai lui $g(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$.

E.ef. are cea mai mică împrăștiere în jurul parametrului $\boldsymbol{\theta}$ sau a lui $g(\boldsymbol{\theta})$ (fig. 108). Cu alte cuvinte, dacă $U(X_1, \ldots, X_n)$ este *e.ef.* al lui $g(\boldsymbol{\theta})$ și dacă

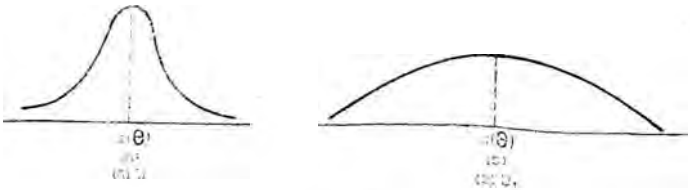


Fig. 108

$U_1 = U_1(X_1, \ldots, X_n)$ este un alt e. nedeplasat al lui $g(\boldsymbol{\theta})$, atunci :

$$\text{Var}_{\boldsymbol{\theta}} U_1 \geqslant \text{Var}_{\boldsymbol{\theta}} U, \text{ pentru orice } \boldsymbol{\theta} \in \Theta.$$

Observație. Valoarea minimă V_0 a dispersiei unui *e.ef.* este dată de *inegalitatea Rao-Cramér-Fréchet*:

$$\mathrm{Var}_{\theta} U \geqslant V_0 = \frac{1}{nE\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2} = \frac{1}{n\int\limits_{-\infty}^{+\infty}\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2 f(x,\theta)\,dx} =$$

$$= 1/I(\theta),$$

unde

$$I(\theta) = nE\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2$$

se numește *informația* (sau *cantitatea de informație*) conținută în eșantionul corespunzător de volum n.

Expresia $E\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2$ se numește *informația lui Fisher*.

Exemplu. Media de selecție

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

a unui eșantion $(X_1,\ldots,X_n)$ provenind dintr-o populație normală corespunzătoare v.a. $X \sim N(m, \sigma^2)$ (m — parametrul necunoscut; σ^2 — cunoscut) este un *e.cf.* al mediei teoretice necunoscute, m.

Într-adevăr, $\overline{X}$ este e. nedeplasat al lui m, avînd:

$$\mathrm{Var}\,\overline{X} = \sigma^2/n$$

adică un e. pentru care este atinsă V_0, marginea inferioară dată de inegalitatea Rao-Cramér-Fréchet:

$$E_{\theta(=m)}[\partial/\partial m \ln f(x,m)]^2 = E_m\left[\frac{\partial}{\partial m}\left(\ln\frac{1}{\sqrt{2\pi}\sigma} - \frac{(x-m)^2}{2\sigma^2}\right)\right]^2 =$$

$$= E_m\left(\frac{1}{\sigma^2}\left(\frac{(x-m)}{\sigma}\right)^2\right) = 1/\sigma^2,$$

deoarece v.a. $(X-m)/\sigma \sim N(0,1)$ și deci:

$$E_m\left(\frac{(X-m)}{\sigma}\right)^2 = 1.$$

Rezultă că:

$$V_0 = \frac{1}{n\cdot 1/\sigma^2} = \frac{\sigma^2}{n}.$$

Informația conținută în eșantionul $(X_1, \ldots, X_n)$ este :

$$I(m) = n/\sigma^2.$$

Observații

1) În cazul în care σ^2 este parametrul necunoscut și m—cunoscut, un *e.ef.* al lui $\sigma(= g(\theta))$ este S/c_n, unde :

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - m)^2,$$

$$c_n = c_n'/\sqrt{n},$$

$$c_n' = E_{\theta(=\sigma^2)} \left(\frac{\sqrt{n}\, S}{\sigma} \right).$$

2) În cazul în care ambii parametri — m, σ^2 sînt necunoscuți, se notează vectorul parametrilor :

$$\theta = (\theta_1, \theta_2)' = (m, \sigma^2)'.$$

E. ef. ai componentelor lui θ sînt :

$$U_1 = \overline{X} \text{ (media de selecție)},$$

respectiv

$$U_2 = nS^2/(n-1), \text{ cu } S^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2,$$

deoarece sînt e. nedeplasați ai lui m, σ^2 respectiv :

$$E_\theta U_1 = m,$$

$$E_\theta U_2 = \sigma^2,$$

iar statistica $(\overline{X}, S^2)$ este o *statistică suficientă* a lui θ, întrucît :

$$f(x_1, \ldots, x_n; \theta) = \left(\frac{1}{\sqrt{2\pi\theta_2}} \right)^n \exp \left[-\frac{1}{2\theta_2} \sum_{i=1}^{n} (x_i - \bar{x})^2 - \frac{n}{2\theta_2} (\bar{x} - \theta_1)^2 \right].$$

3) În aceleași condiții de mai sus, se găsește un *e.ef.* pentru α — *cuantila superioară* * a legii $N(0,1)$, z_α definită astfel încît :

$$\alpha = P_\theta(X_1 \geq Z_\alpha)$$

sub forma

$$U = \overline{X} - \frac{S}{c_n} \Phi^{-1}(1 - \alpha),$$

unde c_n este definit la pct. 1. iar $\Phi(x)$ este funcția de repartiție a legii $N(0, 1)$.
Exemplu. Fie $X_1, \ldots, X_n$ un eșantion de volum n corespunzător unei v.a. $X \sim Po(\lambda)$, $\lambda \rhd 0$ având d.p.:

$$f(x; \theta(=\lambda)) = e^{-\lambda} \frac{\lambda^x}{x!}.$$

Media de selecție:

$$\overline{X} = \frac{X_1 + \ldots + X_n}{n}$$

este un *e.ef.* al parametrului λ (= media teoretică). Într-adevăr, $\overline{X}$ este un **e.** nedeplasat al lui λ, având

$$\operatorname{Var} \overline{X} = \lambda/n,$$

adică un **e.** pentru care este atinsă V_0, marginea inferioară Rao-Cramér-Fréchet:

$$E_\lambda \left(\frac{\partial}{\partial \lambda} \ln f(x; \lambda) \right)^2 = E_\lambda \left[\left(\frac{\partial}{\partial \lambda} (-\lambda + x \ln \lambda - \ln x!) \right)^2 \right] =$$

$$= E_\lambda \left(1 + \frac{1}{\lambda^2} x^2 - \frac{2}{\lambda} x \right) = -1/\lambda,$$

ținînd seama că:

$$E_\lambda X = \lambda; \quad E_\lambda X^2 = \lambda(1 + \lambda).$$

Deci

$$V_0 = \frac{1}{n \cdot 1/\lambda} = \frac{\lambda}{n}.$$

Eficiență absolută. *E.a.* a unei **e.** nedeplasat U este raportul dintre $V_0 = 1/I(\theta)$ — dispersia minimă dată de inegalitatea Rao-Cramér-Fréchet — și dispersia $V = \operatorname{Var}_\theta U$ a **e.** nedeplasat U:

$$e.a. = \frac{V_0}{V} = \frac{V_0}{\operatorname{Var}_\theta U}.$$

Eficiență relativă. Fie U un estimator eficient al lui $g(\theta)$ și U^* un alt estimator nedeplasat al lui $g(\theta)$.
E.r. a lui U^* relativ la U este raportul dintre dispersia lui U și dispersia lui U^*, adică:

$$e.r. = \frac{\operatorname{Var}_\theta U}{\operatorname{Var}_\theta U^*}.$$

E.r. ia valori în $(0, 1]$.

Estimator suficient (exhaustiv). Un **e.** $\mathbf{T}(X_1, \ldots, X_n)$ al lui $\boldsymbol{\theta}$ care asigură folosirea completă a informației cuprinse în eșantionul de volum n relativ la parametrul $\boldsymbol{\theta}$ al legii v.a. $X \sim f(\cdot\,; \theta)$ corespunzătoare.

Un **e.** absolut corect $\mathbf{T}(X_1, \ldots, X_n)$ al lui $\boldsymbol{\theta}$ este *e.s.* (*e.*) dacă $\mathbf{T}(X_1, \ldots, X_n)$ este o *statistică suficientă* pentru familia de densități $f(x; \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subseteq R^r$, adică d.p. a vectorului $(X_1, \ldots, X_n)'$ este de forma:

$$f(x_1, \ldots, x_n; \boldsymbol{\theta}) = g(\mathbf{T}(x_1, \ldots, x_n); \boldsymbol{\theta}) \cdot h(x_1, \ldots, x_n)$$

(*Teorema de factorizare Fisher-Neyman*), unde g depinde de x_i — valorile observate (datele) ale v.a. X_i, $1 \leqslant i \leqslant n$ numai prin intermediul lui $\mathbf{T}$ și h este independentă de $\boldsymbol{\theta}$.

Cu alte cuvinte, probabilitatea condiționată a ansamblului $(X_1, \ldots, X_n)$ pentru $\mathbf{T}(X_1, \ldots, X_n)$ dat este independentă de $\boldsymbol{\theta}$.

Se spune că **e.** $\mathbf{T}(X_1, \ldots, X_n)$ conține în mod exhaustiv toată informația referitoare la parametrul necunoscut $\boldsymbol{\theta}$.

Exemplu. Media de selecție $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ a unui eșantion de volum n, $(X_1, \ldots, X_n)$, provenind dintr-o populație normală $N(m, \sigma^2)$; σ^2 — cunoscut, este un *e.s.* (*e*) al lui m, media teoretică.

Observație. Orice **e.** eficient este și *e.s.* (*e*).

ESTIMAȚIE. Valoarea $U(x_1, \ldots, x_n)$ a unei statistici $U = U(X_1, \ldots, X_n)$ pentru valorile observate x_i ale v.a. independente identic repartizate X_i, $1 \leqslant i \leqslant n$. Valoarea numerică a estimatorului obținută pe baza unor date de observație, adică valoarea furnizată de metoda de estimare folosită.

FUNCȚIE PARAMETRICĂ. Funcție liniară de parametrii modelului, adică o funcție de forma $\psi = \mathbf{c}'\boldsymbol{\beta}$, unde $\mathbf{c} \in R^p$ este un vector cunoscut, $\boldsymbol{\beta} \in R^p$ — *spațiul parametrilor**, vector necunoscut.

FUNCȚIE ESTIMABILĂ. *Funcție parametrică** ψ pentru care există un vector constant $\mathbf{a} \in R^n$ astfel încît $E(\mathbf{a}'\mathbf{y}) = \psi$ (identitate în raport cu $\boldsymbol{\beta}$), adică există un *estimator linear nedeplasat** al său, $\mathbf{a}'\mathbf{y}$. **F.e.** se mai numește și *funcție care admite estimație.*

Proprietăți ale **f.e.**

1. O *funcție parametrică** $\psi = \mathbf{c}'\boldsymbol{\beta}$ este **f.e.** dacă și numai dacă există un vector constant $\mathbf{a}$ astfel încît: $\mathbf{c}' = \mathbf{a}'\mathbf{X}'$.

2. Pentru orice funcție $\psi = \mathbf{c}'\boldsymbol{\beta}$, $\mathbf{c} \in R^p$, **f.e.**, există o singură *estimație nedeplasată** de forma $\psi^* = \mathbf{a}^{*\prime}\mathbf{y}$, $\mathbf{a}^* \in R^r$. Dacă $\boldsymbol{\alpha}'\mathbf{y}$ este o estimație nedeplasată oarecare a lui ψ, atunci $\mathbf{a} = \mathrm{pr}_{R^r}(\boldsymbol{\alpha})$, $\boldsymbol{\alpha} \in R^n$ ($\mathbf{a}$ este proiecția lui $\boldsymbol{\alpha}$ pe R^r).

3. (*Teorema Gauss-Markov*). În condițiile Ω (→ *Model linear general*) orice funcție parametrică $\psi = \mathbf{c}'\boldsymbol{\beta}$, **f.e.** are o estimație liniară nedeplasată unică $\hat{\psi}$ cu cea mai mică dispersie, de forma $\hat{\psi} = \mathbf{c}'\hat{\boldsymbol{\beta}}$, unde $\hat{\boldsymbol{\beta}}$ este *m.m.p.-estimatorul** lui $\boldsymbol{\beta}$.

4. Dacă $\psi_1, \ldots \psi_q$ sînt **f.e.**, atunci orice combinație liniară de forma $\psi = \sum_i h_i \psi_i$ este de asemenea **f.e.** și m.m.p.-estimația $\hat{\psi}$ a combinației liniare ψ este de forma $\hat{\psi} = \sum_i h_i \hat{\psi}_i$, unde $\hat{\psi}_i$ sînt m.m.p.-estimațiile **f.e.** ψ_i, $1 \leqslant i \leqslant q$.

PRINCIPIUL VEROSIMILITĂȚII MAXIME. (*Maximum Likelihood*, în engl.). Metodă de obținere a unui estimator $\hat{\theta}$ al lui θ, astfel încît să realizeze maximul *funcției de verosimilitate**.

Exemplu. Se consideră experimentul din cazul urnei lui Bernoulli, care conduce în n repetări la *legea binomială* *, $Bi(n, p)$, unde p este probabilitatea apariției „stemei" la o aruncare a monedei. Probabilitatea obținerii „stemei" de r ori și a „banului" de $n-r$ ori în n aruncări independente ale monedei este:

$$P(X = r) = p_X = C_n^r p^r q^{n-r}.$$

Funcția de verosimilitate a evenimentului observat este:

$$L = C_n^r p^r (1 - p)^{n-r}.$$

Aplicarea **p.v.m.** constă în determinarea valorii $\hat{p}$ a parametrului p astfel încît să se realizeze:

$$\max L$$

sau

$$\max \ln L = \max \{\ln C_n^r + r \ln p + (n - r) \ln (1 - p)\}.$$

Soluția $\hat{p}$ se obține din ecuația de verosimilitate:

$$\frac{\partial}{\partial p} \ln L = \frac{r}{p} - \frac{(n - r)}{1 - p} = 0$$

și este

$$\hat{p} = r/n$$

p este un *estimator de verosimilitate maximă** al parametrului p.

ESTIMATOR DE VEROSIMILITATE MAXIMĂ. (Maximum Likelihood Estimator, în engl.) Estimator obținut aplicînd *principiul verosimilității maxime* *.

$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(X_1, \ldots, X_n)$ este **e.v.m.** al lui $\boldsymbol{\theta}$, dacă

$$L(x_1, \ldots, x_n; \hat{\boldsymbol{\theta}}) = \max \{L(x_1, \ldots, x_n; \boldsymbol{\theta}); \quad \boldsymbol{\theta} \in \Theta\},$$

unde L este *funcția de verosimilitate* * calculată pentru valorile x_i ale v.a. X_i $1 \leqslant i \leqslant n$ ale unui eșantion de volum n din populația statistică corespunzătoare v.a. $X \sim f(\cdot\,; \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subseteq R^r$. Aici se consideră valorile x_i, $1 \leqslant i \leqslant n$ ca fiind constante și L ca funcție de $\boldsymbol{\theta}$, vectorul parametrilor.

E.v.m. este *consistent*; (*asimptotic*) *eficient* cînd n, volumul eșantionului, este suficient de mare.

E.v.m. are proprietatea de *invarianță*: dacă $\hat{\boldsymbol{\theta}}$ este **e.v.m.** lui $\boldsymbol{\theta}$ și dacă $f(\boldsymbol{\theta})$ este o funcție *bijectivă* de $\boldsymbol{\theta}$, atunci $f(\hat{\boldsymbol{\theta}})$ este **e.v.m.** al lui $f(\boldsymbol{\theta})$.

De exemplu, $\hat{\sigma} = \widehat{(\sigma^2)^{1/2}}$

E.v.m. este de regulă *deplasat*. Deplasarea este înlăturată printr-o transformare simplă.

Exemplu.

1. Fie v.a. $X \sim N(m, \sigma^2)$ și fie $(x_1, \ldots, x_n)$ valorile v.a. X_i, $1 \leqslant i \leqslant n$, care formează un eșantion de volum n relativ la v.a. X. Funcția de verosimilitate core-

punzătoare este :

$$L(x_1, \ldots, x_n : m, \sigma^2) = \left[\frac{1}{\sigma(2\pi)^{1/2}} \exp\left(-\frac{1}{2}\left(\frac{x_1 - m}{\sigma}\right)^2\right)\right] \cdot \ldots \cdot$$

$$\cdot \left[\frac{1}{\sigma(2\pi)^{1/2}} \exp\left(-\frac{1}{2}\left(\frac{(x_n - m)}{\sigma}\right)^2\right)\right].$$

E.v.m. ai parametrilor m, σ^2 se obțin astfel încît

$$\max L$$

sau

$$\max \ln L,$$

dacă soluțiile sistemului de ecuații :

$$\begin{cases} \dfrac{\partial \ln L}{\partial m} = 0 \\ \dfrac{\partial \ln L}{\partial \sigma^2} = 0 \end{cases}$$

și sînt de forma :

$$\hat{m} = \bar{x}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

Ultimul estimator este ușor deplasat, dar deplasarea poate fi înlăturată prin multiplicare cu $n/(n-1)$, adică :

$$\hat{\sigma}'^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

este estimator nedeplasat al lui σ^2.

Așadar, **e.v.m.** ai parametrilor m, σ^2 ai legii $N(m, \sigma^2)$ sînt media, respectiv dispersia, de selecție a valorilor observate x_i ale v.a. X_i ale eșantionului considerat.

2. Fie $X_1, \ldots, X_n$ un eșantion de volum n corespunzător v.a. $X \sim Po(\lambda)$. Funcția de verosimilitate corespunzătoare valorilor observate x_i ale v.a. X_i, $1 \leqslant i \leqslant n$ este :

$$L(\lambda : x_1, \ldots, x_n) = e^{-n\lambda} \frac{1}{\prod_{i=1}^{n} x_i !} \lambda^{\sum_{i=1}^{n} x_i}$$

și deci :

$$\ln L(\lambda ; x_1, \ldots, x_n) = -n\lambda - \ln\left(\prod_{i=1}^{n} x_i !\right) + \left(\sum_{i=1}^{n} x_i\right) \ln \lambda.$$

E.v.m. al parametrului λ se obține astfel încît :

$$\max_{\lambda} \ln L,$$

adică drept soluție a ecuației :

$$\frac{\partial}{\partial \lambda} \ln L(\lambda ; x_1, \ldots, x_n) = -n + n\bar{x} \cdot \frac{1}{\lambda} = 0,$$

având forma :

$$\hat{\lambda} = \bar{x}.$$

Se observă că :

$$\frac{\partial^2}{\partial \lambda^2} L(\lambda ; x_1, \ldots, x_n) = -n\bar{x} \cdot \frac{1}{\lambda^2} < 0, \quad \text{pentru orice } \lambda > 0$$

și deci, în particular, și pentru $\lambda = \hat{\lambda}$. Deci $\bar{x}$ realizează max ln L și este **e.v.m.** al lui λ.

E.v.m. $\hat{\lambda}$ este un *estimator absolut corect* al lui λ :

$$E(\hat{\lambda}) = \lambda$$

$$\text{Var}\,(\hat{\lambda}) = \frac{\lambda}{n} \qquad (\to 0 \text{ cînd } n \to \infty).$$

Întrucît V_0, marginea inferioară dată de *inegalitatea Rao-Cramér-Fréchet*

$$V_0 = \frac{1}{n \sum_{k=0}^{\infty} \left(\frac{\partial \ln P(k ; \lambda)}{\partial \lambda}\right)^2 \cdot P(k ; \lambda)} = \frac{\lambda}{n}$$

este atinsă pentru $\hat{\lambda}$, rezultă că $\hat{\lambda}$ este estimator eficient al lui λ.

FUNCȚIE DE VEROSIMILITATE. Fie $f(x, y, \ldots ; \theta_1, \ldots, \theta_k)$ d.p. a unei v.a. X sau vector aleator $\mathbf{X}$ de una sau mai multe variabile $x, y, \ldots$, depinzînd de unul sau mai mulți parametri $\theta_1, \ldots, \theta_k$.

F.v. este această funcție $f(x, y, \ldots ; \theta_1, \ldots, \theta_k)$ considerată ca funcție de parametrii $\{\theta\}$, pentru $x, y, \ldots$ fixați. De exemplu, în cazul unui eșantion de volum n constînd din n valori independente $x_1, \ldots, x_n$ ale v.a. $X \sim f_X(x ; \theta)$, **f.v.** a parametrului θ este :

$$L(x_1, \ldots, x_n ; \theta) = f(x_1, \theta) f(x_2, \theta) \cdot \ldots \cdot f(x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta),$$

adică este probabilitatea obținerii unui eșantion particular $(x_1, \ldots, x_n)$ în ipoteza că legea (de repartiție) este specificată apriori prin $f(x : \theta)$.

METODA CELOR MAI MICI PĂTRATE. Metodă de obținere a estimatorilor unor parametri ai legii $f(\cdot : \theta)$ astfel încît suma pătratelor abaterilor dintre observația propriu-zisă (O) și valoarea teoretică (E) ce se așteaptă a fi obținută să fie minimă :

$$\sum (O - E)^2 = \min.$$

Mai precis, dacă $y_1, \ldots, y_n$ sînt n v.a. avînd dispersiile constante cunoscute, σ_i^2 și ale căror medii $m_i(\theta_1, \ldots, \theta_r)$, $1 \leqslant i \leqslant n$ depind de r parametri $\theta_1, \ldots, \theta_r$, funcțiile m_i fiind cunoscute, **m.m.m.p.** constă în determinarea estimatorilor parametrilor $\theta_1, \ldots, \theta_r$ astfel încît :

$$\mathcal{S} = \sum_{i=1}^{n} g_i (Y_i - m_i)^2 = \text{minimă}.$$

unde

$$g_i = 1/\sigma_i^2, \quad 1 \leqslant i \leqslant n.$$

Notînd

$$\theta_j = \theta_j^0 + u_j, \quad 1 \leqslant j \leqslant r,$$

unde θ_j^0 sînt valorile aproximative ale parametrilor și u_j — corecțiile respective, $\min \mathcal{S}$ se obține rezolvînd sistemul de ecuații format din derivatele parțiale ale lui $\mathcal{S}$ în raport cu $u_1, \ldots, u_r$:

$$\frac{\partial \mathcal{S}}{\partial u_j} = 0, \quad j = 1, \ldots, r.$$

M.m.m.p. se aplică în *modele liniare**, adică atunci cînd în prezența unor erori întîmplătoare e_i, v.a. Y_i depind linear de parametrii necunoscuți θ_j :

$$Y_i = \sum_{j=1}^{r} x_{ji} \theta_j + e_i, \quad i = 1, \ldots, n$$

sau

$$\mathbf{Y} = \mathbf{X}'\boldsymbol{\theta} + \mathbf{e},$$

unde

$$\mathbf{Y} = (y_1, \ldots, y_n)'$$

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ x_{r1} & x_{r2} & \ldots & x_{rn} \end{pmatrix}$$

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)'$$

$$\mathbf{e} = (e_1, \ldots, e_n)'$$

V.a. $\{e_i\}$ sînt erori aleatoare despre care se presupune că:

$$E(e_i) = 0$$

$$\text{Var } e_i = \sigma^2, \quad i = 1, \ldots, n$$

$$\text{Cov } (e_i, e_j) = 0, \quad i \neq j$$

M.m.m.p. constă în determinarea parametrului $\boldsymbol{\theta}$ astfel încît *suma pătratelor erorilor* să fie minimă, adică:

$$\min \mathcal{S} = \min \|\mathbf{e}\|^2 = \min \|\mathbf{Y} - \mathbf{X}'\boldsymbol{\theta}\|^2.$$

ESTIMATOR MMP (Least Square Estimator, în engl.). Estimator obținut prin *metoda celor mai mici pătrate*.

$\hat{\boldsymbol{\theta}}$ este **e.m.m.p.** al lui $\boldsymbol{\theta}$ dacă minimizează suma pătratelor erorilor:

$$\min \mathcal{S}(\mathbf{Y}, \boldsymbol{\theta}) = \min \|\mathbf{e}\|^2 = \min \|\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}\|^2.$$

E.m.m.p. este soluția ecuațiilor:

$$\frac{\partial \mathcal{S}(\mathbf{Y}, \boldsymbol{\theta})}{\partial \theta_\nu} = 0, \quad \nu = 1, \ldots, r$$

numite *ecuații normale*, adică:

$$\sum_{i=1}^{n} \sum_{j=1}^{r} x_{\nu i} x_{ji} \theta_j = \sum_{i=1}^{n} x_{\nu i} Y_i, \quad \nu = 1, \ldots, r$$

sau

$$\mathbf{X}\mathbf{X}'\boldsymbol{\theta} = \mathbf{X}\mathbf{Y}.$$

$\hat{\boldsymbol{\theta}}$ — **e.m.m.p.** al lui $\boldsymbol{\theta}$ este:

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}\mathbf{X}')^{-1}\mathbf{X}\mathbf{Y}.$$

E.m.m.p. este estimator nedeplasat:

$$E(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$$

și are matricea de covarianță:

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = \sigma^2(\mathbf{X}\mathbf{X}')^{-1}.$$

Exemplu. **E.m.m.p.** al mediei teoretice m a unei v.a. $X \sim f_X(x; m)$ obținut pe baza unui eșantion de volum n: $X_1, \ldots, X_n$ se obține făcînd minimă expresia

$$\min \sum_{i=1}^{n} (X_i - m)^2$$

și este

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i. \quad \text{media de selecție.}$$

METODA MOMENTELOR. Metodă folosită pentru ajustarea unei legi (de repartiție) teoretică specificată pe baza unei legi empirice și care constă în egalarea primelor momente empirice (de selecție) cu momentele corespunzătoare ale legii (de repartiție) teoretice:

Exemplu. Să se ajusteze legea (de repartiție) pe baza observațiilor din tab. 61 (J. H. Pollard. *Numerical and Stat. Techniques*, Cambridge Univ. Press, Cambridge-Melbourne, 1977).

Tabelul 61

Frecvențe ajustate

Nr. de ouă	Frecvențe observate	Legea binomial negativă cu media 6,73 și parametrul k egal cu				Legea Poisson cu media 6,73
		0,41419	0,5	0,6	100	
0	20	27,7	23,7	20,0	0,1	0,1
1	12	10,8	11,0	11,0	0,8	0,7
2	14	7,2	7,7	7,7	2,7	2,4
3	7	5,5	6,0	6,5	5,7	5,5
4	3	4,4	4,9	5,3	9,3	9,2
5	6	3,6	4,1	4,5	12,2	12,4
6	3	3,1	3,5	3,9	13,5	13,9
7	3	2,7	3,0	3,3	12,9	13,3
8	2	2,3	2,6	2,9	10,9	11,2
9	3	2,1	2,3	2,6	8,2	8,4
10—14	4	7,4	8,2	9,0	13,0	12,5
15—19	2	4,5	4,8	5,1	0,5	0,4
20—	11	8,7	8,2	7,8	0,2	0,0
Total	90	90,0	90,0	90,0	90,0	90,0

reprezentînd repartiția numărului de ouă X ale unor paraziți specifici oilor. Se preconizează că legea teoretică este legea binomial negativă sau legea Poisson.

Se observă că

$$\overline{X} = 6{,}73 \text{ (numărul mediu de ouă).}$$

Ajustarea după legea Poisson cu $\lambda = 6{,}73$ nu este suficient de satisfăcătoare. Se încearcă ajustarea după legea binomial negativă folosind m.m., egalînd media și dispersia legii binomial negative cu valorile empirice corespunzătoare calculate pe baza eșantionului considerat de volum $n = 90$:

$$\frac{kq}{p} = 6{,}73$$

$$\frac{kq}{p^2} = 116{,}195.$$

De aici se obține :

$$p = 0{,}057948286$$

$$q = 0{,}942051714$$

$$k = 0{,}414186524.$$

Frecvențele ajustate $\{f_j\}$ sînt :

$$f_0 = 90 \cdot (0{,}057948286)^{0{,}414186524}(0{,}942051714)^0 = 27{,}7$$

$$f_1 = f_0 \cdot (0{,}414186\,524) \cdot (0{,}942051714)/1 = 10{,}8$$

$$f_2 = f_1 \cdot (1{,}414186524) \cdot (0{,}942051714)/2 = 7{,}2$$

$$f_3 = f_2 \cdot (2{,}414186524) \cdot (0{,}942051714)/3 = 5{,}5$$

Ținînd seama că :

$$C_{k-1}^0 = 1$$

$$p_{j+1} = \{(k + j)q/(j - 1)\}p_j .$$

Pentru ca concordanța dintre legea teoretică și cea empirică să fie mai bună se folosește aceeași valoare pentru media de selecție 6,73, dar valori mai mari pentru k (ceea ce implică valori mai mici ale dispersiei).
O ajustare convenabilă se obține pentru $k = 0{,}6$. *)

Deosebirea dintre legea Poisson și binomial negativă dispare practic pentru $k = 100$.

METODA MINIMULUI LUI χ^2. Metodă de obținere a estimatorilor unor parametri astfel încît să minimizeze *funcția* χ^2 ce depinde de observațiile unui eșantion considerat de volum n.

XVI. VERIFICAREA IPOTEZELOR STATISTICE

Ansamblu de metode statistice care permit ca pe baza rezultatelor unor experimente (datele de observație) să se precizeze legea de repartiție a variabilelor considerate din populația studiată, natura unui experiment (proces stochastic), estimarea parametrilor procesului, verificarea anumitor ipoteze statistice. Concluziile sînt de natură probabilistă, adică concordanța diferitelor ipoteze cu realitatea se face în limitele unui risc sau unei erori asociate lor.

În acest capitol sînt incluse noțiunile introductive și procedeele generale pentru testarea ipotezelor statistice. Va fi prezentat un număr de teste statistice referitoare la ipotezele statistice corespunzătoare legilor binomială și multinomială și tabelelor de contingență, testele de concordanță, ipotezele statistice asupra mediilor și dispersiilor populațiilor statistice, asupra corelației și formei legilor statistice.

---

*) Pentru testarea concordanței ajustării se poate folosi *testul de concordanță* χ^2 sau *Testul de concordanță Kolmogorov-Smirnov*.

VERIFICAREA (TESTAREA) IPOTEZELOR STATISTICE. Ansamblu de metode ale statisticii inductive, bazate pe calculul probabilităților, care permit, plecînd de la date experimentale sau de la observații, precizarea formei legilor de repartiție a v.a. referitoare la populațiile considerate sau validitatea anumitor supoziții referitoare la valorile parametrilor acestor legi.

IPOTEZĂ STATISTICĂ. Ipoteză asupra uneia sau mai multor legi ce caracterizează anumite populații, adică asupra valorilor parametrilor acestor legi sau asupra tipului legilor. De exemplu, dacă densitatea de probabilitate f a legii respective depinde de un parametru θ, se poate verifica ipoteza $H_0 : \theta = \theta_0$, conform căreia θ are valoarea θ_0.

Dacă în afară de valoarea θ_0 parametrul θ mai poate avea și una din valorile $\theta_1, \theta_2, \ldots$, atunci ipotezele :

$$H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1, \quad H_2 : \theta = \theta_2, \ldots$$

se numesc *ipoteze admisibile*.

Ipoteză nulă H_0. Ipoteza ce urmează a fi verificată și care se presupune apriori că este adevărată.

Ipoteză alternativă H_1. Orice altă *ipoteză admisibilă* cu care este confruntată i.n. H_0.

Ipoteză simplă. Ipoteza statistică de forma :

$$H_0 : \theta_1 = \theta_1^0, \ldots, \theta_k = \theta_k^0,$$

care specifică complet legea de repartiție sau valorile parametrilor ei.

Ipoteză compusă. Ipoteză statistică de forma : $H_0 : \theta \in \omega$, unde ω este o submulțime care nu se reduce la un punct a lui $\Omega = \{\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)', \textit{spațiul parametrilor}\}$, adică spațiul k-dimensional al tuturor vectorilor parametrilor $\boldsymbol{\theta}$. Submulțimea $\omega \subset \Omega$ conține cel puțin două puncte.

Ipoteză parametrică. Ipoteză statistică ce se referă la valorile parametrilor unei legi de repartiție specificate apriori.

Ipoteză neparametrică. Ipoteză statistică care nu se referă la forma legii de repartiție a v.a. (în engl. : *distribution free methods* sau *non parametric methods*).

Exemple. Se consideră :

1) $X \sim Bi(n, p)$, pentru care

$$H_0 : p < 0{,}2$$

este o ipoteză compusă.

2) $X \sim Bi(n, p)$ pentru care

$$H_0 : p = 0{,}05$$

este o ipoteză simplă (inclusă în cazul precedent).

3) $X \sim$ Normală, fără să se precizeze valorile parametrilor m și σ^2 (verificarea normalității), este o ipoteză neparametrică.

4) X este o v.a. a cărei lege de repartiție are mediana nulă — este ipoteză neparametrică (nu depinde de forma legii de repartiție) :

PRAG DE SEMNIFICAȚIE. Valoarea probabilității sub care se respinge i.n H_0. Se mai numește și *volumul testului*.

EROARE DE GENUL ÎNTÎI (I). Eroarea comisă la respingerea eronată a i.n. H_0, cînd aceasta este adevărată.

EROARE DE GENUL AL DOILEA (II). Eroarea comisă la acceptarea eronată a i.n. H_0 cînd aceasta nu este adevărată (cînd i.a. H_1 este adevărată).

RISCUL ERORILOR ASOCIATE UNUI TEST STATISTIC. Se asociază două riscuri unui test, mai exact deciziei luate.

Riscul erorii de genul I. Este probabilitatea de a respinge i.n. H_0 cînd aceasta este adevărată, adică :

$$P(H_1|\text{dacă } H_0 \text{ adevărată}) = \alpha.$$

Riscul α al erorii de genul întîi se numește *prag de semnificație.**

Riscul erorii de genul II. Este probabilitatea de a accepta i.n. H_0 cînd este adevărată i.n. H_1, adică :

$$P(H_0|\text{dacă } H_1 \text{ adevărată}) = \beta.$$

Schematic, **r.e.a.t.s.** se pot urmări în tabelul următor :

Ipoteza statistică reținută după aplicarea testului	Ipoteza adevărată H_0	Ipoteza adevărată H_1
H_0	adevărată $(1-\alpha)$	eroare de genul II Probabilitatea : β
H_1	eroare de genul I Probabilitatea : α	adevărată $(1-\beta)$

REGIUNE (INTERVAL) DE ACCEPTARE. Domeniul tuturor valorilor posibile ale statisticii t_n se împarte în două regiuni disjuncte. **R(i). a.** este mulțimea tuturor valorilor din *spațiul eșantioanelor* * astfel încît dacă o valoare a statisticii aparține acestei regiuni, se acceptă i.n. H_0. Dacă i.n. H_0 este adevărată, atunci $1-\alpha$ este probabilitatea ca valoarea statisticii să aparțină **r (i) a.**

REGIUNE (INTERVAL) DE RESPINGERE (*Regiune critică*). Mulțimea tuturor valorilor din *spațiul eșantioanelor* * astfel încît, dacă o valoare a statisticii aparține **r. (i). r.** se respinge i.n. H_0. Dacă i.n. H_0 este adevărată, atunci α este probabilitatea ca valoarea statisticii să aparțină **r. (i). r.** (fig. 109).

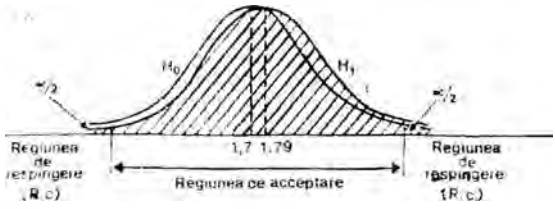


Fig. 109

PUTEREA UNUI TEST. Probabilitatea respingerii i.n. H_0 ca funcție de parametrul θ, notată cu $\pi(W, \theta)$, adică

$$\pi(W, \theta) = P(\mathbf{X} = (X_1, \ldots, X_n) \in W|\theta).$$

P.t. ia valoarea $1-\beta$ și este probabilitatea de a respinge i.n. H_0 cînd i.a. H_1 este adevărată, adică :

$$\pi(W, \theta_1) = 1 - \beta.$$

P.t. ia valoarea α și este probabilitatea de a respinge i.n. H_0, deși H_0 este adevărată, adică este p.s. α al testului:

$$\pi(W, \theta_0) = \alpha.$$

Graficul funcției $\pi(W, \theta)$ se numește *curba puterii testului*.

CARACTERISTICĂ OPERATIVĂ (*Funcția C.O.*). Probabilitatea de acceptare a i.n. H_0 ca funcție de parametrul θ, notată cu $L(W, \theta)$, adică:

$$L(W, \theta) = P(\mathbf{X} = (X_1, \ldots, X_n) \in CW/\theta).$$

C.O. ia valoarea $1-\alpha$ și este probabilitatea de a accepta i.n. H_0 cînd H_0 este adevărată, adică:

$$L(W, \theta_0) = P(\text{accept. } H_0/H_0 \text{ adev.}) = 1 - P(\text{resp. } H_0/H_0 \text{ adev.}) = 1-\alpha.$$

C.O. ia valoarea β și este probabilitatea de a accepta i.n. H_0, deși i.n. H_0 este falsă, adică:

$$L(W, \theta_1) = P(\text{accept. } H_0/H_1 \text{ adev.}) = P(\text{accept. } H_0/H_0 \text{ falsă}) = \beta,$$

deci în punctul $\theta = \theta_1$ funcția **C.O.** ia valoarea riscului de genul II, β.

Graficul funcției **C.O.** se numește *curba de eficacitate*.

FORȚA UNUI TEST. Dacă două teste statistice ale căror riscuri de genul I și riscuri de genul II sînt (α, β), respectiv (α', β'), verifică relațiile:

$$\alpha < \alpha', \quad \beta < \beta',$$

atunci se spune că primul test statistic are forță mai mare decît al doilea.

STATISTICĂ. Fie $(X_1, \ldots, X_n)$ un *eșantion* * de volum n din populația statistică considerată și fie v.a. X asociată populației a cărei lege de repartiție depinde de parametrul θ: $X \sim f(x:\theta)$.

O s. este o funcție $t(X_1, \ldots, X_n)$ de valorile observate ale eșantionului. Această funcție definește o v.a. a cărei lege de repartiție este utilizată pentru verificarea ipotezei nule și determinarea unui *interval de încredere** pentru parametrul θ. S. se mai numește și *funcție discriminată*. (→ *Statistică (m-dimensională)*). De exemplu, *media de selecție** a unui *eșantion** de volum n,

$$\bar{X} = \frac{X_1 + \ldots + X_n}{n},$$

este o s. a cărei lege de repartiție se cunoaște cînd i.n. H_0 este adevărată.

TEST STATISTIC. Criteriu pentru verificarea ipotezei statistice, constînd în calculul unei *statistici* * și stabilirea unei reguli fixate în prealabil de acceptare sau de respingere a i.n. cu o anumită probabilitate de a lua o decizie inexactă cînd i.n. H_0 este confruntată cu i.a. H_1.

În aplicarea unui **t.s.** se disting următoarele etape:

1. *Prezentarea datelor*: observațiile obținute într-un experiment statistic și elementele considerate apriori adevărate sau valabile, cum ar fi: independența, cunoașterea legii de repartiție, cunoașterea unora din parametrii acestor legi etc.
2. *Modelul statistic*: supozițiile fundamentale trebuie să fie în concordanță cu observațiile de care se dispune. Alegerea unui model neadecvat experimentului și observațiilor poate duce la concluzii greșite.
3. Definirea i.n. H_0 și i.a. H_1.
4. Alegerea p.s. α al testului.
5. Alegerea unei statistici, adică a v.a. a cărei lege de repartiție cunoscută să servească drept criteriu de decizie.
6. Stabilirea regiunii critice.
7. Luarea deciziei.

Sintetic, procedeul general de verificare a unei *ipoteze statistice**) constă în următoarele :

Procedeu general	*Exemplu*
1. Prezentarea datelor	În n experimente independente de aruncare a monedei, se obțin X „succese", adică X apariții ale feței „stema" cu $p = P$ (o apariție a feței „stema") $= 1/2$
2. Formularea modelului statistic	În experimentul de aruncare a unei monede de un număr $n = 1000$ de ori se urmărește $X =$ numărul de apariții ale feței „stema", adică realizările v.a. X ce urmează o lege binomială cu parametrii n și p (p este probabilitatea apariției feței „stema" la o aruncare, $p = 1/2$)
3. Definirea i.n. H_0 și i.a. H_1	Se verifică ipoteza conform căreia raportul dintre numărul de apariții ale feței „stema" și numărul de apariții ale feței „ban" este de 1 : 1. Se spune că moneda este echilibrată sau că este perfectă, adică se verifică i.n. : $H_0 : p = \frac{1}{2}$ față de i.a. $H_1 : p \neq \frac{1}{2}$
4. Alegerea p.s. α al t.s.	$\alpha = 0{,}05$
5. Alegerea statisticii testului — o v.a. T a cărei lege de repartiție este cunoscută	V.a. $T = \frac{X - np}{\sqrt{npq}}$ are aproximativ legea normală, cînd n este suficient de mare (⇒ *Legea numerelor mari*: *Legea binomială*.)
6. Stabilirea regiunii critice corespunzătoare pentru i.n. H_0 (dacă valoarea statisticii testului aparține regiunii critice, atunci se respinge i.n. H_0). Cu o probabilitate egală cu α, statistica testului aparține regiunii critice, cînd i.n. H_0 este adevărată (eroare de gradul I)	Utilizînd tabelele repartiției normale, regiunea critică este $\lvert T \rvert > z_{0{,}025} = 1{,}96$ (test bilateral)

*) În cadrul acestui capitol, principalele t.s. vor parcurge cele 7 etape ale procedeului general de verificare a ipotezei statistice.

7. Se ia decizia de a respinge i.n. dacă valoarea statisticii T aparține regiunii critice, adică $|T| > z_{0,025} = 1{,}96$. În caz contrar, dacă $|T| \leqslant z_{0,025}$ se acceptă i.n. H_0 (se spune că statistica testului este nesemnificativă și deci datele de observație nu contrazic ipoteza nulă).

I.n. H_0 conform căreia moneda este perfectă sau echilibrată este respinsă dacă numărul de apariții ale feței „stema" este fie prea mare, fie prea mic (→ *Test bilateral.*) Curba putere este dată de fig. 110.

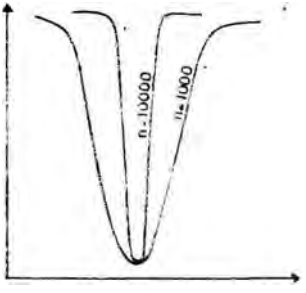


Fig. 110 Curba putere pentru n = *1000 și* n = *10000 pentru verificarea ipotezei* $H_0 : p = 1/2$ (*prag de semnificație* $\alpha = 0{,}05$)

Test de semnificație. T.s. pentru verificarea egalității unui parametru-estimat pe baza unui eșantion — cu o valoare dată. De exemplu, $X \sim N(m, \sigma^2)$ și fie $(X_1, \ldots, X_n)$ un eșantion de volum n; se presupune σ^2 — cunoscut. I.n. $H_0 : m = m_0$ (valoare dată) față de i.a. $H_1 : m \neq m_0$ se verifică cu ajutorul t.s. z

Test de concordanță. T.s. pentru verificarea concordanței între repartiția empirică și repartiția teoretică (modelul teoretic). *T.c.* arată calitatea ajustării repartiției teoretice prin repartiție empirică (în engl. : *goodness-of-fit*).
De exemplu, *testul* χ^2 *de concordanță* *, *testul Kolmogorov-Smirnov**.

Test unilateral (dreapta sau stînga). T.s. a cărui regiune (interval) critică este formată dintr-o singură *coadă a repartiției* * (din dreapta, respectiv din stînga) de arie α, egală cu p.s. al t.s. fixat apriori.

Exemplu. Fie $X \sim N(m, \sigma^2)$, σ^2-cunoscut și fie $(X_1, \ldots, X_n)$ un *eșantion de volum* n. Media de selecție** $\overline{X} \sim N(m, \sigma^2/n)$. Regiunile critice pentru i.n. $H_0 : m \leqslant m_0$ față de i.a. $H_1 : m > m_0$, respectiv pentru i.n. $H_0 : m \geqslant m_0$ față de i.a. $H_1 : m < m_0$, sînt date în tab. 62, iar testul folosit este *testul z unilateral drept*, respectiv *stîng*, unde z_α este determinat astfel încît :

$$P_{m_0}\left(z = \frac{\overline{X} - m_0}{\sigma/\sqrt{n}} > z_\alpha\right) = P_{m_0}(\overline{X} > C = m_0 + \sigma/\sqrt{n} \cdot z_\alpha) = \alpha$$

și $z_{1-\alpha} = -z_\alpha$.

Exemplu. Urmărind retribuția lunară a muncitorilor unei secții, să verificăm ipoteza că salariul mediu este de 1900 lei/lună, adică i.n. $H_0 : m = m_0 = 1900$ lei față de i.a. conform căreia retribuția acestor muncitori este mai mare decît 1900 lei ($H_1 : m > 1900$ lei). Avem de a face cu o ipoteză unilateral dreaptă și *t.u.d.* bazat pe un eșantion de volum $n = 100$ și un p.s. $\alpha = 0{,}01$ (cazul b). Experiența anterioară arată că $\sigma^2 = 10000$.

Funcția C.O. este reprezentată grafic în fig. 111, iar funcția putere în fig. 112 pe baza calculelor din tab. 63.

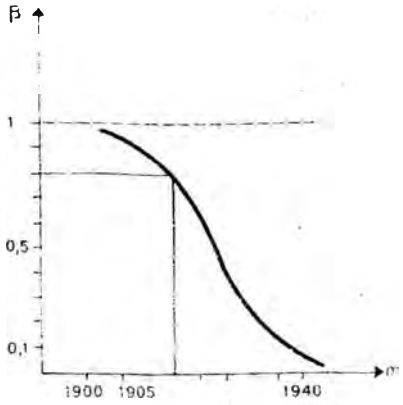


Fig. 111 Graficul caracteristicii operative (test z unilateral drept)

Dacă dorim să verificăm i.n. $H_0 : m = m_0 = 1900$ lei față de i.a. $H_1 : m < 1900$ lei avem de a face cu o ipoteză unilaterală stinga și t.s. folosit este *t.u.stg.z.* Exemplul va fi reluat în cadrul aplicării *testului z**.

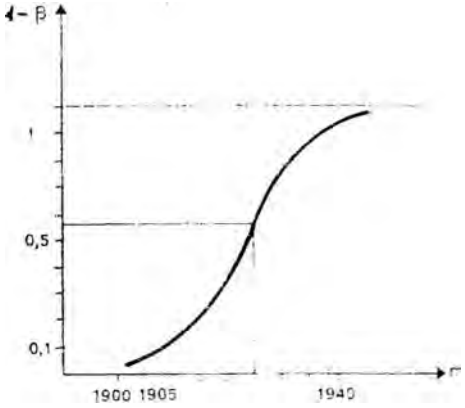


Fig. 112 Graficul puterii (test z unilateral drept)

Pe baza calculelor din tab. 64, funcția C.O. și funcția putere se reprezintă grafic în mod asemănător cu fig. 111, respectiv 112.

Test bilateral. T.s. a cărui regiune (interval) critică este formată din cele două *cozi ale repartiției **, situate de o parte și de alta a regiunii (intervalului) de

acceptare de arie α_1, respectiv α_2, astfel încît $\alpha_1 + \alpha_2 = \alpha$, unde α este p.s. al t.s. fixat apriori. În particular, dacă

$$\alpha_1 = \alpha_2 = \alpha/2,$$

atunci *t.b.* se numește *t.b. central*. (→ fig. 109).

Test robust. T.s. pentru care validitatea concluziilor este puțin afectată de nerespectarea ipotezelor fundamentale ale modelului matematic (normalitatea, dispersie constantă, independență etc.)

De exemplu, t.s. referitoare la mediile populațiilor sînt *t.r.* la încălcarea ușoară a normalității, spre deosebire de t.s. referitoare la dispersii, care nu sînt *t.r.*

Test neparametric. T.s. în cadrul căruia nu se fac presupuneri asupra formei repartiției.

De exemplu, nu se presupune normalitatea cînd nu există o anumită siguranță că este adevărată. Acesta ar fi un avantaj, dar *t.n.* este mai puțin puternic decît t.s. corespunzător bazat pe presupunerea normalității, cînd populația considerată este normală (în engl.: *non-parametric test* sau *distribution free test*). Exemple: *t.n. referitor la mediana populației, testul de concordanță Kolmogorov-Smirnov* *.

Test uniform cel mai puternic. T.s. pentru care puterea este maximă. Dacă β este considerat ca funcție de parametrul θ, $\beta(\theta)$, atunci trebuie găsit minimul funcției $\beta(\theta)$, considerînd toate valorile posibile ale lui θ în i.a. H_1.

Test al raportului de verosimilități. T.s. pentru verificarea i.n. H_0 față de i.a. H_1 bazat pe *raportul de verosimilități* * ale celor două ipoteze, avînd la dispoziție *eșantionul* $(X_1, \ldots, X_n)$ *de volum* n *:

$$\frac{L_0}{L_1} = \frac{L(X_1, \ldots, X_n; \theta_0)}{L(X_1, \ldots, X_n; \theta_1)} = \frac{f(X_1; \theta_0) \cdot f(X_2; \theta_0) \ldots f(X_n; \theta_0)}{f(X_1; \theta_1) \cdot f(X_2; \theta_1) \ldots f(X_n; \theta_1)},$$

unde $f(x, \theta)$ este d.p. a v.a. X_1.

Regiunea critică a i.n. H_0 este definită de condițiile:

$$L_0/L_1 < k < 1,$$

unde k este ales astfel încît probabilitatea de a respinge i.n. $H_0: \theta = \theta_0$, atunci cînd i.n. H_0 este adevărată (probabilitatea de respingere eronată a i.n.), este egală cu α — p.s. al t.s., valoare dată.

Test secvențial (progresiv) al lui Wald. T.s. în care unitățile experimentale sau observațiile sînt extrase succesiv, una cîte una (sau în grupe de același efectiv), din populația considerată. După fiecare nouă extragere se ia una din deciziile următoare: se acceptă i.n. H_0; se respinge i.n. H_0; se continuă testarea, adică se prelevă o nouă unitate experimentală sau un grup de unități experimentale.

T.s.W. se utilizează în *controlul statistic de recepție* *. Este un caz particular al *testului raportului de verosimilitate** : a fost introdus de A. Wald în 1947.

Test nedeplasat. T.s. a cărui regiune critică are proprietatea că pentru valorile posibile ale parametrului θ specificînd i.a. H_1 *funcția putere** are o valoare maximă pentru $\theta = \theta_0$, valoarea parametrului specificată de i.n. H_0.

Test de omogeneitate. T.s. pentru verificarea egalității valorilor a doi sau mai mulți parametri ai legilor de repartiție considerate, pentru care se obțin estimații independente. *T.o.* se mai numesc și *teste de comparație*.

Exemple

1) Într-un *tabel de contingență* se verifică ipoteza conform căreia fiecare linie reprezintă un eșantion dintr-o populație caracterizată de aceeași variabilă:

$$H_0 : p_1 = \ldots = p_k$$

față de i.a.

$$H_1 : p_i \neq p_j, \quad i \neq j \in \{1, \ldots, k\}.$$

2) *T.o.* a mediilor a I populații normale (→ *Analiză de varianță — Model unifactorial ANOVA**), adică i.n.

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_I$$

față de i.a.

$$H_1 : \alpha_i \neq \alpha_j, \quad \forall\, i, j \in \{1, \ldots, I\}.$$

3) *T.o.* a dispersiilor a k populații normale (→ *Testul lui Bartlett*). adică i.n.

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$$

față de i.a.

$$H_1 : \sigma_i^2 \neq \sigma_j^2, \ \forall\, i, j \in \{1, \ldots, k\}.$$

TEST χ^2. T.s. folosit pentru evaluarea discrepanțelor dintre frecvențele observate și cele teoretice, ceea ce permite să se decidă — cu un *prag de semnificație** fiixat — dacă aceste discrepanțe sînt fluctuații întîmplătoare sau nu, adică i.n. H_0 este acceptată sau respinsă.

T. χ^2 este folosit ca:

a) *Test de independență*: se verifică i.n. conform căreia două principii de clasificare aplicate în *tabele de contingență* sînt independente unul de altul.

b) *Test de omogenitate*: se verifică i.n. conform căreia o serie de observații grupate în mulțimi sau submulțimi sînt omogene în raport cu anumite caracteristici definitorii, adică observațiile relevă faptul că unitățile experimentale sînt extrase din aceeași populație de bază (de origine).

c) *Test de concordanță*: se verifică i.n. conform căreia observațiile eșantionului studiat avînd determinate *frecvențele observate* $*F_i$ sînt extrase dintr-o populație în care au anumite *frecvențe teoretice** f_i.

Deoarece expresia *t.* χ^2 este obținută pe baza observațiilor, rezultă că este o *statistică* * și deci nu este un parametru. Așa încît *t.* χ^2 mai este numit și t.s. *neparametric* sau *test liber de repartiție* (în engl. *distribution free test*)*, adică un test care nu depinde de forma legii inițiale de bază.

TEST *z*. T.s. * privind media unei populații statistice. T.s. bazat pe *legea* v.a.

$$z = \frac{\overline{X} - m}{\sigma/\sqrt{n}} \sim N(0,1) \ \text{(la limită)}$$

1) Fie $(X_1, \ldots, X_n)$ un *eșantion de volum n**, adică n observații independente provenind dintr-o populație cu *legea normală* * $N(m, \sigma^2)$ cu media m-parametru necunoscut și dispersia σ^2-parametru cunoscut.

---

* În cadrul acestui test se exemplifică toate cazurile a), b), c), corespunzătoare i.a. H_1. Pentru economie de spațiu, la alte t.s.-au exemplificat numai unele din aceste cazuri, a), b), sau c).

2) V.a. $X \sim N(m, \sigma^2)$.

3) a) $H_0 : m = m_0$ (m_0 — valoare fixată) (test bilateral — t.b.)*

$$H_1 : m \neq m_0$$

b) $H_0 : m \leq m_0$ (test unilateral dreapta — t.u.dr.)*

$$H_1 : m > m_0$$

c) $H_0 : m \geq m_0$ (test unilateral stinga — t.u.stg.)*

$$H_1 : m < m_0.$$

4) $\alpha = 0{,}05$, de exemplu.

5) Criteriu de decizie — statistica :

$$z = \frac{\overline{X} - m_0}{\sigma/\sqrt{n}}$$

este o v.a. $\sim N(0,1)$, unde

$$\overline{X} = \frac{X_1 + \ldots + X_n}{n} \quad \text{(media de selecție)*}$$

este o v.a. $N(m, \sigma^2/n)$.

6) *Regiunea critică*. a) Mulțimea valorilor statisticii z mai mici decît $\alpha/2$ — cuantila inferioară a legii $N(0, 1)$ sau mai mari decît $\alpha/2$ — cuantila superioară a legii $N(0,1)$:

$$|z| > z_{\alpha/2} \Leftrightarrow z < -z_{\alpha/2} = z_{1-\alpha/2},\ z > z_{\alpha/2}$$

sau

$$\overline{X} < m_0 - \sigma/\sqrt{n} \cdot z_{\alpha/2},\ \overline{X} > m_0 + \sigma/\sqrt{n} \cdot z_{\alpha/2},$$

unde

$$P(|z| > z_{\alpha/2}) = \Phi(z_{1-\alpha/2}) + 1 - \Phi(z_{\alpha/2}) = \alpha/2 + \alpha/2 = \alpha$$

cu

$$\Phi(t) = 1/\sqrt{2\pi} \int_{-\infty}^{t} e^{-x^2/2}\, dx \ ,\ \text{f.r. a legii } N(0,1);\ 1 - \Phi(-t) = \Phi(t)$$

Regiunea critică a testului bilateral z este ilustrată în fig. 109.

*Puterea** sub i.a. H_1 este :

$$\pi(W, m_1) = \Phi\left(\frac{m_1 - m_0}{\sigma/\sqrt{n}} + z_{1-\alpha/2}\right) + \Phi\left(\frac{m_0 - m_1}{\sigma/\sqrt{n}} + z_{1-\alpha/2}\right)$$

C.O.* sub i.a. H_1 este :

$$L(W, m_1) = \Phi\left(\frac{m_0 - m_1}{\sigma/\sqrt{n}} + z_{\alpha/2}\right) - \Phi\left(\frac{m_0 - m_1}{\sigma/\sqrt{n}} - z_{\alpha/2}\right)$$

b) Mulțimea valorilor statisticii z mai mari decît α-quantila superioară a legii $N(0,1)$:

$$z > z_\alpha$$

sau

$$\overline{X} > m_0 + \sigma/\sqrt{n} \cdot z_\alpha,$$

unde

$$P(z > z_\alpha) = 1 - \Phi(z_\alpha) = \alpha.$$

Puterea sub i.a. H_1 este:

$$\pi(W, m_1) = \Phi\left(\frac{m_1 - m_0}{\sigma/\sqrt{n}} + z_{1-\alpha}\right) = 1 - \Phi\left(\frac{m_0 - m_1}{\sigma/\sqrt{n}} + z_\alpha\right).$$

C.O. sub i.a. H_1 este:

$$L(W, m_1) = \beta = \Phi\left(\frac{m_0 - m_1}{\sigma/\sqrt{n}} + z_\alpha\right).$$

c) Mulțimea valorilor statisticii z mai mici decît α-quantila inferioară a legii $N(0,1)$:

$$z < z_{1-\alpha}$$

Tabelul 62

$H_0 : m = m_0$		test statistic
H_1 :	Regiunea critică	
$m \neq m_0$ a)	$\lvert z \rvert > z_{\alpha/2}$ sau $\overline{X} < m_0 - \sigma/\sqrt{n} \cdot z_{\alpha/2}$, $\overline{X} > m_0 + \sigma/\sqrt{n} \cdot z_{\alpha/2}$	*t.b.*
$m > m_0$ b)	$z > z_\alpha$ sau $\overline{X} > m_0 + \sigma/\sqrt{n} \cdot z_\alpha$	*t.u.dr.*
$m < m_0$ c)	$z < z_{1-\alpha}$ sau $\overline{X} < m_0 + \sigma/\sqrt{n} \cdot z_{1-\alpha} = m_0 - \sigma/\sqrt{n} \cdot z_\alpha$	*t.u.stg.*

sau

$$\overline{X} < m_0 + \sigma/\sqrt{n} \cdot z_{1-\alpha} = m_0 - \sigma/\sqrt{n} \cdot z_{\alpha},$$

unde

$$P(z < z_{1-\alpha}) = \Phi(z_{1-\alpha}) = \alpha\,;\ z_{1-\alpha} = -z_{\alpha}.$$

Puterea sub i.a. H_1 este :

$$\pi(W, m_1) = \Phi\left(\frac{m_0 - m_1}{\sigma/\sqrt{n}} + z_{1-\alpha}\right) = \Phi\left(\frac{m_0 - m_1}{\sigma/\sqrt{n}} - z_{\alpha}\right).$$

C.O. sub i.a. H_1 este :

$$L(W, m_1) = \beta = 1 - \Phi\left(\frac{m_0 - m_1}{\sigma/\sqrt{n}} - z_{\alpha}\right) = \Phi\left(\frac{m_1 - m_0}{\sigma/\sqrt{n}} + z_{\alpha}\right).$$

Pentru diferite i.a. H_1 ale i.n. H_0, regiunile critice sînt date de tab. 62.

7) Se calculează :

$\overline{X}$, *media de selecție*

$$z = \frac{\overline{X} - m_0}{\sigma/\sqrt{n}}.$$

Regula de decizie

a) Se acceptă H_0 (H_0 adevărată), dacă $|z| \leqslant z_{\alpha/2}$.

b) Se acceptă H_0 (H_0 adevărată), dacă $z \leqslant z_{\alpha}$.

c) Se acceptă H_0 (H_0 adevărată), dacă $z \geqslant z_{1-\alpha}$.

Exemplu

a) Să verificăm i.n. conform căreia media v.a. $X \sim N(m, 9)$ este 1,70 : (G. Ciucu, ș.a. : *Statistica matem. și cerc. oper.*, Ed. did. ped. București, 1974)

$$H_0 : m = 1{,}70$$

față de i.a. H_1 conform căreia media este diferită de 1,70 :

$$H_1 : m \neq 1{,}70$$

știind că $\sigma = 3$, pe baza a $n = 9$ observații independente.
Se alege p.s. al t.s. $\alpha = 0{,}05$.
Deoarece i.a. H_1 include o infinitate de alternative, se alege o singură alternativă, fie $H_1 : m = 1{,}79$. Așadar :

i.n. $H_0 : m = 1{,}70$

i.a. $H_1 : m_1 \neq 1{,}70$.

p.s. al t.s. $\alpha = 0{,}05$.

Statistica $\frac{(\overline{X}-1.70)}{3/\sqrt{9}}$ este o v.a. $N(0,1)$ cînd H_0 este adevărată, adică

$$P\left(\left|\frac{\overline{X}-1,70}{3/3}\right|<1,96\right)=0,95$$

sau

$$P\,(1,70-1,96.3/3<\overline{X}<1,70+1,96\cdot 3/3)=$$
$$=P\,(-0,26<\overline{X}<3,66)=1-0,05=0,95.$$

Intervalul

$$(-0,26,\ 3,66)$$

reprezintă mulțimea valorilor statisticii $\overline{X}$ care conduc la acceptarea i.n. H_0 deci regiunea de acceptare; aceasta, precum și regiunea critică sînt arătate în fig. 109.
Graficele puterii și C.O. sînt date de fig. 113, respectiv fig. 114.

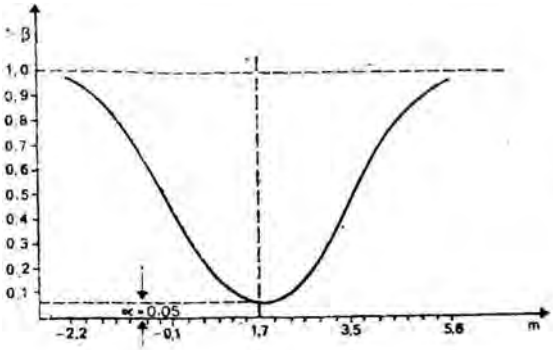


Fig. 113 Graficul puterii (test z bilateral)

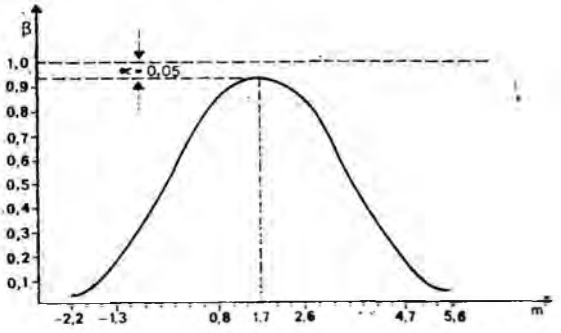


Fig. 114 Graficul caracteristicii operative (test z bilateral)

Curbele ce reprezintă d.p. ale v.a. X în cele două ipoteze, H_0 și H_1, sînt identice ca formă, curba reprezentînd d.p. sub H_1 fiind deplasată spre dreapta (fig. 109). Probabilitatea erorii de genul II în punctul $m = 1{,}79$:

$$\beta = P(\text{accept } H_0 \mid H_1 \text{ adevărată}) = P(-0{,}26 < \overline{X} < 3{,}66/H_1 \text{ adevărată}) =$$

$$= P\left(\frac{-0{,}26-1{,}79}{3/\sqrt{9}} < z < \frac{3{,}66-1{,}79}{3/\sqrt{9}}\right) =$$

$$= P(-2{,}05 < z < 1{,}87) = \Phi(1{,}87) - \Phi(-2{,}05) =$$

$$= 0{,}9693 - 0{,}0202 = 0{,}9491.$$

C.O. în punctul $m = 1{,}79$ (sau sub H_1) este $L(W, m_1) = 0{,}9491$.
Puterea sub H_1 este $\pi(W, m_1) = 1-\beta = 0{,}0509$.
b) Urmărind retribuția lunară a muncitorilor unei centrale industriale, să verificăm ipoteza că retribuția medie lunară este de 1900 lei, adică i.n. H_0 : $m_0 = 1900$ lei, față de ipoteza conform căreia retribuția medie lunară a acestor muncitori este mai mare decît 1900 lei, adică i.a. H_1 : $m > 1900$ lei pe baza unui eșantion de volum $n = 100$ și cu un p.s. $\alpha = 0{,}01$. Experiența anterioară arată că $\sigma^2 = 10\,000$. Avem de a face cu ipoteză unilaterală dreapta și un test unilateral drept z.
Funcția putere și C.O. sînt reprezentate grafic în fig. 112, respectiv fig. 111, pe baza calculelor din tab. 63.

Tabelul 63

m_1	$\frac{m_0-m_1}{\sigma/\sqrt{n}}$	$\Phi\left(\frac{m_0-m_1}{\sigma/\sqrt{n}} + z_\alpha\right)$
1905	$-0{,}5$	$\Phi(1{,}83) = 0{,}9664$
1910	$-1{,}0$	$\Phi(1{,}33) = 0{,}9082$
1915	$-1{,}5$	$\Phi(0{,}83) = 0{,}7967$
1920	$-2{,}0$	$\Phi(0{,}33) = 0{,}6293$
1925	$-2{,}5$	$\Phi(-0{,}17) = 0{,}4325$
1930	$-3{,}0$	$\Phi(-0{,}67) = 0{,}2514$
1933	$-3{,}5$	$\Phi(-1{,}17) = 0{,}1210$
1940	$-4{,}0$	$\Phi(-1{,}67) = 0{,}0475$

c) Dacă se verifică i.n. H_0 : $m = 1900$ lei față de i.a. H_1 : $m < 1900$ lei avem de a face cu o ipoteză unilaterală stîngă și un t.u. stg. z.
Puterea, respectiv funcția C.O. sînt reprezentate grafic asemănător cu fig. 112, respectiv 111, pe baza calculelor din tab. 64.

Observații

1. Graficele C.O. sub i.a. $H_1 : m < m_0$, respectiv $H_1 : m > m_0$ sînt simetrice față de dreapta $m = m_0$.
2. Graficele puterii sub i.a. H_1 menționate la punctul 1) sînt simetrice față de dreapta $m = m_0$.
3. În verificarea ipotezelor se urmărește să se facă puterea cît mai mare, adică i.a. H_1 să se accepte cu o probabilitate mare.

Tabelul 64

m_1	$\frac{m_0-m_1}{\sigma/\sqrt{n}}$	$\Phi\left(\frac{m_0-m_1}{\sigma/\sqrt{n}}-z_\alpha\right)$	β
1895	0,50	$\Phi(-1,83) = 0,0336$	0,9664
1890	1,00	$\Phi(-1,33) = 0,0918$	0,9082
1885	1,50	$\Phi(-0,83) = 0,2033$	0,7967
1880	2,00	$\Phi(-0,33) = 0,3707$	0,6293
1875	2,50	$\Phi(\ 0,17) = 0,5675$	0,4325
1870	3,00	$\Phi(\ 0,67) = 0,7486$	0,2514
1865	3,50	$\Phi(\ 1,17) = 0,8790$	0,1210
1860	4,00	$\Phi(\ 1,67) = 0,9525$	0,0475

4. Valoarea funcției putere sub i.a. H_1 poate fi folosită pentru a rezolva o problemă de *planificare a experimentului* *, și anume determinarea numărului de observații necesare pentru ca t.s. să aibă o anumită putere specificată $1-\beta$:

a) $n \approx (z_\beta + z_{\alpha/2})^2\ (\sigma/(m_1-m_0))^2$

b) și c) $n = (z_\beta + z_\alpha)^2\ (\sigma/(m_0 - m_1))^2$.

TEST t. T.s. privind media unei populații statistice, în care dispersia σ^2 este cunoscută.

T. t se bazează pe *legea de repartiție t Student** cu $n-1$ g.l. a v.a.

$$t = \frac{\overline{X} - m}{s/\sqrt{n}} \sim t_{n-1},$$

unde

$$\overline{X} = \frac{X_1 + \ldots + X_n}{n} \text{ media de selecție*}$$

$$s^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1} \text{ dispersia de selecție*.}$$

Se procedează ca pentru *testul z**, legea de repartiție *t* înlocuind legea normală $N(0,1)$, iar abaterea standard σ din statistica *z* fiind înlocuită prin *s*.

Observații

1. *T.t.* este puțin afectat de o depărtare modestă față de normalitate, adică este *test robust**.

2. Pentru diferitele i.a. H_1 ale i.n. H_0 regiunile critice sînt date în tab. 65, unde :

$t_{\alpha/2;\,n-1}$ este α-cuantila bilaterală a legii *t* cu $(n-1)$ g.l., adică

$$P(|t| > t_{\alpha/2;\,n-1}) = \alpha\,;$$

$t_{\alpha;n-1}$ este α-cuantila superioară a legii t cu $(n-1)$ g.l., adică

$$P(t>t_{\alpha;n-1}) = \alpha$$

și $t_{\alpha;n-1} = -t_{1-\alpha;n-1}$.

3. Pentru eșantioane de volum mic extrase prin populații normale, W. S. Gosset (cu pseudonimul Student) a stabilit repartiția exactă a v.a. $Z = (\overline{X}-m)/s$. Ulterior, R. A. Fisher a stabilit legea v.a. $t = (\overline{X}-m)/s\sqrt{n-1}$ cunoscută sub numele de *legea* t*.

Tabelul 65

$H_0 : m = m_0$		test statistic
H_1	Regiunea critică	
a) $m \neq m_0$	$\lvert t\rvert > t_{\alpha/2;\,(n-1)}$ sau $\overline{X} < m_0 - s/\sqrt{n}\,t_{\alpha/2;\,(n-1)}$ $\overline{X} > m_0 + s/\sqrt{n}\,t_{\alpha/2;\,(n-1)}$	t.b. t
b) $m > m_0$	$t > t_{\alpha;\,(n-1)}$ sau $\overline{X} > m_0 + s/\sqrt{n}\,t_{\alpha;\,(n-1)}$	t.u.dr. t
c) $m < m_0$	$t < t_{1-\alpha;\,n-1} = -t_{\alpha;\,(n-1)}$ sau $\overline{X} < m_0 - s/\sqrt{n}\,t_{\alpha;\,(n-1)}$	t.u.stg. t

TEST STATISTIC PRIVIND EGALITATEA A 2 MEDII. T.s. privind egalitatea mediilor a 2 populații normale $N(m_1, \sigma_1^2)$, respectiv $N(m_2, \sigma_2^2)$.

Test statistic privind egalitatea a 2 medii—dispersiile σ_1^2, σ_2^2 cunoscute (testul z),

1. Se consideră un eșantion de volum n_1, respectiv n_2 din cele 2 populații. Toate observațiile sînt independente.

2. $X_1 \sim N(m_1, \sigma_1^2)$, $X_2 \sim N(m_2, \sigma_2^2)$.

3. a) $H_0 : m_1 - m_2 = \delta_0$ (t.b. z)*

$$H_1 : m_1 - m_2 \neq \delta_0.$$

b) $H_0 : m_1 - m_2 \leqslant \delta_0$ (t.u. dr.)

$$H_1 : m_1 - m_2 > \delta_0.$$

c) $H_0 : m_1 - m_2 \geqslant \delta_0$ (t.u. stg.)

$$H_1 : m_1 - m_2 < \delta_0$$

4. $\alpha = 0.05$, de exemplu

5. $\overline{X}_1 \sim N(m_1, \sigma_1^2/n_1)$ **media de selecție** a primului eșantion.

$\overline{X}_2 \sim N(m_2, \sigma_2^2/n_2)$ **media de selecție** a celui de al doilea eșantion.

$\overline{X}_1$, $\overline{X}_2$ v.a. **independente.**

Se calculează:

$E(\overline{X}_1 - \overline{X}_2) = m_1 - m_2 = \delta_0$ **(valoare specificată)**

$\text{Var}\ (\overline{X}_1 - \overline{X}_2) = \sigma_1^2/n_1 + \sigma_2^2/n_2$.

Statistica:

$$z = \frac{\overline{X}_1 - \overline{X}_2 - (m_1 - m_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

este o v.a. $\sim N(0,1)$ **cînd i.n.** H_0 **este** adevărată.
6. Pentru diferitele **i.a. ale i.n.** H_0 : $m_1 - m_2 = \delta_0$, în tab. 66 sînt trecute regiunile critice.

Tabelul 66

i.n. $H_0 : m_1 - m_2 = \delta_0$		test statistic
i.a. H_1	Regiunea critică	
a) $m_1 - m_2 \neq \delta$.	$\lvert z \rvert > z_{\alpha/2}$ sau $\overline{X}_1 - \overline{X}_2 < \delta_0 - z_{\alpha/2} \cdot \text{Var}\,(\overline{X}_1 - \overline{X}_2)$, $\overline{X}_1 - \overline{X}_2 > \delta_0 + z_{\alpha/2} \cdot \text{Var}\,(\overline{X}_1 - \overline{X}_2)$	t.b.
b) $m_1 - m_2 > \delta_0$	$z > z_\alpha$ sau $\overline{X}_1 - \overline{X}_2 > \delta_0 + z_\alpha \cdot \text{Var}\,(\overline{X}_1 - \overline{X}_2)$	t.u.dr.
c) $m_1 - m_2 < \delta_0$	$z < z_{1-\alpha}$ sau $\overline{X}_1 - \overline{X}_2 < \delta_0 + z_{1-\alpha}\ \text{Var} \cdot (\overline{X}_1 - \overline{X}_2) = \delta_0 - z_\alpha \cdot \text{Var}\,(\overline{X}_1 - \overline{X}_2)$	t.u.stg.

Funcția putere sub i.a. H_1 **este:**

$$\text{a)}\quad \pi\ (W, \delta_1) = \Phi\left(-z_{\alpha/2} + \frac{\delta_1 - \delta_0}{\text{Var}\,(\overline{X}_1 - \overline{X}_2)}\right) + \Phi\left(-z_{\alpha/2} - \frac{\delta_1 - \delta_0}{\text{Var}\,(\overline{X}_1 - \overline{X}_2)}\right)$$

b) $$\pi(W, \delta_1) = \Phi\left(z_{1-\alpha} + \frac{\delta_1 - \delta_0}{\mathrm{Var}(\overline{X}_1 - \overline{X}_2)}\right)$$

c) $$\pi(W, \delta_1) = \Phi\left(z_{1-\alpha} + \frac{\delta_0 - \delta_1}{\mathrm{Var}(\overline{X}_1 - \overline{X}_2)}\right)$$

Exemple

a) Masa cărnii ambalate în pachete a 1000 gr de mașina M_1, respectiv M_2 este o v.a. X_1, respectiv X_2 de abatere medie pătratică 3 g, respectiv 4 g. Se presupune $X_1 \sim N(m_1, 9)$, respectiv $X_2 \sim N(m_2, 16)$. (G. Ciucu, ș.a., *Stat. mat. și cerc. oper*, Ed. did și ped., București, 1974)

Presupunem că în urma cîntăririi a 100 pachete de carne din producția fiecărei mașini s-au obținut mediile

$$\overline{X}_1 = 1007 \text{ g, respectiv } \overline{X}_2 = 1002 \text{ g.}$$

Să verificăm i.n.

$$H_0 : m_1 = m_2$$

față de i.a.

$$H_1 : m_1 \neq m_2$$

Se alege p.s. al t.s. $\alpha = 0{,}01$. Statistica

$$z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} = \frac{5}{\sqrt{9/100 + 16/100}} = 10$$

este o v.a. $\sim N(0,1)$ cînd i.n. H_0 este adevărată.

Deoarece statistica z este semnificativă:

$$|z| = 10 > z_{0{,}005} = 2{,}57,$$

se respinge i.n. H_0 cu p.s. de $\alpha = 0{,}01$.

b) Se verifică ipoteza că mediile repartițiilor v.a. ce dau suma calificativelor acordate studenților a două serii paralele ale aceluiași an de studiu diferă printr-o constantă dată.

Pentru cei 75 (80) studenți ai seriei 1 (2) ai unui an de studiu se obține $\overline{X}_1 = 43$ ($\overline{X}_2 = 40$). Studenții celor două serii sînt ascultați de aceleași cadre didactice, dispersiile fiind 100, respectiv 40. La un p.s. $\alpha = 0{,}01$, se verifică

$$\text{i.n. } H_0 : m_1 - m_2 = 2 \text{ ; i.a. } H_1 : m_1 - m_2 > 2.$$

Deoarece

$$\overline{X}_1 - \overline{X}_2 = 3 < \delta_0 + z_{0{,}01}\,\mathrm{Var}(\overline{X}_1 - \overline{X}_2) = 2 + (2{,}33)\cdot(1{,}3) = 5$$

se acceptă i.n. H_0.

Test statistic privind egalitatea a 2 medii- dispersiile σ_1^2, σ_2^2 necunoscute (testul t)

$\sigma_1^2 = \sigma_2^2$ (necunoscute) $(= \sigma^2)$.

TESTE PRIVIND VALOAREA SPECIFICATĂ A UNUI PARAMETRU (TESTE DE SEMNIFICAȚIE)

Ipoteza nulă H_0	Ipoteza alternativă H_1	Condiții	Statistica testului	Repartiția statisticii testului	Regiunea critică
1. $m = m_0$	$m \neq m_0$ $m > m_0$ $m < m_0$	σ-cunoscut	$z = \dfrac{(\overline{X} - m_0)}{\sigma/\sqrt{n}}$	Legea normală	$\lvert z\rvert > z_{\alpha/2}$ $z > z_{\alpha}$ $z < -z_{\alpha}$
2. $m = m_0$	$m \neq m_0$ $m > m_0$ $m < m_0$	σ-necunoscut	$t = \dfrac{(\overline{X} - m_0)}{s/\sqrt{n}}$	t-Student cu $(n-1)$ g.l.	$\lvert t\rvert > t_{\alpha/2;\,(n-1)}$ $t > t_{\alpha;\,(n-1)}$ $t < -t_{\alpha;\,(n-1)}$
3. $\sigma = \sigma_0$	$\sigma \neq \sigma_0$ $\sigma > \sigma_0$ $\sigma < \sigma_0$		$\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$	χ^2 cu $(n-1)$ g.l.	$\chi^2 < \chi^2_{1-\alpha/2;\,(n-1)}$ sau $\chi^2 > \chi^2_{\alpha/2;\,(n-1)}$ $\chi^2 > \chi^2_{\alpha;\,(n-1)}$ $\chi^2 < \chi^2_{1-\alpha;\,(n-1)}$
4. $p = p_0$	$p \neq p_0$ $p > p_0$ $p < p_0$	Eșantion mare (pentru eșantioane mici există teste exacte bazate pe tabelele legii binomiale)	$z = \dfrac{X/n - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ (Corecția de continuitate — valoarea lui z se micșorează dacă se înlocuiește X prin $X - 1/2$ sau $X + 1/2$)	Legea normală	$\lvert z\rvert > z_{\alpha/2}$ $z > z_{\alpha}$ $z < -z_{\alpha}$

1. Se consideră un eșantion de volum n_1, respectiv n_2 din cele 2 populații:

$$(X_{11}, \ldots, X_{n_1 1}), \text{ respectiv } (X_{12}, \ldots, X_{n_2 2})$$

Toate observațiile sînt independente.

2. $X_1 \sim N(m_1, \sigma_1)$; $X_2 \sim N(m_2, \sigma_2^2)$.

3. a) $H_0 : m_1 - m_2 = \delta$ (t.b.t)*

$$H_1 : m_1 - m_2 \neq \delta$$

b) $H_0 : m_1 - m_2 \leq \delta$ (t.u.dr.t)*

$$H_1 : m_1 - m_2 > \delta$$

c) $H_0 : m_1 - m_2 \geq \delta$ (t.u. stg. t)*

$$H_1 : m_1 - m_2 < \delta.$$

4. $\alpha = 0{,}05$, de exemplu.

5. Statistica

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (m_1 - m_2)}{\sqrt{(n_1 - 1)\, s_1^2 + (n_2 - 1) s_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

este o v.a. $\sim t_{n_1+n_2-2}$, cînd i.n. H_0 este adevărată:
$\overline{X}_1$, $\overline{X}_2$ sînt *mediile de selecție** ale primului, respectiv ale celui de al doilea eșantion;
s_1^2, s_2^2 sînt *estimațiile nedeplasate** ale lui σ_1^2, respectiv σ_2^2 — dispersiilor populațiilor din care sînt extrase cele două eșantioane.

6. Pentru diferite i.a. H_1 ale i.n. H_0 în tab. 68 sînt trecute regiunile critice.

Tabelul 68

i.n. $H_0 : m_1 - m_2 = \delta$	
i.a. H_1	Regiunea critică
a) $m_1 - m_2 \neq \delta$	$\lvert t \rvert > t_{n_1+n_2-2;\,\alpha/2}$
b) $m_1 - m_2 > \delta$	$t > t_{n_1+n_2-2;\,\alpha}$
c) $m_1 - m_2 < \delta$	$t < t_{n_1+n_2-2;\,1-\alpha} = -t_{n_1+n_2-2;\,\alpha}$

Exemplu

a) Datele din tabelul de mai jos reprezintă $X_1(X_2)$, nivelul colesterolului (mg/100 ml) pentru două eșantioane de volum $n_1 = 7$ masculi, respectiv

$n_2 = 6$ femele din două populații de animale (J. H. Pollard, *Numerical and Stat. Tech.*, Cambridge Univ. Press, 1977).

Masculi (X_1)	Femele (X_2)
226,5	221,5
224,1	230,2
218,6	223,4
220,1	224,3
228,8	230,8
229,6	223,8
222,5	

Să verificăm ipoteza că nivelul mediu (concentrația medie) a colesterolului este aceeași pentru masculi și femele, adică i.n.

$$H_0 : m_1 = m_2,$$

față de i.a.

$$H_1 : m_1 \neq m_2.$$

Se alege p.s. $\alpha = 0{,}05$.
Se calculează

$$\overline{X}_1 = 224{,}304$$

$$\overline{X}_2 = 225{,}667$$

$$s_1^2 = 17{,}765$$

$$s_2^2 = 14{,}951$$

Din aplicarea *testului lui Barlett** rezultă că se acceptă ipoteza cu privire la egalitatea dispersiilor.
Statistica testului (t.b.t)

$$|t| = \left| \frac{(224{,}314 - 225{,}667)}{\sqrt{6 \cdot 17{,}765 + 5 \cdot 14{,}951}} \cdot \sqrt{\frac{7 \cdot 6 \cdot 11}{13}} \right| = 0{,}599.$$

Deoarece valoarea statisticii t nu este semnificativă:

$$|t| = 0{,}599 < t_{11=(n_1+n_2-2);0{,}025} = 2{,}201,$$

se acceptă i.n. cu p.s. $\alpha = 0{,}05$, conform căreia concentrația medie a colesterolului pentru masculi și femele este aceeași.

c) Pentru două loturi de $n_1 = 23$, respectiv $n_2 = 23$ televizoare se măsoară X_1, respectiv X_2 — sensibilitatea limitată de amplitificare (măsurată în μV)

pe canalul 1, respectiv canalul 2 (televizor tip „Miraj") (E. Ursianu, *Verificarea ipotezelor statistice*, Preprint, Centrul de Stat. Matem., 1973).

Canalul 1 X_1	Canalul 2 X_2
30	24
25	19
22	18
35	38
60	44
30	24
28	30
40	39
22	15
27	34
40	25
50	42
16	20
20	26
32	30
30	19
16	20
12	19
25	30
12	15
21	12
19	18
27	20

Se verifică ipoteza că sensibilitatea medie a canalului 1 este mai bună decît cea a canalului 2, adică i.n.

$$H_0 : m_1 \geqslant m_2$$

față de i.a.

$$H_1 : m_1 < m_2.$$

Se calculează

$$\overline{X}_1 = 27{,}78$$
$$\overline{X}_2 = 25{,}26$$
$$s_1^2 = 136{,}4506$$
$$s_2^2 = 82{,}8379.$$

Se alege p.s. al t.s. : $\alpha = 0{,}05$
Statistica testului (t.u.stg.t):

$$t = \frac{(27{,}78 - 25{,}26)}{\sqrt{22 \cdot 134{,}4506 + 22 \cdot 82{,}8379}} \cdot \sqrt{\frac{23 \cdot 23 \cdot 44}{46}} = 0{,}8223.$$

Deoarece valoarea statisticii t nu este semnificativă:

$$t = 0{,}8223 > t_{0{,}95;\,44} = -t_{0{,}05;44} = -1{,}645,$$

se acceptă i.n. H_0 cu p.s. $\alpha = 0{,}05$, conform căreia canalul 1 al primului lot de televizoare are sensibilitatea mai bună decît canalul 2 al celui de al doilea lot.

$\sigma_1^2 \neq \sigma_2^2$ *(necunoscute) (test Behrens-Fisher).*

Se procedează ca la cazul precedent.

Statistica testului:

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (m_1 - m_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

este o v.a. aproximativ $\sim t_\nu$ sub i.n. H_0, unde

$$\frac{1}{\nu} = \frac{c^2}{n_1 - 1} + \frac{(1-c)^2}{n_2 - 1}, \quad c = \frac{\frac{s_1^2}{n_1}}{s_1^2/n_1 + s_2^2/n_2},$$

adică

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1}\right)^2 \frac{1}{n_1 - 1} + \left(\frac{s_2^2}{n_2}\right)^2 \frac{1}{n_2 - 1}} - 2$$

Exemplu

a) Se consideră un experiment analog cu cel din cazul precedent, pe două loturi de $n_1 = n_2 = 11$ televizoare tip „Saturn".

X_1	X_2
78	55
41	27
25	18
7	12
30	32
20	12
70	50
16	16
30	80
41	26
16	12

Se verifică i.n. $H_0: m_1 = m_2$ față de i.a. $H_1: m_1 \neq m_2$.

Se calculează:

$$\overline{X}_1 = 34$$
$$\overline{X}_2 = 31{,}81$$
$$s_1^2 = 501{,}6$$
$$s_2^2 = 484{,}96$$

Din aplicarea *testului lui Barlett** rezultă că ipoteza cu privire la egalitatea dispersiilor nu se verifică.

Statistica testului :

$$t = \frac{(34-31{,}81)}{\sqrt{\frac{501{,}6}{11} + \frac{484{,}96}{11}}} = 0.232$$

Se calculează :

$$c = 0{,}5084$$

$$\nu = 19{,}99, \quad \text{deci} \quad \nu = 20 \quad \text{g.l.}$$

Deoarece valoarea statisticii t nu este semnificativă :

$$|t| = 0{,}232 < t_{0{,}025}\,;\ \nu = 20 = 2{,}086,$$

se acceptă i.n. conform căreia sensibilitatea medie pe cele 2 canale este aceeași.

Observații

1. Dacă $\delta = 0$, **t.s.e. 2 m.** este un caz particular al *modelului unifactorial ANOVA** pentru $k = 2$.
2. **T.s.e. 2 m.** este t.s. *robust** la o depărtare moderată față de ipoteza de normalitate.
3. **T.s.e.2 m.** este t.s. *robust** la o încălcare moderată a ipotezei de egalitate a dispersiilor : $\sigma_1^2 = \sigma_2^2$, cînd volumele eșantioanelor n_1, n_2 sînt aproximativ egale.
4. Dacă $n_1 + n_2 - 2 > 30$, se poate folosi *aproximarea legii t prin legea normală.*
5. Cînd observațiile din cele 2 eșantioane sînt observații perechi se aplică *testul t pentru observații perechi* *.
6. Din cauza proprietății de simetrie a *legii t**, în cazul 3 a) se compară valoarea statisticii $|t|$ cu $\alpha/2$ — quantila superioară a legii t sau cu α — quantila bilaterală a legii t.

Observații perechi. T.s. pentru compararea a 2 populații bazat pe studiul diferențelor $d_i = X_{i1} - X_{i2}$ corespunzătoare observațiilor perechi (X_{i1}, X_{i2}), $i = 1, 2, \ldots, n$.

În fiecare pereche (X_{i1}, X_{i2}) factorii de variație ai v.a. X_1 și X_2 corespunzătoare celor 2 populații intervin în același mod, diferența semnificativă a mediilor fiind provocată de un factor străin a cărui acțiune vrem să o testăm.

Pentru a înlătura efectele factorilor de variație se consideră observații perechi, asigurîndu-se ca cei doi membrii ai perechii să fie similari din toate punctele de vedere, cu excepția celui pe care dorim să îl studiem.

De exemplu, într-un experiment ce constă în a verifica care din două tipuri de îngrășămînt (A sau B) este mai bun, în k întreprinderi agricole se plantează cîte două parcele. Din fiecare întreprindere se alege la întîmplare o parcelă pe care se aplică îngrășămîntul de tip A. Pe celelalte k parcele rămase se aplică îngrășămîntul de tip B. Dacă se compară recolta de grîu de pe cele două parcele, diferența poate rezulta și din faptul că grîul s-a dezvoltat în condiții de sol și climă diferite. Asigurînd condiții similare din punctul de vedere al solului și al climei, se testează acțiunea factorului îngrășămînt și se elimină factorii de variație sol și climă prin considerarea perechilor de observații de pe cele două parcele.

T.s. constă în a vedea dacă media celor n diferențe observate între recoltele celor două loturi diferă semnificativ de zero.

1. Se consideră un eșantion de volum n_1, respectiv n_2 din cele două populații :

$$(X_{11}, X_{21}, \ldots, X_{n1}), \text{ respectiv } (X_{21}, X_{22}, \ldots, X_{n2})$$

avînd mediile m_1, m_2 respectiv.

Fie (X_{i1}, X_{i2}) perechea de ordin i și fie n perechi de observații :

$$(X_{11}, X_{12}), \quad (X_{21}, X_{22}) \ldots (X_{n1}, X_{n2}).$$

2.—4. *(→ Test statistic privind egalit. 2 medii, dispersiile nec.)*

5. Statistica :

$$t = \frac{\bar{d} - (m_1 - m_2)}{\sqrt{\sum_{i=1}^{n} (d_i - \bar{d})^2}} \sqrt{n(n-1)} = \frac{\bar{d} - (m_1 - m_2)}{s_{\bar{d}}} \cdot \sqrt{n}$$

este o v.a. $\sim t_{n-1}$ sub i.n. H_0, unde

$$d_i = X_{i1} - X_{i2}$$

$$\bar{d} = \frac{\sum_i d_i}{n} \; ; \; s^2_{\bar{d}} = 1/(n-1) \sum_{i=1}^{n} (d_i - \bar{d})^2.$$

6. *(→ Test statistic privind egalitatea a 2 medii, dispersii nec.)*

Exemplu. Următoarele date dau efectele coroziunii în diferite terenuri asupra conductelor de oțel acoperite și neacoperite.

Tabelul 69

Acoperite (X_{i1})	42	47	51	52	59	54	32	47	66
Neacoperite (X_{i2})	40	43	50	44	57	56	44	37	70

Să vedem dacă acoperirea conductelor de oțel micșorează efectul coroziunii, adică i.n. $H_0 : m_1 - m_2 = 0$, față de i.a. $H_1 : m_1 - m_2 < 0$.

Se alege $\alpha = 0{,}01$.

Se calculează :

Tabelul 70

$d_i = X_{i1} - X_{i2}$	2	4	1	8	2	−2	−12	10	−4	$\bar{d} = 1$
$d_i - \bar{d}$	1	3	0	7	1	−3	−13	9	−5	
$(d_i - \bar{d})^2$	1	9	0	49	1	9	−169	81	25	

Statistica testului (t.u. stg.t) :

$$t = 0{,}457.$$

Deoarece valoarea statisticii t nu este semnificativă :

$$t = 0{,}457 > -t_{0{,}01\,;8} = t_{0{,}99\,;8} = -2{,}821,$$

se acceptă i.n. H_0.

Tabelul 71

Teste de semnificație (compararea a două populații)

Ipoteza nulă H_0	Ipoteza alternativă H_1	Condiții	Statistica testului	Legea de repartiție a statisticii testului	Regiunea critică	Test statistic
1	2	3	4	5	6	7
1. $m_1 = m_2$	$m_1 > m_2$ $m_1 < m_2$ $m_1 \neq m_2$	σ_1, σ_2 — cunoscute	$z = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$	$N(0,1)$	$z > z_\alpha$ $z < z_{1-\alpha} = -z_\alpha$ $\lvert z \rvert > z_{\alpha/2}$	t.s.u.dr.z t.s.u.stg.z t.s.b.z
2. $m_1 = m_2$	$m_1 > m_2$ $m_1 < m_2$ $m_1 \neq m_2$	$\sigma_1 = \sigma_2$ — necunoscute	$t = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$	t_{n-1}	$t > t_{n_1+n_2-2;\,\alpha}$ $t < -t_{n_1+n_2-2;\,\alpha}$ $\lvert t \rvert > t_{n_1+n_2-2;\,\alpha/2}$	t.s.u.dr.t t.s.u.stg.t t.s.b.t
3. $m_1 = m_2$	$m_1 > m_2$ $m_1 < m_2$ $m_1 \neq m_2$	$\sigma_1 \neq \sigma_2$ — necunoscute	$t = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$	t_ν	$t > t_{\nu;\,\alpha}$ $t < -t_{\nu;\,\alpha}$ $\lvert t \rvert > t_{\nu;\,\alpha/2}$ unde ν-g.l. este cel mai apropiat întreg pt. valoarea $\nu = -2 + \dfrac{(s_1^2/n_1 + s_2^2/n_2)^2}{\dfrac{(s_1^2/n_1)^2}{n_1+1} + \dfrac{(s_2^2/n_2)^2}{n_2+1}}$	t.s.u.dr.t t.s.u.stg.t t.s.b.t

Tabelul 71 (continuare)

1	2	3	4	5	6	7
4. $m_1 = m_2$	$m_1 > m_2$ $m_1 < m_2$ $m_1 \neq m_2$	Observații perechi	$t = \dfrac{\bar{d}\sqrt{n}}{s_{\bar{d}}}$ $d_i = X_{i1} - X_{i2}$; $\bar{d} = (\sum_i d_i)/n$ $s_{\bar{d}}^2 = 1/(n-1)\sum_i (d_i - \bar{d})^2$	t_{n-1}	$t > t_{n-1;\,\alpha}$ $t < -t_{n-1;\,\alpha}$ $\lvert t \rvert > t_{n-1;\,\alpha/2}$	t.s.u.dr.t t.s.u.stg.t t.s.b.t
5. $\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$ $\sigma_1^2 \neq \sigma_2^2$	Statistica testului bilateral F are la numărător cea mai mare dintre s_1^2, s_2^2.	$F = s_1^2/s_2^2$	F_{n_1-1,n_2-1}	$F > F_{\alpha;\,n_1-1,\,n_2-1}$ $F > F_{\alpha/2;\,n_1-1,\,n_2-1}$ $(s_1^2 > s_2^2)$ sau $F > F_{\alpha/2;\,n_2-1,\,n_1-1}$ $(s_1^2 < s_2^2)$	t.s.u.dr.F t.s.b.F
6. $p_1 = p_2$	$p_1 > p_2$ $p_1 < p_2$ $p_1 \neq p_2$		$z = \dfrac{x_1/n_1 - x_2/n_2}{\sqrt{\dfrac{x_1/n_1(1-x_1/n_1)}{n_1} + \dfrac{x_2/n_2(1-x_2/n_2)}{n_2}}}$ Corecția de continuitate — valoarea lui z se micșorează dacă se înlocuiește x prin $x^{-1/2}$ sau $x^{+1/2}$	$N(0,1)$	$z > z_\alpha$ $z < z_{1-\alpha} = -z_\alpha$ $\lvert z \rvert > z_{\alpha/2}$	t.s.u.dr.z t.s.u.stg.z t.s.b.z

Test neparametric Mann-Whitney-Wilcoxon. T.s. pentru verificarea egalității între 2 medii, cînd ipotezele fundamentale necesare pentru t.s.*t* nu sînt respectate (de exemplu, cele 2 populații nu sînt normale).
Se mai folosește atunci cînd datele sînt sub formă de ranguri ale unor *statistici de ordine* *.
Se verifică i.n. conform căreia cele două populații — din care provin eșantioanele cercetate — sînt identice :
1. Fie 2 eșantioane de volum n_1, respectiv n_2 din cele 2 populații și rangurile lor $\{r_{ij}\}$, $j = 2$; $i = 1, \ldots, n_1$, respectiv $i = 1, \ldots, n_2$. Datele de observație $(X_{11}, \ldots, X_{n_1 1}, X_{12}, \ldots, X_{n_2 2})$

	Eșantion 1 volum n_1	Eșantion 2 volum n_2	
Observația	Rangul	Observația	Rangul
X_{11}	r_{11}	X_{12}	r_{12}
X_{21}	r_{21}	X_{22}	r_{22}
.	.	.	.
.	.	.	.
.	.	.	.
$X_{n_1 1}$	$r_{n_1 1}$	$X_{n_2 2}$	$r_{n_2 2}$

se consideră ca un singur eșantion de volum $n = n_1 + n_2$. Se ordonează de la 1 la n, adică se consideră statisticile de ordine :

$$Y_{(1)} \quad Y_{(2)} \ldots Y_{(n)}.$$

2. Toate observațiile sînt independente și observațiile din interiorul unui eșantion provin din aceeași populație.
3. H_0 : cele două populații sînt identice.
H_1 : cele două populații nu sînt identice.
4. $\alpha = 0{,}05$, de exemplu :
5. Statistica *t.n. M.W.W.* :
Eșantioane de volum mic

$$W_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$W_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

$$W = \max(W_1, W_2) \quad \text{(statistica Mann-Whitney)},$$

unde

$$R_1 = \sum_{i=1}^{n_1} r_{i1} \quad \text{(statistica Wilcoxon)}$$

$$R_2 = \sum_{i=1}^{n_2} r_{i2}$$

Statistica W este o v.a. $\sim$ *legea Mann-Whitney* * de parametrii n_1, n_2.
Tabele cu α-cuantilele unilaterale, w_α respectiv $\alpha/2$-cuantilele bilaterale $w_{\alpha/2}$ ale legii Mann-Whitney sînt date în J. H. Pollard, *Numerical and Stat. Tech.*, Cambridge Univ. Press, 1977.
Eşantioane de volum mare
Dacă $n_1, n_2 \geqslant 8$, v.a. $W \sim$ aproximativ $N\left(\frac{n_1 n_2}{2}, \frac{n_1 n_2 (n_1+n_2+1)}{12}\right)$, (fig. 115).

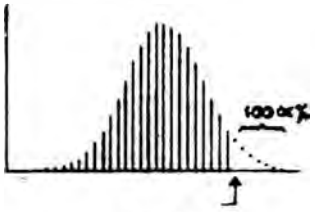


Fig. 115

Pentru $n_1, n_2 \to \infty$, statistica :

$$Z = \frac{W - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

este v.a. $\sim N(0,1)$; $z_{\alpha/2}$, z_α sînt $\alpha/2$, α-cuantilele bilaterală şi superioară $N(0,1)$ (fig. 116).

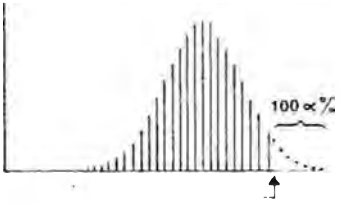


Fig. 116

6. Regiunea critică :
i) $|W| > w_{\alpha/2;\, n_1, n_2}$ (t.b.)
sau $W > w_{\alpha;\, n_1, n_2}$ (t.u.dr.)
ii) $|Z| > z_{\alpha/2}$ (t.b.)
sau
$Z > z_\alpha$ (t.u.dr.)

Exemple. (→ *Test stat. privind egalit. 2 medii, disp. nec.**, Ex.)
i) Se ordonează (în ordine descrescătoare) observaţiile.
Se alege p.s. $\alpha = 0{,}05$.

Se calculează :

$$R_1 = 54$$

$$R_2 = 37$$

$$W_1 = 16$$

$$W_2 = 26.$$

deci :

$$W = \max(W_1, W_2) = 26.$$

Valoarea statisticii nu este semnificativă : $W = 26 < 36 = w_{0,025;7,6}$ ($= 0,025$ — cuantila bilaterală a legii Mann-Whitney de parametri $n_1 = 7$, $n_2 = 6$), valoare obținută prin interpolare.
Se acceptă i.n.H_0 conform căreia mediile sînt aceleași.
ii) Operatorul $A(B)$ face $n_1 = 14$ ($n_2 = 11$) măsurători asupra unui obiect oarecare. Să verificăm dacă rezultatele obținute de acești doi operatori diferă semnificativ. Notăm măsurătorile făcute de operatorul A prin $(X_{11}, \ldots, X_{n_1 1})$, iar ale operatorului B, prin $(X_{12}, X_{22}, \ldots, X_{n_2 2})$. Dacă contopim măsurătorile și considerăm statisticile de ordine :

$$Y_{(1)}, Y_{(2)}, \ldots, Y_{(n_1+n_2)}$$

corespunzătoare, găsim că suma rangurilor statisticilor $Y_{(i)}$, avînd drept elemente măsurătorile operatorului A, este :

$$R_1 = 193.$$

Se alege p.s. $\alpha = 0,01$.
Se calculează :

Media : $$M(W) = \frac{n_1 n_2}{2} = 77.$$

Dispersia : $$\text{Var}(W) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12} = 333,66 ; \quad \sqrt{\text{Var } W} = 18,2.$$

Statistica Mann-Whitney-Wilcoxon are valoarea :

$$W = 14 \cdot 11 - 193 + \frac{14(14+1)}{2} = 66.$$

Statistica testului $Z = \dfrac{66 - 77}{18 \cdot 2} = -0,6.$

Deoarece statistica $Z = -0,6$ nu este semnificativă : $|Z| = 0,6 < 2,57 = z_{0,025}$, se acceptă i.n. conform căreia cei doi operatori au aceeași îndemînare.

Observații

1. *T.n. M-W-W.* a fost modificat sau extins pentru cazul mai multor eșantioane de volume egale de către Colin, Kruskal și Wallis, precum și de Sved și Eisenhart.

2. Mai general, se poate testa ipoteza că eșantioanele provin din aceleași populații identice, cu excepția mediilor lor, eventual.

Cu alte cuvinte, se verifică:

$$\text{i.n. } H_0 : m_1 - m_2 = \delta \text{ (constantă specificată)}$$

față de:

$$\text{i.a. } H_1 : m_1 - m_2 \neq \delta \quad \text{(t.b.)}$$

sau

$$\text{i.a. } H_1 : m_1 - m_2 > \delta \text{ (t.u.dr.)}$$

Se scade valoarea δ din fiecare observație, se ordonează, se determină rangurile observațiilor și se procedează ca în cazul tratat anterior.

3. *T.n.M.-W.-W.* nu se folosește pentru observații-perechi.

4. În cazul în care sînt îndeplinite condițiile din cadrul *testului privind egalitatea a două medii, dispersiile* σ_1^2, σ_2^2 *necunoscute**, se recomandă aplicarea *testului t** care este mai puternic decît *t.n. M.-W.-W.*

TEST STATISTIC PRIVIND EGALITATEA A k MEDII. T.s. privind egalitatea mediilor a k legi normale $N(m_1, \sigma^2), \ldots, N(m_k, \sigma^2)$ avînd dispersiile egale, adică t.s. de omogenitate a mediilor $m_1, \ldots, m_k$:

$$\text{i.n. } H_0 : m_1 = \ldots = m_k$$

față de i.a.

$$H_1 : m_i \neq m_j, \ \forall i, j \in \{1, \ldots, k\}.$$

T.s.e.km. se bazează pe raportul dintre estimația dispersiei mediilor a k eșantioane independente din k populații normale $N(m_i, \sigma^2)$, $i = 1, \ldots, k$ avînd aceeași dispersie σ^2, *dispersia între eșantioane sau între grupe (clase)* — și estimația dispersiei în interiorul eșantioanelor (grupelor), *dispersia reziduală*. (→ *Model unifactorial ANOVA.*)

TEST DE SEMNIFICAȚIE PENTRU DISPERSIE. T.s. privind egalitatea dispersiei σ^2 a unei populații normale — estimată pe baza unui eșantion de volum n — cu o valoare specificată σ_0^2.

1. Fie $(X_1, \ldots, X_n)$ un eșantion de volum n.

Observațiile sînt independente și provin dintr-o populație normală $N(m, \sigma^2)$

2. V. a. $X \sim N(m, \sigma^2)$.

3. a) $H_0 : \sigma^2 = \sigma_0^2$ (σ_0^2 — valoare fixată) (t.b.)

$H_1 : \sigma^2 \neq \sigma_0^2$

b) $H_0 : \sigma^2 \leq \sigma_0^2$ (t.u.dr.)

$H_1 : \sigma^2 > \sigma_0^2$

c) $H_0 : \sigma^2 \geq \sigma_0^2$ (t.u.stg.)

$H_1 : \sigma^2 < \sigma_0^2$.

4. $\alpha = 0{,}05$, de exemplu.

5. Se calculează

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}.$$

Statistica

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

este o v.a. $\sim \chi^2_{n-1}$, cînd i.n. H_0 este adevărată.

6. Pentru diferitele i.a. H_1 ale i.n. H_0 în tab. 72 sînt trecute regiunile critice.

Tabelul 72

i.n. $H_0 : \sigma^2 = \sigma_0^2$		Test statistic
i.a. H_1	Regiunea critică	
a) $\sigma^2 \neq \sigma_0^2$	$(n-1)\,s^2/\sigma_0^2 > \chi^2_{\alpha/2;\,(n-1)}$ sau $(n-1)\,s^2/\sigma_0^2 < \chi^2_{1-\alpha/2;\,(n-1)}$	t.b.
b) $\sigma^2 > \sigma_0^2$	$(n-1)s^2/\sigma_0^2 > \chi^2_{\alpha;\,(n-1)}$	t.u.dr.
c) $\sigma^2 < \sigma_0^2$	$(n-1)s^2/\sigma_0^2 < \chi^2_{1-\alpha;\,(n-1)}$	t.u.stg.

Observații

1. T.s. este afectat de o depărtare față de normalitate, adică nu este t.s. *robust*. Înainte de aplicarea **t.s. d.** este necesară *verificarea normalității* *.

Exemplu

a) Se consideră un eșantion de volum $n = 30$ arbori, care intră în componența demarorului D_2, prelucrați pe o mașină de rectificat, cărora li se măsoară diametrul, X_i, $i = 1, \ldots, 30$.

Să verificăm dacă abaterea standard a diametrelor este de 0,064 mm, adică

$$\text{i.n. } H_0 : \sigma^2 = \sigma_0^2 = 0{,}004096$$

față de

$$\text{i.a. } H_1 : \sigma^2 \neq \sigma_0^2 = 0{,}004096.$$

Se alege p.s. $\alpha = 0{,}05$.

Se calculează

$$\overline{X} = 15{,}96 \text{ mm}$$

$$s^2 = 0{,}0017.$$

Statistica

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = 12{,}3$$

este o v.a. $\sim \chi^2_{29}$ cînd i.n. H_0 este adevărată.
Deoarece statistica χ^2 nu este semnificativă :

$$\chi^2 = 12{,}3 < \chi^2_{0{,}025;29} = 45{,}72,$$

se acceptă i.n. H_0 cu p.s. $\alpha = 0{,}05$.
b) Timpul mediu de producere a unei reacții chimice este de 4,01 secunde și abaterea standard este $s = 0{,}07$ secunde.
Se consideră un eșantion de volum $n = 11$ privind timpii de reacție :

i	1	2	3	4	5	6	7	8	9	10	11
X_i	4,21	4,03	3,99	4,05	3,89	3,98	4,01	3,92	4,23	3,85	4,20

Să verificăm i.n.

$$H_0 : \sigma^2 \leqslant 0{,}0049$$

față de i.a. :

$$H_1 : \sigma^2 > 0{,}0049.$$

Se alege p.s. $\alpha = 0{,}05$.
Se calculează

$$\overline{X} = 4{,}033$$

$$s^2 = 0{,}017.$$

Statistica

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{10 \cdot 0{,}0170}{0{,}0049} = 34{,}7.$$

Deoarece statistica χ^2 este (mult) semnificativă :

$$\chi^2 = 34{,}7 > \chi^2_{0{,}05;10} = 18{,}31,$$

se respinge i.n. H_0 cu p.s. $\alpha = 0{,}05$.
Funcția putere sub i.a. $H_1 : \sigma^2 = 0{,}02$, de exemplu, are valoarea :

$$\pi(W, \sigma^2) = P\left(\chi^2_{10} > \frac{18{,}31 \cdot 0{,}0049}{0{,}02} = 3{,}4859\right) = 0{,}966.$$

ținînd seama că $1/\sigma_0^2 \sum_{i=1}^{11} (X_i - \overline{X})^2 \sim \chi^2_{n-1} = \chi^2_{10}$ cînd $\sigma^2 = \sigma_0^2$ este adevărata dispersie.

TEST STATISTIC PRIVIND EGALITATEA A DOUĂ DISPERSII. T.s. privind egalitatea dispersiilor a două populații normale $N(m_1, \sigma_1^2)$, respectiv $N(m_2, \sigma_2^2)$ din care se consideră două eșantioane.

Testul F al lui Snedecor—Fisher

1.—2. (→ *Test stat. privind egalit. 2 medii, dispersii necun.*)

3. a) $H_0 : \sigma_1^2 = \sigma_2^2$ (t.b.F)
   $H_1 : \sigma_2^2 \neq \sigma_2^2$
   b) $H_0 : \sigma_1^2 \leqslant \sigma_2^2$ (t.u.dr.F)
   $H_1 : \sigma_1^2 > \sigma_2^2$
   c) $H_0 : \sigma_1^2 \geqslant \sigma_2^2$ (t.u.stg.F)
   $H_1 : \sigma_1^2 < \sigma_2^2$

4. $\alpha = 0{,}05$, de exemplu.

5. Statistica :

$$F = \frac{s_1^2}{\sigma_1^2} \Big/ \frac{s_1^2}{\sigma_2^2}$$

este o v.a. $\sim F_{\nu_1,\nu_2}$ sub i.n. H_0, unde $\nu_1 = n_1 - 1$ și $\nu_2 = n_2 - 1$.

6. Pentru diferite i.a. H_1 ale i.n. H_0 în tab. 73 sînt trecute regiunile critice :

Tabelul 73

$H_0 : \sigma_1^2 = \sigma_2^2$		test statistic
H_1	Regiunea critică	
a) $\sigma_1^2 \neq \sigma_2^2$	$\dfrac{s^2_{\max}}{s^2_{1(2)}} > F_{\alpha/2;\, n_1-1,\, n_2-1}$	t.b.
b) $\sigma_1^2 > \sigma_2^2$	$\dfrac{s^2_{\max}}{s^2_{1(2)}} > F_{\alpha;\, n_1-1,\, n_2-1}$	t.u.dr.
c) $\sigma_1^2 < \sigma_2^2$	$\dfrac{s^2_{\max}}{s^2_{1(2)}} > F_{\alpha;\, n_2-1,\, n_1-1}$	t.u.stg.

unde:

$$s^2_{max} = \max(s_1^2,\ s_2^2),$$

$F_{\alpha/2;\nu_1,\nu_2}$ este $\alpha/2$-cuantila superioară a legii F (= α-cuantila bilaterală a legii F),

$F_{\alpha;\nu_1,\nu_2}$ este α-cuantila superioară a legii F.

Exemple

a) Se consideră datele experimentale de la *Exemplu c), Test stat. privind egalit. 2 medii, dispersiile* σ_1^2, σ_2^2 *necunoscute**.

Statistica testului:

$$F = s_1^2/s_2^2 = 1{,}62305,$$

Deoarece valoarea statisticii F nu este semnificativă:

$$F = 1{,}62305 < F_{0{,}025;\ 22,22} = 2{,}05.$$

se acceptă i.n.H_0 cu p.s. $\alpha = 0{,}05$.

c) Pe două strunguri se prelucrează aceeași piesă, variabila sub cercetare fiind diametrul piesei respective. Din producția primului (celui de al doilea) strung se extrage un eșantion de volum 16 (11). Să verificăm i.n. $H_0: \sigma_1^2 \geqslant \sigma_2^2$ față de i.a.

$$H_1: \sigma_1^2 < \sigma_2^2.$$

Se calculează

$$s_1^2 = 25.$$

$$s_2^2 = 20.$$

Statistica:

$$F = \frac{s_1^2}{s_2^2} = \frac{25}{20} = 1{,}25.$$

Deoarece statistica F nu este semnificativă:

$$F = 1{,}25 < F_{0{,}05;10,15} = 2{,}54,$$

se acceptă i.n. H_0 cu p.s. $\alpha = 0{,}05$.

Observații

1) În cazul a), t.b.F. dacă statistica F are valoarea mai mare decît 1, se alege $\nu_1 = n_1 - 1$ și $\nu_2 = n_2 - 1$; dacă statistica F are valoare mai mică decît 1, se folosește inversa acestei valori, luînd $\nu_1 = n_2 - 1$ și $\nu_2 = n_1 - 1$. (→ *Legea F*).

2) Mai general, pot fi verificate ipotezele statistice de forma:

a') i.n. $H_0: \sigma_1^2 = k\sigma_2^2$, k — constantă specificată

i.a. $H_1: \sigma_1^2 \neq k\sigma_2^2$.

Se procedează ca în cazul a) înlocuind $\sigma_1^2 = k\sigma_2^2$ în expresia statisticii F.

b') i.n. $H_0: \sigma_1^2 \leqslant k\sigma_2^2$.

i.a. $H_1: \sigma_1^2 > k\sigma_2^2$.

Se procedează ca în cazul b) înlocuind $\sigma_1^2 = k\sigma_2^2$ în expresia statisticii F.

Testul Z al lui Fisher. T.s. Z este un caz particular al t.s. F precedent introdus de R. Fisher și bazat pe legea v.a. Z a lui Fisher :

$$Z_{\nu_1, \nu_2} = \frac{1}{2} \ln \frac{s_1^2}{s_2^2}.$$

TEST STATISTIC PRIVIND EGALITATEA A k DISPERSII (k > 2). T.s. privind egalitatea dispersiilor a k populații normale $N(m_1, \sigma_1^2), \ldots, N(m_k, \sigma_k^2)$, indiferent dacă mediile m_i, $i = 1, \ldots, k$ sînt egale sau nu, adică t.s. de omogenitate a k dispersii : $\sigma_1^2, \ldots, \sigma_k^2$, estimate pe baza a k eșantioane din populațiile normale respective.

Textul lui Bartlett. T.s. pentru omogenitatea dispersiilor.

1. Se consideră k eșantioane de volume : $n_1, \ldots, n_k$ din k populații normale $N(m_1, \sigma_1^2), \ldots, N(m_k, \sigma_k^2)$, respectiv :

eșantionul 1 (volumul n_1)	eșantionul 2 (volumul n_2)	$\cdots$	eșantionul k (volumul n_k)
x_{11}	x_{12}		x_{1k}
x_{21}	x_{22}		x_{2k}
.	.		.
.	.		.
.	.		.
$x_{n_1 1}$	$x_{n_2 2}$		$x_{n_k k}$

Toate observațiile sînt independente.

2. $X_1 \sim N(m_1, \sigma_1^2), \ldots, X_k \sim N(m_k, \sigma_k^2)$.

3. i.n. $H_0 : \sigma_1^2 = \ldots = \sigma_k^2$

i.a. $H_1 : \sigma_i^2 \neq \sigma_j^2, \forall i, j \in \{1, \ldots, k\}, i \neq j$ (dispersiile nu sînt toate egale)

4. $\alpha = 0{,}05$, de exemplu.

5. Se calculează *mediile* și *dispersiile de selecție** ale celor k eșantioane : x_{i*} și s_i^2, $i = 1, \ldots, k$, adică :

$$x_{i*} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, \quad 1 \leqslant i \leqslant k$$

$$s_i^2 = \frac{\sum_{j=1}^{n_i} x_{ij}^2 - n_i x_{i*}^2}{n_i - 1}, \quad 1 \leqslant i \leqslant k$$

precum și

$$s^2 = \frac{1}{n-k} \sum_{i=1}^{k} (n_i - 1)s_i^2 = \overline{SS_e}(\rightarrow \textit{Model unifactorial ANOVA})$$

$$n = \sum_{i=1}^{k} n_i$$

$$C = 1 + \frac{1}{3(k-1)} \left[\sum_{i=1}^{k} \frac{1}{n_i - 1} - \frac{1}{\sum_{i=1}^{k} (n_i - 1)} \right].$$

Statistica :

$$\chi_B^2 = \frac{1}{C} \left[\sum_{i=1}^{k} (n_i - 1) \ln s^2 - \sum_{i=1}^{k} (n_i - 1) \ln s_i^2 \right]$$

sau

$$\chi_B^2 = \frac{2{,}306}{C} \left[\sum_{i=1}^{k} (n_i - 1) \log s^2 - \sum_{i=1}^{k} (n_i - 1) \log s_i^2 \right]$$

este o v.a. $\sim \chi_{k-1}^2$ cînd i.n. H_0 este adevărată.

6. Regiunea critică este :

$$\chi_B^2 > \chi_{\alpha;\, k-1}^2,$$

unde $\chi_{\alpha;\, k-1}^2$ este α-cuantila superioară a legii χ^2 cu $(k-1)$ g.l.

Exemplu. Pentru alcătuirea unor planuri de control statistic de recepție pentru fire tip relon s-a studiat stabilitatea procesului tehnologic relativ la caracteristica de calitate „alungire" a produsului respectiv, adică se verifică ipoteza statistică de omogenitate a dispersiilor eșantioanelor culese pe o perioadă de 11 luni, fiecare eșantion avînd volumul n_i, $i = 1, \ldots, 11$.

Se calculează :

Nr.	ν_i $(n_i - 1)$	s_i^2	s_i
1	151	6,8306	2,61
2	159	7,3437	2,70
3	159	8,9359	2,98
4	159	7,4673	2,73
5	151	6,5592	2,56
6	159	7,9302	2,81
7	127	6,4025	2,53
8	103	8,5293	2,92
9	167	6,9790	2,64
10	119	6,1055	2,47
11	103	9,4061	3,06

unde :

$$\nu = \sum_{i=1}^{11} \nu_i = \sum_{i=1}^{11} (n_i - 1) = 1557 ; \quad \sum_{i=1}^{11} \frac{1}{\nu_i} = 0{,}080081 ;$$

$$\nu \ln s^2 = 2{,}009193 ; \quad \sum_{i=1}^{11} \nu_i \ln s_i^2 = 3115{,}496553.$$

$$C = 1{,}0029$$

Valoarea statisticii $\chi_B^2 = 13{,}165788$ nu este semnificativă :

$$\chi_B^2 = 13{,}165788 < \chi^2_{0{,}05;\,10} = 18{,}3 ;$$

se acceptă i.n. cu p.s. $\alpha = 0{,}05$, adică dispersiile sînt omogene. Aşadar, se poate trage concluzia că procesul tehnologic este stabil pentru această caracteristică. *Observaţii.*

1. Verificarea egalităţii dispersiilor permite estimarea dispersiilor teoretice prin valoarea lor comună, adică estimaţia abaterii standard teoretice este :

$$\hat{\sigma} = \sqrt{\frac{\nu_1 s_1^2 + \ldots + \nu_k s_k^2}{\nu_1 + \ldots + \nu_k}} = \sqrt{\frac{151 \cdot 6{,}8306 + \ldots + 103 \cdot 9{,}4061}{151 + \ldots + 130}} = 2{,}73.$$

2. *T. B.* este un test aproximativ.

3. Pentru $k = 2$ se foloseşte *testul exact F.*

4. Cînd i.a. H_1 este adevărată, valoarea statisticii F este foarte mare. Valoarea statisticii F este mică atunci cînd dispersiile de selecţie sînt aproximativ egale.

5. Mai general, pot fi verificate ipotezele statistice de forma i.n. $H_0 : \sigma_1^2 = C\sigma_2^2$ $\sigma_2^2 = \sigma_3^2 = \ldots = \sigma_k^2$, unde C = constantă. În *t.B.* de mai sus se înlocuieşte s_1^2 prin s_1^2/C.

6. *T.B.* este afectat de nerespectarea ipotezei fundamentale privind normalitatea, adică nu este t.s. *robust.*

Înainte de aplicare *t.B.* este necesară *verificarea normalităţii* datelor eşantionului.

Testul lui Hartley. T.s. pentru omogenitatea dispersiilor în cazul în care avem de a face cu eşantioane de acelaşi volum. *T. H.* este bazat pe raportul dintre estimaţia dispersiei cea mai mare şi cea mai mică.

1.–5. (→ *Testul lui Bartlett*) cu condiţia

$$n_1 = n_2 = \ldots = n_k = n.$$

6. Statistica *t.H.* este :

$$F_{\max} = \frac{s^2_{\max}}{s^2_{\min}},$$

unde $s^2_{\max}$ ($s^2_{\min}$) este cea mai mare (mică) dispersie estimată pe baza datelor, eşantioanelor respective.

Statistica $t.H$ este o v.a. $\sim F_{k,n-1}$ cînd i.n. H_0 este adevărată.
Regiunea critică a $t.H.$ este :

$$F_{\max} > F_{\max;\alpha;k,\nu},$$

unde $F_{\max;\alpha;k,\nu}$ este α-cuantila superioară a legii $F_{\max}$ tabelate de H. O. Hartley pentru $\alpha = 0.05$ și $\alpha = 0,01$ cu k și $\nu = n-1$ g.l. și date în tab. 74.
Observație. Tabela valorilor aproximative ale α-cuantilelor superioare ale legii $F_{\max}$ au fost întocmite ținînd seama că :
V.a. $\ln s_i^2 \sim N(\ln \sigma^2, 2/(\nu-1))$ (aproximativ) ;

$$\ln \frac{s^2_{\max}}{s^2_{\min}} = \ln s^2_{\max} - \ln s^2_{\min} = R_{\ln s_i^2},$$

adică amplitudinea logaritmilor naturali ai estimațiilor dispersiilor s_i^2, $i = 1, \ldots, k$.
Așadar, quantilele aproximative ale legii $F_{\max}$ se calculează precum quantilele v.a. amplitudinea unui eșantion dintr-o populație normală $N(m, \sigma^2)$.
Exemplu. Pentru cele 12 canale ale televizorului Saturn se măsoară caracteristica de calitate X — sensibilitatea limitată de amplificare (măsurată în μV) pentru 12 eșantioane formate din cîte 11 aparate, adică $n_1 = \ldots = n_{12} = 11$ aparate :

Eșantionul 1	*Eșantionul 2*	...	*Eșantionul 12*
78	55		56
41	27		30
25	.		.
.	.		.
.	.		.
.	.		.
16	12		18

Se calculează

$x_{1*} = 34$ $s_1^2 = 501,60$ $x_{5*} = 33.09$ $s_5^2 = 263,49$ $x_{9*} = 29,45$ $s_9^2 = 335,47$

$x_{2*} = 31,81$ $s_2^2 = 484,96$ $x_{6*} = 25.54$ $s_6^2 = 218,67$ $x_{10*} = 26,09$ $s_{10}^2 = 448,09$

$x_{3*} = 31,54$ $s_3^2 = 361.47$ $x_{7*} = 25$ $s_7^2 = 227,80$ $x_{11*} = 27,63$ $s_{11}^2 = 481,25$

$x_{4*} = 30.27$ $s_4^2 = 310.01$ $x_{4*} = 26,90$ $s_8^2 = 220.09$ $x_{12*} = 25.45$ $s_{12}^2 = 248,27.$

Se observă că :

$$s^2_{\max} = s_1^2 = 501,60 \qquad k = 12$$

$$n = 11$$

$$s^2_{\min} = s_6^2 = 218,67 \qquad \nu = 11-1=10.$$

Se alege p.s. al $t.H.$ $\alpha = 0,05$. de exemplu.

0,05-cuantilele superioare ale legii v.a. $s^2_{max}/s^2_{min} = F_{max}$ cu k, ν g.l.

ν \ k	2	3	4	5	6	7	8	9	10
2	39,0	87,5	142	202	266	333	403	475	550
3	15,4	27,8	39,2	50,7	62,0	72,9	83,5	93,9	104
4	9,60	15,5	20,6	25,2	29,5	33,6	37,5	41,1	14,6
5	7,15	10,8	13,7	16,3	18,7	20,8	22,9	24,7	26,5
6	5,82	8,38	10,4	12,1	13,7	15,0	16,3	17,5	18,6
7	4,99	6,94	8,44	9,70	10,8	14,8	12,7	13,5	14,3
8	4,43	6,00	8,18	8,12	9,03	9,78	10,5	11,1	11,7
9	4,03	5,34	6,31	7,11	7,80	8,41	8,95	9,15	9,91
10	3,72	4,85	5,67	6,34	6,92	7,42	7,87	8,28	8,66
12	3,28	4,16	4,79	5,30	5,72	6,09	6,42	6,72	7,00
15	2,86	3,54	4,01	4,37	4,68	4,95	5,19	5,40	5,59
20	2,46	2,95	3,29	3,54	3,76	3,94	4,10	4,24	4,37
30	2,07	2,40	2,61	2,78	2,91	3,02	3,12	3,21	3,29
60	1,67	1,85	1,96	2,04	2,11	2,17	2,22	2,26	2,30
∞	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

0,01-cuantilele superioare ale legii v.a. $s^2_{max}/s^2_{min} = F_{max}$ cu k, ν g.l.

k \ ν	2	3	4	5	6	7	8	9	10	11	12
2	199	448	729	1036	1362	1705	2063	2432	2813	3204	3605
3	47,5	85	120	151	184	21(6)	24(9)	28(1)	31(0)	33(7)	36(1)
4	23,2	37	49	59	69	79	89	97	106	113	120
5	14,9	22	28	33	38	42	46	50	54	57	60
6	11,1	15,5	19,1	22	25	27	30	32	34	36	37
7	8,89	12,1	14,5	16,5	18,4	20	22	23	24	26	27
8	7,50	9,9	11,7	13,2	14,5	15,8	16,9	17,9	18,9	19,8	21
9	6,54	8,5	9,9	11,1	12,1	13,1	13,9	14,7	15,3	16,0	16,6
10	5,85	7,4	8,6	9,6	10,4	11,1	11,8	12,4	12,9	13,4	13,9
12	4,91	6,1	6,9	7,6	8,2	8,7	9,1	9,5	9,9	10,2	10,6
15	4,07	4,9	5,5	6,0	6,4	6,7	7,1	7,3	7,5	7,8	8,0
20	3,32	3,8	4,3	4,6	4,9	5,1	5,3	5,5	5,6	5,8	5,6
30	2,63	3,0	3,3	3,4	3,6	3,7	3,8	3,9	4,0	4,1	4,2
60	1,96	2,2	2,3	2,4	2,4	2,5	2,5	2,6	2,6	2,7	2,7
∞	1,00	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0

Sursa: *Handbook of Tables for Probability and Statistics*, The Chemical Rubber Co, Ohio, 1968.

Deoarece valoarea statisticii

$$F_{\max} = \frac{s^2_{\max}}{s^2_{\min}} = \frac{501{,}60}{218{,}67} = 2{,}3$$

nu este semnificativă:

$$F_{\max} = 2{,}3 < F_{\max;\, 0{,}05;\, 12,\, 10} = 9{,}34,$$

se acceptă i.n. H_0 cu p.s. $\alpha = 0{,}05$, adică dispersiile sînt omogene.

Testul lui Cochran. T.s. privind omogenitatea dispersiilor în cazul în care avem de a face cu eşantioane de volume egale: $n_1 = \ldots = n_k = n$.

T.C. se bazează pe raportul dintre cea mai mare estimaţie a dispersiei şi suma estimaţiilor dispersiilor celor k eşantioane, bazate pe n date.

1.–5. (→ *Testul lui Hartley.*)

6. Statistica *t.C.* este

$$Q = \frac{s^2_{\max}}{\sum_{i=1}^{k} s_i^2}.$$

Regiunea critică a *t.C.* este:

$$Q > Q_{\alpha;\, k,\, \nu},$$

unde $Q_{\alpha;\, k,\, \nu}$ este α-cuantila superioară a legii v.a. Q tabelată de Cochran pentru $\alpha = 0{,}05$ şi $0{,}01$ şi k, $\nu = n-1$ g.l. şi sînt date de tab. 75.

Exemplu. Pentru datele din exemplul precedent (→ *Testul lui Hartley.*) se observă că:

$$s^2_{\max} = s_1^2 = 501{,}60\,; \quad \sum_{i=1}^{n} s_i^2 = 4101{,}17.$$

Deoarece valoarea statisticii Q:

$$Q = 0{,}12 < Q_{0{,}05;\, 12,\, 10} = 0{,}2098$$

nu este semnificativă, se acceptă i.n. cu p.s. $\alpha = 0{,}05$ privind omogenitatea dispersiilor σ_i^2.

TESTE DE OMOGENEITATE GENERALĂ. T.s. ce permite compararea globală a 2 sau mai multe populaţii normale, adică testarea simultană a egalităţii mediilor, respectiv dispersiilor populaţiilor considerate.

T.o.g. se aplică pentru k eşantioane, toate de acelaşi volum n, provenind din k populaţii normale pentru a vedea dacă populaţiile nu sînt omogene, în sensul că au fie mediile diferite, fie dispersiile diferite. **T.o.g.** se mai numesc şi *teste Lambda Neyman-Pearson*

Testul Lambda Λ.

1. Se consideră k eşantioane de volume n:

$$(X_{i1}, \ldots, X_{in}), \quad i = 1, \ldots, k.$$

0,01-cuantilele superioare ale legii v.a. $Q = s^2_{\max} / \sum_{i=1}^{k} s_i^2$

k \ ν	2	3	4	5	6	7	8	9	10	11	17	37	145	∞
2	0.9999	0.9950	0.9794	0.9586	0.9373	0.9172	0.8988	0.8823	0.8674	0.8539	0.7949	0.7067	0.6062	0.5000
3	0.9933	0.9423	0.8831	0.8335	0.7933	0.7606	0.7335	0.7107	0.6912	0.6743	0.6059	0.5153	0.4230	0.3333
4	0.9676	0.8643	0.7814	0.7212	0.6761	0.6410	0.6129	0.5897	0.5702	0.5536	0.4884	0.4057	0.3251	0.2500
5	0.9279	0.7885	0.6957	0.6329	0.5875	0.5531	0.5259	0.5037	0.4854	0.4697	0.4094	0.3351	0.2644	0.2000
6	0.8828	0.7218	0.6258	0.5635	0.5195	0.4866	0.4608	0.4401	0.4229	0.4084	0.3529	0.2858	0.2229	0.1667
7	0.8376	0.6644	0.5685	0.5080	0.4659	0.4347	0.4105	0.3911	0.3751	0.3616	0.3105	0.2494	0.1929	0.1429
8	0.7945	0.6152	0.5209	0.4627	0.4226	0.3932	0.3704	0.3522	0.3373	0.3248	0.2779	0.2214	0.1700	0.1250
9	0.7544	0.5727	0.4810	0.4251	0.3870	0.3592	0.3378	0.3207	0.3067	0.2950	0.2514	0.1992	0.1521	0.1111
10	0.7175	0.5358	0.4469	0.3934	0.3572	0.3308	0.3106	0.2945	0.2813	0.2704	0.2297	0.1811	0.1376	0.1000
12	0.6528	0.4751	0.3919	0.3428	0.3099	0.2861	0.2680	0.2535	0.2419	0.2320	0.1961	0.1535	0.1157	0.0833
15	0.5747	0.4069	0.3317	0.2882	0.2593	0.2386	0.2228	0.2104	0.2002	0.1918	0.1612	0.1251	0.0934	0.0667
20	0.4799	0.3297	0.2654	0.2288	0.2048	0.1877	0.1748	0.1646	0.1567	0.1501	0.1248	0.0960	0.0709	0.0500
24	0.4247	0.2871	0.2295	0.1970	0.1759	0.1608	0.1495	0.1406	0.1338	0.1283	0.1060	0.0810	0.0595	0.0417
30	0.3632	0.2412	0.1913	0.1635	0.1454	0.1327	0.1232	0.1157	0.1100	0.1054	0.0867	0.0658	0.0480	0.0333
40	0.2940	0.1915	0.1508	0.1281	0.1135	0.1033	0.0957	0.0898	0.0853	0.0816	0.0668	0.0503	0.0363	0.0250
60	0.2151	0.1371	0.1069	0.0902	0.0796	0.0722	0.0668	0.0625	0.0594	0.0567	0.0461	0.0344	0.0245	0.0167
120	0.1225	0.0759	0.0585	0.0489	0.0429	0.0387	0.0357	0.0334	0.0316	0.0302	0.0242	0.0178	0.0125	0.0083
∞	0	0	0	0	0	0	0	0	0	0	0	0	0	0

0,05-cuantilele superioare ale legii v.a. $Q = s^2_{max} / \sum_{i=1}^{k} s_i^2$

k \ ν	2	3	4	5	6	7	8	9	10	11	17	37	145	∞
2	0.9985	0.9750	0.9392	0.9057	0.8772	0.8534	0.8332	0.8159	0.8010	0.7880	0.7341	0.6602	0.5813	0.5000
3	0.9669	0.8709	0.7977	0.7457	0.7071	0.6771	0.6530	0.6333	0.6167	0.6025	0.5466	0.4748	0.4031	0.3333
4	0.9065	0.7679	0.0841	0.6287	0.5895	0.5598	0.5365	0.5175	0.5017	0.4884	0.4366	0.3720	0.3093	0.2500
5	0.8412	0.6838	0.5981	0.5441	0.5065	0.4783	0.4564	0.4387	0.4241	0.4118	0.3645	0,3066	0.2513	0.2000
6	0.7808	0.6161	0.5321	0.4803	0.4447	0.4184	0.3980	0.3817	0.3682	0.3568	0.3135	0.2612	0.2119	0.1667
7	0.7271	0.5612	0.4800	0.4307	0.3974	0.3726	0.3535	0.3384	0.3259	0.3154	0.2756	0.2278	0.1833	0.1429
8	0.6798	0.5157	0.4377	0.3910	0.3595	0.3362	0.3185	0.3043	0.2926	0.2829	0.2462	0.2022	0.1616	0.1250
9	0.6385	0.4775	0.4027	0.3584	0.3286	0.3067	0.2901	0.2768	0.2659	0.2568	0.2226	0.1820	0.1446	0.1111
10	0.6020	0.4450	0.3733	0.3311	0.3029	0.2823	0.2666	0.2541	0.2439	0.2353	0.2032	0.1655	0.1308	0.1000
12	0.5410	0.3924	0.3264	0.2880	0.2624	0.2439	0.2299	0.2187	0.2098	0.2020	0.1737	0.1403	0.1100	0.0833
15	0.4709	0.3346	0.2758	0.2419	0.2195	0.2034	0.1911	0.1815	0.1736	0.1671	0.1429	0.1144	0.0889	0.0667
20	0.3894	0.2705	0.2205	0.1921	0.1735	0.1602	0.1501	0.1422	0.1357	0.1303	0.1108	0.0879	0.0675	0.0500
24	0.3434	0.2354	0.1907	0.1656	0.1493	0.1374	0.1286	0.1216	0.1160	0.1113	0.0942	0.0743	0.0567	0.0417
30	0.2929	0.1980	0.1593	0.1377	0.1237	0.1137	0.1061	0.1002	0.0958	0.0921	0.0771	0.0604	0.0457	0.0333
40	0.1370	0.1576	0.1259	0.1082	0.0968	0.0887	0.0827	0.0780	0.0745	0.0713	0.0595	0.0462	0.0347	0.0250
60	0.1737	0.1131	0.0895	0.0765	0.0682	0.0623	0.0583	0.0552	0.0520	0.0497	0.0411	0.0316	0.0234	0.0163
120	0.0998	0.0632	0.0495	0.0419	0.0371	0.0337	0.0312	0.0292	0.0279	0.0266	0.0218	0.0165	0.0120	0.0087
α	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Sursa : *Handbook of Tables for Probability and Statistics*, The Chemical Ruber Co., Ohio, 1968.

Teste observațiile sînt independente.

2. $X_i \sim N(m_i, \sigma_i^2)$, $i = 1, \ldots, k$.

3. i.n. $H_0 : m_1 = \ldots = m_k$
$\sigma_1^2 = \ldots = \sigma_k^2$
i.a. $H_1 : m_i \neq m_j \quad \forall i, j \in \{1, \ldots, k\}$.
$\sigma_i^2 \neq \sigma_j^2$

4. $\alpha = 0{,}05$, de exemplu.

5. Statistica *t.L.* :

$$\Lambda_0 = \frac{s_0^2}{(s_1^2 s_2^2 \ldots s_k^2)^{1/k}},$$

unde

$$s_0^2 = \frac{1}{k \cdot n} \sum_{i=1}^{k} \sum_{j=1}^{n} (X_{ij} - \overline{X})^2 = \frac{\sum_i \sum_j X_{ij}^2 - (\sum_i \sum_j X_{ij}^2)/k \cdot n}{k \cdot n}$$

$$s_i^2 = \frac{1}{n} \sum_{j=1}^{n} (X_{ij} - \overline{X}_i)^2 = \frac{\sum_j X_{ij}^2 - (\sum_j X_{ij})^2/n}{n}, \quad i = 1, \ldots, k$$

$$\overline{X} = \frac{1}{k \cdot n} \sum_i \sum_j X_{ij}$$

$$\overline{X}_i = \frac{1}{n} \sum_j X_{ij}, \quad i = 1, \ldots, k$$

Pentru p.s. $\alpha = 0{,}05$ și diferite valori ale k, n, $\Lambda_{0;\alpha;k;n}$ α-cuantilele legii v.a. Λ_0 au fost date de D. J. Cowden, în 1957 (tab. 76).

6. Regiune critică a *t.L.* :

$$\Lambda_0 > \Lambda_{0;\alpha;k;n}$$

Observație. Dacă s_i^2, $i = 1, \ldots, k$ sînt *estimațiile nedeplasate** ale dispersiilor σ_i^2 ale populațiilor din care sînt extrase cele k eșantioane, atunci statistica *t. L.* devine :

$$s_i^2 = \frac{\sum_j (X_{ij} - \overline{X}_i)^2}{n-1}, \quad i = 1, \ldots, k$$

$$\Lambda_0 = \frac{s_0^2}{\frac{n-1}{n}(s_1^2 \ldots s_k^2)^{1/k}}.$$

Tabelul 76

0,05 — cuantilele superioare ale legii Λ (k-numărul populațiilor; n — numărul de observații din fiecare populație)

k	n											
	2	3	4	5	6	7	8	9	10	12	15	20
2	52.6	5.82	3.05	2.26	1.99	1.78	1.62	1.51	1.41	1.33	1.25	1.17
3	46.1	5.73	3.12	2.33	2.06	1.84	1.68	1.54	1.43	1.35	1.26	1.18
4	38.8	5.42	3.00	2.26	2.01	1.81	1.66	1.54	1.43	1.35	1.26	1.18
5	33.2	5.14	2.92	2.22	1.97	1.78	1.62	1.50	1.42	1.34	1.25	1.18
10	21.3	4.36	2.64	2.06	1.84	1.66	1.54	1.44	1.38	1.31	1.23	1.16
20	15.2	3.79	2.41	1.93	1.74	1.58	1.41	1.38	1.32	1.26	1.20	1.14
25	13.9	3.64	2.38	1.87	1.68	1.54	1.42	1.36	1.31	1.25	1.19	1.13
50	11.4	3.29	2.15	1.74	1.59	1.48	1.38	1.32	1.27	1.21	1.17	1.12

Sursa: D. J. Cowden, *Statistical Methods in Quality Control*, New York, 1957, p. 687.

Exemplu. Pentru studiul omogeneității generale a unui proces tehnologic, se consideră reperele produse într-o secție de prelucrări mecanice prin așchiere pe 10 strunguri, alese la întîmplare și se măsoară greutățile a cîte 5 piese, de la fiecare strung, pe parcursul unui schimb.
În condițiile de funcționare normale a procesului tehnologic, greutățile pieselor prelucrate pe toate strungurile secției trebuie să fie aceleași.
Se verifică ipoteza conform căreia variabilitatea procesului tehnologic este constantă.
Se calculează $\overline{X}_i$ — *mediile de selecție** și s_i^2 — *estimațiile nedeplasate** ale dispersiilor σ_i^2 pentru cele $k = 10$ eșantioane de cîte $n = 5$ piese:

$$s_1^2 = 0{,}68 \qquad s_6^2 = 1{,}05$$

$$s_2^2 = 2{,}05 \qquad s_7^2 = 3{,}07$$

$$s_3^2 = 1{,}01 \qquad s_8^2 = 1{,}01$$

$$s_4^2 = 1{,}37 \qquad s_9^2 = 1{,}22$$

$$s_5^2 = 2{,}70 \qquad s_{10}^2 = 3{,}32$$

precum și

$$s_0^2 = 3{,}12.$$

Statistica testului:

$$\Lambda_0 = \frac{3.12}{\frac{4}{5}(0{,}68 \cdot 2{,}05 \cdot 1{,}01 \cdot 1{,}37 \ldots 3{,}32)^{1/10}} = 2{,}55.$$

Deoarece valoarea statistică $\Lambda_0 = 2{,}55$ este semnificativă:

$$\Lambda_0 = 2{,}55 > \Lambda_{0{,}05;\ 10;\ 5} = 2{,}06,$$

se respinge i.n. de omogeneitate generală.
Observații.
1) Testul Λ este folosit ca procedeu preliminar de verificare a omogeneității procesului.
Valoarea semnificativă a statisticii Λ_0 impune verificarea i.n. privind *egalitatea mediilor* și i.n. privind *egalitatea dispersiilor.*
Dacă ulterior se verifică omogeneitatea dispersiilor, de exemplu, rezultă că variabilitatea procesului se datorează numai variațiilor semnificative între medii.
2) În cazul în care nu există siguranță prea mare că procesul se desfășoară normal, adică există dubii în ceea ce privește legea normală a v.a. X, valoarea semnificativă a statisticii Λ_0 poate indica nerespectarea ipotezei de normalitate, mai degrabă decît neomogeneitatea mediilor și dispersiilor.
3) Așadar, testul Λ este afectat de nerespectarea ipotezei de normalitate, deci nu este t.s. *robust.*

Teste L Neyman-Pearson. T.o.g. introduse de J. Neyman și K. Pearson în 1933.

În condițiile anterioare (→ *Testul Lambda* Λ) se stabilește *Testul* L_0. **T.o.g.** pentru verificarea ipotezei conform căreia populațiile statistice considerate au aceleași medii și aceleași dispersii.
Statistica *t.* L_0 :

$$L_0 = \frac{s_g^2}{s_0^2},$$

unde :

$$s_g^2 = (s_1^2 \cdot s_2^2 \cdot \ldots \cdot s_k^2)^{1/2},$$

adică media geometrică a celor k dispersii de selecție, s_i^2, $i = 1, \ldots, k$.
Tabelele cuantilelor legii L_0 au fost calculate de Mahalanobis și publicate în „Sankhyá". 1963.
Testul L_1. **T.o.g.** pentru verificarea ipotezei conform căreia populațiile au aceleași dispersii, oricare ar fi mediile lor.
Statistica *t.* L_1 :

$$L_1 = \frac{s_g^2}{s_a^2},$$

unde

$$s_a^2 = \frac{1}{k} \sum_{i=1}^{k} s_i^2,$$

adică media aritmetică a celor k dispersii de selecție, s_i^2, $i = 1, \ldots, k$.
Observație. În *controlul statistic al calității**. **t.o.g.** sînt folosite pentru a arăta că un proces urmînd *legea normală* * este *sub control* sau este *controlabil*. Față de *fișa de control* folosită în controlul statistic al calității, **t.o.g.** prezintă un dezavantaj, și anume : dau numai o idee asupra variabilității generale a procesului, fără a putea preciza cauza și fără a avea o evidență grafică a acestei variabilități (pentru medii, respectiv dispersii) ca în cazul fișelor de control.
În plus, în prezent se dispune încă de tabelări „sărace", lipsind, de exemplu, cuantilele corespunzătoare lui $\alpha = 0.001$, valoare folosită curent în controlul calității produselor.
TESTE DE CONCORDANȚĂ. T.s. pentru verificarea concordanței ajustării unei repartiții teoretice pe baza datelor de observație ale unui eșantion. **T.e.** se aplică pentru verificarea normalității, exponențialității, caracterului Poisson, caracterul Weibull etc.
Test de concordanță χ^2 *(caz parametric).* **T.c.** pentru a stabili dacă un eșantion provine dintr-o populație cu lege specificată depinzînd de s parametri necunoscuți, $\theta_1, \ldots, \theta_s$ care nu au valori predeterminate și urmează să fie estimați pe baza datelor eșantionului considerat.
1. Se consideră un eșantion de volum n. Toate observațiile sînt independente. Se alcătuiește o *tabelă de frecvență*, prin împărțirea valorilor eșantionului în k clase, în fiecare înregistrîndu-se $O_i (= n_i)$ observații $\left(i = 1, \ldots, k; \sum_{i=1}^{k} O_i = \sum_{i=1}^{k} n_i = n \right)$, astfel încît $n_i \geqslant 5$:

	Numărul de observații în clasa 1 2 ..... k	Total
Numărul observat F_i (frecvențe observate)	O_1 O_2 ..... O_k	
Numărul estimat f_i (frecvențe teoretice)	e_1 e_2 ..... e_k	

$$e_i = n\hat{p}_i, \quad i = 1, \ldots, k.$$

Fiecare observație are probabilitatea p_i de a cădea în clasa i.
2. V. a. k-dimensională $(n_1, \ldots, n_k)$ sau $(O_1, \ldots, O_k)$. (→ *Legea multinominală**.)
3. $H_0 : p_i = \bar{p}_i(\theta_1, \theta_2, \ldots, \theta_s)$, $i = 1, \ldots, k$,
adică

$$H_0 : p_i = e_i/n, \quad i = 1, \ldots, k$$

față de i.a.

H_1 : i.n. H_0 nu este adevărată.
4. Se alege p. s. al testului statistic $\alpha = 0{,}05$, de exemplu.
5. Statistica :

$$\chi^2 = \frac{(O_1 - e_1)^2}{e_1} + \ldots + \frac{(O_k - e_k)^2}{e_k},$$

adică :

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$$

urmează aproximativ legea χ^2 cu $\nu = k - s - 1$ g.l., unde $\hat{p}_i = p_i(\hat{\theta}_1, \ldots, \hat{\theta}_s)$ sînt estimațiile probabilităților teoretice p_i, $\hat{\theta}_i$ fiind *estimațiile prin metoda verosimilității maxime** sau prin *metoda minimului lui* χ^2* ale parametrilor necunoscuți $\theta_1, \ldots, \theta_s$.
6. Regiunea critică a *t.c.* χ^2 :
se respinge i.n. H_0 cu p.s. α dacă :

$$\chi^2 < \chi^2_{\alpha;\, k-s-1}.$$

i) *T.c.* χ^2 *pentru verificarea normalității.*
Se alcătuiește un tabel astfel :
1. Valorile observate $\{x_i\}$ ale caracteristicii sub cercetare X se împart în k intervale (clase) :

$$(-\infty, x_1], (x_1, x_2], \ldots, (x_{k-2}, x_{k-1}], (x_{k-1}, \infty),$$

unde

$$x_1 < x_2 < \ldots < x_{k-1},$$

care se trec în prima coloană a tabelului.
2. n_i – frecvențele absolute ale intervalelor (claselor), adică numărul de observații din clasa $(x_{i-1}, x_i]$ se trec în coloana a doua a tabelului.

3. În coloana a treia se trec extremitățile x_i din dreapta intervalelor.

4. Se calculează $\frac{x_i - \bar{x}}{\sqrt{m_2}}$, unde $m_2 = \frac{1}{n} \sum_i (x_i - \bar{x})^2$ și se trec în coloana a patra a tabelului.

5. Se calculează valorile *funcției lui Laplace* Φ în punctele $\frac{x_i - \bar{x}}{\sqrt{m_2}}$ și se trec în coloana a cincea a tabelului.

6. Dacă i.n. este adevărată probabilitățile p_i ca o valoare observată să aparțină clasei i sînt date de :

$$\hat{p}_1 = P(-\infty < X \leqslant x_1) = \Phi\left(\frac{x_1 - \bar{x}}{\sqrt{m_2}}\right)$$

$$\hat{p}_i = P(x_{i-1} < X \leqslant x_i) = \Phi\left(\frac{x_i - \bar{x}}{\sqrt{m_2}}\right) - \Phi\left(\frac{(x_{i-1} - \bar{x}}{\sqrt{m_2}}\right), 2 \leqslant i \leqslant k-1$$

$$\hat{p}_k = P(x_{k-1} < X < \infty) = 1 - \Phi\left(\frac{x_{k-1} - \bar{x}}{\sqrt{m_2}}\right)$$

și sînt trecute în coloana a șasea a tabelului, unde

$$\Phi(x) = 1/\sqrt{2\pi} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

$$\Phi(-x) = 1 - \Phi(x).$$

7. În coloana a 7-a se înregistrează produsul dintre n și valorile $\hat{p}_i$.

8. În coloana a 8-a se calculează $(n_i - n\hat{p}_i)^2$.

9. În coloana a 9-a se calculează $(n_i - n\hat{p}_i)^2/n\hat{p}_i$.

Sumînd rezultatele acestei coloane se obține χ^2.

Exemplu. Ne propunem să verificăm normalitatea datelor unui eșantion de volum $n = 124$:

Tabelul 77

1,97	1,77	1,61	1,61	1,63	1,76	1,80	1,64	1,86	1,67
1,69	1,85	1,75	1,68	1,78	1,93	1,98	1,65	1,85	1,75
1,65	1,77	1,77	1,73	1,93	1,95	1,91	1,81	1,67	2,01
1,69	1,78	2,01	1,95	2,02	1,86	1,82	1,72	1,78	1,87
1,72	1,69	1,74	1,86	1,92	1,90	1,74	1,86	1,80	1,90
1,81	1,68	1,90	1,83	1,77	1,85	1,85	2,03	1,68	1,93
1,70	1,74	1,87	1,61	1,98	1,94	1,74	1,70	1,89	1,76
1,77	1,71	1,76	1,77	1,86	1,97	1,93	1,73	1,71	
1,84	1,69	1,77	1,94	1,62	1,69	1,80	1,84	1,99	
1,81	1,92	1,82	2,02	1,73	1,72	1,79	1,91	1,63	
1,96	1,74	1,86	1,83	1,81	1,90	2,01	1,96	1,92	
1,81	2,03	1,88	1,84	1,70	1,97	1,64	1,99	1,94	
1,87	1,88	1,99	1,82	1,62	1,74	2,02	1,73	1,67	

care reprezintă rezistențele în $k\Omega$ a unor tronsoane de ceramică acoperite cu carbon.

În acest scop se întocmește tab. 78.

Se calculează :

$$\bar{x} = 1{,}81$$

$$\sqrt{m_2} = 0.113.$$

Însumînd ultima coloană a tab. 78 se obține :

$$\chi^2 = \sum_{i=1}^{9} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} = 4{,}11253.$$

Se alege p.s. al testului statistic, $\alpha = 0{,}05$.

Ținînd seama că am estimat doi parametri (m și σ^2) urmează că $s = 2$ și cum $k = 9$ din tabelele legii χ^2 se citește 0.05-cuantila legii χ^2 :

$$\chi^2_{\alpha;\, k-s-1} = \chi^2_{0{,}05;\, 6} = 12{,}6.$$

Deoarece valoarea statisticii χ^2 nu este semnificativă :

$$\chi^2 = 4{,}11253 < 12{,}6 = \chi^2_{0{,}05;\, 6},$$

se acceptă i.n cu p.s. $\alpha = 0{,}05$ conform căruia rezistența chimică a tronsoanelor de ceramică acoperite cu carbon este o variabilă normală.

ii) *T.c.* χ^2 *pentru verificarea exponențialității*

Se testează concordanța repartiției empirice a eșantionului cu *legea exponențială* *, avînd funcția de repartiție :

$$F(x) = 1 - e^{\theta x}.$$

1—3. (→ i).

4. Se calculează valorile funcției de repartiție exponențială în punctele x_i : $:F_X(x_i\,;\,\hat{\theta})$.

5. În coloana a 5-a se calculează probabilitățile $\hat{p}_i$:

$$\hat{p}_i = F_X(x_i\,;\,\hat{\theta}) - F_X(x_{i-1}\,;\,\hat{\theta}) = e^{x_{i-1}/\bar{x}} - e^{x_i/\bar{x}}$$

$$\hat{p}_1 = 1 - e^{-a_1/\bar{x}} = F_X(x_1\,;\,\hat{\theta})$$

$$\hat{p}_k = 1 - \sum_{i=1}^{k} \hat{p}_i,$$

unde

$$\hat{\theta} = -1/\bar{x}.$$

6. În coloana a 6-a se calculează $n\hat{p}_i$.

7. Se calculează $n_i - n\hat{p}_i$ și se trec în coloana a 7-a a tab. 79.

Tabelul 78

Clasa i $1 \leqslant i \leqslant 9$	n_i	x_i	$\frac{x_i - \bar{x}}{\sqrt{m_2}}$	$\varphi\left(\frac{x_i - \bar{x}}{\sqrt{m_2}}\right)$	$\hat{p}_i$	$n\hat{p}_i$	$(n_i - n\hat{p}_i)^2$	$\frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$
$(-\infty ; 1,65]$	11	1,65	$-1,41$	0,0793	0,0793	9,8332	1,36142221	0,13845
$(1,65 ; 1,70]$	14	1,70	$-0,97$	0,166	0,0867	10,7508	10,55730064	0,98200
$(1,70 ; 1,75]$	17	1,75	$-0,53$	0,2981	0,1321	16,3804	0,38390416	0,02344
$(1,75 ; 1,80]$	17	1,80	$-0,09$	0,4641	0,1660	20,5840	12,845056	0,62403
$(1,80 ; 1,85]$	18	1,85	0,35	0,6338	0,1697	21,0428	9,25863184	0,43999
$(1,85 ; 1,90]$	16	1,90	0,79	0,7852	0,1514	18,7736	7,69285696	0,40977
$(1,90 ; 1,95]$	13	1,95	1,23	0,8907	0,1055	13,0820	0,006724	0,00051
$(1,95 ; 2,00]$	10	2,00	1,68	0,9535	0,0628	7,7872	4,89648384	0,62879
$(2,00 ; +\infty)$	8	$+\infty$	$+\infty$	1,0000	0,0465	5,7660	4,99075600	0,86555

$n = 124$

$$4,11253 = \sum_i \frac{(n_i - n\hat{p}_i)}{n\hat{p}_i}$$

8. Se calculează $(n_i - n\hat{p}_i)^2$.
9. Se calculează

$$(n_i - n\hat{p}_i)^2 / n\hat{p}_i,$$

se însumează coloana a 9-a a tabelului și se obține:

$$\chi^2 = \sum_{i=1}^{k} (n_i - n\hat{p}_i)^2 / n\hat{p}_i.$$

Exemplu. Se experimentează $n = 50$ motoare electrice pînă la defectarea ultimului element. Se observă X-numărul de ore de funcționare pînă la defectare. Fie x_i momentul la care se defectează al i-lea element. Datele de observație apar sub forma șirului ordonat al valorilor: $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_{50}$:

1 200 ore	25100	72400	144100	253100
2200	31000	73600	147600	304100
4900	35800	76800	150600	
5000	38900	83000	152600	
6800	39000	95100	164200	
7000	45000	97900	166800	
12100	47900	99600	178600	
13700	48400	102800	185200	
15100	49300	108500	187100	
15200	53200	120700	203000	
23900	55600	128700	204300	
24300	62700	133600	229500	

Se urmărește testarea concordanței repartiției empirice a valorilor observate cu repartiția exponențială cu $F(x) = 1 - e^{-\theta x}$.
Valorile grupate pe intervale cu frecvențele intervalelor respective pot fi urmărite în tab. 79.
Se calculează

$$\bar{x} = 109{,}590 \text{ ore}.$$

Se alege p.s. al t.s. $\alpha = 0{,}05$, de exemplu.
Ținînd seama că am estimat 1 parametru (θ) urmează că $s = 1$ și cum $k = 14$, din tabelele legii χ^2 se citește 0,05-quantila legii χ^2:

$$\chi^2_{0{,}05;\,12} = 21{,}03.$$

Deoarece valoarea χ^2 nu este semnificativă:

$$\chi^2 = 7{,}2211 < \chi^2_{0{,}05;\,12} = 21{,}03$$

se acceptă i.n. cu p.s. $\alpha = 0{,}05$, adică legea v.a. X este legea exponențială.
Test de concordanță χ^2 (*caz neparametric*). Test de concordanță a repartiției empirice cu o lege teoretică $F_X(x;\theta)$, θ fiind parametru cunoscut.

Tabelul **79**

Interval	i	a_i	$F_X(a_i, \hat{\theta})$	$\hat{p}_i$	$n\hat{p}_i$	$n_i - n\hat{p}_i$	$(n_i - n\hat{p}_i)^2$	$(n_i - n\hat{p}_i)^2 / n\hat{p}_i$
$(-00, 3000]$	2	3000	0,0198	0,0198	0,9900	0,01	1,0201	1,0304
(3000, 23000]	8	23000	0,1894	0,1696	8,48	−0,48	0,2304	0,0271
(23000, 43000]	7	43000	0,3229	0,1335	6,675	0,325	0,1056	0,0158
(43000, 63000]	7	63000	0,4345	0,116	5,58	1,42	2,0164	0,3613
(63000, 83000]	7	83000	0,5276	0,0931	4,655	−0,655	0,4290	0,0921
(83000, 103000]	4	103000	0,6094	0,0818	4,09	−0,09	0,0081	0,0019
(103000, 123000]	4	123000	0,6737	0,0643	3,215	−1,215	1,4762	0,4591
(123000, 143000]	2	143000	0,7275	0,0538	2,69	−0,69	0,4761	0,1769
(143000, 163000]	2	163000	0,7724	0,0449	2,245	1,755	3,0800	1,3719
(163000, 183000]	4	183000	0,8099	0,0375	1,875	1,125	1,2656	0,6749
(183000, 203000]	3	203000	0,8428	0,0329	1,645	1,355	1,8360	1,1161
(203000, 223000]	3	223000	0,8687	0,0259	1,295	−1,295	1,6770	1,2949
(223000, 243000]	1	243000	0,8903	0,0216	1,08	−0,08	0,0064	0,0059
(243000, 263000]	1	350000	0,9500	0,0685	3,425	−1,425	2,0306	0,5928
(263000, 350000]	2							$\chi^2_{calc.} =$
	$n = 50$							$= \sum_i \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} =$
								$= 7,2211$

1.—2. (→ *Test de concordanță χ^2, caz parametric.*)

3. Se testează concordanța legii empirice cu legea teoretică $F_X(x\,;\,\theta)$, adică se verifică i.n. : $H_0 : p_i = p_i^0$ (valoarea specificată), unde p_i este probabilitatea ca o observație să aparțină clasei i ($i = 1, \ldots, k$).

4. Se alege p.s. $\alpha = 0{,}05$, de exemplu.

5. Statistica :

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0}$$

este o v.a. cu legea $\sim \chi^2$ cu $(k-1)$ g.e. cînd i.n. H_0 este adevărată.

6. Regiunea critică a *t.c.* χ^2 : se respinge i.n. H_0 cu p.s. α dacă :

$$\chi^2 > \chi^2_{\alpha;\,(k-1)},$$

unde $\chi^2_{\alpha;\,(k-1)}$ este α-cuantila superioară a legii χ^2 cu $(k-1)$ g.l.

i) *T.c. χ^2 (c.n.) pentru verificarea normalității.*

Se procedează analog cu cazul precedent (→ *T.c. χ^2, caz parametric* pct. i). Se alcătuiește un tabel astfel :

1. Valorile observate $\{x_i\}$ ale v.a. X se impart în k intervale (clase) :

$$(-\infty, x_1],\ (x_1,\ x_2],\ \ldots,\ (x_{k-2}, x_{k-1}], (x_{k-1},\ +\infty),$$

unde

$$x_1 < x_2 < \ldots < x_{k-1},$$

care se trec în prima coloană a tabelului.

2. n_i — frecvențele absolute ale claselor, adică numărul de observații din clasa $(x_{i-1},\ x_i]$ — se trec în coloana a doua a tabelului.

3. e_i — mediile numărului de observații căzînd în clasa $(x_{i-1},\ x_i]$:

$$e_i = np_i^0,$$

unde :

$$p_1^0 = P(-\infty < X \leqslant x_1) = \Phi\left(\frac{x_1 - m}{\sigma}\right)$$

$$p_i^0 = P(x_{i-1} < X \leqslant x_i) = \Phi\left(\frac{x_i - m}{\sigma}\right) - \Phi\left(\frac{x_{i-1} - m}{\sigma}\right), \quad 2 \leqslant i \leqslant k-1$$

$$p_k^0 = P(x_{k-1} < X < +\infty) = 1 - \Phi\left(\frac{x_{k-1} - m}{\sigma}\right)$$

se trec în coloana a treia a tabelului.

4. În coloana a patra se calculează $(n_i - np_i^0)^2/np_i^0$ și sumînd elementele acestei coloane se obține χ^2.

Exemplu. Se consideră un eșantion de volum $n = 20$ observații dintr-o populație *) :

0,33	−0,41	−1,82	−0,44
−0,38	−2,41	1,47	−0,21
−1,22	1,29	0,46	−0,43
−0,52	−0,95	−1,23	−0,97
0,48	−1,93	−0,83	2,66

Să verificăm i.n. H_0 conform căreia populația din care provine eșantionul este normală cu media 0 și dispersia 1, adică X-v.a. corespunzătoare are legea $N(0,1)$:

$$H_0 : X \sim N(0,1).$$

Se alcătuiește tabelul :

Intervalul (clasa)	n_i or. de observ. în clasa i	$e_i = np_i^0$	$\frac{(n_i - np_i^0)^2}{np_i^0}$
$(-\infty ; -1,5]$	3	1,3362	2,0717
$(-1,5 ; -0,5]$	6	4,8346	0,2809
$(-0,5 ; 0]$	5	3,8292	0,3580
$(-0 ; 0,5]$	3	3,8292	0,1796
$(0,5 ; 1,5]$	2	4,8346	1,6620
$(1,5 ; +\infty]$	1	1,3362	0,0846
	$\sum_i n_i = 20$	$\sum_i e_i = 20$	$\sum_i \frac{(n_i - np_i^0)^2}{np_i^0} = 4,6368$

Se alege p.s. $\alpha = 0,05$. Din tabelele legii χ^2 pentru $k = 6$ (numărul de intervale) se determină 0,05-cuantila superioară a legii

$$\chi^2_{0,05;\,5} = 11,07.$$

Deoarece valoarea statisticii χ^2 nu este semnificativă :

$$\chi^2 = 4,6368 < 11,07 = \chi^2_{0,05;\,5},$$

se acceptă i.n. H_0 și conchidem că legea v.a. X este $N(0,1)$.

Observații

T.c. χ^2 se aplică în următoarele condiții :

1. Numărul total al observațiilor trebuie să fie suficient de mare, adică $n > 25$.

*) J. H. Pollard, *Numerical and Stat. Techniques*, Cambridge Univ. Press, 1977.

2. Gruparea pe clase (intervale) trebuie să fie astfel încît numărul de observații din fiecare clasă: $n_i \geqslant 5$. În cazul în care $n_i < 5$ se recomandă gruparea claselor (intervalelor), acest fapt avînd drept consecință reducerea numărului g.l. cu 1.
3. Numărul de clase trebuie să fie astfel încît

$$10 \leqslant k \leqslant 15.$$

4. În cazul unui eșantion de volum mare, numărul de clase k se alege astfel:

Pentru	$100 < n < 200$	se alege	$15 < k < 18$
	$n = 200$		$18 < k < 20$
	$n = 400$		$25 < k < 30$
	$n = 1000$		$35 < k < 40.$

5. În cazul legii binomiale, *t.c.* χ^2 este echivalent cu cazul tratat la p. 480.

Test de concordanță Kolmogorov-Smirnov. T.s. neparametric pentru verificarea concordanței repartiției empirice cu cea teoretică.
1. Se consideră un eșantion de volum n, cu observații independente ordonate astfel încît $x_1 < x_2 < \ldots < x_n$ (*șirul statisticilor de ordine*).
2. Cele n observații independente provin din aceeași populație caracterizată de v.a. X avînd legea de repartiție teoretică $F(x) = F(x, \theta)$, depinzînd de parametrul θ.
3. H_0: legea de repartiție a populației este $F(x)$ (specificată).
H_1: i.n. H_0 nu este adevărată.
4. Se alege p.s. al t.s. $\alpha = 0.05$, de exemplu.
5. a) Se calculează *funcția empirică de repartiție**

$$F_n(x) = k/n \qquad x_i \leqslant x \leqslant x_{i+1},$$

unde k este numărul de observații mai mici sau egale cu x, adică

$$F_n(x) = P(X \leqslant x) = \begin{cases} 0, & \text{dacă} \quad i \leqslant 0 \\ i/n, & \text{dacă} \quad 1 \leqslant i \leqslant n-1 \\ 1, & \text{dacă} \quad i > n. \end{cases}$$

b) Se calculează diferențele:

$$d_1 = F_1(x) - F(x_1), \ldots, d_n = F_n(x) - F(x_n),$$

adică

$$d_1 = 1 - nF(x_1); \; d_2 = 2 - nF(x_2), \ldots$$

c) Se alege $|d_i|_{\max}$, cea mai mare diferență în valoare absolută, adică

$$|d_i|_{\max} = \max |F_n(x) - F(x)|.$$

d) Statistica *t.c.* *K.-S.* este:

$$d = (|d_i|_{\max}).$$

6. Regiunea critică a *t.c. K.-S.* este mulțimea valorilor din spațiul de selecție mai mari sau egale cu α-cuantila superioară a legii *K.-S.*
Așadar, se respinge i.n. H_0 cu p.s. α dacă :

$$d \geqslant \lambda_0/\sqrt{n},$$

unde valoarea λ_0 se determină astfel încît :

$$P(d < \lambda_0/\sqrt{n}) = 1 - \alpha.$$

7. Forma calculatorie a statisticii *t.c. K.-S.* este :

$$d_n = \max_i \delta_i,$$

unde

$$\delta_i = \max\left[i/n - F(x_i, \theta) ; F(x_i, \theta) - \frac{i-1}{n} \right].$$

Test de concordanță Lilliefors. În cazul în care legea de repartiție teoretică $F(x, \theta)$ depinde de parametrul θ — necunoscut și care urmează să fie estimat pe baza eșantionului considerat, H. W. Lilliefors (1969) a dat o o variantă a *t.c.* Kolmogorov-Smirnov pentru cazul parametric.
1.—4. (→ *Test de concordanță Kolmogorov-Smirnov.*)
5. Forma calculatorie a *statisticii* d_n^* *dată de Lilliefors* este :

$$d_n^* = \max_i \delta_i^*,$$

unde

$$\delta_i^* = \max\left[i/n - F^*(x_i) ; F^*(x_i) - \frac{i-1}{n} \right]$$

cu

$$F^*(x_i) = F(x_i ; \hat{\theta}),$$

unde $\hat{\theta}$ este estimatorul parametrului θ, obținut pe baza eșantionului considerat.
6. Regiunea critică a *t.c. Kolmogorov - Smirnov - Lilliefors* : se respinge i.n. H_0 cu p.s. α pentru valori mari ale statisticii d_n^*, adică se respinge i.n. H_0 dacă :

$$d_n^* > \lambda_0/\sqrt{n},$$

unde λ_0 este valoarea determinată ca mai sus.
T.c. pentru verificarea exponențialității. Presupunem că repartiția teoretică postulată este *repartiția exponențială** cu densitatea de repartiție

$$f_X(x, \theta) = \begin{cases} \theta e^{-\theta x}, & \theta,\ x > 0 \\ 0, & \text{în rest} \end{cases}$$

și funcția de repartiție teoretică :

$$F_X(x, \theta) = 1 - e^{-\theta x}.$$

Se știe că :

$$\bar{x} = E(X) = \frac{1}{\theta}.$$

Atunci

$$F(x_i, \hat{\theta}) = 1 - e^{-x_i/\bar{x}}.$$

Se calculează

$$\delta_i = \max\left[\frac{i}{n} - (1 - e^{-\theta x_i}) ; \ (1 - e^{-\theta x_i}) - \frac{i-1}{n}\right].$$

Dacă θ este necunoscut, se înlocuiește media teoretică a legii $F(x)$, $E(X) = \frac{1}{\theta}$, prin estimatorul său dat de *media de selecție* * $\bar{x}$:

$$E(X) = 1/\theta = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Exemplu. Se consideră un eșantion de $n = 8$ motoare electrice pentru care se observă numărul de ore de funcționare pînă la defectarea ultimului element. Fie x_i — momentul la care se defectează al i-lea element. Se consideră *șirul statisticilor de ordine* $x_1 < \ldots < x_n$.
Se urmărește testarea concordanței repartiției empirice a eșantionului cu repartiția exponențială avînd funcția de repartiție

$$F(x) = 1 - e^{\theta x}.$$

Calculele pot fi urmărite în tab. 80.
Se observă din ultima coloană a tabelului că

$$d_n^* = 0{,}40961.$$

Din tabelele legii Kolmogorov-Smirnov, pentru $n = 8$, p.s. al testului $\alpha = 0{,}05$ se găsește : $\lambda_0 = 1{,}36$, de unde $\lambda_0/\sqrt{n} = 0{,}489$.

Tabelul 80

i	x_i	$x_i/\bar{x}$	$e^{-x_i/\bar{x}}$	$F^*(x_i)=1-e^{-x_i/\bar{x}}$	i/n	$F^*(x_i)-\frac{i-1}{n}$	$\frac{i}{n}-F^*(x_i)$	δ_i^*
1	2000	0,039	0,96175	0,03824	0,0909	0,05265	0,03821	0,05265
2	4300	0,084	0,91943	0,08056	0,4545	0,37397	−0,28306	0,37397
3	7500	0,146	0,86415	0,13584	0,5454	0,40961	0,31870	0,40961
4	20100	0,397	0,67233	0,32766	0,6363	0,30869	−0,21778	0,30869
5	22800	0,443	0,64210	0,35789	0,7272	0,36937	−0,27817	0,36937
6	110910	2,159	0,11544	0,88455	0,8181	−0,06637	0,15728	0,15729
7	151900	2,957	0,05197	0,94802	0,9090	−0,03893	0,12984	0,12984
8	231900	4,503	0,01107	0,98892	1	−0,011075	0,07983	0,79830

Deoarece $d_n^* = 0{,}40968 < 0{,}489$ urmează că repartiția empirică concordă cu repartiția teoretică exponențială.

Test de concordanță Finklestein-Schafer. Test de concordanță analog cu *testul Lilliefors**, obținut prin modificarea statisticii d_n^*, dată de J. M. Finklestein și R. E. Schafer în 1971 („Biometrika" 58).

1.—4. (→ *Test de concordanță Kolmogorov-Smirnov.*)

5. Statistica *t.c. F.-S.* este:

$$S_n^* = \sum_{i=1}^{n} |\delta_i^*|.$$

6. Regiunea critică a *t.c. F.-S.*: i.n. privind concordanța repartiției empirice cu repartiția teoretică se respinge cu p.s. α dacă:

$$S_n^* > S_{n;\alpha}.$$

unde $S_{n;\alpha}$ sînt valori critice ale statisticii S_n^*, tabelate.

Exemplu.

1) Pentru cazul considerat anterior, se calculează

$$S^* = \sum_{i=1}^{8} |\delta_i^*| = 2{,}59972.$$

Deoarece

$$S^* = 2{,}59972 > S_{8;\,0{,}05} = 1{,}56$$

rezultă că se respinge i.n. privind concordanța repartiției empirice cu *legea exponențială**.

T.c. F.-S. se poate aplica pentru testarea concordanței repartiției empirice cu *repartiția Weibull** pentru valori particulare ale parametrului de formă, și anume $\beta = 2$, respectiv $\beta = \frac{1}{2}$, în cazul în care i.a. este *legea lognormală**, respectiv *legea Gamma**, adică:

$$H_0: F(x) \sim 1 - e^{-\left(\frac{x-\gamma}{\eta}\right)^2}, \text{ respectiv } H_0: F(x) \sim 1 - e^{-\left(\frac{x-\gamma}{\eta}\right)^{1/2}}$$

cu i.a.

$$H_1: F(x) \sim \frac{1}{\sqrt{2\pi}\,\sigma x} e^{-\frac{(\lg x - m)^2}{2\sigma^2}}, \text{ respectiv } H_1: f(x) \sim \beta^\alpha/\Gamma(\alpha)\, x^{\alpha-1} e^{-\beta x},$$

Test de concordanță Gnedenko. B. Gnedenko ș.a. (1969) au dat un test bazat pe repartiția F, care pare să fie mai puternic decît alte teste de concordanță bazate pe *date (eșantioane) cenzurate*, adică un eșantion de volum n pentru

care observațiile s-au făcut pînă la momentul T, înregistrîndu-se doar primele r observații :

$$x_1 \leqslant x_2 \leqslant \ldots \leqslant x_r \quad (r \leqslant n).$$

1.—4. (→ *Test de concordanță Kolmogorov-Smirnov.*)
5. Statistica Q a *t.c.G.* este :

$$Q(r_1, r_2) = \sum_{i=1}^{r} (S_i/r_1) \Big/ \sum_{i=r_1+1}^{r} (S_i/r_2),$$

unde

$$S_i = (n - i + 1)(x_i - x_{i-1})$$

cu $x_0 = 0$ și $r_2 = r - r_1$: testul este terminat la a r-a observație ($r \leqslant n$).

Se arată că S_i sînt v.a. independente cu *legea exponențială** și $\sum_{i=1}^{2k} 2S_i/\theta$ este o v.a. cu *legea* χ^{2*} cu $2k$ g.l. Atunci v.a. $Q(r_1, r_2)$ are *legea* F^* cu $2r_1$ și $2r_2$ g.l.
6. Regiunea critică a *t.c.G.* : Se respinge i.n. de concordanță cu legea exponențială cu p.s. α, dacă :

$$Q(r_1, r_2) > F_{1-\alpha/2}(2r_1, 2r_2) \text{ sau } Q(r_1, r_2) < F^{-1}_{1-\alpha/2}(2r_2, 2r_1),$$

unde $F_u(k_1, k_2)$ este u%-cuantila legii F cu k_1, k_2 g.l.

Observații.

1. Dacă valoarea calculată a statisticii $Q(r_1, r_2)$ este mare, se obține indicația că legea studiată are *rata de defectare** crescătoare.
Dacă valoarea calculată a statisticii $Q(r_1, r_2)$ este mică, se poate trage concluzia că rata de defectare este descrescătoare ($\beta < 1$).
2. W.W. Fercho și L. J. Ringer (1972) („Technometrics" nr. 14) au folosit :

$$r_1 = r_2 = \frac{r}{2}, \text{ dacă } r \text{ este par}$$

$$r_1 = \left[\frac{r}{2}\right], \; r_2 = \left[\frac{r}{2}\right] + 1, \text{ dacă } r \text{ este impar,}$$

unde $[x]$ este cel mai mare întreg mai mic decît x.
3. *T.c. G.* este un test bilateral. Forma particulară a statisticii testului de concordanță Q rezultă din proprietățile v.a. $X \sim F_{k_1, k_2}$, și anume : dacă v.a. X are legea F cu k_1, k_2 g.l., atunci v.a. $1/X$ are legea F cu k_2, k_1, g.l.

Test de concordanță Cramer — von Mises. T.s. neparametric folosit pentru testarea concordanței repartiției empirice $F_n(x)$ cu legea teoretică $F(x)$.
1.—4. (→ *Test de concordanță Kolmogorov-Smirnov.*)
5. Statistica *t.c. C.-v.M.* este :

$$W^2 = \int_{-\infty}^{+\infty} (F_n(x) - F(x))^2 \, dx$$

sau forma modificată propusă de Smirnov ;

$$W^2 = \int_{-\infty}^{+\infty} (F_n(x) - F(x))^2 \, dF(x).$$

6. În rest se procedează ca la testul Kolmogorov-Smirnov.

TESTE PENTRU VERIFICAREA NORMALITĂȚII. Fie un eșantion de n observații independente $X_1, \ldots, X_n$ provenind dintr-o populație necunoscută. Se verifică i.n. conform căreia legea de repartiție este legea normală $N(m, \sigma^2)$; m, σ^2-parametri necunoscuți, ce urmează a fi estimați pe baza datelor eșantionului considerat.

Cele mai cunoscute **t.v.n.** sînt : *testul χ^2 de concordanță* * cu legea normală și *testul Kolmogorov-Smirnov**.

Testul de normalitate al lui Lilliefors. H. W. Lilliefors (1967, JASA, vol. 62) a dat un test analog cu testul de concordanță Kolmogorov-Smirnov pentru verificarea normalității, cînd m, σ^2 sînt parametri necunoscuți. (→ *Test de concordanță Lilliefors*).

1.—4. (→ *Testul Kolmogorov-Smirnov* (1.—4.).)

5. Forma calculatorie a statisticii Lilliefors este :

$$\hat{D}_n = \max_{1 \leqslant i \leqslant n} \hat{\delta}_i,$$

unde

$$\hat{\delta}_i = \max \left[F_0(X_i ; \overline{X}, s) - \frac{(i-1)}{n} ; \frac{i}{n} - F_0(X_i ; \overline{X}, s) \right]$$

cu $(X_i)_{1 \leqslant i \leqslant n}$ șirul *statisticilor de ordine* corespunzătoare eșantionului considerat ;

$s^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$, *estimatorul nedeplasat** al dispersiei teoretice σ^2 ;

$F_0(X_i ; \overline{X}, s)$ este funcția de repartiție a legii normale $N(0, 1)$, adică :

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-t^2/2} \, dt$$

cu proprietatea :

$$\Phi(-y) = 1 - \Phi(y)$$

unde

$$y = (X - \overline{X})/s.$$

6. Regiunea critică a *t.n.L.* : se respinge i.n. H_0 cu p.s. α, fixat dacă valoarea statisticii $\hat{D}_n$ este semnificativă, adică :

$$\hat{D}_n > \hat{D}_{n;\alpha},$$

unde $\hat{D}_{n;\alpha}$ sînt valorile critice tabelate.

Exemplu. Se consideră un eșantion de volum $n = 8$ privind duratele de viață ale unui echipament electric. Valorile ordonate (exprimate în logaritmi naturali) sînt :

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
0,26	0,53	0,88	1,22	1,76	2,44	3,41	4,90

Se verifică i.n. conform căreia duratele de viață provin dintr-o populație *log-normală* cu 2 parametri.

Pe baza datelor eșantionului se calculează :

$$\overline{X} = 1{,}92$$

$$s^2 = \frac{\sum_{i=1}^{8} (X_i - \overline{X})^2}{8-1} = \frac{47{,}30 - 8 \cdot (1{,}92)^2}{8-1} = 2.55.$$

Pentru fiecare valoare X_i se calculează :

$$y_i = (X_i - \overline{X})/s, \quad = \mathfrak{Z}$$

adică

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
−1,03	−0,87	−0,65	−0,44	−0,10	0,33	0,93	1,86

Din tabelele *legii normale** $N(0, 1)$ se calculează :

$$F_0(X_i; \overline{X}, s) = \Phi(y_i), \quad i = 1, \ldots, 8 :$$

0,152	0,192	0,258	0,330	0,460	0,629	0,824	0,968

Deoarece valoarea statisticii $\hat{D}_n$:

$$\hat{D}_8 = \frac{4}{8} - 0.330 = 0.170$$

nu este semnificativă :

$$\hat{D}_8 = 0.170 < \hat{D}_{8;\,\alpha=0,05} = 0{,}285$$

se acceptă i.n. privind legea normală a logaritmilor naturali ai duratelor de viață, adică legea lognormală a populației din care se extrage eșantionul studiat. I.n. se acceptă chiar și pentru p.s. $\alpha = 0.2$, conform valorilor critice ale statisticii $\hat{D}_n$.

Testul lui Massey. T.s.v.n. în cazul unui eșantion de volum mic ($10 \leqslant n \leqslant 30$) 1.—4. (→ *Test de concordanță* χ^2, *caz parametric i*)).*

5. Se calculează diferențele :

$$d_i = |F_n(z_i) - \Phi(z_i)|,$$

unde

$F_n(z_i)$ este *funcția de repartiție empirică**

$\Phi(z_i)$ este *funcția lui Laplace*, calculată în punctele

$$z_i = \frac{X_i - \overline{X}}{\sqrt{m_2}}, \quad m_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Se alege

$$d_s = \max_i d_i$$

6. Regiunea critică a *t.M.* : se respinge i.n. H_0 conform căreia $X \sim N(m, \sigma^2)$ dacă valoarea statisticii d_s este semnificativă, adică

$$d_s > d_{g,n;\alpha},$$

unde $d_{g,n;\alpha}$ este valoarea critică tabelată în funcție de volumul eșantionului n și de p.s. α ales și sînt date în tab. 81.

Tabelul 81

n	α		n	α	
	0,05	0,10		0,05	0,10
8	0,140	0,163	20	0,117	0,133
9	0,134	0,158	21	0,115	0,131
10	0,130	0,156	22	0,113	0,129
11	0,129	0,155	23	0,112	0,128
12	0,128	0,154	24	0,110	0,126
13	0,128	0,153	25	0,109	0,124
14	0,128	0,151	26	0,108	0,121
15	0,127	0,148	27	0,107	0,120
16	0,126	0,144	28	0,105	0,118
17	0,124	0,142	29	0,104	0,116
18	0,122	0,138	30	0,102	0,114
19	0,120	0,136	32	0,099	0,111

Exemplu. Ne propunem să verificăm, la un p.s. $\alpha = 0{,}05$, i.n. conform căreia rezistența chimică (în $K\Omega$) a unor tronsoane de ceramică acoperite cu carbon este o v.a. $\sim N(m, \sigma^2)$. Vom face această verificare pe baza unui eșantion de 10 măsurători : 1,68 ; 1,74 ; 1,82 ; 1,60 ; 1,72 ; 1,90 ; 1,79 ; 1,98 ; 1,85 ; 1,93. Din primele două coloane ale tab. 82, găsim

$$\overline{X} = 1{,}801\ ; \quad m_2 = \frac{1}{10}\sum_{i=1}^{10}(X_i - \overline{X})^2 = 0{,}012669\ ; \sqrt{m_2} = 0{,}1125.$$

Calculele se pot urmări în tab. 82.

Tabelul 82

X_i	$(X_i-\overline{X})^2$	$z_i=\frac{X_i-\overline{X}}{\sqrt{m_2}}$	Frecv.	Funcția empirică de rep. $F_n(z_i)$	$\Phi(z_i)$	$\lvert F_n(z_i)-\Phi(z_i)\rvert$
1,60	0,040401	−1,78	1	0,1	0,03754	0,06246
1,68	0,014641	−1,07	1	0,2	0,1423	0,05770
1,72	0,006561	−0,72	1	0,3	0,2358	0,06420
1,74	0,003721	−0,54	1	0,4	0,2916	0,10540
1,79	0,000121	−0,09	1	0,5	0,4641	0,03590
1,82	0,000361	0,17	1	0,6	0,5675	0,03250
1,85	0,002401	0,43	1	0,7	0,6664	0,03360
1,90	0,009801	0,88	1	0,8	0,8106	0,01060
1,93	0,016641	1,14	1	0,9	0,8729	0,02710
1,98	0,032041	1,59	1	1,0	0,94408	0,05592

$\Sigma=18.01$ $0,126690$

Deoarece valoarea statisticii $d_s=\max_i d_i=0,10540$ nu este semnificativă:

$$d_{max}=0,10540<d_{g,\,10;\,0,05}=0,130\,,$$

se acceptă i.n. conform căreia v.a. X – rezistența chimică – urmează o lege normală $N(m,\sigma^2)$.

Testul Shapiro-Wilk (Test W). T.s.v.n. unei populații caracterizată de v.a. X.

1. Se consideră un eșantion de volum n și fie *statisticile de ordine* $X_1, X_2, \ldots, X_n$ ale eșantionului corespunzător v.a. X. $\{X_i\}$ independente.

2.–4. → (*Test de concordanță Kolmogorov-Smirnov.*)

5. Dacă $n=2l$ se calculează:

$$b=\sum_{i=1}^{l} a_{n-i+1}[X_{n-i+1}-X_i]=\sum_{i=1}^{l} a_{n-i+1}W(i).$$

unde coeficienții a_{n-i+1} se citesc din tab. 83.

Dacă $n=2l+1$, deoarece $a_{l+1}=0$, se calculează $b=a_n[X_n-X_1]+\ldots$ $\ldots+a_{l+2}[X_{l+2}-X_l]$, unde valoarea X_{l+1} – mediana eșantionului – nu intră în calculul lui b. Se calculează

$$S^2=\sum_{i=1}^{n}(X_i-\overline{X})^2$$

și statistica $W=b^2/S^2$.

Coeficienții $\{a_{n-i+1}\}$ folosiți în **t.n.W** pentru $n=3$ (1) 50

i \ n	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.5888	0.5739	0.5601	0.5475	0.5359	0.5251	0.5150	0.5056	0.4968	0.4886
2		0.1677	0.2413	0.2806	0.3031	0.3164	0.3244	0.3291	0.3315	0.3325	0.3325	0.3318	0.3306	0.3290	0.3273	0.3253
3				0.0875	0.1401	0.1743	0.1976	0.2141	0.2260	0.2347	0.2412	0.2460	0.2495	0.2521	0.2540	0.2553
4						0.0561	0.0947	0.1224	0.1429	0.1586	0.1707	0.1802	0.1878	0.1939	0.1988	0.2027
5								0.0399	0.0695	0.0922	0.1099	0.1240	0.1353	0.1447	0.1524	0.1587
6										0.0303	0.0539	0.0727	0.0880	0.1005	0.1109	0.1197
7												0.0240	0.0433	0.0593	0.0725	0.0837
8														0.0196	0.0359	0.0496
9																0.0163

i \ n	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1	0.4808	0.4734	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407	0.4366	0.4328	0.4291	0.4254	0.4220	0.4188	0.4156	0.4127
2	0.3232	0.3211	0.3185	0.3156	0.3126	0.3098	0.3069	0.3043	0.3018	0.2992	0.2968	0.2944	0.2921	0.2898	0.2876	0.2854
3	0.2561	0.2565	0.2578	0.2571	0.2563	0.2554	0.2543	0.2533	0.2522	0.2510	0.2499	0.2487	0.2475	0.2463	0.2451	0.2439
4	0.2059	0.2085	0.2119	0.2131	0.2139	0.2145	0.2148	0.2151	0.2152	0.2151	0.2150	0.2148	0.2145	0.2141	0.2137	0.2132
5	0.1641	0.1686	0.1736	0.1764	0.1787	0.1807	0.1822	0.1836	0.1848	0.1857	0.1864	0.1870	0.1874	0.1878	0.1880	0.1882
6	0.1271	0.1334	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563	0.1584	0.1601	0.1616	0.1630	0.1641	0.1651	0.1660	0.1667
7	0.0932	0.1013	0.1092	0.1150	0.1201	0.1245	0.1283	0.1316	0.1346	0.1372	0.1395	0.1415	0.1433	0.1449	0.1463	0.1475
8	0.0612	0.0711	0.0804	0.0878	0.0941	0.0997	0.1046	0.1089	0.1128	0.1162	0.1192	0.1219	0.1243	0.1265	0.1284	0.1301
9	0.0303	0.0422	0.0530	0.0618	0.0696	0.0764	0.0823	0.0876	0.0923	0.0965	0.1002	0.1036	0.1066	0.1093	0.1118	0.1140
10		0.0140	0.0263	0.0368	0.0459	0.0539	0.0610	0.0672	0.0728	0.0778	0.0822	0.0862	0.0899	0.0931	0.0961	0.0988
11				0.0122	0.0228	0.0321	0.0403	0.0476	0.0540	0.0598	0.0650	0.0697	0.0739	0.0777	0.0812	0.0844
12						0.0107	0.0200	0.0284	0.0358	0.0424	0.0483	0.0537	0.0585	0.0629	0.0669	0.0706
13								0.0094	0.0178	0.0253	0.0320	0.0381	0.0435	0.0485	0.0530	0.0572
14										0.0084	0.0159	0.0227	0.0289	0.0344	0.0395	0.0441
15												0.0076	0.0144	0.0206	0.0262	0.0314
16														0.0068	0.0131	0.0187
17																0.0062

i \ n	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	0.4096	0.4068	0.4040	0.4015	0.3989	0.3964	0.3940	0.3917	0.3894	0.3872	0.3850	0.3830	0.3808	0.3789	0.3770	0.3751
2	0.2834	0.3813	0.2794	0.2774	0.2755	0.2737	0.2719	0.2701	0.2684	0.2667	0.2651	0.2635	0.2620	0.2604	0.2589	0.2574
3	0.2427	0.2415	0.2403	0.2391	0.2380	0.2368	0.2357	0.2345	0.2334	0.2323	0.2313	0.2302	0.2291	0.2281	0.2271	0.2260
4	0.2127	0.2121	0.2116	0.2110	0.2104	0.2098	0.2091	0.2085	0.2078	0.2072	0.2065	0.2058	0.2052	0.2045	0.2038	0.2032
5	0.1881	0.1883	0.1883	0.1881	0.1880	0.1876	0.1876	0.1874	0.1871	0.1868	0.1865	0.1862	0.1859	0.1855	0.1851	0.1847
6	0.1673	0.1678	0.1683	0.1686	0.1689	0.1691	0.1693	0.1694	0.1695	0.1695	0.1695	0.1695	0.1695	0.1693	0.1692	0.1691
7	0.1487	0.1498	0.1505	0.1513	0.1520	0.1526	0.1531	0.1535	0.1539	0.1542	0.1545	0.1548	0.1550	0.1551	0.1553	0.1554
8	0.1317	0.1331	0.1344	0.1356	0.1366	0.1376	0.1384	0.1392	0.1398	0.1405	0.1410	0.1415	0.1420	0.1423	0.1427	0.1430
9	0.1160	0.1179	0.1196	0.1211	0.1225	0.1237	0.1249	0.1259	0.1269	0.1278	0.1286	0.1293	0.1300	0.1306	0.1312	0.1317
10	0.1013	0.1036	0.1056	0.1075	0.1092	0.1108	0.1122	0.1136	0.1149	0.1160	0.1170	0.1180	0.1189	0.1197	0.1205	0.1212
11	0.0873	0.0900	0.0924	0.0947	0.0967	0.0986	0.1004	0.1020	0.1035	0.1049	0.1062	0.1073	0.1085	0.1095	0.1105	0.1113
12	0.0739	0.0770	0.0798	0.0824	0.0848	0.0870	0.0891	0.0909	0.0927	0.0943	0.0959	0.0972	0.0986	0.0998	0.1010	0.1020
13	0.0610	0.0645	0.0677	0.0706	0.0733	0.0759	0.0782	0.0804	0.0824	0.0842	0.0860	0.0876	0.0892	0.0906	0.0919	0.0932
14	0.0484	0.0523	0.0559	0.0592	0.0622	0.0651	0.0677	0.0701	0.0724	0.0745	0.0765	0.0783	0.0801	0.0817	0.0832	0.0846
15	0.0361	0.0404	0.0444	0.0481	0.0515	0.0546	0.0575	0.0602	0.0628	0.0651	0.0673	0.0694	0.0713	0.0731	0.0748	0.0764
16	0.0239	0.0287	0.0331	0.0372	0.0409	0.0444	0.0476	0.0506	0.0534	0.0560	0.0584	0.0607	0.0628	0.0648	0.0667	0.0685
17	0.0119	0.0172	0.0220	0.0264	0.0305	0.0343	0.0379	0.0411	0.0442	0.0471	0.0497	0.0522	0.0546	0.0568	0.0588	0.0608
18		0.0057	0.0110	0.0158	0.0203	0.0244	0.0283	0.0318	0.0352	0.0383	0.0412	0.0439	0.0465	0.0489	0.0511	0.0532
19				0.0053	0.0101	0.0146	0.0188	0.0227	0.0263	0.0296	0.0328	0.0357	0.0385	0.0411	0.0436	0.0459
20						0.0049	0.0094	0.0136	0.0175	0.0211	0.0245	0.0277	0.0307	0.0335	0.0361	0.0386
21								0.0045	0.0087	0.0126	0.0163	0.0197	0.0229	0.0259	0.0288	0.0314
22										0.0042	0.0081	0.0118	0.0153	0.0185	0.0215	0.0244
23												0.0039	0.0076	0.0111	0.0143	0.0174
24														0.0037	0.0071	0.0104
25																0.0035

6. Regiunea critică a *t.S.* – $W.$: dacă $W < W_{n;\,1-\alpha}$, se respinge i.n. H_0 privind normalitatea.

Valorile critice ale lui $W_{n;\,1-\alpha}$ pentru $1-\alpha = 1,\ 2,\ 5,\ 10,\ 90,\ 98,\ 99\%$ și $n = 3, \ldots, 50$, sînt tabelate.

Observații

1) Testul W se bazează pe statisticile de ordine.

2) Testul W necesită tabele speciale atît pentru calculul statisticii pe care se bazează testul cît și pentru luarea deciziei de acceptare sau respingere a ipotezei normalității v.a. X asupra căreia s-a considerat eșantionul.

3) Valorile critice ale statisticii W au fost obținute prin simulare pe o mașină electronică de calcul.

Exemplu. În aceleași condiții tehnologice ca și în exemplul de la p. 443 s-a determinat rezistența chimică a 18 tronsoane de ceramică acoperite cu carbon obținîndu-se următoarele valori:

1,65 1,72 1,70 1,84 1,96 1,58 1,85 1,78 1,88

1,75 1,92 1,60 1,75 2,05 1,90 1,76 1,82 1,68

Se verifică i.n. conform căreia rezistența chimică este o v.a. $\sim N(m, \sigma^2)$.

Se alege p.s. $\alpha = 0{,}05$.

Se calculează:

$$\overline{X} = \frac{1}{18}\,32{,}18 = 1{,}786; \quad \sum_{i=1}^{18} (X_i - \overline{X})^2 = 0{,}2743.$$

Aplicarea *t.S.*-$W.$ se urmărește în tab. 84.

Tabelul 84

i	X_{n-i+1}	X_i	W_i	a_{n-i+1}	$a_{n-i+1}W_i$
1	2,05	1,58	0,47	0,4886	0,229642
2	1,96	1,60	0,36	0,3253	0,117108
3	1,92	1,65	0,27	0,2553	0,068931
4	1,90	1,68	0,22	0,2027	0,044594
5	1,88	1,70	0,18	0,1587	0,028536
6	1,85	1,71	0,14	0,1197	0,016758
7	1,84	1,72	0,12	0,0837	0,010044
8	1,82	1,75	0,07	0,0496	0,003472
9	1,78	1,76	0,02	0,0163	0,000326

Primele 9 observații în ordinea mărimii au fost trecute în coloana a treia a tab. 84, celelalte 9 observații au fost trecute în coloana a doua a aceluiași tabel începînd de jos în sus. Elementele coloanei a 5-a sînt extrase din tab. 83.

Sumînd ultima coloană a tab. 84 se obține :

$$b = \sum_{i=1}^{9} a_{n-i+1} W_i = 0{,}51941 \approx 0{,}52.$$

Deci statistica W are valoarea 0,985.
Deoarece valoarea $W = 0{,}985$ nu este semnificativă :

$$W = 0{,}985 > W_{18;\,0{,}95} = 0{,}982,$$

se acceptă i.n. privind normalitatea, cu p.s. $\alpha = 0{,}05$.
Teste pentru verificarea normalității prin măsurarea asimetriei și excesului (*bazate pe coeficienții Beta*). **T.s.v.n.** bazat pe coeficienții β_1 și β_2 definiți de

$$\beta_1 = \nu_3'^2/\nu_2'^3 = \gamma_1^2; \quad \beta_2 = \nu_4'/\nu_2'^2$$

care dau o descriere a formei curbei de frecvență.
Coeficientul β_1 sau echivalentul

$$\gamma_1 = \nu_3'/(\nu_2')^{3/2}.$$

(rădăcina pătrată din β_1 cu același semn cu ν_3') este o măsură a *asimetriei*.
Pentru o repartiție simetrică $\beta_1 = 0$.
În general, o repartiție este cu *asimetrie pozitivă* sau cu *asimetrie negativă*, după cum ν_3' este pozitiv sau negativ.
Coeficientul β_2 sau echivalent $\gamma_2 = \beta_2 - 3$ este o măsură a *boltirii*, comparația făcîndu-se cu curba normală pentru care $\beta_2 = 3$.
Dacă $\beta_2 < 3$, repartiția se numește *platicurtică* și este mai turtită spre vîrf decît o curbă normală.
Dacă $\beta_2 > 3$, repartiția se numește *leptocurtică* și este mai ascuțită spre vîrf decît curba normală.
Pentru *legea normală**

$$\beta_1 = 0 \quad \text{și} \quad \gamma_2 = 0.$$

Așadar, acești indicatori pot fi folosiți pentru verificarea normalității.
1. Fie un eșantion de volum n asupra unei v.a. $X \sim N(m, \sigma^2)$.
2.–4. (→ *Test de concordanță. i*) : *Testul de concordanță Kolmogorov-Smirnov.*)
5. Fie b_1 și b_2 valorile coeficienților β_1 și respectiv β_2 obținute pe baza datelor eșantionului respectiv.
Pentru valori mari ale lui n, v.a.

$$z_1 = \pm \sqrt{\frac{\beta_1 (n+1)(n+3)}{6(n-2)}} \approx \pm \sqrt{\frac{n\beta_1}{6}}$$

(avînd semnul momentului μ_3) este repartizată aproximativ $N(0, 1)$.
Pentru valori mari ale lui n, v.a.

$$z_2 = \left(\beta_2 - 3 + \frac{6}{n+1}\right) \sqrt{\frac{(n+1)^2 (n+3)(n+5)}{24\, n(n-2)(n-3)}} \approx (\beta_2 - 3) \sqrt{\frac{n}{24}}$$

este repartizată aproximativ $N(0, 1)$.

6. Regiunea critică: *Asimetria** observată în eșantion este semnificativă la p.s. α, dacă

$$|z_1| > z_{\alpha/2}$$

unde z_1 este valoarea statisticii Z_1 pentru $\beta_1 = b_1$.
*Boltirea** observată în eșantion este semnificativă, la p.s. α dacă:

$$|z_2| > z_{\alpha/2},$$

unde z_2 este valoarea statisticii Z_2 pentru $\beta_2 = b_2$.
Testul lui Sarkady. T.s.v.n. în cazul unui eșantion de volum mic.
T.S. se aplică în astfel de situații și atunci cînd micșorarea riscului (p.s. al testului) α duce la scăderea puterii testului.
1. Fie $(X_1, \ldots, X_n)$ un eșantion de volum n provenind dintr-o populație normală $N(m, \sigma^2)$.
2. $X \sim N(m, \sigma^2)$.
3. I.n. $H_0: X \in N(m, \sigma^2)$ față de i.a. $H_1: X \notin N(m, \sigma^2)$.
4. $\alpha = 0{,}05$, de exemplu.
5. Se calculează statisticile

$$f_i = X_i - A \quad (i = 1, \ldots, n-1),$$

unde

$$A = \frac{\overline{X}\sqrt{n} + X_1}{\sqrt{n} + 1}, \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

și

$$\xi_i = \frac{f_i\sqrt{n-i-1}}{\sqrt{\sum_{k=1}^{n-1} f_k^2}}, \quad (i = 1, 2, \ldots, n-2).$$

Se calculează:

$$\delta_{\nu_i} = t_{\nu_i}(\xi_i), \quad (i = 1, 2, \ldots, n-2).$$

unde t_{ν_i} este funcția de repartiție a *legii Student** t cu

$$\nu_i = n - i - 1 \text{ g.l.}, \quad i = 1, 2, \ldots, n-2.$$

6. Dacă i.n. H_0 este adevărată, atunci v.a. δ_{ν_i} au *legea uniformă** pe segmentul (0,1).
Pentru verificarea acestei ipoteze se folosește *criteriul* ω^2 *(von Mises-Smirnov)*, și anume:
Se calculează statistica von Mises—Smirnov:

$$n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left[F(\eta_i) - \frac{2i-1}{2n}\right]$$

unde n este volumul eșantionului, iar $F(\gamma_i)$ este funcția de repartiție teoretică calculată pentru elementele eșantionului ordonate crescător (pentru *șirul statisticilor de ordine*).

În cazul de față există $n-2$ elemente ale eșantionului, iar $F(\delta_i) = \delta_i$ (*proprietatea repartiției uniforme*)*, unde prin δ_i s-au notat valorile δ_{γ_i} ordonate (crescător).

Statistica testului Mises-Smirnov devine :

$$(n-2)\,\omega^2_{n-2} = \frac{1}{12(n-2)} + \sum_{i=1}^{n-2}\left[\delta_i - \frac{2i-1}{2(n-2)}\right]^2.$$

Regiunea critică : se acceptă i.n.H_0 cu p.s. $\alpha = 0{,}05$, de exemplu, dacă :

$$(n-2)\,\omega^2_{n-2} \leqslant 0{,}4164.$$

Exemplu. Se studiază X — caracteristica „tenacitate" a firului de relon. Pentru verificarea stabilității procesului tehnologic, relativ la această caracteristică, este necesară verificarea i.n. privind normalitatea v.a. X.

Se consideră un eșantion de volumul $n = 96$. Dacă aplicăm testul χ^2 se respinge i.n. pentru p.s. ales $\alpha = 0{,}05$ (sau $\alpha = 0{,}025$ și $\alpha = 0{,}01$).

S-a constatat însă că pentru $\alpha = 0{,}0005$ (ceea ce reprezintă un risc foarte mic) i.n. H_0 se acceptă.

În această situație testul χ^2 nu este concludent.

Se aplică *t.S.* folosind criteriul ω^2 (von Mises-Smirnov).

În cazul caracteristicii X — „tenacitate", s-a calculat :

$$\overline{X} = 4{,}7, \quad A \approx 4{,}6, \quad 94\,\omega^2_{94} \approx 0{,}4.$$

Deoarece valoarea statisticii nu este semnificativă :

$$94 \cdot \omega^2_{94} = 0.4 < 0{,}4164,$$

se acceptă i.n. privind normalitatea legii v.a. X — „tenacitatea" firului de relon.

Teste de normalitate Shapiro-Wilk, Shapiro-Francia, De Wet-Venter, d'Agostino. Teste de concordanță a repartiției empirice cu legea normală bazate pe statistici de forma :

$$W = \frac{\left(\sum_{i=1}^{n} c_i X_i\right)^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2},$$

unde $\{X_i\}$ este *șirul statisticilor de ordine*, adică șirul valorilor ordonate crescător ale eșantionului de volum n considerat :

$$X_1 \leqslant X_2 \leqslant \ldots \leqslant X_n.$$

Constantele $\{c_i\}$ și valorile critice ale statisticii W au fost tabelate pentru $n \leqslant 50$ în cazul testului Shapiro-Wilk (→ *Testul de normalitate Shapiro-Wilk*) : $50 \leqslant n \leqslant 99$ în cazul testului de normalitate Shapiro-Francia ; pentru testul d'Agostino și testul de Wet-Venter au fost obținute constantele și valorile critice asimptotice.

Testul de normalitate Shapiro-Francia și testul de normalitate de Wet-Venter sînt *teste consistente**, în timp ce testul d'Agostino nu este test consistent. Testul Shapiro-Wilk folosit ca *test de concordanță** cu legea exponențială nu este test consistent.

Testul de normalitate Sarkady. Test de concordanță a repartiției empirice cu legea normală în cazul în care nu dispunem de eșantioane de volum mare, ci de mai multe eșantioane de volum mic.

1. Fie X_{ij}, $i = 1, \ldots, r$; $j = 1, \ldots, n_i$ v.a. independente.
2. Cele r eșantioane provin din populații caracterizate de legile $F((x - m_i)/\sigma_i)$; m_i, σ_i — parametri necunoscuți, $i = 1, \ldots, r$.
3. $H_0 : F(x) \sim \Phi\left(\dfrac{x - m}{\sigma}\right)$,

unde $\Phi(x)$ este funcția de repartiție a legii $N(0, 1)$.
$H_1 : F(x)$ nu este legea normală.
4. Se alege p.s., α.
5. Se disting două cazuri : *cazul homoscedastic*, adică dispersii egale $\sigma_1^2 = \ldots = \sigma_r^2 = \sigma^2$ și *cazul heteroscedastic*, adică dispersii oarecare σ_i^2, $i = 1, \ldots, r$.

i) *Cazul homoscedastic.*

Se calculează $\left(\sum_{i=1}^{r} n_i - r\right)$ v.a. Y_{ij} :

$$Y_{ij} = X_{ij} - U_i,$$

unde

$$U_i = \frac{\sum_{j=1}^{n_i - 1} X_{ij}}{n_i + \sqrt{n_i}} - \frac{X_{in_i}}{\sqrt{n_i}} \qquad \begin{matrix} i = 1, \ldots, r \\ j = 1, \ldots, n_i - 1 \\ n_i \geqslant 2. \end{matrix}$$

ii) *Cazul heteroscedastic*

Se calculează $\left(\sum_{i=1}^{r} n_i - 2r\right)$ v.a. Y_{ij} :

$$Y_{ij} = (X_{ij} - U_i)\frac{S_{Y_i}}{S_i},$$

unde

$$U_i = \frac{\sum_{j=1}^{n_i-2} X_{ij}}{n_i + \sqrt{2n_i}} + \frac{X_{in_i-1} + X_{in_i}}{\sqrt{2n_i}}$$

$$S_i = \sqrt{\sum_{j=1}^{n_i-2} (X_{ij} - U_i)^2}$$

$$S_{Y_{ij}} = \psi_{n_i}\left(\frac{S_i}{|X_{in_i-1} - X_{in_i}|}\right) \quad \begin{matrix} i = 1, \ldots, r \\ j = 1, \ldots, n_i - 2 \end{matrix}$$

cu $\psi_n(x)$ funcție monoton crescătoare, astfel încît legea v.a. $S_{Y_{ij}}$ este aceeași cu legea v.a. S_i/σ_i cînd i.n. H_0 este adevărată. Valorile funcției $\psi_n(x)$ sînt calculate cu ajutorul valorilor critice ale *legilor* χ^2 și *t a lui Student*.

6. Regiunile critice se stabilesc ținînd seama că:

Cazul homoscedastic. Dacă i.n. H_0 este adevărată, atunci v.a. independente $U_i \sim N(0, \sigma^2)$.

Se testează normalitatea acestora, ceea ce implică normalitatea v.a. inițiale (folosind o teoremă cunoscută a lui H. Cramér).

Cazul heteroscedastic. Dacă i.n. H_0 este adevărată, atunci v.a. independente $Y_{ij} \sim N(0, 1)$.

Se testează normalitatea acestora, ceea ce implică normalitatea v.a. inițiale (folosind o teoremă cunoscută a lui H. Cramér, o teoremă a lui A. M. Kagan, Yu. V. Linnik, C. R. Rao (1973), o teoremă a lui K. Sarkadi (1980)).

Observații

1. Aplicarea transformărilor asupra v.a. inițiale folosite în *t.S.* este condiționată de *caracterul aleator* al observațiilor $\{X_{ij}\}$. Dacă există îndoieli în această privință se recomandă schimbarea ordinei observațiilor din interiorul fiecărui eșantion, prin alegerea la întîmplare a ultimului element X_{in_i} al fiecărui eșantion, $i = 1, \ldots, r$, respectiv ale ultimelor două elemente X_{in_i-1}, X_{in_i} ale fiecărui eșantion, $i = 1, \ldots, r$.

2. Forme simplificatorii ale *t.S.* se obțin pentru $n_i = 2$ și $n_i = 3$.

Testul d'Agostino. T.s.v.n. pentru eșantioane de volum moderat ($n \geqslant 50$). A fost elaborat în 1971 și se bazează pe statistic a lui F. Downton („Biometrika" 53/1966) folosită pentru estimarea abaterii standard s a unei populații normale.

1. Se consideră un eșantion de volum n; se ordonează datele experimentale și fie *statisticile de ordine* * :

$$X_1 \leqslant X_2 \leqslant \ldots \leqslant X_n.$$

2. Eșantionul provine dintr-o populație caracterizată de v.a. $X \sim N(m, \sigma^2)$

3. I.n. H_0 : $X \sim N(m, \sigma^2)$

I.a. H_1 : i.n. H_0 nu este adevărată.

4. Se alege p.s. al t.s., α.

5. Se calculează :

$$T = \sum_{i=1}^{n} \{i - 1/2(n+1)\} X_i = \sum_{i=1}^{n} iX_i - \frac{n(n+1)}{2} \overline{X}$$

$$s_1 = \left[\frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 \right]^{1/2}$$

$$D = \frac{T}{n^2 s_1}.$$

Statistica testului este :

$$Y = \frac{D - 0{,}282095}{0{,}029986} \sqrt{n}.$$

6. Regiunea critică : dacă $Y_{n;\alpha/2} \leqslant Y \leqslant Y_{n;1-\alpha/2}$, atunci se acceptă ipoteza normalității ; dacă $Y < Y_{n;\alpha/2}$ sau $Y > Y_{n;1-\alpha/2}$, atunci se respinge ipoteza normalității, unde α-cuantilele $Y_{n;\alpha/2}$ și $Y_{n;1-\alpha/2}$ pentru $\alpha = 0{,}02$ și $\alpha = 0{,}05$ sînt date în tab. 85.

Tabelul 85

Valorile α-cuantilelor testului D'Agostino pentru verificarea normalității

	$\alpha = 0{,}05$		$\alpha = 0{,}02$	
	$Y_{n;0{,}025}$	$Y_{n;0{,}975}$	$Y_{n;0{,}01}$	$Y_{n;0{,}99}$
50	−2,757	1,038	−3,442	1,140
60	−2,699	1,115	−3,360	1,236
70	−2,652	1,176	−3,293	1,312
80	−2,613	1,226	−3,237	1,374
90	−2,580	1,260	−3,100	1,426
100	−2,552	1,303	−3,050	1,470
150	−2,452	1,423	−3,009	1,623
200	−2,391	1,496	−2,922	1,715
250	−2,348	1,545	−2,861	1,779
300	−2,316	1,528	−2,816	1,826

Exemplu. În procesul fabricării aluminiului Al, unul dintre parametrii care se urmăresc este așa-numitul oxid de sodiu lavabil $(Na_2O)_L$, al cărui conținut se exprimă în procente. Un procent ridicat de astfel de substanță influențează negativ conținutul de Al_2O_3, materia primă preliminară în procesul fabricării Al. În tab. 86 sînt date valorile $(Na_2O)_L$ în procente și frecvența lor de apariție în cadrul unui experiment efectuat în scopul stabilirii unei toleranțe superioare pentru oxidul de sodiu lavabil.

Tabelul 86

Datele experimentale $(Na_2O)_L$ în procente și frecvența lor de apariție

$(Na_2O)_L$%	0,03	0,06	0,09	0,10	0,11	0,12		0,13	
Frecvența	1	1	3	4	8	9		10	
$(Na_2O)_L$%	0,14	0,15	0,16	0,17	0,18	0,19	0,20		0,22
Frecvența	7	7	6	4	3	3	3		1

Se verifică normalitatea datelor eșantionului.

Se alege p.s., $\alpha = 0{,}05$.

Se calculează $\sum_i iX_i$.

Se calculează:

$$\frac{1}{2}(n+1)n = 2485; \quad \overline{X} = 0{,}13742; \quad \frac{1}{2}(n+1)n \cdot \overline{X} = 341{,}4887;$$

$$s_1 \approx 0{,}033: \quad \sum_i i \cdot X_i = 387{,}26; \quad T = 45{,}7713;$$

$$D = 0.28306$$

Statistica t. $D'A$. este:

$$Y = \frac{0.28306 - 0.28203}{0{,}02998} \cdot 8{,}366 \approx 0{,}27247.$$

Deoarece valoarea statistică Y nu este semnificativă:

$$Y_{70;\,0{,}025} = -2.652 < 0{,}27247 < 1{,}176 = Y_{70;\,0{,}975}$$

se acceptă i.n. H_0 cu p.s. $\alpha = 0{,}05$, adică ipoteza normalității se acceptă.

Metode grafice pentru testarea concordanței repartiției empirice cu unele repartiții teoretice. Reprezentarea grafică a datelor în coordonate carteziene sau pe rețele special gradate. Sînt folosite pentru detectarea valorilor aberante, testarea concordanței repartiției empirice cu unele repartiții teoretice, pentru estimarea parametrilor care intervin în forma analitică a legii de repartiție teoretice.

*Rețeaua probabilistică** sau *hîrtia probabilistică* (*probability paper*, în engl.) este o hîrtie special gradată (simplu sau dublu logaritmică) pe care se reprezintă punctele ale căror coordonate sînt: (valorile observate ale v.a. X; estimația funcției de repartiție F), adică $(X_i, F_{i,n})$ sau $(\ln X_i, F_{i,n})$. Se obține liniarizarea funcției de repartiție empirice corespunzătoare v.a. $X \sim F(x; \theta)$.

Se recunoaște astfel repartiția teoretică respectivă și se estimează parametrii ei, fără a apela la metode analitice.

Exemple: rețele Weibull, gama, normală, lognormală etc. pentru repartițiile statistice respective.

L. Nelson a propus în lucrarea : „Weibull probability paper" — publicată în „Industrial Quality Control" 23, nr. 9 /1967, p. 452—453, o rețea probabilistică de tip Weibull care permite estimarea parametrilor direct din rețea. Pentru întocmirea *hîrtiei probabilistice Weibull* (de tip Nelson) se reprezintă punctele de coordonate : *abscisa* — datele de observație ordonate X_i (timpii de funcționare, duratele de viață etc.) ; *ordonata*—estimația funcției de repartiție empirice $F_X(X_i) = F_{i,n}$:

$$F_{i,n} = \frac{i}{n+1}\,\%, \quad i = 1, 2, \ldots, n.$$

sau

$$F_{i,n} = \left(i - \frac{1}{2}\right) \Big/ n\%$$

$$F_{i,n} = (i - 3/8)\,(n + 1/4)\,\%.$$

Abscisa X_i este transformată de hîrtia probabilistică în $t_i = \ln X_i$.
Ordonata $F_X(X_i) = F_{i,n}$ este transformată de hîrtia probabilistică în $v_{i,n} = \ln\{\ln[1/(1-F_{i,n})]\}$.
Așadar, la baza gradării hîrtiei probabilistice Weibull stă relația de dependență liniară între $\ln X$ și $\ln\{\ln[1/(1-F_X(x))]\}$:

$$\ln\{\ln[1/(1-F_X(x))]\} = \beta[\ln(x-\mu) - \ln\delta],$$

unde $F_X(x) = 1 - e^{-\left(\frac{x-\mu}{\delta}\right)^{\beta}}$ este *funcția de repartiție a legii Weibull**
Pentru liniarizarea graficului, se scade din fiecare observație X_i o valoare $\hat{\mu} < X_1$ aleasă prin încercări, unde $\hat{\mu}$ este o estimație a *parametrului de locație* μ *al legii Weibull**.

După linearizare, panta liniei astfel obținută din graficul $(X_i, F_X(X_i))$ dă o estimație a *parametrului de formă* β al legii Weibull.
Pentru obținerea acestei estimații se duce o paralelă, prin originea 0 la linia obținută și se citește intersecția acestei paralele cu scala din stînga a hîrtiei probabilistice Weibull.
Intersecția liniei obținute cu ordonata corespunzătoare originei (0, 0) a axei $t = \ln x$ citită pe scala din dreapta dă o estimație a valorii $-\beta \ln \delta$.
Valoarea de pe axa orizontală $t = \ln x$ corespunzătoare intersecției liniei obținute cu abscisa originei (0,0) de pe scala din dreapta dă o estimație a parametrului $\ln \delta$, unde δ este *parametrul de scală al legii Weibull*.
Exemplu. Pentru testarea concordanței repartiției empirice cu unele repartiții teoretice (cu repartiția Weibull) se folosește un eșantion format din 6 probe din același tip de aluminiu supuse unui test de coroziune accelerată.
Momentele la care s-a atins pragul de coroziune admis sînt X_i : 3,45 ; 5,00 ; 5,90 ; 7,30 ; 8,60 ; 10,20.
Folosind hîrtia probabilistă Weibull se reprezintă grafic punctele :

$$\left(X_i, \left[100\left(i - \frac{1}{2}\right) \Big/ 6\right]\%\right), \quad i = 1, \ldots, 6,$$

conform datelor din tab. 87.

Tabelul 81

	X_i	$F_{i,n} = \left\lfloor 100\left(i - \frac{1}{2}\right)/6 \right\rfloor$ %
1	3,45	8,33
2	5,00	25
3	5,90	41,66
4	7,30	58,33
5	8,60	75
6	10,20	91,66

Scăzînd valoarea 1 din observațiile X_i și reprezentînd grafic punctele : $(X_i - 1, [100\ (i-1/2)/6]\%)$ se obține linearizarea graficului, adică dreapta (d).
O paralelă la dreapta (d) dusă prin originea O intersectează scala din stînga aproximativ în 2,6 ; deci $\hat{\beta} = 2{,}6$.
Intersecției A dintre dreapta (d) și ordonata corespunzătoare originei (0,0) a axei t îi corespunde pe scala din dreapta valoarea -5 : deci $-\hat{\beta} \ln \hat{\delta} = -5$.
Intersecției B dintre dreapta (d) și abscisa originei (0, 0) a scalei din dreapta îi corespunde pe axa orizontală $t = \ln x$, valoarea 1,9 : deci $\ln \hat{\delta} = 1{,}9$.
Diagrama Barlow-Campo.
1. Fie un eșantion de volum n corespunzător unei v.a. X nenegative cu funcția de repartiție teoretică F_X (necunoscută) și fie șirul *statisticilor de ordine* corespunzătoare primelor r observații :

$$X_{n1} \leqslant X_{n2} \leqslant \ldots \leqslant X_{nr}, \quad i = 1, \ldots, r$$

$$r < n$$

2. $$X \sim F_X(x)$$

cu :

Densitatea de probabilitate : $f(x)$.

Media : $E(X) = m$.

Rata de defectare : $r(x) = \dfrac{f(x)}{1 - F(x)}$.

Funcția empirică de repartiție :

$$F_n(u) = \begin{cases} 0, & u \leqslant X_{n1} \\ i/n, & X_{ni} \leqslant u < X_{n,i+1} \\ 1, & u \geqslant X_{nn} \end{cases}$$

3. i.n. H_0 : v.a. X are funcția de repartiție teoretică $F_X(x)$ (de formă analitică specificată).
i.a. H_1 : i.n. H_0 nu este adevărată.
4. Se alege p.s. α.

5. Se definesc:
Timpul total de testare pînă la a r-a defectare (sau *cădere*) ca fiind suma tuturor *duratelor de viață* observate:

$$T(X_{nr}) = nX_{n1} + (n-1)(X_{n2} - X_{n1}) + \ldots + (n-r+1)(X_{nr} - X_{n,r-1})$$

sau

$$T(X_{nr}) = \sum_{i=1}^{r} X_{ni} + (n-r)X_{nr};$$

Transformata timpului total de testare corespunzătoare funcției empirice de repartiție F_n:

$$H_n^{-1}\left(\frac{r}{n}\right) = \frac{1}{n} T(X_{nr})$$

sau

$$H_n^{-1}(r/n) = \int_0^{F_n^{-1}(r/n)} (1 - F_n(u))\,du,$$

unde $F_n^{-1}(u)$ este definită prin relația:

$$F_n^{-1}(u) = \inf\{x;\ F_n(x) \geqslant x\};$$

Transformata timpului total de testare (transformata TTT) corespunzătoare funcției teoretice de repartiție $F_X(x)$:

$$H_F^{-1}(t) = \int_0^{F^{-1}(t)} (1 - F(u))\,du;$$

Transformata normată a timpului total de testare:

$$T_s(t) = H_F^{-1}(t)/H_F^{-1}(1),$$

unde

$$H_F^{-1}(1) = \int_0^{F^{-1}(1)} (1 - F(u))\,du = m$$

$T_s(t)$ este funcție crescătoare pe $[0, 1]$

$$T_s(0) = 0$$

$$T_s(1) = 1.$$

Observații

1. Determinarea lui $H_F^{-1}(t)$ conduce la obținerea inversei F^{-1} și deci a funcției de repartiție F, deoarece :

$$\frac{\mathrm{d}}{\mathrm{d}t} H_F^{-1}(t) = (1-t)\frac{\mathrm{d}}{\mathrm{d}t} F^{-1}(t)$$

$$\frac{\mathrm{d}}{\mathrm{d}t} F^{-1}(t) = \frac{1}{1-t}\frac{\mathrm{d}}{\mathrm{d}t} H^{-1}(t).$$

Panta la graficul transformatei TTT, calculată în punctul $t = F(x)$, este e iversa *ratei de defectare** în punctul considerat :

$$\frac{\mathrm{d}}{\mathrm{d}t} H_F^{-1}(t)\Big|_{t=F(x)} = \frac{1-t}{f(F^{-1}(t))}\Big|_{t=F(x)} = \frac{1-F(x)}{f(x)} = \frac{1}{r(x)}.$$

Această proprietate permite clasificarea unei repartiții cu ajutorul ratei de defectare.

3. Dacă $F(x)$ este o funcție de repartiție pentru care rata de defectare este *crescătoare* (IFR), respectiv *descrescătoare* (DFR), atunci transformata TTT $T_s(t)$ este *concavă*, respectiv *convexă* pentru $t \in [0, 1]$.

4. Pentru *legea exponențială**, care are rata de defectare constantă

$$r(x) = \lambda = \text{constant}$$

și funcția de repartiție :

$$F(x) = 1 - \mathrm{e}^{-x/a} \quad x \geqslant 0,\ a \geqslant 0,$$

se observă că transformata normată TTT este :

$$T_s(t) = \frac{H_F^{-1}(t)}{H_F^{-1}(1)} = t, \quad t \in [0, 1].$$

Așadar, în pătratul de latura 1 ce poate fi construit pe segmentul [0,1], transformata normată TTT este prima bisectoare a acestuia. Pentru alte legi decît cea exponențială, poziția curbelor transformatei normale TTT, corespunzătoare diverselor valori ale parametrilor, se judecă în raport cu prima bisectoare a pătratului de latură 1.

Astfel, pentru *legea Gamma* * cu parametrul de formă k, $k > 1$, se obțin curbe concave (situate deasupra bisectoarei) și concavitatea se accentuează pe măsură ce k crește.

Pentru *legea Weibull* *, curbele transformatei normate TTT sînt concave (situate deasupra bisectoarei) pentru $\beta > 1$, respectiv convexe (situate sub bisectoare) pentru $\beta < 1$.

5. *Întocmirea diagramelor TTT.* Se consideră șirul *statisticilor de ordine* corespunzătoare primelor r observații :

$$X_{n1} \leqslant X_{nr} \leqslant \ldots \leqslant X_{nr}$$

testarea oprindu-se la momentul X_{nr}. Se reprezintă grafic — în sistem de axe rectangular — punctele de coordonate

$$(i/r\ ;\ T(X_{ni})/T(X_{nr})),\quad i = 1, \ldots, r.$$

Diagrama obținută este analoagă cu cea întocmită folosind transformata TTT cu factorul de scală $H_F^{-1}(u)$, adică diagrama obținută reprezentînd grafic punctele :

$$(t\ ;\ H_F^{-1}(t)/H_F^{-1}(u))$$

pentru n suficient de mare, astfel încît $i/n \to t$ și $r/n \to u$ (conform *teoremei lui Glivenko*).

Exemplu. Se consideră 10 observații reprezentînd duratele de funcționare pînă la prima defectare (ordonate crescătoare) ale unui eșantion de 10 piese dintr-un lot de $n = 100$ piese :

Tabelul 88

i	X_{ni}	$i/n = \hat{F}(x)$	$n - i$	$\sum_{i=1}^{10} X_{ni}$	$T(X_{ni})$	$T(X_{ni})/T(X_{n,10})$
1	2800	0,01	99	2800	280000	0,465
2	3200	0,02	98	6000	319600	0,531
3	3400	0,03	97	9400	339200	0,563
4	3800	0,04	96	13200	378000	0,628
5	4200	0,05	95	17400	416400	0,692
6	4500	0,06	94	21900	444900	0,739
7	4750	0,07	93	36650	468400	0,778
8	5200	0,08	92	31850	510250	0,847
9	5800	0,99	91	37650	565450	0,939
10	6200	0,1	90	43850	601850	1

S-a calculat în prealabil :

$$T(X_{n,10}) = \sum_{j=1}^{10} X_{n,j} + 90\, X_{n,10} = 601859.$$

Se reprezintă grafic punctele de coordonate $(i/r\ ;\ T(X_{ni})/T(X_{nr}))$, $i = 1, \ldots, r = 10$. Prin confruntare cu diagramele TTT pentru repartiția Weibull se conchide că datele de observație provin dintr-un eșantion corespunzător repartiției Weibull de parametru $\beta = 1{,}85$.

Observație. Diagramele Barlow-Campo se folosesc pentru studiul datelor nenegative incomplete sau cenzurate (adică în perioada de observare nu ajung să se defecteze sau să se supună observației toate elementele eșantionului).

TESTE PENTRU VERIFICAREA CARACTERULUI ALEATOR. T.s. pentru verificarea caracterului întîmplător sau omogenității datelor eșantionului folosit. Respingerea i.n. privind caracterul întîmplător (aleator)

al datelor eşantionului arată una din următoarele trăsături ale eşantionului : a) prezenţa unor valori extreme ; b) prezenţa unor fluctuaţii periodice ; c) prezenţa unor discontinuităţi ; d) prezenţa unor tendinţe în evoluţia fenomenului studiat.

Testul semnelor. T.s. neparametric bazat pe semnele anumitor cantităţi şi nu pe valorile lor. De exemplu, se consideră semnele diferenţelor *observaţiilor-perechi* :

$$d_i = x_{i2} - x_{i1}$$

din două eşantioane $(x_{11}, \ldots, x_{n1})$, respectiv $(x_{12}, \ldots, x_{n2})$ de volume egale provenind din populaţiile P_1, respectiv P_2.

Testul semnelor se foloseşte ca t.s. pentru :

i) verificarea i.n. H_0 privind o valoare specificată a medianelor populaţiilor P_1, P_2 ;

ii) verificarea tendinţei unei *serii cronologice** (*serii de timp*) pe baza raportului dintre numărul diferenţelor pozitive şi negative între termenii succesivi ai seriei ;

iii) test de omogeneitate a datelor, adică se verifică i.n. :

$$H_0 : F_1(x) = F_2(x),$$

unde F_1, F_2 sînt funcţiile de repartiţie ale populaţiilor P_1, P_2.

Testul iteraţiilor. T.s. bazat pe numărul de iteraţii dintr-un eşantion faţă de *mediana** eşantionului.

1. Se consideră un eşantion de volum $2n$: $X_1, \ldots, X_{2n}$ asupra v.a. X, avînd funcţia de repartiţia $F(x)$. Fie

$$X_{(1)}, X_{(2)}, \ldots, X_{(n)}, X_{(n+1)}, \ldots, X_{(2n)}$$

şirul *statisticilor de ordine* corespunzătoare şi fie *mediana** eşantionului.

2. Se alcătuieşte şirul *iteraţiilor*, adică şirul succesiunilor de semne „a", respectiv „b" folosite pentru a desemna o valoare $X_i \leqslant Me$, respectiv $X_i > Me$. V.a. U — numărul de iteraţii astfel obţinut printre cele $2n$ observaţii considerate — are media şi dispersia :

$$M(U) = n + 1$$

$$\operatorname{Var}(U) = \frac{n(n-1)}{2n-1}.$$

3. I.n. H_0 : eşantionul are un caracter întîmplător.
H_1 : i.n. H_0 nu este adevărată.

4. Se alege α, p.s. al testului.

5. Statistica testului este U.

6. Regiunea critică : se respinge i.n. H_0 pentru valori mici ale lui U, deoarece U semnificativ mic indică tendinţa de grupare a datelor eşantionului şi deci lipsa unui caracter întîmplător în variaţia lor. Variaţia mare a valorilor eşantionului faţă de Me implică U semnificativ mare şi certifică caracterul aleator al eşantionului.

Pentru n suficient de mare, v.a.

$$\frac{U - M(U)}{\sqrt{\operatorname{Var} U}} \sim N(0, 1),$$

dacă i.n. H_0 este adevărată.

Se procedează în continuare ca la *testul z**.
Exemplu. Se consideră eșantionul de volum $n = 18$:

$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(4)}$	$X_{(5)}$	$X_{(6)}$	$X_{(7)}$	$X_{(8)}$	$X_{(9)}$	$X_{(10)}$	$X_{(11)}$
1,58	1,60	1,65	1,68	1,70	1,72	1,75	1,75	1,76	1,78	1,82

$X_{(12)}$	$X_{(13)}$	$X_{(14)}$	$X_{(15)}$	$X_{(16)}$	$X_{(17)}$	$X_{(18)}$
1,84	1,85	1,88	1,90	1,92	1,96	2,05

Se alege

$$Me = \frac{X_{(9)} + X_{(10)}}{2} = 1{,}77$$

și se alcătuiește șirul iterațiilor „a", respectiv „b" după cum $X_i < 1{,}77$, respectiv $X_i > 1{,}77$:

U : a a a b b a b b b a b a a b b a b a
1 2 3 4 5 6 7 8 9 10 11

care are $n = 9$ valori $a = X_{(i)} < Me = 1{,}77$ și $n = 9$ valori $b = X_{(i)} > Me = 1{,}77$. Numărul total de iterații este $U = 11$.
Se calculează :

$$M(U) = 9 + 1 = 10$$

$$\text{Var } U = \frac{9 \cdot 8}{18 - 1} = 4{,}23.$$

Se alege p.s. al testului, $\alpha = 0{,}05$.
Deoarece statistica (→ *test z (bilateral)*) :

$$\left| \frac{U - M(U)}{\sqrt{\text{Var } U}} \right| = 0{,}5$$

nu este semnificativă :

$$\left| \frac{U - M(U)}{\sqrt{\text{Var } U}} \right| = 0{,}5 < z_{\alpha/2} = 2{,}34$$

se acceptă i.n. privind caracterul aleator al repartiției datelor eșantionului studiat.

Test von Neumann (-Kent-Bellinson-Hart). T.s, bazat pe statistica :

$$\delta^2 = \frac{1}{n-1} \sum_{j=1}^{n-1} (x_{i+1} - x_i)^2,$$

numită *media pătratică a diferențelor succesive*, introdusă de R. L. Young în 1941, folosită pentru a detecta una din cele patru trăsături a)—d) specifice caracterului nealeator al datelor eșantionului.

1. Se consideră eşantionul de volum n : $x_1, \ldots, x_n$.
2. Testarea caracterului aleator se bazează pe compararea a doi estimatori ai dispersiei teoretice, presupunînd că eşantionul considerat provine dintr-o populaţie $N(m, \sigma^2)$.
3.—4. (→ *Testul iteraţiilor.*)
5. Statistica testului :

$$M = \delta^2/s^2,$$

unde δ^2 este *media pătratică a diferenţelor succesive*, iar $s^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$ este *dispersia de selecţie** (estimatorul nedeplasat al dispersiei teoretice σ^2). Valorile critice (α-cuantilele inferioare şi superioare) ale statisticii M au fost tabelate de Hart (1942) pentru p.s. $\alpha = 0{,}01$ şi $0{,}05$.
6. Regiunea critică : mulţimea valorilor de selecţie mai mici decît α-cuantila inferioară sau mai mari decît α-cuantila superioară a legii v.a. M; adică se respinge i.n. dacă

$$M \leqslant M_{1-\alpha;\,n} \quad \text{sau} \quad M \geqslant M_{\alpha;\,n},$$

unde $M_{1-\alpha;\,n}$, $M_{\alpha;\,n}$ sînt date în tab. 89.
Valorile α-cuantilelor inferioare, $M_{1-\alpha;\,n}$, respectiv α-cuantilelor superioare, $M_{\alpha;\,n}$

Tabelul 89

n	α-cuantila inferioară	$M_{1-\alpha;\,n}$	α-cuantila superioară	$M_{\alpha;\,n}$
	$\alpha = 0{,}05$	$=0{,}01$	$=0{,}05$	$=0{,}01$
4	0,78	0,63	3,22	2,37
5	0,82	0,54	3,18	3,46
6	0,89	0,56	3,11	3,44
7	0,94	0,61	3,06	3,39
8	0,98	0,66	3,02	3,34
9	1,02	0,71	2,98	3,29
10	1,06	0,75	2,94	3,25
11	1,10	0,79	2,90	3,21
12	1,13	0,83	2,87	3,17
15	1,21	0,92	2,79	3,08
20	1,30	1,04	2,70	2,96
25	1,37	1,13	2,63	2,87

Observaţie. Pentru $n > 25$, statistica

$$M' = 1 - M/2$$

este o v.a. $\sim N\left(0, \dfrac{n-2}{n^2-1}\right)$ dacă i.n. H_0 este adevărată.

Se calculează statistica :

$$M_1 = \frac{\delta^2}{2s^2}.$$

Regiunea critică a t.s. : se respinge i.n. H_0 față de i.a. H_1 dacă

$$M_1 < r_{\alpha;\, n} = 1 - z_\alpha \sqrt{\frac{n-2}{n^2-1}},$$

unde z_α este α-cuantila legii $N(0, 1)$.

Exemplu

a) Un lot format din 16 scule aşchietoare este supus unor încercări de durabilitate. În tab. 90 sînt date durabilităţile obţinute (în minute), precum şi elementele de calcul necesare*) :

Tabelul 90

Nr. crt.	Durabilitatea (minute) x_i	Diferenţele succesive $(x_{i+1} - x_i)$	Diferenţele faţă de medie $(x_i - \bar{x})$	$(x_{i+1} - x_i)^2$	$(x_i - \bar{x})^2$
1	7,16	4,18	−7,86	17,4724	61,7736
2	11,34	−6,29	−3,68	39,5641	13,5424
3	5,05	7,15	−9,97	2,1025	99,4009
4	6,50	7,76	−8,52	60,2176	72,5904
5	14,26	− 4,86	−0,76	23,6196	0,5776
6	9,40	− 2,20	−5,62	4,8400	31,5844
7	7,20	4,98	−7,82	24,8004	61,1524
8	12,18	5,36	−2,84	28,7296	8,0656
9	17,54	− 1,44	2,52	2,0738	6,3504
10	16,10	6,96	1,08	48,4416	1,6664
11	23,06	− 1,78	8,04	3,1684	64,6516
12	21,28	9,16	6,26	83,9056	39,1876
13	30,44	−14,14	15,42	199,9396	237,7764
14	16,30	1,72	1,28	2,9584	1,6384
15	18,02	6,56	3,00	43,0336	9,0000
16	24,58	—	9,55	—	91,3936

* Sursa : D. V. Iliescu, V. Gh. Vodă, *Statistică şi toleranţe*, Ed. tehnică, Bucureşti, 1977.

Se calculează :

$$\bar{x} = 15{,}02$$

$$\sum_{i=1}^{16} (x_i - \bar{x})^2 = 799{,}8477$$

$$\delta^2 = \frac{1}{15} \sum_{i=1}^{15} (x_{i+1} - x_i)^2 = 38{,}9911 ; \quad s^2 = 53{,}3231.$$

$$M_{calculat} = \frac{38{,}9911}{53{,}3231} \approx 0{,}7312.$$

Se alege p.s. al t.s., $\alpha = 0{,}05$.

Deoarece valoarea statisticii M este semnificativ mai mică decît α-cuantilă inferioară : $M < M_{0{,}05;\,16} \approx 1{,}22$, se respinge i.n. H_0 cu p.s. $\alpha = 0.05$. Rezultă că eşantionul respectiv prezintă un caracter ncaleator. Experimentatorul poate suspecta faptul că eşantionul este influenţat de valori extreme, de exemplu $x_3 = 5{,}05$, care diferă suficient de mult faţă de restul valorilor. Practic, aceasta înseamnă că au avut loc anumite „căderi timpurii" ale sculelor, care s-au rupt mult înaintea limitei normale de uzură.

b) În tab. 91 sînt înregistrate $n = 32$ de valori reprezentînd tensiunea electromotoare (e) a unui termocuplu Pt Rh – Pt, la punctul de solidificare al stibiului, precum şi elementele de calcul necesare.

Se calculează :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (e_i - \bar{e})^2 = \frac{1}{31} \sum_{i=1}^{32} (e_i - \bar{e})^2 = 0{,}00000135 \text{ mV}^2.$$

$$\delta^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} (e_{i+1} - e_i)^2 = \frac{1}{31} \sum_{i=1}^{31} (e_{i+1} - e_i)^2 = 0{,}000002064 \text{ mV}^2.$$

Întrucît $n > 25$, se calculează statistica :

$$M_1 = \frac{\delta^2}{2s^2} = \frac{0{,}000001032}{0{,}000001350} \approx 0{,}974.$$

Se alege p.s. al testului, $\alpha = 0{,}05$.

Deoarece valoarea statisticii M_1 nu este semnificativă :

$$M_1 = 0{,}974 > r_{0{,}05;\,32} = 1 - 1{.}645 \left(\frac{30}{31 \times 33} \right)^{1/2} \approx 0{,}720$$

se acceptă i.n. H_0, deci împrăştierea datelor obţinute este întîmplătoare, eşantionul respectiv fiind reprezentativ.

*Tabelul *) 91*

n_i	e_i	$e_i-\bar{e}$	$(e_i-\bar{e})^2$	$e_{i+2}-e_i$	$(e_{i+1}-e_i)^2$
1	5,538	0,001	0,000001	—	—
2	5,536	−0,001	0,000001	0,002	0,000004
3	5,538	0,001	0,000001	0,002	0,000004
4	5,538	0,001	0,000001	0,000	0,000000
5	5,536	−0,001	0,000001	−0,002	0,000004
6	5,537	0,001	0,000000	0,001	0,000001
7	5,537	0,000	0,000000	0,000	0,000000
8	5,538	0,001	0,000001	0,001	0,000001
9	5,536	−0,001	0,000001	−0,002	0,000004
10	5,536	−0,001	0,000001	0,000	0,000000
11	5,537	0,000	0,000000	0,001	0,000001
12	5,537	0,000	0,000000	0,000	0,000000
13	5,538	0,001	0,000001	0,001	0,000001
14	5,535	−0,002	0,000001	0,003	0,000009
15	5,536	−0,001	0,000001	0,001	0,000001
16	5,536	−0,001	0,000001	0,000	0,000000
17	5,538	0,001	0,000001	0,002	0,000004
18	5,538	0,001	0,000001	0,000	0,000000
19	5,538	0,001	0,000001	0,000	0,000000
20	5,536	−0,001	0,000001	−0,002	0,000004
21	5,536	−0,001	0,000001	0,000	0,000000
22	5,538	0,001	0,000001	0,002	0,000004
23	5,538	0,001	0,000001	0,000	0,000000
24	5,536	−0,001	0,000001	−0,002	0,000004
25	5,536	−0,001	0,000001	0,000	0,000000
26	5,536	−0,001	0,000001	0,000	0,000000
27	5,539	0,002	0,000004	0,003	0,000009
28	5,540	0,003	0,000009	0,001	0,000001
29	5,538	0,001	0,000001	−0,002	0,000004
30	5,536	−0,001	0,000001	−0,002	0,000004
31	5,536	−0,001	0,000001	0,000	0,000000
32	5,536	−0,001	0,000001	0,000	0,000000

$$\sum_{i=1}^{32} e_i = 171{,}183;\quad \sum_{i=1}^{32} (e_i-\bar{e})=0;\quad \sum_{i=1}^{32} (e_i-\bar{e})^2=0{,}000042;\quad \sum_{i=1}^{31} (e_{i+1}-e_i)^2 = 0{,}000064$$

* Sursa : Idem, ca la tab. 90.

TESTE PRIVIND COEFICIENTUL DE CORELAȚIE

Test de independență. 1. Se consideră un eșantion de volum n : n puncte aleatoare (x_i, y_i), $i = 1, \ldots, n$ dintr-o populație avînd *legea normală bidimensională* * cu coeficientul de corelație ρ.

2. Vectorul aleator (X, Y) corespunzător eșantionului dat are *legea normală bidimensională* * cu coeficientul de corelație ρ, avînd d.p. :

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{1}{2}\frac{1}{1-\rho^2}\left[\left(\frac{x-m_1}{\sigma_1}\right)^2 - 2\rho\frac{x-m_1}{\sigma_1}\frac{y-m_2}{\sigma_2} + \left(\frac{y-m_2}{\sigma_2}\right)^2\right]}.$$

3. V. a. X, Y sînt independente dacă și numai dacă sînt necorelate, adică se verifică :

a) i.n. $H_0 : \rho = 0$ (t.b.)
i.a. $H_1 : \rho \neq 0$

b) $H_0 : \rho \leqslant 0$ (t.u.dr.)
$H_1 : \rho > 0$

c) $H_0 : \rho \geqslant 0$ (t.u.stg.)
$H_1 : \rho < 0$

4. Se alege p.s. al testului, α.

5. Se calculează *coeficientul de corelație de selecție*

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

Statistica $t.i.$

$$t = \frac{\sqrt{n-2}\, r}{\sqrt{1-r^2}}$$

este o v.a. $\sim t_{n-2}$, dacă i.n. H_0 este adevărată.

6. Regiunile critice corespunzătoare diferitelor i.a. H_1 sînt date în tab. 92

Tabelul 92

i.n. H_0: $\rho = 0$		
i.a. H_1	Regiunea critică	Test statistic
a) $\rho \neq 0$	$\lvert t\rvert > t_{n-2;\,\alpha/2}$ sau $t < -t_{n-2;\,\alpha/2} = t_{n-2;\,\alpha/2}$; $t > t_{n-2;\,\alpha/2}$	t.b.
b) $\rho > 0$	$t > t_{n-2;\,\alpha}$	t.u.dr.
c) $\rho < 0$	$t < t_{n-2;\,1-\alpha} = -t_{n-2;\,\alpha}$	t.u. stg.

Exemplu
a) Pentru un eşantion de volum $n = 38$ dintr-o populaţie în care caracteristica sub cercetare este un vector bidimensional, se obţine coeficientul de corelaţie de selecţie $r = 0{,}8$. Se verifică i.n. $H_0 : \rho = 0$ faţă de i.a. $H_1 : \rho \neq 0$.
Se calculează statistica *t.i.* :

$$t = \sqrt{38 - 2} \cdot \frac{0{,}8}{\sqrt{1 - 0{,}64}} = 8.$$

Se alege p.s. al testului, $\alpha = 0.05$.
Deoarece valoarea statisticii este semnificativă :

$$t = 8 > t_{36;\,0{,}025} = 2{,}021$$

se respinge i.n. privind independenţa v.a. X, Y cu p.s. $\alpha = 0{,}05$. Aşadar, v.a. X, Y sînt corelate.

Test privind coeficientul de corelaţie bazat pe transformata Z a lui Fisher
1.—4. (→ *Test de independenţă* (1.—4.).)
5. Se calculează r. (→ *Test de independenţă* (5.).)
R. A. Fisher a arătat că v.a. transformată :

$$z = \frac{1}{2} \ln \frac{1 + r}{1 - r} \sim N\left(\frac{1}{2} \ln \frac{1 + \rho_0}{1 - \rho_0}, \frac{1}{n - 3}\right)$$

dacă i.n. $H_0 : \rho = \rho_0$ este adevărată.
Statistica testului :

$$Z = \frac{\frac{1}{2} \ln \frac{1 + r}{1 - r} - \frac{1}{2} \ln \frac{1 + \rho_0}{1 - \rho_0}}{\sqrt{n - 3}}$$

este o v.a. $\sim N(0,1)$ cînd i.n. H_0 este adevărată.
6. Regiunile critice corespunzătoare diferitelor i.a. H_1 sînt trecute în tab. 93.

Tabelul 93

i.n. $H_0 ; \rho = \rho_0$		
i.a. H_1	Regiunea critică	Test statistic
a) $\rho \neq \rho_0$	$\lvert Z \rvert > z_{\alpha/2}$	t.b.
b) $\rho > \rho_0$	$Z > z_\alpha$	t.u.dr.
c) $\rho < \rho_0$	$Z < z_{1-\alpha} = -z_\alpha$	t.u.stg.

Exemplu
c) Un eşantion de volum $n = 28$ dintr-o populaţie în care caracteristica este vectorul bidimensional (X, Y) dă $r = 0{,}65$. Se verifică i.n. : $H_0 : \rho = 0{,}72$ faţă de i.a. $H_1 : \rho < 0{,}72$.

Se calculează statistica :

$$Z = \frac{\frac{1}{2}\ln\frac{1+0{,}65}{1-0{,}65} - \frac{1}{2}\ln\frac{1+0.72}{1-0.72}}{\sqrt{25}} = 0{,}665.$$

Se alege p.s. al t.s., $\alpha = 0.05$.
Deoarece valoarea statistică Z nu este semnificativă :

$$Z = -0{,}665 > z_{0,95} = -z_{0,05} = -1{,}64$$

se acceptă i.n.H_0 cu p.s. $\alpha = 0.05$.
Observații
1. *T.c.c.t. ZF.* este un test aproximativ. Legea de repartiție exactă a v.a. r depinde de ρ și n și este destul de complicată.
2. *Robustețea t.c.c.t.ZF. : v.a.* transformată Z are asimptotic legea normală și cînd vectorul aleator (X, Y) nu are legea normală bidimensională, cu condiția ca n să fie suficient de mare.
Asimetria repartițiilor marginale ale vectorului aleator (X, Y) are efect relativ mic asupra testului, în timp ce nerespectarea valorii *excesului* v.a. normale ($\beta_2 = 3$) are efect asupra testului.
3. Valorile transformatei Z sînt tabelate.
4. Un *interval de încredere** pentru $\frac{1}{2}\ln\frac{1+\rho}{1-\rho}$ cu coeficientul de încredere $1-\alpha$ este :

$$\left(\frac{1}{2}\ln\frac{1+r}{1-r} - z_{\alpha/2}\sqrt{\frac{1}{n-3}};\ \frac{1}{2}\ln\frac{1+r}{1-r} + z_{\alpha/2}\sqrt{\frac{1}{n-3}}\right).$$

Test statistic referitor la egalitatea a doi coeficienți de corelație.
1. Se consideră 2 eșantioane ($\rightarrow$*Test de independență*) formate din n_1, respectiv n_2 puncte aleatoare.
2. Vectorii aleatori (X_1, Y_1) respectiv (X_2, Y_2) au *legea normală bidimensională** cu coeficienții de corelație ρ_1, respectiv ρ_2.
Observațiile sînt independente, fiecare observație fiind compusă din două valori : o valoare x și o valoare y. Perechile (x, y) sînt presupuse independente, dar valorile dintr-o pereche nu sînt presupuse a fi independente.

3) a) i.n. $H_0 : \rho_1 = \rho_2$ (t.b.)
i.a. $H_1 : \rho_1 \neq \rho_2$

b) $H_0 : \rho_1 = \rho_2$ (t.u.dr.)
$H_1 : \rho_1 > \rho_1$

c) $H_0 : \rho_1 = \rho_2$ (t.u.stg.)
$H_1 : \rho_1 < \rho_2$

4. Se alege p.s. al testului, α.

5. Se calculează coeficienții de corelație de selecție r_1, respectiv r_2 pentru cele două eșantioane. (→ *Test de independență* pct. (5)).
Se calculează :

$$z_1 = \frac{1}{2} \ln \frac{1 + r_1}{1 - r_1} \quad \text{(transformatele } z_1, z_2 \text{ ale lui Fisher)}$$

$$z_2 = \frac{1}{2} \ln \frac{1 + r_2}{1 - r_2}$$

$$S^2 = \frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}.$$

Statistica testului :

$$Z = \frac{z_1 - z_2}{S}.$$

6. Regiunile critice pentru diferite i.a. H_1 sînt date în tab. 94.

Tabelul 94

i.n. $H_0 : \rho_1 = \rho_2$

i.a. H_1	Regiunea critică	Test statistic
a) $\rho_1 \neq \rho_2$	$\lvert Z \rvert > z_{\alpha/2}$ sau $Z < -z_{\alpha/2} = z_{1-\alpha/2}$ $Z > z_{\alpha/2}$	t.b.
b) $\rho_1 > \rho_2$	$Z > z_\alpha$	t.u.dr.
c) $\rho_1 < \rho_2$	$Z < z_{1-\alpha} = -z_\alpha$	t.u.stg.

Exemplu
a) Înălțimile și greutățile a $n_M = 10$ masculi și $n_F = 8$ femele aleși întîmplător dintr-o populație de animale sînt date în tab. 95.
Se verifică dacă coeficientul de corelație între înălțime și greutate este același pentru masculi și femele, adică :

$$\text{i.n. } H_0 : \rho_M = \rho_F$$

față de i.a.

$$H_1 : \rho_M \neq \rho_F.$$

Tabelul 95

Masculi		Femele	
Înălțime (cm)	Greutate (kg)	Înălțime (cm)	Greutate (kg)
166	63	161	60
187	98	158	61
170	82	165	68
171	80	154	52
182	86	166	67
188	85	170	68
175	84	167	69
183	79	177	80
179	75		
178	76		

Se calculează :

$$r_M = 0{,}6744 \qquad z_M = 0{,}8188$$

$$r_F = 0{,}9665 \qquad z_F = 2{,}0362$$

$$S^2 = 0{,}34286$$

corespunzătoare celor două eșantioane de volume $n_M = 10$, respectiv $n_F = 8$
Se calculează statistica testului :

$$Z = \frac{0{,}8188 - 2{,}0362}{\sqrt{0{,}34286}} = -2{,}08.$$

Se alege p.s. al testului, $\alpha = 0{,}05$.
Deoarece valoarea statisticii Z este semnificativă :

$$|Z| = 2{,}08 > z_{0{,}025} = 1{,}96$$

se respinge i.n. H_0 cu p.s. $\alpha = 0{,}05$ și se conchide că : $\rho_M \neq \rho_F$.

Test statistic referitor la egalitatea a k coeficienți de corelație

1. Se consideră k eșantioane (→ *Test de independență*) formate din $n_1, \ldots, n_k$, respectiv puncte aleatoare.
2. Vectorii aleatori $(X_1, Y_1), \ldots, (X_k, Y_k)$ au *legea normală bidimensională** cu coeficienții de corelație $\rho_1, \rho_2, \ldots, \rho_k$, respectiv.
Observațiile sînt independente. (→ *Test stat. egal. 2 coef. corel.*)*
3. i.n. $H_0 : \rho_1 = \ldots = \rho_k$
i.a. H_1 : i.n. H_0 nu este adevărată.
4. Se alege p.s. al testului, α.
5. Se calculează coeficienții de corelație de selecție $r_1, \ldots, r_k$, respectiv, pentru cele k eșantioane. (→ *Test de independență*, (5)).

Se calculează :

$$z_i = \frac{1}{2} \ln \frac{1 + r_i}{1 - r_i}, \; i = 1, \ldots, k \; \textit{(transformatele } z_i \textit{ ale lui Fisher*)}$$

$$N = (n_1 + n_2 + \ldots + n_k) - 3k$$

$$\bar{z} = \{(n_1 - 3)\, z_1 + \ldots + (n_k - 3)\, z_k\}/N.$$

Statistica testului :

$$\chi^2 = \sum_{i=1}^{k} (n_i - 3) z_i - \frac{1}{N} \bar{z}^2 =$$

$$= \sum_{i=1}^{k} (n_i - 3) \left(\frac{1}{2} \ln \frac{1 + r_i}{1 - r_i} \right)^2 - \frac{1}{\sum_{i=1}^{k} (n_i - 3)} \left[\sum_{i=1}^{k} (n_i - 3) \frac{1}{2} \ln \frac{1 + r_i}{1 - r_i} \right]^2$$

este o v.a. $\sim \chi^2_{k-1}$ cînd i.n. H_0 este adevărată.
6. Regiunea critică a t.s. : se respinge i.n. H_0 dacă

$$\chi^2 > \chi^2_{\alpha;\, k-1}.$$

Exemplu. Se consideră 5 eşantioane independente de volume 21, 30, 39, 26 şi 35 din populaţii în care caracteristica sub cercetare este un vector aleator bidimensional. Se calculează coeficienţii de corelaţie de selecţie r_i, $i = 1, \ldots, 5$: 0,39 ; 0,61 ; 0,43 ; 0,54 ; 0,48.
Se verifică i.n.

$$H_0 : \rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_5$$

faţă de i.a. H_1 : i.n. H_0 nu este adevărată.
Se alege p.s. al testului, $\alpha = 0{,}05$.
Calculele pot fi urmărite în tab. 96

Tabelul 96

r_i	$\frac{1}{2} \ln \frac{1+r_i}{1-r_i}$	$n_i - 3$	$(n_i-3) \frac{1}{2} \ln \frac{1+r_i}{1-r_i}$	$(n_i-3) \left(\frac{1}{2} \ln \frac{1+r_i}{1-r_i} \right)^2$
0,39	0,412	18	7,416	3,055
0,61	0,709	27	19,143	3,572
0,43	0,460	36	16,560	7,618
0,54	0,604	23	13,892	8,391
0,48	0,524	32	16,768	8,786
		136	73,799	41,422

Se calculează statistica testului:

$$\chi^2 = 1{,}4$$

Deoarece valoarea statisticii χ^2 nu este semnificativă:

$$\chi^2 = 1{,}4 < \chi^2_{0{,}05;\,4} = 9.49$$

se acceptă i.n. privind egalitatea coeficienților de corelație pentru cele 5 eșantioane.

Observație. Pentru $k = 2$ se obține *t.s. referitor la egalitatea a 2 coef. corel**. Valoarea statisticii χ^2 pentru $k=2$ este pătratul valorii statisticii Z a testului menționat.

TESTE PRIVIND REPARTIȚIA POISSON. Teste pentru verificarea concordanței unei repartiții empirice cu *legea Poisson**; teste referitoare la parametrul λ al legii Poisson, adică teste referitoare la medii.

Testarea concordanței cu legea Poisson. Verificarea ipotezei statistice că populația din care provine eșantionul studiat are legea Poisson. Se consideră un eșantion de volum n, format din n observații independente, $X_1, \ldots, X_n$, având valori întregi nenegative și provenind dintr-o populație cu legea necunoscută.

Test de concordanță χ^2 cu legea Poisson. Se aplică *testul de concordanță χ^2, cazul parametric**, după ce s-a obținut un estimator al parametrului λ al legii Poisson sub forma:

$$\hat{\lambda} = \overline{X},$$

unde $\overline{X}$ este *media de selecție** calculată pe baza eșantionului de volum n considerat.

Test de concordanță Kolmogorov-Smirnov cu legea Poisson. Se aplică cu unele precauții, și anume: decizia luată în testul de concordanță Kolmogorov-Smirnov poate să nu fie corectă atunci cînd valoarea statisticii testului de concordanță Kolmogorov-Smirnov este doar puțin mai mică decît valoarea critică tabelată.

În plus, nu se cunoaște efectul estimării parametrului λ al legii Poisson asupra statisticii testului, așa cum se întîmplă în cazul *testului de concordanță χ^2, cazul parametric**. Se știe doar că adevăratul p.s. al testului Kolmogorov-Smirnov este mai mic decît valoarea nominală a p.s. α ales și că probabilitatea erorii de genul II crește.

Aceste aspecte nedorite ale aplicării testului sînt atenuate atunci cînd numărul parametrilor estimați este mic față de volumul eșantionului considerat.

Test preliminar bazat pe verificarea egalității mediei și dispersiei v.a. Poisson. Testul concordanței repartiției empirice cu legea Poisson bazat pe proprietatea *legii Poisson** de a avea media și dispersia egale. Se știe că v.a. $X \sim Po(\lambda)$ are:

$$E(X) = \mathrm{Var}\ X = \lambda.\ (\rightarrow \textit{legea Poisson})^*$$

Dacă estimatorii medici și dispersiei obținuți pe baza eșantionului considerat nu diferă foarte mult, se poate conchide că eșantionul provine dintr-o populație Poisson.

1. Se consideră un eșantion de volum n: $X_1, \ldots, X_n$. Toate observațiile sînt independente și sînt numere întregi nenegative.
2. Legea v.a. X este $Po(\lambda)$.

Modelul statistic postulează că eșantionul considerat provine dintr-o populație statistică de caracteristici întregi nenegative, având o lege *unimodală* cu media m și dispersia σ^2, apropiată de legea normală $N(m, \sigma^2)$

2. H_0 : $\sigma^2 = m$ (t.b.)

H_1 : $\sigma^2 \neq m$

4. Se alege p.s. α al testului.

5. Statistica testului :

$$\chi^2 = (n-1)\, s^2/\overline{X}$$

unde $\overline{X}$ este *media de selecție**, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

s^2 este *dispersia de selecție**, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ este o v.a. $\sim \chi^2_{n-1}$, cînd i.n. H_0 este adevărată.

6. Regiunea critică este formată din mulțimea valorilor de selecție mai mari decît $\alpha/2$ — cuantila superioară sau mai mici decît $\alpha/2$ — cuantila inferioară a legii χ^2 cu $(n-2)$ g.l. :

$$|\chi^2| > \chi^2_{\alpha/2\,;\,(n-2)}$$

Așadar, se respinge i.n. H_0 conform căreia populația teoretică este Poisson dacă valoarea statisticii χ^2 este semnificativă.

Observații :

1) *T.p.P.* propus este un test aproximativ.

Se aplică numai dacă $X > 10$.

2) În expresia statisticii χ^2 a testului se înlocuiește σ^2 — dispersia teoretică necunoscută prin $\overline{X}$ — estimatorul mediei și dispersiei v.a. $Po(\lambda)$.

3) *T.p.P.* presupune validitatea unei legi unimodale, care să fie aproximativ normală.

T.p.P. arată mai degrabă împrăștierea valorilor eșantionului decît forma legii teoretice postulate.

4) Pentru mai multă siguranță a inferenței statistice se recomandă *testul de concordanță cu legea Poisson* *.

Teste privind variabila aleatoare ~ Po(λ).

Legea Poisson depinde de un singur parametru λ (media sau dispersia v.a. $\sim Po(\lambda)$). Așadar, *t.c. l.P.* adică *v.a.* $\sim Po(\lambda)$ vor fi teste referitoare la medii. Cînd $\lambda \geqslant 9$ și volumul eșantionului este suficient de mare se folosește aproximarea legii Poisson prin *legea normală**.

Test unilateral pentru media λ. Test unilateral pentru verificarea i.n. :

$$H_0 : E(X) = \lambda \leqslant \lambda_0$$

față de i.a.

$$H_1 : \lambda > \lambda_0 \qquad \text{(t.s.u.dr.)}$$

unde v.a. $X \sim Po(\lambda)$.

1. Se consideră un eșantion de volum n : $X_1, X_2, \ldots, X_n$. Toate observațiile sînt independente și sînt numere nenegative.

2. V.a. $X \sim Po(\lambda)$, λ parametru necunoscut.

3. $H_0 : \lambda \leqslant \lambda_0$ (t.s.u.dr.)

$H_1 : \lambda < \lambda_0$,

respectiv

$H_0 : \lambda \geqslant \lambda_0$ (t.s.u.stg.)

$H_1 : \lambda < \lambda_0$

4. Se alege p.s. al testului, α.
5. Se calculează :

$$s = \sum_{i=1}^{n} X_i = n\overline{X} = n\hat{\lambda}.$$

6. I.n. H_0 se respinge cu p.s. α, față de i.a. H_1 dacă :

$$P = \sum_{k=s}^{\infty} \frac{(n\lambda_0)^k}{k!} e^{-n\lambda_0} < \alpha,$$

respectiv :

$$Q = \sum_{k=0}^{s} \frac{(n\lambda_0)^k}{k!} e^{-n\lambda_0} < \alpha.$$

Valorile statisticii P, respectiv statisticii Q pot fi tabelate pentru diverse valori ale lui λ_0, n, s.

Exemplu. Numărul mediu de accidente pe lună, datorită vitezei, este 10. În prima lună după introducerea unor restricții de viteză, acest număr scade la 4. Presupunem că numărul de accidente pe lună, după introducerea restricțiilor de viteză, este o v.a. Poisson de medie λ.
Se alege p.s. $\alpha = 0{,}05$, de exemplu.
Se verifică i.n.

$$H_0 : \lambda = 10$$

față de i.a.

$$H_1 : \lambda \neq 10.$$

Deoarece $n = 1$,

$$Q = \sum_{k=0}^{s} e^{-n\lambda_0} \frac{(n\lambda_0)^k}{k!} = \sum_{k=0}^{4} e^{-10} \frac{10^k}{k!} = 0{,}029$$

și ținînd seama de regiunea critică a testului unilateral sting se respinge i.n. H_0. conform căreia noile restricții conduc la micșorarea numărului de accidente. Pentru valori mari ale lui n, v.a.

$$z = \frac{s - n\lambda}{\sqrt{n\lambda}} = \frac{\overline{X} - \lambda}{\sqrt{\lambda/n}}$$

are legea $N(0,1)$ și servește pentru verificarea diferitelor ipoteze alternative ipotezei $H_0 : \lambda = \lambda_0$.

În tab. 97 sînt trecute i.a. și regiunea critică corespunzătoare unde $z = \frac{\overline{X} - \lambda}{\sqrt{\lambda/n}} \sim$

$\sim$ (N (0,1) dacă i.n. este adevărată : $\lambda = \lambda_0$.

Tabelul 97

$H_0 : \lambda = \lambda_0$		
H_1	Regiunea critică	Test statistic
$\lambda \neq \lambda_0$	$\lvert z \rvert > z_{\alpha/2}$	t.b.
$\lambda < \lambda_0$	$z < z_{1-\alpha}$	t.u.stg.
$\lambda > \lambda_0$	$z > z_\alpha$	t.u.dr.

Pentru n suficient de mare, se utilizează aproximarea legii $Po(\lambda)$ prin legea normală.
V.a.

$$z = \frac{s - n\lambda_0}{\sqrt{n\lambda_0}} = \frac{X - \lambda_0}{\sqrt{\lambda_0/n}}$$

este v.a. $\sim$ N (0,1).
Se respinge i.n. H_0 față de i.a. H_1, cu p.s. α dacă :

$$z > z_\alpha, \qquad \text{(t.s.u.dr.)}$$

unde z_α este α-cuantila superioară a legii $N(0,1)$.
Test bilateral pentru media λ. Test bilateral pentru verificarea valorii specificate a parametrului λ al legii $Po(\lambda)$
1.—2. (→ *Test unilateral pentru media* λ).
3. Se verifică i.n.

$$H_0 : \lambda = \lambda_0$$

față de i.a. :

$$H_1 : \lambda \neq \lambda_0.$$

4.—5. (→*Test unilateral pentru media* λ).
6. Pentru n suficient de mare se utilizează aproximarea legii Po (λ) prin legea normală :
Regiunea critică a t.s. bilateral este : se respinge i.n. H_0 cu p.s. față de i.a. H_1 dacă :

$$\lvert z \rvert = \left| \frac{X - \lambda_0}{\sqrt{\lambda_0/n}} \right| > z_{\alpha/2},$$

unde $z_{\alpha/2}$ este $\alpha/2$-cuantila bilaterală a legii $N(0,1)$.
Test pentru egalitatea a două medii λ_1, λ_2.
1 —2. Fie v.a. $X_1 \sim Po\ (\lambda_1)$ și $X_2 \sim Po(\lambda_2)$, v.a. independente.
Atunci v.a. $X_1 - X_2 \sim N\ (0, 2\lambda)$.

3. Se verifică i.n.

$$H_0 : \lambda_1 = \lambda_2$$

față de i.a.

$$H_1 : \lambda_1 > \lambda_2 \quad \text{(t.u.dr.)}$$

4. Se alege p.s. α al testului.
5. Statistica testului :

$$z = \frac{\left(x_1 - \frac{1}{2}\right) - \left(x_2 + \frac{1}{2}\right)}{x_1 + x_2},$$

unde x_1, x_2 sînt valorile observate ale v.a. X_1, X_2.
6. Regiunea critică este formată din mulțimea valorilor eșantionului mai mari decît α-cuantila superioară a legii $N(0,1)$.
Așadar, se respinge i.n. H_0 față de i.a. H_1 dacă :

$$z = \frac{\left(x_1 - \frac{1}{2}\right) - \left(x_2 + \frac{1}{2}\right)}{\sqrt{x_1 + x_2}} > z_\alpha.$$

Exemplu
1.—2. Numărul de morți în accidente auto, într-un oraș, în două săptămîni consecutive, a fost de $x_1 = 6$, respectiv $x_2 = 3$. Presupunînd că numărul de morți în accidente auto este o variabilă Poisson de medie λ, ne propunem să verificăm i.n.
3. $H_0 : \lambda_1 = \lambda_2$
față de i.a.

$$H : \lambda_1 > \lambda_2.$$

4. Se alege p.s. al testului, $\alpha = 0{,}01$ de exemplu.
5. Statistica testului :

$$z = \frac{\left(6 - \frac{1}{2}\right) - \left(3 + \frac{1}{2}\right)}{\sqrt{9}} = 0{,}66.$$

Regiunea critică : $z < z_{0,01} = 2.33$ și acceptă i.n. H_0.
Test pentru egalitatea a k medii a k legi $Po(\lambda_i)$, $i = 1, k$; $k \geqslant 2$.
1.—2. Fie $X_1, X_2, \ldots, X_k$ v.a. independente repartizate Poisson de parametri $\lambda_1, \lambda_2, \ldots, \lambda_k$; $k \geqslant 2$.
3. Se verifică i.n.

$$H_0 : \lambda_1 = \lambda_2 = \ldots = \lambda_k$$

față de i.a.

$$H_1 : \lambda_i \neq \lambda_j, \quad i, j \in \{1, \ldots, k\}, \quad i \neq j.$$

4. Se alege p.s. α al testului.

5. Statistica testului :

$$X^2 = \sum_{i=1}^{k} \frac{(x_i - \bar{x})^2}{\bar{x}} .$$

6. Regiunea critică : $X^2 > \chi^2_{\alpha;(k-1)}$, unde x_i sînt valorile observate ale v.a. X_i, $i = 1, \ldots, k$.

În eșantioane a cîte 1000 piese fiecare s-au găsit x_i = 15, 13, 8, 6, 11, 9, 14, 10, 16, 9, 12 piese defecte, respectiv.

Deoarece probabilitatea ca să obținem o piesă defectă este foarte mică, frecvențele observate $x_{i;\,i=1,\ldots,11}$ pot fi considerate ca provenind din populații Poisson.

Fie X_i — v.a. Poisson ce caracterizează populația din care provine eșantionul i.

$$X_i \sim Po(\lambda_i), \quad i = 1, \ldots, k.$$

Se verifică i.n. privind egalitatea mediilor λ_i ale v.a. X_i, adică :

$$H_0 : \lambda_1 = \ldots = \lambda_{11}.$$

Se calculează :

$$\bar{x} = 11.2 ; \sum_{i=1}^{11} (x_i - \bar{x})^2 = 97.6 : \sum_{i} \frac{(x_i - \bar{x})^2}{\bar{x}} = 8.7.$$

Se alege p.s. al testului $\alpha = 0{,}05$, de exemplu.

Din tabelele legii χ^2 pentru $k = 11$ (numărul eșantioanelor) se determină 0,05 — cuantila superioară a legii χ^2 :

$$\chi^2_{0{,}05;\,10} = 18{,}3.$$

Deoarece valoarea statisticii X^2 nu este semnificativă :

$$X^2 = 8.7 < \chi^2_{0{,}05\,;\,10} = 18{,}31$$

se acceptă i.n. H_0 și putem considera că cele 11 eșantioane a cîte 1000 piese fiecare provin din populații omogene din punctul de vedere al pieselor defecte.

TESTE PRIVIND REPARTIȚIA BINOMIALĂ. T.s. pentru verificarea concordanței unei repartiții empirice cu *legea binomială* Bi (x, p)* ; t.s. referitoare la parametrul p al legii binomiale etc.

Teste de concordanță cu legea binomială Bi (x, p).

Test de concordanță χ^2 cu legea binomială (cazul neparametric).

Fie X o v.a. reprezentînd numărul de realizări („succese") ale unui anumit eveniment în n probe independente pentru care probabilitatea obținerii unui „succes" este p.

Repartiția v. a. X este *legea binomială**.

T.c. l.b. revine la verificarea i.n.

$$H_0 : X \sim Bi(x, p)$$

față de i.a. :

$$H_1 : X \text{ nu are legea } Bi(x, p).$$

De exemplu, v.a. X reprezintă numărul de produse defecte dintr-un eșantion de volum n extras dintr-un lot de volum N ($n \leqslant N$), p fiind probabilitatea ca un produs să fie defect la o verificare.

Fie F_i, $1 \leqslant i \leqslant k$ frecvențele observate, adică numărul de realizări ale evenimentului corespunzător v.a. X, în k clase și fie f_i frecvențele teoretice respective, conform i.n. H_0.
Dacă i.n. H_0 este adevărată, statistica :

$$\chi^2 = \sum_{i=1}^{k} \frac{(F_i - f_i)^2}{f_i}$$

adică :

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} \quad (\rightarrow \textit{Test de concordanță } \chi^2)$$

este o v.a. χ^2_ν cu ν g.l., unde $\nu = k -$ nr. de relații dintre F_i și f_i, folosite pentru a determina pe f_i) $= k-1$, ținînd seama că

$$\sum_i f_i = \sum_i F_i.$$

Se procedează în continuare ca la *testul de concordanță** χ^2.
Exemplu. Pentru a verifica rezistența la îndoire, un număr de $N = 80$ bare metalice sînt supuse unor încercări. Experiența se face îndoind în $n = 5$ puncte fiecare bară și notînd X — numărul de puncte în care fiecare bară a prezentat rupturi ($X = 0, 1, 2, 3, 4, 5$).
Se verifică ipoteza legii binomiale a v.s. X, știind din verificări anterioare că probabilitatea producerii unei rupturi la îndoire este $p = 0.4$ în fiecare dintre încercări.
Fie F_i — numărul barelor ce au $X = 0,1, \ldots, 5$ sau mai multe rupturi și fie f_i — frecvențele teoretice corespunzătoare.
Se calculează

$$P = P(X = n_1/n = 5) = C_5^{n_1}(0,4)^X(1 - 0,4)^{5-X}, \quad n_1 = 0,1, \ldots, 5.$$

Fiecare valoare P multiplicată cu $N = 80$ dă f_i — frecvențele teoretice (coloana a 8-a a tab. 98)

$$f_i = 80 \cdot P(X = n_1/n = 5).$$

Calculele pot fi urmărite în tab. 98.
Se alege p.s. al testului, $\alpha = 0,05$, de exemplu.
Ținînd seama că există o singură relație de legătură :

$$\sum_i f_i = \sum_i F_i = 80$$

și $k = 6$ clase de valori, din tabelele legii χ^2 se citește 0.05-cuantila legii χ^2 cu $6-1 = 5$ g.l. :

$$\chi^2_{0,05;5} = 11.1.$$

Deoarece valoarea statisticii χ^2 este semnificativă :

$$\chi^2 = 31,215 > \chi^2_{0,05;5} = 11,1$$

se respinge i.n. H_0 cu p.s. $\alpha = 0,05$, adică legea v.a. X nu concordă cu legea binomială.

Tabelul 98

X	Frecvența observată F_i	$P = P(X = n_1/n = 5)$	Coeficienții binomiali	$(0,4)^x$	$(0,6)^{5-x}$	P	Frecvența teoretică f_i	$F_i - f_i$	$\frac{(F_i - f_i)^2}{f_i}$
0	13	$C_5^0 (0,4)^0 (0,6)^5$	1	1	0,07776	0,07776	6,2208	6,7792	7,388
1	18	$C_5^1 (0,4)^1 (0,6)^4$	5	0,4	0,1296	0,25920	20,7360	−2,7360	0,361
2	20	$C_5^2 (0,4)^2 (0,6)^3$	10	0,16	0,2160	0,34560	27,6480	−7,6480	2,116
3	18	$C_5^3 (0,4)^3 (0,6)^2$	10	0,064	0,3600	0,23040	18,4320	−0,4320	0,010
4	6	$C_5^4 (0,4)^4 (0,6)^1$	5	0,0256	0,6	0,07680	6,1440	−0,1440	0,003
5	5	$C_5^5 (0,4)^5 (0,6)^0$	1	0,01024	1	0,01024	0,8192	4,1808	21,337

$\sum_i F_i = 80$

$\sum_i f_i = 80 \quad \sum_i \frac{(F_i - f_i)^2}{f_i} = 31,215$

Test de concordanță χ^2 cu legea binomială (cazul parametric). *Test de concordanță** cu legea binomială cînd nu se cunoaște valoarea parametrului p-probabilitatea obținerii unui „succes" și urmează a fi estimat pe baza datelor eșantionului considerat, adică :

$$\hat{p} = \sum_i XF_i(X)/n.N,$$

ceea ce duce la pierderea a 2 g.l.
Se procedează ca în cazul i) anterior ținînd seama că statistica χ^2 :

$$\chi^2 = \sum_{i=1}^{k} \frac{(F_i - f_i)^2}{f_i}$$

este o v.a. $\sim \chi^2_\nu$ cu $\nu = k-1-1 = k-2$ g.l. cînd H_0 este adevărată.
Exemplu. Reluînd exemplul din cazul i) anterior și presupunînd că p este necunoscut, se estimează valoarea acestui parametru prin :

$$\hat{p} = \frac{\sum_i XF_i(X)}{5 \cdot 80} = \frac{161}{400} = 0{,}4025.$$

În continuare se procedează în mod analog, alcătuind un tabel similar cu tab. 98, utilizînd $\hat{p} = 0{,}4025$ și ținînd seama că χ^2 este o v.a. χ^2 cu $6-2=4$ g.l. cînd i.n. H_0 este adevărată.

Test pentru o valoare specificată p_0 a parametrului legii binomiale Bi(n, p)

i) *Test exact*

1. Se consideră un eșantion de volum n obținut din n probe independente, în care se obțin $X = r$ realizări („succese"). În fiecare probă probabilitatea de a obține un „succes" este p. Proporția observată de „succese" este r/n.
2. $X \sim Bi\ (n, p)$.
3. a) $H_0 : p = p_0$ (valoare specificată a proporției).

   $H_1 : p \neq p_0$ (t.b.)

   b) $H_0 : p \leqslant p_0$ (t.u.dr.)

   $H_1 : p > p_0$

   c) $H_0 : p \geqslant p_0$ (t.u.stg.)

   $H_1 : p < p_0$.

4. Se alege p.s. al t.s. : $\alpha = 0{,}05$, de exemplu.
5. Se calculează statistica :

$$P = \sum_{r=x}^{n} C_n^r p_0^r (1 - p_0)^{n-r}.$$

Probabilitățile binomiale :

$$P(X = r/n, p_0) = C_n^r p_0^r (1 - p_0)^{n-r}$$

se determină prin calcul direct sau folosind *tabelele legii binomiale* (tab. 99), iar

$$P = \sum_{r=x}^{n} C_n^r p_0^r (1 - p)^{n-r}$$

se obține prin adunarea acestora.
6. Regiunea critică a t.s. :
b) se respinge i.n. H_0 cu p.s. dacă :

$$P < \alpha$$

sau folosind v.a. $X = r$ drept statistica testului :
— regiunea critică ce include $\alpha/2$-cuantila inferioară și $\alpha/2$-cuantila superioară a legii binomiale $Bi(n, p_0)$.
— regiunea critică ce include α-cuantila superioară a legii binomiale $Bi(n, p_0)$.
— regiunea critică ce include α-cuantila inferioară a legii binomiale $Bi(n, p_0)$, respectiv a legii binomiale $Bi(n, 1/2)$. Regiunea critică cuprinde valoarea tabelată :

Tabelul 99

n	$\alpha = 0{,}005$	0,01	0,025	0,05
5	—	—	—	5
6	—	—	6	6
7	—	7	7	7
8	8	8	8	7
9	9	9	8	8
10	10	10	9	9
11	11	10	10	9
12	11	11	10	10
13	12	12	11	10
14	13	12	12	11
15	13	13	12	12
16	14	14	13	12
17	15	14	13	13
18	15	15	14	13
19	16	15	15	14
20	17	16	15	15

Exemplu. Cu ocazia unor săpături se descoperă 10 fosile, dintre care 6 sînt clasificate ca aparținînd unor femei și 4 unor bărbați. Se verifică ipoteza conform căreia raportul dintre sexe este 1 : 1 sau sînt mai multe fosile de femei, adică i.n. :

$$H_0 : p = \frac{1}{2}$$

față de i.a.

$$H_1 : p > \frac{1}{2},$$

unde p este proporția de femei.
Se calculează :

$$P = \sum_{r=6}^{10} C_{10}^{r} \left(\frac{1}{2}\right)^{r} \left(\frac{1}{2}\right)^{10-r} = \frac{1}{2^{10}} (210+120+45+10+1) = 0{,}377$$

folosind tabelele binomiale ;

$$P = P\left(X = 6/10, \frac{1}{2}\right) + P\left(X = 7/10, \frac{1}{2}\right) + \ldots + P\left(X = 10/10, \frac{1}{2}\right) =$$

$$= 0{,}0251 + 0{,}1172 + 0{,}0439 + 0{,}0098 + 0{,}0010 = 0{,}377.$$

Se alege p.s. al t.s., $\alpha = 0{,}01$, de exemplu.
Deoarece valoarea lui P nu este semnificativă :

$$P = 0{,}377 > 0{,}01$$

se acceptă i.n.H_0 cu p.s. $\alpha = 0{,}01$, adică adevărata proporție de fosile femei este 1/2 și deci raportul între sexe este 1 : 1.
ii) *Test exact bazat pe legea F*
Folosirea tabelelor legii binomiale este incomodă, de aceea se recomandă folosirea legii F pentru t.s. corespunzător parametrului p al legii $Bi(n, p)$.
1.—4. (→ i) *Test exact*) (cazul precedent).
5. Se calculează statistica :

$$Y = \frac{r}{n+1-r} \frac{1-p_0}{p_0}$$

sau

$$Z = \frac{n-r}{r+1} \frac{p_0}{1-p_0}.$$

Dacă i.n. H_0 este adevărată, v.a. $Y \sim F_{2(n+1-r),\,2r}$, iar v.a. $Z \sim F_{2(r+1),\,2(n-r)}$
6. Regiunea critică a testului pentru diferite forme ale i.a. H_1 :
a) se respinge i.n. H_0 dacă :

$$Y > F_{\alpha/2;\,2(n-1-r),\,2r}$$

sau

$$Z > F_{\alpha/2;\,2(r+1),\,2(n-r)},$$

unde $F_{\alpha/2;\,\nu_1,\,\nu_2}$ este $\alpha/2$-cuantila superioară a legii F^* cu ν_1, ν_2 g.l.
b) se respinge i.n.H_0 dacă

$$Y > F_{\alpha;\,2(n+1-r),\,2r}$$

unde $F_{\alpha;\,\nu_1,\,\nu_2}$ este α-cuantila superioară a legii F cu ν_1, ν_2 g.l.

c) se respinge i.n.H_0 dacă

$$Z > F_{\alpha;\, 2(r+1),\, 2(n-r)}$$

ţinîndu-se seama de relaţia între cuantilele legii F, şi anume :

$$F_{\alpha;\, \nu_1,\, \nu_2} = \frac{1}{F_{1-\alpha;\, \nu_2,\, \nu_1}}.$$

Exemplu. Într-un lot de $n = 100$ difuzoare s-au găsit $r = 8$ difuzoare care prezintă defecte : piese nelipite, impedanţă necorespunzătoare etc.
Pentru difuzoarele de tipul studiat se ştie din experienţa anterioară că proporţia de produse defecte este :

$$p = 0{,}091$$

şi există motive să bănuim că pentru perioada urmărită proporţia de produse defecte a crescut.
Se verifică i.n. :

$$H_0 : \; p = p_0 = 0{,}091$$

faţă de i.a.

$$H_1 : \; p > p_0 \text{ (cazul b).}$$

Se calculează :

$$Y = \frac{8}{100 + 1 - 8} \cdot \frac{0{,}909}{0{,}091} = 0.86.$$

Se alege p.s. al t.s., $\alpha = 0{,}05$.
Deoarece valoarea statisticii Y nu este semnificativă :

$$Y = 0{,}086 < F_{0{,}05;\, 186{,}16} = 2{,}01$$

se acceptă i.n. H_0 cu p.s. $\alpha = 0{,}05$, adică se acceptă şi pentru perioada urmărită o proporţie de piese defecte egală cu 0,091.
iii) *Test bazat pe aproximarea legii binomiale Bi (n, p) cu legea normală* $N(0,1)$ 1.—4. (→ i) *Test exact*) (→ tab. 67).
5. Pentru valori mari ale lui n, $(n > 20)$ v.a.

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{r - np}{\sqrt{np(1-p)}} = \frac{\dfrac{X}{n} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$$

are, aproximativ, legea normală $N(0,1)$ cînd i.n. H_0 este adevărată.
Se procedează în continuare ca la *testul Z**.
6. Regiunea critică corespunzătoare diferitelor i.a. sînt trecute în tab. 100. unde

$$Z = \frac{r - np_0}{\sqrt{np_0(1-p_0)}}.$$

Tabelul 100

	$H_0 : p = p_0$	Test statistic
H_1	Regiunea critică	
a) $p < p_0$	$\|Z\| > z_{\alpha/2}$	t.s.b.
b) $p > p_0$	$Z > z_\alpha$	t.s.u.dr.
c) $p < p_0$	$Z < z_{1-\alpha}$	t.s.u.stg.

Exemplu. Într-un lot de 80 de difuzoare C_4 s-au găsit 10 produse care prezintă diferite defecte : blocat, întrerupt, vibrații, lipsuri, neglijențe etc. Dintr-un studiu anterior al producției difuzorului (I chenzină) se știe că proporția de produse defecte este $P_0 = 0{,}091$ și de asemenea există motive să bănuim că pentru perioada urmărită proporția de produse defecte a crescut.
Se verifică i.n.

$$H_0 : p = p_0 = 0{,}091$$

față de i.a.

$$H_1 : p > p_0 \text{ (cazul b).}$$

Pentru $n = 80$, $r = 10$, $p_0 = 0{,}091$ se calculează

$$Z = \frac{10 - 80 \cdot 0{,}091}{\sqrt{80 \cdot 0{,}091 \cdot 0{,}909}} = \frac{2{,}22}{2{,}57} = 0{,}86$$

Se alege p.s. al testului, $\alpha = 0{,}05$.
Deoarece valoarea statisticii Z este semnificativă :

$$Z = 0{,}86 > z_{0{,}05} = -z_{0{,}95} = -1{,}65,$$

se respinge i.n. H_0 cu p.s. $\alpha = 0{,}05$, adică proporția de produse defecte nu a crescut față de perioada I chenzină.
c) Într-un proces de fabricație proporția articolelor defecte este $p = 10\%$. După introducerea unor noi metode de muncă, proporția de defecte a scăzut, astfel încît într-un eșantion de 100 de articole se găsesc 15 articole defecte. Se verifică dacă noile metode conduc la reducerea proporției articolelor defecte, adică se verifică i.n.

$$H_0 : p = p_0 = 0{,}1 \text{ față de i.a. } H_1 : p < 0.1,$$

unde am notat prin p proporția articolelor defecte.
Pentru $n = 100$, $r = 15$, $p_0 = 0{,}1$ se calculează statistica :

$$Z = \frac{15 - 100(0{,}10)}{\sqrt{100(0{,}1)(0{,}9)}} = 1{,}66.$$

Se alege p.s. al testului, $\alpha = 0{,}01$, de exemplu.

Deoarece valoarea statisticii Z este semnificativă:

$$Z = 1{,}66 < z_{0{,}99} = 2{,}33$$

se respinge i.n. H_0 cu p.s. $\alpha = 0{,}01$, adică noile metode tehnologice nu au dus la reducerea proporției de articole defecte.
Pentru calcularea *puterii* t.s. pentru verificarea i.n. H_0 față de i.a. H_1 se procedează ca la *testul* z^* și se obține:
a)

$$\pi(W, p_1) = \Phi\left(z_{1-\frac{\alpha}{2}}\sqrt{\frac{p_0(1-p_0)}{p_1(1-p_1)}} - \sqrt{n}\,\frac{p_0 - p_1}{\sqrt{p_1(1-p_1)}}\right) -$$

$$- \Phi\left(z_{1-\frac{\alpha}{2}}\sqrt{\frac{(p_0(1-p_0)}{p_1(1-p_1)}} + \sqrt{n}\,\frac{p_0 - p_1}{\sqrt{p_1(1-p_1)}}\right), p_1 \neq p_0$$

b)

$$\pi(W, p_1) = 1 - \Phi\left(\frac{n(p_0 - p_1)}{\sqrt{np_1(1-p_1)}} + z_\alpha\sqrt{\frac{p_0(1-p_0)}{p_1(1-p_1)}}\right), p_1 > p_0$$

c)

$$\pi(W, p_1) = \Phi\left(\frac{n(p_0 - p_1)}{\sqrt{np_1(1-p_1)}} + z_{1-\alpha}\sqrt{\frac{p_0(1-p_0)}{p_1(1-p_1)}}\right), \quad p_1 < p_0.$$

Observație. Relațiile ce dau puterile testului pot fi folosite pentru determinarea numărului de observații astfel încît testul să fie de putere $1-\beta$, specificată.
De exemplu, egalînd relația b), respectiv c) cu $1-\beta$ se obține:

$$n = \frac{1}{(p_1 - p_0)^2}\,(z_\alpha\sqrt{(p_0(1-p_0)} + z_\beta\sqrt{p_1(1-p_1)})^2,$$

respectiv:

$$n = \frac{1}{(p_0 - p_1)^2}\,(z_\alpha\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p_1(1-p_1)})^2.$$

iv) *Test bazat pe aproximarea legii binomiale Bi (n, p) cu legea Poisson Po(λ).*
Se aplică atunci cînd parametrul p este mic și n este mare.
Se procedează în continuare ca în cazul aplicării t.s. referitoare la v.a. $X \sim Po(\lambda)$ unde $\lambda = np$
v) *Test pentru parametrul p al legii binomiale (n, p) folosind transformări unghiulare.*
1. → (i) *Test exact)**
Se consideră v.a. r/n — proporția observată de „succese".

2. $E\left(\frac{r}{n}\right) = p$

$$\mathrm{Var}\left(\frac{r}{n}\right) = \frac{p(1-p)}{n}.$$

Se observă că Var (r/n) depinde de p și n, cei doi parametri ai legii binomiale $Bi\,(n,\ p)$.
Se aplică o transformare unghiulară v.a. r/n pentru a obține o v.a. cu dispersia independentă de parametrul p.
Fie v.a. transformată :

$$\nu_1\left(\frac{r}{n}\right) = 2\sqrt{n}\ \text{arc sin}\ \sqrt{\frac{r}{n}}$$

sau

$$\nu_2\left(\frac{r}{n}\right) = 2\ \text{arc sin}\ \sqrt{\frac{r}{n}}$$

având dispersiile :

$$\text{Var}\ \nu_1 = 1$$

$$\text{Var}\ \nu_2 = 1/n,\ \text{respectiv.}$$

3.—4. (→ *Test exact*, pct. i).
5. V.a.

$$\nu_2\left(\frac{r}{n}\right) = 2\ \text{arc sin}\ \sqrt{\frac{r}{n}}$$

are aproximativ legea $N(2\ \text{arc sin}\ \sqrt{p},\ 1/n)$ (unghiul este măsurat în radiani). Dacă se aplică corecția de continuitate, se folosește v.a.

$$\nu_3(r/n) = 2\ \text{arc sin}\ \sqrt{r/n + 1/2\,n}.$$

6. Se procedează în continuare ca la pct. i) ținînd seama că :

$$P\left(\frac{r}{n} \leqslant h\right) \approx \Phi\left(\frac{2\ \text{arc sin}\ \sqrt{h + 1/2n} - 2\ \text{arc sin}\ \sqrt{p}}{\sqrt{\frac{1}{n}}}\right).$$

unde $\Phi\,(u)$ este *funcția lui Laplace.*
Observație
1. Folosirea transformărilor unghiulare ușurează calculul puterii t.s. și al volumului n al eșantionului față de t.s. folosit la pct. iii), (→ *Test bazat pe aproximarea legii Bi (n, p) cu legea normală.*)*.
Exemplu
c) Se consideră *Exemplul b), pct.* ii) anterior (→ p. 483).
Se estimează proporția de produse defecte :

$$\frac{r}{n} = \frac{8}{100} = 0{,}08.$$

Se verifică i.n.

$$H_0 :\ p = p_0 = 0{,}091$$

față de i.a.

$$H_1 : p < p_0 = 0{,}091$$

folosind **transformările unghiulare**.
Se alege p.s. $\alpha = 0{,}05$.
Se calculează :

$$\frac{2 \arcsin \sqrt{h + \frac{1}{200}} - 2 \arcsin \sqrt{0{,}091}}{\frac{1}{\sqrt{100}}} = z_{0{,}05} = -1{,}65,$$

de unde :

$$2 \arcsin \sqrt{h + 0{,}005} = 0{,}9943.$$

Din tabelele funcției $y = 2 \arcsin \sqrt{x}$ se obține :

$$h + 0{,}005 = 0{,}227,$$

de unde

$$h = 0{,}222.$$

Deoarece

$$\frac{r}{n} = 0{,}08 < h = 0{,}222$$

se respinge i.n. H_0 cu p.s. $\alpha = 0{,}05$.
Puterea t.s. în punctul $p_1 = 0{,}089$ este :

$$\pi(\tau_1) = \Phi\left(\frac{\tau_{i1} - \tau_{i0}}{1/\sqrt{n}} + z_{1-\alpha}\right),$$

unde

$$\tau_{i0} = \arcsin \sqrt{p_0} = 0{,}3064$$

$$\tau_{i1} = \arcsin \sqrt{p_1} = 0{,}3029.$$

Se obține :

$$\pi(\tau_1) = 0{,}053.$$

Se poate determina volumul eșantionului pentru a obține o putere dorită, de exemplu egală cu 0,95. (→ *Test z*).
Se obține :

$$n \approx 8000.$$

Test pentru parametrii p_1, p_2 a două legi binomiale Bi(n_i, p_i), i = 1,2.
1. Se consideră 2 eșantioane de volume n_1, respectiv n_2, obținute din n_1 (n_2) probă în care se obțin $X_1 = X_1$, respectiv $X_2 = x_2$ „succese". În fiecare probă probabilitatea de a obține un „succes" este $p_1(p_2)$.

2. V.a.

$$Z_1 = \frac{\frac{X_1}{n_1} - p_1}{\sqrt{\frac{p_1(1-p_1)}{n_1}}} \sim N(0,1)\,;\quad Z_2 = \frac{\frac{X_2}{n_2} - p_2}{\sqrt{\frac{p_2(1-p_2)}{n_2}}} \sim N(0,1)$$

(aproximativ).
Z_1, Z_2 sînt v.a. independente.
V.a.

$$\frac{X_1}{n_1} \sim N\left(p_1, \frac{p_1(1-p_1)}{n_1}\right);\quad \frac{X_2}{n_2} \sim N\left(p_2, \frac{p_2(1-p_2)}{n_2}\right).$$

$$\text{V.a. } \frac{X_1}{n_1} - \frac{X_2}{n_2} \sim N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right).$$

V.a.

$$Z = \frac{\left(\frac{X_1}{n_1} - \frac{X_2}{n_2}\right) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim N(0,1).$$

3. a) $H_0: p_1 = p_2$ (t.b.) (→ tab. 71)

$H_1: p_1 \neq p_2$

b) $H_0: p_1 = p_2$ (t.u.dr.)

$H_1: p_1 > p_2$

c) $H_0: p_1 = p_2$ (t.u. stg.)

$H_1: p_1 < p_2$.

Mai general :

a) $H_0: p_1 - p_2 = \delta$ (t.b.)

$H_1: p_1 - p_2 \neq \delta$

b) $H_0: p_1 - p_2 = \delta$ (t.u.dr.)

$H_1: p_1 - p_2 > \delta$

c) $H_0: p_1 - p_2 = \delta$ (t.u.stg.)

$H_1: p_1 - p_2 < \delta$.

4. Se alege p.s. al testului, α.

5. Dacă i.n. H_0 este adevărată, atunci v.a. Z devine :

$$Z = \frac{\frac{X_1}{n_1} - \frac{X_2}{n_2}}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

unde $p = p_1 = p_2$.
Se calculează statistica :

$$Z = \frac{\frac{X_1}{n_1} - \frac{X_2}{n_2}}{\sqrt{\frac{x_1 + x_2}{n_1 + n_2}\left(1 - \frac{x_1 + x_2}{n_1 + n_2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

sau, mai general, statistica

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - \delta}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \quad \text{cu } \delta = p_1 - p_2,$$

unde $\hat{p}_1$, $\hat{p}_2$, $\hat{p}$ sînt *m.l.-estimatorii* * parametrilor p_1, p_2, p, respectiv :

$$\hat{p}_1 = \frac{x_1}{n_1}$$

$$\hat{p}_2 = \frac{x_2}{n_2}$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}.$$

6. Regiunile critice corespunzătoare diferitelor i.a., în cazul general, sînt trecute în tab. 101.

Tabelul 101

i.n. H_0 : $p_1 - p_2 = \delta$		
i. a. H_1	Regiunea critică	Test statistic
a) $p_1 - p_2 \neq \delta$	$\lvert Z \rvert > z_{\alpha/2}$	t.b.
b) $p_1 - p_2 > \delta$	$Z < z_{1-\alpha} = -z_\alpha$	t.u.dr.
c) $p_1 - p_2 < \delta$	$Z > z_\alpha$	t.u.stg.

Exemplu

a) S-au controlat $n_1 = 80$ produse (Difuzor C_4) şi s-au găsit $x_1 = 9$ produse defecte. La sfîrşitul chenzinei s-au controlat $n_2 = 157$ produse din care s-au găsit $x_2 = 16$ produse defecte.

Există motive să presupunem că pe parcursul celor 15 zile diferă calitatea la Difuzor C_4.

Cu alte cuvinte, se verifică i.n. asupra egalităţii proporţiilor de defecte ;

$$H_0 : p_1 = p_2$$

faţă de i.a.

$$H_1 : p_1 \neq p_2.$$

Se calculează *m.l.-estimatorii* * proporţiilor de defecte :

$$\hat{p} = \frac{x_1}{n_1} = \frac{9}{80} = 0{,}1125$$

$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{16}{157} = 0{,}1019$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{25}{237} = 0{,}1054.$$

Se calculează statistica testului :

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0{,}1125 - 0{,}1019}{\sqrt{0{,}1054 \cdot 0{,}9946\ (1/80 + 1/157)}} = 0.2465$$

Se alege p.s. al testului, $\alpha = 0{,}05$.

Deoarece valoarea statisticii Z nu este semnificativă :

$$|Z| = 0{,}2465 < z_{\alpha/2} = z_{0{,}025} = 1{,}96,$$

se acceptă i.n. H_0 cu p.s. $\alpha = 0{,}05$.

Observaţie. Prin aplicarea *corecţiei de continuitate statistica* Z devine :

$$Z = \frac{|(\hat{p}_1 - 1/2n_1) - (\hat{p}_2 - 1/2n_2)|}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}.$$

În cazul considerat : $Z = 0{,}1744$ şi decizia t.s. este aceeaşi.

c) Un nou tratament al unei boli este comparat cu tratamentul folosit obişnuit. Materialul clinic obţinut sub un control atent este trecut în tab. 102. Se cercetează dacă tratamentele sînt de aceeaşi eficacitate sau tratamentul nou este mai eficace decît cel vechi, adică i.n. $H_0 : p_1 = p_2$ faţă de $H_1 : p_1 < p_2$.

Tabelul 102

Rezultatul / Tratamentul	Vindecat („succes")	Nevindecat	Total
Vechi	$5(x_1)$	8	$13(n_1)$
Nou	$9(x_2)$	3	$12(n_2)$
Total	$14(x_1 + x_2)$	11	$25(n_1 + n_2)$

Se calculează *m.l.-estimatorii* * proporţiilor de vindecaţi :

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{5}{13},\ \hat{p}_2 = \frac{x_2}{n_2} = \frac{9}{12},\ \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{14}{25}.$$

Se calculează statistica testului :

$$Z = \frac{\frac{5}{13} - \frac{9}{12}}{\sqrt{\frac{14}{25}\left(1 - \frac{14}{25}\right)\left(\frac{1}{13} + \frac{1}{12}\right)}} = -1{,}845.$$

Se alege p.s. al testului, $\alpha = 0{,}01$.
Deoarece valoarea statistică Z nu este semnificativă :

$$Z = -1{,}845 > z_{1-\alpha} = z_{0{,}99} = -2{,}57$$

se acceptă i.n.H_0 cu p.s. $\alpha = 0{,}01$, adică tratamentele sînt echivalente din punct de vedere al eficacităţii vindecării loturilor considerate de bolnavi.

Test pentru egalitatea a k parametri p_i a k legi binomiale $Bi(n_i, p_i)$, $i = 1 \ldots k$, $k \geqslant 2$

1. Se consideră k eşantioane de volume n_i, $i = 1, \ldots, k$, $k \geqslant 2$ obţinute din n_i probe în care se obţin $X_i = x_i$ „succese". În fiecare probă probabilitatea „succesului" fiind p_i, $i = 1, 2, \ldots, k$. Repartiţia statistică a v.a. X este *legea multinomială**
2. V.a.

$$Z_i = \frac{X_i - n_i p_i}{\sqrt{n_i p_i (1 - p_i)}} = \frac{\frac{X_i}{n_i} - p_i}{\sqrt{\frac{p_i(1 - p_i)}{n_i}}} = \frac{\hat{p}_i - p_i}{\sqrt{\frac{p_i(1 - p_i)}{n_i}}} \sim N(0,1)$$

Z_i v.a. independente

V.a. $$\sum_{i=1}^{k} Z_i^2 \sim \chi_k^2.$$

3. $H_0 : p_1 = p_2 = \ldots = p_k = p.$

$H_1 : p_i \neq p_j, \forall i, j = 1, \ldots, k, \; k \geqslant 2, \; i \neq j.$

4. Se alege p.s. al testului, α.

5. Se calculează statistica testului :

$$X^2 = \sum_{i=1}^{k} Z_i^2 \simeq \sum_{i=1}^{k} \left(\frac{\hat{p}_i - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_i}}} \right)^2 = \sum_{i=1}^{k} \frac{n_i(\hat{p}_i - \hat{p})^2}{\hat{p}(1-\hat{p})} =$$

$$= \frac{\sum_{i=1}^{k} n_i \hat{p}_i^2 - n\hat{p}^2}{\hat{p}(1-\hat{p})},$$

unde $\hat{p}_i$, $\hat{p}$ sînt *m.l.-estimatorii* * parametrilor p_i ai legii binomiale $Bi(n_i, p_i)$, $i = 1, \ldots, k$, $k \geqslant 2$, respectiv p (valoarea comună) :

$$\hat{p}_i = \frac{x_i}{n_i}, \; i = 1, \ldots, k, k \geqslant 2$$

$$\hat{p} = \frac{\sum_i x_i}{\sum_i n_i}$$

V.a. $X^2 \sim \chi^2_{(k-1)}$, cînd i.n. H_0 este adevărată cu condiția ca $n_i \hat{p}_i > 5$.

Regiunea critică : se respinge i.n. ori de cîte ori diferența dintre proporțiile p_i este mare, adică

$$X^2 > \chi^2_{\alpha;(k-1)}.$$

Exemplu. Din 3 loturi de articole produse de 3 unități diferite s-au considerat eșantioane, înregistrîndu-se numărul de articole defecte în fiecare eșantion. S-au obținut rezultatele din tab. 103.

Tabelul 103

	Unitatea i			Total
	1	2	3	
Volumul eșantionului	150	120	130	400
Nr. de art. defecte	40	15	20	75

Se verifică i.n. $H_0 : p_1 = p_2 = p_3$ față de i.a. $H_1 : p_i \neq p_j, \forall\, i, j = 1, 2, 3,$ unde p_i este proporția de articole defecte din articolele produse de unitatea i. Se calculează *m.l. estimatorii* * proporțiilor p_i de articole defecte :

$$\hat{p}_1 = 0{,}2666\,;\quad \hat{p}_2 = 0{,}1250\,;\quad \hat{p}_3 = 0{,}1538\,;\quad \hat{p} = 0{,}1875.$$

Se calculează X^2-statistica testului : $\chi^2 = 10{,}3$.

Se alege p.s. al testului, $\alpha = 0{,}05$.

Deoarece valoarea statisticii X^2 este semnificativă

$$X^2 = 10.3 > \chi^2_{0{,}05;\,2} = 5{,}99$$

se respinge i.n. H_0 cu p.s. $\alpha = 0{,}05$, adică loturile nu sînt omogene din punctul de vedere al proporțiilor de produse defecte.

TESTE PRIVIND PARAMETRII LEGII MULTINOMIALE. T.s. referitoare la parametrii p_i, $i = 1, \ldots, k$ ai *legii multinomiale**, adică probabilitățile p_i de realizare a unui „succes" în fiecare probă.

Test privind valorile specificate ale probabilităților p_i.

Caz neparametric

1. Fie $X_i = n_i$ numărul de realizări („succese") ale evenimentului A_i, $1 \leqslant i \leqslant k$ în n probe (încercări) independente constînd în n repetări ale experienței și fie p_i- probabilitatea realizării evenimentului A_i în fiecare dintre încercări (probe). Se alcătuiește o *tabelă de frecvență*.

2. Vectorul aleator $\mathbf{X} = (X_1, \ldots, X_k')$ are *legea multinomială** cu componentele $X_i \sim Bi\ (n, p_i)$ (*legea binomială*)*

3. I.n. H_0 conform căreia probabilitățile p_i au valori specificate p_0^i :

$$H_0 : p_i = p_i^0,\ i = 1, \ldots, k$$

față de i.a.
H_1 : i.n. H_0 nu este adevărată.

4. Se alege α — p.s. al testului.

5. Se calculează statistica (→ *Test de concordanță χ^2, caz parametric*)

$$\chi^2 = \sum_i \frac{(O_i - e_i)^2}{e_i},$$

unde $0_i = n_i$ (sau F_i) sînt frecvențele observate ale „succeselor", iar f_i frecvențele teoretice (estimațiile) :

$$e_i(f_i) = np_i^0$$

V.a.

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0} \sim \chi^2_{\alpha;\,(k-1)}$$

cînd i.n. H_0 este adevărată.

6. Regiunea critică a t.s. : se respinge i.n. H_0 cu p.s. α dacă :

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0} > \chi^2_{\alpha;(k-1)}.$$

Exemplu. Într-un experiment*) legat de încrucișarea a două specii de mazăre relativ la culoarea și forma bobului de mazăre se poate face următoarea clasificare pe sorturi a $n = 556$ boabe de mazăre :

Tabelul 104

Sortul	1	2	3	4
Frecv. obs. n_i	315	101	108	32

Conform teoriei lui Mendel asupra eredității, probabilitatea obținerii unui bob de mazăre din primul, al doilea, al treilea sau al patrulea sort este 9/16, 3/16, 3/16, 1/16, respectiv.

Se verifică ipoteza conform căreia există concordanță între rezultatele experimentale și teoria lui Mendel, adică i.n. :

$$H_0 : p_1 = p_1^0 = \frac{9}{16};\ p_2 = p_2^0 = \frac{3}{16};\ p_3 = p_3^0 = \frac{3}{16};\ p_4 = p_4^0 = \frac{1}{16}$$

față de i.a. :

H_1 : i.n. H_0 nu este adevărată.

Se alege p.s. al testului, $\alpha = 0{,}05$.

Numărul estimat de boabe de sortul i ($i = 1, \ldots, 4$) este :

$$np_1^0 = 312{,}75\ ;\ np_2^0 = 104{,}25\ :\ np_3^0 = 104{,}25 \text{ și } np_4^0 = 34{,}75.$$

Se calculează statistica :

$$\chi^2 = \sum_{i=1}^{4} \frac{(n_i - np_i^0)^2}{np_i^0} = 0{,}47.$$

Deoarece valoarea statisticii χ^2 nu este semnificativă :

$$\chi^2 = 0{,}47 < \chi^2_{0{,}05;3} = 7{,}81$$

se acceptă i.n. H_0 și conchidem că rezultatele experimentale confirmă teoria lui Mendel.

Caz parametric

1.—2. (→ *Caz neparametric precedent.*)

*) G. Mendel, *Experiments on Plant-Hybridisation*, Harvard Univ. Press, Cambridge, 1946.

3. I.n. H_0 specifică probabilitățile p_i care sînt funcții de s ($s<k-1$) parametrii $\theta_1, \theta_2, \ldots, \theta_s$ necunoscuți, adică

$$H_0 : p_i = p_i^0(\theta_1, \theta_2, \ldots, \theta_s), \quad 1 \leqslant j \leqslant k$$

față de i.a.
H_1 : i.n. H_0 nu este adevărată.

4.—6. (→ *Test de concordanță χ^2, caz parametric.*)

Test privind egalitatea probabilităților multinomiale p în două sau mai multe experimente.

Caz parametric.

Nu se specifică valorile probabilităților p_j, ce urmează a fi estimate pe baza datelor eșantionului.

1. Se fac n încercări (probe) independente.

Se urmărește $X_{ij} = n_{ij}$ — numărul de încercări care conduc la realizarea evenimentului A_j (rezultatului de tip j), $j = 1, \ldots, k$ în cel de al i-lea experiment, $i = 1, \ldots, r$ din cele r ($r \geqslant 2$) experimente efectuate.

Fie p_{ij} — probabilitatea realizării rezultatului de tip j în experimentul i.

Se alcătuiește următoarea *tabelă de frecvențe* :

	Numărul de încercări care conduc la rezultatul de tip j			Numărul total de încercări
	1	2	.... k	
Experimentul 1	n_{11}	n_{12}	$\ldots . n_{1k}$	$n_{1\cdot}$
Experimentul 2	n_{21}	n_{22}	$\ldots . n_{2k}$	$n_{2\cdot}$
⋮				⋮
Experimentul r	n_{r1}	n_{r2}	$\ldots . n_{rk}$	$n_{r\cdot}$
Toate experimentele	$n_{\cdot 1}$	$n_{\cdot 2}$	$\ldots . n_{\cdot k}$	$n_{\cdot\cdot} = nr.$ total de probe independente

unde

$$n_{i\cdot} = n_{i1} + n_{i2} + \ldots + n_{ik}, \; i = 1, \ldots, r$$

$$n_{\cdot j} = n_{1j} + n_{2j} + \ldots + n_{rj}, \; j = 1, \ldots, k$$

$$n_{\cdot\cdot} = \sum_{i=1}^{r} n_{i\cdot} = \sum_{j=1}^{k} n_{\cdot j}.$$

2. Repartiția rezultatelor de tip j, $1 \leqslant j \leqslant k$ pentru fiecare experiment, adică repartiția vectorului aleator $\mathbf{X} = ((X_{11}, \ldots, X_{r1}), \ldots (X_{1k}, \ldots . X_{2k}))'$ este *legea multinomială* *.

3. I.n. H_0 : $p_{11} = p_{21} = \ldots = p_{r1} = p_1$, valoarea comună)

$p_{12} = p_{22} = \ldots = p_{r2} = p_2$, valoarea comună)

. . . . . .

$p_{1k} = p_{2k} = \ldots = p_{rk} = p_k$, valoarea comună)

față de i.a.

H_1 : i.n.H_0 nu este adevărată.

4. Se alege α — p.s. al testului.

5. Se calculează :

$$\chi^2 = \sum_{\substack{\text{în } rk \\ \text{celule}}} \frac{(O-e)^2}{e} = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{\left(n_{ij} - \dfrac{n_{i\cdot}\, n_{\cdot j}}{n_{\cdot\cdot}}\right)^2}{\dfrac{n_{i\cdot}\, n_{\cdot j}}{n_{\cdot\cdot}}},$$

unde

$$\hat{p}_j = \frac{n_{\cdot j}}{n_{\cdot\cdot}},\ j = 1, \ldots, k$$

este estimatorul valorii comune p_j, $j = 1, \ldots, k$, dacă i.n. H_0 este adevărată ;

$$n_{i\cdot}\, \hat{p}_j = \frac{n_{i\cdot}\, n_{\cdot j}}{n_{\cdot\cdot}},\ \forall i, j,$$

adică pentru fiecare din cele rk celule se estimează $n_{i\cdot}$ — numărul de încercări necesare pentru experimentul i prin $n_{i\cdot} \cdot \hat{p}_j$ — numărul acestor încercări ce conduc la rezultatul de tip j. Pentru aceasta se înmulțesc totalurile marginale pe linii, $n_{i\cdot}$, respectiv pe coloane, $n_{\cdot j}$, împărțindu-se la numărul total de probe, $n_{\cdot\cdot}$.

Regiunea critică este formată din mulțimea tuturor valorilor de selecție mai mari decît α-cuantila superioară a legii χ^2 cu $(r-1)(k-1)$ g.l. :

$$\chi^2 > \chi^2_{\alpha;\,(r-1)(k-1)}.$$

Observații

1. Modelul statistic este similar — dar nu identic — cu cel folosit în cazul unei *tabele de contingență**.

2. Aproximarea legii multinomiale cu legea χ^2 se poate face numai pentru eșantioane de volum suficient de mare.

Dacă numărul de observații dintr-o celulă este mic se recomandă cumularea lor cu observațiile din alte celule ale experimentului.

3. Pentru $k=2$ se obține cazul binomial. (→ *Test param. p_1, p_2 a 2 legi binomiale.*) În cazul a 2 experimente binomiale, se verifică i.n.

a) $H_0 : p_{11} - p_{21} = \delta$ (const. dată) (t.b.)

față de i.a. :

$$H_1 : p_{11} - p_{21} \neq \delta$$

b)
$$H_0 : p_{11} - p_{21} \leqslant \delta \text{ (t.u.dr.)}$$

$$H_1 : p_{11} - p_{21} > \delta$$

Statistica testului :

$$Z = \frac{\hat{p}_{11} - \hat{p}_{21} - \delta}{\sqrt{\hat{p}_{11}\hat{q}_{11}/n_{1\cdot} + \hat{p}_{21}\hat{q}_{21}/n_{2\cdot}}},$$

unde

$$\hat{p}_{11} = n_{11}/n_{1\cdot},\ \hat{q}_{11} = 1-\hat{p}_{11}$$

$$\hat{p}_{21} = n_{21}/n_{1\cdot},\ \hat{q}_{21} = 1-\hat{p}_{21}$$

Regiunea critică : se respinge i.n. cu p.s. α dacă :

a)
$$|Z| > z_{\alpha/2}$$

sau

$$Z > z_{\alpha/2}$$

$$Z < -z_{\alpha/2} = z_{1-\alpha/2},$$

unde $z_{\alpha/2}$ este $\alpha/2$-cuantila superioară a legii normale $N(0,1)$:

b)
$$Z > z_{\alpha}$$

unde z_{α} este α-cuantila superioară a legii normale $N(0,1)$.

Exemplu. Un studiu întreprins asupra duratei de spitalizare în urma unei anumite intervenții chirurgicale făcute în 3 spitale diferit localizate a condus la următoarele concluzii : au fost aleși la întîmplare 50 de pacienți din cele 3 locuri și clasificați după durata de spitalizare : mai scurtă de o săptămînă ; între 1—2 săptămîni ; între 2—3 săptămîni ; între 3—4 săptămîni și mai lungă de 4 săptămîni.

Locul	Durata de spitalizare (exprimată în săptămîni)					
	1	1—2	2—3	3—4	>4	Total
	Numărul de pacienți					
1	8	16	16	6	4	50
2	6	13	14	7	10	50
3	5	12	10	10	13	50
Total	19	41	40	23	27	150

Se studiază dacă durata de spitalizare este aceeaşi pentru pacienţii din cele 3 locuri, adică se verifică :

H_0 : pentru fiecare durată de spitalizare proporţia este aceeaşi în fiecare din cele 3 locuri.

H_1 : i.n. H_0 nu este adevărată.

Dacă i.n. H_0 este adevărată, se calculează :

$\frac{n_{\cdot 1}\, n_{1\cdot}}{n_{\cdot\cdot}} = \frac{19 \cdot 50}{150} = 6{,}3$ — estimaţia numărului de pacienţi din localitatea 1 care sînt spitalizaţi postoperatoriu mai puţin de o săptămînă.

$\frac{n_{\cdot 2} n_{1\cdot}}{n_{\cdot\cdot}} = \frac{41 \cdot 50}{150} = 13{,}7$ — estimaţia numărului de pacienţi din localitatea 1 care sînt spitalizaţi postoperatoriu între 1—2 săptămîni.

. . . . . .

$\frac{n_{\cdot 3} n_{3\cdot}}{n_{\cdot\cdot}} = \frac{27 \cdot 50}{150} = 9$ — estimaţia numărului de pacienţi din localitatea 3 care sînt spitalizaţi postoperatoriu mai mult de 4 săptămîni.

Se calculează statistica :

$$\chi^2 = \frac{(8-6{,}3)^2}{6{,}3} + \frac{(16-13{,}7)^2}{13{,}7} + \ldots + \frac{(13-9)^2}{9} = 8{,}57.$$

Deoarece valoarea statisticii χ^2 nu este semnificativă :

$$\chi^2 = 8{,}57 < \chi^2_{0{,}05;\,8} = 15{,}51$$

se acceptă i.n. cu p.s. $\alpha = 0{,}05$.

Test privind valorile specificate ale probabilităţilor p_{ij} şi probabilităţilor marginale $p_{i\cdot}$, $p_{\cdot j}$ în tabele de contingenţă.

Caz parametric

Se specifică p^0_{ij} — valorile probabilităţilor p_{ij}, precum şi $p^0_{i\cdot}$, respectiv $p^0_{\cdot j}$ — valorile probabilităţilor marginale $p_{i\cdot}$, respectiv $p_{\cdot j}$.

1. Se consideră un eşantion de volum $n_{\cdot\cdot}$ observaţii clasificate după 2 atribute A şi B, conform cărora fiecare individ din eşantion aparţine uneia din categoriile :

$$A_1, A_2, \ldots, A_l$$

în raport cu atributul A şi simultan, uneia din categoriile :

$$B_1, B_2, \ldots, B_k$$

în raport cu atributul B.

Fie $X_{ij} = n_{ij}$ numărul de observaţii din (A_i, B_j) — celula de ordin (i, j), $1 \leqslant i \leqslant l$; $1 \leqslant j \leqslant k$. Fiecare individ poate aparţine uneia din cele $l\ k$ celule.

Datele de observație se sumează într-o *tabelă de contingență**.

Tabelul 105

A \ B	$B_1 \ldots B_j \ldots B_k$	Total
A_1	$n_{11} \ldots n_{1j} \ldots n_{1k}$	$n_{1\cdot}$
⋮	⋮	⋮
A_i	$n_{i1} \ldots n_{ij} \ldots n_{ik}$	$n_{i\cdot}$
⋮	⋮	⋮
A_l	$nl_1 \ldots n_{lj} \ldots n_{lk}$	$n_{l\cdot}$
Total	$n_{\cdot 1} \ldots n_{\cdot j} \ldots n_{\cdot k}$	$n_{\cdot\cdot}$ = nr. total de observații

unde

$$n_{i\cdot} = \sum_{j=1}^{k} n_{ij}$$

$$n_{\cdot j} = \sum_{i=1}^{l} n_{ij}$$

$$n_{\cdot\cdot} = \sum_{i=1}^{l} n_{i\cdot} = \sum_{j=1}^{k} n_{\cdot j} = \sum_{i=1}^{l} \sum_{j=1}^{k} n_{ij}.$$

2. Repartiția numărului de indivizi peste cele l k celule, adică repartiția vectorului aleator $\mathbf{X} = (X_{ij})$, $1 \leqslant i \leqslant l$; $1 \leqslant j \leqslant k$ este *legea multinomială**.
Fie p_{ij} — probabilitatea ca un individ (o observație) să aibă categoriile A_i, respectiv B_j după atributul A, respectiv B în tabela de contingență corespunzătoare și $p_{i\cdot}$, $p_{\cdot j}$ probabilitățile marginale respective, unde:

$$p_{i\cdot} = \sum_{j=1}^{k} p_{ij}\,; \quad p_{\cdot j} = \sum_{i=1}^{l} p_{ij}\,;$$

$$\sum_{i=1}^{l} \sum_{j=1}^{k} p_{ij} = \sum_{i=1}^{l} p_{i\cdot} = \sum_{j=1}^{k} p_{\cdot j} = 1.$$

3. a) I.n. H_0, conform căreia probabilitățile p_{ij} au valorile specificate p^0_{ij}
$H_0 : p_{ij} = p^0_{ij}$, $1 \leqslant i \leqslant l, 1 \leqslant j \leqslant k$ față de i.a.
H_1 : i.n. H_0 nu este adevărată.
b) I.n. H_0, conform căreia probabilitățile marginale $p_{i\cdot}$ corespunzătoare diferitelor categorii ale atributului A au valorile specificate $p^0_{i\cdot}$; $H_0 : p_{i\cdot} = p^0_{i\cdot}$, $i = 1, \ldots, l$ față de i.a.

H_1 i.n.: H_0, nu este adevărată.

c) I.n. H_0, conform căreia probabilităţile marginale $p_{\cdot j}$ corespunzătoare diferitelor categorii ale atributului B au valorile specificate $p^0_{\cdot j}$,
$H_0 : p_{\cdot j} = p^0_{\cdot j}$ $1 \leqslant j \leqslant k$
faţă de i.a.
H_1 : i.n. H_0 nu este adevărată.

4. Se alege α — p.s. al testului.

5. Se calculează :

a) $$X_1^2 = \sum_i \sum_j \frac{(n_{ij} - n_{\cdot\cdot} p^0_{ij})^2}{n_{\cdot\cdot} p^0_{ij}}$$

b) $$X_2^2 = \sum_i \frac{(n_{i\cdot} - n_{\cdot\cdot} p^0_{i\cdot})^2}{n_{\cdot\cdot} p^0_{i\cdot}}$$

c) $$X_3^2 = \sum_{j=1}^{k} \frac{(n_{\cdot j} - n_{\cdot\cdot} p^0_{\cdot j})^2}{n_{\cdot\cdot} p^0_{\cdot j}}.$$

6. Regiunea critică :

a) se respinge i.n. H_0 dacă :

$$X_1^2 > \chi^2_{\alpha;kl-1}$$

b) se respinge i.n. H_0 dacă :

$$X_2^2 > \chi^2_{\alpha;l-1}$$

c) se respinge i.n. H_0 dacă ;

$$X_3^2 > \chi^2_{\alpha;k-1}.$$

Test de independenţă (de asociere) în tabele de contingenţă.

Caz parametric

1.—2. (→ *Test val. spec. ale prob.* p_{ij} *şi prob. marg.* $p_{i\cdot}$, $p_{\cdot j}$ *în tablele de contingenţă.* (pct. 1—2.)).
T.i. (a.) se bazează pe repartiţia condiţionată a numărului de observaţii din fiecare celulă (i, j), dîndu-se totalurile marginale.
În cazul unei tabele 2×2 această repartiţie condiţionată devine *legea hipergeometrică* *.

3. I.n. conform căreia atributele A şi B sînt independente, adică nu există nici o asociere între atributele A şi B pe fiecare observaţie a eşantionului :

$H_0 : p_{ij} = p_{i\cdot} p_{\cdot j}$ pentru $1 \leqslant i \leqslant l$; $1 \leqslant j \leqslant k$
faţă de i.a. H_1 : i.n. H_0 nu este adevărată.

4. Se alege α — p.s. al testului.

5. Fie probabilităţile marginale $p_{i\cdot}$, $p_{\cdot j}$ cu valorile specificate $p^0_{i\cdot}$ $(1 \leqslant i \leqslant l)$ şi $p^0_{\cdot j}$, $(1 \leqslant j \leqslant k)$.

Se calculează:

$$X^2 = X_1^2 - X_2^2 - X_3^2,$$

unde X_1^2, X_2^2 şi X_3^2 sînt date de cazurile a), b), c,) (→ p. 500).

Regiunea critică: se respinge i.n. conform căreia atributele A şi B sînt independente la p.s. α, dacă $X^2 > \chi^2_{\alpha;\,(l-1)(k-1)}$.

Caz neparametric

1.—4. Se procedează ca în *cazul parametric* precedent.

5. În cazul în care probabilităţile marginale nu sînt specificate, se vor estima cu ajutorul datelor din *tabelul de contingenţă** corespunzător, prin *metoda verosimilităţii maxime**, folosind funcţia de verosimilitate

$$P = \prod_{i=1}^{l} \prod_{j=1}^{k} (p_{i\cdot}\, p_{\cdot j})^{n_{ij}} = \prod_{i=1}^{l} p_{i\cdot}^{n_{i\cdot}} \prod_{j=1}^{k} p_{\cdot j}^{n_{\cdot j}}$$

dacă i.n. H_0 este adevărată.

*M.l.-estimaţiile** probabilităţilor marginale sînt:

$$\hat{p}_{i\cdot} = \frac{n_{i\cdot}}{n_{\cdot\cdot}},\ 1 \leqslant i \leqslant l$$

$$\hat{p}_{\cdot j} = \frac{n_{\cdot j}}{n_{\cdot\cdot}},\ 1 \leqslant j \leqslant k.$$

Se calculează:

$$X^2 = \sum_{i=1}^{l} \sum_{j=1}^{k} \frac{\left(n_{ij} - \dfrac{n_{i\cdot} n_{\cdot j}}{n_{\cdot\cdot}}\right)^2}{\dfrac{n_{i\cdot} n_{\cdot j}}{n_{\cdot\cdot}}}.$$

6. Regiunea critică: se respinge ipoteza de independenţă, la p.s. α dacă $X^2 > \chi^2_{\alpha;\,(l-1)(k-1)}$.

Exemplu (*caz neparametric*). Se face o anchetă privind preferinţele pentru diferite tipuri de muzică: muzică naţională, muzică străină, muzică preclasică într-o colectivitate formată din 750 ascultători împărţiţi pe 4 grupe de vîrstă.

Se obţin datele din tab. 106.

Se verifică dacă preferinţele pentru diferite tipuri de muzică sînt influenţate de vîrstă, adică i.n. privind independenţa celor două atribute: A — vîrsta

și B — tipul de muzică preferat:

$$H_0 : p_{ij} = p_{i\cdot}\, p_{\cdot j}, \qquad i = 1,\ 2,\ 3$$

$$j = 1,\ 2,\ 3,\ 4.$$

În ipoteza conform căreia cele două atribute (vîrsta și preferința) sînt independente frecvența ultimei celule este 12 ($= 100 \times 900/750$). Analog, se obțin și celelalte frecvențe din paranteze.

În tab. 106 se dau n_{ij} — numărul de ascultători din celula (i, j) clasificați simultan după cele 2 atribute — A și B.

Tabelul 106

Tipul de muzică preferat	Grupa de vîrstă				Total
	20—25	25—30	30—35	35 și mai mult	
1. Muzică națională	40 (42,666)	53 (64)	37 (32)	30 (21,333)	160
2. Muzică străină	150 (133,333)	225 (200)	75 (100)	50 (66,666)	500
3. Muzică preclasică	10 (24)	22 (36)	38 (18)	12 (20)	90
	200	300	150	100	750

Se alege p.s. al testului, $\alpha = 0{,}01$.

Cifrele din paranteze dau estimațiile probabilităților p_{ij} din fiecare celulă (i, j), obținute prin înmulțirea totalurilor marginale pe linii, $n_{i\cdot}$ cu totalurile marginale pe coloane, $n_{\cdot j}$ și împărțind la numărul total de probe, $n_{\cdot\cdot} = 750$. Se calculează:

$$X^2 = \sum_{\substack{\text{pt. 12}\\ \text{celule}}} \frac{(O-e)^2}{e} = \frac{(40-42{,}666)^2}{42{,}666} + \frac{(53-64)^2}{64} + \ldots +$$

$$+ \frac{(20-12)^2}{12} \approx 63{,}518$$

pentru 12 celule.

Deoarece valoarea statistică X^2 este semnificativă

$$X^2 = 63{,}518 > \chi^2_{0{,}01;6} = 16{,}81$$

și se respinge i.n. conform căreia vîrsta și tipul de muzică sînt independente și se conchide că există o asociere între cele 2 tipuri de atribute după care s-au clasificat observațiile.

Observații

1) *T.i.* (*a.*) este un test aproximativ și se aplică numai pentru valori suficient de mari ale lui n , și anume $n > 20$.

2) Frecvențele teoretice (estimațiile) pentru celula (i, j) nu trebuie să fie foarte mici ($\geqslant 5$).

3) Pentru o tabelă de contingență 2×2 se recomandă *testul exact al lui Fisher**, care poate fi folosit în locul *t.i.* (*a.*) anterior în cazul $n.. < 20$ sau $20 \leqslant n.. < 40$ și cea mai mică estimație pentru celula (i, j) mai mică decît 5.

4) Statistica *t.i.* (*a*) aplicat într-o tabelă de contingență 2×2 se modifică cu ajutorul *corecției de continuitate* * (a lui Yates), adică se calculează :

$$\sum_{\substack{\text{pt. } ik \\ \text{celule}}} \frac{\left(0 - e - \frac{1}{2}\right)^2}{e}.$$

5) Pentru legătura dintre *t.i.* (*a.*) și *testul egalit. probab.** a se vedea Observația 1 p. 496 și *testul de independență*, pct. 1—2 (p. 500).

Testul exact al lui Fisher pentru verificarea independenței în tabele de contingență.

Test introdus de R. A. Fisher pentru o tabelă de contingență 2×2 în locul *testului de independență* (*asociere*)* care este un t.s. aproximativ și pentru care statistica testului este o v. a. $\sim \chi_1^2$ dacă i.n. H_0 este adevărată.

1.—2. Repartiția numărului de observații din cele 4 celule este *legea multinomială* *.

Repartiția condiționată a numărului de observații din fiecare celulă, dîndu-se totalurile marginale pe linii și pe coloane este *legea hipergeometrică**

3. I.n. conform căreia atributele A și B sînt independente.

$H_0 : p_{ij} = p_{i\cdot}\, p_{\cdot j},\ 1 \leqslant i, j \leqslant 2$

față de i.a.

H_1 : i.n. H_0 nu este adevărată.

4. Se alege α — p.s. al testului.

5. Dacă i.n. H_0 este adevărată, probabilitatea de a obține cu cei $n..$ indivizi o tabelă avînd elementele n_{11}, n_{12}, n_{21}, n_{22} dîndu-se frecvențele marginale $n_{\cdot 1}$, $n_{\cdot 2}$, $n_{1\cdot}$, $n_{2\cdot}$ și numărul total $n..$ este egală cu :

$$p(n_{11}) = \frac{n_{\cdot 1}!\ n_{\cdot 2}!\ n_{1\cdot}!\ n_{2\cdot}!}{n..!\ n_{11}!\ n_{12}!\ n_{21}!\ n_{22}!}$$

În acest caz tabela de contingență 2×2 este dată de tab. 107 și este complet determinată dacă se știe numărul.

Se notează cu M și m valoarea maximă și minimă a lui n_{11}, respectiv.

Tabelul 107

A \ B	B_1	B_2	Total pe linie
A_1	$\frac{n_{1\bullet}n_{\bullet 1}}{n_{\bullet\bullet}}$	$\frac{n_{1\bullet}n_{\bullet 2}}{n_{\bullet\bullet}}$	$n_{1\bullet}$
A_2	$\frac{n_{2\bullet}n_{\bullet 1}}{n_{\bullet\bullet}}$	$\frac{n_{2\bullet}n_{\bullet 2}}{n_{\bullet\bullet}}$	$n_{2\bullet}$
Total (pe coloană)	$n_{\bullet 1}$	$n_{\bullet 2}$	n = nr. total de observații

i) Se presupun că

$$n_{11} \leqslant \frac{n_{1\bullet}\, n_{\bullet 1}}{n_{\bullet\bullet}}$$

și că $n_{1\bullet} < n_{2\bullet}$, $n_{\bullet 1} < n_{\bullet 2}$ (liniile și coloanele pot fi totdeauna rearanjate pentru a îndeplini ultimile două relații).
Se calculează :
a) coada din stînga a repartiției, adică probabilitățile

$$p(n_{11}),\ p(n_{11}-1),\ \ldots,\ p(m)\,;$$

b) coada din dreapta a repartiției, adică probabilitățile $p\,(M)$, $p(M-1)$, ..., $p\,(M^*)$, unde M^* este un întreg determinat astfel încît :

$$p(M^*) < p(n_{11}) < p(M^*-1)\,;$$

c) P — statistica testului este formată din suma probabilităților din cele 2 cozi ale repartiției.
ii) Pentru cazul în care

$$n_{11} > \frac{n_{1\bullet}n_{\bullet 1}}{n_{\bullet\bullet}}$$

se procedează asemănător.
Se calculează :
a) coada din dreapta a repartiției, adică probabilitățile :

$$p\,(n_{11}),\ p\,(n_{11}+1),\ \ldots,\ p(M).$$

b) coada din stînga a repartiției, adică probabilitățile : $p\,(m)$, $p\,(m+1)$, ..., $p\,(m^*)$, unde m^* este un întreg determinat astfel încît :

$$p\,(m^*) < p\,(n_{11}) < p\,(m^*+1).$$

c) P — statistica testului este formată din suma probabilităților din cele 2 *cozi ale repartiției* *.

6. Regiunea critică a testului: se respinge i.n. H_0 conform căreia atributele A și B sînt independente dacă:

$$P < \alpha.$$

Observație

1. Testul de independență χ^2 se poate aplica în condițiile Observației 3 (p. 730).

2. Pentru calculul probabilităților corespunzătoare celor 2 cozi ale repartiției se pot folosi tabele special alcătuite (J. H. Zar, *Biostatistical Analysis*, Pretince-Hall, 1974).

Exemplu. Un număr de $n = 15$ indivizi sînt clasificați după două atribute: A — culoarea părului, avînd două categorii A_1, A_2; B — culoarea pielii, avînd două categorii B_1, B_2.
Observațiile sînt înregistrate în tabela de contingență 2×2:

Tabelul 108

Categorii de culoarea părului	Categorii de culoarea pielii		Total
A_1	(n_{11}) 4	(n_{12}) 2	$n_{1\cdot} = 6$
A_2	(n_{21}) 11	(n_{22}) 8	$n_{2\cdot} = 9$
Total	$n_{\cdot 1} = 5$	$n_{\cdot 2} = 10$	$n_{\cdot\cdot} = 15$ numărul total de indivizi

Se pune întrebarea dacă datele de observație sugerează o asociere între culoarea părului și culoarea pielii, adică se verifică i.n.:

H_0: atributele A, B sînt independente, adică nu există asociere între culoarea părului și culoarea pielii.

H_1: există o asociere între culoarea părului și culoarea pielii.

Se alege p.s. al testului, $\alpha = 0{,}05$.

Fie n_{11} frecvența absolută a celulei (1,1) în tabela de contingență 2×2 avînd totalurile marginale date $n_{\cdot 1} = 5$, $n_{1\cdot} = 6$, fixate.
Pentru o astfel de tabelă n_{11} poate lua valoarea minimă:

$$m = 0$$

și valoarea maximă

$$M = 5$$

(păstrînd totalurile marginale la valorile date $n_{\cdot 1} = 5$, respectiv $n_{1\cdot} = 6$ și numărul total de observații $n_{\cdot\cdot} = 15$).

Deoarece valoarea înregistrată din celula (1,1) este

$$4 > \frac{n_1 . n_{.1}}{n_{..}} = \frac{5 \cdot 6}{15} = 2$$

se aplică metoda de calcul din cazul ii).
Se calculează probabilitățile corespunzătoare cozii din dreapta a [repartiției :

$$p(4.) = \frac{6!\,9!\,5!\,10!}{15!} \cdot \frac{1}{4!\,2!\,1!\,8!} = 0{,}044955$$

$$p(4+1) = p(M) = p(5) = \frac{6!\,9!\,5!\,10!}{15!} \cdot \frac{1}{5!\,1!\,0!\,9!} = 0.001998,$$

respectiv cozii de stînga a repartiției :

$$p(m) = p(0) = \frac{6!\,9!\,5!\,10!}{15!} \cdot \frac{1}{0!\,6!\,5!\,4!} = 0{,}041958$$

$$p(0+1) = p(1) = \frac{6!\,9!\,5!\,10!}{15!} \cdot \frac{1}{1!\,5!\,4!\,5!} = 0{,}251748.$$

Statistica testului este :

$$P = p(4) + p(5) + p(0) + p(1) = 0{,}044955 + 0{,}001998 + 0{,}041958 + 0{,}251748 = 0{,}0889$$

Deoarece valoarea statisticii P este nesemnificativă :

$$P = 0.0889 > \alpha = 0{,}05$$

se acceptă i.n. H_0, adică nu există asociere între atributele A-culoarea părului și B-culoarea pielii, deci A și B sînt caractere genetice independente.

CONTROL STATISTIC AL CALITĂȚII. Ansamblu de acțiuni (previziune, analiză, coordonare, realizare) bazat pe metodele statisticii matematice pentru stabilizarea, menținerea și îmbunătățirea calității producției și reducerea cheltuielilor de producție.

Control statistic preventiv (Control în curs de fabricație sau **Control pe flux tehnologic).** Ansamblu de metode ale statisticii matematice pentru controlarea stabilității unui proces tehnologic și urmărirea calității produsului în timpul fabricației cu scopul limitării numărului de produse necorespunzătoare din punctul de vedere al calității dorite.

După natura caracteristicii de calitate considerate, *c.s.p.* poate fi :

1) *C.s.p. pe bază de măsurare. C.s.p.* aplicat unor caracteristici de calitate cantitative, *v.a. continue*. C.s.p.* efectuat simultan asupra a doi parametri statistici pentru caracterizarea poziției cîmpului de împrăștiere (de exemplu, media de selecție, mediana etc.) și mărimii cîmpului de împrăștiere (de exemplu, abatere medie pătratică, amplitudine etc.) corespunzători unei caracteristici de calitate măsurabilă, v.a. continuă.

2) *C.s.p. pe atribut. C.s.p.* pe bază de verificare la „corespunzător", „necorespunzător" sau număr de defecte etc., adică atunci cînd caracteristica de calitate este *v. a. discretă**. De exemplu, v.a. X = numărul de exemplare defecte dintr-o probă (fişă de control *C*) sau numărul de defecte pe exemplar (fişa de control *u*) sau fişa de control pentru fracţiunea defectă (fişa *p*).

Etapele c.s.p.

1. *Analiza statistică premergătoare a procesului tehnologic.* Fază premergătoare introducerii **c.s.c.** Studiul procesului tehnologic din punct de vedere static şi dinamic. Etapele ce urmează a fi parcurse :
— determinarea eşantionului şi prelevarea probelor ;
— detectarea şi eliminarea cauzelor sistematice care produc abaterea de la calitatea dorită (de exemplu, uzura sculelor, oboseala muncitorului, materia primă necorespunzătoare etc.) astfel incît variaţiile caracteristicii de calitate să fie datorate numai unui ansamblu de cauze întîmplătoare, nesistematice (→ *Verificarea caracterului aleator*) ;
— prezumarea unei repartiţii (legi) teoretice a v.a. X-caracteristica de calitate ;
— estimarea parametrilor legii (de repartiţie) teoretice pe baza datelor eşantionului determinat ;
— verificarea concordanţei repartiţiei empirice cu legea teoretică prezumată (*stabilitatea statică* a procesului tehnologic) (→ *Teste de concordanţă.*) ;
— *stabilitatea dinamică* a procesului tehnologic ca *reglaj* : poziţia de centrare a cîmpului de împrăştiere a valorilor caracteristicii de calitate (caracterizată, de exemplu, prin media de selecţie sau mediana eşantionului corespunzător considerat) să fie menţinută constantă (→ *Teste de semnificaţie pentru medie*) şi ca *precizie* : mărimea cîmpului de împrăştiere (caracterizată, de exemplu, prin dispersia de selecţie, s^2, sau abaterea medie pătratică, s, sau amplitudinea eşantionului corespunzător) să fie menţinută constantă. (→ *Teste de semnificaţie pentru dispersie*).
— calculul *fracţiunii defecte (defective)*, adică raportul dintre numărul de defecte sau materiale necorespunzătoare prescripţiilor şi numărul total de unităţi considerate, şi dacă ea poate fi acceptată din punct de vedere economic.
— determinarea *limitelor de control*, adică limitele de variaţie ale caracteristicii de calitate, v.a. X, corespunzătoare desfăşurării stabile a procesului tehnologic şi determinării *intervalului de încredere* pentru *statisticile** corespunzătoare v.a. X. (→ *Interval de acceptare.*)

2. *Întocmirea fişei de control.* Întocmirea unui formular tip în care sînt sintetizate calculele prelucrării datelor de observaţie ale eşantionului, limitele de control şi se ia decizia : se continuă sau se opreşte procesul tehnologic după cum parametrii statistici (statisticile) ce caracterizează *reglajul* sau *precizia* se găsesc sau nu între limitele de control stabilite.

3. *Efectuarea controlului statistic preventiv.* Prelevarea, la intervale de timp fixate, a unor eşantioane de volum fixat, în ordinea obţinerii rezultatelor din procesul de fabricaţie şi prelucrarea observaţiilor conform fişei de control folosite.

Exemple.

1. *Fişa de control pentru medie şi abatere medie pătratică* ; fişă de control simultan al mediei (care determină poziţia (centrarea) cîmpului de împrăştiere a valorilor caracteristicii de calitate X) şi al abaterii medii pătratice (care determină mărimea cîmpului de împrăştiere al acesteia). Este specifică caracteristicilor măsurabile (cantitative), v.a. X continue (fig. 117).

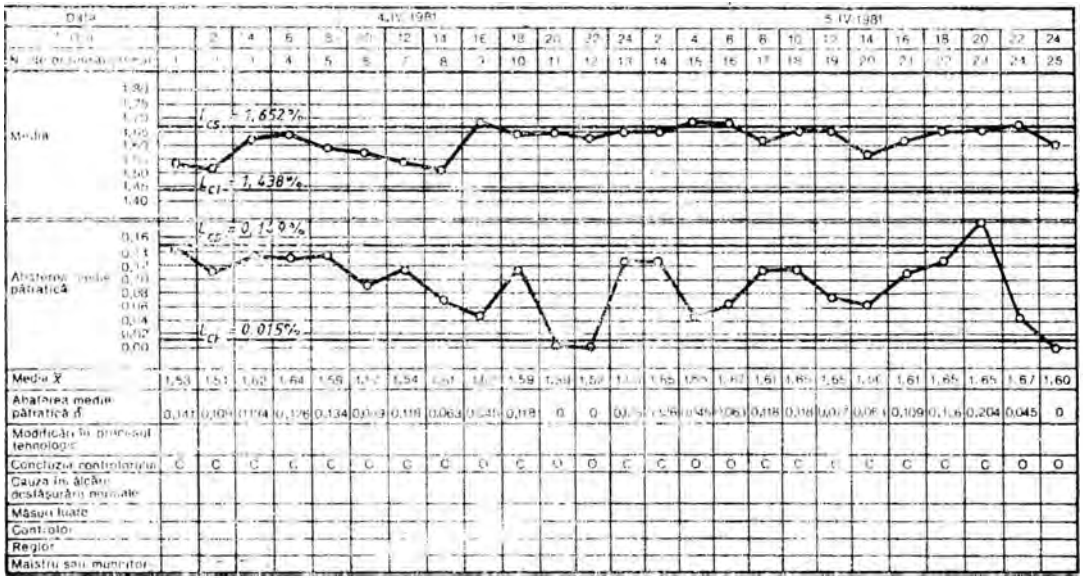


Fig. 117 Fișă de control pentru medie și abatere medie pătratică

2. *Fișa de control al numărului de exemplare defecte din probă (fișa c)* : fișă de control prin atribute (calitativ) bazată pe numărul mediu c de exemplare defecte găsite în k probe examinate (fig. 118).

Controlul statistic de recepție (Controlul loturilor de produse). Ansamblu de metode ale statisticii matematice aplicat loturilor de produse, după ce operația de fabricație a fost terminată, cu scopul de a decide acceptarea sau respingerea lotului controlat.

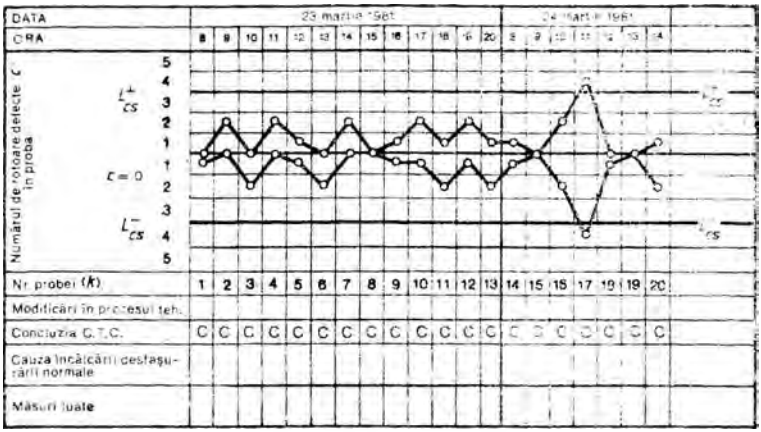


Fig. 118 Fișă de control c.

Probabilitatea de acceptare P_a a unui lot. Probabilitatea de a accepta un lot de produse controlat în funcție de calitatea reală a lotului. *P.a.* P_a este termen sinonim cu *curba caracteristicii operative* * a planului de control aplicat pentru verificarea calității lotului. (→ *Curba CO.*)

Plan de control. Ansamblu de reguli după care se efectuează **c.s.e.**. *P.c.* precizează felul caracteristicii de calitate X : *caracteristică atributivă (calitativă)*, adică v.a. discretă sau *caracteristică măsurabilă (cantitativă)*, adică v.a. continuă ; modul de prelevare a unităților eșantionului studiat : volumul eșantionului ; regula de decizie privind calitatea etc.

Nivelul de calitate admis (acceptat) (*NQA* sau *AQL*). *Frecțiunea defectă** a lotului considerată ca admisibilă de către beneficiar.

AQL se exprimă ca *frecțiune defectivă* sau ca număr de defecte pe 100 de unități de produs (în cazul caracteristicilor de calitate atributive) sau numai ca frecțiune defectivă (în cazul caracteristicilor de calitate măsurabile).

N.c.a. este nivelul de calitate care corespunde unei *probabilități de acceptare** P_a specificate.

Riscul furnizorului α. Termen sinonim cu *riscul erorii de genul I** sau p.s. al unui test statistic. (→ *Riscul erorii de genul I.*).

Riscul beneficiarului β. Termen sinonim cu *riscul erorii de genul II**.

Exemple

1. *Planuri de tip Dodge-Romig* : planuri de control de recepție pentru caracteristici de calitate atributive, calitatea loturilor de produse fiind exprimată

prin *fracțiunea defectivă*. Aceste planuri presupun verificarea 100% a loturilor respinse și înlocuirea sau remanierea produselor defective. Nu pot fi aplicate în cazul verificării distructive.

Planurile Dodge-Romig presupun fie *eșantionare simplă* (decizia de acceptare sau refuz este luată după controlul unui singur eșantion), fie *eșantionare dublă* (decizia se ia după controlul a două eșantioane). Sînt destinate protecției beneficiarului.

2. *Planuri bazate pe AQL* : planuri de control pentru verificarea loturilor prin atribute sau prin măsurare. Aceste planuri presupun acceptarea sau respingerea integrală a loturilor pe baza informației obținute din eșantionul sau eșantioanele verificate. Sînt concepute pentru protecția furnizorului, specificînd un risc mic al acestuia.

(FUNCȚIE DE) FIABILITATE (în engl. : *reliability*, în fr. : *fiabilité*). Pentru o v.a. reală X, **(f) f** se definește pe dreapta reală prin relația

$$R_X(x) = P(X > x) = 1 - F_X(x),$$

unde $F_X(.)$ este *funcția de repartiție** a v.a. X.

Termen sinonim cu *siguranța în funcționare* sau *funcția de supraviețuire*.

Dacă v.a. X reprezintă *durata de viață* sau *de funcționare* a unui element sau a unui sistem (ansamblu), atunci **f(f)** reprezintă probabilitatea de bună funcționare pînă la momentul x a elementului respectiv (probabilitatea cu timpul de funcționare a elementului să fie mai mare decît x, adică elementul sau sistemul să nu se defecteze (să nu „cadă") pînă la momentul x).

(F.) **f** este un indicator global al comportării elementului sau sistemului într-un interval de timp.

Rată de defectare (Rată de hazard). Probabilitatea ca elementul sau ansamblul să se *defecteze* (*„să cadă"*) în momentul imediat următor lui x, adică în intervalul $(x, x+\mathrm{d}x)$, condiționată de faptul că nu s-a defectat pînă la acest moment (că a funcționat bine pînă la momentul x). *R.d.* (*r.h.*) a v.a. X este :

$$r_X(x) = \frac{f_X(x)}{1 - F_X(x)} = \frac{f_X(x)}{R_X(x)} = -\frac{1}{R_X(x)} \frac{\mathrm{d}R_X(x)}{\mathrm{d}x},$$

unde $f_X(x)$, respectiv $F_X(x)$ este densitatea de probabilitate a v.a. X, respectiv funcția de repartiție ; $R_X(x)$ este **(f).f** între **(f).f** și *r.d.* (*r.h*) există relația

$$R_X(x) = e^{-\int_0^x r_X(t)\,\mathrm{d}t}.$$

Termeni sinonimi : *rata căderii* ; *intensitate de mortalitate* ;

Exemple. 1. Fie v.a. $X \sim f_X(x) = \lambda e^{-\lambda x}$. (*legea exponențială**).
(F)f. este

$$R_X(x, \lambda) = e^{-\lambda x},\ x > 0,\ \lambda > 0$$

și *r.d.* (*r.h*) este

$$r_X(x) = \lambda = \text{const.}$$

2. Fie v.a. $X \sim$ *legea Weibull**. (F)f este: $R_X(x; \beta, \gamma, \eta) = \exp\left[-\left(\frac{x-\gamma}{\eta}\right)^{\beta}\right]$, $\beta, \eta > 0$; $\gamma \geqslant 0$; $x \geqslant \gamma$.

XVII. ANALIZA DE VARIANȚĂ (ANOVA). PLANIFICAREA EXPERIMENTELOR

ANALIZĂ DE VARIANȚĂ (ANOVA). Ansamblu de metode ale statisticii matematice pentru analiza datelor de observație (realizări ale unor v.a.) care depind de mai mulți factori cu acțiune concomitentă cu scopul de a stabili pe cei mai importanți dintre ei și de a estima efectele lor. Metodă de descompunere a varianței totale în cel puțin două componente, reflectînd sursele de variație induse în experiment. (→ *Corelația transformată a lui Fisher, R.A.**) Estimînd dispersia unei v.a. ținînd seamă de influența unui factor iar apoi înlăturînd influența acestuia, compararea celor două dispersii ne dă informații referitoare la efectele factorului asupra v.a. studiate: dacă prima dispersie este mai mare, atunci factorul exercită influență.

Generalizare a *testului t** apărută din compararea a mai mult de două medii ale unor populații normale. (Termen sinonim: *analiza dispersională.*)

TRATAMENT. Termen originar din experimentele din agricultură. **T.** s-a extins pentru a desemna orice condiții experimentale ce urmează a fi comparate.

FACTOR. Grup de *tratamente** similare. **F.** are un număr de nivele.

Nivelul (varianta) unui factor. Clasă de valori ale unui **f.** Termen originar din experimentele din agricultură. Modul de manifestare a unui **f.** *N.f.* este uneori termen similar cu *tratament**.

Efectul factorului (efectul tratamentului). Termen sinonim cu efectul principal al **f.** Parte din variabilitatea totală indusă de un **f.** considerînd că acționează independent de alți **f.** Rezultatul acțiunii **f.** asupra v.a. studiate.

Interacțiune. Parte din variabilitatea totală datorată modificărilor acțiunii unui **f.** legată de variabilitatea unui alt **f.** sau unei combinații de alți **f.** *I.* este efectul combinat a 2 sau mai mulți **f.** De exemplu, într-un plan experimental factorial cu 3 **f.**, *A*, *B*, *C*, cu *n* repetiții — adică *n* observații pentru fiecare combinație $A_i B_j C_k$ — intervin: *efectele principale* ale **f.**: *A*, *B*, *C*, respectiv *interacțiuni de primul ordin* între cîte 2 **f.**: *AB*, *AC*, *BC*, precum și *interacțiunea de al doilea ordin ABC*.

Adevărata medie η. Media teoretică a v.a. *y* observate, corespunzătoare celulei *i* sau (*i*, *j*). *A.m.* admite, de obicei, o descompunere liniară în: medie generală, efecte principale ale factorului, interacțiuni între factori. De exemplu, într-un *model I bifactorial ANOVA**, *a.m.* a celulei (*i*,*j*) este:

$$\eta_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij},$$

unde μ este media generală, α_i, β_j sînt efectele principale ale **f.** *A*, *B*; γ_{ij} — interacțiunile între nivelele **f.** *A* și nivelele **f.** *B*.

Celulă. Sistem de *nivele** $(i_1, \ldots, i_p)$ ale unui **f.**, unde i_j, $1 \leqslant j \leqslant p$ este nivelul fixat al **f.** *j*.

Bloc. Sistem de *celule**. În aplicații, *b.* este termen sinonim cu *celulă**.
Factor (variabilă) calitativ(ă) (variabilă indicator). F. care își manifestă prezența sau absența, indicate prin 1 sau 0.
Factor (variabilă) cantitativ(ă). F. care ia valori continue într-un anumit interval.
Factor controlat. F. a cărei influență poate fi scoasă în evidență prin analiză statistică.
Factor de eterogenitate (factor necontrolat). F. care poate să introducă în populația statistică studiată în afară de variații întîmplătoare și variații sistematice. Analiza și interpretarea statistică a rezultatelor necesită organizarea experimentului astfel încît *f.e.* să nu influențeze în mod sistematic vreunul din *factorii controlați **.

MODEL LINIAR GENERAL. (→ *Regresie*; *Ecuația de regresie*; *Dreapta de regresie*; *Metoda M.M.P.*). Relație de dependență liniară a observațiilor efectuate într-un experiment, realizări ale v.a. $\{y_i\}$ față de parametrii necunoscuți, β_j în prezența unor erori întîmplătoare $\{e_i\}$.

$$y_i = \sum_{j=1}^{p} x_{ji}\beta_j + e_i,\ i = 1, \ldots, n,$$

unde $\{x_{ji}\}$ sînt coeficienți constanți necunoscuți; $\{\beta_j\}$ — *efectele factorului** *j* (parametri necunoscuți); v.a. $\{e_i\}$ sînt erorile corespunzătoare observației y_i. $i = 1, \ldots, n$ sau, în notații matriciale:

$$\mathbf{y} = \mathbf{X}'\boldsymbol{\beta} + \mathbf{e}$$

verificînd anumite *ipoteze fundamentale** (v.a. necorelate, cu mediile nule și dispersiile constante necunoscute) ale modelului general. (→*Metoda M.M.P.*) În **m.l.g.** se estimează dispersia necunoscută σ^2, se estimează efectele $\{\beta_j\}$ și se verifică ipoteze statistice asupra lor (de exemplu, i.n. $H_0: \beta_1 = \ldots = \beta_I$, adică nu există diferențe semnificative între cele *I* nivele ale factorului considerat), presupunînd în plus normalitatea v.a.
Clasificarea **m.l.g.** *după rangul matricei* **X**:
a) **M.l.g.**, *cazul rangului complet*: r = rang $\mathbf{X} = p$ = numărul de parametri. Situație specifică modelelor de regresie.
b) **M.l.g.**, *cazul rangului incomplet*: r = rang $\mathbf{X} \leqslant p$ = numărul de parametri. Situație specifică modelelor ANOVA. Se mai numesc și modele singulare.
Clasificarea **m.l.g.** *după natura valorilor* $\{x_{ji}\}$:
a) Constantele $\{x_{ji}\}$ iau valorile 0 sau 1. Ele au sensul de *variabile-indicator*, arătînd absența sau prezența influenței diferiților factori asupra variabilei studiate. **M.l.g.** este un *model de analiza variantei (model ANOVA)* sau *model de analiză dispersională*.
b) Constantele $\{x_{ji}\}$ descriu o mulțime continuă de valori (de exemplu, timpul *t*, temperatura *T* etc.). **M.l.g.** este un *model de analiză de regresie*. (→*Dreaptă de regresie*).
c) În cazul în care o parte din $\{x_{ji}\}$ sînt variabile-indicator și altele iau valori într-o mulțim continuă de valori, se obține un *model de analiză de covarianță**.
Clasificarea **m.l.g** *după modul în care se efectuează experimentele*:
a) *Model (clasificare) încrucișat(ă)* (În engl.: *Cross classification.*) **M.l.g.** în care fiecare nivel al unui factor apare împreună cu toate nivelele celorlalți factori din experiment.

Observațiile pot fi dispuse într-un tabel cu dublă intrare, în celula (i, j), situată la intersecția liniei i (nivelul i al factorului A) și coloanei j (nivelul j al factorului B), se înscriu K observații y_{ijk} (caz echilibrat) :

Factor A \ Factor B	1	2 ..... J
1	$y_{111}, \ldots, y_{11K}$	$y_{121}, \ldots, y_{12K} \cdots y_{1J1}, \ldots, y_{1JK}$
2	$y_{211}, \ldots, y_{21K}$	$y_{221}, \ldots, y_{22K} \cdots y_{2J1}, \ldots, y_{2JK}$
⋮	⋮	⋮
I	$y_{I11}, \ldots, y_{I1K}$	$y_{I21}, \ldots, y_{I2K} \cdots y_{IJ1}, \ldots, y_{IJK}$

Observații

(Model bifactorial ANOVA, clasificare încrucișată, caz echilibrat).

De exemplu, se studiază recolta obținută folosind două tipuri de semințe (factor A cu I nivele), cărora li se administrează îngrășământ chimic în J concentrații diferite (factor B cu J nivele), având la dispoziție IJK parcele. (Fiecare din cele IJ combinații tip de sămânță-îngrășământ se administrează la întîmplare pe cîte K parcele).

Răspunsul-producția (recolta) este y_{ijk} și reprezintă cele K determinări corespunzătoare celulei (i, j), pentru care se admite reprezentarea :

$$\Omega \begin{cases} y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}, & i = 1, \ldots, I \\ & j = 1, \ldots, J \\ & k = 1, \ldots, K \\ \{e_{ijk}\} \text{ v.a. independente} \sim N(0, \sigma^2) \end{cases}$$

(model bifactorial ANOVA clasificare încrucișată, caz echilibrat.)*

b) Model (clasificare) cuibărit(ă) (ierarhic(ă). (În engl. : *nested or hierarchical classification*) **M.l.g.** în care fiecare nivel al unui factor apare împreună numai cu unele nivele ale celuilalt (celorlalți) factor. Se spune că nivelele unui factor sînt cuibărite în nivelele celuilalt factor.

Observațiile se efectuează conform următoarei diagrame ;

Factor A	1			2.....			I		
Factor B în interiorul nivelelor factorului A	1	2	... J_1	1	2	... J_2	1	2	... J_I
Observații	y_{111}	y_{121}	$\cdots y_{1J_1^1}$	y_{211}	y_{221}	$\cdots y_{2J_2^1}$	y_{I11}	y_{I21}	$\cdots y_{IJ_I^1}$
	y_{112}	y_{122}	$\cdots y_{1J_1^2}$	y_{212}	y_{222}	$\cdots y_{2J_2^2}$	y_{I12}	y_{I22}	$\cdots y_{IJ_I^2}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	y_{11K}	y_{12K}	$\cdots y_{1J_1^K}$	y_{21K}	y_{22K}	$\cdots y_{2J_2^K}$	y_{I1K}	y_{I2K}	$\cdots y_{IJ_I^K}$

(Model bifactorial ANOVA, clasificare cuibărită, caz echilibrat).
De exemplu, dacă y_{ijk} este cea de a k-a observație corespunzătoare nivelului i al factorului A și nivelului j al factorului B în interiorul nivelului i al factorului A, pentru care se admite următoarea reprezentare, atunci modelul este :

$$\Omega \begin{cases} y_{ijk} = \mu + \alpha_i + \beta_{i(j)} + e_{ijk}, \; i = 1, \ldots, I \\ \qquad\qquad\qquad\qquad\qquad\quad j = 1, \ldots, J_i \\ \qquad\qquad\qquad\qquad\qquad\quad k = 1, \ldots, K \\ \{e_{ijk}\} \text{ v.a. independente } \sim N(0, \sigma^2). \end{cases}$$

Efectele α_i, $\beta_{i(j)}$ pot fi constante sau v.a. presupuse de obicei necorelate (*model bifactorial I sau II ANOVA, clasificarea ierarhică cuibărită, caz echilibrat**).
Clasificarea **m.l.g.** *după numărul de observații din celule.*

a) *Modele cu o singură observație în celulă* : modele în care se face un singur experiment pentru toate combinațiile posibile ale factorilor.

b) *Modele echilibrate* (*balansate*) : modele în care se face același număr de observații pentru toate combinațiile de factori, adică modele cu un număr egal (constant) de observații în celulă.

c) *Modele neechilibrate* (*nebalansate*) ; modele cu un număr inegal de observații în celulă, adică modele în care numărul de observații diferă de la celulă la celulă, pentru diferite combinații ale nivelelor factorilor.

Clasificarea **m.l.g.** (*cazul în care coeficienții* $\{x_{ji}\}$ *iau valorile 0 sau 1, modelul ANOVA*) *după natura lui* $\{\beta_j\}$;

a) *Modelul ANOVA cu efecte constante sau modelul I de analiză a variației* (*modelul I al lui Eisenhart*) : modelul ANOVA în care $\{\beta_j\}_{1 \leqslant j \leqslant I}$ sînt parametri adică efectele tratamentelor sînt constante necunoscute, numărul de tratamente I fiind fixat.

b) *Modelul ANOVA cu efecte aleatoare sau modelul II de analiză a variației* : modelul ANOVA în care $\{\beta_j\}$ sînt variabile aleatoare, tratamentele fiind alese în mod aleator dintr-o populație infinită de tratamente similare.

c) *Modelul mixt ANOVA sau modelul mixt de analiză a variației* : modelul ANOVA în care o parte din parametri sînt constanți și o parte sînt variabile aleatoare.

Observații. 1. Dacă se notează cu $\mathcal{I}$ — mulțimea tuturor tratamentelor atunci modelul I respectiv modelul II ANOVA corespunde cazului $\mathcal{I} = I$, respectiv $I < \mathcal{I} = \infty$. Uneori dacă $I < \mathcal{I} < \infty$ (populație finită de tratamente) se mai folosește denumirea de *model III ANOVA*.
2. Folosind notații matriciale, modelul I ANOVA se scrie sub forma unui **m.l.g.** :

$$\Omega \begin{cases} \mathbf{y} = \mathbf{X}'\boldsymbol{\beta} + \mathbf{e} \\ E(\mathbf{y}) = \boldsymbol{\eta} = \mathbf{X}'\boldsymbol{\beta}, \; \text{rang}\mathbf{X} = r \leqslant p \\ E(\mathbf{e}) = \mathbf{0} \\ E(\mathbf{e}\mathbf{e}') = \sigma^2 \mathbf{I} = (\text{Cov } e_i, e_j))_{1 \leqslant i, \, j \leqslant n}, \end{cases}$$

unde :

$$\mathbf{y}^{n \times 1} = (y_1, \ldots, y_n)' \; ; \; \boldsymbol{\beta}^{p \times 1} = (\beta_1, \ldots, \beta_p)' \; ; \; \mathbf{e}^{n \times 1} = (e_1, \ldots, e_n)'.$$

$\mathbf{X}^{p\times n} = (x_{ji})_{1\leqslant i\leqslant n;\ 1\leqslant j\leqslant p}$, elementele x_{ji} sînt constantele 0 sau 1.

Exemplu. Se măsoară procentul de alcool dintr-un lichid omogen folosind trei metode de măsurare, făcîndu-se pentru fiecare metodă cîte $J_i (i = 1, 2, 3)$ determinări. Măsurătorile dau adevăratul procent de alcool corespunzător metodei (tratamentului) folosite, influențat de eroarea de măsurare. Se admite că există o relație de dependență stochastică de forma:

$$y_{ij} = \mu + \alpha_i + e_{ij},\ i = 1, 2, 3;\ j = 1, \ldots J_i,$$

unde y_{ij} este a j-a determinare obținută prin metoda de măsurare i; μ este adevăratul procent de alcool din lichid; α_i, $i = 1, 2, 3$ este eroarea sistematică datorată metodei i, considerată ca fiind efectul metodei de măsurare sau efectul tratamentului; $\{e_{ij}\}$ sînt v.a. reprezentînd erorile de măsurare. Se impune condiția: $\sum_{i=1}^{n} \alpha_i = 0$. Cele trei metode de măsurare sînt suspecte a nu fi la fel de exacte sau că cele trei tratamente nu sînt omogene.

Variația totală între observațiile $\{y_{ij}\}$ poate fi descompusă în două componente: variația datorată diferitelor metode de măsurare (variația între tratamente) și variația între cele $j (j = 1, \ldots, J_i)$ determinări făcute cu aceeași metodă de măsurare (variația în interiorul tratamentului).

Dacă sîntem interesați în analizarea numai a celor trei metode de măsurare, α_i — efectele tratamentului sînt considerate constant (*Modelul I ANOVA, clasificare simplă (unifactorială)*).

Dacă se aleg întîmplător I metode de măsurare sau I tratamente dintr-o mulțime infinită de tratamente similare și cu fiecare metodă se fac J_i determinări în scopul generalizării analizei, atunci observațiile (măsurătorile) $\{y_{ij}\}$ pot admite reprezentarea:

$$y_{ij} = \mu + a_i + e_{ij},\ i = 1, \ldots, I;\ j = 1, \ldots, J_i$$

În acest caz efectele tratamentului, reflectînd diferența între cele I metode de măsurare selectate, sînt variabile aleatoare (*modelul II ANOVA, clasificarea simplă (unifactorială)*).

Mai general, dacă fiecare din cele I metode de măsurare alese întîmplător din mulțimea metodelor de măsurare posibile a fi utilizate este folosită de J specialiști care fac K determinări, modelul ANOVA poate admite reprezentarea:

$$y_{ijk} = \mu + a_i + \beta_j + e_{ijk},\ i = 1, \ldots, I;\ j = 1, \ldots, J;\ k = 1, \ldots K,$$

în care μ este media corespunzătoare unei metode de măsurare de referință; a_i — v.a., abaterea de la media corespunzătoare metodei i: β_j — abaterea de la media corespunzătoare specialistului j (const.); $\{e_{ijk}\}$ — v.a., erori de măsurare (*modelul mixt ANOVA*).

FORMĂ CANONICĂ A IPOTEZELOR FUNDAMENTALE. Forma *ipotezelor fundamentale** (→ *Modelul linear general*)* transformate cu ajutorul unei transformări lineare ortogonale: $z = Py$, care pot fi scrise:

$$\Omega_c \begin{cases} \mathbf{z} = (z_1, \ldots z_n)' \\ E(z_i) = \begin{cases} \xi_i, & 1 \leqslant i \leqslant r \\ 0, & r+1 \leqslant i \leqslant n \end{cases} \\ (\mathrm{Cov}\ (z_i, z_j)_{1\leqslant i,\ j\leqslant n} = \mathbf{I}\ (\mathbf{I} - \text{matricea unitate de ordinul } n), \end{cases}$$

unde $\{\xi_1, \ldots, \xi_r\}$, σ^2 sînt parametri necunoscuți; v.a. $\{z_i\}$ se obțin din v.a. $\{y_i\}$ prin transformarea ortogonală de matrice **P**, care nu depinde de acești parametri.

IPOTEZĂ NULĂ ÎN MODELUL LINIAR. I.n. H_0 care se referă la vectorul parametrilor

$$H_0 : \boldsymbol{\beta} = \mathbf{0} \text{ sau } \mathbf{k} \text{ (vector cunoscut)}$$

sau la o *funcție estimabilă** în β:

$$H_0 : \mathbf{C}'\boldsymbol{\beta} = \mathbf{0} \text{ sau } \mathbf{k} \text{ (vector cunoscut)}$$

față de i.a.

$$H_1 : \boldsymbol{\beta} \neq 0 \text{ sau } \mathbf{C}'\boldsymbol{\beta} \neq 0 \ (\neq k).$$

Exemplu. În modelul de regresie liniară simplă (→ *Dreapta de regresie*):

$$y_i = \mu + \beta x_i + e_i, \ i = 1, \ldots, n$$

vectorul parametrilor este:

$$\boldsymbol{\beta} = \begin{pmatrix} \mu \\ \beta \end{pmatrix}.$$

Se pot verifica următoarele ipoteze statistice:

1) **I.n.m.l.** de forma:

$$H_0 : \mu = 0$$

față de i.a.

$$H_1 : \mu \neq 0$$

sau, în scriere matriceală:

$$(1, 0)\begin{pmatrix} \beta \\ \mu \end{pmatrix} = 0$$

$$\mathbf{C}'\boldsymbol{\beta} = 0, \quad \nu_H = 1.$$

2) $$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

sau, în scriere matricială:

$$(0, 1)\begin{pmatrix} \mu \\ \beta \end{pmatrix} = 0$$

$$\mathbf{C}'\boldsymbol{\beta} = 0, \ \nu_H = 1.$$

3) $H_0 : \mu = 0,\ \beta = 0$ (simultan)

H_1 : cel puțin un parametru μ sau β să fie diferit de zero sau, în scriere matriceală :

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \beta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\mathbf{C}'\boldsymbol{\beta} = \mathbf{0},\ \nu_H = 2.$$

SUMĂ TOTALĂ A PĂTRATELOR (SUMĂ TOTALĂ DE PĂTRATE). Suma pătratelor tuturor observațiilor. În notația matriceală folosită în cadrul *modelului linear general**, **s.t.p.** este :

$$SS_{tot} = \mathbf{y}'\mathbf{y} = \|\mathbf{y}\|^2$$

SUMĂ A PĂTRATELOR ERORILOR. (→*Metoda M.M.P.*) Valoarea minimă SS_e a sumei pătratelor erorilor într-un model linear general, obținută înlocuind valoarea parametrului $\boldsymbol{\beta}$ prin m.m.p.-estimatorul său $\hat{\boldsymbol{\beta}}$:

$$SS_e = \hat{\mathbf{e}}'\hat{\mathbf{e}} = \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'(\mathbf{Xy}).$$

Termen sinonim cu *suma pătratelor erorilor de estimare.*

Observații

1. **S.p.e.** se obține scăzind din *suma totală de pătrate** SS_{tot} produsul scalar al m.m.p-estimației vectorului parametrilor, $\hat{\boldsymbol{\beta}}$ și al vectorului $\mathbf{Xy}$, partea dreaptă a *ecuațiilor normale**.

2. **S.p.e.** depinde numai de modelul linear general corespunzător și este independentă de orice formă a ipotezei nule H_0 referitoare la valorile unor funcții estimabile, ce urmează a fi testată în cadrul acestui model.

3. Dacă v.a. $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, atunci v.a. $SS_e/\sigma^2 \sim \chi^2_{\nu_e}$ și este independentă de *suma pătratelor datorată ipotezei H**, pentru orice formă a i.n. G.1. $\nu_e = n -$ — numărul parametrilor independenți, adică n — volumul eșantionului, minus numărul parametrilor independenți din model. (→ *Legea formelor pătratice de v.a. normale.*)

4. $s^2 = SS_e/\nu_e = \overline{SS_e}$ este o *estimație nedeplasată** a dispersiei necunoscute σ^2 :

$$E(s^2) = \sigma^2$$

și se numește *media pătratelor erorilor.*

SUMĂ A PĂTRATELOR DATORATĂ (PENTRU VERIFICAREA) IPOTEZEI NULE H_0. Suma pătratelor erorilor, SS_H (sau SS_{par}) calculată în cazul în care *ipoteza nulă în modelul linear** este adevărată.

Folosind valorile ajustate $\mathbf{Y}$ ale v.a. y din modelul linear general

$$\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}},$$

unde $\hat{\boldsymbol{\beta}}$ este estimatorul m.m.p. al lui $\boldsymbol{\beta} \in R^p$, se obține:

$$SS_H = \mathbf{Y}'\mathbf{Y} = \hat{\boldsymbol{\beta}}'\mathbf{X}\mathbf{X}'\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}'\mathbf{X}\,\mathbf{y}$$

sau

$$SS_H(SS_{\text{par}}) = \hat{\boldsymbol{\beta}}'\mathbf{C}\,(\mathbf{C}'\,(\mathbf{X}\mathbf{X}')^{-1}\mathbf{C})^{-1}\mathbf{C}'\hat{\boldsymbol{\beta}}$$

conform formei i.n. H_0.

Observații

1. **S.p.i.n.** H_0 este produsul scalar al m.m.p-estimatorului vectorului parametrilor. $\hat{\boldsymbol{\beta}}$ și al vectorului $\mathbf{Xy}$, partea dreaptă a *ecuațiilor normale.*

2. Dacă v.a. $\mathbf{y} \sim N(\mathbf{X}'\boldsymbol{\beta}, \sigma^2 I)$, atunci v.a. $SS_H/\sigma^2 \sim \chi^2_{\nu_H(=p)}$ cînd i.n. H_0 este adevărată și este independentă de *suma pătratelor erorilor**.

3. Se calculează:

$$E(SS_H) = p\,\sigma^2 + \boldsymbol{\beta}'XX'\boldsymbol{\beta},$$

adică media v.a. SS_H este $p\,\sigma^2$ plus o formă pătratică în parametrii β_i, $1 \leqslant i \leqslant p$. Așadar, prin compararea v.a. SS_H/ν_H și SS_e/ν_e sub forma raportului $SS_H/\nu_H : SS_e/\nu_e$ se obțin informații asupra parametrilor β_i diferiți de zero.

4. Se observă că:

$$SS_{\text{tot}} = SS_e + SS_H.$$

(→ *Suma totală a pătratelor*; *Suma pătratelor erorilor*; *legea formelor pătratice de v. a. normale.*)

TESTUL IPOTEZEI NULE. Testul F (→ *Test F al lui Snedecor — Fisher*) pentru verificarea *ipotezei nule în modelul linear**. În *ipotezele fundamentale* ale modelului linear și dacă i.n. H este adevărată, atunci v.a.

$$F = \frac{SS_H/\nu_H}{SS_e/\nu_e}$$

este v.a. $\sim F_{\nu_H, \nu_e}$, unde $SS_H \sim \sigma^2\chi^2_{\nu_H}$ este v.a. repezentînd *suma pătratelor datorată i.n.H** și $SS_e \sim \sigma^2\chi^2_{\nu_e}$ este v.a. reprezentînd *suma pătratelor erorilor (de estimare)**.

T.i.n. constă în: se respinge i.n. H cu p.s. α dacă și numai dacă

$$F > F_{\alpha;\,\nu_H,\,\nu_e},$$

unde $F_{\alpha;\,\nu_H,\,\nu_e}$ este *α-cuantila superioară** a v.a. F cu ν_H, ν_e g.l.

Observații

1. **T.i.n.** este un test LR (*test al raportului de verosimilități**) cu *funcția de verosimilitate**:

$$L\,(y_{11}, \ldots, y_{1J_1}; \beta_i, \sigma^2) = \prod_{i=1}^{I}\prod_{j=1}^{J_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(y_{ij} - \beta_i)^2\right].$$

(→ *Model unifactorial ANOVA.*)

2. Testarea semnificației regresiei simple (→ *Dreaptă de regresie*) : se verifică i.n. :

$$H_0 : \beta_1 = 0$$

față de i.a. :

$$H_1 : \beta_1 \neq 0.$$

Dacă $\beta_1 \neq 0$, se spune că regresia este semnificativă.
Statistica testului :

$$F = \frac{\overline{SS}_{\text{regr.}}}{\overline{SS}_e}$$

este o v.a. $\sim F_{1,\, n-2}$. Se poate aplica *testul t**, ținînd seama de legătura dintre v.a. F și t. (↔ *Legi de repartiție.*)

TABEL ANOVA. Tabel care sumează calculele necesare aplicării *testului ipotezei nule** și cuprinde descompunerea *sumei totale a pătratelor** observațiilor, indicînd astfel sursele de variație din *modelul liniar general**, g.l. corespunzătoare legilor v.a. respective și valorile lor medii. Pe baza elementelor din t. **ANOVA** se aplică *testul ipotezei nule** în modelul liniar (tab. 109).

Tabelul 109

Sursa de variație	Suma de pătrate (SS)	g.l.	$\overline{SS}$	$E(SS_{(\cdot)})$
Datorată modelului ajustat (de exemplu, datorată regresiei)	$SS_{H(\text{par})} = \hat{\boldsymbol{\beta}}'\mathbf{X}\mathbf{y}$	$\nu_H (= p)$	$SS_H/\nu_H = \overline{SS}_H$	$\sigma^2 + \frac{1}{p}\boldsymbol{\beta}'\mathbf{X}\mathbf{X}'\boldsymbol{\beta}$
Abaterea de la model (reziduală)	$SS_e = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}\mathbf{y}$	$\nu_e (= n-p)$	$SS_e/\nu_e = \overline{SS}_e$	σ^2
Totală	$\mathbf{y}'\mathbf{y}$	$\mathbf{n}$	—	—

MODEL I UNIFACTORIAL ANOVA. (În engl. *one-way classification, one-way layout, one-way, ANOVA.*) Model linear pentru compararea mediilor mai multor colectivități (populații) statistice, în ipoteza că v.a. ce caracterizează aceste colectivități urmează legea normală.

Model linear pentru studierea *efectelor unui factor* (*tratament*)* (= const. cunoscute) asupra rezultatelor experimentului.

Observațiile pot fi dispuse într-un tabel cu o intrare format din I linii și o coloană, în celula (i) situată la intersecția liniei i cu coloana considerată înscriindu-se observațiile obținute prin cercetarea *nivelului i** al factorului. Se admite

că observațiile din celula i sînt *eșantioane bernoulliene** din populația acestei celule și că comportă următoarea descompunere :

$$\text{Observația} = (\text{media generală}) + \left(\begin{array}{c}\text{abaterea de la}\\ \text{media generală dato-}\\ \text{rată tratamentului}\end{array}\right) + \left(\begin{array}{c}\text{eroarea}\\ \text{întîmplătoare}\end{array}\right)$$

(const.) (efectul nivelului i al factorului) (const.) v.a.

adică :

$$\Omega \begin{cases} y_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1, \ldots, I\,;\ j = 1, \ldots, J_i\,;\ n = \sum_{i=1}^{I} J_i \\ \{e_{ij}\} \text{ v.a. independente } \sim N(0, \sigma^2). \end{cases}$$

În m.u. I ANOVA se estimează parametrii μ, α_i, dispersia σ^2 și se verifică ipotezele nule referitoare la efectele factorului, α_i și la media generală, μ :

i.n. $H_{0\alpha} : \alpha_1 = \ldots = \alpha_I (= 0)$ față de i.a. $H_{1\alpha} : \alpha_i \neq \alpha_j$, $\forall\, i, j \in \{1, \ldots, I\}$,

respectiv

i.n. $H_{0\mu} : \mu = 0$ față de i.a. $H_{1\mu} : \mu \neq 0$.

M. I. u. ANOVA este caz particular al *modelului linear general**, unde :

$\mathbf{y} = (y_{11}, \ldots, y_{1J_1}\,;\, y_{21}, \ldots, y_{2J_2}\,;\, \ldots\,;\, y_{I1}, \ldots y_{IJ_I})'$ este vectorul observațiilor ;

$\mathbf{e} = (e_{11}, \ldots, e_{1J_1}\,;\, e_{21}, \ldots, e_{2J_2}\,;\, \ldots\,;\, e_{I1}, \ldots, e_{IJ_I})'$ este vectorul erorilor aleatoare ;

$\boldsymbol{\beta} = (\mu, \alpha_1, \ldots, \alpha_I)'$ este vectorul parametrilor :

$$\underset{(n,\, I+1)}{\mathbf{X}'} = \begin{pmatrix} 1 & 1 & 0 & \ldots & 0 \\ \vdots & & & & \\ 1 & 1 & 0 & \ldots & 0 \\ \vdots & 0 & 1 & \ldots & 0 \\ \vdots & & & & \\ \vdots & 0 & 1 & \ldots & 0 \\ \vdots & & & & \\ 1 & 0 & 0 & \ldots & 1 \\ \vdots & & & & \\ 1 & 0 & 0 & \ldots & 1 \end{pmatrix} \begin{matrix} \\ \} J_1 \text{ linii} \\ \\ \} J_2 \text{ linii} \\ \\ \\ \} J_I \text{ linii} \\ \\ \end{matrix}$$

pentru verificarea *ipotezei nule din modelul liniar**. (De exemplu, de forma i.n. :

$$H_{0\alpha} : \alpha_1 = \ldots = \alpha_I (= 0)$$

față de i.a. :

$$H_{1\alpha} : \text{cel puțin un } \alpha_i \neq 0,\ \forall i = 1, \ldots, I,$$

adică :

$$\underset{\mathbf{C}'}{\begin{pmatrix} 0 & 1 & 0 & \vdots & 0 \\ 0 & 0 & 1 & \vdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \vdots & 1 \end{pmatrix}} \underset{\boldsymbol{\beta}}{\begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_I \end{pmatrix}} = \underset{\mathbf{0}}{\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}$$

$(I, I+1) \qquad (I+1, 1) \quad (I, 1)$

$\nu_H = I < I+1 =$ numărul parametrilor (*cazul rangului incomplet*) (→*Model liniar general.*) Impunindu-se restricția suplimentară :

$$1/I \sum_{i=1}^{I} \alpha_i = 0$$

se obțin m.m.p -estimatorii unici ai parametrilor (*funcțiilor estimabile*)* μ, α_t, dispersiei σ^2 și se construiește *testul ipotezei nule** pentru **m.u. ANOVA**.) Calculele sînt sintetizate în tab. 110, respectiv tab. 111.

Tabelul 110

Sursa	SS suma de pătrate *)	g.l.	$\overline{SS}$	$E(\overline{SS})$
Efecte principale	$SS_\alpha = \sum_i J_i(y_{*i} - \bar{y})^2$	$\nu_\alpha = I-1$	$SS_\alpha/(I-1)$	$\sigma^2 + \sum_i J_i(\alpha_i - \bar{\alpha})^2/(I-1)$
Erorile	$SS_e = \sum_i \sum_j (y_{ij} - y_{i*})^2$	$\nu_e = \sum_i (J_i - 1)$	$\dfrac{SS_e}{\sum_i (J_i - 1)}$	σ^2
Totală	$SS_t = \sum_i \sum_j (y_{ij} - \bar{y})^2$	$\nu = n-1$	—	—

*) y_{i*} desemnează media observațiilor $\{y_{ij}\}$, adică

$$y_{i*} = \frac{1}{J_i} \sum_{j=1}^{J_i} y_{ij},$$

iar

$$\bar{y} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J_i} y_{ij}}{n} = \frac{1}{n} \sum_{i=1}^{I} J_i y_{i*}, \quad n = \sum_i J_i.$$

„*" desemnează operația de medie.

Tabelul 11 1

Sursa	SS (suma de pătrate)	g.l.	$\overline{SS}$	$E(\overline{SS})$
Media generală (μ)	$SS_\mu = \left(\sum_i y_{i*}\right)^2 \Big/ \sum_i J_i^{-1}$	$\nu_\mu = 1$	SS_μ	$\sigma^2 + \dfrac{I^2\mu^2}{\sum_i J_i^{-1}}$
Erorile	$SS_e = \sum_i \sum_j (y_{ij} - y_{i*})^2$	$\nu_e = \sum_i (J_i - 1)$	$\dfrac{SS_e}{\sum_i (J_i - 1)}$	σ^2
Totală	$SS_t = \sum_i \sum_j \left(y_{ij} - y_{i*} + \dfrac{\sum_i y_{i*}}{J_t \sum_i J_i^{-1}} \right)^2$	$\nu_t = n - I + 1$	—	—

Testul F pentru verificarea i.n. $H_{0\alpha}$, respectiv $H_{0\mu}$ constă în : se respinge i.n. $H_{0\alpha}$, respectiv $H_{0\mu}$ cu p.s. α dacă și numai dacă :

$$\overline{SS_\alpha}/\overline{SS_e} > F_{\alpha;\ \nu_\alpha,\ \nu_e},$$

respectiv

$$\overline{SS_\mu}/\overline{SS_e} > F_{\alpha;\ 1,\ \nu_e},$$

unde $F_{\alpha;\ \nu_\alpha,\ \nu_e}$ sînt α-cuantilele legii F cu ν_α, ν_e g.l.

Observații

1. *Suma de pătrate* SS_α *datorată i.n.* $H_{0\alpha}$ reprezintă măsura ponderată a împrăștierii mediei de selecție în cele I populații, iar SS_e *suma pătratelor erorilor**, reprezintă măsura împrăștierii observațiilor în fiecare populație. De aceea, SS_α se numește *suma de pătrate între grupe*, iar SS_e se numește *suma de pătrate în interiorul grupelor*. SS_t se numește *suma de pătrate în jurul mediei generale*.

2. **M.I u. ANOVA** prezentat mai sus corespunde *cazului nebalansat**. Dacă numărul de observații din fiecare celulă este același, adică $J_1 = \ldots J_I = J$, atunci se obține *modelul balansat* (sau *echilibrat*) în care expresiile m.m.p-estimatorilor și statisticilor F au formă mai simplă.

3. **M.I u. ANOVA** se poate prezenta sub forma echivalentă :

$$\Omega \begin{cases} y_{ij} = \beta_i + e_{ij},\ i = 1, \ldots, I;\ j = 1, \ldots, J_i \\ \{e_{ij}\} \text{ v.a. independente} \sim N(0, \sigma^2) \end{cases}$$

pentru care se verifică i.n. :

$$H_0 : \beta_1 = \ldots = \beta_I (= \beta, \text{ nespecificat}).$$

M.I u ANOVA este caz particular al *modelului linear general** (*cazul rangului complet*) cu matricea $\mathbf{X}'$ de forma :

$$\underset{(n,\,I)}{\mathbf{X}'} = \begin{pmatrix} 1 & 0 & \cdot & \cdot & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \cdot & \cdot & 0 \\ 0 & 1 & \cdot & \cdot & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & \cdot & \cdot & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdot & \cdot & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdot & \cdot & 1 \end{pmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}1\\\vdots\\1\end{matrix}}\right\} J_1 \text{ linii} \\ \left.\vphantom{\begin{matrix}1\\\vdots\\1\end{matrix}}\right\} J_2 \text{ linii} \\ \left.\vphantom{\begin{matrix}1\\\vdots\\1\end{matrix}}\right\} J_I \text{ linii} \end{matrix}$$

($J_1 + J_2 + \ldots + J_I = n$), rang $\mathbf{X}' = I$ ($= p$, numărul de parametri) pentru care *tabelul ANOVA** este :

Tabelul 112

Sursa dispersiei	SS suma de pătrate	g.l.	$\overline{SS}$	$E(SS)$
Între grupe	$SS_H = \sum_i J_i (y_{i*} - \bar{y})^2$	$\nu_H = I - 1$	$\overline{SS}_H = SS_H/\nu_H$	$\sigma^2 + \frac{1}{I-1}\sum_i J_i(\beta_i - \bar{\beta})^2$
În interiorul grupelor	$SS_e = \sum_i \sum_j (y_{ij} - y_{i*})^2$	$\nu_e = n - I$	$\overline{SS}_e = SS_e/\nu_e$	σ^2
Totală	$SS_t = \sum_i \sum_j (y_{ij} - \bar{y})^2$	$\nu_t = n - 1$	—	—

4. **M.I u. ANOVA** presupune egalitatea dispersiilor teoretice ale celor I celule ($= \sigma^2$, necunoscută), adică presupune adevărată ipoteza :

$$\sigma_1^2 = \ldots = \sigma_I^2$$

care trebuie verificată în prealabil (de exemplu, → *testul Bartlett**).

5. **M.I u. ANOVA** este termen sinonim cu *model I de analiză unifactorială* sau *clasificare simplă*.

*Observație. Testul statistic privind egalitatea a 2 medii (dispersiile necunoscute)** (→ XVI. *Verificarea ipot. stat.*) referitor la i.n. $H_0 : m_1 = m_2$ este caz particular al *testului ipotezei nule** pentru **m.I u. ANOVA** admițind că observațiile, v.a. y_{ij} comportă 2 nivele și au descompunerea :

$$\Omega = \begin{cases} y_{ij} = m_i + e_{ij}, \quad i = 1, \ 2; \quad j = 1, \ldots, n_i \\ \text{v.a. } \{e_{ij}\} \text{ independente } \sim N(0, \sigma^2). \end{cases}$$

Modelul se încadrează în *modelul linear general** avînd

$$\mathbf{y} = (y_{11}, \ldots, y_{1n_1}, \ y_{21}, \ldots, y_{2n_2})'$$

$$\boldsymbol{\beta} = (m_1, m_2)'$$

$$\underset{(n,\ 2)}{\mathbf{X}'} = \begin{pmatrix} \left.\begin{matrix} 1 & 0 \\ \cdot & \cdot \\ 1 & 0 \end{matrix}\right\} n_1 \text{ linii} \\ \left.\begin{matrix} 0 & 1 \\ \cdot & \cdot \\ 0 & 1 \end{matrix}\right\} n_2 \text{ linii} \end{pmatrix} \quad n = n_1 + n_2$$

Testul statistic pentru verificarea i.n. $H_0 : m_1 = m_2$ coincide cu *testul ipotezei nule** pentru **m.I u. ANOVA** ținînd seamă de proprietățile v.a. t și F. (→ *Legea t Student*; *Legea F Snedecor-Fisher*.)

MODEL I BIFACTORIAL ANOVA (CLASIFICARE ÎNCRUCIȘATĂ)

(În engl. *two-way classification*; *two-way layout*; *two-way ANOVA*.) Model linear pentru analiza datelor clasificate în două direcții sau după două criterii. Model linear pentru studierea efectelor a doi factori (A și B) și interacțiunilor lor asupra rezultatelor experimentului, expuse într-un tabel cu două intrări, ale cărui linii și coloane corespund nivelelor factorilor A, respectiv B. Celula (i, j) situată la intersecția liniei i cu coloana j conține observațiile efectuate sub influența simultană a nivelului i al factorului A și a nivelului j al factorului B și reprezintă un *eșantion bernoullian** din populația corespunzătoare acestei celule.

Model I bifactorial ANOVA cu o observație în celulă M.I b. ANOVA pentru care în celula (i, j) se face observația y_{ij}, care admite descompunerea :

$$\underset{\text{(observația)}}{} = \underset{\text{(const.)}}{\text{(media generală)}} + \underset{\text{(const.)}}{\begin{pmatrix} \text{efectul nivelului } i \\ \text{al factorului } A \end{pmatrix}} +$$

$$+ \underset{\text{(const.)}}{\begin{pmatrix} \text{efectul nivelului } j \\ \text{al factorului } B \end{pmatrix}} + \underset{\text{(v.a.)}}{\text{eroarea întîmplătoare,}}$$

adică *ipotezele fundamentale** sînt

$$\Omega \begin{cases} y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}. \quad i = 1, \ldots, I(=\mathcal{I}) \quad j = 1, \ldots, J(=\mathcal{J}) \\ \sum_i \alpha_i = \sum_j \beta_j = 0 \quad \text{(restricții suplimentare)} \\ \{e_{ij}\} \text{ v.a. independente} \sim N(0, \sigma^2) \end{cases}$$

pentru care se verifică i.n.

$$H_\alpha : \alpha_i = 0, \quad \forall i$$

$$H_\beta : \beta_j = 0 \quad \forall j.$$

M. I b. ANOVA o.c. este caz particular al *modelului liniar general** pentru care : $\mathbf{y} = (y_{11}, \ldots, y_{1J}; y_{21}, \ldots, y_{2J}; \ldots : y_{I1}, \ldots, y_{IJ1})'$ este vectorul observațiilor ; $\boldsymbol{\beta}) = (\mu ; \alpha_1, \ldots, \alpha_I ; \beta_1, \ldots, \beta_J)'$ este vectorul parametrilor necunoscuți ;

$$\underset{(n, I+J+1)}{\mathbf{X}'} = \begin{pmatrix} 1 & 1 & 0 \cdots 0 & 1 & 0 \cdots 0 \\ \vdots & \cdots\cdots & & \cdots\cdots & \\ 1 & 1 & 0 \cdots 0 & 0 & 0 \cdots 1 \\ 1 & 0 & 1 \cdots 0 & 1 & 0 \cdots 0 \\ 1 & 0 & 1 \cdots 0 & 0 & 0 \cdots 1 \\ \cdots\cdots & & & & \\ 1 & 0 & 0 \cdots 1 & 1 & 0 \cdots 0 \\ \vdots & \cdots\cdots & & \cdots\cdots & \\ 1 & 0 & 0 \cdots 1 & 0 & 0 \cdots 1 \end{pmatrix} \begin{matrix} \\ \\ \left.\begin{matrix} \\ \\ \end{matrix}\right\} J \text{ linii} \\ \\ \left.\begin{matrix} \\ \\ \end{matrix}\right\} J \text{ linii} \\ \\ \left.\begin{matrix} \\ \\ \end{matrix}\right\} J \text{ linii} \end{matrix}$$

$n = IJ$; $p = I + J + 1 =$ numărul parametrilor ; rang $\mathbf{X}' = I + J - 1$ *(cazul rangului incomplet)**.
Tabelul ANOVA este tab. 113.

Model I bifactorial ANOVA cu număr egal de observații în celulă (k>1) (clasificare încrucișată) (model echilibrat). *M. I b. ANOVA* pentru care în fiecare celulă (i, j) se fac K observații $\{y_{ijk}\}$ pentru care se admite descompunerea :

$$\text{(observația)} = \text{(media generală)} + \begin{pmatrix}\text{efectul nivelului } i \\ \text{al factorului } A\end{pmatrix} +$$

$$+ \begin{pmatrix}\text{efectul nivelului } j \\ \text{al factorului } B)\end{pmatrix} + \begin{pmatrix}\text{interacțiunea nivelului } i \\ \text{al factorului } A \text{ cu nive-} \\ \text{lul } j \text{ al factorului } B)\end{pmatrix} + \text{(eroarea întîmplătoare)},$$

adică *ipotezele fundamentale** sînt

$$\Omega \begin{cases} y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}, \quad i = 1, \ldots, I ; \; j = 1, \ldots J ; \; k = 1, \ldots, K \\ \sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0 \quad \text{(restricții suplimentare)} \\ \{e_{ijk}\} \text{ v.a. independente} \sim N(0, \sigma^2) \end{cases}$$

Tabelul 113

Sursa	SS	g.l.	$\overline{SS}$	$E(SS)$
Linii (efectul fact. A)	$SS_\alpha = J\sum_i (y_{i*} - y_{**})^2$	$\nu_\alpha = I - 1$	$\overline{SS}_\alpha = SS_{H\alpha}/\nu_\alpha$	$\sigma^2 + \frac{J}{I-1}\sum_i \alpha_i^2$
Coloane (efectul fact. B)	$SS_\beta = I\sum_j (y_{*j} - y_{**})^2$	$\nu_\beta = J - 1$	$\overline{SS}_\beta = SS_\beta/\nu_\beta$	$\sigma^2 + \frac{I}{J-1}\sum_j \beta_j^2$
Erorile (reziduală)	$SS_e = \sum_i \sum_j (y_{ij} - y_{i*} - y_{*j} + y_{**})^2$	$\nu_e = (I-1)(J-1)$	$\overline{SS}_e = SS_e/\nu_e$	σ^2
Totală	$SS_t = \sum_i \sum_j (y_{ij} - y_{**})^2$	$\nu_t = IJ - 1$	—	—

și pentru care se verifică i.n.

$$H_\alpha : \alpha_i = 0, \quad \forall i$$

$$H_\beta : \beta_j = 0, \quad \forall j$$

$$H_\gamma : \gamma_{ij} = 0, \quad \forall i,j$$

M. I *b.* ANOVA *e.* este caz particular al *modelului liniar general** pentru care :

$$\mathbf{y} = (y_{111}, \ldots, y_{11K}; \ldots y_{1J1} \ldots, y_{1JK}; \ldots y_{IJ1} \ldots, y_{IJK})'$$

este vectorul observațiilor ;
$\boldsymbol{\beta} = (\mu ; \alpha_1, \ldots, \alpha_I ; \beta_1, \ldots, \beta_J ; \gamma_{11}, \ldots \gamma_{IJ})'$ este vectorul parametrilor necunoscuți :
$\mathbf{X}' = (x_{ji})$ de ordin $n = IJK$; $p = I+J+IJ+1$ = numărul parametrilor. *Tabelul ANOVA* este tab. 114.

MODEL I BIFACTORIAL ANOVA (CLASIFICARE IERARHICĂ (CUIBĂRITĂ)). Model bifactorial ANOVA în care fiecare nivel al unui factor apare împreună numai cu unele nivele ale celuilalt factor. Efectele principale ale factorului *A* și efectele nivelului *j* al factorului *B* în interiorul factorului *A* sînt constante.

Model I bifactorial ANOVA cu număr egal de observații în subgrupe (K>1), clasificarea cuibărită (ierarhică).

M. I *b. ANOVA* (*c.i.*) cu același număr K ($K > 1$) de observații în fiecare subclasă.

M. I *b.* ANOVA, *clasificare cuibărită* pentru care observațiile admit descompunerea :

$$\text{observația} = \underset{\text{(const.)}}{\text{media generală}} + \underset{\text{(const.)}}{\text{(efectul nivelului } i \text{ al factorului } A)} +$$

$$+ \underset{\text{(const.)}}{\left(\begin{matrix}\text{efectul nivelului } j \text{ al factorului} \\ B \text{ în interiorul factorului } A\end{matrix}\right)} + \underset{\text{(v.a.)}}{\text{(eroare)}},$$

adică *ipotezele fundamentale** sînt :

$$\Omega \begin{cases} y_{ijk} = \mu + \alpha_i + \beta_{(i)j} + e_{ijk}, \quad i = 1, \ldots, I \\ \qquad\qquad\qquad\qquad\qquad\qquad j = 1, \ldots, J_i, n = K \sum_{i=1}^{I} J_i \\ \qquad\qquad\qquad\qquad\qquad\qquad k = 1, \ldots, K \\ \{e_{ijk}\} \text{ v.a. independente} \sim N(0, \sigma^2) \\ \sum_i \alpha_i = 0, \sum_j \beta_{i(j)} = 0, \quad \forall i \end{cases}$$

Se verifică i.n. de forma :

$$H_A : \alpha_i = 0, \qquad \forall i$$

$$H_{B(A)} : \beta_{i(j)} = 0, \quad \forall i, j$$

pe baza calculelor sintetizate în tabelul 115

Tabelul 114

Sursa	SS Suma de pătrate	g.l.	$\overline{SS}$	$E(\overline{SS})$
Linii efectul fact. A)	$SS_\alpha = JK \sum_i (y_{i**} - y_{***})^2$	$\nu_\alpha = I - 1$	$\overline{SS}_\alpha = SS_\alpha/\nu_\alpha$	$\sigma^2 + \frac{JK}{I-1} \sum_i \alpha_i^2$
Coloane (efectul fact. B)	$SS_\beta = IK \sum_j (y_{*j*} - y_{***})^2$	$\nu_\beta = J - 1$	$\overline{SS}_\beta = SS_\beta/\nu_\beta$	$\sigma^2 + \frac{IK}{J-1} \sum_j \beta_j^2$
Interacțiuni	$SS_{\alpha\beta} = K \sum_i \sum_j (y_{ij*} - y_{i**} - y_{*j*} + y_{***})^2$	$\nu_{\alpha\beta} = (I-1)(J-1)$	$\overline{SS}_{\alpha\beta} = SS_{\alpha\beta}/\nu_{\alpha\beta}$	$\sigma^2 + \frac{K}{(I-1)(J-1)} \sum_{i,j} \sigma_{ij}^2$
Erorile (reziduală)	$SS_e = \sum_i \sum_j \sum_k (y_{ijk} - y_{ij*})^2$	$\nu_e = IJ(K-1)$	$\overline{SS}_e = SS_e/\nu_e$	σ^2
Totală	$SS_t = \sum_i \sum_j \sum_k (y_{ijk} - y_{***})^2$	$\nu_t = IJK - 1$	—	—

Tabelul 115

Sursa	SS (suma de pătrate)	g.l.	$\overline{SS}$
Factorul A	$SS_A = K\sum_i J_i(y_{i**} - y_{***})^2$	$\nu_A = I - 1$	$\overline{SS}_A = SS_A/\nu_A$
Factorul $B(A)$	$SS_{B(A)} = K\sum_i\sum_j (y_{ij*} - y_{i**})^2$	$\nu_{B(A)} = \sum_i J_i - K$	$\overline{SS}_{B(A)} = \frac{SS_{B(A)}}{\nu_{B(A)}}$
Erorile	$SS_e = \sum_i\sum_j\sum_k (y_{ijk} - y_{ij*})^2$	$\nu_e = n - \sum_i J_i$	$\overline{SS}_e = SS_e/\nu_e$
Totală	$SS_t = \sum_i\sum_j\sum_k y_{ijk}^2 - ny_{***}^2$	$\nu_t = n - 1$	—

MODEL I ANOVA MULTIFACTORIAL FORMAL (CLASIFICARE ÎNCRUCIȘATĂ). Obținerea formală a unui model I ANOVA, cazul echilibrat (balansat), clasificare încrucișată cu p factori dintr-un model I ANOVA corespunzător unei clasificări inferioare (cu $p-1$ factori). Pe baza unei *reguli generale formale* se pot estima parametrii, se pot forma *sumele de pătrate* și se pot determina g.l. corespunzătoare.

Numărul de factori este p. Fie $\eta_{i_1\dots i_p}$ adevărata medie a celulei $(i_1, \dots, i_p)$ pentru care se admite descompunerea liniară ca suma dintre μ — media generală; C_p^1 efecte principale ale celor p factori; C_p^2 interacțiuni de ordinul 2 între cei p factori; ...; o interacțiune de ordinul p.

Efectele principale ale factorilor sînt: $\alpha_{i_t}^{A_t} = \eta_{**\dots i_t*\dots*} - \eta_{**\dots**}$, $1 \leqslant t \leqslant p$. Interacțiunea de ordinul $r (r \leqslant p)$ între nivele $i_1, \dots, i_r$ ale factorilor este:

$$\alpha_{i_1\dots i_r}^{A_1\dots A_r} = \eta_{i_1\dots i_r**\dots*} + \sum_{j=1}^{r} (-1)^j \eta_{**\dots*(r-j)*\dots*},$$

unde $\eta_{**\dots*(r-j)*\dots*}$ este o sumă în care fiecare termen se obține prin mediere din media adevărată a celulei $(i_1, \dots, i_r, \dots, i_p)$, $\eta_{i_1\dots i_r\dots i_p}$ lăsînd $(r-j)$ indici nemediați.

De exemplu, modelul I ANOVA cu 4 factori A, B, C, D, cazul echilibrat ($M > 1$ experiențe în fiecare celulă (i, j, k, l)) în care adevărata medie a celulei (i, j, k, l) admite descompunerea:

$$\eta_{ijkl} = \mu + \alpha_i^A + \alpha_j^B + \alpha_k^C + \alpha_l^D + \alpha_{ij}^{AB} + \alpha_{ik}^{AC} + \alpha_{il}^{AD} + \alpha_{jk}^{BC} + \alpha_{jl}^{BD} + \alpha_{kl}^{CD} + \alpha_{ijk}^{ABC} + \alpha_{ijl}^{ABD} + \alpha_{ikl}^{ACD} + \alpha_{jkl}^{BCD} + \alpha_{ijkl}^{ABCD}$$

Efectele principale ale factorilor și interacțiunile verifică restricțiile suplimentare :

$$\overset{A}{\alpha}_* = \overset{B}{\alpha}_* = \overset{C}{\alpha}_* = \overset{D}{\alpha}_* = 0$$

$$\overset{AB}{\alpha}_{i*} = \overset{AC}{\alpha}_{i*} = \ldots = \overset{BD}{\alpha}_{*l} = \overset{CD}{\alpha}_{*l} = 0$$

$$\overset{ABC}{\alpha}_{ij*} = \overset{ABD}{\alpha}_{ij*} = \ldots = \overset{BCD}{\alpha}_{*kl} = 0$$

$$\overset{ABCD}{\alpha}_{ijk*} = \ldots = \overset{ABCD}{\alpha}_{*jkl} = 0$$

În cazul în care adevărata medie a celulei (i, j, k, l) admite descompunerea t

$$\eta_{ijkl} = \mu + \overset{A}{\alpha}_i + \overset{B}{\alpha}_j + \overset{C}{\alpha}_k + \overset{D}{\alpha}_l$$

avem de a face cu un model I ANOVA, cazul *aditivității efectelor* (lipsa interacțiunilor între factori).

Modelul I ANOVA cu 4 factori, cazul echilibrat este :

$$\Omega : \begin{cases} y_{ijklm} = \eta_{ijkl} + e_{ijklm}\,;\ i = 1, \ldots, I\,;\ j = 1, \ldots, J\,;\ k = 1, \ldots, K\,; \\ \qquad\qquad l = 1, \ldots, L\,;\ m = 1, \ldots, M \\ \{e_{ijklm}\} \text{ v.a. independente} \sim N(0, \sigma^2)\,. \end{cases}$$

I.n. referitoare la efectele principale sau interacțiuni de forma :

$$H_A : \overset{A}{\alpha}_i = 0\,; \ldots\ H_{AB} : \overset{AB}{\alpha}_{ij} = 0 \ldots\,; H_{ABC} : \overset{ABC}{\alpha}_{ijk} = 0 \ldots \text{ etc.}$$

Se obțin m.m.p-estimatorii parametrilor și se construiesc testele F pentru i.n. *Regula generală formală*. Se introduce un factor „fictiv" care corespunde numărului de observații din fiecare celulă. Se admite că adevărata medie a celulei are o descompunere liniară și că sînt verificate restricțiile suplimentare referitoare la efectele principale și interacțiuni, numiți parametri „esențiali".

Estimatorii m.m.p. ai acestor parametri „esențiali" se obțin ca sume de observațiile mediate ale celulelor respective, nemediind indicii nivelelor factorului respectiv și mediind indicii celorlalte nivele. Semnele termenilor din aceste sume se succed ca semnele din expresia g.l. corespunzătoare.

Suma pătratelor erorilor SS_e se determină ca sumă a sumelor de pătrate SS care se referă la combinațiile factorilor „esențiali" cu factorul „fictiv".

Dacă la ipotezele fundamentale Ω se adaugă unele i.n. H, atunci în calculul sumei pătratelor erorilor SS_e' se adaugă sumele de pătrate corespunzătoare acestor ipoteze.

Suma de pătrate SS cu un număr de indici se obține ca diferență dintre suma pătratelor abaterilor observațiilor, v.a. y care au indicii corespunzători nemediați față de media generală și suma sumelor de pătrate SS care au indici mai puțini aflate deasupra ei în tabelul ANOVA.

Numărul g.l. al unei sume de pătrate se calculează formal făcînd produsul g.l. corespunzătoare tuturor sumelor SS care au un singur indice. Se spune că există o *dualitate* între expresiile g.l., sumelor de pătrate și m.m.p-estimatorii parametrilor, succesiunea semnelor fiind aceeași. Această dualitate justifică formalizarea modelelor ANOVA.

În modelul considerat în exemplul dat, modelul I ANOVA cu 4 factori, cu M observații în celulă se obține formal ca model I ANOVA cu 5 factori și o observație în celulă.

Tabelul ANOVA pentru exemplul dat este :

Analiză de varianță formală ($p = 4$) cu M observații în celulă (caz echilibrat)

Tabelul 116

Sursa	Parametrii estimați $\{\hat{\alpha}\}$	SS	Grade de libertate	$E(\overline{SS})$
μ	$\hat{\mu} = y_{*****}$	—	1	—
Efecte princ. ale factorului A	$\hat{\alpha}_i^A = y_{i****} - y_{*****}$	$SS_A = \sum_i \left(\hat{\alpha}_i^A\right)^2$	$\nu_A = I - 1$	$\sigma^2 + JKLM\sigma_A^2$
Efecte princ. ale factorului E	$\hat{\alpha}_m^E = y_{***m*} - y_{*****}$	$SS_E = \sum_m \left(\hat{\alpha}_m^E\right)^2$	$\nu_E = M - 1$	$\sigma^2 + IJKL\sigma_E^2$
Interacțiunea AB	$\hat{\alpha}_{ij}^{AB} = y_{ij***} - y_{i****} - y_{*j***} + y_{*****}$	$SS_{AB} = \sum_{i,j} \left(\hat{\alpha}_{ij}^{AB}\right)^2$	$\nu_{AB} = (I-1)(J-1)$	$\sigma^2 + KLM\sigma_{AB}^2$
Interacțiunea DE	$\hat{\alpha}_{lm}^{DE} = y_{***lm} - y_{***l*} - y_{***m*} + y_{*****}$	$SS_{DE} = \sum_{l,m} \left(\hat{\alpha}_{lm}^{DE}\right)^2$	$\nu_{DE} = (L-1)(M-1)$	$\sigma^2 + IJK\sigma_{DE}^2$
Interacțiunea ABC	$\hat{\alpha}_{ijk}^{ABC} = y_{ijk**} - y_{ij***} - y_{i*k**} - y_{*jk**} + y_{i****} + y_{*j***} + y_{**k**} - y_{*****}$	$SS_{ABC} = \sum_{i,j,k} \left(\hat{\alpha}_{ijk}^{ABC}\right)^2$	$\nu_{ABC} = (I-1)(J-1)(K-1)$	$\sigma^2 + LM\sigma_{ABC}^2$
Interacțiunea $ABCD$	$\hat{\alpha}_{ijkl}^{ABCD} = y_{ijkl*} - y_{ijk**} - \ldots - y_{*jkl*} + y_{ij***} + \ldots + y_{**kl*} - y_{i***} - \ldots - y_{***l*} + y_{*****}$	$SS_{ABCD} = \sum_{i,j,k,l} \left(\hat{\alpha}_{ijkl}^{ABCD}\right)^2$	$\nu_{ABCD} = (I-1)(J-1)\cdot(K-1)(L-1)$	$\sigma^2 + M\sigma_{ABCD}^2$
Erori	$\hat{\alpha}_{ijklm}^{ABCDE} = y_{ijklm} - y_{ijkl*} - \ldots - y_{*jklm} + y_{ijk**} + \ldots + y_{**klm} - y_{ij***} - \ldots - y_{***lm} + y_{i****} + \ldots + y_{****l} - y_{*****}$	$SS_{ABCDE} = \sum_{i,j,k,l,m} \left(\hat{\alpha}_{ijklm}^{ABCDE}\right)^2$	$\nu_{ABCDE} = (I-1)(J-1)\cdot(K-1)(L-1)(M-1)$	σ^2

MODEL I ANOVA MULTIFACTORIAL FORMAL (CLASIFICARE IERARHICĂ)

Obținerea formală a unui model I ANOVA, cazul echilibrat, clasificare ierarhică cu $p-1$ factori dintr-un model I ANOVA cu p factori, cazul echilibrat, clasificare cuibărită.

De exemplu, modelul I unifactorial al clasificării ierarhice se poate obține din modelul clasificării încrucișate duble (modelul bifactorial I ANOVA — cazul încrucișării duble) : efectele principale ale factorului B în interiorul nivelelor factorului A, $\beta_{(i)j}$ din cadrul modelului unifactorial I al clasificării ierarhice se obțin sumînd efectele factorului B și interacțiunile AB după indicele j din modelul bifactorial I al clasificării încrucișate : $\beta_{i(j)} = \beta_j + (\alpha\beta)_{ij}$.

Sumele de pătrate și gradele de libertate corespunzătoare din tabelul ANOVA pentru clasificarea ierarhică se obțin prin adunarea termenilor corespunzători din modelul bifactorial I ANOVA, clasificarea încrucișată.

În cazul în care nu există interacțiuni în modelul clasificării încrucișate, se combină termenul AB cu termenul rezidual.

Calculele pot fi urmărite în tab. 117.

Modelul bifactorial I al clasificării ierarhice se poate obține din modelul trifactorial I al clasificării încrucișate astfel : dacă toate nivelele factorului C sînt luate cu toate combinațiile posibile ale nivelelor factorilor A și B, atunci avem de-a face cu un model I ANOVA al clasificării triple încrucișate. Dacă $I\ J\ K$ unități experimentale sînt considerate în experiment cu K unități experimentale diferite pentru fiecare combinație a nivelelor $i = 1, \ldots, I$, respectiv $j = 1, \ldots, J$ ale factorilor A și B, se obține modelul clasificării ierarhice duble cu factorul C în interiorul factorului A. Se mai spune că factorul C este cuibărit în interiorul factorului A.

Presupunînd că observațiile admit următoarea descompunere în cazul clasificării triple încrucișate :

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij}^{AB} + (\beta\gamma)_{ik}^{AC} + (\beta\gamma)_{jk}^{BC} + (\alpha\beta\gamma)_{ijk}^{ABC} + e_{ijkl},$$

unde $i = 1, \ldots, I$; $j = 1, \ldots, J$; $k = 1, \ldots, K$; $l = 1, \ldots L$ = numărul de repetiții din fiecare celulă, efectul factorului C în interiorul combinației AB se obține adunînd în relația de mai sus m.m.p. — estimatorii corespunzători combinațiilor C, AC, BC, ABC, adică

$$\gamma_{(ij)k} = \gamma_k + (\alpha\gamma)_{ik}^{AC} + (\beta\gamma)_{jk}^{BC} + (\alpha\beta\gamma)_{ijk}^{ABC}$$

sau

$$\hat{\gamma}_{(ij)k} = (y_{**k*} - y_{****}) + (y_{ij*k*} - y_{ij**} - y_{**k*} + y_{****}) +$$

$$+ (y_{*j*k*} - y_{*j**} - y_{**k*} + y_{****}) + (y_{ij*k*} - y_{ij**} - y_{i*k*} -$$

$$- y_{*j*k*} + y_{i***} + y_{*j**} + y_{**k*} - y_{****}) = y_{ij*k*} - y_{ij**}.$$

Același model poate fi folosit pentru a obține un model bifactorial de două ori cuibărit, adică avînd efectele factorului C în interiorul lui B în interiorul lui A. Modelul de două ori cuibărit este

$$y_{ijkl} = \mu + \alpha_i + \beta_{(i)j} + \gamma_{(ij)k} + e_{(ijk)l},$$

unde

$$\gamma_{(ij)k} = \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$$

$$\beta_{(i)j} = \beta_j + (\alpha\beta)_{ij}.$$

Tabelul 117

Clasificarea încrucișată dublă (cazul echilibrat)					Clasificarea unifactorială ierarhică (cazul echilibrat)	
Lipsa interacțiunilor			Cu interacțiune			
	Suma de pătrate SS	g.l.	Suma de pătrate SS	g.l.	SS	g.l.
A	$JK\sum_i (y_{i**} - y_{***})^2$	$I-1$	$JK\sum_i (y_{i**} - y_{***})^2$	$I-1$	$JK\sum_i (y_{i**} - y_{***})^2$	$I-1$
B	$IK\sum_j (y_{*j*} - y_{***})^2$	$J-1$	$IK\sum_j (y_{*j*} - y_{***})^2$	$J-1$		
AB			$K\sum_i\sum_j (y_{ij*} - y_{i**} - y_{*j*} + y_{***})^2$	$(I-1)(J-1)$	$K\sum_i\sum_j (y_{ij*} - y_{i**})^2$	$I(J-1)$
rez.	$\sum_i\sum_j\sum_k (y_{ijk} - y_{i**} - y_{*j*} + y_{***})^2$	$IJK-I-J+1$	$\sum_i\sum_j\sum_k (y_{ijk} - y_{ij*})^2$	$IJ(K-1)$	$\sum_i\sum_j\sum_k (y_{ijk} - y_{ij*})^2$	$IJ(K-1)$
Totală	$\sum_i\sum_j\sum_k (y_{ijk} - y_{***})^2$	$IJK-1$	$\sum_i\sum_j\sum_k (y_{ijk} - y_{***})^2$	$IJK-1$	$\sum_i\sum_j\sum_k (y_{ijk} - y_{***})^2$	$IJK-1$

MODEL II UNIFACTORIAL ANOVA. (În engl.: *variable effects model, variance components model*). Model unifactorial ANOVA în care nivelele factorului studiat sînt considerate un eşantion de volum I dintr-o populaţie socotită infinită. Efectele principale ale factorului sînt v.a. Se presupune că observaţiile admit o descompunere liniară ca şi în cazul *modelului I unifactorial ANOVA**.

Ipotezele fundamentale ale **m. II u. ANOVA** (*cazul neechilibrat*) sînt :

$$\Omega\begin{cases} y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, I; j = 1, \ldots, J_i; n = \sum_i J_i \\ \qquad\qquad \text{cu } I < \mathcal{S} = \infty \ (\rightarrow \textit{Model liniar general}) \\ \{a_i\}, \{e_{ij}\} \text{ v.a. independente în totalitatea lor} \\ \{a_i\} \sim N(0, \sigma_a^2) \\ \{e_{ij}\} \sim N(0, \sigma_e^2) \qquad (\sigma_e^2 = \sigma^2) \end{cases}$$

De aici rezultă :

$$\text{Var } y_{ij} = \sigma_a^2 + \sigma_e^2$$

ceea ce justifică denumirea **m. II u. ANOVA** de *model pentru componentele dispersiei*.

În **m. II u.** ANOVA se verifică i.n.

$$H_A : \sigma_a^2 = 0$$

(contribuţia celor I eşantioane la variabilitatea totală) pe baza calculelor din *Tabelul ANOVA*.

*Tabelul ANOVA** pentru **m. II u. ANOVA** dat de tab. 118 este acelaşi cu Tabelul ANOVA din cazul *modelului I unifactorial ANOVA* *, singura deosebire fiind în calculul coloanei care dă mediile sumelor de pătrate $\overline{SS}$.

Testul i.n. H_A : se acceptă i.n. H_A cu p.s. α dacă şi numai dacă

$$\overline{SS}_A / \overline{SS}_e \leqslant F_\alpha ; \nu_A, \nu_e$$

unde $F_\alpha ; \nu_A, \nu_e$ este α-*cuantila superioară a legii* $*F$ cu ν_A, ν_e g.l..

Observaţii

1. În cazul echilibrat, $J_1 = \ldots = J_I = J$ se obţin formulele simplificate. De exemplu :

$$E(\overline{SS}_A) = \sigma_e^2 + J\sigma_a^2.$$

2. *Coeficientul de corelaţie intraclase* : coeficient de corelaţie intraclase care măsoară corelaţia între v.a. y_{ij} şi $y_{ij'}(j \neq j')$, observaţii în interiorul aceleiaşi grupe (clase). Este dat de relaţia :

$$\rho_{y_{ij},\, y_{ij'}} = \frac{\text{Cov}\,(y_{ij}, y_{ij'})}{\sqrt{\text{Var } y_{ij}} \cdot \sqrt{\text{Var } y_{ij'}}} = \frac{\sigma_a^2}{\sigma_e^2 + \sigma_a^2}.$$

3. Se pot obţine intervale de încredere pentru μ, σ_e^2, σ_a^2/σ_e^2 cu p.s. α (sau nivelul de încredere $(1-\alpha)$) : de exemplu, intervalul de încredere pentru μ cu nivelul

Tabelul 118

Sursa de variație	SS (Suma de pătrate)	g.l.	$\overline{SS}$	$E(\overline{SS})$ Model I unif. ANOVA	Model II unif. ANOVA
Efect. princ. fact. A (intre grupe)	SS_A ($SS_\alpha \to$ *model I unif. ANOVA*)	$I-1$	$\overline{SS}_A = SS_A/(I-1)$	$\sigma_e^2 + \dfrac{\sum_i J_i \alpha_i^2}{I-1}$	$\sigma_e^2 + \dfrac{1}{I-1} \times \times \left(n - \dfrac{\sum_i J_i^2}{n}\right)\sigma_a^2$
Erorile (reziduală) (in interiorul grupelor)	SS_e ($SS_e \to$ *model I unif. ANOVA*)	$\sum_i (J_i - 1) = n - I$	$\overline{SS}_e = SS_e/(n-1)$	$\sigma_e^2 (= \sigma^2)$	$\sigma_e^2 (= \sigma^2)$
Totală	SS_t ($SS_t \to$ *model I unif. ANOVA*)	$n - 1$	—	—	—

de încredere $(1-\alpha)$ 100% este :

$$(y_{**} - t_\alpha ; {}_{(I-1)}\sqrt{\overline{SS}_A/IJ} ; \quad y_{**} + t_\alpha ; {}_{(I-1)}\sqrt{\overline{SS}_A/IJ})$$

unde $t_\alpha ; {}_{(I-1)}$ este α-cuantila superioară a legii t cu $I-1$ g.l.

MODEL II BIFACTORIAL ANOVA (CLASIFICARE ÎNCRUCIȘATĂ). Model bifactorial ANOVA în care nivelele factorului A luate cu toate nivelele factorului B sînt eșantioane de volum $I (I \leqslant \mathcal{I} = \infty)$ respectiv $J (J < \mathcal{J} = \infty)$ din populații infinite.

Efectele principale și interacțiunile dintre factori sînt v.a.

Model II bifactorial ANOVA cu 1 observație în celulă (clasificare încrucișată *M. II b. ANOVA* pentru care în celula (i, j) se efectuează o singură observație y_{ij} care admite descompunerea liniară : o medie generală, μ (const.) ; efectele principale ale nivelului i al factorului A, a_i(v.a.) ; efectele principale ale nivelului j al factorului B, b_j(v.a.) ; erorile întîmplătoare, e_{ij}(v.a.) (*cazul aditiv-lipsa interacțiunilor între factori*).

Ipotezele fundamentale ale *m. II b. ANOVA* (1 *o.c.* ; *c.l.*)

$$\begin{cases} y_{ij} = \mu + a_i + b_j + e_{ij} \\ \text{v.a. } \{a_i\} \sim N(0 ; \sigma_a^2) \text{ independente ;} \\ \text{v.a. } \{b_j\} \sim N(0 ; \sigma_b^2) \text{ independente} \\ \text{v.a. } \{e_{ij}\} \sim N(0, \sigma_e^2) \text{ independente} \end{cases}$$

$$(\text{Var } y_{ij} = \sigma_a^2 + \sigma_b^2 + \sigma_e^2)$$

Se verifică i.n.:

$$H_A : \sigma_a^2 = 0$$

$$H_B : \sigma_b^2 = 0$$

cu ajutorul unor teste F. Calculele se pot urmări în tab. 119 care este *tabelul ANOVA** pentru *m. II b. ANOVA (1 o.c.; c.i.)*, identic cu tabelul ANOVA pentru *modelul I bifactorial ANOVA (1 observație în celulă; clasificare încrucișată)**, singura deosebire fiind în calculul coloanei $E(\widehat{SS})$.

MODEL II BIFACTORIAL ANOVA (CLASIFICARE IERARHICĂ). Model bifactorial ANOVA în care pentru fiecare nivel al factorului A se aleg la întîmplare subgrupe (subselecții) de unități experimentale. Nivelele factorului A sînt luate numai cu o parte din nivelele factorului B. Atît nivelele factorului A cît și nivelele factorului B sînt eșantioane de volum fixat din populații infinite.

Efectele principale și interacțiunile dintre factori sînt v.a..

Model II bifactorial ANOVA cu număr egal de observații în subgrupe. *M. II b. ANOVA (c. ier.)* în care fiecare nivel al factorului A se alege la întîmplare J subgrupe (subeșantioane) dintr-o populație infinită de nivele ale factorului B pe care se face același număr K de observații.

Pentru a k-observație din nivelul j al factorului B în interiorul nivelului i al factorului A se admite o descompunere liniară în: media generală μ, (constantă); efectul principal al nivelului i al factorului A, a_i(v.a.); efectul principal al nivelului j al factorului B în interiorul nivelului i al factorului A, $b_{(i)j}$ (v.a.); erorile aleatoare $e_{(ij)k}$.

Ipotezele fundamentale ale *m. II b. ANOVA (l.o.c.; c.i.)* sînt:

$$\Omega \begin{cases} y_{ijk} = \mu + a_i + b_{(i)j} + e_{(ij)k}, & i = 1, \ldots, I < \mathcal{I} = \infty; \\ & j = 1, \ldots, J < \mathcal{J} = \infty \\ & k = 1, \ldots K. \\ \text{v.a. } \{a_i\} \sim N(0, \sigma_a^2), \quad \text{independente, } \forall i \\ \text{v.a. } \{b_{(i)j}\} \sim N(0, \sigma_b^2), \text{ independente, } \forall i, j \\ \text{v.a. } \{e_{(ij)k}\} \sim N(0, \sigma_e^2), \text{ independente, } \forall i, j, k \end{cases}$$

$$(\text{Var } y_{ijk} = \sigma_a^2 + \sigma_b^2 + \sigma_e^2).$$

I. n. de forma $H_A : \sigma_a^2 = 0$; $H_B : \sigma_b^2 = 0$,
se verifică cu ajutorul unor teste F. Calculele se fac pe baza tab. 120, identic cu *tabelul ANOVA, modelul I bifactorial ANOVA, K obs. în celulă, clasificarea ierarhică**, singura deosebire fiind în calculul coloanei $E(\widehat{SS})$.

MODEL DE ANALIZĂ DE COVARIANȚĂ. Caz particular al *modelului liniar general**, folosit pentru studiul simultan al mai multor *regresori**.
Se urmărește mișcarea dispersiilor mediilor de selecție ale populațiilor considerate, înlăturarea efectelor unor *factori** (variabile).

Variabilă concomitentă (covariantă). Factor (variabilă) ale cărui efecte urmează să fie eliminate prin intermediul **m.a.c.**

Tabelul 119

Sursa	SS (suma de pătrate)	g.l.	$\overline{SS}$	$E(\overline{SS})$ model I bifact. 1 obs. cel. cl. incr.	model II, bifact. 1 obs. cel. cl. ierarh.
Linii (ef. fact. A)	$SS_A \rightarrow SS_\alpha$, *model I bifact. ANOVA, o obs. cel., clas. incr.*	$\nu_A = I-1$	$\overline{SS}_A = SS_A/\nu_A$	$\sigma_e^2 + J\left(\frac{\sum_i \alpha_i^2}{I-1}\right)$	$\sigma_e^2 + J\sigma_a^2$
Coloane (ef. fact. B)	$SS_B \rightarrow SS_\beta$, *model I bifact. ANOVA, o obs. cel., clas. incr.*	$\nu_B = J-1$	$\overline{SS}_B = SS_B/\nu_B$	$\sigma_e^2 + I\left(\frac{\sum_j \beta_j^2}{J-1}\right)$	$\sigma_e^2 + I\sigma_b^2$
Erorile (reziduală)	$SS_e = SS_t - SS_A - SS_B$	$\nu_e = (I-1)(J-1)$	$\overline{SS}_e = SS_e/\nu_e$	σ_e^2	σ_e^2
Totală	$SS_t \rightarrow SS_t$ *model I bifact. ANOVA, o. obs. cel., clas. incr.*	$IJ-1$	—	—	—

Tabelul 120

Sursa	SS (suma de pătrate)	g.l.	$\overline{SS}$	$E(\overline{SS})$ model I bifact. K obs. cel. cl. ier.	model II bifact. K obs. cel. cl. ier.
Ef. princ. fact. A (între grupe)	$SS_A \rightarrow SS_a^{*)}$, *model I bifact. ANOVA, K obs. cel.; clas. ierarh.*	$\nu_A = I - 1$	$\overline{SS}_A = SS_A/\nu_A$	$\sigma_e^2 + JK\dfrac{\sum_i a_i^2}{I-1}$	$\sigma_e^2 + K\sigma_b^2 + JK\sigma_a^2$
Ef. princ. niv. j al fact. B în interiorul niv. i al fact. A (subgrupe în int. gr. principale)	$SS_{B(A)} \rightarrow SS_{\beta(\alpha)}$, *model I bifact. ANOVA, K obs. cel; clas. ierarh.*	$\nu_{B(A)} = I(J-1)$	$\overline{SS}_{B(A)} = SS_{B(A)}/\nu_{B(A)}$	$\sigma_e^2 + K\dfrac{\sum_i \sum_j b_{i(j)}^2}{I(J-1)}$	$\sigma_e^2 + K\sigma_b^2$
Reziduală	$SS_e \rightarrow SS_e$, *model I bifact. ANOVA, K obs. cel; clas. ierarh.*	$\nu_e = IJ(K-1)$	$\overline{SS}_e = SS_e/\nu_e$	σ_e^2	σ_e^2
Totală	$SS_t \rightarrow SS_t$, *model I bifact. ANOVA. K obs. cel; clas. ierarh.*	—	—	—	—

*) Pentru obținerea acestor sume de pătrate $SS \rightarrow$ *Formalizarea modelelor I ANOVA, clasificarea încrucișată*, respectiv *clasificarea ierarhică*.

Model simplu de analiză de covarianță.
Combinarea *modelului I unifactorial ANOVA** cu *regresia liniară simplă.** (→*Dreaptă de regresie*).
Exemple.
1. În studierea dinamicii înălțimii unor copii clasificați în 3 grupe după 3 faze ale unei boli (factor „tratament"-3 nivele), dorim să eliminăm efectul (factorului) vîrstei, considerîndu-se mai puțin important în cadrul experimentului întreprins.
2. Se consideră un experiment în care două grupe de animale sînt tratate cu 2 diete alimentare diferite, d_1, d_2, urmărindu-se influența lor asupra creșterii în greutate a animalelor.
Pentru aceasta se pot studia greutățile medii finale ale animalelor tratate cu d_1, d_2, folosind un *model I unifactorial ANOVA** (factorul A — dieta alimentară, cu 2 nivele : d_1, d_2).
Dacă experimentatorul are la dispoziție și greutățile animalelor la începutul experimentului, studiul influenței dietei alimentare asupra greutății poate fi întreprins folosind un **m.a.c.**, eliminînd efectul (factorului) greutății inițiale În exemplul 1, se urmărește eliminarea deosebirilor în înălțime între cele 3 grupe, eliminîndu-se efectul vîrstei (înălțimea influențată de factorul vîrstă), Ipotezele fundamentale ale **m.a.c.** sînt cele de la modelul de *regresie simplă** (→ *Dreaptă de regresie* ; *Metoda M.M.P.*) : observațiile sînt variabile aleatoare independente ; legea normală a lui Y pentru valori particulare ale lui X ; dispersii egale**).
În exemplul considerat pentru o vîrstă fixată, înălțimile copiilor fiecăreia din cele 3 grupe sînt variabile aleatoare normale, avînd mediile dispuse după o dreaptă și dispersii egale, adică :

$$\Omega_{ac.}\begin{cases} y_{ij} = \alpha_{i0} + \beta_i x_{ij} + e_{ij},\ i = 1, \ldots I\ (= \text{număr de tratamente, în ex. considerat, } I = 3) \\ \{e_{ij}\} \text{ v.a. independente} \sim N(0, \sigma_e^2),\ j = 1, \ldots, J_j\ (= \text{nivelul de observații din al } i\text{-lea eșantion}) \end{cases}$$

y_{ij}, x_{ij} sînt valorile lui y și x pentru cea de a j-a observație, căreia i se aplică al i-lea tratament.
α_{i0} — înălțimea celei de a i-a grupă pe dreapta de regresie la $x = 0$.
În **m.a.c.** Ω_{ac} se verifică i.n. de forma :

$H_0 : E(y_i/x) = \alpha_{i0} + \beta_i \cdot x$

$H_0 : E(y_i/x) = \alpha_i + \beta x$ (drepte de regresie paralele)

$H_0 : E(y_i/x) = \alpha_0 + \beta x$ (aceeași dreaptă de regresie)

$H_0 : E(y_i/x) = \alpha_{i0}$ (*Model I unifactorial ANOVA** — dreapta de regresie nu depinde de x, paralelă cu Ox).

Diferitele i.n. H_0 se verifică cu ajutorul unui test F.
METODE DE COMPARAȚIE MULTIPLĂ (INFERENȚĂ STATISTICĂ SIMULTANĂ). Ansamblu de metode de inferență statistică pentru studiul unor *funcții parametrice, estimabile**, construcția intervalelor comune (simul-

) Pentru fiecare tratament, mediile populațiilor (grupelor) sînt dispuse după o (*curbă de*) *regresie.

tane) de încredere și teste statistice pentru familii de ipoteze referitoare la funcții parametrice în *modele liniare*, permițînd astfel comparații între parametri sau grupe de parametri.
Studiul situațiilor create în ansamblul ipotezelor în urma respingerii ipotezei nule generale, în scopul determinării *funcțiilor estimabile** care au contribuit la respingerea i.n.
Compararea globală (multiplă) a mediilor în *modele ANOVA**, în scopul depistării mediilor care determină respingerea ipotezei nule privind egalitatea lor. **M.c.m.** au fost fundamentate de J. W. Tukey (1953), H. Scheffé (1953), E. Spjøtvoll, M. Stoline (1973), K. R. Gabriel (1969); O.J. Dunn (1961) ș.a.
Contrast. *Funcție parametrică** $\psi = \sum_i c_i\theta_i$ pentru care coeficienții $\{c_i\}$, constante cunoscute, satisfac condiția $\sum_i c_i = 0$.

C. este o *funcție estimabilă**.
Spațiul q-dimensional al funcțiilor estimabile. Spațiul generat de *q funcții estimabile** liniar independente. *S.f.e.* se notează cu L. În modelul I unifactorială ANOVA*, de exemplu, pentru parametrii $\{\beta_i\}$, $1 \leqslant i \leqslant I$, un contrast este:

$$\beta_i - \beta_j,\ i \neq j \text{ (coeficienți: } 1,\ -1) \text{ sau } k\left(\frac{1}{I_1}\sum_{i=1}^{I_1}\beta_i - \frac{1}{I_2}\sum_{j=1}^{I_2}\beta_j\right);$$

$I_1 + I_2 = I$, $k = \text{const.}$ (coeficienți: 1, ... 1 de I_1 ori, -1, ... -1 de I_2 ori).
Teorema de acoperire simultană. Fie $\{\beta_1, \ldots, \beta_k\}$ parametrii necunoscuți și fie $I_1, \ldots, I_k$ intervale aleatoare care acoperă parametrii respectivi cu probabilitatea $1 - \alpha_i$, $1 \leqslant i \leqslant k$.
Atunci probabilitatea ca aceste intervale să acopere simultan parametrii corespunzători este cel puțin egală cu $1 - \sum_{i=1}^{k}\alpha_i$, adică:

$$P\left(\bigcap_{i=1}^{k} E_i\right) \geqslant 1 - \sum_{i=1}^{k}\alpha_i,$$

unde E_i este evenimentul care constă în acoperirea parametrului β_i prin intervalul I_i, $i = 1, \ldots, k$.
Metoda S a lui H. Scheffé. M.c.m. bazată pe următoarea teoremă:
Fie *modelul liniar general,** cazul ANOVA

$$\Omega \begin{cases} \mathbf{y} \sim N(\mathbf{X}'\ \boldsymbol{\beta},\ \sigma^2\mathbf{I}) \\ \text{rang } \mathbf{X}' = r. \end{cases}$$

Probabilitatea ca simultan pentru orice *funcție estimabilă** ψ să fie satisfăcute inegalitățile

$$\hat{\psi} - S\hat{\sigma}_{\hat{\psi}} \leqslant \psi \leqslant \hat{\psi} + S\hat{\sigma}_{\hat{\psi}},\quad \forall\psi \in \textit{L-spațiul q-dimensional al funcțiilor estimabile*},$$

este egală cu $1-\alpha$, unde :

$\hat{\psi} = \mathbf{C}'\hat{\boldsymbol{\beta}}$ este *estimația nedeplasată** a funcției estimabile $\psi = \mathbf{C}'\boldsymbol{\beta}$

$\hat{\sigma}_{\hat{\psi}}$ — estimația nedeplasată a Var $\hat{\psi}$

s^2 — estimația nedeplasată a lui σ^2 cu ν g.l., adică $\nu s^2/\sigma^2 \sim \chi^2_\nu$

$S^2 = q \cdot F_{\alpha;\, q,\nu}$ — constanta lui Scheffé.

Observație

Se spune că $\hat{\psi}$ este *semnificativ diferit de zero* în raport cu *m.S.* dacă intervalul $(\hat{\psi} - S\hat{\sigma}_{\hat{\psi}},\ \hat{\psi} + S\hat{\sigma}_{\hat{\psi}})$ nu acoperă pe zero sau $|\hat{\psi}| > S\hat{\sigma}_{\hat{\psi}}$.

Metoda T a lui J. W. Tukey. M.c.m. bazată pe următoarea teoremă :

Fie modelul : $\{\hat{\psi}_i\}$ sînt variabile aleatoare independente.

$\hat{\psi}_i \sim N(\psi_i,\ a_i^2\sigma^2 I),\ 1 \leqslant i \leqslant k.$

$\{a_i\}$ — constante cunoscute, $1 \leqslant i \leqslant k.$

$\{\beta_i\}$ — parametrii necunoscuți, $1 \leqslant i \leqslant k$

$\nu s^2/\sigma^2 \sim \chi^2_\nu$ și nu depinde de $\{\hat{\psi}_i\},\ 1 \leqslant i \leqslant k.$

Cu o probabilitate egală cu $1-\alpha$ toate *funcțiile parametrice** $\psi = \mathbf{C}'\boldsymbol{\beta} = \sum_i c_i\beta_i$ satisfac simultan inegalitățile :

$$\hat{\psi} - g_d \cdot q'_{\alpha;\, k,\nu} \leqslant \psi \leqslant \hat{\psi} + g_d \cdot q'_{\alpha;\, k,\nu}$$

unde

$$\hat{\psi} = \sum_i c_i\hat{\beta}_i$$

$q'_{\alpha;\, k,\nu}$ este α-cuantila superioară a *amplitudinii studentizate largite** ;

$$g_d = \max\left(\sum_{i\in P} c_i a_i,\ -\sum_{i\in N} c_i a_i\right)$$

cu

$$P = \{i,\ 1 \leqslant i \leqslant k;\ \ c_i \geqslant 0\};\quad N = \{i,\ 1 \leqslant i \leqslant k;\ \ c_i < 0\}$$

Observații

1. Pentru *modelul I unifactorial ANOVA**, caz echilibrat :

$$\Omega\begin{cases} y_{ij} = \beta_i + e_{ij}, & i = 1, \ldots, I \\ \{e_{ij}\} \text{ v.a. independente} & j = 1, \ldots, J \\ & \sim N(0,\ \sigma^2) \end{cases}$$

sau scris în forma echivalentă :

$$\{\hat{\beta}_i\}_{1\leqslant i\leqslant I} \text{ v.a. } \sim N\left(0,\ \frac{\sigma^2}{J}\right)$$

s^2 este o estimație nedeplasată a lui σ^2, independentă de $\{\hat{\beta}_i\}$, astfel încît $\nu s^2/\sigma^2 \sim \chi^2_\nu$ $(\nu = n - I)$, *m.T.* se bazează pe *teorema* :

Cu o probabilitate de $1-\alpha$, toate *contrastele** $\psi = \sum_i c_i\beta_i,\ \sum_i c_i = 0,$

satisfac simultan inegalitățile :

$$\hat{\psi} - Ts\left(\frac{1}{2}\sum_i |c_i|\right) \leqslant \psi \leqslant \hat{\psi} + Ts\left(\frac{1}{2}\sum_i |c_i|\right),$$

unde :

$\hat{\psi} = \sum_i c_i \hat{\beta}_i$; $\hat{\beta}_i = y_{i*} = \frac{1}{J}\sum_{j=1}^{J} y_{ij}$, $i = 1, \ldots, I$ este media a J observații din a i-a celulă ;

$T = \frac{1}{\sqrt{J}} q_{\alpha; I, \nu}$ este constanta lui Tukey ;

2. Compararea celor două metode pe un domeniu comun de aplicabilitate în vederea stabilirii eficienței lor relative se face prin compararea raportului pătratelor lungimilor intervalelor de încredere pentru o anumită funcție estimabilă, calculate prin metoda S, respectiv T, adică

$$R = L^2_S / L^2_T.$$

Criteriul de eficiență relativă a celor două metode :
Metoda T este preferabilă metodei S dacă :

$$R > 1.$$

În clasa tuturor *contrastelor* :* $\psi = \sum_i c_i \beta_i, \sum_i c_i = 0$ în parametrii modelului eficiența relativă maximală a metodei S față de metoda T se obține în clasa contrastelor de forma :

$$\psi = k\left(\frac{1}{I_1}\sum_{i=1}^{I_1} \beta_i - \frac{1}{I_2}\sum_{j=1}^{I_2} \beta_j\right), \qquad I_1 + I_2 = I,$$

pentru care :

$$R_{\min} = (I-1)\left(\frac{1}{I_1} + \frac{1}{I_2}\right)\frac{F_{\alpha; I-1, n-I}}{q^2_{\alpha; I, n-I}} ;$$

în timp ce metoda T este preferabilă metodei S pentru contraste de forma $\psi = k(\beta_i - \beta_j)$ $i \neq j$, pentru care

$$R_{\max} = 2(I-1)\frac{F_{\alpha; I-1, n-I}}{q^2_{\alpha; I, n-I}} .$$

PLANIFICAREA EXPERIMENTELOR. Ansamblu de tehnici de realizare a unei experiențe pentru a stabili modul în care se efectuează experiențele asupra a p factori $x_1, \ldots, x_p$ *(factori controlați)** care influențează o variabilă — răspuns, în vederea interpretării statistice a observațiilor.
PLAN EXPERIMENTAL. Schemă după care sînt făcute experimentele și înregistrate observațiile, care permite studierea influenței factorilor controlați asupra caracteristicii studiate, cu scopul de a estima influența lor și de a stabili pe cei mai importanți dintre ei.

P.e. permite separarea a două surse de variație a observațiilor : variația datorată fiecăruia dintre *factorii controlați** și variațiile întîmplătoare datorate altor *factori necontrolați*, numiți *factori de eterogenitate**.
De exemplu, în studierea efectului diferitelor îngrășăminte agricole asupra unei recolte de grîu, factorul controlat este îngrășămîntul (felul îngrășămîntului, doza utilizată), iar factorul necontrolat este variabilitatea datorată mediului (sol, climat, paraziți, insecte, etc.).
P.e. stabilește factorii considerați în experiment, numărul de *nivele** al acestor factori, numărul de observații din fiecare *celulă**, alegerea aleatoare sau extragerea la întîmplare a unităților experimentului (obiecte, persoane, loturi de piese, parcele agricole etc.) din populația respectivă, asigurarea efectuării observațiilor — prin *randomizare** — astfel încît factorii necontrolați să nu influențeze în mod sistematic asupra unităților experimentului și să modifice astfel efectele factorilor controlabili.
Exemple : **p.e.** randomizat pentru compararea a k medii ale colectivităților statistice, **p.e.** în perechi randomizate, **p.e.** în blocuri (metoda blocurilor), **p.e.** incomplete (planul pătratului latin, planul blocurilor incomplete, planuri cu grupare, planuri experimentale split-plot (planuri experimentale cu factori controlați secundari = subsidiari), planul pătratului latin incomplet, plan Youden, etc., planuri factoriale, latice, planuri experimentale „confounding").
De exemplu, **p.e.** randomizat pentru compararea timpilor medii de coagulare a sîngelui unui număr de 24 animale alese întîmplător dintr-o populație de animale, cărora li s-au administrat în mod aleator 4 regimuri alimentare A, B, C și D. Se organizează un experiment cu *randomizare completă**, întrucît, pe de o parte, unitățile experimentale (animale) sînt alese întîmplător din întreaga populație de animale aflată la locul organizării experimentului cu ajutorul unei tabele de numere întîmplătoare, iar pe de altă parte, regimurile alimentare se administrează aleator.

Plan experimental complet. Efectuarea experiențelor se poate după un *p.e.c.* (clasificare completă) dacă se consideră toate cele $N = J_1, J_2, \ldots, J_p$ celule sau sisteme de condiții experimentale.
Un **p.e.c.** se numește *echilibrat* dacă în fiecare celulă se efectuează același număr de experiențe.

Plan experimental incomplet. Un *p.e.i.* (clasificare incompletă) este un **p.e.** în care se consideră numai o parte de celule, precizînd totodată modul în care trebuie alese celulele în care vor efectua experiențele și numărul de experiențe sau de observații din celulă.
Plan experimental *„split-plot"* (**p.e.** cu factori controlați secundari). Caz special al **p.e.** cu blocuri incomplete.
Întreaga cantitate de unități experimentale (toate parcelele-ploturile) pe care se aplică nivelele unuia sau mai multor factori principali este împărțită în subunități sau subploturi pe care se aplică nivelele unuia sau mai multor factori suplimentari, considerați factori secundari (subsidiari). Așadar, fiecare unitate experimentală considerată ca unitate întreagă devine un bloc pentru subunitățile tratamentelor.
De exemplu, se testează factorul A cu 4 nivele în 3 blocuri în plan cu randomizare completă.
Factorul B (cu 2 nivele) poate fi introdus împărțind fiecare unitate a lui A în 2 subunități și aplicîndu-le cele 2 tratamente B.
Unitățile factorului A sînt unitățile întregi, iar unitățile factorului B sînt subunități sau subploturi.

După randomizare modelul este :

Bloc 1	Bloc 2	Bloc 3
$a_4b_2a_1b_2a_2b_1a_3b_2$	$a_2b_1a_1b_2a_4b_1a_3b_1$	$a_1b_1a_2b_2a_4b_2a_3b_1$
$a_4b_1a_1b_1a_2b_2a_3b_1$	$a_2b_2a_1b_1a_4b_2a_3b_2$	$a_1b_2a_2b_1a_4b_1a_3b_2$

În cadrul *p.e.s-p.* randomizarea are 2 faze : mai întîi se randomizează nivelele factorului principal A între toate unităţile experimentale ($a_1, \ldots, a_4$) şi apoi se randomizează nivelele factorului secundar B între subunităţile (2 nivele pentru fiecare unitate experimentală a lui A).
Fiecare plot-unitate experimentală poate fi considerată ca un bloc în raport cu B şi ca un bloc incomplet faţă de toată mulţimea de tratamente.
Dacă se fac repetiţii sau replicări ale experimentului, atunci factorul repetiţie este considerat aleator.
Modelul de analiză asociat unui *p.e.s-p.* cu repetiţii este un *model ANOVA mixt** cu doi factori ficşi (factor tratament, factor bloc) şi un factor aleator (factor repetiţie).
Plan experimental factorial. P.e. pentru examinarea efectelor principale şi interacţiunilor a n factori cu acţiune simultană asupra variabilei-răspuns studiate. De exemplu, planul experimental 2^n factorial este planul experimental cu n factori fiecare avînd 2 nivele.
PĂTRAT LATIN. Plan experimental incomplet pentru analiza a trei factori A, B, C prin folosirea unui *model trifactorial**) ($p = 3$) *ANOVA*, cazul aditiv (lipsa interacţiunilor între factori), satisfăcînd următoarele condiţii :
— cei trei factori au acelaşi număr de nivele m ; experienţele se produc numai în m^2 celule distincte din cele m^3 celule cîte ar corespunde *planului experimental complet** cu factorii A, B, C ; factorii A, B, C, fiecare avînd cît m nivele, se aranjează într-un tabel pătratic (cu latura m) în care cele m linii corespund nivelelor factorului A, cele m coloane corespund nivelelor B, iar cele m numere sau litere)**) care apar scrise în patrat sînt nivelele factorului C. Dispunerea acestora se face astfel încît pe orice linie sau coloană a tabelului pătratic un anumit număr să apară o singură dată ;
— planul experimental constă în organizarea experimentului în blocuri, folosind doi factori sau două criterii de clasificare, care dau liniile, respectiv coloanele pătratului latin, în timp ce tratamentele care apar în interiorul tabelului pătratic sînt nivelele celui de al treilea factor.
Observaţii.
1. Organizarea experimentului după schema **p.l.** îşi are originea în experimentele din agricultură. De exemplu, se consideră experimentul în care se compară 5 varietăţi de plante (factor A), utilizînd 5 densităţi diferite (factor B) şi 5 doze de îngrăşăminte (factor C-tratament). Organizarea experimentului după schema **p.l.** presupune împărţirea lotului în $5^2 = 25$ parcele (unităţi experimentale) şi administrarea celor 5 tratamente pe cîte 5 parcele astfel încît pe o linie un tratament (o litera din **p.l.**) să nu apară decît o singură dată. Recolta de pe fiecare parcelă este v.a. y_{ijk} v.a.-răspuns, admiţînd o descompunere lini-

*) Pentru obţinerea modelului trifactorial ANOVA → *Modelul I ANOVA multifactorial formal, clasificare încrucişată*, (pentru $p = 3$)
) Construcţia pătratelor latine se bazează pe *teoria corpurilor Galois*. Denumirea de **p.l. provine de la folosirea literelor din alfabetul latin pentru desemnarea nivelelor factorului C.

ară în efectul liniei (factor A cu nivelele $I, \ldots, V$); efectul coloanei (factor B cu nivelele $1, \ldots J$); efectul tratamentului (factor C cu nivelele $a, \ldots, e$) plus o medie generală, sub influența unor erori întîmplătoare. **P.l.** de ordinul 5 folosit este :

	1	2	3	4	5
I	*a*	*b*	*c*	*d*	*e*
II	*c*	*d*	*e*	*a*	*b*
III	*e*	*a*	*b*	*c*	*d*
IV	*b*	*c*	*d*	*e*	*a*
V	*d*	*e*	*a*	*b*	*c*

2. Pentru fixarea unui plan de experiențe prin metoda pătratului latin trebuie ales la întîmplare un astfel de pătrat din mulțimea transformatelor, adică mulțimea tuturor pătratelor latine obținute prin permutarea liniilor și coloanelor unui pătrat latin dat.

3. **Pătrat latin standard** de ordinul m sau *forma canonică* a unui pătrat latin de ordinul m, P.l. obținut prin permutarea coloanelor astfel încît pe prima linie simbolurile folosite să fie ordonate natural : a, b, c, d, e și apoi prin permutarea liniilor astfel încît prima coloană să fie (a, b, c, d, e)[1].

	1	2	3	4	5
I	*a*	*b*	*c*	*d*	*e*
II	*b*	*c*	*d*	*e*	*a*
III	*c*	*d*	*e*	*a*	*b*
IV	*d*	*e*	*a*	*b*	*c*
V	*e*	*a*	*b*	*c*	*d*

P.l. au fost introduse de MacMahon și utilizate de R. A. Fischer în probleme privind planificarea experimentului în agricultură.

4. Există $m!\,(m-1)!$ **p.l.** diferite care corespund unui același pătrat standard. Numărul **p.l.** din mulțimea transformatelor este $m!\,(m-1)!$ înmulțit cu numărul pătratelor standard de ordinul m.

5. Alegerea la întîmplare a unui **p.l.** din mulțimea transformatelor este asigurată prin alegerea unui **p.l.** de ordinul m din această mulțime și apoi prin *randomizarea** coloanelor, liniilor și simbolurilor înscrise în tabelul pătratic. Randomizarea se poate face folosind tabele de numere întîmplătoare.

6. Modelul analizei de varianță pentru un experiment organizat după schema **p.l.** este *modelul ANOVA trifactorial echilibrat* :

$$y_{ijk} = \mu + \alpha_i^A + \alpha_j^B + \alpha_k^C + e_{ijk} \quad (i, j, k) \in M$$

$\{e_{ijk}\}$ sînt variabile aleatoare independente și repartizate $N(0, \sigma^2)$ pentru orice $(i, j, k) \in M = \{(i, j, k) : 1 \leq i, j, k \leq m\}$,

unde μ este media generală ; $\alpha_i^A, \alpha_j^B, \alpha_k^C$ sînt efectele principale ale factorilor A, B, C, adică efectele liniilor, efectele coloanelor, respectiv efectele tratamentului care intră în schema experimentală aleasă ; e_{ijk} sînt erori tehnice.

PLAN ÎN BLOCURI. Plan experimental în care unitățile experimentale sînt grupate în clase sau *blocuri* și conform căruia unitățile experimentale dintr-un

bloc trebuie să fie cît mai omogene, iar între două blocuri oarecare să existe deosebiri.

Plan în blocuri întîmplătoare. P.b. conform căruia, în fiecare bloc, cele I tratamente sînt repartizate întîmplător în raport cu I unităţi experimentale. Într-un *P.b.i.* fiecare din cele $I!$ moduri de repartizare a tratamentelor pe unităţile experimentale are aceeaşi probabilitate şi repartiţiile experienţelor în blocuri sînt independente. Pentru obţinerea planurilor cu blocuri întîmplătoare este indicat să se folosească tabela permutărilor întîmplătoare.

Plan experimental în blocuri incomplete. P.b. conform căruia numărul de unităţi experimentale din bloc este mai mic decît numărul I al tratamentelor. Se presupune în plus că fiecare tratament apare de acelaşi număr r de ori, că blocurile au acelaşi număr de unităţi experimentale k şi că nu există nici un tratament care să apară de mai multe ori într-un bloc. De aici se deduce că numărul n de experienţe într-un *p.b.i.* satisface condiţia

$$n = rI = kJ.$$

Plan experimental conform căruia o anumită pereche de „sisteme de condiţii" apare întotdeauna într-un număr constant de blocuri.

Dacă notăm că $\lambda_{ii'}$ numărul de blocuri în care apar sistemele de condiţii i şi i', atunci se deduce că pentru un *p.b.i.* (cazul echilibrat), trebuie să avem

$$\lambda_{ii'} = \begin{cases} \lambda & \text{dacă } i \neq i' \\ r & \text{dacă } i = i' \end{cases}$$

unde λ şi r sînt constante. Se poate arăta că

$$\lambda = \frac{r(k-1)}{I-1}.$$

Modelul ANOVA de analiza datelor reale conform *p.b.i* este un caz particular al *modelului I ANOVA bifactorial echilibrat**.

RANDOMIZARE. Procedeu de alocare întîmplătoare a unor combinaţii de tratamente pe unităţi experimentale care reduce sursele de variabilitate, elimină efectele sistematice ale factorilor necontrolabili, înlătură neajunsurile ce ar decurge din erorile de măsurare sau de citire a rezultatelor, asigură validitatea inferenţelor statistice făcute asupra parametrilor modulului ales pentru analiză. Mai general, r. constă din folosirea unei tehnici corespunzătoare pentru a determina repartiţia în cadrul unei mulţimi de elemente asemănătoare (ordonarea pe o dreaptă, locul unei observaţii într-un tabel etc.) sau pentru a desemna un anumit număr de unităţi experimentale (obiecte, persoane, experimente, etc.) alese la întîmplare dintr-o populaţie supusă cercetării.

R. se poate face prin experimentul de aruncare cu moneda, prin extragerea unei bile sau cartele dintr-o urnă, cu ajutorul unei tabele de numere întîmplătoare etc.

Planificarea unui experiment trebuie să ia în considerare şi *problema selectării factorilor*, adică alegerea investigării acelor variabile care au un efect deosebit asupra răspunsului experienţei.

În anumite condiţii — independenţa observaţiilor, repartiţia normală etc. —, asigurate de tehnicile planificării experimentului (de planul experimental folosit), efectele factorilor asupra variabilei-răspuns pot fi determinate prin analizarea *variabilităţii* observaţiilor prin metodologia oferită *de analiza de varianţă**

(analiza dispersională). În acest scop este necesar să se postuleze un model care leagă observațiile de anumite combinații ale valorilor factorilor, adică de *nivelele factorilor**.

Randomizare parțială. Într-un experiment în care intervin mai mulți factori, r. se face numai pentru nivelele unui factor și în acest caz experimentul este cu *r.p.* sau *restrictivă*.

Randomizare completă. Dacă într-un experiment r. se efectuează asupra tuturor factorilor, experimentul este cu *r.c.*

CONFOUNDING. Tehnică de organizare a unui plan experimental factorial cu mai mulți factori, fiecare avînd mai multe nivele, tehnică care se repetă în mai multe blocuri, necesitînd un număr mare de unități experimentale. Din această cauză există riscul de a nu putea realiza condiții de omogenitate sau de uniformitate în interiorul fiecărui bloc.

C. constă în subîmpărțirea fiecărui bloc în blocuri mai mici, care conțin numai o parte din combinațiile factorilor, alese astfel încît să permită studiul *efectelor principale** și al anumitor *interacțiuni**, celelalte interacțiuni de ordin superior fiind parțial sau total influențate de diferențele dintre blocuri. Aceste interacțiuni devin una din componentele comparației dintre blocuri și se spune că sînt „confundate" cu blocurile, în timp ce efectele factoriale (efecte principale, interacțiuni) nu sînt influențate de diferențele dintre blocuri. Se mai spune că sînt „libere" față de efectele blocurilor sau că sînt „neconfundate" cu blocurile. Volumul blocului este redus prin sacrificarea exactității anumitor interacțiuni de ordin superior.

C. a fost propusă de R. A. Fisher și dezvoltată ulterior de D. J. Finney, F. Yates, O. Kempthorne,, J. C. R. Li etc. în legătură cu planurile experimentale din agricultură.

VARIA

XVIII. ISTORICUL STATISTICII DIN ROMÂNIA

Ca şi în celelalte ţări din Europa, evoluţia statisticii în România cunoaşte aproximativ aceleaşi particularităţi. Înregistrări care prefigurează recensămintele au avut loc pe teritoriul României în decurs de secole; statistica, ca disciplină, a început cu faza descriptivă, evoluînd treptat spre statistica cantitativă; în prima perioadă, statistica a evoluat în legătură organică cu demografia; cadrul instituţional al statisticii statului a fost creat în deceniul al şaselea al sec. al XIX-lea, încă în atmosfera caracteristică a „entuziasmului statistic", iar colaborarea internaţională a statisticii româneşti datează de asemenea din această epocă. Pe măsura dezvoltării sale, statistica ca metodă pătrunde treptat în diferite ştiinţe: economie, sociologie, istorie, geografie, biologie şi medicină, precum şi în altele, în aşa fel încît odată cu progresul metodologic, realizat de calculul probabilităţilor şi statistica matematică, se lărgeşte şi se diversifică cîmpul de aplicare.

1. *Perioada prestatistică*

După unii cercetători ai istoriei statisticii [31], pe teritoriul Daciei romane s-au făcut recensăminte ale populaţiei („census"), primul fiind cel dispus chiar de împăratul Traian după cucerirea Daciei şi despre care ne-au rămas relatări de la Ulpian; ultimul „census" s-a efectuat în jurul anului 240. Statele feudale de pe teritoriul României — Ţara Românească, Moldova şi Transilvania — au continuat, în forme specifice, tradiţia „census"-ului roman: s-au practicat catastife, catagrafii şi conscripţii, cărora li se adaugă urbariile şi altele. Toate prezintă interes sub raportul tehnicii de înregistrare şi au o certă valoare ca surse de cunoaştere a situaţiei demografice şi economice.

Menţiuni despre catastife datează din sec. al XV-lea. Cel mai cunoscut este catastiful lui Petru Şchiopul, din 1591, publicat de Nicolae Iorga; se ştie de asemenea despre catastiful lui Mihai Viteazul, din noiembrie 1599. *Catagrafia* succede catastifului, începînd din a doua jumătate a sec. al XVIII-lea. S-a practicat în două forme: catagrafia religioasă şi catagrafia fiscală. Un tip evoluat de catagrafie a fost catagrafia regulamentară, efectuată în condiţiile Regulamentului Organic, introdus în Muntenia la 1 iulie 1831 şi în Moldova la 1 ianuarie 1832.

Conscripţia, ca înregistrare, s-a practicat în Transilvania, sub diferite forme: *fiscală* — încă din sec. al XIV-lea, ca înscriere a capilor de familii contribuabile; *urbarială* — pentru reglementarea raporturilor dintre feudali şi iobagi; *religioasă* — pentru înregistrarea persoanelor aparţinînd diferitelor culte; *militară* — echivalentă cu un adevărat recensămînt al populaţiei.

Pentru Ţara Românească ne este cunoscută înregistrarea făcută de Constantin Brîncoveanu, în anul 1713, pentru plata haraciului de la care nu a fost exceptată nici o categorie fiscală sau socială. Catagrafia lui Constantin Mavrocordat (1741) a fost studiată aprofundat [32] şi publicată în ce priveşte tehnica de înregistrare. Este vorba de elementele pe care le cuprinde Condica de venituri

și cheltuieli rămasă de la C. Mavrocordat din cea de-a doua lui domnie în Moldova (1741—1744) și care este primul act în care tehnica de înscriere catagrafică a populației este arătată clar. Mai puțin cunoscută este catagrafia făcută de N. Mavrocordat în 1725, tatăl lui C. Mavrocordat, ca și aceea efectuată de C. Mavrocordat, în 1739 sau 1740, în Țara Românească. Se apreciază că domnitorul C. Mavrocordat urmărea să întocmească o lucrare descriptivă a Moldovei la care ar fi adăugat și datele rezultate din catagrafia 1741, lucrare asemănătoare cu cele întocmite în Europa occidentală de statistica descriptivă și de aritmetica politică.

Dintre catagrafiile din Moldova, au fost studiate și publicate cele din 1820 și 1831 [13], dar se păstrează în Arhivele de la Iași și alte înregistrări de același tip. Catagrafiile dispuse de Regulamentul Organic au început să fie aplicate în anul 1831, în Moldova; ele aveau o periodicitate de șapte ani (1838; 1845 și ultima 1851/52). Tehnica catagrafiilor regulamentare este mult mai dezvoltată; ea a fost prevăzută [13] în cap. III, secțiunea IV (Chipul facerii catagrafiilor și a împărțirii dăjdiilor și altor dări), art. 69—95 din Regulamentul Organic al Țării Românești și în cap. III, secțiunea IV (Chipul facerii însrierii sau catagrafiei), art. 84—89 din Regulamentul Organic al Moldovei și, în fapt, sintetizează experiența de secole a catagrafiilor ce s-au efectuat în Țara Românească și în Moldova.

În Transilvania se cunosc numeroase conscripții [27]. Se amintesc astfel conscripția fiscală din anii 1721—1722, care nu cuprindea decît o parte a populației, precum și conscripția din 1750. Dintre conscripțiile confesionale, prezintă un mare interes prima care se referă la populația română-ortodoxă și greco-catolică (unită), efectuată în anul 1733 de episcopul Inochentie Micu Klein; a doua este din 1750, efectuată de episcopul Petre Pavel Aron. Cea mai bine organizată a fost conscripția românilor din Transilvania, efectuată de către generalul Buccow, în anii 1760—1762. Conscripția din 1785, pe timpul împăratului Iosif al II-lea, este considerată un recensămînt general al populației. Semnificativ este faptul că pentru acest recensămînt au fost elaborate și instrucțiuni în limba română. Pentru prima dată la recensămîntul din 1785 au fost înregistrați nu numai capii de familie sau persoanele de anumită virstă, pe porți sau pe sesii iobăgești, ci toți locuitorii, indiferent de starea lor socială. Pe baza conscripțiilor din Transilvania s-a putut determina numărul populației pe naționalități la diferite momente de timp, punîndu-se în evidență proporția majoritară a românilor în Transilvania.

Este locul să precizăm că diferitele înregistrări din perioada prestatistică — catagrafii și conscripții — se efectuau în scopuri fiscale și militare. Cu recensămîntul populației și al agriculturii din Țara Românească, efectuat în anul 1838, se inaugurează seria recensămintelor moderne în țara noastră. Acesta este urmat de recensămîntul din Moldova din anii 1859—1860 și cel din Transilvania (1851—1852).

Multă vreme s-a afirmat că primul recensămînt modern al populației este cel din 1859—1860, legat de numele lui Dionisie Pop Marțian, în Muntenia, și al lui Ion Ionescu de la Brad, în Moldova. Cercetări ample de arhivă, făcute de G. Retegan și I. Donat, au dus la descoperirea materialelor recensămîntului populației și agriculturii din 1838, făcut în Țara Românească. Studierea acestuia le-a permis cercetătorilor amintiți să afirme că acest recensămînt întrunește caracteristicile unui recensămînt modern și să-l declare ca primul din această categorie în istoria recensămintelor din România.

Se face precizarea [13] că acest recensămînt nu reprezintă o simplă catagrafie regulamentară; autorii săi l-au denumit „statistica Principatului", iar caracteristicile sale sînt: a) universalitatea înregistrării — înscrierea în liste a întregii populații aflate la data efectuării pe teritoriul țării; b) înregistrarea datelor pentru fiecare persoană; c) inițierea și efectuarea lui de către autoritatea de stat pe un anumit teritoriu; d) nu a avut un caracter fiscal. În organizarea și efectuarea recensămîntului din 1838 s-au adus soluții metodologice și tehnice originale și care devanseză, în multe privințe, recensămintele similare din Europa. Recensămîntul a fost inițiat de Departamentul Treburilor din Lăuntru (secția a 3-a, biroul 1), iar temeiul legal al efectuării lui l-a constituit art. 151 al Regulamentului Organic al Țării Românești. Programul de observare al recensămîntului a fost foarte bogat; unitatea de înregistrare a fost gospodăria, ar unitatea de observare — persoana. S-a aplicat metoda directă de înregistrare: cei însărcinați cu înscrierea au fost obligați „ca cercetarea aceasta să o facă prin chiar vederea ochilor, preumblîndu-se înșile din casă în casă în tot cuprinsul fiecărui sat". Nu a existat un moment critic — caracteristică ce lipsea tuturor recensămintelor din acea vreme din Europa. Rezultatele recensămîntului populației și agriculturii din 1838 s-au publicat parțial în „Almanah al Statului" pentru anul 1842, iar în întregime în „Analele parlamentare ale României", IX, partea 1, p. 1162 și urm. Cu această ultimă informație, figurează și în bibliografiile internaționale privind istoria recensămintelor din lume

2. *Instituții de statistică*

Primele organe ale statisticii statului au fost înființate în anul 1859, curînd după Unirea Principatelor. Totuși, într-o formă elementară, asemenea organe au fost prevăzute în Regulamentul Organic (Țara Românească, 1 iulie 1831 și Moldova, 1 ianuarie 1832). În Țara Românească, atribuțiile de statistică reveneau Departamentului Treburilor din Lăuntru: secția întîia se ocupa de statistica lunară a prețurilor produselor agricole și industriale, secția a doua era însărcinată cu statistica producției și consumului, iar secția a treia — cu statistica fenomenelor demografice și statistica țării (populație, clădiri etc.). În general, în perioada care a precedat Unirea Principatelor și înființarea organelor de statistică, a existat un mare interes pentru statistică, în condițiile transformărilor economice ce au urmat tratatului de la Adrianopole (1829) și s-au efectuat numeroase lucrări statistice.

Actul de naștere al organului de statistică în Țara Românească datează din 12 iulie 1859: „Biroul statistic vremelnic de pe lîngă Departamentul din Lăuntru" a devenit *Oficiul Central de Statistică Administrativă* (Legea din 3 iulie 1859, sancționată de Domnitorul Alexandru Ioan Cuza prin hotărîrea nr. 276 din 12 iulie 1859). Șeful Oficiului Central de Statistică a fost numit *Dionisie Pop Marțian**, al cărui rol în organizarea statisticii statului și promovarea statisticii a fost deosebit de important.

În Moldova, noul organ de statistică a statului a luat ființă la 1 iulie 1859: Alexandru Ioan Cuza a semnat decretul pentru înființarea *Direcției Centrale de Statistică* a Moldovei. Ca director a fost numit *Ion Ionescu de la Brad**. Se recunoaște astăzi că D. P. Marțian și Ion Ionescu de la Brad sînt ctitorii statisticii oficiale românești și, de asemenea, precursori ai statisticii ca știință în România. Atît în 1959, cu prilejul centenarului înființării statisticii, cît și în 1969 și 1979 au fost evocate figurile celor doi ctitori și s-a pus în evidență importanta lor contribuție în dezvoltarea statisticii românești.

În două studii de sinteză [33] și [23], este înfățișată istoria statisticii oficiale în perioada 1859 pînă în anul 1947.

Prima lucrare importantă inițiată de noile organe de statistică a fost recensămîntul populației, care s-a efectuat în anii 1859 și 1860. În Moldova, Ion Ionescu de la Brad a fost curînd înlocuit cu C. Negruzzi, acestuia i-a luat locul Dim. Gusti. Pînă în 1862, cele două instituții au funcționat separat : la 31 iulie 1862, ele sînt unificate — în cadrul acțiunii generale de unificare a administrației de stat — sub denumirea de *Oficiul Central de Statistică*, iar Dionisie Pop Marțian este numit director al instituției unificate, calitate pe care o păstrează pînă la moartea sa, survenită la 2 iulie 1865.

Pentru perioada de început a instituției centrale de statistică trebuie menționate cîteva aspecte. Înființarea ei are loc în perioada în care în întreaga Europă există un mare interes pentru statistică — în 1853 a avut loc, din inițiativa lui A. Quételet și W. Farr, primul Congres internațional de statistică —, știința statistică era într-un real progres de la statistica descriptivă spre cea cantitativă. D. P. Marțian și Ion Ionescu de la Brad erau buni cunoscători ai statisticii din acea vreme și de asemenea ai experienței organizatorice din alte țări din Europa. În concepția lor, statistica trebuia să fie un important instrument pentru conducerea statului. În memoriul înaintat Domnitorului Alexandru Ioan Cuza, D. P. Marțian a formulat astfel acest rol : „Cifrele statistice sînt cele mai bune argumente, mărturia cea mai autentică ce se poate produce în tot minutul în Sfatul principelui și a corpului legislativ și a instrucției celor ce vor să cunoască patria și averile ei, fiind baza evidenței statului și a trebuințelor lui. Statistica, prin urmare, este nu numai un ajutor al administrației, ci este totodată un monument național" [22].

Să mai amintim că noul organ de statistică era conceput ca un organ central, dispunînd de o rețea teritorială și cu atribuții centralizate față de diferitele departamente și ministere, idee care avea să se realizeze în țara noastră abia după 1947. Din această perioadă de început datează o serie de lucrări importante — recensămîntul 1859/60, recensămîntul stabilimentelor industriale din 1863 și altele —, prima revistă de statistică („Analele statistice"), primul manual de statistică teoretică („Povățuiri ..." de Ion Ionescu de la Brad). Evoluția ulterioară a statisticii românești este marcată, pentru o bună perioadă de timp, de regres, de încercări de redresare, pînă la Leonida Colescu, căruia i s-ar cuveni titlul de al doilea ctitor al statisticii românești. După moartea lui D. P. Marțian, la conducerea Oficiului Statistic este numit I. Petrescu, în timpul căruia Oficiul Central devine o simplă secție a Ministerului de Interne. Se inițiază diferite proiecte, de acte normative pentru redresarea statisticii oficiale, dintre care trebuie amintite eforturile lui Al. Pencovici din anii 1871—1873, legea din 29 nov. 1871 (pusă în aplicare în ianuarie 1872) prin care ia ființă *Serviciul General de Statistică a României*. Demn de menționat este faptul că la Congresul Internațional de Statistică de la Budapesta (1876), Al. Pencovici relatează, în comunicarea sa, că legea din 1871 s-a inspirat din recomandările internaționale ca și din experiența organizatorică proprie a țării. În anii 1872—1876 au fost create diferite birouri de statistică pe lîngă unele ministere și instituții centrale ; în 1876 a fost creat primul birou de statistică al Municipiului București—printre primele orașe din Europa. În 1892 a devenit conducător al instituției centrale *Constantin E. Crupenski** ; la 1 aprilie 1892, Biroul de Statistică se transformă în Direcția Statisticii Generale. În această perioadă statistica cunoaște un progres ; se editează „Buletinul statistic general al României". O adevărată înflorire a statisticii românești se înregistrează pe timpul lui *Leonida Colescu** : se efectuează două recensăminte ale populației (1899 și 1912), o anchetă industrială, mai multe anchete agricole. L. Colescu

s-a înconjurat de statisticieni, demografi și economiști de valoare : I. Scărlătescu, *I. N. Angelescu**, Ilie Georgian, cărora li se datorează remarcabile studii și analize ale datelor statistice. Perioada 1891—1916 nu a fost scutită însă de o serie de reorganizări și restructurări ale statisticii centrale și ale celei departamentale, printre care trecerea Direcției Statisticii Generale în subordinea diferitelor ministere.

Odată cu desăvîrșirea statului național român în 1918, statistica oficială a intrat într-o nouă etapă. Prin Decretul-lege nr. 122 din 25 ianuarie 1919, Direcțiunea Generală a Statisticii trece în subordinea Ministerului Industriei și Comerțului și se organizează ca instituție centrală, fiind formată din : Comisia superioară de statistică și Direcția Generală a Statisticii. Leonida Colescu este înlocuit cu Ion Teodorescu. La 1 aprilie 1921, Direcția Generală a Statisticii a fost desființată. Unele schimbări organizatorice se efectuează prin intermediul Bugetului de Stat. În 1925 ia ființă Institutul de Statistică Generală a Statului, subordonat Ministerului Industriei și Comerțului. În 1929, acest institut a trecut în subordinea Ministerului Muncii, Sănătății și Ocrotirilor Sociale. În martie 1930 apare legea privind efectuarea recensămîntului general și înființarea Comisiei pentru recensămînt — președinte fiind numit profesorul Dimitrie Gusti, — iar în 1933 Oficiul de demografie și statistică sanitară se contopește cu Comisia de recensămînt și devine *Institutul de Demografie și Recensămînt al României*. O nouă perioadă se inaugurează în 1936 (Legea din 1936, iar Regulamentul din 1937), prin înființarea *Institutului Central de Statistică*, pe lîngă Ministerul de Interne, avînd ca director pe dr. *Sabin Manuila**. În 1938, Institutul Central de Statistică se scindează în : Institutul demografic al României (Ministerul Sănătății și al Ocrotirilor Sociale) și Oficiul de statistică economică (Ministerul Industriei și Comerțului). Să mai amintim că în 1933 s-a înființat Corpul tehnic al statisticienilor.

Decretul-lege nr. 488 din 1 iunie 1941 decretează Institutul Central de Statistică drept unicul organ al statisticii statului, fiind subordonat Președinției Consiliului de Miniștri, cu următoarea formulare : „Institutul Central de Statistică este singurul organ de statistică a statului. Toate unitățile statistice departamentale cu bugetul, personalul, inventarul și arhivele lor se încadrează în Institutul Central de Statistică, care trece pe data prezentului decret la Președinția Consiliului de Miniștri". Se înființează, de asemenea, Serviciul exterior, cuprinzînd oficianți statistici la comune, censori statistici la plăși și controlori statistici la județe. În anii 1947 și 1948, director general al Institutului Central de Statistică a fost *Anton Golopenția**.

În această formă organizatorică, Institutul Central de Statistică a funcționat pînă la Conferința ministerială pentru reorganizarea statisticii din România (octombrie 1947), care a formulat noi principii de organizare și funcționare a statisticii, adaptate noilor condiții politice și economice.

În iulie 1948, după înființarea Comisiei de Stat a Planificării, Institutul Central de Statistică a trecut în subordinea noului organ central de planificare, iar în 1950 a devenit Direcțiunea Generală a Statisticii, în subordinea Consiliului de Miniștri, schimbîndu-și, în 1951, denumirea în Direcția Centrală de Statistică. În anul 1970, Direcția Centrală de Statistică trece în subordinea Consiliului de Stat al Republicii Socialiste România.

Succintul istoric al statisticii românești în perioada 1859—1947 pune în evidență cîteva aspecte semnificative. Organizarea instituțională a statisticii are loc aproximativ în aceeași perioadă cu celelalte țări din Europa, în unele cazuri cu un decalaj neînsemnat. Lucrările statistice care s-au efectuat au folosit uneori soluții metodologice originale, cum a fost în cazul recensămîntului popu-

lației. Pînă în anul 1918, au fost efectuate recensămintele din 1838, 1859/60, 1899 și 1912 în Țara Românească și Moldova, apoi în Principatele Unite; în Transilvania, recensăminte de populație s-au făcut în 1850—1851, 1857, 1869, 1880, 1890, 1900 și 1910. S-au efectuat diferite alte recensăminte și anchete, în așa fel încît s-au acumulat informații valoroase pentru studiul populației, economiei naționale, industriei, agriculturii și altor domenii. Pe această bază, cu metode apropriate, s-ar putea întreprinde un studiu sistematic al istoriei populației și a economiei naționale.

3. Statistica ca știință: concepte, metode și aplicații

Prima lucrare remarcabilă de etnologie, geografie și istorie și care ar aminti de statistica descriptivă este „Descrierea Moldovei" a învățatului domnitor Dimitrie Cantemir, îmbinînd în persoana sa calități de om de stat, istoric, geograf, filozof, literat, muzicolog. Lucrarea a fost solicitată lui Dimitrie Cantemir de către Academia de Științe de la Berlin, al cărei președinte era filozoful și matematicianul G. W. Leibniz (1646—1716) și al cărei membru a fost ales Dimitrie Cantemir în anul 1714. Scrisă în limba latină, cu titlul „Descriptio antiqui et hodierni status Moldaviae", lucrarea a fost tradusă în limba germană în anii 1769—1770. Ea cuprinde o parte geografică (7 capitole), o parte politică (19 capitole) și o parte eclesiastică și literară (5 capitole). Ea are de asemenea o hartă („Tabula Geographica Moldaviae"). „Descrierea Moldovei" nu este o statistică numerică sau tabelară, ci una descriptivă. Traducerea ei în germană s-a făcut de A. F. Büsching (1724—1793), profesor la Göttingen și considerat drept părinte al statisticii comparative a statului [14] și [16].

De o deosebită valoare este prima monografie sanitară și demografică, datorată doctorului *Constantin Caracaș** și intitulată „Topografia Țării Românești și observații antropologice privitoare la sănătatea și bolile locuitorilor ei", apărută în limba greacă și republicată în 1937. Această monografie [28] se compune din două părți, repartizate pe capitole și paragrafe. Se începe cu prezentarea hotarelor țării și situația fizică, climă și produse, originea poporului român. Se dă cifra de un milion pentru populația Țării Românești, se examinează cauzele depopulării [29], încriminîndu-se, printre altele, „politica slabă și păcătoasă a guvernului Valahiei și poziționea sa geografică, care o face să fie teatru de războaie continui între puterile vecine". Se examinează cauzele mortalității generale și infantile deosebit de mare, se pun în evidență clasele și categoriile sociale din acea vreme. Partea a doua, denumită „Topografia parțială a Munteniei", dă descrierea fizică a celor 17 județe, cu referiri la starea sanitară. Monografia lui C. Caracaș se bazează în special pe informațiile furnizate de catagrafia fiscală a lui Vodă Caragea din anii 1814—1815. Se poate face afirmația că „Topografia ..." este prima lucrare de statistică descriptivă, împreună cu elemente numerice, din istoria statisticii românești.

O altă lucrare de statistică descriptivă, cu elemente de statistică numerică și tabelară, „Noțiuni statistice asupra Moldovei", este datorată lui *Nicolae Șuțu**, care n-a fost „numai primul nostru economist, ci și unul dintre primii învățați și oameni politici din Principatele Române, care și-a dat seama de marea importanță a statisticii ca instrument de cunoaștere socială și ca pîrghie în opera de guvernare a țării" [13]. Mai mult, se face aprecierea că Șuțu este „pe lîngă primul nostru economist, primul statistician din Țările Române, care pleacă de la o concepție științifică despre statistică și care folosește ca mijloc de analiză a dezvoltării economice a țării noastre". Scrisă în limba franceză sub titlul „Notions statistiques sur la Moldavie" (1849), ea este tradusă în

limba română și publicată cu adăugiri de Teodor Codrescu în 1852, sub titlul „Notiții statistice asupra Moldaviei". În colecția Bibliotecii Academiei se află un manuscris datat din 1837, care cuprinde considerații teoretice asupra caracterului științific al statisticii și asupra rolului statisticii în opera de guvernare a țării.

Lui N. Șuțu îi erau familiare lucrările de statistică ale unor autori francezi, de la care împrumută unele concepții, nu însă fără discernământ critic. După părerea sa, statistica face parte din grupa științelor descriptive și nu dinamice, dar „faptele pe care ea le descrie sînt variabile". Statistica este un mijloc de cunoaștere a faptelor care se raportează la „starea fizică, morală, intelectuală și politică a societății într-un alt timp și într-un alt loc, comparate cu faptele dintr-un alt timp și alt loc". Rolul ei în cadrul statului este astfel definit de N. Șuțu : „Concepută pe o scară largă statistica devine cea mai bună busolă în arta guvernanților ca și a guvernaților" și mai departe : „abia prin statistică știința guvernării va apuca pe un drum mai sigur, mai rapid și mai conciliant, iar guvernele vor cîștiga mai multă încredere în ele și mai multă autoritate asupra publicului". Lucrarea lui N. Șuțu, „Noțiuni statistice", prezintă un material informativ bogat asupra stării de lucruri din Moldova la mijlocul veacului al XIX-lea, însoțit de o analiză multilaterală.

O concepție originală asupra statisticii o întîlnim în lucrările lui Dionisie Pop Marțian, expusă în studiile publicate în „Analele statistice și economice", dar mai ales de Ion Ionescu de la Brad, în lucrarea sa „Povățuiri pentru facerea catagrafici Moldovei, precedate de oarecare elemente de statistică", publicată la Iași în anul 1859.

Ion Ionescu de la Brad preia o definiție a statisticii, aceea dată de Moreau de Jonnès, potrivit căreia „statistica este știința faptelor sociale confirmată prin termeni numerici". Ca atare, regăsim cele două elemente ale statisticii : caracterul social și expresia numerică ca trăsătură specifică. În cap. II al „Povățuirilor...", Ion Ionescu de la Brad se ocupă de probleme de metodă recunoscînd două : a) metoda naturală sau de expozițiune, înțelegînd prin aceasta observarea statistică și centralizarea datelor, și b) metoda de inducțiune (aritmetica politică), care constă în „a dobîndi rezultate prin ajutorul îndeminărilor ce dă calculul". Idei interesante formulează Ion Ionescu de la Brad în legătură cu metoda sondajului. „Povățuirile ..." prezintă un mare interes și pentru calitățile de organizator al recensămîntului populației, de care dă dovadă Ion Ionescu de la Brad. Această lucrare trebuie considerată ca primul manual de statistică teoretică din țara noastră. Merite importante are Ion Ionescu de la Brad și prin folosirea statisticii în lucrările sale cu caracter monografic [13]. Este vorba de trei monografii de județe (Dorohoi, Mehedinți, Putna), publicate între anii 1866 și 1869, ceea ce îndreptățește afirmația că Ion Ionescu de la Brad este un precursor al metodei monografice în studiul vieții social-economice din țara noastră.

Concepții în legătură cu statistica sau aplicații ale acesteia la studiul fenomenelor social-economice se întîlnesc la gînditori de seamă din sec. al XIX-lea. În operele lui Nicolae Bălcescu, statistica — datele și metodele ei — este folosită pe scară largă [13].

Alexandru Papiu Ilarian (1828—1877), jurist și istoric, participant la revoluția din 1848, mai tîrziu profesor la Universitatea din Iași, a folosit pe scară largă statistica în lucrările sale, cum ar fi „Istoria românilor din Dacia Superioară" (1851 și 1852). *George Barițiu* (1812—1893), istoric, jurist și om politic, conducător al revoluției din 1848 și fondator al „Astrei", președinte al Academiei

Române, a valorificat cu talent datele statistice comparative în lupta pentru afirmarea drepturilor românilor din Transilvania [13]. Un alt om de știință, *Petre S. Aurelian* (1833—1909), economist și agronom, om politic, a dat remarcabile studii economice („Terra nostra. Schițe economice asupra României", București, 1880), cu folosirea largă a datelor statistice [13].

Toate acestea vădesc climatul instaurat în legătură cu statistica, cu importanța ei, tendință generală caracteristică și celorlalte țări din Europa.

4. *Statistica românească și cooperarea internațională*

În 1853 se inaugurează seria *congreselor internaționale de statistică**, continuată cu *Congresele internaționale de demografie și igienă** și apoi cu activitatea *Institutului Internațional de Statistică* la care statisticienii din România au adus o remacabilă contribuție [20].

Cele opt congrese internaționale de statistică au avut loc astfel : 1853 (Bruxelles), 1855 (Paris), 1857 (Viena), 1860 (Londra), 1863 (Berlin), 1867 (Florența), 1869 (Haga), 1872 (Petersburg). Urmează apoi sesiunile Comisiei Permanente de Statistică : 1873 (Viena), 1874 (Stockholm), 1876 (Budapesta), 1878 (Paris). Între 1876—1885 au avut loc Congresele Internaționale de igienă și demografie. Institutul Internațional de Statistică, fondat la 24 iunie 1885 la Londra, a ținut 14 adunări generale pînă la primul război mondial.

Prima participare internațională este cea a lui Dionisie Pop Marțian, la Congresul de la Viena (1857), ca observator, fiind în acea vreme student la Universitatea din Viena ; se pare că profesorul său, Lorenz von Stein, i-a înlesnit această participare. În schimb, la cel de-al V-lea Congres Internațional de Statistică (Berlin, 1863), Dionisie Pop Marțian a luat parte ca delegat oficial și a prezentat o comunicare asupra lucrărilor de statistică din Principatele Române. La Congresul din Florența (1867) au participat Gr. Bengescu și Gr. Vulturescu, acesta din urmă prezentînd o comunicare. Un moment important este Congresul de la Haga (1869) la care au luat parte Ion Ionescu de la Brad și Al. Pencovici. La Congresul de la St. Petersburg (1872) au participat Al. Lahovari și Al. Pencovici. Aici s-a decis înființarea unei comisii permanente internaționale de statistică, care a ținut patru sesiuni. La cea din Viena (1873) a participat Al. Pencovici ; la cea de la Stockholm (1874) — Al. Lahovari și Al. Pencovici, la cea din Budapesta (1876) — Al. Pencovici și C. Exarcu. După înființarea Institutului Internațional de Statistică, delegații români au început să participe la sesiunile acestuia : la sesiunea a treia de la Viena (1891) au participat T. G. Djuvara și Gr. P. Olănescu : la Berna (1895) participă Gr. P. Olănescu și C. Crupenski, la St. Petersburg (1897) iau parte N. Basilescu și Gr. P. Olănescu, la Budapesta (1901) a participat Gr. P. Olănescu. De la sesiunea a noua, de la Berlin (1903), participarea română devine și mai importantă : la acest congres iau parte Gr. P. Olănescu, *G. D. Creangă** și Leonida Colescu care prezintă comunicarea : „Populațiunea României în ultimii 25 de ani". La sesiunea din Londra (1905), L. Colescu prezintă comunicarea : „Cultura cerealelor în principalele țări din lume" ; la sesiunea din Copenhaga (1907), țara este reprezentată de G. D. Creangă ; la cea din Paris (1909) participă L. Colescu și G. D. Creangă ; la sesiunea a XIII-a de la Haga (1911) îi întîlnim din nou pe L. Colescu și C. Crupenski, iar la sesiunea a XIV-a de la Viena (1913), L. Colescu prezintă o comunicare cu problemele organizatorice și rezultatele recensămîntului general din România (1 ian. 1913).

Specialiștii români au participat și la Congresele internaționale de igienă și demografie. La cel de la Geneva (1882) a participat dr. *I. Felix** ca și la cel

de la Haga (1884); la Congresul din Viena (1887) iau parte dr. I. Felix, dr. V. Babeș și alți 12 delegați; la cel din Paris (1889), dr. I. Felix este ales președinte de onoare și prezidează alături de L. Pasteur, E. Levasseur și alții. La Congresul din Londra (1891), dr. I. Felix, dr. V. Babeș și alți delegați români prezintă referate. Să mai amintim de participarea specialiștilor din România și la congresele internaționale de agricultură și alte reuniuni internaționale, ca și faptul că unii din ei au fost membri ai diferitelor societăți științifice. Astfel, N. Șuțu a fost membru al Societății regale de statistică din Londra, calitate pe care a avut-o mai tîrziu și Ion Ionescu de la Brad, Leonida Colescu a fost membru al Societății de Statistică din Paris. În afară de L. Colescu, C. Crupenski, Gr. Olănescu au mai fost membri ai Institutului Internațional de Statistică, de asemenea, Dimitrie Sturza (1833—1914), Ion Teodorescu (1880—1956), Eugen Grigore Neculcea (1876—1953), iar din 1935 a devenit membru al Institutului Internațional de Statistică acad. Octav Onicescu.

Din succinta prezentare a participării statisticii românești la manifestările internaționale putem să apreciem interesul ca și contribuția statisticienilor români la cooperarea internațională, la schimbul de idei și de experiență în munca statistică.

Este locul să amintim că statistica din România se poate lăuda cu o revistă de statistică, întemeiată încă din 1860, anume „Analele Statistice", înființată de Dionisie Pop Marțian; primul număr se deschide cu un articol al acestuia, unde întîlnim — printre altele — afirmația: „statistica este pentru guvern ceea ce sînt pentru om cele cinci simțuri". În revistă se publicau date statistice și studii de economie. După moartea lui Dionisie Pop Marțian, „Analele Statistice" apar ca publicații separate; în 1869, revista și-a încetat activitatea. În 1918 a fost reluată apariția „Analelor economice și statistice", dar publicația nu mai era legată de activitatea Organului Central de Statistică. Încetarea apariției ei are loc în 1947.

5. *Învățămîntul statistic*

La difuzarea cunoștințelor de statistică și la progresul acesteia ca știință, un aport important se datorează învățămîntului statistic. După studiile de istorie a statisticii [15] și [13], primul curs s-a predat la Academia Mihăileană de la Iași, înființată în 1835 de către George Săulescu, sub denumirea de „Dreptul și statistica" (1837). Reforma învățămîntului, aprobată de Mihai Gh. Sturza (1846) prevede pentru învățămîntul mediu „geografia nouă și statistica în toată a lor întindere în general și a Principatelor in special". Sînt semnalate unele cursuri de statistică în gimnazii. Cert este că odată cu organizarea unei facultăți juridice la Iași, se propune un curs de statistică. El este încredințat lui Alexandru Papiu Ilarian pe care acesta îl deschide la 24 februarie 1856. Manuscrisul cursului se află la Academie [19]. Cursul este de „statistică și drept penal". În Muntenia, un curs liber de statistică este ținut de Dionisie Pop Marțian. Se mai menționează că la Seminarul teologic „Veniamin Costake" de la Iași, se preda un curs de „istorie statistică laică și bisericească". La Colegiul Național din Iași se preda un curs de istorie, geografie și statistică, încredințat în 1863, pentru scurtă vreme, lui B. P. Hașdeu.

Statistica începe să devină obiect de predare la școlile comerciale. Un moment important în istoria statisticii ca disciplină de învățămînt este marcat de înființarea, în anul 1913, a Academiei de Înalte Studii Comerciale și Industriale. Se introduc un curs de statistică (în special economică) și unul de matematică financiară aplicată, teoria și tehnica asigurărilor. Cursul este încredințat lui

*Stanislas Cihoski**, care predase statistica și la Școala Superioară de Științe de Stat și publicase manualul „Statistica generală. Metodologie statistică" (1910). După lucrarea lui Ion Ionescu de la Brad (Povățuiri ..., 1859), acesta este cel de-al doilea manual sau curs de statistică tipărit în România. După moartea lui Stanislas Cihoski, cursul de statistică de la Academia de Înalte Studii Comerciale și Industriale a fost continuat de Ștefan Stănescu (1925). Cursuri de statistică se introduc la facultățile de drept, de obicei sub denumirea de „Finanțe și statistică". Profesori în perioada interbelică au fost Aristide Basilescu (București), Gh. Leon (Cluj), Ion Mircea Nonu (Iași), Sigmond Andrei (Oradea). La Academia de Înalte Studii Comerciale și Industriale din Cluj, înființată în 1919, statistica este predată de N. Ghiulea și D. D. Hașigan. Se mai semnalează un curs de statistică la Institutul de științe administrative din 1928, ținut de E. C. Decusară, care avea să publice mai tirziu : „Elemente de statistică. Partea I. Statistica teoretică și aplicată", București, 1943. În perioada interbelică, care a însemnat o puternică dezvoltare a statisticii și o amplă extindere a aplicațiilor în variate domenii, se constituie școala românească de teoria probabilităților și statistică matematică, înființată de acad. *Octav Onicescu** secondat de fostul său student, acad. *Gheorghe Mihoc** și continuată apoi de o strălucită pleiadă de matematicieni, probabiliști și statisticieni — pînă în zilele noastre. Primele cursuri regulate de teoria probabilităților acad O. Onicescu le inaugurează în 1924, la Facultatea de Științe a Universității din București. Cu prilejul Recensămîntului General din 1930, se înființează Comisia recensămîntului, prezidată de Dimitrie Gusti, cu care prilej se hotărăște întemeierea unei școli de statistică, condusă de O. Onicescu, secretar fiind Paul Sterian. Procesul verbal din 30 mai 1930 stabilea titularii cursurilor : *D. C. Georgescu**, *N. St. Georgescu-Roegen**, Gh. Mihoc, I. Moruzzi și M. Sanielevici, cei mai mulți matematicieni și actuari. Cursul de demografie a fost încredințat lui Leonida Colescu, de la care ne-au rămas două cursuri scrise (1930 și 1932). Din 1935, Școlii de Statistică i se alătură dr. Gh. Banu care va preda statistica demografică. Școala colabora strîns cu Institutul de Statistică, al cărui director a devenit dr. Sabin Manuila. În 1941 se înființează Institutul de Statistică, Actuariat și Calcul în cadrul Facultății de științe a Universității din București, director fiind acad. Octav Onicescu. În prezența rectorului Universității, acad. Horia Hulubei, are loc inaugurarea festivă a noului institut, în ziua de 24 noiembrie 1941. Potrivit statutului [11], materiile de învățămînt sînt următoarele : 1. Elemente de matematică pentru statisticieni (Prof. N. Ciorănescu) ; 2. Calculul probabilităților (Prof. O. Onicescu) ; 3. Statistica matematică și metodologia statistică (N. St. Georgescu-Roegen) ; 4. Biometrie și demografie (dr. Sabin Manuila) ; 5. Statistica economică (Paul Sterian și Roman Moldovan) ; 6. Statistică și realitate națională (A. Golopenția și Mircea Biji) ; 7. Tehnică și organizare statistică (dr. D. C. Georgescu) ; 8. Econometrie (N. St. Georgescu-Roegen) ; 9. Matematici actuariale (prof. Gh. Mihoc) ; 10. Matematici financiare (Șerban Gheorghiu) ; 11. Tehnica asigurărilor private și organizarea lor (Al. Pantazi și V. Protopopescu) ; 12. Tehnica asigurărilor sociale și organizarea lor (I. Argeșanu) ; 13. Principii și tehnica organizării și contabilității pentru secția de actuariat și statistică (Al. Pantazi) ; 14. Calculul grafic și numeric (prof. Gr. C. Moisil) ; 15. Mașini și organizarea birourilor de calcul (ing. C. Petrescu) și 16. Aplicațiile analizei în calcul (N. Teodorescu). Marea majoritate a cadrelor didactice erau absolvenți ai unor prestigioase universități din Franța, Germania, S.U.A. și Italia. Durata cursurilor era de un an ; cursanții erau licențiați. Regulamentul prevedea trei secții :

1. Secția de statistică; 2. Secția de actuariat și 3. Secția de calcul. Programele analitice erau la nivelul progreselor științelor respective. Institutul a funcționat pînă în 1947.

Din perioada amintită datează două lucrări care își păstrează importanța pînă astăzi: N. St. Georgescu-Roegen, „Metoda statistică. Elemente de statistică matematică", cu o prefață de O. Onicescu. Institutul de Statistică Generală a Statului, București, 1933, 507 p. și Gh. Mihoc, „Tratat de matematici actuariale", București, 1943, 315 p. + Anexe.

Să consemnăm că între cele două războaie mondiale statistica pătrunde în numeroase discipline, fiind folosită de remarcabili oameni de știință.

Aplicații ale calculului probabilităților și statisticii matematice au loc în actuariat. Matematicienii actuari N. Praporgescu și M. Sanielevici întocmesc prima tabelă de mortalitate a populației României pe baza recensămîntului populației din 1899. Alte tabele sînt întocmite de acad. Gh. Mihoc. Se cuvine amintit faptul că în matematica actuarială, de care a beneficiat mult demografia și asigurările, au lucrat: acad. O. Onicescu, acad. Gh. Mihoc, Alexandru Pantazi, Șerban Gheorghiu.

Una din contribuțiile cele mai importante la progresul statisticii, l-au adus Școala de medicină socială de la Cluj, ca și cea de la București, precum și cele trei institute de igienă și sănătate publică din București, Cluj și Iași, precum și Școala de antropologie de la București, legate de numele savanților medici Iuliu Moldovan, Gheorghe Banu și Francisc I. Rainer. *Iuliu Moldovan* (1882—1966), doctor docent al Facultății de Medicină din Viena, profesor la Universitatea din Cluj și membru al Academiei Române, fondatorul „Buletinului Eugenic și Biopolitic" (1927) și director al Institutului de Igienă și Sănătate Publică din Cluj, a promovat metodele de statistică în cercetările de medicină socială, antropologie și biologie. Conceptele și metodele statistico-matematice ale lui Fr. Galton și K. Pearson, ale lui R. Pearl și J. L. Reed sînt folosite pe scară largă. Lucrările sale „Igiena Națiunii" (1925) și „Introducere la etnobiologie și biopolitică" (1944), ca și cele ale elevilor și colaboratorilor săi sînt o dovadă în acest sens.

Gheorghe Banu (1889—1957), medic igienist, cu specializare în Franța unde a lucrat cu prof. V. Levaditi, profesor la Facultatea de Medicină din București, întemeietorul „Revistei de Igienă Socială" (1931), a dat de asemenea un puternic impuls statisticii și demografiei. În remarcabile studii precum: „Probleme sanitare ale populației rurale din România", în monografia clasică „Sănătatea poporului român", Editura Fundațiilor, 1935, în „Medicina socială ca știință. Eugenia. Demografia" (5 vol. Casa Școalelor, 1944) și altele, statistica i-a servit să facă aprofundate cercetări de medicină și igienă socială. Cea mai importantă lucrare este, după afirmația dr. V. Săhleanu [35], „Biologia poporului român".

Francisc I. Rainer (1874—1944), profesor de anatomie la facultățile de medicină din Iași și București, întemeietorul Institutului de Antropologie din București (1940), a efectuat cercetări antropologice populaționale la solicitarea lui Dimitrie Gusti (din 1927), introducînd prelucrarea statistică riguroasă a materialelor de observație, după cum atestă „Enquêtes anthropologiques dans trois villages des Carpathes" (1937). Totodată, a inițiat cercetări biometrice și statistice în morfologia microscopică. Această orientare a fost continuată de acad. Ștefan M. Milcu (n. 1903).

Dintre discipolii și colaboratorii prof. Iuliu Moldovan trebuie amintit *Petru Râmneamțu**, al cărui manual „Elemente de biometrie medicală și statistică vitală" (1939) este primul de acest gen în literatura românească.

Un alt domeniu important în care statistica a fost larg aplicată este sociologia și mai ales Școala monografică de la București. *Dimitrie Gusti* (1880—1955) șeful școlii, acorda o atenție deosebită statisticii, considerînd-o drept metodă indispensabilă pentru studiile sociologice. În lucrările sale, ,,Problema sociologici. Sistem și metodă" (1940) și ,,Îndrumări pentru monografiile sociologice" (1940) rolul statisticii în cercetarea socială este bine reliefat. Discipolii și colaboratorii prof. D. Gusti erau cunoscători temeinici și au folosit statistica în cercetările lor, așa cum se observă din opera prof. H. H. Stahl (n. 1901), *A. Golopenția** și Traian Herseni (1907—1980).

Continuînd tradiția din sec. al XIX-lea, economiștii dintre cele două războaie au făcut statisticii un loc important în studiile lor. *Virgil Madgearu* (1887—1940), profesor de economie politică la Academia de Înalte Studii Comerciale și Industriale, a dat exemple remarcabile de folosire a statisticii în lucrările sale: ,,Curs de economie politică" (1938) și mai ales în ,,Evoluția economiei românești după războiul mondial". Victor Slăvescu (1891—1977), profesor la Academia de Înalte Studii Comerciale și Industriale, membru al Academiei Române, a reactualizat figurile unor economiști și statisticieni ca N. Șuțu, Dionisie Pop Marțian.

Pe linia inaugurată de A. D. Xenopol, N. Iorga și V. Pârvan, care au realizat și importante studii de demografie istorică prin aplicarea metodelor statistice, istoricii din perioada interbelică au contribuit sensibil la lărgirea cadrului de aplicare a statisticii. Se cuvin amintiți în acest sens *Alexandru Lapedatu* (1875—1954), fost președinte al Academiei Române, medievist, *Ion Lupaș* (1880—1967), profesor de istoria nouă a Românilor și de istoria Transilvaniei, *Silviu Dragomir* (1888—1962), ale căror studii privind istoria românilor fac apel la metodele statisticii. Demografia istorică și istoria populației au fost ilustrate de *Ștefan Meteș* (1887—1977), cu remcarcabile studii (,,Emigrări românești din Transilvania în secolele XIII—XX, București, 1971) și de *David Prodan* (n. 1902), academician, profesor la Universitatea din Cluj, în special cu studiile privind imigrația, iobăgia, structura socială (,,Teoria imigrației românilor din Principatele Române în Transilvania în veacul al XVII-lea", Sibiu, 1944, și ,,Urbariile Țării Făgărașului", vol. I, 1970).

Contribuții importante la aplicarea metodelor statistice au și psihologii și pedagogii. *Florian Ștefănescu-Goangă* (1881—1958), profesor la Universitatea din Cluj, a înființat Institutul de psihologie experimentală, aplicînd metode statistico-matematice. Elevul său, *Nicolae Mărgineanu* (1905—1980), cu studii de specializare în Germania, Franța și S.U.A., a contribuit la introducerea metodelor statistico-matematice în psihologia și psihotehnica românească (,,Elemente de psihometrie", 1937, ,,Analiza factorilor psihici", 1938, ,,Psihologia persoanei", 1938, retipărită sub titlul ,,Condiția umană").

Un loc deosebit îl ocupă *Gheorghe Zapan* (1897—1976), psiholog și psihotehnician, doctor în matematici la Universitatea din Berlin, profesor universitar. A lărgit domeniul de aplicare a metodelor statistico-matematice în psihologia experimentală, psihotehnică și pedagogie. I se datorează conceptul de entropie taxiologică și o originală concepție despre cibernetica învățării și a activităților umane. (,,Aplicarea teoriei lanțurilor cu legături complete în procesul învățării". În ,,Revista de Pedagogie", nr. 7, 1971; ,,Cibernetica pedagogică", în ,,Revista de Pedagogie", nr. 7—8, 1969; ,,Modelarea matematică a activităților umane și a proceselor evolutive", în ,,Revista de Pedagogie" nr. 5, 1966).

Se poate face încheierea că între cele două războaie mondiale, atît sub raport teoretic și metodologic, cît și al aplicațiilor sale, statistica din România se situa la un nivel comparabil cu cel al țărilor dezvoltate.

6. *Statistica românească contemporană*

În perioada 1944—1984, în condițiile profundelor schimbări politice, sociale și economice, statistica a cunoscut o dezvoltare deosebită, atît pe plan instituțional, cît și sub raport teoretic și metodologic și al cadrului de aplicare. A fost creată o puternică instituție centrală a statisticii statului, cu atribuții și sarcini importante. Din 1950 denumirea Institutului Central de Statistică se schimbă [1], pentru a trece apoi din 1951 la denumirea actuală de Direcția Centrală de Statistică. În 1970, instituția trece în subordonarea Consiliului de Stat [3]. Actualul act normativ care prevede organizarea și funcționarea statisticii este Decretul Nr. 152 din 24 martie 1970 (Bul. Oficial nr. 34 din 14 aprilie 1970) și care capătă putere de lege în același an. Direcția Centrală de Statistică este „organ central al administrației de stat, înfăptuiește politica partidului și statului în domeniul sistemului informațional statistic" (art. 1). Funcția ei de bază se definește astfel: „Direcția Centrală de Statistică asigură informarea conducerii de partid și de stat cu privire la îndeplinirea planului de stat, dezvoltarea vieții social-economice, ritmurile și proporțiile din economie, folosirea mijloacelor materiale și resursele de muncă, în vederea luării de decizii pentru conducerea economiei naționale și a fundamentării programelor ei de dezvoltare" (art. 2).

Atribuțiile Direcției Centrale de Statistică sînt foarte largi: organizează sistemul informațional statistic de stat și departamental, elaborează sistemul dărilor de seamă statistice de stat și coordonează metodologic dările de seamă statistice departamentale, efectuează recensăminte și alte lucrări statistice importante, asigură și păstrează fondul național de date statistice. Direcția Centrală de Statistică dispune de o rețea de organe teritoriale — direcțiile județene de statistică, inclusiv direcția de statistică a Municipiului București —, are în componența sa un Centru electronic de calcul, dotat cu calculatoare și alte mașini moderne. Direcția Centrală de Statistică publică „Anuarul statistic al Republicii Socialiste România", diferite anuare pe domenii — cum este, de pildă, Anuarul demografic —, rezultatele recensămintelor populației și altele. De asemenea, Direcția Centrală de Statistică editează, din 1952, „Revista de Statistică*".

Activitatea științifică și de cercetare se efectuează în Consiliul științific-metodologic, înființat în 1954. O importanță deosebită pentru dezvoltarea statisticii ca știință au consfătuirile științifice de statistică organizate de Direcția Centrală de Statistică: 1961, 1962, 1963, 1964, 1965, 1968, 1969, 1977, 1980, 1984 ale căror lucrări sînt publicate în volume.

Caracteristic pentru sistemul informațional statistic actual este folosirea cu precădere a sistemului dărilor de seamă statistice de stat, intrat în vigoare la 1 ianuarie 1949 [38]. Rolul recensămintelor a fost considerabil redus, excepție făcînd recensămîntul populației. Dintre anchetele statistice amintim cercetarea bugetelor de familie — anchetă prin sondaj, inițiată în 1948 — și care asigură informații multilaterale privind nivelul de trai al populației. De cea mai mare importanță este crearea Centrului electronic al Direcției Centrale de Statistică care asigură prelucrarea automată a datelor. Direcția Centrală de Statistică participă la activitatea internațională statistică: la Comisia de statistică O.N.U.,

la Conferința Statisticienilor Europeni, F.A.O., U.N.E.S.C.O., O.M.S.; de asemenea, este membră a Comisiei de Statistică a C.A.E.R.

Învățământul statistic pus pe baze noi, în 1948, s-a dezvoltat și s-a diversificat. Au fost create catedre de statistică la Academia de Studii Economice; discipline ca statistica matematică, statistica teoretică, statistica economică, statistica industrială, agricolă și altele au intrat în planurile permanente de învățământ în învățământul economic — A.S.E. și facultățile economice ale universităților din țară. Cursuri de statistică — într-o formă sau alta — se predau și în învățământul universitar, medical, politehnic, agricol.

Cercetarea științifică în domeniul statisticii a înregistrat o extensiune și o diversificare remarcabile, împreună cu lărgirea cadrului de aplicare a metodelor statistice și matematice.

Școala românească de teoria probabilităților, fondată de academicienii Octav Onicescu și Gheorghe Mihoc în deceniul al patrulea, a cunoscut o impresionantă dezvoltare în ultimele trei decenii. Cea mai valoroasă realizare a fondatorilor școlii românești de teoria probabilităților a fost introducerea, în 1935, a unui nou tip de dependență stohastică, „dependența cu legături complete" care generalizează extensiv dependența markoviană. Tot ei au adus contribuții remarcabile la teoria lanțurilor markoviene. Contribuțiile recente se referă la funcțiile-sumă, procesele Markov cu parametru continuu și sistemele aleatoare cu legături complete.

În această activitate, un rol important îl au : catedra de teoria probabilităților de la Facultatea de Matematică a Universității București și Centrul de Statistică Matematică. Printre realizările importante ale ultimei instituții se numără organizarea ultimelor cinci ediții ale Conferinței internaționale de teoria probabilităților de la Brașov (1968, 1971, 1974, 1979, 1982). Bibliografia privind teoria probabilităților și statistica matematică numără cîteva zeci de tratate, manuale și monografii și sute de studii, comunicări și memorii, publicate în țară și în străinătate. Printre domeniile în care statistica matematică a fost aplicată pe scară largă sînt : agrobiologia, unde trebuie consemnată remarcabila activitate a acad. Nichifor Ceapoiu (n. 1911), fiabilitatea și controlul statistic al calității, ilustrate de numeroși cercetători, biologia, medicina etc.

Statistica teoretică și economică a urmat aceeași evoluție ascendentă. Generațiilor mai vechi (prof. R. Moldovan (n. 1911) ; prof. M. Biji (n. 1913)) li se alătură cohorte numeroase de statisticieni — cadre didactice și cercetători — care au îmbogățit substanțial statistica. Statistica populației și demografia sînt și ele beneficiare ale progresului general.

Reviste cum sînt : „Revista de Statistică", editată de Direcția Centrală de Statistică, „Revue Roumaine de Mathématiques Pures et Appliquées" și „Studii și Cercetări Matematice", editate de Academia Republicii Socialiste România, „Cercetări și studii de calcul economic și cibernetică economică", editată la Academia de Studii Economice, contribuie eficient la progresul statisticii românești.

În condițiile în care asistăm la o lărgire considerabilă a cadrului de aplicare a metodelor matematice, la apariția de noi discipline și la afirmarea principiului inter- și multidisciplinarității în știință, statistica continuă să dețină un loc proeminent în peisajul științei românești.

BIBLIOGRAFIE

1. Decretul nr. 246/1950 pentru reorganizarea statisticii în Republica Populară Română.
2. H.C.M. nr. 1225/1957 privind organizarea Direcției Centrale de Statistică.
3. Decretul nr. 152/1970 privind organizarea și funcționarea Direcției Centrale de Statistică.
4. Decretul nr. 414/1970 privind trecerea Direcției Centrale de Statistică ca organ al Consiliului de Stat al Republicii Socialiste România.
5. *Recensământul general al populației României. Rezultatele definitive (1899). Precedate de introducțiune cu explicațiuni* de L. Colescu, vol. I și II, București, 1905.
6. *Statistica demografică a României. Populația Regatului Român după recensământul făcut la 19 decembrie 1912. (I. Scărlătescu)*, București, 1921.
7. *Recensământul general al populației României din 29 decembrie 1930*, 9 vol., 1938—1940.
8. *Recensământul populației din 21 februarie 1956*, Direcția Centrală de Statistică, Rezultate generale (1959),vol. I (1960), vol. II (1961), vol. III (1961).
9. *Recensământul populației și locuințelor din 15 martie 1966*, Direcția Centrală de Statistică, vol. I, Rezultate generale, partea întîi, Populație, 1969; vol. II—VI, vol. VII, partea I și II; vol. VIII (1970).
10. *Recensământul populației și al locuințelor din 5 ianuarie 1977*, vol. I, Populație — Structura demografică; vol. II. Populație — Structura social-economică, Direcția Centrală de Statistică, 1980.
11. *Institutul de Statistică, Actuariat și Calcul, Universitatea din București, Facultatea de Științe, Statut, Programe*, București, 1942.
12. *Dicționar statistic — economic*, Direcția Centrală de Statistică, București, 1969 (cap. XXVI. Istoria statisticii).
13. *Din istoria statisticii românești*, culegere de articole, Direcția Centrală de Statistică, 1969 (Studiile semnate de : Al. Bărbat, M. Biji, R. C.Demetrescu, D. Hașigan, C. Ionescu, V. Lupu, I. Measnicov, Al. Șesan, L. Moldovan, G. Retegan, Gh. Ungureanu, I. Veverca).
14. *Handwörterbuch der Staatswissenschaften*, Ed. a III-a, Dritter Band, Verlag von Gustav Fischer, Jena, 1909, p. 358.
15. BULGARU, Valeriu, *Contribuții la istoria învățământului statistic din România I, II și III*, în „Revista de Statistică" nr. 4, 5 și 6, 1969.
16. BULGARU, Valeriu, *Dimitrie Cantemir, reprezentant de seamă al statisticii descriptive*, în „Revista de Statistică" nr. 11, 1973.
17. CSILLAG, A., *O perioadă de reînnoire a statisticii românești. Activitatea lui L. Colescu (I) și (II)*, în „Revista de Statistică" nr. 11 și 12, 1970.
18. CSILLAG, A., *Evoluția statisticii de stat oglindită în Buletinul statistic al României*, în „Revista de Statistică" nr. 6 și 7, 1971.
19. DEMETRESCU, R. C., *Un curs de statistică din anul 1856*, în „Revista de Statistică" nr. 1, 1970.
20. DEMETRESCU, R. C., *Prezențe românești la reuniunile internaționale de statistică* (I) și (II), în „Revista de Statistică" nr. 4, 6 și 12, 1975.
21. DOBRE, Gh., *Reorganizarea administrativă a statisticii de stat în România în anii 1866—1871*, în „Revista de Statistică" nr. 12, 1969.
22. DONAT, I., *Înființarea Oficiului Central de Statistică Administrativă al Țării Românești (1859)*, în „Analele Institutului Statistic al României", vol. I, 1942, p. 303.

23. IORDACHE, V., *Cadrul instituțional al statisticii în România* (I) și (II), în „Revista de Statistică" nr. 3 și 4, 1969.
24. MARȚIAN, Dionisie POP, *Recensiunea din 1860*, în „Analele statistice" an I, nr. 2, 186, p. 39—109.
25. MIHOC, Gh., *Dezvoltarea statisticii matematice în România*, în „Revista de Statistică" nr. 5—6, 1960.
26. MIHOC, Gh., *Probability, mathematical statistics, operational research*, în *Noesis*, Travaux du Comité Roumain d'Histoire et de Philosophie des Sciences I, Editions de l'Académie de la République Socialiste de Roumanie, Bucarest., 1973, p. 323—329.
27. MOLDOVAN, L., *Conscripții și recensăminte în Transilvania (sec. XVII—XIX). Metodologie și valoare documentară*, în „Revista de Statistică" nr. 8, 1970.
28. NEGRU, I., *Despre valoarea primei monografii medicale a Munteniei (1830) și autorul ei Doctorul Caracaș*, în „Munca Sanitară" nr. 8, 1966, p. 504.
29. NEGRU, I., *Doctorul Constantin Caracaș (1773—1828) despre cauzele depopulării Munteniei la începutul secolului al XIX-lea* în „Munca Sanitară" nr. 10, 1967, p. 624.
30. NEGRUZZI, C., *Lucrări statistice făcute în anii 1859—1860*, Iași, 1861, 335 p.
31. RETEGAN, G., *De la census-ul roman la recensămintele moderne ale populației* în „Viața economică" nr. 2 (126), 4 (128), și 6 (130), 1966.
32. RETEGAN, G., *Dezvoltarea demografiei românești în perioada 1859—1916*, în „Studii de statistică", Lucrările celei de-a șaptea consfătuiri științifice de statistică 11—14 iunie 1969, vol. II, Direcția Centrală de Statistică, București, 1972, p. 2040—2056.
33. RETEGAN, G., *Fragmente mai semnificative dintr-o istorie a instituției centrale de statistică*, în „Revista de Statistică" nr. 6, 1980.
34. SĂHLEANU, V., *Evoluția gîndirii statistice*, în „Revista de Filozofie" nr. 5, 1979, p. 569—574.
35. SĂHLEANU, V., *Începuturile medicinii sociale în România: G. Banu*, Apimondia, București, 1979.
36. STOICA, Gh., *Începuturile statisticii oficiale românești*, în „Revista de Statistică" nr. 1, 1969.
37. TREBICI, V., *Zece ani de muncă științifică în statistică (1947—1957)*, în „Revista de Statistică" nr. 12, 1957.
38. TREBICI, V., *Zece ani de la introducerea sistemului dărilor de seamă statistice în România*, în „Revista de Statistică" nr. 1, 1959.
39. TREBICI, V., *Mică enciclopedie de demografie*, Ed. științifică și enciclopedică, București, 1975.

XIX. PERSONALITĂȚI

ACHENWALL, Gottfried (1719—1772). Statistician german, reprezentant de frunte al statisticii descriptive, profesor la Universitatea din Göttingen. A înlocuit denumirea disciplinei de „teoria statului" cu cea de statistică. Esența concepției sale a fost formulată astfel: statistica „este cunoașterea aprofundată a situației (status) respective și comparative a fiecărui stat", „știința care indică tot ce se găsește efectiv într-o colectivitate politică, într-o țară, într-un loc oarecare". În istoria statisticii este uneori numit „tatăl statisticii", avîndu-se în vedere în special faptul că el a pus în circulație expresia de „statistică".

Op. : „Staatsverfassung der heutigen vornehmsten Reiche", 2 Aufl., Göttingen, 1752 (apărută pentru prima dată în 1749 sub titlul „Abriss der Staatswissenschaft der europäischen Reiche").

ANGELESCU, Ion Nicolae (1884—1930). Economist și om politic român. Doctor în științe economice al Universității din München. Profesor și rector al Academiei de Înalte Studii Comerciale și Industriale din București (1924—1929). Fondator (1927) și președinte al Asociației Economiștilor din România, fondator (1918) și coeditor al revistei „Analele statistice și economice".
A folosit datele statistice și metodele statistice în analiza proceselor economice și demografice. Studiul său asupra mișcării naturale a populației în anii 1904—1905 (publicat în 1915), precedat de o privire retrospectivă pe anii 1860—1913, este printre primele încercări de analiză a fenomenelor demografice în raport cu factorii economici.

Op. : „Histoire économique des Roumains", Tome 1, Geneva (f. an); „Avuția națională a României", București, 1915; „Evoluția economică a țărilor românești", București, 1918.

BAYES, Thomas (1702—1761). Pastor englez, devenit celebru postum prin două studii de probabilități (memorii) care au creat domeniul probabilităților subiective (→ XII *Formula lui Bayes)* și așa-numita „inferență bayesiană". În primul memoriu a examinat seriile Stirling (de Moivre) în (?!); cel de-al doilea, care i-a adus celebritatea, se intitulează "An Essay Towards Solving a Problem in the Doctrine of Chances" (prin "chance" **B.** înțelege probabilități). Ideile sale au fost dezvoltate de alți probabiliști și statisticieni; contribuția sa a fost apreciată de Boole, J. M. Keynes, de R. A. Fisher și alții. Inferența bayesiană sau statistica bayesiană este un demers al *inferenței statistice**, care se bazează pe teoria probabilităților subiective. O dezvoltare importantă a acestei teorii este legată de teoria deciziei și aplicațiile ei în economie, marketing etc.

Op. : Cele două memorii publicate postum (1764) în "Philosophical Transactions of the Royal Society of London", reeditate în 1963 : "Facsimiles of Two Papers by Bayes".

Despre **B.** : M.C. Demetrescu, „Analiza bayesiană", în „Revista de Statistică" nr. 2, 1970.

BERNOULLI, Jakob (Jacques) (1654—1705). Matematician elvețian, unul din clasicii teoriei probabilităților, dintr-o celebră familie de matematicieni. Profesor de matematică (1687) la Universitatea din Basel, rector al acesteia; membru al Academiei de științe din Paris (1699). A fost influențat de Leibniz și calculul infinitesimal. A pus bazele calculului variațional, împreună cu fratele său mai mic I. Bernoulli. Autor al celebrei teoreme a legii numerelor mari. Lucrarea sa principală „Ars conjectandi", apărută postum (1713), cuprinde patru părți : în partea întîi comentează tratatul lui Huygens, „Despre calcule în jocurile de noroc" (apărută în 1657); partea a doua expune teoria combinărilor și permutărilor; a treia — serii de jocuri, a patra este consacrată diferitelor probleme legate de calculul probabilităților — legea numerelor mari, repartiția binomială. În anexă se ocupă de „jeu de paume" (un fel de joc de tenis).

Op.: „Ars conjectandi, opus posthumum: acced. de seriebus infinitis etc.", Basel, 1713. Prima ediție a operelor complete a apărut în 2 vol., la Geneva (1744).

BERNȘTEIN, Serghei Natanovici (1880—1968). Matematician probabilist sovietic. Studii universitare: Sorbona (1898—1902) și la Göttingen (1902—1903); doctorat în matematici la Universitatea din Paris, avîndu-l în comisie pe H. Poincaré (1904). Membru al Academiei de Științe a U.R.S.S. (din 1929), profesor la Universitatea din Moscova; în anii 1932—1968 a lucrat la Institutul de Matematică „V. A. Steklov" din Moscova. Laureat al Premiului de Stat. Domenii: teoria ecuațiilor diferențiale, teoria funcțiilor, teoria probabilităților. Încercări de a sintetiza școala franceză a probabilităților cu școala rusă (Cebîșev, Leapunov, Markov).
A dezvoltat formula lui Laplace cu privire la probabilitatea binomială, a perfecționat formula lui Cebîșev („Inegalitatea lui Cebîșev"), a generalizat inegalitatea lui Kolmogorov. S-a ocupat de aplicații ale teoriei probabilităților în statistică și biologie.

Op. „Sobranie socinenia", vol. I (1952), vol. II (1954), vol. III (1960) și vol. IV (1964), Editura „Nauka", Moscova.

BOREL, Emil (1871—1956). Probabilist francez. Absolvent al École Normale Supérieure, doctor în matematici (1893); profesor la École Normale Supérieure, la Facultatea de Științe a Universității din Paris. Fost președinte al Societății de Matematică din Franța, al Societății de Statistică din Paris; membru al Academiei de Științe din Franța, membru de onoare al Institutului Internațional de Statistică. Domenii: teoria mulțimilor, teoria funcțiilor, geometrie, istoria și filozofia științei, psihologie și economie politică, mecanica statistică; primul care a formulat „teoria jocurilor psihologice", cu 7 ani înaintea lui J. v. Neumann. Contribuții fundamentale la teoria probabilităților, de care a început să se ocupe din 1915; a dat un tratat clasic de teoria probabilităților și aplicații. Circa 300 lucrări științifice. Printre elevii săi: *M. Fréchet**.

Op: „Leçons sur les fonctions de variables réelles et les développements en séries de polynomes", Gauthier—Villars, Paris, 1905 (numeroase reeditări); „Elements de la théorie des probabilités", Hermann, Paris, 1909 (mai multe ediții); „Traité du calcul des probabilités et de ses applications", Gauthier—Villars, Paris, 4 volume, 1925, 1926, 1934, 1939.
Despre **B.**: M. Fréchet, „Emile Borel. Viața și opera" (prefața, acad. O. Onicescu), Editura științifică, București, 1972.

BORTKIEWICZ, Ladislaus von (1868—1931). Statistician, matematician și economist german, născut în Rusia. Studii la Göttingen, cu *W. Lexis**. Profesor la Universitatea din Berlin. Contribuții în teoria probabilităților, actuariat, statistică matematică, demografie și statistică economică. A redescoperit legea numerelor rare (S. D. Poisson); a propus o nouă interpretare a dispersiei, în studiul indicilor a propus corelația. Membru al Institutului Internațional de Statistică (1913), al Academiei Suedeze de Științe.

Op.: „Das Gesetz der kleinen Zahlen", Leipzig, 1898: „Die Iterationen. Ein Beitrag zur Wahrscheinlichkeitstheorie", Springer, Berlin, 1907; „Zweck und Struktur einer Preisindexzahl", în „Nordisk Statistik Tidskrift", B. 2, 1923; B. 3, 1924; B. 4, 1925).

BOWLEY, sir Arthur Lyon (1869—1957). Statistician, matematician, economist și demograf englez. Studii la Cambridge, prieten cu Alfred Marshall a cărui influență a resimțit-o. Profesor la London School of Economics (din 1895), unde a predat cursuri de statistică. Membru al lui Royal Society (1922); director al Institutului de Statistică ale Universității din Oxford. Studii privind dinamica salariilor și prețurilor, contribuții la teoria sondajului, participînd la faimoasa dispută în legătură cu sondajul propus de statisticianul norvegian Anders N. Kiaer (1838—1919); a aplicat sondajul la studiul clasei muncitoare din Anglia; estimații asupra venitului național; studii de conjunctură pentru London and Cambridge Economic Service; studii de demografie. A făcut parte din grupul care a prezentat, în 1924, Institutului Internațional de Statistică "Report on the Representative Method in Statistics", care a dat cîștig de cauză metodei sondajului. Autor al unor manuale de statistică de largă circulație.

Op.: "Elements of Statistics", 1901; "An Elementary Manual of Statistics", 1910; "The Mathematical Groundwork of Economics: an Introduction Treatise", 1924; "Studies in the National Income: 1924—1928", în „Economic Journal".

BURKHARDT, Felix (1888—1973). Matematician, statistician și demograf din R. D. Germană. Studii de matematici și statistică la universitățile din Leipzig și Frankfurt/Main. Profesor de matematici și statistică la Universitatea Tehnică din Dresda. Director al statisticii populației la Oficiul de Statistică al Saxoniei. Membru al Institutului Internațional de Statistică.
A publicat peste 250 lucrări. Pînă în 1945 a editat „Deutsches Statistisches Zentralblatt", coeditor la „Archiv für Matematische Wirtschafts—und Sozialforschung"; coeditor la „Biometrische Zeitschrift" și la „Metrika". Internationale Zeitschrift für theoretische und angewandte Statistik".

CARACAȘ, Constantin (1773—1828). Medic și demograf român. Studii la Facultatea de Medicină din Viena. Medic șef al orașului București; a înființat spitalul „Filantropia" (1815); s-a preocupat de introducerea măsurilor de asistență sanitară și igienă publică. În 1800, la numai doi ani de la descoperirea vaccinului lui Ed. Jenner, introduce vaccinarea antivariolică. A întocmit prima monografie cu caracter medical și demografic din țara noastră, sub denumirea de „Topografia Țării Românești și observații antropologice privitoare la sănătatea și bolile locuitorilor ei" (1830), în limba greacă, în care, folosind datele catagrafiei fiscale a lui Caragea Vodă (1814—1815), face o descriere amănunțită a Țării Românești, a celor 17 județe, analizează cauzele depopulării populației, mortalitatea generală și infantilă, pe clase sociale, prezintă date asupra numărului populației pe județe și pe orașe etc. Monografia s-a bucurat de aprecierea lui N. Iorga (1919).

Op.: „Topografia Țării Românești și observații antropologice privitoare la sănătatea și bolile locuitorilor ei" (1830), tradusă de G. Sion și republicată de P. P. Samarian: „O demografie sanitară a Munteniei de Constantin Caracaș". Tip. Bucovina, București, 1937.

CEBÎȘEV, Pafnuti Lvovici (1821—1894). Matematician rus, profesor la Universitatea din Petersburg, membru al Academiei de Științe din Petersburg și al altor academii. Este fondatorul școlii matematice din Petersburg, care a dat personalități ca *A. A. Markov** și *A. M. Leapunov**. Contribuții remarcabile

la teoria probabilităților, la teoria aproximării funcțiilor prin polinoame. Autor al inegalității care îi poartă numele; a demonstrat teorema centrală în teoria probabilităților, prin metoda momentelor, construind o teorie generală a polinoamelor ortogonale, dezvoltate și de *V. S. Nemcinov**. A dat forma cea mai generală legii numerelor mari. Printre primii a aplicat matematica în lucrări practice. A exercitat o puternică influență asupra evoluției teoriei probabilităților.

Op.: „Elementarnoe dokazatelstvo odnogo obșcego predlojenia v teorii veroiatnostei", 1846; „O dvuh teoremah otnositelno veroiatnostei", 1877. Toate studiile au fost grupate în „Polnoie sobranie socinenii", vol. 1 — 5, Moscova — Leningrad, 1944 — 1951.

CIHOSKI, Stanislas (1868 — 1924). Economist și statistician român. Profesor la Academia de Înalte Studii Comerciale și Industriale, unde a predat statistica. Rector al Academiei (1918 — 1924). Cursul său „Statistica generală" (1910) este primul curs modern de statistică în învățământul superior din România.

Op.: „Statistica generală. Metodologie statistică", Atelierele grafice SOCEC București, 1910, VII + 192 p.

CIUPROV, Aleksandr Aleksandrovici (1874 — 1926). Statistician, matematician și economist rus. Profesor de statistică la Secția de economie a Institutului Politehnic din Petersburg. S-a ocupat în special de teoria corelației, căreia i-a adus importante contribuții ca și teoriei stabilității valorilor statistice tipice. A încercat o sinteză a principalelor curente ale statisticii timpului său: a școlii germane (W. Lexis și L. v. Bortkiewicz) și a statisticii engleze (Fr. Galton și K. Pearson).

Op.: „Ocerki po teorii statistiki", 1909; „Osnovniie problemî teorii korelații", 1926.

COCHRAN, William Gemmell (1909 — 1980). Matematician, statistician american de origine engleză. A studiat matematica la Universitatea din Glasgow și statistica la Cambridge, cu Wishart; a lucrat cinci ani la Stațiunea Experimentală din Rothamsted. În S.U.A., a fost profesor și cercetător la diferite universități: Iowa (cu G. Cox), la Princeton, la North Carolina, la Johns Hopkins (Biostatistics Department in the School of Hygiene and Public Health), iar între anii 1949 — 1957, la Universitatea din Harvard. Unul din cei mai reputați specialiști în tehnica sondajului și planificarea experimentelor, îmbinînd strălucit teoria cu practica. Președinte al lui Institute of Mathematical Statistics; președinte al lui American Statistical Association, al Institutului Internațional de Statistică (1967 — 1971); membru al lui National Academy of Sciences (1974). A lăsat circa 100 lucrări științifice cu referire la analiza dispersională, analiza regresiei, hi-pătrat, sondaj, planificarea experimentelor, teoria statisticii.

Op.: "The use of the analysis of various in enumeration by sampling", 1939; "Experimental Design", 1950 (în colab. cu G. M. Cox); "Sampling Techniques", 1953; coautor cu G. W. Snedecor la lucrarea clasică: "Statistical Methods", 1967.

COLESCU, Leonida (1872—1949. Statistician, demograf și economist român. Studii la Universitățile din Paris și München; doctor în științe economice și financiare. Între anii 1899—1922, director general al Serviciului Statistic General al Statului; președinte al Consiliului Superior al Statisticii. Profesor la Școala de Statistică, înființată în 1930, unde a predat demografia, elaborînd primul curs de demografie în literatura română. A organizat și a condus recensămintele populației din 1899 și 1912, Ancheta industrială 1901—1903; a editat Buletinul statistic al României 1899—1922 și a înființat „Bibliografia economică română", 1910. Membru al Institutului Internațional de Statistică (1910); a participat la sesiunile acestuia cu comunicări. Sub conducerea sa, instituția centrală de statistică a statului a cunoscut o puternică înflorire, fiind considerat cel de-al doilea etitor al statisticii după Dionisie Pop Marțian. A publicat remarcabile studii statistice și demografice, în special pe baza celor două recensăminte.

Op.: „Recensămîntul general al populațiunei României. Rezultate definitive (1899). Precedate de o introducțiune cu explicațiuni de L. Colescu". Institutul de arte grafice „Eminescu", vol. I și II, București, 1905 (reeditat de Institutul Central de Statistică, 1944): „Ancheta industrială a României". 2 vol., București, 1904; „Statistica știutorilor de carte din România întocmită pe baza rezultatelor definitive ale recensămîntului general al populației din 19 decembrie 1912" (reeditată de Institutul Central de Statistică, 1947); „La population de la Roumanie. Résumé démographique présenté à la session de l'Institut International de Statistique", Berlin, 1903; „Dicționarul statistic al României", 2 vol., București, 1914; „Recensămîntul populației" — tratat practic întocmit după cursul ținut în 1930 la Școala de Statistică, București, 1930; „Curs de demografie" după notițele și prelegerile prof. L. Colescu, text de M. Maxim, București, 1932 (litografiat).

CONRING, Hermann (1606—1681). Statistician german, unul dintre precursorii statisticii descriptive. Profesor la Universitatea din Helmstedt. În concepția sa, descrierea statului cuprindea patru părți: *causa materialis* (poporul, pămîntul, producția); *causa finalis* (scopul statului și mijloacele pentru a-l atinge); *causa formalis* (forma de stat și de guvernămînt); *causa efficiens* (puterea politică, suveran, guvern, armată, administrație).
Influența lui Conring a fost importantă la constituirea statisticii descriptive și a continuat pînă la începutul sec. al XIX-lea, prin reprezentanți ca L. v. Schlözer și alții.
Operele sale au fost publicate de Gabel în 1730 la Braunschweig (7 vol.)

CREANGĂ, Gheorghe D. (1874—1940). Economist și statistician român. Doctor în științe economice și financiare al Universității din Berlin. Profesor de finanțe și statistică la Universitatea din Iași (1912—1920). A folosit statistica în analiza economică, în special în ce privește proprietatea rurală și țărănimea, pledînd pentru împroprietărirea țăranilor.

Op.: „Proprietatea rurală și chestiunea țărănească", București, 1905: „Proprietatea rurală în România", București, 1907.

CRUPENSKI, Constantin E. (1860—1935). Statistician și economist român. Conducător al instituției centrale de statistică din România (din 1892), a marcat începutul unei perioade de înflorire a statisticii. Membru al Institutului Internațional de Statistică (din 1895), a participat la sesiunile acestuia cu comunicări științifice.

DARMOIS, Georges (1888—1960). Matematician și statistician francez, elev al lui Emile Borel. Profesor la Institutul de Statistică al Universității din Paris și director al acestuia din 1946, după retragerea lui M. Huber. Membru al Institutului Franței, președinte al Institutului Internațional de Statistică. Contribuții la analiza factorială propusă de *C. Spearman**; a inventariat exemple paradoxale de legături stohastice.

Op.: „Statistique mathématique", Paris, 1928; „Statistique et applications", Paris, 1934; Studii în „Les Mathématiques et la Psychologie".

DIVISIA, François (1898—1964). Statistician, inginer și economist francez. Studii la École Polytechnique și la École Nationale des Ponts et Chaussées. Membru al Institutului Internațional de Statistică (din 1933); președinte al Société de Statistique de Paris (1939), membru la Academia Nazionale dei Lincei (Roma), din 1950. Profesor de economie politică și socială la École Polytechnique. A publicat studii de econometrie și statistică („Indicele lui Fr. Divisia").

Op.: „L'indice monétaire et la théorie de la monnaie", 1925 și 1926; „L'épargne et la richesse collective", 1928; „Traitement économetrique de la monnaie, l'intérêt, l'emploi", Librarie Dunod, 1962.

EDGEWORTH, Francis Ysidro (1845—1926). Statistician, matematician și economist englez. Profesor de economie politică la Oxford (1891) și la Londra. Membru al lui Royal Society; editor, împreună cu J. M. Keynes, al publicației "The Economic Journal" (1891—1926). A fost influențat de concepția bayesiană. Contribuții la teoria indicilor unde a propus un indice ce-i poartă numele, la aplicarea statisticii matematice în economia politică. Numeroase studii publicate în "Journal of the Royal Statistical Society".

Op.: "Mathematical Psychics. An Essay on the application of mathematics to the Moral Sciences", 1881; "Defense of Index-Numbers", în "The Economic Journal", vol. VI, 1896; "Methods of representing Statistics of Wages" (în colab. cu A. L. Bowley), în "Journal of the Royal Statistical Society", 1902.

ENGEL, Ernst (1821—1896). Statistician și inginer german. Studii la École des Mines (Paris); influențat de anchetele lui Frédéric Le Play. A condus biroul de statistică al Saxoniei (Dresda) și al Prusiei. A inițiat și studiat bugetele de familie, stabilind o regularitate numită astăzi „legea lui E" și care constă în aceea că pe măsura creșterii veniturilor scade proporția cheltuielilor pentru alimentație, altfel spus, elasticitatea veniturilor în raport cu cererea pentru alimente este subunitară. Această lege, ca și reprezentarea grafică („curba lui E", cum a fost numită mai tîrziu) a fost studiată și dezvoltată, printre alții, de J. M. Keynes. Este considerat unul din precursorii econometriei. A participat activ la congresele internaționale de statistică și a contribuit la perfecționarea metodelor de anchetă și la comparabilitatea internațională.

Op. : „Productions — und Consumationsverhältnisse des Königreiches Sachsen", inițial în limba germană (1857) și apoi în Bulletin de l'Institut International de Statistique, 9, Nr. 1, Supl. 1. (1895) ; „Das Rechnungsbuch der Hausfrau und seine Bedeutung im Wirtschaftsleben der Nation", Berlin, 1882. ; „Der Kostenwert des Menschen", Berlin, 1883.

FARR, William (1807—1887). Statistician, demograf și medic englez, studii de medicină ; a profesat medicina pînă în 1838 ; în 1839 a fost angajat la "Registrar General's Office" în calitate de statistician ("Compilar of Abstracts"), unde a lucrat pînă în1880. A înființat, în cadrul acestei instituții, departamentul de statistică pe care l-a condus. A avut o importantă contribuție la organizarea congreselor internaționale de statistică, a Institutului Internațional de Statistică și la dezvoltarea metodelor statistice în special cu aplicație la demografie și sănătatea publică. A perfecționat metodele de întocmire a tabelelor de mortalitate.

Op : "On the Construction of Life Tables, illustrated by a new Life Table of the Healthy Districts of England", Londra, 1859 ; "English reproduction table", Londra, 1880 : "Annual Reports of the Registrar General of Births, Deaths and Marriages in England" (sub redacția lui F) : "Vital statistics, a memorial volume of selections from the reports and writtings of William Farr", edited by Noel A Humphreys, Londra, 1885.

FELIX, Iacob (1832—1905). Medic, demograf și statistician român. Profesor la Universitatea din București : medic-șef al orașului ; decan al Facultății de Medicină. Membru al Academiei Române. A fost membru al diferitelor societăți de medicină, igienă și statistică ; a participat la congresele internaționale de igienă și demografie. A studiat aspectele demografice și sanitare ale populației României pe bază de date statistice.

Op. : „Despre alimentațiunea poporului român", 1861 ; „Tratatul de igienă și politică sanitară", 2 vol., 1870 și 1889. „Despre mișcarea populațiunii României", 1880 ; „Geografia medicală a României", 1897.

FERMAT, Pierre de (1601—1665). Matematician francez, precursor al teoriei probabilităților, împreună cu B. Pascal (1632—1662). A dezvoltat, înaintea lui Descartes, teoria coordonatelor. Contribuții la dezvoltarea geometriei, a fizicii. Considerațiile privind probabilitățile au fost generate de jocurile de noroc, ca „împărțirea mizei", probleme puse de cavalerul de Méré și care a solicitat părerea unor matematicieni, printre care B. Pascal.
Lucrările lui F. sînt cuprinse în „Varia opera mathematica" (1679).

FISHER, Irving (1867—1947). Statistician, matematician și economist american. Profesor de matematică, apoi de economie politică la Universitatea din Yale. A introdus matematica în economie ; unul din întemeietorii econometriei. Inventator în domeniul cartelelor perforate și fondator al Corporației Remington Rand, Inc., specializată în fabricarea mașinilor de calcul. Contribuții la teoria monedei și la teoria indicilor, unde a propus un indice ce-i poartă numele.

Op. : "Mathematical Investigations in the Theory of Value and Prices", 1892 ; "The Purchasing Power of Money", 1911 ; "The Best Form of the Index Number",

1921 : "The Making of Index Number. A Study of their Varieties, Tests, and Reliability", 1922.

FISHER, sir Ronald Aylmer (1890—1962) Matematician, statistician, genetician englez. Strălucit reprezentant al școlii anglo-saxone de statistică matematică, continuator al lui K. Pearson. A dezvoltat tehnica statistică și aplicațiile acesteia în biologie. Între anii 1919—1933 a lucrat la celebra "Rothamsted Experimental Station", institut de cercetări agricole unde s-au aplicat cele mai multe metode de statistică matematică. În 1933 i-a succedat lui K. Pearson ca profesor la Universitatea din Londra ; membru al lui Royal Society (1929) ; în 1947 a înființat The Biometric Society ; din 1957 pînă în 1962, președinte de onoare al Institutului Internațional de Statistică. În 1912 a propus metoda verosimilității maxime ; numeroase alte contribuții fundamentale la teoria sondajului, a estimației statistice, a planificării experimentale, a analizei dispersionale. A exercitat o influență profundă asupra evoluției statisticii matematice.

Op. : "Statistical Methods for Research Workers", 1925 ; "Theory of Statistical Estimation", 1925, care a cunoscut numeroase ediții ; "The Genetical Theory of Natural Selection", 1930 ; "The Design of Experiments" ; "Statistical Tables for Biological Agricultural and Medical Research" (în colab. cu F. Yates), 1938 ; "Statistical Methods and Scientific Inference", 1956.

FORTUNATI, Paolo (1906—1980). Statistician și demograf italian. Studii la Universitatea din Padova (licențiat în statistică, 1927) ; profesor de statistică la Universitatea din Palermo, a condus Institutul de Statistică de la Universitatea din Bologna timp de 35 ani. Membru al Institutului Internațional de Statistică ; al Uniunii Internaționale pentru Studiul Științific al Populației ; al Societății de Statistică din Paris. Editor și fondator al revistei „Statistica". Fost președinte al Societății Italiene de Statistică. Preocupări de metodologie statistică, în special în domeniul variabilității ; de asemenea, aplicații statistice la economie. A publicat studii de demografie generală și de demografie istorică. A dezvoltat unele din metodele lui C. Gini, al cărui discipol s-a considerat.

Op. : „Lezioni di demografia", Scuola di Statistica dell'Università di Padova, 1931 ; „Statistica e politica economica", în „Statistica" nr. 2, 1943, Ferrara. ; „Sulle distribuzione massimanti della variabilità" în : Actes de la 28e session de l'Institut International de Statistique, Roma, 1954 ; „Statistica e ricerca scientifica", în „Statistica" nr. 3, 1958, Bologna ; „Contributi alle ricerche sulle relazioni tra le misure della variabilità con riferimento alle ricerche economiche", Roma, 1960 ; „Considerazioni sugli schemi di Pareto e di Gini a proposito di misura della disuguaglianza", în „Statistica" nr. 3, 1971, Bologna, 1971. ; „Tecnica conoscitiva statistica e astrazione determinata", în „Statistica" nr. 3, 1976, Bologna, 1976.

FRÉCHET, Maurice (1878—1973). Matematician, statistician francez. Profesor la École Normale Supérieure. Membru al Academiei de Științe din Franța, membru al Institutului Internațional de Statistică (din 1931) și membru de onoare (din 1959), membru de onoare al Institutului Actuarilor Francezi. Contribuții la noțiunea de variabilă aleatoare, studiul tuturor formelor de con-

vergență, studiul probabilităților în lanț, teoria spațiilor abstracte. În cadrul Enciclopediei calculului probabilităților a publicat două cărți.

Op.: „Integrala abstractă a unei funcții abstracte de variabilă abstractă". Curs universitar. 1948—1949. „Les espaces abstraits en statistique théorique et même appliquée".

FRIȘCII, Ragnar (1894—1973). Statistician, economist și econometrician norvegian. Profesor la Universitatea din Oslo. Întemeietor al econometriei: împreună cu Charles F. Roos și Irving Fisher a creat Econometric Society (1930). În 1969 a fost distins cu Premiul Nobel (științele economice) pentru lucrările sale consacrate modelelor dinamice necesare analizei proceselor economice. Membru al Institutului Internațional de Statistică (din 1937). S-a ocupat de teoria cererii, teoria producției, estimarea ecuațiilor teoriei economice, analiza regresiei, modele econometrice.

Op.: "Pitfalls in the Statistical Construction of Demand and Supply Curves", Leipzig, 1933; "Statistical Confluence Analysis by Means of Complete Regression Systems", Oslo, 1934. Despre F.: K. J. Arrow, "The work of Ragnar Frisch econometrician", în „Econometrics", 28, 1969, p. 175—182.

GALTON, sir Francis (1822—1911). Statistician, matematician, genetician englez. Studii de matematică la Cambridge; văr cu Gh. Darwin. Împreună cu K. Pearson este considerat întemeietorul școlii anglo-saxone de statistică matematică. Fondator al biometriei: în 1901, împreună cu K. Pearson, întemeiază revista „Biometrika", în cuprinsul căreia s-au publicat cele mai importante studii de statistică matematică. A fost influențat de Quetelet. Membru al lui Royal Society. Cercetări în domeniul eredității, antropometriei, psihometriei. A adus contribuții la teoria corelației și regresiei (termenul îi aparține); a introdus termenul de "eugenics", interpretat adeseori greșit. Funcția lui G., „ogiva lui G." se numără de asemenea printre contribuțiile sale.

Op.: "Hereditary Genius: An Inquiry into its Laws and Consequences", 1869; "Inquiries Into Human Faculty and its Development", 1883; "Natural Inheritance", 1889, considerată ca cea mai importantă.

GAUSS, Carl Friedrich (1775—1855). Matematician, fizician și astronom german. Considerat drept „prinț al matematicienilor" și pus alături de Arhimede și Newton pentru contribuția sa la dezvoltarea matematicii. Profesor la Universitatea din Göttingen. A dat numeroase demonstrații riguroase și complete în diferite ramuri ale matematicii. Este cunoscut mai ales prin metoda celor mai mici pătrate, de mare utilitate în prelucrarea datelor experimentale, prin contribuții la teoria erorilor; a definit legea normală a repartiției numită și curba normală Gauss-Laplace; a găsit formula erorii probabile a mediei (1809).

Op.: "Disquisitiones arithmeticae", 1799—1801.

GEORGESCU, Dumitru C. (1904—1974) — medic, statistician și demograf român. Doctor în medicină la Facultatea de Medicină din București (1939); licențiat în drept, subdirector general la Institutul Central de Statistică (până

în 1948): a organizat și efectuat cercetări statistice, recensăminte, anchete. Cercetări științifice la Institutul de Oncologie din București (1949—1971). Studii de demografie și de statistică sanitară.

Op.: „Populația României", Institutul Central de Statistică, 1938 (în colab. cu S. Manuila); „Mortalitatea prin tumori maligne", Institutul Central de Statistică, 1939; „La fécondité différentielle de la population de la Roumanie", Institutul Central de Statistică, 1940; "Territorial Variations of the Mortality from Malignant Tumors in R.P.R." (în colab.), Acta Un. Inst. Cancer, 20, 3, 652—655, 1964.

GEORGESCU-ROEGEN, Nicholas (n. 1906). Statistician, matematician și economist american, de origine română. Licențiat în matematică al Universității București (1926), având printre profesori pe acad. O. Onicescu. Doctor al Universității din Paris, Institutul de Statistică (1930), având ca profesor pe G. Darmois. Cursuri postdoctorale la Universitatea din Londra, cu Karl Pearson, precum și la Universitatea Harvard, cu J. A. Schumpeter.

Între 1932—1946, profesor la Școala de Statistică din București condusă de acad. O. Onicescu, apoi la Institutul de Statistică, Calcul și Actuariat; de asemenea, a lucrat la Institutul Central de Statistică. Profesor la Universitatea Vanderbilt (S.U.A.), între anii 1947—1976. Profesor emeritus (1971) al acestei universități. Membru al Institutului Internațional de Statistică (1970) și al diferitelor academii și instituții internaționale. Contribuții la teoria probabilităților, statistica matematică, econometrie, economie matematică, economie. Primele studii au fost publicate în „Journal de la Société de Statistique de Paris", în „Biometrika". După părerea acad. Gh. Mihoc, **G.R.** este primul cercetător român care a publicat lucrări originale în domeniul statisticii matematice. A introdus o interpretare nouă a economiei, prin așa-numita „bioeconomie".

Op.: „Le problème de la recherche des composantes cycliques d'un phénomène", în „Journal de la Société de Statistique de Paris", 1930; „Metoda statistică. Elemente de statistică matematică", cu o prefață de Octav Onicescu. Institutul de Statistică Generală a Statului, București, 1933, XV + 507 p.; „Enciclopedia României" 1939—1943. Studii de **G.R.**: Inventarul agricol al României. Comerțul exterior al României. Costul vieții în România. Avuția națională a României; "The Entropy Law and the Economic Process". Harvard University Press, Cambridge, Mass. 1971, trad. în rom.: „Legea entropiei și procesul economic". Cuvînt înainte al autorului. Studiu introductiv A. Jaeu, Gh. Mihoc, B. Zaharescu, Editura politică, București, 1979, 688. p.

GINI, Corrado (1884—1965). Statistician, demograf, economist și sociolog italian. Profesor la Universitatea din Roma; între 1926—1932 a condus Institutul Central de Statistică al Italiei. A înființat revista „Metron" (1920) și „Genus", revista Comitetului italian pentru studiul populației. Membru de onoare al Academiei Române (1938), al altor academii, doctor honoris causa al Universităților din Geneva și Harvard; membru de onoare al Institutului Internațional de Statistică. Contribuții la biometrie, antropologie, statistică, demografie. A creat noțiunile de indici de conexiune, homofilie, cograduare concentrare („raportul lui Gini"); contribuții la teoria dispersiei, a asociației,

între două variabile, a mediilor și diferenței dintre acestea. A efectuat comparația internațională în legătură cu avuția națională și venitul național.

Op. : „Memorie di metodologia statistica", vol. 1. „Variabilità e concentrazione", 1912 ; „Quelques considérations au sujet de la construction des nombres indices des prix et des questions analogues" („Metron". 1924) ; „Sur la théorie de la dispersion et sur la vérification et l'utilisation des schémas théoriques ; „Le medie", 1958 : „La logica nella statistica", 1962.

GOLOPENȚIA, Anton (1909—1951). Sociolog, statistician și demograf român. Studii la Universitatea din București : licențiat în drept (1930), licențiat în filozofie (1933), studii la Berlin, Hamburg și Leipzig (1935—1936) ; doctor în filozofie al Universității din Leipzig (1936) cu teza „Die Information des Staatsführung und die überlieferte Soziologie" („Informația conducerii de stat și sociologia tradițională"). Asistent la catedra de sociologie, etică și politică a Universității din București (1936—1939), condusă de prof. Dimitrie Gusti. În perioada 1940—1948, la Institutul Central de Statistică ca inspector general, director al Oficiului de Studii, director general. Preocupări referitoare la valoarea sociologiei tradiționale, la raportul dintre teoria sociologică generală și cercetările sociologice concrete. Colaborator a lui Dimitrie Gusti la cercetări monografice și studii sociologice regionale. A participat și a condus înregistrări statistice (recensăminte, anchete inventarieri).

Op. : „Rolul actual al sociologiei", în „Sociologie Românească", an II, no. 1 ian.-febr. 1937 ; „Starea culturală și economică a populației rurale din România ", în „Revista de igienă socială", an 10, nr. 6, 1940 ; „60 de sate românești cercetate de echipele studențești în vara anului 1938". Anchetă sociologică condusă de A. Golopenția și dr. D. C. Georgescu, Institutul de științe sociale al României, Biblioteca de sociologie, etică și politică. Sociologia României, 4, București, 1942 ; „Dîmbovnicul — o plasă din sudul județului Argeș" (cu Mihai Pop), Institutul de științe sociale al României, Buc., 1942 ; „Populația Republicii Populare Române la 25 ianuarie 1948. Rezultatele provizorii ale Recensământului" (împreună cu dr. D. C. Georgescu), în „Probleme economice", nr. 2, martie 1948 și ca Extras , Institutul Central de Statistică, București, 1948.

GOSSET, William Sealy (1876—1937). Matematician și statistician englez, cunoscut sub pseudonimul „Student" ; reprezentant de frunte al școlii anglo-saxone de statistică matematică. A condus laboratorul unei fabrici de bere, ajungînd chiar „șef de fabricație a berii", la firma Park Royal, împrejurare care i-a determinat să aplice metodele statisticii matematice. Un rol important în dezvoltarea sa l-a avut întîlnirea cu K. Pearson (1905), care i-a oferit posibilitatea de a lucra la Laboratorul biometric de la University College din Londra, și cu R. A. Fisher (1922), la celebra "Rothamsted Experimental Station". După caracterizarea lui R. A. Fisher, G. este „Faraday al Statisticii", pentru intuiția și ingeniozitatea sa. Cele mai multe studii le-a publicat în „Biometrika". Contribuțiile sale se referă la teoria sondajelor de volum redus, la legea de repartiție a mediilor, teoria erorii probabile a mediei, a erorii probabiliste a coeficientului de corelație. Repartiția „Student" este una din repartițiile fundamentale.

Op. : "The Probable Error of a Mean", 1908 (testul *t* Student) ; "The Probable Error of a Correlation Coefficient", 1908.

GRAUNT, John (1620—1674). Statistician și demograf englez. Împreună cu *W. Petty**, întemeietor al școlii de „aritmetică politică". În lucrarea sa din 1662, considerată ca primă lucrare științifică de demografie, a valorificat o mare cantitate de informații statistice, determinînd cele mai importante regularități și legități în domeniul fenomenelor demografice : mortalitatea ca funcție de vîrstă și sex, echilibrul dintre sexe la naștere ; a întocmit prima tabelă de mortalitate a populației. Considerat întemeietor al demografiei.

Op. : "Natural and Political Observations upon the Bills of Mortality Chiefly with Reference to the Government, Religion, Trade, Growth, Air, Diseases etc. of the City of London", London, 1662.

HINCIN, Aleksandr Iakovlevici (1894—1959). Matematician probabilist sovietic. A studiat la Facultatea de fizico-matematică de la Universitatea din Moscova. Profesor de teoria probabilităților la Universitatea din Moscova. Membru corespondent al Academiei de Științe a U.R.S.S. Împreună cu A. N. Kolmogorov a găsit, în 1925, condițiile necesare și suficiente ale convergenței seriilor ai căror termeni sînt mărimi aleatoare independente.
Contribuții importante : demonstrarea legii logaritmului iterat pentru schema Bernoulli, teoria însumării variabilelor aleatoare independente, dezvoltarea teoremei lui Leapunov, teorema ergodică a lui Hincin, mecanica statistică, teoria informației și, mai ales, contribuția la construirea bazelor teoriei proceselor aleatoare staționare.

Op. : Khinchin, A. I., Gnedenko, B. V., "An Elementary Introduction to the Theory of Probability", New York, 1962 (trad. după originalul în rusă : „Elementarnoe vvedenie v teoriiu veroiatnostei", ed. a V-a, 1945).

HOTELLING, Harold (1890—1973). Matematician, statistician și economist matematician american. Studii la Universitățile din Washington și Princeton. Profesor de economie la Columbia University (1931), la University of North Carolina at Chapel Hill (1946), unde a organizat Departamentul de statistică matematică. În 1929 a lucrat cu R. A. Fisher la celebra "Rothamsted Agricultural Station" din Anglia. Membru al Institutului Internațional de Statistică (din 1948), al Academiei Naționale de Științe a S.U.A., la Accademia Nazionale dei Lincei (Roma), membru de onoare al lui Royal Statistical Society. Președinte al lui Econometric Society (1936—1937). Contribuții la generalizarea testului Student, la analiza factorială, teoria corelațiilor economice, distribuția coeficienților de corelație a rangurilor Spearman.

IONESCU DE LA BRAD, Ion (1818—1891). Agronom, economist și statistician român. Studii în Franța. Profesor la Academia Mihăileană (1842). Între 1 iulie și 28 octombrie 1859 a condus Direcția Centrală de Statistică a Moldovei și a pregătit Recensămîntul populației și agriculturii din 1859—1860. A elaborat „Povățuiri pentru facerea catagrafiei Moldovei", care este și primul curs de statistică (1859). A participat la congresele internaționale de statistică. Precursor al cercetării monografice din România. Împreună cu Dionisie Pop Marțian, ctitor al statisticii românești.

Op. : „Povățuiri pentru facerea catagrafiei Moldovei. Precedată de oarecare elemente de statistică", Iași, 1859; „Agricultura română din județul Dorohoi", Imprimeria statului, București, 1866; „Agricultura română din județul Mehedinți", Imprimeria statului, București. 1868; „Agricultura română din județul Putna", Imprimeria statului, București 1869; „Rapport du délégué officiel du gouvernement roumain au septième Congrès International de Statistique sur les progrès statistiques accomplis en Roumanie depuis le dernier Congrès de Florence jusqu'à celui de La Haye, 1869".

IONESCU, Haralambie M. (1909—1974). Matematician, statistician și actuar român. A urmat cursurile Facultății de Științe a Universității din București, licențiat în matematici (1930), doctor în matematici la Universitatea din București (1948), cu teza „Schema Taucer generalizată și ecuațiile liniare transgresive cu limite variabile"; doctor docent în științe matematice.

A fost statistician și actuar expert la Casa Centrală a Asigurărilor Sociale, unde a funcționat pînă în 1948, cînd a trecut la Academia de Studii Economice, unde a fost profesor și șef al catedrei de matematici pînă la sfîrșitul vieții. A participat la manifestări științifice din țară și din străinătate, prezentînd comunicări și ținînd conferințe științifice.

Domenii principale abordate : actuariat, teoria probabilităților, statistica matematică, programare matematică, cercetări operaționale, fiind unul din promotorii metodelor matematice moderne cu aplicații în economie. Autor a numeroase manuale și tratate de teoria probabilităților și de statistică matematică, destinate învățămîntului universitar, unde a activat aproape trei decenii. Autor și coautor la numeroase lucrări de economie matematică. În actuariat a dat studii în legătură cu teoria rezervelor matematice în asigurări sociale ; aplicații ale matematicii actuariale în demografie și ocrotirea sănătății. A publicat (în colab. cu Gh. Mihoc) prima monografie consacrată programării liniare, altele consacrate teoriei grafurilor și cercetărilor operaționale.

Op. : „Asupra rezervelor matematice din Asigurările Sociale", în Buletinul Asigurărilor Sociale, 1945.; „Calcule numerice", 1954 : „Curs de matematici pentru statisticieni și economiști" (în colab.), 1956; „Teoria matematică în operațiunile financiare", vol. I și II (în colab. cu Gh. Mihoc's.a.). 1960 : „Statistica matematică". 1962 ; „Matematici superioare", vol. I și II, 1963. (în colab.) ; „Bazele matematice ale programării liniare" (în colab. cu Gh. Mihoc), 1965 ; „Teoria grafurilor cu unele aplicații în economie", 1969 ; „Probleme ale cercetării operaționale", 1972.

KENDALL, sir Maurice (1907—1983). Matematician și statistician englez, reprezentant strălucit al școlii anglo-saxone de statistică matematică. Statistician la Ministerul agriculturii al Angliei, șeful catedrei de statistică de la London School of Economics (1949—1961), președinte al lui Royal Statistical Society, membru al Institutului Internațional de Statistică, președinte de onoare al acestuia. Director al Anchetei Mondiale a Fertilității (1972—1980), cea mai amplă anchetă demografică, desfășurată sub auspiciile O.N.U. și executată de Institutul Internațional de Statistică, anchetă pentru care K. a primit Medalia Păcii O.N.U. (1980). Contribuții la statistica neparametrică, autor de

manuale şi tratate de statistică matematică. Numeroase lucrări de statistică. *Op.*: în colab. cu Yule, G. U., "An Introduction to the Theory of Statistics", 1911, Griffin, ed. XIV, Londra, 1960 (în rom. „Introducere în teoria statisticii" (după ed. XIV), Ed. ştiinţifică, Bucureşti 1969; "Dictionary of Statistical Terms" (în colab. cu W. R. Buckland), Boyd and Oliver, Edinburg, 1966; "Time-Series", Second Edition, Griffin, Londra, 1980; "Multivariate Analysis", Second Edition, Griffin, Londra, 1980; "Rank Correlation Methods", ed. IV. Griffin, 1981; "The Advanced Theory of Statistics" (în colab. cu A. Stuart); "Distribution theory", vol. II; „Inference and relationship", vol. II; "Design and analysis, and time-series" (în colab. cu J. Keith Ord), Griffin, ed. IV, vol. III, 1983.

KIAER, Anders Nikolai (1838—1919). Statistician norvegian. A condus Biroul Central de Statistică al Norvegici (1867—1913). Meritele sale principale în statistică sînt legate de aplicarea cercetărilor prin sondaj, fiind primul care a folosit-o la studiul repartiţiei populaţiei după venit şi avere şi a încercat să o teoretizeze, în ciuda unor critici vehemente din partea altor statisticieni ai vremii. Metoda a fost discutată în mai multe sesiuni ale Institutului Internaţional de Statistică şi acceptată cu greu, cu sprijinul unor statisticieni ca *A. L. Bowley* şi *C. Gini* (→ şi III. *Ancheta statistică*).

Op.: „Observations et Expériences concernant les dénombrements représentatifs", în Bull. de l'Institut International de Statistique, vol. 9, Livraison 2, 176—183 (1895—1896); „Die represäntative Untersuchungsmethode", în Allgemeines Statistisches Archiv, vol. 5, 1—22, 1899.
Despre K.: W. Kruskall and Frederick Mosteller: "Representative Sampling", IV: "The History of the Concept in Statistics", 1895—1939, în „International Statistical Review", 48, 1980.

KING, Gregory (1648—1712). Economist şi statistician englez, reprezentant al şcolii de aritmetică politică. Precursor al econometriei; a analizat relaţia dintre preţul grîului şi abaterile de la recolta normală de grîu; a efectuat comparaţii internaţionale privind venitul naţional şi cheltuielile Angliei, Franţei şi Olandei. Din analiza cererii a stabilit o relaţie numită astăzi „legea lui Gregory King". Estimaţii privind populaţia şi venitul naţional.

Op.: "National and Political Observations and Conclusions upon the State and Conditions of England", 1696.

KOLMOGOROV, Andrei Nikolaevici (n. 1903). Matematician sovietic. Membru al Academiei de Ştiinţe a U.R.S.S., membru de onoare al Academiei Republicii Socialiste România şi al altor academii. Membru de onoare al Institutului Internaţional de Statistică. A adus contribuţii fundamentale la multe ramuri ale matematicii şi în special la teoria probabilităţilor, fiind autorul axiomatizării acesteia şi unul din creatorii statisticii neparametrice. A folosit metodele teoriei funcţiilor de variabilă reală la axiomatizarea teoriei probabilităţilor, împreună cu *A. I. Hincin** şi a dezvoltat teoria proceselor Markov. A influenţat considerabil teoria probabilităţilor; printre elevii şi colaboratorii săi se numără acad. B. V. Gnedenko.

Op.: "Foundations of the Theory of Probability", New York, 1956 (apărută prima dată în 1933, în limba germană); „Osnovnîie poniatiia teorii veroiatnostei", Moscova-Leningrad, 1936.

LANGE, Oskar (1904—1965). Economist, statistician și sociolog polonez. Doctor în economie; profesor la Universitatea Jagellonă din Cracovia. Contribuții la teoria economică a socialismului, la econometrie, la statistica teoretică. A aplicat metodele matematice la studiul fenomenelor economice. Printre primii a formulat cibernetica economică. Lucrările sale au fost traduse în numeroase limbi. Membru al diferitelor asociații și instituții internaționale.

Op.: „Teoria statisticii", Curs universitar (l. polonă), 1952; "Statistical Estimation of Parameters in Markov Process", 1955; „Economia politică" (l. polonă) vol. I, 19,59; "Introduction to econometries", ed. 2, Pergamon Press, London, 1962; „Decizii optime", 1962. (în rom. „Decizii optime. Bazele programării", Ed. științifică, București, 1970); „Cibernetica economică", 1965 (în rom. „Introducere în cibernetica economică", Ed. științifică, București, 1967); În 1964, cu prilejul împlinirii vîrstei de 60 de ani, volumul omagial: "On political Economy and Econometrics. Essays in honour of Oskar Lange".

LAPLACE, Pierre-Simon marquis de (1749—1827). Matematician, astronom și fizician francez. Discipol al lui D'Alembert. Membru al Academiei Franceze (1816); a fost ministru de interne în timpul Consulatului lui Napoleon Bonaparte. Contribuții la mecanica cerească, fiind numit „al doilea Newton". A ilustrat aproape toate domeniile matematicii cu lucrări fundamentale. Un clasic al teoriei probabilităților; s-a ocupat de jocurile de noroc dezvoltînd unele idei ale lui A. de Moivre și Lagrange; s-a ocupat de inferența bayesiană, de funcții generatoare. A reformulat legea repartiției normale, formulată de A. de Moivre în 1733 și de asemenea a introdus teorema limită centrală în domeniul probabilităților. Aplicații originale în demografie; estimații privind populația Franței și formularea unei noi metode de întocmire a tabelei de mortalitate.

Op.: „Théorie analytique des probabilités", 1812; „Essai philosophique sur les probabilités", 1814.

LASPEYRES, Etienne (1834—1913). Statistician și economist german, profesor la universitățile din Basel, Riga, Karlsruhe. A publicat studii în „Jahrbücher für Nationalökonomie und Statistik" (editor Bruno Hildebrand). În statistica prețurilor a formulat un indice (1864) care îi poartă numele, avînd, alături de indicele Paasche, o largă utilizare.

Op.: „Hamburger Warenpreise 1851—1863 und die californisch-australischen Goldentdeckungen seit 1848", 1864; „Die Berechnung einer mittleren Warenpreissteigerung", în „Jahrbücher für Nationalökonomie und Statistik", vol. XVI.

LEAPUNOV, Aleksandr Mihailovici (1857—1918). Matematician rus, membru al Academiei de științe din Petersburg: elev al lui *P. L. Cebîșev**. Contribuții la teoria probabilităților. A demonstrat teorema limită centrală în condiții mai generale decît predecesorii săi și a folosit pentru prima dată metoda funcției caracteristice care a devenit o achiziție importantă a teoriei moderne a probabilităților.

Op.: „Izbranîe trudî" (sub redacția acad. V. I. Smirnov), Leningrad, 1948.

LEXIS, Wilhelm (1837—1914). Statistician, economist, matematician și demograf german. Profesor la universitățile din Strasbourg, Freiburg și Göttingen, unde a predat științele politice. Editor al monumentalei „Handwörterbuch der Staatswissenschaften".
Contribuții la statistică: teoria fenomenelor de masă, analiza dispersiei, omogenitatea statistică, semnificația testelor („criteriul de divergență L."); mai tîrziu, R. A. Fisher a analizat legătura dintre hi-pătrat și criteriul L. Contribuții importante în demografie („graficul lui L."). A exercitat o mare influență asupra statisticienilor.

Op.: „Einleitung in die Theorie der Bevölkerungsstatistik", 1875; „Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft", 1877; „Uber die Theorie des Stabilität statistisches Reihen", 1879; „Uber die Wahrscheinlichkeitsrechnung und deren Anwendung auf die Statistik", 1886.

LOTKA, Alfred James (1880—1949). Matematician, statistician și biolog american. A condus (din 1924) secția matematică a societății "Metropolitan Life Insurance". A scris 6 cărți și 95 de studii, marcînd începutul epocii moderne a demografiei. A anticipat unele idei ale ciberneticii: a interpretat cibernetic comportamentul ființelor vii. A introdus noțiunea de populație stabilă și staționară, a interpretat populația ca agregat cu procese de reînnoire, noțiune fundamentală în demografia matematică.
De asemenea, cu L. J. Dublin, a formulat „rata intrinsecă (adevărată) a creșterii naturale".

Op.: "On the true rate of natural increase of a Population" (în colab. cu L. J. Dublin), 1925; „Theorie analitique des associations biologiques", part. I, "Principes", part. II, „Analyse démographique avec applications particulières à l'espèce humaine", Paris, 1934—1939.

MACKENROTH, Gerhard (1903—1955). Statistician, demograf și economist german. A studiat psihologia, filozofia și economia politică: studii de specializare la universitățile din Stockholm, Londra și Cambridge; doctor în științe politice. Profesor la universitățile din Marburg și Kiel. Contribuții la metodologia statistică și aplicațiile ei în demografie și economie. Autor de tratate de statistică și demografie. În tratatul său de demografie ("Bevölkerungstheorie", 1953) încearcă o amplă sinteză a demografiei cu sociologia, într-o viziune modernă. A dezvoltat teoria „comportamentului generativ" și a „structurilor generative", de mare actualitate.

Op.: „Methodenlehre der Statistik", Göttingen, 1949; „Bevölkerung und Wirtschaft", studiu în H. f. Soz. Wiss., 1, Göttingen, 1950; „Bevölkerungslehre". Theorie, Soziologie und Statistik", Springer—Verlag, Berlin—Göttingen—Heidelberg 1953; „Die generative Struktur von Bevölkerungen und Sozialschichten", studiu în Weltw. Archiv, 75, 1955, Hamburg.

MAHALANOBIS, Prasanta Chandra (1893—1972). Matematician, statistician și economist indian. A contribuit la înființarea Institutului Indian de Statistică (1931): a proiectat și realizat Ancheta Națională prin Sondaj (N.S.S. — National Sample Survey), în 1950. Fost președinte al Comisiei de Statistică a O.N.U., Consilier onorific al guvernului indian. Membru al lui Royal Society (1945), al lui Econometric Society (1951), membru de onoare al lui Royal Sta-

tistical Society (1954), președinte de onoare al Institutului Internațional de Statistică (1957—1972): membru corespondent al Academiei de Științe a U.R.S.S. (1958). Organizator de instituții și anchete statistice. A publicat cinci cărți și peste 200 de studii. Contribuții la teoria sondajului cu eșantioane mari, la taxonomie ("Mahalanobis Distance"), la analiza factorială. Creatorul școlii de statistică din India, cu numeroși discipoli de valoare internațională; contribuția sa este considerată drept „epoca de aur a statisticii din India" (C. Rao).

MANUILA, Sabin (1894—1964). Medic, statistician și demograf american, de origine română. Doctor în medicină la Universitatea din Budapesta (1919). Studii de specializare la Universitatea Johns Hopkins din Baltimore (S.U.A.), în 1926. Activitate didactică și de cercetare la Cluj, sub îndrumarea prof. Victor Babeș: șef de lucrări la Catedra de igienă și igienă socială la Cluj, șef de secție la Institutul de igienă și sănătate publică al Facultății de medicină din București (1927—1946). Cursuri de biometrie la Facultatea de Medicină și la Institutul de statistică, calcul și actuariat al Universității din București. A organizat și condus lucrările Recensămîntului general din 1930, ale cărui rezultate au fost publicate în 9 volume. Director general al Institutului Central de Statistică (pînă în 1947). Activitate internațională: vicepreședinte al sesiunii Comisiei Internaționale pentru Studiul Populației (1932 și 1933): membru al Comisiei Populației la Societatea Națiunilor: membru al Institutului Internațional de Statistică. A fost ales membru corespondent al Academiei Române (1938) la propunerea lui Dimitrie Gusti, care l-a caracterizat ca fiind „organizatorul statisticei științifice din România".

A publicat studii privind statistica sanitară a României, structura pe naționalități a populației României, fertilitatea diferențială, funcțiile statisticii în statul modern.

Op.: „Populația României", Institutul Central de Statistică, 1938 (în colab. cu D. C. Georgescu).

MARKOV, Andrei Andreevici (1866—1922). Matematician rus. Profesor la Universitatea din Petersburg, membru al Academiei de Științe din Petersburg. Contribuții deosebite la teoria probabilităților, definind pentru prima dată șirurile de variabile dependente care îi poartă numele și stau la baza teoriei moderne a proceselor stohastice. Noțiunea de lanț M. a deschis un cîmp important de cercetări cu aplicații în cele mai diferite domenii. Școala românească de teoria probabilităților a dezvoltat mult acest concept modern.

Op.: „Iscislenie veroiatnostei", 1910: „Izbranîie trudi. Teoria cisel. Teoriia veroiatnostei", Moscova, 1951.

MARȚIAN, Dionisie Pop (1829—1865). Statistician și economist român. Studii de economie și statistică la Universitatea din Viena, cu Lorenz von Stein. A fost primul șef al Oficiului Central de Statistică din Muntenia (1859): după contopirea celor două instituții de statistică din Moldova și Muntenia (1862), M. este numit director al Oficiului Central de Statistică, calitate pe care o deține pînă la moartea sa.

A organizat rece nsământul din 1859—1860, recensământul stabilimentelor industriale din 18 63. Spirit organizator, sub conducerea sa, instituția centrală

de statistică a efectuat importante lucrări. A înființat prima revistă de statistică din România : ,,Analele statistice. Pentru cunoștința părții muntene din România" (1860), publicație trimestrială. A publicat studii, a sistematizat date statistice pe care le-a publicat în ,,Analele statistice". **M.** a participat la Congresul Internațional de Statistică din 1863 (Berlin) cu o comunicare asupra lucrărilor de statistică din Principatele Române. Considerat drept ctitor al statisticii românești.

MIHOC. Ghеorghe (1906—1981). Matematician, statistician și actuar român. Licențiat în matematici al Universității București (1928); doctor în științe statistice și actuariale al Universității din Roma (1930), doctor în matematici al Universității București (1934); doctor docent în științe matematice. Asistent la Facultatea de științe a Universității București (1930—1943). Conferențiar (1943—1946); profesor șef de catedră din 1946 pînă în 1976. Rector al Universității București (1963—1968), director al Centrului de Statistică Matematică al Academiei R.S.R. (1963—1976). Director general al Direcției Centrale de Statistică (1948—1951). Academician (1963), președinte al Academiei Republicii Socialiste România (din martie 1980 pînă la încetarea din viață).
Membru al Institutului Internațional de Statistică (1969), al Societății Internaționale de Biometrie (1970).
Lucrări de teoria probabilităților și statistică matematică, a publicat peste 30 de tratate și monografii. A introdus, împreună cu acad. O. Onicescu., noțiunea de lanț cu legături complete și metoda funcției caracteristice în studiul proceselor markoviene. Contribuții la actuariat și demografie (a întocmit tabele de mortalitate pentru populația României), studii de teoria siguranței, controlul statistic al calității, programare stohastică.

Op.: ,,Calculul probabilităților", 1939 (în colab. cu O. Onicescu); ,,Tratat de matematici actuariale", 1943; ,,Calculul probabilităților și aplicații", 1956 (în colab. cu O. Onicescu și C. T. Ionescu Tulcea); ,,Lecții de statistică matematică", 1957 (în colab. cu O. Onicescu); ,,Matematici aplicate în statistică" (în colab. cu V. Urseanu), 1962; ,,Sondaje și estimații statistice. Teorie și aplicații" (în colab. cu V. Urseanu), 1977; ,,Procese stochastice. Elemente de teorie și aplicații" (în colab. cu C. Bergthaler și V. Urseanu), 1978; ,,Modele de analiză statistică" (în colab. cu V. Urseanu și E. Ursianu), 1982.

MISES, Richard von (1883—1953). Matematician și inginer german. Profesor de mecanică și matematici aplicate; în 1933, profesor la Istanbul; în 1939, profesor la Harvard. A înființat, în 1921, Zeitschrift für angewandte Mathematik und Mechanik — prima de acest gen din lume. S-a preocupat de mecanică și geometrie, analiză și teoria probabilităților și statistica matematică. A introdus conceptul de ,,colectiv", a tratat noțiunea de probabilitate ca limită a frecvenței, astăzi depășită de axiomatizarea datorată lui Kolmogorov.

Op.: ,,Fundamentalsätze der Wahrscheinlichkeitsrechnung", 1919: ,,Grundlagen der Wahrscheinlichkeitsrechnung" 1919: Opera sa, reeditată de soția sa, Hilda Geiringer, "Mathematical Theory of Probability and Statistics" a fost publicată în 1964, la New York.

MOIVRE, Abraham de (1667—1754). Matematician englez de origine franceză. Membru al lui Royal Society. Contribuții importante la teoria probabilităților; a formulat legea normală, forma limită a termenului binomial, funcția genera-

toare de probabilitate. Are contribuții și la actuariat și la demografie (speranța de viață și probabilitatea de supraviețuire).

Op.: "The Doctrine of Chances: or, a Method of Calculating the Probabilities of Events in Play", 1718: "Annuties upon Lives: or, the Valuation of Annuities upon Any Number of Lives, as also, of Reversions", 1725.

NEMCINOV, Vasili Serghecvici (1894—1964). Statistician și economist sovietic. Membru al Academiei de Științe a U.R.S.S., membru al Institutului Internațional de Statistică. A lucrat vreme îndelungată la Direcția Centrală de Statistică a U.R.S.S., în cadrul căreia a efectuat calcule economice, considerate astăzi clasice. A adus contribuții la statistica matematică, dezvoltînd polinoamele lui Cebîșev. Promotor al aplicării metodelor matematice în economie; în 1958 creează Laboratorul de aplicare a metodelor matematice în economie, devenit mai tîrziu Institutul de Economie Matematică al Academiei de Științe a U.R.S.S. Laureat al Premiului Lenin.

Op.: „Selskohozeastvennaiia statistika s osnovami obșcei teorii", 1946; "Polinomî Cebișeva i matematiceskaia statistika", 1946: "Ekonomiko-matematiceskie metodî i modeli", 1962.

NEYMAN, Jerzy (1894—1981). Matematician și statistician american de origine poloneză. Membru al Institutului Internațional de Statistică (din 1948); președinte de onoare al acestuia. Membru al diferitelor academii și instituții științifice. Domeniul de predilecție: testarea ipotezelor statistice, împreună cu *E. S. Pearson**. Aplicații în teoria sondajelor. Profesor la Department of Statistics University of California, Berkeley.

Op.: "On the Problem of the most efficient tests of statistical hypotheses" (în colab. cu E. G. Pearson), în Phil. Trans. Roy. Soc. 231, 289—337, 1933; "On the two different aspects of the representative method — the method of stratified sampling and the method of purposive selection", in Journal of the Royal Statistical Society, vol. 97, 558—606, 1934; "Lectures and Conferences in Mathematical Statistics and Probability". U.S. Dept. of Agr., Washington, 1952; "Joint Statistical Papers of J. Neyman and E. S. Pearson", University of California Press Berkeley and Los Angeles, 1966.
Despre **N.**: Bartoszynski, L. and Klonecki, W., "Proceedings of the Symposium to honour Jerzy Neyman", held in Varsaw, April 3—10. 1974. Varșovia, 1977.

ONICESCU, Octav (1892—1983). Matematician, statistician, mecanician și filozof român. Licențiat în matematică și filozofie al Universității București (1913); doctor în matematică al Universității din Roma (1920) cu teza „Sopra gli spazi einsteinieni a gruppi continui di transformazioni", susținută în fața comisiei prezidate de T. Levi-Città și avînd ca membri pe G. Castelnuovo V. Volterra. Profesor la Liceul militar „Mănăstirea Dealu". Carieră didactică universitară din 1922; profesor de mecanică teoretică și apoi de calculul probabilităților. Creator al Școlii românești de teoria probabilităților, ajutat de fostul său student, Gheorghe Mihoc. În 1930 a înființat și condus Școala de Statistică, transformată, în 1941, în Institutul de Statistică, Actuariat și Calcul de pe lîngă Universitatea din București. Președinte al Comisiei Centrale a recensămîntului din 1941. În anii 1948—1962, profesor de mecanică și titular la ca-

tedra de calculul probabilităților de la Universitatea București. Din 1962, profesor consultant. A inițiat și a condus o serie de seminarii științifice, printre care cel de la Direcția Centrală de Statistică (1965), avînd ca obiect aplicarea metodelor statistico-matematice în economie. Membru corespondent al Academiei Române (1933); membru titular (1965). O. a fost unul din inițiatorii creării Uniunii Interbalcanice a Matematicienilor (1934). Din 1968 a fost unul din rectorii Centrului Internațional de științe mecanice din Udine (Italia). Membru al Institutului Internațional de Statistică (1935), membru de onoare al acestuia pînă la moarte; membru al altor instituții și asociații internaționale. Membru al Consiliului Științific — Metodologic al Direcției Centrale de Statistică. Contribuții teoretice în special la teoria probabilităților prin crearea teoriei dependenței cu legături complete (în colab. cu acad. Gheorghe Mihoc), teoria funcțiilor holotrope, teoria mecanicii invariantive, statistica informațională (,,energia informațională O."), econometria informațională (împreună cu M. C. Botez). Autor al unui interesant model demoeconomic, care îi poartă numele.
Din vasta sa activitate științifică și didactică publicată, menționăm cîteva din domeniul teoriei probabilităților și statisticii.

Op.: ,,Théorie générale des chaînes à liaisons complètes", 1938; ,,Calculul probabilităților", 1956; ,,Lecții de statistică matematică" (în colab. cu Gh. Mihoc). 1957; ,, Teoria probabilităților și aplicații", 1963; ,, Numere și sisteme aleatoare", 1964; ,,Mecanica", 1969; ,,Mecanica statistică" (în colab.), 1971; ,,Mecanică invariantivă și cosmologie", 1974; ,,Probabilități și procese aleatoare", 1977; ,,Mecanică invariantivă", 1976; ,,Elemente de statistică informațională cu aplicații" (în colab.), 1979.
Despre O.: *Papers in Honour of Octav Onicescu on his 90th birthday,* Studies in Probability and Related Topics (Eds. M. C. Demetrescu, Marius Iosifescu), Editrice Nagard, Roma, 1983; V. Trebici, *Academicianul Octav Onicescu și știința econon ică*, în ,,Revista economică" nr. 16, septembrie 1983.

PAASCHE, Hermann (1851—1925). Economist și statistician german; s-a ocupat de statistica prețurilor, elaborînd, în 1874, un indice ce-i poartă numele (→ *Indice Paasche*), care se deosebește de *indicele Laspeyres** prin sistemul de ponderare.

Op.: ,,Über die Preisentwicklung der letzten Jahre nach der Hamburger Preisnotierung", în ,,Jahrbücher für Nationalökonomie und Statistik", 1874; ,,Studien über die Natur der Geldwertung und ihre praktische Bedeutung in den letzten Jahrzehnten auf Grund statistischen Detailmaterials entnommen der Stadt Halle a/S", Jena, 1878.

PEARSON, Egon Sharpe (1895—1980). Matematician și statistician englez, fiul lui *K. Pearson**. Studii la Cambridge. Profesor de statistică la Universitatea din Londra; șef al Departamentului de Statistică; profesor emerit în 1960. Membru al Institutului Internațional de Statistică; fost președinte al lui Royal Statistical Society; membru de onoare al Institutului Actuarilor. Editor al revistei ,,Biometrika" (1936—1966). Contribuții la probleme de testare a ipotezelor statistice împreună cu *J. Neyman**, în cadrul colaborării lor în perioada 1928—1938, care a avut ca urmare studii apărute în cinci reviste. Ideea a fost sugerată de o controversă între *E. Borel** și J. Bertrand în legătură cu testele ipotezelor. Primul rezultat au fost erorile de gradul 1 și de gradul 2 în testarea

ipotezelor, dezvoltat apoi în alte studii (de exemplu, testul λ). Strălucit reprezentant al şcolii anglo-saxone de statistică matematică. (→ *K. Pearson, R. A. Fisher, W. S. Gosset (Student).*)

Op.: Pearson, E. S. and Neyman, J., "On the Problem of the most efficient tests of statistical hypotheses", în Phil. Trans. Roy. Soc. 231, 289—337, 1933; Pearson, E. S. and Hartley, H. O., "Biometrika Tables for Statisticians", vol. 1, ed. 3, Cambridge Univ. Press, 1966; "Selected Papers of ... University of California Press, Berkely, 1966; "Memories of the Impact of Fisher's Work in the 1920 s", în „International Statistical Review", vol. 42, Number 1, April 1974; "Joint Statistical Papers of J. Neyman and E. S. Pearson", University of California Press, Berkely and Los Angeles, 1966; Pearson, E. S. and Kendall, M. G. (eds.), "Studies in the History of Statistics and Probability", Griffin, Londra, 1970.
Despre **P.**: Jerzy Neyman, "Egon S. Pearson" (1895—1980). "An appreciation" în "The Annals of Statistics", vol.9, No. 1, January 1981.

PEARSON, Karl (1857—1936). Matematician, statistician şi biolog englez. Profesor la universităţile din Cambridge şi Londra. Unul din întemeietorii şcolii anglo-saxone de statistică matematică. A întemeiat, împreună cu Fr. Galton şi W. F. R. Weldon, "Biometrika. A Journal for the Statistical Study of Biological Problems" (1900), pe care a condus-o până la moarte, după care a fost preluată de fiul său, *E. S Pearson**. A editat: "Annals of Eugenics" din 1925, ca publicaţie a Laboratorului de Eugenie. A adus contribuţii la toate capitolele statisticii matematice şi a lărgit cadrul ei de aplicaţie: teoria repartiţiilor statistice, momente, corelaţii, hi-pătrat, teoria erorilor, funcţiile beta şi gamma şi altele. Influenţa sa a fost considerabilă; după unii, **P.** este „întemeietorul ştiinţei statisticii".

Op.: "The Grammar of Science", 1892; "Contributions to the Mathematical Theory of Evolution", 1894—1896; "Tables for Statisticians and Biometricians", 2 vol. 1914; "On the Relationship of Health to the Physical and Psychical Characters in School Children", 1923.

PETTY, sir William (1623—1687). Economist şi statistician englez; întemeietor al şcolii de aritmetică politică, împreună cu J. Graunt. Medic, profesor de anatomie la Universitatea din Oxford, inventator, om de afaceri, a dat dovadă de o ingeniozitate extraordinară. După aprecierea lui K. Marx, **P.** este „întemeietorul economiei politice moderne, unul din cei mai geniali şi mai originali cercetători ai economiei" şi „într-o oarecare măsură inventatorul statisticii".
În lucrarea sa, „Aritmetica politică" (1676) se cuprind bazele statisticii. A efectuat calcule asupra populaţiei, comerţului exterior, avuţiei naţionale. Contribuţii importante şi la demografie.

Op.: „A Treatise of Taxes and Contributions", 1662; "Political Arithmetick, or a Discourse concerning the Extent and Value of Lands, People, Buildings", 1676; "Five Essays in Political Arithmetick", 1687.

POISSON, Siméon Denis (1781—1840). Matematician francez. Profesor de mecanică la Paris. A adus contribuţii importante la fizica matematică, hidromecanică, electrostatică. În teoria probabilităţilor, a descoperit repartiţia

care îi poartă numele , care se manifestă în cazul evenimentelor rare („legea P.", „procese P.). De asemenea, a generalizat teorema lui Bernoulli cu privire la legea numerelor mari, creînd cu acest prilej și termenul ca atare „loi de grands nombres". Contribuții și la demografie.

Op. „Recherches sur les probabilités des jugements, en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités", 1837.

QUETELET, Lambert Adolphe Jacques (1796—1874). Statistician, matematician, astronom belgian. A contribuit la crearea Comisiei Naționale de Statistică a Belgiei (1841), al cărei președinte a fost pînă la moarte. A aplicat teoria probabilităților la studiul fenomenelor sociale. În concepția sa, statistica era o „fizică socială". A introdus noțiunea de „om mediu", controversată de-a lungul timpului. Un mare merit îl are Q. în organizarea Congreselor internaționale de statistică (1853) și activitatea de asigurare a comparabilității internaționale. Teoria „omului mediu" a fost reluată de sociologul M. Halbwachs, de Corrado Gini, M. Fréchet. Cu prilejul centenarului morții sale (1974) s-au reeditat lucrările sale și s-au organizat sesiuni științifice, punîndu-se în evidență rolul și contribuția excepțională a lui Q. la progresul statisticii.

Op.: „Instructions populaires sur le calcul des probabilités", 1928; „Sur l'homme et le développement de ses facultés ou essai de Physique sociale", 1835; „L'anthropologie ou mesure des différentes facultés de l'homme", 1871.

RÉNYI, Alfred (1921—1970). Matematician, statistician maghiar. Studii la Universitatea din Szeged, specializare la Universitatea din Leningrad, cu prof. Linnik. Director al Institutului de Matematică al Academiei de Științe din R. P. Ungară; profesor la Universitatea din Budapesta și șef de catedră la teoria probabilităților. A publicat circa 220 de cărți, studii și memorii privind analiza complexă și reală, teoria numerelor, teoria probabilităților, teoria informației, statistica matematică, teoria grafurilor. Membru al diferitelor societăți științifice. Vicepreședinte al Institutului Internațional de Statistică (1965—1969).

Op.: "Foundations of Probability Theory". Holden Day. San Francisco, 1970; "Probability Theory", North-Holland, Amsterdam, 1970.
Despre P.: P. Revesz, "The work of Alfred Rényi. II. "The work in probability theory", în „Mat. Lapok", 21 (1970), p. 211—231; A. Rényi: „Dialog despre calculul probabilităților", Editura enciclopedică, București, 1973.

RIMNEAMȚU, Petru (1902—1981). Medic, statistician și demograf român. Doctor în medicină și chirurgie al Facultății de medicină din Cluj (1927). Specializare în igienă și sănătate publică și în statistică la Universitatea Johns Hopkins din Baltimore (S.U.A.) în anii 1932—1933, la prof. L. J. Reed. Preparator, asistent și șef de secție la Institutul de Igienă și Sănătate Publică din Cluj (1924—1943), profesor de igienă la Institutul de Medicină și Farmacie din Timișoara. Membru al Consiliului Sanitar Superior al României (1942); membru al American Statistical Association (1934) și al altor societăți străine. A efectuat numeroase cercetări biometrice, auxologice, de medicină socială, de epidemiologie, de demografie; printre primii, a întocmit tabele de morta-

litate; a contribuit esențial la introducerea metodelor statistico-matematice în medicina și demografia românească.

Op.: „Cercetări asupra originii etnice a populației din Sud-Estul Transilvaniei pe baza compoziției serologice a sîngelui", 1936 (Premiul „I. Oroveanu"); „Studii asupra depopulării Banatului", 1937 (Premiul „Gh. Chițu"); „Elemente de biometrie medicală și statistică vitală". Cu o prefață de prof. dr. Iuliu Moldovan, Editura Institutului Central de Statistică, București, 1939.

SAUVY, Alfred (n. 1898). Demograf, sociolog, statistician și economist francez. Diplomat al lui École Polytechnique (1922). Profesor la Collège d, France (din 1959). Doctor honoris causa al universităților din Geneva, Bruxellese Utrecht, Liège. Membru al Consiliului economic și Social al Franței. Reprezentant al Franței la Comisia O.N.U. pentru populație. Statistician la Direcția de Statistică Generală a Franței (1922—1937); director al Institutului de Conjunctură al Franței (1937—1945); director al Institutului Național de Studii Demografice de la Paris (1945—1962). Profesor la Institutul de Științe Politice (1940—1958); director al Institutului de Demografie al Universității din Paris (1957—1969). Membru al Institutului Internațional de Statistică (din 1948); membru de onoare al I.I.S.; membru al Uniunii Internaționale pentru Studiul Științific al Populației; președinte de onoare al U.I.S.S.P.

Domenii: demografie, teorie demografică, istorie economică, economie mondială, metode și modele în demografie și economie. Aplicații ingenioase ale statisticii la studiul fenomenelor economice, sociale și demografice.

Op.: „Essai sur la conjoncture et la prévision économique", 1938; „Le plan Sauvy", 1960; „Les limites de la vie humaine", 1961; „Théorie générale de la population", vol. I (1963), vol. II (1966); „Histoire économique de la France entre les deux guerres", 4 vol., 1965, 1967, 1972 și 1978.; "La fin des riches", Paris, 1975; „Coût et valeur de la vie humaine", Paris, 1977.

SMIRNOV, Nikolai Vasilievici (1900—1966). Matematician probabilist sovietic. Studii la Universitatea din Moscova (1921—1926); doctor în matematici cu teza „Cu privire la aproximarea legilor de repartiție a mărimilor aleatoare" (1938). Profesor la Academia agricolă „Timiriazev" și la Universitatea din Moscova. În anii 1938—1966 a lucrat la Institutul de matematică „V.A. Steklov" din Moscova. Membru corespondent al Academiei de Științe a U.R.S.S.; laureat al Premiului de Stat. A pus bazele teoriei metodelor neparametrice ale statisticii matematice care i-a adus celebritate mondială *(Testul* S.); a publicat manuale pentru aplicarea teoriei probabilităților și statisticii matematice; a continuat biblioteca de tabele de statistică matematică a lui E. E. Sluțki. S-a preocupat de genetica matematică („Evoluția unei populații finite în condițiile unei încrucișări întîmplătoare").

Despre S.: Academia de științe a U.R.S.S., „Teoriia veroiatnostei i eio primeneniia", tom XI, fascicola 3, 1966, Moscova.

SNEDECOR, George W. (1881—1974). Statistician, matematician american. Studii de matematică și fizică la Universitatea Alabama (S.U.A.). Profesor de matematică și statistică. Activitatea științifică desfășurată în principal la Iowa State College din 1931, cînd este numit profesor de matematică, și mai ales din 1933, cînd Laboratorul de Statistică este declarat Institut de Cercetări

pe lîngă Iowa State College, iar *S.* este numit director. Staţiunea experimentală, replică la celebra Rothamsted Experimental Station din Anglia, unde au lucrat generaţii de matematicieni statisticieni, a aplicat şi a perfecţionat metodele statistice experimentale în agrobiologie. Printre colaboratori: Gertrude Cox, G. Tintner şi *W. G. Cochran** Din 1950, S. a fost membru al Institutului Internaţional de Statistică; în 1948 — preşedinte al American Statistical Association; membru de onoare al lui Royal Statistical Society. După aprecierea, lui W. G. Cochran, S. face parte din aceeaşi familie cu Quételet, Galton, K. Pearson, R. Pearl şi R. A. Fisher.

Op.: "Statistical Methods applied to Experiments in Agriculture and Biology", Collegiate Press Inc., Arnes. Iowa (ediţii: 1937, 1938, 1940, 1946, 1954). Trad. rom. după ed. V (tiraj 7): „Metode statistice aplicate în cercetările de agricultură şi biologie", Editura didactică şi pedagogică, Bucureşti, 1968.

SPEARMAN, Charles Edward (1883—1945). Psiholog englez. A contribuit la crearea analizei factoriale, a statisticii neparametrice. A introdus coeficientul de corelaţie a rangurilor care îi poartă numele. Aplicaţii în psihologie; contribuţii la dezvoltarea psihometriei. Metodele elaborate de S. se aplică în cele mai variate domenii.

Op.: "General Intelligence", în: "American Journal of Psychology" ,1904; "The Abilities of Man. Their Nature and Measurement", 1927.

SÜSSMILCH, Johann Peter (1707—1767). Demograf şi statistician german. Membru al Academiei de Ştiinţe din Berlin. Adversar al şcolii de statistică descriptivă a lui Achenwall; unul din fondatorii demografiei ca ştiinţă de-sinestătătoare (Levasseur). După Knapp: „corpul demografiei a fost creat de S., sufletul ei de Halley, iar spiritul ei de Malthus". În disputa dintre statistica descriptivă, cea „tabelară" şi „aritmetica politică", S. se situează pe poziţiile acesteia din urmă. A analizat o importantă cantitate de date statistice şi a formulat regularităţi şi legi în producerea evenimentelor demografice, dar interpretarea pe care a dat-o este teologică. Printre primii care a arătat efectul legii numerelor mari la construcţia unei tabele de mortalitate. A exercitat o influenţă profundă asupra evoluţiei ulterioare a statisticii şi demografiei.

Op.: „Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechts aus der Geburt, dem Tode und der Fortpflanzung desselben erwiesen", ed. I, 1741, ed. II, 2 vol., Berlin, 1761—1762; Monografia franceză: Johann Peter Süssmilch (1707—1767). „L'ordre divin aux origines de la démographie", traduction originale, avec les études et commentaires rassemblés par Jacqueline Hecht, vol. I şi II, Institut National d'Etudes Démographiques, Paris, 1979.

ŞUŢU, Nicolae (1798—1871). Economist şi statistician român; om politic. Studii de economie politică pe baza datelor statistice, fiind considerat primul nostru economist. A cules o mare cantitate de date statistice pe care le-a analizat, ajungînd să caracterizeze multilateral situaţia economică a Moldovei. Membru al lui Royal Statistical Society.

Op.: „Aperçu sur les causes de la gêne et de la stagnation du commerce et sur les besoins industriels de la Moldavie", Iassy, 1838;

„Notions statistiques sur la Moldavia", Ed. Henning, 1884, XII + 184 p (tradusă de Teodor Codrescu sub titlul „Noțiții statistice asupra Moldaviei" Iași, 1852); "Quelques observations sur la statistique de la Roumanie", 1867; „N. Șuțu. Opere economice" (traducerea și îngrijirea lui Ion Veverca), 1957 (cuprinzînd viața și opera economică și cele trei lucrări publicate, traduse în limba română).

THIELE, Thorvald Nicolai (1838—1910). Matematician probabilist și statistician danez. Licențiat în astronomie (1860) și doctor în matematici (1866). A lucrat la Observatorul din Copenhaga; în 1875 a fost numit profesor de astronomie și director al Observatorului din Copenhaga (pînă în 1907). În 1895 a devenit membru corespondent al Institutului Actuarilor din Londra; a întemeiat Societatea Daneză a Actuarilor și a fost președintele acesteia pînă la moarte.

Preocupat în special de matematici aplicate, **Th.** a adus contribuții importante în diferite domenii. În teoria repartițiilor asimetrice, el a definit cumulanții și a studiat proprietățile lor. A inventat forma canonică a modelului liniar și a redus modelul liniar general cu ajutorul unei transformări ortogonale, creînd o justificare nouă pentru metoda celor mai mici pătrate. S-a ocupat și de analiza variației. A dezvoltat modelul liniar adăugîndu-i o mișcare browniană și a îmbogățit teoria estimației și predicției pentru această serie cronologică.

Op.: "En mathematisk Formel for Dodeligheden", Copenhaga, 1871; „Sur la compensation de quelques erreurs quasisystématiques pour la méthode des moindres carrés", Copenhaga, 1880.; "Theory of Observations", Londra, 1903; "Adjustement of tables of mortality", în „Aktuaren" 1—10, 1904; „Interpolationsrechnung", Teubner, Leipzig, 1903.

Despre **Th.**: A. Hald, "T. N. Thiele's Contributions to Statistics", în "International Statistical Review", volume 49,Number 1, April 1981, 1—20.

URSEANU, Veniamin (1914—1983). Statistician și economist român. Studii la Universitatea din Cernăuți; licențiat al Facultății de Drept; doctor în economie, la Academia de Studii Economice (1971), cu teza: „Metode statistice de cercetare și mișcare a prețurilor". Magistrat (1939—1949); statistician la Direcția Centrală de Statistică (1950—1953), redactor principal la „Revista de Statistică", organ al Direcției Centrale de Statistică (1953—1963); cercetător la Institutul de Matematică al Academiei R.S.R. (1963—1964); cercetător principal și șef al sectorului de economie matematică la Centrul de Statistică Matematică al Academiei R.S.R. (1964—1977). Membru al Consiliului Științific al Direcției Centrale de Statistică. Domenii: matematica economică, statistica economică, statistica financiară, modele matematice ale prețurilor, teoria indicilor, cu referire specială la indici teritoriali, manuale de statistică. Colaborator al acad. Gh. Mihoc, cu care a publicat lucrări de matematici aplicate în statistică, teoria sondajului, probabilități și statistică matematică. Opera științifică cuprinde cîteva zeci de titluri (cărți, comunicări la sesiuni științifice, articole, recenzii). A tradus lucrări de statistică din l. engleză și rusă.

Op.: „Statistica financiară" (cu pseudonimul V. Martin, în colab. cu A. Anatol), Tratat, 1956; „Elemente de statistică matematică și aplicațiile ei" (sub red. lui V. Urseanu, în colab. cu I. Marinescu, C. Moineagu, R. Niculescu și N. Ran

cu), 1966 (trad. rusă 1970); „Elemente de teoria probabilităților și aplicațiile ei" (în colab. cu Gh. Mihoc, M. Iosifescu), 1966; „Matematici aplicate în statistică" (în colab. cu Gh. Mihoc), 1962; „Metode statistice de cercetare a legităților de formare și mișcare a prețurilor", 1973 (premiul „Simion Stoilov" al Academiei R.S.R.); „Procese stohastice" (în colab. cu Gh. Mihoc, C. Bergthaller), 1978; „Modele de analiză statistică" (în colab. cu Gh. Mihoc și E. Ursianu), 1982.

YULE, George Udny (1871—1951). Statistician și matematician englez. Colaborator al lui *K. Pearson**. Profesor la Cambridge; președinte al lui Royal Statistical Society; membru al lui Royal Society. Contribuții la teoria corelației parțiale a caracteristicilor calitative. Autor al unui manual clasic de teorie a statisticii. Cea mai importantă contribuție este la dezvoltarea teoriei seriilor dinamice.

Op.: în colab. cu Kendall M., "An Introduction to the Theory of Statistics", 1911, Griffin, ed. XIV, Londra, 1960 (în rom.),„Introducere în teoria statisticii' (după ed. XIV). Ed. științifică, București, 1969; "On the Methods of Measuring Association Between Two Attributs", în "Journal of the Royal Statistical Society", 1912; "The Growth of Population and the Factors Which Control It", 1925.

WALD, Abraham (1902—1950). Matematician, statistician american. Născut la Cluj-Napoca, s-a stabilit la Viena (1930), iar în 1938 cu sprijinul lui *H. Hotelling**, a emigrat în S.U.A. unde a fost profesor la Universitatea Columbia. Continuator al activității științifice a lui *R. A. Fisher**, *J. Neyman** și *E. S. Pearson**. Contribuții importante la teoria deciziei statistice; creator al analizei secvențiale.

Op.: "Sequential Analysis", John Wiley, New York, 1947; "Statistical Decision Functions", John Wiley, New York, 1950.

WESTER GAARD, Harald (1853—1936), Statistician, demograf și economist danez. Profesor la Universitatea din Copenhaga. A contribuit la aplicarea metodelor matematice în demografie. Autorul unei istorii a statisticii.

Op.: „Die Lehre von der Mortalität und Morbidität", 1882; „Die Grundzüge zur Theorie der Statistik", 1890; "Contributions to the History of Statistics", 1932.

WILLCOX, Walter Francis (1861—1964). Statistician, demograf, economist și sociolog american. Studii la Columbia University și la Universitatea din Berlin. Profesor de economie și statistică la Cornell University (1891—1931). Președinte de onoare al Institutului Internațional de Statistică. A publicat lucrări de statistică demografică și socială. Autor al unei estimații clasice a populației globului.

Op.: "Increase in the Population of the Earth and of the Continents since 1650", în *International Migrations*, vol. II, 1931; "Introduction to the Vital Statistics of the United States 1900 to 1930" (1933), în "Studies in the American Demography", 1940.

WINKLER, Wilhelm (1884—1984). Statistician și demograf austriac. Doctor în drept la Universitatea germană din Praga (1907), abilitat în 1922 la Universitatea din Viena, doctor honoris causa al universităților din München și Viena. Activitate ca statistician la Oficiul Federal de Statistică al Austriei. Profesor la Universitatea din Viena. Membru al Academiei Austriece de Științe; fondator și președinte de onoare al Societății Austriece de Statistică. Membru al Institutului Internațional de Statistică (din 1927), membru de onoare al I.I.S.; membru al Uniunii Internaționale pentru studiul științific al populației (din 1948). Studii, cercetări, manuale de statistică, demografice, econometrie. A introdus (împreună cu H. Hyrenius) disciplina numită „demometrie".

Op.: „Grundriss der Statistik", în „Theoretische Statistik", Berlin, 1931; "Gesellschaftsstatistik", Berlin, 1933; „Grundriss der Statistik", Bd. I și II, 1948; „Mehrsprachiges demographisches Wörterbuch", Deutsche Ausgabe, Augsburg, 1960; „Demometrie", Berlin, 1969.

XX INSTITUȚII ȘI PUBLICAȚII DE STATISTICĂ

Institutul Internațional de Statistică *(International Statistical Institut; Institut International de Statistique)*. Organ internațional cu caracter științific în domeniul statisticii, înființat în anul 1885, cu sediul permanent la Haga (Olanda). Potrivit statutelor sale (1983), „Institutul este o societate autonomă având drept obiectiv dezvoltarea și îmbunătățirea metodelor statistice și a aplicațiilor acestora pe plan mondial..." (art. 1). Pentru realizarea acestui obiectiv I.I.S. încurajează asocierea internațională a statisticienilor, schimbul de cunoștințe profesionale în rîndul statisticienilor, dezvoltarea interesului colectiv pentru aceste cunoștințe, organizarea de relații între societățile de statistică și alte organe oficiale și neoficiale ce au interese statistice, se preocupă de formarea unor statisticieni competenți, încurajează cercetarea științifică în statistică, se preocupă de asigurarea comparabilității internaționale în statistică".
I.I.S. are în componența sa ca secții: Societatea Bernoulli pentru Statistică Matematică și Probabilități (Bernoulli Society for Mathematical Statistics and Probability, prescurtat B.). Asociația Internațională de Calcul Statistic (International Association for Statistical Computing, prescurtat I.A.S.C.) și Asociația Internațională a Statisticienilor de Anchete (International Association of Survey Statisticians, prescurtat I.A.S.S.). Membrii I.I.S. sînt individuali: membri de onoare, membri titulari și membri *ex officio*. Numărul membrilor de onoare și al membrilor titulari se ridică la 1220 (ianuarie 1983). Din rîndul membrilor de onoare fac parte (alfabetic): M. S. Bartlett (n. 1910; Anglia); H. Cramér (n. 1893; Suedia); M. H. Hansen (n. 1910; S.U.A.); A. N. Kolmogorov (n. 1903; U.R.S.S.); O. Onicescu (1892—1983; România); C. R. Rao (n. 1920; India); A. Sauvy (n. 1898; Franța); W. Winkler (1884—1984; Austria). Membrii I.I.S. *ex officio* sînt reprezentanți ai O.N.U., F.A.O., B.I.T., UNESCO, O.M.S., Banca Mondială și șefii birourilor (oficiilor) naționale de statistică, reprezentînd 96 țări. Următoarele societăți fac parte din categoria organizațiilor internaționale afiliate ale I.I.S.; Biometric Society, Econometric Society, Inter American Statistical Institute, International Association for Research in Income and Wealth; Union Internationale pour l'Étude Scientifique de la Population; Institut of Mathematical Statistics; European Organization for Quality Control; American Society for Quality Control; International Actuarial Association și Institute of Management Sciences. Este de asemenea afiliată International Association for regional and urban statistics.

De asemenea sînt afiliate la I.I.S. 27 organizaţii statistice naţionale, printre care : Société de Statistique de Paris, Indian Statistical Institute, Royal Statistical Society, American Statistical Association.
Structura organizatorică a I.I.S. : Adunarea generală a membrilor ce se întruneşte cu prilejul sesiunilor bianuale ; Consiliul, Comitetul executiv şi Oficiul permanent al I.I.S. (cu sediul la Voorburg, lîngă Haga). Forma principală de activitate ştiinţifică a I.I.S. sînt sesiunile ştiinţifice, publicaţiile, diferitele comitete speciale. La sesiunea din Buenos Aires (1981) s-a decis înfiinţarea unui Centru Internaţional de Cercetări Statistice (International Statistical Research Center). În cursul existenţei I.I.S. s-au ţinut 44 de sesiuni ; în perioada postbelică (din 1948) sesiunile se ţin la doi ani ; cele mai recente : Buenos Aires (1981), a 44-a sesiune, la Madrid (1983).
I.I.S. este condus de un comitet executiv compus din preşedinte, preşedinte ales şi trei vicepreşedinţi. În 1983 : preşedinte — E. Cansado (Spania) ; preşedinte ales (din septembrie 1983) J. Durbin (Anglia). Ultimii preşedinţi de onoare ai I.I.S. : Jerzy Neyman (1894—1981) şi sir Maurice Kendall (1907—1983).
Principalele publicaţii ale **I.I.S.** :
International Statistical Review (Revue Internationale de Statistique).. cu trei apariţii anuale. Întemeiată în anul 1932. Iniţial, se intitula "Review of the International Statistical Institute" („Revue de l'Institut International de Statistique"). editată de **I.I.S.** (Haga). În această formă a apărut 39 de ani (vol. 1—39). În 1972 (vol. 40, nr. 1), titlul se schimbă în "International Statistical Review" („Revue Internationale de Statistique"), cu subtitlul "A Journal of the International Statistical Institute" (ed. Longman—Londra). În anul 1981 (vol. 49, nr. 1) revista îşi modifică titlul şi programul tematic, devenind "International Statistical Review" („Revue Internationale de Statistique", cu subtitlul "A Journal of the International Statistical Institute . Bernoulli Society for Mathematical Statistics and Probability, International Association for Statistical Computing, International Association of Survey Statisticians". Editor : O. E. Barndorff — Nielsen (Danemarca), cu 10 editori asociaţi din diferite ţări. Revista publică : a) comunicări ştiinţifice originale, de interes general ; b) studii cuprinzînd analize critice integrate referitoare la anumite domenii ale statisticii şi calculului probabilităţilor sau expuneri referitoare la cele mai importante dezvoltări metodologice ; c) studii privind istoria statisticii şi a probabilităţilor. Revista mai publică note şi informaţii referitoare la progresele calculatoarelor în statistică, programe de anchete, metode de învăţămînt etc.
Bulletin of the International Statistical Institute (Bulletin de l'Institut International de Statistique), cu subtitlul "Proceedings of the ... th session" („Actes de la ... e session"), prezentate în volume şi fascicole (livraisons). Fiecare volum al buletinului cuprinde comunicările prezentate la sesiunile **I.I.S.** ; acesta este editat de ţara care a găzduit sesiunea respectivă. Ultimul buletin "Bulletin of the International Statistical Institute". Proceedings of the and session. Volume XLIX, Books 1, 2, 3), 1981, Buenos Aires, (tipărit în 1983). **I.I.S.** a editat, în decursul istoriei sale şi alte publicaţii.

Istoric. Primele forme organizate de cooperare internaţională în statistcă au fost *Congresele internaţionale de statistică*, în număr de nouă (1853—1877). Primul a fost iniţiat de *A. Queletet* şi *W. Farr* şi a avut loc la Bruxelles (1853). Celelalte congrese : Paris (1855) ; Viena (1857) ; Londra (1860), Berlin (1863), Florenţa (1867), Haga (1869), Petersburg (1872) şi Budapesta (1877). Principala

preocupare a congreselor a fost asigurarea comparabilității internaționale a datelor statistice, adoptarea de recomandări privind recensămintele de populație etc.

În 1884, cu prilejul sărbătoririi a 50 de ani de la înființarea lui Royal Statistical Society, guvernul englez a invitat guvernele străine să trimită (în 1885, la Londra) reprezentanți spre a se înființa o asociație internațională de statistică, idee promovată în special de *W. Farr**. Astfel a luat ființă Institutul Internațional de Statistică. Inițial, sesiunile (congresele) aveau loc la doi ani: Roma (1887), Paris (1889), Viena (1891), periodicitate întreruptă de cele două războaie mondiale.

I.I.S. a jucat un rol deosebit în promovarea statisticii, în elaborarea de instrumente de comparabilitate internaționale (nomenclatoare, clasificări etc.), în aplicarea de noi metode, în cooperarea internațională în domeniul statisticii. România a participat atît la congresele internaționale de statistică cît și la activitatea **I.I.S.** (→ *Istoricul statisticii din România*). În 1983, numărul membrilor români al **I.I.S.** se ridica la 8, din care: un membru onorar, un membru *ex officio* și 6 membri titulari.

Bibliografie

1. Zahn, Fr., *50 années de l'Institut International de Statistique*. I.I.S. München, 1934.
2. Nixon J. W., *A History of the International Statistical Institute*, 1885—1960, I.I.S., The Hague, 1960.
3. *Dicționar statistic-economic*, D.C.S., 1969, p. 493, 516—517.

Societatea Americană pentru Controlul Calității (*American Society for Quality Control*). Fondată în 1946, sediul: Milwaukee (S.U.A.). Număr de membri: peste 20 000; sesiuni anuale.
Publicații: „Industrial Quality Control" din 1944); alte publicații ocazionale.

Asociația statistică americană (*American Statistical Association*). Fondată în 1839, sediul: Washington (S.U.A.). Număr de membri: peste 9000; sesiuni anuale. *Publicații*: „The American Statistician' (din 1947, 5 apariții anuale); „Journal of the American Statistical Association" (din 1959; trimestrial); „Technometrics" (din 1959, trimestrial) editată împreună cu The American Society of Quality Control.

Institutul de statistică matematică (*The Institute of Mathematical Statistics*). Fondat: 1935, sediul: University of North Carolina, Chapel Hill (S.U.A.); sesiuni anuale. *Publicații*: „Annals of Mathematical Statistics" (din 1930 trimestrial). Alte publicații ocazionale.

Societatea mexicană de geografie și statistică (*Sociedad Mexicana de Geografia y Estadistica*). Fondată, în 1833, sediul: Mexico; sesiuni săptămînale. *Publicații*: „Boletin de la Sociedad Mexicana de Geografia y Estadistica" (din 1839 bimensuală).

Institutul Indian de Statistică (*Indian Statistical Institute*). Fondat în 1931 sediul: Calcutta (India). *Publicații*: „Sankhya: The Indian Journal of Statistics' (din 1933, trimestrial).

Societatea germană de statistică (*Deutsche Statistische Gesellschaft*). Fondată în 1911, reorganizată în 1948, sediul: Wiesbaden (R.F.G.). *Publicații*: „Allgemeines Statistisches Archiv" (din 1980, trimestrial). Publicații ocazionale.

Uniunea statisticienilor germani de orașe (*Verband Deutscher Städterstatistisker*) Fondată în 1878 (1903), sediul: Braunschweig. *Publicații*: Statistisches Jahrbuch deutscher Gemeinden (din 1890, anual).

Societatea belgiană de statistică (*Société Belge de Statistique*; *Belgische Vereniging voor de Statistiek*). Fondată în 1937, sediul : Bruxelles (Belgia). *Publicații* : „Revue belge de Statistique et de Recherche opérationnelle".
Societatea daneză de statistică (*Dansk Statistisk Forening*). Fondată în 1922, sediul : Copenhaga.
Societatea de Statistică din Paris (*Société de Statistique de Paris*). Fondată : 1860, sediul : Paris. *Publicații* : „Journal de la Société de Statistique de Paris" (din 1860, trimestrial).
Societatea de statistică și cercetări sociale din Irlanda (*Statistical and Social Inquiry Society of Ireland*). Fondată în 1847, sediul : Dublin. *Publicații* : „Journal of the Statistical and Social Inquiry Society of Ireland" din 1847, anual)
Societatea italiană de economie, demografie și statistică (*Società Italiana di Economia, Demografia e Statistica*). Fondată în 1938, sediul : Roma. *Publicații* : „Rivista italiana di economia, demografia e statistica" (din 1947. trimestrial).
Societatea de statistică (*Vereniging voor Statistiek*). Fondată în 1945, sediul : Roterdam (Olanda). *Publicații* : „Statistica Neerlandica" (din 1946, trimestrial).
Institutul statisticienilor (*Institute of Statisticians*). Fondat în 1948, sediul : Londra. *Publicații* : „The Statistician" (din 1950. trimestrial).
Societatea de Statistică din Manchester (*Manchester Statistical Society*). Fondată în 1833, sediul : Manchester. *Publicații* : „Transactions of the Manchester Statistical Society" (din 1833, anual).
Societatea Regală de Statistică (*Royal Statistical Society*). Fondată în 1834, sediul : Londra. Număr de membri : circa 4000 membri. *Publicații* : „Journal of the Royal Statistical Society", Series A (General), (din 1838; trimestrial) : „Applied Statistics" din 1952, o dată la patru luni) : „Journal of the Royal Statistical Society", Series B (Methodological), (din 1934, de două ori pe an).
Societatea elvețiană pentru statistică și economie (*Schweizerische Gesellschaft für Statistik und Volkwirtschaft*; *Société Suisse de Statistique et d'Economie Politique*). Fondată în 1864, sediul : Berna. *Publicații* : „Schweizerische Zeitschrift für Volkswirtschaft und Statistik" (din 1865; trimestrial).
International Actuarial Association.
International Union for the Scientific Study of Population.
Econometric Society.
Inter-American Statistical Institute.
Biometric Society.
International Association for Research in Income and Wealth.
The Institute of Management Sciences.
International Federation of Operational Research Societies.
European Organisation for Quality Control.
International Association of Municipal Statisticians.
International Association of Survey Statisticians.

INDICE

A

	Pagina
Abatere cuartilă (interval intercuartil, intercuartilă)	164
Abatere mediană (abatere probabilă, abatere echiprobabilă)	166
Abatere medie (abatere absolută medie)	165
Abatere pătratică medie (abatere standard, abetere tip)	166
Abatere medie pătratică (→ şi abaterea pătratică medie)	311
Ajustarea seriei cronologice	247
Ajustarea seriei cronologice cu ajutorul curbei Gompertz	275
Ajustarea seriei cronologice cu ajutorul curbei logistice	279
Ajustarea seriei cronologice cu ajutorul dreptei	252
Ajustarea seriei cronologice cu ajutorul exponenţialei	264
Ajustarea seriei cronologice cu ajutorul exponenţialei modificate	274
Ajustarea seriei cronologice cu ajutorul parabolei	258
Amplitudine (interval de variaţie)	163
Amplitudinea clasei	124
σ — algebră	302
Analiză armonică a seriilor cronologice	293
Analiza seriilor cronologice	239
Analiza spectrală a seriei cronologice	295
Analiza tendinţei generale (seculare) a seriei cronologice	251
Analiza variaţiilor (fluctuaţiilor) ciclice ale seriei cronologice	292
Analiza variaţiilor (fluctuaţiilor) sezoniere ale seriei cronologice	281
Analiza de varianţă (Anova)	511
Analiză a periodogramei	293
Anchetă prin sondaj (statistic)	57
Anchetă prin sondaj (aleator)	58
Anchetă prin sondaj nealeator	58
Anchetă (sondaj) de opinie	58
Anchetă statistică	57
Autocorelaţie de ordine 1	294
Autocorelaţie de ordinul k	294

B

Bancă de date statistice	74
Bază de date statistice	73

	Pagina
Bază de metadate statistice	73
Bloc	512

C

Calculator electronic (în prelucrarea datelor statistice)	71
Canalul de transmisie a datelor statistice	70
Caracteristică operativă (funcția C.O.)	382
Caracteristică statistică	50
Cartodiagramă	120
Cartogramă	120
Cîmp de probabilitate	305
Celulă	511
Centile	154
Centrul (mijlocul) clasei	127
Chestionar	58
Ciclu economic (ciclu de afaceri)	297
Clasă	124
Clasă deschisă	124
Coada unei repartiții	316
Coeficient bugetar	235
Coeficient de concentrare (indicele lui Gini)	180
Coeficientul de corelație (→ și p. 197)	311
Coeficient de corelație (coeficient de corelație liniară coeficient de corelație Bravais-Pearson)	197
Coeficient de corelație biserial	205
Coeficient de corelație informațională (Onicescu)	38
Coeficient de corelație tetrachoric	205
Coeficient de regresie	193
Coeficient de variație	171
Coeficientul lui Fisher	176
Coeficientul lui Kendall	210
Coeficientul lui Spearman	208
Coeficientul lui Yule	175
Coeficienții β și γ	178
Coeficienții pearsonieni	175
Colectivitate statistică (→ și populație statistică)	49
Concepția analitică a indicelui	219
Concepția sintetică a indicelui	219
Condițiile lui Yule	159
Confounding	547
Conjunctură economică	297
Control statistic al calității	506
Control statistic de recepție (controlul loturilor de produse)	509
Contrast	540
Corelația lui Sheppard	161
Corelație (statistică)	181
Corelație (covarianță) (→ și p. 181)	311
Corelație aparentă (corelație falsă)	183
Corelație informațională (Onicescu)	37

	Pagina
Corelaţie multiplă	205
Corelaţie a rangurilor	208
Corelaţie serială a seriei cronologice	294
Corel ogramă	122
Covariaţă (→ şi p. 311)	196
Criterii de alegere a curbelor de ajustare a tendinţei generale (teste)	281
Cuantile (de ordinul n)	152
Cuantilă (α % — cuantilă)	316
α % — cuantilă bilaterală	317
α % — cuantilă inferioară	317
α % — cuantilă superioară	316
Cuartile	152
Cumulant de ordinul n	162
Curbă de concentrare (curbă Lorenz-Gini)	179
Curbă cumulativă (curbă de repartiţie, curbă a frecvenţelor cumulate)	315
Curbă de frecvenţă	315
Curbă a frecvenţelor (curbă de repartiţie)	105
Curbele lui Pearson	179

D

Date statistice	18
Decile	152
Densitatea de probabilitate (a variabilei aleatoare) (→ şi p. 307; p. 315)	52
Descompunere a seriilor cronologice	245
Descriere statistică	18
Determinarea contribuţiei factorilor la modificarea unui indicator complex	233
Diagrama diferenţială	102
Diagrama integrală (cumulativă)	104
Diagramă Barlow-Campo	122, 456
Diagramă în batoane (bare)	85
Diagramă circulară	88
Diagramă prin coloane (diagramă prin tuburi de orgă)	86
Diagramă cumulativă (→ şi p. 104)	86
Diagramă logaritmică (diagramă log-log)	98
Diagramă polară (radială)	109
Diagramă semilogaritmică	91
Diagramă prin suprafeţe	89
Diagramă triunghiulară	111
Diagramă în Z	112
Diferenţă medie a lui Gini	172
Dispersie brută	163
Dispersie relativă	163
Dispersie de selecţie	366
Dispersie (a repartiţiei)	52
Dispersie (varianţă)	310

	Pagina
Dreaptă de regresie (dreaptă de estimare)	189
Dreapta de regresie a probiturilor	192

E

Echipamentul electronic central	69
Echipamentul pentru integrarea/utilizarea datelor statistice la nivel central	69
Echipamentul pentru prelucrarea datelor statistice la nivelul centralelor, departamentelor şi ministerelor	69
Echipamentul pentru prelucrarea datelor statistice la nivelul teritorial	69
Echipamentul pentru realizarea publicaţiilor statistice	69
Echipamentul pentru teletransmisia automată a datelor statistice	69
Econometrie	297
Ecuaţie de regresie (ecuaţie de estimaţie)	187
Ecuaţii normale	190, 193, 203, 207, 377
Efectul factorului (efectul tratamentului)	511
Eficienţă absolută	371
Eficienţă relativă	371
Energie informaţională (Onicescu)	37, 312
Entropie	131
Eroare accidentală	55
Eroare de genul întîi (I)	384
Eroare de genul al doilea (II)	380
Eroare sistematică	55
Eroare statistică	55
Estimator	367
Estimator absolut corect	367
Estimator consistent	367
Estimator corect	367
Estimator eficient (estimator uniform nedeplasat de regresie minimă)	368
Estimator nedeplasat	367
Estimator MMP (metoda celor mai mici pătrate)	377
Estimator suficient (exhaustiv)	372
Estimator de verosimilitate maximă	373
Estimaţie	372
Eşantion (de volum *n*)	362
Eşantion (statistic)	49
Eveniment (aleator)	300
Eveniment imposibil	300
Eveniment sigur	300
Evenimente dependente	305
Evenimente independente	305
Experiment	58, 362
Experiment aleator	362
Experiment Bernoulli	305

	Pagina
Experiment determinist	362
Extrapolare a seriei cronologice	296
F	
Factor	511
Factor (variabilă) calitativ(ă) (variabilă indicator)	512
Factor (variabilă) cantitativ(ă)	512
Factor controlat	512
Factor de eterogenitate (factor necontrolat)	512
Fenomen de masă	43
Flux informațional statistic	66
Flux informațional statistic cu canal unic	66
Flux informațional statistic cu canale multiple	66
Foarfecele prețurilor	237
Formă (a repartiției)	52
Formă canonică a ipotezelor fundamentale	515
Formula lui Bayes (probabilitățile cauzelor)	306
Formula Bortkiewicz	226
Formula condiționării repetate	306
Formula Gompertz-Makeham	279
Formula probabilității totale	306
Forța unui test	382
Frecvență	125
Frecvență dedusă	125
Frecvență relativă	125
Frecvență totală	125
Funcție caracteristică	311
Funcție empirică de repartiție (funcție de repartiție de selecție)	364
Funcție estimabilă	372
Funcție de fiabilitate	510
Funcție de frecvență (→ și p. 125)	307
Funcție parametrică	372
Funcție de repartiție (→ și p. 315)	308
Funcție de verosimilitate	375
G	
Grad de libertate	316
Grafic Gantt	113
Graficul lui Lexis	116
Grupare statistică	61
Grupare statistică analitică	62
Grupare statistică combinată	62
Grupare statistică simplă	62
Grupare statistică tipologică	
H	
Histogramă	99
Historiogramă	89

I

	Pagina
Indicator statistic	53
Indice	213
Indice agregat	222
Indice analitic	230
Indice aritmetic mediu ponderat	227
Indice armonic mediu ponderat	227
Indice cu bază fixă	223
Indice al costului vieţii	235
Indice Divisia	224
Indice Edgeworth (Marshal-Edgeworth)	224
Indice elementar (individual)	221
Indice general	222
Indice „Ideal" Fisher	224
Indice în lanţ	223
Indice Laspeyres	224
Indice median	223
Indice mediu	223
Indice mediu aritmetic	223
Indice mediu armonic	223
Indice mediu geometric	223
Indice modal	223
Indice Paasche	224
Indice parţial (de grup)	222
Indice ponderat	222
Indice cu ponderi constante	224
Indice cu ponderi variabile	224
Indice al preţurilor	234
Indice al preţurilor la consumator	235
Indice al puterii de cumpărare a monedei	237
Indice sezonier (de sezonalitate)	284
Indice sezonier ajustat	287
Indice sezonier simplu neajustat	284
Indice simplu	222
Indice simplu de covariaţie diferenţială	295
Indice simplu de covariaţie tendenţială	295
Indice sintetic (total, general)	221
Indice standardizat	232
Indice cu structură fixă	231
Indice cu structură variabilă	231
Indice de structură (al schimbărilor structurale)	231
Indice teritorial	233
Indice cu utilitate constantă	236
Indice al valorii medii	230
Indice al volumului fizic al producţiei (indice al cantităţilor)	236
Inferenţă statistică	18
Integrarea sistemului informatic statistic în sistemul informatic naţional	76
Interacţiune	511
Interpolarea seriei cronologice	251

Interval modal 155
Ipoteza nulă în modelul liniar 516
Ipoteză alternativă H_1 380
Ipoteză compusă 380
Ipoteză nulă H_0 380
Ipoteză neparametrică 380
Ipoteză parametrică 380
Ipoteză simplă 380
Ipoteză statistică 380

L

Lanț 314
Lanț cu legături complete 315
Lanț Markov (finit) 314
Lanț OM (Onicescu — Mihoc) 315
Lege dinamică (deterministă) 43
Lege de repartiție 305
Lege statistică (stohastică) 43
Legea amplitudinii studentizate 340
Legea amplitudinii studentizate mărite 341
Legea arc-sin 333
Legea Beta 330
Legea Beta multidimensională (Repartiția Dirichlet) 353
Legea binomial negativă (repartiția lui Pascal sau Polya) 320
Legea binomială $Bi(n, p)$ 320
Legea Cauchy 332
Legea exponențială 343
Legea F necentrată 339
Legea F Snedecor-Fisher 335
Legea formelor pătratice de variabile aleatoare normale 328
Legea Gamma 329
Legea geometrică 321
Legea Gram-Charlier 345
Legea Gram-Charlier, Edgeworth generalizată 357
Legea Helmert 334
Legea χ^2 (hi-pătrat) (Helmert-Pearson) 333
Legea χ^2 (hi-pătrat) necentrată 338
Legea hipergeometrică 322
Legea hipergeometrică generalizată 360
Legea (Prima...) a lui Laplace 325
Legea logaritmică 324
Legea logaritmului iterat 314
Legea lognormală 327
Legea maximului studentizat al abaterii față de media de selecție 341
Legea maximului studentizat al abaterii față de media teoretică 343
Legea maximului studentizat al modulului 340
Legea multinominală 359
Legea normală (legea Laplace-Gauss) 326

Pagina

Legea normală bidimensională 352
Legea normală multidimensională (n-dimensională) 350
Legea numerelor mari 45, 313
Legea Pareto 349
Legea Poisson 323
Legea Poisson multidimensională 361
Legea Rayleigh 345
Legea Rayleigh generalizată 345
Legea t a lui Student 334
Legea t necentrată 339
Legea t multidimensională (O. J. Dunn) 355
Legea T^2 Hotelling 355
Legea triunghiulară 325
Legea uniformă 325
Legea (discretă) uniformă 318
Legea Weibull 344
Legea Wishart 356
Legea Z (a lui Fisher) 337
Legi de repartiţie continue (repartiţia variabilei aleatoare continue) 325
Legi de repartiţie discretă (repartiţia variabilei aleatoare discrete) 318
Legi de repartiţie multidimensională (n-dimensionale) 350
Legi de repartiţie multidimensionale discrete 359
Legi de tip Pearson (I—XII) 348
Limitele clasei 124

M

Matrice de corelaţie (covarianţă) 311
Media de selecţie 365
Medială 151
Mediană 148
Medie 310
Medie (adevărata...) 511
Medie aritmetică 135
Medie aritmetică simplă 135
Medie aritmetică ponderată 135
Medie armonică 140
Medie armonică ponderată 142
Medie cronologică 158, 241
Medie cronologică simplă 158
Medie cronologică ponderată 159
Medie geometrică 143
Medie geometrică ponderată 144
Medie mobilă 242
Medie pătratică 147
Medie pătratică simplă 147
Medie pătratică ponderată 147
Metoda agregată a indicelui 234

Pagina

Metoda Calot 292
Metoda celor mai mici pătrate (→ și p. 190) 376
Metoda centrării mediilor mobile 290
Metoda minimului lui 379
Metoda momentelor 378
Metoda relativă a indicelui 234
Metoda S a lui H. Scheffé 540
Metoda T a lui J. W. Tukey 541
Metodă statistică 18
Metoda de ajustare analitică a seriei cronologice 249
Metode de ajustare grafică a seriei cronologice 247
Metode de ajustare mecanică a seriei cronologice 247
Metode de comparație multiplă (inferență statistică simultană) 539
Metode grafice pentru testarea concordanței repartiției empirice cu unele repartiții teoretice 454
Metode statistice 18
Microrecensământ 57
Mod 154
Model (statistic) 46
Model aditiv al seriei cronologice 246
Model de analiză de covarianță 536
Model I bifactorial ANOVA (clasificare ierarhică) (cuibărită) 527
Model I bifactorial ANOVA (clasificare încrucișată) 524
Model I bifactorial ANOVA cu număr egal de observații în celulă ($k > 1$) (clasificarea încrucișată) (model echilibrat) 525
Model I bifactorial ANOVA cu număr egal de observații în subgrupe ($k > 1$). clasificare cuibărită (ierarhică) 527
Model I bifactorial ANOVA cu observație în celulă 524
Model II bifactorial ANOVA (clasificare ierarhică) 536
Model II bifactorial ANOVA (clasificare încrucișată) 535
Model II bifactorial ANOVA cu observație in celulă (clasificare încrucișată) 535
Model II bifactorial ANOVA cu număr egal de observații în subgrupe 536
Model (clasificare) cuibărit(ă) (ierarhic(ă)) 513
Model determinist 47
Model (clasificare) încrucișat(ă) 512
Model liniar general 512
Model I ANOVA multifactorial formal (clasificare ierarhică) 532
Model I ANOVA multifactorial formal (clasificare încrucișată) 529
Model multiplicativ al seriei cronologice 246
Model simplu de analiză de covarianță 539
Model statistic (stochastic) → și p.46) 47

Pagina

Model I unifactorial ANOVA 519
Model II uniffactorial ANOVA 534
Modelul ANOVA cu efecte aleatoare (modelul II de analiză a varianței) 514
Modelul ANOVA cu efecte constante (modelul I de analiză a varianței) 514
Modelul mixt ANOVA (modelul mixt de analiză a varianței) 514
Moment absolut de ordinul m 310
Moment aleator de oprire 364
Moment centrat 161
Moment centrat absolut de ordinul m 310
Moment centrat de ordinul m 310
Moment empiric centrat (moment centrat de selecție) de ordin k 366
Moment empiric (moment de selecție) de ordin k 365
Moment de ordinul k al variabilei X în raport cu o valoare arbitrară x 160
Moment de ordinul m 310
Moment simplu 161
Momente 160
Momente uzuale 161

N

Netezirea serie cronologice 250
Nivel (al repartiției) 52
Nivelul de calitate admis (acceptat) 509
Nivelul (varianta) unui factor 511
Nomenclator statistic unitar 66

O

Observare curentă 55
Observare directă 55
Observare indirectă 55
Observare parțială 54
Observare periodică 55
Observare statistică (înregistrare statistică) 54
Observare totală (exhaustivă) 54
Observații perechi 402
Ogiva lui Galton 109
Operații cu evenimente 301
Operații cu variabile aleatoare 309
Organizarea centralizată (a prelucrării datelor statistice) 64
Organizarea descentralizată (a prelucrării datelor statistice) 64
Organizarea mixtă (a prelucrării datelor statistice) 64
Organizarea prelucrării datelor statistice 63

Pagina

P

Panel 58
Parametrii dispersiei 163
Parametrii formei (repartiției) 174
Parametru statistic 52
Pătrat latin 544
Pătrat latin standard 545
Perioadă de bază (de referință) 223
Perioada curentă 223
Periodogramă 293
Piramidă a vîrstelor 114
Plan în blocuri 545
Plan în blocuri întîmplătoare 546
Plan de control 509
Plan experimental 542
Plan experimental în blocuri incomplete 546
Plan experimental complet 543
Plan experimental factorial 544
Plan experimental incomplet 543
Plan experimental „split-plot" 543
Planificarea experimentelor 542
Poligon al frecvențelor 102
Pondere 222
Ponderi standard 233
Populație (statistică) 49
Postulatele corelației lui Pearson 202
Prag de semnificație 380
Predicatul tabelului statistic 77
Prelucrare automată a datelor recensămîntului populației și al locuințelor 74
Prelucrare automată a datelor statistice operative 74
Prelucrare centralizată 61
Prelucrare descentralizată 61
Prelucrare integrată a datelor demografice 75
Prelucrare primară 61
Prelucrare secundară 61
Prelucrarea datelor statistice 61
Prelucrarea datelor statistice cu ajutorul mijloacelor tehnicii de calcul 62
Preț absolut 235
Preț relativ 235
Principiul verosimilității maxime 372
Proba (testul, verificarea) lui Charlier 177
Probabilitate 32, 302
Probabilitate condiționată 305
Probabilitate subiectivă 33
Probabilitatea de acceptare P a unui lot 509
Proces aleator (stohastic) cu parametru continuu 315

	Pagina
Proces aleator (stohastic) cu parametru discret	314
Programare și sisteme de programare standard pentru prelucrarea dărilor de seamă și fenomenelor statistice în mod lot	72
Programe și sisteme de programe standard pentru prelucrarea datelor statistice	72
Programe standard pentru prelucrarea dărilor de seamă/formularelor statistice în mod interactiv	72
Puterea unui test	381

R

Racordare a indicilor	237
Randomizare	546
Randomizare completă	547
Randomizare parțială	547
Rapoarte în lanț	292
Raport de corelație (indice de corelație)	202
Rata de defectare	510
Recensământ	56
Recensământ economic	57
Recensământ experimental (de probă)	57
Recensământ al populației	56
Registru statistic automat	68
Registru (interval) de acceptare	381
Regiune (interval) de respingere	381
Regresie	182
Relații cauzele statistice (legături cauzale)	44
Repartiție (a unei variabile aleatoare) (→ și p. 127, 304 și 305)	308
Repartiție (de probabilitate)	304
Repartiție aplatizată (boltită)	176
Repartiție condiționată (repartiție legată)	195
Repartiție de frecvențe	127
Repartiție leprocurtică	176
Repartiție marginală	194
Repartiție platicurtică	176
Repartiție statistică	52
Repartiții pluvimodale (multimodale)	157
Repartiții unimodale	157
Reprezentare grafică	81
Rețeaua probabilistică	85, 454
Riscul beneficiarului	509
Riscul erorii de genul I	381
Riscul erorii de genul II	381
Riscul erorilor asociate unui test statistic	381
Riscul furnizorului	509

Pagina

S

Schemă aditivă 233
Schemă multiplicativă 233
Selecţie 363
Selecţie bernoulliană (selecţie cu revenire sau sondaj cu revenire) 363
Selecţie nebernoulliană (selecţie fără revenire sau sondaj fără revenire) 363
Serie cronologică (serie în timp, serie temporală, cronică) 239
Serie cronologică de intervale 240
Serie cronologică de momente 240
Serie de indici 223
Serie statistică (repartiţie statistică) 123
Serie statistică de repartiţie 123
Serie statistică spaţială 123
Sistem informatic statistic 68
Sistem informatic statistic integrat 71
Sistem informaţional statistic 65
Sisteme de gestiune a bazelor de date statistice 72
Sistemul informatic naţional 76
Sistemul programelor de analiză a datelor statistice 70
Sistemul programelor pentru crearea/întreţinerea, livrarea, copierea pe suport magnetic a nomenclatoarelor unitare statistice 70
Sistemul programelor de gestiune a bazelor de date/metadate statistice 70
Sistemul programelor pentru prelucrarea automată a datelor statistice în loturi 70
Sistemul programelor pentru teleprelucrarea datelor statistice operative şi curente 70
Sondaj 363
Sondaj aleator 363
Sondaj dirijat 363
Sondaj din materiale în vrac 364
Sondaj mixt 363
Sondaj cu revenire (selecţie barnoulliană) 363
Sondaj fără revenire (selecţie nebernoulliană) 363
Sondaj secvenţial 364
Sondaj stratificat simplu fără revenire 363
Sondaj stratificat (tipic) fără revenire în două faze 363
Sondaj stratificat fără revenire proporţional 363
Spaţiu de evenimente elementare 300
Spaţiul q-dimensional al funcţiilor estimabile 540
Spaţiul eşantioanelor 362
Spaţiul parametrilor 364
Statistică 17
Statistică (m-dimensională) 364
Statistică actuarială 42

Pagina

Statistică analitică (inductivă sau inferențială) 40
Statistică descriptivă 36
Statistică economică 39
Statistică informațională Onicescu 37
Statistică matematică 34
Statistică de ramură 41
Statistică teoretică (teorie a statisticii) 41
Stereogramă 122
Subiectul tabelului statistic 77
Subsistemul automatizării pregătirii publicării informației statistice 68
Subsistemul bazelor de date statistice 68
Subsistemul metainformatic statistic 68
Subsistemul modelelor matematice macrocconomice statistice 68
Subsistemul prelucrării datelor recensămintelor și anchetelor statistice 68
Subsistemul prelucrării datelor statistice operative și curente 68
Sumă a pătratelor datorată (pentru verificarea) ipotezei nule H_0 517
Sumă a pătratelor erorilor 517
Sumă totală a pătratelor (sumă totală de pătrate) 517

T

Tabel ANOVA, 519
Tabel de asociație 80
Tabel de corelație 183
Tabelul cu dublă intrare (→ și p. 183) 78
Tabel statistic 76
Tabel statistic combinat 78
Tabel statistic pe grupe 78
Tabelul statistic simplu (tabel cu simplă intrare) 77
Tabelă (statistică) 80
Tendință generală (tendință secundară) 245
Teorema de acoperire simultană 540
Teorema limită centrală 313
Teoria deciziilor statistice 34
Teoria economică a indicelui 219
Teoria estimației 34
Teoria probabilităților 31
Teoria sondajului 34
Teoria statistică a indicelui 219
Teorie statistică 18
Test bilateral 385
Test circular al indicelui 228
Test privind coeficientul de corelație pe transformata Z a lui Fischer 467
Test de concordanță 384

Pagina

Test de concordanță cu legea binominală $Bi(x, p)$ 477
Test de concordanță Cramér-von Mises 440
Test de concordanță Finkelstein-Schafer 439
Test de concordanță Gnedenko 439
Test de concordanță χ^2 (hi-pătrat) (caz neparametric) 431
Test de concordanță χ^2 (hi-pătrat) (caz parametric) 426
Test de concordanță Kolmogorov-Smirnov 435
Test de concordanță Lilliefors 436
Test de determinare a indicelui 228
Test pentru egalitatea a k parametri p_i a k legi binomiale $Bi(n_i, p_i)$ 491
Test χ^2 (hi-pătrat) 387
Test de identitate a indicelui 228
Test de independență 466
Test nedeplasat 386
Test neparametric 386
Test neparametric Mann-Whitney-Wilcoxon 406
Test von Neumann (Kent-Bellinson-Hart) 461
Test de omogenitate 386
Test de omogenitate (comensurabilitate) a indicelui 228
Test pentru parametri p_1, p_2 a două legi binominale $Bi(n_i, p_i)$, $i = 1,2$ 487
Test de proporționalitate a indicelui 228
Test al raportului de verosimilități 386
Test de reversibilitate a factorilor indicelui 228
Test de reversibilitate în timp a indicelui 228
Test robust 386
Test al schimbării bazei indicelui 228
Test secvențial (progresiv) al lui Wald 386
Test de semnificație 384
Test de semnificație pentru dispersie 409
Test de semnificație privind egalitatea a două dispersii 412
Test statistic 382
Test statistic referitor la egalitatea a doi coeficienți de corelație 468
Test statistic referitor la egalitatea a k coeficienți de corelație 470
Test statistic privind egalitatea a k dispersii ($k > 2$) 414
Test statistic privind egalitatea a 2 medii 394
Test statistic privind egalitatea a 2 medii — dispersiile cunoscute (testul Z) 394
Test statistic privind egalitatea a medii — dispersiile σ_1 și σ_2 necunoscute (testul t) 396
Test statistic privind egalitatea a k medii 409
Test t 393
Test uniform cel mai puternic 386
Test unilateral (dreapta sau stînga) 384
Test pentru o valoare specificată p_0 a parametrului legii binomiale $Bi(n, p)$ 480

Pagina

Test pentru verificarea normalității 441
Test Z 387
Testarea concordanței cu legea Poisson 472
Teste privind coeficientul de corelație 466
Teste de concordanță 426
Teste privind egalitatea probabilităților multinomiale p în două sau mai multe experimente 495
Teste de independență (de asociere) în tabele de contingență 500
Teste ale indicilor 227
Teste ale indicelui elementar (individual) 228
Teste ale indicelui sintetic 229
Teste de normalitate Shapiro-Wilk, Shapiro-Francia De Wet-Venter, D'Agostino 450
Teste de omogenitate generală 420
Teste privind parametrii legii multinomiale 493
Teste privind repartiția binomială 477
Teste privind repartiția Poisson 472
Teste privind valorile specificate ale probabilităților p_i 493
Teste privind valorile specificate ale probabilităților p_{ij} și probabilităților marginale p_i. p_j în tabele de contingență 498
Teste privind variabila aleatoare — $Po(\lambda)$ 473
Teste privind verificarea caracterului aleator 451
Teste pentru verificarea normalității prin măsurarea asimetriei și excesului 448
Testul D'Agostino 452
Testul lui Barlett 414
Testul lui Cochran 420
Testul exact al lui Fisher pentru verificarea independenței în tabele de contingență 502
Testul Z al lui Fisher 412
Testul lui Hartley 416
Testul ipotezei nule 518
Testul iteraților 460
Testul Lambda 420
Testul lui Massey 442
Testul L Neyman-Pearson 425
Testul de normalitate al lui Lilliefors 441
Testul de normalitate Sarkady 451
Testul lui Sarkady 449
Testul semnelor 460
Testul Shapiro-Wilk(test W) 444
Testul F al lui Snedecor — Fischer 412
Transformata lui Fisher 197
Tratament 511

U

Unitate de informatică statistică 65
Unitate statistică (individ) 49

Pagina

V

Valori medii 309
Valori medii condiționate 312
Variabilă aleatoare 51, 306
Variabilă aleatoare discretă 307
Variabilă aleatoare continuă 307
Variabilă calitativă 50
Variabilă cantitativă 50
Variabilă concomitentă (covarianță) 536
Variabilă continuă 50
Variabilă dependentă 47
Variabilă discretă 50
Variabilă endogenă 47
Variabilă exogenă 47
Variabilă explicată 47
Variabilă explicativă 47
Variabilă independentă 47
Variabilă statistică 50
Variabile aleatoare independente 313
Varianță (dispersie sau fluctuație) 52, 171, 310
Variația concomtentă a seriei cronologice 295
Variația ciclică (mișcare ciclică) 245
Variație periodică (mișcare periodică) 246
Variație reziuală(mișcare aleatoare sau accidentală) 246
Variație sezonieră (mișcare sezonieră) 246
Vector aleator (*n*-dimensional) 307
Verificarea (testarea) ipotezelor statistice 34, 38

Redactor: MARIA BLAGA. Tehnoredactor: OLIMPIU POPA. Coli de tipar: 38,25. Bun de tipar 20.02.1985



C. 680 – I. P. INFORMAȚIA
Str. Brezoianu Nr. 23-25
București