

Gheorghe Caralicea–Mărculescu

Daniela -Emanuela Dănăcică

Dragoș-Pătru Covei

STATISTICA PENTRU EDUCAȚIE FIZICĂ ȘI SPORT

Referenți Științifici:

Tehnoredactare computerizată:

Daniela- Emanuela Dănăcică

Dragoș-Pătru Covei

Coperta:

Costin Covei

Editura: *AGER*

ISBN: 973-8086-00-0

2005

Prefață

Lucrarea aprofundează programa analitică a cursului de statistică ținut la secția de educație fizică din cadrul Universității Constantin Brâncuși din Tg-Jiu și este structurată pe 7 capitole având în mare parte un caracter elementar, conținând noțiuni și rezultate indispensabile unui viitor profesor de educație fizică și sport din învățământul preuniversitar.

Ca aplicații la partea teoretică am selectat un anumit număr de exerciții cărora le prezentăm soluții complete.

Dorim ca lucrarea să fie un pretext pentru cititorul interesat să-și cristalizeze propriile opțiuni în abordarea acestui domeniu, iar, pe de altă parte, un argument pentru studiul aprofundat al statisticii, care prin ramurile sale relativ disparate, mai mult sau mai puțin abstracte, reușește o modelare naturală eficientă a diferitelor fenomene ale realității.

Capitolul I

Statistica – instrument de cunoaștere a fenomenelor sociale

Statistica constituie o componentă esențială a pregătirii profesionale a oricărui om fiind considerată o știință având ca obiect fenomenele și procesele sociale.

1. Rolul statisticii în studiul fenomenelor sociale

Pornind de la considerentul că știința constituie factorul primordial al progresului, că era în care trăim este tot mai mult guvernată de valorificarea și implementarea în practică a descoperirilor științifice, de ansamblul revoluției tehnico-științifice, se poate afirma că un rol hotărâtor în definirea instrucției în învățământ revine științelor fundamentale. În edificiul acestora, una dintre cele mai vechi științe–alături de astronomie, și cu rădăcini adânc înfipte în necesitățile reale ale omului–plurivalenta și omniprezenta statistică se bucură de o atenție specială datorită și inepuizabilului său potențial formator.

În ceea ce privește profesiunile se poate spune că nu există domeniu de preocupare în care statistica să nu fie necesară, ba chiar în aproape toate categoriile de calificare ea a devenit indispensabilă.

Spre exemplu, în educația fizică aplicând un model pentru dezvoltarea rezistenței la proba de 3000 m alergare, timp de 20 de zile pe un lot de sportivi seniori, dintr-o tabără de pregătire, vom obține creșterea semnificativă a rezistenței. Această lege statistică confirmată de practică nu se manifestă obligatoriu în fiecare caz particular. Nu se exclude faptul ca la un subiect rezistența să scadă în cazul în care el este testat pe distanță mai mare. Creșterea rezistenței va fi o caracteristică ce depinde de fiecare individ, însă depinzând de un număr mare de factori, uneori necontrolați, ca gradul inițial de pregătire, modul cum tratează antrenamentul, regimul de odihnă și alimentar, particularități neuro–musculare și neuro–hormonale etc. Creșterea rezistenței nerealizându-se obligatoriu pentru fiecare subiect în parte, ci în majoritatea cazurilor. După studiul a suficiente cazuri particulare suntem în măsură să stabilim legitatea statistică.

Cartea își propune să înlăture ideea falsă că științele care se joacă cu cifre sunt un domeniu greu, inaccesibil omului de rând, că trebuie să ai predispoziție etc. Printr-o muncă

perseverență, oricine are cunoștințe de matematici elementare, le poate aborda și pe cele superioare. Statistica trebuie să rămână un bun pentru cei mulți nu numai pentru "elitele intelectuale". Statistica pătrunzând în mase, va contribui în largă măsură la ridicarea generală a nivelului intelectual, ridicând nivelul calificării profesionale a unei însemnate categorii de oameni ai muncii, învățători, profesori, tehnicieni etc..

La baza statisticii moderne sunt:

- ✓ *statistica practică* în sensul unor înregistrări sistematice, ce pot fi asimilate unor observări statistice utilizate și azi, ca servind unor scopuri fiscale, demografice și administrative;
- ✓ *statistica descriptivă*, necesară conducerii de stat, se rezumă la descrierea statelor sau a unor părți ale acestora fără a se ocupa și de cunoașterea regularităților care se manifestă în interiorul lor, a legilor care guvernează aceste fenomene;
- ✓ *aritmetica politică și calculul probabilităților*, ca bază conceptuală și ca mod de interpretare a fenomenelor în statistica modernă, tind spre exactitate și spre cunoașterea socială, obiectivul de bază fiind găsirea regulilor ce se produc în manifestările sociale.

Pentru statistica din mijlocul secolului al XVIII^{-lea} și secolul al XIX^{-lea} este specifică folosirea tot mai frecventă a metodelor matematice și, în special, a calculului probabilităților în investigarea și interpretarea rezultatelor privind fenomenele și procesele din societate, utilizându-le ca instrument de cercetare în producerea fenomenelor, la cunoașterea legilor care le determină și pe această bază, la efectuarea de previziuni științifice.

Ceea ce are sens pentru un anumit domeniu al statisticii poate să rămână o “abstracție pură” fără nici o reprezentare reală pentru un alt domeniu. După părerea unor autori “Lăsând la o parte latura sa fabuloasă, conceptul de om mediu se cuvine evocat pentru ideile implicate: repartiție; medie; dispersie; observare de masă; regularitate; noțiuni esențiale în cercetările statistice”.

2. Concepte de bază folosite în statistică

În cele ce urmează ne vom ocupa în mod deosebit de acele concepte ale statisticii care pot fi studiate mai aprofundat cu instrumentele matematice uzuale, pe de-o parte, iar pe de altă parte de cele ce modelează fenomene precum cele din educație fizică și sport.

Colectivitatea statistică (populația) reprezintă mulțimea tuturor elementelor cu aceleași caracteristici esențiale.

Exemple.

1) Studenții de la cursurile de zi de la secția de Educație Fizică și Sport din Tg-Jiu la data de 1 octombrie 2004;

2) Cetățenii României la 1 ianuarie 2005 ;

3) Cadrele didactice de la secția de Educație Fizică și Sport din Tg-Jiu la data de 1 octombrie 2004.

Astfel de exemple sugerează împărțirea colectivităților statistice în două categorii :

- ✓ *colectivități statice*, sunt acelea care exprimă o stare și au o anumită întindere în spațiu, formând împreună un existent la un moment dat (spre exemplu: populația din Tg-Jiu la 1 ianuarie 2005 ; câștigătorii la Prono-Sport din data de);
- ✓ *colectivități dinamice*, sunt formate din fluxuri de evenimente și exprimă un proces, o devenire în timp. Caracterizarea lor presupunând înregistrarea elementelor componente pe un interval de timp (spre exemplu: performanțele obținute de sportivi din Tg-Jiu, în luna ianuarie 2005; studenții restanțieri în anul 2004).

Unitățile statistice reprezintă elementele constitutive ale colectivităților și se separă în statice, care compun efectivul

de persoane, de animale, de mărfuri în stoc etc., și dinamice, pentru care unitățile aparțin aceleiași structuri organizatorice, însă în condiții diferite de timp. Unitățile mai pot fi simple, definite de elementele constitutive ale colectivității (spre exemplu: persoana, profesorul etc.), și complexe, definite ca fiind rezultatul organizării sociale și economice a colectivității (spre exemplu : echipa, secția, anul de antrenament etc.)

Reținem că unitățile statistice sunt unități obiective și independente ce pot fi studiate separat, pe subgrupuri mai mici sau pe întregul grup.

Caracteristicile statistice (numite și variabile statistice sau variabile aleatoare), sunt acele criterii pe baza cărora se caracterizează unitățile colectivității. Formele concrete de manifestare ale caracteristicilor la nivelul fiecărei unități a colectivității se numesc variante sau valori. Numărul de unități la care se înregistrează aceeași variantă sau valoare se numește frecvență sau pondere.

Privind conținutul, caracteristicile pot fi :

- ✓ *caracteristici de timp*, sunt acelea care arată apartenența unităților la un moment sau o perioadă de timp;
- ✓ *caracteristici de spațiu*, sunt acelea care se exprimă prin cuvinte pe baza unui nomenclator al unităților teritoriale și arată situarea în teritoriu a unității;

- ✓ *caracteristici atributive*, sunt acelea care servesc la definirea fenomenelor studiate.

După modul de exprimare caracteristicile statistice se grupează în :

- ✓ *caracteristici calitative*, sunt cele exprimate prin cuvinte, de exemplu profesia respectivei persoane (această caracteristică nu se cantifică) ;
- ✓ *caracteristici cantitative*, sunt cele exprimate prin numere, de exemplu vârsta respectivei persoane asupra căreia se face referire.

După natura variației, cele numerice se împart în *caracteristici cu variație* continuă (spre exemplu performanțele unui atlet vitezist poate fi măsurată manual în zecimi, sutimi sau miimi de secundă) și cu variație discontinuă (spre exemplu: numărul de victorii ale unui sportiv).

Datele statistice reprezintă caracteristici numerice ale unităților, grupelor și colectivității, obținute din observare și prelucrare, având ca mesaj informația.

Indicatorii statistici reprezintă expresia numerică a unei determinări calitative obiective, obținută în urma efectuării unei cercetări statistice raportată la condiții specifice de timp, spațiu și organizatorice.

Capitolul II

Elemente de teoria probabilităților

Capitol al teoriei probabilității, cu o problematică precis conturată, statistica are scopul de a confrunta modelele probabilistice teoretice cu realitatea obiectivă. Această confruntare se face prin intermediul conceptului fundamental de selecție (eșantionare), care poartă numai o informație parțială asupra realității examinate.

1. Noțiuni de teoria elementară a probabilităților.

Câmp de evenimente. Orice disciplină științifică s-a conturat în jurul experienței acumulate de om de-a lungul timpului. Realizările unei experiențe au fost de fapt probe ale experienței având ca rezultat *evenimentul* asociat experienței.

Experiențele care conduc la același rezultat (eveniment) se numesc experiențe deterministe.

Experiențele care conduc la rezultate (evenimente) diferite se numesc experiențe aleatoare.

În cele ce urmează Ω reprezintă mulțimea tuturor rezultatelor posibile ale experienței iar $\mathcal{P}(\Omega)$ mulțimea tuturor submulțimilor mulțimii Ω . Orice element din $\mathcal{P}(\Omega)$ constituie un eveniment asociat experienței.

Notăm cu $\emptyset$ evenimentul imposibil din $\mathcal{P}(\Omega)$, iar prin Ω evenimentul sigur, care apare la fiecare probă a experienței.

Evenimentul a cărei realizare constă în nerealizarea evenimentului E îl notăm prin $\bar{E}$ și-l numim evenimentul contrar evenimentului E , iar probabilitatea să se întâmple evenimentul E prin P_E .

Exemple clasice în teoria probabilității sunt cele oferite de aruncarea monedei, a zarului, extragerea din urnă etc..

În cazul aruncării monedei o singură dată vom avea două evenimente posibile : stema (S) sau banul (B). Înaintea probei putem afirma că se va produce un eveniment B sau S; probabilitatea producerii unui eveniment este 1 (certitudinea).

Deci:

$$P_B + P_S = 1.$$

Evenimentele sunt egal probabile (echiprobabile) adică :

$$P_S = \frac{1}{2}, P_B = \frac{1}{2}; \frac{1}{2} + \frac{1}{2} = 1$$

Aceasta sugerează:

$$P_E = \frac{\text{Nr. total de cazuri favorabile}}{\text{Nr. total de cazuri posibile}}$$

Arătăm că aruncând o monedă de două ori evenimentele $A=\{SS, SB\}$; $C=\{SS, BS\}$; $D=\{SS, BB\}$ sunt independente două câte două dar nu sunt independente.

Într-adevăr, spațiul evenimentelor elementare egal posibile (evenimentul sigur) este

$$\Omega=\{SS, SB, BS, BB\}$$

Atunci

$$P_{A \cap C} = \frac{1}{4} = P_A \cdot P_C$$

$$P_{A \cap D} = \frac{1}{4} = P_A \cdot P_D$$

$$P_{C \cap D} = \frac{1}{4} = P_C \cdot P_D$$

dar

$$P_{A \cap C \cap D} = 0 \neq \frac{1}{8} = P_A \cdot P_C \cdot P_D$$

Analog se poate demonstra că dacă moneda este aruncată de trei ori evenimentele $A=\{SSS, SSB, SBS, SBB\}$; $C=\{SSS, SSB, SBS, BSS\}$; $D=\{SSS, BSB, BBS, BBB\}$ sunt independente trei câte trei dar nu sunt independente două câte două.

Conchidem că, evenimentele care se produc (S sau B) sunt independente, rezultatul așteptat nu depinde de evenimentul anterior.

Considerăm acum experiența ce constă în aruncarea unui zar. În acest caz, mulțimea evenimentelor posibile este

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

În mulțimea $\mathcal{P}(\Omega)$ putem distinge:

i) Evenimentul imposibil $\emptyset$; de exemplu la aruncarea zarului să apară fața cu numărul 16.

ii) Evenimentul sigur Ω .

IV) $A = \{1, 3, 5\}$, la aruncarea zarului să apară una din fețele cu număr impar.

Spre exemplu evenimentul contrar evenimentului E este evenimentul $\bar{E} = \{2, 4, 6\}$.

Considerăm ultimul experiment, ce constă în extragerea bilelor.

Presupunem că urna U_i conține a_i bile albe și b_i bile negre, $i=1,2$. Din fiecare urnă extragem la întâmplare câte o bilă. Să se calculeze probabilitatea ca bila extrasă să fie albă.

Soluție. Observăm că numărul cazurilor egal posibile este $(a_1+b_1)(a_2+b_2)$ întrucât un eveniment elementar posibil este o pereche (m,n) , unde m este bila provenită din U_1 și n este bila provenită din U_2 . Fie acum A evenimentul ca bilele extrase să

fie albe. Numărul cazurilor favorabile producerii lui A este a_1a_2 întrucât eveniment elementar favorabil lui A este o pereche (m,n) formată cu o bilă albă din U_1 și o bilă albă din U_2 .

$$\text{Deci : } P_A = \frac{a_1a_2}{(a_1 + b_1) \cdot (a_2 + b_2)}$$

Astfel de exemple ne sugerează că probabilitatea “teoretică”, “legea statistică” se apropie de realitatea empirică în cazul producerii în număr mare a evenimentelor respective (legea numerelor mari). Cu alte cuvinte nu putem formula concluzii dacă numărul evenimentelor observate nu este suficient de mare.

2. Repartiția normală a evenimentelor întâmplătoare.

Revenind la exemplul *aruncarea monedei*, observăm că producerea în serie “B” este posibilă, dar cu atât mai puțin probabilă cu cât numărul succesiv de apariții este mai mare, cu alte cuvinte nu putem formula concluzii dacă numărul evenimentelor observate nu este suficient de mare.

Astfel, producerea în serie a evenimentelor B, B, B, B, B, B, B, B, B, B are șansa de a se realiza de $\left(\frac{1}{2}\right)^{10} = \frac{1}{1024}$

Dacă înregistram fiecare eveniment din cele 1024 de încercări s-ar fi obținut următoarea distribuție “probabilă” a evenimentelor:

- ✓ o dată 10 B și 10 S;
- ✓ de 10 ori 9B/1S și 1B/9S;
- ✓ de 45 de ori 8B/2S și 2B/8S;
- ✓ de 120 de ori 7B/3S și 3B/7S;
- ✓ de 210 ori 6B/4S și 4B/6S;
- ✓ de 252 de ori 5B/5S;

observându-se o distribuire simetrică a evenimentelor cu frecvența crescândă până la 5B/5S.

Astfel, numărul mare de evenimente aleatoare, la un număr mare de repetări pentru care se obține o repartiție aproape constantă a evenimentelor, compun repartiția normală.

Din experiență, fiecare eveniment în sine este determinat de întâmplare, marea sumă a evenimentelor aleatoare însă fiind repartizate invariabil, astfel că acțiunea simultană a multor evenimente aleatoare conduc la un rezultat nonîntâmplător obținându-se un model „teoretic” al unor evenimente determinate de o sumă de elemente întâmplătoare echiprobabile.

Din reprezentarea unor evenimente reale s-a constatat că distribuția lor se aseamănă cu cea normală, deci se justifică

extinderea proprietăților distribuției teoretice pentru evenimentele reale distribuite analog.

3. Eșantionarea

Eșantionarea (selecția) este metoda prin care putem deduce caracteristicile unei populații întregi, interogând doar o parte din subiecți, parte numită eșantion.

Există două tipuri de eșantioane:

- ✓ *aleatoare*, care se realizează atunci când fiecare individ din populație are o șansă diferită de zero de a fi ales în eșantion;
- ✓ *nealeatoare*, este acea procedură inversă celei aleatoare.

Este de observat faptul că reprezentativitatea eșantionului se poate calcula numai în cazul eșantioanelor probabilistice.

Eșantionarea simplă aleatoare. Subiecții componenți ai eșantionului sunt aleși uniform și cu o probabilitate identică pentru fiecare, fără a mai fi necesară o grupare a subiecților sau repetarea selecției, care la rândul ei implică două proceduri:

- ✓ *procedura loteriei sau a tragerii la sorți*, care constă în introducerea în urnă a fișei fiecărui subiect, și

extragerea numărului necesar de fișe realizării eșantionului.

- ✓ *procedeul tabelului cu numere la întâmplare*, caz în care se alcătuiește un tabel alfabetic care conține 5 coloane și 40 de rânduri (numerotate de la 1 la 5 respectiv de la 1 la 40) în care fiecare subiect al populației este reprezentat printr-un număr de ordine. Utilizarea acestui tabel constă în extragerea din cadrul populației a unităților ale căror numere de ordine stabilite printr-o numărătoare prealabilă, au fost citite după o anumită ordine din tabelul numerelor aleatoare. Direcția în care pornim în tabel se poate stabili prin aruncarea banului (stânga-dreapta, jos-sus în ordine crescătoare a numerelor de ordine ale coloanelor și ale rândurilor).

Există posibilitatea ca în alcătuirea eșantionului să nu se poată aplica procedeele menționate datorită unor diferențe calitative în structura populației, însă în cercetări pilot sau cercetări experimentale cu autocontrol se pot trage concluzii și fără respectarea strictă a regulilor menționate așa cum vor arăta exemplele din secțiunile următoare.

Numărul elementelor care compun eșantionul se stabilește făcând un compromis între cerințele de economicitate

ale cercetării, care tind să reducă numărul subiecților supuși cercetării și cerința de a mări numărul subiecților pentru a obține date semnificative.

Capitolul III

Serii statistice și reprezentări grafice

1.Introducere

Datele statistice sunt utile dacă sunt supuse unor operațiuni de prelucrare ce vizează cunoașterea atât a operațiilor de ordin cantitativ cât și a celor de ordin calitativ, astfel încât specialistul să poată separa tendințele esențiale de cele întâmplătoare, elementele obiective de cele subiective.

Prelucrarea statistică, ca etapă a cercetării statistice, presupune efectuarea unor operații cu ajutorul cărora se realizează trecerea de la date individuale la indicatori derivați, sintetici, care reflectă esența din manifestarea fenomenelor.

Prelucrarea statistică a datelor culese poate fi primară sau secundară. Prelucrarea primară presupune efectuarea unor operații de clasificări, de grupări, comparații și prezentarea centralizată a datelor sub formă de tabele, grafice sau serii statistice. Datele statistice sunt astfel sintetizate pe grupe, se calculează indicatorii absoluți și relativi, iar rezultatele obținute sunt prezentate în formă sintetizată prin tabele, grafice sau serii

statistice. Rezultatele operațiilor de prelucrare primară sunt prima etapă pentru prelucrarea secundară, în urma căreia se estimează valorile tipice, se apreciază omogenitatea și asimetria repartițiilor, intensitatea corelației dintre diferite fenomene analizate, etc.

2.Serii statistice

Definiție. Seria statistică reprezintă o paralelă între două sau mai multe șiruri de date, din care cel puțin unul vizează caracteristica de grupare. O *serie* statistică *simplică* este o paralelă între două șiruri de date, primul șir conținând valorile individuale ale variabilei de grupare iar cel de-al 2^{-lea} valorile individuale ale unei caracteristici oarecare. O *serie complexă* cuprinde trei sau mai multe șiruri paralele, conținând două sau mai multe caracteristici de grupare.

În funcție de tipul caracteristicii de grupare seriile statistice se clasifică în: serii statistice de distribuție, serii statistice de timp (cronologice) și serii statistice de spațiu.

2.1.Seriile de distribuție

Deosebim serii de distribuție simple și serii de distribuție complexe.

Seria de distribuție simplă (unidimensională) este o paralelă între șirul variantelor sau intervalelor de variație ale caracteristicii atributive și șirul frecvențelor corespunzătoare fiecărei variante sau interval de variație. Seria de distribuție complexă (bi sau multi-dimensională) este o paralelă între 3 sau mai multe șiruri de date, cu două sau mai multe caracteristici atributive.

Frecvențele specifice acestor serii pot fi:

- ✓ *Frecvențe absolute*: evidențiază numărul de apariții din

fiecare grupă. Se notează cu f_i , iar $\sum_{i=1}^n f_i = N$; N-

numărul de unități statistice ale colectivității;

- ✓ *Frecvențe relative*: evidențiază ponderea frecvențelor absolute în totalul colectivității din care fac parte. Se

notează cu f_i^* ; $f_i^* = \frac{f_i}{\sum_{i=1}^n f_i} \cdot 100$.

Pentru determinarea mărimii intervalului de variație, în cazul seriilor de distribuție pe intervale de variație se folosește formula lui H.A. Stuhges:

$$K = \frac{x_{\max} - x_{\min}}{1 + 3,322 \lg N}, \text{ unde } x_{\min} \text{ și } x_{\max} \text{ reprezintă varianta}$$

minimă și respectiv cea maximă a caracteristicii de grupare.

Pentru determinarea centrului de interval se vor utiliza formulele:

$$x_i^c = \frac{x_i^{\text{inf}} + x_i^{\text{sup}}}{2}, \text{ dacă } x_i^{\text{sup}} = x_{i+1}^{\text{inf}} \quad \text{și} \quad x_i^c = x_i^{\text{inf}} + \frac{k}{2}, \text{ dacă}$$

cele două limite sunt distincte, fiind distanțele cu o unitate egală cu cifra de cel mai mic ordin.

x_i^c – centru de interval;

x_i^{inf} – limita inferioară a intervalului;

x_i^{sup} – limita superioară a intervalului;

k – lungimea intervalului.

Exemplu de distribuție unidimensională:

Distribuția studenților grupei 101, secția educație fizică și sport din Tg-Jiu, după numărul de factori:

Nr. de factori	Nr. de studenți	Frecvențe relative
→2	10	0,1
1-4	30	0,3
4-6	10	0,1
6-8	18	1,8
8-10	20	0,2
10-12	5	0,05
12-14	5	0,05
14-16	2	0,02
Total	100	1

Exemplu de distribuție bidimensională. Distribuția studenților grupei 101, secția Educație fizică și sport după numărul de flotări efectuate și sexul studenților:

Nr. fl. Sex	1-5	5-10	10-15	15-20	20-25	25-30
Feminin	20	5	10	15	5	5
Masculin	5	10	15	25	30	20
Total	25	15	25	40	35	25

2.2.Serii cronologice

Sunt de două tipuri: serii cronologice simple și serii cronologice complexe.

Seria cronologică simplă este o paralelă între două șiruri de date, unul corespunzător caracteristicii de timp iar celălalt unei caracteristici oarecare.

Seria cronologică complexă este o paralelă între șirul caracteristicii de timp și mai multe șiruri de date, aferente unei caracteristici oarecare.

O altă clasificare a seriilor cronologice este:

a)*seria cronologică de momente*, în cadrul căreia variantele caracteristicii de timp sunt momente sau intervale foarte mici. De exemplu: ziua, data, etc.

Seria cronologică de momente evidențiază evoluția unei fenomen supus observării statistice până la un moment dat de timp sau de la un moment la altul.

Exemplu de serie cronologică de momente de timp:

Numărul de mingi de baschet existente în depozitul secției de educație fizică și sport este:

Data	Nr. de mingi
1.1.2004	10
1.2.2004	20
1.3.2004	25
1.4.2004	30
1.5.2004	40
1.6.2004	40
1.7.2004	42
1.8.2004	44
1.9.2004	45
1.10.2004	50
1.11.2004	45
1.12.2004	40

b) *Serii cronologice de intervale de timp* în cadrul cărora intervalele de timp variantele caracteristicii de timp sunt intervalele mari de timp, de *exemplu*: luna, trimestrul, anul etc.

Exemplu de serie cronologică de intervale de timp:

Situația numărului de studenți înscriși la specializarea Educație fizică și sport a Universității Constantin Brâncuși din Tg-Jiu.

Anul	Nr. studenți
1995	100
1996	120
1997	150
1998	110
1999	200
2000	205
2001	210
2002	250
2003	300
2004	320
2005	350

2.3.Serii de spațiu

Seria de spațiu simplă este o paralelă între șirul variantelor caracteristicii de spațiu și șirul nivelurilor individuale ale unei caracteristici oarecare. Seria de spațiu complexă este o paralelă între șirul variantelor caracteristicii de spațiu și 2 sau mai multe șiruri de variante ale unor variabile oarecare.

Exemplu de serie de spațiu. Distribuția cluburilor sportive de baschet și volei în județele Gorj, Dolj, Olt, Vâlcea.

Județul	Nr. cluburi baschet	Nr. cluburi volei
Gorj	10	15
Dolj	25	30
Olt	15	30
Vâlcea	5	40

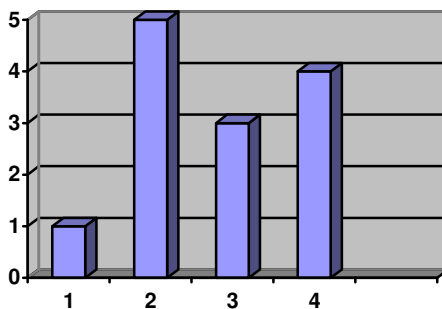
3.Grafica seriilor de distribuție

Reprezentările grafice specifice seriilor de distribuție sunt: histograma, poligonul frecvențelor și curba frecvențelor cumulate (ogiva).

3.1.Histograma

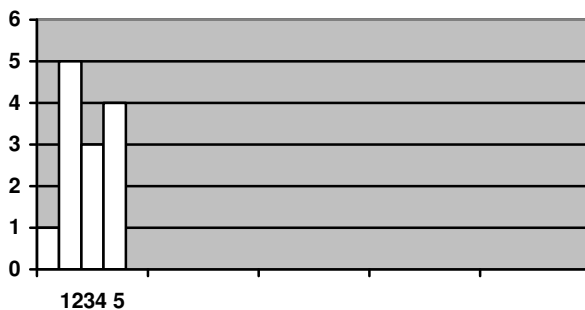
Deosebim:

- ✓ Histograma prin batoane-se folosește în cazul seriilor de distribuție; după variante ale caracteristicii atributive, poate fi folosită și în cazul distribuțiilor după intervale de variație pe Ox, folosindu-se centrele intervalelor.



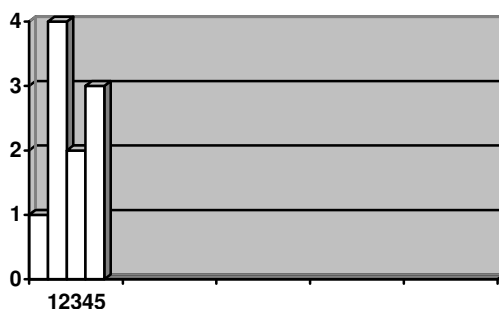
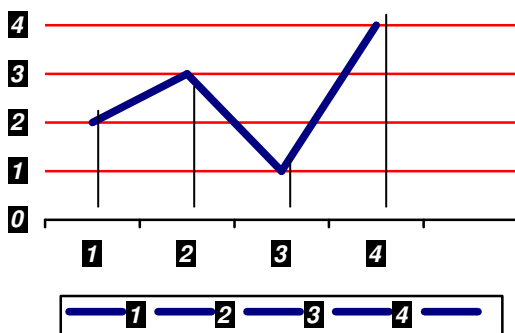
- ✓ *Histograma prin dreptunghiuri*: se folosește pentru reprezentarea grafică a distribuțiilor după intervale de variație. Pe axa Ox se trec intervalele de variație iar pe Oy frecvențele corespunzătoare. De pe abscisă, din dreptul diviziunilor scării se ridică dreptunghiuri a căror înălțime este proporțională cu frecvențele corespunzătoare fiecărui interval.

Histograma prin dreptunghiuri se folosește și la determinarea grafică a dominantei.



III.3.2. Poligonul frecvențelor

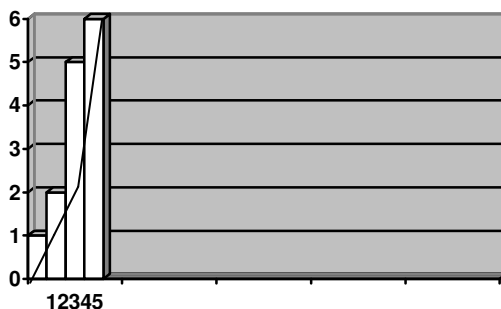
Metodologia de construcție este asemănătoare histogramei, cu deosebirea că se vor uni vârfurile batoanelor sau mijlocul bazelor superioare ale dreptunghiurilor, obținându-se astfel poligonul frecvențelor.



3.3. Curba frecvențelor cumulate (ogiva)

Metodologia de construcție este asemănătoare histogramei, cu deosebirea că pe ordonată sunt trecute, în conformitate cu scara de reprezentare, frecvențele cumulate. Cumularea frecvențelor poate fi făcută de la începutul seriei spre sfârșit, obținându-se curba ascendentă a frecvențelor, sau de la sfârșitul seriei spre început, obținându-se curba descendentă a frecvențelor. Prin unirea extremităților din

dreapta ale bazelor superioare ale dreptunghiurilor se obține ogiva.



Curba frecvențelor cumulate poate fi folosită, așa cum vom vedea în capitolul *Indicatorii tendinței centrale*, la determinarea grafică a medianei.

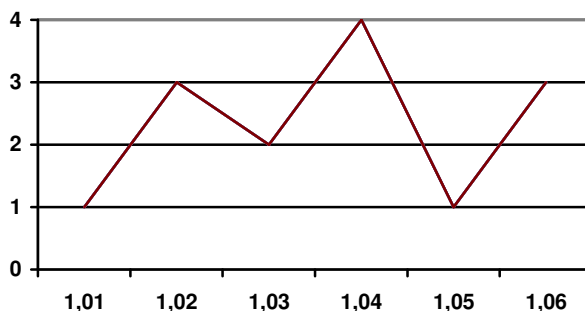
4. Grafica seriilor cronologice

Reprezentările grafice specifice seriilor cronologice sunt cronograma, histograma, diagrama prin coloane și diagrama polară.

4.1. Cronograma

Metodologia de construcție a cronogramei este următoarea: pe axa Ox se trec valorile caracteristicii de timp, în dreptul diviziunilor scării, dacă seria este de momente sau în dreptul intervalelor dintre diviziuni dacă seria este de intervale

de timp; pe axa Oy se trec valorile scării de reprezentare corespunzătoare fenomenului urmărit. De pe axa Ox, din dreptul diviziunilor sau centrelor intervalelor se ridică perpendiculare a căror înălțime este proporțională cu indicatorii reprezentați în grafic. Prin unirea vârfurilor acestor perpendiculare se va obține cronograma.



4.2.Histograma

Metodologia de construcție a acesteia este asemănătoare cu cea a cronogramei, deosebirea constând în faptul că pe orizontală sau pe verticală va apare un canal de întrerupere, ce ilustrează renunțarea la unele porțiuni din scara de reprezentare.

4.3.Diagrama prin coloane

Se folosește pentru reprezentarea grafică a seriilor cronologice de intervale de timp. Metodologia de construcție este asemănătoare cu cea a cronogramei, deosebirea constând

în faptul că de pe abscisă se ridică dreptunghiuri din dreptul intervalelor dintre diviziuni, a căror înălțime este proporțională cu nivelurile indicatorului supus observării. Figurile geometrice folosite în construcția diagramei prin coloane sunt: dreptunghiul, paralelipipedul, cilindrul, conul, etc.

4.4.Diagrama polară

Pentru reprezentarea grafică a seriilor cronologice ce evidențiază o evoluție sezonieră, ciclică, se folosește diagrama polară.

Deosebim:

a)*diagrama radială*. Metodologia de construcție presupune folosirea sistemului axelor polare.

Suportul diagramei este un cerc, iar raza cercului este elementul care asigură proporționalitatea valorilor reprezentate grafic. Mărimea razei se determină ca raport între nivelul mediu al seriei și valoarea prestabilită a unui centimetru. Variantele caracteristicii de timp se trec în dreptul razelor în jurul cercului suport. Nivelurile efective ale indicatorului supus observării se trec în funcție de scara de reprezentare pe raze sau pe prelungirile acestora. Prin unirea cu segmente de dreaptă a acestor niveluri vom obține diagrama radială.

b)*Diagrama sectorială*. Metodologia de construcție este aceeași cu deosebirea că variantele caracteristicii de timp

se trec în dreptul sectoarelor de cerc, în jurul cercului, iar nivelurile efective ale indicatorului urmărit se trece de asemenea în dreptul sectoarelor. Se vor trasa arce de cerc din punctele corespunzătoare acestor niveluri, paralele cu cercul folosit ca suport, determinându-se astfel sectoare de cerc a căror arie e proporțională cu nivelurile reprezentate grafic. Prin unirea capetelor acestor

5. Grafica diverselor comparații

Pentru grafica diverselor comparații se folosesc:

a) *Dreptunghiul*. Deosebim 2 situații:

- ✓ când aria dreptunghiului este funcție doar de mărimea unei laturi, cealaltă menținându-se constantă;
- ✓ când aria dreptunghiului este funcție de ambele laturi, situație întâlnită atunci când sunt reprezentați grafic indicatorii bifactoriali.

b) *Pătratul*. Latura pătratului este element de proporționalitate.

$$L = \sqrt{A}$$

A-aria pătratului;

c) *Cercul*. Elementul de proporționalitate îl constituie raza. Ea se determină cu ajutorul relației:

$$R = \sqrt{\frac{A}{\pi}} = \sqrt{\frac{Q}{\pi}}$$

A-reprezentând aria cercului;

Q-nivelul indicatorului supus observării statistice.

d)*Paralelipipedul*. Este folosit pentru reprezentarea grafică a indicatorilor trifactoriali (fiecare latură reprezentând un indicator, volumul proporțional cu nivelul indicatorului general).

e)*Cilindrul*. Se folosește pentru reprezentarea indicatorilor bifactoriali. Suprafața bazei este proporțională cu nivelul unui factor, iar înălțimea cilindrului este proporțională cu celălalt factor. Volumul cilindrului reliefează nivelul indicatorului general supus observării statistice.

6.Grafica structurilor

Diagramele de structură se folosesc pentru reprezentarea grafică a elementelor structurale ale unui indicator supus observării statistice. Figurile geometrice folosite sunt divizate în sectoare a căror dimensiune este proporțională cu nivelurile indicatorilor structurali reprezentați grafic. Metodologia de construcție a diagramelor de structură este următoarea:

- ✓ se alege figura geometrică construirea acesteia făcându-se pe baza nivelurilor totale ale indicatorilor reprezentați grafic;
- ✓ Se divizează figura geometrică în funcție de scara de reprezentare aleasă și de nivelurile indicatorilor structurali;
- ✓ în cazul în care figura aleasă este cercul, $1\% = 3,6$ grade. Mărimea unui sector de cerc este dată de produsul dintre 3,6 grade și valoarea indicatorului structural exprimat în procente;
- ✓ legenda este obligatorie.

7.Diagrama prin coloane în aflux

Dacă se urmărește compararea mai multor indicatori ce fac parte din aceeași grupă, reprezentarea grafică a acestora este realizată cu ajutorul diagramei prin coloane în aflux. Metodologia de construcție este următoarea: pe axa Ox se vor construi coloane a căror înălțime este proporțională cu cea a indicatorului reprezentat grafic; coloanele sunt dispuse în ordine crescătoare una în spatele celeilalte, cu o deplasare spre dreapta egală cu jumătate din mărimea bazei coloanelor.

Legenda este obligatorie. Indicatorii reprezentați grafic trebuie exprimați în aceeași unitate de măsură.

7.Aplicații

Aplicația 1. Se dau următoarele date înregistrate pentru proba de exercițiu în 4 timpi, grupa 202 fete, Facultatea de Drept, Universitatea Constantin Brâncuși din Tg-Jiu.

Nr. crt.	Nume și prenume	Exercițiu în patru timpi
1	A.C.	9
2	A.I.	12
3	B.I	9
4	B.M.A.	10
5	B.M.C.	10
6	B.R.	11
7	B.C.	10
8	B.M	9
9	B.R.	10
10	C.D.P.	11
11	C.I.V.	8
12	C.M.	9
13	D.E.	10
14	V.I.	10
15	D.C.	11
16	D.A.	11
17	F.M.C.	9
18	H.A.M.	11
19	J.C.6.	9
20	C.I.A.	10
21	M.C.R.	11

22	M.E.P.	11
23	P.A.	11
24	P.D.E.	8
25	P.A.M.	9
26	P.C.	8
27	S.E.A.	9
28	S.E.D.	11
29	T.L.S.	11
30	U.T.	11

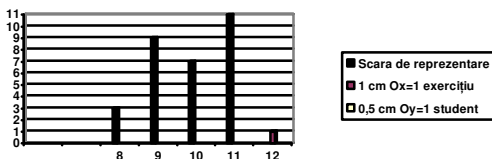
Să se construiască seria corespunzătoare și să se reprezinte grafic.

Rezolvare.

Vom construi distribuția pe variante a studenților

Variante	Nr. studenți
8	3
9	8
10	7
11	11
12	1
Total	30

Vom reprezenta grafic seria prin histograma prin batoane:

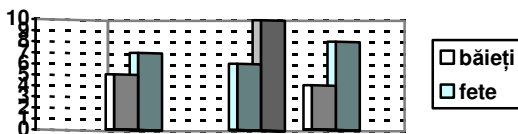


Aplicația 2. Se dau următoarele date: privind numărul băieților și numărul fetelor de la Universitatea Constantin Brâncuși din Tg-Jiu ce au câștigat titlul de campion național.

Anii	Nr. băieți	Nr. fete
2000	5	7
2001	10	6
2002	4	8

Să se reprezinte grafic datele prezentate.

Rezolvare. Vom folosi diagrama prin coloane în aflux.

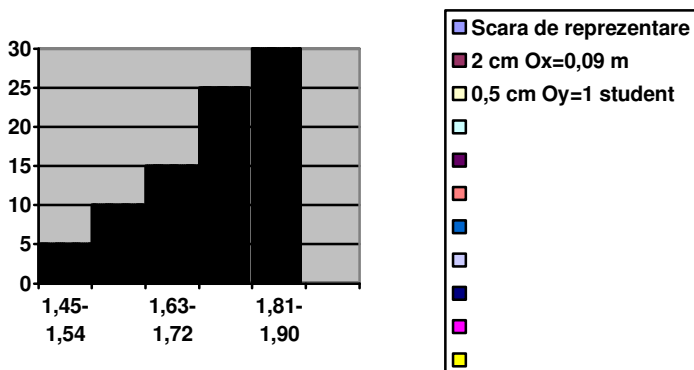


Aplicația 3. Să se reprezinte grafic distribuția de mai jos folosind curba frecvențelor cumulate.

Intervale de grupare	Nr. studenți	Frecvențe cumulate crescător
1,45-1,54	2	2
1,54-1,63	8	10
1,63-1,72	5	5
1,72-1,81	9	24
1,81-1,90	6	30
Total	30	*

Distribuția fețelor după mărimea săriturii de pe loc (m).

Rezolvare. După cumularea frecvențelor vom reprezenta grafic distribuțiile:



Capitolul IV

Indicatorii tendinței centrale

Așa cum am văzut în *capitolul* I, obiectul de studiu al statisticii îl reprezintă fenomenele de masă. Caracteristica principală a fenomenelor de masă este variabilitatea formelor de manifestare, determinată de un complex de factori sistematici sau întâmplători, obiectivi sau subiectivi, esențiali sau neesențiali.

1. Importanța utilizării indicatorilor tendinței centrale.

Legile statistice acționează la nivelul colectivității, ca tendință. Este deosebit de important deci, cunoașterea la nivelul colectivității supuse observației statistice a tendinței, a ceea ce este esențial, firesc și obiectiv în formele de manifestare individuală a unităților statistice ce compun colectivitatea, cunoaștere ce se realizează prin determinarea indicatorilor sintetici adecvați.

Indicatorii tendinței centrale se determină ca indicatori medii sau indicatori de poziție, în funcție de tipul variabilelor

de grupare. După Zule, condițiile ce trebuie îndeplinite de un astfel de indicator sunt:

- ✓ să fie definit obiectiv, independent de dorința utilizatorului;
- ✓ să aibă o semnificație concretă, accesibilă;
- ✓ să fie simplu și rapid de calculat;
- ✓ determinarea sa să depindă de toate valorile; individuale înregistrate;
- ✓ să poată fi folosit la diferite operații algebrice;
- ✓ să fie puțin sensibil la fluctuațiile de selecție;

În practica statistică indicatorii tendinței centrale nu satisfac simultan condițiile de mai sus, interpretarea lor depinzând de capacitatea de înțelegere a utilizatorului.

2. Indicatorii medii

Media este expresia sintetică a nivelurilor individuale ale unei variabile oarecare, concentrată într-un singur nivel reprezentativ, care evidențiază ceea ce este esențial, firesc, tipic și obiectiv în dezvoltarea unui fenomen.

Media se exprimă în unități concrete de măsură, însă are un caracter abstract, valoarea ei putând să coincidă sau nu cu valoarea individuală a unei variabile numerice urmărite.

O condiție esențială pentru reprezentativitatea mediei, ca măsură a tendinței centrale este verificarea omogenității colectivității în funcție de caracteristica urmărită.

Se deosebesc următoarele tipuri de medii:

- ✓ media aritmetică;
- ✓ media armonică;
- ✓ media pătratică;
- ✓ media geometrică;
- ✓ media cronologică;

Fiecare dintre aceste tipuri are două variante: simplă și ponderată.

2.1. Media aritmetică

Media aritmetică se folosește atunci când variabilitatea unui fenomen se produce aproximativ în progresie aritmetică, înregistrând o tendință lineară. Când numărul variantelor caracteristicii este egal cu numărul unităților statistice supuse observării, adică atunci când nu se repetă nici o variantă, se folosește media aritmetică simplă. Dacă însă cel puțin una din variante se repetă, vom folosi media aritmetică ponderată.

a) Media aritmetică simplă

Metodologia de calcul a mediei aritmetice are la bază proprietatea următoare: înlocuind în cadrul funcției

determinante variantele x_i cu media lor, nivelul general al caracteristicii X va rămâne neschimbat. Vom avea:

$$x_1 + x_2 + x_3 + \dots + x_n = \sum_{i=1}^n x_i$$

$$\bar{x} + \bar{x} + \bar{x} + \dots + \bar{x} = n\bar{x}$$

Dar cum $n\bar{x} = \sum_{i=1}^n x_i$, rezultă că formula de calcul a

mediei aritmetice simple va fi:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

b) Media aritmetică ponderată

Metodologia de calcul este aceeași, cu deosebirea că fiecare variantă x_i este ponderată cu frecvența corespunzătoare f_i .

Vom avea:

$$x_1 f_1 + x_2 f_2 + \dots + x_n f_n = \sum_{i=1}^n x_i f_i$$

$$\bar{x} f_1 + \bar{x} f_2 + \dots + \bar{x} f_n = \bar{x} \sum_{i=1}^n f_i$$

Dar cum $\sum_{i=1}^n x_i f_i = \bar{x} \sum_{i=1}^n f_i$, rezultă că formula de calcul

a mediei aritmetice ponderate este: $\bar{x} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i}$

Proprietățile mediei aritmetice

✓ de verificare a corectitudinii calculului:

$$x_{\min} < \bar{x} < x_{\max}$$

Orice valoare în afara acestui interval semnalează un calcul eronat al mediei.

$$\sum_{i=1}^n (x_i - \bar{x}) = 0 - \text{pentru media aritmetică simplă}$$

$$\sum_{i=1}^n (x_i - \bar{x}) f_i = 0 - \text{pentru media aritmetică ponderată}$$

✓ de simplificare a calculului:

Media calculată din variantele caracteristicii, micșorate în prealabil cu o constantă „a” este mai mică decât media reală cu acea constantă „a”, astfel încât:

$$\frac{\sum_{i=1}^n (x_i - \bar{x}) f_i}{\sum_{i=1}^n f_i} < \bar{x} \text{ cu „a”, adică:}$$

$$\bar{x} = \frac{\sum_{i=1}^n (x_i - \bar{x}) f_i}{\sum_{i=1}^n f_i} + a \quad (1)$$

Media calculată pe baza variantelor caracteristicii micșorate în prealabil de „k” ori (k este constantă) este mai mică decât media reală de „k” ori.

$$\frac{\sum_{i=1}^n (\frac{x_i}{k}) f_i}{\sum_{i=1}^n f_i} < \bar{x} \quad \text{de „k” ori, adică:}$$

$$\bar{x} = \frac{\sum_{i=1}^n (\frac{x_i - a}{k}) f_i}{\sum_{i=1}^n f_i} \cdot k + a$$

Precizări:

- constanta „a” este de regulă varianta caracteristicii cu frecvența cea mai mare;
- „k” reprezintă mărimea intervalului de variație
- odată stabilită poziția lui „a” în cadrul seriei de distribuție, pentru $\frac{x_i - a}{k}$ se vor trece valorile “ 0 “ în

dreptul poziției lui “ a “, -1, -2, -3, ...etc., deasupra de 0, 1, 2, 3...etc. sub 0.

Metodologia de calcul simplificat a mediei aritmetice se aplică doar pentru seriile de distribuție cu intervale de variație egale.

2. Media armonică

Media armonică, ca indicator al tendinței centrale, se definește ca valoarea inversă a mediei aritmetice a inverselor valorilor individuale înregistrate. Aplicarea mediei armonice pentru caracterizarea tendinței centrale are sens numai dacă însumarea inverselor valorilor individuale este obiectivă.

a) Media armonică simplă

Funcția determinantă este tot de tip aditiv, ca și în cazul mediei aritmetice:

$$\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} = \sum_{i=1}^n \frac{1}{x_i}$$

și

$$\frac{1}{\bar{x}} + \frac{1}{\bar{x}} + \dots + \frac{1}{\bar{x}} = \frac{n}{\bar{x}}$$

Dar cum $\sum_{i=1}^n \frac{1}{x_i} = \frac{n}{\bar{x}}$, rezultă că formula mediei armonice simple va fi:

$$\bar{x}_h = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

b) Media armonică ponderată

Vom avea:

$$\frac{1}{x_1} f_1 + \frac{1}{x_2} f_2 + \dots + \frac{1}{x_n} f_n = \sum_{i=1}^n \frac{1}{x_i} f_i$$

$$\frac{1}{\bar{x}} f_1 + \frac{1}{\bar{x}} f_2 + \dots + \frac{1}{\bar{x}} f_n = \frac{1}{\bar{x}} \sum_{i=1}^n f_i$$

Cum însă $\sum_{i=1}^n \frac{1}{x_i} f_i = \frac{1}{\bar{x}} f_i$ rezultă formula mediei armonice ponderate :

$$\bar{x}_h = \frac{\sum_{i=1}^n f_i}{\sum_{i=1}^n \frac{1}{x_i} f_i}$$

Dacă se calculează pentru același set de date atât media aritmetică cât și media armonică, întotdeauna vom avea relația:

$$\bar{x}_h < \bar{x}$$

3. Media pătratică

În cazul în care fenomenele studiate prezintă creșteri aproximativ în progresie exponențială, creșterea fiind lentă la început și din ce în ce mai pronunțată spre sfârșitul seriei, pentru caracterizarea tendinței centrale vom folosi media pătratică.

a) Pentru *media pătratică simplă* avem:

$$x^2_1 + x^2_2 + \dots + x^2_n = \sum_{i=1}^n x^2_i$$

$$\bar{x}^2 + \bar{x}^2 + \dots + \bar{x}^2 = n\bar{x}^2$$

Cum însă $n\bar{x}^2 = \sum_{i=1}^n x^2_i$, formula mediei pătratice

simplă este:

$$\bar{x}_p = \sqrt{\frac{\sum_{i=1}^n x^2_i}{n}}$$

b) *Media pătratică ponderată*

$$x^2_1 f_1 + x^2_2 f_2 + \dots + x^2_n f_n = \sum_{i=1}^n x^2_i f_i$$

$$\bar{x}^2 f_1 + \bar{x}^2 f_2 + \dots + \bar{x}^2 f_n = \bar{x}^2 \sum_{i=1}^n f_i$$

Din $\bar{x}^2 \sum_{i=1}^n f_i = \sum_{i=1}^n x^2_i f_i$ rezultă formula mediei pătratice ponderate:

$$\bar{x}_p = \frac{\sum_{i=1}^n x^2_i f_i}{\sum_{i=1}^n f_i}$$

Pentru aceeași serie statistică întotdeauna $\bar{x} < \bar{x}_p$.

4. Media geometrică

În cazul în care fenomenul studiat înregistrează modificări aproximativ în progresie geometrică, prezentând diferențe mari între variantele caracteristicii la începutul seriei și din ce în ce mai mici spre finalul acesteia vom folosi ca indicator al tendinței centrale media geometrică.

Pentru determinarea mediei geometrice funcția folosită este de tipul produsului.

a) *Media geometrică simplă:*

$$x_1 \cdot x_2 \cdot x_3 \cdot \dots \cdot x_n = \prod_{i=1}^n x_i$$

$$\bar{x} \cdot \bar{x} \cdot \bar{x} \cdot \dots \cdot \bar{x} = \bar{x}^n$$

Cum $\prod_{i=1}^n x_i = \bar{x}^n$, rezultă formula de calcul a mediei geometrice simple:

$$\bar{x}_g = \sqrt[n]{\prod_{i=1}^n x_i}$$

b) *Media geometrică ponderată*

$$x^{f_1}_1 \cdot x^{f_2}_2 \cdot x^{f_3}_3 \cdot \dots \cdot x^{f_n}_n = \prod_{i=1}^n x^{f_i}_i$$

$$\bar{x}^{f_1} \cdot \bar{x}^{f_2} \cdot \bar{x}^{f_3} \cdot \dots \cdot \bar{x}^{f_n} = \bar{x}^{\sum_{i=1}^n f_i}$$

Cum însă $\bar{x}^{\sum_{i=1}^n f_i} = \prod_{i=1}^n x^{f_i}_i$, rezultă că formula de determinare a mediei geometrice ponderate este:

$$\bar{x}_g = \sqrt[\sum_i f_i]{\prod_{i=1}^n x^{f_i}_i}$$

Pentru același set de date vom avea întotdeauna relația:

$$\bar{x}_g < \bar{x} < \bar{x}_p.$$

5. Media cronologică

Se folosește în mod unic doar pentru seriile cronologice de momente. În cazul în care intervalele dintre momente sunt egale, vom folosi media cronologică simplă. Logic, dacă intervalele dintre momente sunt diferite vom folosi media cronologică ponderată.

Calculul mediei cronologice presupune parcurgerea a două etape:

- ✓ calcularea mediilor mobile ca medii aritmetice simple a câte doi sau mai mulți termeni, în cadrul cărora cel puțin un termen se repetă
- ✓ calculul mediei cronologice, simplă sau ponderată, după caz.

a) *Media cronologică simplă*

Fie seria:

Data	x_i
1.01	x_1
1.02	x_2
1.03	x_3
1.04	x_4

Etapa I. Calculul mediilor mobile:

$$\bar{x}_1 = \frac{x_1 + x_2}{2}; \bar{x}_2 = \frac{x_2 + x_3}{2}; \bar{x}_3 = \frac{x_3 + x_4}{2}$$

Etapa a II-a Calculul mediei cronologice simple. Formula de calcul este:

$$\bar{x}_c = \frac{\sum_{i=1}^n \bar{x}_k}{k}, \quad k - \text{numarul mediilor mobile}$$

În cazul de față avem:

$$\bar{x}_c = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3}{3} = \frac{\frac{x_1 + x_2}{2} + \frac{x_2 + x_3}{2} + \frac{x_3 + x_1}{2}}{3}$$

b) *Media cronologică ponderată*

Fie seria

Data	x_0
1.01	x_1
1.03	x_2
1.05	x_3
1.08	x_4
1.12	x_5

Etapa I. Calculul mediilor mobile

$$\bar{x}_1 = \frac{x_1 + x_2}{2}; \bar{x}_2 = \frac{x_1 + x_2}{2}; \bar{x}_3 = \frac{x_3 + x_4}{2}; \bar{x}_4 = \frac{x_4 + x_5}{2}$$

Etapa II. Calculul mediei cronologice ponderate se face cu ajutorul formulei:

$$\bar{x}_c = \frac{\sum_{i=1}^n \bar{x}_k \cdot t_k}{\sum_{i=1}^n t_k}$$

În cazul de față:

$$\bar{x}_c = \frac{\bar{x}_1 \cdot 2 + \bar{x}_2 \cdot 2 + \bar{x}_3 \cdot 3 + \bar{x}_4 \cdot 4}{11}$$

6.Indicatorii de poziție

Pentru a cunoaște tendința centrală în cazul seriilor de distribuție este necesar a determina nu numai valorile individuale ale caracteristicii observate, ci și cunoașterea formei în care se repartizează unitățile colectivității după caracteristica respectivă. Indicatorii de poziție oferă informații utile pentru caracterizarea tendinței centrale a fenomenului supus observației statistice.

6.1. Mediana

Definiție. *Mediana* este dată de acea valoare a caracteristicii care împarte seria în 2 părți egale.

Pentru serii simple distingem cazurile:

a) dacă seria are număr impar de termeni, mediana este acea valoare a caracteristicii ce ocupă poziția $\frac{n+1}{2}$ în cadrul seriei ordonate crescător sau descrescător.

De exemplu: avem următoarele date privind săritura în lungime (m) a unei grupe de 9 studenți:

1,90; 1,6; 2,30; 2,40; 2,50; 1,80; 1,85; 1,70; 1,95;

Ordonăm crescător seria:

1,6; 1,70; 1,80; 1,85; 1,90; 1,95; 2,30; 2,40; 2,50;

$$loc(M_e) = \frac{n+1}{2} = \frac{9+1}{2} = 5 \Rightarrow M_e = 1,90 \text{ m}$$

b) Dacă seria are număr par de termeni, mediana este dată de semisuma termenilor centrali, în seria ordonată crescător sau descrescător.

Exemplu: Folosim aceleași date, dar mai adăugăm o săritură (un rezultat). Seria ordonată crescător este:

1,6; 1,70; 1,80; 1,85; 1,90; 1,95; 2,00; 2,30; 2,40; 2,50;

$$M_e = \frac{1,6 + 2,50}{2} = 2,05 \text{ m}$$

Pentru serii de distribuție există două posibilități de determinare a medianei:

a) *Calculul algebric.*

Formula de calcul este:

$$M_e = l_i + \left(\frac{\sum f_i}{2} - S_m \right) \cdot \frac{k}{f_m}, \text{ unde}$$

l_i - limita inferioară a intervalului median;

$$\frac{\sum_{i=1}^n f_i}{2} \text{ -jumătate din unitățile statistice;}$$

S_m -suma frecvențelor intervalelor ce preced intervalul median;

k - mărimea intervalului de variație;

f_m -frecvența intervalului median.

b) *Determinare grafică.*

Se folosește curba frecvențelor cumulate din dreptul lui

$\frac{\sum_{i=1}^n f_i}{2}$ se duce o paralelă la axa Ox. Din punctul de intersecție

al acesteia cu originea se coboară o perpendiculară pe Ox. Punctul de intersecție al acesteia cu abscisa corespunde, cu o precizie de 100% cu valoarea medianei.

6.2.Modulul (Dominanta)

Definiție. Modulul este acea valoare a caracteristicii cu frecvența cea mai mare.

Exemplu: Dispunem de următoarele date privind numărul flotărilor efectuate de o grupă de 10 studenți:

10; 12; 14; 10; 22; 20; 18; 20; 19; 20;

Valoarea modală este: 20 de flotări.

Pentru seriile de distribuție există de asemenea două posibilități de determinare a valorii modale:

a)*Calculul algebric.* Formula de calcul este:

$$M_o = l_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) \cdot k, \quad \text{unde}$$

l_i -limita inferioară a intervalului modal (intervalul cu frecvența cea mai mare)

Δ_1 - frecvența intervalului modal – frecvența intervalului precedent;

Δ_2 - frecvența intervalului modal – frecvența intervalului următor.

b)*Determinarea grafică.* Se folosește histograma prin dreptunghiuri. Se unesc vârfurile coloanei maxime cu punctele de incidență ale coloanei adiacente. Din intersecția segmentelor se coboară o perpendiculară pe Ox. Punctul de

intersecție reprezintă în proporție de 100% valoarea modului (dominantei).

7.Aplicații

Aplicația 1. La secția de educație fizică și sport din cadrul Universității Constantin Brâncuși din Tg-Jiu s-a înregistrat următoarea situație privind numărul studenților înscriși:

Anul	Nr. studenților	Anul	
1995	35	2000	55
1996	40	2001	60
1997	42	2002	65
1998	45	2003	70
1999	50	2004	80

- a) Să se precizeze tipul caracteristicii de grupare și tipul seriei.
- b) Să se determine numărul mediu anual al studenților înscriși, justificându-se tipul de medie ales.

Rezolvare:

a) Caracteristica de grupare-anual- este o caracteristică statistică de timp, seria prezentată fiind deci o serie cronologică simplă de intervale de timp.

b) Deoarece nivelurile individuale ale șirului numărul de studenți cresc aproximativ cu aceeași valoare, vom folosi pentru determinarea numărului mediu de studenți media aritmetică simplă:

$$\bar{x}_0 = \frac{\sum x_i}{n} = \frac{542}{10} = 54,2 - \text{numărul studenților înscriși.}$$

Aplicația 2. Se dau următoarele date privind rezultatele probei “Săritura în lungime de pe loc”, băieți, de la specializarea Educație fizică și sport din cadrul Universității Constantin Brâncuși din Tg-Jiu.

Nr. crt.	Săritura în lungime de pe loc (cm)	Nr. crt.	Săritura în lungime de pe loc (cm)
1	190	16	260
2	210	17	220
3	225	18	220
4	220	19	215
5	210	20	240
6	215	21	260
7	210	22	200
8	215	23	240
9	210	24	215
10	190	25	210
11	210	26	225
12	210	27	225
13	225	28	220
14	230	29	225
15	235	30	230

a) Să se determine săritura în lungime de pe loc medie a grupei.

b) Să se determine mediana prin:

1) calcul algebric;

2) determinare grafică;

c) Să se determine valoarea modală a grupei prin:

1) calcul algebric;

2) determinare grafică;

d) Să se reprezinte grafic seria folosind histograma prin dreptunghiuri.

Rezolvare:

Determinarea mărimii intervalului de grupare (se folosește formula lui H.A. Sturges):

$$k = \frac{x_{\max} - x_{\min}}{1 + 3,322 \lg N}$$

în cazul nostru $x_{\max}=260$, $x_{\min}=190$, $N=30$, deci vom avea:

$$k = \frac{260 - 190}{5} = \frac{70}{5} = 14 \text{ cm}$$

Cele 5 intervale de grupare se vor determina astfel:

Se pornește de la valoarea minimală și se adaugă succesiv mărimea intervalului de grupare, determinată anterior. Vom obține distribuția:

Intevale de grupare	Nr. studenți	x_i	$x_i f_i$
190-204	3	197	591
204-218	15	211	3165
218-232	8	225	1800
232-246	3	239	717
246-260	1	253	253
Total	30	*	6524

Determinarea lungimii medii a săriturilor se face cu ajutorul formulei mediei aritmetice ponderate:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{6526}{30} = 217,53 \text{ cm/student}$$

b)Determinarea medianei.

1)Calcul algebric.

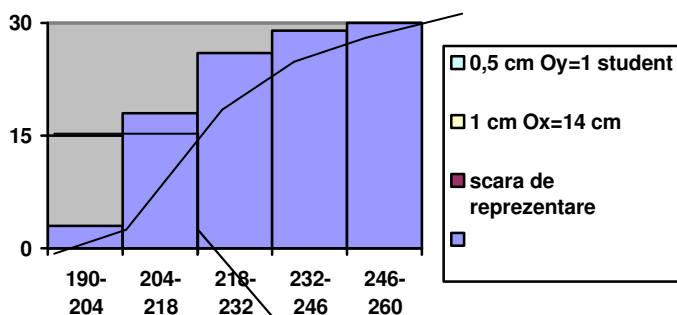
$$M_e = l_i + \left(\frac{\sum f_i}{2} - S_m \right) \cdot \frac{k_m}{f_m} = 204 + (15 - 3) \cdot \frac{14}{15} = 215,2 \text{ cm/student}$$

Intervalul median este: (204-218)

2)Determinare grafică.

Vom construi curba frecvențelor cumulate (ogiva).

Intervale	Frecvențe cumulate
190-204	3
204-218	18
218-232	26
232-246	29
246-260	30



c) Modulul (Dominanta) M_e

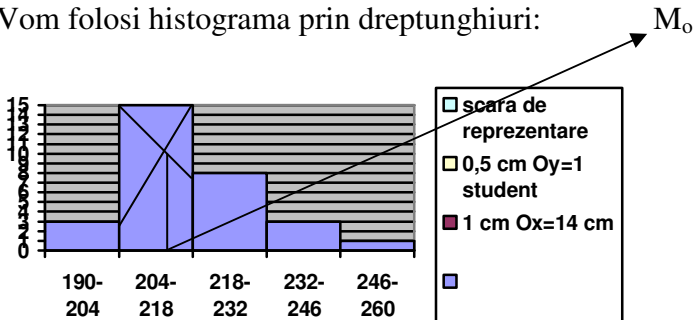
1) Calcul algebric.

$$M_o = l_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) \cdot k_m = 204 + \left(\frac{12}{12 + 7} \right) \cdot 14 = 212,84 \text{ cm}$$

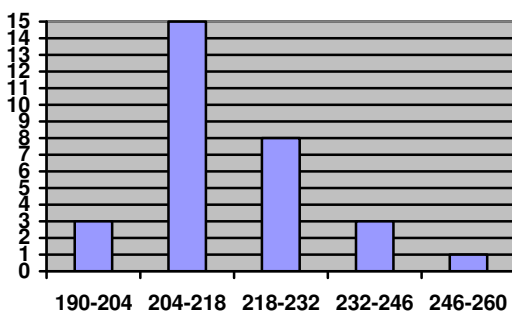
Intervalul modal este intervalul cu frecvența cea mai mare, deci în cazul nostru 204-218.

2) Determinare grafică.

Vom folosi histograma prin dreptunghiuri:



d) Histograma prin dreptunghiuri:



Aplicația 3. În urma inventarierii întreprinse la două depozite A și B ale secției de Educație fizică și sport din cadrul Universității Constantin Brâncuși din Tg-Jiu s-au înregistrat următoarele date statistice privind numărul rachetelor de tenis existând în gestiunea acestora:

Data	A	B
1.01.	38	72
1.02.	32	*
1.03.	50	*
1.04.	48	57
1.05.	38	48
1.06.	30	68
1.07.	32	*
1.08.	45	70

a) Precizați tipul caracteristicii de grupare. Datele prezentate formează o serie statistică? Care este tipul ei?

b) Să se determine numărul mediu de rachete de tenis pentru fiecare deposit.

Rezolvare.

a) Caracteristica de grupare, data, este o caracteristică statistică de timp. Datele statistice prezentate formează o serie statistică cronologică de moment, complexă, alcătuită din două serii cronologice simple de momente de timp. (col. 1+col.2 și col. 1+col.3)

b) Pentru prima serie, deoarece intervalele de momente sunt egale, tipul de medie folosit va fi : media cronologică simplă. Determinarea acesteia presupune parcurgerea celor două etape :

✓ Calculul mediilor mobile:

$$\begin{aligned}\bar{x}_1 &= \frac{38+32}{2} = 35; \bar{x}_2 = \frac{32+50}{2} = 41; \bar{x}_3 = \frac{50+48}{2} = 49; \\ \bar{x}_4 &= \frac{48+38}{2} = 43; \bar{x}_5 = \frac{38+30}{2} = 34; \bar{x}_6 = \frac{30+32}{2} = 31; \\ \bar{x}_7 &= \frac{32+45}{2} = 38,5\end{aligned}$$

✓ Calculul mediei cronologice:

$$\bar{x}_c = \frac{\sum_{i=1}^7 x_i}{7} = \frac{271,5}{7} = 38,75 \approx 39 \text{ rachete tenis}$$

Pentru ce-a de-a doua serie, intervalele dintre momente sunt diferite, deci vom folosi media cronologică ponderată.

Determinarea ei presupune de asemenea parcurgerea celor două etape:

✓ Calculul mediilor mobile:

$$\bar{x}_1 = \frac{72 + 57}{2} = 64,5; \bar{x}_2 = \frac{57 + 48}{2} = 52,5; \bar{x}_3 = 58$$

$$\bar{x}_4 = \frac{48 + 70}{2} = 59;$$

✓ Calculul mediei cronologice:

$$\bar{x}_c = \frac{\bar{x}_1 \cdot 3 + \bar{x}_2 \cdot 1 + \bar{x}_3 \cdot 1 + \bar{x}_4 \cdot 2}{7} = \frac{193,5 + 52,5 + 58 + 118}{7} = 60,29 \approx 60$$

rachete tenis.

Capitolul V

Variația și asimetria

1. Importanța măsurării variabilității valorilor individuale

Pentru a cunoaște și a estima corect principalele tendințe ale unei caracteristici este absolut necesară determinarea domeniului acesteia de variație. Cu cât fenomenele au un grad mai mare de complexitate, cu atât variația (împrăștierea) valorilor individuale este mai mare. Verificarea stabilității și a reprezentativității valorilor înregistrate de indicatorii tendinței centrale este necesară în fundamentarea deciziilor.

Calcularea și analiza indicatorilor variației sau împrăștierea valorilor individuale față de tendința centrală oferă posibilitatea :

a) analiza gradului de omogenitate a datelor din care s-au determinat indicatorii tendinței centrale și verificarea reprezentativității acesteia ;

b) Modul de dispersare a valorilor individuale în interiorul câmpului de variație ;

c) separarea acțiunii factorilor întâmplători de acțiunea factorilor esențiali.

d) compararea în timp și spațiu a mai multor serii de distribuție după caracteristicile independente sau interdependente.

2.Indicatorii variației

Distingem 2 grupe :

1)Indicatori simpli;

2)Indicatori sintetici.

1.)*Indicatorii simplii ai variației.:*

a)*Amplitudinea variației.* Oferă posibilitatea cunoașterii câmpului de variație a unui fenomen. Distingem :

a.1.)*Amplitudinea absolută (A_a).* Se calculează ca diferența între variația maximă și variația minimă a caracteristicii :

$$A_a = x_{max} - x_{min}$$

a.2.)*Amplitudinea relativă (A_r).* Se determină raportând amplitudinea absolută la media caracteristicii analizate, exprimându-se în procente :

$$A_r = \frac{A_a}{\bar{x}} \cdot 100$$

a.3.) *Abaterile individuale absolute* (d_a). Dau posibilitatea cunoașterii structurii variației la nivelul fiecărei unități statistice și se calculează ca diferența între variantele caracteristicii și media lor :

$$d_a = x_i - \bar{x}$$

a.4.) *Abaterile individuale relative* (d_r). Se calculează ca raport între abaterea absolută și medie, exprimându-se în procente:

$$d_r = \frac{d_i}{\bar{x}} \cdot 100 = \frac{x_i - \bar{x}}{\bar{x}} \cdot 100$$

2.) *Indicatorii sintetici ai variației.*

a.1.) *Abaterea medie lineară*. Este calculată ca o medie aritmetică, simplă sau ponderată, a abaterilor absolute ale variantelor caracteristicii de la media lor, în modul, folosind una dintre relațiile :

$$\bar{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} \quad \text{- pentru serii simple}$$

$$\bar{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}| f_i}{\sum_{i=1}^n f_i} \quad \text{- pentru serii de frecvențe absolute}$$

$$\bar{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}| f_i^*}{100} \quad \text{-pentru serii de frecvențe relative}$$

exprimate sub formă de coeficienți.

Cu ajutorul abaterii medii lineare se poate determina intervalul mediu de variație :

$$\bar{x} + \bar{d} \quad \text{-limita superioară}$$

$$\bar{x} - \bar{d} \quad \text{-limita inferioară}$$

a.2.) *Dispersia*. Ca măsură sintetică a variației, dispersia reprezintă media aritmetică, simplă sau ponderată, a pătratului abaterilor valorilor individuale de la tendința lor centrală.

Pentru o serie simplă :

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

Pentru o serie de frecvențe :

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i}$$

Pentru o serie de frecvențe relative:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f_i^*}{100}$$

Proprietățile dispersiei. Deoarece dispersia este calculată ca o medie aritmetică, rezultă că proprietățile mediei aritmetice sunt aplicabile și în cazul dispersiei. Deosebit de importante sunt cele 2 proprietăți care duc la simplificarea calculului :

1.) Dispersia calculată din abaterile variantelor x_i de la o constantă "a" este mai mare decât dispersia reală cu pătratul diferenței dintre medie și constanta "a", adică:

$$\frac{\sum (x_i - a)^2 f_i}{\sum f_i} > \sigma^2 \text{ cu } \left(\bar{x} - a \right)^2$$

sau

$$\sigma^2 = \frac{\sum (x_i - a)^2 f_i}{\sum f_i} - \left(\bar{x} - a \right)^2 \quad (1)$$

2.) Dispersia calculată din abaterile variantelor x_i de la media lor, micșorate în prealabil prin împărțirea cu o constantă k, este mai mică decât dispersia reală, de "k" ori, adică:

$$\frac{\sum \left(\frac{x_i - \bar{x}}{k} \right) \cdot f_i}{\sum f_i} < \sigma^2 \text{ de "k" ori}$$

sau

$$\sigma^2 = \frac{\sum \left(\frac{x_i - \bar{x}}{k} \right) \cdot f_i}{\sum f_i} \cdot k^2 \quad (2)$$

Din (1) și (2) rezultă:

$$\sigma^2 = \frac{\sum \left(\frac{x_i - a}{k} \right) \cdot f_i}{\sum f_i} \cdot k^2 - \left(\bar{x} - a \right)^2$$

După stabilirea poziției “a” în cadrul seriei, pentru rapoartele $\left(\frac{x_i - a}{k} \right)$ se vor trece automat 0, în dreptul poziției lui “a”, -1, -2, -3, etc., deasupra și 1, 2, 3, ... etc., sub zero, analog formulei de calcul simplificat al mediei ponderate.

a.3.) *Abaterea medie pătratică* (Abaterea standard). Se calculează ca medie pătratică, simplă sau ponderată, a abaterilor valorilor individuale de la media lor, după relația:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \quad \text{-pentru serii simple;}$$

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}} \quad \text{-pentru serii de frecvențe absolute;}$$

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x}) \cdot f_i^*}{\sum f_i}} \quad \text{-pentru serii de frecvențe relative,}$$

sub formă de coeficienți;

Vom putea determina intervalul de variație:

$\bar{x} + \bar{d}$ -limita superioară

$\bar{x} - \bar{d}$ -limita inferioară

Intervalul mediu de variație determinat cu ajutorul abaterii standard este mai larg decât cel calculat cu ajutorul abaterii medii lineare, abaterea standard fiind indicatorul preferat în analiza variației fenomenelor sociale.

a.4)*Coeфициentul de variație*. Dă posibilitatea comparării variației valorilor ce nu sunt exprimate în aceeași unitate. Se calculează ca raport între abaterea standard și nivelul mediu al variabilei statistice, exprimându-se în procente:

$$v = \frac{\sigma}{\bar{x}} \cdot 100 \quad \text{sau} \quad v = \frac{\bar{d}}{\bar{x}} \cdot 100$$
$$0 \leq v \leq 100\%$$

Astfel:

$v=0$ rezultă lipsă de variație, valorile sunt egale între ele și egale cu media lor, adică

$$x_1 = x_2 = x_3 = \dots x_n = \bar{x}$$

$v \rightarrow 0$, variația caracteristicii este mică.

$v \rightarrow 100\%$, variația caracteristicii este mare.

Intervalul de valori se poate împărții astfel:

$0 < v \leq 35\%$, variația este mică, media ca indicator al tendinței centrale este semnificativă, abaterile variantelor x_i , de la medie sunt mici, valorile mici gravitează în jurul mediei, colectivitatea este omogenă, adică e formată din unități ce aparțin aceluiași tip calitativ, gruparea ca metodă de sistematizare este bine făcută.

$35\% < v \leq 50\%$, variație relativ mare, ceea ce înseamnă că aspectele prezentate mai sus devin discutabile.

$50\% < v \leq 100\%$, variație foarte mare, media nu este semnificativă, deoarece maschează abateri mari ale termenilor, care sunt poziționați la distanțe mari unul față de altul. Gruparea nu este corect făcută, se recomandă împărțirea colectivității în grupe omogene și determinarea indicatorilor sintetici pentru fiecare grupă.

Coeficientul de variație poate fi folosit ca test de verificare a reprezentativității mediei, distingându-se cazurile:

$0 < v \leq 17\%$, media este strict reprezentativă ;

$17\% < v \leq 35\%$, media este moderat semnificativă;

$35\% < v \leq 50\%$, media este relativ reprezentativă;

$v > 50\%$, media nu este reprezentativă.

3.Indicatorii variației pentru distribuții bidimensionale

În cazul distribuțiilor bidimensionale, pentru analiza variației caracteristicii secundare se folosește ca indicator dispersia.

Variabilitatea valorilor individuale este determinată de acțiunea factorilor esențiali (de grupare) și a factorilor întâmplători (reziduali) care acționează în interiorul fiecărei grupe.

Deosebim următoarele tipuri de dispersie:

a.1.) *Dispersia de grupă*. Se calculează ca o medie aritmetică ponderată a pătratului abaterilor variantelor y_i de la media de grupă, după relația:

$$\sigma_i^2 = \frac{\sum_j \left(y_j - \bar{y}_i \right)^2 \cdot f_{ij}}{f_i}$$

sau

$$\sigma_i^2 = \frac{\sum_j \left(y_j - \bar{y}_i \right)^2 \cdot f_{ij}}{\sum_j f_{ij}}, \text{ unde}$$

y_j – variantele caracteristicii secundare y poziționate în cadrul fiecărei grupe constituită după caracteristica principală x .

$\bar{y}_i$ -media valorilor y_j corespunzătoare fiecărei grupe “ i ”,

$\overline{i = 1, n}$, numărul de grupe constituite după variabila principală x .

$\overline{j = 1, n}$, numărul de grupe constituite după variabila secundară y .

f_{ij} -frecvențele corespunzătoare variabilelor y_j din fiecare grupă i ;

f_i - frecvențele fiecărei grupe “ i ”.

$$\sum_j f_{ij} = f_i$$

$$y_i = \frac{\sum_j y_j \cdot \bar{f}_{ij}}{\sum_j f_{ij}} = \frac{\sum_j y_j f_{ij}}{f_i}$$

Dispersia de grupă sintetizează influența factorilor întâmplători ce determină variația caracteristicii secundare y în cadrul fiecărei grupe constituite după caracteristica principală “ x ”.

a.2.) *Media dispersiilor de grupă*. Sintetizează influența factorilor întâmplători asupra variației secundare y la nivelul colectivității generale. Se calculează ca o medie aritmetică ponderată a dispersiilor de grupă după relația:

$$\sigma^2 = \frac{\sum_i \sigma_i^2 \cdot f_i}{\sum_i f_i}, \text{ unde } f_i \text{ -- frecvențele corespunzătoare grupelor}$$

“ i ”.

$\sum f_i$ – numărul total de unități statistice.

a.3.) *Dispersia dintre grupe*. Sintetizează variația caracteristicii secundare y determinată de acțiunea factorilor esențiali, la nivelul întregii colectivități. Se calculează ca o medie aritmetică ponderată a pătratului abaterilor mediilor de grupă de la media generală, după relația:

$$\delta^2 = \frac{\sum_i \left(y_i - \bar{y}_0 \right)^2 \cdot f_i}{\sum_i f_i}, \text{ unde } \bar{y}_0 \text{ -- media caracteristicii}$$

secundare pentru întreaga colectivitate.

f_i – frecvențele corespunzătoare grupelor “ i ”

$\sum_i f_i = N$ – numărul total al unităților statistice.

a.4) *Dispersia generală*. Sintetizează variația caracteristicii secundare y determinată de acțiunea simultană

atât a factorilor întâmplători cât și a factorilor esențiali, la nivelul colectivității generale. Se calculează ca o medie asimetrică ponderată a pătratului abaterilor variantelor caracteristicii secundare de la media generală, după formula:

$$\sigma_0^2 = \frac{\sum_j \left(y_j - \bar{y}_0 \right)^2 \cdot f_j}{\sum_j f_j}$$

Regula de aur a adunării dispersiilor:

$$\sigma_0^2 = \bar{\sigma}^2 + \delta^2$$

De asemenea putem calcula:

-*coeficientul de determinație*, se evidențiază ponderea variației caracteristicii secundare determinată de acțiunea factorilor esențiali de grupare în variația totală a caracteristicii:

$$R^2 = \frac{\delta^2}{\sigma_0^2} \cdot 100$$

-*coeficientul de nedeterminație*, ce evidențiază problema variației caracteristicii secundare determinată de acțiunea factorilor întâmplători la nivelul întregii colectivități:

$$1 - R^2 = \frac{\bar{\sigma}^2}{\sigma_0^2} \cdot 100$$

4.Indicatorii variabilei aleatoare a lui Bernoulli

În practica statistică se întâlnesc variabile ale căror variante sunt antogonice, una fiind alternativă celeilalte. Pentru prelucrarea și analiza statistică a acestora se folosesc următoarele situații convenționale:

Variantele caracteristicii	Frecvențe absolute f_i	Frecvențe relative f_i^*
$x_1=1$	f_1	$p = \frac{f_1}{N}$
$x_2=0$	f_2	$q = \frac{f_2}{N}$
Total	N	p+q=1

Metodologia de calcul a mediei caracteristicii alternative este aceeași cu cea a mediei aritmetice ponderate, în formula acesteia introducându-se elementele specifice caracteristicii lui Bernoulli.

Deci vom avea:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{x_1 f_1 + x_2 f_2}{f_1 + f_2} = \frac{f_1}{f_1 + f_2} = \frac{f_1}{N} = p$$

Dispersia. În mod analog vom folosi:

$$\begin{aligned}
\sigma^2 &= \frac{\sum_i \left(x_i - \bar{x}\right)^2 f_i}{\sum_i f_i} = \frac{\left(x_1 - \bar{x}\right)^2 f_1 + \left(x_2 - \bar{x}\right)^2 f_2}{f_1 + f_2} = \\
&= \frac{(1-p)^2 f_1 + (0-p)^2 f_2}{f_1 + f_2} = \\
&= \frac{\frac{(1-p)^2 f_1}{f_1 + f_2} + \frac{(0-p)^2 f_2}{f_1 + f_2}}{1} = \frac{q^2 \frac{f_1}{f_1 + f_2} + p^2 \frac{f_2}{f_1 + f_2}}{1} = \\
&= \frac{q^2 \frac{f_1}{N} + p^2 \frac{f_2}{N}}{1} = \frac{q^2 p + p^2 q}{p + q} = \frac{pq(p+q)}{p+q} = pq
\end{aligned}$$

Deci:

$$\sigma^2 = p \cdot q$$

Abaterea standard:

$$\sigma = \sqrt{\sigma^2} = \sqrt{pq}$$

5. Indicatorii asimetriei și curtozisului

Analiza statistică a formelor de distribuție a frecvențelor presupune caracterizarea asimetriei, adică deplasarea valorilor individuale față de anumite valori tipice ale tendinței centrale și a gradului de aplatizarea a curbei frecvențelor.

Momentele sunt indicatori statistici prețioși pentru analiza asimetriei și a gradului de aplatizare a seriilor de

distribuție. Momentele se calculează ca medii aritmetice ponderate ale abaterilor variantelor caracteristicii de la o anumită valoare, folosită ca bază de comparație, abaterile respective fiind luate la puteri diferite.

Deosebim:

a) *Momente ordinare:*

-pentru serii simple

$$\mu_n^* = \frac{\sum_i (x_i - a)^n}{N}$$

-pentru serii de distribuție

$$\mu_n^* = \frac{\sum_i (x_i - a)^n \cdot f_i}{\sum_i f_i}$$

-formula de calcul simplificat:

$$\mu_n^* = \frac{\sum_i \left(\frac{x_i - a}{k} \right)^n \cdot f_i}{\sum_i f_i} \cdot k^n$$

unde n reprezintă ordinul momentului.

a, k-constante alese ca și în cazul formulei de calcul simplificat a mediei aritmetice.

b) *Momente centrate:*

-pentru serii simple:

$$\mu_n = \frac{\sum_i \left(x_i - \bar{x} \right)^n}{N}$$

-pentru serii de distribuție:

$$\mu_n = \frac{\sum_i \left(x_i - \bar{x} \right)^n \cdot f_i}{\sum_i f_i}$$

Avem următoarele relații între cele 2 tipuri de momente:

$$\mu_n = \mu_2^* - (\mu_1^*)^2$$

$$\mu_3 = \mu_3^* - 3\mu_1^* \mu_2^* + 2(\mu_1^*)^3$$

$$\mu_4 = \mu_4^* - 4\mu_1^* \mu_3^* + 6(\mu_1^*)^2 \mu_2^* - 3(\mu_1^*)^4$$

Asimetria. Prin asimetrie înțelegem abaterea de la simetrie a seriilor de distribuție, abatere ce poate fi moderată, generând distribuții moderat asimetrice sau abtere pronunțată, întâlnită în cazul distribuțiilor extrem asimetrice.

Dintre indicatorii folosiți în măsurarea gradului de asimetrie a seriilor de distribuție cei mai des utilizați sunt coeficientul lui Pearson și coeficientul lui Fisher.

Pentru coeficientul lui Pearson avem relația:

$$Kas P = \frac{\bar{x} - \bar{M}_0}{\sigma}$$

$$-1 \leq Kas P \leq +1$$

Interpretare

$Kas P=0$ atunci seria este simetrică;

$Kas P \rightarrow 0$ avem serie cu simetrie mică;

$Kas P \rightarrow \pm 1$ avem asimetrie pronunțată;

$Kas P > 0$ avem asimetrie la stânga sau negativă;

$Kas P < 0$ avem asimetrie la dreapta sau pozitivă.

În practică $-0,3 \leq Kas P \leq +0,3$ indică o asimetrie moderată.

Pentru coeficientul lui Fisher avem:

$$Kas F = \frac{3 \left(\bar{x} - M_0 \right)}{\sigma}$$
$$-3 \leq Kas F \leq +3$$

Cu cât $Kas F \rightarrow 0$ cu atât seria este mai aproape de simetrie.

Alte metode de caracterizare a simetriei:

Dacă:

$\bar{x} = M_e = M_0$ avem serie simetrică;

$M_0 < M_e < \bar{x}$ avem serie asimetrică la stânga;

$\bar{x} < M_e < M_0$ avem serie asimetrică la dreapta;

$\mu_3 = 0$ avem serie simetrică;

$\mu_3 < 0$ avem serie asimetrică la stânga;

$\mu_3 > 0$ avem serie asimetrică la dreapta.

Curtozisul. Reprezentările grafice ale seriilor de distribuție de frecvențe sunt mai mult sau mai puțin aplatizate comparativ cu graficul distribuției normale Gauss-Laplace. Acestea au bolta mai largă sau mai ascuțită, în funcție de gradul de concentrare a frecvențelor în jurul valorilor tipice-medie, mediana, dominanta.

Rădăcinile etimologice ale conceptelor utilizate sunt cuvintele grecești kurtos=cocoșat, platos=larg și léptos=îngust. Gradul de concentrare a frecvențelor în jurul valorilor tipice este cunoscut sub numele de exces.

Aplatizarea este măsurată cu ajutorul următorilor indicatori:

-*coeficientul de boltire*, ce se calculează ca raport între momentul centrat de ordinul 4 și pătratul momentului centrat de ordinul 2, după formula:

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{(\sigma^2)^2}$$

-*curtozisul* (excesul) se determină cu ajutorul relației:

$\gamma_2 = \beta_2 - 3$, 3 fiind valoarea coeficientului de boltire corespunzător repartiției normale.

Vom avea:

$\beta_2 > 3$ implică $\gamma_2 > 0$ implică distribuție lytecustică (cu vârf ascuțit);

$\beta_2 = 3$ implică $\gamma_2 = 0$ implică distribuție normală;

$\beta_2 < 3$ implică $\gamma_2 < 0$ implică distribuție platicustică (cu vârf plat).

6.Aplicații

Aplicația 1. Se dau următoarele date privind rezultatele probei “Săritura în lungime de pe loc”, băieți, de la specializarea Educație fizică și sport din cadrul Universității Constantin Brâncuși din Tg-Jiu.

Nr. crt.	Săritura în lungime de pe loc (cm)	Nr. crt.	Săritura în lungime de pe loc (cm)
1	190	16	260
2	210	17	220
3	225	18	220
4	220	19	215
5	210	20	240
6	215	21	260
7	210	22	200
8	215	23	240
9	210	24	215
10	190	25	210
11	210	26	225
12	210	27	225
13	225	28	220
14	230	29	225
15	235	30	230

- a)Indicatorii simpli și sintetici;
- b)Caracterizați seria d.p.d.v al omogenității;
- c)Calculați coeficientul de asimetrie Pearson și Fisher.

Comentați rezultatele obținute.

Rezolvare.

- a) 1.Indicatorii simpli ai variației.

1.1.Amplitudinea variației:

-absolută: $A_a = x_{max} - x_{min} = 260 - 190 = 70 \text{ cm}$

-relativă: $A_r = \frac{A_a}{\bar{x}} \cdot 100 = \frac{70}{217,53} \cdot 100 = 32,18\%$

1.2.Abaterrea variantelor caracteristicii de la media lor:

-abaterea absolută: $d_a = x_i - \bar{x}$.

Vom avea:

$d_1 = 197 - 217,53 = -20,53 \text{ cm}$

$d_2 = 211 - 217,53 = -6,53 \text{ cm}$

$d_3 = 225 - 217,53 = 7,47 \text{ cm}$

$d_4 = 239 - 217,53 = 21,47 \text{ cm}$

$d_5 = 253 - 217,53 = 35,47 \text{ cm}$

-abterea relativă: $d_r = \frac{d_a}{\bar{x}} \cdot 100$. Valorile se

determină ca mai sus.

2.Indicatorii sintetici ai variației:

2.1.Abaterea medie lineară:

$$\bar{d} = \frac{\sum |x_i - \bar{x}| \cdot f_i}{\sum f_i} = \frac{20,53 \cdot 3 + 6,53 \cdot 15 + 7,47 \cdot 8 + 21,47 \cdot 3 + 35,47 \cdot 1}{30} =$$

$$= \frac{61,59 + 97,95 + 59,76 + 64,41 + 35,47}{30} = 10,63 \text{ cm} \Rightarrow$$

variantele caracteristicii se abat, în medie, cu $\pm 10,63$ cm de la medie.

Intervalul mediu de variație se determină astfel:

$$\bar{x} \pm \bar{d} = \begin{cases} \bar{x} - \bar{d} = 217,53 - 10,63 = 20,69 \text{ cm limita inferioară} \\ \bar{x} + \bar{d} = 217,53 + 10,63 = 228,16 \text{ cm limita superioară} \end{cases}$$

2.2.Dispersia.

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i} = \frac{4991,44}{30} = 166,38. \text{ Datele necesare}$$

calculului sunt centralizate în tabelul:

$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
-20,53	421,48	1264,44
-6,53	42,64	639,6
7,47	55,80	446,4
21,47	460,96	1382,88
35,47	1258,12	1258,12
Total	*	4991,44

3. Abaterea medie pătratică (abaterea standard)

$$\sigma = \sqrt{\sigma^2} = \sqrt{166,38} = 12,89$$

Intervalul mediu de variație calculat pe baza abaterii standard este:

$$\bar{x} \pm \sigma = \begin{cases} \bar{x} - \sigma = 217,53 - 10,63 = 20,69 \text{ cm limita inferioară} \\ \bar{x} + \sigma = 217,53 + 10,63 = 228,16 \text{ cm limita superioară} \end{cases}$$

b) Pentru aprecierea omogenității vom calcula coeficientul de variație:

$$v = \frac{\sigma}{\bar{x}} \cdot 100 = \frac{12,89}{217,53} \cdot 100 = 5,92 < 30\% \Rightarrow \text{colectivitatea este}$$

omogenă, indicatorii calculați sunt reprezentativi.

b) 1. Coeficientul de asimetrie Pearson:

$$K_{as} = \frac{\bar{x} - M_o}{\sigma} = \frac{217,53 - 212,84}{12,89} = 0,36 > 0 \text{ asimetrie}$$

pozitivă moderată.

2. Coeficientul de asimetrie Fisher.

$$K_{as} F = \frac{\left(\bar{x} - M_e \right)}{\sigma} = \frac{3 \cdot (217,53 - 215,2)}{12,89} = 0,54 > 0$$

asimetrie pozitivă moderată.

Aplicația 2.

În cadrul operațiunilor de control a calității produselor dintr-un lot de 5000 clăpări, 250 au fost defecte. Să se determine dispersia și abaterea standard.

Rezolvare.

$$p = \frac{f_1}{N} = \frac{250}{5000} = 0,05$$

$$q = \frac{5000 - 250}{5000} = 0,95$$

Determinarea mediei se face cu ajutorul formulei $\bar{x} = p = 0,05 \Rightarrow$ lotul controlat conține 5% clăpări defecte și respectiv 95% clăpări corespunzătoare din punct de vedere calitativ.

Dispersia

$$\sigma^2 = pq = 0,05 \cdot 0,95 = 0,0475$$

Abaterea standard

$$\sigma = \sqrt{0,0475} = 0,22$$

Aplicația 3.

Se dau următoarele date privind rezultatele probei “alergarea de viteză 30 m” băieți, grupa 101, secția de

Educație fizică și sport din cadrul Universității Constantin Brâncuși din Tg-Jiu.

Rezolvare.

Intervale de grupare (secunde)	Nr. studenți	x_i	$x_i f_i$
4,13-4,37	3	4,25	12,75
4,37-4,62	17	4,493	76,38
4,62-4,90	9	4,735	42,62
4,90-5,11	1	4,983	4,98
Total	30	*	136,73

a) Să se determine timpul mediu al grupei;

b) Să se calculeze mediana și dominantă a grupei;

c) Să se calculeze amplitudinea, dispersia, coeficientul de variație. Interpretare;

d) Să se reprezinte grafic seria.

Rezolvare.

a) Determinarea timpului mediu al grupei.

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{136,73}{30} = 4,56 \text{ secunde/student}$$

b) Mediana.

$$M_e = l_i + \left(\frac{\sum f_i}{2} - \delta_m \right) \cdot \frac{K_m}{J_m} = 4,37 + (15 - 3) \cdot \frac{0,205}{17} = 4,55 \text{ secunde/student}$$

Intervalul median este (4,37-4,62)

Modulul. Intervalul modal este (4,37-4,62)

$$M_o = l_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) \cdot K_m = 4,37 + \left(\frac{14}{14 + 8} \right) \cdot 0,25 = 4,53$$

secunde/student

c) Amplitudinea variației.

-absolută: $A_a = 5,11 - 4,13 = 0,98$ secunde

-relativă: $A_r = \frac{A_a}{\bar{x}} \cdot 100 = 21,49\%$

Dispersia.

$$\sigma^2 = \frac{\sum \left(x_i - \bar{x} \right)^2 \cdot f_i}{\sum f_i} = 0,028$$

Abaterea standard.

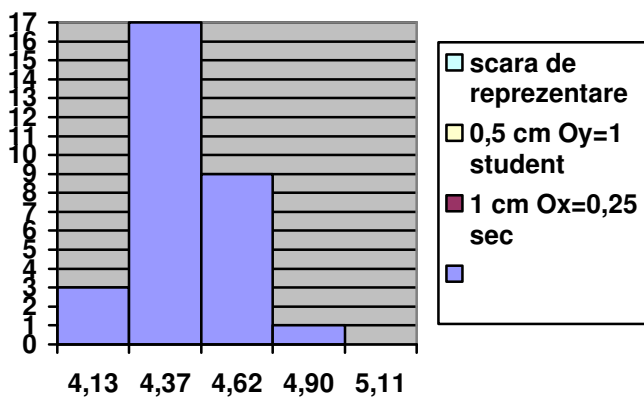
$$\sigma = \sqrt{\sigma^2} = 0,17 \text{ secunde}$$

Coeficientul de variație.

$$v = \frac{\sigma}{\bar{x}} \cdot 100 = \frac{0,17}{4,56} \cdot 100 = 3,73 < 30\% \Rightarrow \text{colectivitatea este}$$

omogenă, media, mediana, dominantă sunt reprezentative.

d) Pentru reprezentarea grafică vom folosi histograma prin dreptunghiuri:



Aplicatia 4. Se dau urmatoarele date privind lungimea săriturilor studenților grupei 202 sport și vârsta acestora. Analizați importanța factorului vârstă în variația lungimii medii a săriturilor.

Tabel 3. Y-caracteristica secundară (lungimea săriturii), X-caracteristica principală (vârsta)

Y \ X	120-134	134-148	148-162	162-176	176-190	Total
19	1	3	4	-	-	8
20	-	-	10	7	1	18
21	1	-	2	-	1	4
Total	2	3	16	7	2	30

Calculul dispersiilor de grupă se face cu ajutorul formulei :

$$\sigma^2_i = \frac{\sum_j (y_i - \bar{y}_i)^2 f_{ij}}{\sum_j f_{ij}}$$

Pentru grupa 1 (19 ani) avem :

$$\bar{y}_1 = \frac{\sum_j y_j f_{ij}}{\sum_j f_{ij}} = \frac{127 \times 1 + 141 \times 3 + 155 \times 4}{8} = \frac{127 + 423 + 620}{8} = 146,25 \text{ cm}$$

lungimea medie a săriturii grupei 1

$$\begin{aligned} \sigma^2_1 &= \frac{\sum_j (y_i - \bar{y}_1)^2 f_{ij}}{\sum_j f_{ij}} = \\ &= \frac{(127 - 146,25)^2 \times 1 + (141 - 146,25)^2 \times 3 + (155 - 146,25)^2 \times 4}{8} = 94,93 \end{aligned}$$

Pentru grupa 2 (20 ani) avem :

$$\bar{y}_2 = \frac{\sum_j y_j f_{ij}}{\sum_j f_{ij}} = \frac{155 \times 10 + 169 \times 7 + 183 \times 1}{18} = 162 \text{ cm} \quad \text{lungimea}$$

medie a săriturii grupei 2

$$\sigma^2_2 = \frac{\sum_j (y_i - \bar{y}_2)^2 f_{ij}}{\sum_j f_{ij}} = \frac{(155 - 162)^2 \times 10 + (169 - 162)^2 \times 7 + (183 - 162)^2 \times 1}{18} = 70,77$$

Pentru grupa 3 (21 ani):

$$\bar{y}_3 = \frac{\sum_j y_j f_{ij}}{\sum_j f_{ij}} = 155 \text{ cm} \quad \text{lungimea medie a sariturii grupei 3}$$

$$\sigma^2_3 = \frac{\sum_j (y_j - \bar{y}_3)^2 f_{ij}}{\sum_j f_{ij}} = 392$$

Calculul mediei dispersiilor de grupă :

$$\bar{\sigma}^2 = \frac{\sum_i \sigma^2_i f_i}{\sum_i f_i} = 120,04$$

Dispersia dintre grupe :

$$\delta^2 = \frac{\sum_i (\bar{y}_i - \bar{y}_0)^2 f_i}{\sum_i f_i} = 46,33$$

$$\bar{y}_0 = \frac{\sum_j y_j f_j}{\sum_j f_j} = 156,87 \text{ cm/student, lungimea medie a săriturii}$$

întregii colectivități

Dispersia generală

$$\sigma^2_0 = \frac{\sum_j (y_i - \bar{y}_0)^2 f_j}{\sum_j f_j} = 166,37$$

Se observă că este verificată regula de aur a adunării dispersiilor :

$$\sigma^2_0 = \delta^2 + \bar{\sigma}^2$$

Ponderea factorului esențial de grupare (vârsta) în variația totală a caracteristicii Y se determină calculând coeficientul de determinație :

$$R^2 = \frac{\delta^2}{\sigma^2_0} \times 100 = 27,85\%$$

Ponderea factorilor întâmplători în variația totală a caracteristicii va fi :

$$1 - R^2 = \frac{\bar{\sigma}^2}{\sigma^2_0} \times 100 = 72,15\%$$

Cum $1 - R^2 > R^2$, rezultă că vârsta nu este factor determinant în variația caracteristicii secundare lungimea săriturilor.

Capitolul VI

Regresia și corelația statistică

1.Introducere

Fenomenele sociale nu evoluează independent, ci sunt în directă legătură cu alte variabile sociale, variabilitatea lor fiind determinată de acțiunea simultană a mai multor factori.

Raporturile de cauzalitate dintre fenomenele sociale pot fi cuantificate, analizate și interpretate cu ajutorul corelației. Cu ajutorul metodelor statistice specifice putem evidenția existența raporturilor de cauzalitate dintre fenomene, putem determina influența fiecărui factor asupra variabilității globale a variabilei efect, putem cunoaște intensitatea legăturilor cauzale dintre fenomen și putem estima tendințele evolutive ale corelației dintre fenomene. Statistica oferă o serie de metode de analiză a dependențelor dintre două sau mai multe fenomene. Cele mai des uzitate sunt regresia și corelația. Acestea studiază dependența dintre o variabilă rezultativă, notată în general cu

Y, și una sau mai multe variabile factoriale independente notate cu X.

Metoda regresiei dă posibilitatea cunoașterii formei analitice a dependenței unei variabile de alte sau alte variabile, pe când corelația evidențiază gradul, intensitatea cu care o variabilă rezultat este dependentă de o altă sau alte variabile cauză.

2.Clasificarea legăturilor statistice.

Legăturile statistice se pot clasifica după mai multe criterii:

a)după numărul variabilelor ce alcătuiesc cuplul corelativ avem:

- ✓ *legături simple*, care studiază dependența dintre o variabilă dependentă Y și o singură variabilă independentă, factorială x.

Exemplu. Dependența dintre numărul abdomenelor efectuate și vârsta sportivilor.

- ✓ *legături multiple*, ce studiază dependența dintre o variabilă rezultativă Y și două sau mai multe variabile factoriale $x_1, x_2, \dots, x_n$.

Exemplu. Dependența dintre numărul flotărilor și sexul, vârsta și greutatea sportivilor.

b) După semnul legăturilor factoriale avem:

- ✓ *legături directe*- când variabila factorială x (sau variabilele factoriale $x_1, x_2, \dots, x_n$).

Exemplu. Corelația dintre săritura în înălțime și detentă.

- ✓ *legături inverse*, când variabila rezultativă y se modifică în sens opus modificării variabilei sau variabilelor factoriale.

Exemplu. Numărul de flotări și vârsta sportivilor.

c) După forma legăturilor cauzale deosebim:

- ✓ *legături lineare* – în cadrul cărora variabila rezultativă y înregistrează o tendință lineară.
- ✓ *legături nelineare*, în cadrul cărora variabilitatea rezultativă y înregistrează o tendință curbilinie (parabolică, exponențială, hiperbolică, etc.)

Cunoașterea tipurilor legăturilor statistice este deosebit de importantă pentru estimarea tendințelor evolutive ale fenomenelor efect în strânsă legătură cu variabilitatea factorilor determinați.

3. Metode elementare de caracterizare a legăturilor dintre variabile.

Principalele metode elementare de analiză a legăturilor dintre variabile sunt: metoda tabelului de corelație, metoda grafică, metoda grupărilor și metoda seriilor paralele independente.

3.1. Metoda tabelului de corelație

Tabelul de corelație este tabelul cu dublă intrare folosit în prezentarea seriilor de distribuție bidimensionale, rolul variabilei rezultative fiind preluat de caracteristica secundară Y, iar cel al variabilei factoriale de caracteristica principală X.

Cu ajutorul acestei metode pot fi reliefate următoarele aspecte:

- ✓ *existența corelației* – este evidențiată de modul de dispunere a frecvențelor f_{ij} în cadrul tabelului. Dacă aceste frecvențe se concentrează de-a lungul uneia dintre diagonalele tabelului înseamnă că între cele două variabile există corelație. Dacă acestea se dispun haotic în interiorul tabelului, între cele două variabile nu există corelație.

- ✓ *semnul corelației*, este evidențiat de direcția și sensul deplasării frecvențelor f_{ij} în interiorul tabelului. Dacă acestea se deplasează din zona valorilor mici ale variabilei rezultative Y spre zona valorilor mari pe măsură ce valorile factorialului X cresc, corelația este directă. Dacă însă frecvențele f_{ij} se deplasează din zona valorilor mari ale rezultativei Y spre zona valorilor mici în timp ce valorile factorialei X cresc, corelația este inversă.
- ✓ *forma corelației*. Este dată de forma fâșiei formată de frecvențele f_{ij} .
- ✓ *intensitatea corelației*. Este evidențiată de dimensiunea fâșiei frecvențelor f_{ij} . Cu cât aceasta este mai îngustă cu atât corelația este mai mare, cu cât fâșia este mai largă, cu atât intensitatea corelației este mai mică. Dacă frecvențele f_{ij} se situează pe una din diagonalele tabelului, atunci corelația este maximă, adică modificările variabilei Y sunt determinate exclusiv de variabila factorială X.

3.2. Metoda grafică.

Presupune folosirea corelogramei în analiza legăturii dintre variabile.

Graficul se construiește astfel: pe axa Ox se reprezintă valorile variabilei factoriale X în funcție de scara de reprezentare aleasă, iar pe axa Oy se trec valorile variabilei rezultative Y . Prin unirea punctelor corespunzătoare coordonatelor xy se obține corelograma.

Cu ajutorul corelogramei se poate reliefa:

a)*existența corelației*. Este evidențiată de existența unghiului format de linia de tendință a corelogramei cu orizontala. Dacă acest unghi are o valoare diferită de 0, atunci între cele două variabile există corelație, dacă însă valoarea acestui unghi este 0, variabilele sunt independente, corelația neexistând.

b)*sensul corelației*. Este dat de sensul corelogramei. Dacă punctele xy se dispun pe o corelogramă a cărei tendință este ascendentă, atunci între cele două variabile există o corelație directă, dacă punctele xy se dispun pe o corelogramă descendentă, atunci între cele două variabile există o corelație inversă.

c)*forma corelației*. Dacă oscilațiile corelogramei au aproximativ aceeași amplitudine, atunci corelația este lineară, dacă amplitudinile acestor oscilații sunt diferite, sugerând o schimbare de tendință, atunci corelația este nelineară, (de tip parabolic, exponențial, logistic etc.).

d) *intensitatea corelației*. Este dată de mărimea unghiului format de linia de tendință a corelogramei cu orizontala. Cu cât mărimea acestui unghi este mai mare cu atât corelația este mai intensă. Dacă acest unghi este egal cu zero, atunci între cele două variabile nu există corelație.

3.3. Metoda grupărilor

Constă în constituirea de grupe omogene în funcție de o variabilă factorială. Pentru fiecare grupă astfel constituită se calculează medii pe baza datelor numerice ce caracterizează variabila rezultativă și mărimi relative. Aprecierea existenței și a formei corelației se face prin compararea variabilității variabilei factoriale cu indicatorii calculați pentru variabila rezultativă.

3.4. Metoda seriilor paralel independente

Pe baza datelor numerice ale variabilelor cuplului corelativ se ordonează datele în funcție de caracteristica factorială Y , crescător sau descrescător, și se observă modul în care se aranjează valorile rezultativei Y .

Dacă variabila factorială X este ordonată crescător și variabila rezultativă Y se ordonează aproximativ crescător, corelația este directă. Dacă variabila Y se ordonează

aproximativ descrescător, în timp ce variabila X este ordonată crescător, corelația este inversă. Dacă însă variabila Y nu înregistrează o tendință de ordonare, fie crescătoare, fie descrescătoare, în timp ce caracteristica factorială X este ordonată crescător, rezultă că între cele două variabile nu există corelație.

4.Regresia și corelația lineară simplă

Metodele analitice folosite în aprecierea legăturii dintre variabile sunt regresia și corelația. Presupunând că avem un cuplu corelativ format dintr-o variabilă Y și o variabilă factorială X vom avea:

4.1.Regresia lineară simplă

Metoda regresiei oferă posibilitatea caracterizării legăturii de cauzalitate dintre variabile (factoriale și rezultative) prin intermediul uneia dintre funcțiile statistico matematice.

În cazul regresiei lineare simple vom folosi o funcție lineară pentru evidențierea legăturii dintre cele două variabile. Vom avea:

$$Y_{x_i} = a + bx_i, \text{ unde } a \text{ și } b \text{ sunt parametrii de regresie.}$$

a-reprezintă nivelul funcției de regresie în punctul $x=0$ (în sens geometric este ordonată la origine). Arată la ce nivel ar fi ajuns valoarea variabilei rezultative dacă toți factorii, mai puțin cel înregistrat, ar fi avut o acțiune constantă asupra ei.

b-Este numit coeficient de regresie și arată cu cât se modifică, în medie, variabila rezultativă la o modificare cu o unitate a factorului X. În sens geometric, b reprezintă panta dreptei de regresie. În sens statistic, dacă $b>0$ vom avea corelație directă, $b<0$ corelație inversă, $b=0$ lipsă de corelație.

x_i -nivelurile variabilei independente (factoriale) X

Y_{x_i} -nivelurile variabilei dependente (rezultativa) Y

Pentru a determina nivelurile parametrilor a și b și cu ajutorul lor valorile ajustate ale variabilei Y, se folosește metoda celor mai mici pătrate. Condiția de minim impusă de această metodă este:

$$\sum_i (y_i - Y_{x_i})^2 \rightarrow \min \Leftrightarrow \sum_i (y_i - a - bx_i)^2 \rightarrow \min$$

$$\text{Rezultă că: } \begin{cases} \frac{\partial f}{\partial a} = 0 \\ \frac{\partial f}{\partial b} = 0 \end{cases} \Leftrightarrow \begin{cases} 2 \sum_i (y_i - a - bx_i)(-1) = 0 \\ 2 \sum_i (y_i - a - bx_i)(-x_i) = 0 \end{cases} \Leftrightarrow$$

$$\begin{cases} -2 \sum_i y_i + 2 \sum_i a + 2 \sum_i b x_i = 0 \\ -2 \sum_i y_i x_i + 2 \sum_i a x_i + 2 \sum_i b (x_i)^2 = 0 \end{cases} \Leftrightarrow \begin{cases} \sum_i y_i = n a + b \sum_i x_i \\ \sum_i x_i y_i = a \sum_i x_i + b \sum_i x_i^2 \end{cases}$$

Pe baza acestui sistem se determină parametrii a și b și implicit valorile ajustate Y_{x_i} .

Regresia simplă lineară se folosește atunci când avem:

- ✓ Număr mic de informații (negrupate) în care se dau valorile perechii (x_i, y_i) sub forma a 2 serii paralele independente.
- ✓ Număr mare de informații.
- ✓ Sistematizare prin grupare simplă

Dacă cuplul corelativ (x_i, y_i) are frecvențe comune f_i , vom avea sistemul de ecuații normale:

$$\begin{cases} \sum_i y_i f_i = a \sum_i f_i + b \sum_i x_i f_i \\ \sum_i x_i y_i f_i = a \sum_i x_i f_i + b \sum_i x_i^2 f_i \end{cases}$$

Dacă avem sistematizare prin gruparea combinată $(x_i, y_i, f_i, f_j, f_{ij})$, sistemul de ecuații normale va fi:

$$\begin{cases} \sum_i y_i f_i = a \sum_j \sum_i f_{ij} + b \sum_i x_i f_i \\ \sum_i \sum_j x_i y_j f_{ij} = a \sum_i x_i f_i + b \sum_i x_i^2 f_i \end{cases}$$

4.2. Corelația lineară simplă

Pentru determinarea intensității corelației simple lineare se folosește coeficientul de corelație lineară al lui Pearson, coeficient ce se determină ca o medie aritmetică simplă a produsului abaterilor normale ale valorilor x_i și y_i după formula:

$$r_{y/x} = \frac{\sum_i \left(\frac{x_i - \bar{x}}{\sigma_x} \right) \left(\frac{y_i - \bar{y}}{\sigma_y} \right)}{n}, \text{ unde } \sigma_x, \sigma_y - \text{abaterile standard}$$

corespunzătoare variabilei rezultative Y și variabilei factoriale X.

$$r_{y/x} = \frac{n \sum_i x_i y_i - \sum_i x_i \cdot \sum_i y_i}{\sqrt{\left[n \sum_i x_i^2 - \left(\sum_i x_i \right)^2 \right] \cdot \left[n \sum_i y_i^2 - \left(\sum_i y_i \right)^2 \right]}}$$

În practica statistică se folosesc frecvent următoarele formule:

✓ pentru serii de distribuție bidimensionale:

$$r_{y/x} = \frac{\sum \sum \left(x_i - \bar{x} \right) \left(y_i - \bar{y} \right) \cdot f_{ij}}{\sum f \sigma_x \sigma_y}$$

sau

$$r_{y/x} = \frac{\sum \sum f_{ij} \sum x_i y_i \cdot f_{ij} - \sum x_i \cdot f_i \sum y_i f_j}{\sqrt{\left[\sum f_i \sum x_i^2 \cdot f_i - (\sum x_i f_i)^2 \right] \cdot \left[\sum f_i \sum y_i^2 \cdot f_i - (\sum y_i f_i)^2 \right]}}$$

$$r_{y/x} \in [-1, 1]$$

Interpretarea coeficientului Pearson este următoarea:

Dacă:

$r_{y/x} > 0$, legătura este directă;

$r_{y/x} < 0$, legătura este inversă;

$r_{y/x} = 0$, legătură de tip funcțional.

Pentru o interpretare cât mai exactă, intervalul $[-1, 1]$ poate fi divizat astfel:

$0 \leq r_{y/x} < 0,2$, legătură foarte slabă;

$0,2 \leq r_{y/x} \leq 0,5$, legătură slabă;

$0,5 \leq r_{y/x} < 0,75$, legătură de intensitate medie;

$0,75 \leq r_{y/x} < 1$, legătură funcțională.

Precizare. Coeficientul de corelație Pearson nu este indicat în cazul legăturilor nelineare, fiind nesemnificativ, în

aceste cazuri folosindu-se raportul de corelație $R_{y/x}$, ca indicator al legăturii dintre variabile. Dacă legătura este lineară, putem calcula ambii indicatori, relația $r_{y/x}=R_{y/x}$ putându-se folosi ca test de verificare a linearității.

5.Regresia și corelația nelineară simplă

5.1.Regresia nelineară simplă

În cazul regresiei lineare variabilitatea rezultativei Y se produce în progresie geometrică exponențială, etc., tendința evolutivă a acesteia fiind de tip parabolic, hiperbolic, exponențial, logaritm, logistic, etc.

5.1.1.Regresia simplă exponențială

Funcția de regresie este:

$$Y_{x_i} = ab^{x_i}, \text{ unde } a \text{ și } b \text{ parametrii de regresie}$$

b -are caracter de indice mediu și arată modificarea (de câte ori crește sau scade) variabila rezultativă ca urmare a modificării factorialei X.

Logaritmând ecuația funcției de regresie vom avea:

$$\log Y_{x_i} = \log a + \log b$$

Pentru determinarea parametrilor a și b vom folosi metoda celor mai mici pătrate:

$$\sum_i (y_i - y_{x_i})^2 \rightarrow \min \Leftrightarrow \sum_i (y_i - ab^{x_i})^2 \rightarrow \min \Leftrightarrow$$

$$\begin{cases} \frac{\partial f}{\partial a} = 0 \\ \frac{\partial f}{\partial b} = 0 \end{cases} \Leftrightarrow \begin{cases} \sum_i \log y_i = u \log a + \log b \sum_i x_i \\ \sum_i x_i \log y_i = \log a \sum_i x_i + \log b \sum_i x_i^2 \end{cases}$$

Pe baza sistemului se determină valorile $\log a$ și $\log b$, nivelurile $\log Y_{x_i}$, urmând ca nivelurile estimative Y_{x_i} să se obțină prin antilogaritmare.

Interpretare.

Dacă:

- ✓ $b \neq 1$, între cele două variabile există corelație;
- ✓ $b = 1$, variabilele sunt independente;
- ✓ $b > 1$, corelația este directă;
- ✓ $b < 1$, corelația este inversă.

1.2.Regresia simplă parabolică de ordinal 2

Folosirea parabolei de ordin 2 este recomandată în cazul analizei regresiiilor în cadrul cărora variabilitatea rezultativei Y înregistrează o valoare maximă sau una minimă în zona centrală, prezentând unul sau mai multe puncte de maxim sau de minim între valorile extreme.

Funcția de regresie este:

$$Y_{x_i} = a + bx_i + cx_i^2, \text{ unde } a, b, c, x_i \text{ --au aceeași}$$

semnificație ca la modelul linear.

Y_{x_i} , sunt valorile ajustate.

Pentru determinarea a, b, c se apelează la metoda celor mai mici pătrate. Vom obține sistemul de ecuații normale :

$$\begin{cases} \sum_i y_i = na + b \sum_i x_i + c \sum_i x_i^2 \\ \sum_i x_i y_i = a \sum_i x_i + b \sum_i x_i^2 + c \sum_i x_i^3 \\ \sum_i x_i^2 y_i = a \sum_i x_i^2 + b \sum_i x_i^3 + c \sum_i x_i^4 \end{cases}$$

Interpretare.

Dacă:

- ✓ $b \neq 0$, există corelație între variabile;
- ✓ $b > 0$, corelația este directă;
- ✓ $b < 0$, corelația este inversă.

5.1.3.Regresia simplă hiperbolică

Folosirea funcției hiperbolice este recomandată pentru corelațiile în cadrul cărora variabila rezultativă Y înregistrează tendințe regresive.

Funcția de regresie este: $Y_{x_i} = a + b \frac{1}{x_i}$ cu aceleași

semnificații ca cele de mai sus.

Aplicând metoda celor mai mici pătrate vom obține sistemul de ecuații normale:

$$\begin{cases} \sum_i y_i = na + b \sum_i \frac{1}{x_i} \\ \sum_i y_i \frac{1}{x_i} = a \sum_i \frac{1}{x_i} + b \sum_i \frac{1}{x_i^2} \end{cases}$$

5.2. Corelația nelineară

Pentru determinarea intensității corelației nelineare se folosește raportul de corelație. Calcularea raportului de corelație presupune descompunerea dispersiei totale a variabilei rezultative în dispersia valorilor empirice față de valorile teoretice calculate cu ajutorul funcției de regresie Y_{x_i} și dispersia valorilor teoretice față de media valorilor empirice.

Vom avea:

$$\left(y_i - \bar{y} \right) = \left(Y_{x_i} - \bar{y} \right) + \left(y_i - Y_{x_i} \right)$$

Însumând și ridicând la pătrat:

$$\sum \left(y_i - \bar{y} \right)^2 = \sum \left(Y_{x_i} - \bar{y} \right)^2 + \sum \left(y_i - Y_{x_i} \right)^2$$

Vom avea astfel:

- ✓ dispersia totală: $\sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n}$
- ✓ dispersia sistemică ce măsoară variația datorată influenței factoriale X:

$$\sigma_{y/x}^2 = \frac{\sum (Y_{x_i} - \bar{y})^2}{n}$$

- ✓ dispersia reziduală ce măsoară variația neexplicată

$$\sigma_{y/r}^2 = \frac{\sum (y_i - Y_{x_i})^2}{n}$$

Între dispersii există relația: $\sigma_y^2 = \sigma_{y/x}^2 + \sigma_{y/r}^2$

denumită regula de aur a dispersiilor.

Ponderea influenței factorialei X în variabilitatea rezultativei Y se determină cu ajutorul coeficientului de determinație:

$$R_{y/x}^2 = \frac{\sigma_{y/x}^2}{\sigma_y^2}$$

Ponderea influenței factorialelor neînregistrate în modelul de regresie asupra rezultativei Y este dată de coeficientul de nedeterminație

$$R_{y/r}^2 = \frac{\sigma_{y/r}^2}{\sigma_y^2}$$

Dacă $R_{y/x}^2 > R_{y/r}^2$, atunci variabila factorială este determinată, indicând o legătură intensă.

$$R_{y/x}^2 + R_{y/r}^2 = 1$$

$$R_{y/x} = \sqrt{R_{y/x}^2} = \sqrt{1 - R_{y/r}^2} = \sqrt{1 - \frac{\sigma_{y/r}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sum (y_i - Y_{x_i})^2}{\sum (y_i - \bar{y})^2}}$$

$$R_{y/x} \in [0,1]$$

Interpretare.

Dacă:

$\sigma_{y/x}^2 = \sigma_y^2 \Rightarrow R_{y/x} = 1$ -legătură funcțională între cele două variabile;

$\sigma_{y/r}^2 = \sigma_y^2 \Rightarrow R_{y/x} = 0$ -lipsă de legătură, variabile independente;

$\sigma_{y/x}^2 \rightarrow 0 \Rightarrow R_{y/x} = 0$ -legătură slabă;

$\sigma_{y/x}^2 \rightarrow 0 \Rightarrow R_{y/x} \rightarrow 0$ -legătură slabă;

$\sigma_{y/r}^2 \rightarrow 0 \Rightarrow R_{y/x} \rightarrow 1$ -legătură puternică, intensă.

Dacă avem ca funcție de regresie o dreaptă de ecuație

$$Y_{x_i} = a + bx_i \text{ vom avea:}$$

$$R_{y/x} = \sqrt{1 - \frac{\sum_i y_i^2 - a \sum_i y_i - b \sum_i x_i y_i}{\sum_i y_i^2 - \left(\frac{\sum_i y_i}{n}\right)^2}} = \sqrt{\frac{a \sum_i y_i + b \sum_i x_i y_i - \left(\frac{\sum_i y_i}{n}\right)^2}{\sum_i y_i^2 - \left(\frac{\sum_i y_i}{n}\right)^2}}$$

În cazul seriilor simple:

$$R_{y/x} = \sqrt{\frac{a \sum y_i f_i - b \sum x_i y_i \cdot f_{ij} - \frac{(\sum y_i f_i)^2}{\sum f_i}}{\sum y_i^2 \cdot f_i - \frac{(\sum y_i f_i)^2}{\sum f_i}}}$$

Pentru distribuții bidimensionale:

$$R_{y/x} = \sqrt{\frac{a \sum y_j f_j + b \sum x_i y_j \cdot f_{ij} - \frac{(\sum y_j f_j)^2}{\sum \sum f_{ij}}}{\sum y_j^2 \cdot f_j - \frac{(\sum y_j f_j)^2}{\sum \sum f_{ij}}}}}$$

În cazul modelului exponențial:

$$R_{y/x} = \sqrt{1 - \frac{\sum [y_i - (ab^{x_i})]^2}{\sum (y_i - \bar{y})^2}}$$

În cazul modelului parabolic:

$$R_{y/x} = \sqrt{1 - \frac{\sum_i \left[y_i - \left(a + bx_i + cx_i^2 \right) \right]^2}{\sum_i \left(y_i - \bar{y} \right)^2}}$$

În cazul modelului hiperbolic:

$$R_{y/x} = \sqrt{1 - \frac{\sum_i \left[y_i - \left(a + \frac{1}{x_i} b \right) \right]^2}{\sum_i \left(y_i - \bar{y} \right)^2}}$$

6. Regresia și corelația multiplă

6.1. Regresia multiplă

Regresia multiplă poate fi exprimată printr-o funcție lineară sau o funcție curbilinie.

În cazul regresiei lineare multiple o variabilă rezultativă Y este exprimată în funcție de două sau mai multe variabile factoriale $x_1, x_2, \dots, x_n$ cu ajutorul funcției lineare:

$$Y_{x_1, x_2, x_3, \dots, x_n} = a + b_1 x_1 + b_2 x_2 + \dots + b_n x_n$$

Pentru corelația hiperbolică vom avea:

$$Y_{x_1, x_2, x_3, \dots, x_n} = a + b_1 \frac{1}{x_1} + b_2 \frac{1}{x_2} + \dots + b_n \frac{1}{x_n}$$

Pentru corelația parabolică:

$$Y_{x_1, x_2, x_3, \dots, x_n} = a + b_1 x_1 + c_1 x_1^2 + b_2 x_2 + c_2 x_2^2 + \dots + b_n x_n + c_n x_n^2$$

Pentru corelația exponențială:

$$Y_{x_1, x_2, x_3, \dots, x_n} = ab_1^{x_1} b_2^{x_2} b_3^{x_3} \dots b_n^{x_n}$$

6.2.1. Regresia multiplă lineară

Modelul linear este:

$$Y_{x_1, x_2, \dots, x_n} = a + b_1 x_1 + b_2 x_2 + \dots + b_n x_n,$$

unde:

a este parametrul funcției de regresie ce arată influența factorilor neînregistrați în model, presupuși a avea o acțiune constantă;

$b_1, b_2, \dots, b_n$ – coeficienții de regresie ce arată modificarea variabilei Y (rezultativa), când una dintre factoriale se modifică cu o unitate.

Pentru a determina parametrii a și $b_1, b_2, \dots, b_n$ vom apela la metoda celor mai mici pătrate:

$$\sum_i (y_i - Y_{x_i})^2 \rightarrow \min$$

deci vom avea:

$$\sum_i (y_i - a - b_1 x_1 - b_2 x_2 - \dots - b_n x_n)^2 \rightarrow \min \Rightarrow$$

$$\left\{ \begin{array}{l} \frac{\partial f}{\partial a} = 0 \\ \frac{\partial f}{\partial b_1} = 0 \\ \frac{\partial f}{\partial b_2} = 0 \\ \dots\dots\dots \\ \frac{\partial f}{\partial b_n} = 0 \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} \sum y_i = na + b_1 \sum x_i + b_2 \sum x_2 + \dots + b_n \sum x_n \\ \sum y_i x_i = a \sum x_i + b_1 \sum x_i^2 + \dots + b_n \sum x_i x_n \\ \dots\dots\dots \\ \sum \sum y_i x_i = a \sum \sum x_i + b_1 \sum \sum x_i x_1 + \dots + b_n \sum \sum x_i x_n \\ \dots\dots\dots \\ \sum x_n y_n = a \sum x_n + b_1 \sum x_1 x_n + b_2 \sum x_2 x_n + \dots + b_n \sum x_n^2 \end{array} \right.$$

Determinând parametrii de regresie vom putea afla valorile ajustate Y_{x_i} .

Interpretare.

Dacă:

$b_1, b_2, \dots, b_n > 0$, legătura dintre variabile este directă ;

$b_1, b_2, \dots, b_n < 0$, legătura dintre variabile este inversă ;

$b_1, b_2, \dots, b_n = 0$, nu există legătură între variabile.

Modelul bifactorial. Este un caz particular al regresiei multiple lineare, cuplu corelativ cuprinzând o variabilă Y și două factoriale X_1, X_2 .

Folosind metoda celor mai mici pătrate vom avea sistemul de ecuații normale:

$$\begin{cases} \sum y_i = na + b_1 \sum x_1 + b_2 \sum x_2 \\ \sum x_1 y_i = a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 \\ \sum x_2 y_i = a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 \end{cases}$$

6.2. Corelație multiplă

Pentru determinarea intensității corelației multiple se folosește raportul de corelație multiplă:

$$R = R_{y/x_1, x_2, \dots, x_n} = \sqrt{1 - \frac{\sum (y_i - Y_{x_1, x_2, \dots, x_k})^2}{\sum (y_i - \bar{y})^2}}$$

$$R \in [0,1]$$

$$R > R_{y/x_i}, i = 1, n$$

Dacă ridicăm la pătrat coeficientul de corelație multiplă obținem coeficientul de determinație multiplă, ce exprimă ponderea influenței simultane a factorialelor incluse în model asupra variabilei rezultative.

$1-R^2$ -coeficientul de nedeterminație. Arată ponderea influenței simultane a celorlalți factori neinduși în model.

Fie x_1, x_2 două variabile factoriale independente. Avem:

$$R_{y/x_1, x_2}^2 = R_{y/x_1}^2 + R_{y/x_2}^2$$

Dar cum pentru legături lineare:

$$R_{y/x} = r_{y/x} \Rightarrow R_{y/x_1, x_2} = \sqrt{r_{y/x_1}^2 + r_{y/x_2}^2} \Rightarrow$$

$$R_{y/x_1, x_2}^2 = r_{y/x_1}^2 + r_{y/x_2}^2 \Rightarrow R_{y/x_1, x_2} = \sqrt{r_{y/x_1}^2 + r_{y/x_2}^2}$$

Cum însă, în practică variabilele factoriale sunt de cele mai multe ori independente, vom folosi formula:

$$R_{y/x_1, x_2} = \sqrt{\frac{r_{y/x_1}^2 + r_{y/x_2}^2 - 2 \cdot r_{y/x_1} \cdot r_{y/x_2} \cdot r_{x_1/x_2}}{1 - r_{x_1/x_2}^2}}$$

7.Corelația parțială

Pentru determinarea intensității legăturii cauzale dintre variabila rezultativă Y și o singură variabilă factorială X, cu excluderea influenței celorlalte factoriale se folosește corelația parțială. Pentru aceasta se folosesc coeficienții de corelație parțială care, pentru un cuplu corelativ cu trei variabile se calculează cu ajutorul formulelor:

- ✓ pentru corelația parțială dintre Y și X_1 :

$$r_{y/x_1, x_2} = \frac{r_{y/x_1} - r_{y/x_2} \cdot r_{x_1/x_2}}{\sqrt{1 - r_{y/x_2}^2} \cdot \sqrt{1 - r_{x_1/x_2}^2}}$$

- ✓ pentru corelația parțială dintre Y și X_2 :

$$r_{y/x_2, x_1} = \frac{r_{y/x_2} - r_{y/x_1} \cdot r_{x_1/x_2}}{\sqrt{1 - r_{y/x_1}^2} \cdot \sqrt{1 - r_{x_1/x_2}^2}}$$

8.Corelația rangurilor

Coeficientul de corelație a rangurilor se utilizează pentru analiza legăturilor statistice între variabile calitative, pentru analiza legăturilor statistice dintre două variabile cantitative pentru care nu se cunosc date suficiente pentru a se stabili forma legăturii, sau pentru analiza corelației dintre o variabilă cantitativă și una calitativă. Valorile variabilelor sunt înlocuite cu numerele de ordine (rangurile) acestora, ordonate crescător sau descrescător. Prin rang se înțelege locul pe care îl ocupă valorile x_i sau y_i în cadrul șirurilor din care acestea fac parte, ordonate crescător sau descrescător.

Pentru calculul coeficientului de corelație a rangurilor se folosesc formulele lui Spearman și Kendall.

8.1.Coefficientul de corelație a rangurilor Spearman

Se calculează cu ajutorul formulei:

$$C_s = 1 - \frac{6 \sum d_i^2}{n^3 - n},$$

unde:

d_i -este diferența de rang (număr de ordine) între variabilele corelate pentru aceeași unitate statistică.

n -numărul perechilor corelative.

$c_s \in [-1,1]$, c_s are aceeași semnificație cu cea a coeficientului lui Pearson.

8.2. Coeficientul de corelație a rangurilor Kendall

Se calculează cu ajutorul formulei:

$$C_K = \frac{2S}{n(n-1)}, \text{ unde } S = \sum_i (P_i - Q_i), \text{ iar :}$$

P_i -numărul rangurilor superioare rangului curent pentru variabila dependentă;

Q_i -numărul rangurilor inferioare rangului curent pentru variabila dependentă.

$$C_K \in [-1,1]$$

Coeficientul de corelație a rangurilor Kendall are aceeași semnificație cu cea a coeficientului de corelație simplu lineară Pearson. Dacă pentru aceleași date se calculează ambii coeficienți vom avea totdeauna:

$$C_K < C_S$$

În analiza corelației rangurilor se poate utiliza corelograma rangurilor. Modul de construcție al acesteia este:

- ✓ pe axa Ox se vor trece valorile rangurilor variabilei factoriale;
- ✓ pe axa Oy se trec trece valorile rangurilor variabilei rezultative;

- ✓ din dreptul diviziunilor de pe axa Ox se ridică perpendiculare, iar din dreptul diviziunilor de pe axa Oy se duc paralele la Ox, obținându-se o rețea dublă uniformă constituită din n^2 pătrățele ;
- ✓ diagonalele rețelei sintetizează corelația, directă sau inversă-maximă.
- ✓ corelograma empirică se obține prin unirea punctelor din rețea corespunzătoare $x_i y_i$ din cele două șiruri de ranguri.

Corelograma rangurilor evidențiază sensul și nivelul estimativ al intensității corelației.

9.Aplicații

Aplicația 1. Se dau următoarele date:

Grupa	101	102	103	104	105	106	107	108	109	110
Achiz.										
Echip.	2	3	5	7	8	9	11	13	16	20
Sportiv										
Niv.										
Total										
cheltuieli	15	13	12	11	5	6	7	9	14	15
(mld. lei)										

Carcaterizați dependența dintre cele două variabile.

Rezolvare.

Din reprezentarea grafică a datelor se observă că legătura dintre cele două variabile este cel mai bine aproximată cu ajutorul funcției parabolice:

$$Y_{x_i} = a + bx_i + cx_i^2$$

Datele necesare calculului sunt prezentate în tabelul următor:

	x_i	y_i	x_i^2	x_i^3	$x_i y_i$	$x_i^2 y_i$	x_i^4
	2	15	4	8	30	60	16
	3	13	9	27	39	117	81
	5	12	25	125	60	300	625
	7	11	49	343	77	539	2401
	8	5	64	512	40	320	4096
	9	6	81	729	54	486	6561
	11	7	121	1331	77	847	14641
	13	9	169	2197	117	1521	28561
	16	14	256	4096	224	3584	65536
	20	15	400	8000	300	6000	160000
T	94	107	1178	17368	1018	13774	282518

Determinarea parametrilor a, b, c se realizează cu ajutorul metodei celor mai mici pătrate. Vom obține sistemul:

$$\begin{cases} \sum_i y_i = na + b \sum_i x_i + c \sum_i x_i^2 \\ \sum_i x_i y_i = a \sum_i x_i + b \sum_i x_i^2 + c \sum_i x_i^3 \\ \sum_i x_i^2 y_i = a \sum_i x_i^2 + b \sum_i x_i^3 + c \sum_i x_i^4 \end{cases}$$

$$\Leftrightarrow \begin{cases} 107 = 10a + 94b + 1178c \\ 1018 = 94a + 1178b + 17358c \\ 13774 = 1178a + 17358b + 282518c \end{cases}$$

Soluția sistemului este:

$$a=17,83, b=-1.90, c=0,09$$

Deci vom avea :

$$Y_{x_i} = a + bx_i + cx_i^2 = 17,83 - 1,90x_i + 0,09x_i^2$$

Pentru caracterizarea tipului de legătură dintre cele două variabile, se folosește raportul de corelație :

$$R_{y/x} = \sqrt{1 - \frac{\sum (y_i - Y_{x_i})^2}{\sum (y_i - \bar{y})^2}}$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{107}{10} = 10,7 \text{ miliarde lei, este nivelul mediu al}$$

cheltuielilor pe întreprindere.

$$R_{y/x} = \sqrt{1 - \frac{\sum (y_i - Y_{x_i})^2}{\sum (y_i - \bar{y})^2}} = 0,77 \text{ ceea ce indică o legătură între}$$

cele două variabile.

Aplicația 2. Se dau următoarele date:

Valorile factorialei X	20	30	40	50	60
Valorile rezultativei Y	90	80	100	120	110

Determinați și interpretați coeficientul de corelație a rangurilor Spearman și coeficientul rangurilor Kendall.

Rezolvare.

În tabelul următor sunt sistematizate datele necesare calculului.

X	Y	R _x	R _y	d _i = R _x - R _y	d _i ²	P	Q	S=P-Q
20	90	1	2	-1	1	3	1	2
30	80	2	1	1	1	3	0	3
40	100	3	3	0	0	2	0	2
50	120	4	5	-1	1	0	1	-1
60	110	5	4	1	1	0	0	0
*	*	*	*	*	$\sum d_i^2 = 4$	*	*	6

Coeficientul Sperman.

$$C_s = 1 - \frac{6 \sum d_i^2}{n^3 - n} = 1 - \frac{6 \cdot 4}{5(25 - 1)} = 0,8 > 0, \text{ deci corelația este}$$

directă și de intensitate mare.

Coeficientul lui Kendall.

$$C_K = \frac{2S}{n(n-1)} = \frac{12}{20} = 0,6 > 0, \text{ cu aceeași interpretare ca a}$$

coeficientului Spearman.

Aplicația 3. Se dau datele:

Coeficientul de corelație lineară (R): 0,957;

Dispersia pentru variabila factorială: 224,1009;

Abaterea standard pentru variabila rezultativă: 1598,12;

Determinați, interpretați și alegeți rezultatul corect care vizează coeficientul de corelație lineară (b):

- a. -45,5; b. 102,233; c. 0,4; d. -67

Rezolvare.

$$b = \frac{r\sigma_x}{\sigma_y} = \frac{0,957 \cdot 1598,12}{14,97} = 102,164 > 0, \text{ deci corelația este}$$

directă, variabila rezultativă crescând cu 102,164 la o creștere a variabilei X egală cu unitatea.

Capitolul VII

Testarea concordanței între repartiția experimentală și repartiția teoretică

Modelul de abordare al acestui capitol constă în studiul unor distribuții experimentale în raport cu distribuțiile teoretice în privința caracterului întâmplător sau nonîntâmplător al acestora.

1. Ipoteza nulă

Este un raționament prin care deducem dacă un indicator statistic este semnificativ.

Ipoteza nulă impune realizarea a 2 etape:

1.) se presupune că indicatorul obținut poate apărea datorită unor variații întâmplătoare;

2.) se verifică în tabelele statistice dacă valoarea indicatorului statistic poate apărea din *întâmplare* cu o probabilitate mai mare de 5% .

3) se interpretează datele astfel:

- ✓ dacă probabilitatea de apariție din *întâmplare* este mai mare de 5% acceptăm ipoteza nulă și respingem ipoteza specifică, care postulează că indicatorul respectiv poate apărea din *întâmplare* în mai mult de 5% din cazuri;
- ✓ dacă cifra obținută este mai mare decât valoarea corespunzătoare din tabele la pragul de 0,05 se respinge ipoteza nulă și se acceptă ipoteza specifică, cu precizarea șanselor de a greși (a pragului de semnificație 0,05).

Este de reținut faptul că nu putem afirma pe baza datelor de care dispunem că rezultatul obținut este semnificativ, deoarece mărimea eșantionului poate deseori transforma o valoare semnificativă în una nesemnificativă. Acceptarea ipotezei specifice (respingerea ipotezei nule) neînsemnând certitudine, ci probabilitatea suficient de ridicată care permite generalizarea cu un anumit risc, mic, asumat de noi.

2. Distribuția Hi-pătrat(χ^2)

Considerăm exemplul:

Să se aprecieze activitatea a trei centre de juniori A, B și C, cunoscând că în urma unui sezon centrul A a “dat” 16 sportivi de categoria I și 33 de categoria a II^a, centrul B a

“dat” 20 sportivi de categoria I și 15 de categoria a II^a, iar centrul C a “dat” 9 sportivi de categoria I și 12 de categoria a II^a.

Datele din problemă se trec într-un tabel astfel:

Categoria	Centrul A	Centrul B	Centrul C	Total
Categoria I	¹ 16(24)	² 20(15)	³ 9(9)	45
Categoria II	⁴ 33(32)	⁵ 15(20)	⁶ 12(12)	60
Total	56	35	21	105

Formula, după care stabilim caracterul întâmplător sau nonîntâmplător al acestor diferențe este:

$$\chi^2 = \sum \frac{(f_e - f_t)^2}{f_t} \quad (1)$$

unde:

f_e = frecvența reală;

f_t =frecvența teoretică, care se calculează după stabilirea gradelor de libertate (notate cu f), ce arată numărul frecvențelor teoretice care trebuie calculate:

$$f=(r-1)(c-1)$$

r =numărul rândurilor

c =numărul coloanelor

În cazul nostru $f=(2-1)(3-1)=2$

Deci, vom calcula frecvența teoretică a 2 celule, celelalte valori determinându-le în funcție de acestea.

În cazul nostru calculăm f_t pentru celula 1 și 2 cu formula

$$f_t = \frac{\text{total rând} \times \text{total coloană}}{\text{total general}}$$

Pentru celula 1 avem :

$$f_{t_1} = \frac{56 \times 45}{105} = 24$$

Pentru celula 2 avem :

$$f_{t_2} = \frac{35 \times 45}{105} = 15$$

Se trec aceste frecvențe în tabel, între paranteze, urmând să calculăm celelalte frecvențe astfel:

Pentru celula 3:

$ft_1 + ft_2 + ft_3 = 45$ (reprezintă totalul sportivilor de categoria I) $\Leftrightarrow$
 $24 + 15 + ft_3 = 45$, astfel $ft_3 = 45 - 24 - 15 = 9$

Pentru celula 4:

$ft_1 + ft_4 = 56$ (reprezintă totalul sportivilor din centrul A) $\Leftrightarrow$
 $24 + ft_4 = 56$, astfel $ft_4 = 56 - 24 = 32$

Pentru celula 5:

$ft_2 + ft_5 = 35$ (reprezintă totalul sportivilor din centrul B) $\Leftrightarrow$
 $15 + ft_5 = 35$, astfel $ft_5 = 35 - 15 = 20$

Pentru celula 6:

$ft_3 + ft_6 = 21$ (reprezintă totalul sportivilor din centrul C) $\Leftrightarrow$
 $9 + ft_6 = 21$, astfel $ft_6 = 21 - 9 = 11$

Rzultatele adunării frecvențelor empirice și frecvențelor teoretice coincid, lucru normal.

Observăm că avem calculate toate datele din formula (1).

Astfel:

$$\chi^2 = \sum \frac{(f_e - f_t)^2}{f_t} = \frac{(16 - 24)^2}{24} + \frac{(20 - 15)^2}{15} + \frac{(9 - 9)^2}{9} + \frac{(33 - 32)^2}{32} + \frac{(15 - 20)^2}{20} + \frac{(12 - 12)^2}{12} = \frac{64}{24} + \frac{25}{15} + \frac{1}{32} + \frac{25}{20} \sim 4,7$$

Interpretarea se realizează folosind ipoteza nulă. Presupunem diferențele întâmplătoare. Pentru a decide asupra acceptării sau respingerii ipotezei nule consultăm valorile variabilei $\chi^2 = 4,7$ în funcție de probabilitatea $\alpha = P$ și numărul gradelor de libertate $f=2$ în tabela lui Fischer (vezi tabele statistice). Constatăm că la pragul minimal de semnificație 0,10 avem $\chi^2 = 4,60$ iar la pragul 0,05, $\chi^2 = 5,99$.

Deducem că rezultatele nu pot fi întâmplătoare decât cu o probabilitate mai mică de 10% dar mai mare de 5%. Se acceptă astfel ipoteza nulă.

BIBLIOGRAFIE

1. **Andrei T., Stancu S.,** *Statistică-teorie și aplicații*, Editura All, București, 1995.
2. **Baron T., Anghelache C., Țițan E.,** *Statistică*, Editura Economică, București, 1996.
3. **Bădiță M., Baron T., Korka M.,** *Statistica pentru afaceri*, Editura efficient, București, 1998.
4. **Bădiță M., Baron T., Korka M., Pecican, E.,** *Statistică teoretică și economică*, Editura Universității, București, 1989.
5. **Bărbat Al.,** *Teoria statisticii sociale*, Editura Didactică și Pedagogică, București, 1972.
6. **Biji E., Baron T.,** (coordonatori) *Statistică teoretică și economică*, Editura Didactică și Pedagogică, București, 1996.
7. **Biji M., Biji E.,** *Curs de teoria statisticii*, A.S.E., București, 1971.
8. **Biji E., Lilea E., Wagner P.,** *Statistică*, Editura Universitatea Titu Maiorescu, București, 1995.
9. **Biji E., Lilea E., Vătui M.,** *Statistică-Studii de caz, probleme propuse*, Univ. Creștină Dimitrie Cantemir, Fac. de Management turistic și comercial, Editura Oscar Print, București, 1997.
10. **Căpanu I., Wagner P., Secăreanu C.,** *Statistică macroeconomică*, Editura Economică, București, 1997.

- 11.Craiu V., Bâscă O., *Teste de omogenitate*, Editura Economică București, 1998.
- 12.Diday E., Lebart L., Pages J.P., *Data Analysis and informatics*, North-Holland, 1980.
- 13.Georgescu P., Georgescu V., Radu C., *Statistica-Modele și cazuri*, Repografia Universității din Craiova, 1994.
- 14.Ionescu C., și colectiv, *Dicționar statistico-economic*, București, 1969.
- 15.Isaic-Maniu Al., Mitruț C., Voineagu, V., *Statistica pentru managementul afacerilor*, Editura Economică, ediția a II^a, revizuită și îmbunătățită, București, 1999.
- 16.Ivănescu I., și colectiv, *Statistică*, Editura Didactică și Pedagogică, București, 1999.
- 17.Neagoescu Gh., Ciobanu R., Bontaș C., *Bazele statisticii pentru afaceri*, Editura All Beck, București, 1999.
- 18.Neagoescu Gh., Ciobanu C., *Probleme de statistică*, Editura Algorithm, Galați, 1998.
- 19.Ștefan T., *Elemente de statistică aplicată*, Ministerul tineretului și sportului, București, 1993
- 20.Tovissi L., Andrei T., Spircu L., *Analiza seriilor de timp și procese dinamice*, Editura All, București, 1995.
- 21.Vasilescu N., *Statistica-sinteze teoretice și lucrări practice*, Litografia Universității Craiova, 1987.

22.Vasilescu N., *Statistică generală-Metodologie*, Editura Sitech, Craiova, 1994.

23.Voineagu V., Mitruț C., Isaic-Maniu Al., Tițan E., Baron T., Matache S., Isaic-Maniu I., Șerban D., Voineagu M., *Statistică teoretică și macroeconomică-Teste, lucrări practice, studii de caz*, Editura Economică, București, 1998.

Cuprins

1.Capitolul I Statistica – instrument de cunoaștere a fenomenelor sociale.....	1
2.Capitolul II Elemente de teoria probabilităților	11
3.Capitolul III Serii statistice și reprezentări grafice	20
4.Capitolul IV Indicatorii tendinței centrale	40
5.Capitolul V Variația și asimetria	65
6.Capitolul VI Regresia și corelația statistică	94
7.Capitolul VII Testarea concordanței între repartiția experimentală și repartiția teoretică	125
8.Bibliografie	130