DUMITRACHE IULIA

COORDONATOR
PROF. UNIV. DR. FLORIN CRISTIAN GHEORGHE

METODE DE ESTIMARE A PARAMETRILOR

Editura Sfântul Ierarh Nicolae
2010

ISBN 978-606-8129-28-0

INTRODUCERE

Preocupări pe linia evaluării unor probabilități au existat şi înainte de anul 1654. Astfel, pe la mijlocul secolului al XIV-lea, Gardano a scris « Liber de Ludo Aleae » (Cartea despre jocul cu zaruri) care însă a fost publicată abia în 1663.

De asemenea, la începutul secolului al XVII-lea, Galilei se arăta interesat de studiul erorilor de măsurare. Tot în acea perioada apar şi primele preocupări de teoria asigurărilor. Dar toate acestea nu au condus la un studiu sistematic al probabilităților. Intr-un fel era firesc ca acest studiu sa pornească de la analizarea jocurilor de noroc, întrucât acestea oferă modele din cele mai simple şi posibilități nelimitate de repetare a experiențelor.

Odată stârnit interesul de lucrările lui Pascal, Fermat, Huygens, teoria probabilităților cunoaşte o dezvoltare vertiginoasă.

Au urmat apoi lucrările lui J. Bernoulli, Poisson, Borel, Cantelli, Kolmogorov, Moivre, Gauss, Laplace.

Următoarea perioadă de dezvoltare este dominată de lucrările lui Cebîşev (1821 – 1894), Leapunov (1857 – 1918), Markov (1856 – 1922), care au adus contribuții importante în aşa-numita teorema centrala a teoriei probabilităților şi au inaugurat studiul variabilelor aleatoare dependente.

În această perioadă, sfera aplicațiilor teoriei probabilităților s-a mărit, cuprinzând şi ştiințele naturii, în special fizica.

Calculul probabilităților şi statistica matematică sunt ramuri deosebit de importante şi actuale ale matematicii. Contactul lor permanent cu alte ştiințe, chiar şi cu cele considerate până nu demult « nematematizabile », a dus la o îmbogățire şi evoluție impresionantă a acestor discipline.

Modelele probabilistice şi metodele statistice de confruntare a acestor modele cu realitatea obiectiva oferă procedee de investigație ştiințifică în sectoare de activitate dintre cele mai diferite.

Aceasta explică importanța deosebită care se acorda teoriei probabilităților şi statisticii matematice în țara noastră. Introducerea în cadrul învățământului de cultură generală a unor elemente de teoria probabilităților şi statistică matematică a corespuns unor cerințe stringente ale unui învățământ modern şi eficient.

În această lucrare, prin varietatea domeniilor din care sunt culese date, se caută să se sugereze o idee asupra ariei deosebit de vaste a aplicațiilor posibile ale statisticii. Noțiunile de teoria probabilităților constituie un subiect puțin abordat în învățământul gimnazial.

În primul capitol sunt prezentate noțiuni introductive de statistică matematică şi sondaj. Sondajul statistic bucurându-se de privilegiul de a fi adeseori singura metoda de cercetare practic posibilă.

În cel de-al II-lea capitol , « Teoria probabilităților », sunt prezentate noțiunile: evenimente; câmp de evenimente; probabilitate, proprietăți; probabilitate condiționată , evenimente independente; formule de calcul pentru probabilități; variabile aleatoare; media; dispersia; repartiții clasice.

În capitolul al III-lea, numit "Metode de estimații", sunt prezentate metoda verosimilității maxime şi metoda celor mai mici pătrate.

În ultimul capitol sunt propuse câteva aplicații practice folosind ca metodă de estimație metoda celor mai mici pătrate.

CAPITOLUL I

Noțiuni introductive de statistică matematică. Sondaj.

1. Populație statistică. Caracteristica

Statistica matematică se ocupa cu gruparea, analiza şi interpretarea datelor referitoare la un anumit fenomen, precum şi cu unele previziuni privind producerea lui viitoare.

În cadrul analizei statistice a unui fenomen acționează mai întâi statistica descriptiva care se ocupa cu culegerea datelor asupra fenomenului respectiv şi cu înregistrarea acestor date apoi intervine statistica matematică, care grupează datele, le analizează şi le interpretează in vederea unor predicții privind comportarea viitoare a fenomenului.

Vom numi *populație statistică* orice mulțime care formează obiectul unei analize statistice.

Elementele unei populații statistice se numesc *unități statistice* sau indivizi.

Trăsătura comună a tuturor unităților unei populații care ne interesează în cadrul unei analize statistice se numeşte *caracteristica*.

Analiza statistică se poate face după una sau mai multe caracteristici.

EXEMPLE:

a) Dacă ne interesează rezultatele obținute, la teza la matematica, de elevii din clasa aVIII-a, a unei scoli, atunci:

- mulțimea tuturor elevilor din clasa aVIII-a din acea şcoală, formează populația statistică;
- fiecare elev din clasa aVIII-a a acestei şcoli, este o unitate statistică;
- nota la teză la matematică este caracteristica studiată.

b) Dacă ne interesează numărul locuitorilor din fiecare oraş al țării
la o anumită dată, atunci:

- mulțimea tuturor oraşelor tarii la data respectivă formează populația statistică;
- fiecare oraş constituie o unitate statistică;
- numărul de locuitori la data respectiva este caracteristica studiata.

c) Daca ne interesează diametrul unor piese de acelaşi fel fabricate într-o întreprindere dată, atunci:

- mulțimea pieselor fabricate în întreprindere este populația statistică;
- o piesă constituie o unitate statistică;
- diametrul piesei este caracteristica studiată.

d) Dacă ne interesează distribuția unui grup de copii după culoarea ochilor şi culoarea părului, atunci:

- mulțimea copiilor grupului considerat formează populația statistica;
-fiecare copil în parte din grupul respectiv este o unitate statistică;

-culoarea ochilor şi culoarea părului sunt caracteristicile care ne interesează.

Se pot da nenumărate exemple de mulțimi care pot constitui obiectul unei analize statistice: distribuția unui grup de persoane după talie, vârstă şi distribuția oraşelor după numărul de sateliți, etc.

Din înseşi exemplele date rezultă existența a doua feluri de caracteristici.

O caracteristică se numeşte *cantitativă* dacă se poate măsura. În caz contrar, caracteristica se numeşte *calitativă*. (răspuns prin DA sau NU, BUN sau PROST, etc.)

Diametrul piesei, vârstă, talia, etc. sunt exemple de caracteristici.

1.1 Gruparea datelor

Sa presupunem ca s-a măsurat înălțimea unui grup de 120 de persoane. Rezultatele obținute (înălțimea în centimetrii) sunt date în tabelul 1, în ordinea în care au apărut.

Tabelul 1

176	173	161	171	174	168	178	166	169	172
181	172	163	174	173	169	172	175	158	182
186	190	173	173	169	171	176	172	188	175
162	170	176	177	171	164	162	175	176	176
170	176	178	164	174	177	180	175	175	180
174	171	175	170	179	186	177	178	169	180
188	173	172	174	183	177	176	174	181	159
183	174	179	167	165	182	176	178	171	169
168	179	177	177	181	178	184	177	173	177
162	177	173	170	176	179	170	168	174	175
173	178	185	185	171	165	167	174	175	172
179	168	171	175	165	178	172	175	166	171

Este clar că sub această formă, tabelul nu ne permite sa tragem prea multe concluzii cu caracter mai general. De aceea, este necesar sa facem o grupare a acestor date. O primă posibilitate de grupare este cea din tabelul 2, după înălțime

.

Tabelul 2

cm	Nr. Pers.	cm	Nr. Pers.	cm	Nr. Pers.	cm	Nr. Pers.	cm	Nr. pers
158	1	165	3	171	8	177	9	183	2
159	1	166	2	172	7	178	7	184	1
161	1	167	2	173	8	179	5	185	2
162	3	168	4	174	9	180	3	186	2
163	1	169	5	175	10	181	3	188	2
164	2	170	5	176	9	182	2	190	1

Tabelul are doua coloane, dar a fost prezentat aşa deoarece are multe linii. În prima coloană este înălțimea în centimetrii, iar în a doua numărul de persoane care au aceeaşi înălțime. Sub aceasta forma tabelul ne ajuta sa tragem unele concluzii: înălțimea căreia ii corespunde cel mai mare număr de persoane este 175 cm, înălțimilor apropiate de 175 cm le corespund un număr mai mare de persoane, decât celor mai depărtate, etc.

În tabelul 3 sunt prezentate rezultatele obținute de elevii unei clase la teza la matematică.

Tabelul 3

nota	Număr elevi	nota	Număr elevi
2	1	7	15
3	2	8	6
4	2	9	3
5	4	10	1
6	7		

Din acest tabel putem trage concluzii referitoare la nivelul la care s-a prezentat clasa respectivă la teza la matematică.

Din aceste exemple rezultă că analiza statistică a unui fenomen, în raport cu o singură caracteristică, ne conduce la o serie de perechi de valori, pe care o vom numi serie statistică. În exemplele de mai sus este vorba de perechi de numere, primul număr al unei perechi reprezentând valoarea caracteristicii (înălțimea în cm, nota la teza), iar cel de-al doilea număr reprezentând numărul de unități statistice corespunzătoare acelei valori a caracteristicii (numărul de persoane, numărul de elevi).

În tabelul 4 este prezentata împărțirea unui grup de 1500 de persoane după culoarea ochilor.

Tabelul 4

Culoarea ochilor	Nr. persoane
Negri	240
Căprui	752
Verzi	302
albaştri	206
	1500

După cum se vede în, acest caz, caracteristica ia patru valori: negru, căprui, verde, albastru, care nu sunt valori numerice. Daca împărțim un grup de persoane după prenume, atunci caracteristica ia ca valoare un prenume, iar în tabel, în dreptul fiecărui prenume trecem numărul de persoane care îl poarta.

În acest capitol ne vom ocupa de serii cu caracteristici cantitative.

Atunci când analiza statistica a unei populații se face după doua caracteristici, rezultatele obținute se trec intr-un tabel cu dubla intrare.

Astfel, în tabelul 5 este redata distribuția unui număr de 1500 de copii după culoarea ochilor şi culoarea parului.

Tabelul 5

Culoarea ochilor / Culoarea parului	negri	căprui	verzi	albaştrii	Total
Negru	145	285	30	471	471
Castaniu	62	431	87	647	647
Blond	33	36	185	382	382
total	240	752	302	206	1500

Din acest tabel reiese ca au fost găsiți 87 copii cu ochi verzi şi par castaniu, 33 de copii cu ochi negri şi par blond, 11 copii cu ochi albaştri şi par negru, etc.

În cazul seriilor statistice cu o singura caracteristica, pentru obținerea cu uşurința a unor concluzii generale asupra fenomenului studiat, tabelele cu doua coloane, aşa cum au fost ele prezentate mai sus, sunt suficiente, daca numărul valorilor pe care le ia caracteristica este în jur de 20.

Când acest număr este depăşit, citirea tabelului devine greoaie, fiind prea voluminoasa. Este cazul tabelului 1. în aceasta situație se impune o noua grupare a datelor. Revenind la tabelul 1, împărțim mulțimea valorilor caracteristice în clase, după cum reiese din tabelul 6.

Tabelul 6

Clase de valori în centimetri	Nr. persoane
< 160	2
160 - 165	7
165 - 165	16
170 - 175	37
175 - 180	40
180 - 185	11
185 - 190	7
	120

La aceste tabele facem convenția ca extremitatea dreapta a fiecărei clase (cu excepția, eventual a ultimei clase) sa nu aparțină clasei. Astfel, clasa 165 - 170 cuprinde valorile x ale caracteristicii, pentru care $165 \leq x < 170$.

În cazul pe care l-am prezentat, intervalele reprezentând clasele de valori au aceeaşi lungime. De la aceasta regula putem excepta, eventual clasele extreme (prima şi ultima). Lungimea acestor intervale nu este impusa de o regula fixa, ea fiind la îndemâna statisticianului, care caută ca împărțirea în intervale sa fie cat mai judicioasa. Acest mod de realizare a tabelelor (cu clase de valori de amplitudini egale sau nu) se impune mereu în cazul caracteristicilor comune. În tabelul 7 este data repartiția elevilor după înălțime.

Tabelul 7

Înălțimea în cm.	Nr. elevi
150 – 154	38
154 – 158	65
158 – 162	175
162 – 166	189
166 – 170	111
Peste 170	62
	640

Aici lungimea intervalelor alese (amplitudinea claselor) este de 4 cm.

1.2 Frecventa absoluta. Frecventa relativa. Frecvente cumulate

Numărul tuturor elementelor unei populații statistice se numeşte volumul acelei populații.

Astfel, în tabelul 7, populația este mulțimea elevilor unei scoli şi are un volum total de 640 de elevi.

Se numeşte frecventa absoluta, a unei valori x a caracteristicii, numărul de unități ale populației corespunzătoare acelei valori. De exemplu, în tabelul 2, valoarea 172 cm a caracteristicii are frecventa absoluta egala cu 5, iar în tabelul 3, nota 5 are frecventa absoluta egala cu 4.

Se numeşte frecventa relativa (sau pe scurt, frecventa) a unei valori x a caracteristicii, raportul dintre frecventa absoluta a valorii x şi volumul populației. Vom scrie:

$$f(x) = \frac{n_x}{n},$$

unde f(x) este frecventa relativa a valorii x, n_x este frecventa absoluta a acestei valori, iar n volumul populației.

În tabelul 2, frecventa relativa a valorii 179 este $\frac{5}{120} = \frac{1}{24}$, a valorii 175, $\frac{10}{120} = \frac{1}{12}$ etc. , iar în tabelul 3, frecventa valorii 5 este $\frac{4}{40}$ = 10% etc.

În tabelele statistice cu doua coloane putem înlocui în coloana a doua frecventele absolute prin frecvente relative. Astfel, tabelul 3, care prezintă rezultatele la teza de la matematica dintr-o clasa în care sunt 40 de elevi, poate fi descrisa în tabelul 8.

Tabelul 8

Nota	Frecventa
2	0,025
3	0,025
4	0,050
5	0,100
6	0,175
7	0,375
8	0,150
9	0,075
10	0,025

Deci, în cazul caracteristicilor cantitative, aceste tabele scot în evidenta o corespondenta intre doua mulțimi de numere: mulțimea valorilor caracteristicii şi mulțimea frecventelor corespunzătoare. Se poate remarca analogia cu distribuția variabilelor aleatoare pe care o scriem sub forma:

$$\begin{pmatrix} x_1 x_2 \ldots x_n \\ p_1 p_2 \ldots p_n \end{pmatrix}$$

În prima linie sunt trecute valorile variabilei, iar în cea de-a doua linie probabilitățile corespunzătoare acestor valori. În mod similar, tabelul corespunzător unei caracteristici cantitative se poate scrie sub forma prezentata în tabelul 9.

Tabelul 9

Valori	Frecventa
x_1 x_2 . . . x_n	f_1 f_2. . . . fn

În prima coloana sunt trecute valorile caracteristicii, iar în cea de-a doua, frecventele corespunzătoare.

De altfel, noțiunile de variabila aleatoare şi probabilitate sunt modele teoretice

ale noțiunilor de caracteristica şi respectiv frecventa care sunt noțiuni cu caracter experimental(statistic).

De multe ori, în loc de caracteristica se spune variabila statistica, sau pe scurt variabila şi vom folosi în cele ce urmează şi aceasta denumire.

Vom spune ca un tabel de genul tabelului 9 defineşte distribuția sau repartiția statistica a variabilei respective.

Este uşor de observat ca suma frecventelor respective ale tuturor valorilor variabilei este 1.

Se numeşte frecventa absoluta cumulata crescătoare a unei valori x a variabilei suma frecventelor absolute ale tuturor valorilor variabilei care apar pana la x inclusiv.

*Se numeşte frecventa absoluta cumulata descrescătoare a unei valori x a variabilei suma frecventelor absolute ale tuturor valorilor care apar de la x inclusiv (*în cazul caracteristicilor cantitative vom considera da acum numai tabele în care valorile variabilei sunt scrise în ordine crescătoare).

Astfel, frecventa absoluta cumulata crescătoare a valorii 6 din tabelul 3 este 15, iar frecventa absoluta cumulata descrescătoare a aceleiaşi valori este 32; pentru valoarea 9 a aceleiaşi variabile, cele doua frecvente sunt respectiv 4 şi 39.

În mod analog, definim frecventa relativa(sau scurt frecventa)cumulata crescătoare şi frecventa cumulata descrescătoare.

Se numeşte frecventa cumulata crescătoare a unei valori x a variabilei, suma tuturor frecventelor valorilor care apar pana la x inclusiv, iar frecventa cumulata descrescătoare suma tuturor frecventelor valorilor care apar de la x inclusiv.

Rezulta ca frecventa cumulata crescătoare este exprimata şi de raportul dintre frecventa absoluta cumulata crescătoare şi volumul total al populației, iar frecventa cumulata descrescătoare de raportul dintre frecventa absoluta cumulata descrescătoare şi volumul total al populației.

În cele ce urmează, *prin frecventa cumulata vom înțelege frecventa relativa cumulata crescătoare*. Un tabel statistic poate fi completat cu aceste frecvente. Tabelul 3 completat devine tabelul 10.

Tabelul 10

nota	Frecventa absoluta	Frecventa relativa	Frecventa absoluta cumulata descrescă-oare	Frecventa absoluta cumulata crescătoare	Frecventa cumulata descres-cătoare	Frecventa cumulata crescătoare
2	1	0,025	40	1	1	0,025
3	1	0,025	39	2	0,975	0,050
4	2	0,050	38	4	0,950	0,100
5	4	0,100	36	8	0,900	0,200
6	7	0,175	32	15	0,800	0,375
7	15	0,375	25	30	0,625	0,750
8	6	0,150	10	10	0,250	0,900
9	3	0,025	4	39	0,100	0,975
10	1	0,025	1	40	0,025	1

0bservatie: frecventa absoluta cumulata crescătoare (descrescătoare) a valorii x ne da numărul de unități corespunzătoare valorilor strict mai mici (mai mari) sau egale cu x. astfel, numărul din linia a 5-a a tabelului10 ne spune ca exista 15 elevi cu note mai mici sau egale cu 6 (deci cu note de 2, 3, 4, 5, 6). Pe acelaşi tabel citim ca exista 30 de elevi cu note mai mici ca 8, sau 10 elevi cu note mai mari decât 7, etc. putem interpreta, de asemenea, datele din coloanele 6 şi 7. astfel putem spune ca 0,1 = 10% din elevi au note mai mici de 5 (deci note de 2, 3, 4), ca 0,1 = 10% au note mai mari decât 8 (deci 9 şi 10), ca 0,2 = 20% au note mai mici ca 6, etc.

Noțiunile introduse în raport cu valorile individuale ale variabilei pot fi extinse şi în cazul tabelelor cu clase de valori. Astfel, frecventa absoluta a unei clase este numărul de unități corespunzătoare valorilor variabilei care aparțin clasei respective, iar frecventa relativa (frecventa) unei clase este raportul dintre frecventa sa absoluta şi efectivul total al populației. Astfel din tabelul 6, re3zulta ca frecvența absoluta a clasei 160 – 165 este 7, iar frecventa sa relativa este $\frac{7}{120}$.

Se numeşte frecventa absoluta cumulata crescătoare a unei clase suma frecventelor absolute ale tuturor claselor care apar pana la clasa considerata inclusiv.

Se numeşte frecventa absoluta cumulata descrescătoare a unei clase suma frecventelor absolute a tuturor claselor care apar de la clasa considerata inclusiv.

Completând cu aceste frecvente tabelul 6 obținem tabelul 11.

Tabelul 11

Clasa de valori	Frecventa absoluta	Frecventa absoluta cumulata crescătoare	Frecventa absoluta cumulata descrescătoare
<160	2	2	120
160-165	7	9	118
165-170	16	25	111
170-175	37	62	95
175- 180	40	102	58
180- 185	11	113	18
185-190	7	120	7

Cum se face citirea unui astfel de tabel ?

Frecventa absoluta cumulata crescătoare a unei clase ne da numărul de unități corespunzătoare valorilor mai mici (strict) decât limita superioara a intervalului, iar frecventa absoluta cumulata descrescătoare ne da numărul unităților corespunzătoare valorilor mai mari sau egale cu limita inferioara a intervalului. Astfel din tabelul 11 rezulta ca din persoanele măsurate 62 au înălțimi sub 175 cm; 95 persoane au cel puțin 170 cm înălțime, etc.

Frecventa relativa cumulata crescătoare (frecventa cumulata) a unei clase este suma frecventelor claselor care apar pana la clasa considerata, inclusiv sau, ceea ce este acelaşi lucru, raportul dintre frecventa absoluta cumulata crescătoare şi efectivul total al populației. În mod analog se defineşte frecventa relativa cumulata crescătoare a unei clase.

Din tabelul 11 rezulta ca frecventa cumulata a clasei 175- 180 este $\frac{102}{120}$=0,85=85%. Vom spune ca 85% din persoanele măsurate au înălțimea sub 180cm. frecventa cumulata descrescătoare a clasei 165-170 este 0,925, deci 92,5% din persoanele considerate au înălțimea de cel puțin 165cm.

1.3 Serii cronologice

Tot în cadrul seriilor statistice sunt incluse şi aşa-numitele *serii cronologice* care prezintă evoluția în timp a unor mărimi.

În cazul unui tabel corespunzător unei serii cronologice, în prima coloana sunt trecute anumite momente sau intervale de timp, iar în coloana a doua valorile corespunzătoare ale mărimii considerate. Astfel sa presupunem ca se măsoară temperatura apei unui lac intr-un anumit punct, în fiecare an la 15 iulie ora 12, iar rezultatele obținute sunt trecute în tabelul 12.

Tabelul12

Data	Temperatura în ^{0}C
15 iulie 2000	20
15 iulie 2001	22
15 iulie 2002	19
15 iulie 2003	18
15 iulie 2004	20
15 iulie 2005	20
15 iulie 2006	21
15 iulie 2007	20

În coloana stânga avem trecute momentele precise de timp, spre deosebire de tabelul 13, unde avem trecute în coloana stângă intervalele de timp.

Tabelul 13

NUMARUL ABSOLVENTILOR UNEI SCOLI

1990-1994	1000
1995-1999	997
1999 -2004	1002
2004-2008	1018

(În tabelele 12 şi 13 datele sunt imaginare.)

2. Reprezentarea grafica a seriilor statistice

În acest paragraf ne vom ocupa de reprezentarea grafica a seriilor statistice cu o singura caracteristica. Reprezentarea grafica a unei serii este uneori foarte sugestiva, ea contribuind la o prima interpretare intuitiva, pe cale vizuala a datelor. Deseori reprezentarea grafica a datelor sugerează însăşi legea pe care o urmează fenomenul studiat.

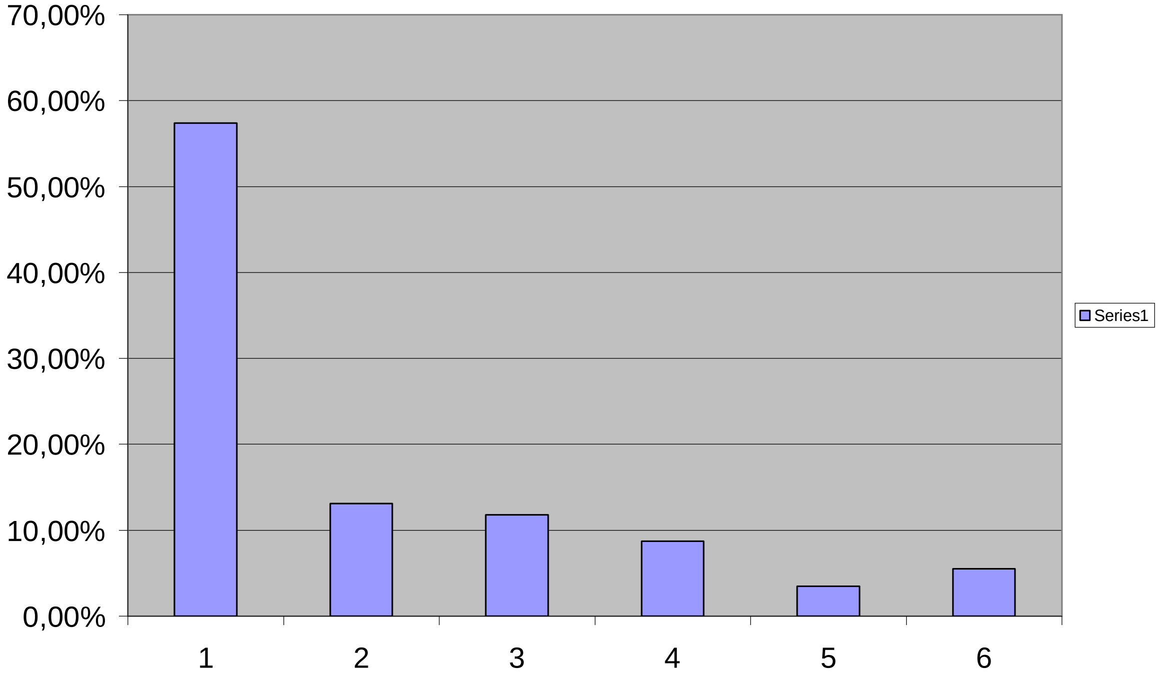
Reprezentarea grafica a seriilor cu caracteristica calitativa

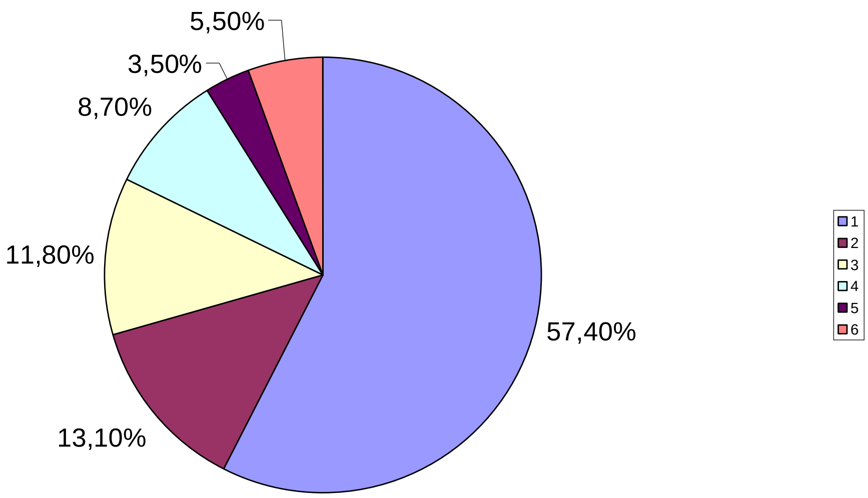
Reprezentarea acestor distribuții constituie un capitol deosebit de important al reprezentării grafice, dat fiind ca ilustrează, prin desene, anumite rapoarte numerice. Graficul corespunzător poarta numele de diagrama.

Sa consideram de exemplu, distribuția investițiilor unei societăți pe diferite ramuri, în anul 2002-2006.

Tabelul 14

Total investiții	100%
I1	57,4%
I2	13,1%
I3	11,8%
I4	8,7%
I5	3,5%
I6	5,5%





Datele pot fi reprezentate ca mai sus.

Reprezentarea seriilor cu caracteristica cantitativa

Seriile cu caracteristica cantitativa se reprezintă grafic în raport cu un sistem de axe rectangulare. Alegerea unității pe fiecare dintre axe este la îndemâna statisticianului, care are grija ca alegerea sa uşureze obținerea concluziilor dorite, cat şi ca desenul sa sugereze histograma fenomenului.

a) *Reprezentarea în batoane*. Aceasta reprezentare se foloseşte mai ales pentru seriile în care variabila ia un număr mic de valori.

Sa consideram datele din tabelul 15.

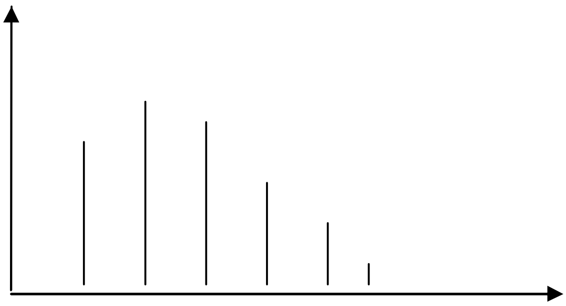
Tabelul 15
DISTRIBUTIA FAMILIILOR DINTR-UN BLOC
DUPA NUMARUL DE COPII

Nr. membri	Frecventa absoluta
0	6
1	18
2	23
3	20
4	14
5	6
6	2
7	1

Obținem distribuția în batoane din figura 2.

Deci pe axa orizontala sunt trecute punctele reprezentând valorile reprezentând valorile variabilei şi din aceste puncte se ridica segmente verticale de lungime egala cu frecventa absoluta sau relativa a valorii respective.

Fig. 2



Segmentele ridicate sunt măsurate cu unitate pe Oy.

b) Histograma. Fiind data o serie cu clase de valori cu amplitudini egale obținem histograma acestei serii, luând pe axa orizontala o succesiune de segmente egale reprezentând amplitudinea claselor, şi ridicând, pe fiecare din aceste segmente considerate ca baze, dreptunghiuri de înălțimi proporționale cu frecventele (relative sau absolute) ale claselor respective.

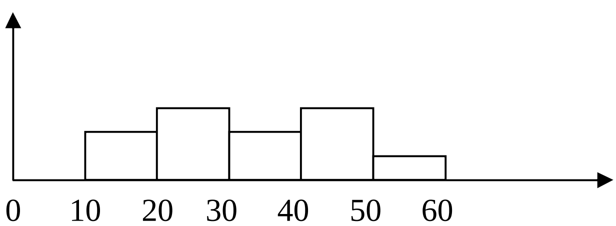


Fig.3

Histograma corespunzătoare din tabelul 16 este data în figura 3.

DISTRIBUTIA UNOR PIESE
DUPA DIAMETRUL LOR
Tabelul 16

Mărimea diametrului mm	Frecventa absoluta	Frecventa cumulata
10 – 20	10	10
20 – 30	15	25
30 – 40	12	37
40 – 50	15	52
50 - 60	18	60

Histograma corespunzătoare tabelului 11 este data în figura 4.

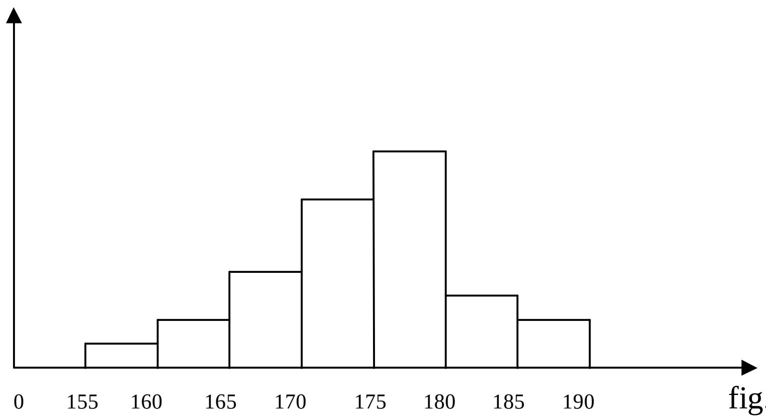


fig. 4

c) Daca din mijlocul fiecărui segment de pe axa orizontala ridicam segmente proporționale cu frecventele claselor corespunzătoare fiecărui segment şi unim printr-o linie poligonala extremitățile superioare ale acestor segmente obținem poligonul frecventelor. Astfel, poligonul frecventelor corespunzătoare tabelului 16 este dat în figura 5.

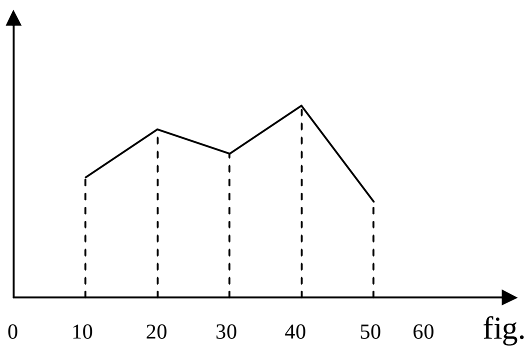


fig. 5

d) Daca aceleaşi puncte de la aliniatul precedent le unim nu printr-o linie poligonala, ci printr-o curba, obținem curba de distribuție empirica a seriei respective.

Poligonul frecventelor cumulate (crescătoare) se obține unind printr-o linie poligonala punctele (x, y), unde x este extremitatea dreapta a intervalului unei clase, iar y frecventa cumulata a clasei respective, la care mai adăugăm punctul (a, 0), unde a este limita inferioara a primei clase.

Astfel, poligonul frecventelor cumulate corespunzător tabelului 11 este dat în figura 6

În mod analog se defineşte curba frecventelor cumulate descrescătoare. Daca punctele din figura le unim nu printr-o linie poligonala, ci printr-o curba, obținem curba cumulativa a seriei considerate.

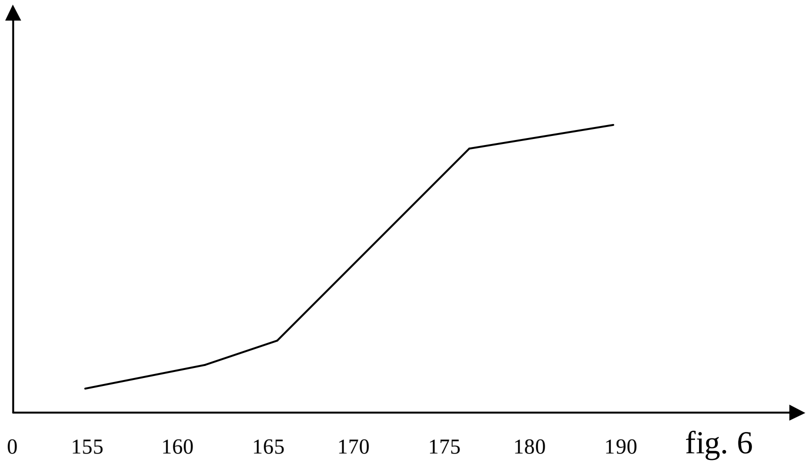


fig. 6

fig. 7

Repartizarea seriilor cronologice

Sa presupunem ca intr-un cămin şcolar se dau note fiecărui dormitor pentru întreținerea curățeniei. Rezultatele obținute de un anumit dormitor intr-o săptămâna sunt date în tabelul 17.

Tabelul 17

L	10
M	9
M	9
J	10
V	8
S	10
D	9

Diagrama corespunzătoare acestui tabel este dat în figura 7.

În acest mod se pot face graficele realizărilor zilnice (pe o perioada de timp) – individuale sau colective – intr-o întreprindere. Abscisele punctelor care se unesc prin linia poligonala sunt mijloacele segmentelor reprezentând intervale de timp (sau punctele reprezentând momentele), iar ordonatele, valorile mărimii considerate, în intervalul de timp corespunzător.

3. Elemente caracteristice ale unei serii statistice

În cele ce urmează *vom numi valoarea centrala a unei clase de variație, media aritmetica a extremităților acestei clase.* Astfel, valoarea centrala a clasei 165 - 170 din tabelul 11 este 167,5.

3.1 Modul

Modulul sau dominanta unei serii se numeşte valoarea caracteristicii corespunzătoare celei mai mari frecvente în cazul în care valorile caracteristicii sunt date individual şi valoarea centrala a clasei corespunzătoare celei mai mari frecvente, în cazul variabilelor continue, când se dau clase de variație. Aceasta noțiune prezintă interes mai ales în cazurile când avem o singura dominanta.

În cazul prezentat în tabelul 3, dominanta este 7, iar în cazul tabelului 11 este 177,5.

3.2 Mediana

Mediana unei serii este un număr x astfel ca exista tot atâtea unități statistice corespunzătoare valorilor > x.

Daca o caracteristica ia valorile:

1, 3, 3, 3, 4, 5, 6, 6, 7, 8, 9,

Atunci 5 este mediana, deoarece exista 5 valori <5 şi 5 valori >5.

Daca avem valorile:

1, 3, 3, 3, 4, 5, 6, 7, 7, 9,

Atunci vom lua ca mediana media aritmetica a numerelor situate la mijloc (daca ele au fost luate în ordinea mărimii).în acest caz, mediana este 4,5. uneori se considera ca mediana oricare din cele doua numere.

Cum se calculează mediana în cazul unei variabile continue, vom arata pe un exemplu. Sa consideram pentru aceasta tabelul 16.

Daca piesele ar fi aranjate în ordinea diametrelor lor, noi vrem sa calculam diametrul celei de-a 30-a. diametrul acestei piese este cuprins intre 30 şi 40mm. clasa 30 – 40 are frecventa absoluta 12. vom presupune ca cele 12 piese corespunzătoare creste uniform de la 30 la 40. deci creşterea diametrului de la o piesa la următoarea este $\frac{40-30}{12}$. Pe de alta parte, a-30-a piesa a populației este a 30 – 25 = a 5-a piesa a clasei (deoarece exista 25 de piese cu diametrul <30). Deci, diametral celei de-a 30-a piesa este

$$30 + (30 - 25) \times \frac{40-30}{12} = 34{,}16\text{mm}.$$

3.3 Media aritmetica

Daca $x_1, x_2, \dots, x_n$ sunt n valori, se ştie ca media lor aritmetica este

$$\frac{x_1 + x_2 + \dots + x_n}{n}.$$

Fiind data distribuția unei variabile x :

valoarea	Frecventa
x_1 x_2 . . . x_n	y_1 y_2 . . . y_n

Valoarea medie a variabilei respective este:

$$\bar{x} = \frac{x_1 y_1 + x_2 y_2 + ... + x_n y_n}{y_1 + y_2 + ... + y_n}. \qquad (1)$$

Daca N = $y_1 + y_2 + \ldots + y_n$ este efectivul total al populației, atunci :

$$\bar{x} = x_1 \frac{y_1}{N} + x_2 \frac{y_2}{N} + ... + x_n \frac{y_n}{N}.$$

Sau, daca notam cu $f_i = \frac{y_i}{N}$ frecventa relativa a valorii x_i (i = 1, 2, …, n), atunci:

$\bar{x} = x_1 f_1 + x_2 f_2 + \ldots + x_n f_n$.

Expresia (1) are într-adevăr semnificația unei medii aritmetice. Variabila x ia după cum reiese din tabel de y_1 ori valoarea x_1, de y_2 ori valoarea x_2, şi aşa mai departe. Deci pentru a calcula valoarea medie a variabilei, calculam media aritmetica a numerelor

$$\underbrace{x_1, x_1, \ldots x_1,}_{y_1 \text{ ori}} \quad \underbrace{x_2, x_2, \ldots x_2,}_{y_2 \text{ ori}} \quad \ldots, \quad \underbrace{x_n, x_n, \ldots x_n}_{y_n \text{ ori}}$$

şi obținem chiar expresia din membrul drept al relației (1). Aceasta expresie se mai numeşte media aritmetica ponderata a numerelor $x_1, x_2, \ldots, x_n$, numerele $y_1, y_2, \ldots, y_n$ fiind ponderile respective ale acestor valori.

Daca vom considera tabelul 3 rezulta ca media pe întreaga clasa la notele la matematica este:

$$\frac{10X1 + 9X3 + 8X6 + 7X15 + 6X7 + 5X4 + 4X2 + 3X1 + 2X1}{40} = 6,625.$$

Cazul seriilor cu variație continua îl reducem la cazul precedent substituind fiecare clasa cu valoarea sa centrala.

Astfel în exemplul precedent în tabelul 16 se obțin datele din tabelul 18.

Tabelul 18

Mărimea diametrului	Frecventa absoluta y_i	Valoarea centrala xi	$x_i y_i$
10 – 20	10	15	150
20 – 30	15	25	375
30 – 40	12	35	420
40 – 50	15	45	675
50 - 60	8	55	440
	60		2060

$$\bar{x} = \frac{2060}{60} = 34,3$$

Daca vrem sa lucram cu numere mai mici decât cele ce ne sunt date în tabele facem următoarele observații. Avem pentru orice I :

$x_i = x_0 + (x_i – x_0)$,
$x_i y_i = x_0 y_i + (x_i – x_0) y_i$.

Dând lui i valorile 1, 2, 3, …, n obținem n relații care adunate termen cu termen ne dau :

$x_1y_1 + x_2y_2 + \ldots + x_ny_n = x_0 (y_1 + y_2 + \ldots + y_n) + (x_1 - x_0)y_1 + (x_2 - x_0)y_2 + \ldots + (x_n - x_0)y_n$,
sau, împărțind cu $N = y_1 + y_2 + \ldots + y_n$,

$$\bar{x} = x_0 + \frac{(x_1 - x_0)y_1 + (x_2 - x_0)y_2 + \ldots + (x_n - x_0)y_n}{N}$$

această relație o mai putem scrie:

$$\bar{x} = x_0 + \overline{x - x_0}.$$

În exemplul precedent, luând $x_0 = 35$, avem calculele prezentate în tabelul 19.

Tabelul 19

Clase	Frecventa absoluta y_i	Valoarea centrala x_i	$x_i – x_0$	$(x_i – x_0)y_i$
10 – 20	10	15	-20	-200
20 - 30	15	25	-10	-150
30 – 40	12	35	0	0
40 – 50	15	45	10	150
50 - 60	8	55	20	160
	60			-40

$$\bar{x} = 35 - \frac{40}{60} = 34,3$$

3.4 Dispersia

Definiție. Fiind date n valori x_1, x_2, …, x_n având media $\overline{x}$, se numeşte dispersie de selecție, mărimea:

$$\sigma^2 = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + ... + (x_n - \overline{x})^2}{n-1}.$$

fiind data seria statistica:

Valori x_i	Frecventa absoluta y_i
x_1 x_2 . . . x_n	y_1 y_2 . . . y_n

Unde $N = y_1 + y_2 + … + y_n$ şi $\overline{x} = \dfrac{x_1y_1 + x_2y_2 + ... + x_ny_n}{N}$

sunt respectiv efectivul total al populației şi valoarea medie, dispersia corespunzătoare este :

$$\sigma^2 = \frac{(x_1 - x)_2 y_1 + (x_2 - x)_2 y_2 + ... + (x_n - x_n)_2 y_n}{N-1}.$$

Caracteristica

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

se numeşte *abaterea* mediei pătratice. Ea se exprima în aceleaşi unități ca şi caracteristica studiata. Prin urmare, valoarea abaterii medii pătratice D(X), şi deci implicit valoarea dispersiei $D^2(X)$, dau o măsura pentru împrăştiere (dispersia) valorilor variabilei X în jurul valorii medii. Când D(X) este mic, atunci valorile lui X sunt strânse în jurul valorii medii M(X), iar daca D(X) este mare, valorile variabilei X sunt împrăştiate. Aceasta observație justifica denumirea de dispersie pentru mărimea $D^2(X)$.

În cazul caracteristicilor continue se substitue fiecare interval de variație prin valoarea sa centrala.

Sa dam şi o alta forma dispersiei. Vom dezvolta expresia:

$$\frac{1}{N-1}[(x_1 - \overline{x})^2 y_1 + (x_2 - \overline{x})^2 y_2 + …(x_n - \overline{x})^2 y_n]$$

şi obținem:

$$\sigma^2 = \frac{1}{N-1}[x_1^2y_1 + x_2^2y_2 + \ldots + x_n^2y_n - 2\,\bar{x}\,(x_1y_1 + x_2y_2 + \ldots +$$

$$+x_ny_n) + \bar{x}^2(y_1 + y_2 + \ldots + y_n)] - \frac{x_1^2y_1 + x_2^2y_2 + \ldots + x_n^2y_n}{N-1} - 2\,\bar{x}^2 + \bar{x}^2 =$$

$$= \frac{x_1^2y_1 + x_2^2y_2 + \ldots + x_n^2y_n}{N-1} - x^2. \qquad (2)$$

Deci

$\sigma^2 = x^2 - (\bar{x})^2$.

Daca am fi înlocuit mărimile x_i prin $x_i - x_0$ unde x_0 este o constanta, am fi obținut:

$$\sigma^2 = \frac{(x_1 - x_0)^2y_1 + (x_2 - x_0)^2y_2 + \ldots + (x_n - x_0)^2y_n}{N} - (x - x)^2 \quad (2')$$

unde prin $\bar{x}^2$ am notat media aritmetica a mărimilor x_1^2, x_2^2, …,x_n^2 cu ponderile y_1, y_2, …,y_n. am regăsit o proprietate a dispersiei unei variabile aleatoare.

Sa completam tabelul 19 cu datele necesare calculului dispersiei σ^2. Vom calcula σ^2 atât direct cat şi folosind formula (2^I).

Luam $x_0 = 35$ şi ştim ca $\bar{x} = 34,3$. se obțin datele din tabelul 20.

Tabelul 20

clase	Frecv. y_i	Val. Centr. x_i	$x_i - \bar{x}_0$	$x_i - \bar{x}$	$(xi - x_0)^2$	$(x_i - \bar{x}_0)^2$	$y_i(x_i - x_0)^2$	$y_i(x_i - x_0)^2$
10-20	10	15	- 20	- 19,3	400	372,49	4000	3724,90
20-30	15	25	- 10	-9,3	100	86,49	1500	1296,45
30-40	12	35	0	0,7	0	0,49	0	5,88
40-50	15	45	10	10,7	100	114,49	1500	1717,35
50-60	8	55	20	20,7	400	428,49	3200	3427,92
	60						10200	10172,50

Deci :

$$\sigma^2 = \frac{10172,5}{60} = 169,5$$

sau folosind (2')

$$\sigma^2 = \frac{1}{60} * 10200 - (34,3 - 35) = 169,5.$$

Se observa ca alegând convenabil valoarea lui x_0, calculele se simplifica.

4. SONDAJ

4.1 Noțiuni introductive

Statistica matematică este una din ramurile moderne ale matematicii care se ocupă cu gruparea, analizarea şi interpretarea datelor referitoare la anumite fenomene, precum şi cu previziuni în ceea ce priveşte producerea lor viitoare.

Pornind de la cunoaşterea modului de repartizare a frecventelor, statistica matematică îşi propune prin metode inductive să obțină informații referitoare la legile de repartiție (probabilitate) ale fenomenului care a produs frecventele.

În cadrul analizei statistice a unui fenomen acționează mai întâi statistica *formala* sau *descriptiva*, care se ocupa cu colectarea datelor asupra fenomenului respectiv. Intervine apoi statistica *matematica* cu ajutorul căreia datele sunt analizate şi interpretate.

Definiție. Numim *populație statistica* sau mai simplu populație, orice mulțime care formează obiectul unei analize statistice.

Rezolvarea numeroaselor probleme practice confrunta pe statisticieni cu necesitatea de a studia populații cu un mare număr de elemente constitutive fără a avea posibilitatea reala de a examina fiecare element în parte. În asemenea cazuri s-a considerat ca cel mai bun lucru pe care îl poate face cel interesat, este acela de a examina un număr limitat de elemente, sperând ca informația primită va fi utilă şi suficient de precisă pentru a caracteriza (cunoaşte) întreaga populație.

Acest număr finit de elemente (de fapt o subpopulație) constituie ceea ce se cunoaşte sub denumirea de eşantion care arata legătura cu noțiunea de populație totala indicând mulțimea finita sau infinita luata în studiu.

Un postulat unanim admis, verificat în numeroase aplicații, este acela ca un eşantion ne va da informații utile despre populația originala din care a fost extras şi ca pe măsura ce volumul eşantionului creste, proprietățile şi structura populației originale vor fi mai fidel reprezentate. Direcțiile de cercetare tot mai ample în acest sens au conturat şi apoi au definit riguros obiectul teoriei sondajului, capitol fundamental al statisticii matematice.

H. Poincare a spus: “ Slăbiciunea noastră nu ne permite sa îmbrățişăm tot universul şi suntem obligați sa-l descompunem în bucăți.”

Evenimentele şi fenomenele din natura şi societate sunt foarte numeroase, prea complexe şi prea variate pentru a permite o observare totala. Suntem deci obligați, în cercetarea ştiințifică, sa recurgem la fracționări, din care sa scoatem adevăruri generale.

4.2 Caracteristici ale sondajului

Vom prezenta câteva noțiuni elementare de statistică. Vom începe cu două exemple care evidențiază noțiunile prezentate.

Exemplul 1

Dorim sa cunoaştem repartiția elevilor dintr-o clasa, după notele obținute la matematica:

- mulțimea elevilor dintr-o clasa formează populația statistică;
- fiecare elev reprezintă o unitate statistică;
- nota obținută la matematică este caracteristica studiată.

Exemplul 2

Consideram repartiția unei colectivități de recruți după înălțimea şi greutatea corporala:

- mulțimea recruților reprezintă populația statistica;
- fiecare recrut reprezintă o unitate statistica;
- exista doua caracteristici: înălțimea şi greutatea.

În exemplele prezentate observăm două tipuri de caracteristici: cantitative (măsurabile) şi calitative (bun sau prost).

Caracteristicile cantitative se prezintă atunci când pot fi măsurate în numere reale (exemplul 2).

Caracteristicile calitative se prezintă atunci când pot fi calificate prin atributul bun sau prost (exemplul 1).

După efectuarea sondajului, de obicei repartizarea observațiilor se face grupând datele în intervale.

Definiție. Totalitatea valorilor care aparţin unui interval dat se numeşte *clasa*.

Pentru fiecare interval se definesc:

- *frecventa absoluta* n_i *(* $i = 1, 2, \ldots k$ *)*, care reprezintă numărul de observații în intervalul *i*.
- *frecventa absoluta cumulata,* care reprezintă numărul tuturor observațiilor cărora le corespund valori mai mici sau egale cu limita superioara a intervalului *i*.
- *frecventa relativa cumulata* $F_n(x_1)$*,* adică suma frecventelor relative corespunzătoare intervalului I şi intervalelor precedente.

Suma frecventelor absolute ne dă numărul total al eşantionului.

Observație. Intervalele nu sunt de lungimi egale. Pe statisticieni ii interesează frecventa corespunzătoare unui interval.

4.3 Populații şi sondaje

În statistica matematică vom întâlni trei tipuri de populație:

- populații finite;
- populații ipotetice;
- populații infinite.

Populațiile finite sunt totodată, şi populații reale în sensul ca sunt construite dintr-un număr finit, efectiv de elemente, cu o anumita determinare calitativa, intr-un loc şi intr-un timp date.

Populațiile ipotetice sunt o extindere logica a conceptului de populație reala. Formal o populație ipotetica o putem defini ca fiind mulțimea tuturor modurilor posibile ce pot fi concepute şi prin care se poate produce un eveniment specificat. De fapt, o populație ipotetică poate fi imaginată în legătură cu orice eveniment specificat. Putem sa ne imaginam ca o populație ipotetica este o populație infinita de populații finite analoage, adică având aceeaşi determinare calitativa cu populația reala considerata.

Populații infinite. În acest caz populațiile considerate au un număr infinit de elemente. În unele cazuri numărul elementelor este atât de mare încât îl putem considera practic infinit.

A. Avantajele şi obiectivele sondajului.

Sondajul statistic se bucura de privilegiul de a fi adeseori singura metoda de cercetare practic posibila. Ne referim, desigur, la împrejurări în care înregistrarea totala nu se poate utiliza fără prejudiciul ireparabil al obiectelor studiate. În anumite situații, când se efectuează controlul statistic al calității producției, de exemplu măsurarea duratei medii de ardere a becurilor, produsul este pierdut ireparabil.

Principalul avantaj al metodei sondajului statistic al calității consta în faptul ca acesta este aplicabil în toate cazurile, când se preconizează obținerea unor importante economii de timp şi, totodată, al unui volum de informații strict necesar şi suficient, adeseori cu un grad de exactitate mai mare decât cel ce se poate realiza cu prilejul înregistrărilor totale. Se ştie ca în cazul acesta din urma sunt posibile numeroase erori, adeseori inevitabile.

Obiectivul fundamental al cercetării, pe baza de sondaj, a populațiilor statistice este acela de a obține cu un efort minim un volum maxim de informație asupra populației originale, luata în calcul.

Valori tipice ca *media, abaterea medie pătratică, momente, mediana,* şi altele, se numesc *parametrii* populațiilor originale, spre deosebire de variabilele corespunzătoare de sondaj care poarta denumirea de *statistici.*

O parte importanta a teoriei sondajului este consacrată găsirii procedeelor adecvate de estimare a valorilor tipice ale populației originale pe baza informației oferite de eşantion.

Un alt obiectiv al cercetării pe baza de sondaj este acela de a determina precizia unei estimații, stabilindu-se în acest scop anumite limite pentru abordarea posibila a estimației bazata pe eşantion, fata de valoarea adevărată a parametrului estimat al populației originale.

B. Tipuri de sondaje.

Un eşantion se formează dintr-un număr determinat de unități elementare observate. După modul în care pot fi prelevate în eşantion aceste unități se cunosc următoarele tipuri de sondaje:

- sondajul pur aleator;
- sondajul dirijat;
- sondajul mixt.

Sondajul pur aleator. În cazul acestui tip de sondaj, unitățile elementare ale populației originale sunt prelevate în eşantion la întâmplare. Formal, alegerea la întâmplare înseamnă ca la fiecare element component al populației originale are aceeaşi probabilitate de a fi ales în eşantion. S-a formulat astfel un principiu fundamental al sondajului aleator.

Vom numi *eşantion aleator,* când acesta a fost ales astfel încât, în procesului sondajului, toate eşantioanele posibile, formate, sa zicem, din n elemente, au aceeaşi probabilitate de a se realiza. (Exista şi sondaje aleatoare cu probabilități variabile.)

În ceea ce priveşte *reprezentativitatea sondajului*, prima problema care apare în procesul sondajului este aceea a modului în care putem obține un eşantion cu adevărat aleator pentru a asigura astfel reprezentativitatea sondajului. Prin aceasta noțiune înțelegem o anumita identitate (mai mare sau mai mica) a structurii eşantionului cu structura populației originale.

Obținem o reprezentativitate perfecta a unui eşantion, când identitatea dintre cele doua structuri este deplina, ceea ce în practica este imposibil de realizat. De obicei se considera satisfăcător acel sondaj care asigura abateri ale valorilor tipice de structurale populației de sondaj cu cel mult 5% fata de parametrii corespunzători ai populației originale.

Există două tipuri de erori de sondaj:

- erori sistematice;
- erori aleatoare (întâmplătoare).

Prin erori reprezentative se înțeleg diferențele dintre caracteristicile studiate în eşantion şi cele din populația originala. Aceste diferențe se produc în toate cazurile, eliminarea lor completa nefiind posibila. Este posibilă însă determinarea cu exactitate a mărimii lor probabile.

Erorile de reprezentativitate sunt, în general, rezultatul încălcării principiului fundamental al sondajului aleator, acela ca fiecărei unități din populația generala sa-I fie asigurata aceeaşi probabilitate de a fi aleasa din eşantion. Menționam faptul ca ființa umana are, ca parte componenta a structurii sale psihologice, tendința de a se îndepărta de caracterul cu adevărat întâmplător în alegerile sale. Chiar şi observatori experimentați din diferite ramuri ale cercetării ştiințifice, efectuând sondaje, adeseori nu pot evita apariția de erori sistematice destul de mari.

Menționam ca anumite tehnici de sondaj valabile în cercetarea populațiilor mici se dovedesc cu totul ineficiente în cazul populațiilor mari. Astfel, putem spune, în măsura în care ne referim la aceste caracteristici, ca nu exista nicio rațiune ca un element sa fie ales mai curând decât altul şi, in consecință, se va respecta principiul conform căruia fiecare element al populației va avea aceeaşi şansă de a fi ales în eşantion. În acest sens au fost concepute procedee, cum ar fi acela al alegerii mecanice şi, mai recent cel al tabelelor întâmplătoare.

Sondajul dirijat. Un asemenea sondaj se efectuează după un anumit principiu prestabilit, atunci când acesta apare ca fiind rațional şi util. lăsând loc opiniilor personale şi inclinațiilor proprii la extragerea eşantionului, ceea ce, după cum se ştie, prezintă pericolul unor erori sistematice. Încrederea pe care o putem acorda acestui tip de sondaj depinde de circumstanțele cazului dat.

Uneori sondajul dirijat formează eşantioane tipice sau reprezentative intr-o măsură mai mare decât un sondaj aleator. Nu este exclus ca un sondaj aleator sa se abată mult față de medie, în timp ce un eşantion dirijat, prin construcția sa, să fie foarte apropiat de medie.

În cazul sondajului de volum mare, datorită erorilor sistematice, rezultatele sondajului dirijat pot fi din ce în ce mai puțin reprezentative, în timp ce rezultatele sondajului aleator sa capete o reprezentativitate tot mai bună.

Sondajul mixt. Cel mai reprezentativ exemplu de sondaj mixt este sondajul stratificat (tipic). El combină principiul sondajului dirijat cu cel al sondajului aleator. Esența acestui tip de sondaj consta în divizarea întregii populații în structuri (grupe tipice în raport cu un principiu oarecare, prestabilit) şi apoi, extragerea cate unui eşantion aleator din fiecare strat. Şi în cazul sondajelor mixte apar erori de reprezentativitate, de regulă, ele sunt însă mai mici.

Sondajul mixt ne sugerează ideea ca formarea unui eşantion reprezentativ se poate face în faze.

Capitolul II
TEORIA PROBABILITATILOR

1. Evenimente. Câmp de evenimente.

Se pot da extrem de multe exemple (practic din orice domeniu) de experiențe aleatoare şi în cadrul fiecăruia din acestea se pot da exemple de evenimente pe care se poate pune problema evaluării şanselor de a se produce. Trecerile de la o stare la alta (elev, student) se produc cu anumite probabilități.

Un exemplu este planificarea nevoilor de învățământ în funcție de dezvoltarea economica. Un astfel de studiu se poate raporta la o populație finită, în permanență mişcare, urmând legi stocastice. Este deci, naturală introducerea noțiunii de probabilitate când se examinează un astfel de model.

Vom accepta deci punctul de vedere conform căruia teoria probabilităților este o parte a studiului naturii.

Menționam ca un eveniment legat de o anumita experiență poate fi *sigur, imposibil* sau *aleator,* în raport cu aceasta experiență. Prin experiență se înțelege realizarea unui complex de condiții S.

Evenimentul sigur este acel eveniment care se produce de fiecare data când sunt realizate condițiile S. de exemplu la aruncarea unui zar sa apară una din fetele numerotate cu: {1}, {2}, {3}, {4}, {5}, {6}.

Evenimentul imposibil este acela care în prezenta condițiilor S nu se poate realiza (produce) niciodată când condițiile S sunt realizate. De exemplu, la aruncarea unui zar sa nu apară nici o față.

Eveniment aleator este acela care în prezenta condițiilor S se poate produce sau nu. În cazul aruncării unui zar ne punem problema daca apare o fata cu un număr par sau impar de puncte.

În exemplul de mai sus este clar ca cele doua evenimente nu se pot realiza împreună.

Definiție. Doua sau mai multe evenimente sunt *incompatibile* daca nu se pot realiza împreună. În caz contrar se numesc *compatibile.*

Exemple:

1. Sa considerăm experiența care consta în aruncarea unui zar şi fie evenimentele:

A: apariția unui număr par de puncte;
B: apariția unui număr impar de puncte.

Evenimentele A şi B sunt *incompatibile.*

2. În cadrul aceleiaşi experiențe considerăm evenimentele:

C: apariția unui număr mai mic decât 4;
D. apariția unui număr intre 2 şi 5.

În cazul în care zarul cade cu fata 2 sau 3 atunci se realizează ambele evenimente.

Fie Ω mulțimea tuturor rezultatelor posibile ale unei experiențe. Mulțimea Ω este numita *mulțimea evenimentelor elementare,* iar elementele sale vor fi numite *puncte,* notate ω. În cele ce urmează vom spune ca $\{\omega\}$ reprezintă un *eveniment elementar*.

Exemplu: În cazul aruncării unui zar, mulțimea Ω este formata din:
Ω = {{1}, {2}, {3}, {4}, {5}, {6}}

în timp ce {1} reprezintă un eveniment elementar.

Notam cu P(Ω) mulțimea tuturor evenimentelor aleatoare. În aceasta mulțime vom reliefa doua evenimente importante.

Definiție. *Evenimentul contrar al lui A* sau *non A* este evenimentul care se realizează daca şi numai daca nu se realizează A.

Exemplu: În cazul aruncării unui zar, avem:

A = {1,2,4}; $\overline{A}$ = {3,5,6}.

Observație. Evenimentele de mai sus sunt incompatibile.

Pe mulțimea evenimentelor P(Ω) vom introduce următoarele operații cu evenimente:

Definiție. Daca realizarea evenimentului A va implica realizarea evenimentului B vom spune ca *evenimentul A implica evenimentul B* şi vom nota $A \subset B$.

Exemplu: Fie A = {1,2} şi B = {1,2,3,5} doua evenimente. Evident $A \subset B$.

Definiție. Fie $A, B \in P(\Omega)$. Evenimentul A sau B numit *reuniunea evenimentelor A şi B* este evenimentul care se realizează daca şi numai daca se realizează cel puțin unul din evenimentele A sau B şi se notează $A \cup B$.

Exemplu. Fie A = {1,2} şi B = {4,5} doua evenimente. Rezulta $A \cup B$ = {1,2,4,5}.

Definiție. Fie A,B$\in P(\Omega)$. Evenimentul A şi B numit *intersecția evenimentelor A şi B* este evenimentul care se realizează daca şi numai daca se realizează ambele evenimentele A sau B şi se notează $A \cap B$.

Exemplu: Fie A = {1,2,3} şi B = {2,4} doua evenimente. Rezulta $A \cap B$ ={2}.

Definiție. Doua evenimente A şi B se numesc *incompatibile* daca ele nu se realizează împreună, adică daca $A \cap B = \phi$.

Exemplu: evenimentele A şi $\overline{A}$ sunt incompatibile.

Definiție. Diferența evenimentelor A şi B este evenimentul care se produce atunci şi numai atunci când se produce A dar nu se produce B. se notează

A-B= $A \cap \overline{B}$.

Definiție. Vom spune ca evenimentele A_1, … ,A_n formează un sistem complet de evenimente daca se realizează cu certitudine unul şi numai unul din aceste evenimente. Se poate scrie:

1) $\bigcup_{i=1}^{n} A_i = \Omega$

2) $A_i \cap A_j = \phi, \forall i \neq j,\ i, j = \overline{1,n}$

Observație. Se foloseşte denumirea de *partiție* a evenimentului Ω.

Corp de părți.

Fie K o familie de evenimente a lui Ω.

Definiție. O familie K de părți ale lui Ω, $K \subset P(\Omega)$ se numeşte *corp* daca sunt verificate următoarele axiome:

1. $\forall A \in K \Rightarrow \overline{A} \in K$;

2. $\forall A, B \in K \Rightarrow A \cup B \in K$.

Observație. Mulțimea K va fi numita *câmp finit sau adiție algebrica* de părți ale lui Ω.

Propoziție. Daca K este un corp de părți ale lui Ω atunci:

a) $\Omega \in K, \phi \in K$;

b) $\forall A, B \in K \Rightarrow A \cap B \in K$.

Demonstrație:

a) Daca $A \in K \Rightarrow A \cup \overline{A} \in K$, deci

$A \cup \overline{A} = \Omega \in K \Rightarrow \overline{\Omega} \in K \Rightarrow \phi \in K$.

b) $\left.\begin{array}{l} A \in K \\ B \in K \end{array}\right\} \Rightarrow \left\{\begin{array}{l} \overline{A} \in K \\ \overline{B} \in K \end{array}\right. \Rightarrow \overline{A} \cup \overline{B} \in K$ sau $\overline{A} \cap \overline{B} \in K$ de unde $A \cap B \in K$.

Observație. Putem spune ca un corp de părți ale lui Ω este o familie nevidă de submulțimi ale lui Ω închisă la reuniune finita, intersecție finita şi trecere la complementara.

Observație. Daca $A, B \in K \Rightarrow A - B = A \cap \overline{B} \in K$ şi $A \Delta B = (A \setminus B) \bigcup (B \setminus A) \in K$.

Definiție. O mulțime K de părți ale lui Ω se numeşte *corp borelian* daca sunt verificate:

1. daca $A \in K \Rightarrow \overline{A} \in K$;

2. daca $\{A_n\}_{n \in N} \subset K \Rightarrow \bigcup\limits_{n \in N} A_n \in K$.

Observație. Un corp borelian este un corp. se mai numeşte *σ – algebra* sau *algebra completa.*

Definiție. Perechea {Ω, K} se numeşte câmp de evenimente.

2. Probabilitate. Proprietăți.

Câmp de probabilitate.

Fie {Ω, K} un câmp de evenimente.

Definiție. Se numeşte *probabilitate pe K* o aplicație P : K → **R**, cu proprietățile:

1) $P(A) \geq 0 \quad \forall A \in K$;

2) P(Ω) = 1;

3) $P(A \cup B) = P(A) + P(B) \quad \forall A, B \in K$ cu $A \cap B = \phi$

Definiție. Se numeşte câmp de probabilitate tripletul {Ω, K, P} unde Ω este mulțimea evenimentelor elementare, K un câmp de părți, iar P o probabilitate pe K.

Definiție. Se numeşte *probabilitate σ – aditiva* sau încă incomplet aditiva pe K o aplicație P : K→ **R**, cu proprietățile:

1. $P(A) \geq 0 \quad \forall A \in K$;
2. P(Ω) = 1;
3. $P\left(\bigcup\limits_{i \in I} A_i\right) = \sum\limits_{i \in I} P(A_i) \quad \forall \{A_i\}_{i \in I} \subset K$ cu $A_i \cap A_j = \phi, i \neq j, i, j \in I$ iar I este o mulțime cel mult numărabilă de indici.

Observație. Proprietatea 3. se numeşte axioma aditivității complete.

Proprietate. *Pentru orice* $A \in K$ *avem* $P(\overline{A}) = 1 - P(A)$.

Demonstrație: Evident

$A \cup \overline{A} = \Omega$ şi $a \cap \overline{a} = \phi \Rightarrow P(A \cup \overline{A}) = P(A) + P(\overline{A}) = 1$

Proprietate. P(ϕ) = 0.

Demonstrație: $\phi = \overline{\Omega} \Rightarrow P(\phi) = 1 - P(\Omega)$

Proprietate. *Pentru orice* $A \in K \Rightarrow 0 < P(A) < 1$

Demonstrație: din egalitatea:

$0 \le P(A \cup \overline{A}) = P(A) + \overline{P(\overline{A}) = 1}$

şi din P($\overline{A}) \ge O \Rightarrow P(A) \le 1$.

Proprietate. *Dacă* $A_1 \subset A_2$ *implica* $P(A_1) \le P(A_2)$.

Demonstrație: Putem scrie:

$A_2 = A_1 \cup (A_2 \cap \overline{A_1})$ şi deoarece $A_1 \cap (A_2 \cap \overline{A_1}) = \phi$ rezulta

$P(A_2) = P(A_1 \cup (A_2 \cap \overline{A_1})) = P(A_1) + P(A_2 \cap \overline{A_1})$

Deoarece $P(A_2 \cap \overline{A_1}) \ge 0$ implica $P(A_2) \ge P(A_1)$.

Proprietate. *Pentru orice* $A_1, A_2 \in$ K implica $P(A_2 - A_1) = P(A_2) - P(A_1 \cap A_2)$.

Demonstrație: Deoarece

$A_2 = (A_2 - A_1) \cup (A_2 \cap A_1)$

şi deoarece

$(A_2 - A_1) \cap (A_2 \cap A_1) = \phi \Rightarrow P(A_2 - A_1) = P(A_2) - P(A_1 \cap A_2)$.

Proprietate. *Daca* $A_1 \subset A_2$ *implica* $P(A_2 - A_1) = P(A_2) - P(A_1)$.

Demonstrație: Deoarece $A_2 = (A_2 - A_1) \cup A_1$ şi $(A_2 - A_1) \cap A_1 = \phi$ de unde rezulta proprietatea.

Proprietate. *Pentru orice* $A_1, A_2 \in K$ *avem:*

$P(A_1) + P(A_2) = P(A_1 \cup A_2) + P(A_1 \cap A_2)$

Demonstrație: Deoarece

$A_1 \cup A_2 = A_1 \cup [A_2 - (A_1 \cap A_2)]$ şi $A_1 \cap [A_2 - (A_1 \cap A_2)] = \phi$

rezulta

$P(A_1 \cup A_2) = P(A_1) \cup P[A_2 - (A_1 \cap A_2)] = P(A_1) + P(A_2) - P(A_1 \cap A_2)$

Proprietate. *Fie* $\{A_n\}_{n \in N}$ *un sir de evenimente din K şi P o probabilitate complet aditiva, atunci:*

$$P\left(\bigcup_{n \in N} A_n\right) \le \sum_{n \in N} P(A_n).$$

Demonstrație: introducem un şir de evenimente din K în felul următor:

$A_1' = A_1$; $A_2' = A_2 - A_1$; $A_3' = A_3 - (A_1 \cup A_2)$; … ; $A_{n+1}' = A_{n+1} - \bigcup_{i=1}^{n} A_i$; …

Evenimentele din şirul $\{A_n'\}_{n \in N}$ prin modul de construcție sunt incompatibile două câte două, iar $A_i' \subset A_i, \forall i \in \mathbf{N}$ şi în plus sa arătăm:

$$\bigcup_{n \in N} A_n' = \bigcup_{n \in N} A_n$$

Evident

$$\bigcup_{n \in N} A_n' \subseteq \bigcup_{n \in N} A_n.$$

Demonstrăm incluziunea contrară.

Fie $\omega \in \bigcup_{n\in N} A_n^{'}$ şi fie n cel mai mic număr natural pentru care $\omega \in A_{n_0}$; rezulta ca

$A_{n_0} \subset \bigcup_{n\in N} A_n^{'}$

Deci

$\bigcup_{n\in N} A_n \subseteq \bigcup_{n\in N} A_n^{'}$.

Deoarece

$$P\left(\bigcup_{n\in N} A_n\right) = P\left(\bigcup_{n\in N} A_n^{'}\right) = \sum_{n\in N} P(A_n^{'} \le \sum_{n\in N} P(A_{n)}$$

Folosind proprietăţile anterioare.

Proprietate. Inegalitatea lui Boole. *Daca* $\{A_n\}_{n\in N}$ *este o familie cel mult numărabilă de elemente din K, atunci:*

$$P\left(\bigcap_{n\in N} A_i\right) \ge 1 - \sum_{i\in I} P(\overline{A_i})$$

Demonstraţie: Se cunoaşte că

$$\bigcap_{i\in I} A_i = \overline{\left(\bigcup_{i\in I} \overline{A_i}\right)}$$

deci

$$P\left(\bigcap_{i\in I} A_i\right) = P\overline{\left(\bigcup_{i\in I} \overline{A_i}\right)} = 1 - P\left(\bigcup_{i\in I} \overline{A_i}\right) \ge 1 - \sum_{i\in I} P(\overline{A_i})$$

Deoarece

$$P\left(\bigcup_{i\in I} \overline{A_i}\right) \le \sum_{i\in I} P(\overline{A_i})$$

Observaţie. Daca i = $\overline{1,n}$ atunci inegalitatea lui Boole devine:

$$P\left(\bigcap_{i=1}^{n} A_i\right) \ge 1 - \sum_{i=1}^{n} P(\overline{A_i})$$

Inegalitatea lui Boole dă o margine inferioară pentru probabilitatea unei intersecţii de evenimente.

3. Probabilitate condiţionată, evenimente independente.

Fie (Ω, K, P) un câmp borelian de probabilitate şi fie A $\in$ K cu P(A) $\ne 0$.

Definiţie. Se numeşte *probabilitatea evenimentului B condiţionată de evenimentul A,* notata cu $P_A(B)$, raportul:

$$P_A(B) = \frac{P(A \cap B)}{P(A)} = P(B/A).$$

Teorema. *Tripletul (Ω, K, P_A) este un câmp borelian de probabilitate.*

Demonstrație:

1. $P_A(B) \geq 0$ deoarece $P(A \cap B) \geq 0$ şi $P(A) > 0$.
2. daca $B = \Omega$ atunci $P_A(\Omega) = \dfrac{P(A \cap \Omega)}{P(A)} = 1$
3. daca $\{A_i\}_{i \in I} \subset K$ este o familie cel mult numărabilă de evenimente incompatibile doua cate doua, atunci evenimentele $\{ A \cap A_i \}_{i \in I}$ sunt incompatibile doua cate doua:

$$P_A\left(\bigcup_{i \in I} A_i\right) = \frac{P[A \cap (\bigcup_{i \in I} A_i)]}{P(A)} = \frac{P[\bigcup_{i \in I}(A \cap A_i)]}{P(A)} = \frac{\sum_{i \in I} P(A \cap A_i)}{P(A)} = \sum_{i \in I} P_A(A_i)$$

Observație. P_A ia valoarea I pentru orice eveniment B pentru care $A \subset B$, rezultă că corpul $K_A = \{A \cap B \mid B \in K\}$ este cel mai mic corp borelian pe care P_A este o probabilitate.

Observație. Din definiție mai rezultă:

$P(A \cap B) = P(A)P_A(B)$.

În mod analog, daca $P(B) \neq 0$, avem:

$P(A \cap B) = P(B)P_B(A)$

Care este legea probabilităților reciproc condiționate a doua evenimente de probabilități diferite de zero.

Evenimente independente.

Fie (Ω, K, P_A) un câmp borelian de probabilitate.

Definiție. Fie $A_1, A_2 \in K$ doua evenimente. Ele se numesc independente dacă:

$P(A_1 \cap A_2) = P(A_1)P(A)$

Observație. Dacă $P(A_1 \cap A_2) \neq 0$ şi evenimentele A_1 şi A_2 sunt independente, atunci

$P_{A2}(A_1) = P(A_1)$

i

$P_{A1}(A_2) = P(A_2)$

i reciproc, daca $P_{A2}(A_1) = P(A_1)$ atunci evenimentele A_1si A_2 sunt independente.

Definiție. Evenimentele $A_K \in K$, $k = \overline{1,n}$ sunt independente daca

$P(A_{k1} \cap \ldots \cap A_{ks}) = P(A_{k1}) \ldots P(A_{ks})$

Pentru orice indici $1 \leq k_1 \leq \ldots \leq k_s \leq n$.

Teorema. Daca evenimentele A, B $\in$ K, k = $\overline{1,n}$ sunt independente atunci sunt independente şi evenimentele A şi $\overline{B}$, $\overline{A}$ şi B, $\overline{A}$ şi $\overline{B}$

.

Demonstrație. Deoarece $A\cap\overline{B} = A\setminus(A\cap B)$, atunci

$P(A\cap\overline{B}) = P[A\setminus(A\cap B)] = P(A) - P(A\cap B) = P(A)-P(A)P(B) = P(A)[1 - P(B)] = \quad = P(A)P(\overline{B})$

Din faptul ca A şi $\overline{B}$ sunt independente pe baza raționamentelor de mai sus.

Observație. Daca evenimentele sunt independente k cate k nu rezultă că ele sunt independente k+1 cate k+1. În acest caz prezentam exemplul lui Bernstein.

Exemplu. Sa consideram un tetraedru omogen cu fetele colorate în felul următor: una în alb, una în roşu, una în negru şi a patra cu toate cele trei culori. Aruncând tetraedrul pe o masa, el se aşează pe una din fete. Ne interesează culoarea acestei fete. Sa notam cu A_1 culoarea alb, A_2 culoarea roşu, A_3 culoarea negru. Rezultă $P(A_1) = P(A_2) = P(A_3) = \frac{1}{2}$ deoarece pentru fiecare culoare sunt patru cazuri posibile şi doua favorabile (față cu culoarea respectivă şi față cu cele trei culori).

Se constata uşor ca:

$$P(A_1\cap A_2) = P(A_1\cap A_3) = P(A_2\cap A_3) = \frac{1}{4}$$

Dar

$$P(A_1\cap A_2\cap A_3) = \frac{1}{4} \neq P(A_1)P(A_2)P(A_3) = \frac{1}{8}$$

de unde rezultă că evenimentele sunt independente două câte două dar nu sunt independente trei câte trei.

4. Formule pentru calculul unor probabilități.

a) Probabilitatea unei reuniuni de intervale.

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i,j=1}^{n} P(A_i\cap A_j) + \ldots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)$$

Demonstrație. Se face prin inducție. În cazul n=2 rezulta:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

Observație. Daca evenimentele $\{A_i\}_{1<i<n}$ sunt independente atunci formula devine:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i,j=1}^{n} P(A_i)P(A_j) + \ldots + (-1)^{n-1} \prod_{i=1}^{n} P(A_i)$$

b) Formula probabilității totale.

Daca A_1, A_2, … , A_n formează un sistem complet de evenimente, atunci pentru orice eveniment A avem:

$$P(A) = \sum_{i=1}^{n} P(A_i)P(A/A_1).$$

Demonstrație: Un eveniment A nu se poate realiza decât împreună cu unul şi numai unul din evenimentele A_1,A_2, …, A_n deci:

$$A = (A \cap A_1) \cup (A \cap A_2) \cup ... \cup (A \cap A_n).$$

Dacă aplicam probabilitatea peste această egalitate:

$$P(A) = \sum_{i=1}^{n} P(A \cap A_1) \sum_{i=1}^{n} P(A_i)P(A/A_1)$$

ceea ce era demonstrat.

Exemplu. Se dau doua urne identice în interior. Una conține trei bile albe şi patru bile negre, iar cealaltă, patru bile albe şi cinci bile negre. Din una din aceste urne aleasă la întâmplare se extrage o bilă. Care este probabilitatea ca această bilă să fie albă?

Rezolvare.

Se consideră evenimentele:
A_1 : Extragerea se face din prima urnă;
A_2 : Extragerea se face din a doua urnă;
A_3 : Bila extrasă este albă.
Se observă imediat că A_1, A_2 formează un sistem complet de evenimente şi

$P(A) = P(A_2) = \frac{1}{2}$; $P(A/A_1) = \frac{3}{7}$; $P(A/A_2) = \frac{4}{9}$.

Aplicând formula avem:

$$P(A) = P(A_1)P(A/A_1) + P(A_2)P(A/A_2) = \frac{65}{126}$$

c) Formula de înmulțire a probabilităților.

Fie $\{A_i\}_{1<i<n} \in K$ şi $P(A_1 \cap A_2 \cap ... \cap A_n) \neq 0$. atunci:

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)P(A_2/A_1)...P(A_n/A_1 \cap ... \cap A_{n-1}).$$

Demonstrație. Din definiția probabilității condiționate:

$$P(A_1) = P(A_1)$$

$$P(A_1 \cap A_2) = P(A_2)P(A_2/A_1)$$

$$P(A_1 \cap A_2 \cap A_3) = P(A_1 \cap A_2)P(A_3/A_1 \cap A_2)$$

$$P(A_1 \cap ... \cap A_n) = P(A_1 \cap ... \cap A_{n-1})P(A_n/A_1 \cap ...A_{n-1})$$

De unde prin înmulțire avem:

$P(A_1 \cap ... \cap A_n) = P(A_1)P(A_2 / A_1)...P(A_n / A_1 \cap ... \cap A_{n-1})$

Exemplu. Intr-o urnă sunt 5 bile albe şi 5 bile negre. Se scot trei bile una câte una fără întoarcerea bilei extrase în urnă. Care este probabilitatea obținerii a trei bile albe? Dar probabilitatea să avem doua bile albe şi o bilă neagră?

Rezolvare.

Notam cu:

A_1: prima bila extrasa este alba;

A_2: a doua bila este alba;

A_3: a treia bila este alba.

Cu aceste notații avem:

$P(A_1) = \frac{1}{2}$; $P(A_2/A_1) = \frac{4}{9}$; $P(A_3/A_1 \cap A_2) = \frac{3}{8}$

Iar

$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2) = \frac{1}{2} * \frac{4}{9} * \frac{3}{8} = \frac{1}{12}$

Pentru a răspunde la întrebarea a doua consideram evenimentul:

$B = (A_1 \cap A_2 \cap \overline{A}_3) \cup (A_1 \cap A_3 \cap \overline{A}_2) \cup (A_2 \cap A_3 \cap \overline{A}_1)$

De unde

$$P(B) = P(A_1 \cap A_2 \cap \overline{A}_3) + P(A_1 \cap A_3 \cap \overline{A}_2) + P(A_2 \cap A_3 \cap \overline{A}_1) =$$
$$= P(A_1)P(A_2/A_1)P(\overline{A}_3 / A_1 \cap A_2) + P(A_1)P(\overline{A}_2 / A_1)P(A_3 / A_1 \cap \overline{A}_2) +$$
$$+ P(\overline{A}_1)P(A_2 / \overline{A}_1)P(A_3 / \overline{A}_1 \cap A_2) = \frac{5}{36} + \frac{5}{36} + \frac{5}{36} = \frac{5}{12}$$

d) Formula lui Baves.

Daca $A_1, A_2, \dots, A_n$ formează un sistem complet de evenimente, atunci pentru orice $A \in K$ avem:

$$P(A_i/A) = \frac{P(A_i)P(A / A_i)}{P(A_1)P(A / A_1) + ... + P(A_n)P(A / A_n)}$$

Demonstrație. Conform regulii de înmulțire a probabilităților:

$P(A \cap A_i) = P(A)P(A_i/A)$

$P(A_i \cap A) = P(A_i)P(A/A_i)$

Din aceste egalități rezultă:

$$P(A_i/A) = \frac{P(A_i)P(A / A_i)}{P(A)}$$

Iar daca scriem P(A) conform formulei probabilității totale:

$$P(A_i/A) = \frac{P(A_i)P(A / A_i)}{\sum_{i=1}^{n} P(A_i)P(A / A_i)}$$

Exemplu. Se dau doua urne identice în interior. Una conține trei bile albe şi patru bile negre, iar cealaltă patru bile albe şi cinci bile negre. Din una din aceste urne aleasa la întâmplare se ia o bila, de asemenea la întâmplare. Daca bila extrasa este alba, care este probabilitatea ca ea sa provină din prima urna?

Rezolvare: este aceeaşi problema ca cea de mai sus cu deosebirea ca aplicam formula lui ***Bayes:***

$$P(A_1/A) = \frac{P(A_1)P(A/A_1)}{P(A_1)P(A/A_1)+P(A_2)P(A/A_2)} = \frac{\frac{3}{7}}{\frac{3}{7}+\frac{4}{9}} = \frac{27}{55}$$

e)Schema lui Poisson.

Daca evenimentele $A_1, A_2, \ldots, A_n$ au probabilități cunoscute:

$P(A_1) = p_1; P(A_2) = p_2; \ldots ; P(A_n) = p_n$

Atunci probabilitatea ca din cele n evenimente sa se realizeze *k* (şi sa nu se realizeze *n – k* evenimente) este coeficientul lui x^k din polinomul:

$Q(x) = (p_1x + q_1)(p_2x + q_2) \ldots (p_nx + q_n)$
Unde $q_i = 1 - p_i$; $i = \overline{1,n}$.

Demonstrație: Luam n = 4, k = 2. în aceste condiții evenimentul a cărui realizare înseamnă realizarea a doua evenimente şi nerealizarea celorlalte doua se scrie:

$$(A_1 \cap A_2 \cap \overline{A_3} \cap \overline{A_4}) \cup (A_2 \cap A_3 \cap \overline{A_1} \cap \overline{A_3}) \cup (A_1 \cap A_3 \cap \overline{A_2} \cap \overline{A_4}) \cup$$
$$\cup (A_1 \cap \overline{A_2} \cap \overline{A_3} \cap A_4) \cup (A_2 \cap A_3 \cap \overline{A_1} \cap \overline{A_4}) \cup (A_3 \cap A_4 \cap \overline{A_1} \cap \overline{A_2})$$

El este reuniunea a şase evenimente incompatibile, fiecare din acestea fiind intersecția a patru evenimente independente. Probabilitatea acestuia este:

$p_1p_2q_3q_4 + p_1p_3q_2q_4 + p_1p_4q_2q_3 + p_2p_3q_1q_4 + p_2p_4q_1q_3 + p_3p_4q_1q_2$

evident, tot atât este şi coeficientul lui x^2din polinomul
$Q(x) = (p_1x + q_1)(p_2x + q_2)(p_3x + q_3)(p_4x + q_4)$.

Exemplu: patru trăgători trag asupra unei ținte, primul cu probabilitatea 2/3, al doilea cu probabilitatea ¾, al treilea cu probabilitatea 4/5, iar al patrulea cu probabilitatea 5/6. care este probabilitatea ca ținta sa fie atinsa de exact trei ori?

Rezolvare:
A_1 : primul trăgător atinge ținta;
A_2 : al doilea trăgător atinge ținta;
A_3 : al treilea trăgător atinge ținta;
A_4 : al patrulea trăgător atinge ținta.

$p_1 = P(A_1) = \frac{2}{3}$; $p_2 = P(A_2) = \frac{3}{4}$; $p_3 = P(A_3) = \frac{4}{5}$; $p_4 = P(A_4) = \frac{5}{6}$

$q_1 = 1/3 \quad q_2 = 1/4 \quad q_3 = 1/5 \quad q_4 = 1/6$

probabilitatea ca din aceste patru evenimente, sa se realizeze trei şi numai trei şi unul nu, este coeficientul lui x^3 din polinomul:

$$Q(x) = (\frac{2}{3}x + \frac{1}{3})(\frac{3}{4}x + \frac{1}{4})(\frac{4}{5}x + \frac{1}{5})(\frac{5}{6}x + \frac{1}{6})$$

i. Schema lui Bernoulli

Se considera o experienţă şi un eveniment legat de aceasta experienţă, de probabilitate cunoscuta p. în acest caz probabilitatea ca acest eveniment sa se realizeze de *k* ori când se efectuează de *n* ori experienţă este:

$P_{n,k} = C_n^k p^k q^{n-k}$

Unde $q = 1- p$.

Demonstraţie: În general, fiind date n evenimente independente având aceeaşi probabilitate p. probabilitatea ca din acestea sa se realizeze exact k este, conform schemei lui Poisson, coeficientul lui k din polinomul $(px + q)^n$.

Exemplu: Se arunca perechea de zaruri de şase ori. Care este probabilitatea ca de exact trei ori sa se obţină şapte puncte?

p = 1/6; q = 5/6.

$$P_{6,3} = C_6^3 p^3 q^3 = C_6^3 \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^3$$

ii. Problema coincidentelor.

O urna conţine bile numerotate 1, 2, 3, …, n. se extrag la întâmplare una cate una toate aceste bile. Spunem ca la extragerea k s-a produs o coincidenta daca în urma acestei extrageri se obţine bila cu numărul k. care este probabilitatea a cel puţin unei coincidenţe?

Rezolvare: Notam cu A_i ($i = \overline{1,n}$) evenimentul obţinerii unei concordante la extragerea i. din cele n! moduri în care ar putea ieşi cele n bile (n-1)! Sunt favorabile evenimentului A. Intr-adevăr, cele n-1 bile: 1, 2, 3, …, i-1, i+1, …, n pot ieşi intr-o ordine oarecare. Deci :

$$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}, \quad (i = \overline{1,n})$$

Daca fixam doi indici oarecare I, j, 1<i<j<n atunci din cele n! cazuri posibile, (n-2)! cazuri sunt favorabile lui $A_i \cap A_j$ (deoarece intr-un caz favorabil (n-2) bile pot ieşi intr-o ordine oarecare).

$$P(A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k}) = \frac{(n-k)!}{n!}$$

Evenimentul A “aparţine unei coincidente” se poate scrie:

$A = A_1 \bigcup A_2 \bigcup \ldots \bigcup A_n$
Iar prin formula a) avem:

$$P(A) = \sum_{i=1}^{n} P(A_i) - \sum_{\substack{i \neq j \\ i,j=1}}^{n} P(A_i \bigcap A_j) = \ldots + (-1)^n P\left(\bigcap_{i=1}^{n} A_i\right).$$

În cazul nostru, prima suma are n termeni fiecare egal cu $\frac{(n-1)!}{n!}$, a doua are C_n^2 termeni, fiecare egal cu $\frac{(n-2)!}{n!}$, a treia are C_n^3 termeni, fiecare egal cu $\frac{(n-3)!}{n!}$ etc.
Rezulta:
P(A) = $C_n^1 \frac{1}{n} - C_n^2 \frac{(n-2)!}{n!} + C_n^3 \frac{(n-3)!}{n!} + \ldots + (-1)^n C_n^n \frac{1}{n!}$
Dar
$$C_n^k \frac{(n-k)!}{n!} = \frac{1}{k!}$$
De unde rezulta:
P(A) = $1 - \frac{1}{2!} - \frac{1}{3!} + \ldots + (-1)^n \frac{1}{n!}$
Probabilitatea de a nu avea nici o concordanta este:

$$P(\overline{A}) = 1 - P(A) = \frac{1}{2!} - \frac{1}{3!} + \ldots + (-1)^n \frac{1}{n!}$$

Observație. Se ştie ca
$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots$$
Iar pentru x = - 1 obținem:
$$e^{-1} = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} + \ldots = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{(k+1)!}.$$

5 Variabile aleatoare. Definiție şi proprietăți.

Fie $\{\Omega, \mathrm{K}\}$ câmp borelian de evenimente, **R** dreapta reala şi $\mathcal{B}$ corpul borelian din **R**.

Definiție. Se numeşte *variabila aleatoare definita pe un câmp borelian de evenimente* $\{\Omega, \mathrm{K}\}$ o aplicație X : **Ω** → **R** cu proprietatea ca $X^{-1}(A) \in \mathrm{K}$, $\forall A \in \mathfrak{R}$.

Observație. Pentru ca o aplicație X : **Ω** → **R** sa fie variabila aleatoare este suficient ca pentru orice $x \in$ **R**, $\{\omega : X(\omega) < x\} \in \mathrm{K}$. Aceasta deoarece mulțimea intervalelor $(-\infty, x)$, generează pe $\mathcal{B}$. De fapt este suficient ca aceasta condiție sa fie verificata pentru orice $x \in$ D, unde D este o parte densa a lui **R**.

În cele ce urmează vom studia variabila aleatoare simpla. În acest sens definim:

Definiție. Se numeşte *variabila aleatoare simpla* o variabila aleatoare care ia un număr finit de valori finite. Cu alte cuvinte X : **Ω** → **R** este o variabila simpla aleatoare daca:

1. ia un număr finit de valori (distincte) finite $\{x_1, \ldots, x_n\}$;
2. $\{\omega : X(\omega) = x_i\} \in K, \forall 1 \le i \le n$.

Sa notam cu $\Omega_j = \{\omega : X(\omega) = x_j\} \in K, \forall 1 \le i \le n$ unde evident:

$\Omega_i \bigcap \Omega_j = \varphi, \forall i \ne j$

$$\bigcup_{j=1}^{n} \Omega_j = \Omega$$

În acest caz variabila aleatoare simpla X se poate scrie:

$$X = \sum_{i=1}^{n} x_i \aleph_{\Omega_i}$$

unde $\aleph_{\Omega_i}$ este indicatorul evenimentului Ω_i.

Daca notam cu S(Ω, K) mulțimea variabilelor aleatoare simple, atunci avem evident următoarele proprietăți:

1. daca X, Y $\in S(\Omega, K)$ atunci X+Y $\in S(\Omega, K)$;
2. daca X, Y $\in S(\Omega, K)$ atunci XY $\in S(\Omega, K)$;
3. daca X$\in S(\Omega, K)$ şi $c \in \mathbf{R}$ atunci cX$\in (\Omega, K)$.

Teorema. *Fie X o variabila aleatoare; exista atunci un sir* $\{X_n\}_{n\in N}$ *de variabile aleatoare simple care converg către X, în plus daca X≥0, atunci şirul* $\{X_n\}_{n\in N}$ *este crescător.*

Demonstrație: Sa presupunem ca $X \ge 0$. Atunci pentru orice n$\in$**N**, avem:

$$X_n(\omega) = \begin{cases} \dfrac{i-1}{2^n} & \text{daca} \quad \dfrac{i-1}{2^n} \le X(\omega) < \dfrac{I}{2^n}, i = \overline{1.2''} \\ n \end{cases}$$

daca $X \ge n$

Evident, fiecare X_n este o variabila aleatoare simpla nenegativa, iar şirul $\{X_n\}_{n\in N}$ este prin construcție crescător. Daca X<∞ atunci pentru orice $n \in \mathbf{N}$ avem:

$$0 \le X - X_n \le \frac{i}{2^n} - \frac{i-1}{2^n} = \frac{1}{2^n}$$

şi deci:

$$\lim_{n\to\infty} X_n = X$$

daca X = +∞, atunci X_n = n pentru orice n $\in$ **N** şi deci şi în acest caz avem:

$$\lim_{n\to\infty} X_n = X.$$

6. Operații cu variabile aleatoare simple.

În toate cazurile când se vor face operații cu variabile aleatoare se va presupune ca toate variabilele aleatoare sunt ataşate aceleiaşi experiențe, deci toate variabilele aleatoare sunt definite pe aceeaşi mulțime de evenimente elementare, fără sa mai precizam acest lucru. De asemenea în toate calculele ce urmează vom presupune ca variabilele aleatoare X şi Y au următoarele tablouri de distribuție:

X: $\begin{pmatrix} x_1 ... x_n \\ p_1 ... p_n \end{pmatrix}$

Y: $\begin{pmatrix} x_1 ... x_m \\ q_1 ... q_m \end{pmatrix}$

a) *Adunarea variabilelor aleatoare.*

Variabila aleatoare X+Y are tabloul de distribuție următor:

X+Y: $\begin{pmatrix} x_1 + y_1 ... x_i + y_j ... x_n + y_m \\ p_{11} ... p_{ij} ... p_{nm} \end{pmatrix}$

Unde

$P_{ij} = P(X+Y = x_i+y_j) = P[(X = x_i) \cap (Y = y_j)]$

Iar

$\sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} = 1$

Dacă variabilele aleatoare sunt independente, atunci

$p_{ij} = P(X + Y = x_i + y_j) = p_i q_j$

iar daca variabilele nu sunt independente, atunci:

$p_{ij} = P(X + Y = x_i + y_j) = P(X = x_i) \cap (Y = y_j)) =$
$= P(X = x_i) P(Y = y_j \mid X = x_i) = P(Y = y_j) P(\mid X = x_i \mid Y = y_j)$

Variabila aleatoare X + Y se numeşte suma variabilelor aleatoare X şi Y.

În cazul particular când Y = a (a = constant) se obține suma dintre o variabila aleatoare şi o constanta care devine

X + a : $\begin{pmatrix} x_1 + a ... x_n + a \\ p_1 ... p_n \end{pmatrix}$

b) *Înmulțirea variabilelor aleatoare.*

Variabila aleatoare XY are tabloul de distribuție:

XY: $\begin{pmatrix} x_1 y_1 ... x_i y_j ... x_n y_m \\ p_{11} ... p_{ij} ... p_{nm} \end{pmatrix}$

Cu

$p_{ij} = P(XY = x_i y_j) = P[(X = x_i) \cap (Y = y_j)]$

Daca variabilele aleatoare X şi Y sunt independente atunci $p_{ij} = p_i q_j$, iar daca nu sunt independente avem:

$p_{ij} = P(XY = x_i y_j) = P[(X = x_i) \cap (Y = y_j)]$

daca variabilele aleatoare X şi Y sunt independente atunci $p_{ij} = p_i q_j$, iar daca nu sunt independente avem:

$p_{ij} = p_i P(Y = y_j \mid X = x_i) = q_i P(\mid X = x_i \mid Y = y_j)$

variabila aleatoare XY se numeşte produsul variabilelor X şi Y. În cazul în care X = Y avem

$$X^2 : \begin{pmatrix} x_1^2 \ldots x_n^2 \\ p_1 \ldots p_n \end{pmatrix}$$

7. Funcția de repartiție a unei variabile aleatoare. Definiție şi proprietăți.

Fie X o variabila aleatoare.

Definiție. Se numeşte *funcție de repartiție* a variabilei aleatoare X, funcția F : **R** → **R** definită prin relația:
$F(x) = P(X < x)$

Observație. Deoarece F se defineşte cu ajutorul unei probabilități, pentru orice $x \in \mathbf{R}$, evident
$0 \le F(x) \le 1$

Funcția de repartiție F se bucura de următoarele proprietăți:
1. $F(x_1) \le F(x_2)$ daca $x_1 < x_2$
2. $F(x - 0) = F(x)$ pentru orice $x \in \mathbf{R}$
3. $F(-\infty) = 0$
4. $F(+\infty) = 1$

Observație. Cu alte cuvinte funcția de repartiție a unei variabile aleatoare este o funcție nedescrescătoare, continua la ştanga şi ale cărei valori în $-\infty$ şi $+\infty$ sunt 0 şi 1.

În cazul unei variabile aleatoare discrete care are tabloul

$$X : \begin{pmatrix} x_1 \ldots x_n \\ p_1 \ldots p_n \end{pmatrix}$$

funcția de repartiție F corespunzătoare este:

$$F(x) = \begin{cases} 0 & \text{daca} \quad x \le x_1 \\ p_1 & \text{daca} \quad x_1 < x \le x_2 \\ p_1 + p_2 & \text{daca} \quad x_2 < x \le x_3 \\ \dots\dots\dots & \\ P_1 + p_2 + \dots + p_{n-1} & \text{daca} \quad x_{n-1} < x \le x_n \\ 1 & \text{daca} \quad x > x_n \end{cases}$$

Daca X este o valoare aleatoare continua, atunci exista o funcție f(x) numita *densitate de probabilitate* care are proprietățile:

1. $f(x) \ge 0, \ \forall x \in \mathbf{R}$;
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$.

În acest caz, definim *funcția de repartiție* în felul următor:

$$F(x) = \int_{-\infty}^{+\infty} f(t)dt$$

daca integrala converge. Evident

$F'(x) = f(x)$
Daca exista derivata lui F(x).

8. Caracteristicile numerice ale unei variabile aleatoare. Media şi dispersia.

Fie X o variabila aleatoare simpla având distribuția

$$X : \begin{pmatrix} x_1 \dots x_n \\ p_1 \dots p_n \end{pmatrix}$$

Definiție. Se numeşte *valoare medie* sau speranța matematica a variabilei aleatoare X numărul:

$$M(X) = \sum_{i=1}^{n} p_i x_i$$

Prezentam în continuare câteva proprietăți ale variabilelor aleatoare simple. În cele ce urmează variabilele aleatoare X şi Y au legile de probabilitate date de tablourile următoare:

$$X:\begin{pmatrix} x_1 \ldots x_n \\ p_1 \ldots p_n \end{pmatrix}$$

$$Y:\begin{pmatrix} x_1 \ldots x_n \\ q_1 \ldots q_n \end{pmatrix}$$

Proprietate. M(X+Y)=M(X)+M(Y).

Proprietate. M(Cx)=Cm(x), $\forall$ c $\in$ **R**.

Proprietate. Daca X<Y atunci M(X)<M(Y).

Proprietate. Daca X şi Y sunt variabile aleatoare independente, atunci M(XY)=M(X)M(Y).

Observație. Daca variabila aleatoare X are o mulțime numărabilă de valori x_1, x_2, …, x_n, …. Cărora le corespund probabilitățile p_1, p_2, …, p_n,…. Atunci numărul:

$$M(X) = \sum_{i\geq 1} p_i x_i$$

se numeşte *valoare medie* a variabilei aleatoare X daca seria $\sum_{i=1}^{\infty} p_i x_i$ este absolut convergenta.

Daca $\sum_{i=1}^{\infty} p_i x_i = \infty$ spunem ca X nu are valoare medie.

Daca X este o variabila aleatoare de tip continuu având densitatea de probabilitate f(x), atunci numărul:

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx$$

Se numeşte *valoare medie* a variabilei aleatoare X daca integrala din membrul drept exista. Proprietățile din cazul discret rămân valabile şi în cazul continuu.

Momente.

Definiție. Se numeşte *moment de ordin p, p*$\in$**N** , al variabilei aleatoare X numărul:

$$M_p(X) = \sum_{i=1}^{n} p_i x_i^p$$

în cazul discret , şi:

$$M_i(X) = \int_R x^p f(x)dx$$

în cazul continuu.

Definiție. Se numeşte *moment centrat de ordin p, p*$\in$ **N**, numărul:

$$\mu_p = M_p[X - M(X)] = M[(X - M(X))^p]$$

Observație. Intre momente şi momentele centrate au loc relațiile:

$$\mu_p(X) = \sum (-1)^i C_p^i M_{p-1}(X) M^i(X), p \in N - \{0\}$$

$$M_p(X) = \sum_{i=1}^{n} C_p^i \mu_{p-1} M^i(X), p \in N - \{0\}$$

Definiție. Se numeşte *dispersie* a variabilei aleatoare X şi se notează $D^2(X)$ sau σ^2 numărul

$$\sigma^2 = D^2(X) - M_2[X - M(X)] = M[(X - M(X))^2]$$

Numărul D(X) = $\sqrt{D^2(X)}$ se numeşte *abatere medie pătratică* şi joaca un rol important în aplicații.

Proprietăți:

1. $D^2(X) \geq 0$ *oricare ar fi variabila aleatoare X;*
2. $D^2(X) = 0$ daca şi numai daca x = m (sau mai exact P(x = m = 1);
3. $D^2(X) = M(X^2) - [M(X)]^2$;
4. $D^2(c X) = c^2 D^2(X)$;
5. daca variabilele aleatoare $X_1, \ldots, X_n$ sunt independente doua cate doua, atunci:

$$D^2\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} D^2(X_i)$$

Observație. Gradul de împrăştiere a valorilor variabilelor aleatoare X în jurul valorii sale medii se măsoară nu prin $D^2(X)$, ci prin abaterea medie pătratică : D(X) = $\sqrt{D^2(X)}$ care este exprimata în aceleaşi unități de măsura ca şi variabila aleatoare X.

9. Inegalitatea lui Cebîşev

Daca variabila aleatoare X ara valoare medie m şi dispersia σ^2, atunci: $P(|X - m| \geq k\sigma) \leq \frac{1}{k^2}, k \in R, k \geq 0$.

Observație. Fie X o variabila aleatoare, iar ε un număr pozitiv oarecare, atunci:

$$P(|X - M(X)| > \varepsilon) < \frac{D^2(X)}{\varepsilon^2}$$

sau trecând la evenimentul contrar:

$$P(|X - M(X)| < \varepsilon) \geq 1 - \frac{D^2(X)}{\varepsilon^2}.$$

În cazul notației : ε = kD(X), avem :

$$P(|X - M(X)|) < kD(X) \geq 1 - \frac{1}{k^2}$$

sau :

P[M(X) – k D(X) < X < M(X) + $kD(X)] \geq 1 - \frac{1}{k^2}$.

Inegalitatea lui Cebîşev da o margine inferioara pentru probabilitatea P $(|X - M(X)|) < \varepsilon$ şi se interpretează în felul următor : cu o probabilitate cel puțin egală cu 1 - $\frac{D^2(X)}{\varepsilon}$, respectiv 1 - $\frac{1}{k^2}$, variabila aleatoare X ia valori în intervalul [M(X) – ε, M(X) + ε], respective (M(X) – $kD(X), M(X)+kD(X))$ sau ceea ce revine la acelaşi lucru, variabila aleatoare X – M(X), adică abaterea lui X de la valoarea medie ia valori în intervalul (-ε, ε) corespunzător (-$kD(X), kD(X))$.

Se vede ca 1 - $\frac{1}{k^2}$ creste odată cu k şi ca deja pentru k = 6,1 - $\frac{1}{k^2}=\frac{35}{36}$, deci ia o valoare apropiata de 1.

Daca variabila aleatoare X are distribuție binomiala, atunci M(X) = np, $D^2(X)$ = npq şi inegalitatea lui Cebîşev are forma : $P(|X-np|<\varepsilon)\geq 1-\frac{npq}{\varepsilon^2}$, iar pentru variabila aleatoare $Y=\frac{X}{n}$ (frecventa relativa) obținem :

$P(|Y-p|\geq\varepsilon)\geq 1-\frac{npq}{\varepsilon^2}$ deoarece M(Y) = p şi $D^2(Y)$ = $\frac{pq}{n}$.

10. Repartiții clasice : binomială, Poisson, normală.

Distribuția binomială.

Definiție. Spunem că variabila aleatoare X are *distribuție binomială* cu parametrii n şi p (n întreg pozitiv, 0 < p < 1) dacă pentru orice k =$\overline{0,n}$:

$P(X=k) = C_n^k p^k q^{n-k}$

Dacă A_1, A_2, … , A_n sunt evenimente independente şi $P(A_i)$ = p, i = $\overline{1,n}$, iar X numărul evenimentelor care se realizează, atunci X are distribuție binomială cu parametrii n şi p (schema lui Bernoulli). În particular, daca A este un eveniment legat de o anumita experiență aleatoare şi probabilitatea ca A sa se producă când efectuam o singura data experiența este P(A) = p, atunci numărul X al realizărilor lui A când efectuam de n ori experiența are distribuție binomiala cu parametrii n şi p.

În acest caz: M(X) = np; $D^2(X)$ = npq; $\varphi_x(t) = (pe^{it} + q)^n$.

Distribuție binomiala cu exponent negativ.

Definiție. Spunem ca variabila aleatoare X are distribuție binomiala cu exponent negativ de parametrii m şi p (n număr întreg negativ, 0 < p <1) adică poate lua valorile m, m+1, … şi P(X=k) = $C_{k-1}^{m-1}p^m q^{k-m}$ **,** $k\geq$ m unde q = 1 – p.

experiența se efectuează pana la cea de a m-a realizare a unui eveniment A legat de ea. Daca probabilitatea acestui eveniment, când se face o singura data experiența este p, atunci numărul X de efectuări ale experienței este o variabila aleatoare care are distribuție binomiala cu exponent negativ de parametrii m şi n. intr-adevăr, evenimentul {X = k} se scrie ca intersecție de doua evenimente: “ în primele k-1 efectuări ale experienței A se produce de m-1 ori “ şi “ în a k-a efectuare a experienței se produce A “. Probabilitatea primului este $C_{k-1}^{m-1} p^{m-1} q^{k-m}$ (schema Bernoulli), iar probabilitatea celui de-al doilea este p. în acest caz:

P(X = k) = $C_{k-1}^{m-1} p^m q^{m-k}$

Media şi dispersia variabila aleatoare distribuite binomial negativ sunt:

M(X) = $\frac{m^2}{p^2}$; $\varphi_x(t) = \left(\frac{pe^{it}}{1-qe^{it}}\right)^n$.

Distribuția hipergeometrica.

Definiție. Variabila aleatoare X are distribuție hipergeometrica cu parametrii a, b, n (a, b, n – întregi pozitivi, $n \leq a+b$) şi min(n,a) şi:

P(X = k) = $\frac{C_a^k C_b^{n-k}}{C_{a+b}^n}$

Pentru orice k întreg, max(0, n-b) < k < min(n, a). se arata uşor ca:

$$\sum_k P(X=k) = 1$$

Daca dintr-o urna ce conține a bile albe şi b bile negre se extrag n bile una cate una, fără întoarcerea bilei extrase în urna (sau toate n simultan) iar cu X notam numărul de bile albe extrase, atunci X are distribuție hipergeometrica cu parametrii a, b, n. valoarea medie şi dispersia lui X sunt:

M(X) = np

unde

$$p = \frac{a}{a+b},$$

$$q = \frac{b}{a+b}$$

$$D^2(X) = npq\frac{a+b-n}{a+b-1}$$

Demonstrarea acestei afirmații se face cu ajutorul identificării:

$$\sum_{k=0}^{n} C_a^k C_b^{n-k} = C_{a+b}^n$$

(adică $\sum_k P(X=k) = 1$) care se mai poate deduce calculând coeficientul lui x^n în fiecare membru al relației:

$(1+x)^a(1+x)^b = (1+x)^{a+b}$

$$M(X) = \sum_k k\frac{C_a^k C_b^{n-k}}{C_{a+b}^k} = \frac{a}{C_{a+b}^k}\sum_k C_{a-1}^{k-1} C_b^{n-k} = \frac{a}{C_{a+b}^k} C_{a+b-1}^{n-1} = \frac{an}{a+b}$$

Pentru dispersie calculam:

$M(X)^2 =$

$$\sum k^2 \frac{C_a^k C_b^{n-k}}{C_{a+b}^k} = \sum k(k-1)\frac{C_a^k C_b^{n-k}}{C_{a+b}^k} + \sum k \frac{C_a^k C_b^{n-k}}{C_{a+b}^k} = \frac{a(a-1)}{C_{a+b}^n}\sum C_{a-2}^{k-2} C_b^{n-k} + M(X) =$$

$$= \frac{a(a-1)}{C_{a+b}^n} C_{a+b-2}^{n-2} + np$$

$D^2(X) = M(X^2) - [M(X)]^2 - npq \dfrac{a+b-n}{a+b-1}$

Distribuția Poisson.

Definiție. Spunem ca variabila X are *distribuție Poisson* cu parametrul λ (λ > 0) daca poate lua orice valoare întreagă negativa şi

$P(X = k) = \dfrac{\lambda^k}{k!} e^{-\lambda}$, k = 0,1,…

Evident:

$$\sum_{k=0}^{\infty} P(X = k) = e^{-\lambda} e^{\lambda} = 1$$

$M(\lambda) = \lambda$;
$D^2(X) = \lambda$;
$\varphi_X(t) = e^{\lambda(e^{it}-1)}$

Distribuția normala (legea normala).

Definiție. Spunem ca variabila aleatoare X are *distribuție normala* sau ca urmează legea normala cu parametrii m şi σ daca are densitatea de repartiție:

$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-m)^2}{2\sigma^2}}$, $(\ x \in R; \sigma > 0\ )$

Daca notam cu:

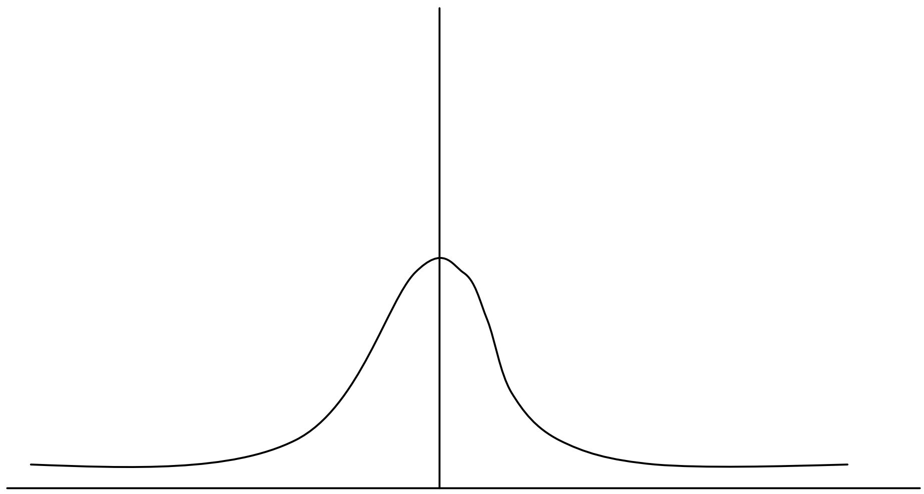
$$y = \frac{x-m}{\sigma\sqrt{2}}; dx = \sigma\sqrt{2}dy$$

rezulta

$$\int_{-\infty}^{\infty} f(x)dx = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-y^2} dy = \frac{2}{\sqrt{\pi}}\int_{0}^{\infty} e^{-y^2} dy$$

$$unde \int_0^{\infty} e^{-y^2} dy = \frac{\sqrt{\pi}}{2}$$

Graficul funcției lui f are forma de clopot. Dreapta x = m este o axa de simetrie a acestui grafic. Pentru x = m se obține maxima care este $\dfrac{1}{\sigma\sqrt{2\pi}}$. Punctele x = m + σ sunt puncte de inflexiune.



În cele ce urmează vom nota cu f(x;m;σ) densitatea de repartiție normala cu parametrii m şi σ. Funcția de repartiție F(x;0;1) este:

F(x;0;1) = $\int_{-\infty}^{x} f(t;0;1)dt = \int_{-\infty}^{0} f(t;0;1)dt + \int_{0}^{x} f(t;0;1)dt = \frac{1}{2} + \Phi(x)$

Unde:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} f(t;0;1)dt.$$

Pentru funcția integrala a lui Laplace sunt întocmite tabele.
Aceasta funcție este folosita nu numai pentru cazul legii normale cu 0 şi 1 ci şi pentru orice pereche de valori ale parametrilor m şi $\sigma, (\sigma > 0)$. Aceasta deoarece daca variabila aleatoare X urmează legea normala cu parametrii m şi σ atunci variabila aleatoare:

$$Y = \frac{1}{\sigma}(X - m)$$

Urmează legea normala cu parametrii 0 şi 1. intr-adevăr, daca F este funcție de repartiție a lui Y:
F(x) = P(Y < x) = P(X < m +σx) = F(m + σx; m,σ).

Densitatea de repartiție a variabilei aleatoare Y:

F(x) = F'(x) = σf((m + σx); m, σ) = $\frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}} = f(x;0,1)$

Daca X urmează legea naturala cu parametrii m şi σ şi a, b $\in$ R, atunci:

P(a < X < b) = $\Phi\left(\frac{b-m}{\sigma}\right)-\Phi\left(\frac{a-m}{\sigma}\right)$

Intr-adevăr:

P(a < X < b) = $P\left(\frac{a-m}{\sigma}<\frac{X-m}{\sigma}<\frac{b-m}{\sigma}\right)=F\left(\frac{b-m}{\sigma};0,1\right)-F\left(\frac{a-m}{\sigma};0,1\right)=$

$$=\left[\frac{1}{2}+\Phi\left(\frac{b-m}{\sigma}\right)\right]-\left[\frac{1}{2}+\Phi\left(\frac{a-m}{\sigma}\right)\right]=\Phi\left(\frac{b-m}{\sigma}\right)-\Phi\left(\frac{a-m}{\sigma}\right)$$

În particular:

P(│X – m │< kσ) = 2Φ(k)

Care rezulta din:

P(│X – m │<kσ) = P(m – kσ < X < m + kσ) =

= Φ(k) – Φ(-k) = 2Φ(k)

Deoarece

Φ(-k) = -Φ(k).

Sa calculam media şi dispersia lui X :

M(X) = $\int_{-\infty}^{\infty} xf(x;m,\sigma)dx=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} xe^{\frac{(x-m)^2}{2\sigma^2}}dx$

şi notând cu

$\frac{x-m}{\sigma}=y;dx=\sigma dy$

rezulta:

M(X) = $m\int_{-\infty}^{\infty} f(y;0,1)dy+\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty} ye^{\frac{-y^2}{2}}dy=m$

În mod similar:

D^2(X) = M[(X – m)^2] = $\int_{-\infty}^{\infty}(x-m)^2 f(x;m,\sigma)dx=\int_{-\infty}^{\infty}\frac{(x-m)^2}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}dx$

Facem din nou schimbarea:

$\frac{x-m}{\sigma}=y;dx=\sigma dy$

şi obţinem:

D^2(X) = $\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y^2e^{-\frac{y^2}{2}}dy$

Integrând prin părţi obţinem:

u = y; $du=ye^{-\frac{y^2}{2}}dy$

D^2(X) = $\sigma^2\int_{-\infty}^{\infty}(y;0,1)dy-\frac{\sigma^2}{\sqrt{2\pi}}ye^{-\frac{y^2}{2}}\Big|_{-\infty}^{\infty}=\sigma^2$

Luând k = 3 în egalitatea:

$P(|X - m| < 3\sigma) = 2\Phi(3) = 0,9974$

Aceasta egalitate exprima faptul ca practic aproape toate valorile variabilei X cad în intervalul (m - 2σ; m+ 3σ). Aceasta este aşa numita “regula a celor şase sigma”. Sa găsim acum şi celelalte momente centrate ale legii normale cu parametrii m şi σ. Fie k > 1 şi

$$\mu_k = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-m)^k e^{\frac{(x-m)^2}{2\sigma^2}} dx$$

facem aceeaşi schimbare

$$\frac{x-m}{\sigma\sqrt{2}} = y$$

de unde rezulta:

$$\mu_k = \frac{(\sigma\sqrt{2})^k}{\sqrt{\pi}} \int_{-\infty}^{\infty} y^k e^{-y^2} dy$$

Integrând prin părţi:

$$u = y^{k-1}; du = ye^{-\frac{y^2}{2}} dy$$

obţinem formula de recurenta:

$$\mu_k = (k-1)\sigma^2 \mu_{k-2}$$

Ştiind ca $\mu_1 = 0$, $\mu_2 = \sigma^2$ rezulta ca: $\mu_{2n-1} = 0$, $\mu_{2n} = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-1)\sigma^{2n}$.

Observaţie. Daca X şi Y urmează legea normala şi sunt independente, atunci X ± Y urmează de asemenea legea normala.

Observaţie. Legea normala reprezintă “cazul limita” a multor legi de probabilităţi.

Capitolul III

Metode de estimații

1. Estimații. Proprietăți.

Fie x o variabila aleatoare care are funcția de repartiție F(x;θ). Forma funcțională a funcției de repartiție f(x;θ) este specificată, însa θ este un parametru real a cărui valoare adevărata θ_0 este necunoscuta. θ_0 aparține unei mulțimi de valori reale Θ, numita spațiul parametrilor. În vederea găsirii unei valori care sa aproximeze pe θ folosim un eșantion

$$X_1 = x_1,\ X_2 = x_2,\ \dots,\ X_n = x_n,$$

şi ne propunem sa determinam o funcție $f(x_1, x_2, \dots, x_m)$, care sa poată fi luata ca valoare a parametrului θ.

Aceasta funcție o vom numi funcție de estimație sau *estimator*.

Definiție. Daca

$$\lim_{n\to\infty} f(x_1, x_2, \dots, x_m) \overset{P}{=} \theta \qquad (1)$$

$f(x_1, x_2, \dots, x_m)$ *se numeşte estimator corect.*

În relația (1) convergenta are loc în probabilitate.

Exista o infinitate de funcții care îndeplinesc condiția (1).

Dintre acestea prezintă cel mai mare interes cele pentru care convergenta este cat mai puternica.

Daca relația (1) are loc, aceasta înseamnă ca pentru valorile mari ale lui n, funcția de estimație $f(x_1, x_2, \dots, x_m)$ ia valori apropiate de θ, cu o probabilitate foarte mare.

Aceste valori aproximează valoarea parametrului θ deşi este natural ca funcția considerata sa fie luata drept un estimator al lui θ.

Definiția se păstrează şi în cazul când exista mai mulți parametri a căror valoare este necunoscuta. Vom da un exemplu din teoria erorilor. Sa presupunem ca am măsurat o lungime de n ori şi ca am obținut valorile:

$$x_1, x_2, \dots, x_n. \qquad (2)$$

Din experiențe anterioare şi din teoria generala a erorilor avem motive pentru a admite ca repartiția acestor erori este normala:

$$F(x; m, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Prin urmare, funcția de repartiție este cunoscuta dar nu cunoaştem valorile numerice ale parametrilor m şi σ.

Daca vom putea găsi doua funcții $f(x_1, x_2, \dots, x_n)$, $g(x_1, x_2, \dots, x_n)$ astfel încât

$$\lim_{n\to\infty} f(x_1, x_2, \dots, x_n) \overset{P}{=} m,$$

$$\lim_{n\to\infty} g(x_1, x_2, \ldots, x_n) \overset{P}{=} \sigma \;.$$

Problema este rezolvata întrucât cu o probabilitate atât de mare, cat dorim noi, pentru valorile lui n > N, unde N este un număr mare pe care îl putem determina, diferențele dintre $f(x_1, x_2, \ldots, x_n)$ şi $g(x_1, x_2, \ldots, x_n)$ devin arbitrar de mici.

Stabilirea faptului ca o funcție converge în probabilitate către o constanta poate prezenta uneori dificultăți. De aceea, este preferabil sa se recurgă la condiții mai simple. Astfel, daca

$$M(f(x_1, x_2, \ldots, x_n)) = \theta + \alpha(n), \quad \lim_{n\to\infty} \alpha(n) = 0 \;,$$

(3)

$$D^2(f(x_1, x_2, , \ldots, x_n)) \to 0$$

Rezulta imediat prin aplicarea inegalității lui Cebîşev, ca $f(x_1, x_2, \ldots, x_n)$ este un estimator corect pentru θ.

Din relațiile (3), α (n) este o funcție de n, care tinde către 0, când n creste necontenit. Daca α (n) = 0, vom spune ca $f(x_1, x_2, \ldots, x_n)$ este un estimator absolut corect pentru θ. Uneori, în acest caz, se spune ca $f(x_1, x_2, \ldots, x_n)$ este un estimator nedeplasat.

Vom da un exemplu simplu. Consideram un eveniment cu probabilitatea *p* de a se realiza intr-o proba. Supunem evenimentul la n probe independente şi cunoscând rezultatele fiecărei probe, ne propunem sa evaluam valoarea probabilităților necunoscute *p*. Este o problema de estimație cu un singur parametru *p*. Notam prin α numărul de realizări a evenimentului în n probe. Avem:

$$M\left(\frac{\alpha}{n}\right) = p\,, \quad D^2\left(\frac{\alpha}{n}\right) = \frac{pq}{n}.$$

Condițiile (3) sunt îndeplinite şi deci $\frac{\alpha}{n}$ este un estimator corect al lui *p*. Se vede ca $\frac{\alpha}{n}$ este un estimator nedeplasat.

Sa consideram funcția de selecție

$$\frac{\alpha+1}{n+1}.$$

Un calcul simplu ne arata ca

$$M\left(\frac{\alpha+1}{n+1}\right) = p + \frac{1-p}{n+1}$$

(4)

$$D^2\left(\frac{\alpha+1}{n+1}\right) = \frac{npq}{(n+1)^2} \;.$$

Condițiile (3) sunt îndeplinite deoarece:

$$\lim_{n\to\infty} \frac{1-p}{n+1} = 0\,, \qquad \lim_{x\to\infty} \frac{npq}{(n+1)^2} = 0\,.$$

Prin urmare şi funcția

$$\frac{\alpha+1}{n+1}$$

este un estimator, însa un estimator deplasat, în prima relație din (4) exista termenul $\alpha(n) = \frac{1-p}{n+1} \neq 0$.

Referitor la exemplul de mai sus, admitem ca am împărțit grupul de n experiențe în doua grupuri: n_1 şi n_2. Notam prin α_1 numărul de realizări ale evenimentului în primele n_1 probe şi prin α_2 numărul de realizări în ultimele n_2 probe ($n_1 + n_2 = n$).

Alegem ca estimator pentru *p*, expresia

$$\frac{k_1\alpha_1 + k_2\alpha_2}{k_1 n_1 + k_2 n_2}.$$

unde k_1 şi k_2 sunt doua numere reale, pozitive.

Efectuând calculele obținem:

$$M\left(\frac{k_1\alpha_1 + k_2\alpha_2}{k_1 n_1 + k_2 n_2}\right) = p;$$

(5)

$$D^2\left(\frac{k_1\alpha_1 + k_2\alpha_2}{k_1 n_1 + k_2 n_2}\right) = \frac{k_1 n_1^2 + k_2 n_2^2}{(k_1 n_1 + k_2 n_2)^2} \; .$$

Din aceasta relație, rezulta:

$$D^2\left(\frac{k_1\alpha_1 + k_2\alpha_2}{k_1 n_1 + k_2 n_2}\right) = D^2\left(\frac{\alpha}{n}\right) + \frac{(k_1 - k_2)^2 n_1 n_2}{n(k_1 n_1 + k_2 n_2)} \; .$$

Prin urmare:

$$D^2\left(\frac{k_1\alpha_1 + k_2\alpha_2}{k_1 n_1 + k_2 n_2}\right) \geq D^2\left(\frac{\alpha}{n}\right) ,$$

Egalitatea neavând loc decât pentru $k_1 = k_2$ când estimatorul $\frac{k_1\alpha_1 + k_2\alpha_2}{k_1 n_1 + k_2 n_2}$ devine egal cu $\frac{\alpha}{n}$.

Din (5) rezulta ca toți estimatorii

$$\frac{k_1\alpha_1 + k_2\alpha_2}{k_1 n_1 + k_2 n_2} , \quad k_1 \in N, k_2 \in N$$

sunt absolut corecți (nedeplasați). Printre ei se afla şi estimatorul $\frac{\alpha}{n}$, obținut pentru $k_1 = k_2$, care se bucura de proprietatea ca are cea mai mica dispersie. Aceasta proprietate de minim are importanta din punct de vedere statistic. Ea da naştere la cea mai mare concentrare de masa a probabilității în jurul lui *p* şi deci estimatorul de dispersie minima este întotdeauna preferabil celorlalți estimatori.

2. Metoda verosimilității maxime

Vom demonstra o teorema generala privind minimul dispersiei estimatorilor. În acest scop vom introduce următoarea

Definiție. Daca

$x_1, x_2, \dots, x_n$

este un eşantion asupra unei variabile aleatoare X, având densitatea de probabilitate f((x); θ), funcția

$$P(x;\theta) = P(x_1, x_2, \dots, x_n ; \theta) = \prod_{i=1}^{n} f(x_1;\theta)$$

ca funcție de θ, adică P(· ;θ), poarta numele de **funcție de verosimilitate.**

Este evident, conform definiției ca:

$$\int\int\dots\int_{R^n}\prod_{i=1}^{n} f(x_i;\theta)dx_1, dx_2,\dots,dx_n = 1$$

sau

$$\int_{R^n} Pdv = 1$$

unde

$$dv = \prod_{i=1}^{n} dv_i$$

Definiție : Numim ecuație de verosimilitate maxima, ecuația :

$$\frac{\partial \ln P}{\partial \theta} = 0 \ ;$$

iar soluțiile acestei estimații se numesc *estimații consistente pentru parametrul* θ.

În cazul repartiției normale avem :

$$P = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-m)^2}{2\sigma^2}} = \sigma^{-n}(\sqrt{2\pi})^{-n} \cdot e^{-\sum_{i=1}^{n}\frac{(x_i-m)^2}{2\sigma^2}}$$

$$\ln P = -\frac{n}{2}\ln\sigma^2 - n\ln\sqrt{2\pi} - \sum_{i-1}^{n}\frac{(x_i-m)^2}{2\sigma^2}$$

$$\frac{\partial \ln P}{\partial m} = \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m) = 0$$

Dacă

$$\sum x_i - n \cdot m = 0 \qquad \text{şi} \qquad \overline{m} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

atunci

$$\frac{\partial \ln P}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{1}{2\sigma^n}\sum_{i=0}^{n}(x_i - \overline{x})^2 = 0$$

$$\overset{\wedge}{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$$

este o estimație consistentă.

Iar

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

este o estimație absolut corectă.

Teorema Rao – Cramer.

Dacă

1. mulțimea valorilor lui θ formează un interval D al dreptei reale.
2. $\frac{\partial P(x;\theta)}{\partial \theta} < \infty$, oricare ar fi $\theta \in D$ şi $x \in R^n$
3. $\frac{\partial}{\partial \theta}\int_{R^n} P(x;\theta)dv = \int_{R^n}\frac{\partial P(x;\theta)}{\partial \theta}d\theta$ oricare ar fi $\theta \in D$.
4. pentru orice estimator t_n nedeplasat : $\frac{\partial}{\partial \theta}\int_{R^n} t_n P dv = \int_{R^n}\frac{\partial}{\partial \theta} t_n P dv$
5. $M\left[\left(\frac{\partial \ln P(x;\theta)}{\partial \theta}\right)^2\right] < \infty$, pentru toți $\theta \in D$,
6. t este un estimator nedeplasat al lui $\tau(\theta)$, atunci:
7. $D^2(t) \geq [\tau'(\theta)]^2 / M\left(\frac{\partial \ln P(x,\theta)}{\partial \theta}\right)^2$,

egalitatea având loc numai daca exista o constanta A, astfel încât

$$A[t - \tau(\theta)] = \frac{\partial \ln P(x;\theta)}{\partial \theta}.$$

Demonstrație. Din

$$\int_{R^n} P dv = 1$$

deducem

$$\int_{R^n}\frac{\partial \ln P}{\partial \theta} P dv = 0.$$

Derivând mai departe în raport cu θ:

$$\int\left[\frac{\partial^2 \ln P}{\partial\theta^2}+\left(\frac{\partial \ln P}{\partial\theta}\right)^2\right]Pdv=0\,.$$

Aceasta înseamnă ca:

$$M\left(\frac{\partial^2 \ln P}{\partial\theta^2}\right)=-M\left[\left(\frac{\partial \ln P}{\partial\theta}\right)\right]^2$$

t fiind un estimator nedeplasat pentru τ(θ), avem:

$$M(t)=\int_{R^n} tPdv=\tau(\theta)$$

şi prin derivare:

$$\int t\frac{\partial \ln P}{\partial\theta}Pdv=\tau'(\theta)$$

sau

$$\tau'(\theta)=\int_{R^n}(t-\tau'(\theta))\frac{\partial \ln P}{\partial\theta}Pdv\,.$$

Aplicând inegalitatea lui Schwartz, se obține:

$$\tau'^2(\theta)\le\int_{R^n}[t-\tau(\theta)]^2Pdv\int_{R^n}\left(\frac{\partial \ln P}{\partial\theta}\right)^2Pdv$$

sau

$$\tau'^2(\theta)\le M[t-(\theta)]^2M\left(\frac{\partial \ln P}{\partial\theta}\right)^2.$$

Însa

M[t – τ (θ)]^2 = D^2(t).

Prin urmare:

$$D^2(t)\ge\frac{\tau'^2(\theta)}{M\left(\frac{\partial^2 \ln P}{\partial\theta^2}\right)}.$$

Teorema este deci demonstrata. Mai rămâne sa se examineze cazul când inegalitatea de mai sus se transforma în egalitate.

Pentru aceasta, trebuie ca t – τ (θ) sa fie proporțională cu $\frac{\partial \ln P}{\partial\theta}$, adică $A[t-\tau(\theta)]=\frac{\partial \ln P}{\partial\theta}$

Unde A este independenta de valorile observate de selecție, depinzând numai de θ. Prin ridicare la pătrat:

$$A^2(\theta)[t-\tau(\theta)]^2=\left(\frac{\partial \ln P}{\partial\theta}\right)^2,$$

$$M\left(\frac{\partial \ln P}{\partial\theta}\right)^2=\int_{R^n}A(\theta)[t-\tau(\theta)]\frac{\partial \ln P}{\partial\theta}Pdv=A(\theta)\tau'(\theta),$$

$D^2(t) = \tau'(\theta) / A(\theta)$,

dispersia în acest caz căpătând aceasta forma simpla.

Definiție. *Un estimator pentru care inegalitatea (6) se transforma în egalitate se numeşte eficient.*

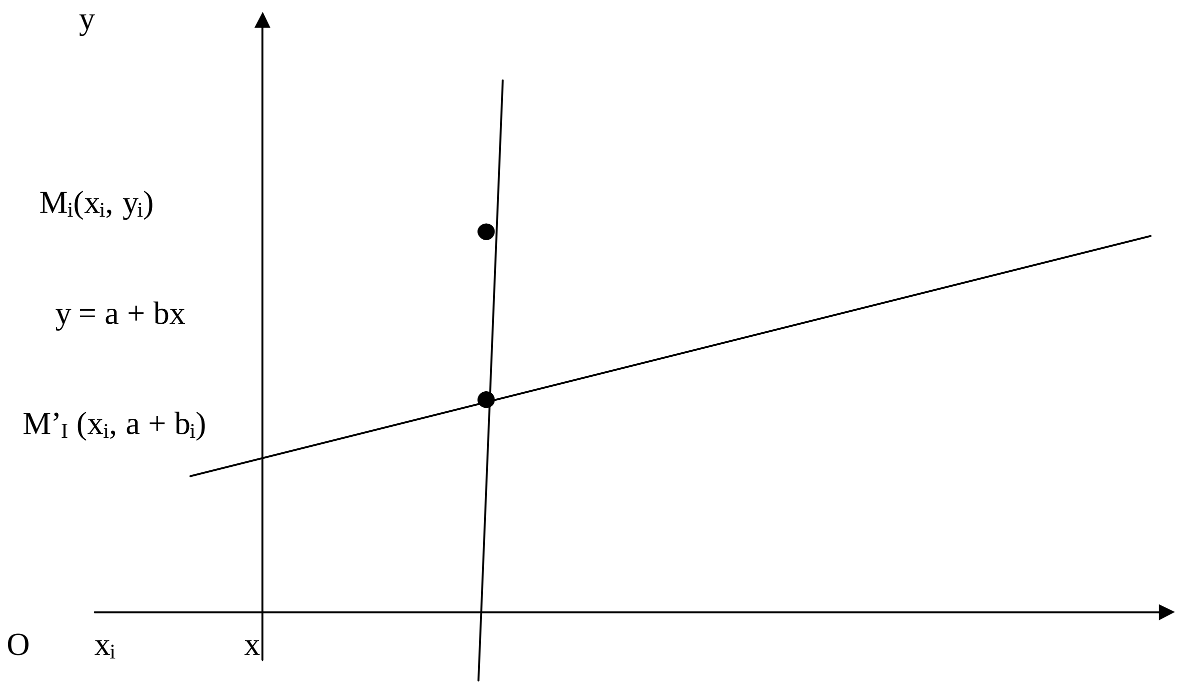
3. Metoda celor mai mici pătrate

Sa consideram doua axe OX şi OY şi o serie de puncte $M_1(x_1, y_1), \dots, M_n(x_n, y_n)$ care corespunde unui anumit fenomen. Punctele M_i, ($i = \overline{1,n}$) se dispun cu o oarecare neregularitate întâlnita în aproape toate seriile statistice. Scopul ajustării este sa găsim o curba care se apropie cel mai mult de punctele date M_i $\left(i = \overline{1,n}\right)$ şi care din aceasta cauza, sa putem admite ca nu se poate indica de dezvoltare a fenomenului respectiv.

Sa consideram seria de puncte M_i şi dreapta $y = a + by$.

Vom considera coeficienții necunoscuți a şi b, astfel încât expresia

$$\sum_{i=1}^{n} \left(a + bx_i - y_i\right)^2 = \min \qquad (1)$$



Din figura reiese imediat semnificația acestei metode. Fie M_i (x_i, y_i) un punct al seriei statistice date. Paralela Oy dusa prin M_i (x_i, y_i) taie dreapta $y = a + bx$ în punctul M'_i (x_i, $a + b + x_i$). în acest caz, relația (1) ne arata ca trebuie sa determinam pe a şi b, astfel încât sa avem:

$$\sum_{i=1}^{n} \overline{M_i M_i'}^2 = \min$$

Suma pătratelor diferențelor dintre ordonatele teoretice şi ordonatele empirice trebuie sa fie minim. Din acest motiv procedeul de mai sus poarta numele de “metoda celor mai mici pătrate “.

Pentru determinarea lui a şi b, vom considera funcția:

$$F(a,b) = \sum_{i=1}^{n} \left(a + bx_i - y_i\right)^2$$

şi vom obține prin derivare sistemul:

$$\frac{\partial F}{\partial a} = 0$$

$$\frac{\partial F}{\partial b} = 0$$

care poate fi scris sub forma:

$$\begin{cases} na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases}$$

care se numeşte sistem de ecuații normale.
Prin rezolvare se obțin estimații pentru parametrii a şi b.

Formula de interpolare Lagrange

Consideram din nou seria statistica în plan M_i (x_i, y_i), $i = \overline{1,n}$ şi ne propunem sa determinam polinomul în x de grad n-1 care pentru x_1, …, x_n ia valorile y_1, …, y_n. coeficienții acestui polinom sunt funcții liniare de y_i.

Putem scrie P(x) = $y_1 f_1(x) + \ldots + y_n f_n(x)$ unde $f_i(x)$ sunt polinoame de grad n-1 pe care urmează sa le determinam.

Observam ca, pentru x = x_i trebuie sa avem $P(x_i) = y_i$, de unde rezulta conditiile: $f_i(x_i) = 1$ şi $f_k(x_i) = 0$, $\forall i \neq k$; i, k = $\overline{1,n}$.

În acest caz polinomul $f_1(x)$ se anulează pentru x_2, x_3, …, x_n şi deci admite aceste valori ca rădăcini, obținându-se:

$$f_1(x) = k\,(x - x_2) \ldots (x - x_n)$$

unde k este o constanta. Însa pentru x = x_1, $f_1(x) = 1$ de unde rezulta:

$$k = \frac{1}{(x_1 - x_2)\ldots(x_1 - x_n)}.$$

În mod analog, se determina şi celelalte polinoame $f_i(x)$.
Rezulta formula de interpolare a lui Lagrange sub forma:

$$P(x) = y_1 \frac{(x-x_2)\dots(x-x_n)}{(x_1-x_2)\dots(x_1-x_n)} + \dots + y_i \frac{(x-x_1)(x-x_2)\dots(x-x_{i-1})(x-x_{i+1})\dots(x-x_n)}{(x_i-x_1)(x_i-x_2)\dots(x_i-x_{i-1})(x_i-x_{i+1})\dots(x_i-x_n)} +$$

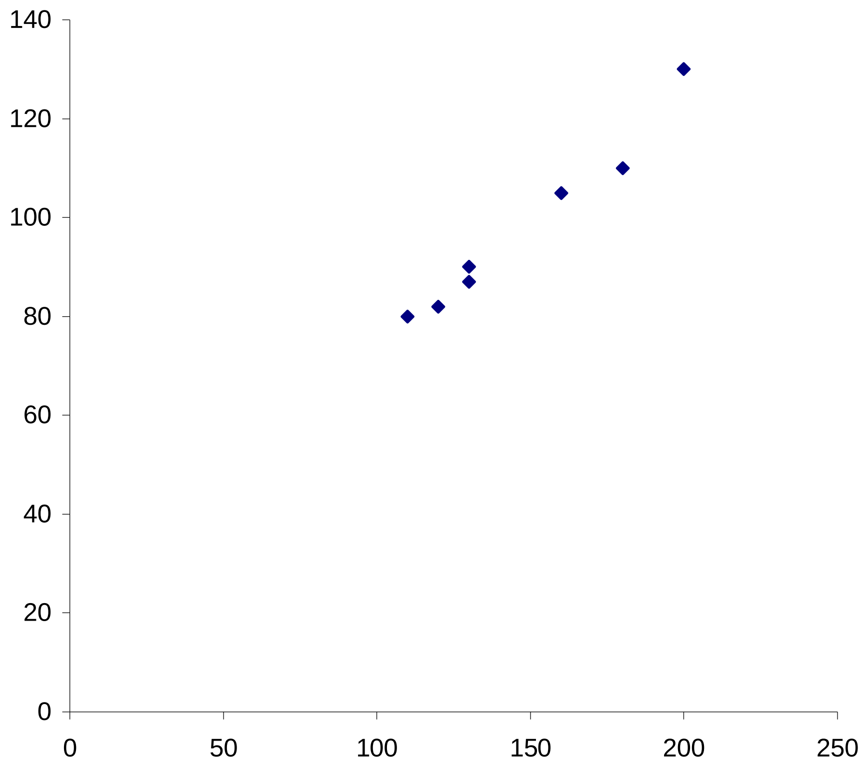
$$\dots + y_n \frac{(x-x_1)\dots(x-x_{n-1})}{(x_n-x_1)\dots(x_n-x_{n-1})}.$$

Capitolul IV
Aplicații

1. Pentru un lot de sportivi din atletism s-au făcut verificări pentru genoflexiuni cu bare şi ridicarea halterei la stilul smuls, obținându-se rezultatele din tabelul 1, exprimate în kilograme.

Tabelul 1

Nr. Crt.	Genoflexiuni x_i	Stilul smuls y_i
1	200	130
2	180	110
3	160	105
4	130	87
5	130	90
6	120	82
7	110	80



Sa se reprezinte punctele $(x_i\ ,\ y_i)$ şi sa se determine dreapta de ajustare prin metoda celor mai mici pătrate.

Rezolvare: Se obține figura 1 iar

$\overline{x}$ = 147,14 $\overline{y}$ = 97,71 n=7, deci

$$a=\frac{\sum_{i=1}^{7}x_i y_i - n\overline{x}\,\overline{y}}{\sum_{i=1}^{7}x_i^{\,2} - n\overline{x}^2} = \frac{3610,66}{6748,75} = 0,53\,.$$

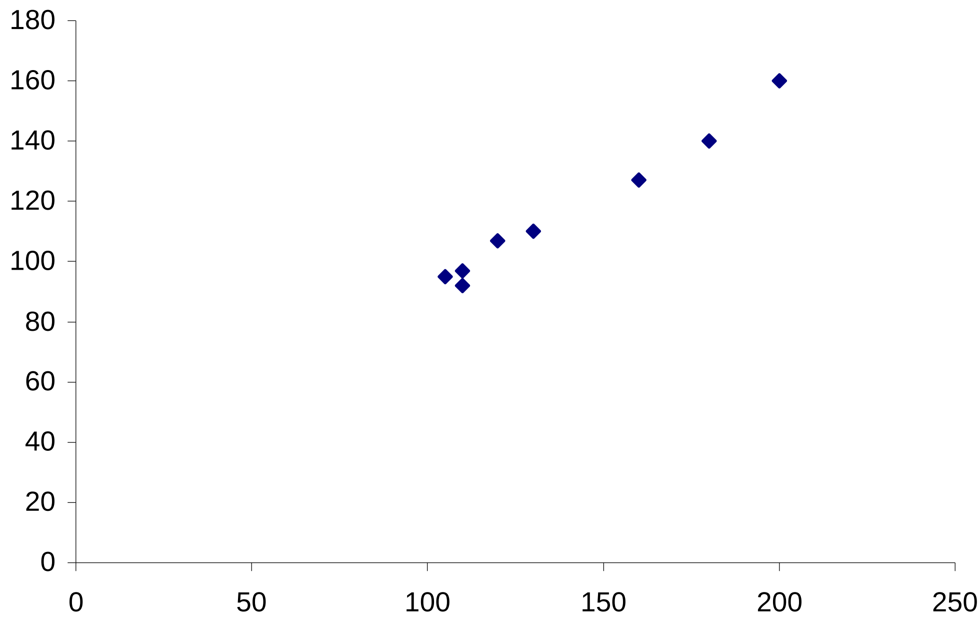
Iar $b=\overline{y}-a\overline{x}$ = 19,73, deci y = 0,53x + 19.73.

Rezultatul la acest stil dinamic este puternic influențat de forța picioarelor, aceasta din urma fiind indicata şi prin rezultatele la genoflexiuni, ceea ce justifica aşezarea punctelor (x_i, y_i), în jurul unei drepte.

2. Pentru un lot de sportivi din atletism s-au făcut verificări pentru genoflexiuni cu bare şi ridicarea halterei la stilul aruncat. Rezultatele obținute exprimate în kilograme sunt cuprinse în tabelul 2.

tabelul 2

Nr. Crt.	Genoflexiuni x_i	Stilul aruncat y_i
1	200	160
2	180	140
3	160	127
4	130	110
5	120	107
6	110	97
7	110	92
8	105	95



Sa se reprezinte punctele (x_i, y_i) şi sa se determine dreapta de ajustare prin metoda celor mai mici pătrate.

Rezolvare: Se obține figura 2.

$\bar{x}$ =139,37; $\bar{y}$ = 116 ; n=8, deci

$$a = \frac{6089{,}44}{9133{,}03} = 0{,}66 \qquad b=24{,}02.$$

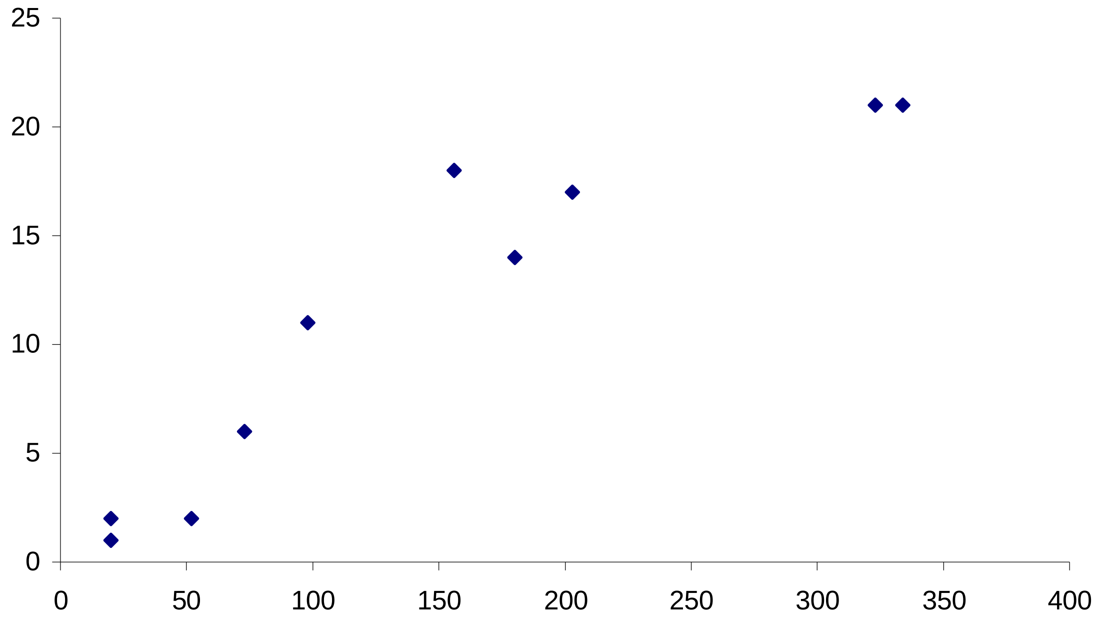
Deci y= 0,66x + 24,02.

Punctele (x_i, y_i) au putut fi ajustate printr-o dreapta, deoarece exista o legătura foarte puternica intre acest stil dinamic şi genoflexiuni cu bara.

3. Numărul de copii înscrişi şi numărul de cadre didactice din 10 unități şcolare şi preşcolare este prezentat în tabelul 3.

Tabelul 3

Nr. crt. al unității şcolare	Nr. copii înscrişi (x_i)	Nr. cadre didactice (y_i)	$\bar{y}_i$
	persoane	persoane	
1	20	2	3
2	323	21	23
3	156	18	12
4	180	14	14
5	98	11	8
6	73	6	7
7	334	21	24
8	20	1	3
9	52	2	5
10	203	17	15
Total	1459	113	113



Rezolvare:
Se notează cu x- numărul de copii înscrişi şi cu y - numărul de cadre didactice. Pentru a analiza legătura dintre cele doua variabile, aplicam un model de regresie liniara simpla de forma:

$y_i = a + b*x_i + e_i$

$\overline{y} = a\overline{x} + b$

Prin aplicarea metodei celor mai mici pătrate, rezulta sistemul:

$$\begin{cases} a\sum x_i = \sum y_i - nb \\ a\sum x_i^2 + b\sum x_i = \sum x_i y_i \end{cases}$$

$$a = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} = \frac{10*24256 - 1459*113}{10*332267 - (1459)^2} = \frac{77693}{1193989} = 0{,}06507$$

$$b = \frac{\sum y_i * \sum x_i^2 - \sum x_i \sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2} = \frac{113*332267 - 1459*24256}{10*332267 - 1459^2} = \frac{2156667}{1193989} = 1{,}80627\,.$$

1) De unde: $\overline{y}$ = 1,80627*x_i + 0,06507

Valorile ajustate ale numărului de cadre didactice în funcție de numărul de copii înscrişi sunt calculate în coloana $\overline{y}$ din tabelul 3. Intre cele doua variabile stabilite la început este o legătura directa (b > 0).Daca numărul de copii înscrişi creste cu 100 persoane, atunci numărul de cadre didactice va creste în medie cu ~ 7 persoane.

4. Ni se dau frecventele corespunzătoare, în procente, pentru următoarele grupe de vârsta ; prin x notând vârsta :

$12 \leq x < 15$ … 7,9%
$15 \leq x < 20$ … 8,9%
$20 \leq x < 30$ … 18,0%.

Se cere procentul pentru grupa $15 \leq x < 16$.

Grupele nu corespund întotdeauna la intervale egale. De multe ori ne interesează frecventa corespunzătoare unui interval care nu se găseşte cuprins în tabel, dar care poate fi dedus din prin interpolare.

Rezolvare : Plecam de la polinomul
$f(x) = a_0 + a_1x + a_2x^2$,
pentru care punem condițiile :
$f(0) = 7{,}9$, $f(1) = 16{,}8$, $f(3) = 34{,}8$.

Pentru determinarea lui f(x) vom folosi formula de interpolare a lui Lagrange, care ne da expresia unui polinom de gradul n, care pentru $x = x_i$ (i = 0, 1, 2, …, n).

Acest polinom se poate pune sub forma :

$$\sum_{k=o}^{n} \frac{(x-x_0)(x-x_1)...(x-x_{k-1})(x-x_{k+1})...(x-x_n)}{(x_x-x_0)(x_k-x_1)...(x_k-x_{k-1})(x_k-x_{k+1})...(x_k-x_n)} y_k \ .$$

Pentru k=2 şi ținând seama de valorile lui f(x) în punctele x = 0, x = 1, x = 3, avem :

$$f(x) = 7{,}9\frac{(x-1)(x-3)}{3} - 16{,}8\frac{x(x-3)}{2} + 34{,}8\frac{x(x-1)}{6} \ .$$

Rezulta :
$f(0{,}2) = 9{,}675$;
$f(0{,}2) - f(0) = 1{,}78$.

Aceasta problema nu admite o soluție unica. Folosind alte metode de interpolare, vom găsi alte valori, diferite intre ele dar în general apropiate.

BIBLIOGRAFIE

i. Gheorghe Florin Cristian, « Curs de statistică », Editura Universitatea din Piteşti
ii. Gheorghe Florin Cristian – Obreja George « Curs de matematici aplicate în economie », Editura Sirona, 2002
iii. Sânboan A., Reischer C., « Culegere de probleme de teoria probabilităților şi statistică matematică », Editura Didactică şi Pedagogică, 1972

Cuprins