

Elemente de statistica

Pe teritoriul tarii noastre, lucrarea lui Dimitrie Cantemir-”*Descriptio Moldaviae*”(1716) poate fi considerata ca o prima *lucrare de statistica*. Ea a fost scrisa la cererea Academiei din Berlin si continea toate cunostintele acumulate in domeniu la acea data.

Statistica este disciplina care se ocupa cu culegerea, inregistrarea, gruparea, analiza si interpretarea datelor referitoare la un anumit fenomen precum si cu formularea unor previziuni privind comportarea viitoare a acestuia.

Activitatea de culegere si inregistrare a datelor referitoare la un fenomen face obiectul *statisticii descriptive* sau *statisticii formale*.

Activitatea de grupare, de analiza si de interpretare a datelor precum si formularea unor previziuni privind comportarea viitoare a unui fenomen reprezinta obiectul *statisticii matematice*.

Elemente de limbaj in statistica. Date statistice.

- ✦ Multimea pe care se realizeaza un studiu statistic se numeste **populatie statistica**.
- ✦ Elementele componente ale unei populatii se numesc **unitati statistice** sau **indivizi**.
- ✦ Numarul total de unitati statistice se numeste **efectivul total** al populatiei statistice.
- ✦ O parte a populatiei statistice aleasa special pentru a fi studiată se numeste **esantion**.
- ✦ Proprietatea sau indicatorul in functie de care se cerceteaza o populatie statistica se numeste **caracteristica** sau **variabila statistica**.
- ✦ Exemplu: Sa presupunem ca ne intereseaza studiul privind mediile generale obtinute de elevii unei clase la sfarsitul anului scolar. In acest caz: populatia statistica o constituie multimea

elevilor clasei;unitatile statistice sunt elevii clasei;caracteristica este media generala;efectul total al populatiei este numarul tuturor elevilor clasei.

- ✦ *Din exemplul de mai sus se observa ca exista caracteristici care sunt exprimate numeric si altele care se nu exprima numeric.*
- ✦ *O caracteristica se numeste **caracteristica calitativa** daca nu poate fi masurata(valoarea ei nu se exprima numeric): o stare civila, calificativul, profesia.*
- ✦ *O caracteristica se numeste **cantitativa** daca se poate exprima numeric: media generala, salariul, inaltimea.*
- ✦ *O caracteristica cantitativa se numeste **discontinua** sau **discreta** daca nu poate lua decat valori numerice izolate.*
- ✦ *O caracteristica cantitativa care poate lua orice valoare dintr-un interval de lungime finita sau infinita se numeste **caracteristica cantitativa continua**.*
- ✦ *Intervalele in care o caracteristica ia valori se numesc **grupe** sau **clase de valori**.*
- ✦ *Exemplu: Nota de la teza, media generala sunt caractreristici cantitative discrete.Inaltimea este o caracteristica cantitativa continua in fuctie de care poate fi studiat un grup.*

Culegerea, inregistrarea si clasificarea datelor statistice.

Sa consideram studiul efectuat asupra unui grup de sportivi dupa inaltime (exprimata in centimetri).

Rezultatele masuratorii sunt inregistrate In ordinea in care a decursmasuratoarea si sunt asezate in urmatorul tabel:

165	168	177	195	172	198	196	190	201	168
172	168	168	196	173	199	182	195	196	196

185	205	184	192	178	165	174	182	177	172
196	192	188	195	175	192	175	184	192	170
184	205	190	200	188	176	184	174	188	170
170	180	184	199	192	184	170	175	184	188

cm	Nr. Sportivi	cm	Nr. Sportivi
165	2	174	1
168	4	175	3
170	4	176	1
172	3	177	2
173	1	178	2

Cu ajutorul acestei clasificari a datelor se pot obtine niste concluzii mai rapide privind particularitatile acestui grup: inaltimea cea mai mica, cea mai mare, numarul de indivizi care intrunesc aceeasi caracteristica, care este inaltimea domonanata si altele.

Rezultatele valorilor caracteristicii stidiate in clase bde valori. analizei statistice uneori pot fi obtinute mai usor daca se practica o alta grupare a datelor statistica.

De exemplu, se poaate face o grupare a

Clase de valori	Nr. Sportivi
{165,170)	6
{170,175)	9
{175,180)	8

{180,185)	10
{185,190)	5
{190,195)	7
{195,200)	11
{200,205)	1

Cu ajutorul acestui tabel de date se pot obtine cu usurinta diferite informatii despre grupul de studiu, ca de exemplu:-sunt 60 de sportivi; cei mai multi sportivi au inaltimea in clasa de valori {195,200), urmand cei cu inaltimea in clasa de valori {180,185);noua sportivi au inaltimea in clasa de valori {170,175) si reprezinta 15% din efectivul populatiei statistice; exista 45 de sportivi cu inaltimea sub 195 cm, etc.

Valoarea absoluta a diferentei extremitatilor unnei clase de valori se numeste ***amplitudinea clasei***.

Amplitudinea se stabileste de cel care realizeaza studiul.

O clasa de valori este in general un interval semideschis{a,b), exceptier facand eventual ultima clasa care poate fi luata de forma {cd}.

Clase de valori	{x1,x2)...{xp-1,xp)
Frecventa absoluta	n1.....np
Valorile caract.	x1 x2...xp
Frecventa absoluta	n1 n2...np

Frecventa absoluta cumulata crescatoare a valorii x a variabilei statistice este suma tuturor frecventelor absolute ale valorilor variabilei care apar pana la xi, inclusiv.

Frecventa absoluta cumulata descrescatoare a valorii xi a variabilei statistice este suma tuturor frecventelor absolute ale valorilor variabile care apar dwe la xi, inclusiv.

Analog se definesc ***frecventele absolute cumulate ale claselor de valori*** ale variabilei.

Raportul dintre frecventa absoluta a unei valori xi sau a unei clase de valori a variabilei statistice si efectivul total al populatiei se numeste

frecventa relativa a valorii xi, respectiv frecventa relativa a clasei de valori.

*Se numeste **frecventa relativa cumulata crascatoare** a valorii xi a variabilei statistice, suma tuturor frecventelor relative ale valorilor care apar pana la xi, inclusiv.*

*Se numeste **frecventa relativa cumulata crescatoare** a valorii xi a variabilei statistice, xsuma tuturor frecventelor relative ale valorilor care apar de la xi, inclusiv.*

Nota xi	Frecventa absoluta ni	Frecventa crescatoare	Frecventa descresc.	Frecventa relativa	Frecventa relativa ctresc	Frecventa relativa descr
4	1	1	50	0.02=2%	0.02=2%	1=100%
5	4	5	49	0.08=8%	0.1=10%	0.98=98%
6	5	10	45	0.1=10%	0.2=20%	0.9=90%
7	7	17	40	0.14=14%	0.34=34%	0.8=80%
8	13	30	33	0.26=26%	0.6=60%	0.66=66%
9	14	44	2	0.28=28%	0.88=88%	0.4=40%
10	6	50	6	0.12=12%	1=100%	0.12=12%

Reprezentarea garfica a datelor statistice.

Graficul unei serii statistice se numeste diagrama structurala.

Cercul de structura sau diagrama circulara este u cerc a carui arie reprezinta efectivul total al populatiei statistice. Valorile variabileise reprezinta prin sectoare de cerc ale caror arii sunt proportionalecu frecventelerelative ale valorilor variabilei.Cu ajutorul regulii de trei simpla sevdetermina masura unghiului la centru corespunzator fiecaei frecvente.

Cu regula de trei simpla se obtine urmatoarea corespondenta intre frecventa relativa f_1 si m , asura unghiului de la centru corespunzator:

f_1	40%	20%	35%	5%
n^*	144*	72*	126*	18*

Pentru desenarea dreptunghiului de structura se considera un reper cartezian in plan.

Cu baza pe axa orizontala se deseneaza un dreptunghi cu inaltimea de 100 de unitati. Se divizeaza dreptunghiul prin linii orizontale obtinand dreptunghiuri cu ariile proportionale cu frecventele f_1 .

Diagrama structurala cu ajutorul batoanelor se obtine astfel:

-se alege un reper cartezian in plan

-pe axa orizontala se reprezinta valorile x_1 ale variabilei statistice

-pe axa verticala se reprezinta frecventele absolute n_1 .

Acest tip de reprezentare grafica foloseste dreptunghiuri cu latimi egale si lungimile proportionale cu frecventele absolute. Daca dreptunghiurile sunt asezate vertical, reprezentarea grafica se numeste **diagrama prin coloane**, iar daca sunt asezate orizontal se numeste **diagrama prin benzi**.

O modalitate de vizualizare a datelor unei serii statistice este poligonul frecventelor care permite reprezentarea grafica sub forma unei curbe. Pentru a construi poligonul frecventelor absolute se unesc printr-o linie poligonala punctele de coordonate (x_1, n_1) .

Se considera o serie statistica cu variabila cantitativa continua si clasele de valori de amplitudini egale: distributia unui grup de tineri dupa inaltimea exprimata in centimetri:

Inaltime	Nr. De tineri	Frecventa absoluta cresc.	Frecventa absoluta descresc.
[155,160)	5	5	63

[160,165)	12	17	58
[165,170)	15	32	46
[170,175)	20	52	31
[175,180)	8	60	11
[180,185)	3	63	3

Modulul unei serii statistice.

Modulul sau dominanta unei serii statistice reprezinta valoarea sau clasa de valori a variabilei care corespunde celui mai mare efectiv si se noteaza cu M_o .

Exemplu: -pentru seria statistica din tabelul 11. $M_o=5$.

-fie distributia dupa vechimea in munca a unui grup socio-profesional.

Sa consideram o serie statistica cu variabila cantitativa de tip continuu. Dupa modul este o clasa de valori , acesta poate fi inlocuita cu valoarea ei centrala.

$M_o = x_k = x_{k=1/2}$, $1 \leq k \leq p-1$.

Pentru a masura gradul de imprastiere a datelor unei serii statistice fata de medie se folosesc urmatoorii parametri de pozitie > dispersia si abaterea medie patratica.

Fiind data seria statistica , **dispersia** valorilor este media aritmetica ponderata a patratelor abaterilor la medie ale valorilor variabilei.

Fiind data seria statistica se numeste **abatere medie** patratica a valorilor variabilei numarul O , unde S la a doua este dispersia seriei.

Raportul dintre abaterea medie patraticca si valoarea medie a unei serii statistice se numeste **coeficient de variatie**.