

L3 Rezumat

3.5. Indicatori de împrăștiere

Se raportează la indicatorii de localizare. Există indicatori de împrăștiere bazați pe:

- indicatori de tendință extremă și anume *amplitudinea*,
- indicatori de tendință intermediară și anume *intercuartila* și
- indicatori de tendință centrală și anume *dispersia*, *abaterea standard*, *coeficientul de variație*.

3.5.1. Amplitudinea

1° Notății: A , ω . 2° Definiție:

Amplitudinea = diferența dintre valoarea maximă și valoarea minimă din serie:

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

- Nu se pretează la calcule algebrice, ceea ce o descalifică în teoria și practica statistică.

3.5.2. Intercuartila

1° Notăție: IQ . 2° Definiție

Intercuartila = diferența dintre cuartila superioară și cea inferioară:

$$IQ = Q_3 - Q_1$$

- Dacă $IQ \leq A / 2$ distribuția este considerată **intens concentrată**; altfel - **intens dispersată**.
- Nu se pretează la calcule algebrice.

Reprezentare sub formă de "boxplot" = un segment având drept extremități *valorile extreme*, peste interiorul segmentului fiind suprapuse centrat orizontal două "parcele" (plot-uri) sub formă de "cutii" (box-uri) alipite, cotele orizontale ale acestora fiind cele trei *cuartile*.

3.5.3. Indicatori de împrăștiere în jurul tendinței centrale reprezentate de medie

$x_i - M$ sn abaterea valorii x_i față de media aritmetică M .

3.5.4. Dispersia

Sinonime: varianța, sigma pătrat, fluctuația.

1° Notății

S^2 (pentru populații în general), σ^2 (pentru populații teoretice), s^2 (pentru eșantioane), *Disp.*

2° Definiție

Dispersia = media aritmetică a pătratelor *abaterilor (valorilor seriei) față de media lor aritmetică.*

3° Mod de calcul

În cazul <i>seriei statistice</i> a	În cazul grupării seriei într-o distribuție de frecvențe	
	absolute	relative
celor N valori distincte sau nu, x_i	a celor p ($\leq N = \sum_{j=1}^p N_j$) valori distincte x_j	

$$S^2 = \frac{\sum_{i=1}^N (x_i - M)^2}{N} \quad S^2 = \frac{\sum_{j=1}^p N_j \cdot (x_j - M)^2}{N} \quad S^2 = \sum_{j=1}^p F_j \cdot (x_j - M)^2$$

Are extrem de importanta proprietate algebrică denumită:

4° Proprietatea de aditivitate a dispersiei

Proprietatea de aditivitate a varianței (a dispersiei) înseamnă, de fapt, **aditivitatea variațiilor** (acestea fiind numărătoarele varianțelor):

Variația totală = *variația intragrupări* + *variația intergrupări*.

5° Alte proprietăți

- $S^2 \geq 0$. $S^2 = 0 \Leftrightarrow$ șirul este constant.
- Se utilizează pentru compararea seriilor de date exprimate în aceeași unitate de măsură și cu valori de același ordin de mărime (și, deci, medii apropiate).
- Are alt ordin de mărime față de datele inițiale și medie și se exprimă în unitatea de măsură a datelor ridicată la pătrat.

3.5.5. Abaterea standard

Sinonime: abaterea pătratică medie, abaterea medie pătratică, deviația standard, σ -ul seriei, abaterea tip, *SD*-ul seriei

1° Notății

S (pentru populații statistice în general), σ (pentru populații statistice teoretice), s (pentru eșantioane).

2° Definiție $S = \sqrt{S^2}$.

3° Proprietăți

- $S \geq 0$. $S = 0 \Leftrightarrow$ șirul este constant (ca și S^2).
- Se utilizează pentru compararea seriilor de date exprimate în aceeași unitate de măsură și cu valori de același ordin de mărime (și, deci, medii apropiate), ca și S^2 .
- Are ACELAȘI ordin de mărime față de datele inițiale și medie și se exprimă în ACEEAȘI unitate de măsură a datelor. (Calitate + defect).

3.5.6. Coeficientul de variație

1° Notății: $CV\%$, CV , $Cv\%$, Cv . 2° Definiție:

Pentru o serie de date pe o *scală raport*, $CV\% = \frac{S \cdot 100}{M} \%$.

3° Proprietăți

- $CV\% \geq 0$.
- $CV\% = 0 \Leftrightarrow$ șirul de date este constant.
- Indicator universal de variabilitate pentru cele mai frecvente scale, scalele raport.

4° Reguli empirice

Pentru biomedicină:

- $Cv\% < 10\% \Leftrightarrow$ populație omogenă;
- $Cv\% > 30\% \Leftrightarrow$ populație eterogenă;
- $Cv\%$ cuprins între $10\% - 20\% \Leftrightarrow$ populație relativ omogenă sau chiar omogenă (în funcție de variabilă).
- $Cv\%$ cuprins între $20\% - 30\% \Leftrightarrow$ populație relativ eterogenă.

3.6. Formule de calcul pentru medie și dispersie

3.6.1. Formule de calcul simultan rapid și precis al mediei și dispersiei

În cazul unei serii statistice

$$(x_i)_{i=1,2,\dots,N}$$

(de valori distincte sau nu)

În cazul grupării seriei într-o distribuție de frecvențe absolute

(de valori distincte x_j):

$$\begin{pmatrix} x_j \\ N_j \end{pmatrix}_{j=1,2,\dots,p} \text{ cu } \sum_{j=1}^p N_j = N$$

Notăm:

$$T_1 = \sum x_i \quad T_2 = \sum x_i^2$$

$$T_1 = \sum N_j \cdot x_j \quad T_2 = \sum N_j \cdot x_j^2$$

1° și 2° Formule de calcul rapid și precis pentru medie și dispersie:

$$M = \frac{T_1}{N} \quad S^2 = \frac{T_2}{N} - M^2.$$

+3.6.2. Metode de calcul manual simplificat prin artificii

1° Translatarea datelor (sau introducerea unei medii provizorii x_0)

$$x' = x - x_0 \Rightarrow M = M' + x_0, \quad S^2 = (S')^2$$

2° Dilatarea ori contractarea datelor (sau introducerea unui factor de scară sau schimbarea unității de măsură)

$$x'' = x' / h = (x - x_0) / h \Rightarrow \begin{cases} M = M' + x_0 = M'' \cdot h + x_0 \\ S^2 = (S')^2 = (S'')^2 \cdot h^2 \end{cases}$$

4 3 Teste, exerciții și probleme

TG3. Durata 90'' pe calculator.

Alegeti afirmatia ERONATA:

1. amplitudinea se preteaza la calcule algebrice
2. amplitudinea ia in considerare numai valorile extreme
3. amplitudinea este sensibila la valorile extreme

$S^2 = \sum F_j \cdot (X_j - M)^2$ reprezinta:

1. formula de calcul a dispersiei pentru date grupate in distributie de frecvente absolute
2. formula teoretica a dispersiei pentru date grupate in distributie de frecvente relative
3. formula de calcul rapid si precis a dispersiei

Dispersia este utilizata pentru:

1. compararea imprastierii a doua sau mai multe caractere cantitative ale aceleiasi populatii daca acestea sunt exprimate in aceeasi unitate de masura si au ordine de marime apropiate
2. compararea imprastierii a doua sau mai multe caractere cantitative ale aceleiasi populatii daca valorile au ordine de marime diferite
3. compararea imprastierii aceluasi caracter cantitativ in doua sau mai multe populatii daca datele au ordine de marime diferite

Coeficientul de variatie este indicatorul universal de comparare a variabilitatii pentru variabile pe scale:

1. interval
2. ordinala si interval
3. ordinala, interval si raport
4. raport

TC3. Durata 160''.

1. Există indicatori de împrăștiere bazați pe indicatori de tendință centrală și anume: _____, _____ și _____.
2. Dacă $IQ \leq A/2$, distribuția este considerată _____.
3. _____ se utilizează pentru compararea variabilității unui același caracter în două sau mai multe populații diferite pentru care datele au același ordin de mărime.

4. _____ are același ordin de mărime cu datele inițiale și cu media, precum și aceeași unitate de măsură.
5. Pentru compararea variabilității a două sau mai multe caractere pe scale raport în aceeași populație dacă acestea sunt exprimate în aceeași unitate de măsură, dar au ordine de mărime diferite, se folosește _____.
6. În științele biologice, o populație eterogenă este caracterizată de un CV% _____.

Exerciții sau probleme rezolvate

EGA 1.

Sa se calculeze:

cuartilele, mediana, amplitudinea, media, dispersia, abaterea standard si coeficientul de variatie pentru urmatoarea serie de 10 valori:

1 11 2 2 2 3 10 12 11 11

Rezultatele finale se vor rotunji la doua zecimale.

1. Ce parametru exprima mai bine tendința centrala pentru acest sir si de ce ?
2. Prin ce parametru se exprima corect si cel mai adecvat imprastierea, daca valorile sirului reprezinta temperaturi in grade Celsius ?
3. Prin ce parametru se exprima corect si cel mai adecvat imprastierea, daca valorile sirului reprezinta temperaturi in grade Kelvin ?
4. Sa se interpreteze variabilitatea sirului in cazul in care valorile sale sunt temperaturi in grade Kelvin.

Rezolvare:

LEGENDA: \$ =.5* puncte. \$\$ din oficiu. Numarul de \$ se imparte la 2.

Rangurile si sirul ordonat ascendent sunt:

1 2 3 4 5 6 7 8 9 10

1 2 2 2 3 10 11 11 11 12

\$ Q 1 lasa la stanga, in seria ordonata crescator, 2.5 termeni(T). Deci este T de rang 3.

Q 1 = 2

\$ Q 2 lasa la stanga, in seria ordonata, 5 termeni(T). Deci este semisuma T de rang 5 si

6. Q 2 = 6.5[†]

\$ Q 3 lasa la stanga, in seria ordonata crescator, 7.5 termeni(T). Deci este T de rang 8. Q 3 = 11

*.5 este forma sub care afișează calculatorul numărul 0,5.

† 6.5 înseamnă 6,5

\$ Mediana $Me=Q_2= 6.5$

\$ Amplitudinea= $x_{max}-x_{min}= 12 - 1 = 11$

Calcul simultan medie, dispersie, apoi abatere standard si Cv

i	x(i)	x(i)^2*
1	1	1
2	2	4
3	2	4
4	2	4
5	3	9
6	10	100
7	11	121
8	11	121
9	11	121
10	12	144
T1= 65		T2= 629

\$ Media $M = T1 / N = 65 / 10 = 6.5$

\$ Dispersia:

$$S^2 = T2 / N - M^2 =$$

$$629 / 10 - 6.5^2 =$$

$$62.9 - 42.25 = 20.65$$

\$ Abaterea standard :

$$S = \text{RAD}^\dagger(S^2) = \text{RAD}(20.65) = 4.544227 \approx 4.54$$

\$ Coeficientul (procentual) de variatie :

$$cv\% = S * 100 / M \% = 4.544227 * 100 / 6.5 \% = 69.91119 \% \approx 69.91 \%$$

RASPUNSURI

\$\$\$ 1.Nici unul, seria fiind bimodala (conform DIAGRAMEI IN BATOANE urmatoare).

\$ Distribuțiile de frecvențe si DIAGRAMA IN BATOANE pusa orizontal sunt:

j	Val. distincta xj	Frecv. abs. Nj	BATON	Frecventa rel. Fj	Frecv. proc.Pj	Frecv. proc. cum. PCj	
1	1	1	□	0.1	10%	10%	
2	2	3	□□□	0.3	30%	40%	Q1 = 2
3	3	1	□	0.1	10%	50%	
4	10	1	□	0.1	10%	60%	Me = (3
5	11	3	□□□	0.3	30%	90%	+ 10) / 2
6	12	1	□	0.1	10%	100%	Q3 = 11

* $x(i)^2$ înseamnă $x(i)^2$

† $\text{RAD}(S^2)$ înseamnă $\sqrt{S^2}$

\$\$ 2. Abaterea standard

\$\$ 3. CV%

\$\$ 4. Sir eterogen deoarece $CV\% > 30\%$

Observație:

Dacă o serie de volum mare este formată din puține valori distincte, este mai convenabil să se calculeze totalurile T_1 și T_2 după gruparea seriei în *distribuția de frecvențe absolute*. Pentru aceasta, vom utiliza un tabel de forma alăturată.

N_j	x_j	x_j^2	$N_j \cdot x_j$	$N_j \cdot x_j^2$
1	1	1	1	1
3	2	4	6	12
1	3	9	3	9
1	10	100	10	100
3	11	121	33	363
1	12	144	12	144

$$T_1 = 65 \quad T_2 = 629$$

2.

Determinați gradul de concentrare / dispersare al distribuției următoare:

x_j	N_j
7	3
8	4
9	5
10	9
11	13
12	10
13	4
14	3
15	2
$N = 53$	

Fc Rezolvare:

3 ← Calculăm coloana de frecvențe cumulate pentru identificarea rangurilor termenilor.
 7
 12 $Q_1 =$ punctul care lasă la stânga sa $N \cdot (1 / 4) = 53 / 4 =$
 → 21 13,25 termeni în seria ordonată ascendent. $Q_1 =$ termenul
 34 de rangul 14 = 10.
 44 → $Q_3 =$ punctul care lasă la stânga sa $N \cdot (3 / 4) = 53 \cdot (3 /$
 48 4) = 39,75 termeni în seria ordonată ascendent. $Q_3 =$
 51 termenul de rangul 40 = 12
 53 $IQ = Q_3 - Q_1 = 12 - 10 = 2$; $A = x_{max} - x_{min} = 15 - 7 = 8$;
 $A / 2 = 4$.
 $IQ < A / 2 \Rightarrow$ distribuție intens concentrată.

3.

Utilizând artificii de calcul adecvate, calculați media și dispersia seriei:

21 23 25 29 33.

Rezolvare:

Translatăm datele cu $x_0 = 25$, valoare care se află aproximativ la "mijlocul" seriei.

x	$x' = x - 25$	x'^2
21	-4	16
23	-2	4
25	0	0
29	4	16
33	8	64
Sume:	$T'_1 = 6$	$T'_2 = 100$

Volumul seriei, $N = 5$.

$$M' = T'_1 / N = 6 / 5 = 1,2 \Rightarrow M = M' + x_0 = 1,2 + 25 = 26,2$$

$$(S')^2 = \frac{T'_2 - \frac{(T'_1)^2}{N}}{N} =$$

$$= \frac{100 - \frac{6^2}{5}}{5} = \frac{100 - 7,2}{5} = \frac{92,8}{5} = 18,56$$

$$\Rightarrow S^2 = (S')^2 = 18,56.$$

4.

Utilizând artificii de calcul adecvate, calculați media și dispersia seriei: 0,1 0,3 0,4 0,6 0,7.

Rezolvare:

Dilatăm datele cu factorul de scară $h = 1 / 10$, pentru a nu mai lucra cu zecimale, adică împărțim cu $1 / 10$, ceea ce revine la înmulțirea cu 10.

x	$x'' = x \cdot 10$	x''^2
0,1	1	1
0,3	3	9
0,4	4	16
0,6	6	36
0,7	7	49
	$T''_1 = 21$	$T''_2 = 111$

Volumul seriei este $N = 5$.

$$M'' = T''_1 / N = 21 / 5 = 4,2 \Rightarrow M = M'' \cdot h = 4,2 \cdot (1 / 10) = 4,2 / 10 = 0,42$$

$$S''^2 = 4,56 \Rightarrow S^2 = S''^2 \cdot h^2 = 4,56 \cdot (1/10)^2 = 4,56 / 100 = 0,0456.$$

5.

Utilizând artificii de calcul adecvate, calculați media și dispersia seriei: 10 18 20 24 30.

Rezolvare:

Translatăm datele cu $x_0 = 20$, valoare centrală a seriei și apoi le comprimăm cu factorul de scară $h = 2$.

x	$x' = x - 20$	$x'' = x'/2$	x''^2	
10	-10	-5	25	Volumul seriei, $N = 5$.
18	-2	-1	1	$M'' = T''_1 / N = 1 / 5 = 0,2 \Rightarrow M =$
20	0	0	0	$= M'' \cdot h + x_0 = 0,2 \cdot 2 + 20 = 20,4$
24	4	2	4	$S''^2 = 10,96 \Rightarrow S^2 = S''^2 \cdot h^2 =$
30	10	5	25	$= 10,96 \cdot 2^2 = 43,84.$
	Sume:	$T''_1 = 1$	$T''_2 = 55$	

Exerciții sau probleme propuse

EGA 6.

Să se calculeze:

cuartilele, mediana, amplitudinea, media, dispersia, abaterea standard și coeficientul de variație pentru următoarea serie de 8 valori:

4 4 3 7 4 5 4 6

Rezultatele finale se vor rotunji la două zecimale.

1. Ce parametru exprimă mai bine tendința centrală pentru acest sir și de ce ?
2. Prin ce parametru se exprimă corect și cel mai adecvat împrăștierea, dacă valorile sirului reprezintă temperaturi în grade Celsius ?
3. Prin ce parametru se exprimă corect și cel mai adecvat împrăștierea, dacă valorile sirului reprezintă temperaturi în grade Kelvin ?
4. Să se interpreteze variabilitatea sirului în cazul în care valorile sale sunt temperaturi în grade Kelvin.

EGA 7.

Același enunț ca în exercițiul anterior, dar pentru seria: 4 3 4 4 4 5 4.

EGA

Să se calculeze, rotunjindu-se la primele două zecimale, cuartilele, mediana, amplitudinea, media, dispersia, abaterea standard și coeficientul de variație pentru următoarele serii de valori:

8. 3 2 6 3 4 3 5.

9. 1 11 2 2 2 3 10 12 11 11.

10. 7 8 2 9.

11. 1 4 3 5 5.

12. 1 9 10 3 5 7 2 13.

Pentru fiecare din distribuțiile următoare determinați gradul de concentrare /
dispersare:

13.

x_j	0,25	0,26	0,27	0,28	0,29	0,30	0,31	0,32	
N_j	1	2	3	5	6	5	3	2	$N=27$

14.

x_j	0,25	0,26	0,27	0,28	0,29	0,30	0,31	0,32	0,33	
N_j	4	4	4	3	3	3	3	3	1	$N=28$
F_c	4	8	12	15	18	21	24	27	28	

Pentru fiecare din următoarele serii calculați cât mai rapid și cât mai exact
media și dispersia:

15.

5 10 15 25 35.

16.

30 32 34 36 38 44.

17.

1 7 10 16 22 25.