Lecția 7. Dispersia unei serii statistice

Cât de împrăștiate sunt datele în jurul mediei

Material didactic. Sursa: mateexamene.ro

Acest material este licențiat sub Creative Commons Atribuire–Necomercial–FărăDerivate 4.0 Internațional (CC BY-NC-ND 4.0): poate fi distribuit gratuit, cu menționarea autorului și a sursei, fără modificări și fără scop comercial.

1. Ce este dispersia?

Media, mediana și modulul ne arată „centrul” unei serii statistice. Ele nu spun însă nimic despre cât de aproape sau cât de departe de acest centru se află datele. Doi elevi pot avea aceeași medie anuală, unul cu note de 8 și 9, celălalt cu note de la 5 la 10. Pentru a descrie această diferență avem nevoie de un indicator al *împrăștierii*: dispersia.

Definiție

Fie seria statistică $x_1, x_2, \ldots, x_n$, cu media $\overline{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$.

- Diferența $x_i - \overline{x}$ se numește **abaterea** valorii x_i de la medie.
- **Dispersia** (sau **varianța**) seriei este media pătratelor abaterilor:

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

- **Abaterea medie pătratică** (abaterea standard) este $\sigma = \sqrt{\sigma^2}$.

Ce ne spune valoarea dispersiei:

- $\sigma^2 = 0$ — toate valorile sunt egale (nu există împrăștiere);
- σ^2 **mică** — valorile sunt grupate strâns în jurul mediei (serie omogenă);
- σ^2 **mare** — valorile sunt răspândite departe de medie (serie eterogenă);
- σ se exprimă în aceeași unitate de măsură ca datele, iar σ^2 în unitatea la pătrat.

2. Dispersia unei serii date prin valori

Se calculează media, apoi abaterile, pătratele lor și, în final, media acestor pătrate. Un tabel cu trei rânduri face calculul ușor de urmărit.

Exemplul 1. Calcul direct

La cinci teste de fizică, un elev a obținut notele: $6,\ 8,\ 9,\ 7,\ 10$. Calculați dispersia și abaterea medie pătratică.

Rezolvare. Media este $\overline{x} = \dfrac{6 + 8 + 9 + 7 + 10}{5} = \dfrac{40}{5} = 8$.

x_i	6	8	9	7	10	Total
$x_i - \overline{x}$	-2	0	1	-1	2	0
$(x_i - \overline{x})^2$	4	0	1	1	4	10

Așadar $\sigma^2 = \dfrac{10}{5} = 2$ și $\sigma = \sqrt{2} \approx 1{,}41$.

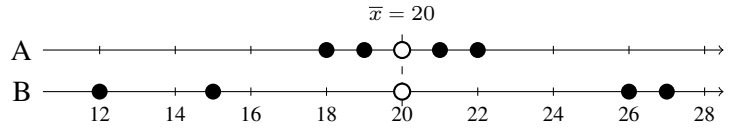
Observați că suma abaterilor este 0 — de aceea abaterile se ridică la pătrat înainte de a face media.

Exemplul 2. Aceeași medie, împrăștiere diferită

Temperaturile maxime (în °C) din cinci zile au fost: în orașul A: $18, 20, 22, 19, 21$; în orașul B: $12, 26, 20, 15, 27$. Comparați cele două serii.

Rezolvare. Ambele serii au suma 100, deci $\overline{x}_A = \overline{x}_B = 20\,^\circ\text{C}$.

- A: abaterile $-2, 0, 2, -1, 1$; $\sigma_A^2 = \dfrac{4+0+4+1+1}{5} = 2$, $\sigma_A \approx 1{,}41\,^\circ\text{C}$;
- B: abaterile $-8, 6, 0, -5, 7$; $\sigma_B^2 = \dfrac{64+36+0+25+49}{5} = 34{,}8$, $\sigma_B \approx 5{,}90\,^\circ\text{C}$.



Interpretare. Mediile sunt egale, dar $\sigma_A^2 < \sigma_B^2$: vremea din orașul A a fost mult mai stabilă.

3. O formulă mai comodă

Formula de calcul a dispersiei

Dezvoltând pătratele din definiție, se obține

$$\sigma^2 = \frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n} - \overline{x}^{\,2},$$

adică **media pătratelor minus pătratul mediei**. Este utilă mai ales când media nu este un număr întreg.

Exemplul 3. Două metode, același rezultat

Calculați dispersia seriei $2,\ 3,\ 5,\ 6,\ 9$.

Rezolvare. $\overline{x} = \dfrac{25}{5} = 5$.

Metoda 1 (definiția): $\sigma^2 = \dfrac{9+4+0+1+16}{5} = \dfrac{30}{5} = 6$.

Metoda 2 (formula de calcul): $\dfrac{4+9+25+36+81}{5} - 5^2 = \dfrac{155}{5} - 25 = 31 - 25 = 6$.

4. Dispersia unei serii date prin tabelul frecvențelor

Dacă valoarea x_i apare de n_i ori ($i = 1, \ldots, k$) și $n = n_1 + n_2 + \cdots + n_k$, fiecare pătrat se înmulțește cu frecvența sa:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i, \qquad \sigma^2 = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \overline{x}^{\,2}.$$

Exemplul 4. Media întreagă

Într-un sondaj, 20 de elevi au declarat câte cărți au citit în vacanța de iarnă:

Cărți citite x_i	0	1	2	3	4
Număr de elevi n_i	2	4	8	4	2

Rezolvare. $\overline{x} = \dfrac{0+4+16+12+8}{20} = \dfrac{40}{20} = 2$, iar

$$\sigma^2 = \frac{2\cdot 4 + 4\cdot 1 + 8\cdot 0 + 4\cdot 1 + 2\cdot 4}{20} = \frac{24}{20} = 1{,}2, \qquad \sigma \approx 1{,}10.$$

Exemplul 5. Media neîntreagă

Numărul de goluri marcate de o echipă în 20 de meciuri:

Goluri x_i	0	1	2	3
Meciuri n_i	5	8	4	3

Rezolvare. $\overline{x} = \dfrac{0\cdot 5 + 1\cdot 8 + 2\cdot 4 + 3\cdot 3}{20} = \dfrac{25}{20} = 1{,}25$. Folosim formula de calcul:

$$\frac{1}{n}\sum n_i x_i^2 = \frac{0+8+16+27}{20} = \frac{51}{20} = 2{,}55,$$

$$\sigma^2 = 2{,}55 - 1{,}25^2 = 2{,}55 - 1{,}5625 = 0{,}9875.$$

Deci $\sigma \approx 0{,}99$ goluri.

Atenție! Greșeală frecventă

- Se împarte la n (numărul total de date), **nu** la k (numărul de valori distincte din tabel).
- Nu uitați să înmulțiți fiecare pătrat cu frecvența lui.
- În formula de calcul se scade $\overline{x}^{\,2}$, nu $\overline{x}$; iar σ se obține abia după extragerea radicalului.

5. Dispersia unei serii cu date grupate pe intervale

Fiecare interval este reprezentat prin **centrul** său (media capetelor), iar calculul se face ca la o serie cu frecvențe.

Exemplul 6. Masa coletelor

Masele (în kg) a 20 de colete expediate într-o zi:

Masa (kg)	$[0,2)$	$[2,4)$	$[4,6)$	$[6,8)$
Număr de colete n_i	2	5	9	4
Centrul x_i	1	3	5	7

Rezolvare. $\overline{x} = \dfrac{2\cdot 1 + 5\cdot 3 + 9\cdot 5 + 4\cdot 7}{20} = \dfrac{90}{20} = 4{,}5$ kg.

$$\frac{1}{n}\sum n_i x_i^2 = \frac{2\cdot 1 + 5\cdot 9 + 9\cdot 25 + 4\cdot 49}{20} = \frac{468}{20} = 23{,}4,$$

$$\sigma^2 = 23{,}4 - 4{,}5^2 = 23{,}4 - 20{,}25 = 3{,}15, \qquad \sigma \approx 1{,}77 \text{ kg}.$$

6. Proprietăți ale dispersiei

Proprietăți

Dacă seria $x_1, \dots, x_n$ are media $\overline{x}$ și dispersia σ^2, iar $a, b \in \mathbb{R}$, atunci seria $ax_1 + b,\ \dots,\ ax_n + b$ are

$$\text{media } a\,\overline{x} + b, \qquad \text{dispersia } a^2\sigma^2, \qquad \text{abaterea medie pătratică } |a|\,\sigma.$$

Adunarea unei constante doar „mută" datele, fără să le împrăștie; înmulțirea cu a înmulțește dispersia cu a^2.

Exemplul 7. Verificarea proprietăților

Seria $2, 4, 6, 8$ are $\overline{x} = 5$ și $\sigma^2 = \dfrac{9+1+1+9}{4} = 5$.

- $12, 14, 16, 18$ (s-a adunat 10): media 15, dispersia tot 5;
- $6, 12, 18, 24$ (s-a înmulțit cu 3): media 15, dispersia $\dfrac{81+9+9+81}{4} = 45 = 3^2 \cdot 5$.

7. Probleme cu parametru

Exemplul 8. Frecvență necunoscută

Valorile $1, 2, 3$ apar cu frecvențele $2,\ k,\ 2$ ($k \in \mathbb{N}^*$). Arătați că media nu depinde de k și determinați k pentru care dispersia este $0{,}5$.

Rezolvare. $\overline{x} = \dfrac{2 + 2k + 6}{4 + k} = \dfrac{2(4+k)}{4+k} = 2$, pentru orice k.

$$\sigma^2 = \frac{2\cdot(1-2)^2 + k\cdot 0 + 2\cdot(3-2)^2}{4+k} = \frac{4}{4+k} = 0{,}5 \iff 4 + k = 8 \iff k = 4.$$

Interpretare. Cu cât valoarea centrală apare mai des, cu atât dispersia scade.

8. Medie și dispersie — interpretare

Dispersia se compară direct doar între serii cu aceeași unitate de măsură și medii apropiate. Altfel se folosește **coeficientul de variație**

$$v = \frac{\sigma}{\overline{x}} \cdot 100\% \quad (\overline{x} > 0),$$

care arată cât de mare este împrăștierea *față de medie*.

Exemplul 9. Comparare relativă

Într-o clasă, înălțimea elevilor are $\overline{x} = 165$ cm și $\sigma = 6{,}6$ cm, iar masa are $\overline{x} = 55$ kg și $\sigma = 4{,}4$ kg. Care caracteristică variază mai mult?

Rezolvare. $v_{\text{înălțime}} = \dfrac{6{,}6}{165} \cdot 100\% = 4\%,\ v_{\text{masă}} = \dfrac{4{,}4}{55} \cdot 100\% = 8\%$. Masa variază relativ mai mult, deși σ ei este mai mic.

De reținut

- Media arată centrul seriei, dispersia arată cât de împrăștiate sunt datele în jurul lui.
- $\sigma^2 = \dfrac{1}{n}\sum (x_i - \overline{x})^2 = \overline{x^2} - \overline{x}^2;\ \sigma = \sqrt{\sigma^2};\ \sigma^2 \geq 0.$
- La tabele de frecvențe se înmulțește cu n_i, iar la intervale se lucrează cu centrele.
- Seria $ax_i + b$ are dispersia $a^2\sigma^2$.

9. Exerciții propuse

1. Calculați media, dispersia și abaterea medie pătratică ale seriei $3,\ 5,\ 7,\ 9,\ 11$.

2. Determinați media și dispersia seriei:

x_i	5	6	7	8
n_i	3	4	2	1

3. Notele a două grupe la un test: grupa I: $7, 7, 8, 9, 9$; grupa a II-a: $5, 6, 8, 10, 11$. Arătați că mediile sunt egale și stabiliți care grupă are rezultatele mai omogene.

4. Seria $4,\ 6,\ x$ are media 6. Aflați x și dispersia seriei.

5. Timpii (în minute) în care 20 de elevi au rezolvat o temă: $[10, 20)$ — 4 elevi, $[20, 30)$ — 10 elevi, $[30, 40)$ — 6 elevi. Calculați media și abaterea medie pătratică.

6. Seria $x_1, \ldots, x_n$ are dispersia 4. Calculați dispersia și abaterea medie pătratică ale seriei $3x_1 + 5,\ \ldots,\ 3x_n + 5$.

7. Calculați coeficientul de variație al seriei $10,\ 12,\ 14,\ 16,\ 18$.

8. Valorile $0, 1, 2$ apar cu frecvențele $3,\ m,\ 3$. Determinați $m \in \mathbb{N}^*$ pentru care dispersia este $0{,}6$.

Răspunsuri

1. $\overline{x} = 7$, $\sigma^2 = 8$, $\sigma = 2\sqrt{2} \approx 2{,}83$. **2.** $\overline{x} = 6{,}1$, $\sigma^2 = 0{,}89$. **3.** $\overline{x} = 8$; $\sigma_I^2 = 0{,}8$, $\sigma_{II}^2 = 5{,}2$ — grupa I este mai omogenă. **4.** $x = 8$, $\sigma^2 = \frac{8}{3}$. **5.** $\overline{x} = 26$ min, $\sigma = 7$ min. **6.** $\sigma^2 = 36$, $\sigma = 6$. **7.** $\overline{x} = 14$, $\sigma = 2\sqrt{2} \approx 2{,}83$, $v \approx 20{,}2\%$. **8.** $m = 4$.

Bibliografie: M. Burtea, G. Burtea, *Matematică. Manual pentru clasa a X-a*, Editura Carminis, Pitești, 2004.