



# **BAZELE STATISTICII**

---

# *Programa analitică*

---

1. Noțiuni introductive
2. Analiza unei serii statistice univariate, folosind metode grafice și numerice (*variabile cantitative*: indicatori ai tendinței centrale, indicatori ai dispersiei, indicatori ai formei și ai concentrării; *variabile calitative*).
3. Analiza unei serii statistice bivariate.



# *Programa analitică*

---

- 4. Probabilități și distribuții teoretice
- 5. Estimarea parametrilor unei populații
- 6. Testarea statistică
- 7. Indici statistici

## 5. Estimarea parametrilor unei populații

---

### 5.1. *Concepte fundamentale*

#### *a) Populație - Eșantion*

- O populație statistică este definită prin precizarea laturii calitative, spațiului și timpului.
  
- Un eșantion reprezintă o sub-populație sau un sub-ansamblu extras din populația de referință după o anumită procedură.

## 5. Estimarea parametrilor unei populații

---

### *b) Sondajul aleator simplu*

- Sondajele aleatoare permit calcularea a priori a probabilității fiecărei unități din populație de a aparține eșantionului.
- Un sondaj aleator simplu presupune ca fiecare unitate din populație să aibă aceeași probabilitate de a fi inclusă în eșantion.
- Aceasta este  $p = \frac{n}{N}$  .

## 5. Estimarea parametrilor unei populații

---

*c). Numărul de eșantioane care se pot extrage*

Depinde de metoda de eșantionare:

- în cazul eșantionării aleatoare nerepetate (fără revenire):

$$K = C_N^n = \frac{N!}{(N - n)!n!}$$

- în cazul eșantionării aleatoare repetate (cu revenire):

$$K = N^n$$

## 5. Estimarea parametrilor unei populații

---

### 5.2 Parametru – Estimator – Estimație

**Parametrul** reprezintă o valoare fixă și necunoscută, numită și valoare reală sau adevărată, a unei populații studiate după o anumită variabilă. ( $\theta$ )

*Exemplu:*  $\mu, \sigma^2, \sigma, \pi$ .

**Estimatorul** este o statistică, adică o variabilă aleatoare care este determinată de totalitatea eșantioanelor posibile de volum  $n$  care se pot extrage din populația de referință.

## 5. Estimarea parametrilor unei populații

---

Estimatorul este definit ca o funcție a variabilelor de selecție. Se notează cu  $\hat{\theta}$

*Exemplu:*  $\hat{\mu}(\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_k)$

- **Estimația** este o valoare realizată dintre valorile posibile ale estimatorului.
- O estimație se obține la nivelul unui eșantion extras, pe baza datelor culese, și este o funcție a valorilor de sondaj.

*Exemplu:*  $\bar{x}, s^2, s, p.$



## 5. Estimarea parametrilor unei populații

---

### *e) Proprietățile estimatorilor*

- De regulă, există o diferență între estimație și parametru, care reprezintă o eroare de estimare.
- Această eroare poate fi caracterizată prin intermediul proprietăților estimatorilor:

#### 1. Nedeplasarea

$$M(\hat{\theta}) = \theta$$

## 5. Estimarea parametrilor unei populații

---

### 2. Convergența:

$$V(\hat{\theta}) \rightarrow 0, \text{ când } n \rightarrow N$$

- *convergența în probabilitate* impune o condiție de volum al eșantionului: dacă acesta este suficient de mare, atunci orice valoare posibilă a estimatorului (orice estimație) converge către parametru.

Această proprietate este o expresie a *legii numerelor mari*.

- *convergența în repartiție (teorema limită centrală)* impune o condiție de volum pentru estimatorul transformat prin

## 5. Estimarea parametrilor unei populații

---

operația de standardizare:

$$\tilde{\hat{\theta}} = \frac{\hat{\theta} - M(\hat{\theta})}{\sqrt{V(\hat{\theta})}}$$

Dacă volumul eșantionului crește peste o anumită limită, atunci variabila aleatoare obținută prin standardizarea estimatorului urmează o lege de repartiție normală standard:

$$\tilde{\hat{\theta}} \longrightarrow Z \sim N(0,1)$$

3. Eficiența:  $V(\hat{\theta}) = \min$ .

## 5. Estimarea parametrilor unei populații

---

### 5.3. *Statistici uzuale în inferența statistică*

#### a) **Media de selecție**

- Estimatorul numit medie de selecție este obținut ca o medie aritmetică a variabilelor aleatoare de selecție  $X_i$ .
- O valoare posibilă a estimatorului este media de sondaj.
- Variabila media de selecție se caracterizează prin legea normală - teorema limită centrală bazată pe legea numerelor mari.

$$\hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

## 5. Estimarea parametrilor unei populații

---

□ Caracteristici ale estimatorului  $\hat{\mu}$  :

- nedeplasat;
- convergent;
- eficient.

### b) Dispersia de selecție

- Este un estimator **deplasat**.
- Ca o corecție la acest estimator, se construiește dispersia de selecție modificată sau corectată. O valoare posibilă a acestui estimator este dispersia de sondaj modificată:

## 5. Estimarea parametrilor unei populații

---

$$s'^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

### c) **Proporția de selecție**

- are aceleași proprietăți cu media de selecție.

$$\hat{\pi} \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

## 5. Estimarea parametrilor unei populații

---

### 5.4 Estimarea punctuală a parametrilor unei populații

#### a) Definire

- presupune calculul unei estimații la nivelul unui eșantion, ca o valoare a unui estimator convenabil ales, care respectă proprietățile de nedeplasare și convergență.

#### b) Estimarea punctuală a mediei unei populații

- presupune calculul unei estimații la nivelul unui eșantion

$$\bar{x} = \frac{\sum x_i}{n}$$

## 5. Estimarea parametrilor unei populații

---

### b) Estimarea punctuală a varianței unei populații

- La nivelul unui eșantion se calculează varianța corectată (nedeplasată) astfel:

$$s'^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

### b) Estimarea punctuală a proporției unei populații

$$p = \frac{n_A}{n}$$



## 5. Estimarea parametrilor unei populații

---

### 5.4 Estimarea prin interval de încredere (IC) a parametrilor unei populații

#### a) Definire

- a estima prin IC un parametru presupune a identifica două variabile aleatoare,  $L_i$  și  $L_s$ , care, pentru o anumită probabilitate  $(1 - \alpha)$ , numită nivel de încredere, respectă condiția:

$$P( L_i \leq \theta \leq L_s ) = ( 1 - \alpha ) \quad , \text{cu} \quad \alpha \in ( 0, 1 )$$

## 5. Estimarea parametrilor unei populații

---

- estimarea prin IC se bazează pe estimatori nedeplasați și convergenți, cărora li se aplică Teorema limită centrală.

### b) Estimarea prin IC a mediei unei populații

- când se cunoaște parametrul  $\sigma$  :

$$\hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow Z \sim N(0,1)$$

## 5. Estimarea parametrilor unei populații

---

$$z = \frac{\hat{\mu} - \mu}{\sigma / \sqrt{n}}$$

$$P\left(-z_{\alpha/2} \leq \frac{\hat{\mu} - \mu}{\sigma / \sqrt{n}} \leq +z_{\alpha/2}\right) = (1 - \alpha)$$

$$P\left(\hat{\mu} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \hat{\mu} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = (1 - \alpha)$$

## 5. Estimarea parametrilor unei populații

---

- la nivelul unui eșantion extras:

$$\left[ \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \quad \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

- când nu se cunoaște parametrul  $\sigma$  :

Variabila Z devine o variabilă student:

$$t = \frac{\hat{\mu} - \mu}{\frac{\hat{\sigma}'}{\sqrt{n}}} \sim t(n-1)$$

## 5. Estimarea parametrilor unei populații

---

$$P( -t_{\alpha/2} \leq t \leq t_{\alpha/2} ) = ( 1 - \alpha )$$

- valoarea  $t_{\alpha/2}$  se citește din tabelul Student pentru:

$$P( t \geq t_{\alpha/2} ) = \alpha / 2$$

$$\left[ \bar{x} - t_{\alpha/2;n-1} \frac{s'}{\sqrt{n}}, \quad \bar{x} + t_{\alpha/2;n-1} \frac{s'}{\sqrt{n}} \right]$$



## 5. Estimarea parametrilor unei populații

---

### Observație:

Precizia estimării crește (mărimea intervalului de încredere este mai mică), atunci când:

- **volumul eșantionului ( $n$ ) crește.**

## 5. Estimarea parametrilor unei populații

---

df	$t_{0.1}$	$t_{0.05}$	$t_{0.025}$
1	3,078	6,314	12,706
2	1,886	2,920	4,303
3	1,638	2,353	3,82



## 5. Estimarea parametrilor unei populații

---

- **varianța eșantionului** este mică (valorile aberante afectează mărimea intervalului de încredere).



## *Exemplu*

---

La nivelul unui eșantion format din 20 de persoane, extras aleator simplu repetat, s-au obținut următoarele rezultate privind vârsta (ani):

$$\bar{x} = 20 \text{ ani}; s' = 2 \text{ ani}.$$

Să se estimeze prin interval de încredere vârsta medie a întregii populații din care a fost extras eșantionul, considerând un risc de 0,05.

## Exemplu

---

- În estimarea prin IC a mediei populației se folosește statistica *t Student*. Din Tabela Student se citește valoarea:

$$t_{0,025;20-1}=2,093.$$

- IC este :

$$\left[ \bar{x} \pm t_{0,025;19} \cdot \frac{s'}{\sqrt{n}} \right] = \left[ 20 \pm 2,093 \cdot \frac{2}{\sqrt{20}} \right] = [19,064; 20,936]$$

- *Interpretare:* Se poate garanta cu o probabilitate de 0,95 că vârsta medie a întregii populații din care a fost extras eșantionul este acoperită de intervalul: 19,064 ~ 19 ani și 20,936 ~ 21 ani.

## 5. Estimarea parametrilor unei populații

---

### c) Estimarea prin IC a proporției unei populații

- când se cunoaște varianța variabilei alternative:

$$\left[ \hat{\pi} - z_{\alpha/2} \frac{\sigma_{\pi}}{\sqrt{n}}, \quad \hat{\pi} + z_{\alpha/2} \frac{\sigma_{\pi}}{\sqrt{n}} \right]$$

- când nu se cunoaște varianța variabilei alternative:

$$\left[ p - t_{\alpha/2; n-1} \frac{\sqrt{p(1-p)}}{\sqrt{n}}, \quad p + t_{\alpha/2; n-1} \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$$

## *Exemplu*

---

- În urma realizării unui sondaj electoral la nivelul unui eșantion format din 1500 persoane, se observă că 840 persoane au votat pentru candidatul A. Să se estimeze prin interval de încredere proporția persoanelor care votează pentru candidatul A la nivelul întregii populații, considerând un risc de 0,05.

## *Rezolvare*

---

- proporția persoanelor care au votat pentru candidatul A, la nivelul eșantionului, este:  $p=840/1500=0,56$ .
- I.C. se calculează astfel:

$$\left[ p \pm t_{\alpha/2; n-1} \cdot \frac{\sqrt{p \cdot (1-p)}}{\sqrt{n}} \right] = \left[ 0,56 \pm 1,96 \cdot \frac{\sqrt{0,56 \cdot (1-0,56)}}{\sqrt{1500}} \right] = [0,53; 0,59]$$

# 5. Estimarea parametrilor unei populații

## 5.5. Estimarea prin IC în SPSS

Descriptives

			Statistic	Std. Error
rata_som_2010	Mean		7.7442	.32041
	95% Confidence	Lower Bound	7.0976	
	Interval for Mean	Upper Bound	8.3908	
	5% Trimmed Mean		7.8339	
	Median		8.0000	
	Variance		4.414	
	Std. Deviation		2.10105	
	Minimum		2.30	
	Maximum		11.80	
	Range		9.50	
	Interquartile Range		2.60	
	Skewness		-.605	.361
	Kurtosis		.418	.709

<i>rata divort</i>	
Mean	1,25
Standard Error	0,06356
Median	1,3
Mode	#N/A
Standard Deviation	0,201
Sample Variance	0,0404
Kurtosis	0,27489
Skewness	-0,90155
Range	0,63
Minimum	0,84
Maximum	1,47
Sum	12,5
Count	10
Confidence Level(95,0%)	0,14378

## 5. Estimarea parametrilor unei populații

---

### 5.6. Estimarea volumului eșantionului ( $n$ )

- În cazul unui sondaj de opinie electoral, se utilizează ca variabilă de bază o variabilă alternativă, pentru care parametrul  $\pi$  este proporția de voturi pentru un candidat.
- În practică, se fixează probabilitatea sau nivelul de încredere cu care dorim să garantăm rezultatul (de regulă, 0,95) și eroarea maxim admisibilă  $\Delta_\pi$  (de exemplu,  $\pm 3\%$ ).
- Având aceste date, se poate estima volumul eșantionului care estimează parametrul  $\pi$ .

## 5. Estimarea parametrilor unei populații

---

- Știind că:

$$\Delta_{\pi} = z_{\alpha/2} \frac{\sigma_{\pi}}{\sqrt{n}}$$

- Se află  $n$ :

$$n = \frac{z_{\alpha/2}^2 \cdot \sigma_{\pi}^2}{\Delta_{\pi}^2}$$

- Parametrul  $\sigma_{\pi}^2$ , care exprimă gradul de omogenitate al populației, nu se cunoaște, însă se poate utiliza valoarea lui maximă, care este egală cu 0,25.



## 5. Estimarea parametrilor unei populații

---

### Exemplu

- Pentru o probabilitate de 0,95 și o eroarea maxim admisibilă de  $\pm 3\%$  să se calculeze volumul eșantionului.
- Ce se întâmplă dacă se utilizează o eroare de  $\pm 2\%$  ?