BAZELE STATISTICII

Programa analitică

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

Programa analitică

4. Probabilităţi şi distribuţii teoretice
5. Estimarea parametrilor unei populaţii
6. Testarea statistică
7. Indici statistici

5. Estimarea parametrilor unei populaţii

5.1. Concepte fundamentale

a) Populaţie - Eşantion

- O populaţie statistică este definită prin precizarea laturii calitative, spaţiului şi timpului.
- Un eşantion reprezintă o sub-populaţie sau un sub-ansamblu extras din populaţia de referinţă după o anumită procedură.

5. Estimarea parametrilor unei populaţii

b) Sondajul aleator simplu

- Sondajele aleatoare permit calcularea a priori a probabilităţii fiecărei unităţi din populaţie de a aparţine eşantionului.
- Un sondaj aleator simplu presupune ca fiecare unitate din populaţie să aibă aceeaşi probabilitate de a fi inclusă în eşantion.
- Aceasta este $p = \frac{n}{N}$.

5. Estimarea parametrilor unei populaţii

c). Numărul de eşantioane care se pot extrage

Depinde de metoda de eşantionare:

- în cazul eşantionării aleatoare nerepetate (fără revenire):

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

- în cazul eşantionării aleatoare repetate (cu revenire):

$$K = N^n$$

5. Estimarea parametrilor unei populaţii

5.2 Parametru – Estimator – Estimaţie

Parametrul reprezintă o valoare fixă şi necunoscută, numită şi valoare reală sau adevărată, a unei populaţii studiate după o anumită variabilă. (θ)

Exemplu: μ, σ^2, σ, π.

Estimatorul este o statistică, adică o variabilă aleatoare care este determinată de totalitatea eşantioanelor posibile de volum *n* care se pot extrage din populaţia de referinţă.

5. Estimarea parametrilor unei populaţii

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

Exemplu: $\hat{\mu}\left(\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots, \bar{x}_k\right)$

- **Estimaţia** este o valoare realizată dintre valorile posibile ale estimatorului.
- O estimaţie se obţine la nivelul unui eşantion extras, pe baza datelor culese, şi este o funcţie a valorilor de sondaj.

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

5. Estimarea parametrilor unei populaţii

e) Proprietăţile estimatorilor

- De regulă, există o diferenţă între estimaţie şi parametru, care reprezintă o eroare de estimare.
- Această eroare poate fi caracterizată prin intermediul proprietăţilor estimatorilor:

1. Nedeplasarea

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

5. Estimarea parametrilor unei populaţii

2. Convergenţa:

$$V(\hat{\theta}) \to 0, \textit{când} \;\; n \to N$$

- *convergenţa în probabilitate* impune o condiţie de volum al eşantionului: dacă acesta este suficient de mare, atunci orice valoare posibilă a estimatorului (orice estimaţie) converge către parametru.

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

- *convergenţa în repartiţie (teorema limită centrală)* impune o condiţie de volum pentru estimatorul transformat prin

5. Estimarea parametrilor unei populaţii

operaţia de standardizare:

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

Dacă volumul eşantionului creşte peste o anumită limită, atunci variabila aleatoare obţinută prin standardizarea estimatorului urmează o lege de repartiţie normală standard:

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

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

5. Estimarea parametrilor unei populaţii

5.3. Statistici uzuale în inferenţa statistică

a) **Media de selecţie**

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

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

5. Estimarea parametrilor unei populaţii

- Caracteristici ale estimatorului $\hat{\mu}$:
- nedeplasat;
- convergent;
- eficient.

b) **Dispersia de selecţie**

- Este un estimator deplasat.
- Ca o corecţie la acest estimator, se construieşte dispersia de selecţie modificată sau corectată. O valoare posibilă a acestui estimator este dispersia de sondaj modificată:

5. Estimarea parametrilor unei populaţii

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

c) **Proporţia de selecţie**

- are aceleaşi proprietăţi cu media de selecţie.

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

5. Estimarea parametrilor unei populaţii

5.4 Estimarea punctuală a parametrilor unei populaţii

a) **Definire**

- presupune calculul unei estimaţii la nivelul unui eşantion, ca o valoare a unui estimator convenabil ales, care respectă proprietăţile de nedeplasare şi convergenţă.

b) **Estimarea punctuală a mediei unei populaţii**

-presupune calculul unei estimaţii la nivelul unui eşantion

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

5. Estimarea parametrilor unei populaţii

b) **Estimarea punctuală a varianţei unei populaţii**

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

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

b) **Estimarea punctuală a proporţiei unei populaţii**

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

5. Estimarea parametrilor unei populaţii

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

a) **Definire**

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

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

5. Estimarea parametrilor unei populaţii

- estimarea prin IC se bazează pe estimatori nedeplasaţi şi convergenţi, cărora li se aplică Teorema limită centrală.

b) **Estimarea prin IC a mediei unei populaţii**

- când se cunoaşte parametrul σ :

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

5. Estimarea parametrilor unei populaţii

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

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

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

5. Estimarea parametrilor unei populaţii

- la nivelul unui eşantion 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 cunoaşte parametrul σ :

Variabila Z devine o variabilă student:

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

5. Estimarea parametrilor unei populaţii

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

- valoarea $t_{\alpha/2}$ se citeşte 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 populaţii

Observaţie:

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

- **volumul eşantionului** (n) creşte.

5. Estimarea parametrilor unei populaţii

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 populaţii

- **varianţa eşantionului** este mică (valorile aberante afectează mărimea intervalului de încredere).

Exemplu

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

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

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

Exemplu

- În estimarea prin IC a mediei populaţiei se foloseşte statistica *t Student*. Din Tabela Student se citeşte 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 populaţii din care a fost extras eşantionul este acoperită de intervalul: 19,064 ~ 19 ani şi 20,936 ~ 21 ani.

5. Estimarea parametrilor unei populaţii

c) **Estimarea prin IC a proporţiei unei populaţii**

- când se cunoaşte varianţa 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 cunoaşte varianţa 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 eşantion format din 1500 persoane, se observă că 840 persoane au votat pentru candidatul A. Să se estimeze prin interval de încredere proporţia persoanelor care votează pentru candidatul A la nivelul întregii populaţii, considerând un risc de 0,05.

Rezolvare

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

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

5. Estimarea parametrilor unei populaţii

5.5. Estimarea prin IC în SPSS

Descriptives

			Statistic	Std. Error
rata_som_2010	Mean		7.7442	.32041
	95% Confidence Interval for Mean	Lower Bound	7.0976	
		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

rata divort	
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 populaţii

5.6. Estimarea volumului eşantionului (*n*)

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

5. Estimarea parametrilor unei populaţii

- Ştiind 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 σ_π^2, care exprimă gradul de omogenitate al populaţiei, nu se cunoaşte, însă se poate utiliza valoarea lui maximă, care este egală cu 0,25.

5. Estimarea parametrilor unei populaţii

Exemplu

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