BAZELE STATISTICII

6. Testarea statistică

6.1. Aspecte generale ale testării statistice

6.1.1. Obiectivele testării statistice

6.1.2. Demersul testării statistice

6.1.3. Teste parametrice versus teste neparametrice

6.2. *Testarea ipotezelor asupra unui eşantion*

6.2.1 Testarea ipotezelor asupra mediei: testul t, testul Z

6.2.2 Testarea ipotezelor asupra proporţiei: testul binomial

6.1. Aspecte generale ale testării statistice

6.1.1. Obiectivele testării statistice

- verificarea ipotezelor asupra unui parametru al unei populaţii;
- verificarea ipotezelor privind legea de distribuţie a unei populaţii;
- verificarea ipotezelor privind două sau mai multe populaţii;
- verificarea existenţei legăturii dintre două variabile.

6.1. Aspecte generale ale testării statistice

6.1.2. Demersul testării statistice

a)***Formularea ipotezelor statistice***

O ipoteză este o presupunere cu privire la valoarea unui parametru, legea de distribuţie a variabilei studiate etc.

Ipoteza nulă H_0 : se presupune egalitatea unui parametru cu o valoare fixă sau se face o presupunere cu privire la legea de repartiţie a unei variabile.

Ipoteza alternativă H_1: este opusul ipotezei nule.

6.1. Aspecte generale ale testării statistice

Test bilateral:

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$

θ - parametru

θ_0 – valoare fixă

Test unilateral la dreapta:

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta > \theta_0$$

Test unilateral la stânga:

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta < \theta_0$$

6.1. Aspecte generale ale testării statistice

b) Alegerea testului statistic

- există două categorii de teste statistice: teste parametrice şi teste neparametrice.

Alegerea testului statistic se face în funcţie de mai multe criterii.

Pentru variabile cantitative se folosesc teste parametrice şi neparametrice, iar pentru variabile calitative se folosesc neparametrice.

6.1. Aspecte generale ale testării statistice

c) **Alegerea pragului de semnificaţie α al testului şi citirea valorii critice din tabelul repartiţiei statisticii test**

- riscul (pragul de semnificaţie) α reprezintă probabilitatea de a respinge ipoteza nulă, atunci când aceasta este adevărată.

d) **Calculul valorii statisticii test,** folosind datele observate la nivelul eşantionului.

e) **Regiunea de respingere/acceptare a ipotezei nule**

6.1. Aspecte generale ale testării statistice

Regiunea de respingere – intervalul dintr-o distribuţie de probabilitate în care se respinge ipoteza nulă, acest interval este acoperit de probabilitatea α

Regiunea de acceptare (interval de încredere) – intervalul în care nu se respinge ipoteza nulă şi este acoperit de probabilitatea 1- α

f) Decizia statistică

Erori de testare

Decizia testului se ia cu o anumită eroare, care poate fi:

- eroare de tip I (eroare de primă speţă, notată α)
- eroare de tip II (eroare de a doua speţă, notată β)

		Realitate	
		H_0 adevărată	H_0 falsă
Decizia	Se acceptă H_0	Decizie corectă $(1-\alpha)$	Eroare de tip II (β)
	Nu se acceptă H_0	Eroare de tip I (α)	Decizie corectă $(1-\beta)$

6.1. Aspecte generale ale testării statistice

- $1 - \alpha$ măsoară nivelul de siguranţă al testului (siguranţa statistică)
- $1 - \beta$ măsoară puterea testului

6.1.3 Teste parametrice şi teste neparametrice

Teste parametrice:

- ✓ presupun o serie de ipoteze restrictive (de ex. ipoteza de normalitate a distribuţiei populaţiei din care a fost extras eşantionul analizat), care nu sunt întotdeauna reale/adevărate;
- ✓ variabila analizată este măsurată pe o scală interval sau raport;
- ✓ mărimea eşantionului trebuie să fie suficient de mare *(ex. $n>30$).*

Exemplu: testul z, testul t, testul F, testul χ^2

6.1.3 Teste parametrice şi teste neparametrice

Teste neparametrice:

- ✓ puţine ipoteze restrictive privind legea de distribuţie a populaţiei din care a fost extras eşantionul analizat (“distribution free methods”);
- ✓ adecvate pentru date calitative;
- ✓ mărimea eşantionului poate fi mică, până la n=6;
- ✓ datele sunt transformate în ranguri sau în semne (pozitive, negative), ceea ce duce la pierderea de informaţii.

Exemplu: runs test

6.2. Testarea ipotezelor asupra unui eşantion

6.2.1**. Testarea ipotezelor asupra mediei unei populaţii**

a) *Formularea ipotezelor* $H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

b) *Alegerea testului statistic*

1. Dacă se cunoaşte σ^2 se foloseşte statistica Z, $Z \sim N(0,1)$

$$Z_{calc} = \frac{\hat{\mu} - \mu_0}{\sigma / \sqrt{n}}$$

2. Dacă nu se cunoaşte σ^2, se foloseşte statistica t, $t \sim t(n-1)$

$$t_{calc} = \frac{\hat{\mu} - \mu_0}{\hat{\sigma}' / \sqrt{n}}$$

6.2.1. Testarea ipotezelor asupra mediei unei populaţii

c. Alegerea pragului de semnificaţie şi citirea din tabel a valorii critice a statisticii test

d. Calculul valorii statisticii test pe baza datelor eşantionului

$$z_{calculat} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \qquad t_{calculat} = \frac{\bar{x} - \mu_0}{s' / \sqrt{n}}$$

6.2.1. Testarea ipotezelor asupra mediei unei populaţii

e. Regula de decizie

$|z_{calculat}| > z_{\alpha/2}$ *sau* $Sig < \alpha$ se respinge ipoteza nulă, pentru un risc α și se acceptă ipoteza alternativă

$|z_{calculat}| \leq z_{\alpha/2}$ *sau* $Sig \geq \alpha$ ipoteza nulă nu se respinge

6.2.1. Testarea ipotezelor asupra mediei unei populaţii

e. Regula de decizie

$\left|t_{calculat}\right| > t_{\alpha/2;n-1}$ *sau* $Sig < \alpha$ se respinge ipoteza nulă, pentru un risc α și se acceptă ipoteza alternativă.

$\left|t_{calculat}\right| \leq t_{\alpha/2;n-1}$ *sau* $Sig \geq \alpha$ ipoteza nulă nu se respinge

f. Compararea valorii calculate a statisticii testului cu valoarea critică (teoretică)

Exemplu

- În urma prelucrării datelor privind veniturile familiilor dintr-o regiune înregistrate la nivelul unui eşantion de volum n=625, s-au obţinut următoarele rezultate:

  $\bar{x} = 12\, mii\, lei$, $s'^2 = 4$. Să se testeze dacă există diferenţe semnificative între veniturile medii ale familiilor la nivelul populaţiei din care a fost extras eşantionul şi venitul mediu pe ţară, $\mu_0 = 13\, mii\, lei$, considerând un risc de 5%.

Exemplu

$\bar{x} = 12\, mii\ lei$ $\alpha = 0,05$

$\mu_0 = 13\, mii\ lei$

n=625, $s'^2 = 4$

1. Formularea ipotezelor:

$$H_0 : \mu = \mu_0 \qquad H_0 : \mu = 13 \text{ mii lei}$$

sau

$$H_1 : \mu \neq \mu_0 \qquad H_1 : \mu \neq 13 \text{ mii lei}$$

2. Alegerea pragului de semnificație

$$\alpha = 0,05$$

Exemplu

3.Alegerea și calcularea statisticii test

$$t_{calc} = \frac{\overline{x} - \mu_0}{s' / \sqrt{n}}$$

$$t_{calc} = \frac{12 - 13}{2 / \sqrt{625}} = \frac{-1}{2 / 25} = \frac{-25}{2} = -12,5$$

4. Regula de decizie

Dacă $\left| t_{calc} \right| > t_{\alpha/2; n-1}$ cu un risc se respinge ipoteza H_0 și se acceptă ipoteza alternativă

Dacă $\left| t_{calc} \right| \leq t_{\alpha/2; n-1}$ nu se respinge ipoteza H_0

Exemplu

5. Decizia statistică

Deoarece $\left(\left|t_{calc}\right|=12,5\right)>\left(t_{0,025;624}=1,96\right)$ cu un risc 0,05 se respinge ipoteza H_0 și se acceptă ipoteza alternativă. Veniturile medii ale familiilor la nivelul populației din care a fost extras eșantionul diferă semnificativ d.p.d.v. statistic de 13 mii lei.

6.2.2 Testarea ipotezelor asupra proporţiei

Demersul testării:

a) *Formularea ipotezelor statistice*

$$H_0 : \pi = \pi_0$$

$$H_1 : \pi \neq \pi_0$$

- De ex. $H_0 : \pi = 0.5$

  $H_1 : \pi \neq 0.5$

(*cele două categorii ale variabilei au şanse egale să apară)*

6.2.2 Testarea ipotezelor asupra proporţiei

b) *Alegerea pragului de semnificaţie* α

c) *Testul statistic*

$$t_{calculat} = \frac{p - \pi_0}{\sqrt{p(1-p)} / \sqrt{n}}$$

d) *Regula de decizie*

$\left| t_{calculat} \right| > t_{\alpha/2;n-1}$ *sau* $Sig < \alpha$ se respinge ipoteza nulă, pentru un risc α și se acceptă ipoteza alternativă.

$\left| t_{calculat} \right| \leq t_{\alpha/2;n-1}$ *sau* $Sig \geq \alpha$ ipoteza nulă nu se respinge