SEMINAR BAZELE STATISTICII

SEMINAR 9

ESTIMAREA PARAMETRILOR UNEI POPULAȚII

Concepte fundamentale

a) Populație – Eşantion

- Inferența statistică: are ca obiectiv cunoașterea indirectă a unei populații prin extragerea și prelucrarea datelor la nivelul unui eșantion.
- O populație statistică este definită prin precizarea naturii sale, a caracteristicilor intrinseci, spațiului şi timpului. ->N
- Un eşantion reprezintă un sub-ansamblu extras din populația de referință după o procedură anume. ->n

b) Sondajul aleator simplu repetat

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

c) Numărul de eşantioane care se pot extrage depinde de metoda de esantionare

- În cazul eşantionării aleatoare repetate: $k = N^n$
- În cazul eşantionării aleatoare nerepetate: $k = C_N^n$

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

Exemplu:

Dintr-o populaţie de adultă de 5000 de persoane, se doreşte extragerea unui eşantion de 50 de persoane.

- In cazul eşantionării aleatoare repetate:

  $k = N^n = 5000^{50} = 8.8818E+184$ *esantioane de 50 de persoane.*

- In cazul eşantionării aleatoare nerepetate:

  $k = C_N^n = C_{5000}^{50}$

  $C_N^n = \frac{N!}{n!(N-n)!} = \frac{5000!}{50!*(5000-50)!} = 2.28\ \mathrm{E} + 120\ eșantioane.$

Exemplu:

As a specific example, one can compute the number of five-card hands possible from a standard fifty-two card deck as:

$$\binom{52}{5} = \frac{52 \times 51 \times 50 \times 49 \times 48}{5 \times 4 \times 3 \times 2 \times 1} = \frac{311{,}875{,}200}{120} = 2{,}598{,}960.$$

Alternatively one may use the formula in terms of factorials and cancel the factors in the numerator against parts of the factors in the denominator, after which only multiplication of the remaining factors is required:

$$\begin{aligned}\binom{52}{5} &= \frac{52!}{5!47!} \\ &= \frac{52 \times 51 \times 50 \times 49 \times 48 \times \cancel{47!}}{5 \times 4 \times 3 \times 2 \times \cancel{1} \times \cancel{47!}} \\ &= \frac{52 \times 51 \times 50 \times 49 \times 48}{5 \times 4 \times 3 \times 2} \\ &= \frac{(26 \times \cancel{2}) \times (17 \times \cancel{3}) \times (10 \times \cancel{5}) \times 49 \times (12 \times \cancel{4})}{\cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2}} \\ &= 26 \times 17 \times 10 \times 49 \times 12 \\ &= 2{,}598{,}960.\end{aligned}$$

Parametru – Estimator – Estimație

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

Parametrii se vor nota cu litere greceşti: μ, σ^2, σ, π.

Estimatorul ($\hat{\theta}$) **este o statistică**, adică o variabilă aleatoare care este determinată de totalitatea eşantioanelor posibile de volum *n* care se pot extrage din populația de referință de volum N.

Estimatorul este definit ca o funcție a variabilelor de selecție.

Estimatorii se notează astfel: $\hat{\mu}, \hat{\sigma}^2, \hat{\sigma}, \hat{\pi}.$

$$\hat{\mu}(\overline{x_1}, \overline{x_2}, \ldots, \overline{x_k})$$

-> media de selectie

EXEMPLU

Estimaţia este o valoare realizată dintre valorile posibile ale estimatorului**.**

O estimaţie se obţine la nivelul unui eşantion extras, pe baza datelor culese.

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

	Estimatii *-valori calculate la nivel de esantion-*	***Estimatori*** *-variabile de selectie care au ca valori estimatii-*	***Parametri*** *-valori estimate la nivel de populatie-*
Media	$\bar{x}$	$\hat{\mu}$	μ
Varianța	s^2, s'^2	$\widehat{\sigma^2}$	σ^2
Abaterea standard	s, s'	$\hat{\sigma}$	σ
Proporția	p	$\hat{\pi}$	π

Date de intrare, cunoscute la nivel de **esantion.**

Estimare

Date de iesire, (necunoscute) estimate la nivel de **populatie.**

Observatie: *s ’^{2} – varianta corectata sau modificata*

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

Proprietățile estimatorilor

De regulă, există o diferență între estimație şi parametru, care reprezintă o eroare de estimare.

Această eroare poate fi măsurată cu ajutorul proprietăților estimatorilor:

1. Nedeplasarea $M(\hat{\theta}) = \theta$

2. Convergența: $V(\hat{\theta}) \to 0, \text{ când } n \to N$

-pentru un volum al eşantionului suficient de mare, orice estimație converge către parametru. (conform teoremei limită centrală, bazată pe legea numerelor mari).

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

Intre doi estimatori nedeplasaţi, un estimator este eficient dacă are dispersia cea mai mică

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.

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

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

- nedeplasat;
- convergent;
- eficient.

b) **Dispersia de selecție**

- Este un estimator deplasat.
- Ca o corecție la acest estimator, se construieşte dispersia de selecție modificată sau corectată:

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

c) **Proporția de selecție**

- are aceleași proprietăți cu media de selecție.

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

Estimarea punctuală a parametrilor unei populații

a) **Definire**

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

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

c) **Estimarea punctuală a varianței unei populații**

d) **Estimarea punctuală a proporției unei populații**

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 şi L_s , care, pentru o anumită probabilitate, numită nivel de încredere, respectă condiția:

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

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

b) Estimarea prin IC a mediei unei populații

- când se cunoaşte parametrul σ:

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

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

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

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

- 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:

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

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

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

v=(*n*-1) grade de libertate ,

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

APLICATII

1. Distribuții de selecție (media de selectie)

Se consideră o populație formată din 4 unități statistice care înregistrează următoarele valori: 2, 4, 6, 8.

Se cere:

a) să se calculeze numărul de eșantioane de volum $n=2$ care se pot extrage din această populație aleatoriu nerepetat.

b) să se formeze eșantioanele și să se calculeze media lor.

c) să se formeze distribuția mediei de selecție (variabilei $\hat{\mu}$).

d) să se calculeze media distribuției formate.

e) să se precizeze caracteristicile variabilei media de selecție.

2. Estimarea mediei unei populații (μ)

Pentru un eșantion format din 4 persoane, se cunosc următoarele rezultate privind vârsta (ani): 19, 21, 22, 26. Se cere:

a) să se estimeze punctual vârsta medie a întregii populații din care a fost extras eșantionul.

b) să se estimeze prin IC vârsta medie a întregii populații din care a fost extras eșantionul, garantând rezultatele cu o probabilitate de 95%.

c) să se estimeze prin IC vârsta medie a întregii populații din care a fost extras eșantionul, garantând rezultatele cu o probabilitate de 90%.

d) să se compare rezultatele obținute.

Tabela Student

n\p	0.10	0.05	0.025	0.01	0.005
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
…					