

**STATISTIKA U ISTRAŽIVANJU OBRAZOVANJA**

**STATISTIČKI OBRASCI KOJE STUDENTI TREBA DA  
ZNAJU – nije ih dozvoljeno nositi na ispit**

## (1) Osnovne oznake

1. Operator sabiranja ili sumacioni operator, u oznaci  $\Sigma$ :

$$\sum_{i=1}^n x_i = x_1 + x_2 + x_3 \cdots x_n$$

2. Operator proizvoda, u oznaci  $\Pi$ :

$$\prod_{i=1}^n x_i = x_1 * x_2 * \cdots * x_n$$

*Napomena: Znak \*, ako nije posebno drugačije naznačeno, u ovom tekstu predstavlja oznaku množenja.*

## (2) Deskriptivna statistika

3. Relativna frekvencija za vrednost  $k$  ili kategoriju  $k$  neke varijable, u oznaci  $p_k$  :

$$p_k = \frac{f_k}{n}$$

U ovom obrascu  $f_k$  je frekvencija ili učestalost za vrednost  $k$  ili kategoriju  $k$  određene varijable, a  $n$  je veličina uzorka.

4. Relativna frekvencija u procentima za vrednost  $k$  ili kategoriju  $k$  varijable, u oznaci  $P_k$

$$P_k = \frac{f_k}{n} * 100$$

5. Aritmetička sredina uzorka, u oznaci  $M$ ,

$$M = \frac{\sum_{i=1}^n x_i}{n}$$

$x_i, i = 1, \dots, n$  predstavlja sirovi, izvorni rezultat jedinice posmatranja  $e_i$ .

6. Medijana

Iz sortiranih  $n$  vrednosti na varijabli, tj.  $n$  sirovih mera uređenih po veličini, mesto na kojem je medijana određuje se po obrascu:

$$\frac{n+1}{2}$$

7. Raspon, u oznaci  $R$ :

$$R = x_{\max} - x_{\min}$$

pri čemu je  $x_{\max}$  najveći, a  $x_{\min}$  najmanji rezultat, tj.

$$x_{\min} = \min_i x_i, \quad i = 1, \dots, n$$

$$x_{\max} = \max_i x_i, \quad i = 1, \dots, n$$

8. Interkvartilni raspon, u oznaci  $IQR$ :

$$IQR = P_{75} - P_{25}$$

pri čemu je  $P_{75}$  percentil 75, a  $P_{25}$  je percentil 25.

9. Varijansa uzorka, u oznaci  $S^2$ :

$$S^2 = \frac{\sum_{i=1}^n (x_i - M)^2}{n-1}$$

10. Standardna devijacija uzorka, u oznaci  $S$ :

$$S = + \sqrt{\frac{\sum_{i=1}^n (x_i - M)^2}{n-1}}$$

(oznaka  $+$  ispred korena znači da se uzima samo pozitivni kvadratni koren iz potkorene veličine)

### (3) Linearne transformacije

11. Ako je  $t = a + b \cdot x$ , tada je:

$$M_t = a + b \cdot M$$

$$S_t = b \cdot S$$

U ovim jednačinama  $a$  je aditivna, a  $b$  multiplikativna konstanta u linearnoj transformaciji.  $M$  i  $S$  predstavljaju aritmetičku sredinu i standardnu devijaciju izvornih rezultata, a  $M_t$  i  $S_t$  označavaju aritmetičku sredinu i standardnu devijaciju rezultata koji su iz izvornih dobijeni linearnom transformacijom.

12. Standardni ili standardizovani skor (z-skor) iz sirovog rezultata, u oznaci  $z_i$ :

$$z_i = \frac{x_i - M}{S}$$

13. Sirovi, izvorni, netransformisani rezultat na osnovu standardnog skora, u oznaci  $x_i$ :

$$x_i = M + z_i \cdot S$$

#### (4) Kovarijansa i linearna korelacija

14. Kovarijansa uzorka, u oznaci  $S_{XY}$ :

$$S_{XY} = \frac{\sum_{i=1}^n (x_i - M_X)(y_i - M_Y)}{n - 1}$$

15. Brave-Pirsonov koeficijent linearne korelacije uzorka, u oznaci  $r_{XY}$ :

$$r_{XY} = \frac{S_{XY}}{S_X S_Y}$$

pri čemu je  $S_{XY}$  kovarijansa uzorka, a  $S_X$  i  $S_Y$  standardne devijacije uzorka na varijablama  $X$  i  $Y$ , tim redom.

## (5) Verovatnoća

16. Osnovno pravilo kombinatorike: ako rezultati eksperimenta obuhvataju  $k$  događaja, od kojih se prvi može desiti na  $n_1$  načina, drugi na  $n_2$  načina i tako redom, onda se svih  $k$  događaja mogu desiti na  $n_1 * n_2 * \dots * n_k$  načina.

17. Broj mogućih permutacija bez ponavljanja od ukupno  $n$  elemenata, u oznaci  $nP_n$ : jednak je  $n$  faktorijelu (u oznaci  $n!$ ):

$$nP_n = n!$$

pri čemu je  $n!$  ( $n$  faktorijel) proizvod celih brojeva od  $n$  do 1.

18. Broj mogućih kombinacija bez ponavljanja sa  $r$  elemenata od ukupno  $n$  elemenata, u oznaci  ${}_nC_r$ :

$${}_nC_r = \frac{n!}{r!(n-r)!} = \frac{n(n-1)(n-2)\dots(n-r+1)}{r!}$$

19. Verovatnoća događaja  $A$ , u oznaci  $P(A)$ :

$$P(A) = \frac{f_A}{n}$$

pri čemu je  $f_A$  broj ishoda obuhvaćenih događajem  $A$ , a  $n$  ukupan broj ishoda.

20. Šanse događaja  $A$ , u oznaci  $O(A)$ :

$$O(A) = \frac{P(A)}{1 - P(A)}$$

pri čemu je  $P(A)$  verovatnoća događaja  $A$ .

21. Uslovna verovatnoća događaja  $A$ , pod uslovom da se desio događaj  $B$ , u oznaci  $P(A|B)$ :

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

22. Statistička nezavisnost događaja  $A$  i  $B$ :

Događaji  $A$  i  $B$  su statistički nezavisni, ako je

$$P(A|B) = P(A)$$

tj.

$$P(A \cap B) = P(A) * P(B)$$

## (6) Ocenjivanje parametara

23. Standardna greška za aritmetičku sredinu, u oznaci  $\sigma_M$ :

$$\sigma_M = \frac{\sigma}{\sqrt{n}}$$

pri čemu je  $\sigma$  standardna devijacija populacije a  $n$  veličina uzorka.

24. Ocena standardne greške za aritmetičku sredinu, u oznaci  $SE_M$ :

$$SE_M = \frac{S}{\sqrt{n}}$$

pri čemu je  $S$  standardna devijacija uzorka a  $n$  veličina uzorka.

25. Interval poverenja za aritmetičku sredinu, u oznaci  $100(1 - \alpha)\%$  CI (engl. Confidence Interval):

$$100(1 - \alpha)\% \text{ CI: } M - t_{1-\alpha/2} * SE_M \leq \mu \leq M + t_{1-\alpha/2} * SE_M$$

pri čemu je  $\mu$  parametar, tj. aritmetička sredina populacije,  $M$  aritmetička sredina uzorka,  $SE_M$  ocena standardne greške za aritmetičku sredinu,  $t_{1-\alpha/2}$  vrednost kvantila  $1 - \alpha/2$  iz Studentove raspodele sa  $n - 1$  stepeni slobode, a  $\alpha$  odabrani nivo rizika greške tipa I. Ako je odabrani nivo rizika greške tipa I jednak 0.05 onda se pravi 95% interval poverenja, ako je odabrani nivo rizika greške tipa I jednak 0.01 pravi se 99% interval poverenja i slično.

## (7) t-test

26. t-statistik za jedan uzorak:

$$t = \frac{M - \mu_0}{SE_M}; \quad df = n - 1$$

pri čemu je  $M$  aritmetička sredina uzorka,  $\mu_0$  pretpostavljena vrednost aritmetičke sredine populacije,  $SE_M$  ocena standardne greške za aritmetičku sredinu, a  $df$  stepeni slobode.

27. t-statistik za dva nezavisna uzorka:

$$t = \frac{M_1 - M_2}{SE_{M_1 - M_2}}; \quad df = n_1 + n_2 - 2$$

pri čemu su  $M_1$  i  $M_2$  aritmetičke sredine uzoraka,  $SE_{M_1 - M_2}$  ocena standardne greške za razliku između aritmetičkih sredina, a  $df$  stepeni slobode.

28. t- statistik za dva zavisna uzorka (metoda diferencija):

$$t = \frac{M_D}{SE_D}; \quad df = n - 1$$

pri čemu je  $M_D$  aritmetička sredina diferencija (razlika između sparenih rezultata),  $SE_D$  standardna greška za aritmetičku sredinu diferencija, a  $n$  broj diferencija, tj. broj parova rezultata.

### (8) Hi-kvadrat test

Test distribucije verovatnoća jedne kategoričke varijable

29. Pirsonov hi-kvadrat statistik

$$\chi^2 = \sum_{k=1}^r \frac{(f_k - \phi_k)^2}{\phi_k}; \quad df = r - 1$$

pri čemu  $f_k$  označava empirijske, a  $\phi_k$  očekivane, teorijske frekvencije.

Test nezavisnosti dveju kategoričkih varijabli

30. Pirsonov hi-kvadrat statistik

$$\chi^2 = \sum_{k=1}^c \sum_{j=1}^r \frac{(f_{jk} - \phi_{jk})^2}{\phi_{jk}}; \quad df = (r - 1)(c - 1)$$