

Statistika u bibliotekama

Oznake i obrasci za pismeni ispit

N — veličina uzorka

f — (apsolutne) frekvencije

f_r — relativne frekvencije

$f_{\%}$ — (relativne) frekvencije u procentima

f_k — kumulativne (apsolutne) frekvencije

$f_{k,\%}$ — kumulativne (relativne) frekvencije u procentima

odnos apsolutnih frekvencija i veličine uzorka — $\sum f = N$

i — dužina intervalnog razreda

D i G — donja i gornja granica intervalnog razreda

D' i G' — donja i gornja egzaktna granica intervalnog razreda

X' ili SMI — srednje mesto intervalnog razreda, $X' = \frac{D+G}{2}$

M — aritmetička sredina (srednja vrednost)

M — **iz sirovih mera** $M = \frac{\sum X}{N}$

M — **iz jedinične raspodele** ($i = 1$) $M = \frac{\sum f \cdot X}{N}$

M — **iz nejedinične raspodele** ($i > 1$, $X' = SMI$) $M = \frac{\sum f \cdot X'}{N}$

M — **iz nejedinične raspodele (kodiranje)** $M = PAS + i \cdot \frac{\sum f \cdot d}{N}$

$$\sum x = \sum (X - M) = 0$$

σ — standardna devijacija

σ — **iz sirovih mera, za velike uzorke** $\sigma = \sqrt{\frac{\sum (X-M)^2}{N}}$, $\sigma = \frac{\sqrt{N \cdot (\sum X^2) - (\sum X)^2}}{N}$
($N \geq 100$)

σ — **iz jedinične raspodele** ($i = 1$), **za velike uzorke** $\sigma = \frac{\sqrt{N \cdot (\sum f \cdot X^2) - (\sum f \cdot X)^2}}{N}$
($N \geq 100$)

σ — **iz nejedinične raspodele** ($i > 1, X' = SMI$), **za velike uzorke** $\sigma = \frac{\sqrt{N \cdot (\sum f \cdot X'^2) - (\sum f \cdot X')^2}}{N}$ ($N \geq 100$)

σ — **iz nejedinične raspodele (kodiranje), za velike uzorke**

$$\sigma = i \cdot \frac{\sqrt{N \cdot (\sum f \cdot d^2) - (\sum f \cdot d)^2}}{N}, \quad (N \geq 100)$$

$PR_{(X)}$ — percentilni rang za sirovu meru X u intervalnom razredu (D', G') .

$$PR_{(X)} = \frac{100}{N} \cdot (F + f \cdot \frac{X - D'}{i})$$

pri čemu je f apsolutna frekvencija koja odgovara intervalnom razredu (D', G') , a F kumulativna apsolutna frekvencija koja odgovara D' (tj. prethodnom intervalnom razredu).

Ako su f_r i F_r odgovarajuće relativne frekvencije, obrazac glasi

$$PR_{(X)} = 100 \cdot (F_r + f_r \cdot \frac{X - D'}{i})$$

P_p — percentil kome odgovara percentilni rang p . Specijalno, P_{50} je medijana. Oznake u obrascima imaju isto značenje kao u obrascima za računanje percentilnog ranga.

$$P_p = D' + i \cdot \frac{p\% \cdot N - F}{f}$$

$$P_p = D' + i \cdot \frac{p\% - F_r}{f_r}$$

svođenje sirovih skorova na standardne skorove :

$$z = \frac{X - M}{\sigma}$$

df — broj stepena slobode

r — koeficijent linearne korelacije (Pirsonov koeficijent)

$$r = \frac{N \cdot \sum(X \cdot Y) - (\sum X) \cdot (\sum Y)}{\sqrt{N \cdot (\sum X^2) - (\sum X)^2} \cdot \sqrt{N \cdot (\sum Y^2) - (\sum Y)^2}} \quad (df = N - 2)$$

Statistička značajnost koeficijenta korelacije se proverava u tablici 4.

ρ — koeficijent korelacije Ro, koeficijent razlike rangova (Spirmanov koeficijent):

$$\rho = 1 - \frac{6 \cdot (\sum D^2)}{N(N^2 - 1)}$$

gde su D razlike odgovarajućih rangova.

Φ — koeficijent korelacije Fi:

$$\Phi = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}, \quad N = a + b + c + d$$

C — koeficijent kontingencije:

$$C = \pm \sqrt{\frac{S - N}{S}}, \quad S = \sum \frac{f_o^2}{f_e}, \quad f_e = \frac{(\sum \text{reda}) \cdot (\sum \text{kolone})}{N}$$

gde su f_o dobijene (empirijske, uočene ili opažene, eng. *observed*) frekvencije, a f_e teorijske ili očekivane (eng. *expected*) frekvencije.

χ^2 — Hi kvadrat. Ako je R broj redova sa frekvencijama (opaženim i očekivanim), a K broj kolona koje odgovaraju kategorijama čija se frekvencija meri, onda je

$$df = (R - 1)(K - 1).$$

χ^2 — Hi kvadrat za test značajnosti koeficijenta korelacije Φ :

$$\chi^2 = N \cdot \Phi^2$$

χ^2 — Hi kvadrat za test značajnosti koeficijenta kontingencije C :

$$\chi^2 = \frac{C^2 \cdot N}{1 - C^2}$$

χ^2 — Hi kvadrat za test značajnosti razlike između dobijenih (empirijskih, uočenih, opaženih) frekvencija f_o i teorijskih (očekivanih) frekvencija f_e :

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

σ — iz sirovih mera, za male uzorke

$$\sigma = \sqrt{\frac{\sum (X - M)^2}{N - 1}}, \sigma = \sqrt{\frac{N \cdot (\sum X^2) - (\sum X)^2}{N(N - 1)}} \quad (N < 100)$$

σ — iz jedinične raspodele ($i = 1$), za male uzorke

$$\sigma = \frac{\sqrt{N \cdot (\sum f \cdot X^2) - (\sum f \cdot X)^2}}{N(N - 1)} \quad (N < 100)$$

σ — iz nejedinične raspodele ($i > 1, X' = SMI$), za male uzorke

$$\sigma = \frac{\sqrt{N \cdot (\sum f \cdot X'^2) - (\sum f \cdot X')^2}}{N(N - 1)} \quad (N < 100)$$

σ — iz nejedinične raspodele (kodiranje), za male uzorke

$$\sigma = i \cdot \frac{\sqrt{N \cdot (\sum f \cdot d^2) - (\sum f \cdot d)^2}}{N(N - 1)}, \quad (N < 100)$$

$\sigma_M = \sigma_{SE}$ — standardna greška aritmetičke sredine

$\sigma_M = \sigma_{SE}$ **za velike uzorke** $\sigma_M = \sigma_{SE} = \frac{\sigma}{\sqrt{N}}$ (obrazac važi i za male uzorke pod uslovom da je standardna devijacija σ izračunata po obrascu za male uzorke)

$\sigma_M = \sigma_{SE}$ **za male uzorke** $\sigma_M = \sigma_{SE} = \frac{\sigma}{\sqrt{N-1}}$. Obrazac se primenjuje pod uslovom da je standardna devijacija σ izračunata po obrascu za velike uzorke. Ako je σ izračunata po obrascu za male uzorke, onda se koristi obrazac za velike uzorke, tj. $\sigma_M = \sigma_{SE} = \frac{\sigma}{\sqrt{N}}$.

i.p. (M) — intervali poverenja (pouzdanosti) aritmetičke sredine za velike uzorke ($N \geq 100$):

$$95\% \text{ i.p. } M \pm 1,96 \cdot \sigma_M$$

$$99\% \text{ i.p. } M \pm 2,58 \cdot \sigma_M$$

i.p. (M) — intervali poverenja (pouzdanosti) aritmetičke sredine za male uzorke ($N < 100$):

$$95\% \text{ i.p. } M \pm t_{0,05} \cdot \sigma_M$$

$$99\% \text{ i.p. } M \pm t_{0,01} \cdot \sigma_M$$

t -vrednosti iz tablice 5 se određuju za broj stepeni slobode $df = N - 1$.

t -test za značajnost razlike između aritmetičkih sredina za nezavisne uzorke

$$t = \frac{D_M}{\sigma_{D_M}}; D_M = |M_1 - M_2|; \sigma_{D_M} = \sqrt{\sigma_{M_1}^2 + \sigma_{M_2}^2}$$

Broj stepeni slobode $df = N_1 - 1 + N_2 - 1$. Kao granične vrednosti za velike nezavisne uzorke ($df \geq 100$) se koriste 1,96 (95%), odnosno 2,58 (99%). Za male nezavisne uzorke ($df < 100$) se koriste vrednosti iz tablice 5: $t_{0,05}$ (95%), odnosno $t_{0,01}$ (99%).

σ_P — standardna greška procenta $P\%$

$$\sigma_P = \sqrt{\frac{PQ}{N}}; Q = 100 - P$$

Primenjuje se ako važi $N \geq 50$, $10\% \leq P\% \leq 90\%$ i ako su procenti normalno raspodeljeni.

t -test za značajnost razlike između procenata za velike nezavisne uzorke

$$t = \frac{D_P}{\sigma_{D_P}}; D_P = |P_1 - P_2|; \sigma_{D_P} = \sqrt{\frac{P_1 Q_1}{N_1} + \frac{P_2 Q_2}{N_2}}$$

Kao granične vrednosti se koriste 1,96 (95%), odnosno 2,58 (99%).

H_0 — nulta hipoteza testa značajnosti:

koeficijenta korelacije (r, Φ, C) — „nema korelacije (koeficijent korelacije nije statistički značajan)”

razlike između aritmetičkih sredina dva uzorka — „razlika između aritmetičkih sredina je slučajno nastala (razlika aritmetičkih sredina nije statistički značajna)”

razlike između procenata — „razlika između procenata je slučajno nastala (razlika procenata nije statistički značajna)”

razlike između frekvencija (χ^2) — „razlika između dobijenih (empirijskih, opaženih) i teorijskih (očekivanih) frekvencija je slučajno nastala (razlika frekvencija nije statistički značajna), tj. raspodela dobijenih i teorijskih frekvencija je ista”