

Статистика у библиотекама (10. VIII 2025)

Ознаке и обрасци за писмени испит

N — величина узорка

f — (апсолутне) фреквенције

f_r — релативне фреквенције

$f_{\%}$ — (релативне) фреквенције у процентима

f_k — кумулативне (апсолутне) фреквенције

$f_{k,\%}$ — кумулативне (релативне) фреквенције у процентима

однос апсолутних фреквенција и величине узорка — $\sum f = N$

i — дужина интервалног разреда

D и G — доња и горња граница интервалног разреда

D' и G' — доња и горња егзактна граница интервалног разреда

X' или SMI — средње место интервалног разреда, $X' = \frac{D+G}{2}$

M — аритметичка средина (средња вредност)

M — из сирових мера $M = \frac{\sum X}{N}$

M — из јединичне расподеле ($i = 1$) $M = \frac{\sum f \cdot X}{N}$

M — из нејединичне расподеле ($i > 1$, $X' = SMI$) $M = \frac{\sum f \cdot X'}{N}$

M — из нејединичне расподеле (кодирање) $M = PAS + i \cdot \frac{\sum f \cdot d}{N}$

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

σ — стандардна девијација

σ — из сирових мера, за велике узорке $\sigma = \sqrt{\frac{\sum (X-M)^2}{N}}$, $\sigma = \frac{\sqrt{N \cdot (\sum X^2) - (\sum X)^2}}{N}$
($N \geq 100$)

σ — из јединичне расподеле ($i = 1$), за велике узорке $\sigma = \frac{\sqrt{N \cdot (\sum f \cdot X^2) - (\sum f \cdot X)^2}}{N}$
($N \geq 100$)

σ — из нејединичне расподеле ($i > 1, X' = SMI$), за велике узорке

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

σ — из нејединичне расподеле (кодирање), за велике узорке

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

$PR_{(X)}$ — перцентилни ранг мере X из интервалног разреда (D', G') .

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

при чему је f апсолутна фреквенција која одговара интервалном разреду (D', G') , а F кумулативна апсолутна фреквенција која одговара D' (тј. претходном интервалном разреду).

Ако су f_r и F_r одговарајуће релативне фреквенције, образац гласи

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

P_p — перцентил коме одговара перцентилни ранг p . Ознаке имају исто значење као у обрасцима за рачунање перцентилног ранга.

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

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

свођење сирових скорова на стандардне скорове: $z = \frac{X-M}{\sigma}$

проценти $1\% = \frac{1}{100}$

$$p = P\% = \frac{P}{100} \Leftrightarrow P = 100 \cdot p = 100 \cdot P\%$$

(на пример, $0,64 = 64\% = \frac{64}{100} \Leftrightarrow 64 = 100 \cdot 0,64 = 100 \cdot 64\%$)

df — број степена слободе

r — коефицијент линеарне корелације (Пирсонов коефицијент)

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

Статистичка значајност коефицијента r се проверава у табели 4.

ρ — (Спирманов) коефицијент ранг-корелације Ро:

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

где су D разлике одговарајућих рангова.

Φ — коефицијент корелације Фи:

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

C — коефицијент контингенције Це:

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

где су f_o добијене (емпиријске, уочене или опажене, енгл. *observed*) фреквенције, а f_e теоријске или очекиване (енгл. *expected*) фреквенције.

χ^2 — Хи квадрат. Ако је R број редова са фреквенцијама (опаженим и очекиваним), а K број колона које одговарају категоријама чија се фреквенција мери, онда је

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

χ^2 — Хи квадрат за тест значајности коефицијента корелације Φ :

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

χ^2 — Хи квадрат за тест значајности коефицијента контингенције C :

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

χ^2 — Хи квадрат за тест значајности разлике између добијених (емпиријских, уочених, опажених) фреквенција f_o и теоријских (очекиваних) фреквенција f_e :

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

σ — из сирових мера, за мале узорке

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

σ — из јединичне расподеле ($i = 1$), за мале узорке

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

σ — из нејединичне расподеле ($i > 1, X' = SMI$), за мале узорке

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

σ — из нејединичне расподеле (кодирање), за мале узорке

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

$\sigma_M = \sigma_{SE}$ — стандардна грешка аритметичке средине

$\sigma_M = \sigma_{SE}$ за велике узорке $\sigma_M = \sigma_{SE} = \frac{\sigma}{\sqrt{N}}$ (образац важи и за мале узорке под условом да је стандардна девијација σ израчуната по обрасцу за мале узорке)

$\sigma_M = \sigma_{SE}$ за мале узорке $\sigma_M = \sigma_{SE} = \frac{\sigma}{\sqrt{N-1}}$. Образац се примењује под условом да је стандардна девијација σ израчуната по обрасцу за велике узорке. Ако је σ израчуната по обрасцу за мале узорке, онда се користи образац за велике узорке, тј. $\sigma_M = \sigma_{SE} = \frac{\sigma}{\sqrt{N}}$.

σ_P — стандардна грешка процента $p = P\% \Leftrightarrow P = 100 \cdot p$

$$\sigma_P = \sqrt{\frac{pq}{N}}, \quad q = 100\% - p = 1 - p$$

$$100 \cdot \sigma_P = \sqrt{\frac{PQ}{N}}, \quad Q = 100 - P$$

$$\sigma_P = \left(\sqrt{\frac{PQ}{N}} \right) \%, \quad Q = 100 - P$$

Примењује се ако важи $N \geq 50$, $10\% \leq P\% \leq 90\%$ и ако се проценти добијени из свих узорака нормално расподељују око свог параметра.

i.p. (M) — интервали поверења (поузданости)

аритметичка средина	($N \geq 100$)	($N < 100$)
95% i.p.	$M \pm 1,96 \cdot \sigma_M$	$M \pm t_{0,05} \cdot \sigma_M$
99% i.p.	$M \pm 2,58 \cdot \sigma_M$	$M \pm t_{0,01} \cdot \sigma_M$
разлика аритметичких средина	($N \geq 100$)	($N < 100$)
95% i.p.	$D_M \pm 1,96 \cdot \sigma_{D_M}$	$D_M \pm t_{0,05} \cdot \sigma_{D_M}$
99% i.p.	$D_M \pm 2,58 \cdot \sigma_{D_M}$	$D_M \pm t_{0,01} \cdot \sigma_{D_M}$
($N \geq 50$)	проценат	разлика процената
95% i.p.	$P\% \pm 1,96 \cdot \sigma_P$	$D_P\% \pm 1,96 \cdot \sigma_{D_P}$
99% i.p.	$P\% \pm 2,58 \cdot \sigma_P$	$D_P\% \pm 2,58 \cdot \sigma_{D_P}$

t -вредности из таблице 5 се одређују за $df = N - 1$.

t -тест значајности разлике аритметичких средина два независна узорка ($N_1, N_2 \geq 30$)

$$t = \frac{D_M}{\sigma_{D_M}}$$

$$D_M = |M_1 - M_2|$$

$$\sigma_{D_M} = \sqrt{\sigma_{M_1}^2 + \sigma_{M_2}^2}$$

Број степени слободe $df = (N_1 - 1) + (N_2 - 1) = N_1 + N_2 - 2$.

Граничне вредности су 1,96 (95%), односно 2,58 (99%), за велике независне узорке ($df \geq 100, N_1, N_2 \geq 30$).

Ако су независни узорци мањи ($df < 100, N_1, N_2 \geq 30$), користе се вредности из таблице 5: $t_{0,05}$ (95%), односно $t_{0,01}$ (99%).

t -тест значајности разлике процената два независна узорка

$$t = \frac{D_P}{\sigma_{D_P}}$$

$$100 \cdot D_P = |P_1 - P_2| \Leftrightarrow D_P = |P_1\% - P_2\%| = |P_1 - P_2|\%$$

$$\sigma_{D_P} = \sqrt{\sigma_{P_1}^2 + \sigma_{P_2}^2} = \left(\sqrt{\frac{P_1 Q_1}{N_1} + \frac{P_2 Q_2}{N_2}} \right) \% = \sqrt{\frac{P_1\% \cdot Q_1\%}{N_1} + \frac{P_2\% \cdot Q_2\%}{N_2}}$$

$P_1\%$ и $P_2\%$ су проценти у узорцима ($10\% \leq P_1\%, P_2\% \leq 90\%$, $N_1, N_2 \geq 50$). Као граничне вредности се користе 1,96 (95%), односно 2,58 (99%).

H_0 — нулта хипотеза теста значајности:

коефицијента корелације (r, Φ, C) — „Нема корелације
(коефицијент корелације није статистички значајан).”

разлике између аритметичких средина два узорка —
„Разлика између аритметичких средина је случајно настала
(разлика аритметичких средина није статистички значајна).”

разлике између процената — „Разлика између процената је
случајно настала (разлика процената није статистички значајна).”

разлике између фреквенција (χ^2) —
„Разлика између добијених (емпиријских, опажених) и теоријских (очекиваних) фреквенција је случајно настала (разлика фреквенција није статистички значајна), тј. расподела добијених и теоријских фреквенција је иста.”