# VAJE 5: Parametrični in neparametrični preizkusi značilnosti

Na računalniških vajah se za urejanje in prikazovanje statističnih podatkov uporabi statistični programski paket SPSS in podatkovne datoteke *parametricni.sav*, *neparametricni.sav*, *vzorec_1.sav*, *vzorec_2.sav* in *vzorec_3.sav*.

NALOGE:

1. Za statistično spremenljivko *KolicinaTD* opravi naslednje preizkuse značilnosti, vse na stopnji značilnosti $\alpha = 0,05$.

   (a) Testiraj ničelno hipotezo $H_0$, ki pravi, da je povprečna količina popite tekočine na populaciji 3100 $ml$, proti alternativi $H_1$, da ni 3100 $ml$. Ali lahko zavrneš hipotezo $H_0$ tudi na stopnji tveganja $\alpha = 0,01$?
   Pomoč: velik vzorec, $H_0(E(X) = \mu)$, $Z = \frac{X-\mu_0}{S}\sqrt{n}$.
   Rezultat preveri s programom SPSS. Uporabi postopek *Analyze - Compare Means - One-Sample T Test* vstavi *KolicinaTD* v *Test Variable(s)* in 3100 v *Test Value*.
   **Opomba:** program uporablja testno statistiko $T = \frac{X-\mu_0}{S}\sqrt{n}$ za male vzorce!

   (b) Predpostavimo, da je količina popite tekočine X pri osebah, ki se ukvarjajo s športom na populaciji porazdeljena normalno $N(\mu, \sigma)$ in da je tudi količina popite tekočine Y pri osebah, ki se ne ukvarjajo s športom porazdeljena normalno $N(\nu, \tau)$. Testiraj enakost varianc.
   Pomoč: test $H_0(\sigma = \tau) : H_1(\sigma \neq \tau)$, testna statistika $F = \frac{S_X^2}{S_Y^2}$.
   Rezultat lahko preveriš s programom SPSS. Uporabi postopek *Analyze - Compare Means - One-Way ANOVA* vstavi *KolicinaTD* v *Dependet List* in *Sport* v *Factor* in v Options odkljukaj *Homogeneity of variance test.*
   **Opomba:** program uporablja za test homogenosti varianc Leveneovo testno statistiko, ki je primerljiva s klasičnim $F$ testom!

   (c) Glede na predpostavko o enakosti varianc testiraj hipotezo, da sta povprečni količini popite tekočine pri osebah, ki se oz. se ne ukvarjajo s športom, na populaciji enaki.
   Pomoč: velik vzorec, $H_0(\mu = \nu) : H_1(\mu \neq \nu)$, $Z = \frac{\overline{X}-\overline{Y}}{S}$ test.
   Rezultat preveri s programom SPSS. Uporabi postopek *Analyze - Compare Means - Independent - Samples T Test* vstavi *KolicinaTD* v *Variable(s)* in *Sport* v *Grouping Variable* ter v *Define Groups* vstavi vrednosti

0 in 1.
Opomba: program uporablja testno statistiko $T = \frac{\overline{X}-\overline{Y}}{S}\sqrt{\frac{nm}{n+m}}$ za male vzorce! Hkrati program izračuna test homogenosti varianc z Leveneovo testno statistiko. Na osnovi testa o enakosti varianc potem razberemo ustrezne podatke!

(d) Predpostavimo, da je $p$ delež oseb, ki se ukvarjajo s športom pri beli rasi in da je $q$ delež pri drugih rasah. Na osnovi danega vzorca testiraj hipotezo, da je ukvarjanje s športom pri beli rasi in drugih rasah enako pogosto.
Pomoč: test $H_0(p = q) : H_1(p \neq q)$, testna statistika $Z = \frac{\overline{p}-\overline{q}}{\sqrt{\hat{p}(1-\hat{p})}}\sqrt{\frac{nm}{n+m}}$.

2. Avtomat sortira antibiotik v stekleničke po 100 $mg$ v vsako. Zaradi naključnih dogodkov, ki jih ni možno odpraviti, odmerki v steklenicah nekoliko nihajo, porazdeljeni so normalno $N(\mu, \sigma)$. Z avtomatom smo zadovoljni, če je povprečna vrednost odmerka vsaj 100 $mg$. Na osnovi vzorca (*Vzorec_1*) na stopnji tveganja $\alpha = 0,05$ testiraj hipotezo $H_0(\mu \geq 100) : H_1(\mu < 100)$.

3. Proizvajalec zdravil proti nespečnosti zagotavlja, da sta njegovi zdravili $A$ in $B$ enako učinkoviti. Da bi preverili njegova zagotovila, smo zdravili testirali na 10. bolnikih. Pri tem smo dobili naslednje podatke (*Vzorec_2*), kjer je $X$ število dodatnih ur spanja pri zdravilu $A$ in $Y$ število dodatnih ur spanja pri zdravilu $B$. Ali lahko na osnovi tveganja $\alpha = 0,01$ proizvajalčevo zagotovilo zavržemo?
Pomoč: smemo predpostaviti, da je razlika $T = X - Y$ med dodatnimi urami spanja na populaciji porazdeljena normalno $N(\mu, \sigma)$. Testiraj hipotezo $H_0(\mu = 0) : H_1(\mu \neq 0)$.
Rezultat preveri s programom SPSS. Uporabi postopek *Analyze - Compare Means - Paired-Samples T Test* vstavi spremenljivki $X$ in $Y$.
**Opomba:** na ta način se primerja populacijsko povprečje dveh odvisnih vzorcev z metodo razlik!

4. (a) Porazdelitev zveznih statističnih spremenljivk *Starost*, *TezaO* in *KolicinaTD* prikaži s histogramom.
Uporabi postopek *Analyze - Descriptive Statistics - Frequencies - Charts - Histograms* in odkljukaj *With normal curve.*

(b) S testom Kolmogorova s stopnjo značilnosti $\alpha = 0,05$ za vsako omenjeno statistično spremenljivko preizkusi hipotezo, da je na populaciji porazdeljena normalno. Kaj ugotoviš?

5. Igralno kocko vržemo 1200 krat. Pri tem smo dobili naslednje rezultate:

| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 183 | 211 | 170 | 220 | 200 | 216 |

   Na stopnji tveganja $\alpha = 0,05$ preiskusimo hipotezo, da smo metali pošteno igralno kocko.
   Pomoč: uporabi Pearsonov hi kvadrat test.
   Rezultat preveri s programom SPSS. Uporabi postopek *Analyze - Nonparametric Test - Chi-Square*.

6. Na stopnji značilnosti $\alpha = 0,05$ testiraj ničelni hipotezi,

   (a) da je količina popite tekočine neodvisna od ukvarjanja s športom.

   (b) da je ukvarjanje s športom neodvisno od rase.

   Pomoč: uporabi preizkus neodvisnosti s kontingenčno tabelo za nominalni spremenljivki *KolicinaTD* in *Sport* oziroma *Sport* in *Rasa*.
   Rezultat preveri s programom SPSS. Uporabi postopek *Analyze - Descriptive Statistics - Crosstabs* in v *Statistics* označi *Chi-square*.

7. Z neparametričnima testoma za neodvisne vzorce (test Smirnova, inverzijski test) na stopnji tveganja $\alpha = 0,05$ preizkusi ničelno hipotezo o enakosti porazdelitve količine popite tekočine glede na ukvarjanje s športom.
   V SPSS uporabi postopek *Analyze - Nonparametric Test - Two-Independent-Samples Tests* vstavi *KolicinaTD* v *Test Variable List* in *Sport* v *Grouping Variable* ter v *Define Groups* vstavi vrednosti 0 in 1 ter označi testa *Mann-Whitney U* in *Kolmogorov-Smirnov*.

8. Proizvajalec zdravil proti nespečnosti zagotavlja, da sta njegovi zdravili $A$ in $B$ enako učinkoviti. Da bi preverili njegova zagotovila, smo zdravili testirali na 30. bolnikih (*Vzorec_3*). Z neparametričnima testoma za odvisne vzorce (test z znaki, test z rangi) s stopnjo tveganja $\alpha = 0,05$ preizkusi ničelno hipotezo, da sta zdravili enako učinkoviti.
   V SPSS uporabi postopek *Analyze - Nonparametric Test - Two-Related-Samples Tests* ter označi testa *Wilcoxon* in *Sign*.

## Teoretično ozadje

### Parametrični preizkusi značilnosti

*Parametrični preizkusi značilnosti* so namenjeni testiranju *parametričnih hipotez*, to je domnev o vrednostih neznanih parametrov statistične spremenljivke $X$. Na primer praviloma testiramo ničelno hipotezo $H_0$, ki pravi, da je parameter $q = q_0$, proti alternativni hipotezi $H_1$, ki pravi $q \neq q_0$, na *stopnji značilnosti* testa $\alpha$. Najpogosteje uporabimo značilnost testa $\alpha = 0,05$ ali $\alpha = 0,01$. Na osnovi tega pri preizkusu značilnosti ničelno hipotezo $H_0$ :

- bodisi zavrnemo,
- bodisi ne zavrnemo.

V prvem primeru rečemo, da med hipotetičnimi in eksperimentalnimi podatki obstaja značilna razlika (ali razlika je *signifikantna*) in hipotezo $H_0$ zavrnemo, v drugem primeru pa razlika med hipotetičnimi in eksperimentalnimi vrednostmi ni značilna oz. ni statistično pomembna, zato hipoteze $H_0$ ne zavrnemo. Pri testu značilnosti lahko naredimo samo t.i. *napako prve vrste*, to pomeni, da smo zavrnili pravilno hipotezo $H_0$. Verjetnost za to napako je predpisana s stopnjo značilnosti $\alpha$ in znaša običajno $0,05$ ali $0,01$.

Velja si posebej zapomniti, da pri preizkusu značilnosti $H_0$ proti alternativi $H_1$, ničelno hipotezo $H_0$ ali zavrnemo (torej sprejmemo $H_1$) ali o njej ne odločimo!

Parametrični preizkusi značilnosti potekajo vedno na naslednji način:

1. Postavimo ničelno in alternativno hipotezo. Opravka imamo
   (a) bodisi z dvostranskim testom
   $$H_0(q = q_0) \qquad \text{proti} \qquad H_1(q \neq q_0)$$
   (b) bodisi z enim od enostranskih testov
   $$\begin{array}{lll} H_0(q = q_0) \text{ oz. } H_0(q \leq q_0) & \quad\text{proti}\quad & H_1(q > q_0) \\ H_0(q = q_0) \text{ oz. } H_0(q \geq q_0) & \quad\text{proti}\quad & H_1(q < q_0) \end{array}$$
2. Izberemo stopnjo značilnosti testa $\alpha$ (običajno $0,05$ ali $0,01$).
3. Glede na velikost vzorca in obravnavanega problema izberemo primerno testno statistiko $U$.

4. Glede na porazdelitev statistike $U$ in parameter $\alpha$ določimo kritično območje testa $w_0$, to je podmnožica realnih števil izbrana tako, da je verjetnost dogodka, da ob pravilni hipotezi $H_0$ vrednost testne statistike $U$ leži v njej, manjša ali enaka $\alpha$. Torej zapišemo $P(U \in w_0|H_0) \leq \alpha$.

5. Izračunamo eksperimentalno vrednost testne statistike $u_e$. Če je $u_e \in w_0$, potem hipotezo $H_0$ zavrnemo. Če $u_e \neq w_0$, potem hipoteze $H_0$ ne zavrnemo.

## Neparametrični preizkusi značilnosti

Z *neparametrični preizkusi značilnosti* preskušamo neparametrične hipoteze, to se pravi domneve o tipu porazdelitvenega zakona ene ali več slučajnih spremenljivk. Tudi pri neparametričnih preizkusih postopamo natanko tako, kot pri parametričnih preizkusih zančilnosti. Oglejmo si nekaj najbolj uporabnih neparametričnih testov.

### Prilagoditveni testi

So namenjeni preizkušanju ničelne hipoteze $H_0$, da je neznana porazdelitev $F_X$ statistične spremenljivke $X$ enaka neki znani porazdelitvi $F_0$ proti alternativni hipotezi $H_1$, da je ta porazdelitev različna od $F_0$, t.j. $H_0(F_X = F_0) : H_1(F_X \neq F_0)$. Za testiranje tovrstnega problema uporabimo test *Kolmogorova* ali *Pearsonov hi kvadrat.*

- **Test Kolmogorova** lahko uporabimo pri zveznih porazdelitvah in to najbolj zanesljivo v primeru velikih vzorcev. Ker je sam test matematično precej zapleten, predvsem njegova porazdelitev, ga tukaj ne bomo posebej predstavili. Ta test ima vgrajen tudi program SPSS pod imenom *One-Sample Kolmogorov-Smirnov Test* in ga uporablja za testiranje ali je porazdelitev normalna, eksponentna, enakomerna, Poissonova.

- **Pearsonov hi kvadrat** je uporaben tako pri zveznih kot diskretnih porazdelitvah. Ideja pri tem testu je naslednja.
  Zalogo vrednosti statistične spremenljivke $X$ razdelimo na $r$ različnih razredov $S_1, S_2, \ldots, S_r$. Za vsak $k = 1, 2, \ldots, r$ naj bo $p_k$ verjetnost, da statistična spremenljivka $X$ ob pravilni hipotezi $H_0$ zavzame vrednost iz razreda $S_k$. Če je $n$ velikost vzorca, potem je $np_k$ hipotetična frekvenca razreda $S_k$ in naj bo $N_k$ eksperimentalna (vzorčna) frekvenca rezreda $S_k$. Potem se izkaže, da je za velike $n$ statistika, ki ji pravimo Pearsonov hi kvadrat

$$\chi^2 = \sum_{k=1}^{r} \frac{(N_k - np_k)^2}{np_k} \approx \chi^2(r-1)\,,$$

porazdeljena aproksimativno po zakonu hi kvadrat z $r-1$ prostostnimi stopnjami. Če je hipoteza $H_0$ pravilna so vrednosti statistike $\chi^2$ majhne. Hipotezo $H_0$ zavrnemo, če je izračunana vrednost $\chi^2$ večja od kritične $\chi^2_\alpha$, kjer je $P(\chi^2 > \chi^2_\alpha) = \alpha$.
**Opomba** 1. Če je potrebno predhodno oceniti $m$ parametrov (npr., če testiramo, da je nekaj porazdeljeno normalno $N(\mu, \sigma)$ z neznanima parametroma $\mu$, $\sigma$, potem moramo predhodno oceniti vrednosti teh dveh parametrov, zato je $m = 2$), potem je ta statistika porazdeljena približno po zakonu $\chi^2(r-m-1)$ z $r-m-1$ prostostnimi stopnjami.
**Opomba** 2. Pearsonov hi kvadrat test se lahko uporabi zmeraj, ko je $np_k \geq 5$, sicer je potrebno združiti posamezne razrede.

### Primerjalni testi

So namenjeni preizkušanju ničelne hipoteze $H_0$, da sta porazdelitvi $F_X$ in $F_Y$ dveh statističnih spremenljivk $X$ in $Y$ enaki proti alternativni hipotezi $H_1$, da sta porazdelitvi različni, t.j. $H_0(F_X = F_Y) : H_1(F_X \neq F_Y)$. Za testiranje tovrstnega problema najpogosteje uporabimo *test Smirnova, test z znaki, test z rangi, inverzijski test* in *iteracijski test.*

- **Test Smirnova** lahko uporabimo pri zveznih porazdelitvah in to najbolj zanesljivo v primeru velikih vzorcev na dveh neodvisnih vzorcih. Ta test je matematično precej zapleten, predvsem njegova porazdelitev, zato ga ne bomo posebej izpostavili. Test Smirnova ima vgrajen program SPSS pod imenom *Kolmogorov-Smirnov Z.*

- **Test z znaki** se uporablja za testiranje enakosti porazdelitev statističnih spremenljivk $X$ in $Y$ pri zveznih porazdelitvah na isti populaciji. Naj bo

$$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$$

vzorec velikosti $n$. Iz njega naredimo slučajni vzorec razlik

$$X_1 - Y_1, X_2 - Y_2, \ldots, X_n - Y_n\,.$$

Naj bo $K^+$ število pozitivnih razlik v vzorcu. Če je hipoteza $H_0$ pravilna, potem je verjetnost, da je posamezna razlika $X_i - Y_i$ pozitivna, enaka $\frac{1}{2}$. Zato je spremnljivka $K^+$ porazdeljena po binomskem zakonu $b(n, \frac{1}{2})$. Za velike $n$ lahko dano binomsko porazdelitev aproksimiramo z normalno porazdelitvijo in dobimo, da je statistika

$$Z = \frac{2K^+ - n}{\sqrt{n}} \approx N(0,1)$$

porazdeljena približno standardizirano normalno. Hipotezo $H_0$ lahko zavrnemo, če je izračunana vrednost $|z|$ večja od kritične $z_\alpha$, kjer je $P(|Z| \geq z_\alpha) = \alpha$.

- **Test z rangi ali Wilcoxonov test** se uporablja pri enakih pogojih kot test z znaki. V tem primeru absolutne razlike $|X_i - Y_j|$ rangiramo, enakih razlikam damo povprečen rang. Vsota rangov pozitivnih razlik, je vrednost statistike $W^+$ in vsota rangov, ki pripada negativnim razlikam je vrednost $W^-$ Za testno statistiko izberemo manjšo od obeh, torej

$$W = \min\{W^+, W^-\}.$$

Če je hipoteza $H_0$ pravilna, potem je vsota rangov blizu polovične vsote vseh rangov, to je $\frac{1}{4}n(n+1)$. Če je odstopanje večje, je to znak, da porazdelitvi nista enaki. Za velike $n$ se izkaže, da lahko statistiko $W$ aproksimiramo z normalno porazdelitvijo $N(\mu_w, \sigma_w)$, kjer je

$$\mu_w = \frac{1}{4}n(n+1) \qquad \sigma_w = \sqrt{\frac{1}{24}n(n+1)(2n+1)}\,.$$

Zato je v tem primeru statistika

$$Z = \frac{W - \mu_w}{\sigma_w} \approx N(0,1)$$

porazdeljena približno standardizirano normalno. Hipotezo $H_0$ lahko zavrnemo, če je izračunana vrednost $|z|$ večja od kritične $z_\alpha$, kjer je $P(|Z| \geq z_\alpha) = \alpha$.
**Opomba.** Ta test lahko uporabljamo tudi za testiranje enakosti matematičnih upanj, torej $H_0(E(X) = E(Y))$ pri dveh statističnih spremenljivkah, ki nista nujno normalno porazdeljeni.

- **Inverzijski test ali Mann-Whitneyjev test** uporabimo za testiranje enakosti porazdelitev statističnih spremenljivk $X$ in $Y$ pri dveh neodvisnih vzorcih. Naj bosta

$$X_1, X_2, \ldots, X_m \qquad \text{in} \qquad Y_1, Y_2, \ldots, Y_n$$

dva neodvisna vzorca velikosti $m$ in $n$. Oba vzorca združimo in uredimo po velikosti v zaporedje

$$Z_1 \leq Z_2 \leq \ldots \leq Z_{m+n}$$

in to zaporedje rangiramo. Kadar se v zgornjem zaporedju pojavi vrednost $Y_j$ pred vrednostjo $X_i$, pravimo, da je nastopila *inverzija*. Naj bo $R$ vsota vseh rangov, ki jih zavzame spremenljivka $X$. Potem je število inverzij

$$U = R - \frac{m(m+1)}{2}$$

in za velike vzorce (dovolj že $n+m \geq 20$ in oba $n, m \geq 4$) je statistika

$$Z = \frac{2U - mn}{\sqrt{mn(m+n-1)}}\sqrt{3} \approx N(0,1)$$

porazdeljena aproksimativno standardizirano normalno. Hipotezo $H_0$ lahko zavrnemo, če je izračunana vrednost $|z|$ večja od kritične $z_\alpha$, kjer je $P(|Z| \geq z_\alpha) = \alpha$.

- **Iteracijski test ali Wald-Wolfowitzev test** se uporablja pri enakih pogojih kot inverzijski test in temelji na urejenem nizu vrednosti iz obeh vzorcev. Tokrat preštejemo število skupin (iteracij), ki pripadajo isti spremenljivki. Ob pravilni domnevi $H_0$ o enaki porazdelitvi so iteracije kratke in jih je veliko. Za majhne $m$ in $n$ je porazdelitev števila $K$ vseh iteracij znana, a ni preprosta. Za velike vzorce (oba $m, n \geq 20$) uporabljamo normalno aproksimacijo. Testna statistika

$$Z = \frac{(m+n)K - 2mn}{2mn}\sqrt{m+n} \approx N(0,1)$$

je porazdeljena približno standardizirano normalno.

### Testiranje neodvisnosti

Veliko krat nas zanima ali sta statistični spremenljivki $X$ in $Y$ na populaciji porazdeljeni neodvisno. Torej testiramo ničelno hipotezo $H_0$, da sta $X$ in $Y$ neodvisni proti alternativi $H_1$, da nista neodvisni. Neodvisnost testiramo praviloma vedno s *Spearmanovo korelacijo rangov* ali s *kontingenčno tabelo.*

- **Test neodvisnosti s kontingenčno tabelo**. Vrednosti spremenljivke $X$ razdelimo na $r$ razredov $A_1, A_2, \ldots, A_r$ in vrednosti spremenljivke $Y$ razdelimo na $s$ razredov $B_1, B_2, \ldots, B_s$. Denimo, da dobimo iz populacije velik vzorec velikosti $n$. Naj $N_{ik}$ označuje frekvenco dogodka $A_iB_k$ v tem vzorcu. Frekvence $N_{ik}$ predstavimo s kontingenčno tabelo

| $X\backslash Y$ | $B_1$ | $\cdots$ | $B_k$ | $\cdots$ | $B_s$ | |
|---|---|---|---|---|---|---|
| $A_1$ | $N_{11}$ | $\cdots$ | $N_{1k}$ | $\cdots$ | $N_{1s}$ | $N_{1.}$ |
| $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | $\vdots$ |
| $A_i$ | $N_{i1}$ | $\cdots$ | $N_{ik}$ | $\cdots$ | $N_{is}$ | $N_{i.}$ |
| $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | $\vdots$ |
| $A_r$ | $N_{r1}$ | $\cdots$ | $N_{rk}$ | $\cdots$ | $N_{rs}$ | $N_{r.}$ |
| | $N_{.1}$ | $\cdots$ | $N_{.k}$ | $\cdots$ | $N_{.s}$ | n |

in definiramo robne vrstične in stolpčne frekvence

$$N_{i.} = N_{i1} + N_{i2} + \ldots + N_{is}\,, \quad \text{za } i = 1,2,\ldots,r\,,$$
$$N_{.k} = N_{1k} + N_{2k} + \ldots + N_{rk}\,, \quad \text{za } k = 1,2,\ldots,s\,.$$

Ob predpostavki, da sta $X$ in $Y$ porazdeljeni neodvisno se izkaže, da je statistika

$$\chi^2 = n\left(\sum_{i=1}^{r}\sum_{k=1}^{s}\frac{N_{ik}^2}{N_{i.}N_{.k}} - 1\right) \approx \chi^2((r-1)(s-1))$$

porazdeljena aproksimativno po hi kvadrat porazdelitvi z $(r-1)(s-1)$ prostostnimi stopnjami. Hipotezo $H_0$ lahko zavrnemo, če je izračunana vrednost $\chi^2$ večja od kritične $\chi^2_\alpha$, kjer je $P(\chi^2 \geq \chi^2_\alpha) = \alpha$.
**Opomba.** V primeru, ko je $r = s = 2$

| $X\backslash Y$ | $B_1$ | $B_2$ | |
|---|---|---|---|
| $A_1$ | $N_{11}$ | $N_{12}$ | $N_{1.}$ |
| $A_2$ | $N_{21}$ | $N_{22}$ | $N_{2.}$ |
| | $N_{.1}$ | $N_{.2}$ | n |

moramo imeti dodatno izpolnjeno predpostavko, da je $N_{i.}N_{.k} \geq 50n$ za $i,k = 1,2$. Če temu ni tako in je $n \geq 40$ ter so vse vrednosti $N_{i.}N_{.k} \geq 5n$ lahko uporabimo *Yatesovo korekturo*

$$\chi^2 = \frac{n(|N_{11}N_{22} - N_{21}N_{12}| - \frac{n}{2})^2}{N_{.1}N_{.2}N_{2.}N_{1.}} \approx \chi^2(1)\,.$$

- **Spearmanovo korelacijo rangov** lahko uporabimo za testiranje neodvisnosti, če sta $X$ in $Y$ vsaj ordinalni spremenljivki. Naj imata statistični spremenljivki $X$ in $Y$ na vzorcu $\{e_1, e_2, \ldots, e_n\}$ velikosti $n$ vrednosti

  $$X_1, X_2, \ldots, X_n \quad \text{in} \quad Y_1, Y_2, \ldots, Y_n\,.$$

  Dane vrednosti rangiramo

  $$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)} \quad \text{in} \quad Y_{(1)} \leq Y_{(2)} \leq \ldots \leq Y_{(n)}\,.$$

  Naj $I_k$ označuje rang $X_k$ in naj $J_k$ označuje rang $Y_k$ ter naj bo $D_k = I_k - J_k$ razlika rangov. Število

  $$R_S = 1 - \frac{6}{n^3 - n}\sum_{k=1}^{n} D_k^2$$

imenujemo *Spearmanov korelacijski koeficient*, ki zavzame vrednosti z intervala $[-1, 1]$. Ob hipotezi, da sta $X$ in $Y$ neodvisni, mora biti vrednost $R_S$ blizu 0. Če je $n$ velik, je statistika

$$U = R_S\sqrt{n-1} \approx N(0,1)$$

porazdeljena aproksimativno standardizirano normalno.

# Literatura


[1] . Benkovič, Vaje iz biostatistike, Medicinska fakulteta Univerze v Mariboru.

[2] Š. Adamič: Temelji biostatistike, Medicinska fakulteta Univerze v Ljubljani, Ljubljana 1995.

[3] R. Jamnik: Verjetnostni račun in statistika, DMFA, Ljubljana 1995.

[4] B. R. Kirkwood, J. A. C. Sterne: Essential medical statistics, Blackwell Publishing company, Malden 2004.

[5] B. Sluban: Uporaba statističnih metod v tekstilstvu, Fakulteta za strojništvo Univerze v Mariboru, Maribor 2004.