

STATISTIKA ZA PSIHologe 2019/20

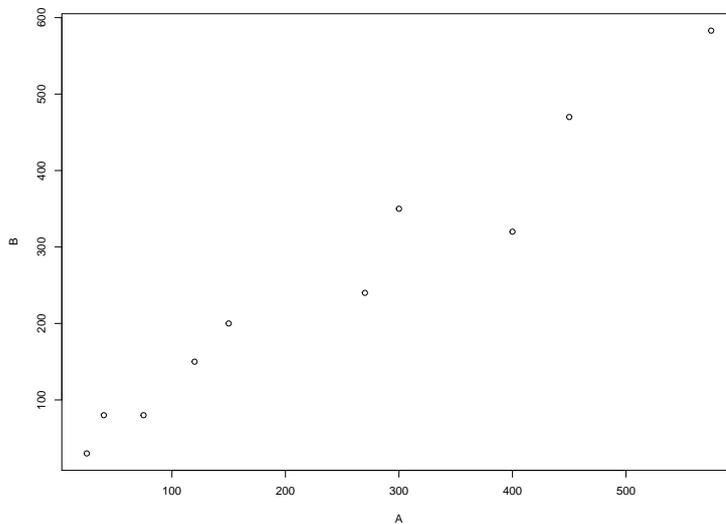
VAJA 10 - REŠITVE

3. ¹ Koncentracijo nitrata so merili s klasično metodo A. To metodo so želeli nadomestiti z novo metodo B. Da bi ugotovili ali dajeta metodi skladne rezultate, so nekaj časa uporabljali obe metodi. Rezultati so podani v spodnji tabeli.

Vzorec št.	A	B
1	25	30
2	40	80
3	120	150
4	75	80
5	150	200
6	300	350
7	270	240
8	400	320
9	450	470
10	575	583

Ali sta metoda A in metoda B pozitivno korelirani pri stopnji značilnosti $\alpha = 0,01$?

Rešitev: Če podatke predstavimo v razsevnem grafikonu, dobimo naslednjo sliko



Že iz slike se vidi, da sta metodi A in B pozitivno korelirani. Preverimo še s statističnim testom. Spremenljivka X naj bo metoda A in spremenljivka Y metoda B.

$$\mu_X = \frac{25 + 40 + 120 + \dots + 575}{10} = 240,5$$

$$\mu_Y = \frac{30 + 80 + 150 + \dots + 583}{10} = 250,3$$

¹K. Košmelj: Uporabna statistika

$$\begin{aligned}\sigma_X &= \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_{10}^2}{10} - (\mu_X)^2} \\ &= \sqrt{\frac{25^2 + 40^2 + \dots + 575^2}{10} - (240,5)^2} \\ &= 179,55\end{aligned}$$

$$\begin{aligned}\sigma_Y &= \sqrt{\frac{y_1^2 + y_2^2 + \dots + y_{10}^2}{10} - (\mu_Y)^2} \\ &= \sqrt{\frac{30^2 + 80^2 + \dots + 583^2}{10} - (250,3)^2} \\ &= 171,17\end{aligned}$$

In še kovarianca

$$\begin{aligned}K_{X,Y} &= \frac{x_1y_1 + x_2y_2 + \dots + x_{10}y_{10}}{10} - \mu_X\mu_Y \\ &= \frac{25 \cdot 30 + 40 \cdot 80 + \dots + 575 \cdot 583}{10} - 240,5 \cdot 250,3 \\ &= 30050,35\end{aligned}$$

Sedaj lahko izračunamo Pearsonov korelacijski koeficient

$$r = r_{X,Y} = \frac{K_{X,Y}}{\sigma_X\sigma_Y} = \frac{30050,35}{179,55 \cdot 171,17} = 0,98$$

Zapišemo ničelno in alternativno hipotezo:

H_0 : X in Y v populaciji nista korelirani.

H_1 : X in Y v populaciji sta pozitivno korelirani.

Glede na našo alternativno domnevo izvedemo enostranski T -test v desno z $n - 2$ stopinjami prostosti. To pomeni, da bomo ničelno domnevo zavrnilo natanko tedaj, ko bo veljalo

$$T > t_{1-\alpha}(n - 2).$$

Izračunamo vrednost testne statistike, ki je enaka

$$T = \frac{r_{X,Y}}{\sqrt{1 - r_{X,Y}^2}} \cdot \sqrt{n - 2}.$$

Vstavimo podatke in dobimo

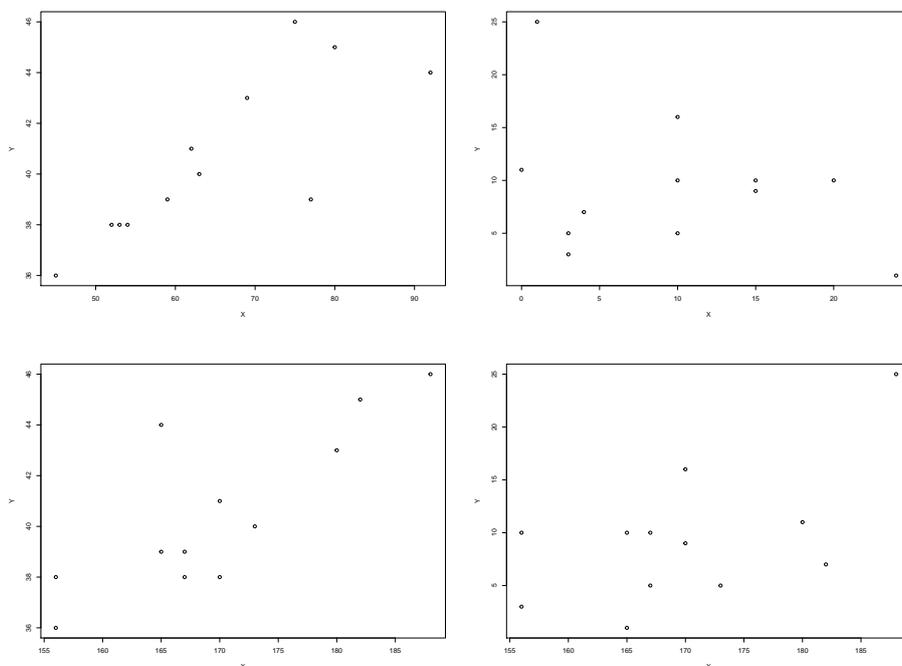
$$T = \frac{0,98}{\sqrt{1 - 0,98^2}} \cdot \sqrt{10 - 2} = 13,93.$$

Sedaj vrednost testne statistike primerjamo z vrednostjo iz tabele

$$t_{1-0,01}(8) = t_{0,99}(8) = 2,90.$$

Ker je $13,93 > 2,90$ lahko ničelno domnevo zavrnilo in potrdimo alternativno domnevo. Torej glede na naše podatke lahko rečemo, da sta metodi A in B za merjenje koncentracije nitrata pozitivno korelirani.

4. Izvedli smo študentsko anketo in zbarli podatke o telesni višini, teži, številki čevlja, število ur interneta na teden in število prebranih knjig na leto. Na spodnjih razsevnih diagramih določite sprmenljivki X in Y izmed prej naštetih in povejte ali sta po vašem mnenju spremenljivki X in Y korelirani.



Podatki iz anket so naslednji

Teža	62	75	54	77	52	53	69	63	80	92	59	45
Višina	170	188	156	165	170	167	180	173	182	165	167	156
Št. čevlja	41	46	38	39	38	38	43	40	45	44	39	36
Internet (ure/teden)	9	25	10	10	16	10	11	5	7	1	5	3
Knjige (na leto)	15	1	10	20	10	15	0	10	4	24	3	3

Za drugi diagram preverite ali sta pri $\alpha = 0,05$ spremenljivki X in Y korelirani v populaciji študentov.

Rešitev:

Diagram 1: $X =$ teža, $Y =$ št. čevlja

Diagram 2: $X =$ knjige, $Y =$ internet

Diagram 3: $X =$ višina, $Y =$ št. čevlja

Diagram 4: $X =$ višina, $Y =$ internet

Za drugi diagram torej rabimo podatke

Knjige (na leto)	15	1	10	20	10	15	0	10	4	24	3	3
Internet (ure/teden)	9	25	10	10	16	10	11	5	7	1	5	3

$$\mu_X = \frac{15 + 1 + 10 + \dots + 3}{12} = 9,58$$

$$\begin{aligned}\mu_Y &= \frac{9 + 25 + 10 + \dots + 3}{12} = 9,33 \\ \sigma_X &= \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_{12}^2}{12} - (\mu_X)^2} \\ &= \sqrt{\frac{15^2 + 1^2 + \dots + 3^2}{12} - (9,58)^2} \\ &= 7,41 \\ \sigma_Y &= \sqrt{\frac{y_1^2 + y_2^2 + \dots + y_{12}^2}{12} - (\mu_Y)^2} \\ &= \sqrt{\frac{9^2 + 25^2 + \dots + 3^2}{12} - (9,33)^2} \\ &= 6,10\end{aligned}$$

In še kovarianca

$$\begin{aligned}K_{X,Y} &= \frac{x_1y_1 + x_2y_2 + \dots + x_{12}y_{12}}{12} - \mu_X\mu_Y \\ &= \frac{15 \cdot 9 + 1 \cdot 25 + \dots + 3 \cdot 3}{12} - 9,58 \cdot 9,33 \\ &= -14,71\end{aligned}$$

Sedaj lahko izračunamo Pearsonov korelacijski koeficient

$$r = r_{X,Y} = \frac{K_{X,Y}}{\sigma_X\sigma_Y} = \frac{-14,71}{7,41 \cdot 6,10} = -0,33$$

Zapišemo ničelno in alternativno hipotezo:

H_0 : X in Y v populaciji nista korelirani.

H_1 : X in Y v populaciji sta korelirani.

Glede na našo alternativno domnevo izvedemo dvostranski T -test z $n - 2$ stopinjami prostosti. To pomeni, da bomo ničelno domnevo zavrnil natanko tedaj, ko bo veljalo

$$|T| > t_{1-\frac{\alpha}{2}}(n-2).$$

Izračunamo vrednost testne statistike, ki je enaka

$$T = \frac{r_{X,Y}}{\sqrt{1 - r_{X,Y}^2}} \cdot \sqrt{n-2}.$$

Vstavimo podatke in dobimo

$$T = \frac{-0,33}{\sqrt{1 - (-0,33)^2}} \cdot \sqrt{12-2} = -1,1.$$

Sedaj vrednost testne statistike primerjamo z vrednostjo iz tabele

$$t_{1-0,025}(10) = t_{0,975}(10) = 2,23.$$

Ker NE velja $|-1,1| > 2,23$, ne moremo zavrniti ničelne domneve. Torej glede na naše podatke ne moremo reči, da sta spremenljivki Št. ur interneta na teden in Št. knjig na leto korelirani v populaciji.