

POVEZANOST DVEH SPREMENLJIVK

Ordinalne spremenljivke

Mere povezanosti ločimo glede na tip spremenljivk:

- NOMINALNI tip para spremenljivk:
 - ena od spremenljivk je nominalna: χ^2 test
- ORDINALNI tip para spremenljivk:
 - ena spremenljivka je ordinalna, druga pa ordinalna ali boljša: Spearmanov koeficient korelacije rangov
- RAZMERNOSTNI oz. INTERVALNI tip para spremenljivk:
 - obe spremenljivki sta številski: Pearsonov koeficient korelacije

Spearmanov korelacijski koeficient ρ_s

Meri jakost povezave dveh ordinalnih spremenljivk. Če med spremenljivkama ni povezave, je vrednost ρ_s enaka 0.

Izračun:

$$\rho_s = 1 - \frac{6 \cdot \sum_{i=1}^N d_i^2}{N \cdot (N^2 - 1)}$$

kjer je: ρ_s oznaka za Spearmanov koef. korelacije rangov,
 d_i razlika med rangoma v i -ti enoti,
 N število enot.

Vrednosti koeficienta so med -1 in 1:

Predznak pove, ali je povezanost negativna ali pozitivna.

Vrednost pove, ali povezanost in obstaja, in če obstaja, ali je ta šibka, srednje močna ali močna.

POVEZANOST DVEH SPREMENLJIVK

Intervalne/razmernostne spremenljivke

Pearsonov korelacijski koeficient ρ

Meri jakost povezave dveh številskih (intervalnih / razmernostnih) spremenljivk. Če med spremenljivkama ni povezave, je vrednost ρ enaka 0.

Izračun:

$$\rho_{XY} = \frac{C_{XY}}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^N (x_i - \mu_X)(y_i - \mu_Y)}{\sqrt{\sum_{i=1}^N (x_i - \mu_X)^2 \cdot \sum_{i=1}^N (y_i - \mu_Y)^2}}$$

kjer je: ρ_{XY} oznaka za Pearsonov korelacijski koeficient,

C_{XY} kovarianca med spremenljivkama X in Y ,

σ_X standardni odklon spremenljivke X ,

σ_Y standardni odklon spremenljivke Y

Vrednosti koeficienta so med -1 in 1:

Predznak pove, ali je povezanost negativna ali pozitivna.

Vrednost pove, ali povezanost in obstaja, in če obstaja, ali je ta šibka, srednje močna ali močna.

Parcialna korelacija $\rho_{XY,Z}$

Če želimo iz zveze med spremenljivkama X in Y odstraniti vpliv tretje spremenljivke Z , potem je koeficient parcialne korelacije enak:

$$\rho_{XY,Z} = \frac{\rho_{XY} - \rho_{XZ} \cdot \rho_{YZ}}{\sqrt{(1 - \rho_{XZ}^2) \cdot (1 - \rho_{YZ}^2)}}$$

Vrednosti koeficienta so med -1 in 1:

Predznak pove, ali je povezanost negativna ali pozitivna.

Vrednost pove, ali povezanost in obstaja, in če obstaja, ali je ta šibka, srednje močna ali močna.

Za vse obravnavane korelacijske koeficiente velja:

Absolutna vrednost korelacijskega koeficienta	Pomen
od 0.5 do 1	močna povezanost
od 0.3 do 0.5	srednje močna povezanost
od 0.1 do 0.3	šibka povezanost
od 0 do 0.1	ni povezanosti

ODVISNOST DVEH SPREMENLJIVK

Intervalne/razmernostne spremenljivke

Odvisnost med pari spremenljivk računamo samo za *številске* spremenljivke, izmerjene na **intervalni** ali **razmernostni** merski lestvici.

Odvisnost med dvema številskima spremenljivkama ocenimo s pomočjo **linearne regresije**.

Linearna regresija

Metoda, ki omogoča pojasnjevanje in napovedovanje vrednosti odvisnih spremenljivk s pomočjo vrednosti neodvisnih spremenljivk.

Linearna premica: $Y' = a + bX$

Koeficienta a in b izračunamo:

$$a = \mu_Y - \frac{c_{XY}}{\sigma_X^2} \mu_X \quad b = \frac{c_{XY}}{\sigma_X^2}$$

Kakovost regresijskega modela:

- determinacijski koeficient: $R = \frac{\sigma_{y'}^2}{\sigma_y^2} = \frac{\sum_{i=1}^N (y_i' - \mu_Y)^2}{\sum_{i=1}^N (y_i - \mu_Y)^2} = \rho^2$
- standardna napaka ocene: $\sigma_e = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y_i')^2} = \sigma_Y \sqrt{1 - \rho^2}$

NALOGE

1. Dva evalvatorja sta razvrstila 10 programov glede na to, koliko prispevajo k večanju kakovosti življenja svojih uporabnikov. Rezultati so podani v naslednji tabeli:

Program	A	B	C	D	E	F	G	H	I	J
<i>Evalvator 1</i>	3	10	1	9	5	6	7	8	2	4
<i>Evalvator 2</i>	5	9	3	6	7	2	10	4	1	8

Izračunajte Spearmanov korelacijski koeficient rangov. V kolikšni meri se evalvatorja strinjata o prispevku programov k večanju kakovosti življenja svojih uporabnikov?

2. Na fakultetnem tekmovanju iz matematike so udeleženci reševali 4 naloge, pri čemer je bila vsaka naloga vredna 10 točk. Za 9 tekmovalcev smo zapisali doseženo število točk ter letnik:

<i>Št. točk</i>	3	5	7	8	15	17	23	25	30
<i>Letnik</i>	2	1	3	3	1	2	3	1	2

Izračunajte Spearmanov korelacijski koeficient med letnikom in številom doseženih točk. Opredelite smer in jakost povezanosti spremenljivk!

3. Naslednja tabela prikazuje število zaposlenih v prodajni službi in v vsem podjetju za 10 manjših podjetij:

<i>Št. zaposl. v podjetju (x_i)</i>	58	39	81	105	78	95	159	125	68	120
<i>Št. zaposl. v prodaji (y_i)</i>	7	4	11	10	7	13	15	14	8	12

Podani sta tudi obe srednji vrednosti in oba standardna odklona:

$$\mu_X = 92.80, \sigma_X = 33.75 \quad \text{in} \quad \mu_Y = 10.10, \sigma_Y = 3.36$$

- a) Podatke prikažite v razsevnem diagramu in izračunajte Pearsonov koeficient. Ali podatki nakazujejo linearno povezanost med številom zaposlenih v podjetju in številom zaposlenih v prodaji? Če je odgovor pritrdilen, kakšno moč in kakšno smer povezanosti nakazujejo?
- b) Če podatki nakazujejo linearno povezanost, izračunajte še regresijsko premico odvisnosti števila zaposlenih v prodajni službi od števila zaposlenih v vsem podjetju.
- c) Izračunajte še determinacijski koeficient in ga obrazložite.

4. Voznik osebnega avtomobila je za 10 voženj beležil število prevoženih kilometrov in porabo goriva. Rezultati so prikazani v tabeli:

Št. kilometrov	20	35	60	35	65	50	40	25	25	45
Poraba goriva (l)	2.5	4.5	6	4.5	5.5	4	3.5	2.5	3	4

- a) Podatke prikažite v razsevnem diagramu in izračunajte Pearsonov koeficient. Ali podatki nakazujejo linearno povezanost med številom prevoženih kilometrov in količino porabljenega goriva? Če je odgovor pritrdilen, kakšno moč in kakšno smer povezanosti nakazujejo?
- b) Če podatki nakazujejo linearno povezanost, izračunajte še regresijsko premico odvisnosti porabljenega goriva od števila prevoženih kilometrov. S pomočjo regresijske premice napovejte, približno koliko goriva bo porabil voznik v povprečju za 42 km.
5. Raziskati želimo, ali je prodajna cena izdelka linearno odvisna od neposrednih materialnih stroškov. Naključno smo izbrali 6 izdelkov in dobili naslednje rezultate (v EUR):

Materialni stroški	1.2	5.6	3.8	4.9	12.3	7.9
Prodajna cena	4.6	7.8	6.1	5.0	11.9	8.5

- a) Podatke prikažite v razsevnem diagramu in izračunajte Pearsonov koeficient. Ali podatki nakazujejo linearno povezanost med prodajno ceno izdelka in neposrednimi materialnimi stroški? Če je odgovor pritrdilen, kakšno moč in kakšno smer povezanosti nakazujejo?
- b) Izračunajte determinacijski koeficient in ga obrazložite.
- c) Če podatki nakazujejo linearno povezanost, izračunajte še regresijsko premico odvisnosti prodajne cene izdelka od neposrednih materialnih stroškov.