

Korelacija – predavanje

Doc. dr. sc. Lidija Bilić-Zulle, dipl. inž. specijalist medicinske biokemije

Zavod za laboratorijsku dijagnostiku KBC Rijeka
Katedra za medicinsku informatiku Medicinskog fakulteta Sveučilišta u Rijeci



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Podsjetnik



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Testiranje hipoteze

- postavljanje hipoteze
- odabir statističkog testa
- određivanje razine značajnosti
- izračun statistike testa
- zaključak



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Istraživačko pitanje – hipoteza

- Postoji li razlika (u istoj varijabli) između dvaju (ili više) skupina?
- Postoji li povezanost dvaju varijabli?
- Može li se iz jedne (ili više) varijable predvidjeti druga?



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Odabir statističkog testa

Ovisi o:

- vrsti podataka/ljestvice mjerenja
- raspodjeli (normalnost)
- uzorku (veličina, zavisnost)
- broju pokazatelja (uni/bi/multivarijatni testovi)
- istraživačkom pitanju



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Osobine obilježja – mjerne ljestvice

	ljestvica mjerenja	primjeri podataka	postupci statističke obrade
smjer prirode podataka	omjerna x + <i>omjer vrijednosti</i>	koncentracija glukoze (mmol/L), broj bolesnika na odjelu.	Kvantitativni podaci: Studentov t-test, ANOVA, Pearsonova korelacija, regresijska rješavanja.
	intervalna + - <i>količina razlike</i>	temperatura (°C), pH-vrijednost.	Kvantitativni podaci: Studentov t-test, ANOVA, Pearsonova korelacija, regresijska rješavanja.
	ordinalna > < <i>smjer razlike</i>	stupanj opetina (I-IV), Ann-Arbor klasifikacija limfoma (I-IV).	Kvalitativni podaci: Mann-Whitney test, Kruskal-Wallis test, Spearmanova korelacija.
	nominalna = ≠ <i>postojanje razlike</i>	spot, fask, etiologija bolesti.	Kvalitativni podaci: χ ² -test, Fisherov test, logistička regresija.



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Odabir testa: ovisnost o raspodjeli – parametrijski i neparametrijski

- Parametrijski testovi:
 - pokazatelji koji ne odstupaju od normalne raspodjele
 - obrađuju izvorna mjerenja
- Neparametrijski testovi:
 - nisu ograničeni normalnošću raspodjele
 - ne raščlanjuju izvorne podatke već "rankove"



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Ovisnost o osobini uzorka

- Veličina:
 - veliki uzorci – parametrijski testovi
 - mali uzorci – neparametrijski testovi
- Zavisnost
 - nezavisni uzorci – neparni testovi
 - zavisni uzorci – parni testovi



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Ovisnost o broju pokazatelja

- univarijatni:
 - raščlanjuju samo jedan pokazatelj
- bivarijatni
 - raščlanjuju dva pokazatelja
- multivarijatni
 - raščlanjuju više pokazatelja



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Varijable

- biološke varijable – značajke ispitanika
- osnovna značajka – biološka varijabilnost
- potreba za mjerenjem na što većem broju ispitanika (reprezentativni uzorak!)



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Varijable

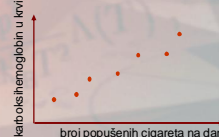
- stalne – ne će se promijeniti tijekom ispitivanja (spol, rasa, genotip...)
- promjenjive – mogu se promijeniti tijekom ispitivanja (broj leukocita, koncentracija glukoze...)



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Varijable

- nezavisne varijable – mjerimo neovisno o ishodu pokusa, prethode zavisnim varijablama, prediktori, rizični čimbenici, varijable izloženosti (os x)
- zavisne varijable – kriteriji, posljedice djelovanja nezavisnih varijabli, predstavljaju mjeru ishoda koju u istraživanju mjerimo (os y)
- primjer:



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Postoji li povezanost dvaju varijabli?

- statistički test

KORELACIJA



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Korelacija

- sukladnost dvaju varijabli
- govori o njihovoj povezanosti
- uvjet pokazatelja: numerički (brojčana mjerna ljestvica)
 - nominalna ☹
 - ordinalna ☺
 - intervalna ☺
 - omjerna ☺
- broj pokazatelja: dva



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Grafički prikaz korelacije

- točkasti dijagram (engl. *scatter diagram*)



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Izračun korelacije

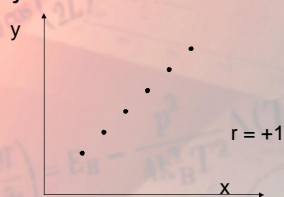
- Matematički postupak za izračunavanje stupnja povezanosti dvaju brojčanih pokazatelja kojim se izračunava:
 - koeficijent korelacije r
 - statistička značajnost koeficijenta korelacije P



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Koeficijent korelacije

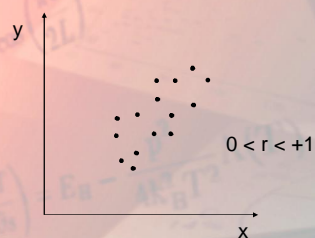
- Pozitivna potpuna (matematička) korelacija



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Koeficijent korelacije

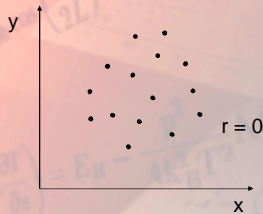
- Pozitivna nepotpuna korelacija



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Koeficijent korelacije

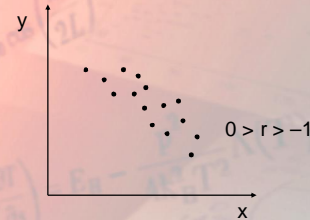
- Nepostojanje korelacije



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Koeficijent korelacije

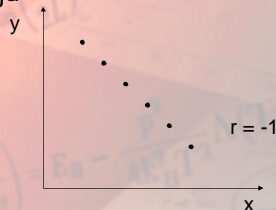
- Negativna nepotpuna korelacija



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Koeficijent korelacije

- Negativna potpuna (matematička) korelacija



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Pearsonov koeficijent korelacije r_p

- parametrijski test (računa s izmjerenim vrijednostima):
 - jednostavna linearna korelacija
 - dva pokazatelja
 - intervalna ili omjerna mjerna ljestvica
 - vrijednosti: $-1 \leq r \leq +1$
 - barem jedna varijabla – normalna raspodjela
 - veliki uzorak ($n > 30$)
 - povezanost varijabli linearna



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

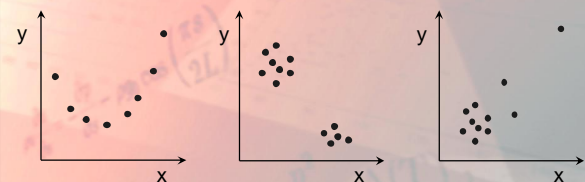
Spearmanov koeficijent korelacije

- neparametrijski test (računa s "rankovima"):
 - linearna korelacija, ρ , r_s
 - dva pokazatelja
 - ordinalna mjerna ljestvica
 - vrijednosti: $-1 \leq r \leq +1$
- kada r_s , a ne r_p :
 - ordinalna mjerna ljestvica
 - mali uzorak ($n < 30$)
 - x i y ne slijede normalnu raspodjelu



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Kada NE računati korelaciju!



- ... i kada se jedna varijabla izračunava iz druge (npr. BMI i tjelesna masa)



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Koeficijent korelacije

- r nema mjerne jedinice i izražava se brojem s dva decimalna mjesta
- ne ovisi o tome koji je pokazatelj x , a koji y
- $r = 0$ znači: nema linearne povezanosti
- ne opisuje uzročno-posljedičnu vezu (samo povezanost)



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Statistička značajnost povezanosti

- uz koeficijent korelacije r potrebno naznačiti i statističku značajnost P
- P – statistička značajnost povezanosti, izražava se brojem s tri decimalna mjesta
- ovisi o veličini uzorka
- statistički značajna korelacija $P < 0,05$

TUMAČITI SE SMIJU SAMO ZNAČAJNI KOEFICIJENTI KORELACIJE!



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Tumačenje koeficijenta korelacije

(za biološke varijable)*

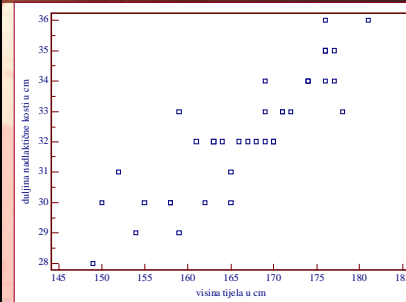
- $r = 0$ do $\pm 0,25$: nema povezanosti,
- $r = \pm 0,26$ do $\pm 0,50$: slaba povezanost,
- $r = \pm 0,51$ do $\pm 0,75$: umjerena do dobra povezanost
- $r = \pm 0,76$ do ± 1 : vrlo dobra do izvrsna povezanost
- $r = \pm 1$: matematička povezanost

*Colton, 1974.



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Primjer – izvrsna povezanost

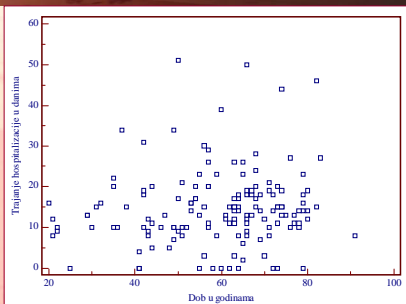


$N=40$
 $r = 0,88$
 $P < 0,001$



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Primjer – nema povezanosti



$N = 149$
 $r = 0,12$
 $P = 0,133$



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Jakost povezanosti

- udio zajedničkih vrijednosti
- koeficijent determinacije – r^2



- Primjer za $r = 0,67$; $r^2 = 0,45$ ili 45% zajedničkih vrijednosti



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Što smo naučili

- Osim ispitivanja razlike statističkom raščlambom može se ispitati i povezanost.
- Razliku ispitujemo među skupinama, a povezanost među varijablama.
- Postojanje i stupanj povezanosti ispituje se korelacijom.
- Pearsonov koeficijent korelacije parametrijski je test dok je Spearmanov neparametrijski.



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Što smo naučili

- Korelacija se tumači s obzirom na vrijednost koeficijenta korelacije, a ne njegove statističke značajnosti.
- Tumačiti se smiju samo statistički značajni koeficijenti korelacije.
- Jakost povezanosti izražava se kao udio zajedničkih vrijednosti, a izračunava se iz koeficijenta determinacije (r^2).



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Hvala na pozornosti



Lidija.Bilic-Zulle@medri.hr



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Pitanja P6

1. Na koje istraživačko pitanje odgovor daje korelacija?
2. Kako se grafički prikazuje korelacija?
3. Što znači potpuna, a što nepotpuna korelacija?
4. Što znači pozitivna, a što negativna korelacija?
5. Koji su uvjeti za izračunavanje Pearsonovog koeficijenta korelacije?
6. Kakav je test Spearmanova korelacija i koji su uvjeti za izračunavanje?
7. O čemu ovisi i kako se tumači statistička značajnost koeficijenta korelacije?
8. U kojim se slučajevima ne izračunava koeficijent korelacije?
9. Kako se tumači vrijednost koeficijenta korelacije za biološke varijable?
10. Kako se tumači jakost povezanosti, kako se izračunava i što predstavlja koeficijent determinacije?



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika