

Korelacija i regresija – seminar

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

Klinički 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

Primjer za raspravu - korelacija*

Je li ispravno računati Pearsonov koeficijent korelacije za stupanj opeklina na tijelu i trajanje bolničkog liječenja izraženo u danima?

* Udovičić M, Baždarić K, Bilić-Zulle L, Petrovečki M. Što treba znati kada izračunavamo koeficijent korelacije? *Biochemia Medica* 2007;17(1):10-5.

<http://www.biochemia-medica.com/content/online-content>



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Primjer za raspravu - korelacija*

U istraživanju povezanosti raspoložnja i količine tekućine (vode) unesene pijenjem tijekom dana dobivena je povezanost $r = 0,12$; $P = 0,003$. Je li ispravno tumačiti kako postoji značajna povezanost raspoložnja i količine popijene tekućine?

* Udovičić M, Baždarić K, Bilić-Zulle L, Petrovečki M. Što treba znati kada izračunavamo koeficijent korelacije? *Biochemia Medica* 2007;17(1):10-5.



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Primjer za raspravu - korelacija*

U istraživanju povezanosti visine tijela i biološke dobi dobivena je korelacija $r = 0,97$; $P = 0,001$. Možemo li zaključiti kako su visina i dob nesumnjivo zaista izvrsno povezani?

* Udovičić M, Baždarić K, Bilić-Zulle L, Petrovečki M. Što treba znati kada izračunavamo koeficijent korelacije? *Biochemia Medica* 2007;17(1):10-5.



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Primjer za raspravu - korelacija*

U istraživanju povezanosti koncentracije alkohola u krvi i prometnih nesreća utvrđeni su $r = 0,78$; $P = 0,002$. Možemo li zaključiti kako uzimanje alkohola nesumnjivo uzrokuje prometne nesreće?

* Udovičić M, Baždarić K, Bilić-Zulle L, Petrovečki M. Što treba znati kada izračunavamo koeficijent korelacije? *Biochemia Medica* 2007;17(1):10-5.



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Primjer za raspravu - korelacija*

Usporedbom katalitičke koncentracije dvaju enzima u krvi ispitanika dobivena je povezanost $r = 0,52$; $P = 0,002$. Možemo li zaključiti kako vrijednosti enzima imaju 52% zajedničkih vrijednosti katalitičke koncentracije?

* Udovičić M, Baždarić K, Bilić-Zulle L, Petrovečki M. Što treba znati kada izračunavamo koeficijent korelacije? *Biochemia Medica* 2007;17(1):10-5.



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Primjer za raspravu - korelacija*

Ispitana je povezanost vremena provedenog u radu s računalom i brzine pisanja teksta na računalu u žena ($n_1=60$) i muškaraca ($n_2=40$). Koeficijent korelacije za žene iznosi $r_1 = 0,70$, a za muškarce $r_2 = 0,50$ i oba su statistički značajna. Možemo li zaključiti kako je $r_1 > r_2$, tj. da je u žena veća povezanost vremena provedenog u radu za računalom i brzine pisanja teksta na računalu?

* Udovičić M, Baždarić K, Bilić-Zulle L, Petrovečki M. Što treba znati kada izračunavamo koeficijent korelacije? *Biochemia Medica* 2007;17(1):10-5.



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Usporedba dvaju koeficijenata korelacije

- za ista obilježja u dva uzorka ispitanika
- utvrđuje se značajnost razlike dvaju "r"
- ovisi o vrijednosti r i n



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Može li se iz jedne varijable predvidjeti druga?

- statistički test

REGRESIJSKA RAŠČLAMBA



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Regresijska raščlamba

... matematički model koji daje što točniju procjenu koliko će se promijeniti promatrana zavisna varijabla (kriterij) kao odgovor na promjenu nezavisne varijable (prediktora).



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Uvjeti za izračunavanje regresije

- postojanje korelacije između ispitivanih varijabli
- ravnomjerno raspršenje rezultata oko pravca regresije
- kvalitetan ispitivani uzorak (nasumično uzorkovanje)



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Regresijska raščlamba

- linearna regresija
- višestruka regresijska raščlamba
- logistička regresija
- Passing-Bablock regresija



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Linearna regresija

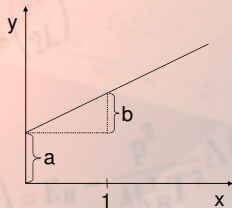
- matematička povezanost...
- ...dvaju pokazatelja (x i y)
- nezavisna mjerenja
- koliko promjena jednog (x) određuje promjenu drugog (y):
 - x: nezavisna varijabla (prediktor)
 - y: zavisna varijabla (kriterij)
- rezultat: jednadžba pravca: $y = a + bx$
tj. $y = \beta_0 + \beta_1 x$



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Linearna regresija

- $y = a + bx$; a = odsječak; b = nagib



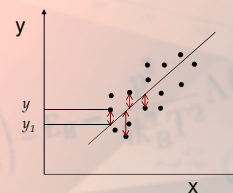
- $y = x$; idealan pravac kroz ishodište, $a = 0$; $b = 1$



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Linearna regresija

- metoda najmanjih kvadrata:
 - način utvrđivanja položaja regresijskog pravca među točkama
 - najmanji mogući zbroj kvadrata odstupanja pojedinačnih točaka od pravca po osi y, tj. suma kvadrata udaljenosti $(y - y_1)^2$ svih točaka



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Linearna regresija

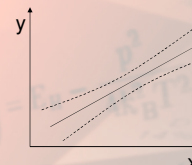
- pogreška prognoze (rezidual) – variranje rezultata oko pravca
- standardna pogreška prognoze – prosjek kvadrata udaljenosti svih točaka od pravca regresije;
 - naziva se i standardna devijacija reziduala (po smislu odgovara standardnoj devijaciji aritmetičke sredine)



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Linearna regresija

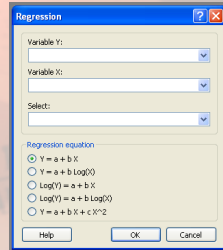
- granice pouzdanosti (CI)
 - raspon vrijednosti rezultata unutar kojeg s određenom sigurnošću nalazimo predviđene rezultate u populaciji (uobičajeno 95% CI)



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

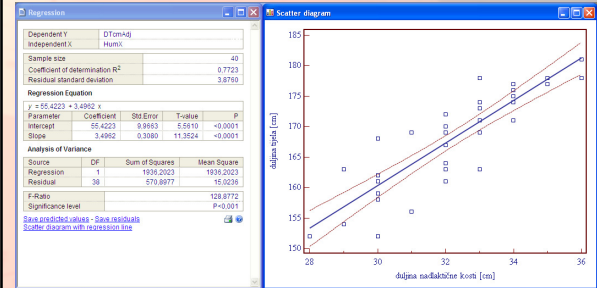
Linearna regresija

- kako se zadaje



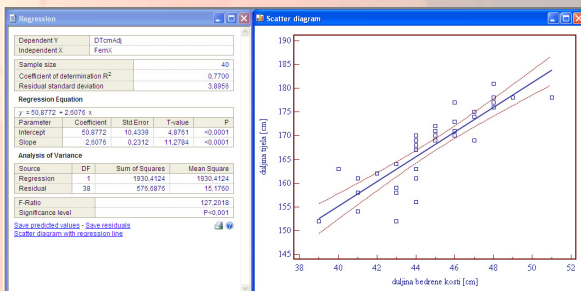
Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Linearna regresija - rezultat



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Linearna regresija - rezultat



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Višestruka regresijska raščlamba

- omogućuje:
 - predviđanje zavisne varijable (kriterija) iz više nezavisnih (pokazatelja, prediktora)
 - nemaju svi pokazatelji istu vrijednost u predviđanju
 - pojedinačni značajni pokazatelji mogu izgubiti značajnost u višestrukoj regresiji



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Višestruka regresija

- rezultat:

Method	Forward		
Enter variable if P <	0.05		
Remove variable if P >	0.1		
Sample size	40		
Coefficient of determination R ²	0.8203		
R ² -adjusted	0.8159		
Multiple correlation coefficient	0.9000		
Residual standard deviation	3.4305		
Regression Equation			
Independent variables	47.3234		
Constant			
HumX	1.8444 0.5580 3.305 0.0021		
FibX	1.5848 0.4955 3.392 0.0017		
Variables not included in the model			
FemX			
RadX			
TibX			
Uinc			
Analysis of Variance			
Source	DF	Sum of Squares	Mean Square
Regression	2	2071.5390	1035.8195
Residual	37	435.4610	11.7692
F-Ratio			88.0109
Significance level			P<0.001
Zero order correlation coefficients			
Variable	r		
FemX	0.877		
HumX	0.875		
FibX	0.880		
RadX	0.855		
TibX	0.819		
Uinc	0.858		



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Logistička regresija

- pronalazi najbolji odgovarajući matematički model za predviđanje dihotomnog ishoda iz niza nezavisnih varijabli
- značajka ishoda: dihotomnost, tj. dva moguća ishoda (0/1)



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Passing-Bablok regresija

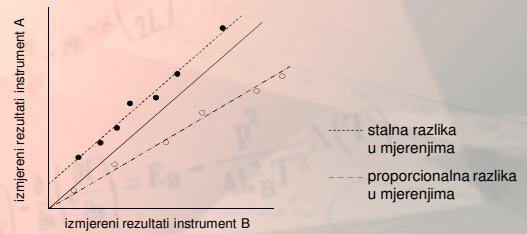
- model linearne regresije koji pretpostavlja kako niti jedna od dvaju varijabli nije niti zavisna niti nezavisna
- najčešće: usporedba dviju metoda mjerenja
- rezultat: vrijednost odsječka i nagiba pravca s intervalom pouzdanosti koji ukazuju na postojanje stalne ili proporcionalne razlike između mjerenja



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Passing-Bablok regresija

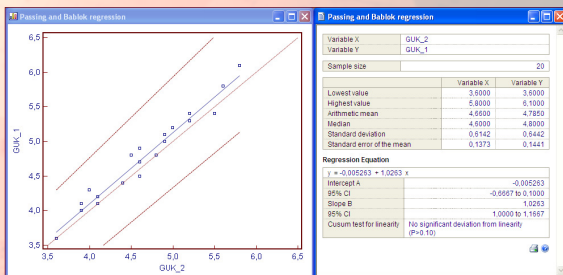
- grafički prikaz



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Passing-Bablok regresija

MedCalc® Version 9.3.9.0
Windows 98/NT/Me/2000/XP/Vista
<http://www.medcalc.be>
Copyright © 1990-2007
Frank Schoorjans



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Vaša pitanja...



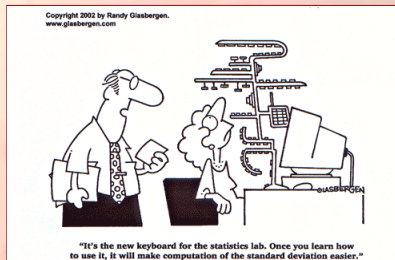
Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Moji odgovori... 😊



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Hvala na pozornosti



Lidija.Bilic-Zulle@medri.hr



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Statistika

Pitanja S6

1. Može li se korelacijom tumačiti uzročno-posljedična veza i zašto?
2. Kako se uspoređuju dva koeficijenta korelacije?
3. Što je regresijska raščlamba?
4. Koji uvjeti trebaju biti zadovoljeni za izračunavanje regresije?
5. Što predstavlja linearna regresija?
6. Objasnite jednadžbu pravca na grafičkom prikazu.
7. Objasnite metodu najmanjih kvadrata i što predstavlja standardna pogreška prognoze?
8. Koje su značajke višestruke regresije i kada se ona upotrebljava?
9. Koje su značajke logističke regresije i kada se ona upotrebljava?
10. Koje su značajke Passing-Bablok regresije i kada se ona upotrebljava?

