

Prof. dr. sc. Mladen Petrovečki
Katedra za medicinsku informatiku
Medicinski fakultet Sveučilišta u Rijeci
Klinička bolnica "Dubrava", Zagreb

Osnove biostatistike u svakodnevnoj praksi


Korelacija i regresija



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku 22. studeni 2008.

Teme i podaci

- podaci (Bolnica milosrdnih sestara, Zgb)
 - TIBC, *total iron binding capacity*
 - UIBC, *unsaturated...*
 - (transferin)+Fe=feritin
 - dob, spol
- dopunski podaci
 - *in vitro* uzgoj matičnih stanica
 - istraživanje polipoidnog melanoma
- korelacija
- pravocrtna regresija
- višestruka regresija
- logistička regresija
- Coxova regresija
- Passing-Bablokova regresija



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Linearni sustavi i kaos



Z. Sardar
I. Abrams

Linearni sustavi

Dakle: jednostavno rečeno, kaos je pojava aperioidičnih, naizgled slučajnih događaja u determinističkom sustavu. U kaosu ima reda, a u redu leži kaos. To dvoje je uzajamno povezano, bliškije no što smo ikada prije pomišljali.

Alli, to se čini nelogičnim, budući da su deterministički sustavi predvidljivi i stabilni. Iz puke navike, ljudi su oduvijek tražili uzorke i linearne odnose u onome što vide.


Linearni odnosi dopuštaju nam da predvidimo što će se dogoditi unutar sustava, i lako se mogu grafički prikazati.




Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Korelacija


- povezanost pokazatelja
- uvjet pokazatelja: numerički (brojčana mjerna ljestvica)
 - nominalna ☹️
 - ordinalna 😐
 - intervalna 😊
 - omjerna 😊
- broj pokazatelja: dva



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Pearsonov koeficijent korelacije


- jednostavna linearna korelacija, r_p
- dva pokazatelja
- intervalna ili omjerna mjerna ljestvica
- vrijednosti: $-1 \leq r \leq +1$



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

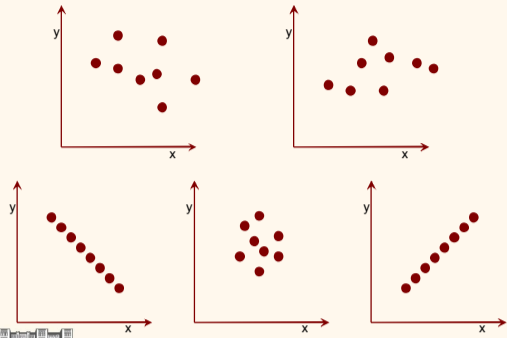
Spearmanov koeficijent korelacije

- neparametrijska linearna korelacija, ρ , r_s
 - dva pokazatelja
 - ordinalna mjerna ljestvica
 - vrijednosti: $-1 \leq r \leq +1$
- zašto r_s , a ne r_p :
 - ordinalna mjerna ljestvica
 - mali uzorak
 - x i y ne slijede normalnu raspodjelu



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Koeficijenti korelacije



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Tumačenje |r|

- 0 – matematička nepovezanost
- ≤0,25 – nema povezanosti
- 0,26-0,50 – slaba povezanost
- 0,51-0,75 – umjerena povezanost
- >0,75 – izvrsna povezanost
- 1 – matematička povezanost

Colton, 1974., *Statistics in Medicine*
(prema: Dawson & Trapp, 2007.)



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Značajnost povezanosti – P

- nulta hipoteza: $\rho = 0$
- statistika: t ovisi o r, N
- P – mjera značajnosti

uvjet	N	r	p
TIBC/UIBC svi	61	0,63	<0,001
izolirana DNA	18	0,66	0,003
dob>62 god.	04	0,97	0,035
Fe/UIBC svi	61	-0,38	0,002
dob>62 god.	04	-0,91	0,088



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Jakost povezanosti – r²

- koeficijent determinacije

uvjet	N	r	p	r ²
TIBC/UIBC svi	61	0,63	<0,001	39,7%
dob>62 god.	04	0,97	0,035	94,1%



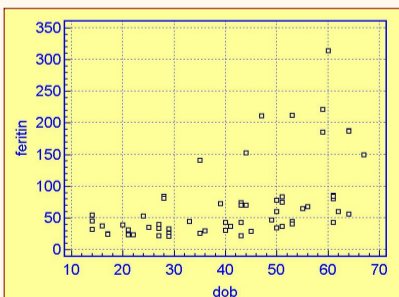
Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Primjer 1

r_P

uvjet	N	r	p
dob : feritin	61	0,53	<0,001
dob : Fe	61	-0,01	0,917
UIBC : TIBC	61	0,63	<0,001



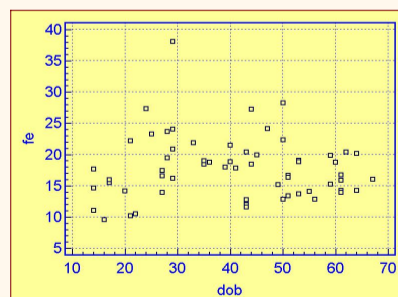
Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Primjer 2

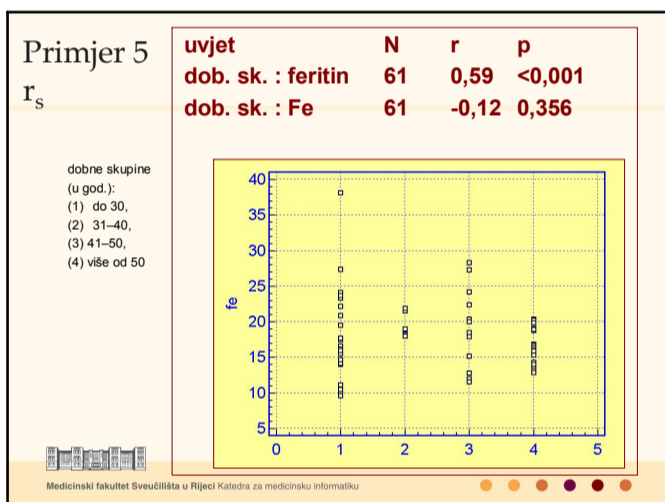
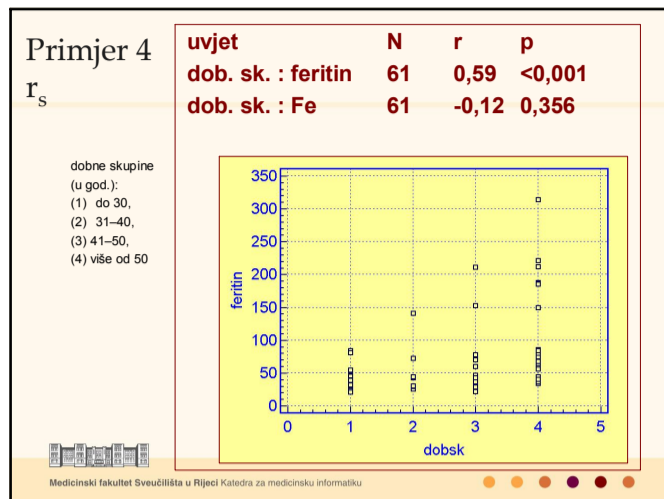
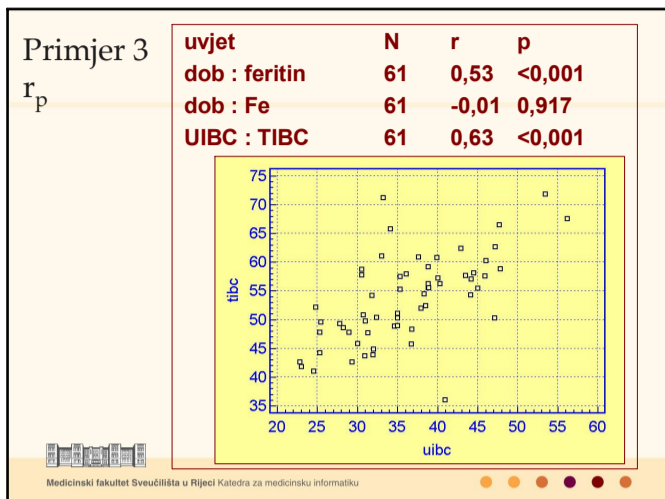
r_P

uvjet	N	r	p
dob : feritin	61	0,53	<0,001
dob : Fe	61	-0,01	0,917
UIBC : TIBC	61	0,63	<0,001



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

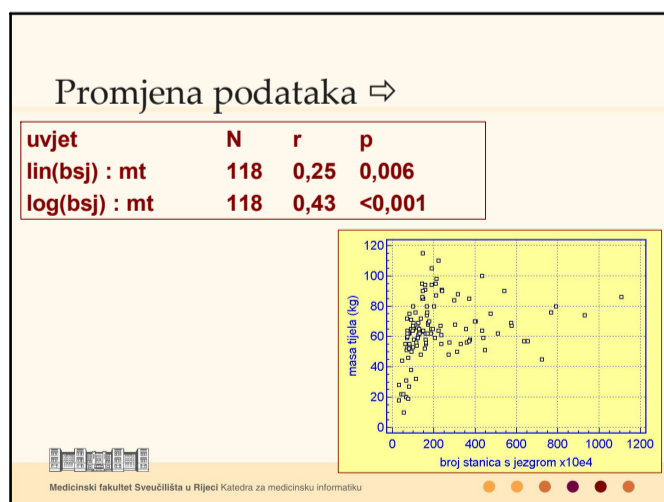
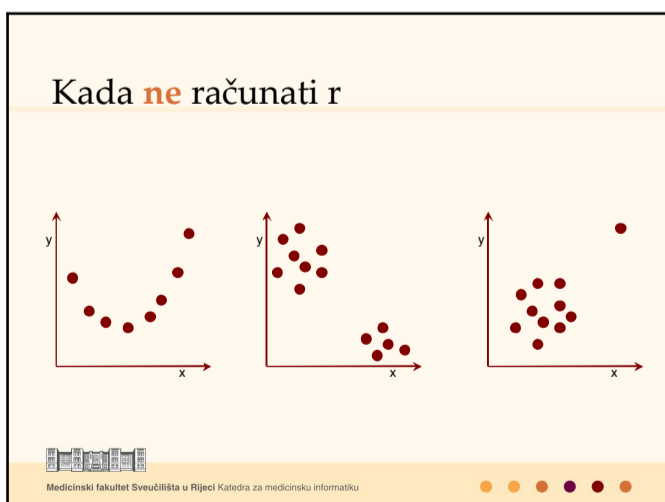




Pet "ne" r

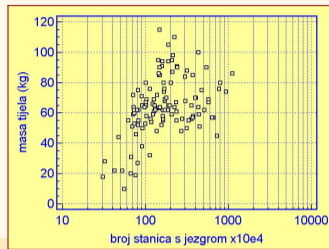
- r **nema** mjerne jedinice
- **ne ovisi** o tome koji je pokazatelj x, a koji y
- r=0 znači: **nema** pravocrtne povezanosti
- **ne smije** se rabiti za ekstrapolacije (samo interpolacije)
- **ne označava** uzročno-posljedičnu vezu (samo povezanost)

Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Promjena podataka

uvjet	N	r	p
lin(bsj) : mt	118	0,25	0,006
log(bsj) : mt	118	0,43	<0,001



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Pravocrtna 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)

$$y = \beta_0 + \beta_1 x$$



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Pravocrtna 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)

NAPOMENA:

$$y = a + bx,$$

$$y = l + kx,$$

$$y = b_0 + b_1 x,$$

$$y = \beta_0 + \beta_1 x \dots$$

a – slobodan član (*intercept*)

b – nagib pravca (*slope*)

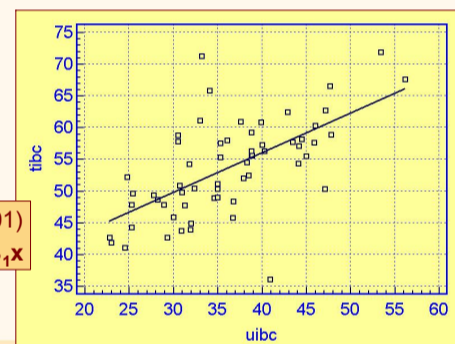
$$y = \beta_0 + \beta_1 x$$



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Primjer: TIBC vs. UIBC ⇔



$$r=0,63 \ (p<0,001)$$

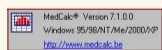
$$y = \beta_0 + \beta_1 x$$



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



primjer ⇔



Regression

Variable Y: tIBC

Variable X: uIBC

Select: []

Regression equation:

$Y = a + bX$

$Y = a + b \text{Log}(X)$

$\text{Log}(Y) = a + bX$

$\text{Log}(Y) = a + b \text{Log}(X)$

$Y = a + bX + cX^2$

Help OK Cancel



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



primjer ⇔

$$r=0,63 \ (p<0,001)$$

$$tIBC = (31,0 \pm 3,7) + (0,6 \pm 0,1) uIBC$$

Regression

Dependent Y : tIBC
Independent X : uIBC

Sample size = 61
Coefficient of determination = 0.3941
Residual standard deviation = 5.9017

--- REGRESSION EQUATION ---
Y = 31.0133 + 0.6244 X

Parameter	Coefficient	Std. Error	T-value	P
Intercept	31.01329	3.71823	8.3409	<0.0001
Slope	0.62438	0.10079	6.1946	<0.0001

--- ANALYSIS OF VARIANCE ---

Source	DF	Sum of Squares	Mean Square
Regression	1	1336.5481	1336.5481
Residual	55	2054.9896	34.8303

F-Ratio = 38.3731 P < 0.001



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



primjer ⇨

$r=0,63$ ($p<0,001$)
 $tibc = (31,0 \pm 3,7) + (0,6 \pm 0,1) uibc$

Ovisnost pokazatelja TIBC o:

UIBC	$\beta \pm s.e.(\beta)$	t	p
β_0	$31,0 \pm 3,7$	8,34	<0,001
β_1	$0,6 \pm 0,1$	6,19	<0,001

Fe

	$\beta \pm s.e.(\beta)$	t	p
β_0	$44,0 \pm 3,4$	12,94	<0,001
β_1	$0,5 \pm 0,2$	2,91	0,005

Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

primjer

$r=0,63$ ($p<0,001$)
 $tibc = (31,0 \pm 3,7) + (0,6 \pm 0,1) uibc$

SPSS for Windows
Release 7.5 (Nov 14 1996)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.628 ^a	.394	.384	5.9017

a. Predictors: (Constant), UIBC

Coefficients^a

Model		Unstandardized Coefficients		t	Sig.	95% Confidence Interval for B	
		B	Std. Error			Lower Bound	Upper Bound
1	(Constant)	31.013	3.718	8.341	.000	23.573	38.453
	UIBC	.624	.101	6.195	.000	.423	.826

a. Dependent Variable: TIBC

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1336.548	1	1336.548	38.373	.000 ^b
	Residual	2054.990	59	34.830		
	Total	3391.538	60			

a. Predictors: (Constant), UIBC
b. Dependent Variable: TIBC

Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

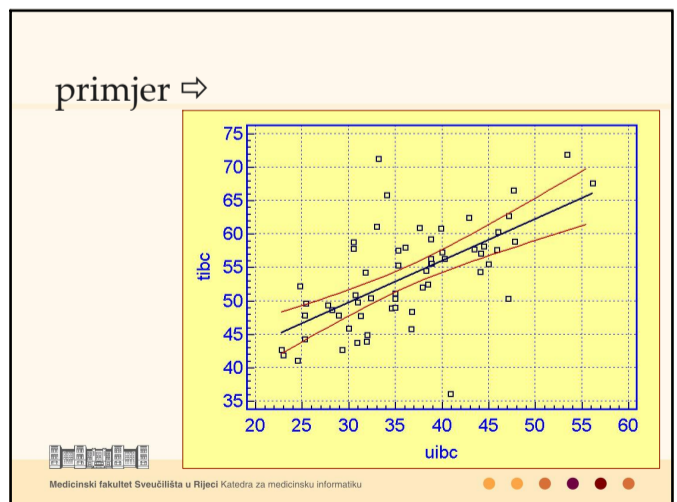
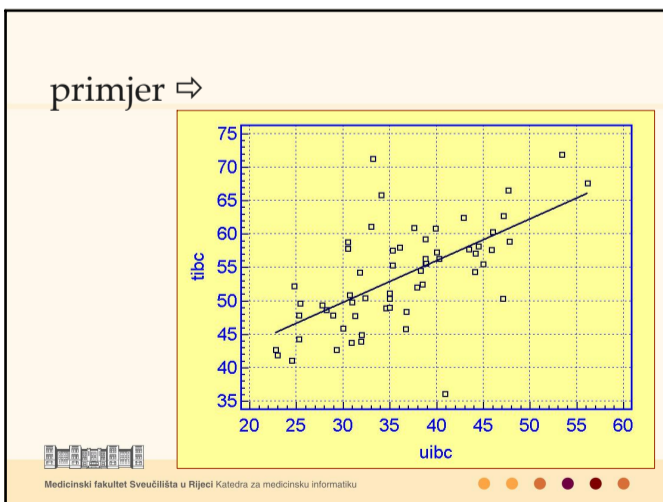
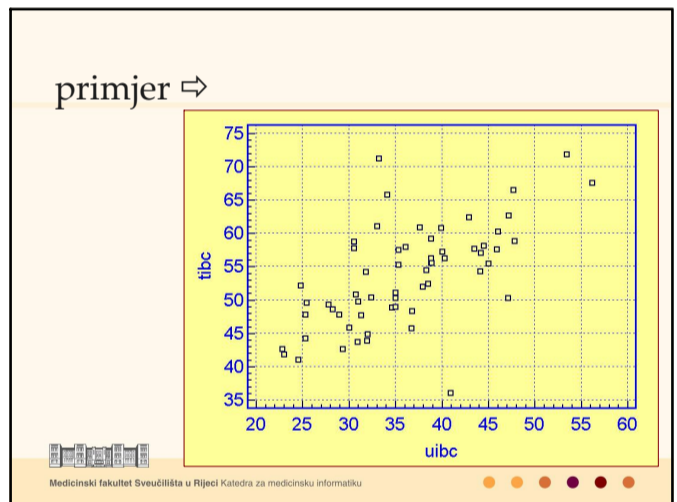
Primjer: granice pouzdanosti

Ovisnost pokazatelja TIBC o:

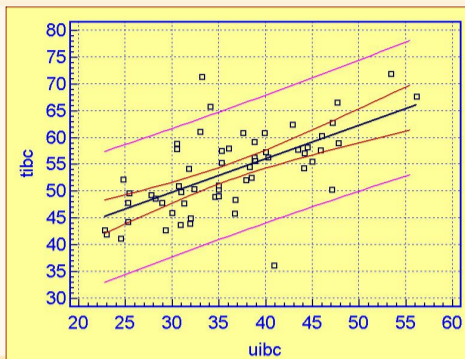
UIBC	$\beta \pm s.e.(\beta)$	95%GP(β)	t	p
β_0	$31,0 \pm 3,7$	23,6-38,5	8,34	<0,001
β_1	$0,6 \pm 0,1$	0,4-0,8	6,19	<0,001

granice pouzdanosti (Confidence Interval Limits):
 95%: $\beta \pm [1,96 \times s.e.(\beta)] = 31 \pm (1,96 \times 3,7) = 31 \pm 7,3 = 23,7-38,3$
 99%: $\beta \pm [2,58 \times s.e.(\beta)] = 31 \pm (2,58 \times 3,7) = 31 \pm 9,5 = 21,5-40,5$

Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



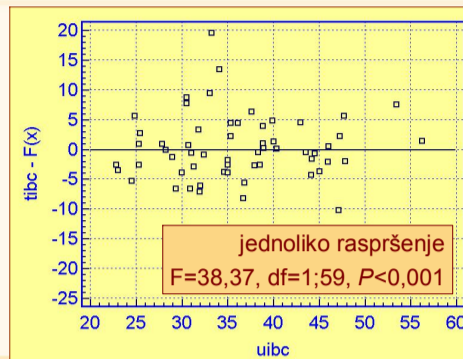
primjer ⇨



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



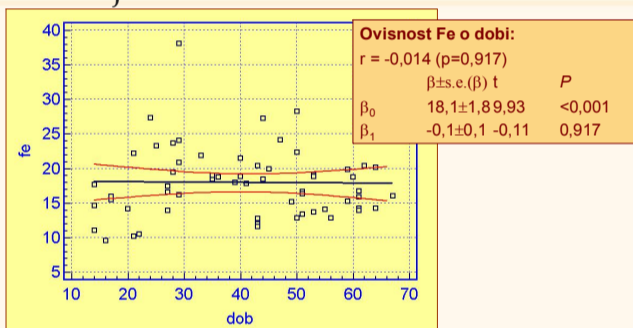
Primjer: reziduali



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Primjer: ovisnost Fe o dobi



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Višestruka regresija

- matematička povezanost više pokazatelja (x_1-x_n i y)
- nezavisna mjerenja
- koliko promjena svakog od x određuje promjenu y :
 - x : nezavisne varijable (prediktori)
 - y : zavisna varijabla (kriterij)

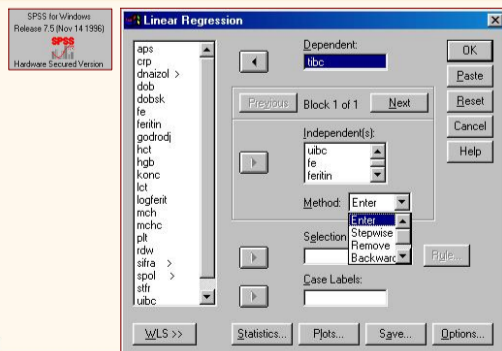
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Primjer: TIBC vs. 4 pokazatelja ⇨



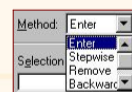
UIBC
Fe
ferritin
dob



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



primjer (univarijatno) ⇨



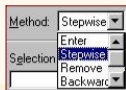
Mjera	$\beta \pm s.e.(\beta)$	t	P	
UIBC	β_0	31,0±3,7	8,34	<0,001
	β_1	0,6±0,1	6,19	<0,001
Fe	β_0	44,0±3,4	12,94	<0,001
	β_1	0,5±0,2	2,91	0,005
ferritin	β_0	57,1±1,4	41,89	<0,001
	β_1	-0,1±0,01	-3,45	<0,001
dob	β_0	62,4±2,4	25,47	<0,001
	β_1	-0,2±0,1	3,85	<0,001



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



primjer (multivarijatno)



Mjera	$\beta \pm s.e.(\beta)$	t	P
konstanta	β_0 4,5±4,3	0,96	0,341
UIBC	β_1 0,8±0,1	11,68	<0,001
Fe	β_2 1,0±0,1	11,01	<0,001
feritin	β_3 0,01±0,01	0,31	0,760
dob	β_4 0,01±0,03	-0,97	0,338

$$TIBC = 0,8 \times UIBC + 1 \times Fe$$



Medicinski fakultet Sveučilišta u Rijeci Katetra za medicinsku informatiku



Logistička regresija

- povezanost više pokazatelja (x_1-x_n i y)
- nezavisna mjerenja
- koliko promjena svakog od x određuje promjenu binarne varijable y :
 - x : nezavisne varijable (prediktori)
 - y : zavisna varijabla (kriterij)
- e^β = omjer izgleda (OR, *odds ratio*)

$$\text{logit}(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$



Medicinski fakultet Sveučilišta u Rijeci Katetra za medicinsku informatiku



Primjer: spol vs. 5 pokazatelja ⇨

Dependent Y: zspol (dummy)

Sample size: 61

Cases with Y=0: 25 (40,98%)

Cases with Y=1: 36 (59,02%)

--- Overall Model Fit ---

Null model -2 Log Likelihood = 82,56946

Full model -2 Log Likelihood = 69,46355

Chi-square = 13,10; df = 5; P = 0,022



Medicinski fakultet Sveučilišta u Rijeci Katetra za medicinsku informatiku



primjer ⇨

Var	β	s.e.	P	OR (95% CI)
tibc	0,05	0,09	0,563	
uibc	-0,05	0,09	0,615	
fe	-0,03	0,11	0,757	
feritin	-0,02	0,01	0,006	0,98 (0,96 – 0,99)
dob	-0,02	0,02	0,421	
cons.	1,60	0,49	0,003	



Medicinski fakultet Sveučilišta u Rijeci Katetra za medicinsku informatiku



primjer

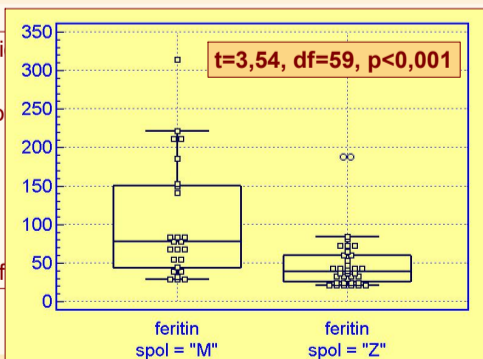
--- Classifi

Actual gro

Y = 0

Y = 1

Percent of



Medicinski fakultet Sveučilišta u Rijeci Katetra za medicinsku informatiku



Coxova regresija

- povezanost više pokazatelja (x_1-x_n i y)
- nezavisna mjerenja
- koliko promjena x određuje y :
 - x : nezavisne varijable (prediktori)
 - y : zavisna varijabla (kriterij) = mjera rizika
- e^β = omjer rizika (RH, *relative hazard*)

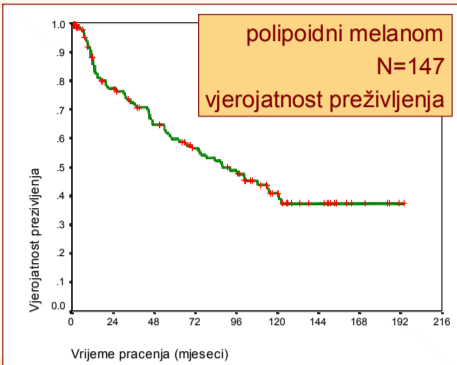
$$\lambda_i(t) = \lambda_0(t) \exp [\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n]$$



Medicinski fakultet Sveučilišta u Rijeci Katetra za medicinsku informatiku



Primjer: utjecaj niza varijabli na $P \Rightarrow$



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



primjer \Rightarrow

Var	β	s.e.	p	RH (95% CI)
dob	0.02	0,011	0,019	1,02 (1,001 – 1,04)
spol_m			0,941	
kl_st			<0,001	
(1)	1,75	0,35	<0,001	5,81 (2,91 – 11,58)
(2)	3,70	0,58	<0,001	40,60 (13,02 – 126,65)
clark_st			0,098	
br_st			0,433	

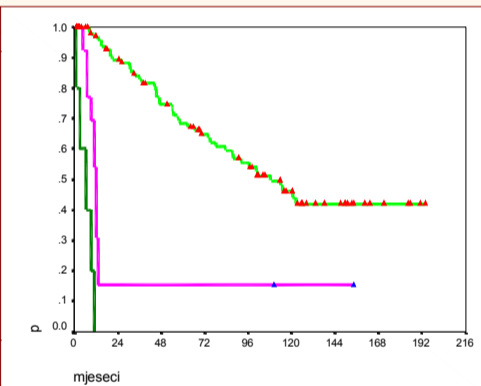


Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



primjer

Var	β
dob	0.02
spol_m	
kl_st	
(1)	1,75
(2)	3,70
clark_st	
br_st	



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Passing-Bablok regresija

- povezanost dvaju istih pokazatelja (x i y)
- nezavisna mjerenja: x/y ili y/x
- koliko promjena bilo kojeg određuje promjenu onog drugog
 - x: nezavisna varijabla (prediktor)
 - y: zavisna varijabla (kriterij) (i obrnuto)

$$y = \beta_0 + \beta_1 x$$

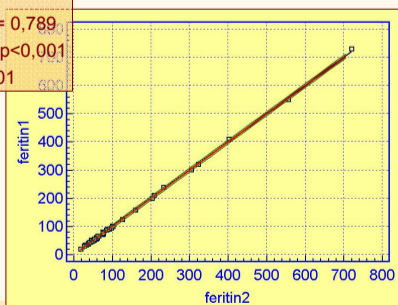


Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Primjer: feritin, 2 postupka mjerenja \Rightarrow

N=46
 $r^2 = 0,997$
 $\beta_0 = 0,17 \pm 0,62$, $t = 0,26$, $p = 0,789$
 $\beta_1 = 1,00 \pm 0,00$, $t = 320,71$, $p < 0,001$
 F-ratio = 102856,65, $p < 0,001$



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

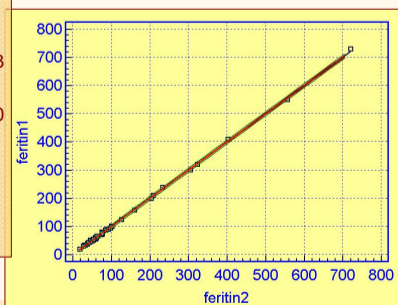


primjer

$$y = 0,956 + 0,991x$$

Intercept (**A=0**): 0,956
 95% CI : 0,500 to 2,163
 Slope (**B=1**): 0,991
 95% CI : 0,979 to 1,000

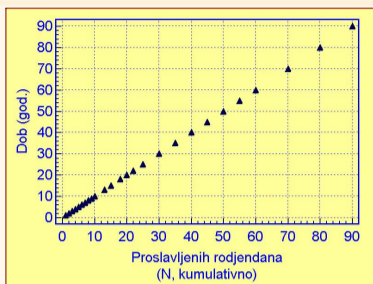
Cusum test for linearity:
 Significant deviation from linearity ($P < 0,01$)



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Zadnji primjer 😊



Regresijski dokaz da je zdravo slaviti rođendan

(S. den Hartog (Ph.D. Thesis), University of Groningen)



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



MEDICINSKA
INFORMATIKA

Hvala na pozornosti

mp@kbd.hr



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

