



ANALIZA POVEZANOSTI

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

1

KORELACIJA

- veza među obilježjima (varijablama)
- obilježja koja "variraju zajedno"

KOEFICIJENT KORELACIJE

- mjera stupnja povezanosti

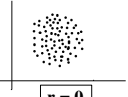
PEARSONOV KOEFICIJENT KORELACIJE r

- mjera stupnja *linearne* povezanosti dviju kvantitativnih varijabli

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku


2

$-1 \leq r \leq 1$

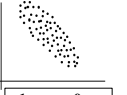


nema povezanosti

$r = 0$

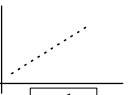


stohastička povezanost




$-1 < r < 0$

$0 < r < 1$



funkcionalna povezanost



$r = -1$

$r = 1$

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

3

x, ynizovi vrijednosti varijabli čiju povezanost ocjenjujemo

POSTUPAK ZA OCJENU KORELACIJE

- crtanje korelacionog dijagrama
- ocjena postojanja povezanosti
- u slučaju da postoji linearna povezanost, računamo koeficijent korelacije r

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}$$

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

4

skraćeni postupak računanja r :

$$r = \frac{\sum_{i=1}^N x_i y_i - \frac{1}{N} \left(\sum_{i=1}^N x_i \right) \left(\sum_{i=1}^N y_i \right)}{\sqrt{\left[\sum_{i=1}^N x_i^2 - \frac{1}{N} \left(\sum_{i=1}^N x_i \right)^2 \right] \left[\sum_{i=1}^N y_i^2 - \frac{1}{N} \left(\sum_{i=1}^N y_i \right)^2 \right]}}$$

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

5

ZNAČAJNOST KOEFICIJENTA KORELACIJE

- testiramo je li r značajno različit od 0
- test statistika

$$t = r \frac{\sqrt{N-2}}{\sqrt{1-r^2}}$$

slijedi t razdiobu uz $df = N - 2$

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

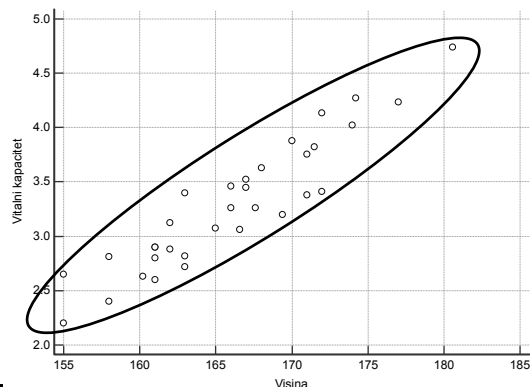
6

Izmjerena je visina u centimetrima i vitalni kapacitet pluća (VC) u litrama 33 studentice prve godine. Dobiveni su sljedeći rezultati:

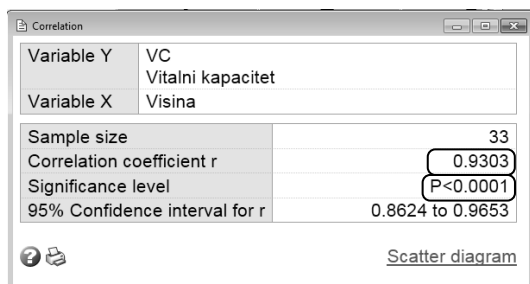
Rbr.	Visina	VC	Rbr.	Visina	VC	Rbr.	Visina	VC
1.	180.6	4.74	12.	155.0	2.20	23.	174.2	4.27
2.	168.0	3.63	13.	171.0	3.38	24.	167.0	3.45
3.	163.0	3.40	14.	171.5	3.82	25.	162.0	2.88
4.	171.0	3.75	15.	167.6	3.26	26.	172.0	4.13
5.	177.0	4.23	16.	160.2	2.63	27.	161.0	2.90
6.	169.4	3.20	17.	166.6	3.06	28.	155.0	2.65
7.	161.0	2.90	18.	167.0	3.52	29.	162.0	3.12
8.	170.0	3.88	19.	163.0	2.82	30.	174.0	4.02
9.	158.0	2.40	20.	172.0	3.41	31.	161.0	2.80
10.	161.0	2.60	21.	158.0	2.81	32.	166.0	3.46
11.	163.0	2.72	22.	165.0	3.07	33.	166.0	3.26

Ocijenite postoji li povezanost visine i vitalnog kapaciteta pluća

Crtanje korelacionog dijagrama (raspršni/"scatter" grafikon)



Izračun koeficijenta korelacije



Interpretacija koeficijenta korelacije

statistička značajnost

- ocjenjuje je li r značajno različit od 0
- ovisi o veličini uzorka - za velike uzorke, mali r će biti značajan

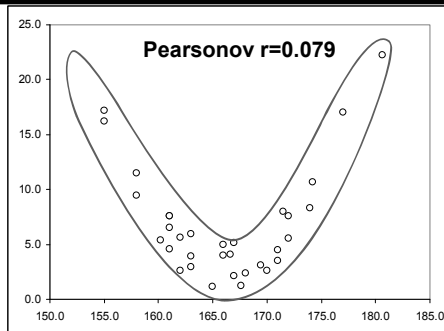
praktična značajnost

- ocjenjuje se pomoću **koeficijenta determinacije r^2**
- koliki udio varijabilnosti je "zajednički"

Interpretacija koeficijenta korelacije

N	Najmanji značajni r (p<0.05)	r^2
10	0.632	0.399
20	0.444	0.197
30	0.361	0.130
40	0.312	0.097
50	0.279	0.078
100	0.197	0.039
200	0.139	0.019
300	0.113	0.013
500	0.088	0.008


VAŽNO: Pearsonov koeficijent korelacije daje stupanj LINEARNE povezanosti dviju varijabli!



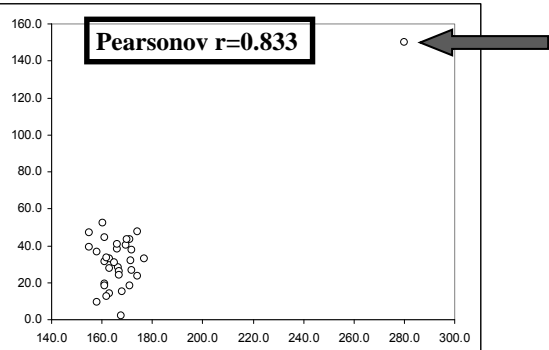


VAŽNO:
Korelacija daje **povezanost**, a ne
UZROČNOST!

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 13

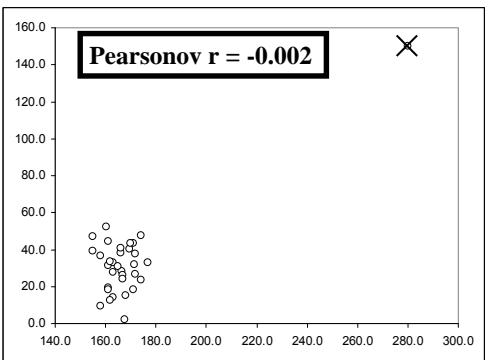


VAŽNO:
Na koeficijent korelacije jako utječu ekstremne vrijednosti!



Pearsonov r=0.833

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 14



Pearsonov r = -0.002

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 15

SPEARMANOV KOEFICIJENT KORELACIJE ρ

- **neparametrijski koeficijent korelacije**

KADA?

- **Ordinalne varijable**
- **Jedna ili obje numeričke varijable nisu normalno distribuirane**
- **Prisustvo ekstremnih vrijednosti**

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 16

"POINT-BISERIJALNI" KOEFICIJENT KORELACIJE

- korelacija između jedne kontinuirane i jedne dihotomne varijable
- računa se kao Pearson-ov r uz numeriranu dihotomnu varijablu

KOEFICIJENT KORELACIJE Φ

- korelacija između dihotomnih varijabli
- izračunava se direktno iz χ^2 prema formuli

$$\Phi = \sqrt{\frac{\chi^2}{N}}$$

– značajnost χ^2 ocjenjuje značajnost koeficijenta Φ

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 17

KOEFICIJENT KONTINGENCIJE C

- korelacija između varijabli od kojih jedna ili obje imaju više kategorija
- izračunava se direktno iz χ^2 prema formuli

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}}$$

- značajnost χ^2 ocjenjuje značajnost koeficijenta C
- prednost: ne zahtijeva simetričnu raspodjelu varijabli
- nedostatak: maksimalna vrijednost C ovisi o broju kategorija

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 18

LINEARNA REGRESIJA

- ako parovi varijabli pokazuju prisustvo korelacije, funkcionalnu vezu prikazuje JEDNADŽBA REGRESIJE

REGRESIJA - prognoza iz jedne varijable u drugu linearni slučaj:

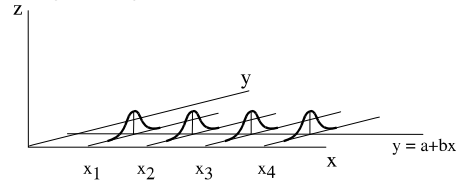
- povezanost varijabli je linearna
- jednadžba regresije je jednadžba pravca** oko kojeg se grupiraju parovi varijabli u korelacionom dijagramu

$$y = a + bx$$

OPĆI OBLIK JEDNADŽBE LINEARNE REGRESIJE

- x ... nezavisna varijabla (prediktorska)
- y ... zavisna varijabla (kriterijska)
- b ... koeficijent smjera

-u realnoj situaciji:



- jednadžba regresijskog pravca dobiva se METODOM NAJMANJIH KVADRATA

$$\text{uz uvjet } \sum_i (y_i - y'_i)^2 = \min$$

y'_i ... vrijednost na regresijskom pravcu koja odgovara x_i

iz normalnih jednadžbi

$$\sum_{i=1}^N y_i = Na + b \sum_{i=1}^N x_i$$

$$\sum_{i=1}^N x_i y_i = a \sum_{i=1}^N x_i + b \sum_{i=1}^N x_i^2$$

$$b = \frac{\sum_{i=1}^N x_i y_i - \frac{1}{N} \left(\sum_{i=1}^N x_i \right) \left(\sum_{i=1}^N y_i \right)}{\sum_{i=1}^N x_i^2 - \frac{1}{N} \left(\sum_{i=1}^N x_i \right)^2}$$

KOEFICIJENT REGRESIJE

a ... odsječak na ordinati

$$a = \bar{y} - b\bar{x}$$

- pravac regresije izražava "prosječni odnos" ("prosječnu vezu") varijabli x i y

LINEARNA REGRESIJA

ocjena modela



87% varijabilnosti vitalnog kapaciteta pluća može se objasniti visinom

poboljšanje u predviđanju zbog korištenja regresijskog modela (razlika sume kvadrata odstupanja od aritmetičke sredine i sume kvadrata odstupanja od vrijednosti predviđenih regresijskim pravcem)

Analysis of Variance

Source	DF	Sum of Squares	Mean Square
Regression	1	9.7037	9.7037
Residual	31	1.5085	0.04866
F-ratio			199.4107
Significance level			P<0.0001

suma kvadrata odstupanja od vrijednosti predviđenih regresijskim pravcem

regresijski model značajno bolje predviđa zavisnu varijablu od predviđanja aritmetičkom sredinom

Regression Equation

$y = -11.5374 + 0.08927 x$

Parameter	Coefficient	Std. Error	95% CI	t	P
Intercept	-11.5374	1.0503	-13.6794 to -9.3953	-10.9851	<0.0001
Slope	0.08927	0.006321	0.07637 to 0.1022	14.1213	<0.0001

VAŽNO:
 Predviđanja se smiju raditi samo za vrijednosti iz postojećeg raspona varijabli!

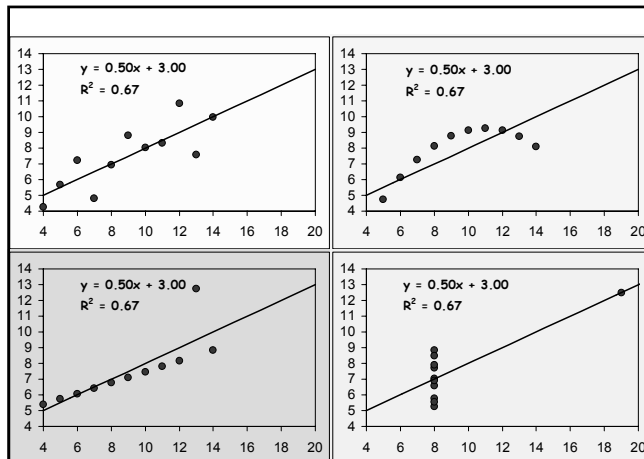
npr. za visinu 175,
 vitalni kapacitet pluća = $-11.537 + 0.089 \times 175 = 4.04$

ZAŠTO MORAMO VIDJETI GRAFIČKI PRIKAZ PODATAKA?

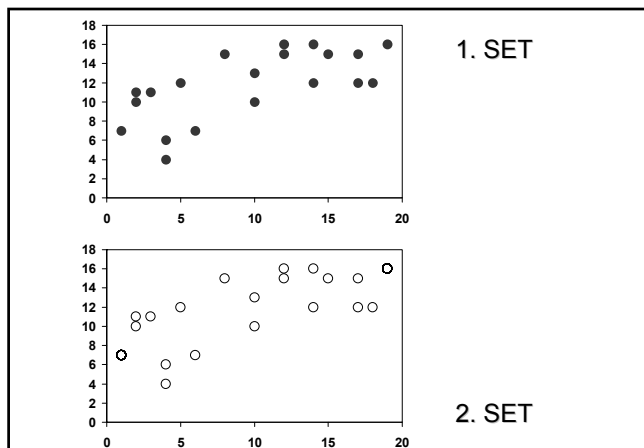
ANSCOMBOVA ČETVORKA

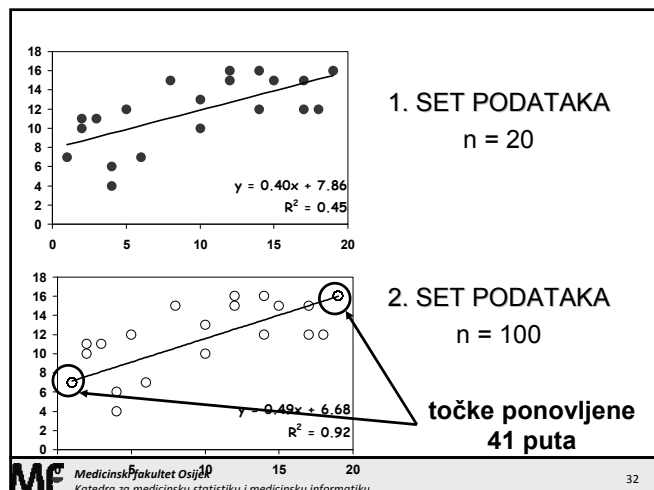
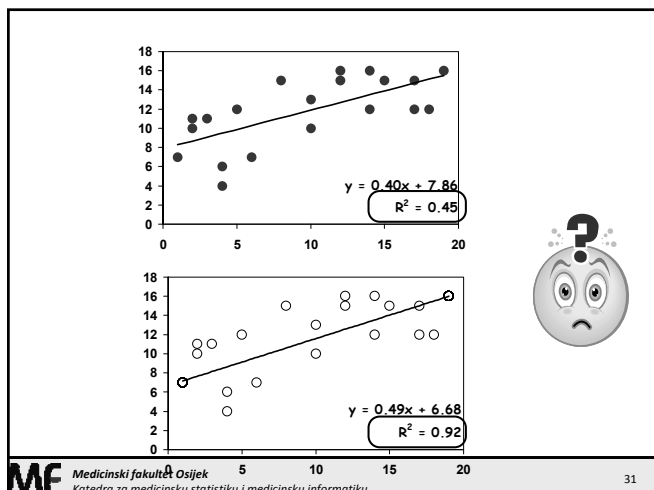
	X1	Y1	X2	Y2	X3	Y3	X4	Y4
1	10	8.04	10	9.14	10	7.46	8	6.58
2	8	6.95	8	8.14	8	6.77	8	5.76
3	13	7.58	13	8.74	13	12.74	8	7.71
4	9	8.81	9	8.77	9	7.11	8	8.84
5	11	8.33	11	9.26	11	7.81	8	8.47
6	14	9.96	14	8.1	14	8.84	8	7.04
7	6	7.24	6	6.13	6	6.08	8	5.25
8	4	4.26	4	3.1	4	5.39	19	12.5
9	12	10.84	12	9.13	12	8.15	8	5.56
10	7	4.82	7	7.26	7	6.42	8	7.91
11	5	5.68	5	4.74	5	5.73	8	6.89
\bar{X}	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
SD	3.32	2.03	3.32	2.03	3.32	2.03	3.32	2.03
r	0.82		0.82		0.82		0.82	

Anscombe FJ. *Graphs in Statistical Analysis. The American Statistician* 1973;27(1):17-21.



ZAŠTO PROMATRANJE GRAFIČKOG PRIKAZA PODATAKA NIJE UVIJEK DOVOLJNO?





1. set podataka

Rb			Rb		
r	X	Y	r	X	Y
1.	1	7	11.	10	10
2.	2	10	12.	12	15
3.	2	11	13.	12	16
4.	3	11	14.	14	12
5.	4	4	15.	14	16
6.	4	6	16.	15	15
7.	5	12	17.	17	15
8.	6	7	18.	17	12
9.	8	15	19.	18	12
10.	10	13	20.	19	16

MF Medicinski fakultet Osijek
 Katedra za medicinsku statistiku i medicinsku informatiku

2. set podataka

Rbr	X	Y	Rbr	X	Y	Rbr	X	Y	Rbr	X	Y	Rbr	X	Y
1.	1	7	21.	1	7	41.	1	7	61.	19	16	81.	19	16
2.	1	7	22.	1	7	42.	2	10	62.	19	16	82.	19	16
3.	1	7	23.	1	7	43.	2	11	63.	19	16	83.	19	16
4.	1	7	24.	1	7	44.	3	11	64.	19	16	84.	19	16
5.	1	7	25.	1	7	45.	4	4	65.	19	16	85.	19	16
6.	1	7	26.	1	7	46.	4	6	66.	19	16	86.	19	16
7.	1	7	27.	1	7	47.	5	12	67.	19	16	87.	19	16
8.	1	7	28.	1	7	48.	6	7	68.	19	16	88.	19	16
9.	1	7	29.	1	7	49.	8	15	69.	19	16	89.	19	16
10.	1	7	30.	1	7	50.	10	13	70.	19	16	90.	19	16
11.	1	7	31.	1	7	51.	10	10	71.	19	16	91.	19	16
12.	1	7	32.	1	7	52.	12	15	72.	19	16	92.	19	16
13.	1	7	33.	1	7	53.	12	16	73.	19	16	93.	19	16
14.	1	7	34.	1	7	54.	14	12	74.	19	16	94.	19	16
15.	1	7	35.	1	7	55.	14	16	75.	19	16	95.	19	16
16.	1	7	36.	1	7	56.	15	15	76.	19	16	96.	19	16
17.	1	7	37.	1	7	57.	17	15	77.	19	16	97.	19	16
18.	1	7	38.	1	7	58.	17	12	78.	19	16	98.	19	16
19.	1	7	39.	1	7	59.	18	12	79.	19	16	99.	19	16
20.	1	7	40.	1	7	60.	19	16	80.	19	16	100.	19	16

MF Medicinski fakultet Osijek
 Katedra za medicinsku statistiku i medicinsku informatiku