

TESTIRANJE RAZLIKA NUMERIČKIH PODATAKA

- razlike mjerenja neke varijable na dvije ili više skupina ispitanika -> nezavisni uzorci
- razlike dva ili više mjerenja neke varijable na istoj skupini ispitanika -> zavisni uzorci

STUDENTOV T-TEST (t-test za nezavisne uzorke)

za što se koristi:

- testiranje razlike sredina između **dvije nezavisne** skupine ispitanika

pod kojim uvjetima:

- varijabla koju testiramo mjerena je **najmanje intervalnom skalom**
- varijabla koju testiramo ima **normalnu razdiobu** u promatranim skupinama
- varijance mjerenja varijable koju testiramo u promatranim skupinama su slične (**homogenost varijanci**)

test statistika:

$$t = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{SE(\bar{x}_A - \bar{x}_B)}$$

ima Studentovu (t) razdiobu s $n_A + n_B - 2$ stupnja slobode

uz $H_0: \mu_A = \mu_B$

$$t = \frac{(\bar{x}_A - \bar{x}_B)}{SE(\bar{x}_A - \bar{x}_B)}$$

$$SE(\bar{x}_A - \bar{x}_B) = \sqrt{s_{zaj}^2 \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}$$

standardna pogreška razlike aritmetičkih sredina

$$s_{zaj}^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{(n_A - 1) + (n_B - 1)}$$

zajednička varijanca

test homogenosti varijanci (F-test):

$$F = \frac{s_A^2}{s_B^2}$$

ima F razdiobu s $n_A - 1, n_B - 1$ stupnjeva slobode

ako koristimo tablice:

- tablice za F sadrže obično samo desnu stranu distribucije \Rightarrow u **brojnik stavljamo veću varijancu**

PRIMJER

Provedeno je istraživanje razine kolesterola u ispitanika s preboljelim infarktomiokarda i zdravih ispitanika. U istraživanju je sudjelovalo 28 bolesnika i 30 kontrolnih ispitanika.

Mjerenjem su dobiveni sljedeći rezultati (mg/dl):

Bolesnici

270	236	210	250	280	272	232
220	226	242	186	266	206	282
294	282	234	224	276	282	306
260	280	278	288	288	244	236

Kontrola

196	232	200	242	206	178
184	198	160	182	182	198
182	238	198	188	166	204
182	178	212	164	230	186
162	182	218	170	200	176

Opis varijabli

Kolesterol:

- numerička varijabla, omjerna ljestvica

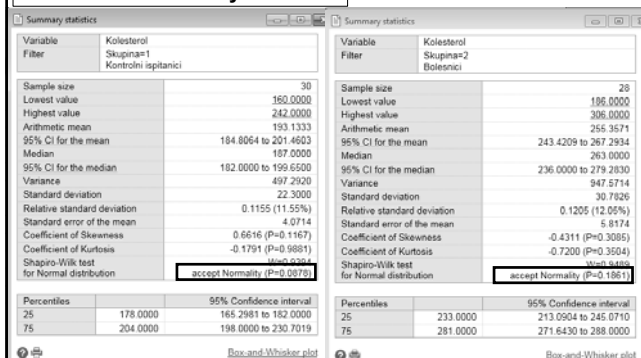
SKUPINA (Ispitivana skupina):

- kategorijska varijabla, nominalna ljestvica
- 1-kontrolni ispitanici
- 2-bolesnici

- ispitati normalnost raspodjele kolesterola u svakoj skupini
- ispitati mjere sredine, raspršenja i oblika za kolesterol u svakoj skupini

Opisna statistika i testiranje normalnosti – MedCalc (skupine posebno)

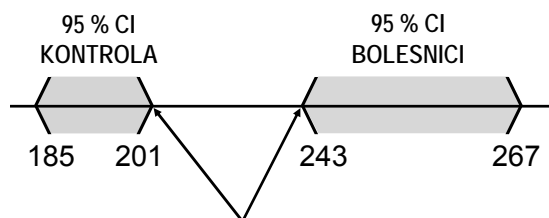
Statistics-> Summary statistics



Opisna statistika i testiranje normalnosti – MedCalc (skupine zajedno)

Statistics-> Create tables -> Summary statistics table...

Skupina	Kolesterol	
	kontrola 1	bolesnici2
N	30	28
Mean	193.133	255.357
95% CI	184.806 to 201.460	243.421 to 267.293
SD	22.3000	30.7826
Minimum	160.000	186.000
Maximum	242.000	306.000
25 - 75 P	178.000 to 204.000	233.000 to 281.000
Normal Distr.	0.0878	0.1861



nema preklapanja =>
očekujemo značajnu razliku!

Testiranje razlika – MedCalc - rezultati

Statistics->T-tests-> Independent samples t-test...

Independent samples t-test		
Sample 1		
Variable	Kolesterol	
Filter	Skupina=1 Kontrolni ispitanici	
Sample 2		
Variable	Kolesterol	
Filter	Skupina=2 Bolesnici	
	Sample 1	Sample 2
Sample size	30	28
Arithmetic mean	193.1333	255.3571
95% CI for the mean	184.8064 to 201.4603	243.4209 to 267.2934
Variance	497.2920	947.5714
Standard deviation	22.3000	30.7826
Standard error of the mean	4.0714	5.8174
F-test for equal variances	varijance su slične	P = 0.092

13

Testiranje razlika – MedCalc - rezultati

T-test (assuming equal variances)

Difference	62.2238
Standard Error	7.0233
95% CI of difference	48.1544 to 76.2932
Test statistic t	8.860
Degrees of Freedom (DF)	56
Two-tailed probability	odbaci H_0 (značajna razlika) P < 0.0001

Welch-test (assuming unequal variances)

Difference	62.2238
Standard Error	7.1006
95% CI of difference	47.9547 to 76.4930
Test statistic t(d)	8.763
Degrees of Freedom (DF)	49.0
Two-tailed probability	P < 0.0001



Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

14

Izveštavanje rezultata – u tekstu

"Vrijednosti kolesterola bile su veće u skupini bolesnika u odnosu na kontrolnu skupinu ispitanika (razlika 62 mg/dl, 95% raspon pouzdanosti razlike od 48 mg/dl do 76 mg/dl, Studentov t-test, $P < 0,001$)."

ili

"Vrijednosti kolesterola bile su veće u skupini bolesnika u odnosu na kontrolnu skupinu ispitanika (razlika 62 mg/dl, 95% raspon pouzdanosti razlike od 48 mg/dl do 76 mg/dl, $t = 8,860$, 56 stupnjeva slobode, $P < 0,001$)."



Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

15

Izveštavanje rezultata – u tablici

"Vrijednosti kolesterola bile su veće u skupini bolesnika u odnosu na kontrolnu skupinu ispitanika (razlika 62 mg/dl, 95% raspon pouzdanosti razlike od 48 mg/dl do 76 mg/dl, Studentov t-test, $P < 0,001$) (Tablica X)."

Tablica X. Vrijednost kolesterola u skupinama bolesnika i kontrolnih ispitanika

Skupina	Broj ispitanika	Kolesterol (u mg/dl)				P*
		Aritmetička sredina (standardna devijacija)	Razlika	95% raspon pouzdanosti		
				Od	Do	
Bolesnici	28	255 (31)	62	48	76	< 0,001
Kontrolna	30	193 (22)				

* Studentov t-test

za što se koristi:

- testiranje razlike sredina između **dvije zavisne** skupine ispitanika (ponavljana mjerenja na istim ispitanicima)

pod kojim uvjetima:

- razlike parova vrijednosti mjerene su **najmanje intervalnom skalom**
- razlike parova vrijednosti imaju **normalnu razdiobu**

T-TEST DIFERENCIJA (t-test za zavisne uzorke)



Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

17



Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

18

test statistika:

$$t = \frac{\bar{d} - (\mu_A - \mu_B)}{\sqrt{\frac{s_d^2}{n}}}$$

ima Studentovu (t) razdiobu s n-1 stupnjeva slobode

uz $H_0: \mu_A = \mu_B$

$$t = \frac{\bar{d}}{\sqrt{\frac{s_d^2}{n}}}$$

$$\bar{d} = \bar{x}_A - \bar{x}_B$$

srednja razlika

$$s_d^2 = \frac{\sum_i (d_i - \bar{d})^2}{n - 1}$$

varijanca razlike

PRIMJER

Ispitivan je utjecaj alkohola na vrijeme reakcije vozača. Izvršeno je mjerenje vremena reakcije 10 vozača na standardne zadatke prije i nakon konzumacije određene količine alkohola.

Mjerenjem su dobiveni sljedeći rezultati:

	prije	poslije
1	0,68	0,73
2	0,64	0,66
3	0,68	0,66
4	0,82	0,92
5	0,58	0,68
6	0,80	0,87
7	0,72	0,77
8	0,65	0,70
9	0,84	0,88
10	0,73	0,79

Opis varijabli

T1 (prije), T2 (poslije):
– numeričke, omjerna ljestvica

kreirati novu varijablu nakon-prije (T2 – T1):

- ispitati normalnost

za obje varijable (T1, T2):

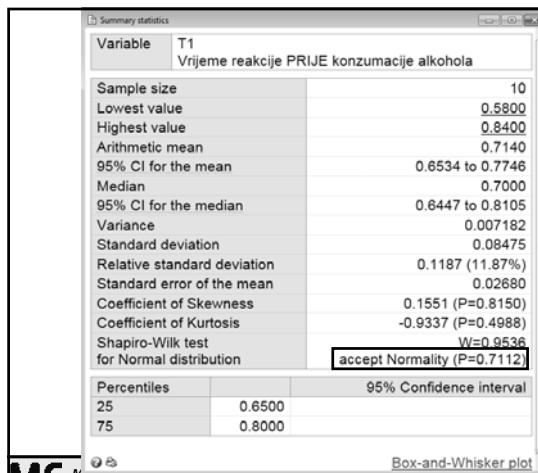
- ispitati mjere sredine, raspršenja i oblika

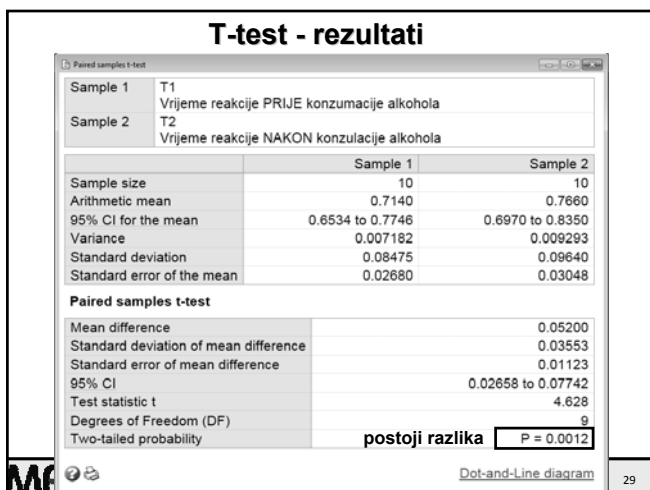
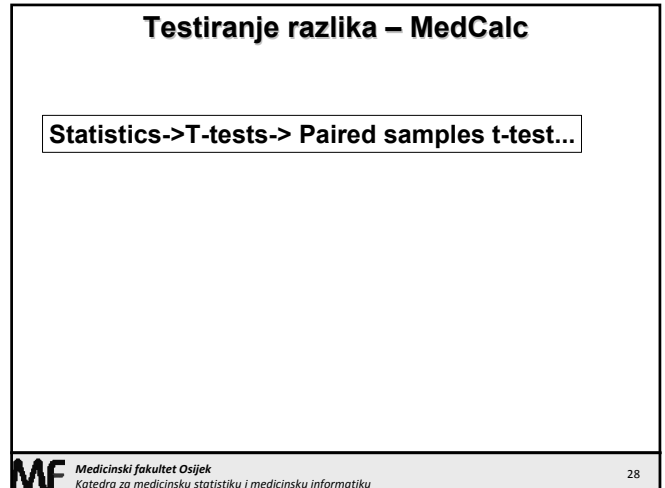
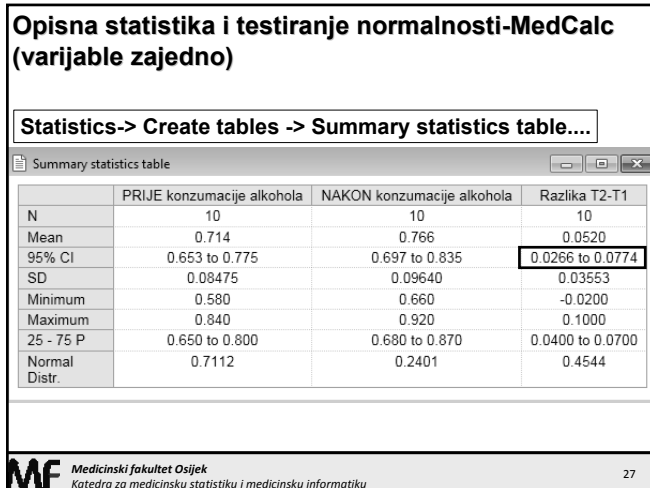
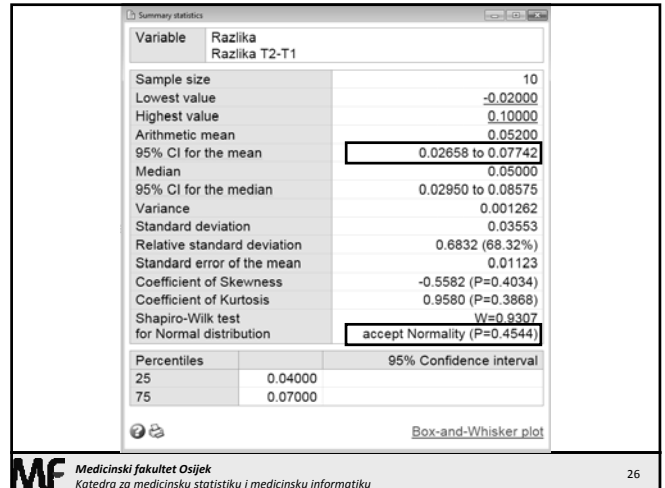
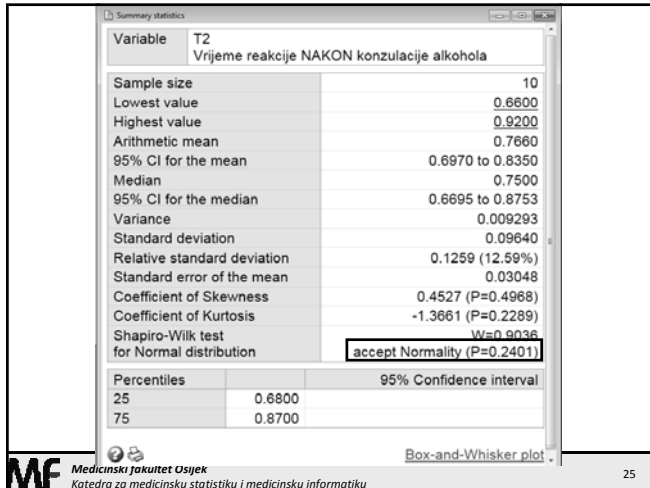
Opisna statistika i testiranje normalnosti-MedCalc (varijable posebno)

Statistics-> Summary statistics....

varijable:

- T1
- T2
- T2 – T1





Izveštavanje rezultata

"Vrijeme reakcije vozača nakon konzumacije alkohola bilo je dulje nego prije konzumacije alkohola (srednja razlika 0,052 s, 95% raspon pouzdanosti razlike od 0,027 s do 0,077 s; t-test za zavisne uzorke, P = 0,001)."

ili

"Vrijeme reakcije vozača nakon konzumacije alkohola bilo je dulje nego prije konzumacije alkohola (srednja razlika 0,052 s, 95% raspon pouzdanosti razlike od 0,027 s do 0,077 s; t = 4,628 s 9 stupnjeva slobode, P = 0,001)."

Izveštavanje rezultata – u tablici

"Vrijeme reakcije vozača nakon konzumacije alkohola bilo je dulje nego prije konzumacije alkohola (srednja razlika 0,052 s, 95% raspon pouzdanosti razlike od 0,027 s do 0,077 s; t-test za zavisne uzorke, $P = 0,001$)(Tablica X)."

Tablica X. Vrijeme reakcije vozača prije i nakon konzumacije alkohola (n = 10)

	Vrijeme reakcije			P*	
	Aritmetička sredina (standardna devijacija)	Razlika	95% raspon pouzdanosti		
			Od		Do
Prije konzumacije	0,714 (0,085)	0,052	0,027	0,077	0,001
Nakon konzumacije	0,766 (0,096)		0,027	0,077	

* T-test za zavisne uzorke

JEDNOSMJERNA ANALIZA VARIJANCE (One-way ANOVA)

za što se koristi:

- testiranje razlike između tri i više skupina

faktor

- kategorička varijabla prema kojoj su definirane skupine

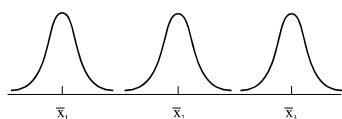
jednosmjerna analiza varijance

- analiza varijance s jednim faktorom

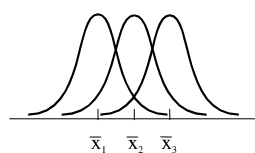
- postupak u kojem donosimo dvije procjene varijance za promatrane grupe (model):
 - procjenu koja odražava **varijabilitet između grupa**
 - procjenu koja odražava **varijabilitet unutar grupa**

OSNOVNA IDEJA:

- utvrditi je li varijabilitet **između grupa** veći od varijabiliteta **unutar grupa**



VARIJABILITET IZMEĐU GRUPE VEĆI JE OD VARIJABILITETA UNUTAR GRUPE



VARIJABILITET UNUTAR GRUPE VEĆI JE OD VARIJABILITETA IZMEĐU GRUPE

pretpostavke:

- varijabla koju testiramo mjerena je **najmanje intervalnom skalom**
- varijabla koju testiramo **ima normalnu razdiobu** u promatranim skupinama
- **varijance** mjerenja varijable koju testiramo u promatranim skupinama **su slične** (homogenost varijanci)

test statistika:

$$F = \frac{MS_{\text{tretman}}}{MS_{\text{pogreška}}}$$

procjena koja odražava varijabilitet **IZMEĐU** grupa

procjena koja odražava varijabilitet **UNUTAR** grupa

- F ima F razdiobu s k-1, N-k stupnjeva slobode

PRIMJER

Bolesnici s uznapredovalim stadijem raka želuca, bronhija i kolona tretirani su novim lijekom. Svrha istraživanja je utvrditi je li preživljavanje bolesnika povezano sa zahvaćenim organom. Vrijeme preživljavanja (u mjesecima) dano je u tablici:

želudac	bronhiji	kolon
11	9	16
6	21	19
5	4	14
7	21	43
20	16	13
7	13	23
23	8	23
7	8	21
10	12	20
21	29	19
12	12	31
18	13	28
20	6	19
	15	13
	12	10
	8	4
	16	17

- nezavisne skupine
3 skupina , ali **2 varijable:**
vrijeme
organ
1 – želudac
2 – bronhiji
3 – kolon

Opisna statistika i ispitivanje normalnosti - MedCalc:

Statistics-> Summary statistics

za sve:

Variable -> vrijeme

Test for normal distribution-> Shapiro-Wilk test

za želudac: **Filter -> organ=1**

za bronhije: **Filter -> organ=2**

za kolon: **Filter -> organ=3**

Variable	Vrijeme
Filter	Vrijeme preživljenja Organ=1 Želudac
Sample size	13
Lowest value	5.0000
Highest value	23.0000
Arithmetic mean	12.8462
95% CI for the mean	8.8617 to 16.8306
Median	11.0000
95% CI for the median	7.0000 to 20.0000
Variance	43.4744
Standard deviation	6.5935
Relative standard deviation	0.5133 (51.33%)
Standard error of the mean	1.8287
Coefficient of Skewness	0.3521 (P=0.5500)
Coefficient of Kurtosis	-1.6977 (P=0.0301)
Shapiro-Wilk test for Normal distribution	W=0.8702 accept Normality (P=0.0526)
Percentiles	95% Confidence interval
25	7.0000 5.0227 to 10.9782
75	20.0000 11.0218 to 22.9548

Variable	Vrijeme
Filter	Vrijeme preživljenja Organ=2 Bronhiji
Sample size	17
Lowest value	4.0000
Highest value	29.0000
Arithmetic mean	13.1176
95% CI for the mean	9.8920 to 16.3433
Median	12.0000
95% CI for the median	8.0161 to 15.9839
Variance	39.3603
Standard deviation	6.2738
Relative standard deviation	0.4783 (47.83%)
Standard error of the mean	1.5216
Coefficient of Skewness	0.9884 (P=0.0729)
Coefficient of Kurtosis	1.2413 (P=0.2209)
Shapiro-Wilk test for Normal distribution	W=0.9309 accept Normality (P=0.2255)
Percentiles	95% Confidence interval
25	8.0000 5.1416 to 12.0000
75	16.0000 12.5751 to 24.4336

Variable	Vrijeme
Filter	Vrijeme preživljenja Organ=3 Kolon
Sample size	17
Lowest value	4.0000
Highest value	43.0000
Arithmetic mean	19.5882
95% CI for the mean	15.0471 to 24.1293
Median	19.0000
95% CI for the median	14.0322 to 22.9678
Variance	78.0074
Standard deviation	8.8322
Relative standard deviation	0.4509 (45.09%)
Standard error of the mean	2.1421
Coefficient of Skewness	0.9783 (P=0.0755)
Coefficient of Kurtosis	2.2563 (P=0.0752)
Shapiro-Wilk test for Normal distribution	W=0.9365 accept Normality (P=0.2792)
Percentiles	95% Confidence interval
25	13.7500 7.4248 to 19.0000
75	23.0000 19.0000 to 36.1504

Jednosmjerna analiza varijance - MedCalc:

Statistics->
Anova->
One-way analysis of variance...

Data: Vrijeme
Factor codes: Organ
Post-hoc test: Student-Newman-Keuls
Significance level: 0,05

Jednosmjerna analiza varijance – MedCalc - rezultati

Data	Vrijeme
Factor codes	Vrijeme preživljenja Organ Zahvaćeni organ
Sample size	47
Levene's test for equality of error variances	
Levene statistic	0.468
DF 1	2
DF 2	44
Significance level	varijance su slične P = 0.629

Source of variation	Sum of Squares	DF	Mean Square
Between groups (influence factor)	471.5317	2	235.7659
Within groups (other fluctuations)	2399.5747	44	54.5358
Total	2871.1064	46	

F-ratio 4.323
Significance level P = 0.019

Student-Newman-Keuls test for all pairwise comparisons

Factor	n	Mean	SD	Different (P<0.05) from factor nr
(1) 1	13	12.8462	6.5935 (3)	
(2) 2	17	13.1176	6.2738 (3)	
(3) 3	17	19.5882	8.8322 (1)(2)	

najmanje jedna skupina značajno je različita od neke od preostalih

Izještavanje rezultata

"Vrijeme preživljenja se razlikovalo između skupina ispitanika s obzirom na lokalizaciju karcinoma (ANOVA, P = 0,019). Najdulje vrijeme preživljenja imali su ispitanici s karcinomom kolona (Tablica X)."

Tablica X. Vrijeme preživljenja u skupinama ispitanika s obzirom na lokalizaciju karcinoma

Lokalizacija	Broj ispitanika	Vrijeme preživljenja [Aritmetička sredina (standardna devijacija)]	P*
Želudac	13	12,8 (6,6)	0,019
Bronhiji	17	13,1 (6,3)	
Kolon	17	19,6 (8,8)†	

* jednosmjerna analiza varijance
† P < 0,05 u usporedbi sa skupinama ispitanika s karcinomom želuca i bronhija (Student-Newman-Keuls post-hoc test)

NEPARAMETRIJSKE METODE

razlozi za primjenu:

- distribucije varijabli očigledno nisu normalne
- uzorak je mali, a postoji sumnja na normalnost
- varijable su mjerene ordinalnom skalom

MODEL	PARAMETRIJSKI	NEPARAMETRIJSKI
dva nezavisna uzorka	<i>Studentov t-test</i>	<i>Mann-Whitney U test</i>
dva zavisna uzorka	<i>t-test diferencija</i>	<i>Wilcoxonov test sume rangova</i>
više nezavisnih uzoraka	<i>jednosmjerna analiza varijance</i>	<i>Kruskal - Wallis test</i>
više zavisnih uzoraka	<i>analiza varijance za ponavljana mjerenja</i>	<i>Friedmanov test (Friedmanova dvosmjerna analiza varijance rangova)</i>

MANN-WHITNEY U TEST (Wilcoxon-Mann-Whitney test)

- testira pripadaju li dva uzorka istoj populaciji (tj. imaju li iste medijane)

MedCalc:

Statistics->

Rank sum tests->

Mann-Whitney test (independent samples)...

MANN-WHITNEY U TEST – primjer za vježbu

Istraživana je vrijednost Apgar skora za djecu rođenu u bolnici i kod kuće. Vrijednosti Apgar skora bili su:

Djeca rođena u bolnici:

8, 7, 8, 6, 6, 9, 8, 5, 7, 9, 7, 8, 7, 8, 8, 8, 4, 8, 7, 9, 7, 4

Djeca rođena kod kuće:

10, 8, 9, 6, 7, 10, 9, 6, 8, 10, 8, 9, 8, 9, 3, 9, 8

WILCOXONOV TEST SUME RANGOVA

- testira hipotezu da uzorak pripada populaciji s određenom vrijednosti medijana, tj. da je medijan populacije razlika parova opažanja dvaju uzoraka jednak 0

MedCalc:

Statistics->

Rank sum tests->

Wilcoxon test (paired samples)...

WILCOXONOV TEST – primjer za vježbu

Dvanaest parova jednojajčanih blizanaca podvrgnuto je psihološkom testiranju agresivnosti s ciljem istraživanja razlika u agresivnosti između prvorođenih i drugorođenih blizanaca. Dobiveni su sljedeći rezultati:

Par	1	2	3	4	5	6	7	8	9	10	11	12
Prv.	86	71	77	68	91	72	77	91	70	71	88	87
Drug.	88	77	76	64	96	72	65	90	65	80	81	72

KRUSKAL-WALLIS TEST

- testira pripada li više uzoraka istoj populaciji

MedCalc:

Statistics->

Anova->

Kruskal-Wallis test...

KRUSKAL-WALLIS TEST – primjer za vježbu

Ispitivana je kvaliteta života u tri skupine ispitanika koji dolaze s različitih zemljopisnih područja. Kvaliteta života je ocijenjena vizualnom samoocjenjivom ljestvicom. Rezultati su kodirani u % skalnog maksimuma.

Skupina 1	80	90	90	80	70	90	90	90	90
Skupina 2	60	60	70	50	70	60	60	70	50
Skupina 3	100	100	90	100	100	90	90	90	

FRIEDMANOV TEST

(Friedmanova dvosmjerna analiza varijance rangova)

- testira pripada li više zavisnih uzoraka istoj populaciji

MedCalc:

Statistics->

Anova->

Friedman test...

FRIEDMANOV TEST – primjer za vježbu

Dvanaest slučajno izabranih studenata bilo je uključeno u pokus učenja. Svaki student je ispunjavao 4 testa znanja iz istog gradiva, jedan test za drugim. Bodovi na testu bili su (20 najviše moguće):

Student	1	2	3	4	5	6	7	8	9	10	11	12
Test 1	18	7	13	15	12	11	15	10	14	9	8	10
Test 2	4	6	14	10	11	9	16	8	12	9	6	11
Test 3	16	5	16	12	12	9	10	11	13	9	9	13
Test 4	20	10	17	14	18	16	14	16	15	10	14	16

ANALIZA POVEZANOSTI



KORELACIJA

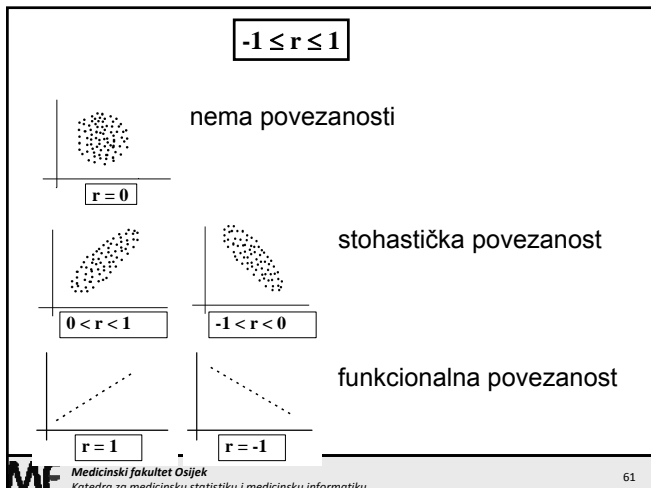
- veza među obilježjima (varijablama)
- obilježja koja "variraju zajedno"

KOEFICIJENT KORELACIJE

- mjera stupnja povezanosti

PEARSONOV KOEFICIJENT KORELACIJE r

- mjera stupnja *linearne* povezanosti dviju kvantitativnih varijabli



x, y ... nizovi 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 62

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 63

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 $df = N - 2$

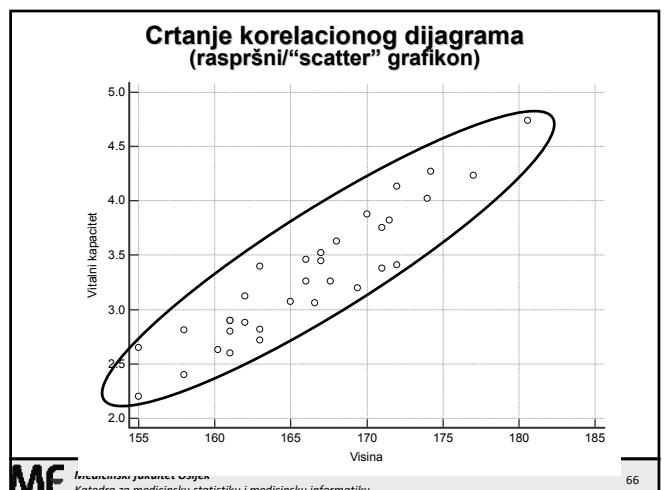
MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku 64

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

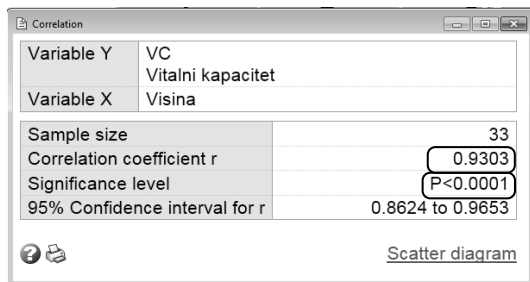
Rbr.	Visina	VC	Rbr.	Visina	VC	Rbr.	Visina	VC
1.	180,6	4,74	12.	155,0	2,20	23.	174,2	4,27
2.	168,0	3,63	13.	171,0	3,38	24.	167,0	3,45
3.	163,0	3,40	14.	171,5	3,82	25.	162,0	2,88
4.	171,0	3,75	15.	167,6	3,26	26.	172,0	4,13
5.	177,0	4,23	16.	160,2	2,63	27.	161,0	2,90
6.	169,4	3,20	17.	166,6	3,06	28.	155,0	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

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku



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

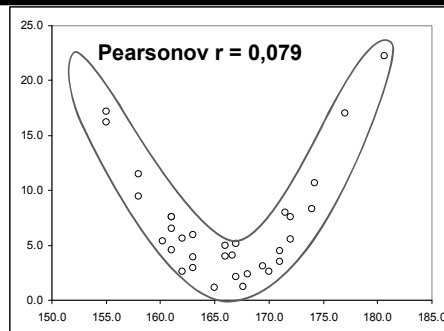
- ocjenjuje se pomoću **koeficijenta determinacije r^2**
- koliki udio varijabilnosti je "zajednički"

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!

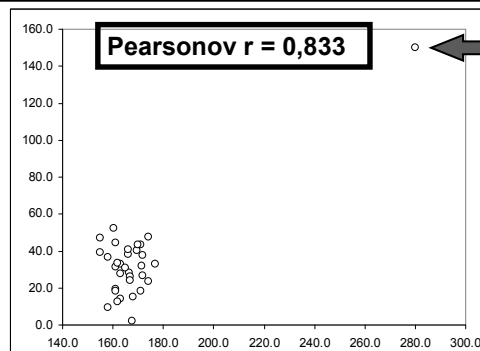


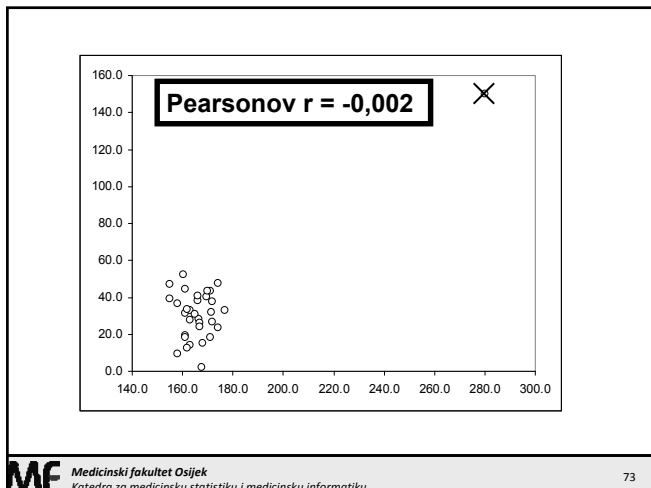
VAŽNO:

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

VAŽNO:

Na koeficijent korelacije jako utječu ekstremne vrijednosti!





SPEARMANOV KOEFICIJENT KORELACIJE ρ

- neparametrijski koeficijent korelacije

KADA?

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

"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

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

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:

- jednažba regresijskog pravca dobiva se METODOM NAJMANJIH KVADRATA

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

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

iz normalnih jednažbi

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

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

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

**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 = $-11.537 + 0.089 \times 175 = 4.04$

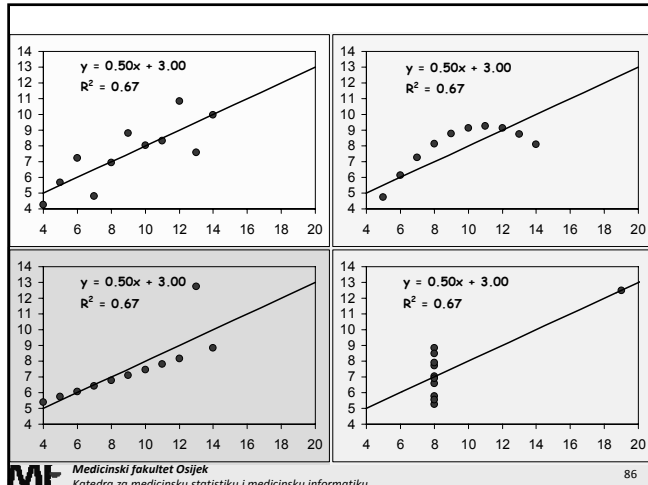
ZAŠTO MORAMO VIDJETI GRAFIČKI PRIKAZ PODATAKA?

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	

Ancombe F.J. *Graphs in Statistical Analysis. The American Statistician* 1973;27(1):17-21.

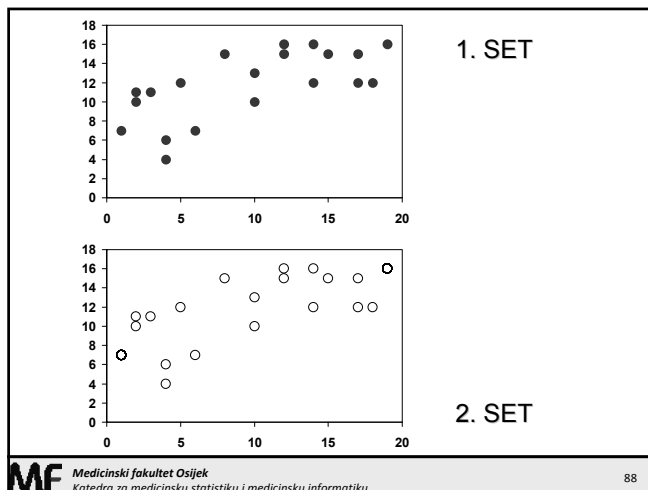
MF Katedra za medicinsku statistiku i medicinsku informatiku



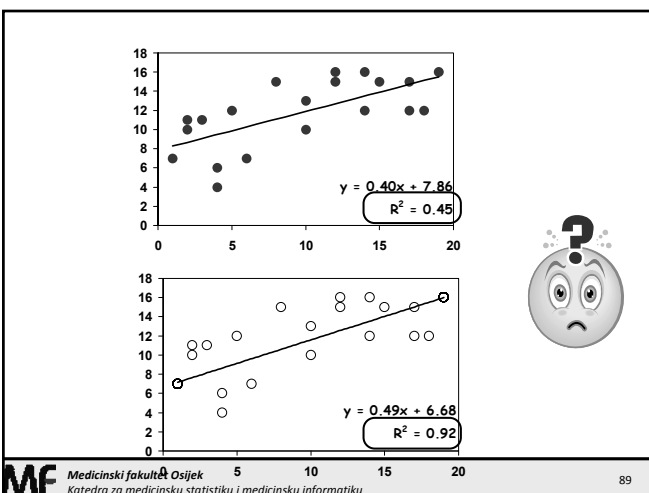
ZAŠTO PROMATRANJE GRAFIČKOG PRIKAZA PODATAKA NIJE UVIJEK DOVOLJNO?

MF Medicinski fakultet Osijek
Katedra za medicinsku statistiku i medicinsku informatiku

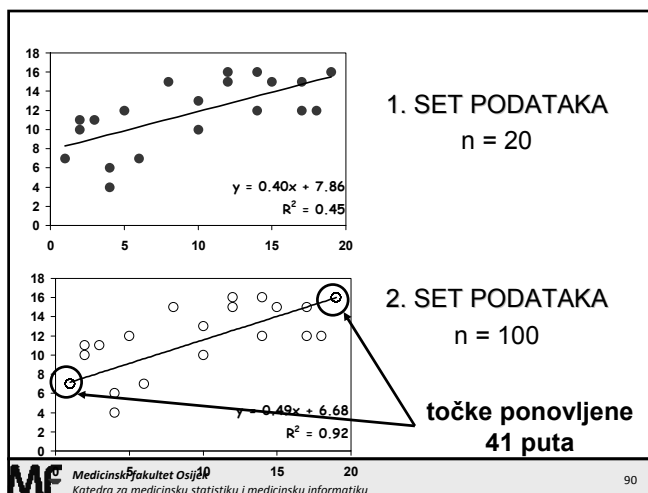
87



88



89



90

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

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