

Analiza brojčanih podataka

Nora Nikolac

Klinički zavod za kemiju
KB "Sestre milosrdnice"

Kolegij: Počela biostatistike

*Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij
„Molekularne bioznanosti“*

Statistička hipoteza - postupak testiranja

1. postavljanje hipoteze: H_0 , H_1
2. odabir statističkog testa
3. određivanje razine značajnosti: α
4. izračunavanje
5. zaključivanje: P

*Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij
„Molekularne bioznanosti“*

1. Postavljanje hipoteze

Nulta hipoteza H_0 - **nema** statistički značajne razlike među skupinama

Ukoliko se odbaci nulta hipoteza, prihvaća se alternativna hipoteza

Alternativna hipoteza H_1 - **postoji** statistički značajna razlika među skupinama

*Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij
„Molekularne bioznanosti“*

2. Odabir statističkog testa

- Za odabir testa moramo poznavati podatke koje testiramo:
 - Veličina uzorka
 - Broj skupina
 - Normalnost raspodjele
 - Parni/Neparni uzorci

*Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij
„Molekularne bioznanosti“*

2. Odabir statističkog testa

Parametrijski testovi

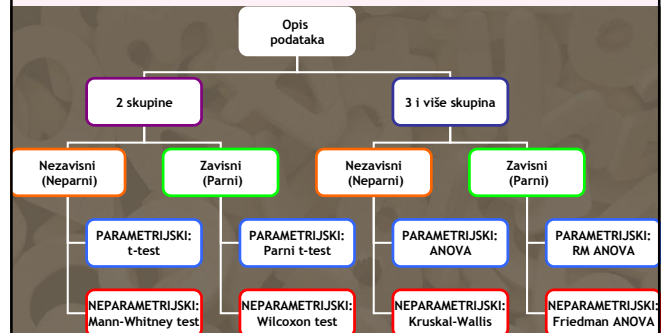
- Veliki uzorci
- Normalna distribucija

Neparametrijski testovi

- Mali uzorci
- Nema normalne distribucije ili se ne može odrediti

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

2. Odabir statističkog testa



Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

3. Određivanje razine značajnosti

- Najčešće: α - razina značajnosti: 0,05
- Kod preliminarnih istraživanja može se odabrati veća razina značajnosti: 0,1 ili 0,2
- Kod pojedinih istraživanja razina značajnosti može biti i manja: 0,01 ili 0,001

Značenje razine značajnosti: vjerojatnost odbacivanja istinite H_0 .

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

4. Izračunavanje

	E	F	G	H	I				
Dijagnoza	0	0	75	2	4,1	1,92	5		
Spol	0	1	63	1	3,4	1,1	8		
Dob	0	0	89	1	5,5	2,1	6		
Pretilost	0	1	65	3	4,2	3,5	8		
Trajanje	0	0	74	1	6,1	2,2	6		
CHOL	0	0	62	3	5,5	1,39	7		
TG	0	0	71	2	6,9	4,52	6		
HbA1c	0	1	88	2	5,6	2,72	8		
	0	1	73	1	4,2	1,73	6		
	0	0	74	1	6,1	1,3	7		
	11	122	0	1	79	2	5,6	1,6	4
	12	137	0	1	80	2	5,6	2,8	6

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

5. Zaključivanje

$P > 0,05$ (ili druga postavljena razina značajnosti)



Prihvaćamo nultu hipotezu i zaključujemo:

Ne postoji statistički značajna razlika među skupinama

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

5. Zaključivanje

$P < 0,05$ (ili druga postavljena razina značajnosti)



Odbacujemo nultu hipotezu, prihvaćamo alternativnu hipotezu i zaključujemo:

Postoji statistički značajna razlika među skupinama

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

Podaci za vježbu: IRB podaci za studente.mc1

- Provedeno je istraživanje u koje su bile uključene skupina ispitanika sa šećernom bolesti i kontrolna skupina ispitanika.
- Prikupljeni su anamnestički podaci o spolu, dobi, pretilosti i trajanju šećerne bolesti.
- Izmjereni su biokemijski parametri: koncentracije CHOL, TG, GUKn, GUKpp i HbA1c.
- Nakon 3 mjeseca ponovno su izmjerene koncentracije GUKpp i HbA1c.

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

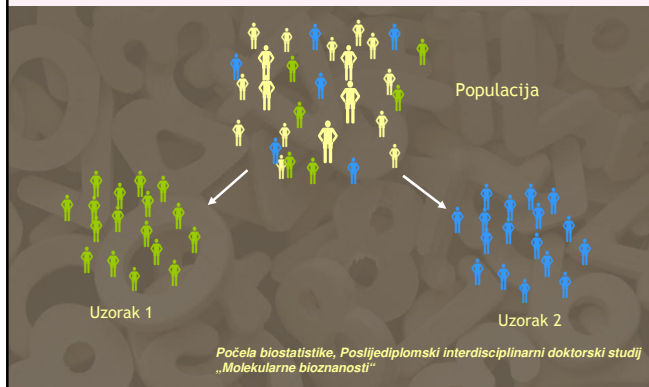
Podaci za vježbu: IRB podaci za studente.mc1

Šifriranje - Legenda:

- **Dijagnoza:** kontrolna skupina (0); šećerna bolest (1)
- **Spol:** žene (0); muškarci (1)
- **Pretilost:** normalna težina (1); umjerena pretilost (2); izrazita pretilost (3)
- **Trajanje:** <5 god. (1); 5-10 god. (2); >5 god. (3)

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

Nezavisni uzorci - dvije skupine



Nezavisni uzorci - dvije skupine

- Studentov t-test (*Statistics* → *t-tests* → *Independent t-test*) ili Mann-Whitney test (*Statistics* → *Rank sum tests* → *Mann-Whitney test*)
- Ispitati koji ćemo test upotrijebiti: vrsta podataka, normalnost raspodjele, homogenost varijanci
- Uvjeti za korištenje studentovog t-testa:
 - Varijabla je mjerena najmanje intervalnom ljestvicom
 - Normalna razdioba
 - Varijance su homogene

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij
„Molekularne bioznanosti“

Primjer #1

Iz podataka iz tablice ispitati postojanje statistički značajne razlike u koncentraciji kolesterola između bolesnika s šećernom bolesti i kontrolnih ispitanika.

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij
„Molekularne bioznanosti“

Normalnost razdiobe

- *Statistics* → *Summary statistics*
- *Variable* → CHOL
- *Options* → *Categorical variable to identify subgroups* → Dijagnoza

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij
„Molekularne bioznanosti“

Normalnost razdiobe - podskupine

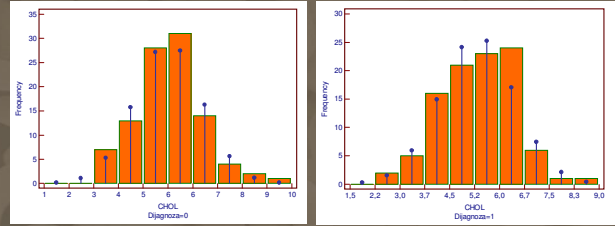
Subgroup: Dijagnoza=0			Subgroup: Dijagnoza=1		
Sample size	100		Sample size	99	
Lowest value	3,1099		Lowest value	2,7000	
Highest value	9,4000		Highest value	8,7000	
Arithmetic mean	5,5200		Arithmetic mean	5,3283	
95% CI for the mean	5,2818 to 5,7582		95% CI for the mean	5,1045 to 5,5480	
Median	5,5000		Median	5,3000	
95% CI for the median	5,3000 to 5,7000		95% CI for the median	5,0383 to 5,7617	
Variance	1,4410		Variance	1,2364	
Standard deviation	1,2004		Standard deviation	1,1120	
Relative standard deviation	0,2175 (21,75%)		Relative standard deviation	0,2088 (20,88%)	
Standard error of the mean	0,1200		Standard error of the mean	0,1118	
Coefficient of Skewness	0,4590 (P=0,0586)		Coefficient of Skewness	0,03345 (P=0,8667)	
Coefficient of Kurtosis	0,5794 (P=0,0495)		Coefficient of Kurtosis	0,30099 (P=0,1278)	
Kolmogorov-Smirnov test for Normal distribution	accept Normality (P=0,619)		Kolmogorov-Smirnov test for Normal distribution	accept Normality (P=0,766)	
Percentiles			Percentiles		
		95% Confidence Interval			95% Confidence Interval
10	4,1000	3,5000 to 4,3626	10	3,9000	3,3678 to 4,2000
25	4,8000	4,2000 to 5,1313	25	4,6250	4,2000 to 5,0000
75	6,1000	5,8000 to 6,5000	75	6,1000	5,8000 to 6,2000
90	7,2000	6,5000 to 7,6456	90	6,5200	6,2000 to 7,0483

Raspodjela je normalna ako je $P > 0,05$

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

Grafički prikaz

- Statistics → Distribution plot → Histogram
- Variable: CHOL



Select: Dijagnoza=0

Select: Dijagnoza=1

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

t-test

- Statistics → T-test → Independent t-test

Independent samples t-test

Sample 1
Variable: CHOL
Select: Dijagnoza=0

Sample 2
Variable: CHOL
Select: Dijagnoza=1

Options
 Log transformation
Correction for unequal variances: Assume equal variances

Buttons: Help, OK, Cancel

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

t-test

Ako je $P > 0,05$ varijance su homogene i možemo koristiti t-test

$P > 0,05$ Prihvaćamo nultu hipotezu i zaključujemo da nema statistički značajne razlike među skupinama.

Sample 1		Sample 2	
Variable	CHOL	Variable	Dijagnoza=0
Select	Dijagnoza=0	Select	Dijagnoza=1
Sample size	100	Sample size	99
Arithmetic mean	5,5200	Arithmetic mean	5,3283
95% CI for the mean	5,2818 to 5,7582	95% CI for the mean	5,1045 to 5,5480
Variance	1,4410	Variance	1,2364
Standard deviation	1,2004	Standard deviation	1,1120
Standard error of the mean	0,1200	Standard error of the mean	0,1118
F-test for equal variances			P = 0,449
T-test (assuming equal variances)			
Difference			-0,1937
Standard Error			0,1641
95% CI of difference			-0,5173 to 0,1298
Test statistic t			-1,164
Degrees of Freedom (DF)			197
Two-tailed probability			P = 0,2391

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

Primjer #2

Iz podataka iz tablice ispitati postojanje statistički značajne razlike u koncentraciji triglicerida između muškaraca i žena.

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

Normalnost razdiobe

- *Statistics* → *Summary statistics*
- *Variable* → *TG*
- *Options* → *Categorical variable to identify subgroups* → *Spol*

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

Normalnost razdiobe

Subgroup Spol 0		Subgroup Spol 1		
Sample size	114	Sample size	85	
Lowest value	0,8000	Lowest value	0,2000	
Highest value	10,8000	Highest value	8,2000	
Arithmetic mean	2,4887	Arithmetic mean	2,1568	
95% CI for the mean	2,1672 to 2,7702	95% CI for the mean	1,8877 to 2,4259	
Median	2,2000	Median	2,2000	
95% CI for the median	1,9032 to 2,4900	95% CI for the median	1,8000 to 2,2912	
Variance	2,8403	Variance	1,5564	
Standard deviation	1,6249	Standard deviation	1,2475	
Relative standard deviation	0,8592 (85,92%)	Relative standard deviation	0,5784 (57,84%)	
Standard error of the mean	0,1522	Standard error of the mean	0,1353	
Coefficient of Skewness	2,8284 (P<0,0001)	Coefficient of Skewness	2,5640 (P<0,0001)	
Coefficient of Kurtosis	10,4658 (P<0,0001)	Coefficient of Kurtosis	11,5303 (P<0,0001)	
Kolmogorov-Smirnov test for Normal distribution	reject Normality (P<0,001)	Kolmogorov-Smirnov test for Normal distribution	accept Normality (P=0,121)	
Percentiles		Percentiles		
	95% Confidence Interval		95% Confidence Interval	
10	1,1000	1,0171 to 1,3000	0,9000	0,7176 to 1,1671
25	1,4700	1,3000 to 1,7736	1,3150	1,0803 to 1,6308
75	2,8800	2,6185 to 3,0283	2,6000	2,3153 to 2,8175
90	4,0000	2,9721 to 5,8195	3,2000	2,8000 to 4,4790

Mann-Whitney test

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

Ispitivanje značajnosti

Mann-Whitney test (independent samples)

Sample 1
Variable: TG
Select: Spol=1

Sample 2
Variable: TG
Select: Spol=0

Help OK Cancel

P = 0,1594

Nema statistički značajne razlike u koncentraciji triglicerida između muškaraca i žena.

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

Primjer #3

Iz podataka iz tablice ispitati postoji li statistički značajna razlika u dobi između ispitanika sa šećernom bolesti i kontrolne skupine.

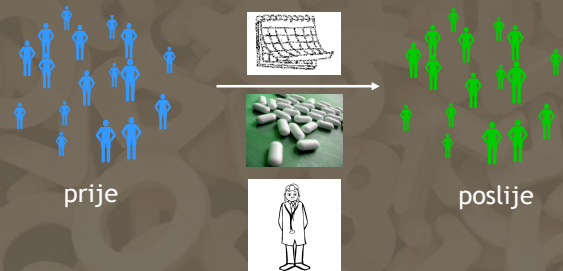
Mann-Whitney test

$P = 0,3021$

Nema razlike u dobi između ispitanika sa šećernom bolesti i kontrolne skupine.

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij
„Molekularne bioznanosti“

Zavisni uzorci - dvije skupine



Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij
„Molekularne bioznanosti“

Zavisni uzorci - dvije skupine

- **Parni t-test** (*Statistics* → *t-tests* → *Paired samples t-tests*) ili **Wilcoxonov test** (*Statistics* → *Rank sum tests* → *Wilcoxon test*)
- Ispitati koji ćemo test upotrijebiti: vrsta podataka, normalnost raspodjele
- Uvjeti za korištenje parnog t-testa:
 - Varijabla je mjerena najmanje intervalnom ljestvicom
 - Normalna razdioba

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij
„Molekularne bioznanosti“

Zavisni uzorci

- Koje biste podatke iz tablice testirali testovima za zavisne uzorke?
- HbA1c1 i HbA1c2
- GUKpp1 i GUKpp2

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij
„Molekularne bioznanosti“

Primjer #4

Iz podataka iz tablice ispitati postoji li kod bolesnika s šećernom bolesti promjena u koncentraciji HbA1c između prvog i drugog posjeta liječniku.

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

Normalnost razdiobe

Variable	HbA1c1	HbA1c2
Select	Dijagnoza=1	Dijagnoza=1
Sample size	99	99
Lowest value	4.8000	4.1000
Highest value	13.4000	10.8000
Arithmetic mean	7.5091	7.1444
95% CI for the mean	7.5806 to 8.2355	6.8892 to 7.3896
Median	7.6000	7.0000
95% CI for the median	7.3000 to 7.9000	6.7383 to 7.3617
Variance	2.8956	1.6372
Standard deviation	1.5418	1.2795
Relative standard deviation	0.2076 (20.76%)	0.1791 (17.91%)
Standard error of the mean	0.1650	0.1288
Coefficient of Skewness	0.6044 (P=0.0150)	0.5283 (P=0.2610)
Coefficient of kurtosis	0.1721 (P=0.9472)	-0.2162 (P=0.3397)
Kolmogorov-Smirnov test for Normal distribution	accept Normality (P=0.117)	accept Normality (P=0.542)
Percentiles	95% Confidence Interval	95% Confidence Interval
10	5.9000	5.5678 to 6.5000
25	6.8000	6.3874 to 7.1383
75	9.1500	8.1000 to 9.9000
90	10.2800	9.7006 to 10.7181

Parni t-test

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

Ispitivanje značajnosti

Paired samples t-test

Sample 1: HbA1c1

Sample 2: HbA1c2

Select: Dijagnoza=1

Options:
 Log transformation

Help OK Cancel

P < 0,0001

Postoji statistički značajna razlika u koncentraciji HbA1c kod bolesnika sa šećernom bolesti između prvog i drugog posjeta liječniku.

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

Primjer #5

Iz podataka iz tablice ispitati postoji li u cijeloj skupini ispitanika promjena u postprandijalnoj koncentraciji glukoze između prvog i drugog posjeta liječniku.

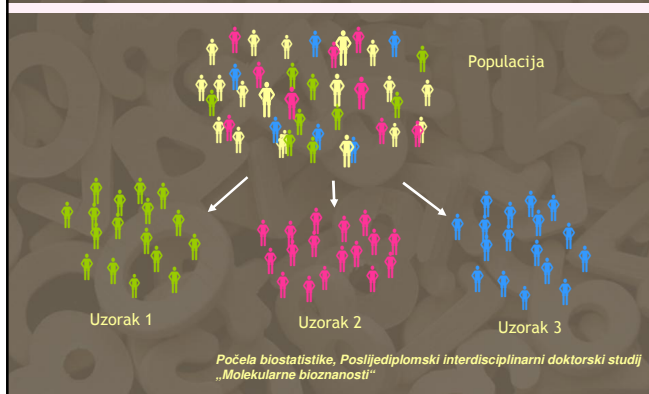
Wilcoxon test

P = 0,1871

Ne postoji statistički značajna razlika u postprandijalnoj koncentraciji glukoze između prvog i drugog posjeta liječniku.

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

Nezavisni uzorci - 3 ili više skupina



Nezavisni uzorci - 3 ili više skupina

- Jedna od najčešćih pogreški u statističkoj analizi: “svaki sa svakim”!!!
- Koristiti testove za ispitivanje više od dvije skupine!!!
- Dobivena razina značajnosti govori o postojanju razlike za “najmanje jednu skupinu”.
- *Post hoc* analiza ispituje kombinacije skupina i otkriva koja je skupina(e) odgovorna za postojanje razlike.

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij
„Molekularne bioznanosti“

Nezavisni uzorci - 3 ili više skupina

- **ANOVA** (*Statistics* → *Anova* → *One way analysis of variance*) ili **Kruskal-Wallis test** (*Statistics* → *Kruskal-Wallis test*)
- Ispitati koji ćemo test upotrijebiti: vrsta podataka, normalnost raspodjele
- Uvjeti za korištenje ANOVE:
 - Varijabla je mjerena najmanje intervalnom ljestvicom
 - Normalna razdioba
 - Varijance su homogene

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij
„Molekularne bioznanosti“

Primjer #6

Iz podataka iz tablice ispitati postoji li razlika u koncentraciji kolesterola između različitih kategorija pretilosti.

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij
„Molekularne bioznanosti“

Normalnost raspodjele

- Statistics → Summary statistics → CHOL → Options → Categorical variable to identify subgroups → Pretilost

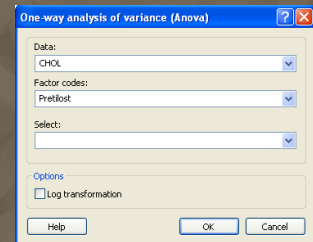
	Pretilost=1	Pretilost=2	Pretilost=3
CHOL (mmol/L)	5,2 ± 0,9	5,3 ± 1,1	5,9 ± 1,4
raspodjela	Norm.	Norm.	Norm.

ANOVA

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

ANOVA

- Statistics → Anova → One way analysis of variance



Data: CHOL
Factor codes: Pretilost

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

Izračunavanje

P>0,05; varijance su homogene i možemo koristiti ANOVU

Barem jedna skupina ima statistički značajno različitu koncentraciju kolesterola.

Post hoc testiranje:
Pretilost3 ≠ Pretilost1
Pretilost3 ≠ Pretilost2
Pretilost1 = Pretilost2

Date	CHOL		
Factor codes	Pretilost		
Sample size	199		
Levene's Test for Equality of Variances			
Levene statistic	2,219		
DF 1	2		
DF 2	196		
Significance level	P = 0,111		
ANOVA			
Source of variation	Sum of squares	D.F.	Mean square
Between groups (influence factor)	11,5720	2	5,7860
Within groups (other restrictions)	254,1270	196	1,2966
Total	265,6990	198	
F-ratio			4,483
Significance level			P = 0,013
Student-Newman-Keuls test for all pairwise comparisons			
Factor	n	Mean	Different (P<0,05) from factor nr
(1) 1	75	5,2360	(3)
(2) 2	70	5,3260	(3)
(3) 3	54	5,6130	(1)(2)

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

Zaključivanje

- Izrazito pretilo osobe (Pretilost=3) imaju veću koncentraciju kolesterola od osoba normalne težine (Pretilost=1)
- Izrazito pretilo osobe (Pretilost=3) imaju veću koncentraciju kolesterola od umjereno pretilih osoba (Pretilost=2)
- Nema razlike u koncentraciji kolesterola između umjereno pretilih osoba (Pretilost=2) i osoba normalne težine (Pretilost=1)

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij „Molekularne bioznanosti“

Primjer #7

Iz podataka iz tablice ispitati postoji li razlika u koncentraciji triglicerida u skupinama sa različitim trajanjem bolesti.

Kruskal-Wallis test

$P = 0,8431$

Nema statistički značajne razlike u koncentraciji triglicerida u skupinama sa različitim trajanjem bolesti.

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij
„Molekularne bioznanosti“

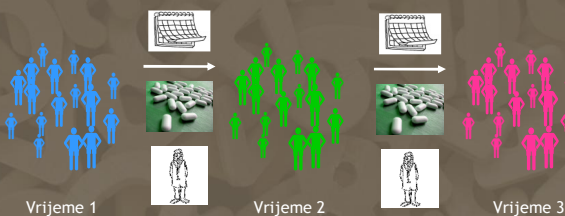
Kruskal-Wallis test



U programu MedCalc Kruskal-Wallis test ne radi *post hoc* testiranje pa ne možemo zaključivati o međusobnom odnosu skupina.

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij
„Molekularne bioznanosti“

Zavisni uzorci - 3 i više skupina



Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij
„Molekularne bioznanosti“

Zavisni uzorci - 3 i više skupina

- Parametrijski RM ANOVA ili neparametrijski Friedman ANOVA
- Program MedCalc ne nudi ove testove!
- Način testiranja isti kao i za nezavisne uzorke
- Način interpretacije isti kao i za nezavisne uzorke.

Počela biostatistike, Poslijediplomski interdisciplinarni doktorski studij
„Molekularne bioznanosti“