

Usporedba brojčanih podataka – 3 i više skupina

Nora Nikolac
Klinički zavod za kemiju
KB "Sestre milosrdnice"

Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu
Biostatistika
24.-27.11.2009.



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

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



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

1. Postavljanje statističke hipoteze

H_0 - nul-hipoteza: $\mu_1 = \mu_2 = \mu_3$
Aritmetičke sredine svih ispitivanih skupina **su jednake**

H_1 - alternativne hipoteze:

$$\mu_1 \neq \mu_2$$

$$\mu_1 \neq \mu_3$$

$$\mu_2 \neq \mu_3$$

Barem jedna aritmetičke sredina se **razlikuje** od ostalih



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

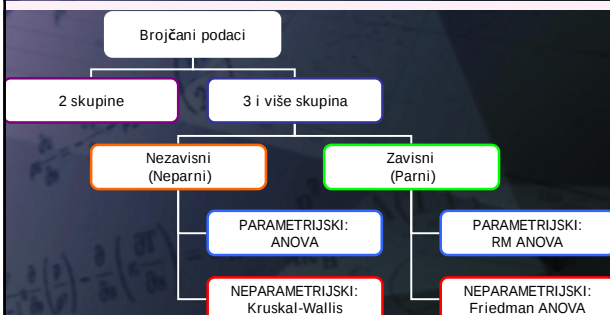
2. Odabir statističkog testa

- Za pravilan odabir testa moramo poznavati podatke koje testiramo:
 - Parni (zavisni)/Neparni (nezavisni) uzorci
 - Broj skupina
 - Veličina uzorka
 - Normalnost raspodjele
 - Homogenost varijanci



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

2. Odabir statističkog testa



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

2. Odabir statističkog testa

Parametrijski testovi

- Veliki uzorci
- Normalna distribucija
- Homogenost varijanci

Neparametrijski testovi

- Mali uzorci
- Nema normalne distribucije ili se ne može odrediti
- Varijance nisu homogene



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Nezavisni uzorci - 3 ili više skupina

Parametrijski: ANOVA
Neparametrijski: Kruskal-Wallisov test

- Uvjeti za korištenje ANOVE:
 - Kvantitativna varijabla
 - Normalna razdioba
 - Varijance su homogene

Prije odabira testa potrebno je ispitati vrstu podataka, normalnost raspodjele, homogenost varijanci.



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Zavisni uzorci - 3 ili više skupina

Parametrijski: RM ANOVA
Neparametrijski: Friedman ANOVA

- Uvjeti za korištenje RM ANOVE:
 - Kvantitativna varijabla
 - Normalna razdioba
 - Varijance su homogene

Prije odabira testa potrebno je ispitati vrstu podataka, normalnost raspodjele, homogenost varijanci.



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

3. Određivanje razine značajnosti: α

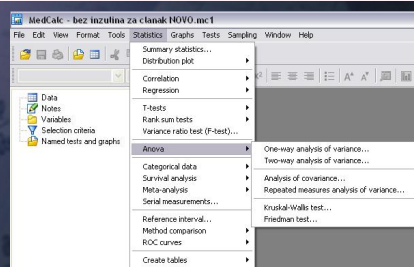
Vjerojatnost odbacivanja istinite nulte hipoteze.

... 0,05



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

4. Izračunavanje



Statistički programi



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

5. Zaključivanje

$P >$ razine značajnosti



Prihvaćamo
nultu hipotezu

$P <$ razine značajnosti



Odbacujemo
nultu hipotezu



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Primjer

Postoji li statistički značajna razlika u koncentraciji kolesterola između triju ispitivanih skupina?



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Podaci

CHOL (mmol/L)	Skupina 1	Skupina 2	Skupina 3
N	75	70	54
$\bar{X}$	5,2	5,3	5,8
SD	1,0	1,1	1,3
Medijan	5,3	5,5	5,6
IR	4,7-5,8	4,5-6,1	4,9-6,5



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Normalnost razdiobe

Ispituje se **statističkim testom**
(npr. Kolmogorov-Smirnovljev test)

Ako je $P \geq 0,05$

Raspodjela je **normalna**
(ne razlikuje se statistički
značajno od teorijske
normalne razdiobe)

Ako je $P < 0,05$

Raspodjela nije **normalna**
(razlikuje se statistički
značajno od teorijske
normalne razdiobe)



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Normalnost razdiobe

Test za ispitivanje normalnosti raspodjele:

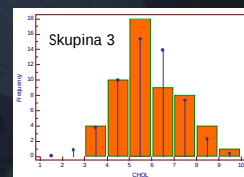
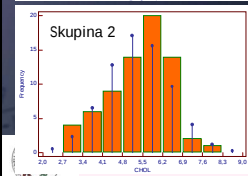
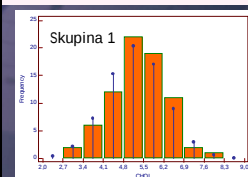
	Skupina 1	Skupina 2	Skupina 3
P	0,880	0,492	0,496
raspodjela	Norm.	Norm.	Norm.
CHOL (mmol/L)	$5,2 \pm 1,0$	$5,3 \pm 1,1$	$5,8 \pm 1,3$

Sve ispitivane razdiobe su normalne!



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Normalnost razdiobe



Sveučilišta u Zagrebu Biostatistika

Homogenost varijanci

Varianca = SD^2

Varijance u ispitivanim skupinama moraju
biti slične!

CHOL (mmol/L)	Skupina 1	Skupina 2	Skupina 3
SD	1,0	1,1	1,3
Varianca	1,0	1,2	1,7



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Homogenost varijanci

Ispituje se **statističkim testom**
(npr. Levenov test)

Ako je $P \geq 0,05$

Varijance su **homogene**

Ako je $P < 0,05$

Varijance **nisu homogene**

Naši podaci:

$P = 0,111$



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Izbor testa

- Kvantitativna varijabla
- Sve ispitivane raspodjele su normalne
- Varijance su homogene

DA

Parametrijski test:
ANOVA

NE

Neparametrijski test:
Kruskal-Wallisov test



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

ANOVA

Dobivena P vrijednost govori o postojanju statistički značajne razlike za **barem jednu** ispitivanu skupinu.

Ako je $P \geq 0,05$
Nema statistički značajne razlike među skupinama

Ako je $P < 0,05$
Barem jedna skupina se statistički značajno razlikuje od ostalih

Naši podaci: $P = 0,013$



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Post hoc testiranje

	Skupina 1	Skupina 2	Skupina 3
CHOL (mmol/L)	5,2 ± 1,0	5,3 ± 1,1	5,8 ± 1,3

Odgovara na pitanje koje se skupine međusobno razlikuju

Post hoc testiranje:

Skupina 1 = Skupina 2
Skupina 3 ≠ Skupina 1
Skupina 3 ≠ Skupina 2

Nema statistički značajne razlike u koncentraciji kolesterola između skupina 1 i 2.

Ispitanici u skupini 3 imaju statistički značajno veću koncentraciju kolesterola od ispitanika u skupini 1.

Ispitanici u skupini 3 imaju statistički značajno veću koncentraciju kolesterola od ispitanika u skupini 2.



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Primjer

Postoji li statistički značajna razlika u starosti ispitanika među različitim stadijima bolesti?



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Normalnost razdiobi

NEPARAMETRIJSKI TEST

Subgroup Stadij bolesti umjereno		Subgroup Stadij bolesti srednje		Subgroup Stadij bolesti tesko	
Sample size	80	Sample size	65	Sample size	65
Lowest value		Lowest value	35,0000	Lowest value	35,0000
Highest value		Highest value	94,0000	Highest value	94,0000
Arithmetic mean		Arithmetic mean	64,0769	Arithmetic mean	64,0769
95% CI for the mean		95% CI for the mean	61,7228 to 66,4310	95% CI for the mean	61,7228 to 66,4310
Median		Median	64,0000	Median	64,0000
95% CI for the median		95% CI for the median	62,0000 to 67,0000	95% CI for the median	62,0000 to 67,0000
Variance		Variance	90,2596	Variance	90,2596
Standard deviation		Standard deviation	9,5005	Standard deviation	9,5005
Relative standard deviation		Relative standard deviation	0,1483 (14,83%)	Relative standard deviation	0,1483 (14,83%)
Standard error of the mean		Standard error of the mean	1,1784	Standard error of the mean	1,1784
Coefficient of Skewness		Coefficient of Skewness	-0,4515 (P=0,1249)	Coefficient of Skewness	-0,4515 (P=0,1249)
Coefficient of Kurtosis		Coefficient of Kurtosis	0,2044 (P=0,3245)	Coefficient of Kurtosis	0,2044 (P=0,3245)
Kolmogorov-Smirnov test for Normal distribution		Kolmogorov-Smirnov test for Normal distribution	accept Normality (P=0,924)	Kolmogorov-Smirnov test for Normal distribution	accept Normality (P=0,924)



Farmaceutsko-biokemijski fakultet

Izračunavanje

Kruskal-Wallis test

Data: Dob

Factor codes: Stadij_bolesti

Select:

Help OK Cancel



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Zaključivanje

	Dob (god)	Umjereno	Srednje	Teško
	74 (65-88)	68 (61-74)	64 (59-71)	

Data	Dob
Factor codes	Stadij bolesti Stadij bolesti
Sample size	228
Test statistic	22.7592
Corrected for ties Ht	22.8072
Degrees of Freedom (DF)	2
Significance level	$P < 0.0001$

Factor	n	Average Rank	Different (P<0.05) from factor nr
(1) SREDNJE	83	107.43	(3)
(2) TESKO	65	90.52	(3)
(3) UMJERENO	80	141.31	(1)(2)



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Zadatak 1

Ispitali smo djelovanje lijeka za snižavanje koncentracije kolesterola. Koncentraciju kolesterola smo mjerili prije uzimanja lijeka te 3 i 6 mjeseci nakon terapije.

Kojim ćemo testom ispitati postojanje razlike:

- | | |
|------------------------------|----------------------------|
| a) Nezavisni parametrijski | c) Zavisni parametrijski |
| b) Nezavisni neparametrijski | d) Zavisni neparametrijski |



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Zadatak 2

Ispitali smo postoji li statistički značajna razlika u duljini trajanja šećerne bolesti između tri skupine bolesnika: bolesnici iz Istre, Slavonije i Dalmacije. Dobiveni podaci su prikazani u tablici:

	Skupina 1	Skupina 2	Skupina 3
N	77	104	66
SV $\pm$ SD	10 $\pm$ 6	11 $\pm$ 7	9 $\pm$ 7
M (IR)	9 (5-13)	10 (7-15)	7 (5-12)
P (normalnost)	0,270	0,120	0,002
P (varijance)	0,804		

Kojim ćemo testom ispitati postojanje razlike:

- | | |
|------------------------------|----------------------------|
| a) Nezavisni parametrijski | c) Zavisni parametrijski |
| b) Nezavisni neparametrijski | d) Zavisni neparametrijski |



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Zadatak 3

Testom ANOVA ispitivali smo razlikuje li se među studentima 1.-4. godine FBF-a prosječna duljina vremena dnevno provedenog u učenju. Dobili smo sljedeći rezultat. Što možemo zaključiti?

	1. god	2. god	3. god	4. god	P (ANOVA)
Vrijeme (min)	61 $\pm$ 12	72 $\pm$ 18	75 $\pm$ 15	92 $\pm$ 21	0,032

- | |
|--|
| a) Studenti prve godine najmanje vremena provode učeći |
| b) Studenti četvrte godine najmanje vremena provode učeći |
| c) Studenti druge i treće godine podjednako vremena provode učeći |
| d) Postoji statistički značajna razlika u vremenu provedenom u učenju među studentima 1.-4. godine |



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Pitanja

- Postupak testiranja statističke hipoteze.
- Kako glase H_0 i H_1 za ispitivanje razine značajnosti razlike između 3 skupine?
- Što je varijanca? Normalnost razdiobe.
- O čemu ovisi izbor testa za ispitivanje razine značajnosti razlike između 3 i više skupina?
- Kada koristimo parametrijske testove?
- Kada koristimo neparametrijske testove?
- Kako u odnosu na dobivenu P vrijednost statističkog testa zaključujemo o normalnosti razdiobe?
- Kako u odnosu na dobivenu P vrijednost statističkog testa zaključujemo o homogenosti varijanci?
- Kakav zaključak možemo donijeti s obzirom na dobivenu P vrijednost statističkog testa za ispitivanje razine značajnosti razlike između 3 skupine?
- Na koji način zaključujemo o postojanju razlike među podskupinama?



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika