

Usporedba brojčanih podataka - 3 i više skupina

Nora Nikolac
Klinički zavod za kemiju
KB "Sestre milosrdnice"

Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu
Biostatistika
1.-3.12.2010.



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

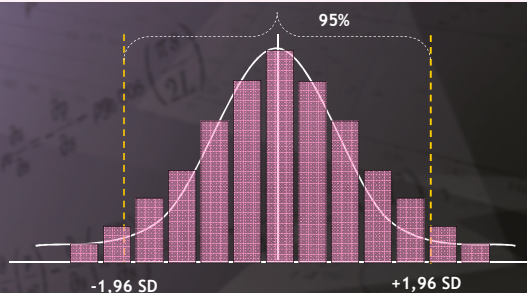
2. Odabir statističkog testa

- 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 Normalnost raspodjele



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

2. Odabir statističkog testa Normalnost raspodjele

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

2. Odabir statističkog testa

Homogenost varijanci

$$\text{Varijanca} = SD^2$$

Varijance u ispitivanim skupinama moraju biti slične!

Ispituje se statističkim testom
(npr. Levenov test)

Ako je $P \geq 0,05$

Varijance su **homogene**

Ako je $P < 0,05$

Varijance **nisu homogene**



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

2. Odabir statističkog testa

Parametrijski testovi

- Veliki uzorci
- Normalna distribucija
- Varijance homogene

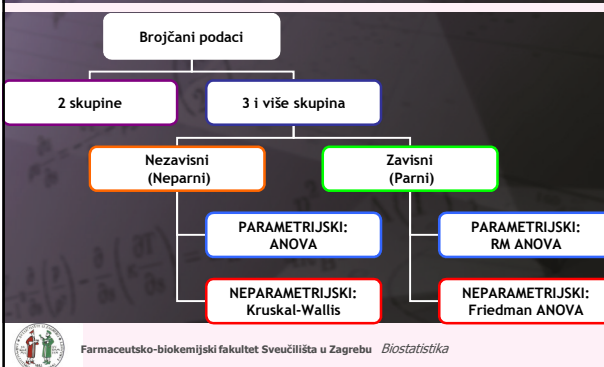
Neparametrijski testovi

- Mali uzorci ($N < 30$) ili
- Nema normalne distribucije ili se ne može odrediti ili
- Varijance nisu homogene



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

2. Odabir statističkog testa



2. Odabir statističkog testa

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

2. Odabir statističkog testa

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

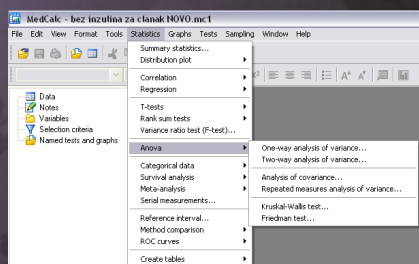
Vjerojatnost odbacivanja istinite nulte hipoteze.

... **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 (0,05)



Prihvaćamo nultu hipotezu

Nema statistički značajne razlike među skupinama

Aritmetičke sredine svih ispitivanih skupina su **jednake**.



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

5. Zaključivanje

$P <$ razine značajnosti (0,05)



Odbacujemo nultu hipotezu

Postoji statistički značajna razlika među skupinama

Barem jedna aritmetičke sredine **se razlikuje** od ostalih.



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Primjer #1

Postoji li statistički značajna razlika u koncentraciji kolesterola između skupina studenata prvih triju godina studija FBF-a?



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

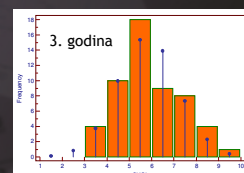
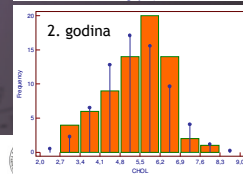
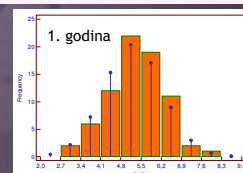
Podaci

CHOL (mmol/L)	1. godina	2. godina	3. godina
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

Ispitivanje normalnosti razdiobe



Sveučilišta u Zagrebu Biostatistika

Ispitivanje normalnosti razdiobe

$P > 0,05$ normalna

Test za ispitivanje normalnosti raspodjele:

	1. godina	2. godina	3. godina
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

Ispitivanje homogenosti varijanci

$P > 0,05$ homogene
 $P < 0,05$ nisu homogene

CHOL (mmol/L)	Skupina 1	Skupina 2	Skupina 3
SD	1,0	1,1	1,3
Varijanca	1,0	1,2	1,7
P (Lavenov test)	0,111		

Varijance ispitivanih skupina su slične -
varijance su homogene



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Izbor testa

- Kvantitativna varijabla
- $N > 30$
- 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

	1. godina	2. godina	3. godina
CHOL (mmol/L)	$5,2 \pm 1,0$	$5,3 \pm 1,1$	$5,8 \pm 1,3$

Odgovara na pitanje koje se skupine međusobno
razlikuju

Post hoc testiranje:
1. godina = 2. godina
1. godina ≠ 3. godine
2. godina ≠ 3. godine

Ispituje odnose "svaki
sa svakim"

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

Studenti 3. godine imaju statistički
značajno višu koncentraciju
kolesterola od studenata 1. godine.

Studenti 3. godine imaju statistički
značajno višu koncentraciju
kolesterola od studenata 2. godine.



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Post hoc testiranje



Samo ako postoji
statistički značajna
razlika među
skupinama!!!



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Primjer #2

Postoji li statistički značajna razlika u dobi ispitanika iz različitih gradova?



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Normalnost razdiobi (Kruskal-Wallisov test)

Subgroup Grad Dubrovnik			
Sample size		Sample size	
Lowest value		Lowest value	
Highest value		Highest value	
Arithmetic mean		Arithmetic mean	
95% CI for the mean		95% CI for the mean	
Median		Median	
95% CI for the median		95% CI for the median	
Variance		Variance	
Standard deviation		Standard deviation	
Relative standard deviation		Relative standard deviation	
Standard error of the mean		Standard error of the mean	
Coefficient of Skewness		Coefficient of Skewness	
Coefficient of Kurtosis		Coefficient of Kurtosis	
Kolmogorov-Smirnov test for Normal distribution		Kolmogorov-Smirnov test for Normal distribution	
Percentiles		Percentiles	
25	5	25	6
75	7	75	7

Subgroup Grad Zagreb			
Sample size		Sample size	
Lowest value		Lowest value	
Highest value		Highest value	
Arithmetic mean		Arithmetic mean	
95% CI for the mean		95% CI for the mean	
Median		Median	
95% CI for the median		95% CI for the median	
Variance		Variance	
Standard deviation		Standard deviation	
Relative standard deviation		Relative standard deviation	
Standard error of the mean		Standard error of the mean	
Coefficient of Skewness		Coefficient of Skewness	
Coefficient of Kurtosis		Coefficient of Kurtosis	
Kolmogorov-Smirnov test for Normal distribution		Kolmogorov-Smirnov test for Normal distribution	
Percentiles		Percentiles	
25	6	25	64.5000
75	7	75	88.0000

Subgroup Grad Osijek			
Sample size		Sample size	89
Lowest value		Lowest value	45.0000
Highest value		Highest value	88.0000
Arithmetic mean		Arithmetic mean	73.5125
95% CI for the mean		95% CI for the mean	70.7443 to 76.2807
Median		Median	74.0000
95% CI for the median		95% CI for the median	67.7608 to 79.0000
Variance		Variance	154.7340
Standard deviation		Standard deviation	12.4392
Relative standard deviation		Relative standard deviation	0.1692 (16.92%)
Standard error of the mean		Standard error of the mean	1.3907
Coefficient of Skewness		Coefficient of Skewness	-0.2764 (P=0.2924)
Coefficient of Kurtosis		Coefficient of Kurtosis	-1.1186 (P=0.0251)
Kolmogorov-Smirnov test for Normal distribution		Kolmogorov-Smirnov test for Normal distribution	reject Normality (P=0.011)
Percentiles		Percentiles	
25	6	25	58.1813 to 67.0000
75	7	75	81.0000 to 88.0000



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Izračunavanje

Kruskal-Wallis test

Data: Dob

Factor codes: Grad

Select:

Help OK Cancel



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Izračunavanje

Kruskal-Wallis test			
Data	Dob		
Factor codes	Grad		
Sample size		228	
Test statistic		22.7592	
Corrected for ties Ht		22.8072	
Degrees of Freedom (DF)			
Significance level		P < 0.0001	
Posthoc analysis:			
Factor	n	Average Rank	Different (P<0.05) from factor nr
(1) DUBROVNIK	65	90.52	(2)
(2) OSIJEK	80	141.31	(1)(3)
(3) ZAGREB	83	107.43	(2)

Multiple comparison graph

Grad	Dob (god)
Dubrovnik	64 (36-84)
Osijek	74 (45-88)
Zagreb	68 (33-2)

• Ispitanici iz Os stariji su od ispitanika iz Du.

• Ispitanici iz Os stariji su od ispitanika iz Zg.

• Između ispitanika iz Du i Zg nema razlike u dobi.

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:

	Istra	Slavonija	Dalmacija
N	77	104	66
SV ± SD (god)	10 ± 6	11 ± 7	9 ± 7
M (IR) (god)	9 (5-13)	10 (7-15)	7 (5-12)
P (normalnost)	0,270	0,120	0,002
P (varijanca)		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 ± 12	72 ± 18	75 ± 15	92 ± 21	0,032

- Studenti prve godine najmanje vremena provode učeći
- Studenti četvrte godine najmanje vremena provode učeći
- Studenti druge i treće godine podjednako vremena provode učeći
- 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