

Najčešće statističke pogriješke

doc. dr. sc. Ana-Maria Šimundić, specijalist medicinske biokemije

Klinički zavod za kemiju, Odjel za molekularnu dijagnostiku
Klinička bolnica "Sestre milosrdnice", Zagreb



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

1. POGRIJEŠKA UZORKOVANJA

- slučajni uzorak
- reprezentativan

Primjer:



cilj: ustanoviti prosječnu visinu
studenata 1. godine studija u Zagrebu

ispitivana skupina: 35 studenata ~~1. godine~~
Kineziološkog fakulteta



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

2. POGRIJEŠKA pri pripremi podataka

- provjeriti jesmo li ispravno unijeli podatke u tablice
- ispitati značajke raspodjele i načiniti deskriptivnu analizu

Primjer:



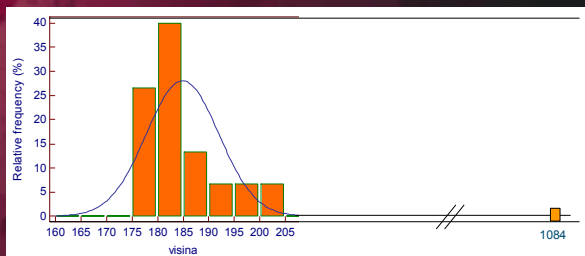
prikupili ste podatke (visina studenata 1.
godine KIF-a) i unijeli u MedCalc



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

	A	B	C	D	E	F	G	H	I	J	K
1	visina										
2	178										
3	179										
4	180										
5	181										
6	182										
7	183										
8	184										
9	185										
10	186										
11	187										
12	188										
13	189										
14	190										
15	191										
16	192										
17	193										
18	194										
19	195										
20	196										
21	197										
22	198										
23	199										
24	200										
25	201										
26	202										
27	203										
28	204										
29	205										
30	206										

kako izgleda raspodjela?



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

deskriptivna analiza ?

Variable: visina		Variable: visina	
Sample size	30	Sample size	30
Lowest value	178.0000	Lowest value	178.0000
Highest value	201.0000	Highest value	201.0000
Arithmetic mean	184.6667	Arithmetic mean	184.6667
95% CI for the mean	182.2141 to 187.5193	95% CI for the mean	182.2141 to 187.5193
Median	183.0000	Median	183.0000
95% CI for the median	180.0000 to 186.0000	95% CI for the median	180.0000 to 186.0000
Variance	29995.6713	Variance	50.4644
Standard deviation	173.1636	Standard deviation	7.1038
Relative standard deviation	0.7647 (76.47%)	Relative standard deviation	0.03843 (3.84%)
Standard error of the mean	29.9982	Standard error of the mean	1.2370
Coefficient of Skewness	5.4609 (P<0.0001)	Coefficient of Skewness	1.2657 (P=0.0063)
Coefficient of Kurtosis	29.8759 (P<0.0001)	Coefficient of Kurtosis	0.5918 (P=0.3725)
Kolmogorov-Smirnov test for Normal distribution	reject Normality (P<0.001)	Kolmogorov-Smirnov test for Normal distribution	accept Normality (P=0.108)
Percentiles		Percentiles	
2.5	178.0000	2.5	178.0000
5	178.0000	5	178.0000
10	178.0000	10	178.0000
25	179.0000	25	179.0000
75	187.0000	75	187.0000
90	200.0000	90	199.0000
95	201.0000	95	201.0000
97.5	201.0000	97.5	201.0000



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

3. POGRJEŠKA u prikazu rezultata

- ✓ podatke valja prikazati s razumnom preciznosti



Primjer:
prosječna visina studenata iznosi ~~184,3667~~ cm



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

4. POGRJEŠKA pri prikazu rezultata

- ✓ ispravan izbor mjera središnjice i rasapa



Primjer:
prosječna visina studenata iznosi:

~~$184 \pm 1,3$ cm~~
aritmetička sredina \pm SEM

184 ± 7 cm
aritmetička sredina \pm SD



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

5. POGRJEŠKA pri izboru testa

- ✓ osnovne pretpostavke za izvođenje testa moraju biti zadovoljene

Primjer:



cilj: ustanoviti razliku li se prosječne visine studenata 1. godine studija u Zagrebu i Splitu

~~parni t-test~~

t-test



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

IZBOR STATISTIČKOG TESTA OVISI O:

- ✓ pitanju na koje želimo odgovoriti
- ✓ mornoj ljestvici
- ✓ vrsti obilježja
- ✓ značajkama raspodjele
- ✓ broju skupina
- ✓ broju mjerenja



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

OSNOVNE VRSTE TESTOVA



Pitanje	Statistička analiza
Što možemo reći o populaciji temeljem podataka iz uzorka?	Opisna analiza
Razlikuje li se neko obilježje između dvije ili više skupina?	Značajnost razlike
Jesu li neka dva obilježja povezana?	Korelacija
Možemo li predvidjeti jedno obilježje s pomoću drugih?	Regresija
Postoji li razlika u preživljavanju između dvije skupine?	Analiza preživljenja



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

ZNAČAJNOST RAZLIKE

Mjerna ljestvica	intervalna	ordinalna	nominalna
Raspodjela	normalna	ne-normalna	ne-normalna
NEZAVISNE SKUPINE			
2	t-test	Mann-Whitney	Fisherov exact
≥ 3	ANOVA	Kruskal-Wallis	χ^2 test
ZAVISNE SKUPINE			
2	parni t-test	Wilcoxon test	McNemar test
≥ 3	RM ANOVA	Friedman ANOVA	Cochrane Q



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

6. POGRIJEŠNO tumačenje P vrijednosti

- ☑ nulta hipoteza: nema razlike
- ☑ P vrijednost – vjerojatnost da opažena razlika postoji uz pretpostavku da je nulta hipoteza točna

➔ statističko zaključivanje

	H_0 točna	H_1 točna
prihvaćamo H_0	ispravna odluka	pogrješka 2. vrste (β)
odbacujemo H_0	pogrješka 1. vrste (α)	ispravna odluka



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Primjer:



cilj: ustanoviti razlikuju li se prosječne visine studenata 1. godine studija u Zagrebu i Splitu

Sample 1	Split	Sample 1	Split
Variable		Variable	
Sample 2	Zagreb	Sample 2	Zagreb
Variable		Variable	
Sample size	30	Sample size	30
Arithmetic mean	175.3200	Arithmetic mean	183.0180
95% CI for the mean	172.6170 to 178.0230	95% CI for the mean	180.3919 to 185.6441
Variance	52.3984	Variance	49.4001
Standard deviation	7.2387	Standard deviation	7.0328
Standard error of the mean	1.3216	Standard error of the mean	1.2840
F-test for equal variances	P = 0.920	F-test for equal variances	P = 0.957
T-test (assuming equal variances)		T-test (assuming equal variances)	
Difference	9.5467	Difference	1.8467
Standard Error	1.9517	Standard Error	1.9251
95% CI of difference	5.8401 to 13.2532	95% CI of difference	-1.8046 to 5.5019
Test statistic t	5.155	Test statistic t	1.013
Degrees of Freedom (DF)	58	Degrees of Freedom (DF)	58
Two-tailed probability	P = 0.0001	Two-tailed probability	P = 0.3153

mala je vjerojatnost da je opažena razlika slučajna

nema dokaza da postoji razlika



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

snaga?

7. POGRIJEŠAN prikaz P vrijednosti

Visina studenata u Splitu i Zagrebu se statistički značajno razlikovala ($P=0,000001$).

Visina studenata u Splitu i Zagrebu se statistički značajno razlikovala ($P<0,05$).

Visina studenata u Splitu i Zagrebu se statistički značajno razlikovala ($P=NS$).

Visina studenata u Splitu i Zagrebu se statistički značajno razlikovala ($P<0,000$).

? ispravno



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

7. POGRIJEŠAN prikaz P vrijednosti

Visina studenata u Splitu i Zagrebu se statistički značajno razlikovala ($P=0,000001$).

Visina studenata u Splitu i Zagrebu se statistički značajno razlikovala ($P<0,05$).

Visina studenata u Splitu i Zagrebu se statistički značajno razlikovala ($P=NS$).

Visina studenata u Splitu i Zagrebu se statistički značajno razlikovala ($P<0,000$).

ispravno = $P<0,001$



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

8. POGRIJEŠKA: smislenost rezultata

- ☑ osim statističke značajnosti razlike moraju biti i smislene, odnosno klinički značajne.

Primjer:

Izmjerili smo prosječne visine studenata 1. godine studija u Zagrebu i Splitu.

Zagreb (N=10 000) 183 ± 7
Split (N=10 000) 184 ± 7
 $P<0.001$



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Primjer:

Izmjerili smo prosječnu koncentraciju glukoze natašte u mladića i djevojaka na 1. godini studija.

mladići (N=7) $6,4 \text{ mmol/L}$
djevojke (N=6) $4,6 \text{ mmol/L}$
 $P<0.085$



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Male razlike mogu biti statistički značajne, a potpuno besmislene.

ako je uzorak veliki

Velike razlike mogu biti klinički značajne, a ne moraju ujedno biti statistički značajne.

ako je uzorak mali



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

9. POGRIJEŠNO tumačenje korelacije

☒ **r ne** označava uzročno-posljedičnu vezu već samo povezanost

Primjer:

Ustanovili smo da je koncentracija glukoze natašte u pozitivnoj korelaciji s visinom studenata ($r = 0,72$ i $P = 0,003$).

Visina uzrokuje porast ~~koncentracije~~ glukoze natašte.



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

tumačenje koeficijenta korelacije



$r = 0,12$; $P = 0,003$

$r = 0,95$ i $P = 0,650$

$r = 0,75$ i $P = 0,003$

$r = 0,23$ i $P = 0,057$

$r = 0,00 - 0,25$: nema povezanosti,
 $r = 0,26 - 0,50$: slaba povezanost,
 $r = 0,51 - 0,75$: umjerena do dobra povezanost,
 $r = 0,76 - 1,00$: vrlo dobra do izvrsna povezanost.

Colton, 1974.



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

10. VIŠESTRUKO TESTIRANJE

Ako testiramo niz hipoteza, u konačnici ćemo negdje pronaći statistički značajnu razliku.



Je li ona stvarna?
Ili slučajna?

Ako izvedemo 20 testova na istom nizu podataka možemo očekivati barem jednu pogrešku tipa 1 (α).

ako je $\alpha = 0.05$



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika

Seminar



- 3 grupe
- svaka grupa dobije članak
- treba tražiti pogreške
- mogućnost dobivanja dodatnih bodova svima koji se posebno istaknu u aktivnom radu tijekom toga seminara



Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu Biostatistika