

# Testiranje hipoteze za srednju vrijednost



Izv.prof. Rosa Karlić  
Predavanje 7, MZIRuB 2024/2025

11.12.2024.

# Testiranje statističkih hipoteza

- **Statistička hipoteza** – pretpostavka o parametru populacije (može i ne mora biti istinita)
- **Nulta hipoteza** ( $H_0$ ) – hipoteza koju testiramo, uglavnom hipoteza koja govori da su naša opažanja rezultat slučajnosti
- **Alternativna hipoteza** ( $H_1$  ili  $H_a$ ) – hipoteza koja govori da na ispitanike u uzorku utječe neki ne-slučajni (non-random) uzrok
- Ispitujemo nasumičan uzorak iz populacije
- Ako su podaci iz uzorka u skladu s nultom hipotezom (na određenoj razini pouzdanosti) ne odbacujemo ju, u suprotnom – odbacujemo nultu hipotezu

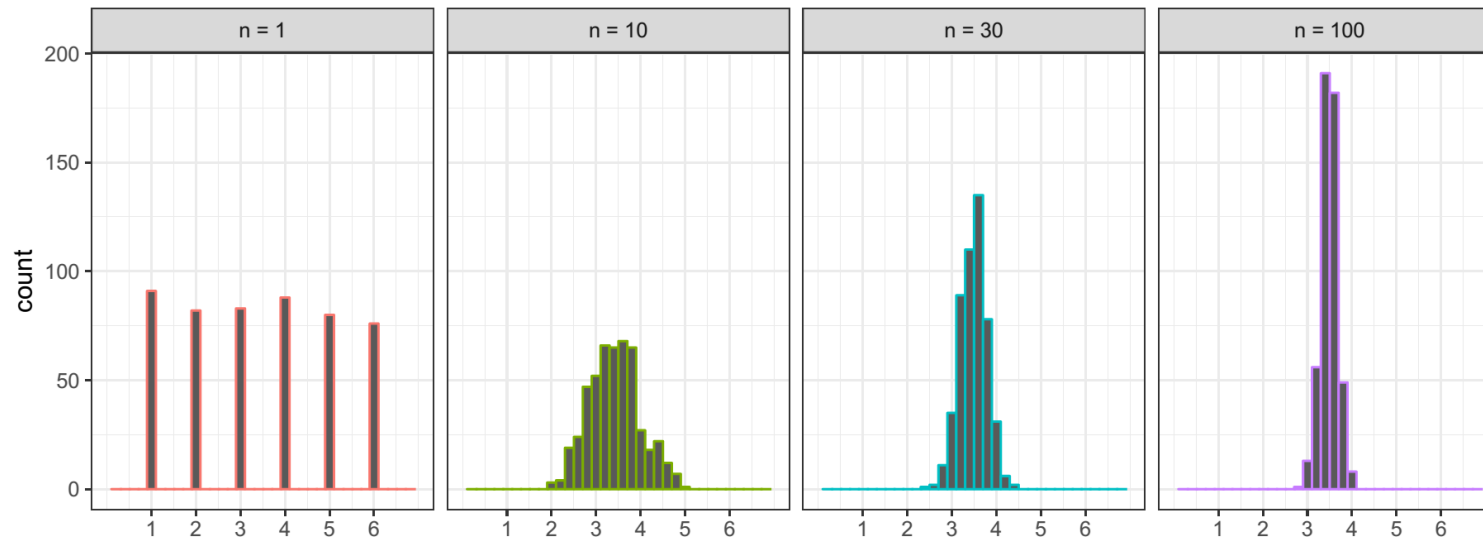
# Logička podloga statističkih testova

$$\text{test statistika} = \frac{\text{sistematska varijacija}}{\text{nesistematska varijacija}} = \frac{\text{učinak}}{\text{greška}}$$

- Poznata nam je distribucija test statistike
- To nam omogućuje da izračunamo vjerojatnost da dobijemo određenu vrijednost test statistike ukoliko je nulta hipoteza točna (P-vrijednost)

# Središnji granični teorem

- Distribucija procjene statistike (zbroja ili srednje vrijednosti) i.i.d. slučajnih varijabli biti će normalna ili gotovo normalna ukoliko je veličina uzorka dovoljno velika



Standardna devijacija statistike

$$\sigma_{\bar{x}} = \sigma / \sqrt{n}$$

Standardna greška statistike

$$SE_{\bar{x}} = s / \sqrt{n}$$

# Studentov t-test (jedan uzorak)

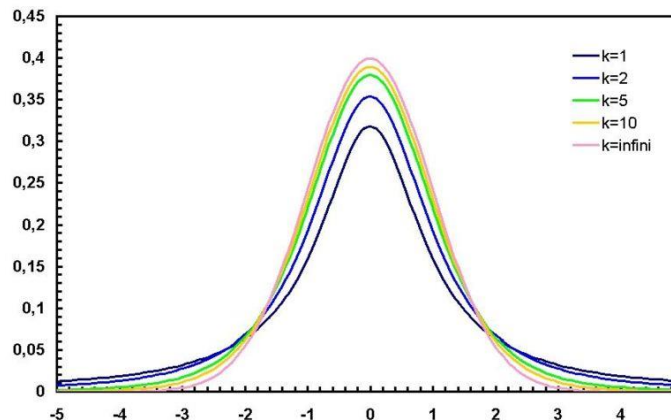
- t-statistika

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

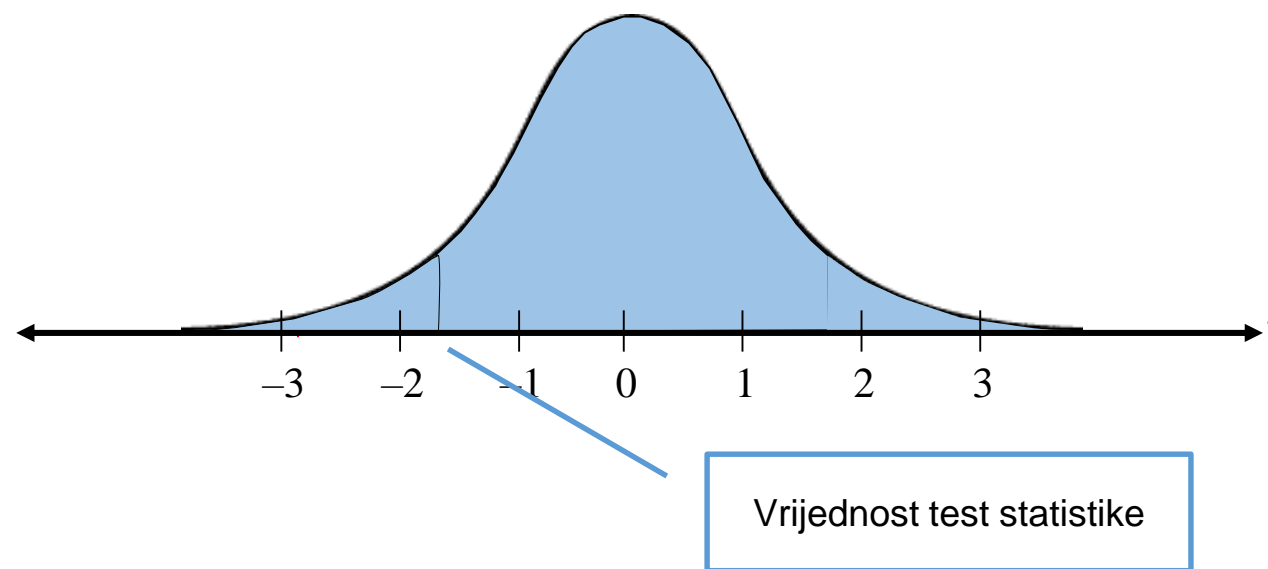
*→ učitavak*  
*→ greška*

$$SE_{\bar{x}} = s / \sqrt{n}$$

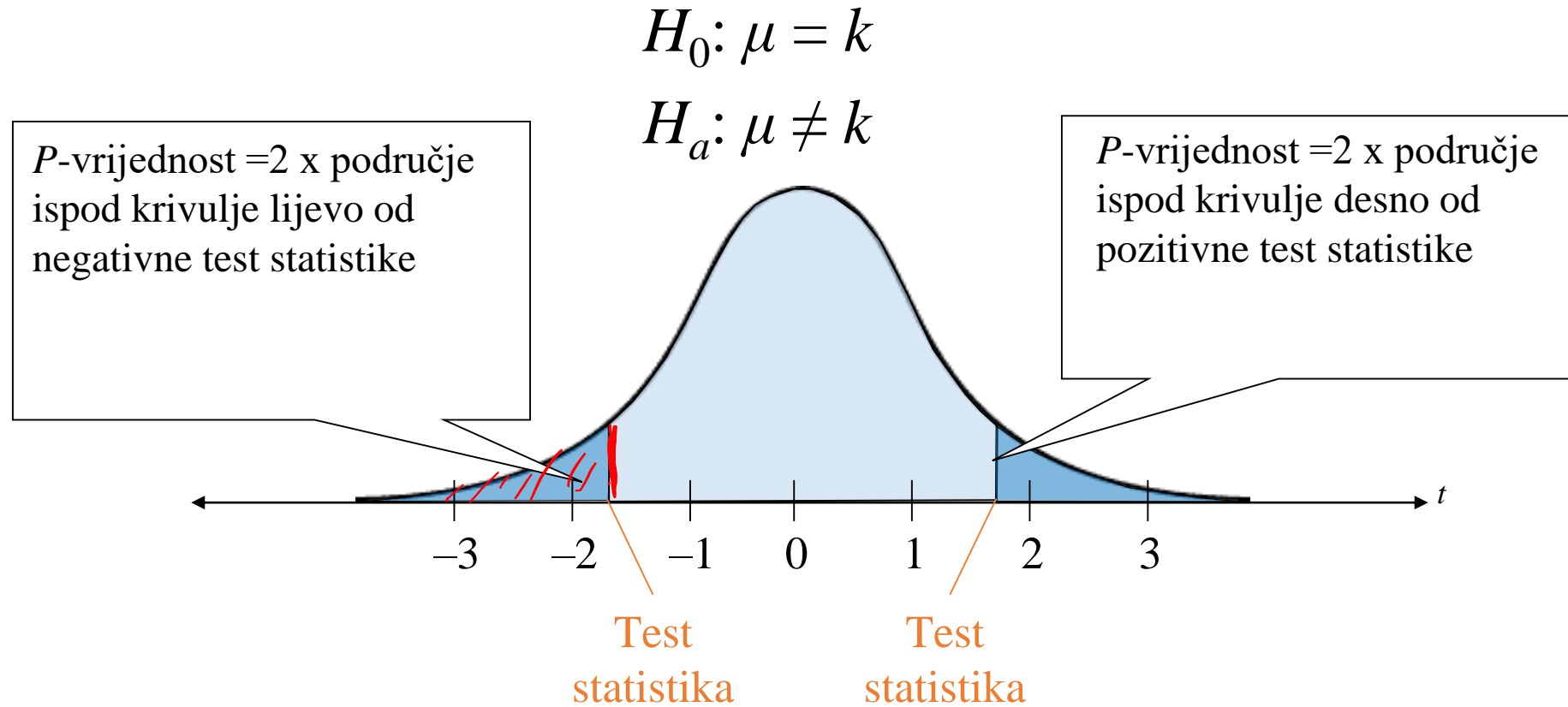
- Ukoliko je nulta hipoteza točna t-statistika će slijediti t-distribuciju sa n-1 stupnjeva slobode



$$f_{\nu}(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}$$



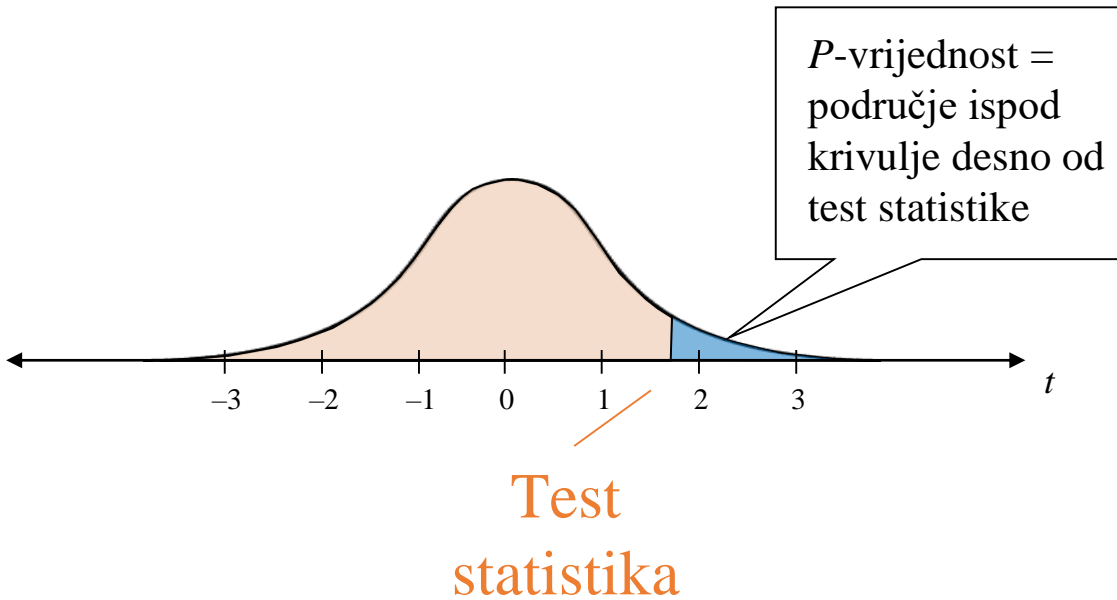
# Studentov t-test (jedan uzorak)



# Studentov t-test (jedan uzorak)

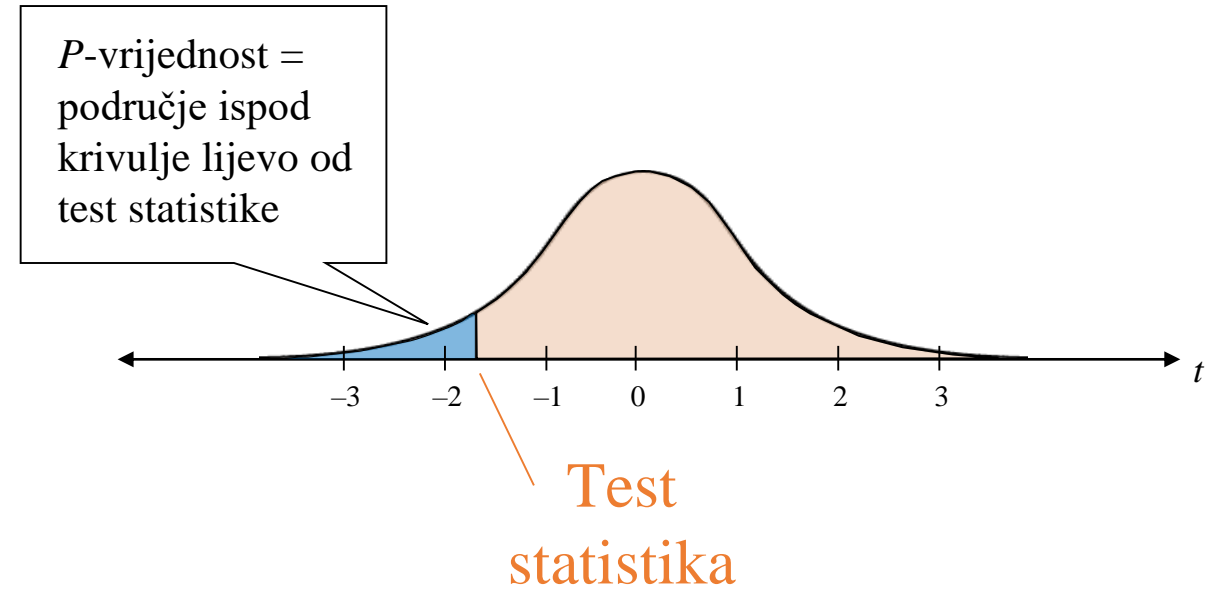
$$H_0: \mu \leq k$$

$$H_a: \mu > k$$



$$H_0: \mu \geq k$$

$$H_a: \mu < k$$



# Studentov t-test (jedan uzorak)

ID	STAROST	SPOL	TRETMAN	GEN1	GEN2	GEN3	STATUS
87	53	F	trt2	17.41	28.23	4.17	zdrav
119	53	M	ctrl	19.84	52.56	47.24	zdrav
67	52	M	trt2	19.19	27.49	12.07	zdrav
62	54	M	trt2	22.77	28.45	9.62	bolestan
131	55	F	ctrl	24.17	49.91	49.55	bolestan
50	54	F	trt1	17.15	15.32	10.67	zdrav
106	54	M	ctrl	17.92	44.95	51.39	zdrav
127	58	F	ctrl	20.06	53.19	49.71	bolestan
30	54	M	trt1	19.97	16.18	13.78	zdrav
72	54	F	trt2	25.44	27.58	11.81	zdrav

$$H_0: \mu = 25$$

$$H_a: \mu \neq 25$$

$$t_{n-1} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

One Sample t-test

```
data: data[, "GEN1"]
```

```
t = -4.3036, df = 9, p-value = 0.00198
```

```
alternative hypothesis: true mean is not equal to 25
```

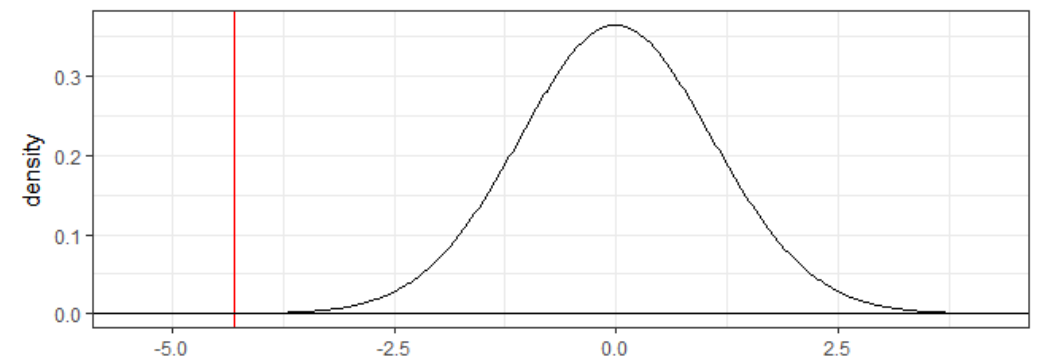
```
95 percent confidence interval:
```

```
18.64107 23.02285
```

```
sample estimates:
```

```
mean of x
```

```
20.83196
```





# Studentov t-test (dva uzorka)

ID	STAROST	SPOL	TRETMAN	GEN1	GEN2	GEN3	STATUS
87	53	F	trt2	17.41	28.23	4.17	zdrav
119	53	M	ctrl	19.84	52.56	47.24	zdrav
67	52	M	trt2	19.19	27.49	12.07	zdrav
62	54	M	trt2	22.77	28.45	9.62	bolestan
131	55	F	ctrl	24.17	49.91	49.55	bolestan
50	54	F	trt1	17.15	15.32	10.67	zdrav
106	54	M	ctrl	17.92	44.95	51.39	zdrav
127	58	F	ctrl	20.06	53.19	49.71	bolestan
30	54	M	trt1	19.97	16.18	13.78	zdrav
72	54	F	trt2	25.44	27.58	11.81	zdrav

$$t_{n_1+n_2-2} = \frac{\bar{X}_1 - \bar{X}_2}{SE_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_P^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

## Two Sample t-test

```
data: data[data$TRETMAN == "trt2", "GEN1"] and data[data$TRETMAN == "ctrl", "GEN1"]
```

```
t = 0.13371, df = 98, p-value = 0.8939
```

```
alternative hypothesis: true difference in means is not equal to 0
```

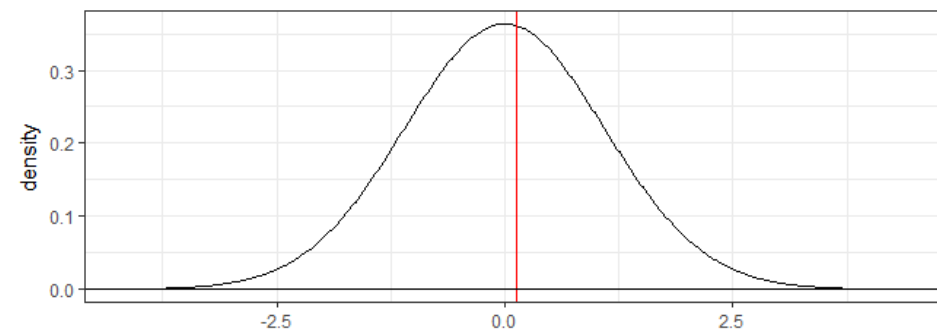
```
95 percent confidence interval:
```

```
-0.7221867 0.8265337
```

```
sample estimates:
```

```
mean of x mean of y
```

```
20.38154 20.32937
```



# Studentov t-test (dva uzorka)

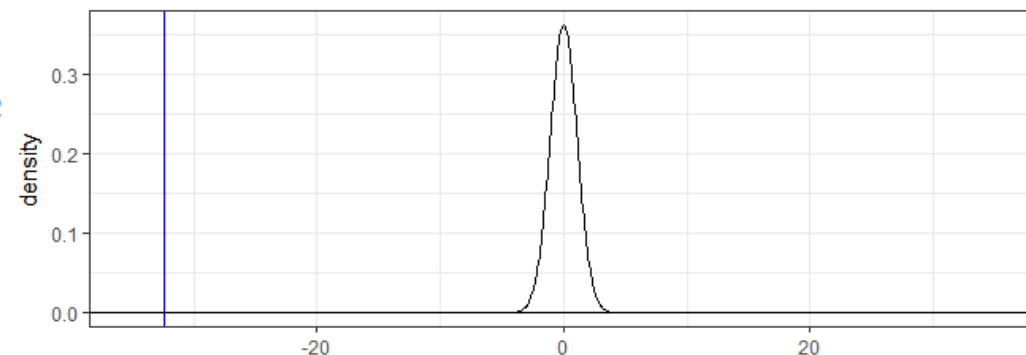
ID	STAROST	SPOL	TRETMAN	GEN1	GEN2	GEN3	STATUS
87	53	F	trt2	17.41	28.23	4.17	zdrav
119	53	M	ctrl	19.84	52.56	47.24	zdrav
67	52	M	trt2	19.19	27.49	12.07	zdrav
62	54	M	trt2	22.77	28.45	9.62	bolestan
131	55	F	ctrl	24.17	49.91	49.55	bolestan
50	54	F	trt1	17.15	15.32	10.67	zdrav
106	54	M	ctrl	17.92	44.95	51.39	zdrav
127	58	F	ctrl	20.06	53.19	49.71	bolestan
30	54	M	trt1	19.97	16.18	13.78	zdrav
72	54	F	trt2	25.44	27.58	11.81	zdrav

$$t_{n_1+n_2-2} = \frac{\bar{X}_1 - \bar{X}_2}{SE_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_P^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$s_P^2 = \frac{\sum_{i=1}^k (n_i - 1) s_i^2}{\sum_{i=1}^k (n_i - 1)} = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2 + \dots + (n_k - 1) s_k^2}{n_1 + n_2 + \dots + n_k - k}$$

Two sample t-test

```
data: data[data$TRETMAN == "trt2", "GEN2"] and data[data$TRETMAN == "ctrl", "GEN2"]
t = -32.433, df = 98, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -20.7178 -18.3287
sample estimates:
mean of x mean of y
 30.10788  49.63114
```

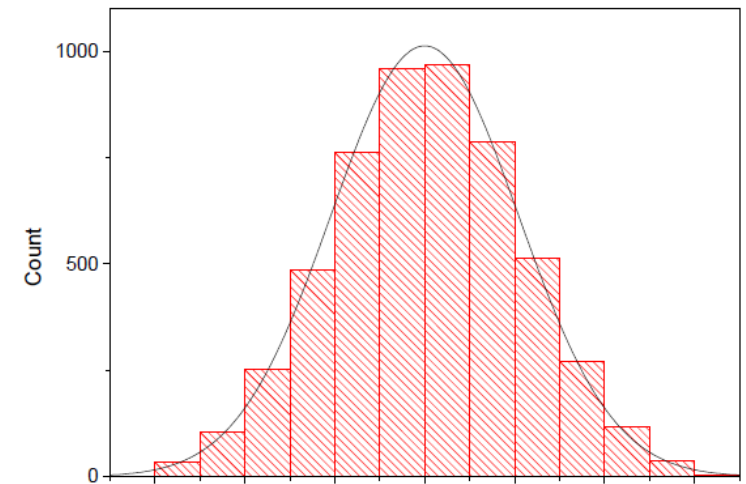


# Pretpostavke za t-test

- Varijabla koju testiramo je normalno distribuirana
- Uzorci imaju jednake varijance
- Uzorci su međusobno neovisni

## Testiranje normalnosti:

- Histogram, druge grafičke metode
- Statistički testovi (Shapiro-Wilk, Kolmogorov-Smirnov, Anderson-Darling)



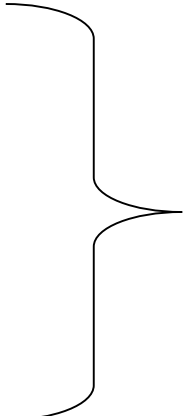
# Što ako naši podaci krše neku od pretpostavki?

- **Welch test** – koristi se u slučaju nejednakih varijanci
- Modificirana t-statistika
- Obično se koristi za uzorke jako različitih veličina
  
- **Studentov t-test za povezane uzorke** – koristi se u slučaju da su podaci u uzorcima međusobno ovisni

# Studentov t-test za povezane uzorke

- Kada uspoređujemo dva povezana uzorka često nas ne zanimaju apsolutne vrijednosti već razlika između izmjerenih vrijednosti za svaki par observacija u povezanim uzorcima

<u>Uzorak 1</u>	<u>Uzorak 2</u>	<u>Razlika</u>
$X_{11}$	$X_{21}$	$X_{11} - X_{21}$
$X_{12}$	$X_{22}$	$X_{12} - X_{22}$
·	·	·
·	·	·
·	·	·
$X_{1n}$	$X_{2n}$	$X_{1n} - X_{2n}$



- Uzorak razlika u mjerenjima uzet iz populacije razlika
- Karakterizira ga srednja vrijednost i standardna devijacija
- Srednja vrijednost razlika u mjerenjima  $\bar{D}$

# Studentov t-test za povezane uzorke

- Nulta hipoteza: Nema razlike među dva međusobno ovisna mjerenja u našem uzorku
- Zbog varijabilnosti očekujemo da će neke razlike biti pozitivne, neke negativne, ali srednja vrijednost tih razlika biti će jednaka nuli

# Studentov t-test za povezane uzorke

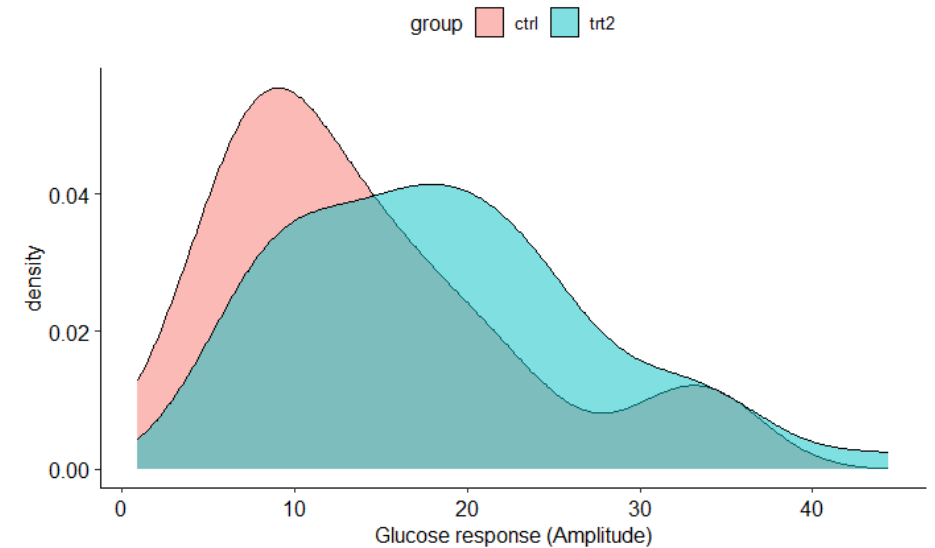
- Koristimo t-test kako bismo utvrdili je li srednja vrijednost razlika značajno različita od nule

$$t_{n-1} = \frac{\bar{D} - \mu_{\bar{D}}}{SE_{\bar{D}}}$$

- Naš uzorak je zapravo stupac s razlikama u mjerenjima na kojem ćemo napraviti t-test na jednom uzorku s nultom hipotezom da je  $\mu = 0$

# Što ako podaci nisu normalno distribuirani?

ID	AGE	SEX	TRETMAN	GEN1	GEN2	GEN3	STATUS	GLUCOSE RESPONSE (Amplitude)
87	53	F	trt2	17.41	28.23	4.17	healthy	6.95
119	53	M	ctrl	19.84	52.56	47.24	healthy	6.41
67	52	M	trt2	19.19	27.49	12.07	healthy	12.85
62	54	M	trt2	22.77	28.45	9.62	sick	7.88
131	55	F	ctrl	24.17	49.91	49.55	sick	6.05
50	54	F	trt1	17.15	15.32	10.67	healthy	9.76
106	54	M	ctrl	17.92	44.95	51.39	healthy	5.16
127	58	F	ctrl	20.06	53.19	49.71	sick	13.51
30	54	M	trt1	19.97	16.18	13.78	healthy	11.41
72	54	F	trt2	25.44	27.58	11.81	healthy	30.85





# Neparametarski testovi

- Testovi koji nemaju pretpostavki o distribuciji proučavanih varijabli
- Testiraju se rangovi podataka, a ne same vrijednosti podataka
- Mogu se koristiti za male uzorke ( $N < 30$ ) i za varijable koje ne slijede normalnu distribuciju (!)
- Manja statistička snaga testa
- **Wilcoxon-Mann-Whitney rank-sum test (Mann-Whitney U)** – usporedba dva neovisna uzorka
- **Wilcoxon signed rank test** – usporedba dva ovisna uzorka

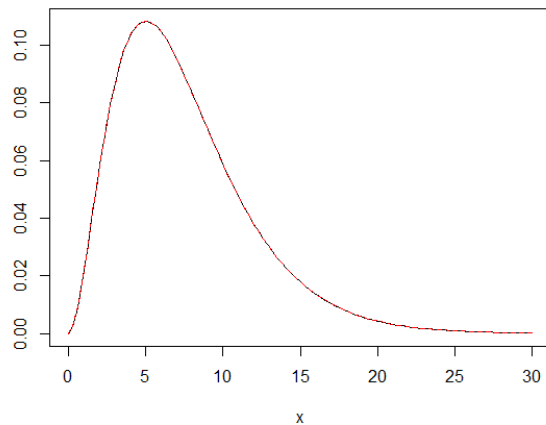
# Wilcoxon-Mann-Whitney Rank Sum Test

- Test statistika računa se na rangovima, a ne sirovim podacima
- Hipoteza
  - $H_0$  – dvije populacije imaju identičnu distribuciju
  - $H_1$  – “lokacije populacija se razlikuju”
- Pretpostavke:
  - Podaci dolaze iz nasumičnih uzoraka
  - Podaci unutar uzoraka su međusobno neovisni
  - Uzorci su međusobno neovisni

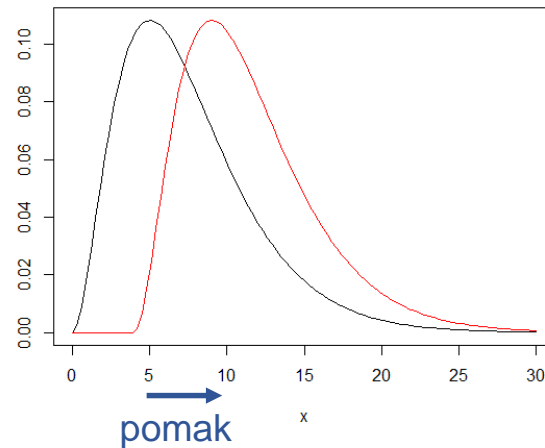
# Wilcoxon-Mann-Whitney Rank Sum Test

- Wilcoxon -Mann-Whitney Rank Sum Test pokušava otkriti pomake u lokaciji
  - $H_1 : A > B$  (A pomaknuta u desno od B)
  - $H_1 : A < B$  (A pomaknuta u lijevo od B)
  - $H_1 : A \neq B$

$H_0: A = B$

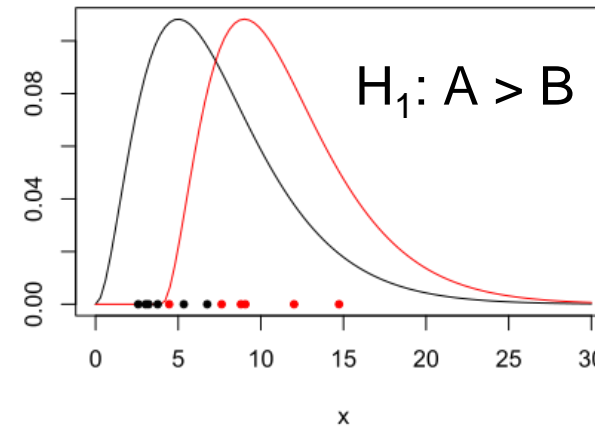
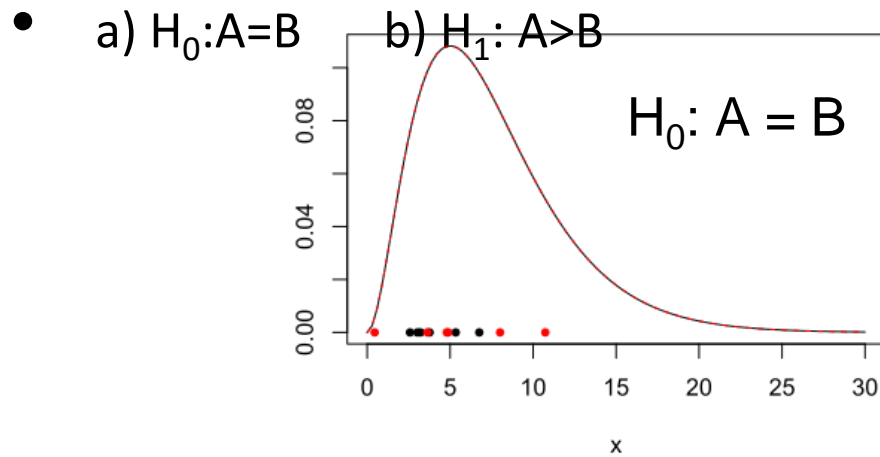


$H_1: A > B$



# Wilcoxon-Mann-Whitney Rank Sum Test

- Staviti sve observacije u jedan uzorak i rangirati  $n_A + n_B$  observacija kombiniranog uzorka
- Wilcoxon rank-sum statistika – suma rangova observacija iz jednog od uzoraka
- $w_A$  = suma rangova observacija iz uzorka A



1 2 4 6 9 10 11 12  
3 5 7 8

$(H_1 : A > B)$

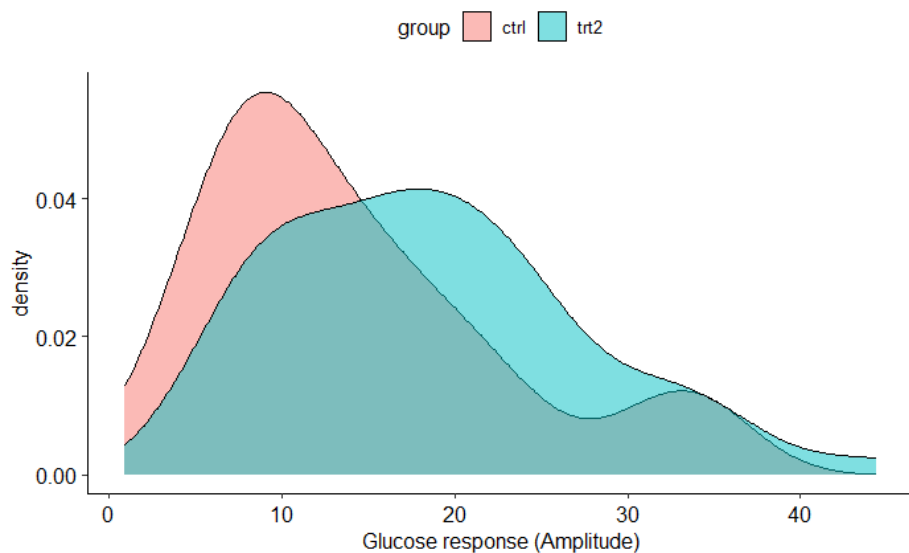
$(H_1 : A < B)$

1 2 4 6 7 9 11 12  
3 5 8 10

$P\text{-value} = \text{pr}(W_A \geq w_A).$

$P\text{-value} = \text{pr}(W_A \leq w_A).$

# Wilcoxon rank-sum test na našem primjeru



```
> wilcox.test(ctrl, trt2)
```

wilcoxon rank sum test with continuity correction

```
data: ctrl and trt2  
w = 851, p-value = 0.006011
```

```
alternative hypothesis: true location shift is not equal to 0
```

```
> t.test(ctrl, trt2, var.equal = F)
```

Two sample t-test

```
data: ctrl and trt2  
t = -2.5267, df = 98, p-value = 0.01311
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
-7.931406 -0.953410
```

```
sample estimates:
```

```
mean of x mean of y
```

```
14.17141 18.61381
```

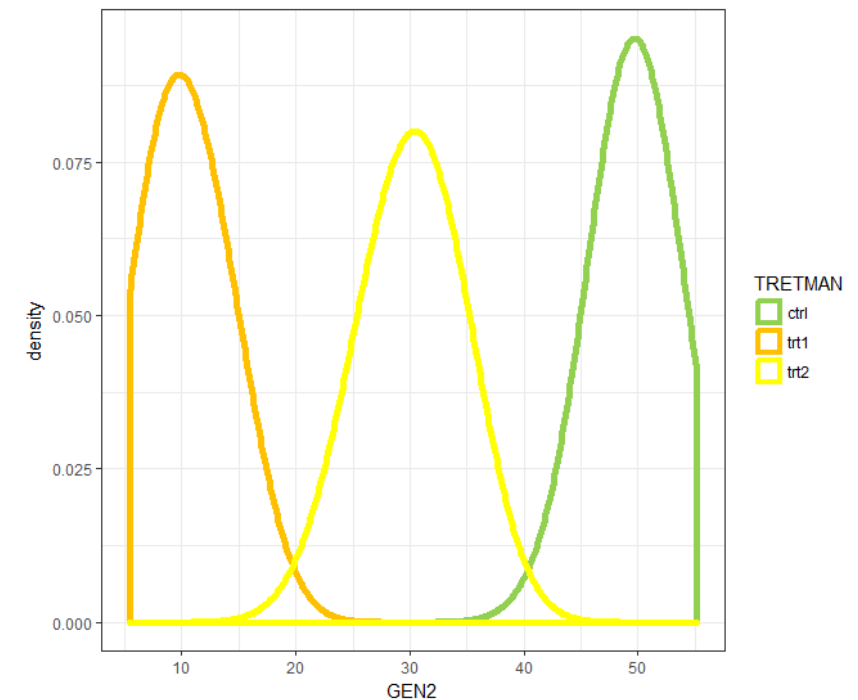
# Wilcoxon signed rank test

- Verzija prethodnog testa gdje se pojedinci mogu mjeriti dvaput ili se mogu uzeti u obzir mjerenja prije i poslije (spareni test)
- Podaci koji imaju isti broj mjerenja.
- Upareni uzorci  $(X_i, Y_i) \rightarrow D_i = X_i - Y_i$
- Ako niti jedan tretman nema učinka, tada ne samo da bi razlike trebale biti jednako raspoređene s obje strane 0, nego bi i to koliko su razlike udaljene od 0 trebale biti isto s obje strane.
- Postupak:
  - Izračunajte  $D_i$  i rangirati
  - Izračunajte  $W+$  = zbroj rangova s pozitivnim predznakom ili  $W-$  = zbroj rangova s negativnim predznakom
  - Ideja: ako je distribucija  $X$  ( $F(x)$ ) ista kao distribucija  $Y$  ( $F(y)$ ), tada je jednako vjerojatno da će  $D_i$  biti pozitivni kao i negativni. Dakle, oko pola rangova je pozitivno, a pola je negativno. ako je  $F(y)$  veći od  $F(x)$ , očekujte da većina rangova ima pozitivne predznake i stoga će  $W+$  biti velik
- Za  $n \geq 20$  približno normalno raspoređen

# ANOVA (Analiza varijance)

- Istovremeno želimo testirati više od dvije skupine ispitanika

ID	STAROST	SPOL	TRETMAN	GEN1	GEN2	GEN3	STATUS
87	53	F	trt2	17.41	28.23	4.17	zdrav
119	53	M	ctrl	19.84	52.56	47.24	zdrav
67	52	M	trt2	19.19	27.49	12.07	zdrav
62	54	M	trt2	22.77	28.45	9.62	bolestan
131	55	F	ctrl	24.17	49.91	49.55	bolestan
50	54	F	trt1	17.15	15.32	10.67	zdrav
106	54	M	ctrl	17.92	44.95	51.39	zdrav
127	58	F	ctrl	20.06	53.19	49.71	bolestan
30	54	M	trt1	19.97	16.18	13.78	zdrav
72	54	F	trt2	25.44	27.58	11.81	zdrav



# ANOVA (Analiza varijance)

- F-statistika – testira omjer varijabilnosti između skupina i unutar skupina

```
      Df Sum Sq Mean Sq F value Pr(>F)
TRETMAN  2  38784   19392    2207 <2e-16 ***
Residuals 147   1292     8.789     9.000 0.0001 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Post-hoc testovi – ako želimo odrediti između kojih skupina postoji statistički značajna razlika

- Tukey's HSD test**

```
Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = GEN2 ~ TRETMAN, data = data)

$TRETMAN
      diff      lwr      upr p adj
trt1-ctrl -39.38705 -40.79070 -37.98339 0
trt2-ctrl -19.52325 -20.92691 -18.11960 0
trt2-trt1  19.86379  18.46014  21.26745 0
```



# Pretpostavke za korištenje ANOVE

- Varijabla koju testiramo je normalno distribuirana
- Uzorci imaju jednake varijance
- Uzorci su međusobno neovisni
  
- Neparametarske alternative:
- **Kruskal-Wallis test** – više od dva neovisna uzorka
- **Friedmanov test** – više od dva ovisna uzorka

# Varijante ANOVE

- Proučavanje efekta dvije nezavisne kategoričke varijable
  - Interakcije među varijablama
- Analiza varijance ponovljenih uzoraka
  - Manji troškovi
  - Može dovesti do veće snage testa, jer varijabilnost unutar ispitanika može biti sistematska

# Two-way ANOVA (dvofaktorska ANOVA)

- 150 ispitanika: 50 ctrl, 50 trt1, 50 trt2
- 50% ispitanika na posebnom režimu prehrane
- Zavisna varijabla: Ekspresija gena GEN2
- Postoji li utjecaj tretmana?  
Postoji li utjecaj prehrane?
- Two-way (dvofaktorski) dizajn

ID	STAROST	SPOL	TRETMAN	GEN1	GEN2	GEN3	STATUS	DIET
144	55	F	ctrl	18.17	58.79	50.80	bolestan	NO
111	54	M	ctrl	20.56	48.08	49.61	bolestan	YES
113	56	F	ctrl	19.66	45.08	51.40	bolestan	NO
11	53	F	trt1	18.69	-0.96	6.00	zdrav	NO
91	56	F	trt2	19.41	30.05	6.03	bolestan	NO
64	56	M	trt2	18.95	11.28	11.71	zdrav	YES
63	50	F	trt2	21.84	36.94	11.26	zdrav	NO
149	50	M	ctrl	17.85	55.66	50.29	zdrav	NO
143	53	M	ctrl	18.71	48.40	45.89	bolestan	YES
38	55	M	trt1	23.62	9.71	14.97	bolestan	YES

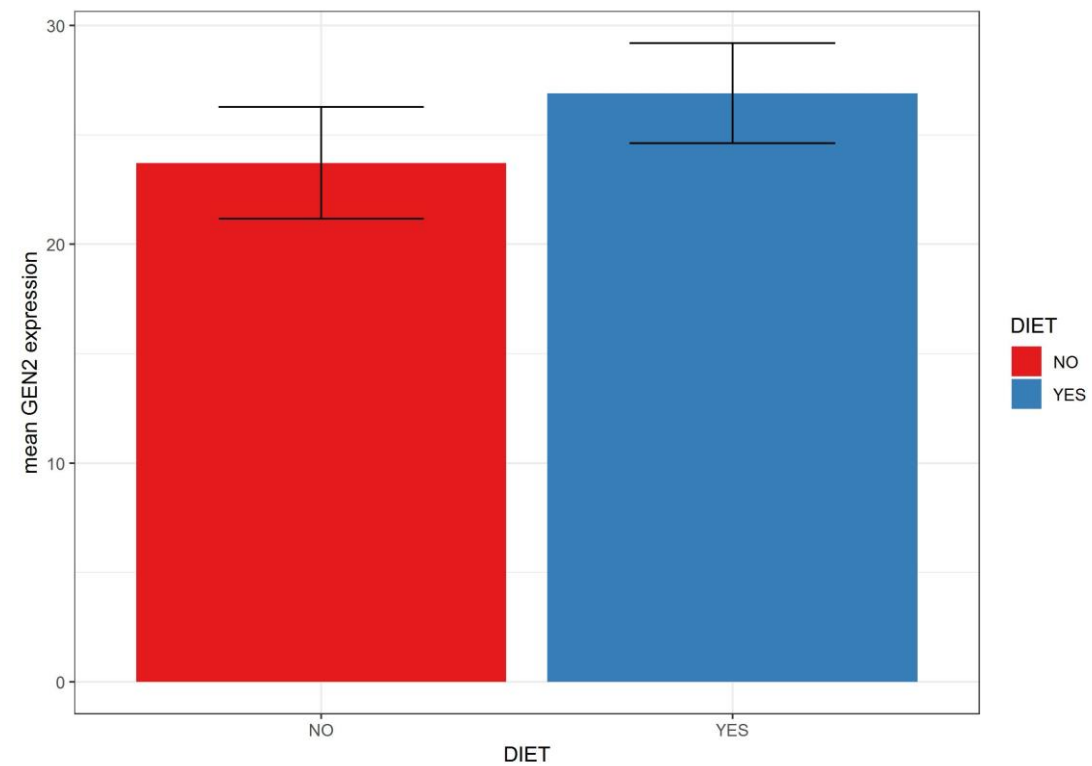
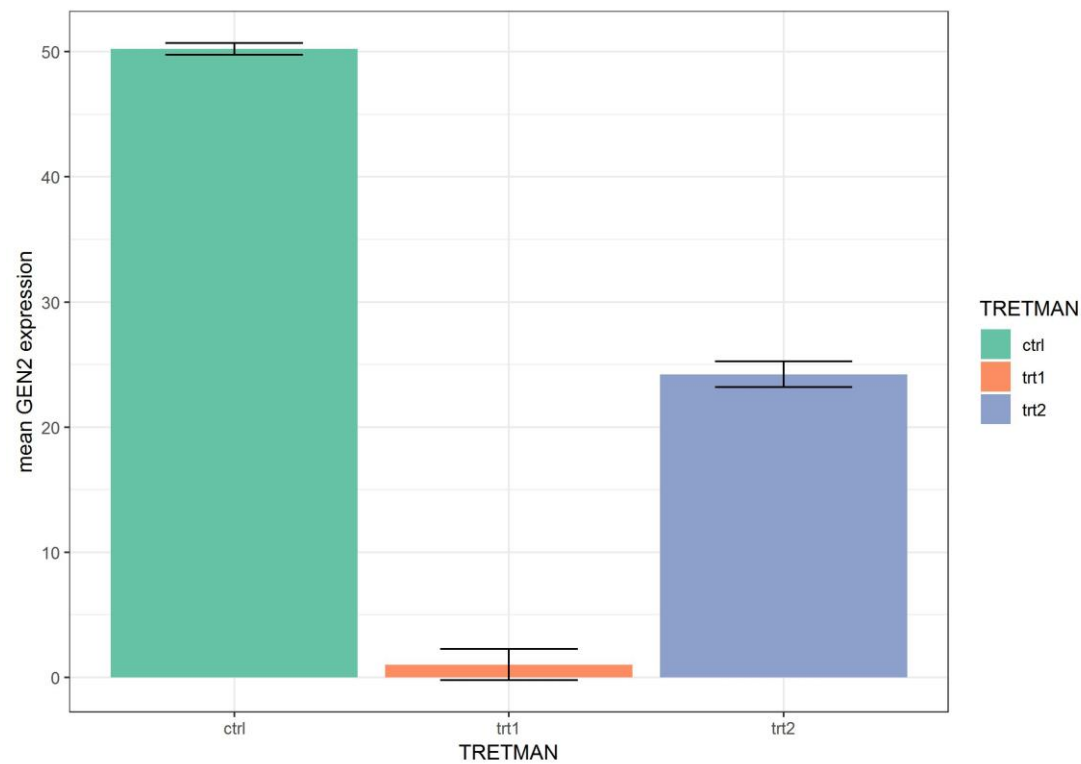
# Testiranje hipoteza

- Simultano testiramo 3 hipoteze:
  - Postoji li učinak tretmana na ekspresiju gena GEN2?
  - Postoji li učinak prehrane na ekspresiju gena GEN2?
  - Ovisi li učinak tretmana na ekspresiju gena GEN2 o prehrani (postoji li interakcija?)
  
- 3 različita učinka
  - Glavni učinak
  - Interakcijski učinak
  - Jednostavni učinak

# Naknadni testovi

- Naknadni testovi za značajne glavne učinke (post-hoc testovi)
- Naknadni testovi za značajne interakcijske učinke – analiza jednostavnih učinaka (učinak jedne varijable na svakoj razini druge varijable )

# Glavni učinci



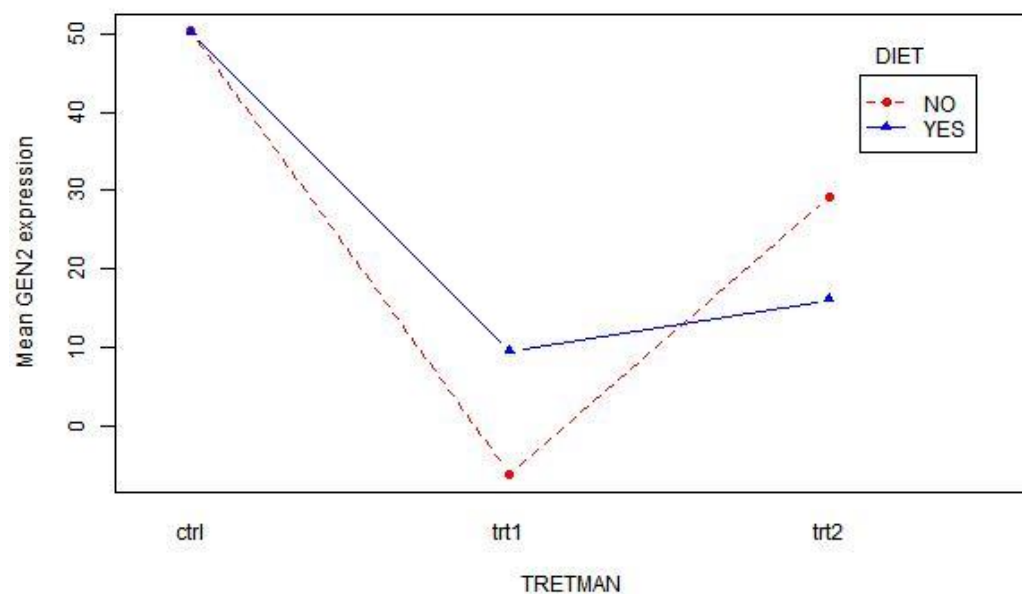
# Prikazi interakcija

## Naš primjer

```
> my.aov = aov(GEN2 ~ TRETMAN + DIET + TRETMAN:DIET,  
+ data = data)  
> summary(my.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
TRETMAN	2	60620	30310	2393.450	<2e-16 ***
DIET	1	43	43	3.374	0.0683 .
TRETMAN:DIET	2	5110	2555	201.754	<2e-16 ***
Residuals	144	1824	13		

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

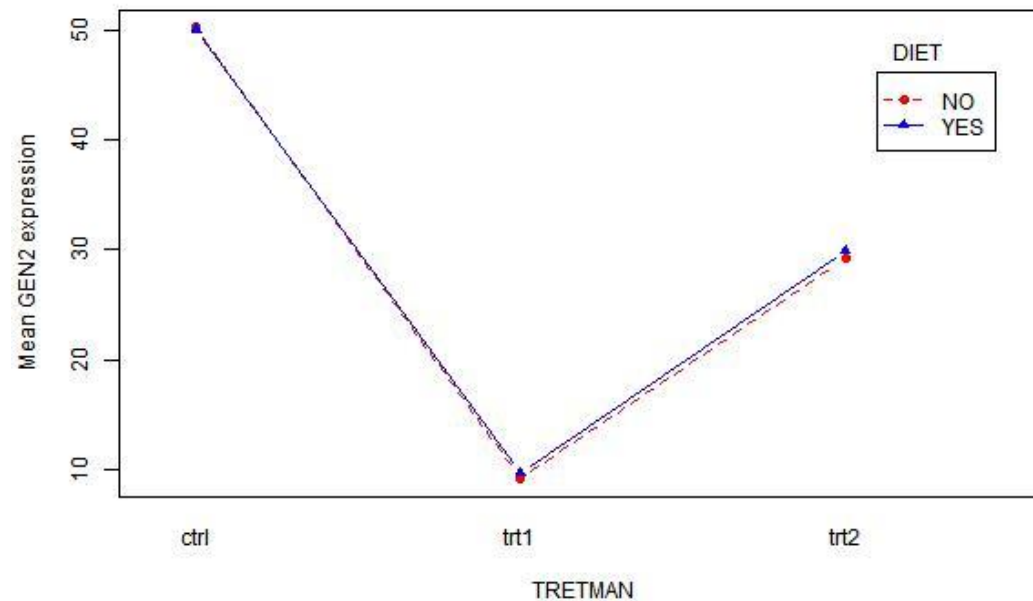


## Bez interakcija

```
> my.aov = aov(GEN2 ~ TRETMAN + DIET + TRETMAN:DIET,  
+ data = data2)  
> summary(my.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
TRETMAN	2	41898	20949	2032.427	<2e-16 ***
DIET	1	4	4	0.406	0.525
TRETMAN:DIET	2	6	3	0.269	0.765
Residuals	144	1484	10		

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



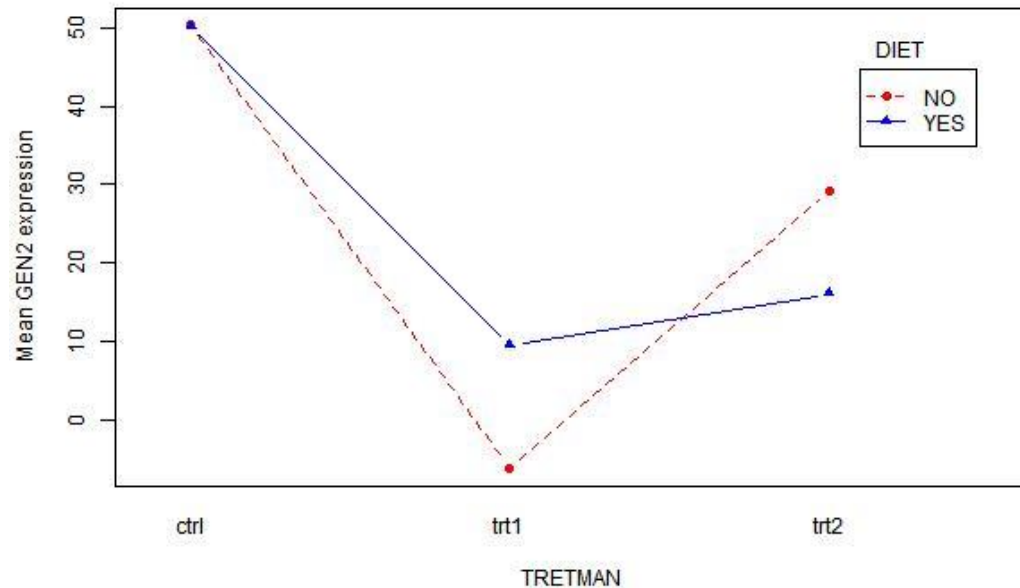
# Prikazi interakcija

## Naš primjer

```
> my.aov = aov(GEN2 ~ TRETMAN + DIET + TRETMAN:DIET,  
+ data = data)  
> summary(my.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
TRETMAN	2	60620	30310	2393.450	<2e-16 ***
DIET	1	43	43	3.374	0.0683 .
TRETMAN:DIET	2	5110	2555	201.754	<2e-16 ***
Residuals	144	1824	13		

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



## Bez interakcija

```
> my.aov = aov(GEN2 ~ TRETMAN + DIET + TRETMAN:DIET,  
+ data = data2)  
> summary(my.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
TRETMAN	2	41898	20949	2032.427	<2e-16 ***
DIET	1	4	4	0.406	0.525
TRETMAN:DIET	2	6	3	0.269	0.765
Residuals	144	1484	10		

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

