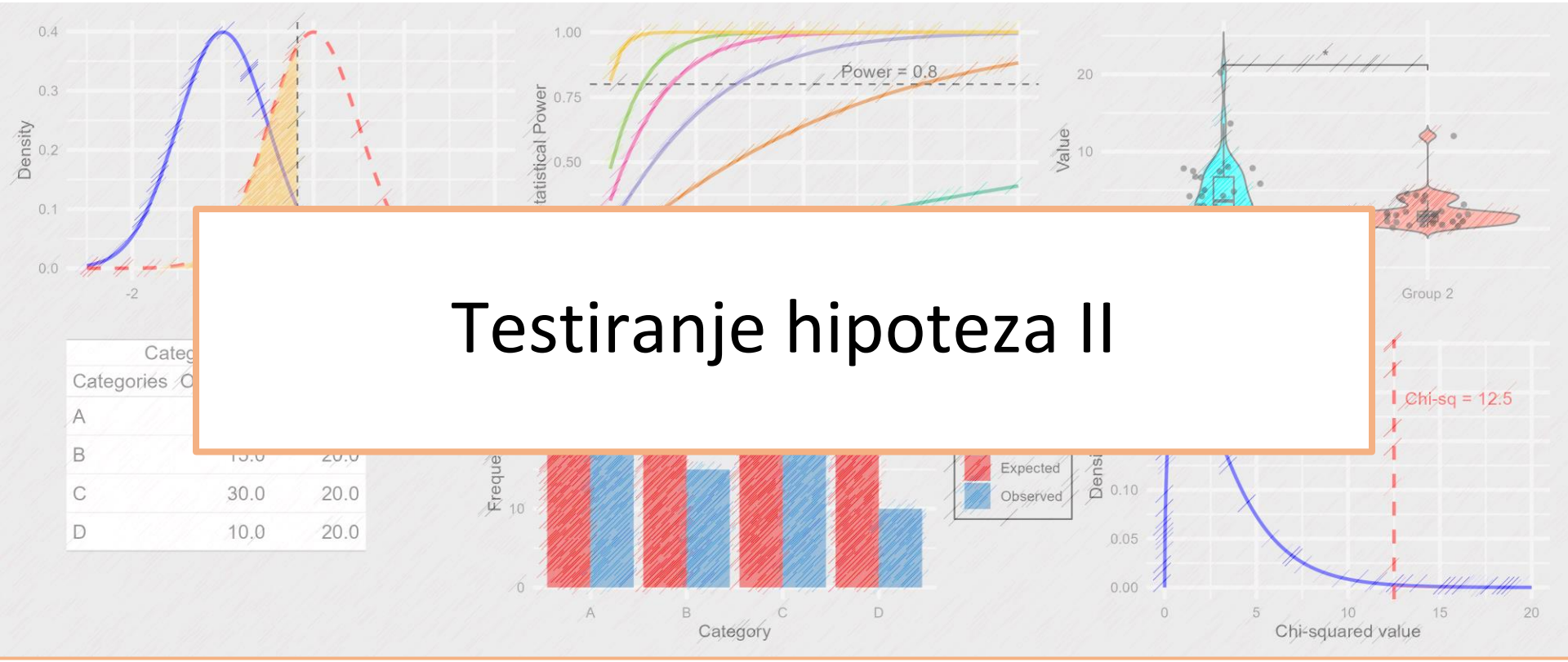


Testiranje hipoteza II

Izv.prof. Rosa Karlić

Predavanje 8, MZIRuB 2025/2026

14.01.2026.

Kategorički podaci

Interval pouzdanosti za proporciju

Interval pouzdanosti = statistika ± margina greške

margina greške= kritična vrijednost x standradna devijacija (ili standardna greška) statistike

- Dihotomni ishodi
- Statistika = proporcija u uzorku

$\bar{p}$ = x/n

x – opaženi broj uspjeha u uzorku; n – broj observacija u uzorku

- Standardna greška statistike

$$SE(\bar{p}) = \sqrt{\frac{\bar{p} \cdot (1 - \bar{p})}{n}}$$

- Interval pouzdanosti

$$\bar{p} \pm z \cdot \sqrt{\frac{\bar{p} \cdot (1 - \bar{p})}{n}}$$

Testiranje hipoteza za proporciju

- Odredite nultu hipotezu H_0 i alternativnu hipotezu H_A
- Izračunajte test statistiku:

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

- Izračunajte p-vrijednost i odredite hoćete li odbaciti nultu hipotezu

- Prikladno za velike uzorke – barem 5 observacija u svakoj kategoriji
- U ostalim slučajevima koristite egzaktne metode (binomni test)

Binomni test

- Testira razlikuje li se promatrani udio uspjeha (ili jedne kategorije) značajno od očekivanog udjela
- Pretpostavke:
  - Binarni podaci (dvije kategorije: uspjeh/neuspjeh ili da/ne).
  - Neovisni eksperimenti

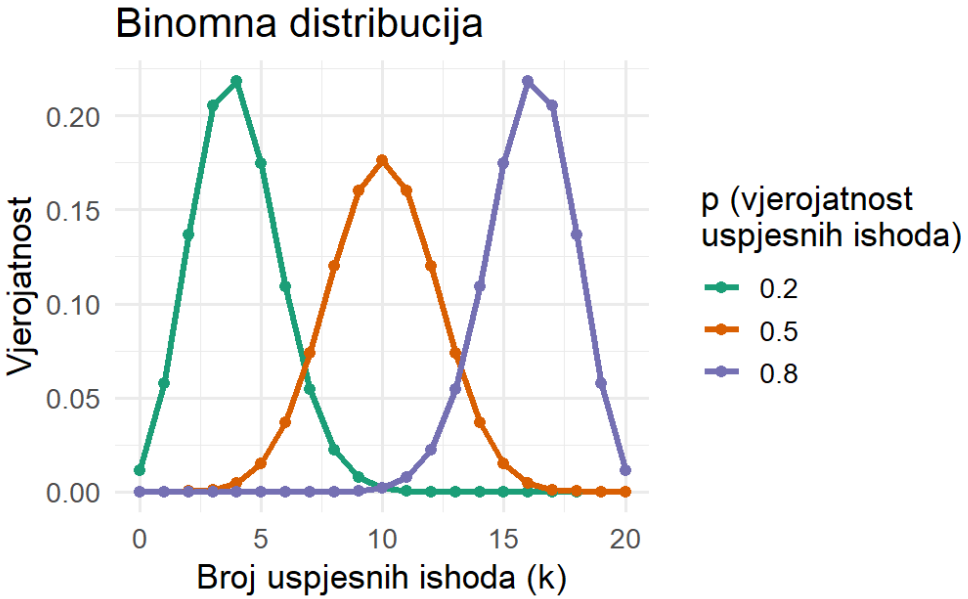
- Uspjeh - jedan od dva moguća ishoda u eksperimentu
- Definicija uspjeha specifična je za studiju koja se provodi i ovisi o istraživačkom pitanju (ishod koji proučavamo)
  - Prisutnost varijante
  - Pacijent koji se oporavio nakon tretmana
  - Detekcija proteina u stanici
  - Uspješan knockout gena

Binomni test

- Nulta hipoteza (H_0): promatrani udio uspješnih ishoda odgovara očekivanom udjelu (p_0).
- p-vrijednost se računa na temelju binomne distribucije
- Koliko su ekstremni opaženi podaci ako je nulta hipoteza točna?

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

n – broj pokušaja (eksperimenata)
k –broj uspješnih ishoda
p – očekivani udio uspješnih ishoda



Koliko je uspješan knockout gena?

- Provodimo eksperiment u kojem želimo napraviti knockout gena CRISPR metodom. Pretpostavljamo da je uspješnost metode 50%
  - Nulta hipoteza (H0): Proporcija stanica sa knockout genom je 0.5 (p=0,5).
  - Alternativna hipoteza (Ha): Proporcija stanica sa knockout genom nije 0.5 (p≠0,5)
- Rezultati: Testiramo 500 stanica, 350 imaju uspješan knockout gena

```
        Exact binomial test

data:  successful_knockouts and total_cells
number of successes = 350, number of trials = 500, p-value < 2.2e-16
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.6577337 0.7398824
sample estimates:
probability of success
                   0.7
```

Hi-kvadrat test za procjenu prikladnosti statističkog modela (*goodness-of-fit*)

Proučavate nasljeđivanje specifičnog genetskog markera u populaciji. Prema Mendelovom nasljeđivanju, očekujete da će potomci slijediti omjer 1:2:1 (25%:50%:25%) za tri genotipa (AA, Aa, aa) u heterozigotnom križanju (Aa×Aa).

- H_0: Učestalost genotipa u populaciji slijedi očekivani omjer od 1:2:1.
- H_A : Učestalost genotipa u populaciji ne slijedi očekivani omjer od 1:2:1.

- U slučajnom uzorku od 200 potomaka uočene frekvencije genotipova su:

	AA	Aa	aa	UKUPNO
Opaženo	40 (20%)	100 (50.0%)	60 (30%)	200
Očekivano	50 (25.0 %)	100 (50.0 %)	50 (25.0 %)	200

Nasljeđuje li se promatrani genetski marker prema Mendelovom nasljeđivanju?

Hi-kvadrat test za procjenu prikladnosti statističkog modela (*goodness-of-fit*)

- Koliko se dvije distribucije prosječno razlikuju jedna od druge

$$\frac{\text{observed frequency} - \text{expected frequency}}{\text{expected frequency}}$$

AA: $\frac{(40-50)}{50} = -0.2$ Aa: $\frac{(100-100)}{100} = +0.0$ aa: $\frac{(60-50)}{50} = 0.2$

$$\frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}}$$

AA: $\frac{(40-50)^2}{50} = 2$ Aa: $\frac{(100-100)^2}{100} = 0$ aa: $\frac{(60-50)^2}{50} = 2$

zbroj: 4

Hi-kvadrat (χ^2) statistika

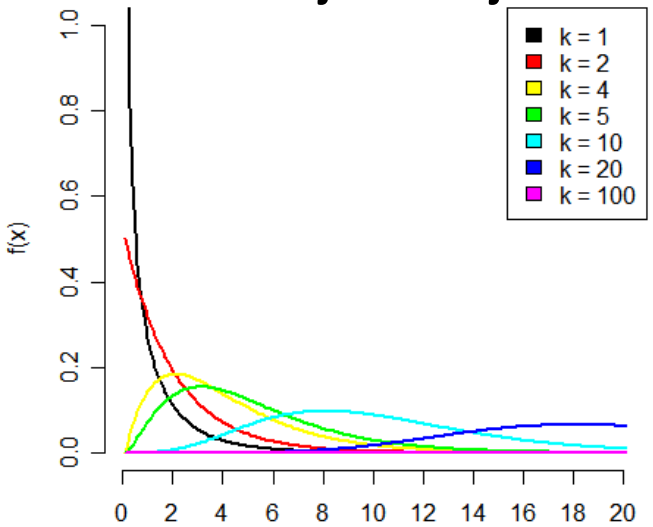
$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i}$$

- O_i – opažena frekvencija (observed frequency)
- E_i – očekivana frekvencija (expected frequency)
- N – broj kategorija

Hi-kvadrat distribucija

- **hi-kvadrat distribucija s *k* stupnjeva slobode je distribucija zbroja kvadrata *k* nezavisnih standardnih normalnih slučajnih varijabli**

$$f_x(X) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2-1)} e^{-x/2}$$



- U slučaju jednodimenzionalne klasifikacije broj stupnjeva slobode = broj klasa (ćelija) -1

Hi-kvadrat test za procjenu prikladnosti statističkog modela (*goodness-of-fit*)

	AA	Aa	aa	UKUPNO
Opaženo	40 (20%)	100 (50.0%)	60 (30%)	200
Očekivano	50 (25.0 %)	100 (50.0 %)	50 (25.0 %)	200

```
> chisq.test(c(40,100,60), p = c(0.25, 0.5, 0.25))

        Chi-squared test for given probabilities

data:  c(40, 100, 60)
X-squared = 4, df = 2, p-value = 0.1353
```

Ne možemo odbaciti nultu hipotezu na razini značajnosti 0.05!

Hi-kvadrat test je uvijek neusmjeren (two-tailed)!

Hi-kvadrat test asocijacije

- Postoji li veza između dvije kategoričke varijable?

		Status		
		Bolestan	Zdrav	
Primio tretman	NE	32	18	50
	DA	23	77	100
		55	95	150

H_0: U populaciji ne postoji povezanost između dvije varijable
H_A : U populaciji postoji povezanost između dvije varijable

- Usporedba očekivane i opažene frekvencije
- Izračun očekivanih frekvencija
- Izračun prikladne vrijednosti stupnjeva slobode

Hi-kvadrat test asocijacije

- Kako izračunati očekivane frekvencije?

		Ishodi za varijablu STATUS		
		Bolestan	Zdrav	
Ishodi za varijablu TRETMAN	NE	$\frac{R1 \times C1}{TOTAL}$	$\frac{R1 \times C2}{TOTAL}$	R1
	DA	$\frac{R2 \times C1}{TOTAL}$	$\frac{R2 \times C2}{TOTAL}$	R2
		C1	C2	TOTAL

- Kako izračunati stupnjeve slobode?

$$df=(r-1)x(c-1)$$

r – broj redova

c – broj stupaca

Hi-kvadrat test asocijacije

		Status		
		Bolestan	Zdrav	
Primio tretman	NE	32 (E = 18.33)	18 (E = 31.67)	50
	DA	23 (E = 36.67)	77 (E = 63.33)	100
		55	95	150

Više od dva stupca i dva reda:

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i}$$

Dva stupca i dva reda:

$$\chi^2 = \sum_{i=1}^{N} \frac{(|O_i - E_i| - 0.5)^2}{E_i}$$

- Hi-kvadrat statistika = 22.396
- Stupnjevi slobode = 1
- P-vrijednost = 2.218×10^{-6}

Pretpostavke za korištenje Hi-kvadrat testa

- Kategoričke varijable su neovisne jedna o drugoj
- Očekivane vrijednosti E su dovoljno velike (>= 5)

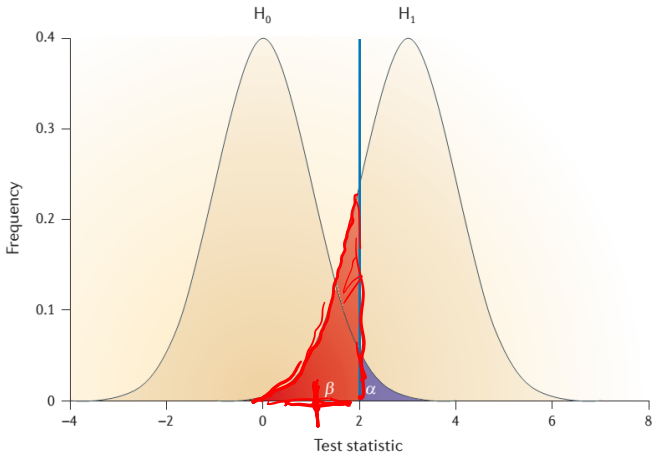
- U slučaju međusobno ovisnih kategoričkih varijabli koristi se **McNemarov test**
- U slučaju malog broja uzoraka (E < 5) koristi se **Fisherov test**

Pogreške u statističkom testiranju

Pogreške u testiranju hipoteza

- **Pogreška tipa I.**
  - Odbacili smo nultu hipotezu koja je istinita
  - Vjerojatnost da ćemo počiniti pogrešku tipa I - α
  - Lažno pozitivni rezultati
- **Pogreška tipa II.**
  - Nismo odbacili nultu hipotezu koja je neistinita
  - Lažno negativni rezultati
  - Vjerojatnost da ćemo počiniti pogrešku tipa II - β
  - Vjerojatnost da nećemo počiniti pogrešku tipa II zove se **snaga testa**

	Nismo odbacili H_0	Odbacili smo H_0
H_0	-	Pogreška tipa I
H_1	Pogreška tipa II	-



Sham & Purcell, 2014, Nat Rev Genetics

Razina značajnosti – α

- Kada je rizik da ćemo napraviti grešku tipa I prevelik?
- Ovisi o okolnostima , uglavnom je prihvaćena vrijednost 5% (p = 0.05)
- Učestalost **greške tipa I** (broj greški tipa I na 100 eksperimenata) zove se *alfa razina*

Snaga testa

- Napravili smo t-test i zaključili da je razlika između srednje vrijednosti našeg uzorka i srednje vrijednosti populacije statistički značajna sa p-vrijednošću $p < 0.05$
- Ali što ako je $p > 0.05$?

- *Koji su mogući razlozi za $p > 0.05$?*
  - Nulta hipoteza je točna

  ili
  - Naš eksperiment nema dovoljno *statističke snage* i napravili smo **grešku tipa II**

Snaga testa

- Zašto p-vrijednost može biti veća od 0.05?
- Podsjetimo se:

$$t_{\bar{x}} = \frac{\bar{x} - \mu_{\bar{x}}}{SE_{\bar{x}}} \quad \text{i} \quad SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Snaga testa

- Zašto p-vrijednost može biti veća od 0.05?

1. Mala veličina učinka:

$\overline{X}$ je blizu srednje vrijednosti populacije

- Ne možemo razlikovati učinak tretmana od varijabilnosti u podacima
- Rješenje: Povećati veličinu učinka (npr. veća doza)

Snaga testa

- Zašto p-vrijednost može biti veća od 0.05?

2. „Šum” u podacima

s a zbog toga i $SE_{\bar{x}}$ je dosta velika

- Veliki broj u nazivniku će „poništiti” mali efekt
- Rješenje: Što je moguće više reducirati pogreške u mjerenjima

Snaga testa

- Zašto p-vrijednost može biti veća od 0.05?

3. Premala veličina uzorka

$SE_{\bar{x}}$ je dosta velika jer je $\sqrt{n}$ malen

- Veliki broj u nazivniku će „poništiti” mali efekt
- Rješenje: povećati broj uzoraka u eksperimentu

Snaga testa

- Koliko smo sigurni da smo mogli otkriti značajan učinak
- SNAGA $\propto \frac{veličina\ učinka\ \ i\ \alpha}{\sigma\sqrt{n}}$
- Točan izračun ovisi o vrsti statističkog testa i alternativnoj hipotezi

- Bitno: to što nismo odbacili nultu hipotezu ne znači da je nulta hipoteza točna!!!

Snaga testa

- Da bismo izračunali potrebnu veličinu uzorka moramo unaprijed odrediti željenu snagu (uglavnom 0.80 ili 0.90), razinu značajnosti α, veličinu učinka i procijeniti standardnu devijaciju
- Distribucija test statistike – ne-centralna t-distribucija, ovisi o ne-centralnom parametru ν i stupnjevima slobode

- Zašto ne uzmemo najveći mogući uzorak: troškovi, etički razlozi i biološka vs. statistička značajnost

Snaga testa

- Snaga testa ovisi o nekoliko različitih faktora
- Problem niske snage testa u biomedicinskim istraživanjima
- Prihvaćena vrijednost je obično 0.8

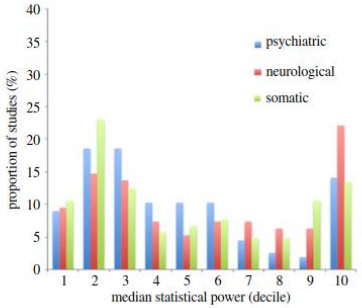
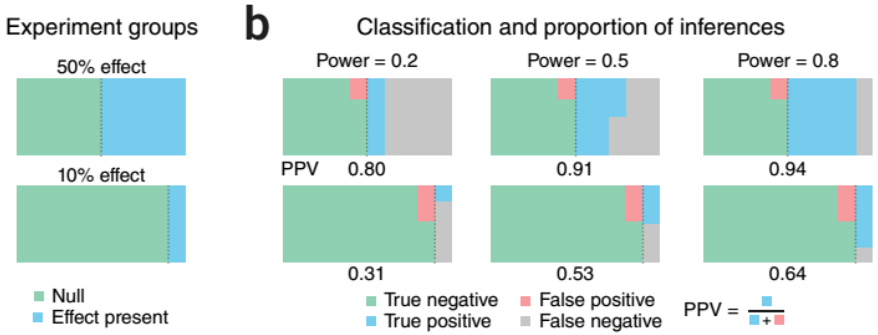


Figure 2. Distribution of statistical power of individual studies (sensitivity analysis). The distribution of the average statistical power of individual studies contributing to meta-analyses across three biomedical domains (psychiatry, neurology and somatic disease) is shown, restricted to meta-analyses indicating a statistically significant pooled effect size estimate only. This indicates a broadly similar bimodal distribution, albeit indicating higher average power overall. This overall pattern again appears to hold across all three domains of interest.

Dumas-Mallet E, R Soc Open Sci. 2017



Krzywinski & Altman, 2013, Nat Methods

Veličina učinka

- Koristi se kako bismo procijenili magnitudu učinka koji proučavamo

- Mogu se koristiti različite statistike:
  - Cohenov *d* (standardizirana razlika srednje vrijednosti)
  - eta-kvadrat (ANOVA)
  - …..

Veličina učinka – razlika srednje vrijednosti

- Cohenov *d*

$$d = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_p^2}} \qquad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1){s_2}^2}{n_1 + n_2 - 2}$$

- 0.2 – mali učinak, 0.5 – srednji učinak, 0.8 – veliki učinak

Alternative:

- Glassov *Δ* – uzorci s različitim varijancama (koristi samo varijancu kontrolne skupine)
- Hedgesov *g* – različite veličine uzoraka

Veličina učinka - ANOVA

- Eta na kvadrat (η^2)

Mjeri udio ukupne varijance u ovisnoj varijabli objašnjenoj nezavisnom varijablom(ama). Obično se koristi u jednosmjernoj i dvosmjernoj ANOVA-i.

- Djelomični eta-kvadrat (η^2_p)

Varijacija eta-kvadrata koja uzima u obzir jedinstveni doprinos svake nezavisne varijable, korisna u ponovljenim mjerenjima i faktorskoj ANOVA-i.

Veličina učinka – kategoričke varijable

- Omjer izgleda (Odds ratio)

a	b
c	d

$$OR = \frac{a/c}{b/d} = \frac{ad}{bc}$$

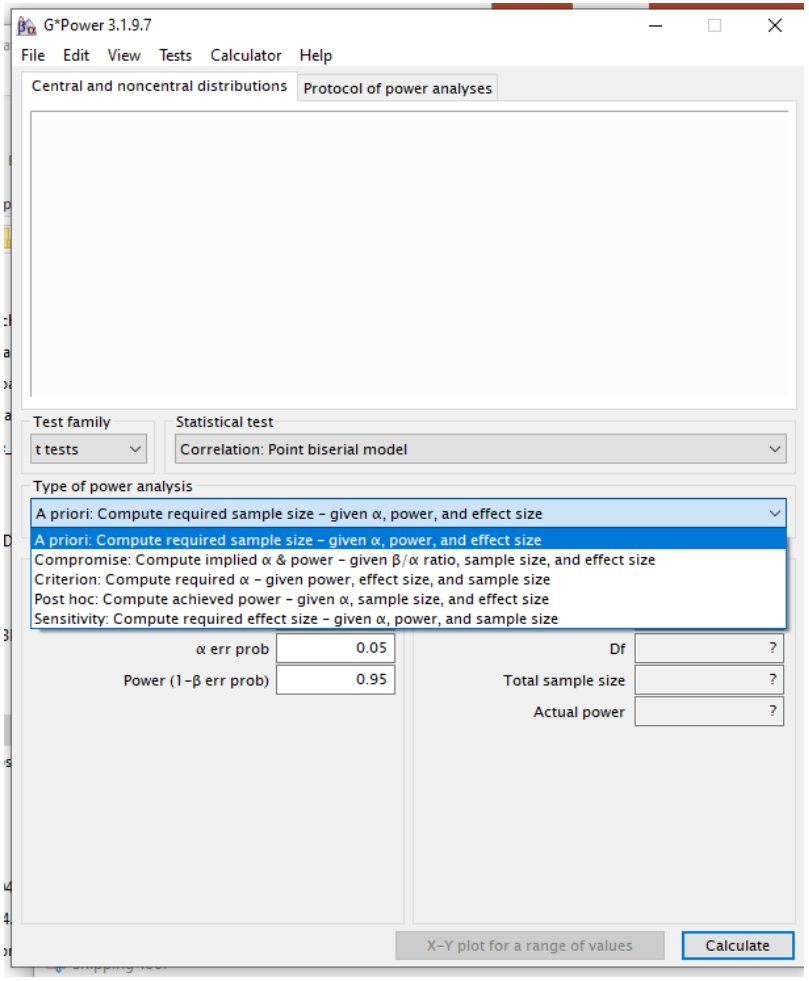
OR = 1 nema asocijacije
OR > 1 pozitivna asocijacija
OR < 1 negativna asocijacija

- Relativni rizik (Risk ratio)
- Phi koeficijent (Phi coefficient)

$$RR = \frac{\text{Risk in group 1}}{\text{Risk in group 2}} = \frac{a/(a+b)}{c/(c+d)}$$

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

GPower


G*Power 3.1.9.7
File Edit View Tests Calculator Help
Central and noncentral distributions
Protocol of power analyses
Test family
t tests
Statistical test
Correlation: Point biserial model
Type of power analysis
A priori: Compute required sample size – given α, power, and effect size
A priori: Compute required sample size – given α, power, and effect size
Compromise: Compute implied α & power – given β/α ratio, sample size, and effect size
Criterion: Compute required α – given power, effect size, and sample size
Post hoc: Compute achieved power – given α, sample size, and effect size
Sensitivity: Compute required effect size – given α, power, and sample size
α err prob
0.05
Power (1–β err prob)
0.95
Df
?
Total sample size
?
Actual power
?
X-Y plot for a range of values
Calculate

Istraživanje treba biti reproducibilno i statistički ispravno

- 17–25% značajnih rezultata u društvenim znanostima (α = 0.05) je vjerojatno krivo, (Johnson, V. E. Proc. Natl Acad. Sci. USA, 2013) – nereproducibilnost rezultata je ozbiljan problem u znanosti
- Nature checklist:

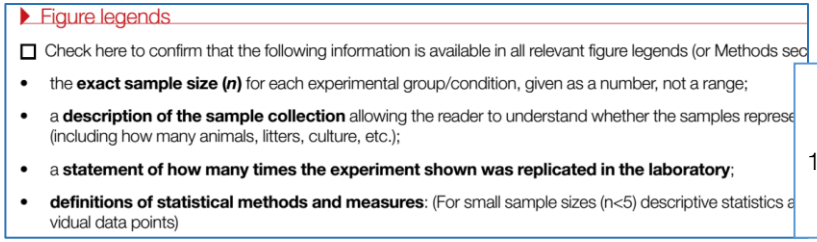
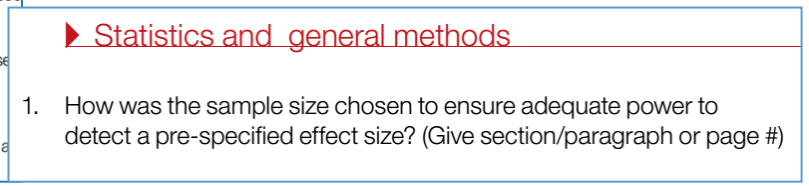


Figure legends

☐ Check here to confirm that the following information is available in all relevant figure legends (or Methods sec

- the **exact sample size (*n*)** for each experimental group/condition, given as a number, not a range;
- a **description of the sample collection** allowing the reader to understand whether the samples represe (including how many animals, litters, culture, etc.);
- a **statement of how many times the experiment shown was replicated in the laboratory**;
- **definitions of statistical methods and measures**: (For small sample sizes (n<5) descriptive statistics a vidual data points)



Statistics and general methods

1. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size? (Give section/paragraph or page #)

Nature statistics collection

http://www.nature.com/collections/qghhqm