

Primjeri korištenja statističkih metoda

## Temelj znanstvene metode – Statističko zaključivanje

Statistical inference - the making of estimates concerning a population from information gathered from samples ([www.merriam-webster.com](http://www.merriam-webster.com))

Kontrola pogreške (*ne znači da netko nešto „krivo” radi!*)

Populacija?

Uzorak?

*Reprezentativan?*

*Dovoljno velik?*

Odabir (statističke) metode (testa)

Pokus - an operation or procedure carried out under controlled conditions in order to discover an unknown effect or law, to test or establish a hypothesis, or to illustrate a known law ([www.merriam-webster.com](http://www.merriam-webster.com))

Kontrolirani uvjeti

Tretmani

Ponavljjanja (kontrola pogreške)

Unaprijed definiran dizajn (i model)

Populacija „poznata”



Poljski pokus

# Uzorkovanje u prirodnim populacijama

Populacije često nepoznate veličine i sastava (reprezentativni uzorci?)

*(možda uzorkujemo samo slabe, stare, bolesne... možda uzorkujemo samo znatiželjne i „agresivne” ili samo one navikle na mjesto gdje uzorkujemo...)*

Sistematsko uzorkovanje (po sezoni, staništu, godini...) teško (reprezentativni uzorci?)

Ponavljanja (kontrola pogreške?)

Unaprijed definiran dizajn (i model) ?

Uzorkovanje često skupo (veličina uzorka?)

Tehnologija pomaže (jeftinije, brže, obuhvatnije uzorkovanje)

Kamere

GPS ogrlice

Dronovi

GIS

Genetika (eDNA, neinvazivno uzorkovanje...)

Statističke metode (za male uzorke, specifične tipove podataka, AI, ML...)

Problem planiranja i zaključivanja ostaje (postaje sve veći?)!

„Forsiranje” metoda – ostavljanje boljeg dojma/pritisak publiciranja

- Koraci u analizi podataka odabrani su ciljem istraživanja
- Cilj treba biti **jasan** i **jednoznačan** – dobro odabrani ishodi i prediktori (sjetimo se priče o hipotezama – cilj proizlazi iz hipoteze)
- Teško bez preliminarnih istraživanja

Problem:

*Kompleksnost pojava u prirodi + želja da se što brže objasni što više = problemi pri analizi i interpretaciji*

Primjer:

*„Koji način čuvanja reznica vinove loze je bolji – u pijesku ili u frižideru?“*

*Kako mjeriti „boljost“??*

*Primjer:*

*„Razlikuje li se količina sekundarnih metabolita u grožđu 2 sorte?”*

*HPLC i druge metode mogu mjeriti cca 40 sek. metabolita 😊*

*Pokus će se ponoviti u 2 godine. ZAŠTO??*

*Primjer:*

*„Razlikuje li se fotoaktivnost biljaka uzgojenih u stresu suše i bez stresa?”*

*Pokus će se raditi na 3 genotipa. ZAŠTO??*

*Primjer:*

*„Razlikuje li se učinak 3 doze fungicida na biljne patogene?”*

*Odabrane su 3 doze. KAKO??*

# 1) Opisne analize (deskriptivna statistika)

Nema definiranih ishoda i prediktora (tretmana)

Cilj je opisati neku pojavu/svojstvo/varijablu ili odnos između varijabli (korelaciju/kovarijancu)

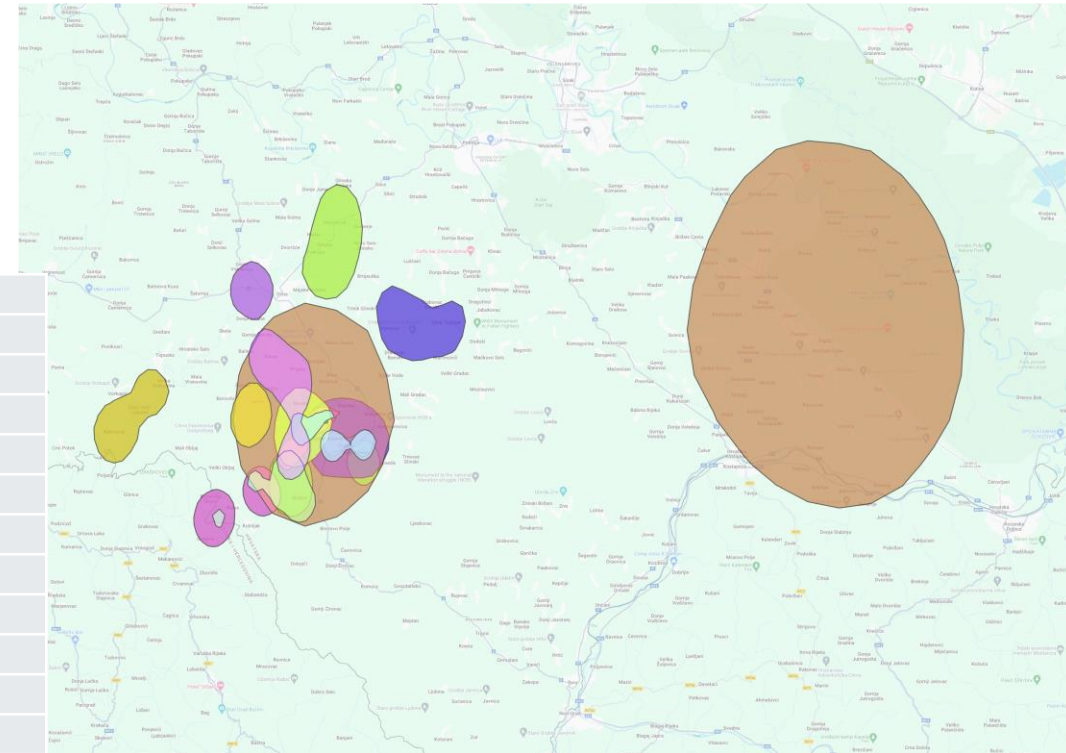
Često nema testa (ako ima, onda su to uglavnom testovi lokacije ili test korelacije/kovarijance)

Primjer:

12 jelena obilježenih GPS-GSM ogrlicama na području oko Gline

Koliki su im home rangevi (područja kretanja)?

	st_area(ud_sf)	id
1	20623314.48	Breza
2	60012104.83	Gordan
3	9143376.387	Hind2
4	19302912	Julija
5	49568287.18	Koka
6	5390486.453	Marijan
7	617226362.8	Mile
8	33188475.61	Nada
9	16690134.04	Pika
10	5039521.501	Strina
11	40033817.92	Tuga
12	7173937.258	Zdravko





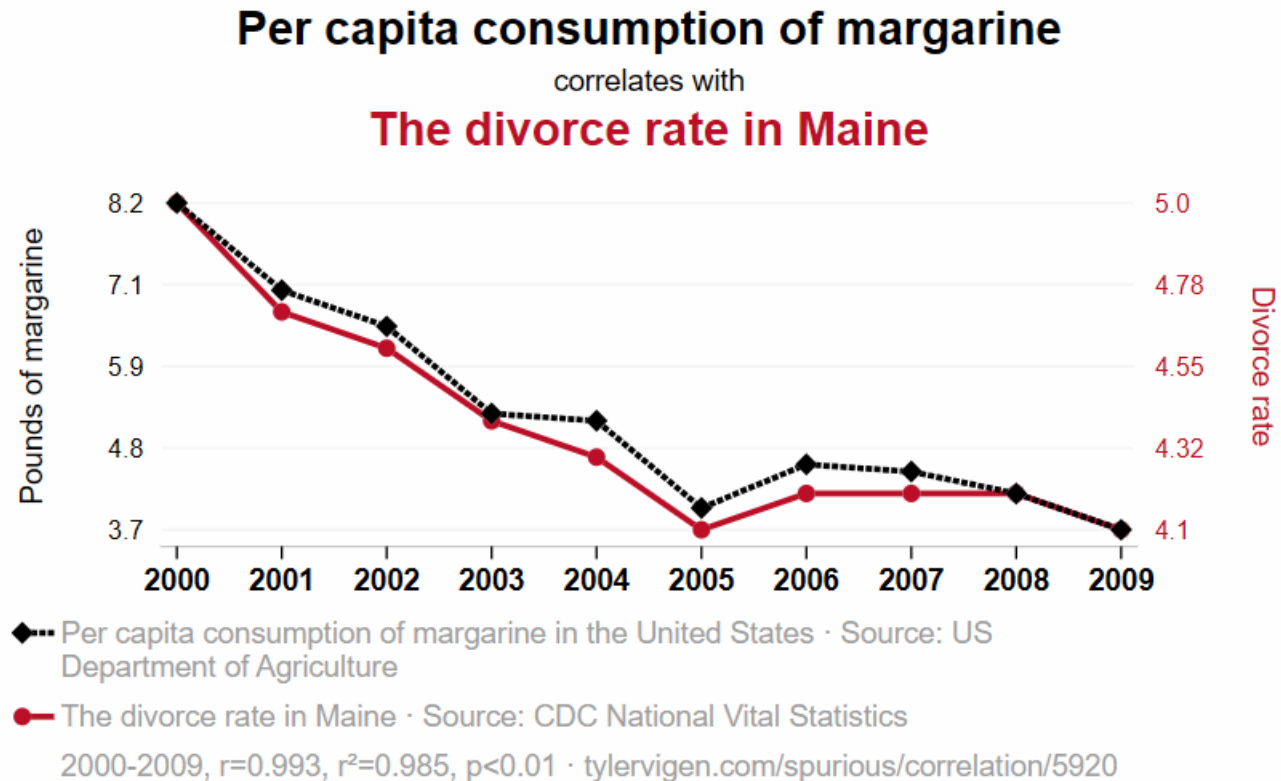
Korelacija – smjer i jačina povezanosti 2 varijable

$$-1 \leq r \leq 1$$

Osnova multivarijatnih analiza

Jako često pogrešno korištena

<https://www.tylervigen.com/spurious-correlations>



## 2) Analiza frekvencija

Ishod je kategorijska (ili binarna) varijabla

Cilj je testirati je li raspodjela jedinki u kategorije slučajna

Metode:

hi kvadrat test (chi-squared)  $\chi^2$

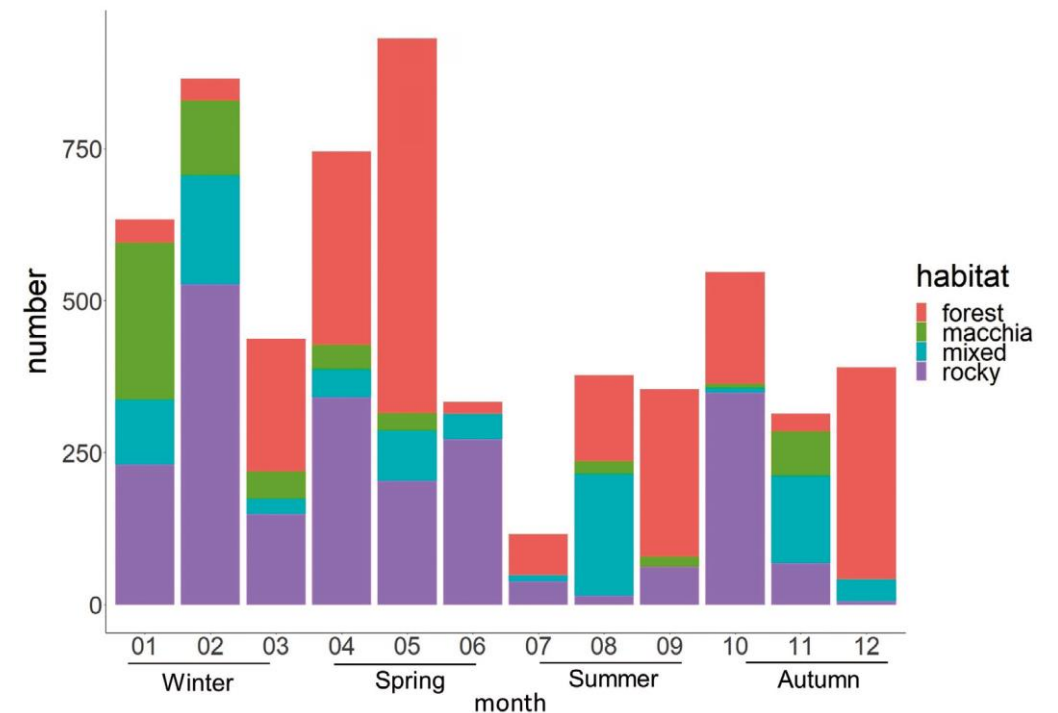
Fisherov ekzaktni test (Fishers exact test)

Primjeri:

*Oboljevaju li jednako pušači i nepušači*

*Mendelova križanja (1:2:1)*

*Broj grivastih skakača snimljenih na kamerama na različitim staništima*



### 3) Odnosi između varijabli (ishoda i prediktora)

Generalni linearni model – poopćenje (regresije, ANOVA, MANOVA, ANCOVA...)

Ishod(i) – kvantitativna varijabla (jedna ili više)

Prediktor(i) – kvantitativne ili kategorijske varijable (jedna ili više)

$$\mathbf{Y} = \mathbf{XB} + \mathbf{U}$$

Y – matrica ishoda; X – matrica prediktora (dizajn matrica); B – matrica parametara (koeficijenata koje procjenjujemo); U- matrica pogrešaka (*noise*)

Pretpostavke: greške su neovisne (*independent*) i multivarijatno normalno distribuirane

Linearna regresija – opisuje (linearnu) ovisnost y o x varijabli

$$Y = \beta_0 + \beta_1 x_1 + \dots + \varepsilon$$

$\beta_0, \beta_1, \dots$  - koeficijenti (intercept + za svaki prediktor u modelu)

$\varepsilon$  - pogreška (reziduali – razlike izmjerenih i procijenjenih y vrijednosti)

Koeficijenti se procjenjuju iz suma kvadrata uzorka

Testira se „vrijede li“ za populaciju



Frekvencionistički - t-test

„Neparametrijski“ – krosvalidacija, poduzorkovanje...

Osim opisa odnosa (testiranje efekata) služi i za interpolaciju (procjenu neizmjerenih vrijednosti y)

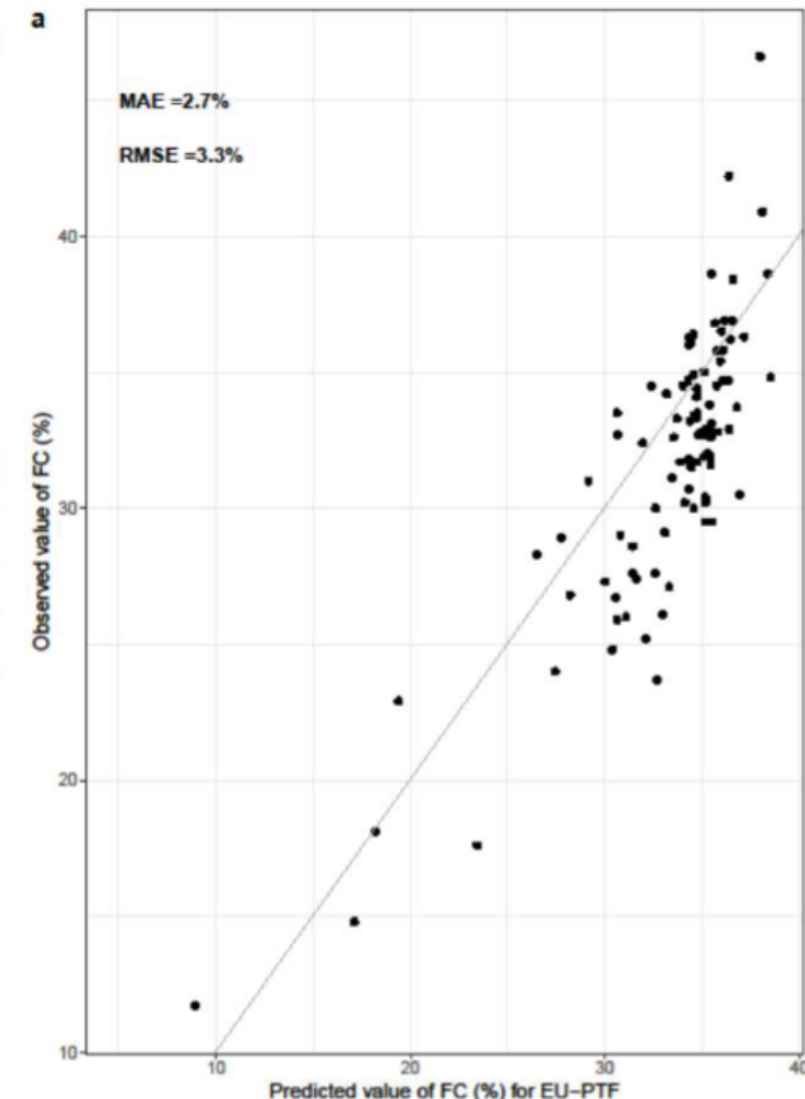
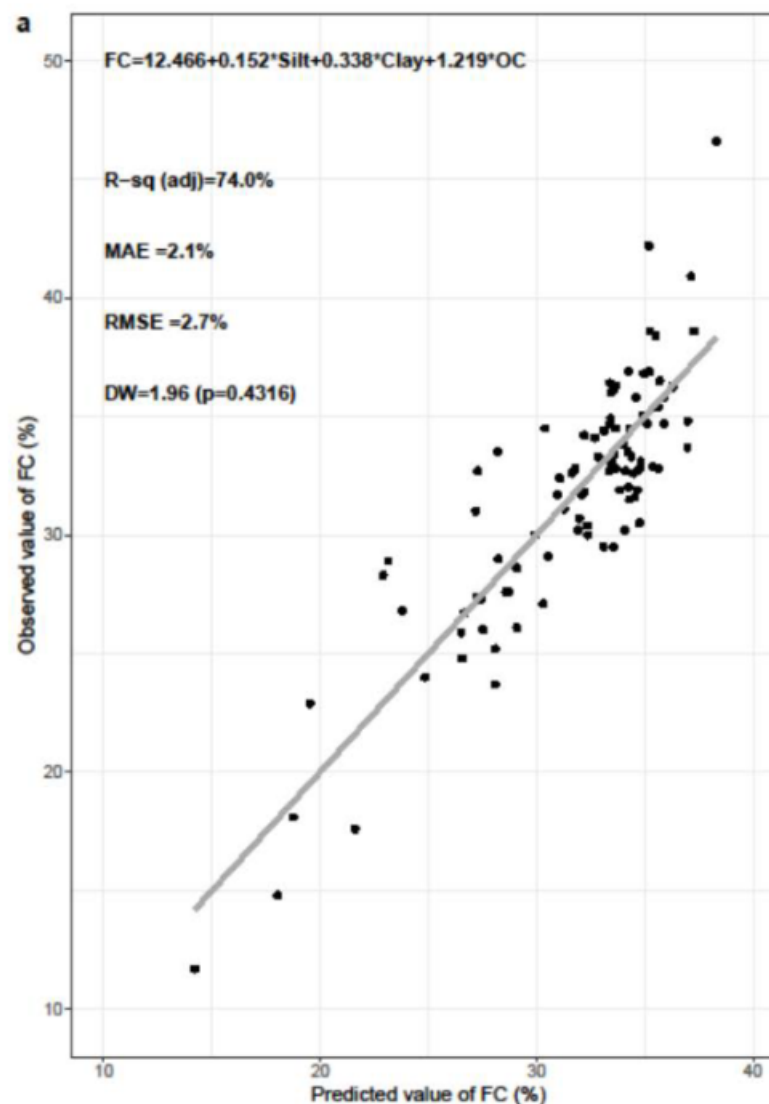
**Predviđanje??**

# Comparison of specific and universal linear regression models for predicting soil water retention in eastern Croatia

Vedran Rubinić  , Stjepan Husnjak, Danijela Jungić, Aleksanc

Procjena poljskog kapaciteta tla za vodu iz sadržaja gline i praha  
n=96  
y=FC(poljski kapacitet)  
x=prah, glina i organski C

*Desno je rezultat krosvalidacije*



Short communication

## Density-dependent decline of early horn growth in European mouflon

Krešimir Kavčić<sup>a</sup>, Luca Corlatti<sup>b,\*</sup>, Toni Safner<sup>c,d</sup>, Ivan Gligora<sup>e</sup>, Nikica Šprem<sup>a</sup>

Rani rast roga muflona

n=341

y=rani rast roga (rast u drugoj i trećoj godini)

x=kohorta (godina okota)

$\beta_1 = -0.44$ , SE = 0.073, t = -6.062, p < 0.001)

**Prediktor je kategorijski!**  
**Ovo je *de facto* ANOVA**

