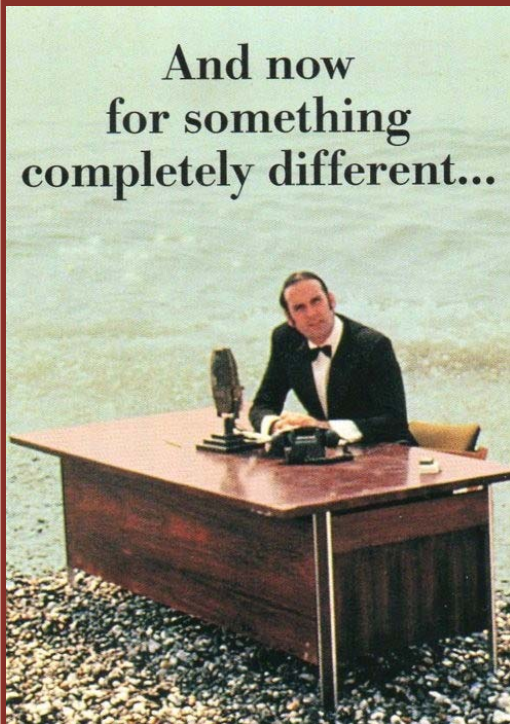


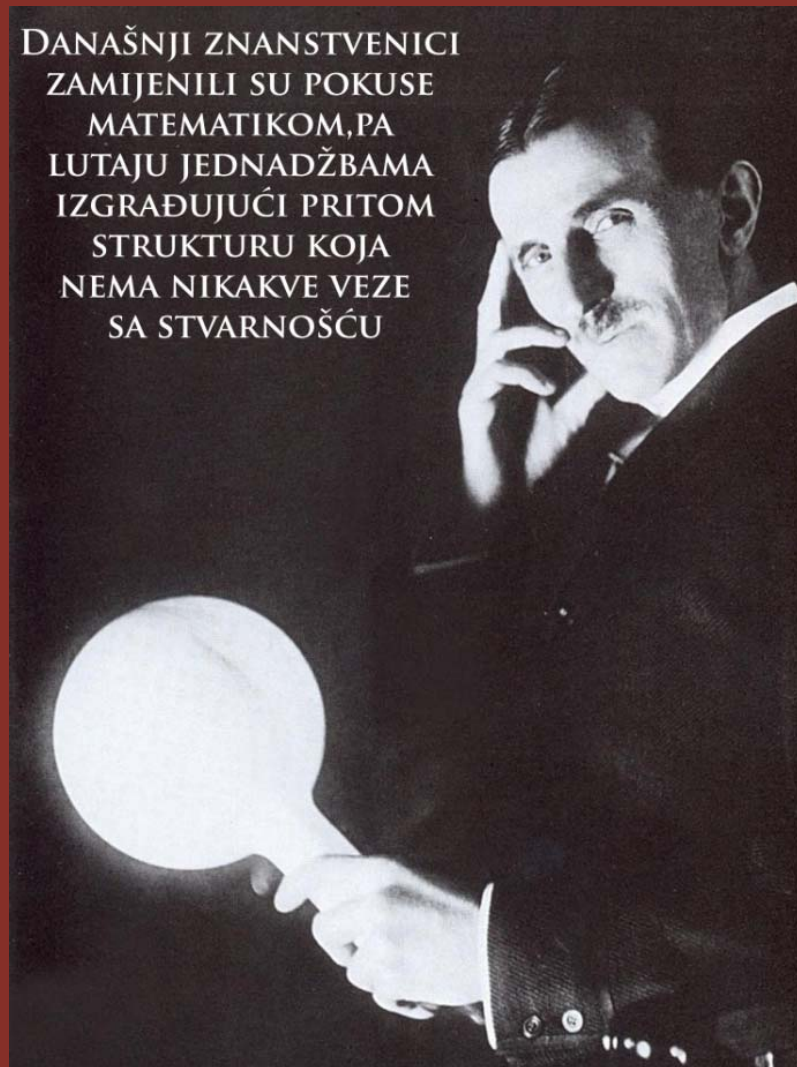
And now
for something
completely different...



...matematika...

Korištenje statističkih metoda u ekološkim istraživanjima

DANAŠNJI ZNANSTVENICI
ZAMIJENILI SU POKUSE
MATEMATIKOM, PA
LUTAJU JEDNADŽBAMA
IZGRAĐUJUĆI PRITOM
STRUKTURU KOJA
NEMA NIKAKVE VEZE
SA STVARNOŠĆU



Korištenje statističkih metoda u ekološkim istraživanjima



Ako je Vašem pokusu potrebna statistika,
trebali ste bolje postaviti pokus

Ernest Rutherford

Potrebe statističkih analiza u istraživanju okoliša:

Utvrditi na razini argumenta (**statistički značajne** vjerojatnosti)

Međusobnu povezanost

(organizama s okolišnim uvjetima i međusobno,
izvora hrane sa životinjama, pokusnih tretmana s
učincima npr. na aktivnost enzima...)

Razlike

(lokacija, staništa, zajednica, tretmani...)

PREDUVJETI: parametri moraju biti mjerljivi (i mjereni)
obilježja se moraju brojčano izraziti

Utvrđivanje uzročne povezanosti promatranih događaja

Posredno - utvrđivanjem korelacije
(koeficijenta korelacije)

Savršena korelacija = 1

(promjenom jednog parametra
drugi se mijenja u istom omjeru)

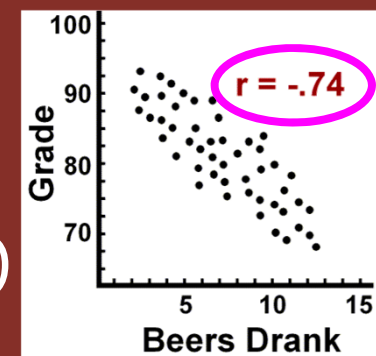
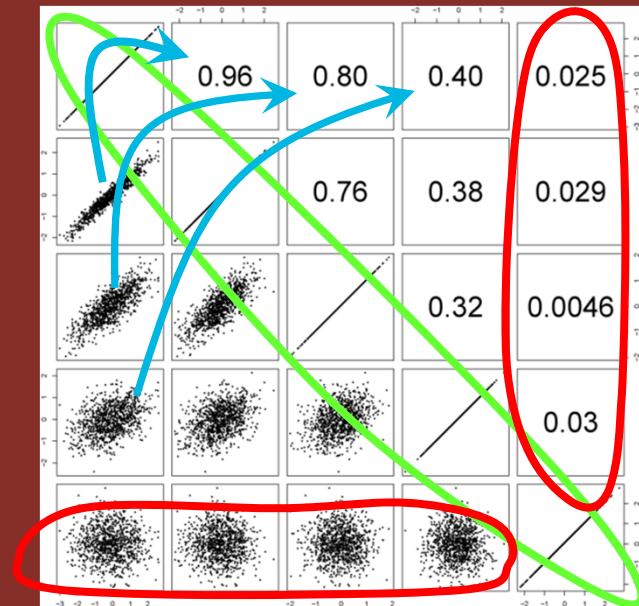
Nema korelacije ≈ 0

(promjenom jednog parametra
izostaje pravilna promjena drugog)

Negativna korelacija

(porastom jednog parametra drugi se smanjuje)

**Većina uzročno povezanih pojava pokazat će i
korelaciju. OBRAT NE VRIJEDI!**



Koeficijenti korelacije - korelacija ne znači uzročnu vezu

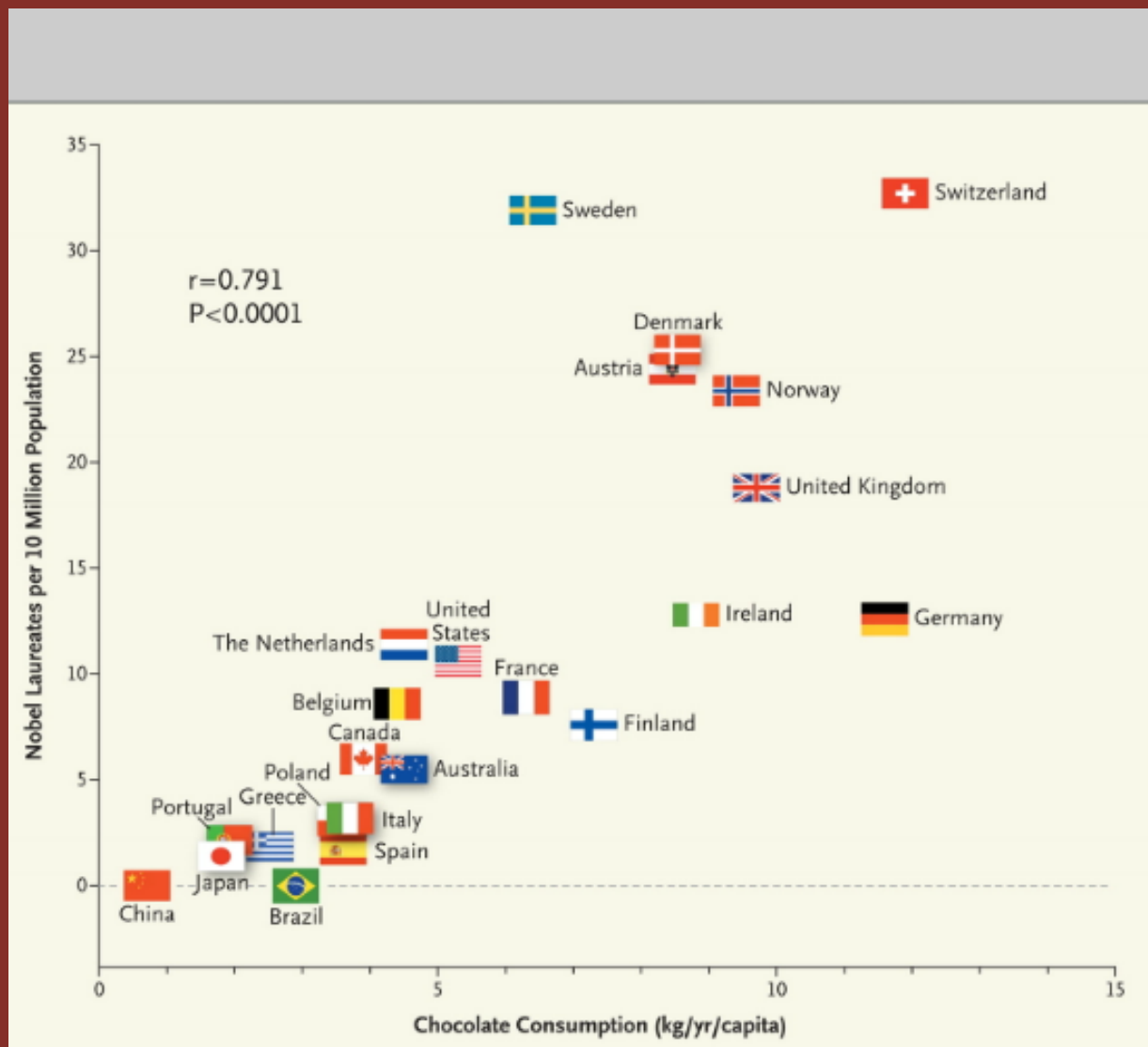


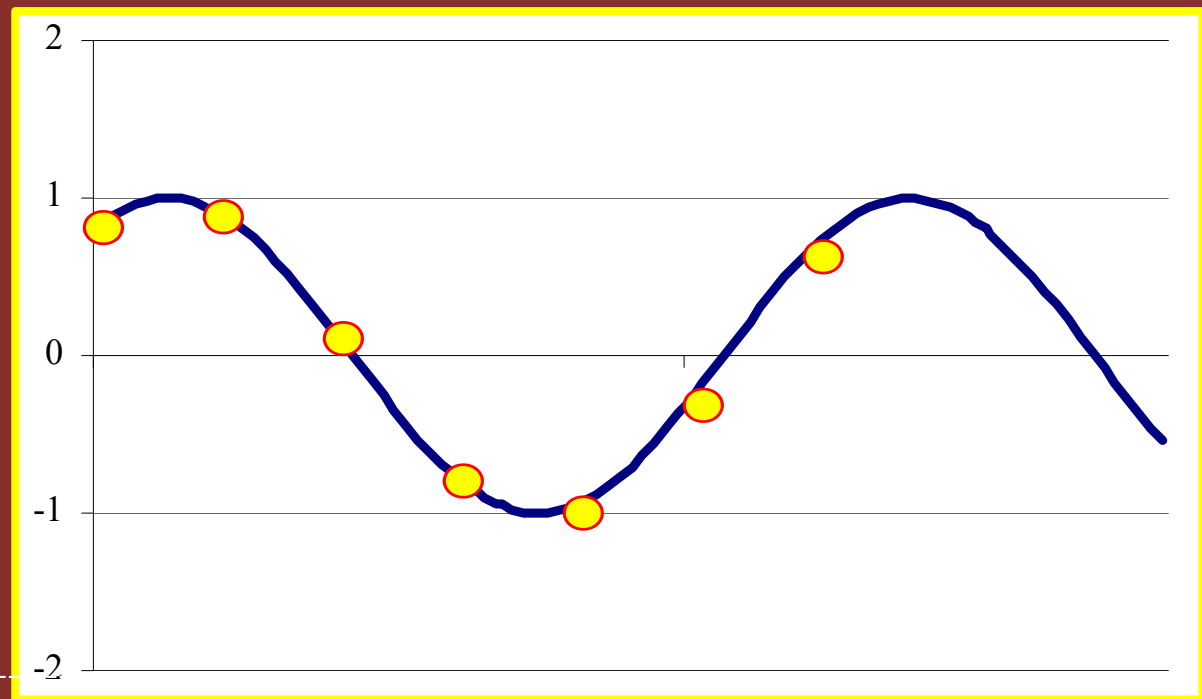
Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Postoje tri vrste laži:
laži,
proklete laži i
statistika

Benjamin Disraeli

Koeficijenti korelacije - izostanak korelacije skriva uzročnu vezu

X	Y
1	0.84
2	0.91
3	0.14
4	-0.76
5	-0.96
6	-0.28
7	0.66



$$r = 0.06$$

$$R_s = 0.04$$

Utvrdjivanje razlika između promatranih događaja

Usporedbom rasapa podataka oko srednje vrijednosti

n	X	X - $\bar{X}$
1	29	-6
2	30	-5
3	33	-2
4	35	0
5	38	3
6	39	4
7	41	6
Σ	245	
$\bar{X}$	35	

n	X	X - $\bar{X}$
1	1	-34
2	10	-25
3	15	-20
4	35	0
5	55	20
6	60	25
7	69	34
Σ	245	
$\bar{X}$	35	

varijanca $\rightarrow s^2 = \frac{\Sigma (X - \bar{X})^2}{n - 1}$

$= 21$

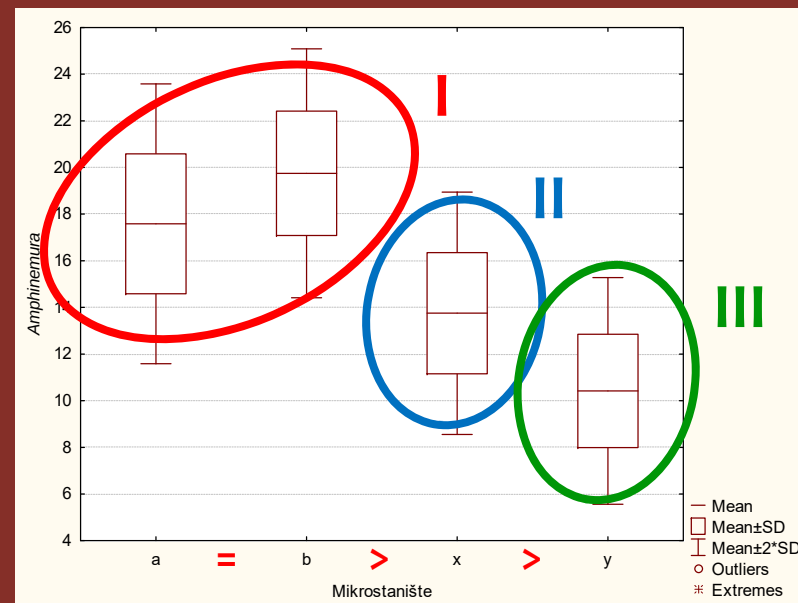
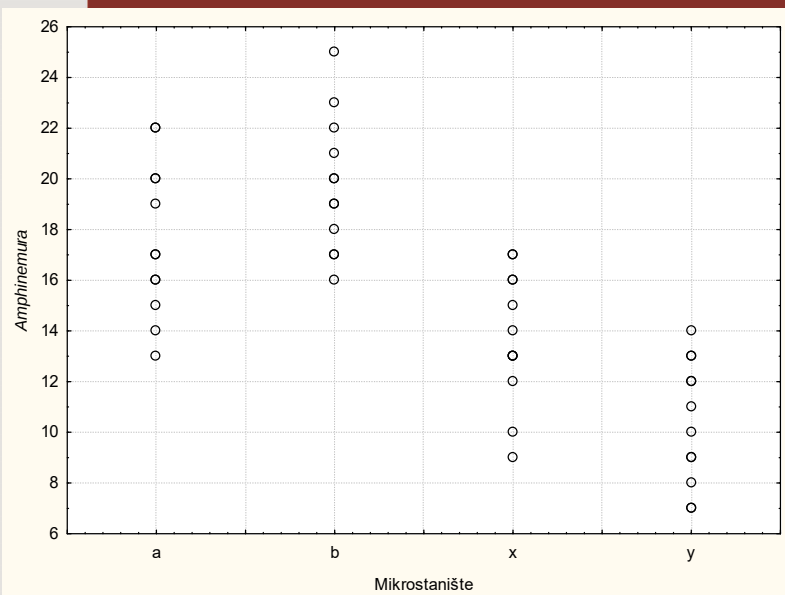
$= 727$

standardna devijacija $\rightarrow \sigma = \sqrt{s^2} = 4.58$

$= 26.96$

Utvrdjivanje razlika između promatranih događaja

Usporedbom rasapa podataka oko srednje vrijednosti



varijanca: $s^2 = \frac{\sum (X - \bar{X})^2}{n - 1}$

Kategorizacija podataka u istraživanju okoliša:

1. Kvalitativni (kategorički) i 2. kvantitativni

1. a) Nominalni

Nisu u vrijednosnom odnosu (ne označavaju se brojkama)

Poprimaju samo određene (opisne) vrijednosti

Spol: M/Ž

Npr?

Lokacija: A/B/C... ili 1/2/3...

Stanište: bentos potoka/bentos jezera/mahovinski pokrov...

1. b) Ordinalni

Jesu u vrijednosnom odnosu, ali nepreciznom

Uzvodno, nizvodno;

Npr?

Brza struja vode, spora struja vode

Ocjena iz Načela... (2,3,4,5)

Podaci u istraživanju okoliša:

2. Kvantitativni

2. a) Diskontinuirani (diskretni)
iskazuju se cijelim brojem

broj jedinki, broj studenata po godini, fotona...

Npr?

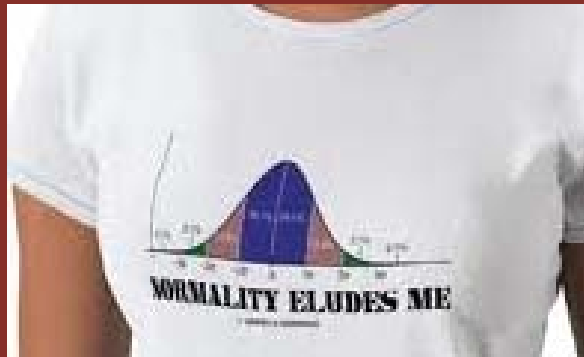
2. b) Kontinuirani
poprimaju bilo koju vrijednost u rasponu
temperatura: $-273.15-10^{17} \text{ }^{\circ}\text{C}$ ili $0-10^{17} \text{ K}$
pH: 0-14

Podaci u istraživanju okoliša:

Raspodjela:

Normalna

Najviše vrijednosti je raspoređeno oko medijana,
a najmanje u ekstremima



Nenormalna



Metode postizanja
normalnosti

$\text{Log}(x+1)$
ili
 $x^3, 2, \frac{1}{2}, \frac{1}{3}...$

Testovi normalnosti Shapiro-Wilk, Kolmogorov-Smirnov...

Razvrstavanje najčešće korištenih metoda u obradi podataka:

Parametrijske:

Podrazumijevaju normalnu raspodjelu

Neparametrijske:

Ne podrazumijevaju
normalnu raspodjelu

za male skupine podataka (10-ak)

Izmjerena vrijednost	Rang
4.361	2
9.553	8
4.588	3
5.451	4
6.855	6
5.807	5
0.983	1
17.736	9
8.239	7

Metode koje se najčešće koriste u obradi ekoloških podataka:

Parametrijske:

Pearsonov koeficijent korelacije

Analiza varijance ANOVA; (t-test)

Neparametrijske:

Spearmanov koeficijent korelacije

Kruskal-Wallis ANOVA; Mann-Whitney U-test

Metode koje se najčešće koriste u obradi ekoloških podataka:

Zadatak	Podataka u pojedinoj skupini	Odgovarajuća vrsta analize	Broj skupina podataka	Odgovarajuća analiza
Razlika	<10	Neparametrijska	3 i više	Kruskal-Walis
			2	Man-Whitney
	>10	Parametrijska*	-	ANOVA
Međudjelovanje	<10	Neparametrijska	-	Spearman R
	>10	Parametrijska*	-	Pearson r (basic)

* Potencijalno potrebno logaritamski transformirati podatke

Nul hipoteza (H_0):

Pretpostavka koju analiza testira

Npr. za:

Korelacije $\rightarrow H_0 =$ NE postoji korelacija

Analize varijance $\rightarrow H_0 =$ srednje vrijednosti se NE razlikuju

Testovi normalnosti $\rightarrow H_0 =$ raspodjela JEST normalna

Alternativna hipoteza (H_a):

Pretpostavka suprotna H_0

Općenito - pretpostavka za koju istraživač vjeruje da je istinita!

Analiza će osim rezultata svojih algoritama dati i

vrijednost 'p' $\rightarrow$ razinu vjerojatnosti da je H_0 istinita!

Statistička značajnost (p):

Dobiveni podatak (skupina podataka) i/ ili odnosi između skupina podataka nije rezultat slučajnosti ((ne)sreće)



Može se očekivati njegovo ponavljanje u istim uvjetima



Podatak ili odnos između podataka je 'istinit'

$$p < 0.05$$

Za analizu varijance $p = 0.01$ znači da je vjerojatnost da nema razlike između uspoređivanih skupina

podataka (točna H_0) samo 1 %;

Slijedi da smo **99 % sigurni da razlike postoje** (točna H_a).

Pogreške

1. vrste (odbijanje točne H_0 - lažno pozitivan rezultat) i
2. vrste (prihvatanje pogrešne H_0 - lažno negativan rezultat)

Knjiga postanka 18: 26 i 32

„Ako nađem u gradu Sodomi pedeset nevinih”, odvrati Jahve, „zbog njih ću poštedjeti cijelo mjesto.”

...

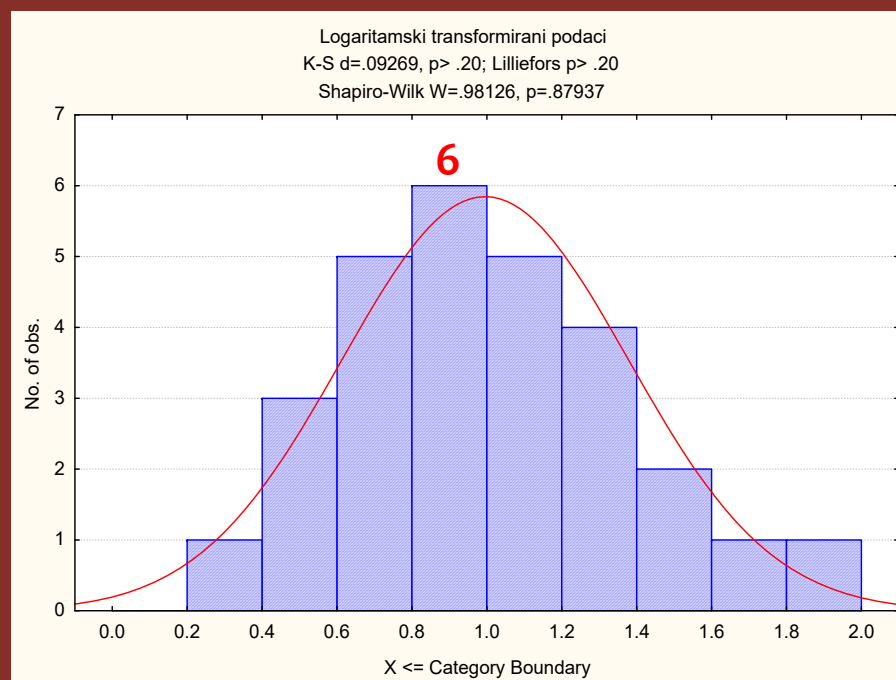
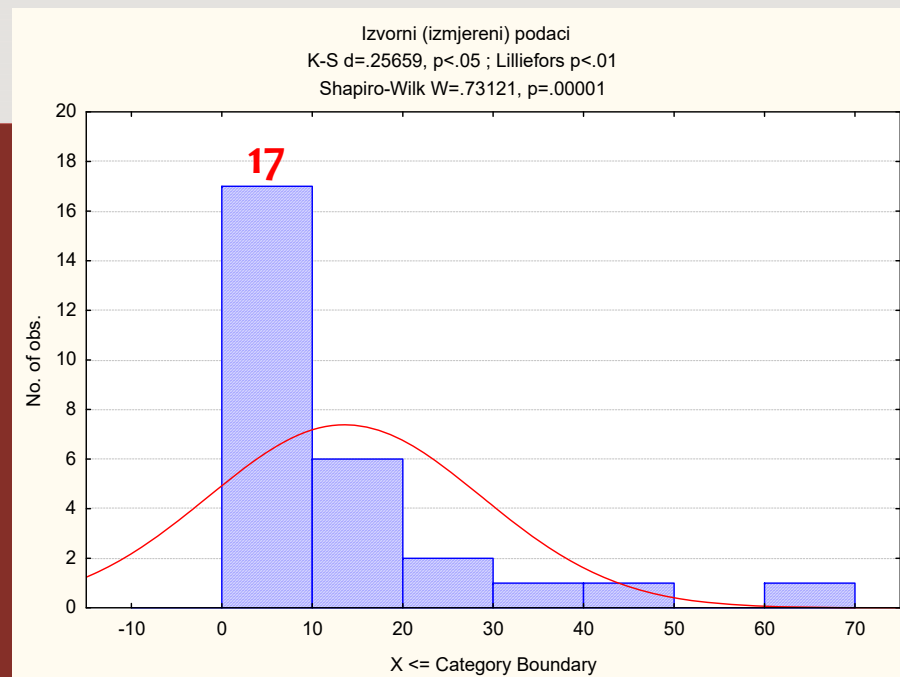
„Neka se Gospodin ne ljuti”, on [Abraham] će opet, „ako rečem još samo jednom: Ako ih je slučajno ondje samo deset?”

„Neću ga uništiti zbog njih deset”, odgovori [Jahve].

Raspodjela primjer:

r.br.	X	log (X + 1)
1	1	0.30
2	2	0.48
3	2	0.48
4	2	0.48
5	3	0.60
6	3	0.60
7	4	0.70
8	5	0.78
9	5	0.78
10	6	0.85
11	7	0.90
12	7	0.90
13	8	0.95
14	9	1.00
15	9	1.00
16	10	1.04
17	10	1.04
18	11	1.08
19	12	1.11
20	12	1.11
21	15	1.20
22	16	1.23
23	20	1.32
24	23	1.38
25	29	1.48
26	34	1.54
27	49	1.70
28	67	1.83

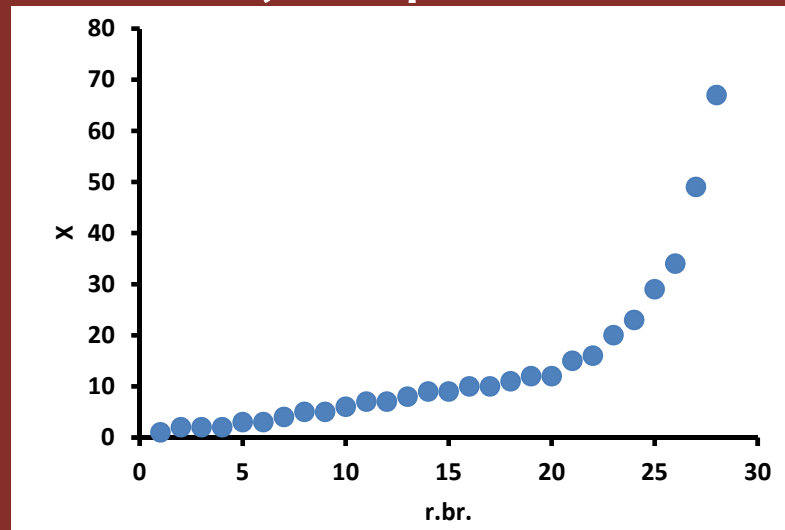
← Medijan



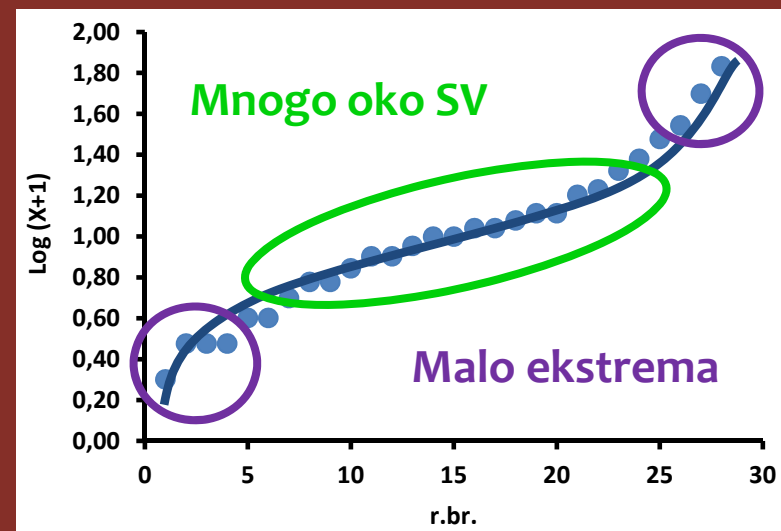
Raspodjela primjer:

r.br.	X	$\log(X + 1)$
1	1	0.30
2	2	0.48
3	2	0.48
4	3	0.60
5	3	0.60
6	3	0.60
7	4	0.70
8	5	0.78
9	5	0.78
10	6	0.85
11	7	0.90
12	7	0.90
13	8	0.95
14	9	1.00
15	9	1.00
16	10	1.04
17	10	1.04
18	11	1.08
19	12	1.11
20	12	1.11
21	15	1.20
22	16	1.23
23	20	1.32
24	23	1.38
25	29	1.48
26	34	1.54
27	49	1.70
28	67	1.83

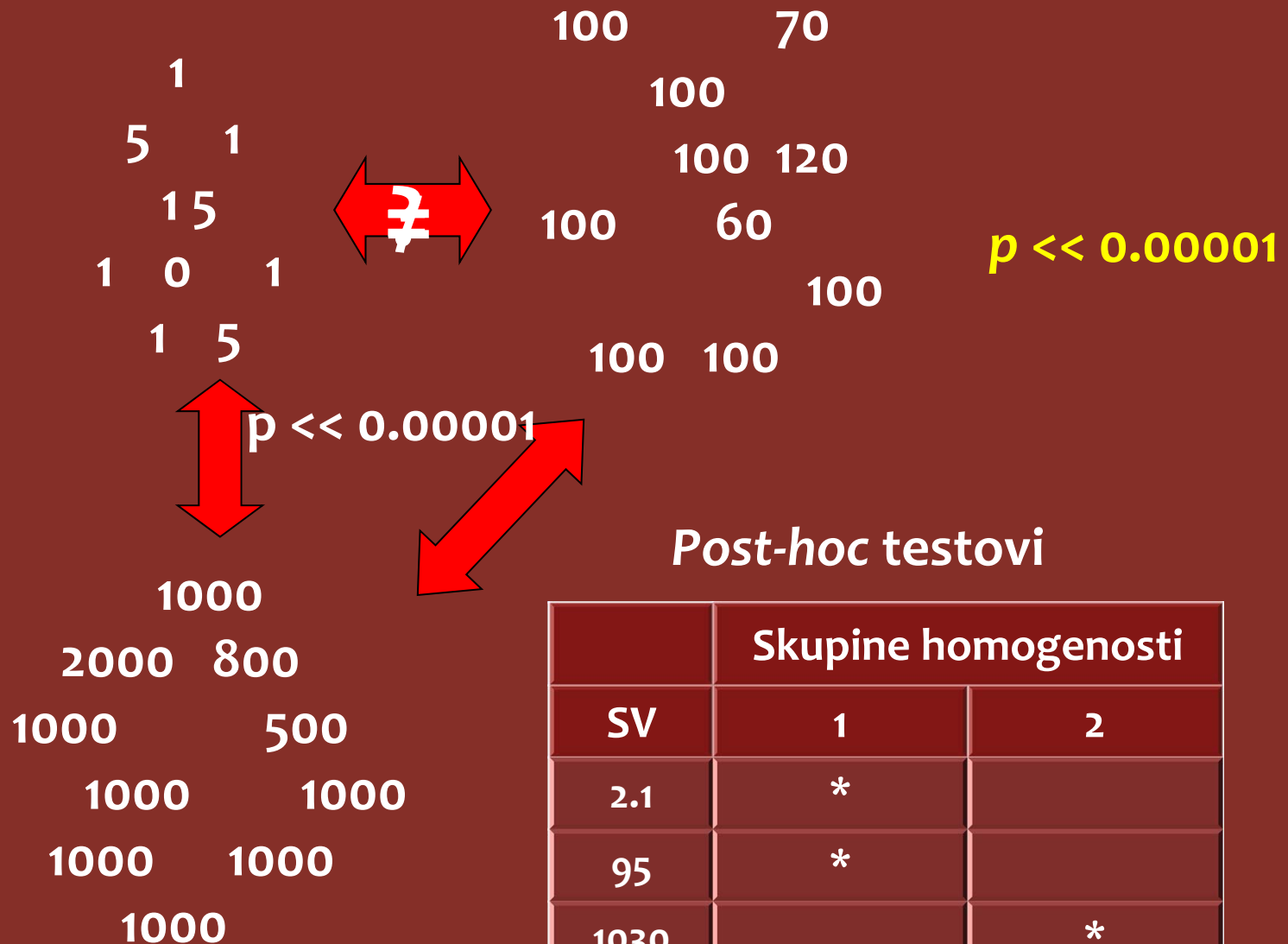
Mjereni podaci



Logaritamski transformirani podaci



Primjer: Analiza varijance



Primjer: Analiza varijance

Transformiranje podataka logaritmiranjem

1	100	70
5	1	100
15	100	120
1	0	1
1		100
	100	100

1000	
2000	800
1000	500
1000	1000
1000	1000
1000	

Primjer: Analiza varijance

Transformiranje podataka logaritmiranjem

0.3	2	1.9
0.7 0.3	2	
0.3 0.7	2 2.1	
0.3 0 0.3	2 1.8	
0.3 0.7		2
	2 2	

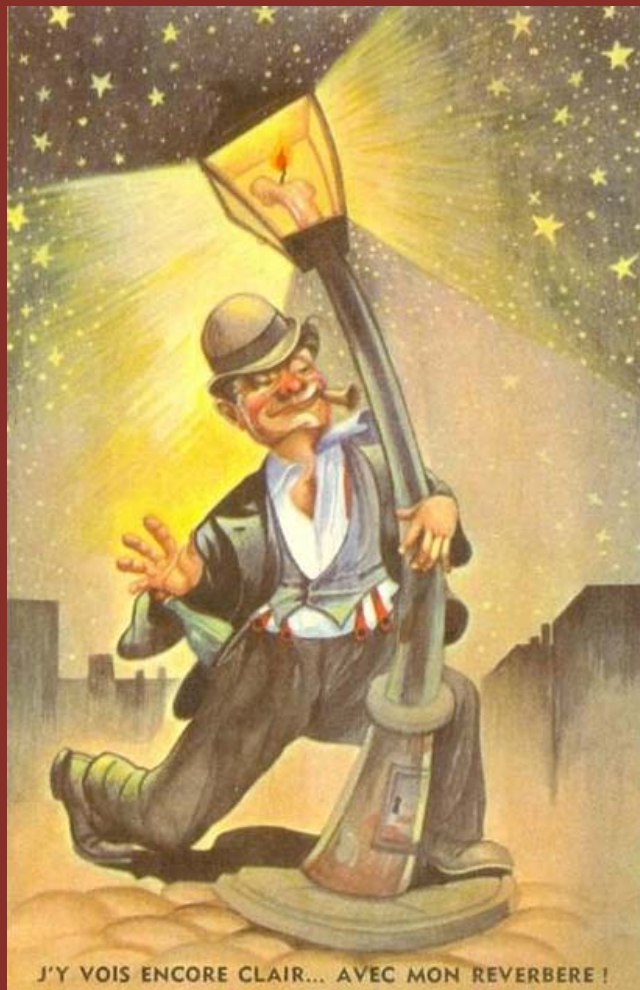
	3
3.3	2.9
3	2.7
3	3
3	3
	3

Post-hoc testovi

	Skupine homogenosti		
SV	1	2	3
0.41	*		
1.97		*	
2.99			*

**Some individuals use statistics as a drunk man uses
lamp post - for support rather than for illumination.**

Andrew Lang



Primjer naslova i interpretacije rezultata

Međudnosi brojnosti svojta i okolišnih čimbenika izraženi Pearsonovim koeficijentom korelacije;
Statistički značajne korelacije ($p < 0.05$) otisnute su masno.

	Lumbriculidae	Ancylus	Amphinemura
T	0.28	0.20	-0.28
O ₂	-0.31	-0.22	0.24
KPK	-0.32	0.24	0.45
pH	-0.32	0.72	0.25
Chl a	-0.14	0.98	0.16

Pearsonovim koeficijentom korelacije na razini statističke značajnosti $p < 0.05$ utvrđeni su negativni odnosi između brojnosti jedinki porodice Lumbriculidae i koncentracije kisika, količine otopljene organske tvari iskazane kemijskom potrošnjom kisika (KPK) i pH. Pozitivni odnosi utvrđeni su između brojnosti populacije roda Ancylus i pH te količine klorofila a, odnosno brojnosti jedinki roda Amphinemura i KPK.

Primjeri naslova i interpretacije rezultata

Usporedba staništa različitih brzina strujanja vode s obzirom na sadržaj detritusa analizom varijance.

	SS	Stupnjevi slobode	MS	F	p
Brzina strujanja vode	0.08	3	0.08	6.56	0.014

Analizom varijance utvrđena je statistički značajna razlika ($p = 0.014$) u količini detritusa između staništa različite brzine strujanja vode.

NAPOMENA: nakon rezultata ove analize ne znamo gdje je više detritusa, tako da je zaključak neodređen - razliku smo utvrdili, ali ne i njena svojstva. Ako se radi o samo dva staništa - i sam navod srednjih vrijednosti bi bio dovoljan da možemo zaključiti određenije npr.: Na staništu x nakupljeno je statistički značajno više detritusa nego na staništu y.

Ali potpuno ispravan način za više staništa jest napraviti post-hoc test.

Primjeri naslova i interpretacije rezultata

Post-hoc Tukey HSD test za utvrđivanje razlike u masi nakupljenog detritusa između četiri staništa različite brzine strujanja vode.

brzina strujanja vode	srednja masa detritusa	skupine homogenosti	
		1	2
< 30 cm s ⁻¹	0.51	x	
30-60 cm s ⁻¹	0.65	x	
60-90 cm s ⁻¹	0.71	x	
>90 cm s ⁻¹	1.25		x

Post-hoc Tukey HSD testom utvrđeno je da je statistički značajno više detritusa u staništu s najbržom strujom vode. Među ostalim staništima nema statistički značajne razlike u količini nakupljenog detritusa (i sva tri Preostala staništa svrstana su u istu skupinu homogenosti).