

Sveučilišni diplomski studij medicinsko-laboratorijske dijagnostike  
Kolegij: Medicinska informatika u kliničko-laboratorijskoj dijagnostici  
(MIKLD 2013./14.)

Mladen Petrovečki  
Lidija Bilić-Zulle

## Obradba podataka



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

## Priprema podataka za obradbu

1. vrsta podataka → izbor testa
2. naziv varijable (plus opis varijable)
3. šifriranje (binariziranje, "nula" = ne)
4. nedostatak podatka → prazno polje
5. preračunavanje složenih podataka



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

## Programska potpora

The screenshot shows the SPSS Data Editor interface with several menu options highlighted. The 'Analyze' menu is open, showing options like 'Descriptive Statistics', 'Regression', 'Loglinear', 'Classify', 'Data Reduction', 'Scale', 'Nonparametric Tests', 'Survival', 'Multiple Response', 'Kaplan-Meier', 'Cox Regression...', and 'Cov. w/ Time-Dep. Cov...'. The 'Survival' menu is also open, showing options like 'Life Tables...', 'Kaplan-Meier...', 'Cox Regression...', and 'Cov. w/ Time-Dep. Cov...'. The 'Survival and Failure Time Analysis: labl...' dialog box is also visible, showing options for 'Life Tables & Distributions', 'Kaplan & Meier product limit method', 'Comparing two samples', 'Comparing multiple samples', 'Regression models', and 'Time-dependent covariates'.

## Dvodimenzijaska tablica podataka

	A	B	C	D	E	F	G	H	I
1	Broj	Doz	Receptol (fmol/mg)	Pratičljenje					
2	biopsije	(god)	estrogen	progest	mjeseci	cenusus	Veličina tumora (cm)	Iskapsina	Histološki tip
3	BRBP	DOB	ER	PR	PRE	CEN	VTVM	VTSK	HTIP
4	1931	55	21	8	156	1	5,4	3	1
5	2017	54	214	31	101	1	3,7	2	2
6	2116	58	58	14	62	1	1,2	1	1
7	2144	44	42	71	38	1	1,0	1	1
8	2213	66	9	5	1	1	3,0	2	1
9	2481	71	216	451	22	1	3,3	2	2
10	2854	60	203	16	59	1	1,3	1	1
11	3084	62	24	20	119	1	4,0	2	2
12	3648	65	102	15	51	1	4,2	2	1
13	4322	68	248	120	118	1	1,1	1	1
14	4497	44	16	40	72	1	6,2	3	1
15	4652	42	15	44	18	0	4,1	2	1
16	4726	36	.	.	71	1	2,8	2	1
17	4890	33	13	38	181	0	4,2	2	1
18	4970	64	52	19	60	1	2,1	2	3
19	4986	54	24	1	59	0	2,0	2	1

Slika 8-1. Struktura podataka dvodimenzionalne tablice. Tablica je izvatak podataka iz studije čimbenika rizika u 164 žene s inozimnim karcinomom dojke. Redovi tablice su bolesnice, a stupci pokazatelji. (Belev B, Alerić I, Vrbanec D, Petrovečki M, Unušić J, Jakić-Razumović J. Nm23 gene product expression in invasive breast cancer – immunohistochemical analysis and clinicopathological correlation. Acta Oncol 2002;41:355-61.)



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

## Prikaz podataka

- prikaz podataka u znanstvenom radu → sažimanje
- tablice i slike (grafikoni)
- kvalitativni podatci
  - apsolutne i relativne frekvencije
  - kontingencijske tablice, stupčasti i okrugli grafikoni
- kvantitativni podatci
  - mjere prosjeka i odstupanja
  - grafikon okvira s ručicama



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

## Tablični prikaz podatka

- stvarna vrijednost podataka
- logički dobro organizirana velika količina podataka → jednostavno snalaženje i čitanje podataka
- tablica – samostalna cjelina:
  - naslov tablice
    - svrha tablice, ustroj istraživanja
    - iznad tablice, numeriran, bez točke na kraju
  - legenda
    - ispod tablice, opis kratica i simbola (paziti na upute časopisa)
  - jasni naslovi stupaca



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

## Slikovni prikaz podataka

- prikaz međusobnog odnosa podataka
- svrha: jasno tumačenje rezultata
- naslov slike (grafikoni, crteži, sheme...)
  - ispod slike, numeriran, s točkom na kraju
  - legenda
  - broj opažanja i vrsta prikazane pogreške
- nije dopušteno:
  - rabiti 3D grafikone
  - prikazivati mjere prosjeka bez mjera odstupanja



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Kvalitativni podatci

- utvrđuje se kvaliteta promatranog obilježja i svrstava se u jednu od kategorija → nominalna ili ordinalna ljestvica
- opis podataka:
  - apsolutna učestalost (broj)
  - relativna učestalost (udio, postotak)
- prikaz podataka:
  - kontingencijske tablice
  - stupčasti grafikoni



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Apsolutna i relativna učestalost

- apsolutna učestalost (broj pojavnosti)
  - NE ukazuju na odnose među skupinama
- relativna učestalost (udjeli)
  - osim ukupnog broja pokazuju i odnos među skupinama

primjer: jednodimenzijska tablica s apsolutnim i relativnim učestalostima

Pušenje	N (%)
da	180 (62)
ne	110 (38)
ukupno	290 (100)



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Kontingencijske tablice

- prikazuju obilježja mjerena nominalnom i ordinalnom ljestvicom (kvalitativni podatci)
- dvodimenzijske: pokazuju podatke dvaju ili više obilježja

primjer: kontingencijska tablica s prikazom apsolutnih i relativnih učestalosti pušača s obzirom na spol u skupini ispitanika (N=290)

Pušenje	spol	
	žene N (%)	muškarci N (%)
da	120 (80)	60 (43)
ne	30 (20)	80 (57)
ukupno	150 (100)	140 (100)

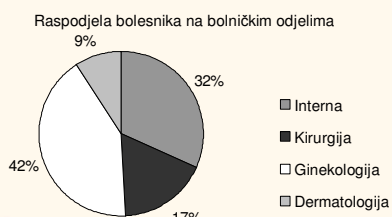


Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Kružni grafikon

- prikaz relativnih učestalosti (ukupan zbroj prikazanih podataka 100%)

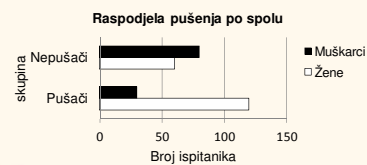


Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Stupčasti grafikon

- koristi se za prikaz apsolutnih učestalosti (histogram frekvencija)
- položeni ili okomiti stupci



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Kvantitativni podatci

- brojni podatci (slijede intervalnu i omjernu ljestvicu)
- prikaz podataka → mjere prosjeka i odstupanja

mjere prosjeka (središnjice)

aritmetička sredina  
medijan (centralna vrijednost)  
mod (dominantna vrijednost)

mjere odstupanja (raspršenja)

standardna devijacija  
raspon, kvartili, percentili  
---

- slikovni prikaz podataka:
  - grafikon okvira s ručicama



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Mjere središta i raspršenja

- najčešće se upotrebljavaju:
  - aritmetička sredina i standardna devijacija
    - normalna raspodjela
    - veliki uzorak
  - medijan i raspon (percentili, interkvartili)
    - ne-normalna raspodjela
    - mali uzorak
  - interval pouzdanosti (i za aritmetičku sredinu i za medijan)
- rjeđe upotrebljavani: varijanca, računski raspon, standardna pogreška



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Aritmetička sredina

ispitanik →	1	2	3	4	5	6	7	8	9
broj popušenih cigareta →	5	7	7	8	8	9	9	9	10

$$\bar{x} = \frac{5 + 7 + 7 + 8 + 8 + 9 + 9 + 9 + 10}{9} = \frac{72}{9} = 8$$



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Medijan

ispitanik →	1	2	3	4	5	6	7	8	9
broj popušenih cigareta →	5	7	7	8	8	9	9	9	10

- vrijednost koja se u nizu podataka nalazi točno u sredini
- dijeli skup na dva jednaka dijela
- ako je N = paran broj
  - medijan = aritmetička sredina dva središnja podatka



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Mod

ispitanik →	1	2	3	4	5	6	7	8	9
broj popušenih cigareta →	5	7	7	8	8	9	9	9	10

primjer: aritmetička sredina = 8; medijan = 8; mod = 9

- najučestalija vrijednost u nizu podataka
- ako su podatci grupirani u razrede:
  - mod = aritmetička sredina razreda s najvećom učestalošću



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Standardna devijacija

$$S = \sqrt{\frac{\sum(X - \bar{X})^2}{N - 1}}$$

- prosječno odstupanje od aritmetičke sredine
- smije se računati i prikazivati SAMO uz aritmetičku sredinu ( $X \pm SD$ )



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Raspon

- raspon od najmanje (minimuma) do najveće (maksimuma) izmjerene vrijednosti ( $X_{\min} - X_{\max}$ )
- moguće izraziti i kao brojčanu vrijednost razlike između najveće i najmanje vrijednosti u nizu:  
 $R = X_{\max} - X_{\min}$
- najmanje informativna mjera odstupanja
- ekstremni rezultati mogu znatno povećati raspon



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Kvartili

1	2	...	...	40	41	42	...	...	81	82
18	30			34		39			63	
min	1. kvartil		2. kvartil		3. kvartil		max			
	25. percentil		medijan		75. percentil					

- kvartil je rezultat koji se nalazi točno na četvrtini raspona
- interkvartilni raspon (IR):
  - raspon 25. i 75. percentila (ili 1. i 3. kvartila)
  - razlika vrijednosti 3. i 1. kvartila



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Percentili

- raspon podijeljen na sto dijelova – percentila
- označava postotak raspodjele
- 1. percentil = najmanja vrijednost (minimum)
- 25. percentil = 1. kvartil
- 50. percentil = medijan = 2. kvartil
- 75. percentil = 3. kvartil
- 100. percentil = najveća vrijednost (maksimum)

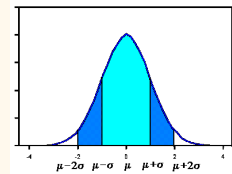


Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Raspodjela podataka

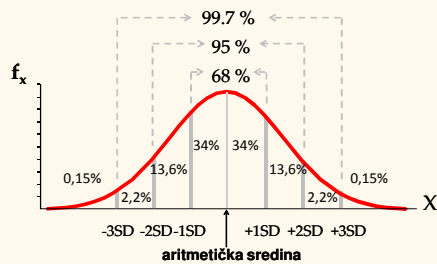
- grafički prikaz mjerenja prikazan samostalno ili u odnosu na druge podatke
- prikazuje:
  - raspon mjerenja
  - jednolikost raspodjele
  - modalnost raspodjele
  - simetričnost raspodjele
- normalna raspodjela



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Normalna raspodjela



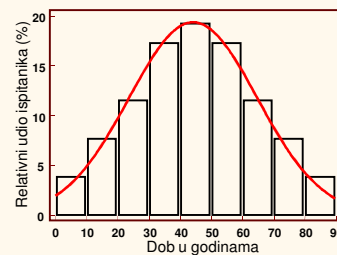
Raspodjela podataka prikazuje se histogramom frekvencije. Os x prikazuje promatrano obilježje, a os y frekvenciju promatranog obilježja.



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Normalna raspodjela



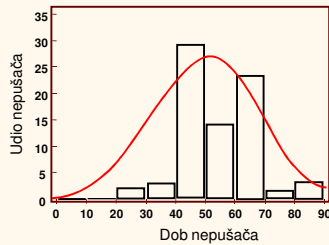
- podatci se opisuju s pomoću aritmetičke sredine i standardne devijacije



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Ne-normalna raspodjela



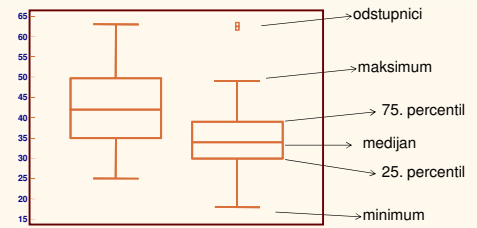
- podaci se opisuju s pomoću medijana i raspona



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Grafikon okvira s ručicama



vrijednosti sedimentacije eritrocita prije (SE0) i tri tjedna poslije terapije (SE3)

	Medijan	Minimum	Maximum	25. - 75. perc.
SE0	42	25	63	35 - 50
SE3	34	18	63	30 - 63



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Odstupnici

- vrijednosti koje odstupaju (engl. *outlier*)
- kriterij za odstupnike:
  - 25. percentil  $-1,5 \times IR$  ( $<$  prosječne vrijednosti)
  - 75. percentil  $+1,5 \times IR$  ( $>$  prosječne vrijednosti)
- "daleki" odstupnici
  - 25. percentil  $-3 \times IR$  ( $<$  prosječne vrijednosti)
  - 75. percentil  $+3 \times IR$  ( $>$  prosječne vrijednosti)
- provjera
  - pogriješka mjerenja i/ili upisivanja?
  - djelovanje neprepoznatog čimbenika?
  - **NE smiju se izbrisati niti zatajiti**



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Odabir testa

- testovi dokazivanja statističkih hipoteza
- kriterij odabira
  - osobine obilježja  $\rightarrow$  mjerne ljestvice
  - osobine uzorka
    - veličina
    - povezanost
  - osobine raspodjele (normalnost)
    - parametrijski testovi
    - neparametrijski testovi
- broj obilježja  $\rightarrow$  uni/bi/multivarijantni testovi



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Ovisnost o broju pokazatelja

- univarijantni
  - raščlanjuju samo jedan pokazatelj
- bivarijantni
  - raščlanjuju dva pokazatelja
- multivarijantni
  - raščlanjuju više pokazatelja



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Ovisnost o raspodjeli

- parametrijski testovi
  - pokazatelji koji slijede normalnu raspodjelu
  - obrađuju izvorna mjerenja
- neparametrijski testovi
  - nisu ograničeni normalnošću raspodjele
  - ne raščlanjuju izvorne podatke nego položaje (rangove)



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Ovisnost o osobini uzorka

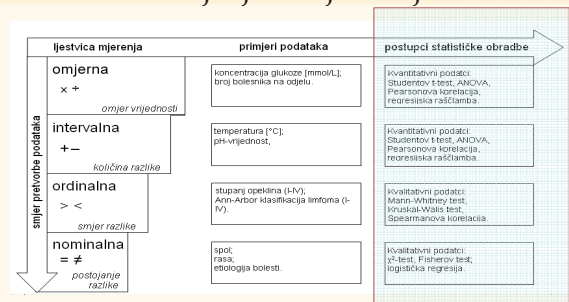
- veličina
  - veliki uzorci → parametrijski testovi
  - mali uzorci → neparametrijski testovi
- zavisnost
  - nezavisni uzorci → neparni testovi
  - zavisni uzorci → parni testovi



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Osobine obilježja – mjerne ljestvice



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Biranje testa

VRSTA UZORKA		VRSTA PODATAKA		
		brojčani		kategorički
		omjerni, intervalni	ordinalni	Nominalni
NEZAVISNI	2 skupine različitih jedinki	t-test za nezavisne uzorke	Mann-Whitneyjev test	t-test za proporcije ili χ² test
	3 ili više skupina različitih jedinki	analiza varijance ANOVA	Kruskal-Wallisov test	
ZAVISNI	prije i poslije	t-test za zavisne uzorke	Wilcoxonov test	McNemarov test
	višestruko ponavljanje	ANOVA za ponavljana mjerenja	Friedmanov test	Cochranov Q test
POVEZANOST PODATAKA		Korelacijska analiza	Spearmanova korelacija	omjer izgleda relativni rizik



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Biranje testa

Pitanje	1. broj poraznuta	2. povećanje ili razlika	3. uključene skupine	4. mjerna ljestvica	5. raspodjela	6. broj skupina	7. broj jedinki	Odgovor
Smjer traženja odgovora	razlika	neuklađene (neparni testovi)	omjerna intervalna ordinalna	nominalna	nije normalna	2	-	Studentov t-test
						>2	-	jednorigma ANOVA
						2	-	Mann-Whitneyjev test
						>2	-	Kruskal-Wallisov test
jedna ili dva	uklađene (parni testovi)	omjerna intervalna ordinalna	nominalna	nije normalna	2	<20	Fisherov test	
					2	≥20	χ²-test	
					2	-	parni t-test	
					2	-	F-oner	
povezanost	-	omjerna intervalna ordinalna	nominalna	-	2	-	Wilcoxonov test parova	
					>2	-	Friedmanov test	
					-	-	McNemarov test	
					-	-	Pearsonov r	
								Spearmanov r
								Cramerov V



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



## Tumačenje i prikaz rezultata

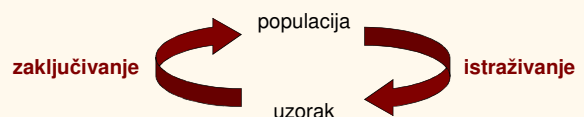
- Poopćavanje s uzorka na populaciju
- Kontingencijske tablice i mjere središnjice
- Tablični i slikovni prikaz podataka
  - frekvencije: stupčasti grafikon
  - središnjice i rasap: kutije i brkovi (*box and whisker*)
- Vrijednosti koje odskaču (*outliers*)
- Raspon pouzdanosti (*confidence interval, CI*)
- **statističko i logičko zaključivanje**



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



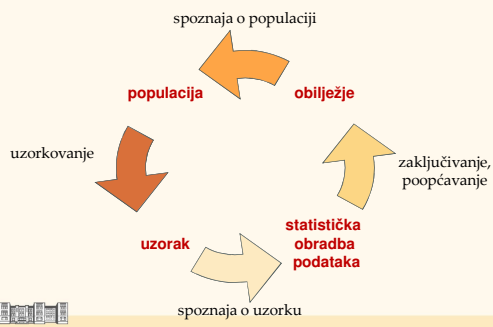
## Populacija, uzorak



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



# Populacija → uzorak → populacija



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



[www.cartoonstock.com](http://www.cartoonstock.com)



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

