

Kolegij: ROLP

Statistička terminologija I. - raspodjela i prikaz podataka

2017.



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Neki temeljni statistički postupci u znanstvenom istraživanju

- odabir uzorka
- prikupljanje podataka
 - određivanje mjerne ljestvice
- obrada podataka
 - raspodjela podataka
 - opis podataka (prikazi, ...)



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

VRSTE PODATAKA (mjerne ljestvice)

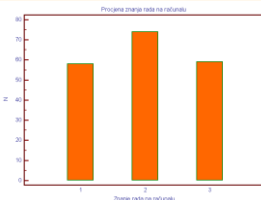
| KVANTITATIVNI (BROJČANI) | |
|---|--|
| OMJERNA BMI Krvni tlak Razina glukoze u krvi | INTERVALNA Broj djece Broj napadaja tjedno Dob (god) |
| KVALITATIVNI (SKUPNI, KATEGORIČKI) | |
| ORDINALNA Stupanj karcinoma Bolje, jednako, lošije Ne slažem se, neopredijeljen/a, slažem se (Likertova ljestvica (stavovi)) | NOMINALNA Spol Pozitivan i negativan nalaz Krvna grupa Pripadnost rasi |



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Kvalitativni podatci - nominalna ljestvica

- frekvencije i relativne frekvencije - $N(\%)$
- grafički prikaz - stupčasti grafikoni



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Kvalitativni podatci - ordinalna ljestvica

- sve kao kod nominalne ljestvice ali i mjere središnjice i raspršenja - medijan, mod



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Medcalc - kategorijski podatci

The screenshot shows the MedCalc interface with the 'Frequency table & Chi-s...' dialog box open. The dialog box has fields for 'Codes X:' and 'Codes Y:', and a 'Subst:' field. Below the dialog box, the 'Frequency table & Chi-square test' window displays the following data:

| Code X | spol | |
|--------|------|--------|
| 1 | 75 | 61.9% |
| 2 | 114 | 93.1% |
| Total | 189 | 100.0% |

Below the table, the Chi-square test results are shown:

| | |
|--------------------|------------|
| Chi-square | 5.990 |
| DF | 1 |
| Significance level | P = 0.0144 |



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

Kada se ispituje vrsta raspodjele?

... SAMO kada podaci vaših istraživanja pripadaju
INTERVALNOJ ili
OMJERNOJ mjerne ljestvici

- kontinuirani broječni podatci:
vrijednosti hemoglobina, broj
leukocita u krvi, tjelesna težina,
stav prema radu...



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Zašto je važno znati kakva je raspodjela podataka?

- za odabir mjere središnjice i mjere varijabilnosti
- za odabir statističkog testa
- u slikovnom prikazu daje jasan uvid u podatke



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Mjere središnjice i odstupanja

Najčešće se upotrebljavaju:

- aritmetička sredina i standardna devijacija (normalna raspodjela, velik uzorak)
- medijan i raspon (percentili, interkvartili)
- interval pouzdanosti (aritmetičku sredina, medijan)

Rjeđe se upotrebljavaju:

- varijanca, raspon, standardna pogreška, mod



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Mjere središnjice

Prikazuju prosječnu vrijednost neke skupine rezultata

- Aritmetička sredina
- Medijan (središnja vrijednost)
- Mod (najčešća vrijednost)



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Aritmetička sredina (M, \bar{X})

- srednja ili prosječna vrijednost

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_N}{N}$$

- Prosječna visina studenata
- Srednja ocjena svih ispitanika na testu znanja



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Izračunavanje aritmetičke sredine

| | | | | |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| 7 | 8 | 7 | 5 | 8 |

$$\bar{X} = \frac{5+7+7+8+8}{5} = \frac{35}{5} = 7$$



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Medijan (C)

- centralna vrijednost
- vrijednost koja se u nizu rezultata poredanih po rastućem iznosu nalazi točno u sredini

| | | | | | | | | |
|----|----|----|-----|----|-----|----|----|----|
| 1 | 2 | 3 | ... | 26 | ... | 49 | 50 | 51 |
| 10 | 18 | 20 | ... | 52 | ... | 80 | 81 | 90 |

- Raspodjela ispitanika prema životnoj dobi



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Mod (D)

- najčešća vrijednost u nizu rezultata poredanih po rastućem iznosu

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 5 | 6 | 6 | 6 | 6 | 7 | 8 | 8 | 9 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|

- Vrijeme provedeno u bolnici



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Mjere raspršenja

...iskazuju odstupanja rezultata oko neke od mjera središnjice

- standardna devijacija
- raspon
- kvartili, percentili



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Standardna devijacija (SD)

- vrijednost prosječnog odstupanja od aritmetičke sredine

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

- prikaz rezultata mjerenja ($\bar{x} \pm 1 SD$)
- 95% granice pouzdanosti ($\bar{x} \pm 1,96 SD$)
- varijanca = SD^2



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Raspon (R)

- površna mjera raspršenja (varijabilnosti)
- razlika između najveće i najmanje vrijednosti rezultata

$$R = X_{\text{najveći}} - X_{\text{najmanji}}$$

| | | | | | | | | |
|----|----|----|-----|----|-----|----|----|----|
| 1 | 2 | 3 | ... | 26 | ... | 49 | 50 | 51 |
| 10 | 18 | 20 | ... | 52 | ... | 80 | 81 | 89 |

- Raspon životne dobi od 10-89; R=79



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Kvartil

| | | | | | | | | |
|-----|------------|-----------------------|------------|-----|-----|----|----|----|
| 1 | 2 | 5 | ... | 26 | ... | 47 | 50 | 51 |
| 10 | 18 | 34 | ... | 52 | ... | 71 | 81 | 89 |
| Min | 1. kvartil | 2. kvartil medijan | 3. kvartil | max | | | | |

- Kvartil je rezultat koji se nalazi točno na četvrtini raspona
- Interkvartilni raspon (IR)
 - Raspon 25 i 75 percentila (1 i 3 kvartila)
 - Razlika vrijednosti 3. i 1. kvartila



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Percentil

- vrijednost odstupanja od medijana
- određena mjestom gdje se nalazi u nizu rezultata

| | | | | | | | | | | |
|----|----|-----|----|-----|----|-----|----|-----|----|----|
| 1 | 2 | ... | 5 | ... | 26 | ... | 47 | ... | 50 | 51 |
| 10 | 18 | ... | 34 | ... | 52 | ... | 71 | ... | 81 | 89 |

10. percentila

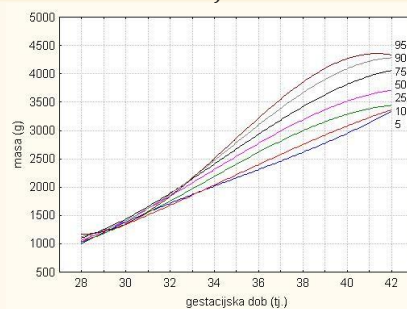
90. percentila



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Percentilna krivulja



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Vrsta raspodjele određuje odabir mjere središnjice i mjere raspršenja

Normalna raspodjela: M i SD

Srednja vrijednost membranskog biopotencijala iznosila je $(70 \pm 3,5)$ mV.

Odstupajuća raspodjela: medijan i percentil

Prosječna životna dob ispitanika iznosila je 52 (34 - 71) godine.



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Prikaz rezultata deskriptivne statistike u programu *MedCalc*

| Variable: Dob | |
|---|----------------------------|
| Sample size | 312 |
| Lowest value | 33,0000 |
| Highest value | 83,0000 |
| Arithmetic mean | 49,2219 |
| 95% CI for the mean | 46,0737 to 50,3351 |
| Median | 53,0000 |
| 95% CI for the median | 51,2518 to 56,0000 |
| Variance | 364,9977 |
| Standard deviation | 19,1349 |
| Relative standard deviation | 0,3864 (38,64%) |
| Standard error of the mean | 1,0816 |
| Coefficient of Skewness | -0,6084 (P<0,0001) |
| Coefficient of Kurtosis | -0,9308 (P=0,1169) |
| Kolmogorov-Smirnov test for Normal distribution | reject Normality (P<0,001) |
| Percentiles | |
| 2,5 | 13,0000 |
| 5 | 14,0000 |
| 10 | 15,0000 |
| 25 | 36,0000 |
| 50 | 53,0000 |
| 75 | 78,0000 |
| 90 | 79,0000 |
| 95 | 71,8000 |
| 97,5 | 74,7000 |
| 95% Confidence Interval | |
| 2,5 | 11,3530 to 14,0000 |
| 5 | 13,0000 to 15,0000 |
| 10 | 14,0000 to 17,0000 |
| 25 | 27,0000 to 41,0000 |
| 50 | 60,8890 to 65,0000 |
| 75 | 68,0000 to 71,0000 |
| 90 | 70,0000 to 74,0673 |
| 95 | 71,8878 to 78,2922 |



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Krivulja normalne raspodjele ili Gaussova krivulja

- simetrična u odnosu na aritmetičku sredinu
- zvonolika
- oblika koji ovisi o aritmetičkoj sredini i standardnoj devijaciji

Svojstva normalne raspodjele:

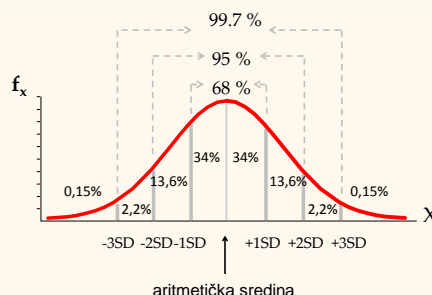
- $(x \pm 1 \text{ SD}) \rightarrow 68,26\%$
- $(x \pm 2 \text{ SD}) \rightarrow 95,44\%$
- $(x \pm 3 \text{ SD}) \rightarrow 99,73\%$



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



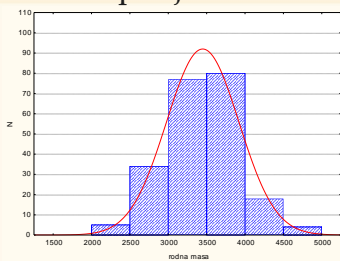
Normalna raspodjela



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Normalna raspodjela



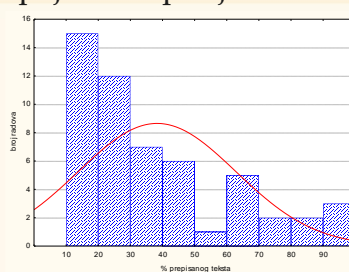
- podaci se opisuju s pomoću aritmetičke sredine i standardne devijacije



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Odstupajuća raspodjela



- podaci se opisuju s pomoću medijana i percentila (kvartila, raspona)



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Kako ispitati normalnost raspodjele?

- Kolmogorov-Smirnovljev test



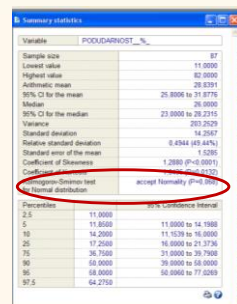
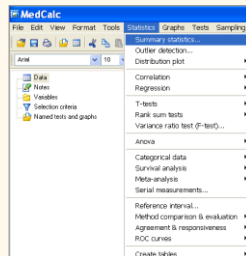
- $P \geq 0,05$ empirijska raspodjela je normalna
- $P < 0,05$ empirijska raspodjela **značajno odstupa od normalne raspodjele**



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Testiranje normalnosti raspodjele



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Vrsta raspodjele određuje odabir statističkog testa

| | podatci | broj skupina | test |
|------------------------|-----------|--------------|--------------|
| Normalna raspodjela | Nezavisni | 2 | t-test |
| | Zavisni | | parni t-test |
| Odstupajuća raspodjela | Nezavisni | 2 | Mann-Whitney |
| | Zavisni | | Wilcoxon |



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Koji prikaz podataka odabrati?

Ovisi o

- podacima
- populaciji kojoj se prezentiraju
- našem znanju i vještinama

Najčešće se koriste

- tablice
- slike (grafikoni)



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Tablični prikaz podataka

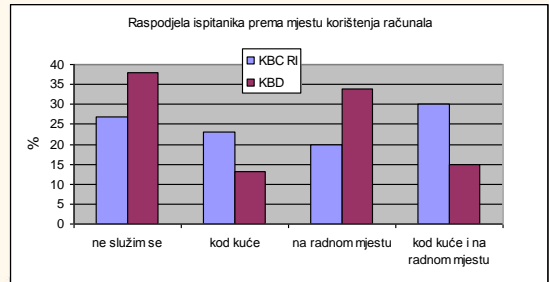
| | | | | |
|-----|------------|-----|------|-----|
| RB | skupina | dob | spol | ... |
| 1 | kontrola | 56 | M | ... |
| ... | ... | ... | ... | ... |
| 1 | ispitanici | 45 | Ž | ... |
| ... | ... | ... | ... | ... |



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Vertikalni stupčasti grafikon (2 skupine)

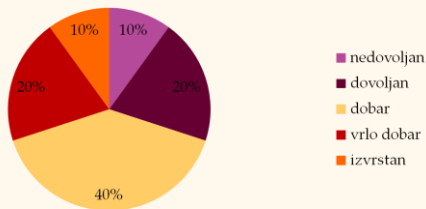


Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Kružni dijagram

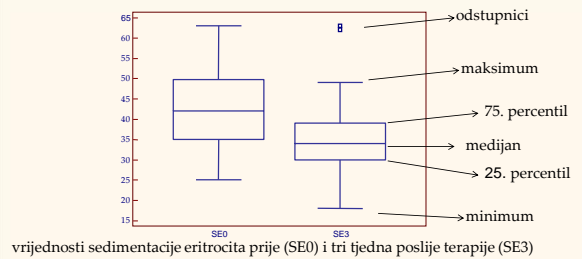
Raspodjela ocjena iz medicinske informatike



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



Grafikon okvira s ručicama



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

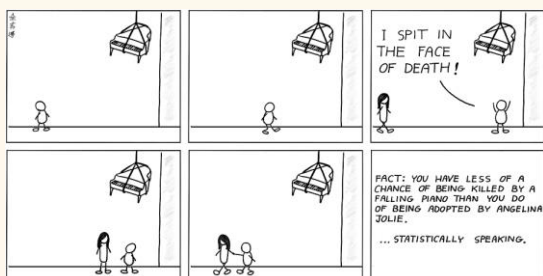
| | Medijan | Minimum | Maximum | 25. - 75. perc. |
|-----|---------|---------|---------|-----------------|
| SE0 | 42 | 25 | 63 | 35 - 50 |
| SE3 | 34 | 18 | 63 | 30 - 63 |



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku



I na kraju ...



Medicinski fakultet Sveučilišta u Rijeci Katedra za medicinsku informatiku

