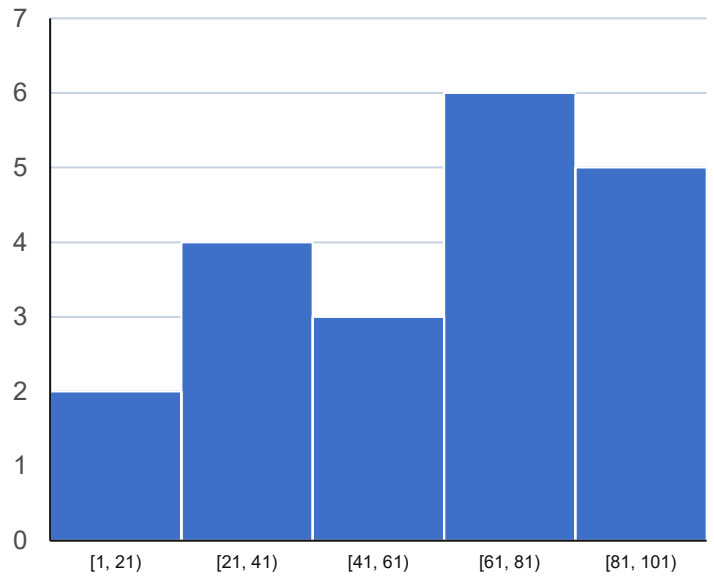
ALATI ZA OBRADU PODATAKA

Vizuelizacija numeričkih podataka

Intervali, tabela frekvencija i histogram



20 vrednosti sažeto u 5 intervala

Predavač: dr Dušan Stefanović

Vrsta promenljive određuje način prikaza

VEZA SA PRETHODNOM LEKCIJOM

KATEGORIJSKI PODACI

Marka automobila
Audi, BMW, Mercedes

Stubovi predstavljaju odvojene kategorije.

NUMERIČKI PODACI

Vreme putovanja
32, 35, 41 minuta

Vrednosti grupišemo u brojčane intervale.

Kod numeričkih podataka posmatramo raspodelu vrednosti, a ne samo naziv kategorije.

Šta treba razumeti

ISHODI UČENJA

1. zašto numeričke vrednosti grupišemo u intervale
2. kako biramo broj intervala i računamo njihovu širinu
3. kako određujemo kojem intervalu pripada granična vrednost
4. kako računamo apsolutnu i relativnu frekvenciju
5. kako gradimo i tumačimo histogram
6. po čemu se histogram razlikuje od stubičastog dijagrama

Veliki broj različitih vrednosti zahteva sažimanje

ZAŠTO GRUPIŠEMO PODATKE

Ako je svaka vrednost različita

20 vrednosti
20 redova u tabeli
frekvencija svake vrednosti = 1

Ako vrednosti grupišemo

20 vrednosti
5 intervala
frekvencije: 2, 4, 3, 6 i 5

Grupisanje smanjuje broj redova i otkriva gde se vrednosti najviše nagomilavaju.

Primer sadrži 20 različitih brojeva

POČETNI SKUP PODATAKA

1	2	3	4	5
1	17	22	25	35
40	43	52	57	62
65	69	75	78	80
83	88	92	97	100

Najmanja vrednost = 1

Najveća vrednost = 100

Tabela po pojedinačnim vrednostima ne otkriva raspodelu

PROBLEM BEZ INTERVALA

Vrednost	Frekvencija
1	1
17	1
22	1
25	1
...	...
100	1

Šta ne vidimo?

gde se nalazi najviše vrednosti

kakav je oblik raspodele

koji deo skale je redak

Intervali pretvaraju dugačak spisak u preglednu raspodelu frekvencija.

Interval grupiše susedne vrednosti

OSNOVNA IDEJA

1 do <21	21 do <41	41 do <61	61 do <81	81 do <101
2 vrednosti	**4 vrednosti**	**3 vrednosti**	**6 vrednosti**	**5 vrednosti**

Brojčana osa je podeljena na susedne delove jednake širine.

Svaka vrednost mora pripadati tačno jednom intervalu.

Broj intervala određuje nivo detalja

IZBOR BROJA INTERVALA

PREMALO INTERVALA

Prikaz je grub.
Razlike u podacima mogu ostati skrivene.

5 DO 20 INTERVALA

Često daje pregledan početni prikaz.
Izbor zavisi od veličine skupa.

PREVIŠE INTERVALA

Mnogi stubovi imaju malu frekvenciju.
Prikaz postaje nestabilan.

Ne postoji jedan broj intervala koji je najbolji za svaki skup podataka.

Širina intervala polazi od raspona podataka

FORMULA

$$\text{širina intervala} = \frac{\text{najveća vrednost} - \text{najmanja vrednost}}{\text{željeni broj intervala}}$$

Raspon = najveća vrednost − najmanja vrednost

Širinu često zaokružujemo naviše na pregledan broj.

Pet intervala daje širinu približno 20

PRIMER RAČUNANJA

Najveća vrednost	Najmanja vrednost	Broj intervala
100	1	5

$$(100 - 1) / 5 = 99 / 5 = 19,8$$

Za pregledne granice biramo širinu 20.

Pravilo granica sprečava dvostruko brojanje

INTERVALNE GRANICE

[21, 41)

uključuje 21

ne uključuje 41

Vrednost	Interval
20	[1, 21)
21	[21, 41)
40	[21, 41)
41	[41, 61)

Koristimo intervale zatvorene sa leve, a otvorene sa desne strane.

Svaka vrednost ulazi u jedan interval

RASPOREĐIVANJE PODATAKA

Interval	Vrednosti u intervalu	Broj
[1, 21)	1, 17	2
[21, 41)	22, 25, 35, 40	4
[41, 61)	43, 52, 57	3
[61, 81)	62, 65, 69, 75, 78, 80	6
[81, 101)	83, 88, 92, 97, 100	5

Kontrola: 2 + 4 + 3 + 6 + 5 = 20 vrednosti.

Tabela frekvencija sažima numeričku raspodelu

APSOLUTNA FREKVENCIJA

Interval	Frekvencija
1 do ispod 21	2
21 do ispod 41	4
41 do ispod 61	3
61 do ispod 81	6
81 do ispod 101	5
Ukupno	20

Najveća frekvencija je 6 u intervalu od 61 do ispod 81.

Frekvencija opisuje broj vrednosti u delu skale

TUMAČENJE TABELE

2

vrednosti su manje od 21
i nisu manje od 1

6

vrednosti su najmanje 61
i manje od 81

5

vrednosti su najmanje 81
i manje od 101

Frekvencija ne predstavlja veličinu broja, već broj posmatranja u intervalu.

Relativna frekvencija pokazuje udeo intervala

FORMULA RELATIVNE FREKVENCIJE

$$\text{relativna frekvencija} = \frac{\text{frekvencija intervala}}{\text{ukupan broj vrednosti}}$$

Interval [61, 81): 6 / 20 = 0,30 = 30%

Zbir relativnih frekvencija mora biti 1, odnosno 100%.

Proširena tabela povezuje broj i procenat

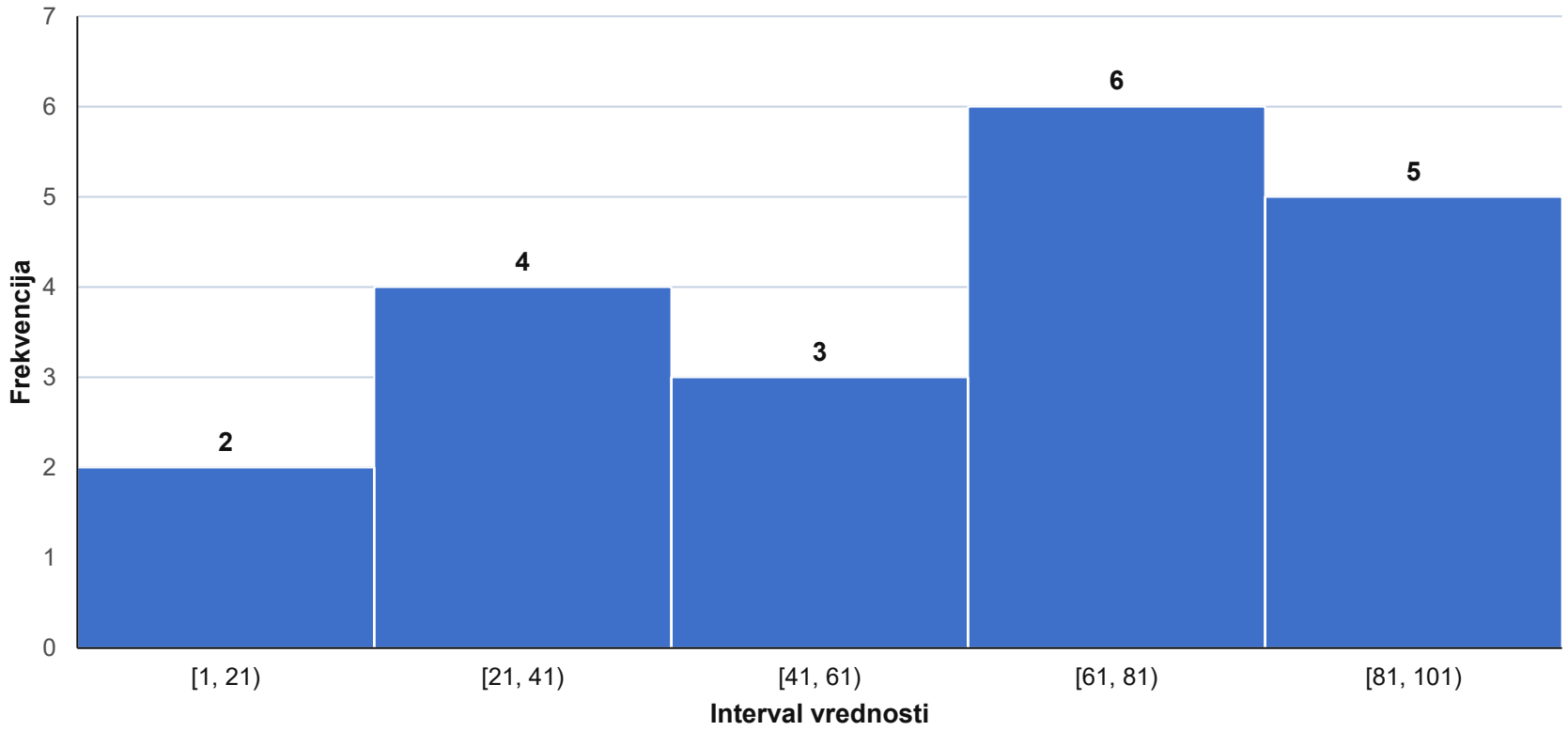
APSOLUTNA I RELATIVNA FREKVENCIJA

Interval	Frekvencija	Relativna frekvencija	Procenat
[1, 21)	2	0,10	10%
[21, 41)	4	0,20	20%
[41, 61)	3	0,15	15%
[61, 81)	6	0,30	30%
[81, 101)	5	0,25	25%
Ukupno	20	1,00	100%

Kontrole: zbir frekvencija je 20, a zbir procenata je 100%.

Histogram prikazuje raspodelu numeričkih podataka

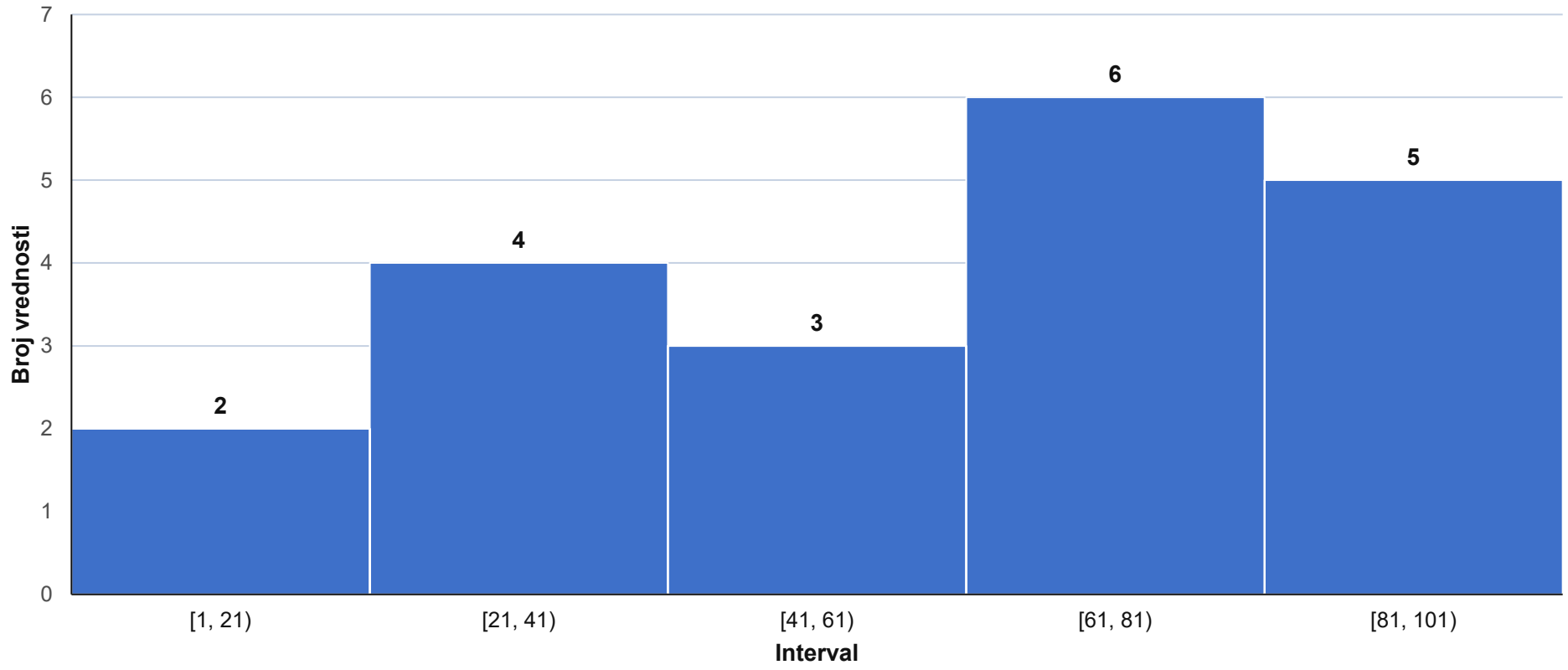
OSNOVNI GRAFIKON



Širina stuba predstavlja interval, a visina stuba njegovu frekvenciju.

Najviše vrednosti nalazi se između 61 i 81

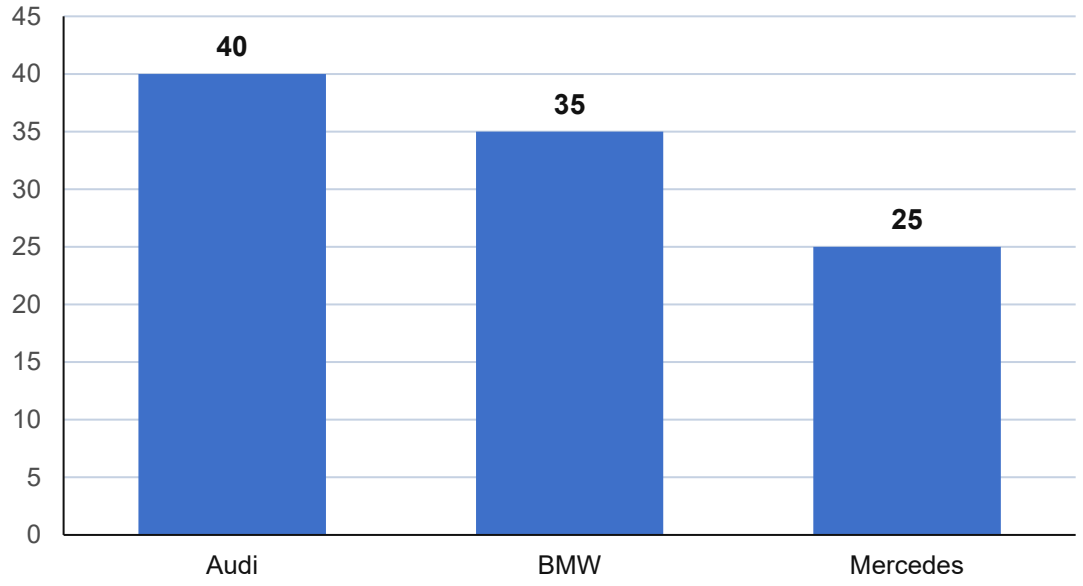
TUMAČENJE APSOLUTNOG HISTOGRAMA



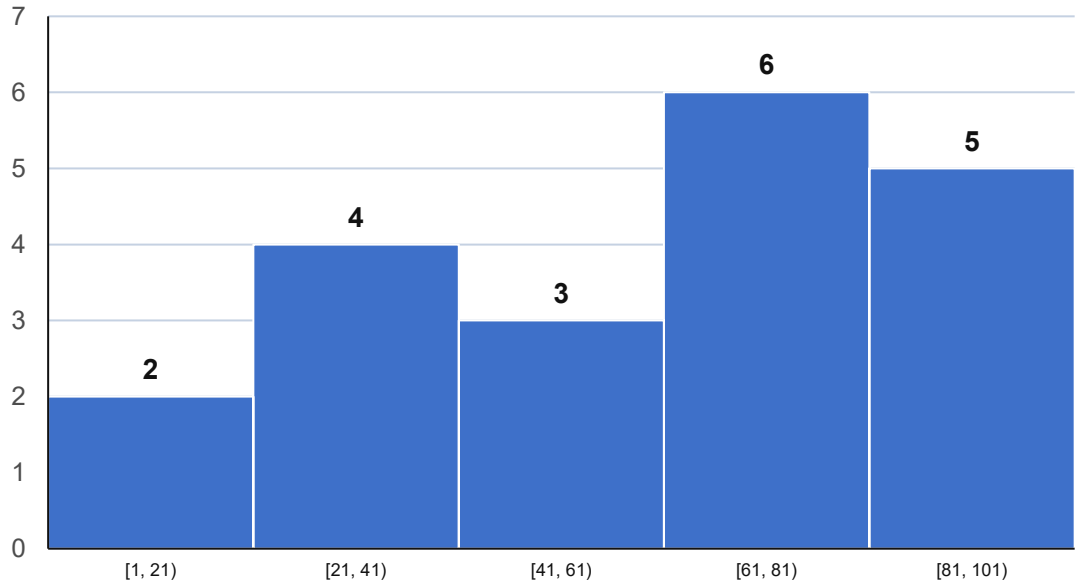
Histogram pokazuje gde se podaci nagomilavaju, ali ne prikazuje pojedinačne vrednosti.

Histogram i stubičasti dijagram prikazuju različite vrste podataka

POREĐENJE GRAFIKONA



STUBIČASTI DIJAGRAM
odvojene kategorije i razmak između stubova

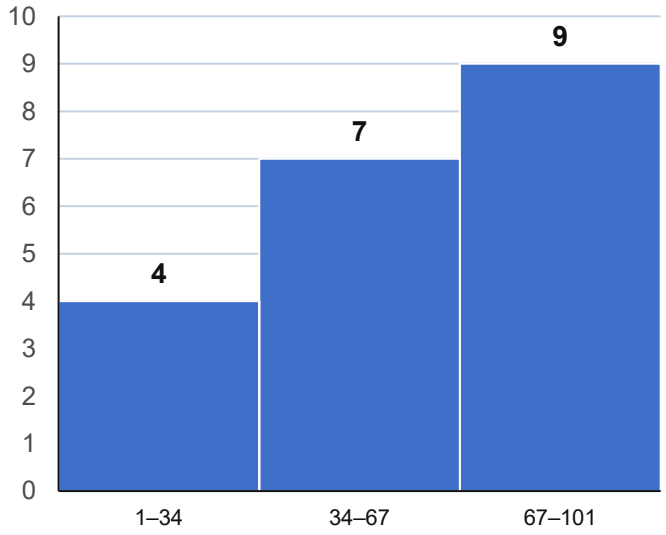


HISTOGRAM
numerički intervali i stubovi koji se dodiruju

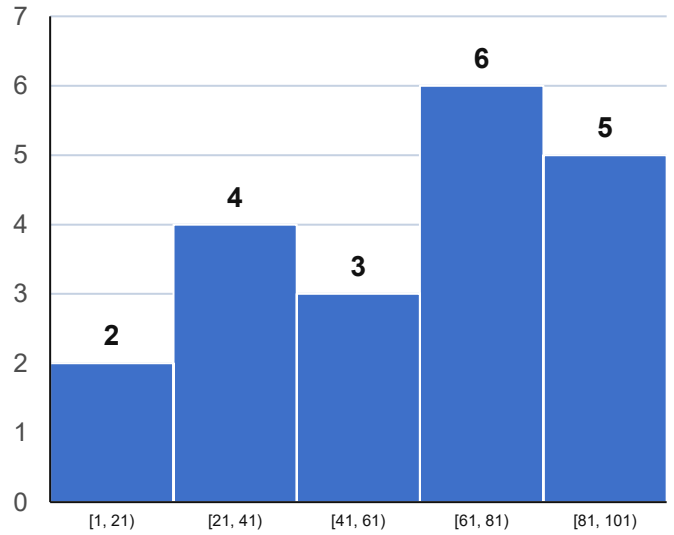
Sličan izgled ne znači isto značenje.

Različit broj intervala može promeniti utisak

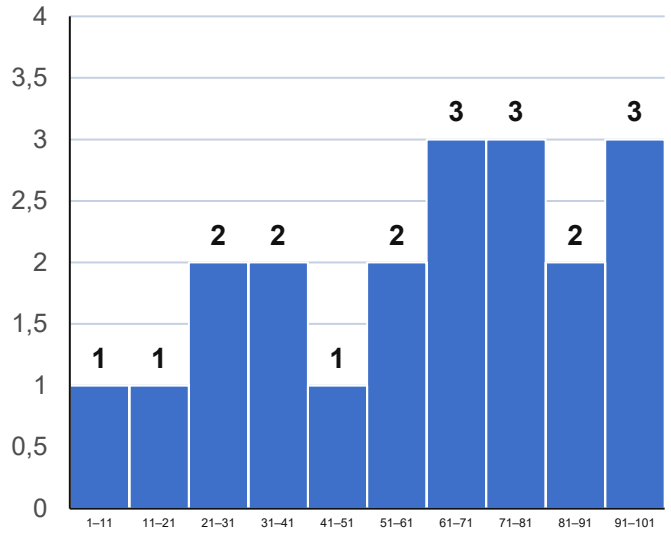
OSETLJIVOST PRIKAZA



3 intervala



5 intervala



10 intervala

Zaključak treba proveriti sa više razumnih podela intervala.

Nejednaki intervali zahtevaju pažljivo tumačenje

OGRANIČENJE HISTOGRAMA

Starosna grupa	Širina intervala	Broj ispitanika
18–25	8 godina	32
26–30	5 godina	25
31–40	10 godina	40
60+	otvoren interval	18

VAŽNO

Kada širine nisu jednake, sama visina frekvencije može da zavara.

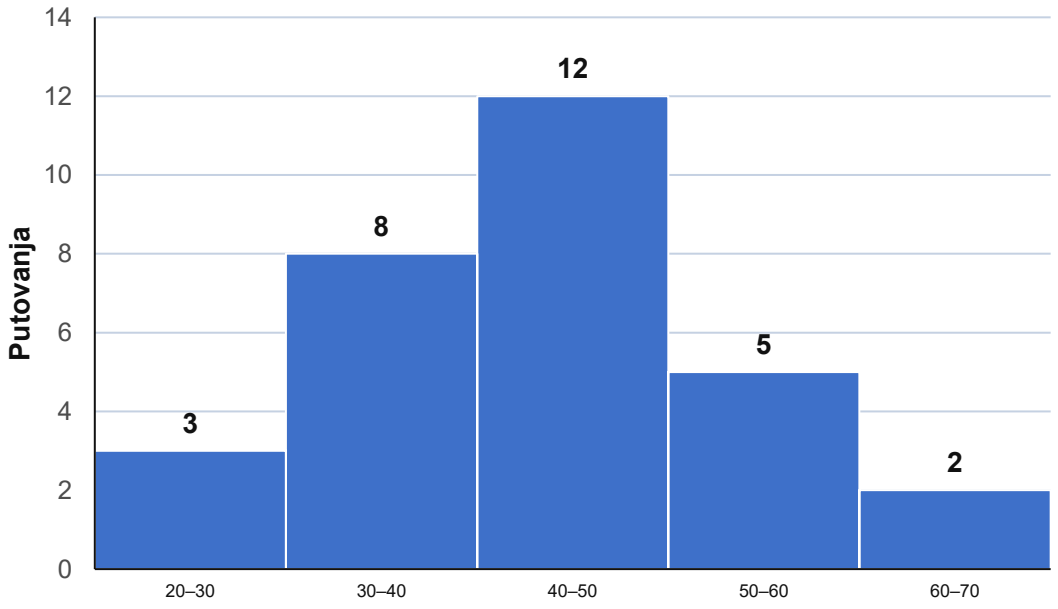
Površina stuba treba da predstavlja frekvenciju, pa se koristi gustina frekvencije.

Za početne analize biramo intervale jednake širine kad god je to smisleno.

Histogram u saobraćaju i radu sa decom

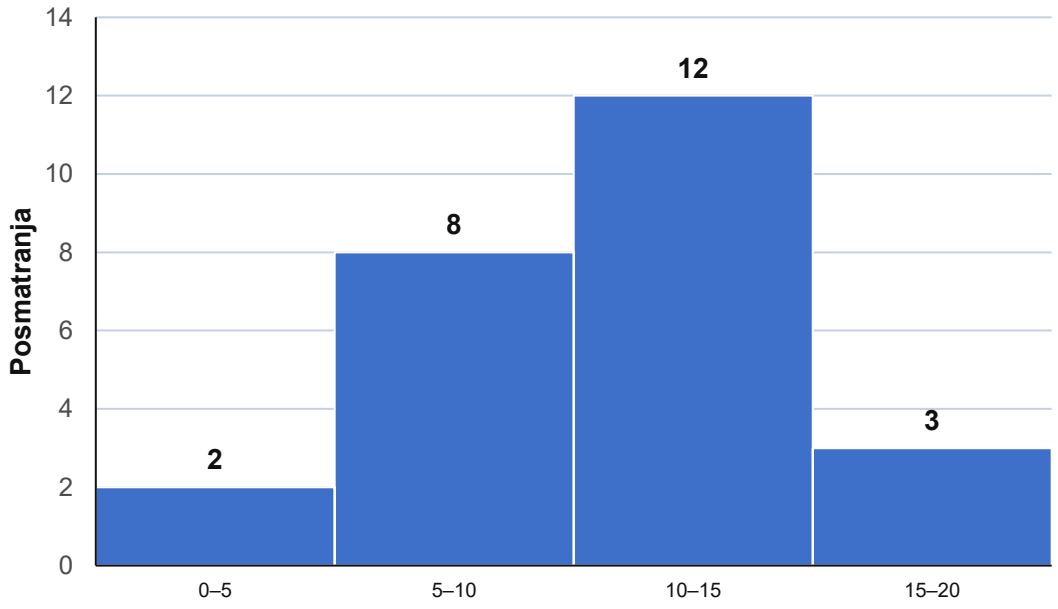
PRIMERI PRIMENE

VREME PUTOVANJA



Najviše putovanja traje 40 do manje od 50 minuta.

TRAJANJE ANGAŽOVANJA U AKTIVNOSTI



Najviše posmatranja nalazi se u intervalu 10 do manje od 15 minuta.

Intervali se biraju prema promenljivoj i pitanju analize.

Postupak od sirovih podataka do histograma

SAŽETAK I ČESTE GREŠKE

Korak	Šta proveravamo
1. Minimum i maksimum	granice skupa podataka
2. Broj intervala	dovoljno detaljan i čitljiv prikaz
3. Širina intervala	svi intervali pokrivaju skalu
4. Granice intervala	svaka vrednost pripada samo jednom intervalu
5. Frekvencije	zbir odgovara broju podataka
6. Histogram	ose, jedinice i dodirivanje stubova

ČESTE GREŠKE

preklapanje intervala

praznina između intervala

dvostruko brojanje granice

zbir frekvencija nije jednak broju podataka

mešanje histograma i stubičastog dijagrama

Dobar histogram počinje pravilno definisanom tabelom frekvencija.

Pitanje 1

PROVERA ZNANJA

Zašto pojedinačna tabela od 20 različitih vrednosti nije naročito korisna?

A Svaka vrednost ima frekvenciju jedan

B Brojevi ne mogu da se sortiraju

C Nema ukupnog broja podataka

D Numerički podaci ne mogu da se tabeliraju

Izaberite jedan odgovor i objasnite izbor.

Pitanje 2

PROVERA ZNANJA

Kolika je približna širina pet intervala ako su minimum 1, a maksimum 100?

A 5

B 19,8, pa biramo 20

C 50

D 99

Izaberite jedan odgovor i objasnite izbor.

Pitanje 3

PROVERA ZNANJA

Kojem intervalu pripada vrednost 41 ako koristimo [21, 41) i [41, 61)?

A Oba intervala

B Nijedan interval

C [21, 41)

D [41, 61)

Izaberite jedan odgovor i objasnite izbor.

Pitanje 4

PROVERA ZNANJA

Šest od dvadeset vrednosti nalazi se u jednom intervalu. Relativna frekvencija je:

A 6%

B 20%

C 30%

D 60%

Izaberite jedan odgovor i objasnite izbor.

Pitanje 5

PROVERA ZNANJA

Koja osobina najjasnije razlikuje histogram od stubičastog dijagrama?

A Histogram uvek ima kružnu legendu

B Histogram prikazuje numeričke intervale čiji se stubovi dodiruju

C Histogram može imati samo tri stuba

D Stubičasti dijagram ne koristi vertikalnu osu

Izaberite jedan odgovor i objasnite izbor.