MERE CENTRALNE TENDENCIJE

Aritmetička sredina, medijana i modus

Predmet:

Alati za obradu podataka

Tema:

Deskriptivna statistika

Ishodi učenja

ŠTA STUDENT TREBA DA RAZUME

Aritmetička sredina

Izračunavanje prosečne vrednosti i razumevanje uticaja ekstremnih podataka.

Medijana

Pronalaženje srednje pozicije u uređenom skupu sa neparnim ili parnim brojem podataka.

Modus

Prepoznavanje najčešće vrednosti kod numeričkih i kategorijskih podataka.

Izbor mere

Poređenje rezultata i obrazložen izbor mere koja najbolje opisuje tipičnu vrednost.

Šta znači „centar” podataka?

JEDAN SKUP, TRI RAZLIČITA PITANJA

Aritmetička sredina

Kolika je prosečna vrednost ako ukupan zbir ravnomerno raspodelimo?

Medijana

Koja vrednost ostaje u sredini nakon sortiranja podataka?

Modus

Koja vrednost se pojavljuje najčešće?

Sve tri mere opisuju skup jednim brojem, ali odgovaraju na različita pitanja.

Polazni primer: cene obroka

BEOGRAD, NOVI SAD I NIŠ

Grad	Sortirane cene obroka (din)	Broj restorana
Beograd	500, 700, 700, 700, 1.000, 1.200, 1.500, 1.800, 2.200, 2.200, 5.000	11
Novi Sad	500, 700, 900, 1.100, 1.200, 1.300, 1.400, 1.600, 1.800, 2.000	10
Niš	400, 600, 700, 800, 1.000, 1.100, 1.200, 1.400, 1.600, 1.700	10

Isto pitanje

Koja cena predstavlja „tipičan” obrok u svakom gradu?

Mogući problem

Cena od 5.000 dinara u Beogradu znatno odstupa od ostalih cena.

Aritmetička sredina

NAJČEŠĆE KORIŠĆENA MERA CENTRA

POPULACIJA

$$\mu = \Sigma x_i / N$$

UZORAK

$$\bar{x} = \Sigma x_i / n$$

1. Saberemo sve vrednosti.

2. Zbir podelimo brojem vrednosti.

3. Rezultat tumačimo u kontekstu.

Σ označava sabiranje svih vrednosti. N ili n označava broj opservacija.

Računanje sredine korak po korak

JEDNOSTAVAN SKUP BEZ EKSTREMNE VREDNOSTI

4	5	6	7	8

Zbir **4 + 5 + 6 + 7 + 8 = 30** → **30 ÷ 5 = 6**

Aritmetička sredina je 6. Kada bi ukupno 30 jedinica ravnomerno raspodelili na pet opservacija, svaka bi dobila 6.

Populacija i uzorak

ISTI RAČUN, RAZLIČITO ZNAČENJE

Populacija

Svi podaci o pojavi koju istražujemo. Primer: vremena putovanja svih 320 zaposlenih.

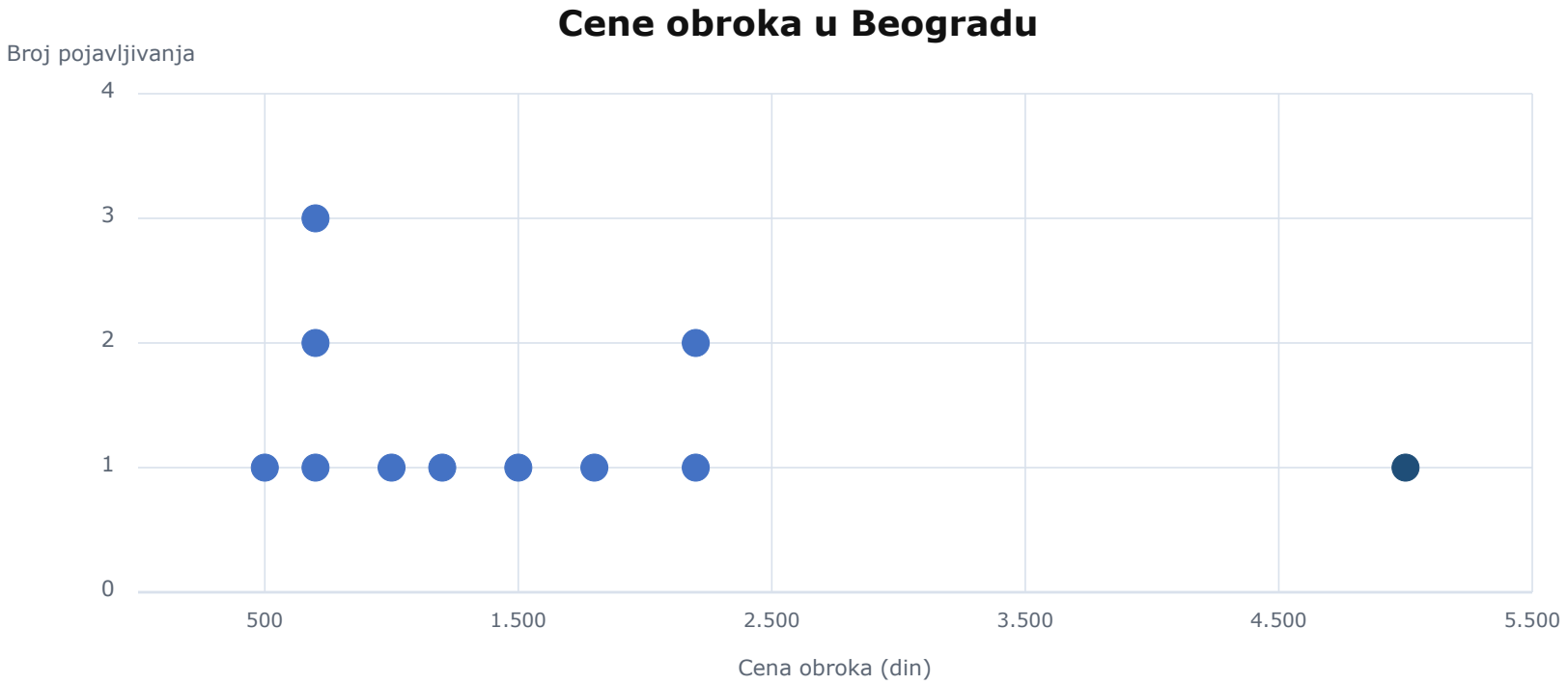
Uzorak

Deo populacije koji smo izabrali. Primer: vremena putovanja 40 zaposlenih.

μ opisuje poznatu populaciju. $\bar{x}$ procenjuje sredinu populacije na osnovu uzorka.

Glavno ograničenje aritmetičke sredine

IZDVOJENA VREDNOST MENJA REZULTAT



Izdvojena vrednost

Cena od 5.000 dinara nalazi se daleko od ostalih podataka.

Aritmetička sredina koristi veličinu svake vrednosti. Zato je osetljiva na ekstremne podatke.

Cena obroka: računanje sredine

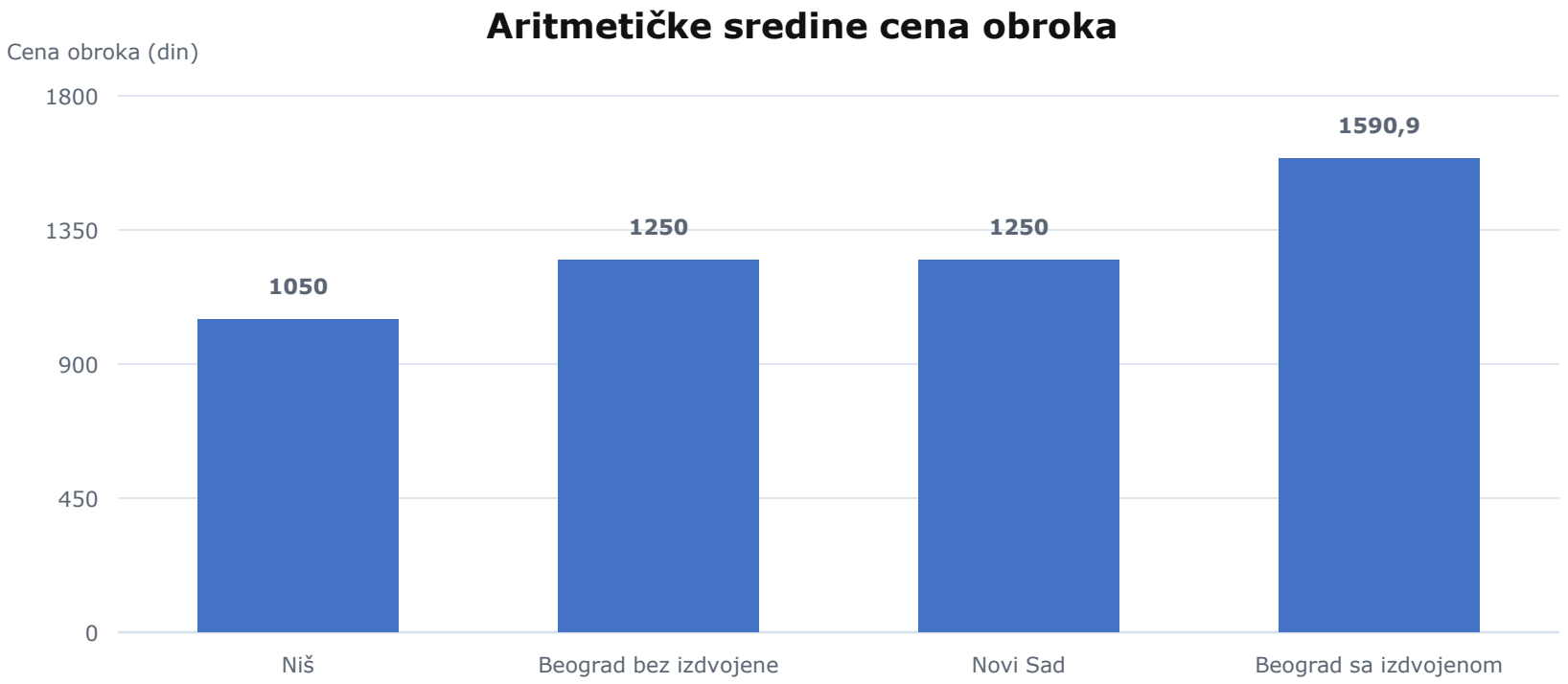
JEDNA VREDNOST MENJA SLIKU SKUPA

Skup	Zbir	Broj cena	Aritmetička sredina
Beograd sa cenom 5.000	17.500 din	11	1.590,91 din
Beograd bez cene 5.000	12.500 din	10	1.250 din
Novi Sad	12.500 din	10	1.250 din
Niš	10.500 din	10	1.050 din

Izdvojena cena podiže beogradsku sredinu; Niš u ovom primeru ima najnižu sredinu.

Uticaj izdvojene vrednosti

POREĐENJE ČETIRI SKUPA PODATAKA



Šta vidimo

Jedna izdvojena cena povećala je aritmetičku sredinu beogradskog skupa za oko 27%.

Sredina je korisna, ali mora se posmatrati zajedno sa raspodelom i drugim merama.

Medijana

SREDNJA VREDNOST UREĐENOG SKUPA

Podatke prvo sortiramo od najmanje ka najvećoj vrednosti.

NEPARAN BROJ PODATAKA

pozicija = (n + 1) / 2

PARAN BROJ PODATAKA

prosek dve srednje vrednosti

Medijana zavisi od položaja vrednosti, a ne od njihove udaljenosti od centra.

Medijana kod neparnog broja podataka

BEOGRAD IMA 11 CENA

500	700	700	700	1000	**1200**	1500	1800	2200	2200	5000
1	2	3	4	5	6	7	8	9	10	11

SREDNJA POZICIJA

$$(11 + 1) / 2 = 6$$

Šesta vrednost je 1.200, pa je medijana beogradskih cena 1.200 dinara.

Medijana kod parnog broja podataka

NOVI SAD IMA 10 CENA

500	700	900	1100	**1200**	**1300**	1400	1600	1800	2000

Srednje pozicije su 5. i 6.

MEDIJANA

(1.200 + 1.300) / 2 = 1.250

Kod parnog broja podataka medijana ne mora biti jedna od zabeleženih vrednosti.

Zašto je medijana stabilnija?

POLOŽAJ JE VAŽNIJI OD VELIČINE EKSTREMA

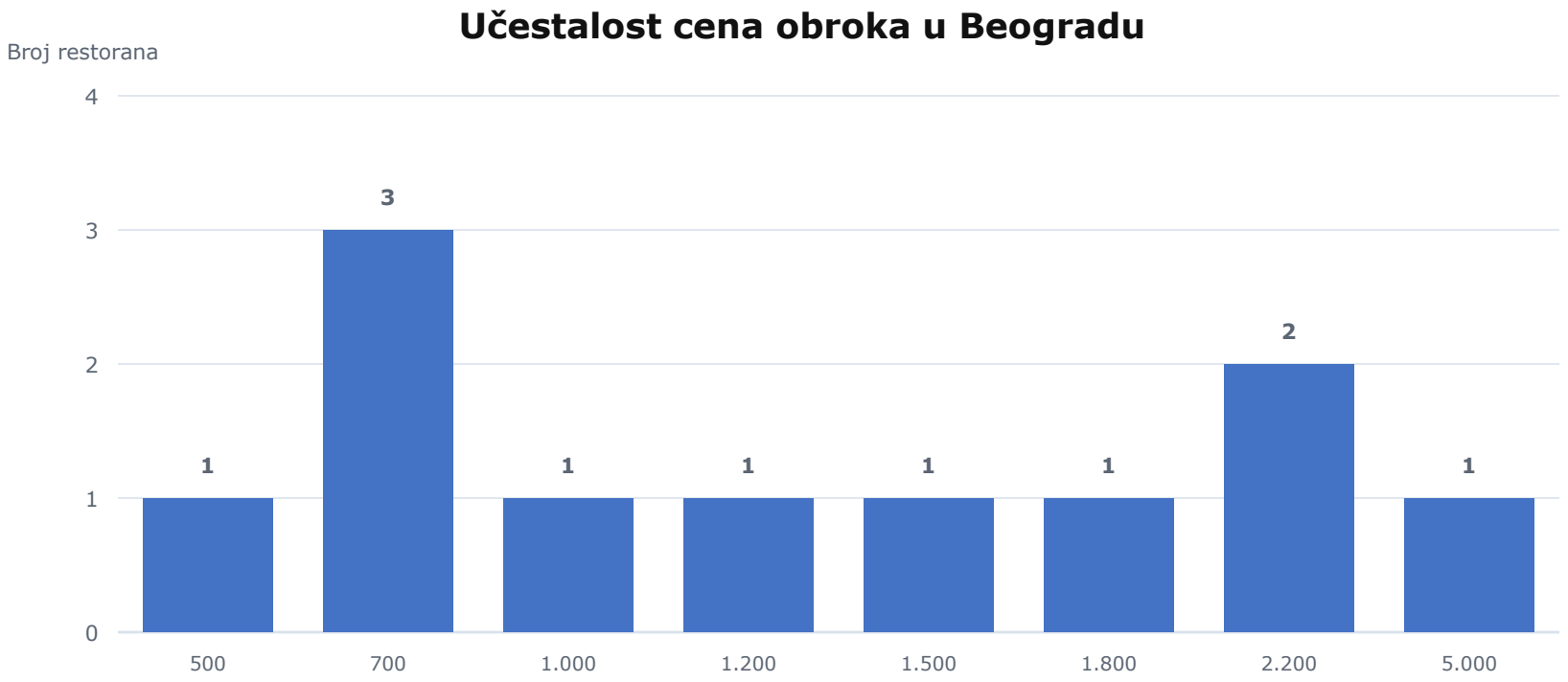
Skup podataka	Aritmetička sredina	Medijana
Beograd sa cenom 5.000	1.590,91 din	1.200 din
Beograd bez cene 5.000	1.250 din	1.100 din
Novi Sad	1.250 din	1.250 din
Niš	1.050 din	1.050 din

Tumačenje

Izdvojena cena snažno menja sredinu, dok medijana ostaje blizu većine podataka.

Modus

VREDNOST KOJA SE POJAVLJUJE NAJČEŠĆE



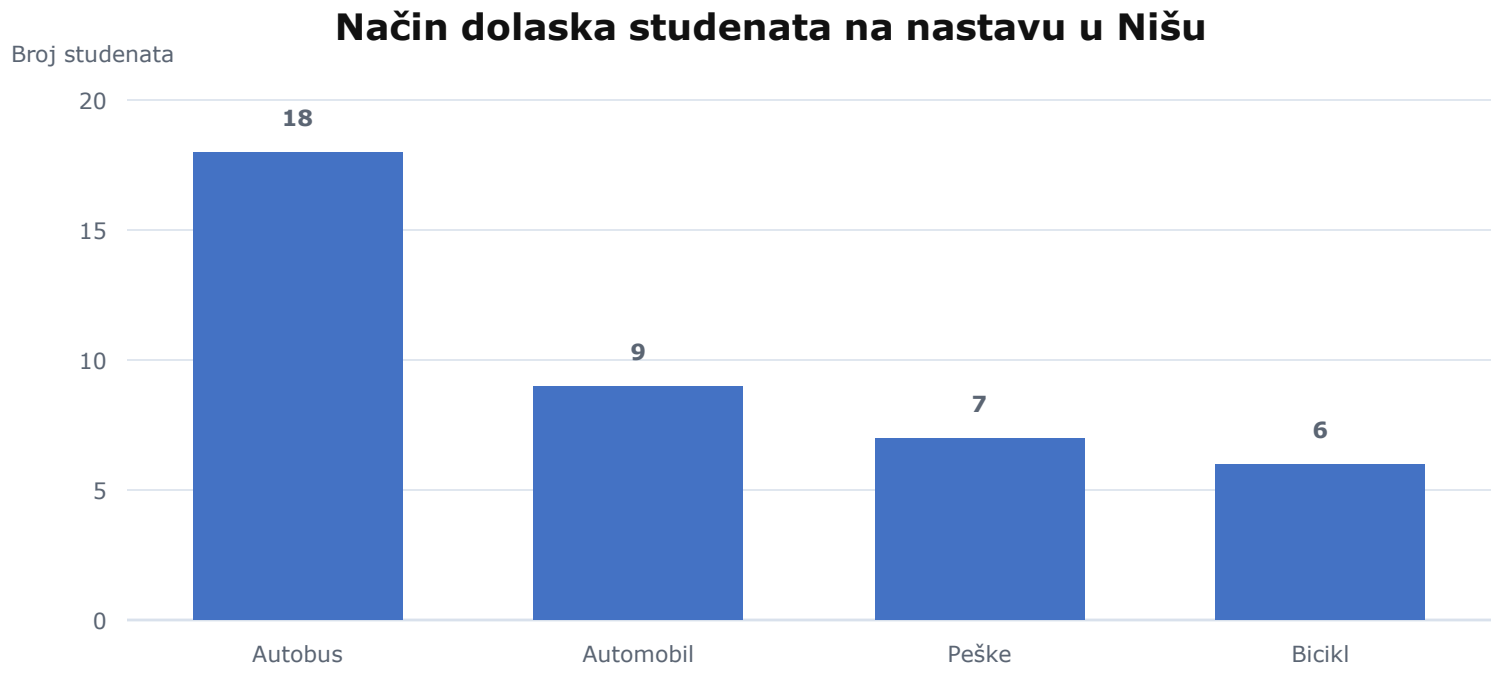
Rezultat

Cena od 700 dinara pojavljuje se tri puta. Modus je 700 dinara.

Modus opisuje najčešću vrednost, a ne sredinu po veličini.

Modus kod kategorijskih podataka

BROJEVI NISU OBAVEZNI



Modus = autobus

Autobus je najčešća kategorija, jer ga koristi 18 studenata.

Za kategorije ne računamo sredinu ili medijanu, ali možemo odrediti modus.

Kada modus ne postoji ili ih ima više

UČESTALOST ODREĐUJE ODGOVOR

Nema modusa

Podaci: 1, 2, 3, 4, 5

Svaka vrednost se pojavljuje jednom. Nijedna nije najčešća.

Dva modusa

Podaci: 2, 2, 3, 4, 4, 5

Vrednosti 2 i 4 pojavljuju se po dva puta. Skup je bimodalan.

Veliki broj „modusa” obično ne daje koristan sažetak podataka.

Poređenje tri mere

ŠTA KORISTI SVAKA MERA

Mera	Kako se određuje	Prednost	Ograničenje
Aritmetička sredina	Zbir ÷ broj podataka	Koristi sve numeričke vrednosti	Osetljiva na ekstreme
Medijana	Srednja pozicija	Stabilna kod izdvojenih vrednosti	Zahteva sortiranje
Modus	Najveća frekvencija	Radi i sa kategorijama	Može ne postojati ili ih biti više

Najbolji opis često dobijamo kada mere posmatramo zajedno.

Izbor odgovarajuće mere

VRSTA PODATAKA I OBLIK RASPODELE

Numerički podaci bez jakih ekstrema

Aritmetička sredina često daje dobar sažetak. Proverite i medijanu.

Asimetrični podaci ili ekstremi

Medijana obično bolje opisuje tipičnu vrednost. Ekstreme analizirajte posebno.

Kategorijski podaci

Koristite modus da utvrdite najčešću kategoriju.

Razlika između sredine i medijane često je prvi signal da raspodelu treba pažljivije pregledati.

Primer iz saobraćaja: kašnjenje linije

JEDAN KVAR MENJA PROSEČNO KAŠNJENJE

Zabeležena kašnjenja (min)	Sredina	Medijana	Modus
6, 7, 7, 8, 8, 9, 45	12,86	8	7 i 8

Tipično putovanje

Medijana od 8 minuta bolje opisuje većinu vožnji.

Operativni rizik

Kašnjenje od 45 minuta ostaje važno i treba ga posebno istražiti.

Česte greške

RAČUN MOŽE BITI TAČAN, A ZAKLJUČAK POGREŠAN

Nesortirani podaci

Medijana se traži tek nakon sortiranja vrednosti.

Pogrešan imenilac

Sredinu delimo brojem opservacija, a ne najvećom vrednošću.

Automatsko brisanje ekstrema

Prvo proveravamo da li je vrednost greška ili važan stvarni događaj.

Jedna mera bez konteksta

Rezultat porediti sa raspodelom, minimumom, maksimumom i drugim merama.

Desna asimetrija: plate i cene nekretnina

MALI BROJ VELIKIH VREDNOSTI PRODUŽAVA DESNI REP



ARITMETIČKA SREDINA

127,7 hiljada

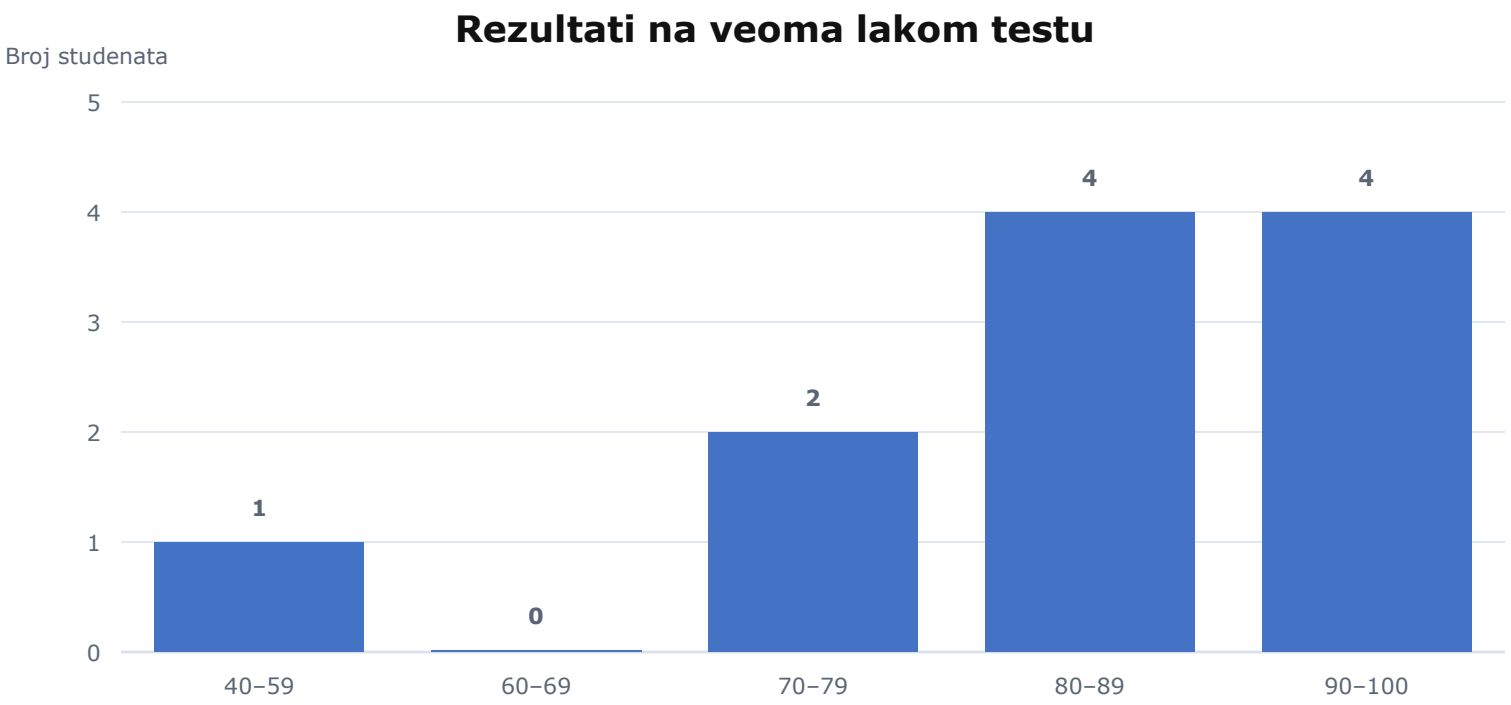
MEDIJANA

95 hiljada

Jedna plata od 450.000 dinara podiže sredinu. Medijana bolje opisuje platu zaposlenog u sredini raspodele.

Leva asimetrija: lak test i plafonski efekat

VEĆINA REZULTATA JE VISOKA, A REP IDE KA MANJIM VREDNOSTIMA



ARITMETIČKA SREDINA

83,7 poena

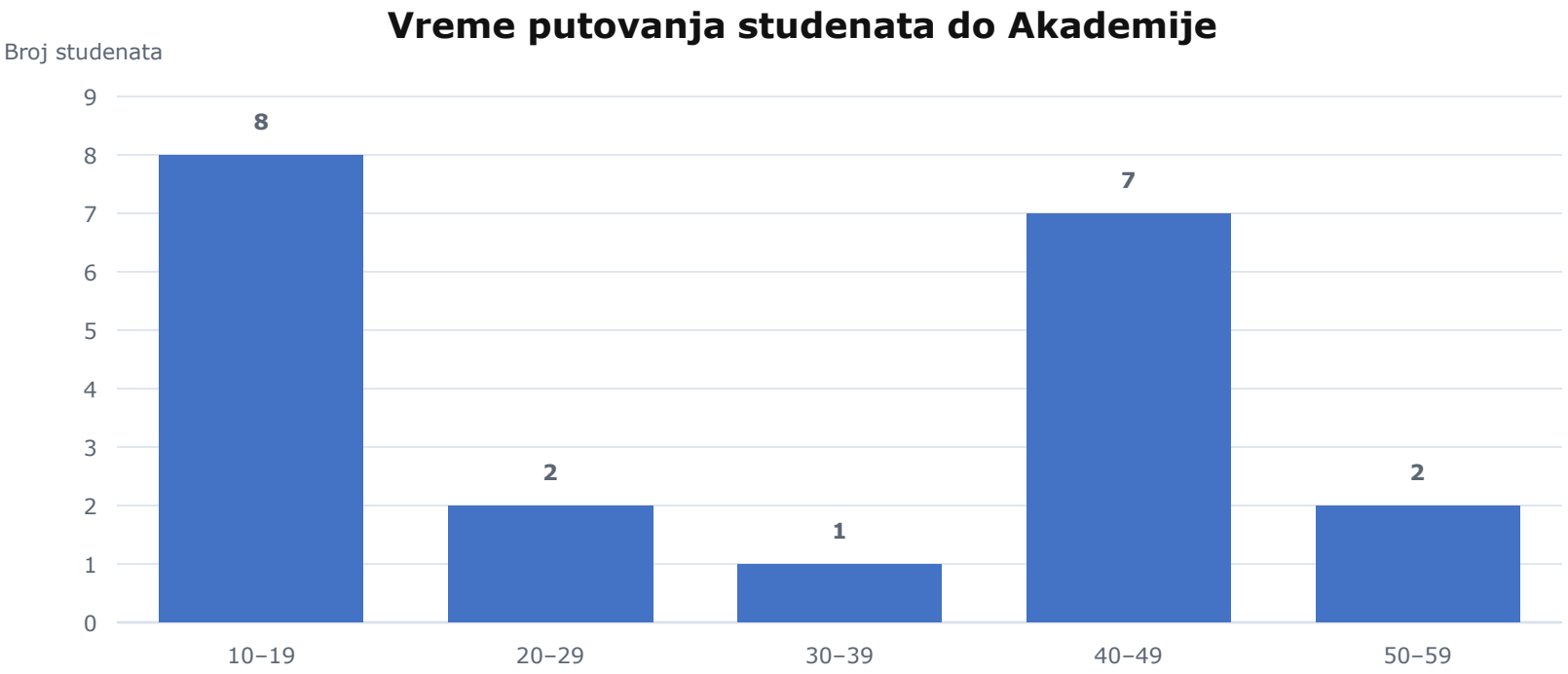
MEDIJANA

88 poena

Nizak rezultat od 48 poena povlači sredinu naniže. Visoki rezultati se gomilaju blizu gornje granice.

Bimodalna raspodela: dve različite grupe

JEDAN CENTAR MOŽE OPISATI MESTO NA KOME SKORO NEMA PODATAKA



Prva grupa

Studenti koji žive blizu Akademije putuju približno 10–20 minuta.

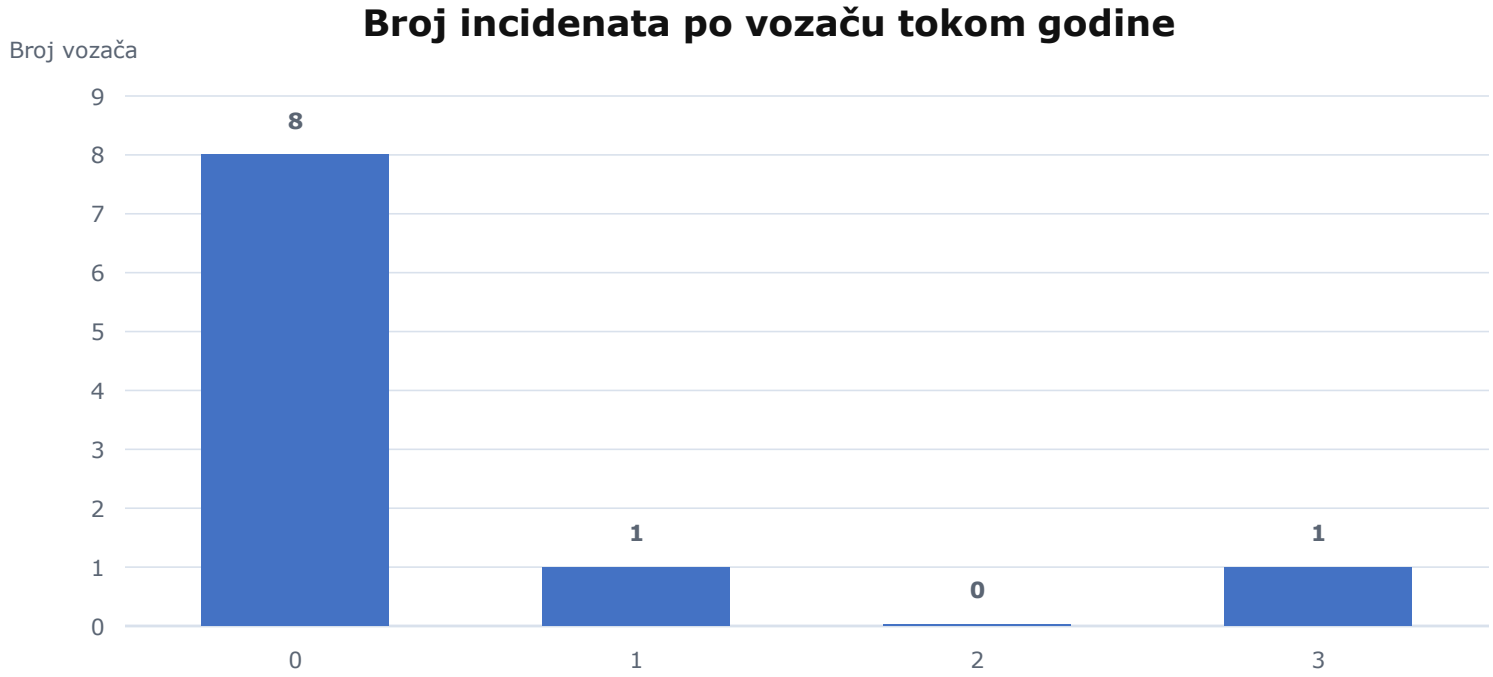
Druga grupa

Studenti iz udaljenijih naselja i okolnih mesta putuju približno 40–55 minuta.

Sredina je oko 31 minut, ali u tom području nalazi se malo studenata.

Mnogo nula: retki događaji

MEDIJANA I MODUS MOGU BITI NULA, A RIZIK IPAK POSTOJI



ARITMETIČKA SREDINA

0,4 incidenta

MEDIJANA I MODUS

0

Tipičan vozač nema incident, ali dva vozača nose celokupan zabeleženi rizik.

Prosek proseka može dati pogrešan rezultat

VELIČINA GRUPA MORA DA UĐE U RAČUN

Grupa	Broj studenata	Prosečna ocena	Ukupan zbir ocena
Grupa A	10	9,0	10 × 9 = 90
Grupa B	30	7,0	30 × 7 = 210
Zajedno	40	?	300

POGREŠNO: GRUPE NISU JEDNAKE

(9 + 7) ÷ 2 = 8,0

TAČNO: PONDERISANA SREDINA

(10×9 + 30×7) ÷ 40 = 7,5

Iste mere centra, različita raspodela

CENTAR NE OPISUJE RASPRŠENOST PODATAKA

Skup	Vrednosti	Sredina	Medijana	Modus	Raspon
A	4, 5, 5, 5, 6	5	5	5	2
B	1, 5, 5, 5, 9	5	5	5	8

Skup A

Vrednosti su tesno grupisane oko 5. Mala odstupanja znače veću ujednačenost.

Skup B

Iste mere centra kriju dve udaljene vrednosti, 1 i 9. Raspon je četiri puta veći.

Najvažnije za pamćenje

ČETIRI KLJUČNE IDEJE

1. Aritmetička sredina koristi sve numeričke vrednosti i zato reaguje na ekstreme.
2. Medijana je srednja pozicija uređenog skupa i stabilnija je kod asimetričnih podataka.
3. Modus je najčešća vrednost i jedini je primenljiv i na nominalne kategorije.
4. **Mere treba tumačiti zajedno sa raspodelom i kontekstom problema.**

Nema univerzalno najbolje mere. Postoji mera koja bolje odgovara konkretnim podacima i pitanju.

Pitanje 1

IZABERITE JEDAN TAČAN ODGOVOR

Koja mera se najviše menja kada skupu dodamo jednu veoma veliku vrednost?

A Aritmetička sredina

B Medijana

C Modus

D Sve tri mere podjednako

Obrazložite izbor pre prikazivanja tačnog odgovora.

Pitanje 2

IZABERITE JEDAN TAČAN ODGOVOR

Skup je 2, 4, 5, 9 i 50. Kolika je medijana?

A 4

B 5

C 9

D 14

Obrazložite izbor pre prikazivanja tačnog odgovora.

Pitanje 3

IZABERITE JEDAN TAČAN ODGOVOR

Koja mera najbolje opisuje tipičnu platu kada mali broj direktora ima veoma visoke plate?

A Medijana

B Maksimum

C Samo aritmetička sredina

D Zbir plata

Obrazložite izbor pre prikazivanja tačnog odgovora.

Pitanje 4

IZABERITE JEDAN TAČAN ODGOVOR

Histogram ima dva jasno odvojena vrha. Koji je najbolji sledeći korak?

A Izračunati samo jednu sredinu

B Istražiti da li postoje dve različite grupe

C Obrisati manji vrh

D Sve vrednosti zameniti medijanom

Obrazložite izbor pre prikazivanja tačnog odgovora.

Pitanje 5

IZABERITE JEDAN TAČAN ODGOVOR

Grupa A ima 10 studenata i prosek 9, a grupa B 30 studenata i prosek 7. Koliki je zajednički prosek?

A 8,0

B 7,5

C 7,0

D 6,5

Obrazložite izbor pre prikazivanja tačnog odgovora.