

PREDAVANJE 1 – I DEO

UVOD U STATISTIKU

Razvoj statistike

- **Stari i srednji vek:** popisi stanovništva, imovine, prebrojavanje rođenih, venčanih, umrlih.
- **XVII vek** početak naučnog zasnivanja statistike, u smislu sistematizacije demografskih i ekonomskih popisa u administrativne svrhe.
- **XVIII vek** u Nemačkoj se prvi put koristi termin *Statistika* (nauka o stanju države), dok se u isto vreme u Engleskoj koristi termin *Politička aritmetika*.
- **XIX vek** nagli razvoj matematičke statistike uslovljen razvojem matematičke analize i teorije verovatnoće.
- **Druga polovina XX veka** razvoj računara i pojava informacionog društva – statističke metode se globalno koriste u analiziranju masovnih pojava, u svim oblastima nauke i društva.
- **XXI vek** – statistika postaje deo nauke o podacima (data science)

Osnovni pojmovi

- **Podaci** obuhvataju informacije dobijene pomoću observacija, prebrojavanja, merenja ili odgovora.
- **Statistika** je grana matematike koja se bavi sakupljanjem, organizovanjem, analizom, interpretacijom i prezentovanjem podataka u cilju njihovog razumevanja i donošenja odluka.
- **Populacija** je skup svih podataka koji su od značaja za dato istraživanje.
- **Uzorak** je podskup populacije.

Populacija i uzorak

Primer:

U nedavnom istraživanju, 360 od 2450 studenata nekog fakulteta je anketirano o tome da li redovno upražnjavaju post, i 28 je potvrdno odgovorilo. Identifikovati populaciju i uzorak.

Odgovori svih studenata fakulteta
(populacija obima $N=2450$)

Odgovori anketiranih
studenata (uzorak obima
 $n=360$)

Parametri i statistike

- **Obeležje** je svojstvo tj karakteristika koja je predmet istraživanja.
- **Parametar** je numerički opis nekog obeležja *populacije*.
- **Statistika** (u užem smislu reči) je numerički opis nekog obeležja *uzorka*.

Parametar —————→ **Populacija**

Statistika —————→ **Uzorak**

Parametri i statistike

Primer

Odlučite da li sledeći numerički izrazi opisuju parametre populacije ili statistike uzorka:

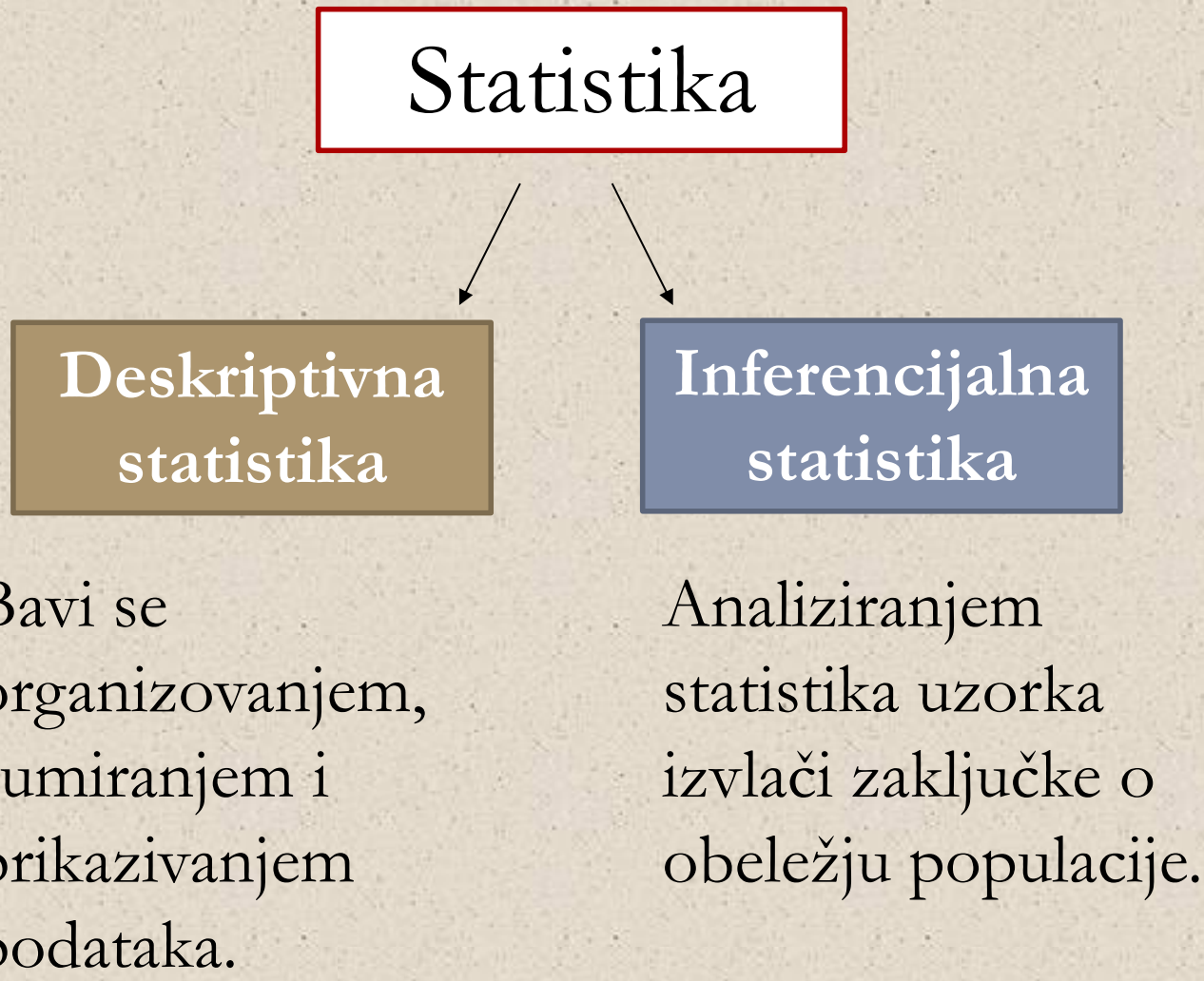
- a.) Ispitivanjem 100 studenata nekog fakulteta utvrđeno je da u proseku mesečno troše 23 450 dinara.

Statistika uzorka

- b.) Prosečan mesečni trošak svih studenata nekog fakulteta je 19895 dinara.

Parametar populacije

Podela statistike



Deskriptivna i inferencijalna statistika

Primer

U sprovedenom istraživanju, ispitanici koji su spavali manje od 6h su tri puta češće grešili na testu pažnje od onih koji su spavali bar 8h. Koji deo je deskriptivna statistika, i kakav bi se zaključak mogao doneti o populaciji?

Tvrđenje “tri puta češće grešili” je deskriptivna statistika. Eventualni zaključak bi mogao biti da sve individue koje spavaju manje od 6h greše češće od onih koji spavaju bar 8h, **ali za njega je potrebno uraditi statistički test, zaključak se ne može doneti samo na osnovu deskripcije.**

Oblasti statistike

- Ocenjivanje parametara populacije
- Testiranje statističkih hipoteza
- Korelaciona i regresiona analiza
- Analiza varijanse
- Analiza vremenskih serija; Prostorna statistika
- Moć testa i veličina uzorka
- Klaster analiza; faktorska analiza
- Kontrolne karte i analiza kvaliteta
- Dizajn eksperimenta
- ...

Statistika vs Teorija verovatnoće

Statistika:

- primenjena oblast matematike,
- bavi se analizom masovnih pojava iz realnog sveta
- događaji iz prošlosti
- Koristi alate teorije verovatnoće (raspodele, momente, itd...)

Teorija verovatnoće :

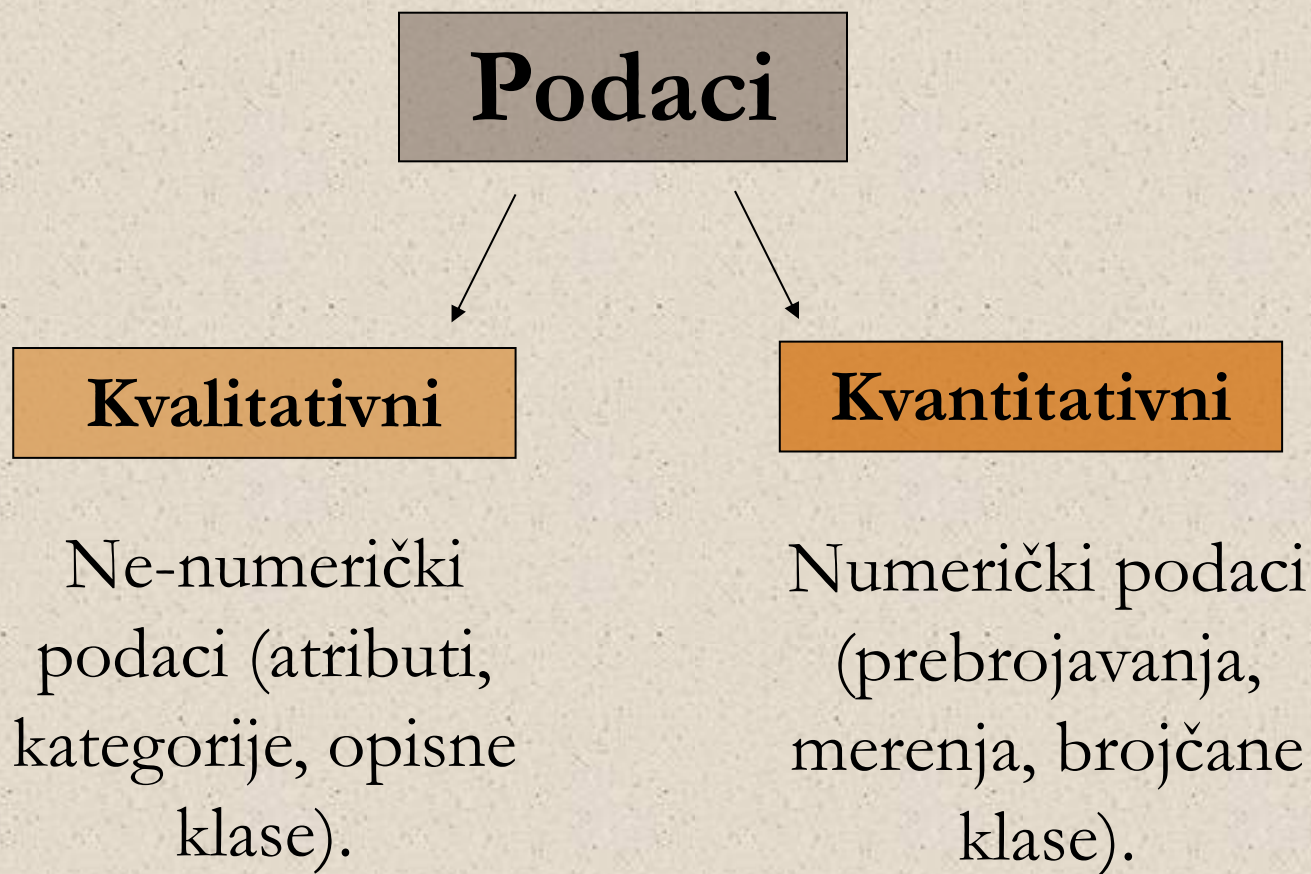
- teorijska oblast matematike,
- bavi se kvantitativnom analizom slučajnih događaja
- događaji u budućnosti
- Koristi informacije dobijene statističkom obradom (statistička definicija verovatnoće)



PREDAVANJE 1 – II DEO

KLASIFIKACIJA PODATAKA

Tipovi podataka



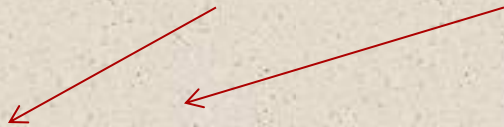
Kvalitativni i kvantitativni podaci

Primer

Koje tipove podataka sadrži tabela?

Student	pol	Godina studija	Prosečna ocena
Ana	Ž	IV	8.22
Boban	M	I	7.98
Vera	Ž	II	7.75
Gavrilo	M	I	9.24
David	M	III	8.84

kvalitativni

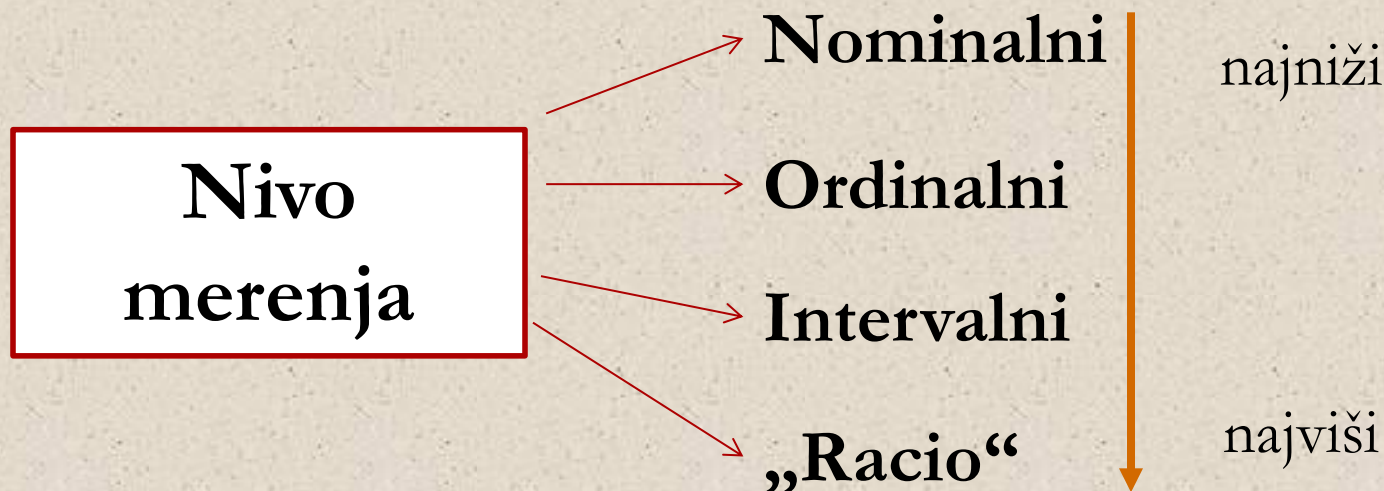


kvantitativni



Nivoi merenja

Nivo merenja podataka određuje koje statističke metode se mogu primeniti na tim podacima.



Nominalni nivo merenja

- Podaci koji pripadaju **nominalnom nivou merenja** su isključivo kvalitativni.
- U ove podatke spadaju imena, kategorije, kvaliteti.

Primeri nominalnih podataka:



Ordinalni nivo merenja

- Podaci koji pripadaju **ordinalnom nivou merenja** mogu biti kvalitativni ili kvantitativni.
- **Uređeni** su, u smislu postojanja **poretka**, ali se ne može smisleno izračunati razlika između kategorija.

Primeri ordinalnih podataka:

Stručna sprema

Godina
studiranja

Koliko nešto
volite (na skali
od 1 do 10)

Veličina preduzeća
(M, S, V)

Intervalni nivo merenja

- Podaci u **intervalnom nivou merenja** su kvantitativni i dobijeni **merenjem**.
- Nula vrednost predstavlja jednu od ravnopravnih pozicija na skali merenja, a ne apsolutnu nulu.
- Razlika je smisljena, ali količnik nije.

Primeri intervalnih podataka:

Temperatura

Godišnja dobit
nekog preduzeća

Godine na
vremenskoj skali

„Racio“ nivo merenja

- Podaci iz **racio nivoa merenja** su slični intervalnim podacima, izuzev što na skali 0 predstavlja apsolutnu nulu.
- Količnik vrednosti je smislen, kao i razlika.

Primeri racio podataka:

starost

(modifikovana)
prosečna ocena

Težina

Sumirani nivoi merenja

Nivo merenja	Stavlja podatke u kategorije	Uređuje podatke u poredak	Smisleno je oduzimanje vrednosti	Smisleno je deljenje vrednosti
Nominalni	DA	NE	NE	NE
Ordinalni	DA	DA	NE	NE
Intervalni	DA	DA	DA	NE
Racio	DA	DA	DA	DA

- Često se intervalni i racio podaci objedinjavaju pod terminom **neprekidni**, a ordinalni se zovu i **diskretni**.

Dimenzija podataka

- Podaci mogu biti **jednodimenzionalni** i **višedimenzionalni**
- Višedimenzionalni podaci su uređene n -torke, kod kojih svaka komponenta predstavlja jednodimenzionalni podatak.
- Njihovom analizom se bavi posebna grana statistike.

Primer

(Starost, BMI, obim struka)

PREDAVANJE 1 – III DEO

TEHNIKE UZORKOVANJA

Kako se vrši statističko istraživanje?

UPUTSTVO

1. Identifikuj jedno ili više obeležja od interesa (fokus) i populaciju. Formuliši ciljeve i hipoteze.
2. Razvij detaljan plan sakupljanja podataka. Ukoliko koristiš uzorkovanje, osiguraj reprezentativnost uzorka.
3. Sakupi podatke; napravi matricu podataka.
4. Opiši podatke. Prikaži podatke.
5. Interpretiraj podatke i donesi odluke o populaciji koristeći metode inferencijalne statistike, u skladu sa postavljenim ciljevima i hipotezama.
6. Identifikuj moguće greške i slabosti istraživanja.

Metode sakupljanja podataka

U **opservacionom (opažajnom)** istraživanju, istraživač opaža i meri karakteristiku od interesa za deo populacije.

U **eksperimentu**, neki tretman se primenjuje na delu populacije i beleži se reakcija (odgovor na tretman).

Simulacija je upotreba matematičkog ili fizičkog modela u cilju reprodukovanja uslova neke situacije ili procesa.

- **Cenzus** je merenje *cele* populacije.
- **Uzorkovanje** je merenje *dela* populacije.

Reprezentativnost uzorka

Uzorak je **reprezentativan** ukoliko dobro oslikava osobine populacije.

Ukoliko uzorak nije reprezentativan, inferencijalnom statistikom se dobijaju nepouzdana zaključci o populaciji.

Na reprezentativnost uzorka utiče:

1. **Tehnika uzorkovanja** (stratifikovani, klaster, sistematski, dostupni, itd...)
2. **veličina tj. obim uzorka.**

Obim zavisi od:

- Obima populacije
- Ciljeva istraživanja \ primenjenih statističkih metoda
- Cene \ vremena istraživanja

Stratifikovan uzorak

Stratifikovan uzorak sadrži elemente iz svakog segmenta populacije. Time se obezbeđuje da je populacija ravnomerno reprezentovana u uzorku.

Primer: istraživanje navika studenata.



I godina



II godina



III godina

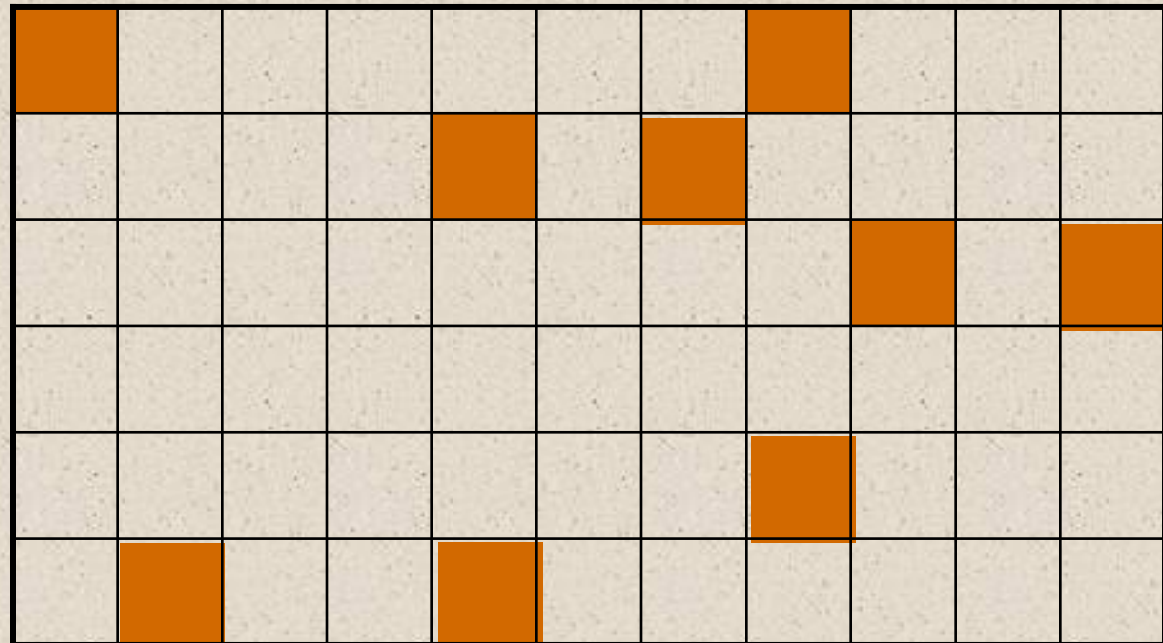


IV godina

Klaster uzorak

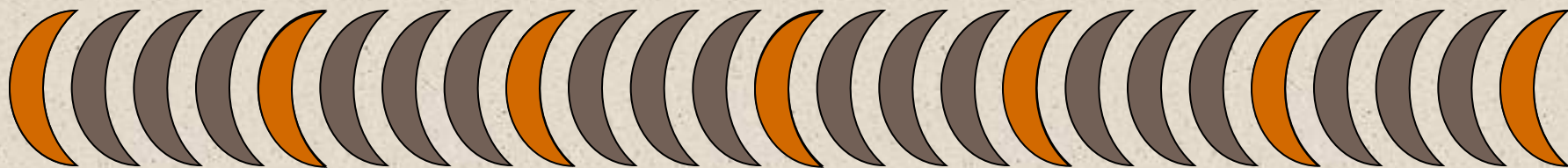
Klaster uzorak sadrži sve elemente slučajno odabranih segmenata cele populacije. Ovo se koristi kada je populacija prirodno podeljena na segmente (npr. geografski – pojedina sela, leksikografski – neke stranice imenika).

Svi elementi iz
odabranih
blokova se
uključuju.



Sistematski uzorak

Sistematski uzorak podrazumeva da se svakom elementu populacije dodeli broj, zatim se na slučajan način bira prvi element uzorka, i počevši od njega se u uzorak uključuje svaki n -ti element.



Ovde je svaki četvrti element uzet.

Dostupan uzorak

Dostupan uzorak se sastoji od onih elemenata koji su istraživaču na raspolaganju. Najmanje je reprezentativan.

Primer:

Cilj istraživanja je utvrditi godine radnog iskustva nastavnika sa FTN-a. Koje tehnike uzorkovanja su korišćene?

- 1.) Nasumično su odabrana 3 departmana i ispitani svi nastavnici koji im pripadaju. **klaster**
- 2.) Ispitani su nastavnici koji vam trenutno predaju. **dostupan**
- 3.) Ispitano je po nekoliko nastavnika sa svakog departmana. **stratifikovani**
- 4.) Iz računovodstva je uzet spisak svih nastavnika i anketiran je svaki deseti. **sistematski**