

Skripta iz Uvoda u teoriju uzoraka

18. februar 2020

1 nedelja

1.1 Naučno istraživanje

Naučno istraživanje je sistematsko, plansko i objektivno ispitivanje nekog problema, prema određenim metodološkim pravilima, čija je svrha da se pruži pouzdan i precizan odgovor na unapred postavljeno pitanje.

Može se shvatiti kao kritički, kontrolisani i ponovljivi proces sticanja novih znanja, neophodnih (a ponekad i dovoljnih) za identifikovanje, određivanje i rešavanje naučnih (teorijskih i empirijskih) problema.

Teorijsko istraživanje vs **Empirijsko (iskustveno)** istraživanje.

Svako naučno istraživanje ima više međusobno logično povezanih faza.

Faze su:

- identifikovanje i određivanje problema
- određivanje cilja istraživanja
- definisanje ključnih izraza
- postavljanje hipoteze i izvođenje logičkih posledica iz hipoteze
- izbor istraživačke strategije i plana istraživanja
- razvijanje mernih i drugih instrumenata istraživanja
- određivanje onovnog skupa (populacije) i odabir uzorka
- sprovođenje istraživanja i prikupljanje relevantnih podataka
- obrađivanje i analiza podataka dobijenih istraživanjem
- tumačenje rezultata istraživanja i izvođenje zaključ(a)ka
- izrada izveštaja o obavljenom istraživanju
- prezentacija rezultata istraživanja

1.2 Osnovni pojmovi

Entitet/jedinica posmatranja (en. 'observation unit') - živo biće ili objekt čija su svojstva predmet istraživanja.

Populacija ('population') - skup / kolekcija entiteta.

Na osnovu broja entiteta, tj. **obima** / veličine **populacije** ('populationsize') N , može biti:

- konačna populacija – N je prirodan broj
- beskonačna populacija – $N \rightarrow +\infty$

Trebalo bi razlikovati:

- ciljnu populaciju ('target population')
- populaciju na kojoj se efektivno sprovodi istraživanje ('study population')

Uzorak ('sample') - podskup populacije; sadrži izvesne entitete koji potiču iz populacije, na bazi čijeg proučavanja će se izvoditi zaključci o čitavoj populaciji

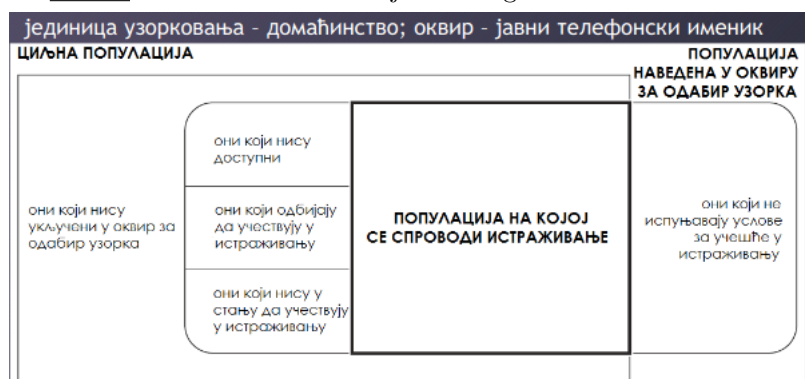
Obim uzorka ('samplesize') n ²

Jedinica uzorkovanja ('sampling unit') ³

Okvir za odabir uzorka ('sampling frame') - popis (ili neka druga specifikacija) svih jedinica uzorkovanja

Npr. svakoj jedinici uzorkovanja pridruži se različit prirodan broj (počevši od 1). Ti brojevi nazivaju se **oznake jedinica**, služe za njihovo identifikovanje i ostaju nepromenjeni sve do kraja istraživanja.

Primer - telefonsko istraživanje biračkog tela.



Zašto uzorkovanje?

Potpuno ispitivanje populacije (proučavanje tzv. **cenzusa**) je, u mnogim slučajevima, neracionalno ili čak principijelno nemoguće. Čak i onda kada postoji

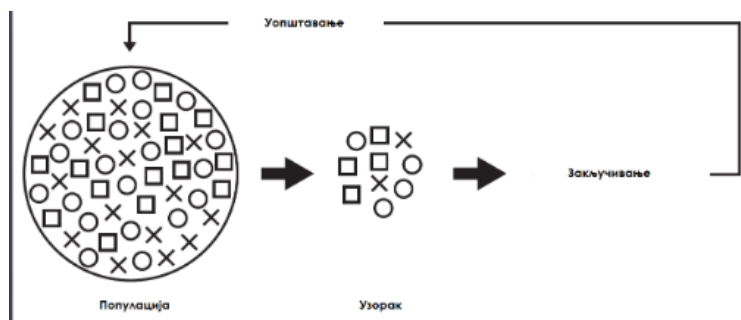
¹Nadalje se pretpostavlja: target population = study population i $N < +\infty$

²Uvek konačna vrednost

³U opštem slučaju nije isto što i jedinica posmatranja, koja predstavlja osnovni objekat posmatranja i prikupljanja informacija. Jedinice uzorkovanja su međusobno disjunktne skupovi entiteta

могућност потпуног испитивања популације истраживач се обично определjuje за **делимично** испитивање (пroučавање узорка) jer je (у односу на потпуно испитивање):

- jeftinije
- brže
- kontrola tačnosti prikupljenih podataka je jednostavnija i lakša



Termin populacija odnosi se na skup entiteta istovrsnih u odnosu na jedno ili više zajedničkih svojstava, koja se mogu posmatrati. Ipak, entiteti, iako **istovrsni, nisu istovetni**.

Одређивање популације представља значајну и, neretko, tešku fazu истраживања. Популација мора бити дефинисана: појмовно (у смислу свог садржаја - шта су ентитети, а шта јединице узorkовања?), просторно и временски.

Оbeležje ('study variable') - посматрана заједничка карактеристика свих ентитета у популацији, тј. прецизније, извесно варијабилно својство од интереса, које је одређено за сваки ентитет у популацији.⁴

1.3 Tipovi obeležja



⁴Obeležje najčešće nije neko od definicionih svojstava populacije.

Primer: tipovi obeležja

- kvalitativna
 - nominalna
 - * boja očiju, krvna grupa
 - * etnička / verska pripadnost
 - * radna mesta na fakultetu
 - * raspoloženje građana Srbije prema pristupanju u EU
 - * posedovanje profila na društvenim mrežama
 - ordinalna
 - * nivo akademskih studija
 - * čin oficira u vojsci
 - * ocena restorana na Tripadvisor
 - * stanje pacijenta
 - * intezitet bola
- kvantitativna
 - diskretna
 - * broj stanovnik sa pravom glasa u određenoj opštini
 - * broj blizanaca rođenih u toku godine u određenoj regiji
 - * broj kućnih ljubimaca u domaćinstvu
 - neprekidna
 - * visina, težina, starost, IQ
 - * dužina lista određene biljne vrste
 - * koncentracija soli u morskoj vodi

Primer: populacija i obeležje

- **Populacija:** skup studenata koji su upisali Uvod u teoriju uzoraka školske 2019/20. godine.
Obeležje: pol; broj položenih ispita, broj položenih ESPB bodova, prosečna ocena svih položenih ispita –zaključno sa rokom Januar 2 ove školske godine; ocena na kursu Statistika
- **Populacija:** skup svih poljoprivrednih gazdinstava u Srbiji(referentni period–oktobar/novembar2018)
Obeležje:površina korišćenog poljoprivrednog zemljišta; broj grla stoke; primenjeni proizvodni metodi
- **Populacija:** skup svih domaćinstava u regionu Šumadije i Istočne Srbije(referentni period –2017. g)
Obeležje: lična potrošnja domaćinstva (mesečni prosek)
- **Populacija:** jedna serija LED sijalica izvesnog proizvođača.
Obeležje: dužina radnog veka sijalice u satima.
- **Populacija:** skup svih meseci u periodu od 2000. do 2016. g.
Obeležje: mesečni broj vetrovitih dana u Vršcu

Obeležje se može shvatiti kao funkcija koja entitetima u populaciji pridružuje realne brojeve ili neke druge vrednosti.

Neka je data populacija sa N jedinica, koje su u okviru za odabir uzorka označene brojevima iz skupa $\omega = 1, 2, \dots, N$ (i time jednoznačno određene) i neka je Y obeležje od interesa. Neka je sa y_k označena vrednost obeležja Y entiteta označenog sa k .

Zadatak pri istraživanju obično se svodi na donošenje zaključaka o (nepoznatoj) vrednosti realne funkcije

$$\theta = f(y_1, y_2, \dots, y_n)$$

, koja se naziva **populacijska vrednost** ('population value') ili **parametar populacije**.

Najčešće funkcije koje se pojavljuju kao parametri populacije:

- Kvantitativna obeležja

- populacijska srednja vrednost ('population mean')

$$m_Y = m = \frac{1}{N} \sum_{k=1}^N y_K$$

- populacijski total ('population total')

$$\tau_Y = \tau = \sum_{k=1}^N y_K = N m_Y$$

- populacijska disperzija ('population variance') / standardno odstupanje

$$\sigma_Y^2 = \sigma^2 = \frac{1}{N-1} \sum_{k=1}^N (y_K - m_Y)^2$$

i

$$\sigma_Y = \sigma = \sqrt{\sigma_Y^2}$$

- Kvalitativna obeležja

- populacijska proporcija ('population proportion')

- populacijska medijana, kvantili, moda...

Ideja je da se zaključci o populacijskim vrednostima donose na osnovu informacija dobijenih ispitivanjem uzorka.

„Dobar“ uzorak ima osobinu **reprezentativnosti**. To je uzorak koji predstavlja „umanjenu“, a nikako „iskrivljenu“, niti „uvećanu“ sliku jednog dela populacije. Uzorak sa ovom osobinom verno odslikava strukturu populacije koju predstavlja, „izgledajući“ kao i populacija u svim aspektima relevantnim za istraživanje.

Na reprezentativnost uzorka utiču:

- tip uzorka (prema metodu odabira)
- veličina uzorka
- varijabilnost posmatranog obeležja

