

Slučajni Uzorak: Teorijski Pristup i Primena

Kristina Zogovic



Naslovni Slajd

- Naslov: Stratifikovani
Slučajni Uzorak u Analizi
Podataka
Podnaslov: Teorijski
pristup, pregled
literature, razlozi za
upotrebu, prednosti, mane
i case study
Autor: [Vaše Ime ili
Dopunski rad]
Datum: 2025-09-01



Sadržaj

- Teorijski pristup stratifikovanom uzorku
- Pregled literature
- Zašto koristiti stratifikovani uzorak?
- Kada koristiti stratifikovani uzorak?
- Prednosti stratifikovanog uzorka
- Mane stratifikovanog uzorka
- Case Study: Primena u analizi retail podataka (2009-2023)
- Kreiranje i priprema podataka u case study-ju
- Eksplorativna analiza u case study-ju
- Rezultati i zaključci case study-ja

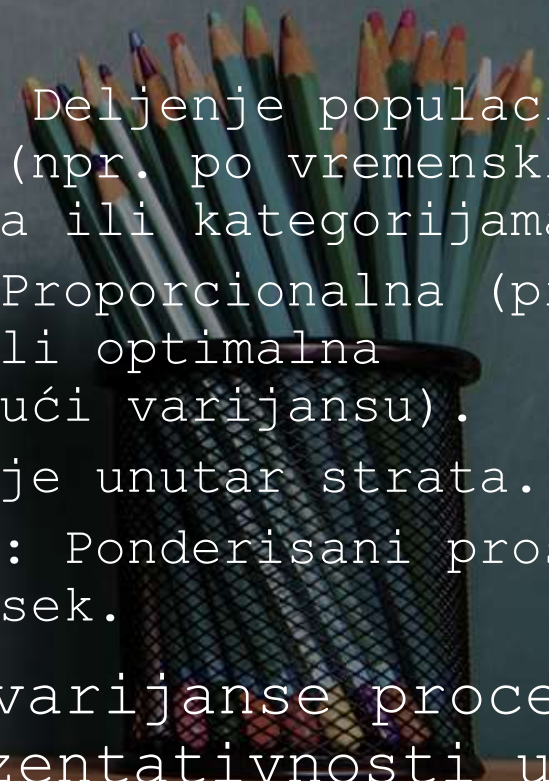


Teorijski pristup stratifikovanom uzorku

- Definicija: Stratifikovani slučajni uzorak je metoda uzorkovanja gde se populacija deli u međusobno isključive i kolektivno iscrpne grupe (strate) na osnovu ključnih karakteristika, a zatim se iz svake strate slučajno bira uzorak.

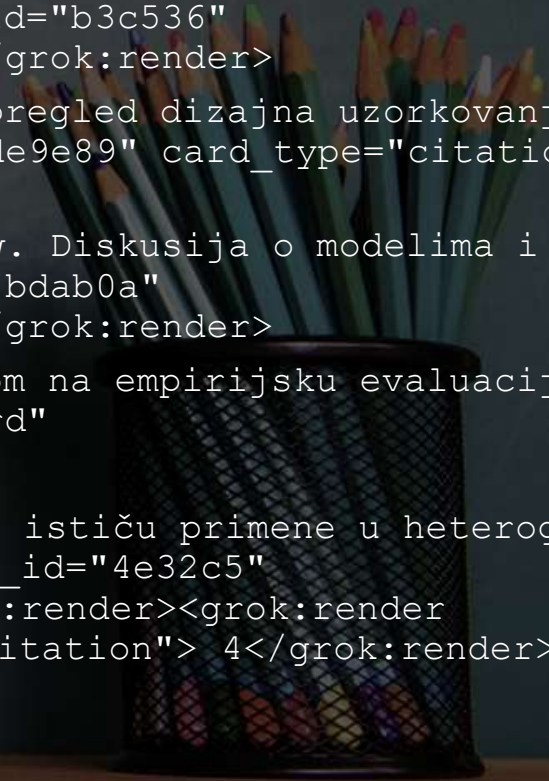
- Osnovni koraci:

- Formiranje strata: Deljenje populacije na homogene podgrupe (npr. po vremenskim periodima, zemljama ili kategorijama).
- Alokacija uzorka: Proporcionalna (prema veličini strate) ili optimalna (Neymanova, uzimajući varijansu).
- Slučajno uzorkovanje unutar strata.
- Procena parametara: Ponderisani proseci za populacioni prosek.
- Cilj: Smanjenje varijanse procene i poboljšanje reprezentativnosti u heterogenim populacijama.



Pregled literature

- Ključne reference:
 - Cochran, W.G. (1977). *Sampling Techniques*. Osnovni tekst o stratifikovanom uzorkovanju, fokus na formiranju strata i varijansama.<grok:render card_id="b3c536" card_type="citation_card" type="render_inline_citation"> 0</grok:render>
 - Lohr, S.L. (2019). *Sampling: Design and Analysis*. Detaljan pregled dizajna uzorkovanja, uključujući primene u istraživanjima.<grok:render card_id="de9e89" card_type="citation_card" type="render_inline_citation"> 10</grok:render>
 - Särndal, C.E., et al. (1992). *Model Assisted Survey Sampling*. Diskusija o modelima i optimalnoj alokaciji u stratifikaciji.<grok:render card_id="bdab0a" card_type="citation_card" type="render_inline_citation"> 2</grok:render>
 - Thompson, S.K. (2012). *Sampling*. Savremeni pristup sa fokusom na empirijsku evaluaciju metoda.<grok:render card_id="116dce" card_type="citation_card" type="render_inline_citation"> 16</grok:render>
- Savremena istraživanja: Članci na ResearchGate i ScienceDirect ističu primene u heterogenim populacijama, poput retaila i socijalnih nauka.<grok:render card_id="4e32c5" card_type="citation_card" type="render_inline_citation"> 1</grok:render><grok:render card_id="a82820" card_type="citation_card" type="render_inline_citation"> 4</grok:render>



Zašto koristiti stratifikovani uzorak?

- Glavni razlozi:
 - Povećava preciznost procena u heterogenim populacijama smanjenjem varijanse.
 - Osigurava reprezentativnost ključnih podgrupa (strata), sprečavajući pristrasnost.
 - Omogućava zasebnu analizu podgrupa (subgroup analysis), korisno za poređenja.
 - Efikasniji od prostog slučajnog uzorka kada postoje poznate razlike u populaciji.
 - Primer: U retail analizi, zašto? Zbog heterogenosti po vremenskim periodima (npr. krize vs. rast), što omogućava bolje razumevanje disrupcija poput COVID-19.



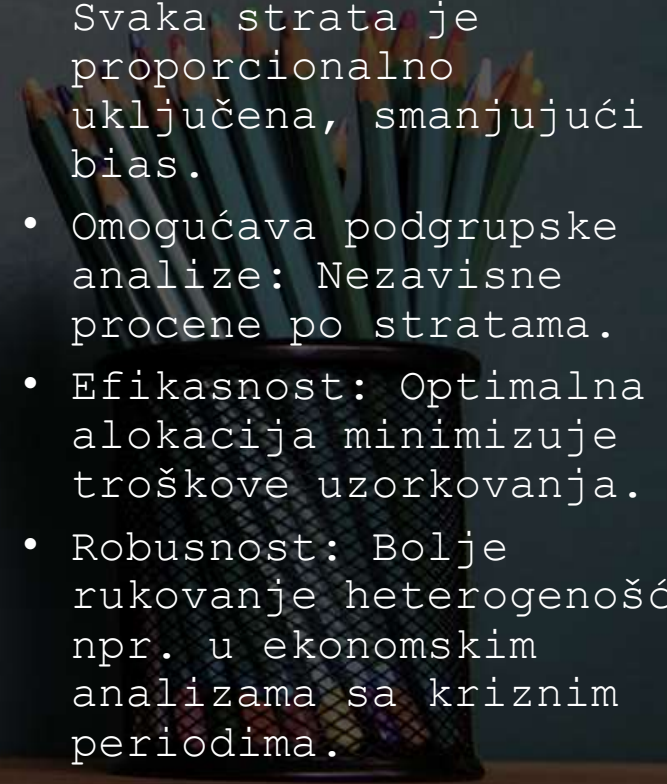
Kada koristiti stratifikovani uzorak?

- Situacije za upotrebu:
- Heterogene populacije sa jasnim podgrupama (npr. vremenski periodi, regioni, demografija).
- Kada je potrebna analiza po podgrupama ili testiranje robusnosti (npr. prirodni eksperimenti poput pandemija).
- Ograničeni resursi: Bolja efikasnost sa manjim uzorcima.
- Poznate populacione karakteristike: Veličine strata i varijanse dostupne.
- Ne koristiti: U homogenim populacijama ili bez informacija o stratama (kompleksnije od prostog uzorka).
- Primer: Analiza retail podataka sa disrupcijama (2009-2023), gde su strate vremenski periodi za bolju reprezentativnost.



Prednosti stratifikog vanog uzorka

- Prednosti:
 - Veća preciznost:
Smanjena varijansa procene (do 35-40% bolje od prostog uzorka po simulacijama).
 - Bolja reprezentativnost:
Svaka strata je proporcionalno uključena, smanjujući bias.
 - Omogućava podgrupske analize: Nezavisne procene po stratama.
 - Efikasnost: Optimalna alokacija minimizuje troškove uzorkovanja.
 - Robusnost: Bolje rukovanje heterogenošću, npr. u ekonomskim analizama sa kriznim periodima.



Mane stratifikovanog uzorka

- Mane:
- Kompleksnost: Zahteva prethodno znanje o populaciji za formiranje strata.
- Potreba za informacijama: Veličine strata i varijanse moraju biti poznate, što može biti izazov.
- Veći troškovi pripreme: Dizajn i implementacija su složeniji od prostog uzorka.
- Rizik grešaka: Nepravilno formiranje strata može dovesti do biasa ili veće varijanse.
- Ograničenja: Manje efikasno u malim populacijama ili bez jasnih podgrupa.



Case Study: Primena u analizi retail podataka (2009- 2023)

- Uvod u case study: Primena stratifikovanog uzorka na analizu vremenskih disrupcija i tržišne otpornosti u retail sektoru.
- Kontekst: Heterogena populacija sa periodima: Post-Crisis (2009-2012), Stable Growth (2013-2019), COVID-19 (2020-2021), Recovery (2022-2023).
- Zašto stratifikacija? Heterogenost po periodima, COVID-19 kao test robusnosti, analiza po podgrupama (zemlje, proizvodi).



Kreiranje i priprema podataka u case study-ju

- Dataset: Ukupan broj transakcija: 25.805, 4 strate (vremenski periodi), 6 zemalja, 5 kategorija proizvoda, 15 godina.
- Karakteristike: Ukupna populacija $N = 25.805$, pravi prosek $\mu = €2.475$.
- Tabela karakteristika strata:

Period	Veličina (N_h)	Ponder (%)	Prosek (€)	Std. dev. (€)	CV
COVID	3300	13.2	1707	251	0.147
Post-Crisis	6580	26.3	1757	299	0.170
Recovery	3316	13.3	2790	396	0.142
Stable Growth	11809	47.2	3002	502	0.167

- Insert Chart: Evolucija



Eksplorativna analiza u case study- ju



Osnovna statistika po periodima:

Period	Broj transakcija	Prosečne prodaje (€)	Medijan (€)	Std. dev.	Min (€)	Max (€)
COVID Period	3300	1707	1704	251	797	2638
Post-Crisis	6580	1757	1755	299	769	2871
Recovery	3316	2790	2784	396	1382	4240
Stable Growth	11809	3002	2995	502	1164	4906

Vizualizacije: Boxplot distribucije prodaja po periodima; Line chart vremenske evolucije.

Uvidi: Stable Growth ima najviše prodaje (~€3.002), COVID pad (~€1.706), heterogenost potvrđena.

Rezultati i zaključci case study-ja

- Monte Carlo simulacija: 1000 iteracija, uzorak 500; Stratified Optimal ima 68.4% poboljšanje varijanse vs. Simple Random.
- COVID analiza: Pad prodaje ~31%, manja varijabilnost u COVID periodu.
- Zaključci: Stratifikacija poboljšava efikasnost u heterogenim podacima; Preporuke: Koristiti za disrupcije poput pandemija; Ograničenja: Kompleksnost implementacije.
- Literatura u case study-ju: Cochran (1977), Lohr (2019), itd.
- *Napomena: Koristite ovu strukturu za PowerPoint/Google Slides. Ubacite grafikone i tabele iz dokumenta gde je naznačeno. Citacije integrisane sa izvorima iz pretrage.*



Thank you
for
watching!







