

Uzorkovanje u analizi sadržaja

Čemu uzorkovanje?

- Populacija tekstova je najčešće previše velika da bi bila analizirana u potpunosti
- Analiziranje dijela tekstova može unijeti pristrasnost u analizu i u naše zaključke
- Teorija uzorkovanja zabrinuta za mogućnost generalizovanja karakteristika uzorka na cijelu populaciju

Uzorkovanje u analizi sadržaja

- Statistička teorija uzorkovanja nastala u istraživanjima javnog mnjenja
- Uzorkovanje teksta razlikuje se na više načina:
 - Uzorkovanje pojedinaca znači da su elementi uzorka nedjeljivi, dok značenje "teksta" može biti drugačije definisano zbog unutrašnje hijerarhije: žanr teksta, tekst, naslov, paragraph...
 - Moguće je uzorkovati slova, riječi, rečenice, tekstove, izdanja...
 - Ankete daju kontrolu oko toga što će biti pitano i koliko će legitimnim smatrati konkretne odgovore. Svi ispitanici tretirani jednake važnosti

Plan uzorkovanja

- Uzorkovanje iz populacije zahtijeva postojanje plana koji oslobađa pristrasnosti
- Isto je kao i sa anketama samo onda kada su svi tekstovi jednako informativni - **tehnike uzorkovanja koristeći nasumičnost/vjerovatnoću**
- Kada jedinice nijesu jednako informativne (češće slučaj) – **tehnike bazirane na našem znanju o distribuciji informacija kroz tekstove**

Plan uzorkovanja

- Da li smo u prilici da pobrojimo sve uzoračke jedinice (populaciju tekstova)?
- Ne možemo unaprijed odrediti ukupan broj tekstova koji čine našu populaciju?

1.1. Nasumično uzorkovanje

- Potrebno je prvo pobrojati i numerički obilježiti svaki pojedinačni tekst u populaciji
- Randomizujemo (nasumično) izvučemo određenu količinu tekstova u uzorak
- Svaki tekst popisani tekst ima istu šansu da bude izvučen u uzorak
- Lako primjenjivo u bilo kojem statističkom programu (npr. Excel)

1.2. Sistematični uzorak

- Istraživač bira svaki N-ti tekst iz populacije
- Nasumično izvuče prvi tekst i onda nakon toga u analizu uzima svaki idući konstantni broj
 - Primjer: Recimo da smo izvukli nasumično da prvi tekst bude treći, da da želimo uzimati svaki peti. Onda bi u uzorak ušli tekstovi po redosljedu: 3, 8, 13, 18, 23, 28...
- Nasumičnost oslobađa pristrasnosti dok konstanta (ritam) smanjuje broj uzorkovanih tekstova

1.3. Stratifikovano uzorkovanje

- Razdvaja populaciju u zasebne kategorije (stratume) i onda zasebno nasumično izvlači određeni broj tekstova iz svake kategorije
- Najčešće garantuje minimalnu prisutnost svih tipova tekstova koje su od važnosti za istraživača
 - Primjer: uzorkovanje medija po geografskoj prisutnosti, tiražu, uzorkovanje žanra teksta: info, kolumna, intervju, govori...

2.1. Uzorkovanje nejednakih vjerovatnoća

- Prepoznaje činjenicu da često tekstualne jedinice nijesu jednako informativne u odnosu na istraživačko pitanje
- Različitim jedinicama analize daje posebne vjerovatnoće da budu uključene u uzorak
 - **Primjer:** veća vjerovatnoća data tekstovima iz najtiražnijih medija, koji navodno imaju više efekta na javno mnjenje
- U slučaju nedostatka objektivnog kriterijuma za određivanje važnosti – ekspertske rangiranje
- Korisno za ispravljanje pristrasnosti u korist poznatih ličnosti u mainstream medijima

2.2. Klaster uzorkovanje

- Koristi se kada istraživač ne može pobrojati sve jedinice analize, može pobrojati listu šire grupe takvih jedinica (klastera)
- Određivanje klastera tekstova i onda odabir unutar te grupe nasumično, stratifikovano ili sistematski
- Vjerovatnoća za odabir zavisi od (nepoznate) veličine klastera
- U analizi sadržaja korisna tehnika jer su kategorije koje istražujemo često previše široke da bi uvijek mogli jasno odrediti listu svih jedinica
- Ekonomična tehnika u vremenu i resursima

2.3. Snowball uzorkovanje

- Tehnika koja se dešava u više faza: počinje sa inicijalnim uzorkom jedinica iz kojeg se iznova uzorkuje
- Svaka novo uzorkovanje kreira novu sekvencu dodatnih uzoračkih jedinica, zbog čega uzorak raste dok god ima novih jedinica koje ispunjavaju uzoračke kriterijume
 - Primjer: Istraživači uzorkuju naučni članak, zabilježe reference i citirane radove iz kojih uzorkuju nove slične
 - Primjer: Povezani članci na portalima
- Proces se ponavlja sve dok pronalazimo nove tekstove (ne dobijamo duplikate već obrađenih tekstova)
- Pretpostavka intertekstualnosti: jedinice teksta su povezane
- Obavezno je oprezno odabrati inicijalni set jedinica

2.4. Uzorkovanje po relevantnosti

- Prethodne tehnike se mogu koristiti bez značajnog čitanja uzorkovanih tekstova
- Uzorkovanje po relevantnosti je **namjerno** zbog odabira na osnovu već postojećeg uvida u to da sadržaj odgovara istraživačkom pitanju
- Koristeći poznate kriterijume sužavati količinu tekstova koje ispunjavaju kriterijume (npr. keywords) i koje je moguće obraditi
- Postoji konceptualna hijerarhija, kroz sistematično smanjenje broja jedinica
- Izabrane jedinice nijesu reprezentativne za populaciju tekstova, već su oni sami populacija relevantnih tekstova
- Kada iscrpimo sve mogućnosti za sužavanje uzorka, možemo primijeniti ostale tehnike uzorkovanja

2.5. Popis

- Potpuni skup tekstova nazivamo popisom (cenzusom)
- Moguće je analizirati sva djela određenog autora, što ne zahtijeva uzorkovanje
- Zahtjevno za sakupljanje ali ne zahtijeva analitičko odlučivanje o tome što da isključimo ili uključimo u analizu

Veličina uzorka

- Nakon što odlučimo o planu uzorkovanja postavlja se pitanje koliko jedinica je dovoljno da bi adekvatno odgovorili na istraživačko pitanje
- Do finalnog broja možemo doći na tri načina:
 - **Uzoračke tehnike:** ako su slučajevi od važnosti rjeđi, potrebno je uzorkovati veći broj jedinica
 - **Uzorački eksperiment:** više uzoraka različitih veličina i pratimo poslije kojeg broja efekat je isti
 - **Tehnika podjele na pola:** nasumično podijelimo uzorak na dva dijela jednake veličine. Ako oba uzorka nezavisno vode do istog rezultata onda je cijeli uzorak adekvatne veličine