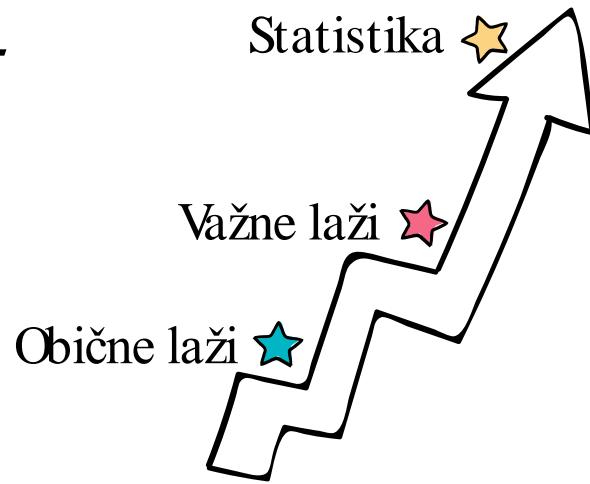


Statistika

Luka Stamenković
Filip Mancić
Marta Kiso
Miloš Bigović
Petar Deljanin
Nina Stamatović
Nađa Lazarević

- Bendžamin Dizraeli (1804 - 1881), britanski državnik:

Postoje tri vrste laži:



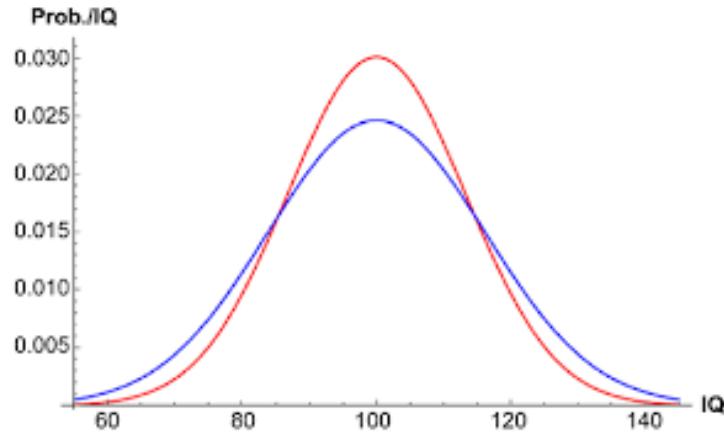
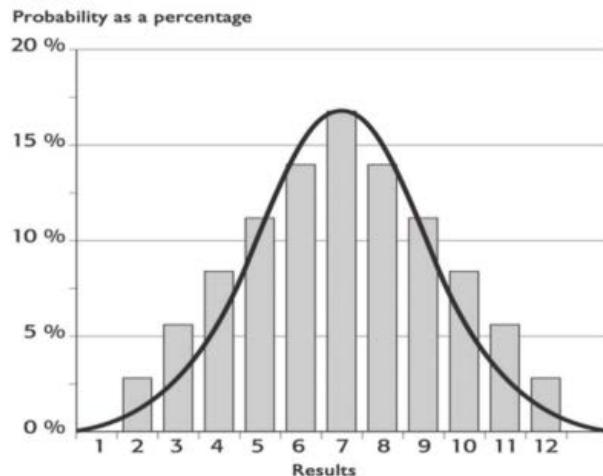
- Hērbert Džordž Vels (1866 - 1946), engleski naučnofantastični književnik:

„Kao što je sada potrebno da znamo da pišemo i čitamo, tako će u 20. veku biti potrebno da znamo statistiku.“

- Reč statistika se koristi na dva načina.
- S jedne strane, odnosi se na kvantitativne podatke – na primer, statistiku o stopi razvoda u Kaliforniji.
- S druge strane, odnosi se na granu matematike koja koristi i razvija metode za prikupljanje, prikazivanje i analizu podataka.

Gausova kriva – primer:

- Zbir koji dobijamo bacanjem dve kockice.
- Normalna raspodela ili Gausova kriva.



• Mere deskriptivne statistike



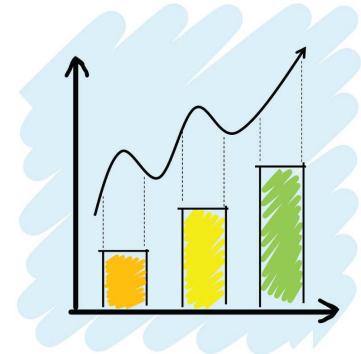
- Najčešće mere koje se koriste u deskriptivnoj statistici:

Aritmetička sredina

Medijana

Modus

Standardna devijacija



- Ove tri mere centralne tendencije ne daju nužno istu vrednost.

S obzirom na to, one se mogu koristiti za obmanu: sve što morate da uradite je da izaberete među njima meru koju želite da se pojavi, a koja možda nije zaista reprezentativna za podatke.

● Aritmetička sredina

Uobičajena aritmetička sredina:

$$\frac{\text{zbir elemenata}}{\text{broj elemenata}}$$

Aritmetička sredina se najčešće koristi.

Uzimaju se svi podaci u razmatranje.

Ipak, osetljiva je na ekstremne vrednosti.

Aritmetička sredina je srednja vrednost svih podataka uključenih u uzorak. Dobija se

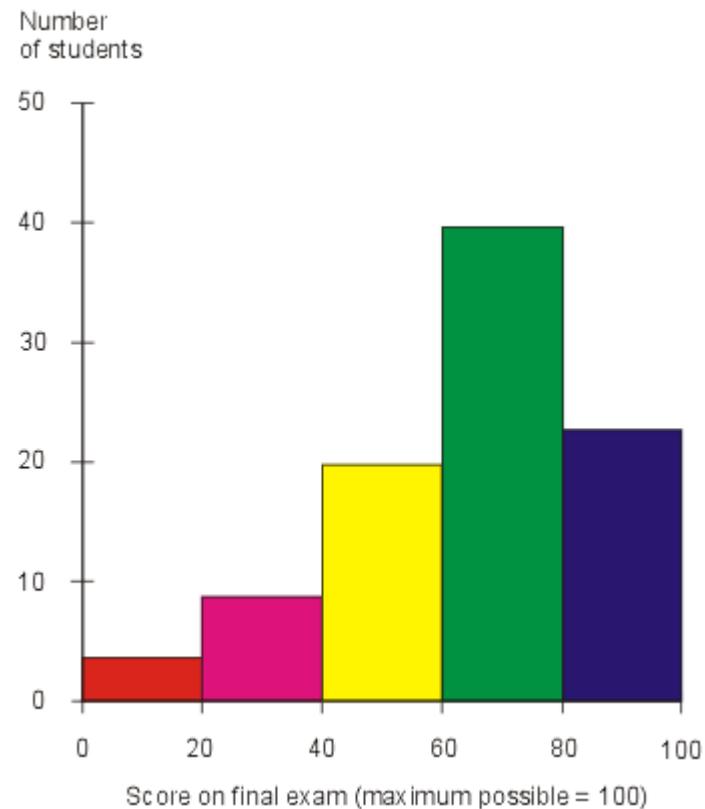
tako što se saberi vrednosti svih podataka, a zatim se taj zbir podeli sa brojem unosa u uzorku.

Na primer, ako u jednom preduzeću:

- 47 radnika zarađuje 1.000 € mesečno
- gazda zarađuje 1.000.000 € mesečno
- sredina je $47 \times 1.000 + 1.000.000 / 48 = 21.182,50 \text{ € mesečno}$
- Iz ove aritmetičke sredine se ne može izvući podatak o raspodeli vrednosti!

○ Histogram, grafik raspodele

- Ne predstavljamo vrednost, već broj subjekata za tu vrednost
- Ovo omogućava da se relativizuje srednja vrednost
- Iz srednje vrednosti se ne vidi maksimum u raspodeli poena...
- Histogram je približna reprezentacija Gausove krive. Takođe je poznat kao zvonasta kriva i normalna distribucija.
- Prikazuje distribuciju mnogih slučajnih ljudskih ili prirodnih pojava.



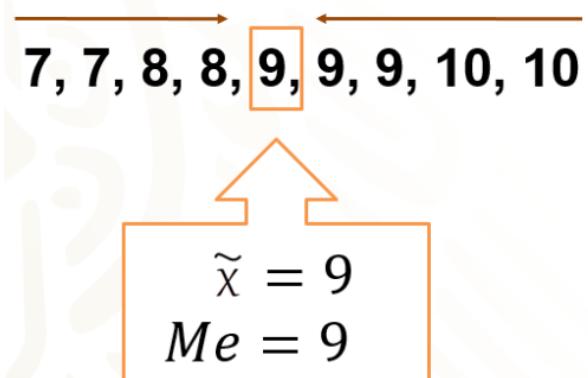
• Medijana

- Medijana – vrednost veća od jedne polovine predstavljenih vrednosti, a manja od druge polovine
- Medijana se u teoriji verovatnoće i statistici opisuje kao broj koji razdvaja gornju polovicu uzorka, populacije ili raspodele verovatnoće od donje polovine.
- Medijana se takođe često koristi. Ne uzima u obzir sve vrednosti i nije osetljiva na ekstremne vrednosti. Ponekad može bolje da opiše podatke od aritmetičke sredine.

Npr: Da bismo pronašli medijanu, sortiramo podatke na sledeći način:

\$109, \$129, \$129, \$135, \$139, \$149, \$159, \$179

• Pošto imamo paran broj podataka (8), uzmemo dve srednje vrednosti (\$135 i \$139), sabiramo ih, delimo sa dva i dobijamo medijanu: \$137.



• Modus

- Modus je vrednost koja se u uzorku ili grupi podataka pojavljuje najčešće.
- Modus se najčešće koristi.
- U nekom uzorku može biti više modusa ili ni jedan.
- Ne posmatra sve vrednosti.

\$109, \$129, \$129, \$135, \$139, \$149, \$159, \$179

Modus = 129

Primer

- Ponekad se oslanjanje samo na jednu od ovih mera centralne tendencije može pokazati zavaravajućim, u smislu da izabrana mera ne daje tačnu ideju o tome šta je tipično za skup podataka.

Zamislite, na primer smer filozofije na fakultetu koji sa ponosom objavljuje da je prosečna godišnja zarada njegovih diplomaca 242.000 dolara.

Pretpostavimo da je jedan od diplomaca takođe hokejaš i da je na kraju studija angažovan da igra za profesionalni tim.

- Naime, srednja vrednost je mera centralne tendencije koja je osetljiva na ekstremne podatke. U slučajevima poput ovih, bolje je osloniti se na drugu mera centralne tendencije.

● Značajnost (standardna devijacija)

Otišli ste da pecate na jezero, ali Vam je rečeno da su se ranije u jezeru izlivali toksini iz obližnje fabrike.

Rekli su Vam da je za čoveka štetno ako u ribi koju pojede ima 7mg toksina. Takođe su Vam rekli da u proseku u ribi ima 4mg toksina.

Da li biste pojeli ribu?

Da li biste pojeli ribu ako znate da je standardna devijacija 1mg?

A u slučaju da je standardna devijacija 4mg?

● Značajnost (standardna devijacija)

- Stepen značajnosti jednog statističkog rezultata jednak je verovatnoći da razmak (devijacija) između : rezultata posmatranja i teorijskog predviđanja bude posledica slučaja.

Standardna devijacija:

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

x_i – i -ti član uzorka

μ – aritmetička sredina uzorka

N – broj elemenata u uzorku

- *Standardna devijacija (σ) nam govori koliko elementi nekog skupa odstupaju od aritmetičke sredine skupa*

● Značajnost (standardna devijacija)

Prvi način: *potreban vam je kalkulator i možete ga dobiti kada pritisnete na dugme namenjeno za to.*

Drugi način:

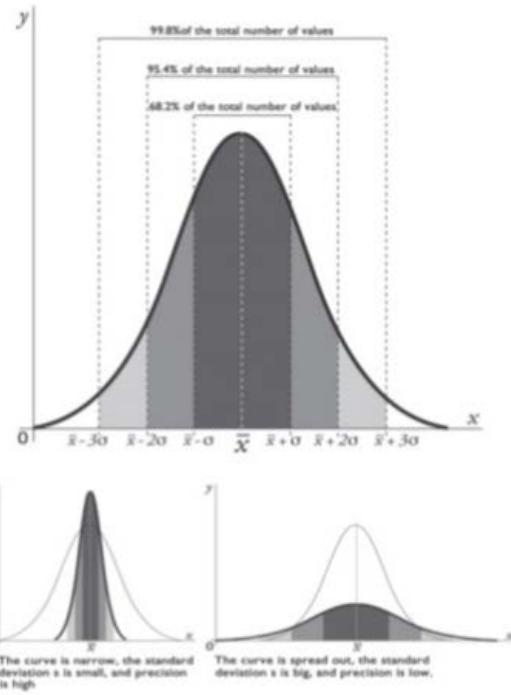
- 1. Izračunajte rasipanje svake vrednosti u odnosu na srednju vrednost koju ste već izračunali.*
- 2. Kvadrirajte svaku razliku i saberite ih.*
- 3. Podelite sumu sa brojem vrednosti: to je varijansa ili disperzija.*
- 4. Uzmite kvadratni koren ove varijanse i dobijete standardnu devijaciju.*

Treći način :

- 1. Uzmite najveću vrednost vašeg uzorka, a zatim oduzmite najmanju. Na taj način ste pronašli numeričku varijaciju rezultata, nazvanu raspon.*
- 2. Zatim podelite dobijeni broj sa četiri.*

● Značajnost (standardna devijacija)

Drugim rečima, ako je srednja vrednost dvanaest, a standardna devijacija tri, oko 68.2 % promatranih vrednosti padaju unutar raspona od devet do petnaest.





Značajnost

Primer: testiranje da li je kocka dobro balansirana?

- *Na 6 bacanja: 3 puta broj 5*

Uzorak je suviše mali pa rezultat nije validan.

- *Na milion bacanja: 900.000 puta broj 5*

Rezultat je realan, ali ga je vrlo teško postići.

- *Na 1.000 bacanja: 248 puta pada broj 5*

Prag značajnosti zavisi od testa:

- *Za kockice, brojimo slučajeve.*
- *Za anketu, moramo voditi računa o uticaju, ne smemo navesti za koga ćemo mi glasati npr.*
- *Za proveru sličnosti između dva pismena zadatka, mora se voditi računa o tome da nema baš 50 mogućih odgovora na neka pitanja.*





Uzorkovanje

- Probleme kao što su procenat radne snage koji je nezaposlen ili šta su stanovnici jedne zemlje gledali na televiziji u najvećem broju, statistika pojednostavljuje koristeći tehniku pod nazivom uzorkovanje.

deo osnovnog skupa - **uzorak**

- Istraživač na osnovu nalaza dobijenog ispitivanjem uzorka nastoji da izvede zaključak o celokupnom osnovnom skupu.

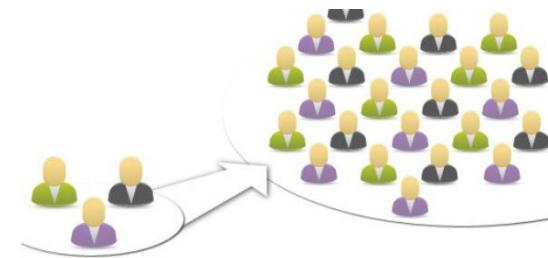
• Umesto da pratimo ponašanja milijardi ili miliona, potrebno je samo da ispitamo hiljade ili stotine.





Uzorkovanje

- Na osnovu nalaza dobijenog is traživanjem, is traživač nastoji da predviđa slučajeve koji nisu obuhvaćeni is traživanjem.
- Definis anje uzoraka obuhvata važan I složen deo statistike.
- Da bi zaključak o populaciji bio validan, analiziran uzorak treba biti reprezentativan za tu populaciju.
- Ovaj kriterijum je ključan, a da bi bio zadovoljen naš uzorak mora biti dovoljno velik i dobro baziran.





Uzorkovanje

Uzorkovanje je izvlačenje (ili niz izvlačenja) pojedinačnog člana osnovnog skupa, pri čemu ono može biti:

slučajno:

- *prost slučajni uzorak*
- *sistematski uzorak*
- *stratifikovani uzorak*
- *grupni (klasterski) uzorak*

namerno:

- *prigodni uzorak*
- *kvotni uzorak*
- *namerni uzorak*





☆ Uzorkovanje ☆

U osnovi teorije uzoraka i statističkog zaključivanja o osnovnom skupu na osnovu uzorka su najmanje dve teoreme:

zakon velikih brojeva

- Šta nam daje pravo da na osnovu samo jednog uzorka zaključujemo o osnovnom skupu?
- Ako smo izvršili ispitivanje na dovoljno velikom uzorku možemo sa velikom verovatnoćom tvrditi da će absolutna razlika između aritmetičke sredine uzorka i aritmetičke sredine osnovnog skupa biti manja.
- Ali na osnovu toga ne možemo neposredno da odredimo verovatnoću pojave greške u zaključivanju o osnovnom skupu na osnovu uzorka (tzv. „greške uzorka“).

centralna granična teorema

- Za slučajne uzorce izvučene iz ma kojeg osnovnog skupa sa konačnom sredinom i varijansom, uzoračna distribucija aritmetičkih sredina slučajnih uzorka biće približno normalna.
- Standardna greška je manja ako je veći uzorak.

**MISTAEKS
HAPPEN**





Istraživanja i uzimanje uzorka

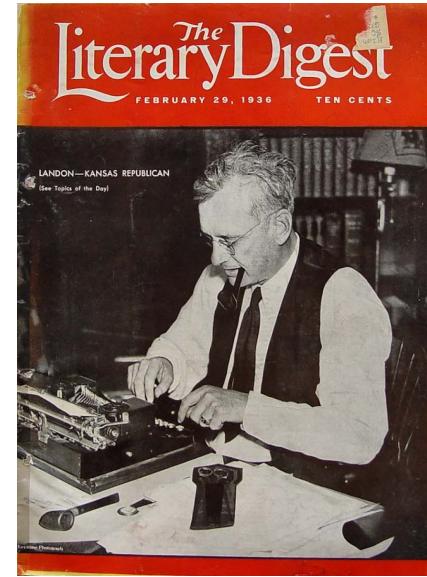
- Statistika nam omogućava da zaključujemo o osobinama neke populacije na osnovu malog dela te populacije koji nazivamo uzorak.
- Uglavnom nije moguće ispitati celu populaciju neke države radi istraživanja. Zato uzimamo uzorak.
- Kreiranje uzorka i zaključivanja na osnovu tih uzorka je najrasprostranjenija i najvažnija primena statistike.
- Da bi zaključak o nekoj populaciji bio validan, uzeti uzorak mora da bude predstavnik te populacije.





⌚ Validnost uzorka ☒

- “Literary Digest“ je od 1920. sprovodio istraživanje o pobedniku Američkih izbora. Iako je prethodnih godina bio veoma uspešan, 1936. godine rezultat istraživanja je bio netačan.
- Časopis je svojim pretplatnicima slao anonimne ankete i poslao je oko 10 miliona anonimnih anketa, a dobio bi oko 2 miliona odgovora.
- Druga kompanija, znatno manja, “Gallup’s“ sprovele istraživanje nad manjim skupom ljudi (4.500) i oni su imali tačnu pretpostavku iako je uzorak ljudi bio dosta manji.
- *Da li možemo da prepostavimo razlog zašto je rezultat manjeg uzorka bio tačniji od većeg uzorka?*





Veličina uzorka

- Kako odrediti dobru veličinu uzorka u zavisnosti od situacije u kojoj se uzorak koristi? Postoji dosta faktora koji utiču na izbor veličine uzorka:
 - Ekonomski mogućnosti onoga koji koristi uzorak za istraživanje.
 - Veličina populacije koju uzorkujemo.
 - Preciznost u zavisnosti od potrebe anketa.
 - Pitanja na koja fokusiramo istraživanje.



Nekada nije potrebno birati ogroman uzorak pre nego manji, zbog tehničkih potreba naših istraživanja, takođe najbolji metod za anketiranje uzorka i saznavanja najtačnijih rezultata jeste slučajni izbor uzorka.



Metode biranja uzorka

Spomenućemo tri metode koje su veoma zastupljene: stratifikacija, klastEROvanje i uzorkovanje kvota.

Primer 1:



Radio stanica sprovodila je anketu koja se bavila temom legalizacije marihuane.

Na anketu je odgovorilo oko 3600 ljudi od kojih je čak 78% odgovorilo da se slaže sa idejom o legalizaciji marihuane.

Zbog čega je toliko velik procenat ljudi koji podržavaju ideju?





Metode biranja uzorka

- Loša strana sistema pozivanja nasumičnih telefonskih brojeva radi istraživanja o nekoj anketi ogleda se u čjenici da beskućnici i siromašniji ljudi možda nemaju telefone, pa se njihovo mišljenje o zadatoj temi čak ni ne dovodi u pitanje iako je odabir brojeva nasumičan.
- Baš iz ovakvih i sličnih razloga je teško napraviti dovoljno nasumičan izbor uzorka.

Margina greške



- Margina greške predstavlja broj koji se odnosi na slučajnu ugrešku uzrokovana.
- U kvalitetnim istraživanjima se navodi kolika je margina greške u procentima.
- Često se definiše kao radius intervala.

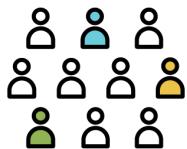


Formulisanje pitanja u anketi



Dobra anketa je ona koja je prvo testirana na manjem uzorku, da bi se proverio način formulacije pitanja kako bi svako pojedinačno razumeo pitanja na jedan način i kako pitanja ne bi bila prisrastna.

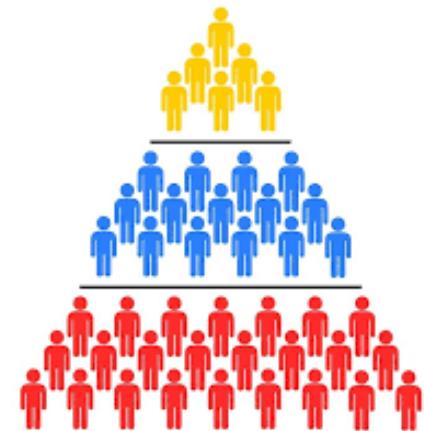
Ukoliko je testirana anketa na manjem uzorku pokazala prisrastnost i lošu formulaciju pitanja, možemo je prilagoditi tako da bude tačna kada se bude koristila na većem uzorku.





★Stratifikacija★

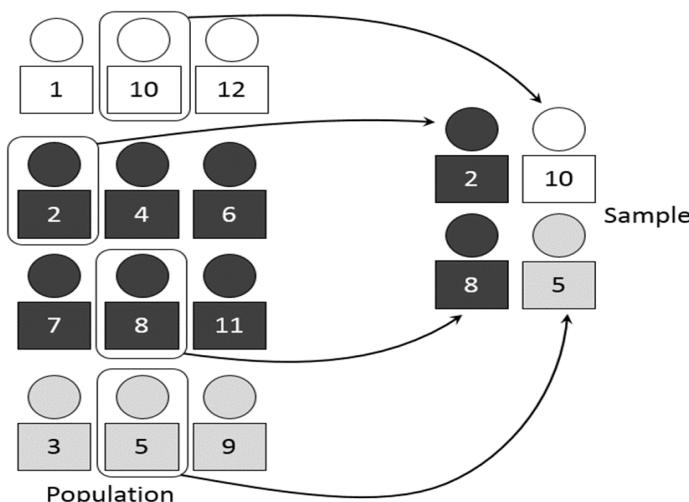
- Stratifikacija (raslojavanje) podrazumeva podelu populacije na delove - stratume (slojeve), disjunktne podskupove čija unija obuhvata celu populaciju, sa zahtevom postizanja što veće homogenosti unutar stratuma (sloja).
- Homogenost se ostvaruje prema nekom zajedničkom svojstvu elemenata populacije, na primer, starosna dob, pol, tip preduzeća i slično.
- Kako je osnovni cilj matematičke statistike ocenjivanje raspodele obeležja posmatranog na populaciji, cilj stratifikacije je da se postigne veća tačnost ocene, ekonomičnost ili jednostavnost ispitivanja i slično.





★ Stratifikacija ★

Primer



- U nekim situacijama je ispitivanje i jedino moguće sprovesti po stratumima.
- Tehnika stratifikacije podrazumeva rešavanje određenih zadataka, odnosno pronalaženje odgovora na sledeća pitanja:
 - *Kako formirati stratume i koliko njih?*
 - *Kako izabrati (raspodeliti, alocirati) ukupan uzorak uočavajući pojedine stratume, tj. alocirati uzorak po stratumima?*
 - *Kako sprovesti statističko zaključivanje na osnovu dobijenog stratifikovanog uzorka?*



★ Stratifikacija ★

- Stratifikovane strategije uzorkovanja:
 - a) Proporcionalna alokacija
 - b) Optimalna alokacija (disproporcionalna alokacija).
- Prednosti
- Mane

★ Korelacija ★

- *Korelacija je međuodnos ili međusobna povezanost između različitih pojava predstavljenih vrednostima dve promenljive. Pri tome povezanost znači da je vrednost jedne promenljive moguće s određenom verovatnoćom predvideti na osnovu saznanja o vrednosti druge.*
- *Korelacija takođe predstavlja i obrazac variranja promenljivih u zavisnosti od načina na koji su povezane, a koji je bitno drugačiji od njihovih izolovanih svojstava ili očekivanog načina reagovanja.*



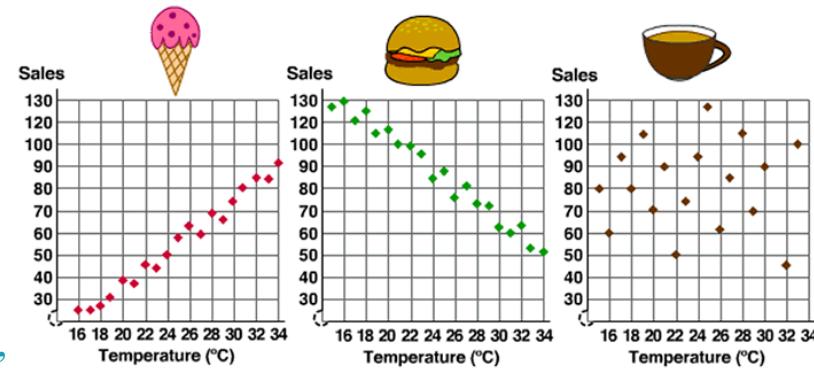
★Korelacija★

- U socijalnom radu se proučavaju korelacije ljudskih odnosa i preduzetih mera, usluga i rezultata.

- Visoke korelacije (izražene koeficijentom korelacije) dobar su pokazatelj uspeha neke aktivnosti. Školski primeri povezanosti su na primer saznanje o uticaju količine saliniteta na brojnost nekih vrsta riba ili školjki,

ili o povezanosti slane hrane i visokog krvnog pritiska i slično.

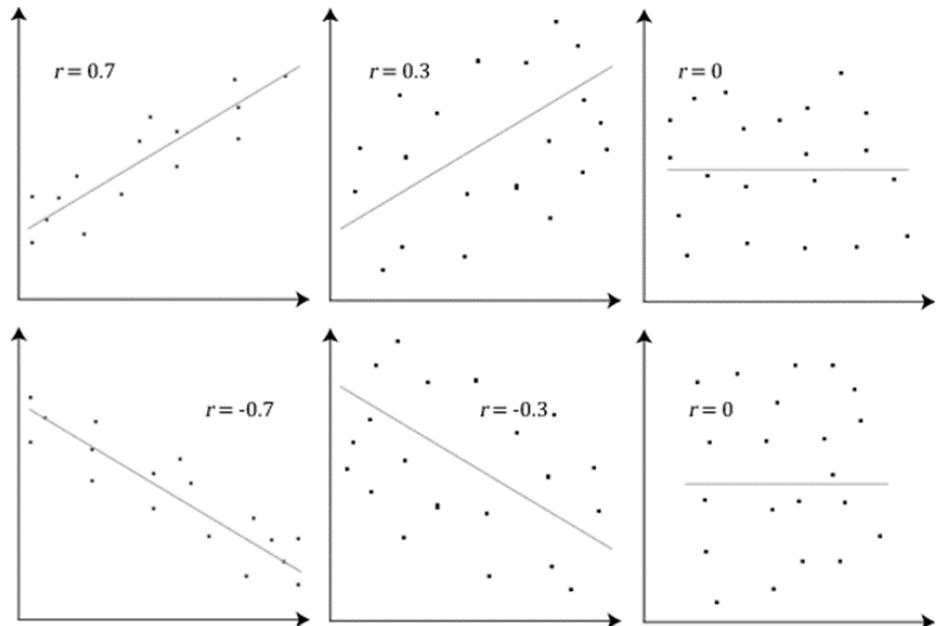
- Promena vrednosti jedne varijable utiče na promene druge varijable.
- Varijabla koja svojom vrednošću utiče na drugu naziva se nezavisna varijabla.
- Varijabla na koju ona utiče naziva se zavisna varijabla.





★ Pirsonov koeficijent korelaciјe ★

- Najpoznatija mera zavisnosti između dve veličine je Pirsonov koeficijent korelaciјe proizvoda i tačaka (PPMCC), ili Pirsonov koeficijent korelaciјe, koji se obično naziva jednostavno koeficijent korelaciјe.
- Matematički je definisan kao kvalitet najmanjih kvadrata koji odgovaraju izvornim podacima.



Raspršeni dijagrami primera različitih skupova podataka sa različitim koeficijentima korelaciјe.

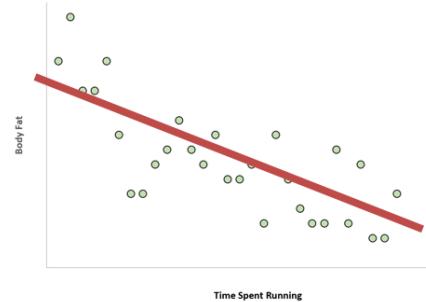


★Tipovi korelacija★

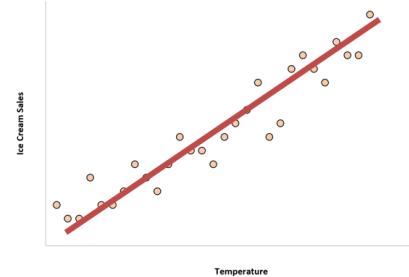
U statistici, korelacija je zapravo prikaz linearne zavisnosti dve varijable koje posmatramo.

Vrednost prethodno pomenutog koeficijenta korelacije može se kretati od -1 do 1, što znači:

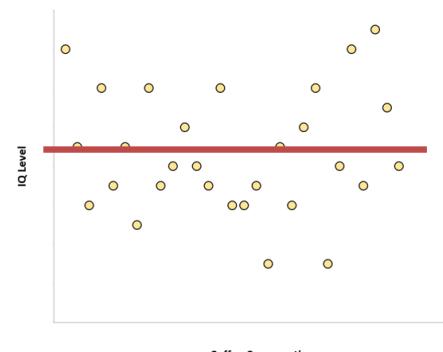
- -1, negativna korelacija
- 0, nepostojanje korelacije
- 1, pozitivna korelacija



Negativna korelacija



Pozitivna korelacija



Nema korelaciјe



★Korelacija★

Korelacija je često pogrešno shvaćena.

Kada kažemo da su dve promenljive A i B u korelaciji, to ne znači da među njima postoji uzročno posledična veza!

Naći uzrok nekog ponašanja jako je teško i predmet je naučnih istraživanja.

A i B su u korelaciji može da znaci sledeće:

- a) A je uzrok pojave B
- b) B je uzrok pojave A
- c) A i B su slučajno povezani bez bilo kakve uzročne veze između njih
- d) A i B su oboje zavisni od nekog trećeg faktora C.



★Korelacija★

Primer 1:

Đaci koji puše cigare – njihove ocene su lošije

Cena kafe u Oregonu – količina kiše koja pada u Oregonu

Broj dimnjaka u kući – broj dece u toj kući

Primer 2:

Monasi u Kini: Pomračenje meseca izaziva nebeski pas koji pojede mesec. Zato moraju da udare u veliki gong da bi psa oterali.

Upravo pogrešno razumevanje korelacije, uzroka i posledice vodi mnogim **sujeverjima**.



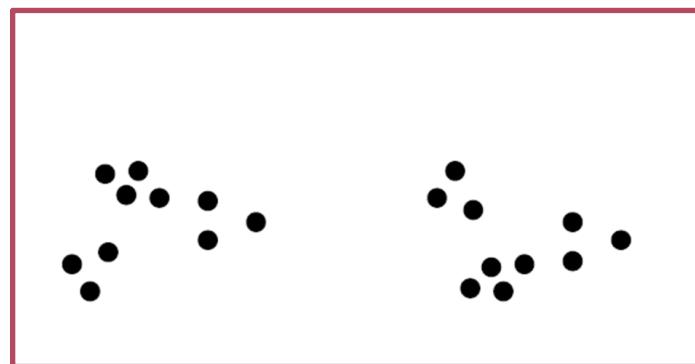


Klasterovanje

Klasterovanje je tehnika istraživanja podataka koja otkriva objekte podataka sličnih osobina i deli ih u grupe - *klastera*, čineći ih pregleđnijim i korisnijim.

Klasterovanje takođe predstavlja pronalaženje grupe objekata takvih da su objekti u grupi međusobno slični (ili povezani), i da su objekti u različitim grupama međusobno različiti (ili nepovezani).

-osnovna načela-



Koliko klastera se može ovde detektovati?



Klasterovanje



2 klastera



6 klastera



4 klastera

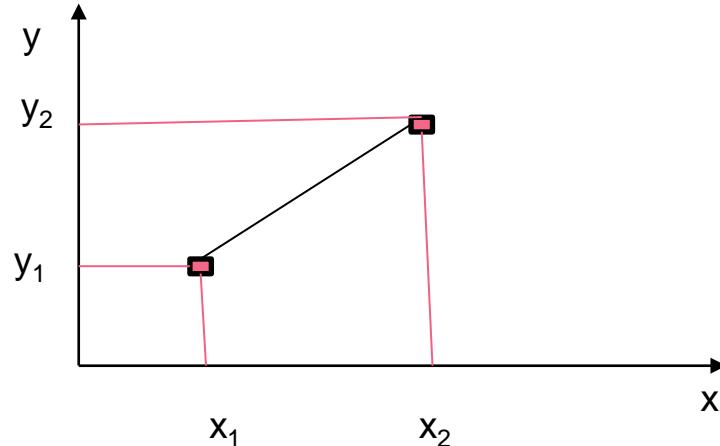


Klasterovanje

Iz prethodnog slučaja se vidi da je važno da se zna koje tačke podataka su blizu jedne drugim.

Takođe je važno i definisati metriku rastojanja koje će te tačke pravilno rasporediti u odgovarajuće grupe podataka koje su međusobno blizu.

U matematici se često koristi Euklidsko rastojanje.



$$E = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$



Klasterovanje

Postoje dva tipa klasterovanja:

Ravna - podela skupa podataka u nepreklopajuće podskupove (klastere) tako da je svaki podatak tačno u jednom podskupu.

Hijerarhijska - skup ugnježdenih klastera organizovanog u obliku hijerarhijskog drveta.

Neke od tehnika kojima se smanjuje količina izračunavanja i potrebne memorije u algoritmima za klasterovanje su:

Metode višedimenzionalnog pristupa

Uzorkovanje

Particionisanje objekata podataka

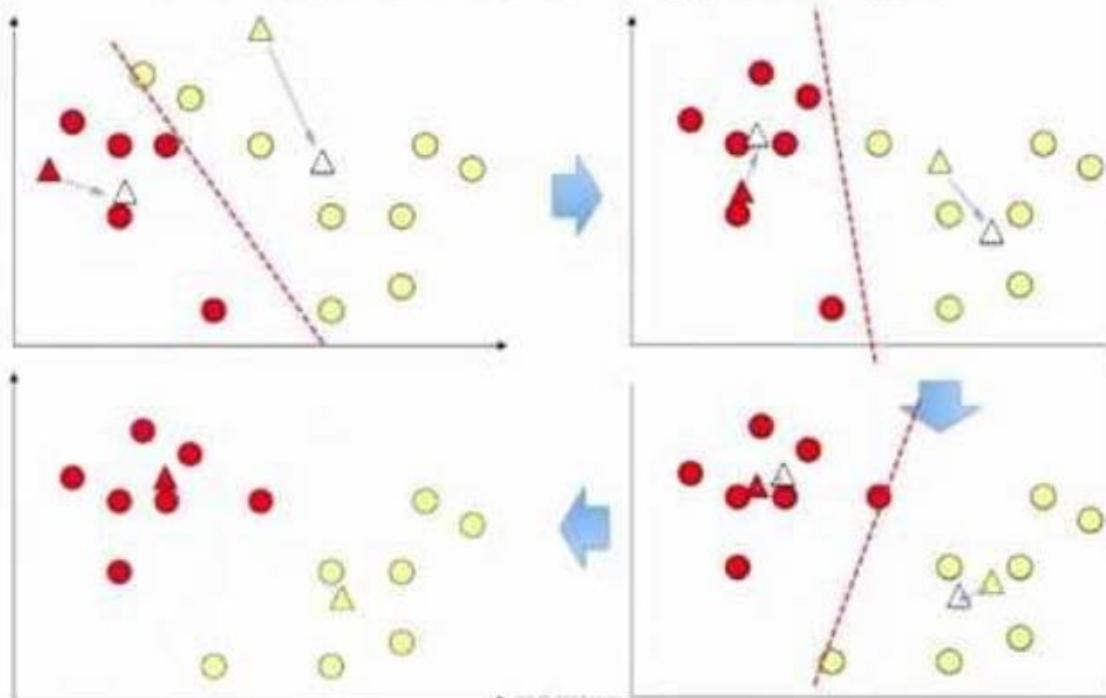
Pregledanje

Paralelno i distribuirano izračunavanje





K-means clustering example





Regresija ka sredini i sujeverje

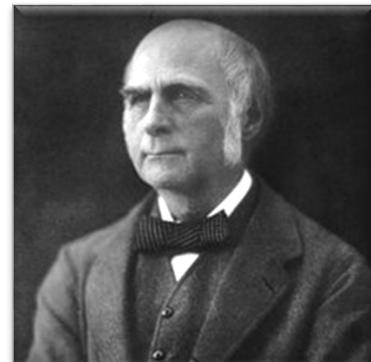


Regresija ka sredini se javlja kada kod dve vrednosti koje nisu u savršenoj korelaciji, a ekstremne vrednosti jedne promenljive su u korelaciji sa neekstremnim vrednostima druge promenljive (često u korelaciji sa srednjom vrednosti druge promenljive).

Ova pojava je sasvim obična, ali ako ne obratimo dovoljno pažnju na nju, možemo pogrešno povezati dve vrednosti u uzročno-posledičnom odnosu.

Ovaj fenomen objašnjava pojavu mnogih sujeverja.

Pojam je uveo engleski naučnik Francis Galton (statističar, sociolog, psiholog, antropolog, ...).





Primer- sinovi i očevi

Frensis Galton je ispitivao odnos visine između očeva i sinova:

Otkrio je očigledno: visoki očevi imaju visoke sinove, a niski očevi imaju niske sinove.

Međutim, otkrio je još nešto iznenadjuće... Veoma visoki očevi imaju sinove koji su niži od njih.

Veoma niski očevi imaju sinove koji su viši od njih.

Šta ovo znači?

Ovo je primer nesavršene korelaciјe dve promenljive.

Mnogo toga utiče na visinu dece: Majčina visina Način odgajanja, ishrane, mesto odgajanja, bavljenje različitim aktivnostima itd. Geni koji određuju dužinu udova, veličinu njegove lobanje itd. Puno faktora mora da se poklopi da bi osoba bila ekstremno visoka ili niska.

Verovatnoća da se svi uslovi poklope je jako mala.

To objašnjava zašto je slučaj ekstremno visokog oca u korelaciјi sa neekstremnim vrednostima visine sina.

Ova pojava je predvidiva i naziva se već poznata regresija ka sredini.



Sujeverje

Sujeverje (praznoverje) predstavlja kvazireligijska i nerazumna verovanja da su zbivanja u budućnosti uzrokovana određenim magijskim radnjama, bez logičnog uzročno/posledičnog odnosa između činjenja tih radnji i događaja u budućnosti.

Svaki narod ili etnička zajednica u svom folkloru imaju određene vrste sujeverja i praznoverja, ali su neka od njih preoblikovana u obeležja savremene pop kulture.

Neke sujeverne radnje:

Kada neko kreće iz kuće na ispit ili drugi važan posao, za njime se prosipa voda

Crna mačka donosi nesreću, te ako pređe nekom preko puta, treba se vratiti tj. ne presecati svojim kretanjem put kojim je prošla mačka.

Ujutru prilikom ustajanja iz kreveta treba prvo stati na desnu nogu, inače će tog dana sve ići loše.





Zanimljivost

Zakonom o javnom redu i miru Republike Srbije koji je donet 2016. godine u članu 15, stavci Uznemiravanje građana vračanjem, proricanjem ili sličnim obmanjivanjem, u šta se može ubrajati astrologija, tarot, numerologija i druga paganska/politeistička (mnogobožačka) verovanja stoji: Ko se bavi vračanjem, proricanjem sudskebine, tumačenjem snova ili sličnim obmanjivanjem na način kojim uznemirava građane ili narušava javni red i mir - kazniće se novčanom kaznom od 10.000 do 50.000 dinara ili radom u javnom interesu od 40 do 120 časova.

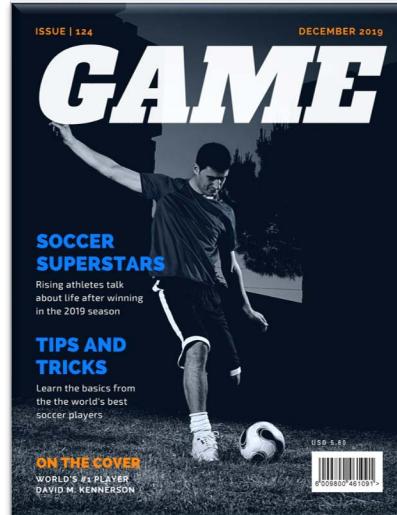
Primer sujeverja



Vrhunski sportisti koji se boje da se pojave na naslovnoj strani časopisa kako ne bi pokvarili svoje rezultate.

Baksuz, sujeverje ili prosto regresija ka sredini?

Međutim ovo je klasičan primer lančanog širenja sujeverja među ljudima.





Ai Logička zabluda

Poznavanje logičkih zabluda (logičkih grešaka) može biti jako važno u svakodnevnom životu.

Primetimo li logičku grešku u nečijem argumentu, odmah možemo zaključiti i da je argument nevaljan, odnosno da ne dokazuje ono što ta osoba želi prikazati istinitim.

Zašto je važno uočiti logičke greške?

Sledećih 10 logičkih grešaka najčešće se javljaju u javnom i svakodnevnom govoru:

- 1) *Ne sledi (lat. non sequitur)*
- 2) *Ignorisanje problema (eng. red herring, lat. ignoratio elenchi)*
- 3) *Zamena teza (eng. strawman argument)*
- 4) *Klizav teren (eng. slippery slope)*
- 5) *Argument ad hominem*
- 6) *Pozivanje na autoritet (eng. appeal to authority)*
- 7) *Lažna dilema (eng. false dilemma)*
- 8) *Uzimanje sleda za uzrok (lat. post hoc ergo propter hoc)*ž
- 9) *Uzimanje dela za celinu (pars pro toto, eng. hasty generalization)*
- 10) *Pozivanje na opšte mišljenje (eng. appeal to popularity or bandwagon fallacy)*

Ispitna pitanja

- Koje mere centralne tendencije se najčešće koriste?
- Objasni šta su aritmetička sredina, medijana, modus?
- Šta je standardna devijacija i zašto je ona značajna?
- Šta se postiže uzorkovanjem?
- Koji su kriterijumi veličine uzorka?
- Kako se definiše uzorak?
- Šta predstavlja klasterovanje? Kako možemo podeliti klasterizaciju po nivoima organizacije?
- Da li može da postoji korelacija između padanja kiše u Etiopiji i cene kafe. Objasniti.
- Koje su nam najbitnije vrednosti Pirsonovog koeficijenta korelacije?
- Kako određujemo kvalitet ankete?
- Šta je reegresija po sredini?
- Šta su sujeverja?
- Ko je Fransis Galton?
- Objasni logičku zabludu „ne sledi“.
- Objasni logičku zabludu „klizav teren“