KVANTITATIVNE METODE U GRAĐEVINSKOM MENADŽMENTU
predavanja 2017/18

MATEMATIČKA STATISTIKA

1. Definicija i osnovni pojmovi statistike: populacija, obilježje i uzorak
2. Prikupljanje i prikaz podataka (vrijednosti obilježja, frekvencije: apsolutne i relativne; kumulativne frekvencije: apsolutne i relativne)
3. Karakteristike empirijske raspodjele: mjere srednjih vrijednosti i odstupanja (srednja vrijednost, standardna devijacija, koeficijent varijacije)

P4

Definicija i osnovni pojmovi

- **Statistika:** matematička disciplina= metoda kvantitativnog istraživanja pojava na nekom skupu
- Statistika obuhvata:
  - deskriptivnu statistiku – prikupljanje, obrada i prikazivanje podataka;
  - induktivnu statistiku (inferencijalna statistika)= statističku analizu -bavi se metodama koje se zasnivaju na teoriji vjerovatnoće i koje omogućavaju dobijanje numeričkih pokazatelja i donošenje odluka o posmatranim masovnim pojavama i njihovim zakonitostima i karakteristikama
    - teorija ocjene (procjene),
    - teorija testiranja statističkih hipoteza,
    - teorija planiranja eksperimenata
  - statističku teoriju – razvoj novih statističkih metoda

Predmet proučavanja i etape

- **Predmet proučavanja:** pojave koje se masovno ispoljavaju sa različitim intenzitetima u pojedinim individualnim slučajevima.
- Ovo proučavanje podrazumijeva sprovođenje statističkog eksperimenta i sprovodi se u tri etape:
  - statističko posmatranje- plansko, sistematsko registrovanje statističkih jedinica, putem neke od metoda: mjerenja, brojanja, ocjenjivanja, evidentiranja i anketiranja. Greška u ovoj fazi može biti sistematska (ponavlja se) i slučajna.
  - grupisanje i sređivanje podataka- izrada statističkih tabela registrovanih i izvedenih vrijednosti (srednje vrijednosti , mjere varijacije, koeficijenti varijacije...)
  - obrada i naučna analiza rezultata- cilj je dobijanje statističkih ocjena obilježja skupa
- Vrste ispitivanja:
  - potpuno ispitivanje osnovne mase (populacije)
  - djelimično ispitivanje dijela ili uzorka- na reprezentativnom (slucajno odabranom i dovoljno brojnom uzorku)

Predmet teorije vjerovatnoće jeste proučavanje slučajnih promjenljivih i njihovih raspodjela, dok je predmet proučavanja statistike kvantitativno istraživanje pojava na mnoštvu uzoraka.

Nepotpune i neistinite prikupljene informacije u fazi statističkog posmatranja značile bi da konačan rezultat statističkog istraživanja sadrži grešku: sistematsku ili slučajnu (ili obje).
Sistematsku grešku je lakše uočiti jer se ona ponavlja (npr. neispravnost određenog mjernog instrumenta, neistinito izjašnjavanje ispitanika).
Slučajnu grešku je teško precizno odrediti, jer se ona ne javlja kod svakog mjerenja i ne javlja se istim intenzitetom. Za slučajnu grešku često se usvaja
pretpostavka da ne utiče na ukupnom nivou posmatranja

Statitički skup (populacija), elementi skupa i obilježja

- **Populacija (osnovni, odnoso statistički skup):** skup (konačan ili beskonačan: prebrojiv ili neprebrojiv) svih pojedinačnih elemenata (elementarnih=statističkih jedinica) iste vrste čija se obilježja (jedno ili više, od mnogih) posmatraju
- **Statistički skup** mora biti određen tako da se precizno odredi:
  - koje se obilježje posmatra (težina, visina, MB)
  - prostor kojem pripadaju elementi statističkog skupa
  - vremenski trenutak ili interval kojim će se obuhvatiti svi elementi koje ulaze i skup
- **Homogenost statističkog skupa (istovrsnost):** sve jedinice skupa su u suštini slične, razlikuju se samo po osnovu obilježja koja se posmatraju
- **Elementarne (statističke) jedinice, odnosno** (**element populacije)** mogu biti nosioci jedne ili više pojava, odnosno obilježja (kvantitativne ili kvalitativne karakteristike) koje se zajedno posmatraju. Koja će se obilježja posmatrati zavisi od toga koliko su ta obilježja važna za proučavanje neke pojave.
- **Statistička obilježja (varijable, promjenljive)** su opšte karakteristike elemenata statističkog skupa, po kojima su ti elementi međusobno slični i po kojima se međusobno razlikuju.

Uzorak

- **Uzorak**: dio populacije koji se posmatra (izabran na unaprijed određen način), kako bi se donio što je moguće tačniji sud o cijeloj populaciji, na osnovu suda o uzorku
- Uzorkom se dolazi do procjene karakteristika osnovnog skupa, a statističkom metodom određuje se pouzdanost i preciznost te procjene
- Adekvatan uzorak mora da ispuni principe:
  – nepristrasnosti- podjednaka vjerovatnoća da svaki od elemenata uđe u uzorak; postiže se načinom i metodom odabiranja uzorka koji su razrađeni u statističkoj praksi, a koji se baziraju na teoriji vjerovatnoće i postavkama slučajnog kombinovanja elemenata
  – reprezentativnosti - uzorak treba da obuhvati one statističke jedinice koje će u sebi nositi sve karakteristike osnovnog skupa, odnosno one jedinice čija obilježja, kada se izbroje ili izmjere i iz njihovih vrijednosti izračunaju odgovarajući parametri (aritmetička sredina, standardna devijacija i dr.), budu ista ili približno ista kao i pravi parametri osnovnog skupa; Reprezentativnost uzorka se postiže adekvatnom veličinom uzorka i objektivnim načinom izbora jedinica uzorka
  – ekonomičnosti- moraju biti prihvatljivi troškovi uzimanja i ispitivanja uzoraka

Na istoj populaciji može se posmatrati jedno ili više obilježja, kako bi se utvrdilo kako su obilježja raspoređena na datom statističkom skupu.
To nije uvijek moguće ispitati na cijelom statističkom skupu: zahtijeva mnogo vremena ili novca, a nekad je nemoguće (utvrditi kvalitet svakog proizvedenog proizvoda, ako se u tojm cilj koriste destruktivne metode ispitivanja, ili ako je skup beskonačan).
Zato se u statistici za određivanje rasporeda pojedinog obilježja ne posmatra uvijek cijela populacija već najčešće samo jedan njen dio, pa se zaključak koji se donese po osnovu posmatranog dijela populacije uopštava (proglašava važećim) za cijelu populaciju. Dakle, tada se vrši reprezentativno posmatranje kojim se obuhvata samo dio jedinica statističkog skupa.

Izvor: Vjerojatnost i statistika , Čuljak
Uzorak je reprezentativan ako po svojim osnovnim karakteristikama nalikuje na osnovni skup, odnosno predstavlja umanjenu sliku osnovnog skupa. Zbog toga je neophodno sastaviti jasan i precizan plan izbora elemenata u uzorak

Plan sadrži:
- ciljeve istraživanja,
- određivanje statističkih skupova - utvrditi što je jedinica skupa, opseg skupa, definirati skup pojmovno, prostorno i vremenski
- određivanje okvira izbora:
  - popis jedinica osnovnog skupa iz kojeg se bira uzorak, npr. registri poslovnih subjekata ili
  - popis jedinica izbora uzoraka koje obuhvataju više elemenata osnovnog skupa
- podatke koje treba prikupiti,
- model uzorka (nacrt, dizajn):
  - troškovi,
  - osobine osnovnih skupova,
  - način izbora elemenata u uzorak za svaki izbor
  - utvrđuju se izrazi za statističko-analitičke veličine iz uzorka, među kojima su i izrazi za veličine grešaka zbog primjene uzorka
- raspoloživa sredstva,
- postupke prikupljanja podataka

Vrste uzoraka- metode uzorkovanja

- **slučajni uzorci** - izbor zasnovan na principima teorije vjerovatnoće- bitan za statističko zaključivanje,
  - svaki član skupa ima vjerovatnoću izbora u uzorak veću od nule
  - podaci se analiziraju prema načelima statistike (procjenjuju se nepoznati parametri i testiraju hipoteze o njima i oblicima rasporeda osnovnih skupova)
  - mogu se izračunati greške nastale primjenom uzorka – važno za ocjenu kvaliteta zaključivanja
    - prost slučajan uzorak
    - sistematski uzorak
    - stratifikovan uzorak
    - klaster uzorci (uzorci skupina)
- **namjerni uzorci** – izbor nije zasnovan na principima teorije vjerovatnoće
    - Prigodni uzorak
    - Uzorak tipičnih jedinica
    - Kvotni uzorci

Slučajni uzorci

- *Prost slučajan uzorak* sastavlja se prema određenim načelima koji odgovaraju zakonu slučaja. Najbolji način je upotreba «tablice slučajnih brojeva» ili korištenje računarskog sistema slučajnog izbora (generator slučajnih brojeva), metoda lutrije itd.
- *Sistematski uzorak* - ako se jedinice iz osnovnog skupa biraju sistematski (na primjer, ako se po redu u uzorak bira svaki N/n-ti element iz osnovnog skupa koji broji ukupno N elemenata, a uzorak broji n elemenata- pogodan kada su statističke jedinice uređene po veličini)
- *Stratificirani ili slojeviti uzorak* – populacija se podijeli u slojeve ili stratume prema nekim karakteristikama (čime se stvaraju homogene klase u odnosu na dato svojstvo) pa se iz svake od grupa uzme slučajni uzorak
  - proporcionalni stratifikovani uzorak: isti procenat članova (svaki stratum u uzorku zastupljen je srazmerno svojoj veličini)
  - disproporcionalni stratifikovani uzorak: različit procenat članova
- *Klaster uzorci (uzorci skupina)* su lošija varijanta slučajnog uzorka kada se sva populacija podijeli na klastere (skupine), a ispituju se obilježja na elementima unutar nekih slučajno izabranih klastera

Izvor:http://www.medfak.ni.ac.rs/PREDAVANJA/2. STOMATOLOGIJA/STATISTIKA/6. predavanje.pdf

„**Prost – slučajan uzorak** se formira metodom tablica slučajnih brojeva i metodom lutrije. Prost – slučajan uzorak predstavlja jednostavan, korektan i relativno lak način uzorkovanja. Nedostatak ovog načina formiranja uzorka ispoljava se kada osnovni skup ima veliki broj jedinica, tada je otežana identifikacija onih jedinica koje ulaze u uzorak.

Najpoznatije tablice slučajnih brojeva su Fisherova, Tippetova, Kendallova i dr. Tablice slučajnih brojeva sadrže niz brojeva koji su na slučajan način raspoređeni u kolone i redove. Kako bi se formirao uzorak na osnovu tablica slučajnih brojeva potrebno je prvo napraviti spisak svih statisitčkih jedinica osnovnog skupa, takozvani «okvir skupa», a zatim numerisati statističke jedinice od 1 do N. Tada se, na slučajan način odredi prvi broj u tablici koji predstavlja i prvu jedinicu u uzorku, a zatim se redom izdvaja n brojeva manjih od N.
Jedan deo Tippetove tablice slučajnih brojeva:
5624 4167 9524 1545 1396 7203 5356
1300 2693 2730 7483 3408 2762 3563
1089 6913 7691 0560 5246 1112 5107
6008 8126 4233 8776 2754 9143 1405

Uzmimo na primer da iz skupa od 50 bolesnika treba formirati uzorak od 10 ispitanika (20%), pod pretpostavkom da smo slučajno izabrali kao prvi broj 5624, zatim krećemo udesno za izborom jedinica za uzorak. Biramo dvocifrene brojeve manje od 50, i to sa poslednje dve cifre iz grupe. Na taj način biramo sledeće jedinice: 24 45 3 30 13 46 12 7 26 33 Kako su se brojevi 24 i 8 ponovili, njih preskačemo, i nastavljamo biranje dva sledeća broja. Na osnovu ove tablice izabrani su redni brojevi bolesnika koji će ući u ispitivanje.

Metoda lutrije je nešto jednostavnija za odabiranje slučajnog uzorka, ali ne ispunjava potpunu zakonitost slučajnosti. Za ovu metodu takođe se formira okvir skupa, a zatim se na karticama ispišu redni brojevi, nakon mešanja svih kartica, vrši se izvlačenje. Kartice sa izvučenim redni brojevima određuju koji će bolesnici ući u ispitivanje. Na ovaj način omogućava se da sve statističke jedinice imaju jednaku verovatnoću, odnosno jednaku šansu da postanu deo uzorka, čime se obezbeđuje reprezentativnost uzorka.

Sistematski uzorak se određuje po nekom unapred određenom sistemu. Formiranje sistematskog uzorka je višefazni proces. Prvo se sastavlja spisak statističkih jedinica (okvir skupa),zatim se određuje stopa ili korak. Stopa se određuje po formuli:
$R = N/n$, R= stopa (korak),
N= ukupan br.statističkih jedinica u osnovnom skupu, n=br.statističkih jedinica u uzorku
Nakon određivanja stope, vrši se određivanje slučajnog početka. Finalni korak je izvlačenje svake R-te jedinice i formiranje uzorka. Sistematski uzorak je posebno pogodan za korišćenje kada su statističke jedinice poređane po veličini.

Primena **stratifikovanog uzorka** je karakteristična za izrazito heterogene skupove. Formiranjem stratuma (=sloja) dobijaju se homogeni podskupovi iz kojih se zatim biraju jedinice koje obrazuju prost slučajan uzorak. Stratumi se određuju prema nekom određenom kriterijumu, time se postiže homogenost i između stratuma se pokazuju prave varijacije.
$ns = n\ Ns/N$
n – br. statističkih jedinica u uzorku stratuma ns – ukupan br.statističkih jedinica u uzorku Ns – ukupan br. statističkih jedinica u stratumu
N – ukupan br. jedinica svih strauma

Stratifikacija osnovnog skupa omogućava formiranje reprezentativnog uzorka, analizu svih delova skupa, smanjuje se slučajna greška izbora jer se izbor vrši iz homogenih skupova. Izazov prilikom obrazovanja stratifikovanog uzorka predstavlja odabir dovoljnog broja stratuma i sam način stratifikacije. Treba težiti da se skup podeli na što više homogenih stratuma, jer su na taj nacin i varijanse ovih podskupova male, samim tim mala je greška ocene aritmetičke sredine i drugih parametara osnovnog skupa. U svakodnevnom radu, stratifikacija se vrši u skladu sa konkretnim problemom uz prethodno proučavanje pojave i njene strukture“

Prednosti i nedostaci primjene slučajnih uzoraka

Tehnika	Prednost	Nedostatak
Slučajni uzorci		
Prost slučajan uzorak – PSU	Jednostavan za razumevanje, rezultati se mogu projektovati na celu populaciju	Zahtevan proces formiranja okvira uzorka, visoki troškovi, nema garancije da je reprezentativan
Sistematski uzorak	Jednostavniji za primenu od PSU, može da poveća reprezentativnost, nije uvek neophodno formirati okvir uzorka	Može da smanji reprezentativnost
Stratifikovan uzorak	Uključuje sve značajne podgrupe (stratume), precizan	Nekada je teško izabrati promenljivu stratifikacije, nije moguće sprovesti stratifikaciju po više promenljivih, visoki troškovi
Uzorak skupina	Jednostavan za primenu, ekonomski efikasan	Teško je računati i interpre-tirati rezultate, manje statistički efikasan u odnosu na stratifikovan uzorak

Izvor:http://www.medfak.ni.ac.rs/PREDAVANJA/2.%20STOMATOLOGIJA/STATISTIKA/6.%20predavanje.pdf

Namjerni (neslučajni) uzorci

- *Prigodni uzorak* - onaj koji se «nađe pri ruci» jer je drugi nedostupan- mnogo nedostataka
- *Uzorak tipičnih jedinica (Namjerni uzorak)* -zasniva se na prosuđivanju istraživača u pravcu cilja istraživanja
- *Kvotni uzorci -* neslučajni stratificirani uzorak. Upotrebljavaju se kod ad hoc organizovanih istraživanja za tržišne potrebe, za prikupljanje mišljenja građana o nekom problemu i slično. Istraživač unaprijed izabere broj ljudi (kvotu) svakog pojedinog stratuma koje mora intervjuisati, stoga se ti uzorci nazivaju kvotni.
- *Uzorak „snježnih grudvi*“-isključivo kada je riječ o ljudima kao ispitanicima; prvo se bira početni broj ispitanika koji će potom ukazati na nove ispitanike koje bi trebalo uključiti u uzorak

Izvor: http://www.ef.uns.ac.rs/Download/metodologija_nir/20_uzorkovanje.pdf

Nacrti namjernog uzorkovanja

ne omogućava određivanje preciznosti uzorka (a time ni tačnosti ocenjivanja)
efikasno se upotrebljavaju u eksplorativnim istraživanjima (čiji cilj nije precizno ocenjivanje parametara osnovnog skupa na osnovu reprezentativnog uzorka)
Prigodni uzorak (sačinjavaju ga raspoložive jedinice (lakoća dobijana); često ostaje nejasno iz kojeg osnovnog skupa uzorak potiče; problematična reprezentativnost (nije od pomoći ni povećanje uzorka); većina članova „ciljanog“ osnovnog skupa nema nikakve izglede da bude uključena u uzorak; ostaju nepoznati smer i veličina razlika između vrednosti koja je nađena ispitivanjem uzorka i vrednosti koja važi za celokupni osnovni skup; nema mogućnosti da se izračuna greška uzorka)
Uzorak tipičnih jedinica (Namerni uzorak) (zasniva se na prosuđivanju istraživača (koji na umu ima cilj istraživanja); precizniji nego prigodni uzorak)
Kvotni uzorak (najznačajniji način jasno definisan osnovni skup deli se na podskupove prema odabranim svojstvima; određuje se veličina svakog od tih podskupova; određuju se potrebna veličina uzorka i kvote (broj članova podskupa koje treba uključiti u uzorak); da bi se izbor članova za uzorak iz svakog podskupa prepušta slobodnom prosuđivanju i odlučivanju istraživača)
Uzorak „snežnih grudvi“ (isključivo kada je reč o ljudima kao ispitanicima; prvo se odabira početni broj ispitanika koji će potom ukazati na nove ispitanike koje bi trebalo uključiti u uzorak; prikladan za ispitivanje i ocenjivanje svojstava koja se u osnovnom skupu retko javljaju)

NAMJERNI NESLUČAJNI UZORAK

izabiru se jedinice prema odluci istraživača (anketara)
ispituju se dostupni članovi skupa
izabiru se jedinice u sklopu kvota

Prednosti i nedostaci primjene neslučajnih uzoraka

Tehnika	Prednost	Nedostatak
Namerni uzorci		
Prigodan uzorak	Najjeftiniji, najpogodniji, najkraće vreme sprovođenja, pogodan za eksplorativna istraživanja	Pristrasnost izbora, nije reprezentativan, nije preporučljiv za opisna i uzročna istraživanja
Uzorak tipičnih jedinica	Niski troškovi, prigodan, nije vremenski zahtevan. Idealan za eksplorativna istraživanja	Naglašena subjektivnost izbora, nema mogućnost generalizacije zaključivanja
Kvota uzorak	Uzorak je moguće kontrolisati prema određenim osobinama	Pristrasnost izbora, nema odlike reprezentativnosti (u statističkom smislu)
Uzorak „*grudve snega*“	Služi za ocenjivanje retkih karakteristika populacije	Veoma vremenski zahtevan

Izvor:http://www.medfak.ni.ac.rs/PREDAVANJA/2.%20STOMATOLOGIJA/STATISTIKA/6.%20predavanje.pdf

Neminovni pratilac statističkih istaživanja su i greške koje mogu biti slučajne i sistematske. Slučajne greške nemaju poseban uticaj na kvalitet podataka, dok sistematske uvek utiču na podatke

Teorijske i empirijske raspodjele obilježja

- Posmatrano obilježje predstavlja slučajnu promjenljivu, koja ima različite vjerovatnoće, pri čemu raspodjela vjerovatnoća može biti:
  - raspodjela vjerovatnoće obilježja na ukupnoj populaciji = teorijska raspodjela
  - raspodjela frekvencija obilježja u posmatranom uzorku (prikupljeni podaci) = empirijska raspodjela

- Kad je velika brojnost (***N)*** uzorka, onda se sa vjerovatnoćom bliskom jedinici empirijska raspodjela malo razlikuje od teorijske (teorema Glivenka)

- Posledica teoreme: za dovoljno veliko ***N*** se i parametri obje raspodjele (teorijske i empirijske) malo razlikuju među sobom, pa se parametri izračunati iz uzorka mogu iskoristiti za ocjenu parametara ukupne populacije.

Statističke tabele, histogrami i poligoni empirijske raspodjele frekvencija

- Statističke podatke koji se dobiju mjerenjem statističkog obilježja X , treba organizovati i prikazati u formi koja je pogodna za ispitivanje i zaključivanje:
- Statističke podatke prikazujemo:
  – u statističkim tabelama: u kojima se prikazuju sve vrijednosti ***xi*** obilježja uređene po veličini (x1<x2<...<xk) i pripadajći broj pojavljivanja te vrijednosti obilježja, odnosno **apsolutne frekvencija vrijednosti obilježja *fi*** , pri čemu je $\sum_{i=1}^{N} f_i = N$. Mogu se prikazati i **relativne frekvencije vrijednosti obilježja *fri=fi/N,*** pri čemu je $\sum_{i=1}^{N} f_{ri} = 1$

X	x_1	x_2	x_3	. . .	x_k	$\sum$
f_i	f_1	f_2	f_3	. . .	f_k	N
$\boldsymbol{fr_i}$	r_1	r_2	r_3	. . .	r_k	1

  – grafički:
    - histogram raspodjele frekvencije (u Dekartovom koordinatnom sistemu - na apscisnoj osi se daju vrijednosti obilježja $\boldsymbol{x_i}$ a nad njima se crtaju pravougaonici čija je visina jednaka frekvenciji f_k ili relativnoj frekvenciji f_{rk})
    - poligonom raspodjele frekvencija (izlomljena linija koja na ortogonalnom koordinatnom sistemu x_iOf_i spaja tačke (x_k, f_k)

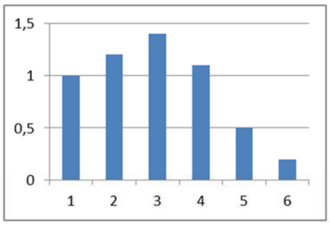
Diskretna promjenljiva uzima konačno ili prebrojivo mnogo vrijednosti xi, a s obzirom da se ovdje radi o prebrojivo mnogo mjerenja, onda su ove izmjerene vrijednosti slučajne promjenljive diskretnog tipa.

Apsolutna frekvencija fj – broj pojava izmjerene vrijednosti u jednom grupnom intervalu
Relativna frekvencija – broj pojava izmjerene vrijednosti u jednom grupnom intervalu /ukupan broj mjerenja n

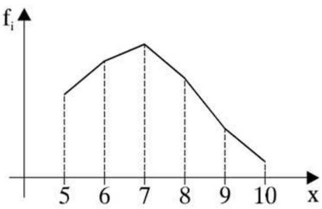
Obično se raspodjela vjerovatnoća osnovnog skupa naziva teorijskom raspodjelom, a raspodjela frekvancija u uzorku – empirijskom raspodjelom. Za dovoljno brojan uzorak (veličine n) sa vjerovatnoćom bliskom jedinici se empirijska raspodjela malo razlikuje od teorijske.

Statističke tabele, histogrami i poligoni empirijske raspodjele frekvencija

– Primjer histograma raspodjele frekvencije



– primjer poligona raspodjele apsolute frekvencije



Statističke tabele, histogrami i poligoni empirijske raspodjele frekvencija za klase podataka

- Ako je broj podataka (N) veliki kao i broj vrijednosti obilježja X, onda se varijacioni interval ($a=x_0, b=x_N$) dijeli na ***n*** klasa (grupnih intervala) i odgovarajuće frekvencije se posmatraju po klasama. Ovo naročito važi za kontinualne slučajne promjenljive X.
- podjela na grupne intervale (grupisanje izmjerenih vrijednosti u intervale sa nedvosmisleno određenim granicama)

N	n
40-60	6-8
60-100	7-10
100-200	8-12
200-500	12-17
> 500	21

  – broj grupnih intervala može se izabrati po nekom od sljedećih pravila:

  $n = \sqrt{N}$ ili $n = \sqrt[3]{N}$ ili u odnosu na broj N

  – širina grupnog intervala *j* je d_j

  – sredina grupnog intervala $x_j=1/2*(u_{j-1}+u_j)$, *gdje su* u_{j-1} *i* u_j granice intervala **ovo nije isto što i aritmetička sredina za interval:** ($\overline{x_j} = \frac{1}{f_j}\sum_{i=j}^{j+1} x_i \cdot f_i$)

  – apsolutna frekvencija f_j- broj pojava izmjerene vrijednosti u jednom grupnom intervalu

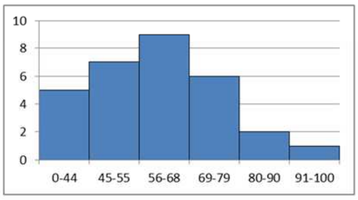
  – relativna frekvencija (statistička vjerovatnoća) $f_{rj}=f_j/N$

Statističke tabele, histogrami i poligoni empirijske raspodjele frekvencija za klase podataka

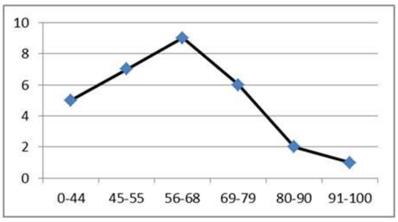
- primjer tabele empirijske raspodjele (raspodjele frekvencija) po intervalima

X	0-44	45-55	56-68	69-79	80-90	91-100	$\sum$
f_i	5	7	9	6	2	1	30
f_{r_i}	5/30	7/30	9/30	6/30	2/30	1/30	1

- primjer histograma raspodjele frekvencija



- primjer poligona raspodjele frekvencija



Pri malom broju posmatranja empirijska raspodjela frekvencija i odgovarajući poligon, odnosno histogram raspodjele frekvencija su najčešće vrlo skokoviti , jer pri posmatranju malg broja pojava preovlađuje djelovanje individualnih faktora i uslova.. Pravilnost u raspodjeli frekvencija bice uocljivija ako povecamo broj posmatranih jedinica, zbog toga sto se smanjuje uticaj faktora koji djeluju na pojedinacne osmatrane elemente, a povecava uticaj faktora koji koji djeluu na sve elemente. U tom slucaju ce se ove stepenaste i izlmljene krive sve vise primicati oblicima neprekidne krive, odnosno teorijske krive raspodjele

Kumulativna relativna ili apsolutna frekvencija

- **kumulativna apsolutna frekvencija** predstavlja se uređenim parovima

$$\left(x_j, \sum_{i=1}^{j} f_i\right)$$

- **kumulativna relativna frekvencija** predstavlja se uređenim parovima

$$F(x) = \left\{\left(x_j, \sum_{i=1}^{j} f_{ri}\right) \middle| \sum_{i=1}^{N} f_{ri} = 1,\ za\ \forall\ (i, j = 1 \dots N)\right\},$$

- Ova kumulativna relativnih frekvencija ima osobine funkcije raspodjele vjerovatnoća

$F(x)=0$, $x<x_0$ i $F(x)=1$, za $x>x_N$

Karakteristike empirijske raspodjele: mjere centra rasturanja

- **aritmetička sredina** ($\bar{x}$, **ili** x')
  - na osnovu potpunog skupa N (gdje su x_i uređeni po veličini i prikazane njihove apsolutne frekvencije f_i)

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \cdot f_i \quad , \quad \sum_{i=1}^{N} f_i = N$$

  - na osnovu grupisanih podataka u n klasa (grupa)

$$\bar{x} = \frac{1}{N}\sum_{j=1}^{n} x_j \cdot f_j$$, gdje je xj sredina klase j, i fj apsolutna frekvencija klase j

- **medijana** – vrijednost x_i za koju je kumulativna relativna frekvencija $F(x_i) = 0,5$
  - na osnovu potpunog skupa N, sa uređenim podacima x_i i pripadajućim apsolutnim frekvencijama f_i, medijana je ona vrijednost x_i koja zadovoljava nejednačine

$$\sum_{i=1}^{j} f_i \le \frac{N}{2} \le \sum_{i=1}^{j+1} f_i$$

  - na osnovu grupisanih podataka u n grupa (klasa), medijana Me zadovoljava nejednakost:

$$L = x_{k+1} - \frac{d}{2} \le \overline{M_e} < x_{k+1} + \frac{d}{2} = L + d, \ odnosno \qquad \overline{M_e} = L + d \cdot \frac{\frac{N}{2} - \sum_{i=1}^{k} f_i}{f_{k+1}} \ ,$$

  - gdje su:
    L=lijeva granica klase (k+1) u kojoj se nalazi medijana
    d=širina grupnog intervala,
    N- broj evidentiranih obilježja
    fi- apsolutna frekvencija

Kao što je (teorijska) raspodjela vjerovatnoća slučajne promjenljive okarakterisana srednjom vrijednošću μ i standardnim odstupanjem σ, empirijsku raspodjelu prvenstveno karakterišu aritmetička sredina x¯ i standardno odstupanje σ (a često se označava sa s)
fj – apsolutne frekvencije pojave izmjerenih vrijednosti unutar intervala j
xj – sredina grupnog intervala j

Geometrijski medijana je predstavljena tačkom na apscisnoj osi odakle polazi vertikalna linija x=Me koja dijeli histogram na dva dijela jednakih površina.

PRIMJER: knjiga Flasar, strana 41-44

Srednje vrijednosti uopšteno možemo podijeliti na dvije grupe:
računske srednje vrijednosti- na njih utiču sve elementarne jedinice; veća je od najmanje a manja od najveće elementarne jedinice, a ako su sve vrijednosti iste onda je i ona ta ista vrijednost
  aritmetička sredina,
  geometrijska sredina,
  harmonijska sredina,
  sredina kvadrata,
  sredina kubova
pozicione srednje vrijednosti- određuju se iz dela vrijednosti posmatranih obilježja prema svojoj poziciji u posmatranom skupu
  medijana
  modus

Karakteristike empirijske raspodjele: mjere centra rasturanja

- **moda-** vrijednost x_i za koju je apsolutna frekvencija najveća, odnosno koja se najviše puta javlja u uzorku (ne mora da postoji, a može ih biti i više):
  - na osnovu potpunog skupa *N:*

$$\overline{M_o} = x_j, tako\ da\ f_j = max$$

  - na osnovu grupisanih podataka u *n* grupa (klasa):

$$\overline{M_o} = L + d \cdot \frac{\Delta_1}{\Delta_1 + \Delta_2}_j$$

  - gdje su:
    L=lijeva granica klase (k+1) u kojoj se nalazi moda
    d=širina grupnog intervala,
    Δ1- razlika modalne frekvencije i susjedne prethodne
    Δ2- razlika modalne frekvencije i susjedne naredne

Karakteristike empirijske raspodjele: mjere odstupanja od centra rasturanja (disperzija, standardna devijacija, koeficijent varijacije)

- **raspon=varijacioni interval** – najjednostavnija mjera rasipanja

$$\bar{R} = xmax - xmin$$

- **srednje apsolutno odstupanje,** na osnovu vrijednosti x_i, matematičkog očekivanja $\bar{x}$, i apsolutnih frekvencija f_i

$$\bar{\theta} = \frac{1}{N}\sum_{i=1}^{n}|x_i - \bar{x}| \cdot f_i \quad , \qquad \sum_{i=1}^{n} f_i = N$$

- **disperzija (varijansa)**- srednji kvadrat odstupanja vrijednosti x_i od aritmetičke sredine $\bar{x}$, za poznate apsolutne frekvencije

$$s^2 = \frac{1}{N}\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot f_i = \frac{1}{N}\sum_{i=1}^{n} x_i^2 f_i - \bar{x}^2$$

Karakteristike empirijske raspodjele: : mjere odstupanja od centra rasturanja (disperzija, standardna devijacija, koeficijent varijacije)

- **standardno odstupanje**

$$s = \sigma = +\sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i \cdot (x_i - \bar{x})^2} \quad \text{odnosno,}$$

$$s = \sigma = +\sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i \cdot x_i^2 - \bar{x}^2}$$

- **koeficijenti varijacije** odnos standardnog odstupanja i aritmetičke sredine

$$k_v = \frac{\sigma}{\bar{x}}$$

Literatura

- Vukadinović, S.: Elementi teorije verovatnoće i matematičke statistike, Privredni pergled, Beograd 1986
- Vukadinović, S.: Zbirka rešenih zadataka iz teorije verovatnoće, Privredni pergled, Beograd, 1983
- Čuljak, V: Vjerojatnost i statistika, Građevinski fakultet, Sveučilište u Zagrebu, 2011, http://www.grad.hr/vera/webnastava/vjerojatnostistatistika/vis-pdf.pdf
- Prof. dr Dušan Joksimović: POSLOVNA STATISTIKA, Megatrend univerzitet primenjenih nauka, Beograd, 2006.
- Pivac, S.: Statističke metode (predavanja, diplomski studij, kolegij "Statističke metode") e-nastavni materijal, Split, 2010.
- http://www.ef.uns.ac.rs/Download/metodologija_nir/20_uzorkovanje.pdf
- http://www.medfak.ni.ac.rs/PREDAVANJA/2.%20STOMATOLOGIJA/STATISTIKA/6.%20predavanje.pdf