Ispitna pitanja:

1. Značenje pojma statistike, oblast primene, značaj, podela, razvoj
2. Statističke zakonitosti
3. Populacije (skup) i uzorci
4. Vrste uzoraka (sve)
5. Statističke serije
6. Statističke tabele
7. Stardžesovo pravilo
8. Grafičko predstavljanje statističkih serija
9. Polarni dijagram
10. Pareto dijagram
11. Mere koncentracije (sve)
12. Mere disperzije (sve)
13. Oblik statističkog skupa
14. Statistička verovatnoća
15. Zakon verovatnoće i funkcija raspodele verovatnoće
16. Osnovni modeli teorijskih raspodela verovatnoće (sve)
17. Testiranje hipoteze o jednakosti empir. i teor. rasporeda (sve)
18. Statističke ocene - tačkasta ocena
19. Interval poverenja
20. Statističke hipoteze, donošenje odluke o hipotezi
21. Statističko upravljanje procesima, ocena sposobnosti procesa
22. Vrste kontrolnih karti
23. Pojam, osobine, vrste i komponente vremenskih serija
24. Modeli vremenskih serija
25. Trend, pojam, osobine, linearni trenad
26. Sezonska kolebanja

SADRŽAJ PROGRAMA

1. Uvod (pojam, predmet, razvoj, podela, značaj, oblasti primene)
2. Populacije i uzorci (statističko zaključivanje o populaciji; greška uzorka)
3. Statističke serije (nastanak i vrste; serije strukture; vremenske serije; grafičko predstavljanje)
4. Mere populacije i uzorka (mere centralne tendencije, varijacije i oblika raspodele)
5. Teorijske raspodele verovatnoće (pojam, osnovni modeli, podudarnost)
6. Teorijske osnove statističkog zaključivanja o parametrima populacije, tj. skupa (statistike uzorka i njihove raspodele; statistička ocena; interval poverenja; statističke hipoteze)
7. Statističko zaključivanje o parametrima populacije (skupa) na osnovu velikog uzorka (zaključivanje o sredini i varijansi)
8. Statistička kontrola procesa (statistička stabilnost procesa; kontrolne karte)
9. Analiza vremenskih serija (statistički model vremenskih serija; metode usklađivanja; ocenjivanje trenda i sezonskih kolebanja)
10. Savremeni statistički programi (odnos statistike i informatike; statistički softveri)

LITERATURA

1. Шолак, Њ., *Статистика у економији и менаџменту,* Младеновац, 1999.
2. Јоветић, С., *Статистика са апликацијом у ЕХЦЕЛ-у,* Крагујевац, 2004.
3. Марковић, М., Петковић, С., *Пословна статистика*, Виша пословна школа, Београд, 2000.

1. UVOD

Reč statistika potiče od latinske reci *status* ili stanje i italijanske reči *stato* ili država. Reč statistika je prvi put upotrebljena u 18.veku u radovima Gottfrida Achenwala iz Getingena, jednog od predstavnika Nemačke deskriptivne škole.

Statistika obuhvata:

- statistiku u užem smislu (deskriptivnu statistiku)
- statističku analizu i
- statističku teoriju

Jedna od brojnih definicija je: **statistika je grana primenjene matematike, čiji principi proisticu iz računa verovatnoće.**
Za druge - to je drustvena nauka koja upoređuje činjenice izražene u brojevima i izvodi opšte zaključke, utvrđuje uzroke i posledice.

DEFINICIJA U ŠIREM SMISLU: Pojam statistika obuhvata statističku teoriju i metodologiju, kao i statistička istraživanja masovnih varijabilnih pojava.
Statistička teorija daje osnove statističkoj metodologiji kao naučnoj proceduri kvantitativnog istraživanja masovnih pojava.
Parcijalna statistčika istraživanja su sva pojedinačna istraživanja masovnih pojava. ali po pravilima koja postavlja statistička metodologija.

DEFINICIJA U UŽEM SMISLU: Pod pojmom statistika podrazumeva se skup metoda sa specifičnim postupkom kvantitativne deskripcije, analize i zaključivanja o masovnim pojavama.

ZADACI STATISTIČKE METODE:

- otkrivanje bitnih karakteristika masovnih pojava
- utvrđivanje činjenica koje uslovljavaju postojeće stanje pojave
- otkrivanje povezanosti karakteristika masovnih pojava
- otkrivanje i objašnjavanje zakonitosti u pojavama
- predviđanje daljeg razvoja pojave

KARAKTERISTIKE STATISTIKE KAO NAUČNE METODE:

- kao metoda pripada OPŠTOJ naučnoj metodologiji
- kvantitativna metoda istraživanja masovnih pojava
- po prirodi je induktivna naučna metoda
- ispituje oblike pojava u masi slučajeva
- doprinosi proveri teorijskih stavova i objašnjenja
- podiže egzaktnost istraživanja primenom matematičkih metoda
- čini kvantitativnu analizu potpunijom
- ne isključuje dedukciju i druge logičko-metododloške postupke

PREDNOSTI STATISTIČKE METODE:

- neutralna je prema stvarnosti
- stroga u oblicima i postupcima
- prikladna za masovna ispitivanja
- raznovrsna po oblicima i postupcima
- primenom matematičkih izraza obezbeđuje tačnost podataka
- precizna u opisivanju pojava i njihovih manifestacija

NEDOSTACI STATISTIČKE METODE:

- nije samostalna, jer zavisi od naučne discipline u kojoj se primenjuje
- ne može se primeniti na sve oblasti i fenomene
- saznajna snaga joj je ograničena
- dobijeni rezultati ne govore o individualnostima u pojavama
- ne otkriva, ne objašnjava svu kompleksnost nastanka, menjanja i razvoja pojave

---

RAZVOJ STATISTIKE

Izdvajaju se 4 etape:

1. prikupljanje podataka za državne potrebe – u Starom veku, pre 4000 do 3000 godine pre Nove ere u starim civilizacijama (Egipat, Grčka, Rim i Kina) prikupljani su podaci o stanovništvu, posebno muškarcima, o zemlji, za potrebe sagledavanja vojne i ekonomske moći države; poznat je rimski *cenzus*, koji se obavlja svake pete godine i značio je popis stanovništa i njihove imovine; u Srednjem veku – popis imovine, poreskih obveznika, stoke, obradivih površina; ova etapa traje do 17.veka; na sakupljenim podacima su vršene najjednostavnije operacije
2. Nemačka deskriptivna statistika ("državopis") i Škola političke aritmetike u Engleskoj – u 17.veku; cilj prve škole (Hermann Conring, pa Gottfried Achenwall) je bio sistematizovanje podataka za vođenje državnih poslova; cilj druge (John Graunt i Willam Peti) je bio naučno saznanje pojava na osnovu utvrđivanja zakonitosti u njima
3. povezivanje statistike i teorije verovatnoće – od 19.veka; zahtev za primenom teorije verovatnoće potiče od belgijskog astronoma i statističara Adolphea Queteleta, koji se smatra osnivačem statistike; statistika je opšta metoda istraživanja masovnih pojava koja se može primeniti u svim naukama; razvoju su doprineli C.Gauss, Sir Francis Galton, R.A.Fiscer, W.S.Gosset, Čebišev, Spierman, M.G.Kendall, Kolmogorov, J.Neyman...
4. razvoj savremene statistike – druga polovina 20.veka; razvoj informacione tehnologije; savremena statistika posebno posvećuje pažnju teoriji uzoraka, savremenim statističkim postupcima, statističkoj analizi, multivarijantnim statističkim postupcima i obradi podataka računarima.

PODELA STATISTIKE

Moguća je podela na osnovu:

- *cilja obrade podataka*
  - evidenciona statistika (prikupljanje, registrovanje i sređivanje podataka)
  - statistika kao naučna metoda (pored prikupljanja, registrovanja i sređivanja podataka obuhvata njihovu analizu i tumačenje)

- *namene i vrste primenjenih metoda*
  - teorijska (opšta) statistika (oblast primenjene matematike; tazvija opšte statističke metode i postupke, objašnjava ih, dokazuje i usavršava; zasniva se na teoriji verovatnoće; čine je 4 oblasti: teorija raspodele, teorija statističkih ocena, teorija testova i teorija povezanosti; posebne grane su: teorija programiranja, teorija diskriminacije itd.)
  - primenjena statistika (posebne statistike; pored opštih postupaka koriste se specifični postupci, prema određenoj naučnoj disciplini; opšta primenjena statistika daje sintezu metodolološkog i iskustvenog u primeni statistike u raznim oblastima naučnog istraživanja; to su pedagoška statistika, psihološka, zdravstvena, demografska, poslovna..)

- *prema ulozi verovatnoće, uzorka i osnovnog skupa u prikupljanju i analizi podataka*
  - deskriptivna statistika (bavi se sređivanjem, klasifikacijom statistisčkih podataka, tabeliranjem, grafičkim prikazivanjem; zove se i statistika u užem smislu)
  - induktivna statistika (ili matematička statistika; bavi se zaključivanjem o zakonitostima i međusobnim odnosima u skupu na osnovu numeričkih podataka dobijenih posmatranjem i merenjem na uzorku)
  - multivarijantna statistika (analizira odnose najmanje tri slučajne promenljive uz primenu linearne algebre; ne sadrži pojmove verovatnoća, uzorak i skup; S.Fajgelj; obuhvata: faktorsku analizu, taksonomijsku analizu, regresionu analizu, multivarijantnu analizu varijanse, diskriminacionu analizu, multivarijantnu analizu kovarijanse itd.)

- *raspodele podataka koji su predmet analize*
  - parametarska statistika (statističke postupke zasniva na modelu normalne krive – Gausove krive verovatnoće; odnosno, osnovna pretpostavka je da podaci o pojavi koja se proučava ne odstupaju značajno od normalne raspodele; postupci parametarske statistike su: t-test, regresiona analiza, analiza varijanse itd.)
  - neparametarska statistika (koristi se kada je nemoguće utvrditi da li je raspodela normalna, kada su podaci izraženi u frekvencijama modaliteta statističkih obeležja, kada se radi o malim uzorcima – do 30 jedinica; postupci ove statistike su hi-kvadrat test, medijana test, test predznaka, test ekvivalentnih parova, test sume rangova, Mek Nemarov test, Kolmogorov-Smirnovljev test itd.)

STATISTIČKA ZAKONITOST

Statistička zakonitost je zakonitost otkrivena u mnoštvu slučajeva. Ona je empirijskog karaktera. Utvrđuje se sa osnovu konkretnih rezultata istraživanja. Njena podloga je zakon velikih brojeva koji se empirijski potvrđuje u masovnim pojavama. Statistički zaključak o utvrđenoj zakonitosti uvek ima karakter verovatnoće. Zakonitost otkrivena u statističkom istraživanju uopštava se na osnovama opštih zakonitosti i statističke teorije.

Masovne pojave mogu biti sa pretežno stabilnim i pretežno dinamičkim odnosima, pa se prema tome statističke zakonitosti dele na:
- stabilne (stacionarne)
- nestabilne ili dinamičke (nestacionarne)

Stacionarnu statističku zakonitost imaju pojave kod kojih se varijacije javljaju pod dejstvom slučajnif faktora. Tada se individualna odstupanja od zakonitosti (sredine) javljaju iznad ili ispod (levo ili desno), tako da se u masi izjednačavaju. Pojedinačno se izražava kao slučajno u odnosu na opštu zakonitost masovne pojave.

Nestacionarnu statističku zakonitost imaju pojave kod kojih se tokom vremena najveši deo varijacija sistematski koncentriše izpod ili iznad (levo ili desno) od srednje (tipične) vrednosti pojave. Time se ispoljava dinamička tendencija u promeni statističkog skupa i, statističke zakonitosti. Ovde se pojedinačno ispoljava kao sistematsko, a statistička zakonitost kao zakonitost razvoja pojave u vremenu.

Kada se statistički istražuje slaganje varijacija više masovnih pojava, statistička zkaonitost se ispoljava kao zakonitost uzajamnog slaganja i funkcionalnog odnosa više pojava (korelacija i regresija).

Zaključujemo da su oblici statističke zakonitosti masovnih pojava sledeći:
- zakonitost raspodele (rasporeda) jedinica unutar skupa (srednja vrednost i mera varijacije oko nje)
- zakonitost promene strukture skupa u vremenu
- zakonitost uzajamnog slaganja i funkcionalnog odnosa više obeležja skupa
- zakonitost razvitka (dinamike) obima pojave

---

2. POPULACIJE I UZORCI (skup i uzorak)

Predmet statističkih istraživanja su masovne varijable pojave, koje se mogu izaziti brojčano i koje se ispoljavaju na individualnim slučajevima.

Skup individualnih slučajeva na kojima se ispoljava masovna pojava je statistički skup (ili statistička masa, ili generalna polulacija, ili osnovni skup, ili jednostavno masa,, ili *populacija*, ili skup). Tako se populacije i statistički skup poklapaju. Zavisno od prirode pojave, skup mogu činiti bića, stvari i događaji.

*Masovnos*t znači da se statistička zakonitost ispoljava u masi pojedinačnih slučajeva (zakon velikih brojeva), gde oni odstupaju naviše ili naniže od utvrđenih zakonitosti (kao rezultat dejstva slučajnih faktora), ali se to odstupanje u masi pojedinačnih slučajeva potire.

Karakteristike skupa:
- **relativna homogenost** – znači da su jedinice koje čine skup istovrsne, slične najmanje po jednom obeležju
- **celovitost** - znači da skup obuhvata sve statističke jedinice na određenom prostorno i u određenom vremenu
- **relativna diferenciranost** - statističke jedinice su istovrsne, ali nisu istovetne, tj. imaju najmanje jedno zajedničko obeležje, a po ostalima se razlikuju; ta obeležja su predmet naše pažnje

Individualni slučajevi na kojima se ispoljava pojava su statističke jedinice (ili elementarne jedinice, ili jedinice posmatranja, ili jednostavno elementi, ili jedinice). One čine konstututivni element skupa, nosioci su pojave i podležu korigovanju u statističkom istraživanju

Prema obimu, statistički skup je:
- konačan (obuhvata ograničen broj jedinica i odnosi se na jedan vremenski trenutak)
- beskonačan (ili neograničen; predstavlja kontinualnu statističku masu, pa je istraživanjem moguće obuhvatiti samo jedan njegov deo)

Po postojanosti, statistički skup je:
- hipotetičan (ili zamišljen; čine ga jedinice koje se definišu nekim pravilom, odnosno modelom)
- stvaran (ili realan; kod njega statističke jedinice postoje u prostoru i vremenu)

Po obuhvatnosti, ststistički skup je:
- osnovni skup (skup statističkih jedinica koji uzima u obzir sve posmatrane jedinice neke pojave; ovaj skup može biti ciljni osnovni skup – onaj na koji istraživač želi da generalizuje svoje zaključke i onaj iz koga je uzorak formiran)
- parcijalni skup (skup koji obuhvata samo deo jedinica neke pojave)

Svaki statistički skup se definiše:
- pojmovno (određuje se obeležje po kome se utvrđuje pripadnost jedinica skupu; može biti *opisno* i *brojčano*)
- prostorno (određuje se prostor sa koga statističke jedinice pripadaju datom skupu; to može biti *uže* i *šire* geografsko područje)
- vremenski (određuje se pripadnost statističkih jedinica nekom vremenskom intervalu ili određenom vremenskom momentu; zato statistički skup može biti *statistički skup stanja* i *statistički skup kretanja*)

STATISTIČKO OBELEŽJE je kvantitativno i kvalitativno svojstvo masovnih pojava, koje se proučavaju. Skup kvalitativnih obeležja pojave pojedinačno registrovanih na svakoj jedinici statističkog skupa čini kvalitativnu (atributivnu) polulaciju. Skup kvantativnih obeležja pojave pojedinačno registrovanih na svakoj jedinici statističkog skupa čini kvantativnu polulaciju.

Osnovni tipovi obeležja:

- numeričko - kada je statističko obeležje dato brojem; numeričko obeležje može biti:
  - prekidno - izražava se celim brojem (npr. broj dece : 0,1, 2...)
  - neprekidno - ovo obeležje se izražava celim brojem ili brojem sa decimalom (npr. visina stanovnika izražena u cm ili mm)
- atributivno - kada je statističko obeležje dato na opisan način, rečima.

S obzirom na činjenicu da se statističko istraživanje teško vrši na celokupnom skupu, iz skupa se na *slučajan način* biraju elementarne jedinice i na njima se vrši istraživanje. Taj izabrani deo skupa je **UZORAK**. Analizom jedinica skupa koje su izabrane u uzorak, dobijaju se parametri kojima se vrši ocenjivanje nepoznatih karakteristika skupa. Postupak statističkog ocenjivanja je postupak statističkog induktivnog zaključivanja: rezultati analize uzorka se pripisuju statističkom skupu. Danas je statističko ocenjivanje skupa na osnovu uzorka postalo vladajući metod statističke analize masovnih pojava.

Izvlačenje elementarnih jedinica u uzorak može biti:

- sa ponavljanjem - znači da se elementarne jedinice posle izvlačenja vraćaju natrag u skup i ravnopravno učestvuju dalje u izvlačenju, čime se tokom celog izvlačenja obezbeđuje ista verovatnoću izbora elementarnih jedinica skupa u uzorak
- bez ponavljanja - znači da se elementarne jedinice posle izvlačenja ne vraćaju natrag u skup, čime se dovodi do promene verovatnoće izbora.

Razni načini formiranja uzroka imaju za cilj da uzorak bude reprezentativni predstavnik skupa iz koga je izdvojen, a da bi to bio, mora da ispuni dva uslova:

- da verno odražava karakteristike skupa
- da je dovoljno brojan (reprezentuje ga utoliko bolje ukoliko je veći - zakon velikih brojeva)

Veličina uzorka zavisi od:

- veličine skupa
- varijabilnosti obeležja
- željene tačnosti rezultata statističke analize

Slučajan izbor elementarinih jedinica u uzorak vrši se korišćenjem Tablice slučajnih brojeva. Ona je sastavljena tako da se u dugom nizu brojeva svaka cifra od 0 do 9 javlja sa verovatnoćom 1/10 parne, odnosno neparne cifre sa verovatnoćom 1/2 itd. Pre korišćenja Tablice, mora se napraviti okvir identifikacije statističkog skupa (npr. spisak radnika preduzeća). Brojevi se mogu uzimati iz bilo kog dela Tablice, s tim da se primenjeni postupak izbora brojeva ne sme menjati do konačnog izbora elementarnih jedinica koje sačinjavaju uzrorak. Izvučeni brojevi se identifikuju sa rednim brojevima iz okvira identifikacije skupa, čime je završen izbor elementarnih jedinica uzorka.

Postoji više načina slučajnog izbora elementarnih jedinica skupa u uzorak i svako od njih opredeljuje vrstu uzorka.

VRSTE UZORAKA:

- **PROST SLUČAJAN UZORAK** - bira se iz relativno homogenog statističkog skupa i to je onaj uzorak kod koga svaka elementarna jedinica ima podjednaku ili unapred poznatu relativnu verovatnoću da će biti izabrana u uzorak. Jednaka verovatnoća izbora elementarnih jedinica skupa u uzorak obezbeđuje se korišćenjem tablica slučajnih brojeva. Ta jednaka verovatnoća izbora obezbeđuje reprezentativnost i objektivnost uzorka.

Uzorak se može birati iz konačnog ili beskonačnog statističkog skupa. Ukoliko se bira iz konačnog, svaka elementarna jedinica mora imati podjednaku verovatnoću da će biti izabrana u uzorak (npr. loto).

Karakteristike ovog uzorka:

- reprezenatativan (svaka elementarna jedinica ima onu osobinu zbog koje se vrši statističko istraživanje)
- elementarne jedinice se biraju slučajnim izborom
- relativno homogen

Prost slučajan uzorak može biti;

- veliki (prost slučajan uzorak koji sadrži >= 30 elementarnih jedinica - kada se govori o prostom slučajnom uzorku uvek se misli na veliki uzorak)
- mali (sadrži < 30 elementarnih jedinica - kada se govori o malom prostom slučajnom uzorku, to se uvek naglašava)

Postoje dve vrste prostog slucajnog uzorka:

- bez ponavljanja (znači da jednom izvučena jedinica iz osnovnog skupa ne učestvuje u daljem izvlačenju; svaka jedinica nema istu verovatnoću izvlačenja; uzastopni izbori jedinica u uzorak su međusobno statistički zavisni)
- sa ponavljanjem (znači vraćanje izbaranih jedinica u osnovni skup; svaka jedinica ima istu verovatnoću izbora; uzastopni izbori jedinica u uzorak su međusobno nezavisni)

U praksi se najviše koristi prost slučajan uzorak bez pobavljanja, jer su po pravilu osnovni skupovi veliki i nejednakost verovatnoća izbora jedinica je zanemarijiva. Ovi uzorci su efikasniji.

Prost slučajni uzorak je kod relativno nehomogenih skupova neefikasan. Zato: 1) što nehomogen skup nema normalnu raspodelu (što je pretpostavka prostog slučajnog uzorka); 2) standardna greška ocene sredine i proporcije kao i sama sredina i proporcija izračunate pomoću prostog slučajnog uzorka izvučenog iz nehomogenog skupa mogu imati nerealnu vrednost zbog izraženije zastupljenosti jedinica izvučenih iz ovog ili onog dela nehomogenog skupa.

- **STRATIFIKOVAN UZORAK** - u slučaju nehomogenog skupa, a uz prihvatanje pretpostavke o normalnoj raspodeli da bi se dobile realne vrednosti parametara, radi se stratifikacija (podela) skupa na relativno homogene stratume, pri čemu jedna jedinica skupa može pripadati samo jednom stratumu. Veća homogenost stratuma dovodi do veće varijacije među stratumima, a manje unutar stratuma. Uzorak primenjen na stratifikovani skup je *stratifikovani uzorak.*

Veličina stratifikovanog uzorka se određuje:
- srazmerno veličini pojedinih stratuma (*metod proporcionalnog izbora* - znači da je učešće elementarnih jedinica iz stratuma u stratifikovanom uzroku proporcionalno učešću stratuma u statističkom skupu)
- pribegava se optimalnom izboru (*metod optimalnog izbora* - polazi od kriterijuma da se dobije minimalna standardna greška)

Stratifikovani uzorak može biti:
- **jednoetapni** (skup se deli na stratume iz iz svakog stratuma se biraju jedinice u uzorak)
- **dvoetapni** (prvo se odrede stratumi, a u okviru svakog stratuma pdstratumi; u prvoj fazi iz svakog stratuma se slučajno bira određeni broj podstratuma; u drugoj fazi se u okviru slučajno izabranih podstratuma biraju, pod jednakom verovatnoćom izbora, jedinice u uzorak)

- KONTROLISANI UZORAK

- PROCESNI UZORAK

- SISTEMSKI UZORAK

To je uzorak kod koga se jedinice iz skupa biraju po nekom sistematskom redu. Najčešće je to izbor sa slučajno izabranom prvom jedinicom, pa se ostale jedinice biraju uvek sa istim korakom biranja. On se odreduje iz odnosa K=N/n. Tako je moguće, posle slučajno izabrane jedinice, primeniti korak da se bira jedanaesta, dvadesetprva, tridesetprva jedinica i slično. Razlog je ograničen broj mogućih uzoraka iz skupa i činjenica da se slučajno bira samo prva jedinica.

U praksi, statističko zaključivanje na osnovu sistematskog uzorka izvodi se po proceduri koja važi za jednostavan uzorak, ili se kombinuju dva sistematska uzorka iz jednog skupa, sa dva slučajna početka, pa se iz podataka za oba uzorka izračunavaju parametri pomoću kojih se zaključuje o parametrima skupa.

- VIŠEETAPNI UZORAK

To je uzorak kod koga se jedinica skupa bira u više etapa. U I etapi se biraju jedinice prve etape koje predstavljaju skupinu jedinica; iz te izabrane skupine jedinica u I etapi biraju se jedinice nižeg ranga. One su jedinice izbora II etape. Postupak se nastavlja sve dok se kao jedinica izbora ne pojavi osnovna jedinica skupa.

Postoje jednoetapni, dvoetapni... Višeetapni izbor se objašnjava preko uslovnih verovatnoća složenih zavisnih događaja. Pouzdanost statist. zaključivanja smanjuje se sa povećanjem etapa izbora.

Smanjena preciznost ne znači malu primenu višeetapnog uzorka. Naprotiv, široku primenu ima zbog minimalnih troškova sprovođenja.

- VIŠEFAZNI UZORAK

Koristi se kada se u okviru glavnog uzorka istovremeno formiraju poduzorci. U glavnom uzorku sve jedinice daju određene osnovne informacije, a od jedinica poduzorka traže samo dopunske informacije. Statističko zaključivanje se vrši na nivou glavnog uzorka, a i po poduzorcima. Pri tome se kod poduzoraka koriste sopstvene informacije, kao i osnovne informacije glavnog uzorka.

- PANEL UZORAK

To je uzorak kod koga se uključuje vremenska dimenzija u praćenju određenih obeležja. Tako se otkriva tendencija razvoja pojave u vremenu. Naročito se koristi u istraživanju motiva, ponašanja i mišljenja (da bi se videla njihova stabilnost i šta izaziva njihove promene). Da bi statističko zaključivanje bilo pouzdano, treba u dužem vremenskom periodu kontinuirano pratiti razvoj pojave na slučajno izabranom reprezentativnom uzroku, mada je reprezenattivnost teško obezbediti.

- KVOTNI UZORAK

Razlikuje se od slučajnog, jer se izbor jedinica u uzorak prepušta licima koja sprovode istraživanje. Ovakav izbor često dovodi do nereprezenativnosti uzorka. Fizionomija uzorka je određena unapred postavljenim kvotama. Znači, kvorni uzorak je stratifikovani uzorak sa neslučajnim izborom jedinica. Zato ima niz slabosti. Nepouzdan je i pristrasan, pa se ograničeno koristi. Koristi se zbog jednostavnosti i brzine sprovodenja. Obično se koristi za iztraživanje javnog mnjenja i promena na tržištu robe široke potrošnje. A baš za ove oblasti teško je primeniti slučajni uzorak, zbog nepostojanja okvira skupa.

---

3. STATISTIČKE SERIJE

STATISTIČKI PODACI

Postupkom merenja, evidentiranja ili klasifikovanja na statističkim jedinicama se pojedinačno registruju utvrđeni oblici obeležja izraženi brojčano. To su podaci. Pošto su obeležja statističkih jedinica promenljiva, statistički podaci izražavaju promenljivost, pa se kaže da je osnovna karakteristika statističkih podataka varijabilnost. Da bi se neki podaci smatrali statističkim, treba da su:

- prikupljeni prema planu posmatranja ili merenja
- da su varijabilni
- da ih ima dovoljno

Prema *poreklu*, podaci su:
- empirijski (dobijeni posmatranjem pojava koje se stvarno dešavaju)
- teorijski (dobijaju se kao rezultat teorijske analize na osnovu očekivanja po nekoj pretpostavci)

Prema *obeležjima koja izražavaju svojstva statističkih jedinica*, podaci su:
- numerički (vrednosti numeričkih obeležja jedinica skupa)
- atributivni (obeležja iskazana opisno rečima)

Prema *izvoru*, podaci su:
- primarni (prikupljeni u skladu sa ciljem istraživanja i planom prikupljanja)
- sekundarni (podaci koji prikupljeni u nekom istraživanju služe za potrebe drugog istraživanja)

Prema *načinu prezentovanja*, podaci su:
- verbalni
- numerički
- tekstualni
- kombinovani
- grafički

PRIKUPLJANJE STATISTIČKIH PODATAKA

Prema *obimu podataka*, prikupljanje je:
- iscrpno (obuhvaćene sve statističke jedinice)
- reprezentativno (obuhvaćene jedinice sa uzorka istraživanja)

S obzorom na *vreme*, prikupljanje je:
- jednokratno (sprovodi se u datom momentu)
- periodično (ponavlja se u određenim vremenskim periodima)
- tekuće (sprovodi se u određenom vremenskom intervalu)

PROVERAVANJE STATISTIČKIH PODATAKA

Proveravanjem se utvrđuje:
- potpunost podataka
- tačnost podataka

S obzirom na karakter, *greške* u podacima mogu biti:
- sistemske (javljaju se na svakom podatku, kao posledica neprecizno definisanih obeležja statiostičkih jedinica ili nestručno primenjenih instrumenata posmatranja i merenja; ove greške utiču na rezultat, pa se moraju otkloniti)
- slučajne (ne javljaju se na svakom podatku i nisu konstantne veličine; nemaju naročit uticaj na rezultat istraživanja, jer se međusobno briši)

STATISTIČKE SERIJE

Statističke serije su rezultat sređivanja, grupisanja statističkih podataka prema određenim kriterijumima u unapred utvrđene grupe. Grupisanje podataka može biti:

- jednostavno (po jednom obeležju)
- kombinovano (po dva ili više obeležja)

Statističke serije mogu biti:

- **SERIJE STRUKTURE** (statistički skup grupisan prema numeričkim obeležjima; mogu biti ***numeričke serije*** – ili tzv. raspodele frekvencija ili distribucije frekvencija i mogu biti u obliku:

+ osnovne serije podataka (nepregledne; npr: X: 1,6,2,3,8,3,2,5)
+ neintervalne raspodele frekvencija (preglednija od osnovne serije)

X	*f*
1	8
2	5
3	3
	16

+ intervalne raspodele frekvencija (Stardžesovo pravilo za određivanje broja i širine intervala)

X	*f*
1-4	8
5-8	5
9-12	3
i=4	16

+ vremenske serije podataka (mogu biti: *momentne* – formiraju se od podataka koji se dobijaju posmatranjem statističkih obeležja u određenim trenucima vremena i *intervalne* – to je niz veličina koje jedno obeležje ima u različitim intervalima vremena)

ili ***atributivne serije*** – ili opisne serije; poredak atributivnih obeležja može biti prema abecednom redu, veličini frekvencija, srodnosti obeležja itd.

status studenta	***f***
budžet	8
sufinans.	5
samofinans.	3
ukupno:	16

- VREMENSKE SERIJE (ili serije dinamike; grupisanje s obzirom na vremenske momente ili vremenske intervale)

- PROSTORNE (geografske) SERIJE (formiraju se prema obeležjima prostora: mestima, teritorijalnim područjima (okrug, opština), područjima karakterističnim s aspekta cilja istraživanja)

FREKVENCIJOM se naziva učestalost pojavljivanja modaliteta (oblika) jednog obeležja u ukupnom broju jedinica posmatranja.

Ako se frekvencija modaliteta obeležja iskazuje u procentima, radi se o ***relativnoj frekvenciji.*** Ona se dobija kada se u odnos stavi broj jedinica koji pripadaju jednom modalitetu obeležja prema ukupnom broju jedinica.

Kumulacija frekvencija se vrši tako što se frekvencija prve klase obeležja prepiše, a zatim se sukcesivno frekvenciji svake naredne klase dodaju frekvencije prethodnih klasa. Poslednja klasa obeležja sadrži sumu svih frekvencija.

Kumulacija relativnih frekvencija se vrši tako što se relativna frekvencija prve klase prepiše, a zatim se sukcesivno relativnoj frekvenciji svake naredne klase dodaju relativne frekvencije prethodnih klasa. Poslednja klasa obeležja sadrži sumu svih relativnih frekvencija.

stepen stručnog obrazovanja	broj zaposlenih (f)	kumulativ f	relatina f (%)	kumulativ relativne f
visoko	362	362	22,80	22,80
više	832	(362+832)1194	52,39	75,19
srednje	184	(1194+184)1378	11,59	86,78
niže	210	(1378+210)1588	13,22	100,00
ukupno:	1588	*	100,00	*

STATISTIČKE TABELE

Statistička tabela je osnovna forma tabelarnog prikazivanja statističkih podataka. Sastoji se iz niza polja, koja se dobijaju kada se povuku horizontalne i vertikalne linije. Horizontalna polja čine **redove,** a vertiklna čine **kolone.** Svaka stat. tabela ima više redova i više kolona. Prvi red je *zaglavlj*e tabele, a prva kolona je *predkolona*. Zadnji red je *zbirni red*, a zadnja kolona je *zbirna kolona.*Često je zbirni red prvi red posle zaglavlja, a zbirna kolona je prva kolona posle predkolone.

Svaka sta. tabela ima *naslov* iz koga se saznaje o pojavi na koju se podaci u tabeli odnose. Ispod naslova, na desnoj strani tabele stoji *jedinica mere* u kojoj se pojava iskazuje. Često ispod tabele stoje tekstualna objašnjenja npr. o izvoru podataka ili o metodologiji.

ŠEMA STATISTIČKE TABELE

					<----- zaglavlje
					red
					red
					red
					<---- zbirni red
pretkolona	kolona	kolona	kolona	kolona	zbirna kolona

Stat. tabele mogu biti:

- proste ili jednostavne (sadrže samo jednu pojavu (obeležje); ima 2 kolone: pretkolona za oznaku modaliteta obeležja i kolonu za brojčane podatke i potreban broj redova ili samo dva reda - zaglavlje i red za podatke)

Stanovništvo po stručnoj spremi

stepen stručnog obrazovanja	broj zaposlenih (f)
visoko	362
više	832
srednje	184
niže	210
ukupno:	1588

- složene (one najčešće sadrže 2 pojave (obeležja))

Rođeni, umrli i venčani (u hiljadama)

godina	živorođeni	umrli	prir.priraštaj	venčani	razvedeni
1951	447	235	212	171	16
1952	499	198	301	177	12
1953	483	212	271	167	16
1954	494	187	305	171	16
1955	472	200	272	162	19

- kombinovane (dobija se odgovarajućim ukrštanjem dva ili više obeležja istog skupa; oznake jednog obeležja se stavljaju u pretkolonu, a oznake drugog u zaglavlje)

STARDŽESOVO PRAVILO (američki naučnik Sturges)

Grupisanje elementarnih jedinica po numeričkom obeležju vrši se pomoću klasifikacije obeležja ili na podatke (modalitete) izražene celim brojem ili na modalitete izražene intervalima (ili intervalnim klasama).

Klasifikacija numeričkog neprekidnog i numeričkog prekidnog obeležja sa velikim brojem modaliteta vrši se podelom obeležja na pregledan broj intervalnih klasa. Poželjno je da sve intervalne klase budu iste veličine (širine), da bi se obezbedila uporedivost intervalnih klasa. Širin aintervalne klase je rastojanje između donje i gornje granice.

Obeležimo sa a_1 donju granicu (prvu vrednost), a sa a_2 gornju granicu (završnu vrednost) intervalne klase (a_1 - a_2). Jasno je da tada elementarna jedinica pripada toj klasi ako joj je vrednost ili jednaka sa a_1 ili je manja od a_2. Znači, *ako intervalne klase nisu međusobno razgraičene, onda elementarna jedinica koja ima graničnu vrednost intervalne klase pripada intervalnoj klasi u kojoj je ta vrednost donja granica*.!!!!

Širina intervalnih klasa zavisi od zahteva preglednosti i cilja istraživanja. Manji broj intervalnih klasa - veća preglednost - manje vođenje računa o osobinama obeležja.

Prva i poslednja intervalna klasa mogu biti otvorene i zatvorene. Otvorene intervalne klase omogućavaju da se obuhvate elementarne jedinice ekstremno malih ili velikih vrednosti obeležja .

Sredina intervalne klase se dobija: $(a_2 - a_1)/2 + a_1$.

Broj intervala klasa određuje se kao:

k=1+3,322 log N

k-broj inervalnih klasa (ili samo intervala)

N-ukupan broj elementarnih jedinica

Nakon određivanja broja intervalnih klasa, određuje se njihova širina kao:

$i=(X_{max} - X_{min}) / k$

X_{max} - elementarne jedinice sa maksimalnom vrednošću obeležja

X_{min} - elementarne jedinice sa minimalnom vrednošću obeležja

Napomena:

sledeći interval ne može započeti istim brojem kojim se prethodni završava.

NPR: N=52, Xmax=84, Xmin=22

k = 1+3,322 log 52=6,7≈7 (7 intervala)

$i=(X_{max} - X_{min})$ / k=(84-22)/7=8,89=9 (širine 9)

22-30	sredina intervala:	(30-22)/2+22=26
31-39		(39-31)/2+31=35
40-48		44
49-57		53
58-66		62
67-75		71
76-84		80

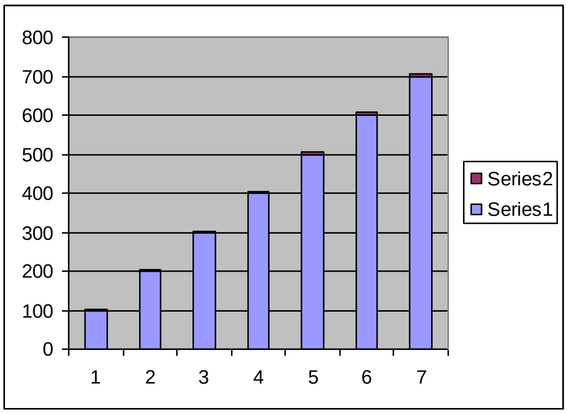
GRAFIČKO PREDSTAVLJANJE STATISTIČKIH SERIJA

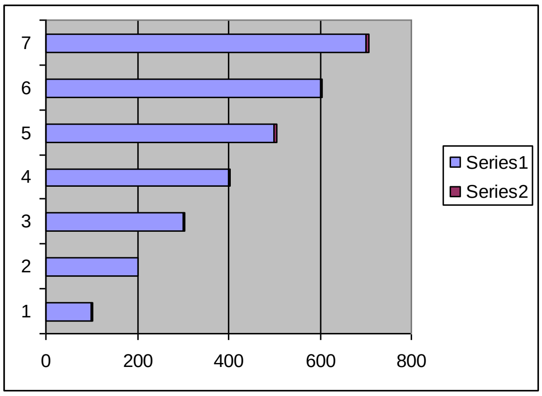
Grafičko prikazivanje se koristi da bi statistički podaci iz tabela bili preglednije i slikovitije prezentirani. U praksi se podaci iz svake tebele ne predstavljaju grafički, već se to radi samo za najvažnije podatke ili kao uvod u statističku analizu. To znači da je osnovni vid prikazivanja podataka tabelarni.

Grafički prikazi moraju biti jasni, jednostavni, pregledni i moraju odražavati srazmeru veličina koje se nalaze u statističkoj tabeli. Mora postojati tumač znakova i zaglavlje grafika (kratak naslov) koje objašnjava podatke.

Strukturne serije sa atributivnim obeležjem se prikazuju pomoću histograma, tako što se frekvencija modaliteta obeležja prikazuje površinom, koja može biti ***pravougaonik*** (to može biti histogram - stub i histogram - gredica; mogu se koristiti dvostruki i višestruku pravougaonici), ***kvadrat*** (dijagrami sa kvadratima se posebno koriste kada se ističe promena obima skupa za određeni period - nacrta se manji i veći kvadrat; površina između manjeg i većeg kvadrata predstavlja povećanje obima skupa za posmatrani vremenski perod) i ***krug*** (isto kao i kod kvadrata - da istakne promenu u obimu skupa; manji i veći krug).

Histogram-stub i histogram-gredica se prikazuju u I kvadrantu koordinatnog sistema tako što se pravougaonici sa istim osnovicama i na istim međusobnim rastojanjima kod histogram-stuba nanose na apscisu (x-osu), a kod histogram-kruga na ordinatu (y-osu). Na drugu koordinatu se nanose frekvencije.

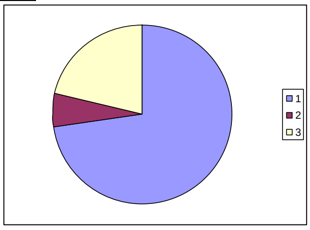




Kumulanta služi za grafički prikaz kumulativnih frekvencija - apsolutnih i relativnih. Prikazuje se u I kvadrantu koordinatnog sistema. Kumulativne frekvencije se nanose na grafik u vidu verikalnih pravih (ordinata) koje polaze iz tačaka na apscisi gde su obeležene sredine grupnih intervala. Vertikalna prava je jednaka kumulativnim frekvencijama. Vrhovi vertikalnih pravih se spajaju pravom, pa se dobija lomljena linija koja počinje o 0 (nule), sve vreme raste dok ne dostigne visinu koja je jednaka opštem zbiru frekvencija. Kumulanta se može dati i pomoću histograma.

Histogram-krug se koristi za prikaz *strukture atributivne serije*. Isečci kruga koji odgovaraju atributima obeležja dobijaju se tako što se procentualna zastupljenost svakog atributa pomnoži sa 3,6 stepeni (jednom procentu strukture pripada 3,6 stepeni jer je 360:100=3,6). Podesan je ako je obeležje dato sa najviše 4 atributa. Može se koristiti i za prikaz strukturnih promena nastalih tokom vremena. Za više od 4 obeležja bolje je koristiti ***histogram-stub ili histogram-gredice.***

vrsta uzorka	broj radnika	struktura u %	stepeni isečka kruga
VSS	257	21.2	76.3
VS	788	65.0	234.1
KV	167	13.8	49.6
ukupno	1212	100.0	360.0



Struktirna serija sa numeričkim prekidnim obeležjem koje je dato celim brojevima, najbolje se prikazuje u I kvadrantu koordinatnog sistema - na apscisu se nanose modaliteti obeležja, a na ordinatu odgovarajuće frekvencije; grafik izgleda: niz vertikalnih linija (ordinata koje imaju odgovarajuće značenje klase obeležja) podignutih iz tačaka na apscisi koje imaju odgovarajuća značenja modaliteta. Visina linija označava veličinu frekvencija datih klasa. Ovakav grafik se naziva stigmogram. Ako se krajevi vertikalnih linija povežu pravom, dobija se linijski dijagram, a ako se krajevi prve i poslednje vertikalne linije povežu sa apscisom, dobija se poligon.

X	f
0	108
1	114
2	45
3	21
4	9
5	3
ukupno	300

POLARNI DIJAGRAM (za prikaz vremenskih serija)

Pored pravougaonog koordinatnog sistema u statistici se ponekad upotrebljava i tzv. polarni koordinatni sistem. Ime je dobio po tome što prikazuje podatke na linijama koje idu iz jednog centra (pola) zrakasto u sve pravce. Linija na kojoj su merni brojevi obično polazi od pola (0) horizontalno na desno u beskonačnost i zove se ***polarna osa***. Položaj ma koje tačke u ovom sistemu određuje se njenim ***potegom*** (rastojanjem od pola) i uglom koji taj poteg zatvara sa polarnom osom (polarnim uglom). Znači, polarne koordinate su poteg i polarni ugao.

U statistici se sistem polarnih koordinata uglavnom koristi za prikaz vremenskih serija koje pružaju mesečne podatke o pojavi i pokazuju sezonske varijacije. Ove serije zahtevaju polarni sistem od 12 potega (znači sa polarnim uglovima od 30^0). Skala se piše na polarnoj osi, pa se odatle rastojanje pojedinih potega, koji odgovara veličini mesečnog podatka, prenosi šestarom. Kad se tačke koje označavaju podatke spoje pravim linijama, dobija se slika varijacije pojave po mesecima. Prosečna mesečna vrednost se ucrtava pomoću kruga odgovarajućeg rastojanja od pola, tako da se svaki mesečni podatak može upoređivati sa prosekom, što polarnom dijagramu daje poseban značaj.

PARETO DIJAGRAM i ABC analiza

To je grafička metoda odvajanja bitnog od manje bitnog, u cilju otkrivanja problema i donošenja korektivnih mera, kako bi se otklonio problem. Pareto metod ukazuje na mali broj važnih problema i veliki broj nevažnih, odnosno nekoliko karakteristika je odgovorno za najveći broj posledica.

Nedostatak metode je što ne govori o uzroku nastalih problema, već samo omogućava statističku analizu.

Razlikovanje najbitnijih karakteristika od manje bitnih, uz pravilno definisanje i sprovođenje korektivnih mera dovodi do optimalnog poboljšanja, racionalnog korišćenja resursa i minimalnih troškova.

Kada se posmatrane karakteristike urede u opadajućem redosledu, kada im se odredi apsolutni i relativni doprinos, uočava se mesto i uloga svake karakteristike.

Služi za prikaz atributivnih serija i to tako što su atributi poređani po veličini s leva u desno, pa atributi sa najvećom frekvencijom zauzimaju krajnju levu poziciju, a atributi sa mnajmanjom frekvencijom krajnje desnu poziciju.

Ime je dobio po italijanskom ekonomisti Paretu (Vilfredo Pareto) koji je prvi konstruisao dijagram preuređene atributivne serije. Analizirao učešće pojedinih grupa stanovnika u ostvarenom dohotku. Utvrdio da samo 20% stanovnika učestvuje u 80% ukupnog dohotka zemlje, a ostalih 80% ostvaruje samo 20% dohotka. On je smatrao da je otkrio univerzalni zakon. Zakon 20-80 nazvan je PARETO PRINCIP.

Omogućava efikasno identifikovanje uzroka problema i na osnovu toga eliminisanje svih gubitaka. Pareto analiza se koristi u analizi pojava sa ciljem:
- rangiranja karakteristika pojava prema stepenu važnosti
- razdvajanja kritičnih karakteristka pojava od nevažnih
- rešavanja problema kako bi se ostvario max efekat

Najvažnija karakteristika Pareto kumulativne krive se može generalizovati na sledeći način:
- prvih 10 izabranih i grupisanih karakteristika predstavlja 60% kumuliranih procentualnih vrednosti; ova oblast se obeležava sa A
- sledećih 30 grupisanih karakteristila predstavlja 30% kumuliranih procentualnih vrednosti; ova oblast se obeležava sa B
- preostalih 60 izabranih i grupisanih karakteristika predstavlja 10% kumuliranih procentualnih vrednosti; ova oblast se obeležava sa C

U oblasti A su koncentrisani najvažnij problemi i ova oblast podleže dejstvu dugoročnih korektivnih mera.

U oblasti B su koncentrisani manje važni problemi i oni podležu dejstvu povremenih mera.

U oblasti C su koncentrisani problemi koji nastaju pod dejstvom slučajnih faktora i oni su nebitni.

Pareto dijagram daje mnoge informacije. Kolone su date u opadajućem redosledu pa ukazuju na koje probleme prvo treba obratiti pažnju, pa onaj ko rešava problem vizuelno lako određuje koji problemi imaju prioritet pri rešavanju. Kumulativna linija pokazuje da je korekcijom dva najveća problema utvrđen najveći procenat problema.

Postupak za dobijanje Pareto karte:

1. korak

+ izbor problema koji se istražuje (koji problem i u kojoj oblasti)
+ određivanje vremenske učestalosti analize
+ prikupljanje podataka iz kritičnog perioda

2. korak

+ kreiranje karte za beleženje podataka

vrsta karakteristika	nivo pojave koja se posmatra
A B C D	
Σ	Σ

3. korak

+ popunjavanje karte

4. korak

+nivo pojave se preuredi tako da se pojave prikazuju po veličini u opadajućem redosledu

+ uređene frekvencije po veličini se grafički prikazuju pomoću histograma

+ izračunavaju se relativna učešća po karakteristikama

+ računaju se kumulativna relativna učešća

+ kumulativ relativnih učešća se prikaže grafički na istom dijagramu pomoću izlomljene linije

5. korak - Konstrukcija ABC dijagrama:

+na X-osi se prikazuju kvalitativne promenljive (sa atributima ili kategorijama), a na Y-osi se prikazuju frekvencije (apsolutna ili relativna)

+ prikažu se kumulativi relativnih učešća

6. korak

+ dobijena kriva se podeli na 3 oblasti da se dobije ABC kriva

7. korak

+ donošenje odluke i korektivne mere; nakon sprovo|enja korektivnih mera dobro je uraditi novi Pareto dijagram iz koga se lako uočavaju efekti preduzetih mera.

Oblast primene Pareto analize

- kvalitet (greška, žalbe, vraćanje proizvoda, popravke)
- troškovi (vrste troškova, planirani troškovi, gubici)
- isporuka (zaliha nekompletnih proizvoda, greške u plaćanju, kašnjenje u isporuci)
-bezbednost saobraćaja (greške, stradanja)
- radna snaga (smena, iskustvo, veština)
- sredstva (oprema, mašine, instrumenti)
- sirovine (proizvođač, fabrika, vrsta)
- metod rada (uslovi, pripreme, metode)

PRIMER

raspodela proizvoda prema uzrocima reklamacija

uzrok reklamacija	broj proizvoda	% (Pi)
A	60	7.5
B	257	32.1
C	70	8.8
D	380	47.5
E	33	4.1
ukupno	800	100.0

(60:800)x100=7.5
(257:800)x100=32.1
....
Preurediti atributivnu seriju i prikazati je na Pareto dijagramu.
Rešenje: nivo pojave sortiramo po veličini i odredimo odgovarajuće kumulative.

uzrok reklamac.	broj proizv.	% (Pi)	kumulativ proizv.	kumulativ % (Pi)
D	380	47.5	380	47.5
B	257	32.1	637	79.5
C	70	8.8	707	88.4
A	60	7.5	767	95.9
E	33	4.1	800	100.0
ukupno	800	100.0	*	*

4. MERE POPULACIJE I UZORKA

Statističke serije su osnova statističke analize. Na tako grupisanim podacima istražuje se zakonitost ponašanja posmatrane pojave i analiziraju odnosi. Pošto nije moguće uraditi statističku analizu na svim jedinicama nekog skupa, bira se uzorak prema definisanim kriterijumima i na njemu se radi statistička analiza. Da bi se dobili podaci o svojstvima skupa (pojave), statistička analiza koristi nekoliko parametara koji daju informacije o koncentaciji, disperziji i obliku tog skupa.

U skupu sve elementarne jedinice teže da se **koncentrišu** oko vrednosti jedne element.jedinice, koja se određuje po strogo definisanoj statist.metodologiji. To grupisanje element.jedinica oko neke vrednosti (iznad i ispod) nastaje pod dejstvom slučajnih faktora.

Grupisanje element.jedinica u grupne intervale pokazuje da statistika ima zadatak da utvrdi zakonitost u masovnim varijabilnim pojavama i da same elementarne jedinice kao jedinke nisu bitne.

Odstupanje ili **disperzija** se takođe izražava i određuje po strogo definisanoj statist.metodologiji. Elementarne jedinice se raspoređuju po nekom modelu (osnov je normalna raspodela). Grafički se pokazuju izgledi tih modela raspodela. Analizom udaljenosti krive od X-ose, analizom rasporeda tačaka u odnosu na Y-osu i udaljenosti tačaka od neke zamišljene linije, dobijaju se podaci o **obliku raspodele** element.jedinica.

Pokazatelji koncentracije, disperzije i oblika nazivaju se mere centralne tendencije (mere koncentracije ili srednje vrednosti), mere disperzije i mere oblika.

MERE KONCENTACIJE (mere proseka - MP)

Težnja koncentracije vrednosti obeležja oko centralne vrednosti (srednje vrednosti, tipične vrednosti) jedne serije podataka naziva se *centralna tendencija*. Na dijagramima, raspodela podataka pokazuje manje ili više izraženu koncentraciju.

Statistički pokazatelji kojima se iskazuje centralna tendencija zovu se ***mere proseka*** (mere srednje vrednosti, mere centralne tendencije). Njihova uloga je da, pošto svaka raspodela frekvencija ima različite vrednosti obeležja kao rezultat razlika u obeležjima posmatranih statističkih jedinica, jednom zajedničkom merom predstave baš te različite vrednosti. Mere proseka su jedna od najvažnijih pokazatelja numeričkih podataka koji izražavaju obeležja statističkih jedinica. Jednom vrednošću, kojom su izražene, mere proseka izražavaju i uopštavaju varijabilnost svih vrednosti jedne serije dajući sažete informacije, izražene jednim brojem.

MP otkrivaju samo jednu karakteristiku raspodele – centralnu tendenciju, omogućavajući:

- zaključivanje o obeležjima pojave koju podaci predstavljaju
- upoređivanje više raspodela koje je nemoguće vršiti samo na osnovu poznavanja sređenih podataka

MP se mogu izračunati:

- iz negrupisanih (sirovih) podataka
- iz grupisanih (sređenih) podataka

MP se mogu grupisati prema sledećim karakteristikama:

- načinu određivanja
- mogućnostima primene

S obzirom na <u>način određivanja</u> MP mogu biti:

- pozicione
    + modus
    + medijana
- izračunate
    + aritmetička sredina (najširu primenu, jer se lako izračunava)
    + geometrijska sredina
    + harmonijska sredina

Zašto pozicione? Srednja vred. se može odrediti kao vrednost koja ima najveću frekvenciju – modus, ili kao vrednost koja se nalazi tačno u sredini raspodele – medijana.

MP se razlikuju po:

- načinu izračunavanja
- matematičkim osobinama
- značaju u procesu istraživanja

<u>Izbor mera proseka</u>

Osnovno je znati da se izabere mera za konkretnu seriju podataka. Izbor zavisi od:

- cilja statističke obrade
- načina raspodele vrednosti obeležja
- vrste statističke serije

U izboru se treba pridržavati sledećih kriterijuma:

- ar.sred. se koristi ako se raspolaže podacima koji mnogo ne odstupaju, tj. mnogo se ne razlikuju međusobno
- medijana se koristi ako su podaci takvi da jedan ili nekoliko mnogo odstupa od ostalih
- modus se koristi ako se nešto želi istaći, nešto što je tipično

ARITMETIČKA SREDINA (ili samo sredina - AS)

Kao mera proseka AS se koristi:
- kada se raspodela vrednosti približava simetričnoj
- kada u daljem postupku obrade treba računati stand.devijaciju i koeficijent korelacije
- kada se zahteva da svaka vrednost raspodele uđe u vrednost mere proseka

Osobine AS:
- najčešće upotrebljavana MP
- predstavlja tačku koja izražava sredinu gustine u raspodeli
- jednostavna, razumljiva i laka za računanje
- najpouzdanija i najpreciznija MP
- bolje predstavlja statistički skup ako je homogeniji i veći
- stabilnija je od drugih MP (ima manju standardnu grešku)
- odstupanja od AS se računaju kada se od svake pojedinačne vrednosti obeležja oduzme AS
- zbir pojedinačnih odstupanja vrednosti serije od AS uvek je 0 (nula)
- ako su elementarne jedinice iste, onda AS ima istu vrednost
- ako se elmentarnim vrednostima pripiše vrednost AS, onda ukupan obim obeležja ostaje nepromenjen
- suma pozitivnih odstupanja elementarnih jedinica od AS jednaka je sumi negativnih odstupanja elementarnih jedinica od AS
- zbir kvadrata odstupanja vrednosti elementarnih jedinica od AS manji je od zbira kvadrata odstupanja vrednosti elementarnih jedinica od bilo koje druge srednje vrednosti

AS se nalazi:
- između najmanje i najveće vrednosti obeležja u seriji
- u svakoj raspodeli frekvencija
- samo kao jedna vrednost u svakom nizu

Bitno je:
- AS je veća od najmanje i manja od najveće vrednosti niza
- računa se iz svih vrednosti numeričkog obeležja
- ako su sve vrednosti obeleža jednake, AS jednaka toj vrednosti

AS se izražava:
- u mernim jedinicama obeležja
- uvek decimalnim brojem

Nedostaci AS:
- ne govori mnogo o homogenosti statističke serije
- umanjuje reprezentativnost serije zbog ekstremnih vrednosti
- ne daje pravu sliku koncentracije vrednosti obeležja kada je raspodela bimodalna i multimodalna (za ovaj slučaj treba povećati uzorak; ako se multimodalnost ne ukloni nakon povećanja uzorka, treba odustati od računanja MP i raspodelu predstaviti grafički)

AS se dobija kada se vrednosti obeležja elementarnih jedinica saberu i zbir podeli sa brojem elementarnih jedinica.

Za **skup** se AS obeležava sa **μ** (″mi″), a za **uzorak** sa $\overline{X}$ (″iks bar″). AS se izračunava za *ponderisane i neponderisane podatke*.

$$\mu = X_1 + X_2 + ... + X_N / N = \sum X / N \quad \text{ili} \quad \mu = \overline{\frac{1}{N}\sum_{i=1}^{N} X_i}$$

$$\overline{X} = X_1 + X_2 + ... + X_N / N = \sum X / N \quad \text{ili} \quad \overline{\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i}$$

Kod raspodela moramo uzeti u račun i frekvencije, jer inače AS ne bi bila reprezentativna. Znači, AS distribucije uvek je jedna ponderisana sredina, koja se dobija pošto se prethodno svaka vrednost obeležja ponderiše njenom frekvencijom, što znači da svaka vrednost dobija važnost (težinu) prema frekvenciji koja joj odgovara. Obrazac za ponderisnu AS:

$$\overline{X} = X_1 f_1 + X_2 f_2 + ... + X_N f_N / f_1 + f_2 + ... + f_N = X_1 f_1 + X_2 f_2 + ... + X_N f_N / \sum f$$

ili kraće:

$$\overline{X} = \sum f X / N \qquad (\sum f = N)$$

frekvencija f – ponder (uloga mu je da istakne značaj nekog obeležja u nizu posmatranih obeležja)

Kada se računa AS kod intervalnih klasa, zapaža se da elementarne jedinice koje pripadaju odre|đnoj intervalnoj klasi često nisu ravnomerno raspoređene sa svojim vrenodnostima unutar intervalne klase. Zato se desi da postoji mala razlika između izračunate i tačne AS. Razlika je *sistematska greška.* Njena veličina zavisi od veličine (širine) intervalnih klasa. manja je kod manjih inetrvalnih klasa i obratno. Ona se ne može eliminisati.

Sredina raspodele (neintervalne i intervalne) relativnih frekvencija se dobija kada se vrednosti klasa (ili sredine intervanih klasa (pomnože odgovarajućim relativnim frekvencijama i dobijeni proizvodi saberu. Sredina skupa je $\mu = \sum X p$, a sredina uzorka je $\overline{X} = \sum X \overline{p}$.

1. PRIMER – AS osnovne serije

Izračunati AS iz negrupisanih (nesređenih) podataka o uspehu studenata na ispitu (N=33)

10,8,9,7,5,6,5,7,6,10,9,8,9,8,7,5,6,6,5,8,7,9,5,7,6,8,6,7,8,8,7,6,7

$\overline{X} = \sum X / N = 235/33 = 7{,}12$

2. PRIMER – AS neintervalne raspodele

ocena	X	10	9	8	7	6	5
br.stud.	f	2	4	7	8	7	5

$\overline{X} = \sum f X / N$

ocena	X	10	9	8	7	6	5	∑
br.stud.	f	2	4	7	8	7	5	33
	fX	20	36	56	56	42	25	235

$\overline{X} = \sum f X / N = 235/33 = 7{,}12$

3. PRIMER – AS intervalne raspodele

X	X_i	f	f X_i
22-30	26	6	156
31-39	35	8	280
40-48	44	9	396
49-57	53	12	636
58-66	62	7	434
67-75	71	6	426
76-84	80	4	320
i=9	*	52	2648

$\overline{X} = \sum f X_i / N = 2648/52 = 50{,}92$

GEOMETRIJSKA SREDINA se retko koristi u statističkoj analizi.
HARMONIJSKA SREDINA se koristi samo kod specifičnih raspodela, koje sadrže podatke o recipročnim vrednostima.

<u>MEDIJANA</u> (Me)

Poziciona srednja vrednost. Deli sumu frekvencija na dva jednaka dela.
Osobine Me:
- računa se jenostavno
- računa se samo za numerička obeležja
- računa se i za grupisane i za negrupisane podatke
- nije pod uticajem ekstremnih vrednosti obeležja
- grublja je procena srednje vrednosti od AS
- ne daje pravu sliku koncentracije vrednosti obeležja ako je raspodela bimodalna ili multimodalna

-Me postoji u svakom skupu ili uzorku
- podesna je kao srednja vrednost kod intervalnih raspodela sa otvorenim početnim ili krajnjim intervalom
- može se odrediti **grafički** pomoću kumulativa frekvencija (kumulanta) - prvo se nacrta kumulanta, a zatim se odredi središna jedinica; na ordinati se pronađe središna jedinica i iz te tačke povuče paralela sa apscisom do preseka kumulante; iz tačke preseka kumulante povuče se paralela sa ordinatom do preseka apscise; vrednost apscise u tački preseka je vrednost medijane

Me se koristi:
- kada nije preporučljivo koristiti AS
- kada je važno da li se podaci nađu u I ili II polovini raspodele, a ne zanima nas koliko su udaljeni od središne tačke
- kada su vrednosti izrazito asimetrično raspoređene oko AS (izrazito velike i izrazito male), na bilo kom kraju niza

Kod normalne raspodele AS=Me=Mo
Središna jedinica se određuje obrascem:
(N+1) / 2 (N - broj podataka)
Broj jedinica može biti paran i neparan. Kada je broj jedinica paran, medijana se određuje kao prosta AS vrednosti obeležja dve središne jedinice rastužć}eg niza.

Medijana intervalnog rasporeda frekvencija se određuje po sledećoj formuli:

$$M_e = A_k + i \cdot \frac{\frac{N}{2} - \sum_{i=1}^{m_e - 1} f_i}{f_{m_e}}$$

A_k-donja (leva) granica medijalnog intervala
i - veličina grupnog intervala
N/2 - polovina frekvencija rasporeda frekvencija
f_{me} -frekvencija medijalnog intervala
Σf_i - kumulativ frekvencija predmedijalnog intervala

1.PRIMER – Me osnovne serije

a) 2,4,4,4,5,7,7,9,10,10 (N=10)
Me = (N+1)/2=10+1/2=5,5
Znači, na 5,5 mestu u seriji nalazi se središna vrednost. Ovde je to između 5 i 7. Pošto je paran broj (N=10), Me se određuje kao AS dve susedne vrednosti:
Me=5+7/2=6

b) 2,4,4,4,5,7,7,9,10,10,11 (N=11)
Me = (N+1)/2=11+1/2=6 Me je na 6. mestu – to je obeležje 7

2. PRIMER – Me neintervalne raspodele

X (ocena na ispitu)	f	∑f
5	5	5
6	7	12
7	8	20
8	7	27
9	4	31
10	2	33
	33	*

Me = (N+1)/2=33+1/2=17
U ∑f se 17 nalazi u 20; to obeležje je ocena 7

3. PRIMER – Me intervalne raspodele

X	X_i	f	∑f
22-30	26	6	6
31-39	35	8	14
40-48	44	9	23
49-57	53	12	35
58-66	62	7	42
67-75	71	6	48
76-84	80	4	52
i=9	*	52	*

N+1/2=52+1/2=26,5

$$M_e = A_k + i \cdot \frac{\frac{N}{2} - \sum_{i=1}^{m_e - 1} f_i}{f_{m_e}} = 49+9 \qquad = 51{,}25$$

Znači, prvih 50,00% studenata je ostvarilo manje od 51,25 poena, a drugih 50,00% rezultat više od toga.

<u>MODUS</u> (MOD - Mo)

Kaže se da modalna ili najčešća vrednost jedne raspodele, koja se zove modus, jeste vrednost koja ima najveću frekvenciju (učestalost pojavljivanja).

Pri određivanju modusa ne koriste se vrednosti obeležja svih jedinica ni sve jedinice. Vrednost sa najvećom frekvencijom je modus.

Osobine modusa:
- modus je poziciona srednja vrednost
- lako se određuje kao gruba mera proseka
- prilikom određivanja modusa ne uzimaju se u obzir sve jedinice
- modus je vrednost obeležja koja ima najveću frekvenciju (tipičan slučaj)

- poželjno je koristiti modus kada se želi da se isključe ekstremne vrednosti obeležja, a nije ga poželjno koristiti ako ekstremna vrednost ima najveću frekvenciju
- raspored frekvencija može biti bez modusa (ako su sve vrednosti serije različite, ali se pojavljuju jednak broj puta) i imati jedan ili više modusa (može biti: unimodalan, bimodalan ili multimodalan)
- pri korišćenju modusa posebno treba obratiti pažnju pri formiranju grupnih intervala, jer različite veličine različito utiču na raspored frekvencija
- antimodus je vrednost obeležja sa najmanjom frekvencijom (raspored frekvencija može imati jedan, dva ili više antimodusa)

Nedostatak Mo:
- uzima u obzir samo obeležje sa najvećom frekvencijom, a ostala zanemaruje
- ne može se uvek odrediti
- nepouzdan za israzito asimetrične raspodele u kojima je modalni interval otvoren

Modus se može odrediti ***grafički***. Tada kod *neintervalnih rasporeda frekvencija* modus je onaj modalite obeležja koji na grafiku ima najveću ordinatu. Za *intervalne rasporede frekvencija* može se pronaći samo približan modus. Tada se određuje na histogramu tako što se dijagonalno spajaju početna i krajnja tačka na gornjoj stranici modalnog pravougaonika sa krajnjom tačkom gornje stranice predmodalnog pravougaonika i početnom tačkom gornje stranice poslemodalnog pravougaonika.

Kod intervalnih rasporeda frekvencija modus je:

$$M_o = A_k + i \cdot \frac{f_m - f_{m-1}}{2f_m - f_{m-1} - f_{m+1}}$$

A_k-donja (leva) granica modalnog intervala
i - veličina grupnog intervala
f_m - frekvencija modalne intervalne klase
f_{m-1} - frekvencija predmodalne intervalne klase
f_{m+1} - frekvencija poslemodalne intervalne klase

1.PRIMER – Mo neintervalne raspodele

X (ocena na ispitu)	f
5	5
6	7
7	8
8	7
9	4
10	2
	33

Najveća frekvencija je 8; odgovara obeležju, tj oceni 7.

2. PRIMER – **Mo intervalne raspodele**

X	X_i	f
22-30	26	6
31-39	35	8
40-48	44	9
49-57	53	12
58-66	62	7
67-75	71	6
76-84	80	4
i=9	*	52

$$M_o = A_k + i \cdot \frac{f_m - f_{m-1}}{2f_m - f_{m-1} - f_{m+1}} = 49+9 \qquad =53{,}3$$

Znači, najčešći broj bodova po učeniku je 53,3

<u>MERE VARIJACIJE</u> (ili disperzije)

Varijabilnost je osnovno svojstvo prirodnih i društvenih pojava.

Mere varijacije (MV) pokazuju koliko je neki statistički skup **heterogen.**

MV pokazuju koliko je neki statist.skup **diferenciran** (koliko vrednosti obeležja odstupaju ispod ili iznad srednje vrednosti i to pod dejstvom slučajnih faktora.

Odstupanje elementarnih jedinica se može meriti u odnosu na bilo koju meru proseka. Međutim, zbog osobine AS da je suma pozitivnih odstupanja vrednosti obeležja od AS jednaka sumi negativnih odstupanja vrednosti obeležja od AS minimalna, i s obzirom da se AS češće koristi od drugih mera proseka, odstupanja se mere u odnosu na AS.

Ako pojedinačne pojave, izražene u vrednostima statističkih jedinica, odstupaju od proseka, iznad i ispod u podjednakom broju i pri tome se potiru, to su *slučajna odstupanja.*

Ako odstupanja idu u suporotnom pravcu, čime menjaju prosek raspodele, pokazujući novu tendenciju, to su *sistematska odstupanja*.

Ako želimo da opišemo neku pojavu, predstavljenu nizom numeričkih podataka, prvo ćemo se opredeliti za neku meru proseka (MP). Ipak MP o tom nizu numeričkih podataka ne mogu reći sve, jer se dešava i da loše reprezentuju osnovni skup.

PRIMER:

četiri raspodele (rezultati na testu znanja)

A	6	8	9	10	12
B	8	9	9	9	10
C	4	6	7	13	15
D	9	9	9	9	9

AS za sve 4 raspodele je 9. Iako je AS jednaka ovo su različite raspodele po disperziji (rasturanju) vrednosti, tj. imaju različit stepen varijabilnosti obeležja. Najveće rasturanje je kod C. Kod raspodele D su svi podaci jednaki, nema rasturanja, pa je i besmisleno računati AS (ovi podaci se i ne smatraju statističkim). Osim toga, one se razlikuju po Me i Mo.

Znači, da bi mogli porediti dve ili više statističkih serija, pored podataka o MP moramo imati informacije o odstupanjima vrednosti od proseka.

Stepen varijabilnosti pojave istovrsnih statističkih podataka može biti veće ili manje.

Mere varijabilnosti (MV) pokazuju kako se individualni rezultati raspodele grupišu oko utvrđenih mera srednje vrednosti. One govore o varijaciji u odnosu na AS i pokazuju reprezentativnost AS.

Varijacije između članova raspodele mogu se meriti kao:

- odstupanje između dva člana u raspodeli
- odstupanje između više članova u raspodeli
- odstupanje između pojedinih članova raspodele i njihove MP

Na mere varijacije utiču sve jedinice koje se istražuju.

Sve mere varijacije mogu imati apsolutni i relativni izraz. Apsolutne mere varijacije imaju isti izraz (meru) kao i obeležje. Ove mere ne mogu služiti za upoređivanje rasporeda frekvencija različitih izraza (mera) obeležja.

Relativne mere varijacije izražavaju varijabilitete u procentima i mogu služiti za upoređivanje varijabiliteta rasporeda frekvencija sa različitim izrazima (merama) obeležja.

Mere varijacije sa apsolutnim izrazom mogu biti:

- pozicione
- izračunate u odnosu na sredinu

Pozicione mere varijacije su:

- razmak varijacije (interval varijacije)
- interkvartilni (ili interdecilni ili interpercentilni) razmak varijacije

Mere varijacije u odnosu na sredinu su:

- srednje apsolutno odstupanje (ili srednje linearno apsolutno odstupanje)
- varijansa
- standardna devijacija

Relativne mere varijacije su:

- koeficijent kvartilne varujacije
- srednji linearni koeficijent varijacije
- koeficijent varijacije

Mera varijacije koja pojedinačni varijabilitet elementarne jedinice izražava u standardnim devijacijama je **standardizovano odstupanje**.

RAZMAK

Razmak je gruba mera varijacije zato što na njegovo određivanje ne utiču sve jedinice. Dobija se kao razmak ekstremnih vrednosti obeležja. Obeležava se sa R i glasi: $\mathbf{R=X_{max}-X_{min}}$ (x - obeležje)

Kod neintervalnih rasporeda frekvencija X_{max} je poslednja vrednost obeležja, a X_{min} je prva.

Kod intervalnih rasporeda frekvencija X_{max} je gornja granica poslednjeg grupnog intervala, a X_{min} je donja granica prvog grupnog intervala.

Nedostaci R:

- ne daje mnogo informacija

- neprecizna mera
- nije pouzdan za raspodele sa otvorenim intervalima
- ne izražava varijacije svih vrednosti obeležja, nego samo ekstremnih

Primena R:

- za preliminarno, na prvi pogled, određivanje varijabilnosti
- kao etapa statističke obrade podataka
- pri izračunavanju ostalih mera varijabilnosti i mera asimetrije
- uz Mo kao meru proseka
- kao pokazatelj homogenosti statističkog niza, koja je veća što je R manji
- kada su podaci rastureni i oskudni

1. PRIMER – R osnovnog skupa

10,20,20,30,50,10,15,30

R= $\mathbf{X_{max}}$-$\mathbf{X_{min}}$ =50-10=40

2. PRIMER – R neintervalne raspodele

X	1	5	10	20

R= $\mathbf{X_{max}}$-$\mathbf{X_{min}}$ =20-1=19

3.PRIMER – R intervalne raspodele

X
22-30
31-39
40-48
49-57
58-66
67-75
76-**84**

R = $\mathbf{X_{max}}$-$\mathbf{X_{min}}$ =84-22=62

INTERKVARTILNI RAZMAK (IR)

IR se formira da bi se izbegao uticaj ekstremnih vrednosti obeležja (max i min).

Raspored frekvencija se može podeliti na više jednakih delova.

Ako se podeli na 10, skup ili uzorak se tada deli na decile, tj. elementerane jedinice koje koje se nalaze na mestu u rastućem nizu u skupu ili uzorku koji je određen procentom 10%, 20%...90%.. IR se tada dobija odstanjivanjem levo i desno od granica razmaka po 10% elementarnih jedinica i sadrži 80% elem.jedinica skupa ili uzorka.

Ako se deli na 100 jednakih delova, skup ili uzorak se tada deli na percentile, elementerane jedinice koje koje se nalaze na mestu u rastućem nizu u skupu ili uzorku koji je određen procentom 1%, 2%...99%... IR se tada dobija odstanjivanjem levo i desno od granica razmaka po 1% elementarnih jedinica i sadrži 98% elem.jedinica skupa ili uzorka.

Ako se podeli na 4, onda se vrednosti obeležja nazivaju *kvartilima*.

Postoje 3 kvartila: prvi (donji), drugi (medijana), treći (gornji).

IR je razlika između III i I kvartila:

$IR = Q_3 - Q_1$

IR govori da je središnjih 50,00% jedinica statističke serije u rasponu između I i III kvartila, što znači da se pri određivanju IR ne uzima u obzir prva i poslednja četvrtina podataka, čime je rešen problem ekstremnih vrednosti varijacije.

IR je nepotpuna MV. Ne može se koristiti za poređenje varijabilnosti raspodela izraženih u različitim mernim jedinicama.

Kod skupa ili uzorka čije su jedinice sređene u rastućem nizu, kvartili su vrednosti obeležja koje se nalaze na mestu u nizu prema formuli:

$Q_1 = (N+1)/4$

$Q_2 = (N+1)/2$

$Q_3 = 3(N+1)/4$

1. PRIMER – IR za neintervalna obeležja

podaci se srede u rastućem redosledu:

11,12,15,15,16,17,18,20,21,23,24,25,25,26,29,30,31,32,33,34

$Q_1 = (N+1)/4=(20+1)/4=5{,}25$ na poziciji 5 (između 16 i 17) nalazi se vrednost I kvartila i iznosi 16,25

$Q_3 = 3(N+1)/4=3(20+1)/4=15{,}75$ na poziciji 15 (između 29 i 30) nalazi se vrednost III kvartila i iznosi 29,75

$Q_2 = (N+1)/2=(20+1)/2=10{,}5$ na poziciji 10 (između 23 i 24) nalazi se vrednost II kvartila i iznosi 23,50

$IR=Q_3 - Q_1=29{,}75-16{,}25=13{,}50$

SREDNJE APSOLUTNO ODSTUPANJE

Mere varijacije (MV) se računaju u odnosu na AS. Jedna od karakteritika AS je da je algebarski zbir odstupanja vrednosti obeležja od nje jednak 0 (nula). Da bi se izbegla 0, uzimaju se apsolutna odstupanja. Zanemaruje se predznak odstupanja i dobija odstupanje bez algebarskih znakova.

Sredina apsolutnih odstupanja (prosečno odstupanje) svih vrednosti obeležja jedinica od zajedničke AS predstavlja MV koja se naziva *srednje apsolutno odstupanje* (SAO).

$|SAO| = \sum |X-AS| / N$

$|SAO| = \sum |X-AS| f / \sum f \qquad \sum f=N$

Karakteristike SAO:

- retko se primenjuje
- nedovoljno pouzdana MV
- pouzdanija MV od razmaka
- manja SAO znači manju varijabilnost pojave
- primenjuje se kao predranja za izračunavanje varijanse i standardne devijacije

Nedostaci SAO:

- ograničena upotreba zbog zanemarivanja osobine AS da je algebarski zbir odstupanja pojedinih članova od njihove AS jednak 0
- ne može se porediti varijabilnost pojava koje imaju različite jedinice mere, različitu AS i različitu veličinu serije

1. PRIMER

X	X_i	f	f X_i	$\lvert$X-AS$\rvert$	$\lvert$X-AS$\rvert$ f
22-30	26	6	156	24,92	149,52
31-39	35	8	280	15,92	127,36
40-48	44	9	396	6,92	62,28
49-57	53	12	636	2,08	24,96
58-66	62	7	434	11,08	77,56
67-75	71	6	426	20,08	120,48
76-84	80	4	320	29,08	116,32
i=9	*	52	2648	*	678,48

$\overline{X} = \sum f X_i$ / N=2648/52=50,92
N=$\sum$ f
$\lvert$SAO$\rvert$=$\sum\lvert$X-AS$\rvert$ f / N=678,48/52=13,04
prosečno odstupanje od AS je 13,04

<u>VARIJANSA</u> (srednje kvadratno odstupanje ili disperzija uzorka)

To je srednje kvadratno odstupanje, odnosno sredina kvadrata odstupanja vrednosti obeležja, od AS (sredine serije).

$$\delta^2 = \mathrm{var}\,ijansa = \frac{\sum (X - sredina)^2}{broj\ \ elementarnih\ \ jedinica}$$

Varijansa usmesto da sve razlike meri u apsolutnim jedinicama, kao SAO, koristi kvadrate odstupanja. Zbir kvadrata tih odstupanja je različit od 0 (nule).

Prednosti varijanse:
- pogodna za dalju matematičku obradu
- izražava se u jedinicama mere koju imaju originalni podaci
- preciznija mera od SAO
- služi za izvođenje standardne devijacije

Nedostaci:
- složena za tumačenje
- nepodesna za poređenje kod različitih serija podataka
- predimenzionisana zbog kvadrata odstupanja

Označava se: σ^2 u skupu, a sa S^2 u uzorku.

1) Kod skupa:
varijansa se računa po jednoj od formula-

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

$$\sigma^2 = \frac{\sum X^2}{N} - \mu^2$$

$$\sigma^2 = \frac{\sum X^2 - N\mu^2}{N}$$

2) Kod uzroka:
varijansa se računa po jednoj od formula-

$$\sigma^2 = \frac{\sum (X-\mu)^2 \cdot f}{N}$$

$$\sigma^2 = \frac{\sum X^2 \cdot f}{N} - \mu^2$$

$$\sigma^2 = \frac{\sum X^2 \cdot f - N\mu^2}{N}$$

1. PRIMER

X	X_i	f	X^2	X^2 f
22-30	26	6	676	4056
31-39	35	8	1225	9800
40-48	44	9	1936	17424
49-57	53	12	2809	33708
58-66	62	7	3844	26908
67-75	71	6	5041	30246
76-84	80	4	6400	25600
i=9	*	52	*	147742

$\overline{X} = \sum f X_i$ / N=2648/52=50,92

$$\sigma^2 = \frac{\sum (X-\mu)^2 \cdot f}{N}$$

$$\sigma^2 = \frac{\sum X^2 \cdot f}{N} - \mu^2 \quad =147742/52-50,92^2=248,34$$

$$\sigma^2 = \frac{\sum X^2 \cdot f - N\mu^2}{N}$$

STANDARDNA DEVIJACIJA (SD)

To je prosečno odstupanje vrednosti numeričkog obeležja statističkog niza od njihove AS, tj. proseka.
Standardna devijacija (SD) ili standardno odstupanje je najvažniji pokazatelj odstupanja od aritmetičke sredine. Ona uvek daje meru odstupanja od aritmetičke sredine (ne od druge sredine). Veliki značaj koji je SD dobila u statistici duguje se njenoj uspešnoj primeni u ispitivanju normalne raspodele.
Označava se: σ u skupu, a sa S u uzorku. $\sigma = +\sqrt{\sigma^2}$

Karakteristike SD:
- iskazuje se u istim jedinicama mere kao i obeležje
- manja SD pokazuje da je skup homogeniji
- najčešće primenjivana MV
- preciznija MV od varijanse
- ne može imati negativnu vrednost

Primena SD:
- kad se zahteva najpouzdanija MV
- za opis raspodele frekvencija (na osnovu vrednosti SD može se proceniti način grupisanja podataka oko srednje vrednosti raspodele)
- za utvrđivanje reprezentativnosti AS
- kao osnova za druga statistička izračunavanja (npr. koeficijent korelacije)
- kada se očekuje da ekstremna odstupanja imaju veći uticaj na MV

Nedostatak SD:
- nepodesna za upoređivanje varijacija u različitim nizovima

Kod računanja SD za intervalnu raspodelu može se javiti mala razlika između dobijene i tačne vrednosti SD. Razlika se naziva *sistemska greška*. Njena veličina zavisi od broja jedinica i od širine (veličine) grupnih intervala.

Sistemska greška se može umanjiti. Engleski naučnik Cheppard – *Šepardova korekcija* dao formulu za otklanjanje sistemske greške:

$\delta' =$

i – širina grupnog intervala

Šepardova korekcija se primenjuje samo kod velikih skupova; ako je broj grupnih intervala manji od 12 (što je najčešći slučaj).

Kod SD uzorka se smatra zadovoljavajućom korekcijom ako se veličina uzorka umanji za 1.

1. PRIMER

$\delta^2=248{,}34$

$\delta=15{,}7$

$\delta'=15{,}5$

$\delta-\delta'=15{,}7-15{,}5=0{,}2$

KOEFICIJENT VARIJACIJE (relativna SD - KV)

Osnovni nedostatak svih MV je nemogućnost korišćenja za poređenje različitih skupova. Nedostatak se prevazilazi upotrebom KV.

KV je relativna MV kojom se u procentima izražava odnos između SD i AS.

Pomoću KV se prikazuje prosečno odstupanje od AS.

Označava se: V_σ kod skupa, a V_s kod uzorka.

$V_\sigma=\sigma / \mu \times 100$ za skup

$V_s = \frac{S}{\overline{X}} \cdot 100$ za uzorak

Karakteristike KV:
- daje informacije o homogenosti niza
- olakšava zaključivanje o stepenu varijabilnosti
- preovladava nedostatke apsolutnih MV
- najpreciznija mera varijabilnosti

Omogućava poređenje varijabilnosti:
- dve ili više serija koje imaju različite jedinice mere

- različitih obeležja istih statističkih jedinica
- nizova istih numeričkih obeležja sa istim jedinicama mere, ali sa različitim AS i SD

Vrednost KV od 30,00% predstavlja graničnu vrednost ispod koje se statistički skup smatra homogenim, a iznad heterogenim.

1. PRIMER

$KV_1=0,173$ i $KV_2=0,4178$

Znači, SD prvog niza predstavlja 1,73% AS, a drugog niza 41,78%, Ovo pokazuje da je varijabilnost kod drugog niza znatno veća nego kod prvog.

2.PRIMER

KV=15,7/50,92x100=30,83

<u>STANDARDIZOVANO ODSTUPANJE</u> (standardizov. normalno odstupanje **- SO)**

Standardizovano odstupanje je mera odstupanja vrednosti obeležja *bilo koje* elentarne jedinice skupa ili uzorka u odnosu na sredinu, izraženu u standardnim devijacijma. Obeležava se sa Z.

$Z=X - \mu / \sigma$ za skup

$Z = \frac{X - \overline{X}}{S}$ za uzorak

X – individualne vrednosti obeležja

Karakteristike SO:

- mogu biti pozitivne i negativne
- mogu se sabirati, oduzimati, množiti, deliti
- ne zavise od mernih jedinica
- AS standardizovanih vrednosti je 0
- SD standardizovanih vrednosti je 1
- zbog preciznosti, zaokružuje se na drugu decimalu

Primena SO:

- za prikazivanje položaja bilo koje vrednosti na normalnoj krivoj
- za prikazivanje udaljenosti neke promenljive od AS, izraženo u SD
- za određivanje položaja nekog pojedinačnog rezultata u grupi rezultata ili za razultat pojedinca u različitim merenjima
- za poređenje individualnih rezultata iz dve ili više serija, sa različitim jedinicama mere, različitim SD i AS.

Vrednosti numeričke promenljive retko odstupaju od AS za više o 3SD levo i desno.

Pošto AS i SD u potpunosti definišu raspodelu nekih rezultata, to se za svaki rezultat može izračunati na koji deo SD ona pada, što znači da se samim tim može i saznati koliko ima većih, a koliko manjih od tog rezultata.

1. PRIMER

SO za vrednosti za 39 i 84 su:

Z=39-50,92/15,7=-0,76

Z=84-50,92/15,7=2,11

<u>MERE OBLIKA RASPODELE</u> (Pirsonovi koeficijenti)

Dve statističke serije mogu imati iste AS, varijanse, SD, KV i različite rasporede frekvencija. **Ovo ukazuje na nedostatak MV, koje opisuju varijacije u vrednostima, ali ne pokazuju smer varijacija u odnosu na AS, kao ni stepen koncentracije vrednosti oko AS, tj. oblik rasporeda frekvencija**.

Postoje dva pravca analize oblika:
- ispitivanje simetrije
- ispitivanje visine rasporeda frekvencija

Simetrija se ispituje pomo}u <u>mera asimetrije</u>.

Visina se ispituje pomo}u <u>mera spljoštenosti</u>.

SIMETRIČNOST EMPIRIJSKOG RASPOREDA –

Empirijski raspored je raspored frekvencija formiran evidentiranjem, grupisanjem i prikazivanjem podataka na osnovu konkretnog ststist.istraživanja (anketa, izveštaj, vremenske serija itd.)

Engleski naučnik-statističar **Karl Pirson** (Pearson) je prvi izvršio analizu oblika rasporeda frekvencija. Po njemu se mere za ispitivanje oblika rasporeda frekvencija nazivaju Pirsonovim koeficijentima. Postoje dva koeficijenta: Prvi Pirsonov koef. meri asimetriju odnosno simetriju, Drugi Pirsonov koef. meri spljoštenost.

Za raspored kod koga je AS jednaka Me i Mo, Prvi Pirsonov koeficijent jednak 0 i Drugi Pirsonov koef, jednak 3, kaže se da je normalan i njegov grafik se naziva Gausova kriva (ili zvono).

Vrednost KA je od -3 do +3. Ako je KA=0, raspored je simetričan, AS je jednaka Me. Ako je vrednost KA negativna, raspored je negativno asimetričan i nagnut ulevo, AS je manja od Me. Ako je vrednost KA pozitivna, AS je veća od Me, raspored je pozitivno asimetričan i nagnut udesno. Asimetričan oblik raspodele znači neravnomeran rast ili pad vrednosti obeležja neke pojave oko srednje vrednosti.

Za ispitivanje oblika koriste se ***centralni momenti rasporeda frekvencija***, koji se stavljaju u odnos sa standardnom devijacijom dignutom na odgovarajući stepen. Iz odnosa se dobijaju ***koeficijenti nulte mere***, koji dopuštaju međusobno upoređivanje oblika različitih rasporeda frekvencija.

MOMENTI RASPODELE FREKVENCIJA

U statistici centralni momenti su veličine koje upotpunjuju sliku o raspodeli frekvencija. Momenti su proseci odstupanja individualnih vrednosti numeričkog obeležja od konstante. Najčešće se izračunavaju kao: momenti oko sredine (to su osnovni momenti), momenti oko nule i momenti oko izabrane vrednosti.

Centralni momenti služe za merenje odstupanja vrednosti obeležja od AS skupa dignuto na određeni stepen.

Drugi moment je jednak varijansi, odnosno to je prosečan kvadrat odstupanja od AS. Drugi moment služi kao mera rasturenosti podataka oko AS.

Treći moment oko sredine može poslužiti za merenje načina raspoređivanja podataka. Može biti veći ili manji od nule.

Četvrti moment, kao prosek odstupanja vrednosti obeležja od AS četvrtog stepena, uvek je pozitivan.

Za analizu empirijskog rasporeda ne koriste se nulti i prvi momenti jer su jednaki za sve raspodele i ne daju nikakvu informaciju o samoj raspodeli.

Izračunati momenti služe za računanje koef.asimetrije i koef.spljoštenosti. Pomoću trećeg momenta definiše se koef.asimetrije. Pomoću četvrtog momenta definiše se koef.spljoštenosti.

Treći i četvrti momenti se iskazuju u istim jedinicama mere kao i empirijski podaci na osnovu kojih se računaju, pa su to zato apsolutne mere oblika raspodele.

Koeficijent asimetrije

Asimetrija se meri tzv. prvim Pirsonovim koeficijentom. On se izražava sledećom formulom:

$$\beta_1 = \frac{M_3^2}{M_2^3}$$ M_2, M_3 - momenti

$$M_0 = \frac{\sum (X - \overline{X})^0}{N} = \frac{1}{n}$$ kod neponderisanih podataka

$$M_1 = \frac{\sum (X - \overline{X})^1}{N} = 0$$

$$M_2 = \frac{\sum (X - \overline{X})^2}{n}$$ varijansa

$$M_3 = \frac{\sum (X - \overline{X})^3}{N}$$

$$M_4 = \frac{\sum (X - \overline{X})^4}{N}$$

Simetrični rasporedi frekvencija imaju istu vrednost za sredinu, medijanu i modus. Koeficijent asimetrije je $\beta_1=0$.

Međutim, empirijski rasporedi frekvencija su, po pravilu, asimetrične sa pozitivnom ili negativnom asimetrijom. Tada je koeficijent asimetrije $\beta_1 \neq 0$.

Rasporedi frekvencija sa pozitivnom asimetrijom imaju veći broj frekvencija ispod (levo) od sredine. Kod njih je sredina > medijane i modusa. Koeficijent asimetrije je $\beta_1>0$.

Rasporedi frekvencija sa negativnom asimetrijom imaju veći broj frekvencija iznad (desno) od sredine. Kod njih je sredina < medijane i modusa. Koeficijent asimetrije je $\beta_1<0$.

Na osnovu vrednosti koeficijenta asimetrije se određuje asimetrija rasporeda frekvencija. Ako je vrednost koeficijenta asimetrije u intervalu **(-2,2)** to su tolerantne granice simetričnosti. Van ovog intervala, raspodela je izrazito asimetrična.

Smatra se da je raspored **umereno simetričan** ako su vrednosti koeficijenta od -0,5 do +0,5

Kod neintervalnih rasporeda frekvencija često je sredina > modusa i medijane koji su jednaki,tj. sredina>medijana=modus.
Važi i obratno: sredina<medijana=modus

Koeficijent asimetrije rasporeda relativnih frekvencija se određuje na isti način kao i koeficijent asimetrije rasporeda frekvencija.

Koeficijent spljoštenosti

Ovo je drugi Pirsonov koeficijent. Predstavlja relativnu meru spljoštenosti rasporeda frekvencija. Dat je formulom:

$$\beta_2 = \frac{M_4}{M_2^2}$$

Kod rasporeda frekvencija sa ″normalnom″ visinom, $\beta_2=3$. Ako je vrednost $\beta_2 \neq 3$ raspored frekvencija nema ″normalnu″ visinu.

Raspored frekvenija je izdužen (raspodela je šiljata) u odnosu na ″normalu″ visinu kada je $\beta_2>3$, a spljošten (raspodela je tupa) u odnosu na ″normalnu″ visinu kada je $\beta_2<3$.

Za procenu homogenosti podataka može poslužiti zaobljenost vrha krive raspodele. Vrednost je oko 3 kada raspodela nije značajno različita od normalne. Oštrije raspodele imaju veću koncentraciju podataka oko AS od normalne raspodele (homogeniji podaci). Tupe raspodele imaju manju koncentraciju podataka oko AS od normalne raspodele (heterogeniji podaci).

Koeficijent spljoštenosti rasporeda relativnih frekvencija se određuje na isti način kao i koeficijent spljoštenosti rasporeda frekvencija.

5. TEORIJSKE RASPODELE VEROVATNOĆE

U poslovnom odlučivanju veliki je problem predvideti ostvarenje neke pojave, ukloniti neizvesnost i rizik od neuspeha. U analizi neizvesnosti i uklanjanja ili bar smanjenja rizika razvio se naučni pristup problemu – osnova odlučivanja i zaključivanja je teorija verovatnoće i na njoj razvijena statistička metodologija.

Osnovni pojam u teoriji verovatnoće je **slučajni događaj**. To je svaka činjenica ili rezultat koji se u izvođenju nekog pokušaja može ostvariti ili neostvariti.

Ti pokušaji mogu biti deterministički ili stohastički. Kod **determinističkih** postoji samo jedan događaj, jedan rezultat. Ti događaji nisu slučajni i njih teorija verov. ne proučava. Kod **stohastičkih**, slučajnih – događaj se može ostvariti ili neostvariti i njih teorija verov. proučava.

Svaki pokušaj mora biti definisan – dati su uslovi izvođenja i predviđeno je šta treba da bude rezultat.

U teoriji verovatnoće se proučava ostvarenje ili neostvarenje događaja. Neodređenost ostvarenja događaja može biti nejednaka, pa se zbog toga može postaviti numerička gradacija verovatnoće događaja. Svaka verovatnoća je data relacijom stohastičkog tipa.

Svaki pokušaj ima određeni broj mogućih rezultata. Svaki rezultat je elementarni događaj. Skup svih elementarnih događaja predstavlja **potpuni skup događaja**.

Postoji nekoliko vrsta događaja:

- **sigurni događaj** (koji se sigurno ostvaruje pri nekom pokušaju)
- **nemoguć događaj** (koji se ne može ostvariti pri nekom pokušaju)
- **slučajan događaj** (koji se može, a ne mora ostvariti pri nekom pokušaju)
- **međusobno isključivi događaji** (koji se ne mogu desiti istovremeno pri nekom pokušaju)
- **događaji koji se međusobno ne isključuju** (koji se mogu desiti istovremeno pri nekom pokušaju)

Vrste slučajnih događaja:

- elementarni
- složen

Verovatnoća se različito definiše.

Ako se analizira **verovatnoća ostvarenja jednog događaja**, verovatnoća se deli na **klasičnu, empirijsku, statističku, geometrijsku** i **aksiomatsku**.

Ako se radi o **uporednoj analizi dva ili više događaja**, verovatnoća se deli na **totalnu, složenu, uslovnu** i **Bajesovu.**

Ako se definiše sa stanovišta **broja ponavljanja**, verovatnoća se deli na **binomnu, Poissonovu** i **hipergeometrijsku**.

<u>Klasična verovatnoća</u>

Bazira se na posmatranju podjednako verovatnih događaja. Naziva se <u>verovatnoća a priori</u>, što znači da se posmatraju podjednako mogući slučajevi i da su verovatnoće unapred poznate. Javlja se kod svih igara na sreću, gde svi događaji imaju istu mogućnost da se dese pri nekom pokušaju.

Klasičnu definiciju verovatnoće je dao ***Laplas*** (Laplace). Ona glasi:

″verovatnoća događaja A [P(A)]je odnos između broja povoljnih (m) i ukupnog broja svih podjednako mogućih slučajeva (N) za neki događaj. Mera ove verovatnoće se izražava razlomkom čiji je brojilac broj povoljnih slučajeva, a imenilac broj svih mogućih slučajeva″:

$$P(A)=m / N$$

1) Kod sigurnog događaja je m = N, pa je: $P(A)=N / N = 1$

pa se kaže da je karakterističan broj izvesnosti ostvarenja očekivanog događaja A jedinica.

2) Kod nemogućeg događaja je m = 0, pa je:

$$P(A)=0 / N = 0$$

pa se kaže karakterističan broj nemogućnosti ostvarenja očekivanog događaja A nula.

3) Kod mogućeg događaja je $0 < m < N$

4) Verovatnoća je uvek ***nenegativna***, tj. uvek je $P(A) \geq 0$.

5) Verovatnoća ostvarenja očekivanog događaja A je **realan broj** u granicama **(0,1), tj.** $0 \leq P(A) \leq 1$.

<u>Statistička verovatnoća</u>

Pored verovatnoće a priori (znači da se posmatraju podjednako mogući slučajevi i da su verovatnoće unapred poznate), postoji i <u>verovatnoća a posteriori</u>. To je verovatnoća do koje se dolazi na osnovu prethodnog posmatranja. Za njeno definisanje se koristi dopunjena definicija nemačkog matematičara ***Misesa*** (Mizes).

Neka se u N ponovljenih pokušaja f puta ostvario događaj A. Odnos f / N je relativna frekvencija događaja A. Granična vrednost relativne frekvencije događaja A kad broj ponovljenih pokušaja neograničeno raste je:

$$P(A) = p = \lim_{N \to \infty} \frac{f}{N}$$ i to je **totalna relativna frekvencija.**

Ako postoji granična vrednost, događaj A ima određenu totalnu frekvenciju. Totalna frekvencija izražava stabilnost frekvencije ostvarenja događaja A. To je *statistička (empirijska) verovatnoća.*

Njena definicija, na osnovu Misesove koncepcije, glasi:

u skupu svih nezavisnih ponovljenih pokušaja, verovatnoća ostvarenja događaja A je njegova totalna relativna frekvencija p, koja ostaje nepromenjena ako se na proizvoljan način odstrani jedan deo nezavisnih pokušaja.

Sva proučavanja statističke verovatnoće vrše se na masovnim pojavama ili događajima koji se ponavljaju. Osnova joj je Zakon velikih brojeva (ZVB), koji je formulisao francuski statističar Puason (Poisson). ZVB dopošta određene generalizacije u zaključivanju, u smislu da je uticaj slučajnosti pri zaključivanju utoliko manji ukoliko se zaključivanje vrši na osnovu posmatranja većeg broja događaja, odnosno da su statistički parametri utoliko bliži svojim "graničnim vrednostima" ukoliko je obuhvaćen veći broj slučajeva ili izvršen veći broj pokušaja.

Nedostaci statističke verovatnoće:

- njeno značenje se poistovećuje sa načinom njenog određivanja
- ona se, kod događaja koji se javlja samo jednom, izjednačava sa relativnom frekvencijom

Slučajna promenljiva

Slučajna promenljiva je kvantitativna ili numerička veličina koja u rezultatu pokušaja može uzeti jednu od mogućih vrednosti sa pripadajućom verovatnoćom.

Vrste slučajnih promenljivih:

- prekidna ili diskretna
- neprekidna ili kontinuirana

Prekidna slučajna promenljiva x uzima vrednosti niza x_1, x_2, ...x_k sa odgovarajućim verovatnoćama p_1, p_2, ...p_k pri čemu je p_1, p_2, ...p_k=1. ***Prekidna slučajna promenljiva*** uzima konačan broj prekidnih vrednosti.

Neprekidna slučajna promenljiva x uzima bilo koju vrednost zatvorenog intervala i može se neprekidno raspoređivanti duž tog intervala.Verovatnoća da će se vrednost neprekidne slučajne promenljive naći u intervalu (a,b) je:

$$P(a \le X \le b) = \int_a^b f(X)dx = 1$$

Za slučaj da neprekidna slučajna promenljiva varira u razmaku (-∞, +∞) biće:

$$P(-\infty \le X \le +\infty) = \int_{-\infty}^{+\infty} f(X)dx = 1$$

<u>Zakon verovatnoće</u>

Zakon verovatnoće slučajne promenljive je odnos između mogućih vrednosti slučajne promenljive i odgovarajućih verovatnoća.

Zakon verovatnoće prekidne slučajne promenljive se definiše kao skup parova vrednosti niza x_1, x_2, ...x_k i odgovarajućih verovatnoća, pri čemu je zbir svih verovatnoća 1 (jedinica). Zakon verovatnoće prekidne slučajne promenljive se daje u vidu tabele:

x	x_1 x_2 ... x_k
p	p_1 p_2 ... p_k

gde je: $\sum_{i=1}^{k} p_i = 1$

Zakon verovatnoće (ZV) prekidne slučajne promenljive mora da ispunjava dva uslova:

1) $0 \leq p \leq 1$

2) $\sum_{i=1}^{k} p_i = 1$

1. PRIMER

ZV pojavljivanja strane kocke sa brojem 5 kod bacanja dve kocke je:

x	0	1	2
p	25/36	10/36	1/36

gde je p=25/36+10/36+1/36=1

2.PRIMER

ZV zbira brojeva koji se mogu dobiti kod bacanja dve kocke je:

x	2	3	4	5	6	7	8	9	10	11	12
p	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

gde je p=1/36+2/36+3/36+4/36+5/36+6/36+5/36+4/36+3/36+2/36+1/36=1

Zakon verovatnoće se može **<u>grafički interpretirati</u>**. Najčešće preko histograma, tako što se na apscisu nanose vrednosti slučajne promenljive, a na ordinatu verovatnoće.

Zakon verovatnoće neprekidne slučajne promenljive je funkcija f(x), gde je x specifična vrednost slučajne promenljive X. Za svaku vrednost neprekidne slučajne promenljive X intervala (a,b) funkcija f(x) je veća od 0. Interval (a,b) se može podeliti na veliki broj jednakoh podintervala. Kad broj podintervala neograničeno raste, tada niz poređanih pravougaonika u histogramu teži graničnoj krivoj f(x), odnosno:

$$f(x) = \lim_{\Delta x \to 0} \frac{p(x \leq X \leq x + \Delta x}{\Delta x}$$

Verovatnoća da će se neprekidna slučajna promenljiva X naći u intervalu (x, x+Δx) jednaka je površini između krive i ose X i ograničena je ordinatama podignutim iz x i x+Δx, tj.

Verovatnoća da će se neprekidna slučajna promenljiva naći u intervalu (a,b) je:

$$P(a \leq x \leq b) = \int_a^b f(x)dx = 1$$

Zakon verovatnoće neprekidne slučajne promenljive, odnosno funkcija f(x), mora da ispuni tri uslova:

1) $0 \le f(x) \le 1$

2) $\int_{-\infty}^{+\infty} f(x)dx = 1$

3) $\int_{x_1}^{x_2} f(x)dx = P(x_1 \le X \le x_2)$ gde su x_1 i x_2 bilo koje vrednosti neprekidne slučajne promenljive X, pri čemu je $x_1 < x_2$, odnosno $x_2=x_1+\Delta x$.

Ovaj treći uslov se javlja samo kod neprekidne slučajne promenljive. On znači da kod neprekidne slučajne promenljive verovatnoća pripada samo intervalima, a pojedinačnim vrednostima x_i pripada verovatnoća 0, jer se nad tačkom ne formira površina.

Razlika prekidne i neprekidne slučajne promenljive:

- kod prekidne slučajne promenljive svakoj njenoj vrenosti pripada odgovarajuća verovatnoća; kod neprekidne slučajne promenljive verovatnoća pripada intervalima
- pošto je skup verovatnoća koje se mogu naći u intervalu neprebrojiv, nemoguće je elementarno prikazati zakon verovatnoće neprekidne slučajne promenljive

Funkcija raspodele (ili rasporeda) verovatnoća

Funkcija raspodele je jedina mogućnost da se prikažu univerzalne karakteristike i prekodne i neprekidne slučajne promenljive. Do tih univerzalnih karakteristika se dolazi proučavanjem verovatnoća $P(X < x)$.

Ako je x je određena vrednost slučajne promenljive X. Verovatnoća da će slučajna promenljiva X biti jednaka ili manja od x je funkcija od x, tj.

$$\mathbf{F(x)=P(X \le x)}$$

Funkcija F(x) je funkcija raspodele verovatnoća. Pun naziv - kumulativna funkcija raspodele verovatnoća slučajne promenljive. F(x) je neopadajuća funkcija po x. Definisana je za svako x u intervalu $(-\infty, +\infty)$. Njena vrednost je uvek između 0 i 1 (jer je to verovatnoća). Granične vrednosti su:

$$\mathbf{F(-\infty) = P(X \le -\infty) = \lim_{X\to+\infty} F(x) = 0}$$

$$\mathbf{F(+\infty) = P(X \le +\infty) = \lim_{X\to+\infty} F(x) = 1}$$

Kod prekidne slučajne promenljive X koja uzima vrednosti $x_1, x_2, ..., x_k$ funkcija raspodele je:

$$\mathbf{F(x)= P(X \le x) = P(X = x_1) + P(X = x_2) +...+ P(X = x_k)} = \sum_{i=1}^{k} P(X = x_i) = 1$$

Dijagram f-je F(x) prekidne slučajne promenljive je stepenastog oblika (kada prekidna slučajna promenljiva X prolazi preko neke od mogućih vrednosti x_i, F-ja raspodele F(x) se skokovito menja sa veličinom skokovite promene koja je jednaka verovatnoći vrednosti x_i. Između ma koje dve uzastopne vrednosti, f-ja F(x) je konstantna. Funkcija F(x) predstavlja ustvari kumulativ relativnih frekvencija).

Verovatnoće prekidne slučajne promenljive X koja uzima vrednosti $x_1, x_2, ..., x_k$ mogu se odrediti iz funkcije F(x).

$$\mathbf{P(X = x_i) = P(X \le x_i) - P(X \le x_{i-1}) = F(x_i)-F(x_{i-1})}$$

Ako verovatnoće označimo sa:

$$\mathbf{P_1 = F(X = x_i)}$$

dobija se: $\mathbf{P_1 = F(x_1)}$

$\mathbf{P_2 = F(x_2) - F(x_1)}$

$\mathbf{P_3 = F(x_3) - F(x_2)}$

....

$\mathbf{P_x = 1 - F(x_{k-1})}$

Kod <u>neprekidne</u> slučajne promenljive X sa zakonom verovatnoće f(x) funkcija raspodele verovatnoća je:

$$\mathbf{F(-\infty < X \leq x) = F(x)} = \underline{\int_{-\infty}^{x} f(x)dx}$$

Granične vrednosti su:

$$\mathbf{F(-\infty) = 0} \qquad \text{i} \qquad \mathbf{F(+\infty)} = \underline{\int_{-\infty}^{+\infty} f(x)dx = \mathbf{1}}$$

Zakon verovatnoće neprekidne slučajne promenljive X se određuje na osnovu f-je raspodele verovatnoća, jer je zakon verovatnoće primitivna f-ja funkcije raspodele verovatnoća, tj.:

$\mathbf{f(x) = F'(x)}$

Prvi izvod funkcije raspodele verovatnoća je zakon verovatnoće neprekidne slučajne promenljive X.

<u>ČEBIŠEVLJEVA TEOREMA</u>

Ruski matematičar Čebišev je dao teoremu za određivanje verovatnoća da će se slučajna promenljiva X naći u intervalu oko aritmetičke sredine (očekivane vrednosti).

Teorema glasi:

kod slučajne promenljive X čiji su parametri μ (sredina) i σ^2 (varijansa) za svako proizvoljno ε ("epsilon") ispinjena je nejednačina: $P(|X-\mu| \geq \varepsilon) \leq \sigma^2/\varepsilon^2$
Verovatnoća da će se slučajna promenljiva X naći u intervalu (μ-z σ, μ+z σ) pri čemu je z proizvoljno odabran broj, je: $P(\mu - z\sigma < X \leq \mu + z\sigma) \geq 1 - 1/z^2$

Mnoge raspodele podataka imaju vrednosti simetrično raspoređene oko aritmetičke sredine. To su raspodele u obliku zvona i za njih važi **empirijsko pravilo**:

- oko 68% svih podataka leži unutar 1 standardne devijacije oko sredine, tj. **(μ±1σ)**
- oko 95% svih podataka leži unutar 2 standardne devijacije oko sredine, tj. **(μ±2σ)**
- oko 99,7% svih podataka leži unutar 3 standardne devijacije oko sredine, tj. **(μ±3σ)**

<u>MODELI TEORIJSKIH RASPODELA VEROVATNOĆA</u>

U ovom slučaju **model** predstavlja **funkcionalnu vezu između vrednosti slučajne promenljive i odgovarajućih verovatnoća, vezu koja je definisana određenim tipom funkcije**.

Za **<u>prekidnu</u>** slučajnu promenljivu X, model raspodele je funkcionalna veza između vrednosti $\mathbf{x_1, x_2, ..., x_n}$ i odgovarajućih verovatnoća $\mathbf{p_1, p_2, ..., p_n}$, odnosno:

$\mathbf{p_i = f(x_i)}$ gde je i=1,2...n

Za **neprekidnu** slučajnu promenljivu X, funkcionalna veza se svodi na zakon verovatnoće definisan funkcijom **f(x)** za svako **x** intervala **(a,b)**.

Teorijske raspodele se dele na:

- **prekidne** (uniformna raspodela, Bernulijeva raspodela, binomna raspodela, Puasonova raspodela)
- **neprekidne** (normalna raspodela, Studentova T-raspodela, Hi-kvadrat raspodela)

Uniformna raspodela

Prekidna slučajna promenljiva X može uzeti konačno mnogo vrednosti (1,2,...n). Zakon verovatnoće izražava podjednaku verovatnoću za sve vrednosti slučajne promenljive.

Zakon verovatnoće je:

X	1	2	...	n
p	1/n	1/n	...	1/n

Uniformna (ravnomerna) raspodela je zasnovana na igrama na sreću. Izražava se (za **prekidn**u slučajnu promenljivu) kao:

$$P(X = x) = p_x = \frac{1}{(b-a)+1}$$

gde su:

a - početna vrednost slučajne promenljive

b- poslednja vrednost slučajne promenljive

Parametri uniformne raspodele su:

1) sredina

$$\overline{X} = \mu = \frac{a+b}{2} \qquad ili \qquad \overline{X} = \mu = \sum_{i=1}^{n} x_i \cdot p_i$$

2) varijansa

$$\sigma^2 = \frac{((b-a)+1)^2 - 1}{12} \qquad ili \qquad \sigma^2 = \sum_{i=1}^{n} x_i^2 \cdot p_i - \mu^2 = \sum_{i=1}^{n} x_i^2 \cdot p_i - \overline{X}^2$$

1. PRIMER

Označena strana kocke je naša slučajna promenljiva X. Zakon verovatnoće je:

X	1	2	3	4	5	6
p	1/6	1/6	1/6	1/6	1/6	1/6

ili: a=1, b=6

$$P_x = p_x = \frac{1}{(b-a)+1} = \frac{1}{(6-1)+1} = \frac{1}{6}$$

sredina je:

$$\overline{X} = \mu = \frac{a+b}{2} = \frac{1+6}{2} = 3.5 \qquad ili$$

$$\overline{X} = \mu = \sum_{i=1}^{n} x_i \cdot p_i = \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 2 + \frac{1}{6}\cdot 3 + \frac{1}{6}\cdot 4 + \frac{1}{6}\cdot 5 + \frac{1}{6}\cdot 6 = \frac{21}{6} = 3.5$$

varijansa je:

$$\sigma^2 = \frac{((b-a)+1)^2 - 1}{12} = \frac{((6-1)+1)^2 - 1}{12} = 2.92 \qquad ili$$

$$\sigma^2 = \sum_{i=1}^{n} x_i^2 \cdot p_i - \mu^2 = \sum_{i=1}^{n} x_i^2 \cdot p_i - \overline{X}^2 = \left(\frac{1}{6} + \frac{4}{6} + \frac{9}{6} + \frac{16}{6} + \frac{25}{6} + \frac{36}{6}\right) - 3.5^2 = 15.17 - 12.25 = 2.92$$

Uniformna raspodela se može definisati i za **neprekidnu** slučajnu promenljivu X koja može uzeti vrednosti iz intervala (a,b). Tada je zakon verovatnoće:

f(x)=1 / b-a

Parametri:

sredina: μ= a+b / 2 varijansa: σ^2= $(b-a)^2$ / 12

<u>Bernulijeva raspodela</u>

Naziv je dobila po švajcarskom matematičaru Bernuliju (Bernoulli).

Prekidna slučajna promenljiva može uzeti vrednosti 0 ili 1. Verovatnoća da uzme vrednost **0** je **q**, a verovatnoća da uzme vrednost **1** je **p** (pri ovom važi da je p+q=1).

Zakon verovatnoće je:

X	0	1
p	q	p

ili:

$p_x = p^x\, q^{n-x}$ za x=0 i x=1, odnosno za (n+1) mogućih vrednosti slučajne promenljive (gde je n=1), gde je 0 < p < 1 i q = 1 - p.

Parametri Bernulijeve raspodele su:

1) sredina

$\mu = p$

2) varijansa

$\sigma^2 = p\, q$

1. PRIMER

Slučajna promenljiva ima Bernulijevu raspodelu kod koje je p = 0,8

a) odrediti zakon verovatnoće

b) izračunati sredinu

c) izračunati varijansu i standardnu devijaciju

Rešenje:

a) p = 0.8

q = 1-0.8=0.2

$p_x = p^x\, q^{n-x} = 0.8^x\, 0.2^{1-x}$

b) μ = p = 0.8

c) σ^2 = p q = (0.8) (0.2)=0.16

stand.dev.:

$\sigma = +\sqrt{0.16} = 0.4$

2. PRIMER

Slučajna promenljiva ima Bernulijevu raspodelu kod koje je p = 0,6

a) odrediti zakon verovatnoće

b)izračunati sredinu

c) izračunati varijansu i standardnu devijaciju

Rešenje:

a) p = 0.6

q = 1-0.6=0.4

X	0	1
p	0.4	0.6

ili:

$p_x = p^x q^{n-x} = 0.6^x \, 0.4^{1-x}$

b) $\mu = p = 0.6$

c) $\sigma^2 = p\, q = (0.6)\,(0.4) = 0.24$

stand.dev.:

$\sigma = +\sqrt{0.24} =$

Binomna raspodela

Binomna slučajna promenljiva je prekidna sa **n+1** mogućih vrednosti (n je prirodan broj).

Zakon verovatnoće je:

X	0	1	...	n
p	q^n	$n\, p\, q^{n-1}$	...	p^n

ili:

$p_x = \quad p^x q^{n-x}$ gde je p+q=1

gde je:

- binomni koeficijent =n! / x! (n-x)! (=1)

Ako je n veliko, zakon verovatnoće se izražava pomoću logaritma. Tada je:

$\log p_x = \log \quad + x \log p + (n-x) \log q$

Parametri binomne raspodele su:

1) sredina

$\mu = n\, p$

2) varijansa

$\sigma^2 = n\, p\, q$

3) koeficijent asimetrije

$\beta_1 = (q-p)^2 / n\, p\, q$

Kada je p = q = 0,5 binomna raspodela je simetrična.

Ako je $p < q$ raspodela je pozitivno asimetrična.

Ako je $p > q$ raspodela je negativno asimetrična.

Binomna raspodela je uvek unimodalan.

Ako $n \to \infty$ koeficijent asimetrije $\beta_1 \to 0$.

Znači, sa porastom n, binomna raspodela smanjuje svoju asimetriju.

4) koeficijent spljoštenosti

β_2 =3+ (1- 6 p q) / n p q

Kada n $\rightarrow \infty$ koeficijent spljo{tenosti $\beta_2 \rightarrow 3$, tj. normalnoj visini

Često nije lako izračunati verovatnoću binomne raspodele. Zato se koristi ***rekurentna veza*** između prethodne i naredne verovatnoće: $P_{x+1} = \frac{n-x}{x+1} \cdot \frac{p}{q} \cdot P_x$

$p_0 = q^n$

$p_1 = n \cdot \frac{p}{q} \cdot p_0$ i tako dalje.

1.PRIMER

Binomna slučajna promenljiva uzima vrednosti 0,1,2,3,4,5. Parametar binomne raspodele je: p=0,6.

Rešenje:

a) jasno je da je n=5

q=1-p=0,4

(kod Bernulijeve raspodele smo rekli da je verovatnoća da slučajna promenljiva uzme vrednost 0 je q, a vrednost 1 je p; a važi p+q=1, pa je q=1-p)

Verovatnoće su:

- za x=0

$p=q^n = 0.4^5 = 0.01024$

- ostale verovatnoće lakše dobijamo is rekurentnog obrasca:

$$p_1 = \frac{p}{q} \cdot n \cdot p_0 = \frac{0.6}{0.4} \cdot 5 \cdot 0.01024 = 0.0768$$

$$p_2 = \frac{p}{q} \cdot \frac{n-1}{2} \cdot p_1 = \frac{0.6}{0.4} \cdot \frac{5-1}{2} \cdot 0.0768 = 0.2304$$

$$p_3 = \frac{p}{q} \cdot \frac{n-2}{3} \cdot p_2 = \frac{0.6}{0.4} \cdot \frac{5-2}{3} \cdot 0.2304 = 0.3456$$

$$p_4 = \frac{p}{q} \cdot \frac{n-3}{4} \cdot p_3 = \frac{0.6}{0.4} \cdot \frac{5-3}{4} \cdot 0.3456 = 0.2592$$

$$p_5 = \frac{p}{q} \cdot \frac{n-4}{5} \cdot p_4 = \frac{0.6}{0.4} \cdot \frac{5-4}{5} \cdot 0.2592 = 0.07776$$

Očigledno je da je:

$$\sum_{i=1}^{n} p_i = p_0 + p_1 + p_2 + p_3 + p_4 + p_5 = 0.01024 + 0.0768 + 0.2304 + 0.3456 + 0.2592 + 0.07776 = 1$$

Parametri raspodele:

1) sredina: μ=n p = 5 x 0.6=3

2) varijansa: σ^2=n p q = 5 x 0.6 x 0.4 = 1.2

3) koeficijent asimetrije: $\beta_1 = \frac{(q-p)^2}{npq} = \frac{(0.4-0.6)^2}{5 \cdot 0.6 \cdot 0.4} = 0.286$

4) koeficijent spljoštenosti: $\beta_2 = 3 + \frac{1-6nq}{npq} = 3 + \frac{1-6 \cdot 0.6 \cdot 0.4}{5 \cdot 0.6 \cdot 0.4} = 2.63$

<u>Puasonova raspodela</u> (francuski statističar Poisson)

Kod ove raspodele, slučajna promenljiva ima prekidne vrednosti, koje mogu biti 0 ili ma koji prirodan broj.

Zakon verovatnoće je:

X	0	1	2	...
p	e^{-m}	$m\ e^{-m}$	$(m^2/2)\ e^{-m}$	...

ili:

$p_x = e^{-m} \cdot \frac{m^x}{x!}$ za svako x=0,1,2... gde je: m=n p i to je konstanta, x=0,1,2...

(često ćete naći umesto m oznaku λ - lambda)

Parametri Puasonove raspodele:

1) sredina: $\mu=m$

2) varijansa: $\sigma^2=m$

3) koeficijent asimetrije: $\beta_1 = \frac{1}{m}$

Poisonov raspored je asimetričan. Kad je $\beta_1>0$ du`i krak rasporeda je sa desne strane. Asimetrija se smanjuje sa porastom n.

4) koeficijent spljoštenosti: $\beta_2 = 3 + \frac{1}{m}$

Kada $n \to \infty$ koeficijent spljo{tenosti te`i normalnoj visini ($\beta_2 \to 3$)

Puasonovu raspodelu je teško direjtno izračunati, pa se koristi rekurentna formula (veza između prethodne i naredne vrednosti):

$$p_{x+1} = \frac{m}{x+1} \cdot p_x$$

1.PRIMER

Puasonova slučajna promeljiva ima vrednosti 0,1,2,3,.... , i m=1.4

Rešenje:

verovatnoća p_0 je: $p_0 = e^{-m} \cdot \frac{m^x}{x!} = e^{-1.4} \cdot \frac{1.4^0}{0!} = e^{-1.4} = 0.24660$

ostale verovatnoće se računaju preko rekurentnog obrasca:

$$p_1 = \frac{m}{1} \cdot p_0 = 1.4 \cdot 0.24666 = 0.34524$$

$$p_2 = \frac{m}{2} \cdot p_1 = \frac{1.4}{2} \cdot 0.34524 = 0.24167$$

$$p_3 = \frac{m}{3} \cdot p_2 = \frac{1.4}{3} \cdot 0.24167 = 0.11278$$

$$p_4 = \frac{m}{4} \cdot p_3 = \frac{1.4}{4} \cdot 0.11278 = 0.03947$$

$$p_5 = \frac{m}{5} \cdot p_4 = \frac{1.4}{5} \cdot 0.03947 = 0.01105$$

$$p_6 = \frac{m}{6} \cdot p_5 = \frac{1.4}{6} \cdot 0.01105 = 0.00257$$

$$p_7 = \frac{m}{7} \cdot p_6 = \frac{1.4}{7} \cdot 0.00257 = 0.00051$$

$p_8 = \frac{m}{8} \cdot p_7 = \frac{1.4}{8} \cdot 0.00051 = 0.00009$

$p_9 = \frac{m}{9} \cdot p_8 = \frac{1.4}{9} \cdot 0.00009 = 0.00001$

$p_{10} = \frac{m}{10} \cdot p_9 = \frac{1.4}{10} \cdot 0.00001 = 0.00001$

Puasonovih raspodela ima beskonačno mnogo, one su manje ako x raste, tj. kada $x \to \infty$ tada $p \to 0$.

Parametri raspodele su:

1) sredina: $\mu = m = 1.4$

2) varijansa: $\sigma^2 = m = 1.4$

3) koeficijent asimetrije: $\beta_1 = \frac{1}{m}$ =1/1.4=0.71

4) koeficijent spljoštenosti: $\beta_2 = 3 + \frac{1}{m}$ =3+1/1.4=3.71

Normalna raspodela (matematičku osnovu normalne raspodele je dao engleski matematičar Abraham de Moivre (1667-1754) i to kao granični slučaj binomne raspodele, pronašao je francuski naučnik Laplas, a detaljno je proučio i izveo matematičku formulu nemački matematičar Karl Fridrih Gauss (1777-1855), pa se često naziva Gausova raspodela ili Gausovo zvono).

Ova raspodela se najviše koristi u statističkom zaključivanju iz sledećih razloga:

- Veliki broj masovnih varijabilnih pojava, prema centralnoj graničnoj teoremi, neograničeno teži normalnom rasporedu i ima karakteristike norm.raspodele.
- Skoro sve prekidne raspodele (Bernulijeva, binomna i Puasonova) teže norm.raspodeli kada broj realizacija neograničeno raste.
- Na osnovu zakona norm.raspodele razvijen je niz neprekidnih raspodela, kao Studentova, Hi-kvadrat, Snedekerova, Fišerova itd.
- Veći deo statističkih zaključaka može se izvesti samo uz pretpostavku da je skup iz koga je izvučen uzorak normalno raspoređen. Ako to nije i ako je skup heterogen, onda se skup deli na homogene podskupove koji su normalno raspoređeni.
- Normalna raspodela je osnov za određivanje intervala poverenja parametara skupa na osnovu velikog uzorka, kada je poznata i nepoznata varijansa skupa; za određivanje parametara u regresionoj i korelacionoj analizi; u analizi vremenskih serija.

Slučajna promenljiva je **neprekidna** promenljiva X koja uzima vrednosti od $-\infty$ do $+\infty$, tj.

$-\infty \le X \le +\infty$

Funkcija f(x) predstavlja **zakon verovatnoće** ove raspodele:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Funkcija f(x) je parna funkcija u odnosu na μ; pozitivna je za $-\infty < X < +\infty$; maksimum ima za $X=\mu$; prevojne tačke su $X=\pm\sigma$; ukupna površina ispod normalne krive jednaka je jedinici; ona je simetrična (Gausovo zvono) u odnosu na pravu $X=\mu$;

dijagram se nalazi iznad x-ose koja predstavlja asimptotu funkcije f(x) i sa leva i sa desna; ona je unimodalna, tj. μ=medijana=modus; maksimum ima u tački X=μ, tj. $\frac{1}{\sigma\sqrt{2\pi}} = 0.3989$.

Parametri od kojih zavisi ova raspodela su aritmetička sredina i varijansa, pa se normalna raspodela označava kao **N(μ,σ²)**

Parametri raspodele su:

1) sredina: $\mu = \int_{-\infty}^{+\infty} x\, f(x)\, dx$

2) varijansa: $\sigma^2 = \int_{-\infty}^{+\infty} x^2\, f(x)\, dx$

3) koeficijent asimetrije: $\beta_1 = 0$

značı da je ova raspodela idealno simetrična.

4) koeficijent spljoštenosti: $\beta_2 = 3$

znači da ova raspodela ima normalnu visinu.

Normalna raspodela koja uvek im aisti oblik, uvek istu vrednost parametara naziva se standardizovanom normalnom raspodelom. Do nje se dolazi tako što se uvodi nova slučajna promenljiva Z, koja se naziva standardizovanom slučajnom promenljivom:

$$Z = \frac{x-\mu}{\sigma}$$

Parametri standardizovane normalne raspodele su:

1) sredina: $\mu=0$

2) varijansa: $\sigma^2=1$

3) koeficijent asimetrije: $\beta_1 = 0$

znači da je ova raspodela unimodalna i idealno simetrična.

4) 4) koeficijent spljoštenosti: $\beta_2 = 3$

znači da ova raspodela ima normalnu visinu.

Funkcija standardizovane normalne raspodele neprekidne slučajne promenljive Z je:

$$F(Z) = \int_{-\infty}^{Z} f(Z)\, dz$$

funkcija F(Z) je kumulativ verovatnoća; za razne vrednosti Z, vrednosti f(Z) su date u tablicama.

Verovatnoća da će Z uzeti neku vrednost u razmaku (Z_1, Z_2) je:

$$\mathbf{p(x_1 \le x \le x_2) = p(Z_1 \le Z \le Z_2) = F(Z_2) - F(Z_1)}$$
$$\mathbf{p(x \le x_1) = p(Z \le Z_1) = F(Z_1)}$$
$$\mathbf{p(x \ge x_1) = p(Z \ge Z_1) = 1 - p(Z \le Z_1) = 1 - F(Z_1)}$$
$$\mathbf{p(-x_0 \le x \le x_0) = p(Z_0 \le Z \le Z_0) = p(Z_0) - (1 - p(Z_0)) = p(Z_0) - 1 + p(Z_0) = 2\,F(Z_0) - 1}$$

Izračunavanje verovatnoća standardizovane normalne raspodele se vrši iz funkcije raspodele F(Z). U statističkoj praksi se najviše za tu svrhu koristi tzv. ″σ-pravilo″ (sigma pravilo).

- za normalnu raspodelu sigma pravilo je:

$p(\mu - \sigma \le x \le \mu + \sigma) = 0.6826$

$p(\mu - 2\sigma \le x \le \mu + 2\sigma) = 0.9545$

$p(\mu - 3\sigma \le x \le \mu + 3\sigma) = 0.9973$

znači da se na rastojanju od 1,2 i 3 standardne devijacije levo i desno od sredine nalazi se 69,26%, 95,45% i 99,73% svih vrednosti normalne neprekidne slučajne promenljive x.

- za standardizovanu normalnu raspodelu sigma pravilo je:

$p(-1 \le Z \le 1)=F(1)-F(-1)= 0.8413 - 0.1587=0.6826$

$p(-2 \le Z \le 2)=F(2)-F(-2)=0.9773 - 0.0228=0.9545$

$p(-3 \le Z \le 3)=F(3)-F(-3)=0.99865 - 0.00135=0.9973$

Iz funkcije raspodele mogu se izračunati verovatnoće za bilo koje vrednosti Z_1 i Z_2 i obratno, mogu se odrediti vrednosti Z_1 i Z_2 ako je verovatnoća već data. Isto tako, Z_1 i Z_2 se mogu odrediti ako su simetrični u odnosu na koordinatni početak.

Obeležavanje:

normalna raspodela: **$N(\mu,\sigma^2)$** (μ-srednina, σ^2-varijansa)

standardizovana normalna raspodela: **N(0, 1)**

PRIMER 1.

data je standardizovana normalna raspodela; naći verovatnoću ako je Z_1=-1.16 i Z_2=0.48.

Rešenje:

$p(-1.16 \le Z \le 0{,}48)=F(0.48)-F(-1.16)=0.6844-0.1230=0.5614$ ili 56.14%

PRIMER 2.

a) data je normalna raspodela $N(\mu,\sigma^2)$ sa parametrima μ=40 i σ^2=2 ,tj. N(40,2). Neka su vrednosti x_1=35 i x_2=45 simetrične u odnosu na μ=40.

Rešenje:

prvo se traže Z_1 i Z_2:

$$Z_1=\frac{x_1-\mu}{\sigma}=\frac{35-40}{2}=-2.5$$

$$Z_2=\frac{x_2-\mu}{\sigma}=\frac{45-40}{2}=2.5$$

verovatnoća je:

$p(35 \le Z \le 45)=p(-2.5 \le Z \le 2.5)=F(2.5) - F(-2.5)=0.9938 - 0.0062=0.9876$

b)Kako se određuju vrednosti x_1 i x_2 koje su simetrične u odnosu na μ ako je unapred data verovatnoća 0.95

Tako što prvo nađu Z_1 i Z_2 u odnosu na zadatu verovatnoću, pa se naizmenično uključuju u relaciju $Z=\frac{x-\mu}{\sigma}$.

Znači:

$p(x_1 \leq x \leq x_2)=0.95$

preostala verovatnoća je $\alpha=0.05$, njena polovina vrednosti služi za određivanje Z_1 i Z_2, pa je:

$Z_1=Z\ (\alpha/2) = Z(0.025)=-1.96$

$Z_2=Z\ (1-\alpha/2) =Z(0.975)=1.96$

ako se vrednosti Z_1 i Z_2 naizmenično stave u relaciju $Z=\frac{x-\mu}{\sigma}$ dobijaju se vrednosti za x_1 i x_2:

iz $Z=\frac{x-\mu}{\sigma}$ je: $x=\mu+Z\,\sigma$ date su $\mu=40$ $\sigma=2$, pa je:

$x_1=\mu+Z_1\,\sigma=40-1.96 \times 2=36.08$

$x_2=\mu+Z_2\,\sigma=40+1.96 \times 2=43.95$

Znači, normalna slučajna promenljiva uzima vrednosti u intervalu (36.08; 43.95) uz verovatnoću 0.95, tj.

$p(36.08 \leq x \leq 43.95)=0.95$

TESTIRANJE HIPOTEZE O JEDNAKOSTI EMPIRIJSKOG I TEORIJSKOG RASPOREDA

Teorijski raspored definiše mogućnosti naučnog tumačenja rezultata statističkog istraživanja. Teorijski raspored pokazuje funkcionalnu (idealnu) vezu između vrednosti slučajne promenljive i odgovarajućih verovatnoća, što znači da daje informacije 1) o vrednosti slučajne promenljive 2) i o verovatnoći pojavljivanja svake vrednosti slučajne promenljive. Slučajna promenljiva može biti prekidna ili neprekidna. Verovatnoća pojavljivanja vrednosti slučajne promenljive proizilazi iz zakona verovatnoće svakog modela teorijske raspodele.

Empirijski raspored daje informacije o obeležju slučajne promenljive i o frekvenciji jedinica sa vrednostima obeležja slučajne promenljive. Obeležje može biti prekidno ili neprekidno, daje se u vidu modaliteta ili u vidu grupnih intervala. Frekvencija pojavljivanja modaliteta obeležja slučajne promenljive je rezultat merenja obeležja na svakoj jedinici i njihovog grupisanja prema izmerenim vrednostima obeležja.

Može se zaključiti :

- da se vrednosti slučajne promenljive (kvantitativne karakteristike) i modaliteti obeležja (vrednosti modaliteta) mogu poistovetiti. Razlika je u verovatnoći pojavljivanja, tj. frekvenciji pojavljivanja vrednosti obeležja.
- rekli smo da teorijski raspored pokazuje funkcionalnu (idealnu) vezu između vrednosti slučajne promenljive i odgovarajućih verovatnoća. Idealne veze su retke u statističkim istraživanjima, pa nema idelane veze kod emprijskog rasporeda. Ali, skoro uvek je moguće utvrditi podudarnost empirijskog rasporeda frekvencija sa teorijskim rasporedom verovatnoća i proučavanjem značajnosti njihove razlike otkriti postojanje ili nepostojanje suštinske sličnosti.

Parametri (mere) teorijske i empirijske raspodele su:

- za centralnu tendenciju - u empirijskoj raspodeli frekvencija, sredina μ u skupu , $\overline{X}$ u uzorku i očekivana vrednost $E(X)=\mu$ u teorijskoj raspodeli
- za varijabilnost - u empirijskoj raspodeli varijansa σ^2 u skupu, S^2 u uzorku i σ^2 u teorijskoj raspodeli
- za oblik raspodele - u empirijskoj raspodeli frekvencije, β_1 i β_2 za skup, b_1 i b_2 za uzorak i β_1 i β_2 za teorijsku raspodelu

Ispitivanje podudarnosti empirijske raspodele frekvencija skupa i teorijske raspodele verovatnoća vrši se poređenjem parametara koji ih karakterišu ili empirijskih i očekivanih frekvencija (apsolutnih ili relativnih).

Statističko zaključivanje o jednakosti empirijskog i teorijskog rasporeda vrši se postupkom testiranja hipoteza. Ono se radi pomoću X^2 (Hi-kvadrat testa) ili pomoću Kolmogorov-Smirnovljevog testa.

<u>Ispitivanje podudarnosti empirijskog rasporeda sa Puasonovim rasporedom</u>

Porede se: sredina i varijansa.

Kada je sredina približna ili jednaka varijansi, zaključuje se da je empirijski raspored podudaran teorijskom Poisonovom rasporedu.

Postupak je sledeći:

1) izračuna se sredina empirijskog rasporeda (μ=m=np)

2) izračuna se varijansa empirijskog rasporeda (σ^2=m=np)

3) ako su $\mu \cong \sigma^2$ zaključuje se da empirijski raspored odgovara modelu teorijskog Poisonovog rasporeda

4) da bi se odredilo srednje kvadratno odstupanja frekvencija empirijskog rasporeda od očekivanih frekvencija Poisonove raspodele, izračunavaju se verovatnoće iz zakona verovatnoće Puasonove raspodele: $p_{teor}=e^{-m}\cdot\frac{m^x}{x!}$ sa parametrom **m** iz empirijskog rasporeda. Očekivane frekvencije se računaju iz: f_{teor}=N p, odnosno: $f_{teor}=N\ e^{-m}\cdot\frac{m^x}{x!}$ gde je N-veličina skupa. Treba voditi računa o tome da slučajna promenljiva može imati beskonačno mnogo vrednosti, a da empirijski raspored ima konačan broj vrednosti obeležja i konačan broj frekvencija. Zato se očekivane frekvencije izračunavaju do zadnje vrednosti obeležja empirijskog rasporeda (koja u sebi sadrži razliku između skupa i kumulativnih frekvencija) do predzadnje vrednosti obeležja, tj.: $f_{teor}=N-\sum_{x=0}^{n-1} f$

5) srednje kvadratno odstupanje se računa po obrascu:

$$X^2=\frac{1}{n+1}\cdot\sum_{x=0}^{n}(f_{teor}-f_{empir})^2$$

VAŽNO - Često u praksi jednom empirijskom rasporedu odgovaraju dva teorijska rasporeda (npr. binomni i Poisonov). Koji se tada raspored bira zavisi od veličine srednjeg kvadratnog odstupanja. Tada je postupak sledeći:

1) odrede se očekivane frekvencije Poisonovog i binomnog rasporeda i uporede se sa empirijskim frekvencijama
2) izračuna se srednje kvadratno odstupanje
3) ako je $X^2_{bin} < X^2_{Pois}$ bira se binomni raspored
4) ako je $X^2_{bin} > X^2_{Pois}$ bira se Puasonov raspored
5) znači, bira se raspored kod koga je srednje kvadratno odstupanje manje

Često se ovaj postupak tumači na sledeći način:
- ako je $X^2_{emp} < X^2_{teor}$ onda se prihvata hipoteza da se empirijski raspored prilagođava Poisonovoj raspodeli (H_0- empir. raspored se prilagođava Poisonovoj raspodeli)
- ako je $X^2_{emp} > X^2_{teor}$ onda se ne prihvata hipoteza da se se empirijski raspored prilagođava Puasonovoj raspodeli (H_1- empir. raspored se ne prilagođava Puasonovoj raspodeli)

Ispitivanje podudarnosti empirijskog rasporeda sa binomnim rasporedom

Porede se: sredina i varijansa i eventualno koeficijenti β_1 i β_2.

Postupak je sledeći:
1) izračuna se sredina empirijskog rasporeda (μ=np) odakle je $p = \mu / n$ i $q = 1 - p$
2) izračunavaju se verovatnoće iz zakona verovatnoće binomne raspodele:

$$p= (\ ) p^x p^{n-x} =(\ ) (\mu / n)^x (1-\mu / n)^{n-x}$$

gde je parametar p iz empirijskog rasporeda
3) odrede se očekivane frekvencije za svaku vrednost obeležja

$$f_{teor}= N p$$

gde je $N=\sum f$

Znači, empirijski raspored frekvencija odgovara idealnom binomnom rasporedu ako su ispunjeni sledeći uslovi:
a) $0 < p = \mu / n < 1$
b) $\sigma^2 \approx n p q = \mu (1-p)= \mu (1- \mu / n)$
4) srednje kvadratno odstupanje očekivanih i empirijskih frekvencija računa se po obrascu:

$$X^2=\frac{1}{n+1}\cdot\sum_{x=0}^{n}(f_{teor}-f_{empir})^2$$

što je srednje kvadratno odstupanje manje, empirijski i teorijski raspored su bliži.

Ispitivanje podudarnosti empirijskog rasporeda sa normalnom rasporedom

Porede se: koeficijenti β_1 i β_2.

Empirijski raspored mora biti dat sa obeležjem u vidu grupnih intervala, čime se vrši aproksimacija neprekidnosti slučajne promenljive (kod normalne raspodele slučajna promenljiva uzima neprekidne vrednosti u intervalu -∞ do +∞). Smatra se da za broj elementarnih jedinica veći od 30 svi prekidni rasporedi teže normalnom, što znači da se svi prekidni rasporedi mogu aproksimirati normalnim rasporedom.

Postupak je sledeći:

1) izračunaju se μ, σ^2, β_1 i β_2. Ako $\beta_1 \rightarrow 0$ i $\beta_2 \rightarrow 3$ onda se zaključuje da datom rasporedu odgovara normalni raspored
2) sredina i standardna devijacija empirijskog rasporeda se uvode u teorijski standardizovan normalni raspored i izračunaju se teorijske verovatnoće. One se računaju za svaki grupni interval iz zakona verovatnoće teorijskog standardizovanog normalnog rasporeda. Označimo sa a_i i a_{i+1} granice grupnog intervala, a sa x_i sredinu njihovog grupnog intervala, tada je verovatnoća da će slučajna promenljiva uzeti neku vrednost tog intervala:

$$p\,(a_i < x < a_{i+1}) = p\,(Z_1 < Z < Z_2)=F(Z_2)-F(Z_1)=p_i$$

očekivana frekvencija je:

$$f_{teor}=N\,p_i$$

zbir teorijskih verovatnoća mora biti 1. Zato se prvoj verovatnoći dodaje izostavljeni kraj sa leve strane, a zadnjoj izostavljeni kraj sa desne strane između normalne krive i apscise.
3) srednje kvadratno odstupanje empirijskih od očekivanih frekvencija je:

$$X^2=\frac{1}{n}\cdot\sum_{x=0}^{n}(f_{teor}-f_{empir})^2$$

6. TEORIJSKE OSNOVE STATISTIČKOG ZAKLJUČIVANJA O PARAMETRIMA SKUPA

Pošto iz tehničkih, finansijskih i drugih razloga nije moguće ispitati čitavu statističku masu, potpun statističko posmatranje se zamenjuje delimičnim, odnosno zaključivanjem na osnovu uzorka. Zato se kaže da je **statističko zaključivanje donošenje zaključaka o parametrima skupa na osnovu parametara slučajno izabranih uzoraka.**

Postoje dve oblasti statističkog zaključivanja:
- *statističko ocenjivanje parametara skupa* (ocenjuju se aritmetička sredina, varijansa, standardna devijacija, relativna frekvencija, parametri regresione analize, koeficijent korelacije, parametri u analizi vremenskih serija)
- *testiranje hipoteze* (testira se hipoteza o parametrima skupa na osnovu pretpostavljenih vrednosti, parametara uzorka i reorije raspodele)

Obe vrste statističkog zaključivanja se zasniva na raspodelama uzorka i na pretpostavkama o raspodeli skupa.

Statističko ocenjivanje se vrši *na sledeći način*: nepoznata vrednost parametra skupa ocenjuje se na osnovu odgovarajućih izračunatih vrednosti slučajne promenljive uzorka; nepoznati parametar skupa se obeležava grčkim slovom θ, a statistika uzorka $\overline{\theta}$

Kriterijum tačnosti statistike uzorka zasniva se na težnji da ocena bude približno jednaka ili da teži vrednosti pravog parametra skupa, da odstupa u malom intervalu oko vrednosti pravog parametra.

Statistička ocena se iskazuje:
I- jednom numeričkom vrednošću (**tačkasta ocena**)
II- intervalom vrednosti (**interval poverenja**)

I-tačkasta ocena:

Statistička ocena mora ispuniti 4 uslova:

1) nepristrasnost

ako je ″teta″ θ nepoznati parametar skupa, a sa $\bar{\theta}$ parametar uzorka; uslov nepristrasnosti je ispunjen kada je: $E(\bar{\theta})=\theta$

kada je: $E(\bar{\theta}) \neq \theta$, onda je to pristrasna ocena (ako je $E(\bar{\theta}) > \theta$, ocena je pozitivno pristrasna; ako je $E(\bar{\theta}) < \theta$, ocena je negativno pristrasna);

Nepristrasnost ocene se postiže povećanjem veličine uzorka.

Apsolutna razlika $|\bar{\theta}-\theta|$ je greška ocene.

PRIMER:

aritmetička sredina prostog slučajnog uzorka je jednaka aritmetičkoj sredini skupa, tj. $E(\bar{\theta})=\mu$; ocene su pristrasne ako je $E(\bar{\theta})-\theta<0$ ili $E(\bar{\theta})-\theta>0$

2) minimalne varijanse

ovaj uslov znači da varijansa izabrane ocene mora imati najmanju vrednost među mogućim ocenama

3) efikasnost

ocena je efikasna ako je nepristarsna i ima manju vrednost od bilo koje druge nepristrasne ocene, odnosno ako je $\lim_{n\to\infty} P(|\bar{\theta}-\theta|<\varepsilon)=1$, tj. da je greška ocene $|\bar{\theta}-\theta|$ manja od proizvoljno malog pozitivnog broja ε; važi i obratno: ocena je efikasna ako je $\lim_{n\to\infty} P(|\bar{\theta}-\theta|>\varepsilon)=0$

4) asimptotske efikasnosti

statistička ocena je asimptotski nepristrasna ako je: $\lim_{n\to\infty}\bar{\theta}=\theta$;

nepristrasnost ocene se postiže povećanjem veličine uzorka

II - interval poverenja:

(Pored statističke ocene (koja se iskazuje jednom numeričkom vrednošću), u statističkom zaključivanju se koristi i interval poverenja (koja se iskazuje intervalima vrednosti)).

Interval poverenja je interval u kome se može očekivati tačna vrednost nepoznatog parametra skupa. U određivanju intervala poverenja učestvuju 3 faktora:

- ocena
- faktor poverenja
- standardna greška ocene

interval poverenja = ocena ± faktor poverenja x standardna greška ocene

Ocena je optimalna vrednosna ocena parametra θ, koja je nepristrasna, saglasna i sa minimalnom varijansom.

Faktor poverenja je teorijska vrednost slučajne promenljive odgovarajuće teorijske raspodele (njegova konkretna vrednost zavisi od nivoa poverenja koji predstavlja verovatnoću tačnog uključivanja parametra skupa u interval koji je proizveden).

Standardna greška ocene je standardna devijacija ocena uzorka.

Cilj istraživanja i *veličina uzorka* određuju primenu teorijske raspodele:

- normalna raspodela kada se radi o određivanju intervala poverenja aritmetičke sredine i relativne frekvencije skupa na osnovu velikog uzorka za slučajeve kada je varijnsa skupa poznata ili nepoznata
- studentova raspodela u slučaju ocene aritmetičke sredine skupa na osnovu malog uzorka za slučaj kada je varijansa skupa nepoznata, kao i u slučaju ocene parametara skupa u regresionoj i korelacionoj analizi i analizi vremenskih serija

Preciznost ocene određuju:
- faktor poverenja
- standardna greška ocene

preciznost ocene (E) = faktor poverenja x standardna greška ocene

Interval koji se opisuje levo i desno (iznad i ispod) od ocene je interval poverenja. Njegova širina je 2E. Interval poverenja ima 2 granice:
- G_1 (donja)
- G_2 (gornja)

G_1=ocena - (faktor poverenja)(standardna greška ocene)

G_2=ocena +(faktor poverenja)(standardna greška ocene)

Širi interval preciznosti (E) daje manje preciznu ocenu, uži daje precizniju ocenu uz porast rizika pogrešne ocene.

Verovatnoća da će se tačna vrednost parametra populacije naći u intervalu poverenja je koeficijent poverenja. Obeležava se : $(1-\alpha)$, a izraženo u procentima: $(100\,(1-\alpha))$

Verovatnoća da će se tačna vrednost parametra skupa (populacije) naći izvan intervala poverenja je koeficijent rizika. Obeležava se: (α) , a izraženo u procentima: $(100\,\alpha)$

PRIMER:

ako je koeficijent poverenja $1-\alpha=95\%$, onda je koeficijent rizika 5%

Vrednost $Z_{\alpha/2}$ je terijska vrednost slučajne promenljive Z standardizovanog (jediničnog) normalnog rasporeda; ova vrednost zavisi od koeficijenta rizika α; najčešći slučajevi su:

$Z_{\alpha/2}$	koefic.rizika α	koefic.poverenja $(1-\alpha)$
±1	0.3174	0.6826
±1.96	0.0500	0.9500
±2	0.0455	0.9545
±3	0.0027	0.9973

Statistička ocena nepoznatog parametra skupa (populacije) data intervalom poverenja zasniva se na teorijskim osnovama jedne od sledećih raspodela:
- Standardizovane normalne raspodele
- Studentove t-raspodele
- Hi-kvadrat raspodele
- Snedekorove F-raspodele

TESTIRANJE STATISTIČKIH HIPOTEZA

Testiranje statističkih hipoteza je provera određenih pretpostavki – uporednom analizom pretpostavljene vrednosti skupa i izračunate statistike uzorka, na osnovu teorije raspodele, donose se zaključci o jednakosti ili razlici ovih veličina.

Testiranje se sprovodi u dva smera:

- testiranje hipoteze o jednakosti parametara skupa i pretpostavljenih vrednosti, na osnovu statistike uzorka
- testiranje hipoteze o jednakosti teorijskih i empirijskih raspodela

Statistička hipoteza se mora ispitati, da bi se dobio odgovor o njenoj istinitosti. Taj postupak ispitivanja se naziva testiranjem hipoteze, proverom istinitosti hipoteze ili verifikacijom hipoteze.

Testovi se dele na:

- parametarske (t-test, F-test)
- neparametarske Hi kvadrat testm medijana test, test predznaka, test ekvivalentnih rangova, test sume rangova, fridmanov test, Kolmogorov-Smirnov test...)

Izbor testa zavisi od istraživačkog cilja, plana istraživanja, statističkog modela, broja i prirode promenljivih, veličine uzorka.

Testiranje počinje formulisanjem <u>nulte hipoteze</u>. Zatim se nastavlja kroz niz uzastopnih faza testiranja i završava se prihvatanjem (verifikovanjem) ili odbijanjem nulte hipoteze. Odluka se donosi uz rizik da se odbaci tačna hipoteza (najčešći slučaj) ili da se prihvati netačna hipoteza (ređi slučaj).

a) <u>Kod parametarskih testova</u>, nultom hipotezom se definiše jednakost stvarne (θ) i hipotetičke vrednosti (θ_0) parametra skupa. Nulta hipoteza se obeležava sa H_0. Postavlja se na 2 načina:

$$H_0 : \theta = \theta_0$$

$$H_0 : \theta - \theta_0 = 0$$

Nulta hipoteza ima svoju antihipotezu. Naziva se alternativnom hipotezom. Ima za cilj da ospori valjanost nulte hipoteze. Obeležava se sa H_1. Važi:

$$H_1 : \theta \neq \theta_0$$

$$H_1 : \theta - \theta_0 \neq 0$$

Postoje 2 moguć nosti realizacije alternativne hipoteze:

$$H_1 : \theta > \theta_0$$

$$H_1 : \theta < \theta_0$$

Tri varijante postavljanja nulte hipteze:

$H_0 : \theta = \theta_0$	protiv	$H_1 : \theta \neq \theta_0$
$H_0 : \theta \leq \theta_0$	protiv	$H_1 : \theta > \theta_0$
$H_0 : \theta \geq \theta_0$	protiv	$H_1 : \theta < \theta_0$

<u>Vrste testova</u>

Pri zaključivanju o parametrima skupa koriste se 4 vrste testova:

- z-test (ako su uzorci veliki, raspodela normalna)

  zaključivanje se vrši na osnovu standardizovane normalne raspodele
- t-test (kada su uzorci mali)

  zaključivanje se vrši na osnovu Studentovog t-rasporeda

- χ_2-test (kada se poredi veći broj promenljivih)
    zaključivanje se vrši na osnovu hi-kvadrat raspodele
- F-test (kada su promenljive kvalitativne)
    zaključivanje se vrši na osnovu Snedekorove F-raspodele
    Sprovode se:
- dvosmerni (zasniva se na pretpostavci da su podaci raspoređeni sa obe strane neke centralne vrednosti. Dozvoljava da se donese zaključak da li je npr. razlika između dve aritmetičke sredine značajna bez obzira da li je razlika + ili -. Test daje odgovor na pitanje kolika je verovatnoća da se tako velika razlika između dve aritmetičke sredine pojavi slučajno).

ako je alternativna hipoteza $H_1 : \theta \neq \theta_0$ sprovodi se dvosmerni test; test ima dve kritične vrednosti (donji i gornji prag testa); oblasti odbacivanja nulte hipoteze nalaze se sa leve i desne strane
- jednosmerni testovi (koristi se kod testiranja razlike kod npr. dve aritmetičke sredine kada se očekuje da je ta razlika samo u jednom smeru, samo pozitivna ili samo negativna, ili kada nas interesuje samo razlika u jednom smeru. Naziva se direktni test)

ako je alternativna hipoteza $H_1 : \theta > \theta_0$ ili $H_1 : \theta < \theta_0$ sprovodi se jednosmerni test sa jednom kritičnom vrednošću testa (prag testa)

b) Kod neparametarskih testova, zaključivanje se vrši poređenjem hipotetičkih i iz uzorka dobijenih frekvencija, ranga i sl. Ako razlike nema, nulta hipoteza se prihvata. Ako postoji razlika, ispituje se statistička značajnost razlike. Za to se koristi određeni statistički izraz (statistika testa). Zaključivanje se vrši poređenjem vrednosti statistike testa u konkretnom ispitivanju i teorijske vrednosti statistike testa. Pa ako je statistička hipoteza istinita, konkretna vrednost statistike testa se javlja sa određenom verovatnoćom u teorijski predviđenom nizu mogućih vrednosti. Granica te verovatnoće je prag testa. Niz mogućih vrednosti sa verovatnoćom do i preko praga testa je oblast prihvatanja, tj. odbacivanja nulte hipoteze.

DONOŠENJE ODLUKE O HIPOTEZI

Svaka statistička hipoteza može biti:
- tačna
- netačna

Pri verifikaciji statičkih hipoteze, hipoteza se može:
- usvojiti
- odbaciti

Tada se donose 2 pravilne odluke:
- nulta hipoteza je tačna i usvaja se
- alternativna hipoteza je tačna i usvaja se

Nepravilne odluke su:
- nulta hipoteza je tačna i odbacuje se
- alternativna hipoteza je tačna i prihvata se

Odluka da se odbaci tačna nulta hipoteza je greška prve vrste. Verovatnoća da se desi greška prve vrste je α. Naziva se i verovatnoćom značajnosti testa. Bira se proizvoljno, ali je uobičajeno da uzima vrednosti:

- $\alpha = 0.05$ (tada se rizikuje da se od 100 slučajeva u 5 slučajeva odbaci hipoteza koja je istinita)
- $\alpha = 0.01$ (tada se rizikuje da se od 100 slučajeva u 1 slučaju odbaci hipoteza koja je istinita)

Odluka o prihvatanju tačne alternativna hipoteze je greška druge vrste. Verovatnoća da se desi greška druge vrste je β.

Odluka o odbacivanju netačne hipoteze je jačina (moć) testa. Moć testa je sposobnost da otkrije neku razliku ako ona zaista i postoji. Iskazuje se verovatnoćom odbacivanja H_0 hipoteze u testu kada je ona pogrešna. Neparametarski testovi imaju manju moć, zaključci na bazi njih su manje precizni. Moć testa raste:

- sa povećanjem veličine uzorka
- većim nivoom merenja

Moć testa je razlika između 1 i verovatnoće greške druge vrste. Što je moć testa veća, test je bolji. Verovatnoća da se odbaci netačna hipoteza je γ.

Odnos verovatnoća je:

$\beta + \gamma = 1$

$\gamma = 1 - \beta$

verovatnoća jačine testa je utoliko veća ukoliko je verovatnoća greške druge vrste manja.

Verovatnoća prihvatanja stvarno tačne hipoteze je:

$p(-Z\alpha/2 < Z < Z\,\alpha/2) = \underline{2F(Z) - 1 = 1 - \alpha}$

Rizik greške postoji i on je $\underline{\alpha}$.

Standardne vrednosti za $1-\alpha$ i α su 0,95 i 0,05, tj. $1-\alpha=0,95$ ili 95% i nivo značajnosti testa α ili rizik greške je $\alpha=0,05$ ili 5%.

$2F(Z) - 1 = 0,95$

$F(Z) = 1,95/2=0,9750$

$Z\alpha/2=1,96$

Kod ispitivanja snačajnosti razlike između hipotetične i stvarne vrednosti parametra skupa (populacije) ocenjene iz uzorka polazi se od 2 pretpostavke:

- razlika nastaje pod dejstvom slučajnih varijacija - ova razlika je dopustiva do određenih granica i meri se standardnom greškom ocene;
- razlika nastaje jer je nulta hipoteza netačna - razlika koja prelazi određene granice naziva se značajnom, tj. visokoznačajnom; nastaje pod dejstvom sistemskih varijacija, tj. zbog netačne nulte hipoteze

Slučajne varijacije se određuju verovatnoćom. To je istovremeno i i verovatnoća usvajanja hipoteze.

Ispitivanje hipoteze se radi pomoću određenog statističkog izraza, koji se naziva statistika testa. Koriste se različite statistike testa, zavisno od statističke metodologije i teorije raspodele.

Slučajna standardizovana normalna slučajna promenljiva Z se koristi kod velikog uzorka ($n \geq 30$) kada je varijansa skupa poznata i nepoznata.

Slučajna Studentova t promenljiva se koristi:

- kod malog uzorka (n<30) i nepoznate varijanse
- kod testiranja hipoteze o statističkoj značajnosti parametara skupa u regresiji i korelaciji

Slučajna χ_2-promenljiva se koristi kod testiranja hipoteze o jednakosti empirijskog i teorijskog rasporeda i testiranja hipoteze o varijansi skupa.

statistika testa=(ocena parametra-hipotetička vrednost parametra)/stand.greška ocene parametra

Poređenjem izračunate statistike testa i teorijske, čija se vrednost dobija na osnovu određenog teorijskog rasporeda, usvaja se nulta ili alternativna hipoteza. Testiranje hipoteze se vrši uz određenu verovatnoću. Ako se usvoji nulta hipoteza, onda se sa određenom verovatnoćom tvrdi da će se vrednost parametra skupa naći u okviru određenih granica teorijske vrednosti. Granica teorijske vrednosti, određena na osnovu verovatnoće, naziva se ***donji i gornji prag značajnosti*** testa, koji se pak opet nazivaju kritičnim vrednostima testa ili samo pragovi testa. Verovatnoća testiranja hipoteze je verovatnoća usvajanja nulte hipoteze ($1-\alpha$). Suprotna verovatnoća (α) se naziva verovatnoća nivoa značajnosti testa ili rizik greške.

7. STATISTIČKO ZAKLJUČIVANJE O PARAMETRIMA POPULACIJE NA OSNOVU VELIKOG UZORKA

Statističko ocenjivanje podrazumeva određeivanje intervala poverenja za parametre skupa na osnovu statistike uzorka. Razlikuju se sledeći slučajevi:

a)- statističko zaključivanje o aritmetičkoj sredini skupa, ako je varijansa skupa poznata

b)- statističko zaključivanje o aritmetičkoj sredini skupa, ako je varijansa skupa nepoznata (ovde se zaključivanje vrši pomoću standardizovane slučajne promenljive Z za velike uzorke i standardizovane Studentove slučajne promenljive t za male uzorke; pošto skup može biti konačan i beskonačan, kod konačnih skupova u standardnu grešku treba uključiti korektivni faktor)

a) Statističko zaključivanje o aritmetičkoj sredini skupa kada je poznata varijansa

Pretpostavke su:

Posmatrano obeležje X je slučajna promenljiva koja ima normalnu raspodelu, tj.

$X:\ N(\mu, \sigma^2)$.

Aritmetička sredina aritmetičkih sredina uzorka ima normalnu raspodelu, tj.

$\overline{X}\ :\ N(\mu,\ \sigma_p^2)$

Standardizovana normalna slučajna promenljiva Z ima normalnu raspodelu sa $\mu=0$, $\sigma^2=1$, tj. kao standardizovana sredina uzoraka, ima standardizovanu normalnu raspodelu, tj.:

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} : \ N(0,1)$$

Varijansa skupa σ^2je poznata.

Aritmetička sredina skupa μ je nepoznata.

Hipotetička vrednost aritmetičke sredine skupa je μ_0.

Uzorak može biti mali i veliki.

Skup može biti konačan i beskonačan. Za konačan se, po određenom pravilu, uključuje korektivni faktor.

Na osnovu statistike uzorka određuje se interval poverenja aritmetičke sredine μ ili testira hipoteza o jednakosti aritmetičkih sredina μ i μ_0.

Statističko zaključivanje se sprovodi po sledećem postupku:

1. odrede se aritmetička sredina i varijansa uzorka

- sredina uzorka je: $\overline{X} = \dfrac{\sum X}{n}$ $\quad i \quad$ $\overline{X} = \dfrac{\sum X f}{\sum f}$

- varijansa uzorka je: $S^2 = \dfrac{\sum_{i=1}^{n}(X - \overline{X})^2}{n} = \dfrac{\sum_{i=1}^{n} X^2}{n} - \overline{X}^2$ $\quad$ i $\quad$ $S^2 = \dfrac{\sum_{i=1}^{n} X f}{\sum f} - \overline{X}^2$

2. odredi se Standardna greška sredine:

$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ za beskonačne skupove

kod konačnih skupova se u standardnu grešku sredine uvodi korektivni faktor $\sqrt{n}$-1.

3. nakon ocene aritmetičke sredine, određuje se interval poverenja

- interval poverenja=ocena ±(faktor poverenja) (standardna greška ocene)
- faktor poverenja je $Z_{\alpha/2}$
- kod standardizovane normalne raspodele je: $P(Z_1<Z<Z_2)=F(Z_2)-F(Z_1)=1-\alpha$

Uvođenjem statistike Z i rešavanjem, dobija se:

$$P(\overline{X} + Z_1 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_2 \cdot \frac{\sigma}{\sqrt{n}}) = F(Z_2) - F(Z_1)$$

Verovatnoća intervala poverenja (koeficijent poverenja) je: $F(Z_2)-F(Z_1)$. Ona je unapred data i iznosi: $\beta=(1-\alpha)$. α je koeficijent rizika ili nivo značajnosti. Određivanje vrednosti Z_1 i Z_2 se vrši u funkciji $\alpha/2$. Zato je:

$Z_1=Z(\alpha/2) \qquad Z_2=Z(1-\alpha/2)$

Zbog uslova:

$P(|Z| \leq Z=\beta$, tj. $2\ \theta(Z) - 1=\beta$ ili $\theta(Z)=(1+\beta)/2$

Verovatnoća intervala poverenja je:

$$P(|\overline{X} - \mu| \leq Z \cdot \frac{\sigma}{\sqrt{n}}) = P(\overline{X} - Z(\alpha/2) \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z(1-\alpha/2) \cdot \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

Pošto su $Z(\alpha/2)$ i $Z(1-\alpha/2)$ simetrične oko nule kao sredine njihove apsolutne vrednosti, one su jednake, tj.:

$|Z(\alpha/2)| = |Z(1-\alpha/2)| = Z_{\alpha/2}$

Pošto je vrednost $Z(\alpha/2)$ uvek negativna, a $Z(1-\alpha/2)$ uvek pozitivna, dobija se:

$Z(\alpha/2) = -Z_{\alpha/2}$

$Z(1-\alpha/2) = Z_{\alpha/2}$

Znači, verovatnoća intervala poverenja je:

$$P(\overline{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

Standardna greška sredine je:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Preciznost ocene je:

$E = Z_{\alpha/2}\, \sigma_x$

Interval poverenja je:

$$(\overline{X} - Z_{\alpha/2} \cdot \sigma_{\overline{X}} < \mu < \overline{X} + Z_{\alpha/2} \cdot \sigma_{\overline{X}})$$

ili

$$\overline{(X - E < \mu < \overline{X} + E)}$$

(interval poverenja je interval levo i desno (iznad i ispod) od sredine uzorka)

Verovatnoća intervala poverenja je 1-α i to je koeficijent poverenja (koeficijent pouzdanosti). Najčešće se izražava u procentima: 100 (1-α). Koeficijent rizika je α. I on se najčešće izra`ava u procentima: 100 α. Rizik pogrešne ocene je dvostran, pa se zato određivanje konkretne vrednosti faktora poverenja vrši u funkciji od α/2. Sa porastom koeficijenta poverenja, interval poverenja se proširuje, tj. smanjuje se preciznost ocene. U praksi se najčešće koristi nivo poverenja od 95% (0,95). Pri istom koeficijentu poverenja interval poverenja možemo da suzimo ako povećamo uzorak, tj. preciznost ocene se povećava sa povećanjem veličine uzorka.

PRIMER1.

Za date nivoe poverenja: 90%, 95%, 98%, 99%, 99,5%, 99,8%, interval poverenja (iz tablica normalne raspodele: 1.645; 1.96; 2.33; 2.582; 2.81; 3.09)) je:

$$(\overline{X} - 1.645 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.645 \cdot \frac{\sigma}{\sqrt{n}})$$

$$(\overline{X} - 1..96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}})$$

$$(\overline{X} - 2.33 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 2.33 \cdot \frac{\sigma}{\sqrt{n}})$$

$$(\overline{X} - 2.58 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 2.58 \cdot \frac{\sigma}{\sqrt{n}}) \qquad (\overline{X} - 2.81 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 2.81 \cdot \frac{\sigma}{\sqrt{n}}) \qquad (\overline{X} - 3.09 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 3.09 \cdot \frac{\sigma}{\sqrt{n}})$$

ili:

koef.poverenja	$Z_{\alpha/2}$	forma interv.ocene
90	1.645	$(\overline{X} - 1.645 \cdot \sigma_{\overline{X}} < \mu < \overline{X} + 1.645 \cdot \sigma_{\overline{X}})$
95	1.96	$(\overline{X} - 1.96 \cdot \sigma_{\overline{X}} < \mu < \overline{X} + 1.96 \cdot \sigma_{\overline{X}})$
99	2.58	$(\overline{X} - 2.58 \cdot \sigma_{\overline{X}} < \mu < \overline{X} + 2.58 \cdot \sigma_{\overline{X}})$

PRIMER 2.

Izvučen je uzorak od 100 jedinica. Standardna devijacija skupa je 10, aritmetička sredina uzorka je 160. Odredi 99% interval poverenja za aritmetičku sredinu skupa.

Rešenje:
Ranije smo naveli uslov:

P(| Z| ≤ Z=β, tj. 2 θ (Z) - 1=β ili θ (Z)=(1+β) / 2

Znači: θ (Z)=(1+β) / 2 = (1+0,99) / 2=0,995

iz tablice normalnog raspodeda se vidi da je Z=2,58

99% interval poverenja je:

$$(\overline{X} - Z \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z \cdot \frac{\sigma}{\sqrt{n}}) = (160 - 2.58 \cdot \frac{10}{\sqrt{100}} < \mu < 160 + 2.58 \cdot \frac{10}{\sqrt{100}}) = (157.42 < \mu < 162.58)$$

PRIMER 3.

Iz skupa sa raspodelom N(μ,6) izvučen je uzorak:

X	f	Xf
0	2	0
1	6	6
2	15	30
3	9	27
4	4	16
	36	79

Naći 99%, 95% i 90% interval poverenja za nepoznatu aritmetičku sredinu skupa μ.

Rešenje:

za date koeficijente poverenja: 0,99; 0,95 i 0,90 iz tablice N(0,1) nalaze se vrednosti: 2.576; 1.96 i 1.645.

Aritmetička sredina uzorka je: $\overline{X} = \frac{\sum Xf}{\sum f} \frac{79}{36} = 2.194$

za intervale poverenja:

- 99%:

$$(\overline{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}) = (2.194 - 2.576 \cdot \frac{6}{\sqrt{36}} < \mu < 2.194 + 2.576 \cdot \frac{6}{\sqrt{36}}) = (-0.382 < \mu < 4.77)$$

- 95%:

$$(\overline{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}) = (2.194 - 1.96 \cdot \frac{6}{\sqrt{36}} < \mu < 2.194 + 1.96 \cdot \frac{6}{\sqrt{36}}) = (0.234 < \mu < 4.154)$$

- 90%:

$$(\overline{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}) = (2.194 - 1.645 \cdot \frac{6}{\sqrt{36}} < \mu < 2.194 + 1.645 \cdot \frac{6}{\sqrt{36}}) = (0.549 < \mu < 3.839)$$

Za najmanji nivo poverenja od 90% dobijen je naju`i ivnerval poverenja, a za najve}i nivo poverenja od 99% dobijen je naj{iri interval poverenja. Ako je interval mnogo {irok, nije koristan kao informacija.

PRIMER 2.

Izvu~en je uzorak od n=81 jedinica iz skupa koji ima N(μ,9). Aritmet.sredina uzorka je 6. Na}i verovatno}u P(5<μ<7).

Re{enje:

$P(Z_1<Z<Z_2)=F(Z_2)-F(Z_1)$

$Z_1=Z(\alpha/2)$ $\quad$ $Z_2=Z(1-\alpha/2)$

Ranije smo naveli uslov:

$P(|Z| \leq Z=\beta$, tj. $2\,\theta(Z) - 1=\beta$ ili $\theta(Z)=(1+\beta)/2$

P(5<μ<7)= $$P(Z_1 < Z < Z_2) = P(\frac{\overline{X_1}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{\overline{X_2}-\mu}{\frac{\sigma}{\sqrt{n}}}) = P(\frac{5-6}{\frac{9}{\sqrt{81}}} < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{7-6}{\frac{9}{\sqrt{81}}}) =$$

$$= P(-1 < Z < +1) = 2\cdot\theta(1) - 1 = 2\cdot 0.8413 - 1 = 0.6823$$

<u>Testiranje hipoteze o vrednosti aritmetičke sredine skupa kada je varijansa skupa poznata</u>

Testiranje se vrši za tri slučaja:

1) *Testira se nulta ipoteza H_0 ($\mu=\mu_0$) protiv alternativne hipoteze H_1 (μ # μ_0) za dati nivo značajnosti α. Skup ima normalnu raspodelu, a uzorak je proizvoljne veličine ili raspored osnovnog skupa nije poznat, a veličina skupa je ≥30.*

Znači, polazi se od činjenice da statistika Z ima standardizovanu normalnu raspodelu, tj. $Z=\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$: N(0,1) pod uslovom da je nulta hipoteza tačna, H_0: $\mu=\mu_0$.

Alternativna hipoteza je H_1: $\mu \neq \mu_0$. U ovom slučaju se sprovodi <u>dvosmerni</u> test. Verovatnoća značajnosti testa je α. Kritične vrednosti, tj. pragovi testa su: $\pm Z_{\alpha/2}$. To su tablične vrednosti standardizovane normalne raspodele. Verovatnoća prihvatanja stvarno tačne hipoteze je:

$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 2F(Z)-1=1-\alpha$

(α-rizik greške; 1-α -interval poverenja; dopunjuju se do 100%, tj. do 1)

Drugi način zaključivanja:

Posmatra se apsolutna vrednost statistike testa:

$P(|Z| \geq Z_{\alpha/2})= \alpha$ (1)

Posmatra se apsolutna vrednost statistike testa i to u odnosu na desni prag testa:

$P(|Z| < Z_{\alpha/2})= 1-\alpha$

Nulta hipoteza se prihvata ako je $|Z| < Z_{\alpha/2}$, tj. nulta se odbacuje ako je $|Z| \geq Z_{\alpha/2}$

Verovatnoća odbacivanja stvarno tačne hipoteze je:

$P(|Z| \geq Z_{\alpha/2})= \alpha$

Odluka o nultoj hipotezi se donosi prema tome da li se dobijena vrednost statistike testa nalazi u zoni prihvatanja ili zoni odbacivanja.

Treći način zaključivanja:

Umesto ovoga, testiranje se može vršiti i primenom statistike $\overline{X}$. Naime, iz uslova $P(|Z| \geq Z_{1-\alpha/2}) = \alpha$ se može dobiti donja kritična vrednost: $\overline{X}_d=\mu_0 - Z_{1-\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$ i gornja kritična vrednost: $\overline{X}_g = \mu_0 + Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$. Pravilo odlučivanja je: ako je $\overline{X}_d \leq \overline{X} \leq \overline{X}_g$ onda se prihvata nulta hipoteza H_0, a suprotno - ne prihvata se nulta hipoteza.

2) *Testira se nulta ipoteza H_0 ($\mu \geq \mu_0$) protiv alternativne hipoteze H_1 ($\mu < \mu_0$) za dati nivo značajnosti α. Skup ima normalnu raspodelu, a uzorak je proizvoljne veličine ili raspored osnovnog skupa nije poznat, a veličina skupa je ≥ 30.*

Za dati nivo značajnosti:

$P(Z \leq Z_\alpha) = \alpha$

dobija se kritična vrednost Z_α iz tablice standizovane normalne raspodele, tj. čita se vrednost $Z_{1-\alpha}$. Ovo je jednosmerni test. Zna se da je $-Z_{1-\alpha} = Z_\alpha$.

3) *Testira se nulta ipoteza H_0 ($\mu \leq \mu_0$) protiv alternativne hipoteze H_1 ($\mu > \mu_0$) za dati nivo značajnosti α. Skup ima normalnu raspodelu, a uzorak je proizvoljne veličine ili raspored osnovnog skupa nije poznat, a veličina skupa je ≥ 30.*

Za dati nivo značajnosti:

$P(Z \geq Z_\alpha) = \alpha$

dobija se kritična vrednost Z_α iz tablice standizovane normalne raspodele, tj. čita se vrednost $Z_{1-\alpha}$. Ovo je jednosmerni test.

8. STATISTIČKA KONTROLA (statist.upravljanje) PROCESA

Osnova statističkog upravljanja je normalna (Gausova) raspodela, gde su levo i desno od aritmetičke sredine tri standardne devijacije, koje obuhvataju 99% procesa. Cilj statističkog upravljanja procesima je da proces bude dosledan, kontrolisan, stabilan i predvidiv.

Osnovna osobina procesa je varijabilnost izazvana dejstvom brojnih faktora, koji mogu biti:

- slučajni (iznenada nastaju, različitog su intenziteta, ne može im se predvideti ni nastanak ni trajanje, ne utiču trajno na proces. U ovom slučaju su elementarne jedinice istovrsne, ali nisu istovetne. Smatra se da odstupanje od nominalne (srednje) vrednosti za 3 standardne devijacije nastaje pod dejstvom slučajnih faktora))
- sistematski (neslučajni, izazivaju znatnije poremećaje u procesu i njegovu nestabilnost, nastaju s vremena na vreme i deluju dok se ne otklone)

Zadatak statističke kontrole procesa je otkrivanje promena u procesima pod dejstvom sistematskih faktora. Procesi su određeni preko parametara. Cilj statističke kontrole je da se postigne nominalna vrednost ($\overline{X}$) i da sve elementarne jedinice u skupu teže nominalnoj (srednjoj) vrednosti. Odstupanje elementarnih jedinica može biti pod dejstvom slučajnih i sistematskih faktora.

Procesi mogu biti:

- statistički stabilni (ako se javljaju samo slučajne varijacije)
- statistički nestabilan (ako se pored slučajnih javljaju i sistematske varijacije)

Uslov da se održi neki proces je njegova statistička stabilnost. Odluka o tome da li je neki proces statist.stabilan donosi se na osnovu procesne karte i kontrolne karte individualnih vrednosti. Osnova stat.kontrole je normalna raspodela i teorija uzoraka. Provera da li elem.jedinice ispunjavaju određene zahteve vrši se preko uzorka, tj. na osnovu uzorka se konstruišu kontrolne karte, na osnovu kontrolišu procesi. Kontrolne karte ne mogu otkriti razlog stat.nestabilnosti procesa.

Procesne karte - pokazuju tok procesa (larta trenda) u toku vremena. Da bi se mogla koristiti za ispitivanje stat.stabilnosti procesa, procesne karte moraju imati najmanje 40 podataka u nizu. Kriterijumi za ocenu stat.stabilnosti procesa su:

- ″vizuelni kriterijum″

- ″serija od 8″
- ″serija od 14″

Konstrukcija *kontrolnih karata*:

Na osnovu uzorka gde treba da bude $k \geq 20$ crta se tabela.

Obeležje	elementi u uzorku	Srednja vred. u uzorku	Razmak varijacije
1 2 . . k			

Elementi kontrolne karte su:

Centralna linija kontrolne karte je CL = μ

Gornja kontrolna granica je GKG = $\mu + 3\sigma$

Donja kontrolna granica je DKG = $\mu - 3\sigma$

Konstante kontrolne karte:

Da bi se pojednostavilo izračunavanje granica kontr.karata, uvedene su konstante (koeficijenti) čija vrednost zavisi od veličine uzorka. Za datu veličinu uzorka, konstante imaju nepromenjenu vrednost, datu u tablicama.

Pravilo za ocenjivanje stat.stabilnosti procesa je: proces je stat.stabilan ako sve tačke leže u intervalu $\mu \pm 3\sigma$. Ako neka od tačak leži izvan tog intervala, proces je nestabilan. (Od ovog pravila može da se javi odstupanje. Naime, može se desiti da stabilan proces ima vrednosti izvan $\mu \pm 3\sigma$. Ako se tada zaključi da je proces nestabilan, a u stvari je u redu, čini se greška prve vrste. Ova greška je malo verovatna i označava se sa α. Isto tako, može se desiti da nestabilan proces postigne vrednosti koje leže unutar $\mu \pm 3\sigma$.. Ako se tada zaključi da je proces stabilan, čini se greška druge vrste i označava se sa β. I ova greška je malo verovatna.)

Nakon što se utvrdilo da li je proces stat.stabilan ili nije, ispituje se da li je raspodela normalna. Kod normalne raspodele je sredina = medijana = modus (sve srednje vrednosti imaju istu vrednost). Koeficijent asimetrije $\beta_1=0$ (normalna raspodela je idealno asimetrična). Koeficijent sploštenosti $\beta_2=3$.

Ocena sposobnosti procesa:

Meri se pomoću dva indeksa:

- indeksa preciznosti (sposobnosti) procesa
- indeksa tačnosti (sigurnosti) procesa

Indeks preciznosti -

$$C_p = \frac{T}{6\sigma}$$

Gde su: T - razmak gornje (maksimalne) i donje (minimalne) propisane vrednosti (tolerancije; $T = X_g - X_d$); σ je stand.devijacija

$$\overline{X} = \frac{\sum X}{n} \qquad \sigma = \sqrt{\frac{\sum (X-\overline{X})^2}{n}}$$

$C_p \leq 1.00$	Nedovoljno precizan proces
$1.00 < C_p \leq 1.33$	Precizan proces
$1.33 < C_p \leq 1.67$	Značajno precizan proces
$C_p > 1.67$	Visoko precizan proces

Indeks tačnosti - $\quad C_{pk} = \dfrac{\Delta}{3\sigma} = \dfrac{\sum \Delta}{T_p}$

Gde je $\Delta = \left|\min(\overline{X} - X_d, X_g - \overline{X}\right|$, bira se ono što je manje

$X_s = \frac{X_d + X_g}{2}$ je položaj sredine tolerancije

$C_{pk} \leq 1.00$	Nedovoljno tačan proces
$1.00 < C_{pk} \leq 1.33$	Tačan proces
$1.33 < C_{pk} \leq 1.67$	Značajno tačan proces
$C_{pk} > 1.67$	Visoko tačan proces

1. ako je $C_p \leq 1$ i $C_{pk} \leq 1$ proces je neprecizan i netačan. Proces treba prekinuti i preregulisati ga, prepodesiti proces za vrednost Δr:

$\Delta r = |\overline{X} - X_s|$

2. ako je $1 < C_p < 1.33$ i $1 < C_{pk} < 1.33$ izvesti naknadnu doradu delova
3. ako je $C_p \geq 1.33$ i $C_{pk} \geq 1.33$ proces je i tačan i precizan.

<u>Kontrolne karte se dele na</u>:

- one za kontrolu numeričkih obeležja
- one za kontrolu atributivnih obeležja

Kontrolne karte za kontrolu numeričkih obeležja su:

- kontrolna karta za sredinu (kontrolna $\overline{X}$)
- kontrolna karta za razmak varijacije (kontrolna R karta)
- kontrolna karta za standardnu devijaciju (kontrolna S karta)

Kontrolna karta za kontrolu atributivnih obeležja je P-karta.

Kombinacijom ove tri karte, dobijaju se kontrolne karte koje imaju najveću praktičnu primenu:

- kontrolna $\overline{X}$ R karta
- kontrolna $\overline{X}$ S karta

1. <u>$\overline{X}$ karta:</u> Redosled izrade $\overline{X}$ karta:

- crta se procesna karta da bi se vizuelno odredilo da li je proces stabilan ili nije; ako jeste, izrađuje se kontrolna karta za sredinu
- izvadi se na slučajan način prvi uzorak od n=5 (ili n=4) element.jedinica, zatim računa sredina i razmak varijacije: $\overline{X} = \frac{\sum X}{n}$ $\quad R = X_{max} - X_{min}$
- izračuna se procesna sredina: $\overline{\overline{X}} = \frac{\sum_{i=1}^{k} \overline{X_i}}{k}$ To je centralna linija na kontrolnoj karti. Kaže se da je to prihvatljiv prosek za proces.
- Računa se srednji procesni razmak: $\overline{R} = \frac{\sum R_i}{k}$
- određuju se kontrolne granice: $GKG = \overline{\overline{X}} + A_2\overline{R}$ i $DKG = \overline{\overline{X}} - A_2\overline{R}$ (A_2 je konstanta kontrolne karte; čita se iz tablice)

- Zatim se konstruiše kontrolna karta, tako što se na x-osu nanose redni brojevi uzoraka u redosledu posmatranja (obeležja), a na Y-osu izračunate karakteristike koje se kontrolišu. Paralelmo sa X-osom se crta centralna linija. Tačkama se nanose sredine uzorka, pa povežu linijama.

- ako nijedna sredina ne prelazi granice kontrolne karte, proces je stat.stabilan
- ako neka sredina pređe granice, proces se zaustavlja, pronalazi se i otklanja uzrok, tj. menja se procesna sredina.

2. <u>R karta:</u>

Za razmak varijacije, prati varijacije u procesu.

Centralna linija je: $CL = \overline{R}$

Kontrolne granice:

$GKG = D_4\ \overline{R}$

$DKG = D_3\ \overline{R}$

D_3 i D_4 su konstante koje se čitaju iz tablice, prema veličini uzorka.

<u>Testiranje hipoteze (proračun kontrolnih granica i zadatih tolerancija)</u>

$T = 6\,\sigma_0$ $\sigma_0 = T/6$

$T = T_g - T_d$ (T-toler)

σ_0 standardna devijacija hipoteze

za $\overline{X}$ karte:

centralna linija je: $CL = X_0$ (X_0 je sredina tolerancije)

kontrolne granice:

$GKG = X_0 + A\,T$ $DKG = X_0 - A\,T$

A je konstanta iz karte, ~ija vrednost zavisi od veli~ine uzorka

za R karte:

centralna linija je: $CL = d_2\,T$

kontrolne granice:

$GKG = D_2\,T$ $DKG = D_1\,T$

D_2, d_2 i D_1 su konstante iz tablice, čije vrednosti zavise od veličine uzorka.

Ako samo jedna tačka izlazi iz kontrolnih granica, proces nije pod kontrolom.

(Proces nije pod kontrolom ako uzastopnih 7 tačaka imaju odgovarajuću tendenciju kretanja koja ukazuje da će proces postati nestabilan.)

PRIMER

Tolerancije su 291 do 293

uzorak	merenje				$\overline{X}$	R
1	292.4	292.4	292.3	292.2	292.32	0.2
2	292.4	292.2	292.4	292.3	292.32	0.2
3	292.3	292.2	292.3	292.4	292.30	0.2
4	292.3	292.8	292.3	292.4	292.45	0.5
5	292.3	292.4	292.3	292.5	292.375	0.2
6	292.4	292.3	292.4	292.4	292.375	0.1
7	292.4	292.3	292.4	292.4	292.375	0.1
8	292.3	292.4	292.3	292.2	292.3	0.2
9	292.2	292.2	292.2	292.2	292.2	0.0
10	292.2	292.4	292.3	292.3	292.3	0.2
11	292.3	292.1	292.2	292.4	292.25	0.3
12	292.4	292.2	292.2	292.2	292.25	0.2
					3507.88	2.4

sredina i razmak varijacije: $\overline{X} = \dfrac{\sum X}{n}$ $R = X_{max} - X_{min}$

(292.4+282.4+292.3+292.2)/4=292.32

(292.4+292.2+292.4+292.3)/4=292.32

----

n=4 k=12 A_2=0.729

procesna sredina: $\overline{\overline{X}} = \dfrac{\sum_{i=1}^{k} \overline{X}_i}{k}$ =3507.88/12=292.32

srednji procesni razmak: $\overline{R} = \dfrac{\sum R_i}{k}$ =2.4/12=0.2

<u>konstrukcija X karte:</u>

Centralna linija kontrolne karte je: $\overline{\overline{X}} = \dfrac{\sum_{i=1}^{k} \overline{X}_i}{k}$ =3507.88/12=292.32

Gornja kontrolna granica je GKG = $\overline{\overline{X}} + A_2\overline{R}$ =292.31+0.729x0.2=292.458

Donja kontrolna granica je DKG = $\overline{\overline{X}} - A_2\overline{R}$ =292.31-0.729x0.2=292.174

konstrukcija R karte:

Centralna linija kontrolne karte je: CL = $\overline{R}$ =0.2

Gornja kontrolna granica je GKG =$D_4\overline{R}$ =2.282x02 =0.4564

Donja kontrolna granica je DKG =$D_3\overline{R}$ =0x0.2=0

Pošto je jedan razmak varijacije izvan granice, proces je nestabilan.

Tolerancije (testiranje hipoteze):

$T = 6\ \sigma_0$ $\sigma_0=T/6$

$T = T_g-T_d$ (T-toler)

T=293-291=2

σ_0=T/6= 2/6=0.33

n=4 A = 1.50 $D_1 = 0$ D_2=4.698 d_2=2.059 $\overline{X_0}$ =292 (izme|u 293 i 201)

za $\overline{X}$ karte:

centralna linija je: CL = X_0 (X_0 je sredina tolerancije)=292

kontrolne granice:

GKG = X_0+A σ_0=292+1.50x0.33=292.459

DKG = X_0 - A σ_0=292-1.50x0.33=291.505

Proces ispunjava standarde, jer se kontrolne granice nalaze u okviru standardom određenih granica tolerancije.

za R karte:

centralna linija je: CL = d_2 T

kontrolne granice:

GKG = $D_2\ \sigma_0$

DKG = $D_1\ \sigma_0$

D_2, D_2 i D_1 su konstante iz tablice, čije vrednosti zavise od veličine uzorka.

Kontrolne karte za atributivne karakteristike - P karte

Služi za praćenje statist.stabilnosti proporcije neispravnih proizvoda u uzorku. Postupak je:

- prati se proizvodni proces putem procesne karte
- ako se utvrdi da je proces stabilan, izvlači se na slučajan način u određenim vremenskim intervalima više uzoraka (k =20 do k=30). Veličina uzorka treba da bude n≥100.
- Računa se procesna proporcija: $\overline{p} = \dfrac{\sum m}{nk}$ odnosno: $\overline{p} = \dfrac{\sum p}{k}$

$\overline{p} = \dfrac{ukupan\ broj\ neispravnih\ \ proizvoda}{ukupan\ broj\ \ proizvoda\ u\ \ uzorku}$ Ovo je ujedno i centralna linija u kontrolnoj karti.

- Kontrolne granice:

GKG = $\overline{p} + 3\sqrt{\dfrac{\overline{p}\ (1-\overline{p})}{n}}$ DKG = $\overline{p} - 3\sqrt{\dfrac{\overline{p}\ (1-\overline{p})}{n}}$

Ako se desi da DKG ima negativne vrednosti, uzima se 0.

- kontrolne granice za zadati standard:

KG_{st}= $\overline{p} \pm 3\sqrt{\dfrac{p_0\,(1-p_0)}{n}}$

- konstruiše se karta
- ako se proporcije uzoraka približavaju DKG ili su ispod nje (ako je DKG >0) kontrolna karta ukazuje na poboljšanje kvaliteta procesa. Ako se proporcije nađu iznad GKG, karta ukazuje na poremećaj kvaliteta procesa.
- Proces je stat.stabilan i pod kontrolom ako su sve proporcije unutar granica.

  Testiranje hipoteze:

$$Z = \frac{\overline{\overline{X}} - \overline{X}}{S_{\overline{X}}} \qquad Z = \frac{\overline{p} - p}{\sqrt{\frac{p\ (1-p\ )}{n}}}$$, F(Z) iz tablice

9. ANALIZA VREMENSKIH SERIJA

Statističke zakonitosti se ispoljavaju na mnoštvu slučajeva, a utvrđuju se na osnovu skupa koji je vremenski opredeljen jednim trenutkom ili jednim periodom. Vreme je znači isključeno iz statističke analize (statička statistička analiza). Međutim, sve masovne pojave, a posebno ekonomske imaju osobinu da se tokom vremana razvijaju i podležu promenama. Otuda zahteva za dinamičkom statističkom analizom. To je primena statist.metoda analize masovnih pojava u toku vremena. Njen cilj je proučavanje i otkrivanje osobina procesa razvoja pojave, tj. obima promena, smera promena i zakonitosti razvoja. Analiza dinamike pojave pora da polazi od prirode pojave i biti u njenoj funkciji. Na razvoj pojave utiče niz faktora, koji se međusobno prepliću, gube ili menjaju dejstvo, tako da se moraju proučavati i kvantitativne i kvalitativne strane pojave, tj. kvalitativna anliza omogućava pravilno definisanje i tumačenje pojave preko njenih pokazatelja.

U dinamičkoj statističkoj analizi proučavano obeležje skupa naziva se pojavom. Obeležava se sa Y (u statičkoj statističkoj analizi se obeležava sa X). Vreme u čijoj se funkciji posmatra pojava obeležava se sa t (u statičkoj analizi se isključuje). Odnos između pojave i vremena u kojem se pojava ispoljava može se dati prostim analitičkim izrazom:

Y=f(t)

Nezavisna promenljiva, vreme, obuhvata sve faktore koji se ispoljavaju u njemu i koji dovode do dinamičkih promena proučavane pojave.

Predmet proučavanja statističke dinamičke analize su vremenske serije (serije dinamike). Vremenske serije su nizovi statističkih podataka o konkretnoj pojavi uređeni po vremenu. Vrem.serija pruža dve informacije: inf. o vremenu i inf. o pojavi. Svaka vremenska jedinica je član vrem.serije. Svakom članu pripada konačan nivo pojave. Sukcesivni niz članova i individualni nivoi pojave po članovima formiraju vrem,seriju. Odnosno, vrem.serija je formirani niz nivoa pojave u vremenu. Prvi član ima početni nivo (stanje) pojave, a poslednji ima krajnji nivo (stanje) pojave.

Svi nivoi (stanja) pojave moraju biti međusobno uporedivi, što je osnov objektivne dinamičke analize.

Vreme u kome se posmatra pojava mora se unapred odrediti. Jednaki vremenski periodi moraju biti sukcesivni i dati u hronološkom redosledu, čime je obezbeđen kontinuitet praćenja pojave.

Podaci o pojavi moraju biti homogeni, tj. moraju odražavati promene jedne iste pojave.

Nivoi (stanja) pojave moraju biti međusobno uporedivi po sadržaju.

Pošto je obim pojave tokom vremena podložan promenama, mora postojati velika obazrivost pri sastavljanju vrem.serija sa stanovišta obima.

Teritorijalni (prostorni) princip formiranja vrem.serije zahteva da se vrem.serija uvek odnosi na precizno definisanu i istu teritoriju (prostor)

(Parametri na osnovu kojih se definiše vrem.serija: - princip suštinske određenosti; - princip varijabilnosti; - princip sveobuhvatnosti; - princip prostorne određenosti; - princip vremenske određenosti)

Vrem.serije mogu sadržati apsolutne, relativne i srednje veličine pojave. Prvo se formira vrem.serija sa apsolutnim izrazom za nivo pojave, pa se na osnovu nje ili poređenjem sa drugom vrem.serijom formira vrem.serija relativnih ili srednjih veličina pojave.

Postoje dve vrste (dva pojavna oblika) vrem.serija, zavisno od karaktera pojave koja se proučava:

- momentne vrem.serije
- intervalne vrem. serije

Momentne vrem.serije pokazuju stanje obima pojave, njene relativne ili srednje veličine u sukcesivnom nizu momenata vremena. One su prekidne. Stanaj pojave se ne mogu sumirati, ni deliti.

Kumulativ ovih serija nema ekonomski smisao. Međutim, razlika stanja pojave u sukcesivnim momentima vremana ima ekonomskog smisla. Razmaci između sukcesivnih momenata vremana u kojima se posmatra pojava moraju biti iste veličine. Razmaci se nazivaju inervalima. Oni mogu biti od mesec dana, kvartali itd.

Intervalne vrem.serije (periodične) pokazuju tok pojave u sukcesivnim vremenskim periodima. One su neprekidne. Sumiranje podataka ima ekonomskog smisla, kao i deljenje nivoa pojave jednog intervala na uže intervale (godina na kvartale, mesece...). Veličina pojave je srazmerna veličini intervala u kome se posmatra.

Izvori statist.podataka za vrem.serije su: izveštaji, ankete i popisi.

Analiza komponenti vrem.serija polazi od varijacija pojave koje nastaju pod dejstvom sistematskih i slučajnih faktora. Kod veličine vrem.serija mogu se izdvojiti 4 grupe faktora koje se de{avaju dovoljno redovno da se mogu formirati odgovarajuća komponente vrem.serija:

- trend (T) (trend je opšti pravac razvoja pojave; promene nastaju pod uticajem faktora koji neprekidno deluju na duži rok; pojave se najčešće prate u godišnjim periodima)
- sezonska kolebanja (S) (pojave se prate u kraćim intervalima od godišnjih - mesec, kvartal)
- ciklična kolebanja (C) (to su oscilatorna gibanja pojave - smenjuju se periodi rasta i periodi opadanja pojave; pojave se prate u godišnjim periodima
- neregularna kolebanja (N) (nastaju pod dejstvom slučajnih faktora koji izazivaju poremećaje u razvoju pojave)

Statistički model vrem.serije može biti:

- multiplikativni (ovaj model polazi od toga da pojava Y predstavlja rezultantu dejstva 4 faktora, tj. Y = T S C N; faktori se mogu množiti; odnos faktora je relativan; komponenta T se iskazuje u veličinama pojave; ostale komponente su date u procentima)
- aditivni (ovaj model polazi od toga da pojava Y predstavlja zbir 4 komponente, tj. Y=T+S+C+N; komponente se sabiraju i međusobno su nezavisne)
- hibridni (kombinacija prethodna dva modela; varijante su: Y=C+T S N; Y=S+T C N)

  Postupak dekompozicije vrem.serija u statističku praksu je uveo američki naučnik W.Persones (Pirson). Vrši se tako što se najpre otkriva trend efekat (T) koji se izražava u jedinicama mere u kojima se izražava i pojava. Zatim se određuje sezonski efekat (S), kao procenat sez.kolebanja. Određivanje cikličnog efekta (C) još nema teorijski model, jer predstavlja višegodišnje udaljavanje iznad i ispod razvojne trend komponente i da nije strogo periodično. Zato se mogu odrediti zajedno sa neregularnim kolebanjima deljenjem originalnih veličina vrem.serije sa proizvodom sezonskog i trend efekta, tj.:

  Y / T S

  Aditivni metod vrem.serije je: Y = T+C+N+S, kako se sezonski efekat gubi za godišnje nivoe pojave, to je: Y = T+C+N. Određivanjem trend efekta i njegovim odstranjivanjem iz originalnih podataka, dobija se: Y - T = C+N. Dobijaju se apsolutne vrednosti cikličnog i neregularnog efekta. Njihove relativne vrednosti su: (Y - T) / T x 100

  Pokazatelji vrem.serija su:

- apsolutni
- relativni

(zavisno od toga da li se iskazuju u apsolutnim jedinicama u kojima se izražava i konkretna pojava ili u relativnim jedinicama u obliku koeficijenata, procenata i promila)

Apsolutni pokazatelji su:

- hronološka sredina
- varijansa i stand.devijacija
- apsolutni prirast
- srednji apsolutni prirast
- apsolutna veličina jednog procenta prirasta

Relativni pokazatelji su:

- koeficijent varijacije
- tempo (stopa) rasta

- tempo (stopa) prirasta
- srednji tempo (stopa) rasta
- srednji tempo (stopa) prirasta

Postupak eliminacije neregularne komponente (N) vrem.serije naziva se <u>usklađivanje.</u> metodi usklađivanja vrem.serije su:

- <u>metod pokretnih sredina</u>
- <u>metod eksponencijalnog usklađivanja</u>

<u>Metod pokretnih sredina</u>

Najčešće korišćeni metod usklađivanja vrem.serija. Omogućava eliminisanje efekata kratkoročnih i slučajnih promena pojave. Suština metode je da se umesto originalnog podatka uzima sredina tog podatka, prethodnih i narenih podataka. Time se dobija prosečno kretanje pojave i ističe se osnovna tendencija (trend) razvoja pojave

Pokretne sredine se mogu formirati na osnovu 3,4,5 i više članova vrem.serije. Broj članova određuje red pokretnih sredina. Slabost metode je što je dozvoljena sloboda izbora reda pokretnih sredina.

a) *Pokretne sredine neparnog reda*:

- pokretna sredina III reda se formira od pojave kod tri člana vrem.serije i to ovako: pojava kod drugog člana vrem.serije se zamenjuje sredinom pojave prvog, drugog i trećeg člana. Pojava kod trećeg člana vrem.serije se zamenjuje sredinom pojave drugog, trećeg i četvrtog člana vrem.serije i tako redom. Pri tome, prvi i poslednji član vrem.serije nemaju pokretnu sredinu.

pokretna sredina III reda je: $\overline{Y_2} = \dfrac{Y_1 + Y_2 + Y_3}{3}$, $\overline{Y_3} = \dfrac{Y_2 + Y_3 + Y_4}{3}$, ...

pokretna sredina V reda je:

$$\overline{Y_3} = \frac{Y_1 + Y_2 + Y_3 + Y_4 + Y_5}{5}, \qquad \overline{Y_4} = \frac{Y_2 + Y_3 + Y_4 + Y_5 + Y_6}{5}, \quad ......$$

b) Pokretne sredine parnog reda:

Ne mogu se pripisati ni jednom članu vrem.serije. Tako je pokretna sredina IV reda:

$$\overline{Y_{2/3}} = \frac{Y_1 + Y_2 + Y_3 + Y_4}{4}, \qquad \overline{Y_{3/4}} = \frac{Y_2 + Y_3 + Y_4 + Y_5}{4}, \quad ......$$

da bi se pripisale određenom članu vrem.serije, vrši se centriranje pokretnih sredina. Centrirana pokretna sredina se dobija zbrajanjem susednih pokretnih sredina parnog reda. Tako je:

$$\overline{Y_3} = \frac{Y_{2/3} + Y_{3/4}}{2}, \qquad \overline{Y_4} = \frac{Y_{3/4} + Y_{4/5}}{2}, \quad ......$$

Često se koriste i *ponderisane pokretne sredine*.

Za pokretne sredine reda III, sistem ponderacije je (1,2,1), a reda V je (1,2,3,2,1) ili (1,2,2,2,1). Uvek je ponderacioni faktor jednak 1.

Za pokretne sredine reda III i V, posle primene ponderacionog faktora imamo:

$$\overline{Y_2} = \frac{Y_1 + 2Y_2 + Y_3}{4}, \qquad \overline{Y_3} = \frac{Y_2 + 2Y_3 + Y_4}{4}, \quad ......$$

$$\overline{Y_3} = \frac{Y_1 + 2Y_2 + 3Y_3 + 2Y_4 + Y_5}{9}, \qquad \overline{Y_4} = \frac{Y_2 + 2Y_3 + 3Y_4 + 2Y_5 + Y_6}{9}, \quad ......$$

ili

$$\overline{Y_3} = \frac{Y_1 + 2Y_2 + 2Y_3 + 2Y_4 + Y_5}{8}, \qquad \overline{Y_4} = \frac{Y_2 + 2Y_3 + 2Y_4 + 2Y_5 + Y_6}{8}, \quad ......$$

PRIMER.

godina	proizvodnja	pok.sr.reda III	pok.sr.reda V	pond.pok.sred.reda V
1987	6045	-	-	-
1988	6671	6721	-	-
1989	7447	7391	7247	7317

1990	8054	7840	8126	7317
1991	8019	8838	8815	8734
1992	10440	8838	8815	8734
1993	10117	10699	10585	10571
1994	11540	11488	11359	11422
1995	12807	12080	11795	12002
1996	11892	12439	12084	12180
1997	12617	12024	-	-
1998	11562	-	-	-

$$\overline{Y_2} = \frac{Y_1 + Y_2 + Y_3}{3} = 6045 + 6671 + 7447/3 = 6721, \quad \overline{Y_3} = \frac{Y_2 + Y_3 + Y_4}{3} = 6671 + 7447 + 8054/3 = 7391$$

$$\overline{Y_3} = \frac{Y_1 + Y_2 + Y_3 + Y_4 + Y_5}{5} = 6045 + 6671 + 7447 + 8054 + 8019/5 = 7247$$

$$\overline{Y_4} = \frac{Y_2 + Y_3 + Y_4 + Y_5 + Y_6}{5} = 6671 + 7447 + 8054 + 8019 + 10440/5 = 8126$$

$$\overline{Y_3} = \frac{Y_1 + 2Y_2 + 3Y_3 + 2Y_4 + Y_5}{9} = 6045 + 2x6671 + 3x7447 + 2x8054 + 8019/9 = 7317$$

$$\overline{Y_4} = \frac{Y_2 + 2Y_3 + 3Y_4 + 2Y_5 + Y_6}{9} = 6671 + 2x7447 + 3x8054 + 2x8019 + 10440/9 = 7317$$

TREND

Kod mnogih dinamičkih pojava izraženih vrem.serijama uočava se stalni rast ili opadanje veličine pojave. To se posebno uočava na duži rok. Trend je opšta razvojna tendencija pojave, nastala pod dejstvom postojanih faktora. Oni obično deluju u dužem vremenskom periodu, u sitom pravcu, istim intenzitetom i mogu se kontrolisati. Zato određivanje trenda omogućava analizu razvojne tendencije pojave i predviđanje u budućnosti.

Trend možemo odrediti matematički pomoću odgovarajuće funkcije. Koji je to oblik funkcije, zavisi od kretanja pojave. Pretpostavke o tipu f-je trenda daju i numerički postupak poznat kao metod diferencija, kojim se ispituje prirast (razlika) pojave po članovima vrem.serije. Ako su povećanja ili smanjenja pojave, od jedne do druge vremenske jedinice približno ista, koristi se f-ja prave. Ako se pojava stalno uvećava ili stalno smanjuje - bira se kriva linija višeg stepena. Ako pojava u svom razvoju pokazuje relativan porast ili pad, tj. menja se po geometrijskoj stopi - bira se eksponencijalna f-ja.

U opštem slučaju, na oblik trenda može ukazati grafički prikaz dinamičke pojave, ali i neki statistički model. Najčešće se koristi *metod pokretnih sredina*.

Pošto se izračuna f-ja trenda, mogu se odrediti vrednosti trenda za ceo period posmatranja pojave. Taj postupak se naziva interpolacija trenda. Na osnovu utvrđene f-je trenda mogu se izračunati i vrednosti trenda za naredni period. To je ekstrapolacija trenda. Realniji podaci se dobijaju ako je period predviđanja kraći. Bitna pretpostavka predviđanja budućeg razvoja pojave je da svi faktori koji su delovali na pojavu nastave da deluju i u budućnosti.

Statistički model trenda ima dva oblika:

- aditivni $Y=f(t)+\varepsilon$
- multiplikativni $Y=f(t)\,\varepsilon$

u oba slučaja su u pitanju regresioni modeli; osnovna forma statist.modela trenda je aditivna; multiplikativna se prevodi u aditivnu (najčešće logaritmovanjem); t-vreme je nezavisna promenljiva, Y-pojava je zavisna; fiksni deo modela f(t) govori o funkcionalnoj vezi specijalnog tipa; nezavisni deo ε definiše mogućnost odstupanja vrednosti pojave i ima sve odlike slučajne promenljive i predstavlja razliku između stvarne vrednosti i vrednosti koja je ušla u funkcionalnu vezu.

(Zamenimo u statist.modelu trenda $Y=f(t)+\varepsilon$, $f(t)=Y_t$. Dobija se $Y=Y_t+\varepsilon$, pa je $\varepsilon=Y-Y_t$; pošto Y_t predstavlja niz sredina, to je, zbog svojstva sredine: $\sum(Y-Y_t)=0$ i $\sum(Y-Y_t)^2=\min$; ispada da je teorijska f-ja trenda najbolje prilagođena funkcija empirijskim podacima, jer minimizira nezavisni deo modela).

Statist.modeli trenda mogu biti: linearni i nelinearni; eksponencijalni; logaritamski...

Linerani trend

Primenjuje se onda kada su povećanja ili smanjenja pojave u vrem.seriji približno ista. Grafički prikaz ovakvih serija izgleda tako da se originalni podaci raspore|uju približno pravolinijski, pa se između njih može povući prava linija koja reprezentuje dinamiku posmatrane pojave. Matematički model ovog tipa funkcije određuje se najčešće metodom najmanjih kvadrata.

Opšti oblik linearne funkcije trenda je: $Y_t=\beta_0+\beta_1 t$, a ocenjena funkcija lineranog trenda je: $Y_t=b_0+b_1 t$. Statistike b_0 i b_1 se nalaze metodom najmanjih kvadrata:

$$b_1=\frac{\sum Yt-\frac{(\sum Y)(\sum t)}{n}}{\sum t^2-\frac{(\sum t)^2}{n}} \qquad b_0=\frac{\sum Y-b\sum t}{n}$$

nezavisna promenljiva t (godišnji vremenski period) uzima vrednosti 1,2,...,n sa stalnim priraštajem $\Delta t=1$. Statistike b_0 nema posebno značenje. Statistike b_1 je koeficijent funkcije trenda, tj. koeficijent pravca prave linije trenda. Pokazuje koliko se povećava ili smanjuje pojava kada se povećava vremenski period za 1 (jedinicu). Zavisno od znaka (+ ili -) izpred statistike b_1 funkcija linearnog trenda je rastuća ili opadajuća.

standardna greška ocenjene funkcije linearnog trenda je:

$$S_e=\sqrt{\frac{\sum(Y-\overline{Y}_t)^2}{n-2}}$$

ocena varijanse pojave Y je: $S_Y^2=\frac{\sum(Y-\overline{Y})^2}{n-1}$

indeks determinacije: $R_Y^2=1-\frac{S_e^2}{S_Y^2}$

Uticaj ostalih (CN) faktora na razvoj pojave je: $(Y\ /\ Y_t)\ 100$. Ako je odnos veći od 100, ostali faktori dejstvuju na povećanje pojave iznad proseka, a ko je odnos manji od 100, ostali faktori dejstvuju na smanjenje pojave ispod proseka.

PRIMER

godine	Y (pojava)	t	Y t	t^2	Y_t	$(Y\text{-}Y_t)$ 100
1977	26.0	1	26.0	1	25.60	101.33
1978	29.5	2	59.0	4	28.90	102.08
1979	33.2	3	99.6	9	32.14	103.30
1980	35.1	4	140.4	16	35.38	99.21
1981	39.5	5	197.5	25	38.62	102.28
1982	40.0	6	240.0	36	41.86	95.56
1983	43.6	7	305.2	49	45.11	96.65
1984	48.6	8	388.8	64	48.35	100.52
1985	51.3	9	461.7	81	51.59	99.4
1986	55.0	10	550.0	100	54.83	100.31
1987	59.4	11	653.4	121	58.07	102.29
1988	60.4	12	724.8	144	61.31	98.52
1989	62.3	13	809.9	169	64.56	96.51
1990	67.6	14	946.4	196	67.79	99.72
1991	73.0	15	1095.0	225	71.03	102.77
1992	74.8	16	1196.8	256	74.28	100.71
1993	77.7	17	1320.9	289	77.52	100.23

ukupno	877.0	153	9215.4	1785	877.00	*

$$b_1 = \frac{\sum Y t - \frac{(\sum Y)(\sum t)}{n}}{\sum t^2 - \frac{(\sum t)^2}{n}} = \frac{9215.4 - \frac{(153)(877)}{17}}{1785 - \frac{153^2}{17}} = 3.24118$$

$$b_0 = \frac{\sum Y - b \sum t}{n} = \frac{877.0 - (3.24118)(153)}{17} = 22.419$$

Ocenjena funkcija linearnog trenda je: $Y_t = b_0 + b_1\, t = 22.419 + 3.24\, t$
za t=1,2,...17, dobijamo (interpolisane vrednosti trenda):
$Y_{1977} = b_0 + b_1\, 1 = 22.419 + (3.24)\, 1 = 25.66$
$Y_{1978} = b_0 + b_1\, 2 = 22.419 + (3.24)\, 2 = 28.90$
....
Uticaj ostalih (CN) faktora: $(Y / Y_t)\, 100$

$$\left(\frac{Y_{1977}}{Y_{1977}}\right)100 = \left(\frac{26.0}{25.66}\right)100 = 101.33$$

$$\left(\frac{Y_{1978}}{Y_{1978}}\right)100 = \left(\frac{29.5}{28.9}\right)100 = 102.08$$

Ekstrapolacija trenda se vr{i kada se u funkciju trenda unose vrednosti za t=18,19...
$Y_{1994} = b_0 + b_1\, 18 = 22.419 + (3.24)\, 18 = 80.76$
$Y_{1995} = b_0 + b_1\, 19 = 22.419 + (3.24)\, 19 = 84.0$

SEZONSKA KOLEBANJA

Sezone su periodi u toku godine. One su strogo periodične. Dele se na mesece i kvartale. Sezone utiču na mnoge pojave. Ispoljavaju se kao varijacije pojava. Sezonske varijacije su posebna komponenta vrem.serije. Da bi se ispitao uticaj sezonskih varijacija, ostale komponente vrem.serije se moraju stabilizovati, a to se postiže formiranjem višegodišnje vrem.serije po mesecima ili kvartalima. Time se otkrivaju zakonitosti razvoja sezonske komponente, a ublažavaju se eventualni slučajni uticaji.

Statističko proučavanje sezonske komponente obuhvata: ispitivanje sezonskog ritma pojave, sezonsko usklađivanje i predviđanje budućeg razvoja pojave.

Sezonski ritam utvrđuju sezonski indeksi. Za ispitivanje sezonskog ritma kod stabilnih pojava najviše se koristi ***Metod odnosa prema opštoj sredini*** kod koga se sezonski indeksi konstruišu tako što se sredine pojedinih istoimenih perioda, meseci ili kvartala, izračunate na osnovu podataka za više godina, stavljaju u odnos sa sredinom svih perioda (svih meseci ili svih kvartala). Isključen je uticaj cikličnih i neregularnih kolebanja. Dobijeni koeficijenti se pomnože sa 100, čime se dobija sezonski indeks. Ako je sez.indeks > 100, pod uticajem sezone dolazi do porasta pojave, a < 100, smanjuje se pojava pod uticajem sezone.

Obeležimo sa Y pojavu, sa i (i=1,2,..12 ili kvartali i=1,2,3,4), sa j (j=1,2,..m) označimo godine. Onda Y_{ij} je nivo pojave u i-tom mesecu (kvartalu) j-te godine.

Sredine za mesec su:

$$\overline{Y_1} = \frac{\sum_{j=1}^{m} Y_{1j}}{m} \qquad \overline{Y_2} = \frac{\sum_{j=1}^{m} Y_{2j}}{m} \qquad \overline{Y_3} = \frac{\sum_{j=1}^{m} Y_{3j}}{m} \ldots\ldots\ldots\ldots \overline{Y_{12}} = \frac{\sum_{j=1}^{m} Y_{12}}{m}$$

Sredine za kvartale su:

$$\overline{Y_I} = \frac{\sum_{j=1}^{m} Y_{1j}}{m} \qquad \overline{Y_{II}} = \frac{\sum_{j=1}^{m} Y_{2j}}{m} \qquad \overline{Y_{III}} = \frac{\sum_{j=1}^{m} Y_{3j}}{m}. \qquad \overline{Y_{IV}} = \frac{\sum_{j=1}^{m} Y_{4j}}{m}$$

Opšta mesečna sredina je:

$$\overline{Y_{mes}} = \frac{\sum_{i-1}^{12}\sum_{j=1}^{m} Y_{ij}}{12m}$$

Opšta kvartalna sredina je:

$$\overline{Y_{k\,\text{var}}} = \frac{\sum_{i-1}^{4}\sum_{j=1}^{m} Y_{ij}}{4m}$$

Sezonski indeksi se obeležavaju sa SI i dobijaju se iz odnosa sredina (zato se zovu sezonski srednji indeksi).

Mesečni sezonski indeksi su:

$$SI_1 = (\frac{\overline{Y_1}}{\overline{\overline{Y_{me\,\text{sec}}}}})100 \ldots\ldots\ldots\ldots\ldots SI_{12} = (\frac{\overline{Y_{12}}}{\overline{\overline{Y_{me\,\text{sec}}}}})100.$$

Kvartalni sezonski indeksi su:

$$SI_I = (\frac{\overline{Y_I}}{\overline{\overline{Y_{k\,\text{var}tal}}}})100 \qquad SI_{II} = (\frac{\overline{Y_{II}}}{\overline{\overline{Y_{k\,\text{var}tal}}}})100 \qquad SI_{III} = (\frac{\overline{Y_{III}}}{\overline{\overline{Y_{k\,\text{var}tal}}}})100 \qquad SI_{IV} = (\frac{\overline{Y_{IV}}}{\overline{\overline{Y_{k\,\text{var}tal}}}})100$$

Zbir izračunatih sez.indeksa po mesecima mora biti 1200, a po kvartalima 400. Ako nije tako, treba izvršiti korekciju sez.indeksa. Koeficijenti za korekciju su **korektivni faktori**. Kor.faktor po mesecima je:

$$c = \frac{1200}{\sum_{i=1}^{12} SI_i}$$

Kor.faktor po kvartalima: $c = \frac{400}{\sum_{i=1}^{4} SI_i}$

Korigovani sez.indeksi su: $SI_i^{,} = (SI_i)c$

Otklanjanje uticaja sezone na pojavu vrši se ***desezonizacijomvrem.serije***. Nivo pojave iz koje je otklonjen uticaj sezone dobija se iz odnosa originalnog nivoa pojave i sez.indeksa, pomnoženo sa 100:

$$Y_{ij}^{,} = (\frac{Y_{ij}}{SI_i})100$$

Najviše se koriste 2 metoda konstrukcije sez.indeksa:

- metod odnosa prema pokretnim sredinama
- metod odnosa prema trend vrednostima

Metod odnosa prema pokretnim sredinama

Pokretna sredina je sredina kvartala ili meseci (12 sukcesivnih meseci, 4 sukcesivna kvartala). Stavljanjem u odnos originalnih vrednosti pojave sa odgovarajućim pokretnim sredinama, dobijaju se sez.koeficijenti, pomnoženi sa 100 daju sez.indekse.

1. Za vrem.seriju po kvartalima

Sredine kvartala su: $\overline{Y_{2/3}} = \frac{Y_1 + Y_2 + Y_3 + Y_4}{4}$, $\qquad \overline{Y_{3/4}} = \frac{Y_2 + Y_3 + Y_4 + Y_5}{4}$,

Pošto su dobijene sredine nastale iz parnog broja kvartala, ne mogu se staviti u odnos sa originalnim vrednostima. Mora se izvršiti njihovo centriranje.

Centrirane pokretne sredine su: $\overline{Y_3} = \frac{Y_{2/3} + Y_{3/4}}{2}$, $\qquad \overline{Y_4} = \frac{Y_{3/4} + Y_{4/5}}{2}$

Stavljanje u odnos originalnih veličina i centriranih pokretnih sredina istoimenih članova, dobijaju se **sez.koeficijenti** (k):

$$-,-,k_3 = \frac{Y_3}{\overline{\overline{Y_3}}}, k_4 = \frac{Y_4}{\overline{\overline{Y_4}}},\ldots,k_{m-2} = \frac{Y_{m-2}}{\overline{\overline{Y_{m-2}}}},-,-$$

Sredine sez.koeficijenata, pomnožene sa 100 daju **sez.indekse**:

$$SI_I = (\overline{k}_I)100 \quad SI_{II} = (\overline{k}_{II})100 \quad SI_{III} = (\overline{k}_{III})100 \quad SI_{IV} = (\overline{k}_{IV})100$$

Zbir sez.indeksa je 400. Ako nije tako, radi se korekcija sez.indeksa. Kor.faktor (koef.za ispravku) je: $c = \frac{400}{\sum_{i=1}^{4} SI_i}$

Kor.faktorom se množe sez.indeksi za svaki kvartal, tj. $SI_i^{\prime} = (SI_i)c$

PRIMER (

godina	I kvartal	II	III	IV
1993	20.4	18.7	28.0	29.3
1994	23.0	21.0	31.0	33.0
1995	24.6	22.6	33.4	35.4
1996	27.2	25.0	36.8	39.0
1997	29.0	26.5	39.1	41.5
1998	31.5	28.2	41.4	44.0

Da bi se metodom pokretnih sredina rešio zadatak, mora se formirati nova radna tabela. Prva kolona su godine, druga su kvartali, treća je redosled kvartala, četvrta su originalne veličine pojave:

godina	kvartali	red.br.kvart	Y_i	$\overline{Y_i}$	k_i
1993	I	1	20.4	-	-
	II	2	18.7	-	-
	III	3	28.0	24.425	1.146
	IV	4	29.3	25.0375	1.170
1994	I	5	23.0	25.7	0.909
	II	6	21.0	26.5375	0.791
	III	7	31.0	27.2	1.140
	IV	8	33.0	27.8	1.204
1995	I	9	24.6	28.1	0.875
	II	10	22.6	28.7	0.787
	III	11	33.4	29.325	1.139
	IV	12	35.4	29.95	1.182
1996	I	13	27.2	30.675	0.887
	II	14	25.8	31.55	0.804
	III	15	36.8	32.225	1.142
	IV	16	39.9	32.6375	1.195
1997	I	17	29.0	33.1125	0.876
	II	18	26.5	33.7125	0.786
	III	19	39.1	34.3375	1.139
	IV	20	41.5	34.8625	1.190
1998	I	21	31.5	35.3625	0.891
	II	22	28.2	35.9625	0.784
	III	23	41.4	-	-
	IV	24	44.0	-	-

Sredine kvartala su:

$$\overline{Y_{2/3}} = \frac{Y_1 + Y_2 + Y_3 + Y_4}{4} = \frac{20.4 + 18.7 + 28.0 + 29.3}{4} = 24.1$$

$$\overline{Y_{3/4}} = \frac{Y_2 + Y_3 + Y_4 + Y_5}{4} = \frac{18.7 + 26.0 + 29.3 + 23.0}{4} = 24.75$$

....

Pošto su dobijene sredine nastale iz parnog broja kvartala, ne mogu se staviti u odnos sa originalnim vrednostima. Mora se izvršiti njihovo centriranje. Centriranje sredine se vrši računanjem sredine iz sukcesivnog niza dve sredine.

Centrirane pokretne sredine su:

$$\overline{Y_3} = \frac{Y_{2/3} + Y_{3/4}}{2} = \frac{24.1 + 24.75}{2} = 24.425$$

$$\overline{Y_4} = \frac{Y_{3/4} + Y_{4/5}}{2}$$

....

Stavljanje u odnos originalnih veličina i centriranih pokretnih sredina istoimenih članova, dobijaju se **sez.koeficijenti** (k):

$$-,-,k_3 = \frac{Y_3}{\overline{Y_3}}, k_4 = \frac{Y_4}{\overline{Y_4}}, \ldots, k_{m-2} = \frac{Y_{m-2}}{\overline{Y_{m-2}}}, -,-$$

$$k_3 = \frac{Y_3}{\overline{Y_3}} = \frac{28.0}{24.425} = 1.146, \quad k_4 = \frac{Y_4}{\overline{Y_4}} = \frac{29.3}{25.0375} = 1.170 \ldots\ldots\ldots\ldots\ldots\ldots$$

Nakon izra~unavanja sez.koeficijenata, dodaju se nove kolone ili se formira nova tabela.

	1993	1994	1995	1996	1997	1998	$\overline{k}$	SI	SI'
I	-	0.908	0.875	0.887	0.876	0.891	0.8874	88.74	88.58
II	-	0.791	0.787	0.804	0.786	0.784	0.7904	79.04	78.90
III	1.146	1.140	1.139	1.142	1.139	-	1.1412	114.12	113.91
IV	1.170	1.204	1.182	1.195	1.190	-	1.1882	118.82	118.61

Sredine sez.koeficijenata po kvartalima su:

$$\overline{k}_I = \frac{k_5 + k_9 + k_{13} + k_{17} + k_{21}}{5} = \frac{0.908 + 0.875 + 0.887 + 0.876 + 0.891}{5} = 0.8874$$

$$\overline{k}_{II} = \frac{k_6 + k_{10} + k_{14} + k_{18} + k_{22}}{5} = \frac{0.791 + 0.787 + 0.804 + 0.786 + 0.784}{5} = 0.7904$$

....

n=5 je broj sez.koef. po kvartalima.

Sez.indeksi se dobijaju kada se sredine sez.koef. pomnože sa 100:

$$SI_I = (\overline{k}_I)100 \quad SI_{II} = (\overline{k}_{II})100 \quad SI_{III} = (\overline{k}_{III})100 \quad SI_{IV} = (\overline{k}_{IV})100$$

Njihov zbir treba da je 400:

$SI_I + SI_{II} + SI_{III} + SI_{IV} = 88.74 + 79.04 + 114.12 + 118.82 = 400.72$

Pošto se razlikuje od 400, mora se uraditi korekcija. Kor.faktor (koef.za ispravku) je:

$$c = \frac{400}{\sum_{i=1}^{4} SI_i} = \frac{400}{400.72} = 0.9982$$

Kor.faktorom se množe sez.indeksi za svaki kvartal, tj. korigovani sez.indeksi su:

$SI_I^{'} = (SI_I)c = 88.74 \cdot 0.9982 = 88.58$

$SI_{II}^{'} = (SI_{II})c = 79.04 \cdot 0.9982 = 78.90$

$SI_{III}^{'} = (SI_{III})c = 114.12 \cdot 0.9982 = 113.91$

$SI_{IV}^{'} = (SI_{IV})c = 118.82 \cdot 0.9982 = 118.61$

2. *Za vrem.seriju po mesecima*

Princip je isti kao i za konstrukciju sez.indeksa po kvartalima.
Sredine meseci su:

$$\overline{Y_{6/7}} = \frac{Y_1 + Y_2 + ... + Y_{12}}{12} \qquad \overline{Y_{7/8}} = \frac{Y_2 + Y_3 + ... + Y_{13}}{12} ....$$

Centrirane pokretne sredine su:

$$\overline{Y_7} = \frac{\overline{Y}_{6/7} + \overline{Y}_{7/8}}{2}, \qquad \overline{Y_8} = \frac{\overline{Y}_{7/8} + \overline{Y}_{8/9}}{2} ....$$

Prvih i poslednjih 6 članova vrem.serije nemaju pokretne centrirane sredine.

Sez.koeficijenti su: $-,-,-,-,-,-,k_7 = \frac{Y_7}{\overline{Y_7}}, k_8 = \frac{Y_8}{\overline{Y_8}}, ...,-,-,-,-,-,-$

(n= je broj sez.koef. po kvartalima)

Sez.indeksi se dobijaju kada se sredine sez.koef. pomnože sa 100:

$SI_1 = (\overline{k}_1)100 \quad SI_2 = (\overline{k}_2)100 \quad ... \quad SI_{12} = (\overline{k}_{12})100$

Njihov zbir treba da je 1200. Ako nije, radi se korekcija.

Kor.faktor (koef.za ispravku) je: $c = \frac{1200}{\sum_{i=1}^{12} SI_i}$

Korigovani sez.indeksi su: $SI_i^{'} = (SI_i)c$. Radi se za svaki mesec.

<u>Metod odnosa prema trend vrednostima</u>

Faze ove metode:

- utvrđuje se funkcija trenda
- u f-ju trenda se unose veličine za mesece ili kvartale, da bi se dobile vrednosti trenda za sve članove vrem.serije
- u odnos se stavljaju originalne veličine pojave i trend vrednosti članova vrem.serije i dobijaju se sez.koeficijenti
- sredina sez.koeficijenata meseci ili kvartala, pomnožena sa 100 daje sez.indekse
- ako ∑SI# 400 ili ∑SI#1200 radi se korekcija sez.indeksa. Kor.faktor za kvartale je: c=400/∑SI, a za mesece je: c=1200/∑SI

PRIMER (

god.	kvartali	Y	t	Y t	t^2	$\overline{Y_t}$
1993	I	140	1	140	1	114.95
	II	97	2	194	4	116.17
	III	90	3	270	9	117.40
	IV	136	4	544	16	118.62
1994	I	143	5	715	25	119.85
	II	102	6	612	36	121.08
	III	94	7	658	49	122.30
	IV	143	8	1144	64	123.53
1995	I	155	9	1395	81	124.75
	II	105	10	1050	100	123.98
	III	97	11	1067	121	127.21
	IV	149	12	1788	144	128.43

1996	I II III IV	161 109 102 154	13 14 15 16	2093 1526 1530 2464	169 196 225 256	129.66 130.88 132.11 133.34
1997	I II III IV	168 114 106 160	17 18 19 20	2856 2052 2014 3200	289 324 361 400	134.56 135.79 137.01 138.24
1998	I II III IV	175 118 109 168	21 22 23 24	3675 2596 2507 4032	441 484 529 576	139.47 140.69 141.92 143.14
$\sum$		3097	300	40122	4900	3097.08

Koristi se linearni trend.
Statistike b_0 i b_1 se nalaze metodom najmanjih kvadrata:

$$b_1 = \frac{\sum Y t - \frac{(\sum Y)(\sum t)}{n}}{\sum t^2 - \frac{(\sum t)^2}{n}} = \frac{40122 - \frac{3097 \cdot 300}{24}}{4900 - \frac{300^2}{24}} = 1.226$$

$$b_0 = \frac{\sum Y - b_1 \sum t}{n} = \frac{3077 - 1.226 \cdot 300}{24} = 113.72$$

Ocenjena f-ja linearnog trenda je: $Y_t=b_0+b_1$ t=113.72+1.226 t
Dajemo t=1,2,...,24 i dobijamo vrednosti trenda po kvartalima:
$Y_{1993}=b_0+b_1$ 1 =113.72+1.226x1=114.95
$Y_{1994}=b_0+b_1$ 2 =113.72+1.226x2=116.17
....
Pravimo novu tabelu (ili da dodamo nove 3 kolone u prethodnoj: $(Y/Y_t)100$, $\bar{k}$, SI).

kvar.	1993	1994	1995	1996	1997	1998	$\bar{k}$	SI
I II III IV	1.218 0.835 0.767 1.147	1.193 0.842 0.769 1.158	1.242 0.833 0.763 1.160	1.242 0.833 0.772 1.155	1.249 0.840 0.774 1.157	1.255 0.839 0.768 1.174	1.233 0.837 0.769 1.159	123.3 83.7 76.9 115.9
								399.8

Stavljamo u odnos originalne vrednosti i trend vrednosti, pa dobijamo sez.koeficijente:
za I kvartal:

140/114.95=1.218, 143/119.85=1.193, 155/124.75=1.242,....

za II:

97/116.17=0.835, 102/121.08=0.842, 105/125.98=0.833,...

....

Sez.indeksi se dobijaju kada se sredine sez.koef. pomnože sa 100:

$SI_I = (\bar{k}_I)100 \quad SI_{II} = (\bar{k}_{II})100 \quad SI_{III} = (\bar{k}_{III})100 \quad SI_{IV} = (\bar{k}_{IV})100$

Njihov zbir treba da je 400:

$SI_I+SI_{II}+SI_{III}+SI_{IV}$=123.3+83.7+76.9+115.9=399.8

Pošto se razlikuje od 400, mora se uraditi korekcija. Kor.faktor (koef.za ispravku) je:

$$c = \frac{400}{\sum_{i=1}^{4} SI_i} = \frac{400}{399.8} = 1.005$$

Kor.faktorom se mno`e sez.indeksi za svaki kvartal, tj. korigovani sez.indeksi su:

$SI'_I = (SI_I)c = 123.3 \cdot 1.005 = 123.4$

$SI'_{II} = (SI_{II})c = 83.7 \cdot 1.005 = 83.7$

$SI'_{III} = (SI_{III})c = 76.9 \cdot 1.005 = 76.9$

$SI'_{IV} = (SI_{IV})c = 115.9 \cdot 1.005 = 116$

Može se reći:

ako podaci pokazuju varijacije u okviru jedne godine, a ne varijacije između godina, koristi se metod odnosa prema opštoj sredini.

ako ako je polarni dijagram kao spirala, koristi se metod odnosa prema trendu