KVANTITATIVNE METODE U GRAĐEVINSKOM MENADŽMENTU

predavanja 2017/18

METODA UZORAKA I TEORIJA OCJENA

1. **Statistike:definicija, raspodjele i standardne greške statistika**
2. **Teorija ocjena; vrste i grupe ocjena: tačkaste (saglasne, centrirane, efikasne) ocjene i intervalne ocjene**

P7

Metoda uzoraka

- **Metoda uzoraka** je dio matematičke statistike usmjeren na rješavanje dvije vrste problema:
  1. Procjena parametara (karakteristika) osnovnog skupa na osnovu rezultata dobijenih statističkom obradom podataka u uzorku
  2. problemi testiranja hipoteza: na osnovu procijenjenih parametara (karakteristika) osnovnog skupa iz uzorka , donijeti odluku ili sud da li prihvatiti ili odbaciti određenu pretpostavku (hipotezu) koja se odnosi na neku karakteristiku osnovnog skupa.
- Primjena: u kontroli kvaliteta, poslovnom odlučivanju, marketingu i sl...
- **Ponovimo:**
  - **Populacija (osnovni, odnoso statistički skup):** skup (konačan ili beskonačan: prebrojiv i neprebrojiv) svih pojedinačnih elemenata (elementarnih=statističkih jedinica) iste vrste čija se obilježja (jedno ili više, od mnogih) posmatraju
  - **Uzorak**: dio populacije koji se posmatra (reprezentativan -izabran na unaprijed određen slučajan način), kako bi se donio što je moguće tačniji sud o cijeloj populaciji, na osnovu suda o uzorku u kome se posmatraju statistička obilježja.
  - **Statistika**: slučajna promjenljiva koja je funkcija uzorka (na primjer: aritmetička sredina, medijana, moda, standardno odstupanje i sl.)

- Statistički skup mora biti određen tako da se precizno odredi:
  - koje se obilježje posmatra (težina, visina, MB)
  - prostor kojem pripadaju elementi statističkog skupa
  - vremenski trenutak ili interval kojim će se obuhvatiti svi elementi koje ulaze i skup
- Homogenost statističkog skupa (istovrsnost): sve jedinice skupa su u suštini slične, razlikuju se samo po osnovu obilježja koja se posmatraju
- Elementarne (statističke) jedinice, odnosno (element populacije) mogu biti nosioci jedne ili više pojava, odnosno obilježja (kvantitativne ili kvalitativne karakteristike) koje se zajedno posmatraju. Koja će se obilježja posmatrati zavisi od toga koliko su ta obilježja važna za proučavanje neke pojave.
- Statistička obilježja (varijable, priomjenljive) su opšte karakteristike elemenata statističkog skupa, po kojima su ti elementi međusobno slični i po kojima se međusobno razlikuju.

Adekvatan uzorak mora da ispuni principe:
- nepristrasnosti- podjednaka vjerovatnoća da svaki od elemanata uđe u uzorak; postiže se načinom i metodom odabiranja uzorka koji su razrađeni u statističkoj praksi, a koji se baziraju na teoriji vjerovatnoće i postavkama slučajnog kombinovanja elemenata
- reprezentativnosti - uzorak treba da obuhvati one statističke jedinice koje će u sebi nositi sve karakteristike osnovnog skupa, odnosno one jedinice čija obilježja, kada se izbroje ili izmjere i iz njihovih vrijednosti izračunaju odgovarajući parametri (aritmetička sredina, standardna devijacija i dr.), budu ista ili približno ista kao i pravi parametri osnovnog skupa; Reprezentativnost uzorka se postiže adekvatnom veličinom uzorka i objektivnim načinom izbora jedinica uzorka
- ekonomičnosti- moraju biti prihvatljivi troškovi uzimanja i ispitivanja uzoraka

Uzorkom se dolazi do procjene karakteristika osnovnog skupa, a statističkom metodom određuje se pouzdanost i preciznost te procjene

Uzorak

- u statističkom istraživanju se bilježe:
  - vrijednosti x_i slučajne promjenljive X (koja predstavlja posmatrano statističko obilježje) na uzorku veličine *n*;
  - odgovarajuće frekvencije pojavljivanja tih vrijednosti *fi*.
- na koliko se načina može izabrati uzorak veličine *n* iz populacije veličine *N*?:
  - ako je odabiranje elemenata iz populacije bez vraćanja, onda je to broj kombinacija bez ponavljanja $\binom{N}{n}$
  - ako je odabiranje elemenata iz populacije sa vraćanjem, onda je to broj varijacija sa ponavljanjem N^n

- Kad se uzorak bira na slučajan način može se smatrati da se *xi* element uzorka mogao izabrati sa određenom vjerovatnoćom na *n* načina, pa se *xi , i=1,n* mogu posmatrati kao pojedinačne vrijednosti niza od *n* nezavisnih slučajnih promjenljivih X1;X2;.....Xn koje imaju istu distribuciju kao i slučajna promjenljiva X
- **slučajni uzorak obima *n* je n-dimenzionalna slučajna promenljiva (X1, X2, ...Xn)**

Raspodjela vjerovatnoće aritmetičkih sredina uzoraka

- Neka za obilježje X i uzorak (X1, X2, ... Xn) , za svako *i=1,n* važi:
  - μ_x- srednja vrijednost obilježja u osnovnom skupu, $M(Xi)=\mu_x$
  - σ_x^2- disperzija obilježja u osnovnom skupu $D(Xi)= \sigma_x^2$,

  <u>Onda je</u>:
  - matematičko očekivanje aritmetičke sredine: $M(\bar{X}) = \mu_{\bar{x}} = \mu_x$
  - disperzija aritmetičke sredine:

$$D(\bar{X}) = \sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n}$$

  - standardno odstupanje aritmetičke sredine (*<u>standardna greška):</u>*
    - ako se uzorak bira iz beskonačnog skupa $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$
    - ako se uzorak bira iz konačnog skupa bez vraćanja $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$
- Specijalno:
  1. ako slučajna promjenljiva X (obilježje osnovnog skupa) ima normalnu raspodjelu $N(\mu_x\ ;\ \sigma_x)$, onda i X1, X2... Xn imaju istu raspodjelu, a njihova aritmetička sredina $\bar{X}$ ima normalnu raspodjelu $N(\mu_{\bar{x}}\ ;\ \frac{\sigma_x}{\sqrt{n}})$
  2. **ako je n≥30, onda se bez obzira na raspodjelu slučajne promjenljive X može prihvatiti da raspodjela aritmetičkih sredina uzoraka ima raspodjelu $N(\boldsymbol{\mu_x}\ ;\ \frac{\boldsymbol{\sigma_x}}{\sqrt{\boldsymbol{n}}})$**

Primjer 1:
Pretpostavimo da su težine 3000 studenata normalno raspoređene N(68kg; 3 kg) i da smo odabrali 80 uzoraka po 25 studenata:
a) koje vrijednosti za aritmetičku sredinu i njeno standardno odstupanje možemo očekivati ako smo formirali uzorke sa vraćanjem i bez vraćanja
b) u koliko uzoraka od uočenih 80 možemo očekivati da će se aritmetička sredina $\bar{X}$ naći u granicama od 66,8 do 68,3 kg.

- **Rješenje**

a) μ_x=68 kg- srednja vrijednost obilježja u osnovnom skupu,
   - σ_x=3 kg- standardno odstupanje obilježja u osnovnom skupu
   - **izvlačenje sa vraćanjem=beskonačan skup studenata**

$$\mu_{\bar{x}} = \mu_x = 68 \quad , \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{3}{\sqrt{25}} = 0{,}6$$

   - **izvlačenje bez vraćanja =ograničen skup studenata**

$$\mu_{\bar{x}} = \mu_x = 68 \quad , \qquad \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{3}{\sqrt{25}}\sqrt{\frac{3000-25}{3000-1}} \approx 0{,}6$$

   - **Dakle raspodjela aritmetičke sredine uzorka $\bar{X}$ ima ima raspodjelu: N(68;0,6)**

b) Treba naći vjerovatnoću da se $\bar{X}$ nađe u intervalu 66,8 do 68,3. odnosno treba naći:

$$P(66{,}8 < \bar{X} < 68{,}3) = P\left(\frac{66{,}8 - 68}{0{,}6} < T < \frac{68{,}3 - 68}{0{,}6}\right) = 0{,}6687$$

   - očekivani broj uzoraka sa aritmetičkom sredinom unutar intervala je 80x0,6687=53

Ocjena parametara osnovnog skupa (parametarska analiza)

- **PARAMETARSKA ANALIZA**=na osnovu utvrđenih veličina i sračunatih statističkih parametara uzorka, treba donijeti sud (ocjenu) o ukupnoj, osnovnoj populaciji odakle su uzeti uzorci, odnosno dati ocjenu parametara raspodjele vjerovatnoća osnovne populacije:
  - tačkaste ocjene – broj (tačka na brojnoj osi) koji se izračunava iz uzorka i služi kao aproksimacija nepoznate vrijednosti parametara raspodjele vjerovatnoće osnovne populacije iz koje je uzet uzorak
    - **saglasne** (stabilne, konzistentne), -ocjena je saglasna (stabilna, konzistentna) ako konvergira u vjerovatnoći ka parametru koga ocjenjuje kada broj elemenata uzorka raste
    - **centrirane** (nepristrasne), ako je njeno matematičko očekivanje jednako parametru populacije koga ocjenjuje
    - **najefikasnije** -saglasna i centrirana ocjena koja ima najmanju varijansu ili asimptotski najefikasnije- saglasna i centrirana koja asimptotski teži ocjeni koja ima najmanju varijansu
  - intervalne – kada su ocjene parametara izražene intervalima sa unaprijed zadatom pouzdanošću.

Izvor: Prof. dr Dušan Joksimović POSLOVNA STATISTIKA, Megatrend univerzitet primenjenih nauka, Beograd, Makedonska 21, 2006.

“5. STATISTIČKO OCENJIVANJE
Suština analize uzoraka neke populacije ogleda se upravo u tome da se iz što manje uzoraka (obično iz jednog) dobiju zadovoljavajuće informacije o ponašanju obeležja u čitavoj populaciji. To ponašanje obeležja u čitavoj populaciji je zadovoljavajuće opisano analizom uzorka, ukoliko su analizom tog uzorka dovoljno dobro procenjeni parametri populacije (aritmetička sredina, proporcija, varijansa) i ukoliko je dovoljno dobro procenjen zakon raspodele tog obeležja u populaciji. Analiza uzorka pomoću koje se procenjuju parametri populacije naziva se parametarska analiza.
Analiza uzorka pomoću koje se procenjuje zakon raspodele obeležja u populaciji naziva se neparametarska analiza
Parametarska analiza se vrši na dva načina, u zavisnosti od raspoloživih informacija o populaciji. Prvi način, koji se zove statističko ocenjivanje, primenjuje se u slučajevima kada parametri populacije nisu poznati, pa ih iz uzetog uzorka procenjujemo, odnosno ocenjujemo.
Drugi način, koji se zove testiranje parametarskih statističkih hipoteza, primenjuje se u slučajevima kada su parametri populacije bili poznati, pa iz uzetog uzorka proveravamo da li su se promenili i u kom smeru je promena išla, odnosno da li su se povaćali ili smanjili.
U postupku statističkog ocenjivanja zaključak o nepoznatoj vrednosti parametra populacije koji ocenjujemo možemo dati u vidu jedne brojne vrednosti i tada se ta ocena zove tačkasta ocena, ili u vidu intervala mogućih vrednosti i tada se ta ocena zove intervalna ocena.

Teorija ocjena (procjena)

- Neka je *X* slučajna promjenljiva (statističko obilježje populacije) s funkcijom distribucije (raspodjele vjerovatnoće) F(x). Slučajni uzorak veličine *n* za slučajnu promjenljivu *X* je slučajni vektor (X1,X2,...Xn), gdje su sve slučajne promjenljive *Xi, i=1,...,n*, nezavisne sa zajedničkom funkcijom distribucije vjerovatnoće F(x).
- Vrijednost slučajnog uzorka je uredena n-torka *(x1, x2,.....xn)* ako je izmjerena vrijednost slučajnih promjenljivih *Xi* jednaka *xi ϵ R(X), i=1,...,n.*
- **Teorija ocjena-** dio induktivne statistike koji se bavi ocjenom *U* parametra *Q* (što može biti *p, λ,σ,μ...)* raspodjele vjerovatnoća osnovne populacije pomoću statistike (funkcije slučajnih promjenljivih X) iz uzorka.
- Dvije vrste ocjena:
  - tačkasta ocjena – broj koji se izračunava iz uzorka i služi kao aproksimacija nepoznate vrijednosti parametra raspodjele vjerovatnoća osnovne populacije
  - intervalna ocjena (interval povjerenja, pouzdanosti)- parametri se ocjenjuju u intervalu sa unaprijed zadatom pouzdanošću (vjerovatnoćom da će se naći u tom intervalu)

Tačkaste ocjene

- **Ocjenom parametra Q sa naziva slučajna promjenljiva U= Un(X1,X2,....Xn**), koja sa vjerovatnoćom bliskom jedinici zadovoljava U≈Q, ako je brojnost uzorka velika, odnosno ako je ispunjeno:

  $$\lim_{n\to\infty} P(|U-Q|<C)=1,$$ gdje je C=const i C>0

  Ocjena je bolja ukoliko je prethodno ispunjeno za što manje vrijednosti C
- **Grupe ocjena:**
  - **Saglasna (stabilna)** je ocjena koja konvergira u vjerovatnoći ka parametru osnovne populacije kada *n* raste, odnosno ako je ispunjeno:

    $$\lim_{n\to\infty} P(|U-Q|<\varepsilon)=1,$$ gdje je ε proizvoljno mali pozitivan broj
  - **Centrirana ili nepristrasna** je ocjena ako je njeno matematičko očekivanje jednako parametru osnovne populacije koji se procjenjuje, odnosno ako je ispunjeno: *M(U)=Q,* a <u>asimptotski centrirana</u> je ona kod koje je ispunjeno:

    $$\lim_{n\to\infty} M(U)=Q$$

    Može se dokazati da je svaka saglasna ocjena asimptotski nepristrasna.
  - **najefikasnija** je ocjena koja je saglasna i centrirana i koja ima najmanju disperziju. Mjera njene efikasnosti je

    $$e(U)=\frac{D(U_o)}{D(U)}\leq 1,$$ gdje je su:

    - *D(Uo)*- disperzija ocjene,
    - *D(U)*- donja granica disperzije svih mogućih centriranih ocjena

    – <u>asimptotski najefikasnija</u> je ona kod koje je ispunjeno: $\lim_{n\to\infty} e(U)=1$

Tačkaste ocjene matematičkog očekivanja i disperzije kod normalne raspodjele

Problem:

- na osnovu izmjerenih i sračunatih vrijednosti $\bar{\mathbf{x}}$ i $\boldsymbol{\sigma}$ za uzorak veličine $\boldsymbol{n}$ odrediti $\boldsymbol{\mu_x}$ (matematičko očekivanje osnovnog skupa), varijansu, odnosno disperziju ***Dx*** i $\boldsymbol{\sigma_o}$ (standardnu devijaciju osnovnog skupa)
  - procjena matematičkog očekivanja $\boldsymbol{\mu_x}$ osnovnog skupa (što je *n* veće, to je procjena tačnija):
    - **uzoračka (empirijska) aritmetička sredina** (aritmetička sredina uzorka): saglasna, centrirana i najefikasnija ocjena parametra *μ:*

$$\mu_x \approx \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

  - procjena disperzije ***Dx*** osnovnog skupa:
    - **uzoračka (empirijska) disperzija** (disperzija uzorka): saglasna, necentrirana=pristrasna ocjena parametra Do*:*

$$D_x \approx s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

    - **korigovana uzoračka (empirijska) disperzija** (disperzija uzorka): saglasna, centrirana=nepristrasna, asimptotski najefikasnija ocjena parametra Do*:*

$$D_x \approx s_1^2 = \frac{n}{n-1}s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Saglasna (stabilna) je ocjena koja konvergira u vjerovatnoći ka parametru osnovne populacije kada n raste, odnosno ako je ispunjeno:
$\lim_{n\to\infty} P(|U - Q| < \varepsilon) = 1,$ gdje je ε proizvoljno mali pozitivan broj
Centrirana ili nepristrasna je ocjena ako je njeno matematičko očekivanje jednako parametru osnovne populacije koji se procjenjuje, odnosno ako je ispunjeno: M(U)=Q, a asimptotski centrirana je ona kod koje je ispunjeno:
$\lim_{n\to\infty} M(U) = Q$
Može se dokazati da je svaka saglasna ocjena asimptotski nepristrasna.
najefikasnija je ocjena koja je saglasna i centrirana i koja ima najmanju disperziju. Mjera njene efikasnosti je
$e(U) = \frac{D(U_o)}{D(U)} \leq 1$, gdje je su:

D(Uo)- disperzija ocjene,
D(U)- donja granica disperzije svih mogućih centriranih ocjena

asimptotski najefikasnija je ona kod koje je ispunjeno: $\lim_{n\to\infty} e(U)$ =1

Primjer 2:
Pretpostavimo da je od populacije od 1546 studenata uzet uzorak od 100 studenata radi utvrđivanja težine studenata. Na osnovu podataka iz uzorka utvrđena je aritmetička sredina uzorka $\overline{x}$= 67,45 kg i disperzija uzorka s^2=8,5275 kg. Odrediti centrirane i efikasne ocjene srednje vrijednosti i disperzije populacije:

Rješenje

- ocjena matematičkog očekivanja (srednje vrijednosti) μ_x osnovnog skupa:
  - **uzoračka (empirijska) aritmetička sredina** (aritmetička sredina uzorka): saglasna, centrirana i najefikasnija ocjena parametra μ_x:
$$\mu_x \approx \bar{x} = 67{,}45$$
- procjena disperzije ***Dx*** osnovnog skupa:
  - **korigovana uzoračka (empirijska) disperzija** (disperzija uzorka): saglasna, centrirana=nepristrasna, asimptotski najefikasnija ocjena parametra Dx:
$$D_x \approx s_1^2 = \frac{n}{n-1} s^2 = \frac{100}{100-1} 8{,}5275 = 8{,}6136$$

Intervalne ocjene

- **Ocjenom parametra Q sa naziva slučajna promjenljiva U= Un(X1,X2,....Xn**), koja sa vjerovatnoćom bliskom jedinici zadovoljava U≈Q, ako je brojnost uzorka velika, odnosno ako je ispunjeno:

$$\lim_{n\to\infty} P(|U-Q|<C)=1, \quad \text{gdje je C=const i C>0}$$

Ocjena je bolja ukoliko je prethodno ispunjeno za što manje vrijednosti C

- **Intervalna ocjena:** određivanje intervala u kojem će se sa unaprijed zadatom vjerovatnoćom bliskom jedinici naći vrijednost parametra Q čija se ocjena traži.
- Kažemo da je interval (u1, u2) **interval pouzdanosti (interval povjerenja)** za parametar Q, ako se tačna vrijednost parametra Q nalazi u tom intervalu sa unaprijed datom vjerovatnoćom 1-α
- vjerovatnoća 1-α je **koeficijent pouzdanosti (nivo povjerenja**) intervala pouzdanosti i zavisi od toga koliki se rizik greške može dopustiti u svakom konkretnom slučaju.
- α – **rizik ocjene**
- Ako treba ocijeniti nepoznati parametar Q zadatak se može i ovako postaviti:

Naći vrijednosti u1 i u2 tako da je P{u1<u2}=1 i P{u1≤Q≤u2}=1-α

- parametar Q koji se procjenjuje u unaprijed zadatom intervalu povjerenja može, između ostalog, biti:
  - A) srednja vrijednost osnovne normalne populacije (čija je disperzija poznata)
  - B) srednja vrijednost osnovne normalne populacije (čija disperzija nije poznata)
  - C) disperzija osnovne normalne populacije (čije je matematičko očekivanje poznato)
  - D) disperzija osnovne normalne populacije (čije matematičko očekivanje nije poznato)
  - parametar p u binomnoj raspodjeli.....

A) Interval pouzdanosti za srednju vrijednost osnovne populacije sa normalnom raspodjelom (sa poznatom disperzijom)

- Prema centralnoj graničnoj teoremi: ako slučajna promjenljiva X (obilježje osnovnog skupa) ima normalnu raspodjelu $N(\mu_x\,;\,\sigma_x)$, onda i X1, X2... Xn imaju istu raspodjelu, a njihova aritmetička sredina $\overline{X}$ (kao slučajna promjenljiva) ima normalnu raspodjelu $N(\mu_x\,;\,\frac{\sigma_x}{\sqrt{n}})$
- Dakle, ako pređemo na standardizovanu slučajnu promjenljivu, problem se svodi na određivanje:

$$P\left(-k<\frac{\bar{X}-\mu_x}{\frac{\sigma_x}{\sqrt{n}}}<k\right)=P(-k<T<k)=1-\alpha, \text{ odakle se odredi } k,$$

a kako je $\bar{x}=\bar{X}$

$$P\left(\bar{x}-k\frac{\sigma_x}{\sqrt{n}}<\mu_x<\bar{x}+k\frac{\sigma_x}{\sqrt{n}}\right)=1-\alpha$$

s vjerovatnoćom ***1-α*** se može očekivati da će se srednja vrijednost $\boldsymbol{\mu_x}$ osnovne populacije naći u intervalu pouzdanosti (kada je na prethodni način određeno k)

$\left(\bar{x}-k\frac{\sigma_x}{\sqrt{n}};\bar{x}+k\frac{\sigma_x}{\sqrt{n}}\right)$ za n≥30 (za uzorak iz beskonačne populacije ili za uzorak uzet iz populacije sa vraćanjem)

$\left(\bar{x}-k\frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}};\bar{x}+\frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right)$ za uzorak koji se izvlači iz konačnog skupa bez vraćanja

ako nije poznato σ_x, a radi se o velikom broju uzoraka (n≥30) može se računati sa uzoračkom ili korigovanom uzoračkom disperzijom s ili s_1

$$P\left(-k<\frac{\bar{X}-\mu_x}{\frac{\sigma_x}{\sqrt{n}}}<k\right)=1-\alpha,$$
$$P\left(-k\frac{\sigma_x}{\sqrt{n}}<\bar{X}-\mu_x<k\frac{\sigma_x}{\sqrt{n}}\right)=1-\alpha,$$
$$P\left(k\frac{\sigma_x}{\sqrt{n}}>-(\bar{X}-\mu_x)>-k\frac{\sigma_x}{\sqrt{n}}\right)=1-\alpha,$$
$$P\left(\bar{x}+k\frac{\sigma_x}{\sqrt{n}}>\bar{x}-(\bar{X}-\mu_x)>\bar{x}-k\frac{\sigma_x}{\sqrt{n}}\right)=1-\alpha, \qquad \bar{x}=\bar{X},$$
$$P\left(\bar{x}+k\frac{\sigma_x}{\sqrt{n}}>\mu_x>\bar{x}-k\frac{\sigma_x}{\sqrt{n}}\right)=1-\alpha,$$
$$P\left(\bar{x}-k\frac{\sigma_x}{\sqrt{n}}<\mu_x<\bar{x}+k\frac{\sigma_x}{\sqrt{n}}\right)=1-\alpha,$$

Primjer 3:

Pretpostavimo da je od populacije od 1546 studenata uzet uzorak od 100 studenata radi utvrđivanja težine studenata. Na osnovu podataka iz uzorka utvrđena je aritmetička sredina uzorka = 67,45 kg i disperzija uzorka s^2=8,5275 kg (s=2,93). Naći 95% interval pouzdanosti za srednju vrijednost osnovnog skupa.

Rješenje

$$P\left(-k < \frac{\overline{X} - \mu_x}{\frac{\sigma_x}{\sqrt{n}}} < k\right) = 1 - \alpha$$

$$P(-k < T < k) = 0{,}95$$

$$P(T < k) - P(T < -k) = 0{,}95$$

$$P(T < k) - [1 - P(T < k)] = 0{,}95$$

$$2P(T < k) = 1 + 0{,}95$$

$$P(T < k) = \frac{1{,}95}{2}$$

iz tablica za normalnu raspodjelu je k=1,96, pa je interval

$$P\left(67{,}45 - k\frac{2{,}93}{\sqrt{100}} < \mu_x < 67{,}45 + k\frac{2{,}93}{\sqrt{100}}\right) = 0{,}95$$

$$(67{,}45 - 1{,}96\frac{2{,}93}{\sqrt{100}}; 67{,}45 + 1{,}96\frac{2{,}93}{\sqrt{100}})$$

$$(66{,}888; 68{,}02)$$

B) Interval pouzdanosti za srednju vrijednost osnovne populacije sa normalnom raspodjelom (sa nepoznatom disperzijom)

- ako nije poznato σ_x , a radi se o velikom broju uzoraka (n≥30) može se računati sa uzoračkom ili korigovanom uzoračkom disperzijom s ili s_1 (na način prikazan u A)
- **ako nije poznato σ_x , a radi se o malom broju uzoraka (n<30) koriste se teoreme koje uspostavljaju vezu između Studentove raspodjele T(sa n-1 stepeni slobode) i granica intervala pouzdanosti, tako da se problem svodi na određivanje intervala:**

$$P(|T| < t_\alpha) = 1 - \alpha \quad \text{odnosno} \quad P(-t_\alpha < T < t_\alpha) = 1 - \alpha$$

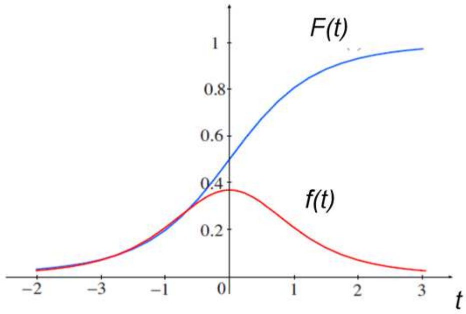
$$P\left(\bar{x} - t_\alpha \frac{s}{\sqrt{n}} < \mu_x < \bar{x} + t_\alpha \frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

Za određivanje t_α koriste se tablice za Studentovu (T) raspodjelu u zavisnosti od stepeni slobode.

Karakteristike gustine ***f(t)*** Studentove raspodjele:

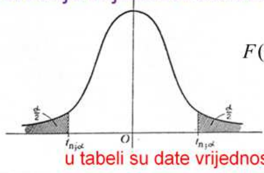
1. kriva gustine pozitivna i simetrična u odnosu na ordinatnu osu
2. apscisna osa je asimptota krive gustine za t→±∞
3. za n→∞, Studentova raspodjela teži normalnoj raspodjeli
4. moda, medijana i matematičko očekivanje =0
5. disperzija D(t)=n/(n-2) za n>2

$$f(t) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{1+n}{2})}{\Gamma(\frac{n}{2})} (1 + \frac{t^2}{n})^{-\frac{n+1}{2}}$$



Primjer 4:

Na osnovu uzorka od n=15 elemenata, s aritmetičkom sredinom $\overline{x}$= 20,4 i standardnim odstupanjem s=0,8, naći 98% interval pouzdanosti (sa rizikom α=0,02) za srednju vrijednost osnovnog skupa μ.



$$F(t)=\int_{-t\alpha}^{t\alpha}\frac{\Gamma(\frac{1+n}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})(1+\frac{x^2}{n})^{\frac{n+1}{2}}}dx$$

Rješenje

broj stepeni slobode k=n-1=15-1=14

rizik α=0,02

iz tablica je $t_\alpha = t_{0{,}02} = 2{,}624$, pa je interval

$$P\left(\bar{x} - t_\alpha \frac{s}{\sqrt{n}} < \mu_x < \bar{x} + t_\alpha \frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(20{,}4 - 2{,}624\frac{0{,}8}{\sqrt{15}} < \mu_x < 20{,}4 + 2{,}624\frac{0{,}8}{\sqrt{15}}\right) = 1 - 0{,}02$$

$$P(19{,}86 < \mu_x < 20{,}94) = 0{,}98$$

u tabeli su date vrijednosti *tα* za koje je P(|t|>tα)=α

	α						
n	0.80	0.60	0.40	0.20	0.10	0.05	0.02	0.01
1	0.325	0.727	1.376	3.078	6.314	12.706	31.823	63.657
2	0.289	0.617	1.061	1.886	2.920	4.403	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.531	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	[illegible]	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.133	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576

C) Interval pouzdanosti disperzije za osnovnu populaciju sa normalnom raspodjelom (sa nepoznatim očekivanjem)

- Prema centralnoj graničnoj teoremi: ako slučajna promjenljiva X (obilježje osnovnog skupa) ima normalnu raspodjelu $N(\mu_x\,;\,\sigma_x)$, onda i X1, X2... Xn imaju istu raspodjelu, a njihova aritmetička sredina $\overline{X}$ ima normalnu raspodjelu $N\left(\mu_x\,;\,\frac{\sigma_x}{\sqrt{n}}\right)$
- Disperzija uzorka s^2 se može izraziti kao realizacija slučajne promjenljive S^2, ako nije poznato matematičko očekivanje osnovne populacije

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

- transformacijom prethodno postaje:

$$S^2 = \frac{\sigma^2}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2 = \frac{\sigma^2}{n-1}\cdot\chi^2_{n-1}$$

- veličina

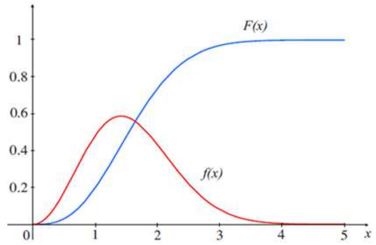
$$\chi^2_{n-1} = \frac{n-1}{\sigma^2}S^2$$

- je slučajna promjenljiva sa χ^2 (hi -kvadrat) raspodjelom sa n-1 stepeni slobode.

Karakteristike funkcije gustine Hi-kvadrat raspodjele:

1. apscisna osa je asimptota krive gustine za $x\to\infty$
2. za $n\to\infty$, Hi-kvadrat raspodjela teži normalnoj raspodjeli (dovoljno n>30)
3. matematičko očekivanje =n (broj stepeni slobode)
4. Moda: Mo=n-2 za n≥2
5. disperzija D=2n

$$f(x) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, \quad (x>0)$$



C) Interval pouzdanosti disperzije za osnovnu populaciju sa normalnom raspodjelom (sa nepoznatim očekivanjem)-nastavak

- Početni zadatak je bio naći interval koji zadovoljava:

$P(u_1 < \sigma^2 < u_2) = 1 - \alpha$, koji na osnovu prethodnih transformacija postaje:

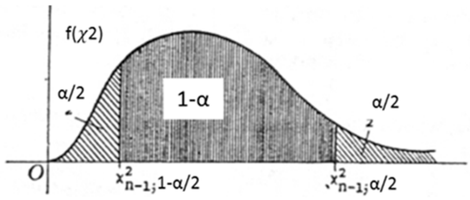
$$P\left(\chi^2_{n-1,1-\frac{\alpha}{2}} < \chi^2_{n-1} < \chi^2_{n-1,\frac{\alpha}{2}}\right) = 1 - \alpha$$

odnosno, kad se uvede smjena $\chi^2_{n-1} = \frac{n-1}{\sigma^2}S^2$, i s^2=S^2

$P\left(\frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right) = 1 - \alpha$, gdje je ***1-α*** koeficijent pouzdanosti, a ***α*** rizik ocjene

s vjerovatnoćom ***1-α*** se može očekivati da će se disperzija D, odnosno S^2 osnovne populacije naći u intervalu pouzdanosti:

$$\left(\frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}}, \frac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}\right)$$



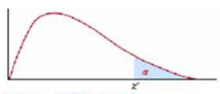
D) Interval pouzdanosti disperzije za osnovnu populaciju sa normalnom raspodjelom (sa poznatim očekivanjem)

- Disperzija uzorka s^2 se može izraziti kao realizacija slučajne promjenljive S^2, ako je poznato matematičko očekivanje osnovne populacije μ
$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$
- transformacijom prethodno postaje:
$S^2 = \frac{\sigma^2}{n}\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{\sigma^2}{n}\cdot\chi_n^2$, pri čemu je veličina
$\chi_n^2 = \frac{n}{\sigma^2}S^2$ slučajna promjenljiva sa χ^2 (hi -kvadrat) raspodjelom sa ***n*** stepeni slobode.
- Početni zadatak je bio naći interval koji zadovoljava:
$P(u_1 < \sigma^2 < u_2) = 1 - \alpha$, koji na osnovu transformacija (kao u slučaju C) postaje:
$$P\left(\chi^2_{n,1-\frac{\alpha}{2}} < \chi_n^2 < \chi^2_{n,\frac{\alpha}{2}}\right) = 1 - \alpha$$
odnosno, kad se uvede smjena $\chi_n^2 = \frac{n}{\sigma^2}S^2$, i s^2=S^2
$P\left(\frac{ns^2}{\chi^2_{n,\frac{\alpha}{2}}} < \sigma^2 < \frac{ns^2}{\chi^2_{n,1-\frac{\alpha}{2}}}\right) = 1 - \alpha$, gdje je ***1-α*** koeficijent pouzdanosti, a ***α*** rizik ocjene
s vjerovatnoćom ***1-α*** se može očekivati da će se disperzija D, odnosno S^2 osnovne populacije naći u intervalu pouzdanosti
$$\left(\frac{ns^2}{\chi^2_{n,\frac{\alpha}{2}}}, \frac{ns^2}{\chi^2_{n,1-\frac{\alpha}{2}}}\right)$$

Primjer 5:

Iz slučajno izabranog uzorka obima n=25 elemenata, izračunata je disperzija s^2=12. od pretpostavkom da je uzorak izvučen iz populacije sa normalnom raspodjelom, intervalno ocijeniti disperziju osnovne populacije s pouzdanošću od 0,9.

χ^2 - distribucija



Rješenje

broj stepeni slobode k=n-1=25-1=24

rizik α=0,1, slijedi: α/2=0,05, 1-α/2=0,95

iz tablica je

$$\chi^2_{n-1,1-\frac{\alpha}{2}} = \chi^2_{24;0,95} = 13{,}848$$

$$\chi^2_{n-1,\frac{\alpha}{2}} = \chi^2_{24;0,05} = 36{,}415$$

pa je interval za disperziju, s pouzdanošću 0,9:

$$\frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}$$

$$\frac{24\cdot 12}{36{,}415} < \sigma^2 < \frac{24\cdot 12}{13{,}848}$$

$$7{,}91 < \sigma^2 < 20{,}79$$

u tabeli su date vrijednosti tα za koje je ispunjen uslov P(T>tα)=α

k	0,99	0,98	0,95	0,90	0,80	0,50	0,20	0,10	0,05	0,02	0,01	0,001
1	0,000	0,001	0,004	0,016	0,064	0,455	1,642	2,706	3,841	5,412	6,635	10,828
2	0,020	0,040	0,103	0,211	0,446	1,386	3,219	4,605	5,991	7,824	9,210	13,816
3	0,115	0,185	0,352	0,584	1,005	2,366	4,642	6,251	7,815	9,837	11,345	16,266
4	0,297	0,429	0,711	1,064	1,649	3,357	5,989	7,779	9,488	11,668	13,277	18,467
5	0,554	0,752	1,145	1,610	2,343	4,351	7,289	9,236	11,070	13,388	15,086	20,515
6	0,872	1,134	1,635	2,204	3,070	5,348	8,558	10,645	12,592	15,033	16,812	22,458
7	1,239	1,564	2,167	2,833	3,822	6,346	9,803	12,017	14,067	16,622	18,475	24,322
8	1,646	2,032	2,733	3,490	4,594	7,344	11,030	13,362	15,507	18,168	20,090	26,124
9	2,088	2,532	3,325	4,168	5,380	8,343	12,242	14,684	16,919	19,679	21,666	27,877
10	2,558	3,059	3,940	4,865	6,179	9,342	13,442	15,987	18,307	21,161	23,209	29,588
11	3,053	3,609	4,575	5,578	6,989	10,341	14,631	17,275	19,675	22,618	24,725	31,264
12	3,571	4,178	5,226	6,304	7,807	11,340	15,812	18,549	21,026	24,054	26,217	32,909
13	4,107	4,765	5,892	7,042	8,634	12,340	16,985	19,812	22,362	25,472	27,688	34,528
14	4,660	5,368	6,571	7,790	9,467	13,339	18,151	21,064	23,685	26,873	29,141	36,123
15	5,229	5,985	7,261	8,547	10,307	14,339	19,311	22,307	24,996	28,259	30,578	37,697
16	5,812	6,614	7,962	9,312	11,152	15,338	20,465	23,542	26,296	29,633	32,000	39,252
17	6,408	7,255	8,672	10,085	12,002	16,338	21,615	24,769	27,587	30,995	33,409	40,790
18	7,015	7,906	9,390	10,865	12,857	17,338	22,760	25,989	28,869	32,346	34,805	42,312
19	7,633	8,567	10,117	11,651	13,716	18,338	23,900	27,204	30,144	33,687	36,191	43,820
20	8,260	9,237	10,851	12,443	14,578	19,337	25,038	28,412	31,410	35,020	37,566	45,315
21	8,897	9,915	11,591	13,240	15,445	20,337	26,171	29,615	32,671	36,343	38,932	46,797
22	9,542	10,600	12,338	14,041	16,314	21,337	27,301	30,813	33,924	37,659	40,289	48,268
23	10,196	11,293	13,091	14,848	17,187	22,337	28,429	32,007	35,172	38,968	41,638	49,728
24	10,856	11,992	13,848	15,659	18,062	23,337	29,553	33,196	36,415	40,270	42,980	51,179
25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
26	12,198	13,409	15,379	17,292	19,820	25,336	31,795	35,563	38,885	42,856	45,642	54,052
27	12,879	14,125	16,151	18,114	20,703	26,336	32,912	36,741	40,113	44,140	46,963	55,476
28	13,565	14,847	16,928	18,939	21,588	27,336	34,027	37,916	41,337	45,419	48,278	56,892
29	14,256	15,574	17,708	19,768	22,475	28,336	35,139	39,087	42,557	46,693	49,588	58,301
30	14,953	16,306	18,493	20,599	23,364	29,336	36,250	40,256	43,773	47,962	50,892	59,703
35	18,509	20,027	22,465	24,797	27,836	34,336	41,778	46,059	49,802	54,244	57,342	66,619
40	22,164	23,838	26,509	29,051	32,345	39,335	47,269	51,805	55,758	60,436	63,691	73,402
45	25,901	27,720	30,612	33,350	36,884	44,335	52,729	57,505	61,656	66,555	69,957	80,077
50	29,707	31,664	34,764	37,689	41,449	49,335	58,164	63,167	67,505	72,613	76,154	86,661
60	37,485	39,699	43,188	46,459	50,641	59,335	68,972	74,397	79,082	84,580	88,379	99,607
70	45,442	47,893	51,739	55,329	59,898	69,334	79,715	85,527	90,531	96,388	100,425	112,317
80	53,540	56,213	60,391	64,278	69,207	79,334	90,405	96,578	101,879	108,069	112,329	124,839
90	61,754	64,635	69,126	73,291	78,558	89,334	101,054	107,565	113,145	119,648	124,116	137,208
100	70,065	73,142	77,929	82,358	87,945	99,334	111,667	118,498	124,342	131,142	135,807	149,449
500	429,388	437,219	449,147	459,926	473,210	499,334	526,401	540,930	553,127	567,070	576,493	603,446

Literatura

- Vukadinović, S.: Elementi teorije verovatnoće i matematičke statistike, Privredni pergled, Beo
- Vukadinović, S.: Zbirka rešenih zadataka iz teorije verovatnoće, Privredni pergled, Beograd, 1983
- Čuljak, V: Vjerojatnost i statistika, Građevinski fakultet, Sveučilište u Zagrebu, 2011, https://portal.uniri.hr/system/resources/docs/000/004/082/original/Skripta_Vera_%C4%8Culjak.pdf?1413283708
- Prof. dr Dušan Joksimović: POSLOVNA STATISTIKA, Megatrend univerzitet primenjenih nauka, Beograd, 2006.
- Pivac, S.: Statističke metode (predavanja, diplomski studij, kolegij "Statističke metode") e-nastavni materijal, Split, 2010.
- http://www.ef.uns.ac.rs/Download/metodologija_nir/20_uzorkovanje.pdf
- http://www.medfak.ni.ac.rs/PREDAVANJA/2.%20STOMATOLOGIJA/STATISTIKA/6.%20predavanje.pdf