STATISTIKA

Predavanja

INFORMACIJE

Ocena
- 60% kolokvija ali izpit
- 15% trije kvizi
- 25% DN
- Dovoljen kalkulator in 1 A4 list (brez rešenih nalog)

Kriterij
6: 50-59 7: 60-69 8: 70-79 9: 80-89 10: 90-100

Snov
- Opisna statistika
- Slučajne spremenljivke
- Vzorčenje

Literatura
- Zapiski s predavanj (dovolj)
- Dytham, C.: Choosing and Usig Statistics ....

2.10.24

UVOD

Statistika je veda, ki preučuje množične podatke/pojave in zajema:
- Zbiranje podatkov - pomembna teorija vzorčenja
- Urejanje in povzemanje podatkov - opisna statistika
- Vrednotenje podatkov in statistično sklepanje - inferenčna statistika in teorija verjetnosti

Osnovni pojmi
- Stat. populacija - zbirka enot, ki jih preučujemo
  ↳ Vzorec je podmn. populacije
- Stat. spremenljivka - predpis, ki vsaki enoti pop. priredi določeno vrednost
  ↳ Spr. označimo z velikimi ($X, Y, \ldots$), vrednosti pa z malimi ($x_1, x_2, \ldots$)
  ↳ Delimo na opisne (kvalitativne) in številske (kvantitativne)
- Merska lestvica - kakšno strukturo imajo vrednosti spr. ter kakšne operacije so smiselne
  ↳ Imenske (nominalne), urejenostne (ordinalne), intervalske, razmernostne
  ↳ eg. stopnja izobrazbe
  ↳ eg. let. rojstva samo razlika
  ↳ tudi sešt, množ, ...

Strukture: populacija razdeljena v k skupin, f_i označuje št. enot v i-ti skupini (frekvenca)
- Delež v iti skupini:

$$f_i^{\circ} = \frac{f_i}{\sum_{i=1}^{k} f_i}$$

- Delež izražen v odstotkih:

$$f_i^{\circ}\% = 100 \frac{f_i}{\sum_{i=1}^{k} f_i}$$

Primer 1

Podatki o št. novih dokt. znanosti v letu 2012

a) $f_M^\circ = \frac{558}{1027}$ $f_Ž^\circ = \frac{469}{1027}$ } struktura po spolu

b) Struktura po področju

$f_{NV}^\circ = \frac{231}{1027} = 22{,}49\%$ ...

c) Struktura po spolu znotraj področja

$f_M^{NV} = \frac{119}{231}$ $f_Ž^{NV} = \frac{112}{231}$

Koeficient - razmerje dveh podatkov

↳ Posebne vrste so t.i. stopnje

Primer 2

Pridelek na hektar $\frac{156771\,t}{26761\,ha} = \dots\ t/ha$

Indeksi (pt. 1)

- S stalno osnovo: Naj bo X številska slučajna spr. in x_i vrednost X v iti skupini $(i = 1 \dots k)$. Zap $x_1, x_2, \dots, x_k$ imenujemo stat. vrsta. Indeks z osnovo x_0 imenujemo stalna osnova.

$$I_{i/0} = 100 \cdot \frac{x_i}{x_0}$$

Primer 3

Št. nasadov v vinogradih v regijah. Indeksi z osnovo „Primorje"

$I_{1/0} = 100 \cdot \frac{17769}{14017} = 126{,}77\%$ $I_{2/0} = 109{,}1\%$

$I_{3/0} = 100\%$ $I_{4/0} = \frac{246}{14017} \cdot 100 = 1{,}76\%$

Indeksi (pt. 2)

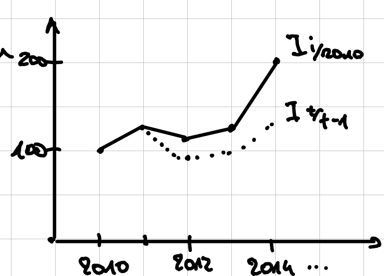
- Verižni indeks: spr. X spremljamo v času, $x_1, \dots, x_T$ je časovna vrsta, št. T je dolžina čas. vrste
- Verižni indeks/indeks s premično osnovo: $I_{t/t-1} = 100 \cdot \frac{x_t}{x_{t-1}}$, $t = 2, \dots, T$
  ↳ Smiselno samo, če je čas. vrsta ekvidistantna
- Stopnja rasti: $S_{t/t-1} = I_{t/t-1} - 100\% = 100 \cdot \frac{x_t - x_{t-1}}{x_{t-1}}$
  ↳ Izraža prirast/upad v primerjavi s prejšnim čas. obdobjem

Primer 4

Pridelava čebule in paradižnika v tonah v RS 2010-2020. Izračunaj indekse s stalno osnovo 2010 in verižne za čebulo.

Leto	$I_{i/2010}$	$I_{t/t-1}$	$S_{t/t-1}$
2010	100%	/	/
2011	$\frac{8333}{4667} \cdot 100 = 135{,}7\%$	135,7%	35,7%
2012	$\frac{5869}{4667} \cdot 100 = 125{,}8\%$	$\frac{5869}{8333} \cdot 100 = 92{,}7\%$	-7,3%

Liniski grafikon



9.10.24

Frekvenčna porazdelitev (pt. 1)

- Opisna spremenljivka
- Št. pojavljanj vrednosti a_j ($j=1 \dots n$) je frekvenca a_j

$$f_j = \#(X = a_j)$$

- Relativna frekvenca j-te enote:

$$f_j^\circ = \frac{f_i}{f_1 + \dots + f_n}$$

- Izraženo v odstotkih:

$$f_j^\circ\% = 100 \cdot \frac{f_i}{f_1 + \dots + f_n}$$

Primer 5 (pt. 1)

	f_i	$f_i\%$
Čokolada	19	31,7
Vanilija	14	23,3
Jagoda	9	15
Lešnik	7	11,7
Pistacija	11	18,3

Frekvenčna porazdelitev (pt. 2)

- Poleg frekvenc f_j lahko izračunamo še komulativne frekvence:

$$F_j = \#(X \leq a_j) = f_1 + \dots + f_n$$

- In relativno komulativno frekvenco:

$$F_j\% = 100 \cdot \frac{F_j}{f_1 + \dots + f_n}$$

Primer 6

	f_i	$f_i\%$	F_i	$F_i\%$
5	22	29,3	22	29,3
6	17	22,7	39	52,0
7	14	18,7	53	70,7
8	9	12	62	82,7
9	6	8	68	90,7
10	7	9,3	75	100

Frekvenčna porazdelitev (pt. 3)

- Št. vrednosti razdelimo v k intervalov
- Opremimo jih še s:
  - Spodnjo in zgornjo mejo $x_{i\,min}$ in $x_{i\,max}$ za $i = 1, \dots, k$:

$$x_{i\,min} = x_{i-1\,max}$$

  - Sredino razredov:

$$x_i = \frac{x_{i\,max} + x_{i\,min}}{2}$$

  - Širino razredov:

$$d_i = x_{i\,max} - x_{i\,min}$$

- Če razredi različno široki, za vsak razred izračunamo gostoto frekvenc:

$$g_i = \frac{f_i}{d_i}$$

Primer 7

	$x_{i\,min}$	$x_{i\,max}$	d_i	x_i	f_i	$f_i\%$	F_i	$F_i\%$
20 - 39	19,5	39,5	20	29,5	2	6,7	2	6,7
40 - 59	39,5	59,5	20	49,5	10	33,3	12	40
60 - 79	59,5	79,5	20	69,5	15	50	27	90
80 - 99	79,5	99,5	20	89,5	3	10	30	100

Podatke lahko prikažemo s histogramom ali poligonom. Poligon dobimo tako, da za vsak razred narišemo točko (x_i, f_i) dodamo še točki $(x_0, 0)$ in $(x_{k+1}, 0)$ ter jih povežemo v linijski grafikon

Prikažemo lahko tudi z ogivo, pri kateri na abcisno os narišemo zg. meje razredov, na ordinatno os pa pripadajoče relativne kumulativne frekvence. Za vsak razred narišemo točko $(x_{i\,max}, F_i\%)$, dodamo točko $(x_{1\,min}, 0)$ in povežemo točke.

Ranžirna vrsta in rangi

- Če vrednosti $x_1, \dots, x_n$ uredimo po velikosti smo jih razvrstili v ranžirno vrsto
$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$
- Elementi $x_{(i)}$ so vrstilni elementi
- Rang dane vrednosti je njen položaj v ranžirni vrsti: rang vrednosti x (oznaka $R(x)$) je enak i, če je $x = x_{(i)}$. Ali je rang enolično določen? Ne. → Primer PPT
- Vezani rang $R(x)$ je aritmetična sredina zg. in sp. ranga:
$$R(x) = \frac{\text{zg. surovi rang} + \text{sp. surovi rang}}{2}$$
- Relativni ali kvantilni rang elementa x je:
$$r(x) = \frac{R(x) - \frac{1}{2}}{n}$$

- Vrstilne statistike razberemo iz komulativnih frekvenc, in sicer
$$x_{(i)} = a_j, \quad \text{če je } 1 + F_{j-1} \leq i \leq F_j$$

- Range določimo iz komulativnih frekvenc
  - vrednost a_j ima surove range od F_{j-1} do F_j
  - Vezani rang
$$R(a_j) = \frac{F_{j-1} + F_j + 1}{2}$$
  - Relativni rang
$$r(a_j) = \frac{F_{j-1} + F_j}{2n}$$

Kvantili

- Kvantil stat. spr. za določen delež je vrednost, pod katero leži približno dani delež podatkov. Naj bolj pomembni:
  - Mediana (Me) je kvantil $\frac{1}{2}$
  - Tercil deleža $\frac{1}{3}$ in $\frac{2}{3}$
  - Kvartil deleži $\frac{1}{4}, \frac{2}{4}, \frac{3}{4}$
  - Decil je delež $\frac{j}{10}$ $(j = 1 \dots 9)$
  - Centil je delež $\frac{j}{100}$ $(j = 1, \dots, 99)$

- Def. kvantila: Vrednost Q_γ je kvantil spremenljivke X za delež γ, če je

$$\frac{\#(x < Q_\gamma)}{n} \leq \gamma \quad \text{in} \quad \frac{\#(x \leq Q_\gamma)}{n} \geq \gamma$$

- Medkvartilni razmik je $Q = Q_{\frac{3}{4}} - Q_{\frac{1}{4}}$. Interval $[Q_{\frac{1}{4}}, Q_{\frac{3}{4}}]$ vsebuje približno polovico sredinskih vrednosti
- Kadar kvantil ni enolično določen lahko predstavnika izberemo z interpolacijo: če kvantil leži med 2 vrednostima in za rang kvantila velja $R(a) < R(Q) < R(b) = R(a)+1$, potem Q določimo z:

$$Q = a + (R(Q) - R(a))(b-a)$$

- Urejenostne spremenljivke lahko grafično predstavimo s škatlo z brki, kjer so prikazani minimalna in maksimalna vrednost ter kvartili

- Kadar so podatki dani le s frekvenčno tabelo, potem osnovnih podatkov nimamo in kvantile izračunamo iz kumulativnih frekvenc s pomočjo interpolacije.
- Če kvantil Q_γ leži v razredu s sp. mejo $x_{0,min}$ in zg. mejo $x_{0,max}$, frekvenco f_0 ter širino d_0, potem Q_γ izračunamo z

$$Q_\gamma = x_{0,min} + \frac{F(Q_\gamma) - F(x_{0,min})}{f_0} \cdot d_0$$

kjer

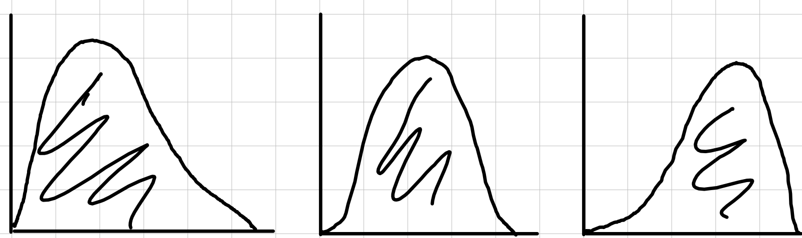
$$F(Q_\gamma) = n \cdot \gamma + 0{,}5$$

23.10.24

Mere centralne tendence

- Najbolj uporabljene mere centralne tendence oz. središčnosti:
  - (tehtana) aritmetična sredina
  - mediana
  - modus
- Tehtana arit. sredina: $\bar{x} = \frac{f_1x_1 + \ldots + f_kx_k}{f_1 + \ldots + f_k} = \frac{f_1x_1 + \ldots + f_kx_k}{f_n}$

Porazdelitve slučajnostnih spremenljivk



Asim. v desno $\bar{x} > Me$ | Prib. simet. $\bar{x} \approx Me$ | Asim. v levo $\bar{x} < Me$

Mere variabilnosti

- Najbolj uporabljene:
  - Variacijski razmik: $VR = x_{max} - x_{min}$
  - Interkvartilni razpon: $IQR = Q_{\frac{3}{4}} - Q_{\frac{1}{4}}$
  - Varianca:
    - ↳ Populacijska (n = št. enot v pop.)

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

    - ↳ Vzorčna (n = velikost vzorca)

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

  - Standardni odklon

Občutljivost mer na skrajne vrednosti

- Za skrajne vrednosti (t.i. osamelce) običajno obravnavamo vrednosti zunaj intervala $[Q_{\frac{1}{4}} - 1{,}5 IQR, Q_{\frac{3}{4}} + 1{,}5 IQR]$

Slučajne spremenljivke

- Neformalna def.: slučajna spr. je vrednost, ki jo dobimo kot rezultat poskusa (dogodka) z več možnimi izidi, npr.:
  - Rezultat posameznega meta kovanca je slučajna spr. X, ki lahko zavzame eno od dveh vrednosti: grb (G) ali cifra (C). Vsak od dogodkov se zgodi z verjetnostjo $\frac{1}{2}$, t.j.:

$$P(X=G) = \frac{1}{2}, P(X=C) = \frac{1}{2}$$

  - Kovanec vržemo 3x. Rezultat je slučajna spr. Y, ki lahko zavzame eno od osmih vrednosti v mn. $\{GGG, GGC, GCG, CGG, GCC, CGC, CCG, CCC\}$, vsak od teh rezultatov se zgodi z verjetnostjo $\frac{1}{8}$.
- Poznamo diskretne in zvezne slučajne spr.

Diskretne sluč. spr.

- Diskretne sluč. spr. zavzamejo diskretne vrednosti.
- Naj bo X disk. sluč. spr., ki zavzame vrednosti v mn. $\{x_1, \dots, x_n\}$. Fja verjetnosti p za sluč. spr. X je:

$$p(x) = P(X=x)$$

  Velja $\sum_{i=1}^{n} p(x_i) = 1$
- Porazdelitvena fja F je def. z:

$$F(x) = P(X \leq x)$$

  $F(x)$ je nepadajoča, $\lim_{x\to\infty} = 1$, $\lim_{x\to 0} = 0$

- Bernulijeva sluč. spr. lahko zavzame le 2 vrednosti
Fja je podana z:

$$P(X=x_1)=p$$
$$P(X=x_2)=1-p$$

- Binomska sluč. spr. beleži št. „uspehov" v n neodvisnih poskusih, kjer je vsak od poskusov Bernulijeva sluč. spr.
Binomsko porazdelitev $B(n,p)$ določata dva parametra: št. poskusov n ter verjetnost p „uspeha" Bernulijevega poskusa. Če ima X binomsko porazdelitev s parametroma n in p, zapišemo kot:

$$X \sim B(n,p)$$

Fja vrjetnosti:

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\ k=0,\dots,n,\ \text{kjer}$$
$$\binom{n}{k}=\frac{n!}{k!(n-k)!}$$

30.10.24

- Pričakovana vrednost slučajne spr. X, ki zavzame vrednosti x_i $(i=1,\dots,n)$ in ima fjo vrjetnosti $p(x)$ je

$$E(X)=\sum_{i=1}^{\infty}p(x_i)x_i$$

Za binom. sluč. spr. $X\sim B(n,p)$ je $E(X)=np$

- Varianca slučajne spr. X s pričakovano vrednostjo $\mu=E(X)$ je

$$Var(X)=E((X-\mu)^2)=\sum_{i=1}^{\infty}p(x_i)(x_i-\mu)^2$$

- Poissonova sluč. spr. X: št. dogodkov v fiksnem čas. intervalu, če se ti dogodki zgodijo neodvisno drug od drugega in z dano stopnjo λ. Poissonova porazdelitev dobimo iz binomske če $n\to\infty$, $p\to 0$ in $\lambda=np=\text{const}$. Poissonova sluč. spr. $X\sim Poisson(\lambda)$ ima fjo vrjetnosti

$$P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!},\ k=0,1,2,\dots$$

Za $Poisson(\lambda)$ je $E(X)=\lambda$ in $Var(X)=\lambda$

- Geometrična sluč. spr. X: za geometrično sluč. spr. imamo v mislih eno od naslednjih sluč. spr.
  - Št. Bernullijevih poskusov do prvega „uspešnega" dogodka (X)
  - Št. „neuspehov" zaporednih Bernullijevih poskusov do prvega „uspešnega" dogodka ($Y = X - 1$)

Omejimo se na prvo def. Če je p vrjetnost uspeha Bernullijevega poskusa je fja vrjetnosti za geom. sluč. spr. $X \sim Geom(p)$ dana z

$$P(X=k) = (1-p)^{k-1} p, \quad k = 1, 2, \dots$$

Velja $E(X) = \frac{1}{p}$, $Var(X) = \frac{1-p}{p^2}$

Zvezne slučajne spremenljivke 6.11.24

- Zvezne slučajne spr. lahko zavzamejo kontinuum vrednosti
- Vlogo fje verjetnosti ima funkcija gostote vrjetnosti f, ki zadošča:
  - $f \geq 0$
  - je odsekovno zvezna
  - $\int_{-\infty}^{\infty} f(x)\,dx = 1$
- Če ima sluč. spr. X gostoto f je

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

verjetnost, da X zavzame vrednosti $[a,b]$

- Porazdelitvena fja slučajne spr. X z gostoto f je

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(s)\,ds$$

- $F(x)$ je verjetnost, da sluč. spr. X zavzame vrednosti do največ x.
- Za porazdelitveno fjo velja:
  - $F(x)$ je naraščujoča fja
  - $\lim_{x\to\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$
  - $F'(x) = f(x)$
- Verjetnost, da X zavzame vrednosti na $[a,b]$ lahko določimo iz porazdelitvene fje

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx = F(b) - F(b)$$

Kvantili za zvezne sluč. spr.

- Naj bo X zvezna sluč. spr. s porazdelitveno fjo F. p-kvantil ($0 \leq p \leq 1$) za spr. X je vrednost x_p, za katero velja

$$P(X \leq x_p) = p$$

oz.

$$F(x_p) = p$$

Zvezne slučajne spremenljivke

- Pričakovana vrednost sluč. spr. X z gostoto f je:
$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$
- Varianca sluč. spr. X s pričakovano vrednostjo $\mu = E(X)$ je:

$$Var(X) = E((X-\mu)^2) = E(X^2) - E^2(X)$$

- Eksponentna porazdelitev sluč. spr.
  - Gostota sluč. spr. X z eksp. poraz. s param. λ ($X \sim Exp(\lambda)$) je:
  $$f(x) = \begin{cases} \lambda e^{-\lambda x} & ; \text{za } x \geq 0 \\ 0 & ; \text{za } x < 0 \end{cases}$$
  - Porazdelitvena fja za $X \sim Exp(\lambda)$ je:
  $$F(x) = \begin{cases} 1 - e^{-\lambda x} & ; \text{za } x \geq 0 \\ 0 & ; \text{za } x < 0 \end{cases}$$
  - Velja:

$$E(X) = \frac{1}{\lambda}$$

$$Var(X) = \frac{1}{\lambda^2}$$

- Normalna (oz. Gaussova) porazdelitev $N(\mu, \sigma^2)$
  - Normalna sluč spr $X \sim N(\mu, \sigma^2)$ ima fjo gostote

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

  - Je simetrična porazdelitev s pričakovano vrednostjo μ ($\mu = E(X)$) in varianco σ^2 ($\sigma = Var(X)$). Param. σ je standardni odklon X.
  - Če $X \sim N(\mu, \sigma^2)$, potem
    - ↳ med $\mu-\sigma$ in $\mu+\sigma$ leži približno 68,3% enot
    - ↳ med $\mu-2\sigma$ in $\mu+2\sigma$ leži približno 95,5% enot
    - ↳ med $\mu-3\sigma$ in $\mu+3\sigma$ leži približno 99,7% enot

- Standardna normalna slučajna spremenljivka $Z \sim N(0,1)$ ima fjo gostote:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

- Standardizacija:

  če $X \sim N(\mu, \sigma^2)$ je $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$
- Verjetnost, da $X \sim N(\mu, \sigma^2)$ zavzame vrednosti na $[a,b]$ je

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

- Te vrednosti določamo numerično oz. s pomočjo tabele za $N(0,1)$ s predhodno standardizacijo (e-učilnica)

Kovarianca in korelacija 13. 11. 24

- Varianca $Var(X)$ je merilo sluč. spr. X okoli pričakovane vrednosti $\mu = E(X)$
- Kovarianca in korelacija merita povezanost dveh sluč. spr.

- Naj bosta X, Y sluč. spr. Kovarianca je:
$$Cov(X,Y) = E((X-\mu_X)(Y-\mu_Y)) = E(XY) - E(X)E(Y)$$
- Korelacija med X in Y je
$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

Vzorčenje
- Način pridobivanja info. o (načeloma veliki) populaciji s preučevanjem dela populacije (vzorca)
- Vzorčenje lahko poteka:
  - Po nekem pravilu
  - Naključno
- Več prednosti

Točkasto ocenjevanje parametrov
- Denimo, da je celotna pop. velikosti N, vrednosti sluč. spr. X v pop. pa $x_1, \dots, x_N$
- Želimo oceniti:
  - Populacijsko povprečje
  - Vsoto
  - Pop. varianco
  - Pop. standardni odklon
- Iz pop. vzamemo naključen vzorec velikosti n. Povp., vsota, pop. varianca in pop. stand. odklon vzorca niso enaki kot v populaciji.
- Porazdelitev $\bar{X}$ imenujemo vzorčna porazdelitev:
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
- Izkaže se, da je:
$$E(\bar{X}) = \mu$$
- Rečemo, da je vzorčenje nepristransko, vzorčno povp. pa nepristranska ocena pop. povp.

- Izkaže se, da je varianca od $\bar{X}$:

$$Var(\bar{X}) = \frac{\sigma^2}{n}\left(1-\frac{n-1}{N-1}\right)$$

- Faktor $\frac{n-1}{N-1}$ je popravek končne populacije. Če je $n \ll N$, lahko popravek zanemarimo

$$Var(\bar{X}) \approx \frac{\sigma^2}{n}$$

- Oceno $Var(\bar{X})$ dobimo s pomočjo pop. variance σ^2, ki v splošnem ni poznana
- Prvi predlog bi bil

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

- Vendar to ni nepristranska ocena ($E(\hat{\sigma}^2) \neq \sigma^2$), zato velja:

$$E(\hat{\sigma}^2) = \sigma^2 \frac{N}{N-1}\cdot\frac{n-1}{n}$$

- Za velike pop. N je nepristranska ocena variance:

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

- Nepristranska ocena za varianco vzorčne poraz. $Var(\bar{X})$:

$$s_{\bar{X}}^2 = \frac{s^2}{n}\left(1-\frac{n}{N}\right)$$

Pop. parameter	Nepristranska ocena
μ	$\frac{1}{n}\sum_{i=1}^{n}\bar{X}_i$
τ	$\frac{N}{n}\sum_{i=1}^{n}X_i$
σ^2	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$

Aproksimacija vzorčne porazdelitve z aproksimacije

- Ob velikem št. vzorcev histogrami vzorčne porazdelitve $\bar{X}$ zgledajo normalno, četudi X ni porazdeljena normalno
- Če je μ pričakovana vrednost spr. X in σ^2 njena varianca, potem $\bar{X}$ aprox. z $N(\mu, \frac{\sigma^2}{n})$
- Centralni limitni izrek:

$$P\left(\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z\right) \rightarrow F(z)$$

Ko se velikost vzorca povečuje, je porazdelitvena fja spr.:

$$\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

vse bližje porazdelitveni fji stand. norm. porazdelitve. Ker porazdelitveno fjo $N(0,1)$ poznamo, lahko sedaj ocenimo, kako tipičen je nek vzorec.

20.11.24

Z-test: preizkušanje domnev o μ pri znanem σ

- Poznamo stand. odklon σ, na podlagi naključnega vzorca želimo preizkusiti domnevo, da pop. povprečje za sluč. spr. μ^*
- Ničelna domneva $H_0: \mu = \mu^*$
- Če $\bar{x}$ preveč odstopa od μ^* H_0 zavrnemo. $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$
- Preveč je odvisen od:
  - Alternativne hipoteze - testiramo proti 1 od treh alternativ
  - Stopnje značilnosti α - določa prag zavrnitve

Interval zaupanja

- Je intervalna ocena parametra. Int. zaupanja za μ pri dani stopnji zaupanja β je tak int. (μ_{min}, μ_{max}), da velja:

$$P(\mu_{min} < \mu < \mu_{max}) = \beta$$

- Za dan β je int. zaupanja za μ:

$$(\mu_{min}, \mu_{max}) = \left(\bar{x} - \frac{\sigma}{\sqrt{n}} z_{(1+\beta)/2}, \bar{x} - \frac{\sigma}{\sqrt{n}} z_{(1+\beta)/2}\right)$$

t-test: preizkušanje domnev o μ kadar σ ni poznan

- Če σ ni poznan, ga nadomestimo z njegovo oceno
- Studentove porazdelitve: družina razdelitev, ki se razlikujejo glede na št. prostih stopenj df
  $df = n-1$
- Testna statistika:
  $$T = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$$
- Naj bo $t_p(df)$ p-kvantil studentove porazdelitve
  Pri stopnji znač. α H_0 zavrnemo v korist 3 alternativ
- Interval zaupanja za dani μ pri dani stopnji zaupanja $0 < \beta < 1$ je sedaj:
  $$(\mu_{min}, \mu_{max}) = \left(\bar{x} - \frac{s}{\sqrt{n}} t_{(1+\beta)/2}(df), \bar{x} + \frac{s}{\sqrt{n}} t_{(1+\beta)/2}(df)\right)$$