

VAJA 8 –OCENA PARAMETROV PRI VELIKIH VZORCIH ter TESTIRANJE HIPOTEZ PRI MALIH VZORCIH

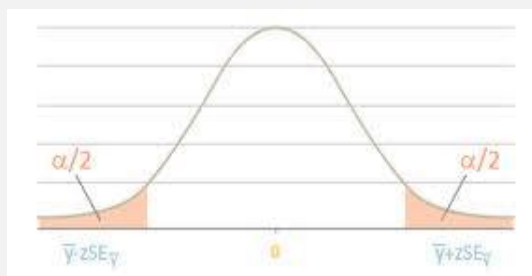
OZNAKE

- Statistične karakteristike populacije označujemo z grškimi črkami.
- Statistične karakteristike vzorca označujemo z latinskimi črkami.

DVOSTRANSKA IN ENOSTRANSKA OCENA PARAMETROV

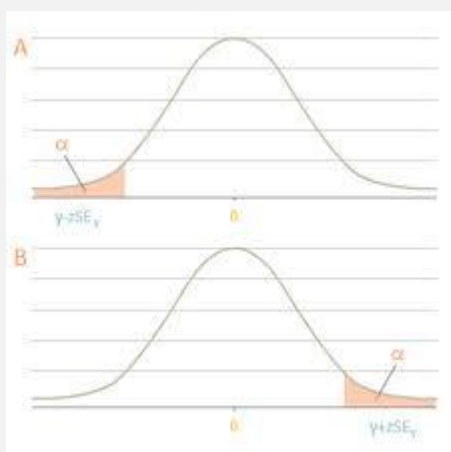
- Veliki, slučajni vzorci brez ponavljanja $n \geq 30$.
- Vzorčna porazdelitev je »normalna« CENTRALNO LIMITNI IZREK.
- Parameter g , katerega želimo oceniti, je $N(E(g), SE(g))$.
- Stopnja tveganja α , stopnja zaupanja $1 - \alpha$.

- Razmična dvostranska ocena $P(g - zSE(g) < G < g + zSE(g)) = 1 - 2 \cdot \frac{\alpha}{2}$



α	z
0,05	1,96
0,01	2,58
0,001	3,29

- Razmična enostranska ocena $P(g - zSE(g) < G) = 1 - \alpha$ ali $P(G < g + zSE(g)) = 1 - \alpha$



α	z
0,05	1,64
0,01	2,33
0,001	3,09

Populacija	Vzorec
ARITMETIČNA SREDINA - μ, σ	$E(\bar{X}) = \mu, SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$ Cenilka standardnega odklona $s_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n - 1}}$
DELEŽ - π	$p = \frac{n_a}{n}; SE(p) = \sqrt{\frac{p(1-p)}{n}}$

Primer 1

Pri analizi proizvodnosti dela, so pri 110 slučajno izbranih delavcev ugotovili naslednjo proizvodnost dela, izraženo s številom proizvodov na delavca. Tabela 1.

- Določite naključni vzorec velikosti 40 (veliki vzorec). Uporabljajte orodje za analizo podatkov *Sampling*.
- Določite: aritmetično sredino vzorca, standardni odklon vzorca (cenilko odklona), varianco vzorca in $SE(\bar{X})$.
- Določite intervalno oceno aritmetične sredine s stopnjo tveganja 1% (5%) in vzorca velikosti 40. Kaj se zgodi, če povečamo velikost vzorca na 65?
- Določite delež populacije, kjer je proizvodnost vsaj 45 proizvodov na delavca. Določite intervalno oceno deleža s stopnjo tveganja 1% (5%) in vzorca velikosti 40.

Tabela 1

30	33	44	37	35	29	37	38	41	45	38
44	36	32	34	36	35	36	43	47	36	42
47	39	30	42	31	20	28	31	35	30	37
33	25	33	46	47	36	36	37	37	38	32
39	38	49	44	37	24	43	26	30	48	48
35	38	37	43	36	44	32	33	44	44	27
38	35	48	41	37	27	53	47	39	59	36
26	37	50	43	44	32	31	39	41	54	34
44	53	50	36	43	39	36	45	43	35	39
35	40	51	28	42	25	27	57	44	34	41

Primer 2

V priloženi datoteki s podatki o letnem prometu neke Luke opravite naslednje analize.

- Določite naključni vzorec velikosti 150 (veliki vzorec).
Uporabljajte orodje za analizo podatkov Sampling.
- Določite: aritmetično sredino vzorca, standardni odklon vzorca, varianco vzorca in $SE(\bar{X})$. Vnesite spodnjo tabelo v Excelu in jo dopolnite.
- Populaciji priredite posebno lastnost in ocenite delež populacije, ki ima posebno lastnost (primerno jo izberite na podlagi vsebine in lastnosti podatkov). Vnesite spodnjo tabelo v Excelu in jo dopolnite.

	μ	$\bar{X}$	$SE(\bar{X})$	Int. oceana AS $\alpha = 1\%$ $n = 150$	π	p	$SE(p)$	Int. oc. deleža $\alpha = 5\%$ $n = 150$
Klicni znak								
Ime ladje								
Tip ladje								
Nosilnost								
BRT								
NRT								
Prihod								
Priplula iz								
Zastava								
Tovor								
Količina U								
Količina I								
Promet								
balast O								
balast T								
balast								
Kapaciteta								
Balast na ladji								
BWM								
Izkrčan ballast								
cona								
balast O								
balast T								

- Dobljene ocene primerjajte z realnimi, ki jih lahko izračunate v populaciji.
- Ugotovite kako se ocene spremenijo če povečamo vzorec na 200, 250 enot.
- Določite vzorec velikosti 10 in ročno izračunajte statistične karakteristike vzorca.

DOLOČANJE VELIKOSTI VZORCA PRI ENOSTAVNEM SLUČAJNEM VZORCU

Določimo napako:

$$\text{Absolutno } |g - G| = z \cdot SE(g)$$

$$\text{Relativno } |g - G| \cdot \frac{100}{G} = z \cdot SE(g) \cdot \frac{100}{G} \text{ v \%}$$

katera more biti manjša od določene vrednosti E , pri določeni stopnji tveganja-zaupanja α .

- **Določanje velikosti vzorca pri oceni aritmetične sredine:**

$$n > \left(\frac{z \cdot \sigma}{E} \right)^2.$$

- **Določanje velikosti vzorca pri oceni deleža:**

$$n > \left(\frac{z}{E} \right)^2 \cdot p(1 - p).$$

Primer 3a

Želimo oceniti povprečno starost podjetnikov v majhnih podjetjih, tako da bo razlika med populacijskim povprečjem in ocenjenim povprečjem manjša od 1 leta.

Določite velikost vzorca, če je standardni odklon 10 let in je stopnja tveganja 5%.

Primer 3b

Želimo oceniti delež populacije, ki bi med dvema kandidatoma, A in B, ki se pojavijo na volitve, izbrali kandidata B. Predvidevamo, da se bo zanj odločilo manj kot 10% volivcev. Zahtevamo, da se ocena ne bo razlikovala od dejanskega deleža volivcev za več kot 5%. Na območju, kjer želimo oceniti velikost vzorca je okoli 400.000 volivcev. Stopnja tveganja naj bo 0,01.

Primer 4

V priloženi datoteki s podatki o letnem prometu Luke določite optimalno velikost vzorca pri oceni aritmetične sredine in deleža za napako $E = 0,05$ ($E = 0,1$).

Primer 5

Opazovana spremenljivka se v populaciji normalno porazdeljuje, vrednost σ za standardni odklon ni znana.

Za aritmetično sredino M postavimo domnevo $H_0 : M = M_0$, ki jo preverimo s pomočjo t -poskusa:

$$t = \frac{\bar{x} - M_0}{se(\bar{x})} = \frac{\bar{x} - M_0}{s_x} \cdot \sqrt{n};$$

$$m = n - 1$$

Skozi neko križišče je v 10 enakodolgh časovnih razmikih, ki so bili slučajno izbrani, peljalo naslednje število avtomobilov:

Časovni razmik	1	2	3	4	5	6	7	8	9	10
Število avtomobilov	32	29	33	31	34	30	31	33	31	36

V smislu prometnih predpisov se mora namestiti semafor na vsakem križišču, kjer je povprečno število avtomobilov, ki gredo skozi križišče v obravnavanih časovnih razmikih večje od 30.

Kaj predlagate na podlagi zgornjih podatkov? ($\alpha = 0,05 - 0,01 - 0,001$)

Primer 5a

Ali lahko iz vzorca 20 delavcev sklepamo, da je povprečna starost delavcev 35 let, če starosti delavcev iz vzorca uredimo v naslednjo frekvenčno tabelo porazdelitev:

starost	Št. delavcev
20-24	3
25-29	4
30-34	7
35-39	3
40-44	2
45-49	1

Upoštevajte 5% stopnjo značilnosti.

Primer 6a – neodvisni pari

Opazovana spremenljivka se v obeh populacijah porazdeljuje normalno, varianci sta v obeh populacijah enaki.

Za aritmetični sredini postavimo domnevo $H_0 : M_1 = M_2$, ki jo preverimo s pomočjo t -poskusa:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (M_1 - M_2)}{se(\bar{x}_1 - \bar{x}_2)}$$

$$m = n_1 + n_2 - 2$$

$$se(\bar{x}_1 - \bar{x}_2) = s_{x_1 - x_2} \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}$$

$$s_{x_1 - x_2} = \sqrt{\frac{(n_1 - 1)s_{x_1}^2 + (n_2 - 1)s_{x_2}^2}{n_1 + n_2 - 2}}$$

Proučevali so ali vpliva nočno delo na povprečno proizvodnost. Slučajno so izbrali v dnevni izmeni 10 delavcev in v nočni izmeni 6 delavcev in za njih ugotovili število proizvodov, ki so jih proizvedli v izmeni:

Dnevna izmena

Delavec	Št. proizvodov
1	16
2	18
3	20
4	17
5	15
6	17
7	14
8	19
9	18
10	16

Nočna izmena

Delavec	Št. proizvodov
1	13
2	17
3	15
4	14
5	18
6	13

($\alpha = 0,05 - 0,01 - 0,001$) pri pogoju enakih varianc v populacijah.

Primer 6b – odvisni pari

$$\begin{aligned} & \{x_i\}, \{y_i\} \\ d_i &= x_i - y_i \\ \bar{d} &= \frac{\sum d_i}{n} \\ s_d^2 &= \frac{\sum (d_i - \bar{d})^2}{n-1} \\ t &= \frac{\bar{d}}{se(\bar{d})} = \frac{\bar{d}}{s_d} \cdot \sqrt{n} \\ m &= n - 1 \end{aligned}$$

Proučevali so spremembe v količini prodaje bencina, ki naj bi nastopile zaradi spremembe v ceni bencina. V ta namen so v 8 bencinskih servisih ugotovili dnevno prodajo bencina pred in po podražitvi:

Črpalka	pred	po
1	11,2	10,8
2	8,9	9,6
3	16,7	17
4	7,2	7,2
5	32,6	31,5
6	12,3	12,8
7	15,6	14,6
8	21,3	22,5

Ali lahko sklepamo, da se je po zvišanju cene bencina spremenila povprečna prodaja bencina?