Distribucija sredina uzoraka

Broj uzoraka veličine n jedinica koji može da se dobije iz jednog osnovnog skupa veličine N jedinica dat je sledećim izrazima:

- uzorci sa ponavljanjem $k = \binom{N+n-1}{n}$

- uzorci bez ponavljanja $k = \binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{N(N-1)(N-2)\cdot.......(N-n+1)}{n!}$

$$\overline{\overline{X}} = \frac{\sum_{i=1}^{k} \overline{X}_i}{k} \qquad \sigma_{\overline{X}}^2 = \frac{\sum_{i=1}^{k} (\overline{X}_i - \mu)^2}{k} \qquad \sigma_{\overline{X}} = \sqrt{\frac{\sum_{i=1}^{k} (\overline{X}_i - \mu)^2}{k}}$$

Aritmetička sredina distribucije sredina uzoraka

$$\overline{\overline{X}} = \frac{\sum_{i=1}^{k} \overline{X}_i}{k} \qquad \overline{\overline{X}} = \mu$$

Varijansa distribucije sredina uzoraka

$$\sigma_{\overline{X}}^2 = \frac{\sum_{i=1}^{k} (\overline{X}_i - \mu)^2}{k} \qquad \sigma_{\overline{X}}^2 = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} \qquad \sigma_{\overline{X}}^2 \langle \sigma^2$$

$\frac{N-n}{N-1}$ *korektivni faktor koji se koristi kod uzoraka bez ponavljanja*

Standardna greška aritmetičke sredine

$$\sigma_{\overline{X}} = \sqrt{\frac{\sum_{i=1}^{k} (\overline{X}_i - \mu)^2}{k}} \qquad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \qquad \sigma_{\overline{X}} \langle \sigma$$

Primer 1. Na osnovu osnovnog skupa od pet elemenata formirati sve kombinacije uzoraka od po dve jedinice bez ponavljanja. Uporediti parametre osnovnog skupa i distribucije sredina uzoraka.

Vrednosti obeležja osnovnog skupa x_i	X_i^2
2	4
4	16
6	36
8	64
10	100
30	220

$$\mu = \frac{\Sigma X_i}{N} = \frac{30}{5} = 6$$

$$\sigma^2 = \frac{\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{N}}{N} = \frac{220 - \frac{(30)^2}{5}}{5} = 8$$

$$\sigma = \sqrt{8} = 2{,}82$$

Broj mogućih kombinacija uzoraka: $k = \binom{N}{n} = \binom{5}{2} = \frac{5 \cdot 4}{1 \cdot 2} = 10$

Kombinacije uzoraka		$\overline{X}_i$	$(\overline{X}_i - \mu)$	$(\overline{X}_i - \mu)^2$
2	4	3	-3	9
2	6	4	-2	4
2	8	5	-1	1
2	10	6	0	0
4	6	5	-1	1
4	8	6	0	0
4	10	7	1	1
6	8	7	1	1
6	10	8	2	4
8	10	9	3	9
		60	0	30

$$\overline{\overline{X}} = \frac{\sum_{i=1}^{k} \overline{X}_i}{k} = \frac{60}{10} = 6$$

$$\overline{\overline{X}} = \mu$$

$$\sigma_{\overline{X}}^2 = \frac{\sum_{i=1}^{k} (\overline{X}_i - \mu)^2}{k} = \frac{30}{10} = 3$$

$$\sigma_{\overline{X}}^2 \langle \sigma^2 \Rightarrow 3 \langle 8$$

$$\sigma_{\overline{X}} = \sqrt{\frac{\sum_{i=1}^{k} (\overline{X}_i - \mu)^2}{k}} = \sqrt{3} = 1{,}73$$

$$\sigma_{\overline{X}} \langle \sigma \Rightarrow 1{,}73 \langle 2{,}82$$

Ocene na osnovu uzorka

Ocena nepoznate aritmetičke sredine osnovnog skupa na osnovu uzorka

Prilikom ocene nepoznate aritmetičke sredine osnovnog skupa na osnovu uzorka razlikujemo dva slučaja. Prvi slučaj kada je poznat varijabilitet osnovnog skupa i drugi slučaj kada varijabilitet osnovnog skupa nije poznat.

I slučaj (poznat varijabilitet osnovnog skupa (σ^2))

Interval poverenja za ocenu nepoznate aritmetičke sredine osnovnog skupa na osnovu uzorka glasi:

$$\bar{X} - Z_\alpha \cdot \sigma_{\bar{x}} < \mu < \bar{X} - Z_\alpha \cdot \sigma_{\bar{x}}$$

$\bar{X}$ - predstavlja aritmetičku sredinu iz uzorka

Z_α- predstavlja kritičnu vrednost iz tablice Normalne raspodele. Za 95% interval poverenja ta vrednost iznosi 1,96 ($Z_{0,05} = 1{,}96$), dok je za 99% interval poverenja ta vrednost 2,58 ($Z_{0,01} = 2{,}58$).

$\sigma_{\bar{x}}$ – prava standardna greška aritmetičke sredine

Standardna greška aritmetičke sredine, ako je poznat varijabilitet osnovnog skupa (ako su poznate vrednosti standardne devijacije ili varijanse) može se izračunati na osnovu sledećeg izraza:

$$\boxed{\boldsymbol{\sigma}_{\bar{\mathrm{X}}} = \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}} \cdot \sqrt{\frac{\boldsymbol{N} - \boldsymbol{n}}{\boldsymbol{N} - 1}}}$$

$\boxed{\dfrac{N-n}{N-1}}$ je korektivni faktor koji se koristi ako je poznata veličina osnovnog skupa N i ako se primenjuje uzorak bez ponavljanja (bez vraćanja).

Total osnovnog skupa je moguće izračunati ukoliko se leva i desna strana prethodno utvrđenog intervala poverenja pomnoži sa ukupnim brojem podataka (N). Sa L_1 označavamo donju granicu intervala dok sa L_2 gornju granicu intervala poverenja.

$$N * L_1 < N\mu < N * L_2$$

Primer 1. U rejonu od 10.000 domaćinstava izabran je uzorak čiji su rezultati dati sledećim vrednostima:

Mesečna potrošnja goveđeg mesa (kg) (X_i)	**Broj domaćinstava (f_i)**	$X_i f_i$
1	10	10
1,6	15	24
1,8	40	72
2,1	20	42
2,6	10	26
2,8	5	14
Ukupno	**100**	**188**

Oceniti prosečnu mesečnu i ukupnu potrošnju goveđeg mesa sa pouzdanošću od 95 procenata ako je od ranije poznat varijabilitet osnovnog skupa $\sigma^2 = 3{,}4$ kg^2.

$$\bar{X} - Z_\alpha \cdot \sigma_{\bar{x}} < \mu < \bar{X} - Z_\alpha \cdot \sigma_{\bar{x}}$$

$$\bar{X} = \frac{\sum_{i=1}^{k} X_i f_i}{\sum_{i=1}^{k} f_i} = \frac{188}{100} = 1{,}88$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}} = \frac{1{,}84}{\sqrt{100}} * \sqrt{\frac{10.000-100}{10.000-1}} = 0{,}1834$$

1,88-1,96*0,1834<μ<1,88+1,96*0,1834

1,5205<μ<2,2395

Total osnovnog skupa glasi:

$$N * L_1 < N\mu < N * L_2$$

$$10.000 * 1{,}5205 < N\mu < 10.000 * 2{,}2395$$

15.205,00<Nμ<22.395,00

II slučaj (nije poznat varijabilitet osnovnog skupa)

U slučaju da nije poznat varijabilitet osnovnog skupa, prilikom izbora odgovarajućeg intervala poverenja posmatra se veličina uzorka.

Za uzorke koji broje više od 30 jedinica (n>30) interval je sledeći:

$$\boxed{\overline{X} - Z_{\alpha} \cdot S_{\overline{X}} \langle \mu \langle \overline{X} + Z_{\alpha} \cdot S_{\overline{X}}}$$

$\bar{X}$ - predstavlja aritmetičku sredinu iz uzorka

Z_{α}- predstavlja kritičnu vrednost iz tablice Normalne raspodele. Za 95% interval poverenja ta vrednost iznosi 1,96 ($Z_{0,05} = 1{,}96$), dok je za 99% interval poverenja ta vrednost 2,58 ($Z_{0,01} = 2{,}58$).

$S_{\bar{x}}$ – ocenjena standardna greška aritmetičke sredine

Za uzorke koji broje manje od 30 jedinica (n<30) interval je sledeći:

$$\boxed{\overline{X} - t_{\alpha(n-1)} \cdot S_{\overline{X}} \langle \mu \langle \overline{X} + t_{\alpha(n-1)} \cdot S_{\overline{X}}}$$

$\bar{X}$ - predstavlja aritmetičku sredinu iz uzorka

$t_{\alpha(n-1)}$- predstavlja kritičnu vrednost iz tablice Studentove raspodele (t-raspodela) za odgovarajući prag značajnosti α i (n-1) broj stepeni slobode.

$S_{\bar{x}}$ – ocenjena standardna greška aritmetičke sredine

Formule za ocenjenu standardnu grešku aritmetičke sredine razlikujemo u skladu sa tim da li su podaci grupisani ili nisu.

Negrupisani podaci:

$$\boxed{S_{\overline{X}} = \sqrt{\frac{\Sigma(X_i - \overline{X})^2}{n(n-1)} \cdot \frac{N-n}{N}}} \qquad \boxed{S_{\overline{X}} = \sqrt{\frac{\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}}{n(n-1)} \cdot \frac{N-n}{N}}}$$

Grupisani podaci:

$$\boxed{S_{\overline{X}} = \sqrt{\frac{\Sigma f_i (X_i - \overline{X})^2}{n(n-1)} \cdot \frac{N-n}{N}}} \qquad \boxed{S_{\overline{X}} = \sqrt{\frac{\Sigma f_i X_i^2 - \frac{(\Sigma f_i X_i)^2}{n}}{n(n-1)} \cdot \frac{N-n}{N}}}$$

Na osnovu ocenjenog intervala poverenja može se oceniti i total osnovnog skupa na osnovu sledećeg izraza:

$$N \cdot L_1 \langle N\mu \langle N \cdot L_2$$

Primer 1. Na osnovu podataka uzorka individualnih gazdinstava kod kojih je posmatrana površina pod voćnjacima i vinogradima utvrditi prosečnu vrednost i odrediti razmak poverenja u kome će se kretati površina pod voćnjacima i vinogradima u osnovnom skupu.

Pov. pod voć.i vinog. X_i	Broj individ. gazdinstava f_i	$f_i X_i$	$f_i X_i^2$
0.6	200	120	72
0.8	400	320	256
1	500	500	500
1.2	600	720	864
1.4	200	280	392
1.6	100	160	256
	2000	2100	2340

Uzorak je veći od 30 (n=2.000), tako da je potrebno koristiti sledeći interval:

$$\overline{X} - Z_\alpha \cdot S_{\overline{X}} \langle \mu \langle \overline{X} + Z_\alpha \cdot S_{\overline{X}}$$

$$\overline{X} = \frac{\Sigma f_i X_i}{N} = \frac{2100}{2000} = 1{,}05ha$$

$$S_{\overline{X}} = \sqrt{\frac{\Sigma f_i X_i^2 - \frac{(\Sigma f_i X_i)^2}{n}}{n(n-1)}} = \sqrt{\frac{2340 - \frac{(2100)^2}{2000}}{2000(2000-1)}} = 0{,}006ha$$

$$\alpha = 0{,}05$$
$$Z_{0,05} = 1{,}96$$

$$1{,}05 - 1{,}96 \cdot 0{,}006 \langle \mu \langle 1{,}05 + 1{,}96 \cdot 0{,}006$$
$$1{,}038ha \langle \mu \langle 1{,}062ha$$

$$\alpha = 0{,}01$$
$$Z_{0,01} = 2{,}58$$

$$1{,}05 - 2{,}58 \cdot 0{,}006 \langle \mu \langle 1{,}05 + 2{,}58 \cdot 0{,}006$$
$$1{,}035ha \langle \mu \langle 1{,}066ha$$

Primer 2. Rezultati uzorka koji se odnose na ispitivanje uticaja jedne vrste pesticida na klijavost semena hibrida kukuruza dati su u tabeli. Utvrditi očekivanu prosečnu klijavost posmatranog hibrida.

Klijavost semena X_i	X_i^2
96	**9216**
98	**9604**
90	**8100**
89	**7921**
92	**8464**
94	**8836**
90	**8100**
88	**7744**
90	**8100**
91	**8281**
918	**84366**

Kako je uzorak manji od 30 (n=10), potrebno izračunati sledeći interval poverenja:

$$\overline{X} - t_{\alpha(n-1)} \cdot S_{\overline{X}} \langle \mu \langle \overline{X} + t_{\alpha(n-1)} \cdot S_{\overline{X}}$$

$$\overline{X} = \frac{\Sigma X_i}{n} = \frac{918}{10} = 91{,}8\%$$

$$S_{\overline{X}} = \sqrt{\frac{\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}}{n(n-1)}} = \sqrt{\frac{84366 - \frac{(918)^2}{10}}{10(10-1)}} = 1{,}02\%$$

$\alpha = 0{,}05$

$t_{0,05(9)} = 2{,}262$

$91{,}8 - 2{,}262 \cdot 1{,}02 \langle \mu \langle 91{,}8 + 2{,}262 \cdot 1{,}02$

$89{,}5 \langle \mu \langle 94{,}1\%$

$\alpha = 0{,}01$

$t_{0,01(9)} = 3{,}250$

$91{,}8 - 3{,}250 \cdot 1{,}02 \langle \mu \langle 91{,}8 + 3{,}250 \cdot 1{,}02$

$88{,}5 \langle \mu \langle 95{,}1\%$

Primer 3. Oceniti prosečan i ukupan prinos šećerne repe na osnovu prinosa dobijenih sa 20 hektara, u regionu koji ima 480 hektara pod ovom kulturom.

Prinos (vag/ha) X_i	**Površina f_i**	$f_i X_i$	$f_i X_i^2$
4.4	**3**	**13.2**	**58.08**
4.6	**4**	**18.4**	**84.64**
5.2	**7**	**36.4**	**189.28**
5.4	**4**	**21.6**	**116.64**
5.9	**2**	**11.8**	**69.62**
	20	**101.4**	**518.26**

$$\boldsymbol{N} = 480$$
$$\boldsymbol{n} = 20$$

$$\overline{X} - \boldsymbol{t}_{\alpha(\boldsymbol{n}-1)} \cdot \boldsymbol{S}_{\overline{X}} \langle \boldsymbol{\mu} \langle \overline{X} + \boldsymbol{t}_{\alpha(\boldsymbol{n}-1)} \cdot \boldsymbol{S}_{\overline{X}}$$

$$\overline{X} = \frac{\Sigma \boldsymbol{f}_i X_i}{\boldsymbol{N}} = \frac{101{,}4}{20} = 5{,}07\,\boldsymbol{vag/ha}$$

$$\boldsymbol{S}_{\overline{X}} = \sqrt{\frac{\Sigma \boldsymbol{f}_i X_i^2 - \frac{(\Sigma \boldsymbol{f}_i X_i)^2}{\boldsymbol{n}}}{\boldsymbol{n}(\boldsymbol{n}-1)} \cdot \frac{\boldsymbol{N}-\boldsymbol{n}}{\boldsymbol{N}}} = \sqrt{\frac{518{,}26 - \frac{(101{,}4)^2}{20}}{20 \cdot (20-1)} \cdot \frac{480-20}{480}} = 0{,}102\,\boldsymbol{vag/ha}$$

$$\boldsymbol{\alpha} = 0{,}05$$
$$\boldsymbol{t}_{0,05(19)} = 2{,}093$$
$$5{,}07 - 2{,}093 \cdot 0{,}102 \langle \boldsymbol{\mu} \langle 5{,}07 + 2{,}093 \cdot 0{,}102$$
$$4{,}857 \langle \boldsymbol{\mu} \langle 5{,}283\,\boldsymbol{vag/ha}$$

$$\boldsymbol{\alpha} = 0{,}01$$
$$\boldsymbol{t}_{0,01(19)} = 2{,}861$$
$$5{,}07 - 2{,}861 \cdot 0{,}102 \langle \boldsymbol{\mu} \langle 5{,}07 + 2{,}861 \cdot 0{,}102$$
$$4{,}778 \langle \boldsymbol{\mu} \langle 5{,}362\,\boldsymbol{vag/ha}$$

Kako je poznata veličina osnovnog skupa (N=480), moguće je izračunati i totale osnovnog skupa za oba intervala poverenja:

$$\boldsymbol{\alpha} = 0{,}05$$
$$480 \cdot 4{,}857 \langle \boldsymbol{N\mu} \langle 480 \cdot 5{,}283$$
$$2331{,}4 \langle \boldsymbol{N\mu} \langle 2535{,}8\,\boldsymbol{vagona}$$

$$\boldsymbol{\alpha} = 0{,}01$$
$$480 \cdot 4{,}778 \langle \boldsymbol{N\mu} \langle 480 \cdot 5{,}362$$
$$2293{,}4 \langle \boldsymbol{N\mu} \langle 2573{,}8\,\boldsymbol{vagona}$$

Standardna greška proporcije i interval poverenja za nepoznatu proporciju osnovnog skupa

Proporcija je specifičan vid izražavanja nekog svojstva u osnovnom skupu ili u uzorku (udeo neispravnih proizvoda u ukupnom broju proizvoda, broj žena u ukupnom broju stanovnika, itd.).

Proporcija izračunata na osnovu uzorka je ocena proporcije osnovnog skupa.

$$p = \frac{A}{N}$$

A- broj jedinica osnovnog skupa koje poseduju traženu karakteristiku
N- ukupan broj jedinica u osnovnom skupu

$$q = \frac{B}{N}$$

B - broj jedinica osnovnog skupa koje ne poseduju traženu karakteristiku
N - ukupan broj jedinica osnovnog skupa

Standardna greška proporcije jednaka je: $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n} \cdot \frac{N-n}{N-1}}$

S obzirom na to da je proporcija osnovnog skupa obično nepoznata ocenjujemo je na osnovu uzorka.

$$\hat{p} = \frac{a}{n} \qquad \hat{q} = \frac{b}{n} \qquad \hat{q} = 1 - \hat{p}$$

a- broj jedinica iz uzorka koje poseduju traženu karakteristiku
b- broj jedinica u uzorku koje ne poseduju traženu karakteristiku
n- ukupan broj jedinica u uzorku (veličina uzorka)

Ocenjena standardna greška proporcije jednaka je: $S_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n} \cdot \frac{N-n}{N}}$

.

- ako se ocena izvodi na osnovu velikog uzorka (n>30): $\hat{p} - Z_{\alpha} \cdot S_{\hat{p}} \langle p \langle \hat{p} + Z_{\alpha} \cdot S_{\hat{p}}$

- ako se ocena izvodi na osnovu malog uzorka (n<30): $\hat{p} - t_{\alpha(n-1)} \cdot S_{\hat{p}} \langle p \langle \hat{p} + t_{\alpha(n-1)} \cdot S_{\hat{p}}$

Na osnovu ocenjenog intervala poverenja može se oceniti i total osnovnog skupa na osnovu sledećeg izraza:

$$\boxed{N \cdot L_1 \langle Np \langle N \cdot L_2}$$

Primer 1. Rezultati uzorka, vezani za prinos šljive, dati su u tabeli. Utvrditi vrednost proporcije stabala sa prinosom od 10 kilograma i odrediti granice u kojima će se uz određenu verovatnoću nalaziti proporcija osnovnog skupa.

Prinos šljive (kg/stablu) X_i	Broj stabala f_i
5	100
8	200
10	300
11	200
15	200
	1000

$$n = 1000$$
$$a = 300$$

$$\hat{p} = \frac{a}{n} = \frac{300}{1000} = 0{,}3$$

$$\hat{q} = 1 - \hat{p} = 1 - 0{,}3 = 0{,}7$$

$$S_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{0{,}3 \cdot 0{,}7}{1000}} = 0{,}0145$$

$$\alpha = 0{,}05$$
$$\hat{p} - Z_\alpha \cdot S_{\hat{p}} \langle p \langle \hat{p} + Z_\alpha \cdot S_{\hat{p}}$$
$$0{,}3 - 1{,}96 \cdot 0{,}0145 \langle p \langle 0{,}3 + 1{,}96 \cdot 0{,}0145$$
$$0{,}272 \langle p \langle 0{,}328$$

$$\alpha = 0{,}01$$
$$\hat{p} - Z_\alpha \cdot S_{\hat{p}} \langle p \langle \hat{p} + Z_\alpha \cdot S_{\hat{p}}$$
$$0{,}3 - 2{,}58 \cdot 0{,}0145 \langle p \langle 0{,}3 + 2{,}58 \cdot 0{,}0145$$
$$0{,}263 \langle p \langle 0{,}337$$

Primer 2. Od 100.000 domaćinstava izabran je uzorak bez ponavljanja od 2.000. Od tog broja 800 domaćinstava gajilo je svinje rase jorkšir, a 1.200 držalo je ostale rase.

a) oceniti proporciju domaćinstava koja gaje svinje rase jorkšir
b) odrediti ukupan broj domaćinstava koja gaje ovu rasu

$$\mathrm{N} = 100000$$
$$\boldsymbol{n} = 2000$$
$$\boldsymbol{a} = 800$$

$$\hat{p} - Z_{\alpha} \cdot S_{\hat{p}} \langle p \langle \hat{p} + Z_{\alpha} \cdot S_{\hat{p}}$$

$$\boldsymbol{N \cdot L_1 \langle Np \langle N \cdot L_2}$$

$$\hat{p} = \frac{a}{n} = \frac{800}{2000} = 0{,}4$$
$$\hat{q} = 1 - 0{,}4 = 0{,}6$$

$$S_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n} \cdot \frac{N-n}{N}} = \sqrt{\frac{0{,}4 \cdot 0{,}6}{2000} \cdot \frac{100000 - 2000}{100000}} = 0{,}0108$$

$$\alpha = 0{,}05$$
$$\hat{p} - Z_{\alpha} \cdot S_{\hat{p}} \langle p \langle \hat{p} + Z_{\alpha} \cdot S_{\hat{p}}$$
$$0{,}4 - 1{,}96 \cdot 0{,}0108 \langle p \langle 0{,}4 + 1{,}96 \cdot 0{,}0108$$
$$0{,}379 \langle p \langle 0{,}421$$

$$\alpha = 0{,}05$$
$$N \cdot L_1 \langle Np \langle N \cdot L_2$$
$$100000 \cdot 0{,}379 \langle Np \langle 100000 \cdot 0{,}421$$
$$37900 \langle Np \langle 42100$$

$$\alpha = 0{,}01$$
$$\hat{p} - Z_{\alpha} \cdot S_{\hat{p}} \langle p \langle \hat{p} + Z_{\alpha} \cdot S_{\hat{p}}$$
$$0{,}4 - 2{,}58 \cdot 0{,}0108 \langle p \langle 0{,}4 + 2{,}58 \cdot 0{,}0108$$
$$0{,}372 \langle p \langle 0{,}428$$

$$\alpha = 0{,}01$$
$$N \cdot L_1 \langle Np \langle N \cdot L_2$$
$$100000 \cdot 0{,}372 \langle Np \langle 100000 \cdot 0{,}428$$
$$32700 \langle Np \langle 42800$$