6 Uzoračka raspodjela

- Statističko ocjenjivanje
- Statistike uzorka kao ocjene parametara skupa
- Uzoračke raspodjele
- Ocjene i njihove osobine
- Stepeni slobode

6-1 Statistike uzorka kao ocjene parametara populacije

- Uzorački prosjek, $\bar{X}$, je ocjena prosjeka populacije, μ.
- Uzoračka varijansa, s^2, je ocjena populacione varijanse, σ^2.
- Uzoračka standardna devijacija, s, je ocjena standardne devijacije populacije, σ.
- Uzoračka proporcija, $\hat{p}$, je ocjena proporcije populacije, p.

Propocije populacije i uzorka

- Proporcija populacije je jednaka broju elemenata u populaciji koji pripadaju traženoj kategoriji, podijeljen sa ukupnim brojem elemenata u populaciji:

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

- Proporcija uzorka je broj elemenata u uzorku koji pripadaju traženoj kategoriji, podijeljen sa veličinom uzorka:

$$\hat{p} = \frac{x}{n}$$

Veze između parametara populacije i uzoračkih raspodjela prosjeka uzoraka

Očekivana vrijednost prosjeka uzoraka je jednaka prosjeku populacije:

$$E(\overline{X}) = \mu_{\overline{X}} = \mu_X$$

Varijansa prosjeka uzoraka je jednaka varijansi populacije podijeljenoj sa veličinom uzorka:

$$V(\overline{X}) = \sigma^2_{\overline{X}} = \frac{\sigma^2_X}{n}$$

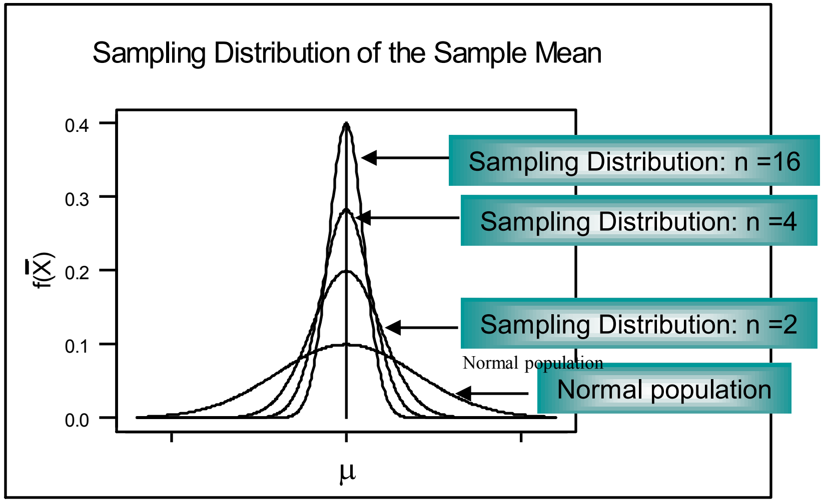
Standardna devijacija prosjeka uzoraka, poznata kao standardna greška prosjeka, jednaka je standardnoj devijaciji populacije podijeljenoj sa kvadratnim korjenom veličine uzorka:

$$SD(\overline{X}) = \sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}$$

Uzorkovanje iz normalne populacije

Kada se vrši uzorkovanje iz normalne populacije sa prosjekom μ i standardnom devijacijom σ, prosjek uzorka, $\bar{X}$, ima normalnu uzoračku raspodjelu:

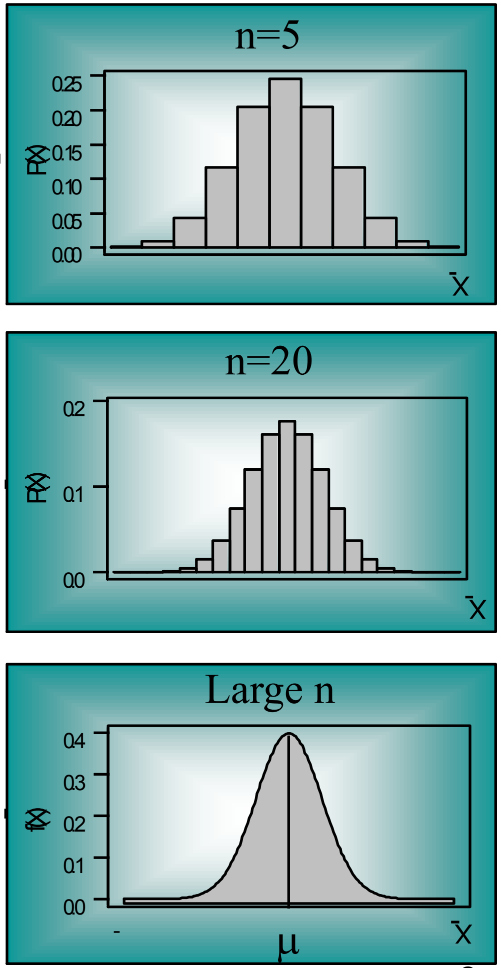
$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$



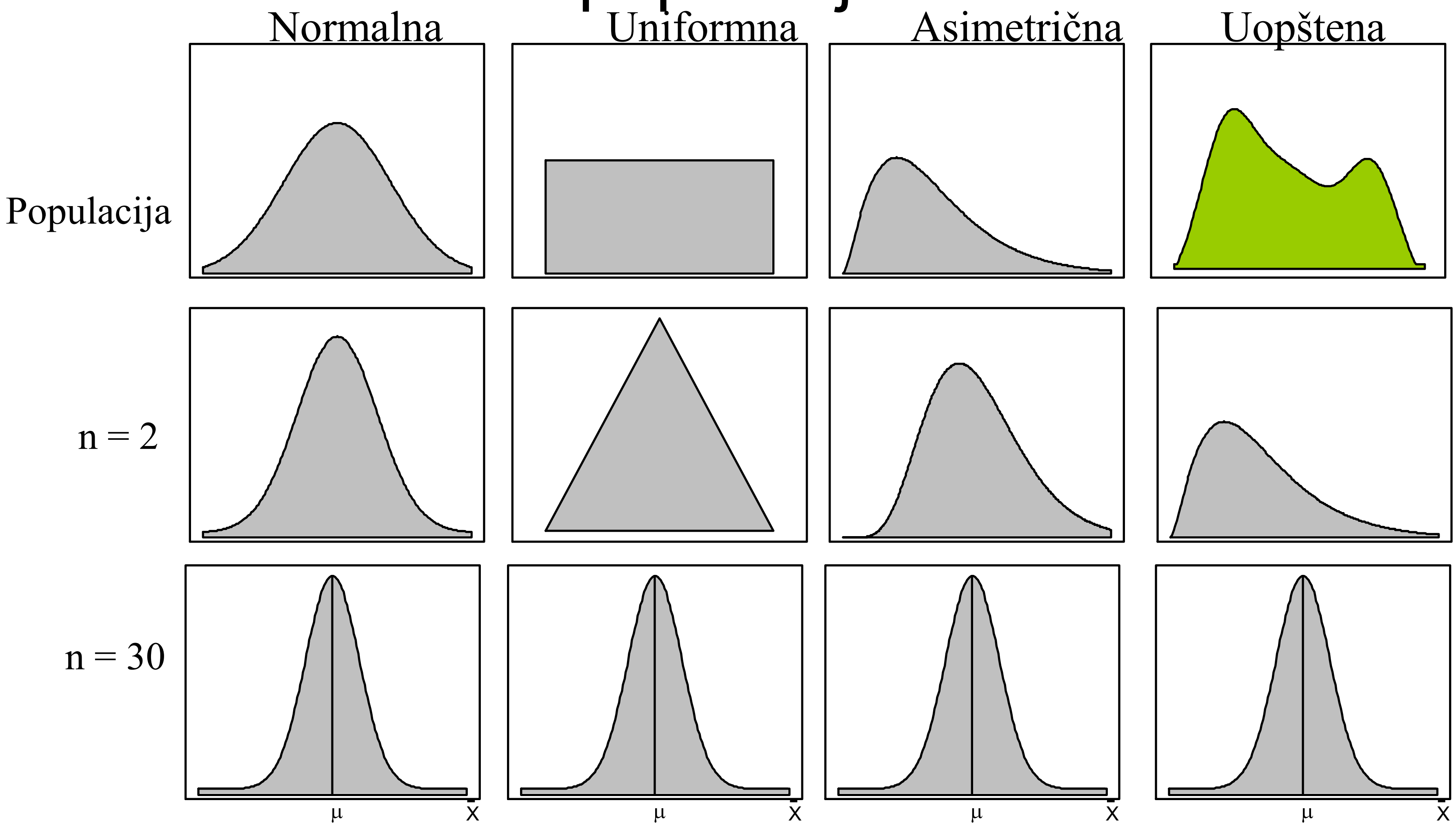
Centralna granična teorema

Kada je uzorak iz populacije sa prosjekom μ i konačnom stand. devijacijom σ, uzoračka raspodjela prosjeka uzorka teži normalnoj raspodjeli sa prosjekom μ i standardnom devijacijom $\frac{\sigma}{\sqrt{n}}$ kako veličina uzorka postaje velika (n >30).

Za “dovoljno veliko” *n*: $\bar{X} \sim N(\mu, \sigma^2 / n)$



Centralna granična teorema se primjenjuje na uzoračke raspodjele iz bilo koje populacije



Centralna granična teorema (Primjer 6-1)

Kompanija Mercury pravi 2.4 litarske V-6 motore, koji se koriste za glisere. Njeni inženjeri vjeruju da motor dostiže prosječnu snagu od 220 konjskih snaga i da je standardna devijacija snage 15 KS. Potencijalni kupac namjerava da uzorkuje 100 motora (svaki motor se provjerava jednom). Koja je vjerovatnoća da će prosjek uzorka biti manji od 217KS?

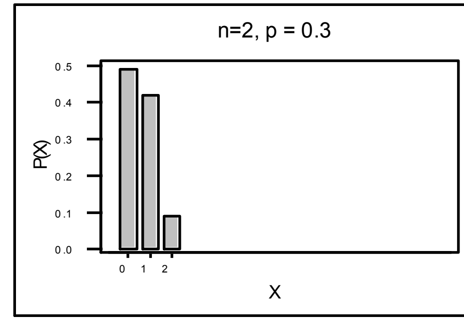
$$P(\overline{X} < 217) = P\left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{217 - \mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$= P\left(Z < \frac{217 - 220}{\frac{15}{\sqrt{100}}}\right) = P\left(Z < \frac{217 - 220}{\frac{15}{10}}\right)$$

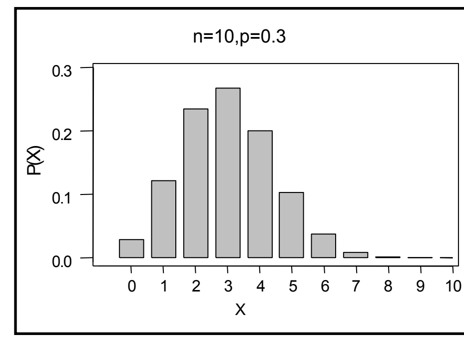
$$= P(Z < -2) = 0.0228$$

Uzoračka raspodjela proporcije uzorka

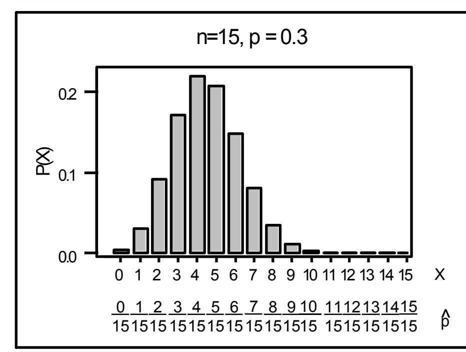
Uzoračka proporcija je procenat uspjeha u n binomnih pokušaja. To je broj uspjeha, X, podijeljen sa brojem pokušaja, n.



Uzoračka proporcija: $\hat{p} = \dfrac{X}{n}$



Sa porastom veličine uzorka, n, uzoračka raspodjela od $\hat{p}$ teži normalnoj raspodjeli sa prosjekom p i stand. devijacijom $\sqrt{\dfrac{p(1-p)}{n}}$



Uzoračka proporcija (Primjer 6-2)

Zadnjih godina, konvertibilna sportska kupe auta su postala popularna u Japanu. Pretpostavimo da je 25% svih Japanaca u određenoj klasi prihoda i stila života zainteresovano da kupi takvo Toyotino auto. Izabran je slučajni uzorak od 100 Japanskih potrošača iz tražene klase. Koja je vjerovatnoća da će najmanje 20% izraziti svoj interes za kupovinom ovakvog tipa vozila?

$$n = 100$$

$$p = 0.25$$

$$np = (100)(0.25) = 25 = E(\hat{p})$$

$$\frac{p(1-p)}{n} = \frac{(.25)(.75)}{100} = 0.001875 = V(\hat{p})$$

$$\sqrt{\frac{p(1-p)}{n}} = \sqrt{0.001875} = 0.04330127 = SD(\hat{p})$$

$$P(\hat{p} > 0.20) = P\left(\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} > \frac{.20 - p}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

$$= P\left(z > \frac{.20 - .25}{\sqrt{\frac{(.25)(.75)}{100}}}\right) = P\left(z > \frac{-.05}{.0433}\right)$$

$$= P(z > -1.15) = 0.8749$$

Standardizacija - proporcija

Prema istraživanju medija, 25,8% muškaraca u SAD-u je mjesečno gledalo TV u udarnom terminu. Ako se bira slučajan uzorak od 225 muškaraca, vjerovatnoća da će udio takvih muškaraca biti veći od 27% je:

- P(Z<0,258)
- **P(Z>0,41)**
- P(p>0,2)
- 0

$$P(\hat{p} > 0,27) = P(\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} > \frac{0,27 - p}{\sqrt{\frac{p(1-p)}{n}}}) = P(z > \frac{0,27 - 0,258}{\sqrt{\frac{0,258(1-0,258)}{225}}}) = P(z > 0,41)$$

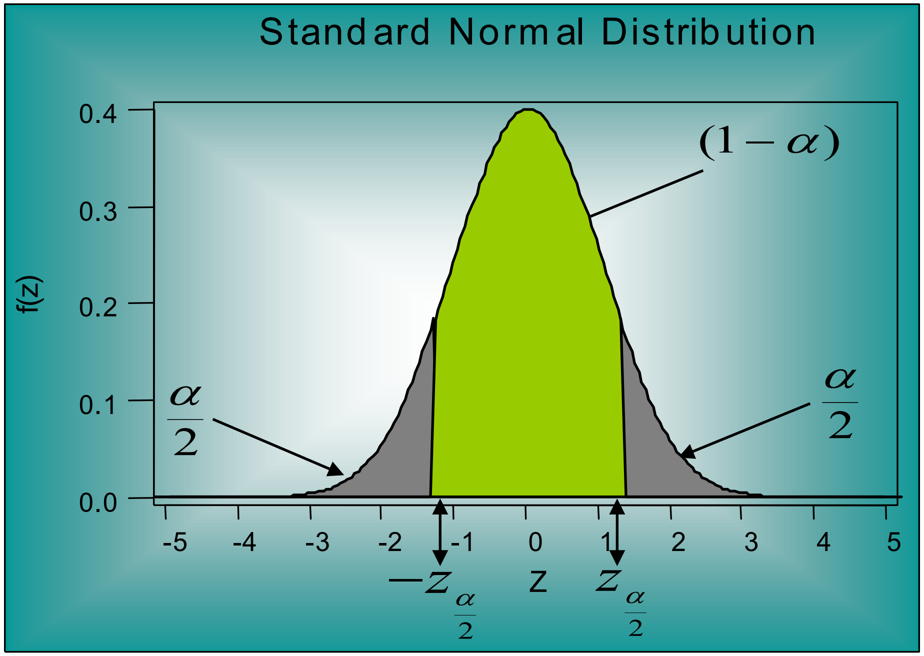
7

Intervali povjerenja

- Interval povjerenja za prosjek populacije kada je standardna devijacija poznata
- Intervali povjerenja za μ kada je σ nepoznata - *t* raspodjela
- Intervali povjerenja velikih uzoraka za proporciju populacije
- Korektivni faktor za konačnu populaciju
- Određivanje veličine uzoraka

(1-α)100% Interval povjerenja

α se zove *vjerovatnoća greške*, a (1-α)100% se zove *nivo povjerenja*.

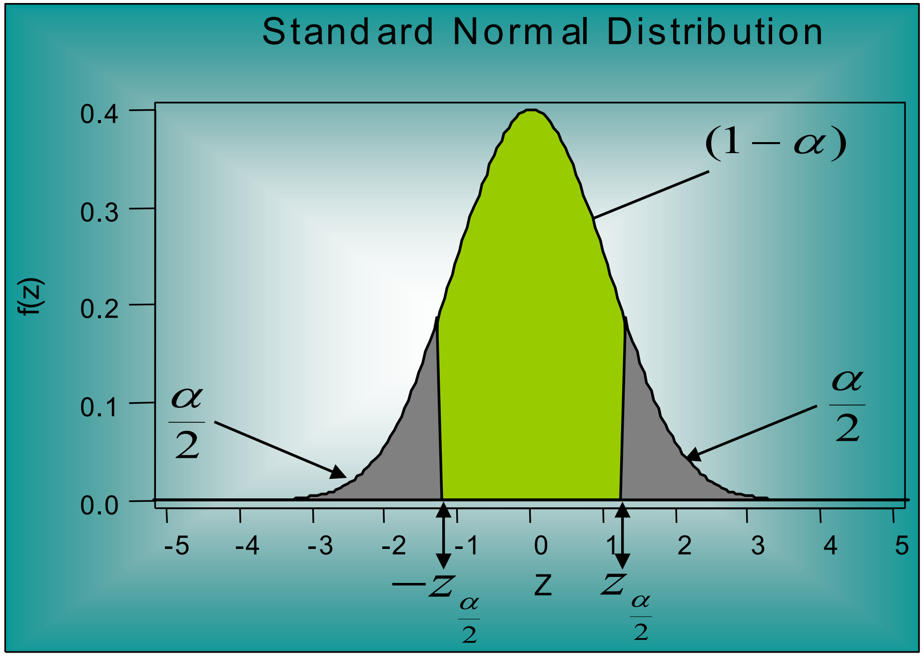


(1- α)100% Interval povjerenja

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Kritične vrijednosti od z i nivoi povjerenja

$(1-\alpha)$	$\frac{\alpha}{2}$	$z_{\frac{\alpha}{2}}$
0.99	0.005	2.576
0.98	0.010	2.326
0.95	0.025	1.960
0.90	0.050	1.645
0.80	0.100	1.282



95% Interval povjerenja za prosjek populacije

95% interval povjerenja za μ kada je σ poznata i uzorkovanje se vrši iz normalne populacije, ili se koristi veliki uzorak:

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

Iznos $1.96 \frac{\sigma}{\sqrt{n}}$ se često zove *marginalna greška* ili *greška uzorka*.

Npr, ako je: $n = 25$

$\sigma = 20$

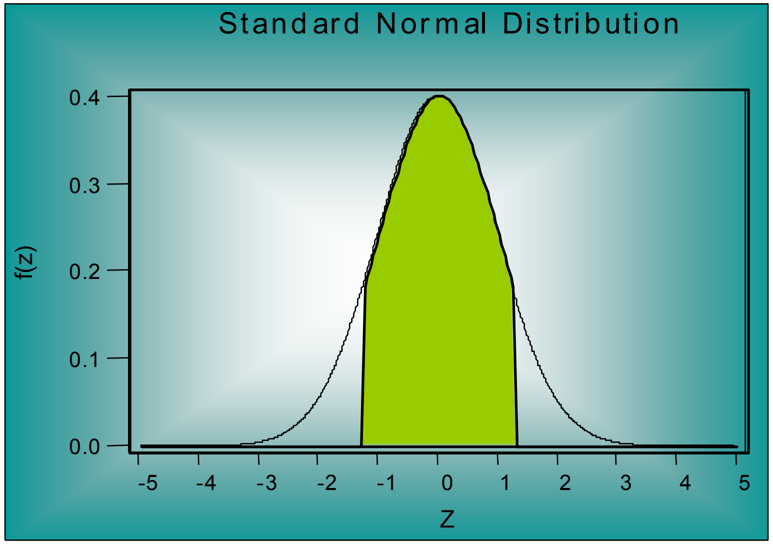
$\bar{x} = 122$

95% interval povjerenja:

$$\begin{aligned}\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} &= 122 \pm 1.96 \frac{20}{\sqrt{25}} \\ &= 122 \pm (1.96)(4) \\ &= 122 \pm 7.84 \\ &= [114.16, 129.84]\end{aligned}$$

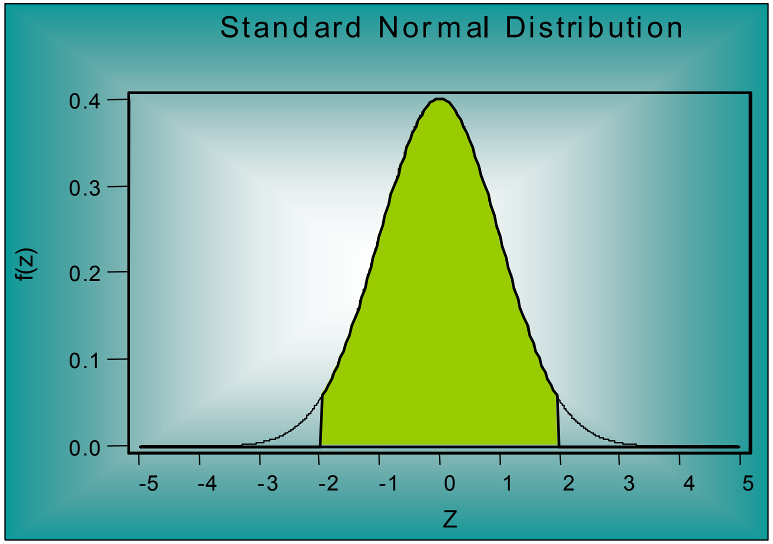
Nivo povjerenja i širina intervala povjerenja

Kada se bira uzorak iz iste populacije, koristeći istu veličinu uzorka, *što je veći nivo povjerenja, širi je interval povjerenja.*



80% Interval povjerenja :

$$\bar{x} \pm 1.28 \frac{\sigma}{\sqrt{n}}$$

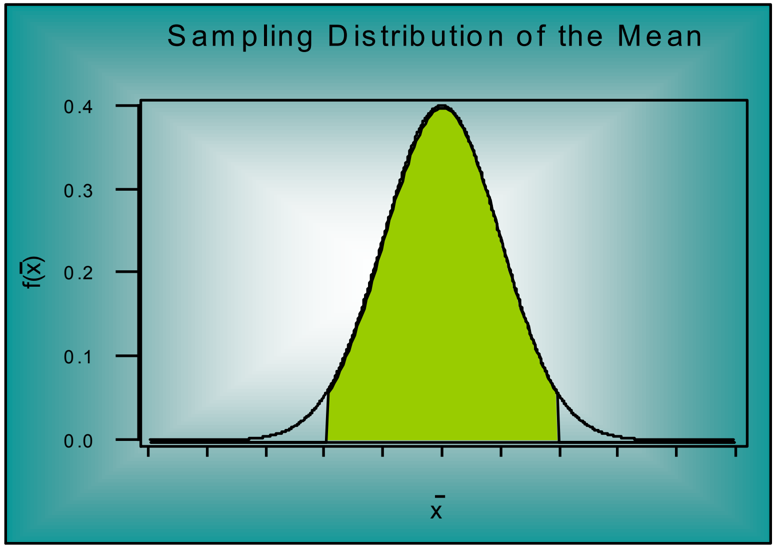


95% Interval povjerenja :

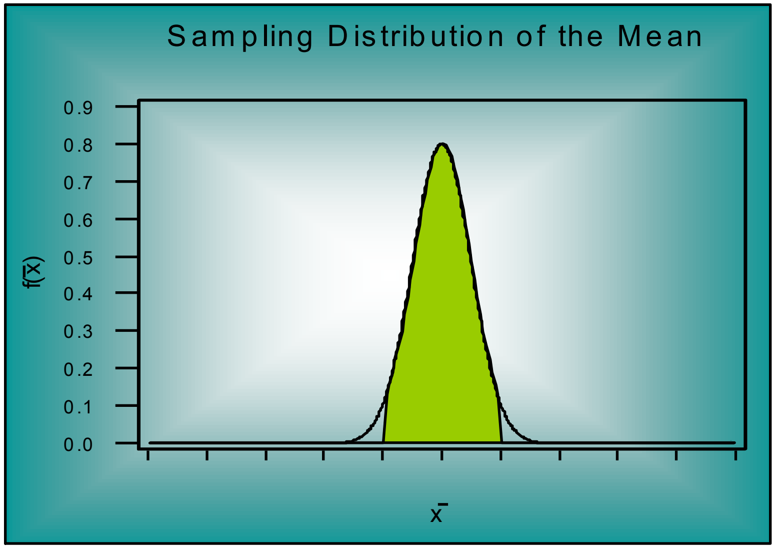
$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

Veličina uzorka i širina intervala povjerenja

Kada se bira uzorak iz iste populacije, uz isti nivo povjerenja, *što je veća veličina uzorka, n, to je uži interval povjerenja.*



95% Interval povjerenja: n = 20



95% Interval povjerenja: n = 40

Primjer 7-1

- Na cjelokupnoj listi kompanija (Fortune Web Site), rangirane po prihodima, želimo da saznamo prosječni prihod kompanija na listi. Standardna devijacija populacije je $15056.37. Slučajan uzorak od 30 kompanija daje prosjek uzorka od $10672.87. Naći 95% i 90% interval povjerenja za prosječan prihod svih kompanija na listi.

Primjer 7-1 *nastavak*

$(1-\alpha)$	$\frac{\alpha}{2}$	$z_{\frac{\alpha}{2}}$
0.99	0.005	2.576
0.98	0.010	2.326
0.95	0.025	1.960
0.90	0.050	1.645
0.80	0.100	1.282

$$P(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}) = 0{,}95$$

$$10672{,}87 - 1.96\frac{15056{,}37}{\sqrt{30}} \le \mu \le 10672{,}87 + 1.96\frac{15056{,}37}{\sqrt{30}}$$

$$10672{,}87 - 1.96 \cdot 2748{,}9 \le \mu \le 10672{,}87 + 1.96 \cdot 2748{,}9$$

$$10672{,}87 - 5387{,}95 \le \mu \le 10672{,}87 + 5387{,}95$$

$$5284{,}92 \le \mu \le 16060{,}82$$

$$P(\bar{x} - 1.64\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 1.64\frac{\sigma}{\sqrt{n}}) = 0{,}9$$

$$10672{,}87 - 1.64\frac{15056{,}37}{\sqrt{30}} \le \mu \le 10672{,}87 + 1.64\frac{15056{,}37}{\sqrt{30}}$$

$$6164{,}67 \le \mu \le 15181{,}06$$

7-2. zadatak

Istraživačko odjeljenje firme je uzelo uzorak od 25 odgovarajućih udžbenika i prikupilo podatke o njihovim cijenama. Njihova prosječna cijena je iznosila 90,5 eura. Standardna devijacija svih ovakvih knjiga je poznata i iznosi 7,5 eura i osnovni skup ima normalnu raspodjelu. 90% interval povjerenja za prosječnu cijenu svih udžbenika je:

$$\alpha = 0.1$$

$$z_{\frac{\alpha}{2}} = z_{0,05} = 1,645$$

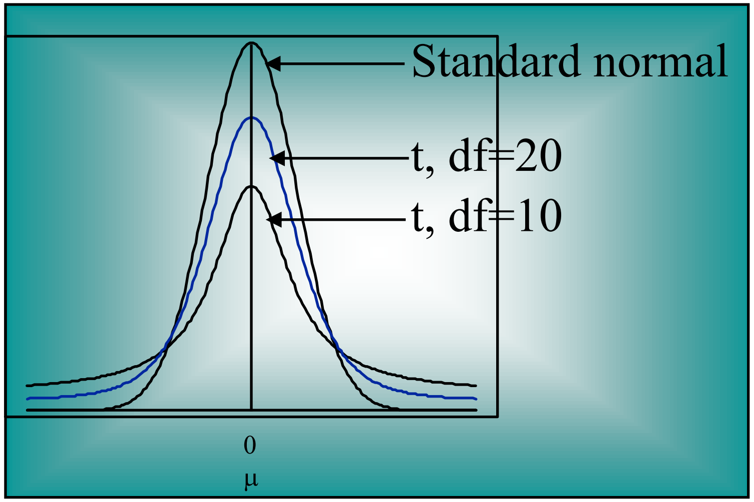
$$\bar{x} - z_{\frac{\alpha}{2}} \sigma_{\bar{x}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}} \sigma_{\bar{x}} = (88.02; 92.98)$$

7-3 Interval povjerenja za μ kada σ nije poznata - t raspodjela

Ako standardna devijacija populacije, σ, nije poznata, mijenjamo σ sa standardnom devijacijom uzorka, *s*. Ako je populacija normalna, tada statistika:

$$t = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}$$

ima *t raspodjelu* sa *(n - 1) stepeni slobode.*



7-3 Interval povjerenja za μ kada σ nije poznata - t raspodjela

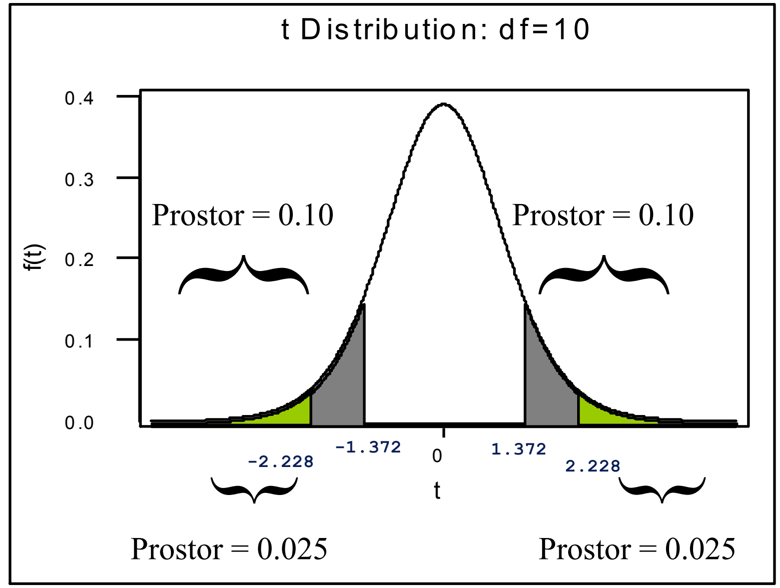
*(1-α)*100% interval povjerenja za μ kada σ nije poznata (pretpostavljajući normalnu raspodjelu populacije):

$$\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

Gdje je $t_{\frac{\alpha}{2}}$ vrijednost *t* raspodjele sa *n-1* stepeni slobode koji odvajaju prostor repa od $\frac{\alpha}{2}$ na desnu stranu.

t raspodjela

df	$t_{0.100}$	$t_{0.050}$	$t_{0.025}$	$t_{0.010}$	$t_{0.005}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	**1.372**	**1.812**	**2.228**	**2.764**	**3.169**
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	**1.282**	**1.645**	**1.960**	**2.326**	**2.576**



t raspodjela (Primjer 7-3)

Analitičar berzanskog tržišta želi da ocijeni prosječan prinos na određenu akciju. Slučajan uzorak od 15 dana daje prosječan (na godišnjem nivou) prinos od $\bar{x}$=10.37% i standardnu devijaciju od s = 3.5%. Uz pretpostavku normalne populacije prinosa, odrediti 95% interval povjerenja za prosječan prinos na ovu akciju.

df	$t_{0.100}$	$t_{0.050}$	$t_{0.025}$	$t_{0.010}$	$t_{0.005}$
1	3.078	6.314	12.706	31.821	63.657
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	**2.145**	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

Kritična vrijednost za t za df = (n-1)=(15-1)=14 i prostor desnog repa od 0.025 je:

$$t_{0.025} = 2.145$$

Odgovarajući interval povjerenja je:

$$\begin{aligned}
&\bar{x} \pm t_{0.025}\frac{s}{\sqrt{n}} \\
&= 10.37 \pm 2.145\frac{3.5}{\sqrt{15}} \\
&= 10.37 \pm 1.94 \\
&= [8.43, 12.31]
\end{aligned}$$

7-4. zadatak

25 slučajno odabranih osoba su upitane koliko obično godišnje troše na kupovinu knjiga. Prosječan iznos u uzorku iznosi 1450 eura sa standardnom devijacijom 300 eura. 99% interval povjerenja za odgovarajuću prosječnu vrijednost troška je:

$$t_{0.005;24} = 2,797$$

$$\bar{x} - t_{n-1;\frac{\alpha}{2}} s_{\bar{x}} \leq \mu \leq \bar{x} + t_{n-1;\frac{\alpha}{2}} s_{\bar{x}} = (1282.18;1617.82)$$

Intervali povjerenja velikih uzoraka za prosjek populacije

df	$t_{0.100}$	$t_{0.050}$	$t_{0.025}$	$t_{0.010}$	$t_{0.005}$
1	3.078	6.314	12.706	31.821	63.657
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
120	1.289	1.658	1.980	2.358	2.617
∞	**1.282**	**1.645**	**1.960**	**2.326**	**2.576**

Kad god σ nije poznata (i pretpostavlja se da je populacija normalna), pravilna raspodjela za upotrebu je *t* raspodjela sa *n-1* stepeni slobode. Primijetimo, takođe, da se za veliki broj stepeni slobode, t raspodjela dobro aproksimira *Z* raspodjelom.

Intervali povjerenja velikih uzoraka za prosjek populacije

$(1-\alpha)100\%$ interval povjerenja za veliki uzorak za μ:

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

Primjer 7-5: Ekonomista želi da ocijeni prosječan iznos na računima u bankama u datom regionu. Slučajan uzorak od 100 računa daje prosjek od $357.60 i s = $140.00. Naći 95% interval povjerenja za μ, prosječan iznos na svim bankovnim računima u datom regionu.

$$\bar{x} \pm z_{0.025} \frac{s}{\sqrt{n}} = 357.60 \pm 1.96 \frac{140.00}{\sqrt{100}} = 357.60 \pm 27.44 = [330.16, 385.04]$$

7-4 Intervali povjerenja velikih uzoraka za proporciju populacije, p

$(1\text{-}\alpha)100\%$ interval povjerenja velikih uzoraka za proporciju populacije, p $je:$

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

gdje je proporcija uzorka, $\hat{p}$, jednaka broju uspjeha u uzorku, x, podijeljenom sa brojem pokusaja (velicina uzorka), n, i $\hat{q} = 1 - \hat{p}$.

Intervali povjerenja velikih uzoraka za proporciju populacije, p (Primjer 7-6)

Firma koja se bavi marketing istraživanjem želi da ocijeni udio koji strane kompanije imaju na američkom tržištu određenih proizvoda. Slučajan uzorak od 100 potrošača je izabran, i nađeno je da 34 ljudi iz uzorka koriste proizvode stranih kompanija; ostali koriste domaće proizvode. Naći 95% interval povjerenja za udio stranih proizvoda na ovom tržištu.

$$
\begin{aligned}
\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} &= 0.34 \pm 1.96\sqrt{\frac{(0.34)(0.66)}{100}} \\
&= 0.34 \pm (1.96)(0.04737) \\
&= 0.34 \pm 0.0928 \\
&= [0.2472, 0.4328]
\end{aligned}
$$

Dakle, firma može biti 95% uvjerena da strani proizvođači kontrolišu negdje između 24.72% i 43.28% tržišta.

7-7. zadatak

Prema jednom istraživanju u kojem je bilo uključeno 1506 osoba, 75% njih je reklo da često imaju problema sa spavanjem. 99% interval povjerenja za odgovarajući procenat takvih ljudi je:

$$\hat{p} = 0.75$$

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = (0.721;0.779)$$

$$\alpha = 0.01$$

$$z_{\frac{\alpha}{2}} = z_{0,005} = 2,58$$