Statistika

Vanja Wagner

4. Procjene parametara

Problem. Tilapija je jedna od najčešćih uzgajanih vrsta slatkovodne ribe u svijetu. Zbog raznih uvjeta uzgoja (gustoća riblje populacije u uzgajalištu, manjak aktivnosti riba, industrijska hrana), ribe vrste Tilapija su često pretile. Smatra se da je riba ove vrste pretila ako njena masa prelazi 700 g, Promatramo uzgajalište na kojem se nalazi oko 50 000 riba vrste Tilapija i želimo odrediti postotak pretilih riba. Kako to učiniti?

Problem. Tilapija je jedna od najčešćih uzgajanih vrsta slatkovodne ribe u svijetu. Zbog raznih uvjeta uzgoja (gustoća riblje populacije u uzgajalištu, manjak aktivnosti riba, industrijska hrana), ribe vrste Tilapija su često pretile. Smatra se da je riba ove vrste pretila ako njena masa prelazi 700 g, Promatramo uzgajalište na kojem se nalazi oko 50 000 riba vrste Tilapija i želimo odrediti postotak pretilih riba. Kako to učiniti?

- Brojanje svih kuglica je prezahtjevan zadatak.
- Slučajno izaberemo određen broj riba (uzorak) i izmjerimo im masu.
- Izračunamo udio pretilih riba u promatranom uzorku.

Podsjetimo se pojmova - populacija, uzorak

Populacija je skup jedinki (objekata) koje imaju neku zajedničku mjerljivu karakteristiku (obilježje) - čini skup jedinki o kojem želimo donijeti zaključak.

Podsjetimo se pojmova - populacija, uzorak

Populacija je skup jedinki (objekata) koje imaju neku zajedničku mjerljivu karakteristiku (obilježje) - čini skup jedinki o kojem želimo donijeti zaključak.

Uzorak je (bilo koji) podskup populacije - na temelju poznavanja vrijednosti obilježja u uzorku želimo donijeti zaključak o cijeloj populaciji.

Podsjetimo se pojmova - populacija, uzorak

Populacija je skup jedinki (objekata) koje imaju neku zajedničku mjerljivu karakteristiku (obilježje) - čini skup jedinki o kojem želimo donijeti zaključak.

Uzorak je (bilo koji) podskup populacije - na temelju poznavanja vrijednosti obilježja u uzorku želimo donijeti zaključak o cijeloj populaciji.

Kako kreiramo uzorak

Svaka jedinica populacije može biti izabrana u uzorak s određenom vjerojatnošću. Razlikujemo:

- (Jednostavni) slučajni uzorak (*random sample*)
- Stratificirani slučajni uzorak (*stratified random sample*)
- Uzorak skupina (*cluster sampling*)
- ...

Podsjetimo se pojmova - populacija, uzorak

Populacija je skup jedinki (objekata) koje imaju neku zajedničku mjerljivu karakteristiku (obilježje) - čini skup jedinki o kojem želimo donijeti zaključak.

Uzorak je (bilo koji) podskup populacije - na temelju poznavanja vrijednosti obilježja u uzorku želimo donijeti zaključak o cijeloj populaciji.

Kako kreiramo uzorak

Svaka jedinica populacije može biti izabrana u uzorak s određenom vjerojatnošću. Razlikujemo:

- (Jednostavni) slučajni uzorak (*random sample*)
- Stratificirani slučajni uzorak (*stratified random sample*)
- Uzorak skupina (*cluster sampling*)
- ...

Jednostavni slučajni uzorak je vjerojatnosni uzorak gdje svaki član skupa ima **jednaku** vjerojatnost izbora u uzorak.

Informativno

Stratificirani slučajni uzorak

Prije izbora uzorka, razvrstavanje elemenata osnovnog skupa u stratume (podskupove populacije koji se međusobno ne preklapaju).

Razvrstavanjem u stratume dobijemo skupove s manjim stupnjem varijabilnosti.

Jedinke unutar istog stratuma imaju istu vjerojatnost izbora u uzorak

Uzorak skupina

Populaciju podijelimo u skupine (clusters)

Slučajno biramo skupine

Slučajni uzorak

(Jednostavni) slučajni uzorak duljine n dobijemo izborom n **nezavisnih** slučajnih jedinki iz populacije. Svakoj jedinki izmjerimo promatrano **obilježje** i vrijednosti obilježja na tih n jedinki u uzorku označimo s

$$X_1, X_2, X_3, \ldots, X_n.$$

Uočimo: $X_1, X_2, X_3, \ldots, X_n$ su slučajne varijable koje su

- jednako distribuirane;
- nezavisne.

Slučajni uzorak

(Jednostavni) slučajni uzorak duljine n dobijemo izborom n **nezavisnih** slučajnih jedinki iz populacije. Svakoj jedinki izmjerimo promatrano **obilježje** i vrijednosti obilježja na tih n jedinki u uzorku označimo s

$$X_1, X_2, X_3, \ldots, X_n.$$

Uočimo: $X_1, X_2, X_3, \ldots, X_n$ su slučajne varijable koje su

- jednako distribuirane;
- nezavisne.

Problem: udio pretilih Tiplapija

Odabrat ćemo na slučajan način $n = 100$ riba iz uzgajališta. Izmjerit ćemo im masu i odrediti jesu li pretile. Obilježje koje promatramo je binarno - je li riba pretila ili ne:

$$X = \begin{cases} 1, & \text{riba je pretila} \\ 0, & \text{riba nije pretila.} \end{cases}$$

Slučajne varijable $X_1, \ldots, X_{100}$ reprezentirat će varijablu koja mjeri pretilost. One su nezavisne i jednakodistribuirane. Koja je njihova distribucija?

Slučajni uzorak

(Jednostavni) slučajni uzorak duljine n dobijemo izborom n **nezavisnih** slučajnih jedinki iz populacije. Svakoj jedinki izmjerimo promatrano **obilježje** i vrijednosti obilježja na tih n jedinki u uzorku označimo s

$$X_1, X_2, X_3, \dots, X_n.$$

Uočimo: $X_1, X_2, X_3, \dots, X_n$ su slučajne varijable koje su

- jednako distribuirane;
- nezavisne.

Problem: udio pretilih Tiplapija

Odabrat ćemo na slučajan način $n = 100$ riba iz uzgajališta. Izmjerit ćemo im masu i odrediti jesu li pretile. Obilježje koje promatramo je binarno - je li riba pretila ili ne:

$$X = \begin{cases} 1, & \text{riba je pretila} \\ 0, & \text{riba nije pretila.} \end{cases}$$

Slučajne varijable $X_1, \dots, X_{100}$ reprezentirat će varijablu koja mjeri pretilost. One su nezavisne i jednakodistribuirane. Koja je njihova distribucija?

$$X_i \sim B(p),$$

gdje je $p \in (0, 1)$ parametar koji označava udio pretilih tilapija u populaciji.

Parametar populacije

Često nas zanima neka djelomična informacija o promatranom obilježju populacije, točnije neki **parametar** distribucije promatranog obilježja. Primjeri parametara populacije su:

- očekivanje,
- varijanca,
- standardna devijacija,
- medijan,
- kvantili,
- intenzitet.

Parametar populacije je najčešće nepoznat. Cilj nam je na temelju slučajnog uzorka dati *najbolju moguću procjenu* traženog parametra distribucije promatranog obilježja za cijelu populaciju.

Parametar populacije

Često nas zanima neka djelomična informacija o promatranom obilježju populacije, točnije neki **parametar** distribucije promatranog obilježja. Primjeri parametara populacije su:

- očekivanje,
- varijanca,
- standardna devijacija,
- medijan,
- kvantili,
- intenzitet.

Parametar populacije je najčešće nepoznat. Cilj nam je na temelju slučajnog uzorka dati *najbolju moguću procjenu* traženog parametra distribucije promatranog obilježja za cijelu populaciju.

Problem: udio pretilih Tiplapija

Zanima nas najbolja moguća procjena za parametar p (udio pretilih Tilapija) na temelju slučajnog uzorka $X_1, \ldots, X_{100} \sim B(p)$. Sjetimo se, parametar p ujedno je i očekivanje Bernoullijeve slučajne varijable, tj.

$$\mathbf{E}[X_1] = \ldots = \mathbf{E}[X_n] = p.$$

Procjenitelj

Statistika je bilo koja funkcija slučajnog uzorka $X_1, \ldots, X_n$, a **procjenitelj** je statistika kojom (točkovno, brojčano) procjenjujemo traženi parametar promatranog obilježja.

Primjeri procjenitelja za parametar:

- očekivanje je

Procjenitelj

Statistika je bilo koja funkcija slučajnog uzorka $X_1, \ldots, X_n$, a **procjenitelj** je statistika kojom (točkovno, brojčano) procjenjujemo traženi parametar promatranog obilježja.

Primjeri procjenitelja za parametar:

- očekivanje je aritmetička sredina, tj. prosjek,

Procjenitelj

Statistika je bilo koja funkcija slučajnog uzorka $X_1, \ldots, X_n$, a **procjenitelj** je statistika kojom (točkovno, brojčano) procjenjujemo traženi parametar promatranog obilježja.

Primjeri procjenitelja za parametar:

- očekivanje je aritmetička sredina, tj. prosjek,
- varijanca je

Procjenitelj

Statistika je bilo koja funkcija slučajnog uzorka $X_1, \ldots, X_n$, a **procjenitelj** je statistika kojom (točkovno, brojčano) procjenjujemo traženi parametar promatranog obilježja.

Primjeri procjenitelja za parametar:

- očekivanje je aritmetička sredina, tj. prosjek,
- varijanca je varijanca uzorka,

Procjenitelj

Statistika je bilo koja funkcija slučajnog uzorka $X_1, \ldots, X_n$, a **procjenitelj** je statistika kojom (točkovno, brojčano) procjenjujemo traženi parametar promatranog obilježja.

Primjeri procjenitelja za parametar:

- očekivanje je aritmetička sredina, tj. prosjek,
- varijanca je varijanca uzorka,
- standardna devijacija

Procjenitelj

Statistika je bilo koja funkcija slučajnog uzorka $X_1, \ldots, X_n$, a **procjenitelj** je statistika kojom (točkovno, brojčano) procjenjujemo traženi parametar promatranog obilježja.

Primjeri procjenitelja za parametar:

- očekivanje je aritmetička sredina, tj. prosjek,
- varijanca je varijanca uzorka,
- standardna devijacija je standardna devijacija uzorka,

Procjenitelj

Statistika je bilo koja funkcija slučajnog uzorka $X_1, \ldots, X_n$, a **procjenitelj** je statistika kojom (točkovno, brojčano) procjenjujemo traženi parametar promatranog obilježja.

Primjeri procjenitelja za parametar:

- očekivanje je aritmetička sredina, tj. prosjek,
- varijanca je varijanca uzorka,
- standardna devijacija je standardna devijacija uzorka,
- medijan je

Procjenitelj

Statistika je bilo koja funkcija slučajnog uzorka $X_1, \ldots, X_n$, a **procjenitelj** je statistika kojom (točkovno, brojčano) procjenjujemo traženi parametar promatranog obilježja.

Primjeri procjenitelja za parametar:

- očekivanje je aritmetička sredina, tj. prosjek,
- varijanca je varijanca uzorka,
- standardna devijacija je standardna devijacija uzorka,
- medijan je medijan uzork,

Procjenitelj

Statistika je bilo koja funkcija slučajnog uzorka $X_1, \ldots, X_n$, a **procjenitelj** je statistika kojom (točkovno, brojčano) procjenjujemo traženi parametar promatranog obilježja.

Primjeri procjenitelja za parametar:

- očekivanje je aritmetička sredina, tj. prosjek,
- varijanca je varijanca uzorka,
- standardna devijacija je standardna devijacija uzorka,
- medijan je medijan uzork,
- kvantili, je

Procjenitelj

Statistika je bilo koja funkcija slučajnog uzorka $X_1, \ldots, X_n$, a **procjenitelj** je statistika kojom (točkovno, brojčano) procjenjujemo traženi parametar promatranog obilježja.

Primjeri procjenitelja za parametar:

- očekivanje je aritmetička sredina, tj. prosjek,
- varijanca je varijanca uzorka,
- standardna devijacija je standardna devijacija uzorka,
- medijan je medijan uzork,
- kvantili, je kvantil uzorka.

Procjenitelj

Statistika je bilo koja funkcija slučajnog uzorka $X_1, \ldots, X_n$, a **procjenitelj** je statistika kojom (točkovno, brojčano) procjenjujemo traženi parametar promatranog obilježja.

Primjeri procjenitelja za parametar:

- očekivanje je aritmetička sredina, tj. prosjek,
- varijanca je varijanca uzorka,
- standardna devijacija je standardna devijacija uzorka,
- medijan je medijan uzork,
- kvantili, je kvantil uzorka.

Problem: udio pretilih Tiplapija

Zanima nas najbolja moguća procjena za parametar p (udio pretilih Tilapija) koji je ujedno očekivanje binarnog obilježja pretilost. Najbolja procjena za očekivanje na temelju uzorka $X_1, \ldots, X_{100}$ je prosjek, pa ćemo za procjenitelj $\hat{p}$ parametra p uzeti statistiku

$$\hat{p} := \overline{X}_{100} = \frac{X_1 + \ldots + X_{100}}{100} = \frac{\text{broj pretilih tilapija u uzorku}}{\text{duljina uzorka}}.$$

Procjenitelj za očekivanje

Dobar procjenitelj za očekivanje μ populacije je aritmetička sredina odnosno prosjek uzorka $X_1, \ldots, X_n$. To je slučajna varijabla

$$\overline{X}_n = \frac{X_1 + \ldots + X_n}{n}.$$

[1]Kako je očekivanje procjenitelja jednako procjenjivanom parametru kažemo da je procjenitelj **nepristran**.

Procjenitelj za očekivanje

Dobar procjenitelj za očekivanje μ populacije je aritmetička sredina odnosno prosjek uzorka $X_1, \dots, X_n$. To je slučajna varijabla

$$\overline{X}_n = \frac{X_1 + \ldots + X_n}{n}.$$

Podsjetimo se:

- očekivanje aritmetičke sredine uzorka jednako je očekivanju populacije:

  $$\mathbf{E}(\overline{X}_n) = \mu,$$ [1]

  a varijanca aritmetičke sredine uzorka jednaka je:

  $$\mathrm{Var}(\overline{X}_n) = \frac{1}{n}\sigma^2,$$

  gdje je σ^2 varijanca populacije.

[1] Kako je očekivanje procjenitelja jednako procjenjivanom parametru kažemo da je procjenitelj **nepristran**.

Procjenitelj za očekivanje

Dobar procjenitelj za očekivanje μ populacije je aritmetička sredina odnosno prosjek uzorka $X_1, \ldots, X_n$. To je slučajna varijabla

$$\overline{X}_n = \frac{X_1 + \ldots + X_n}{n}.$$

Podsjetimo se:

- očekivanje aritmetičke sredine uzorka jednako je očekivanju populacije:

  $$\mathbf{E}(\overline{X}_n) = \mu, \text{ [1]}$$

  a varijanca aritmetičke sredine uzorka jednaka je:

  $$\mathrm{Var}(\overline{X}_n) = \frac{1}{n}\sigma^2,$$

  gdje je σ^2 varijanca populacije.
- aritmetička sredina uzorka je **normalno distribuirana** ako je obilježje (populacija) normalno distribuirano, odnosno

  $$X_1, \ldots, X_n \sim N(\mu, \sigma^2) \implies \overline{X}_n \sim N(\mu, \sigma^2/n).$$

[1] Kako je očekivanje procjenitelja jednako procjenjivanom parametru kažemo da je procjenitelj **nepristran**.

Procjenitelj za očekivanje

Dobar procjenitelj za očekivanje μ populacije je aritmetička sredina odnosno prosjek uzorka $X_1, \ldots, X_n$. To je slučajna varijabla

$$\overline{X}_n = \frac{X_1 + \ldots + X_n}{n}.$$

Podsjetimo se:

- očekivanje aritmetičke sredine uzorka jednako je očekivanju populacije:

  $$\mathbf{E}(\overline{X}_n) = \mu,$$ [1]

  a varijanca aritmetičke sredine uzorka jednaka je:

  $$\mathrm{Var}(\overline{X}_n) = \frac{1}{n}\sigma^2,$$

  gdje je σ^2 varijanca populacije.
- aritmetička sredina uzorka je **normalno distribuirana** ako je obilježje (populacija) normalno distribuirano, odnosno

  $$X_1, \ldots, X_n \sim N(\mu, \sigma^2) \implies \overline{X}_n \sim N(\mu, \sigma^2/n).$$
- aritmetička sredina uzorka je **približno normalno distribuirana** ako je uzorak relativno velik ($n \geqslant 30$), što je posljedica (**centralnog graničnog teorema**).

[1] Kako je očekivanje procjenitelja jednako procjenjivanom parametru kažemo da je procjenitelj **nepristran**.

Procjenitelj za očekivanje

Dobar procjenitelj za očekivanje μ populacije je aritmetička sredina odnosno prosjek uzorka $X_1, \ldots, X_n$. To je slučajna varijabla

$$\overline{X}_n = \frac{X_1 + \ldots + X_n}{n}.$$

Uočimo:

- Za računanje varijance aritmetičke sredine uzorka važna pretpostavka je da su $X_1, X_2, X_3, ..., X_n$ **nezavisne** slučajne varijable.

Procjenitelj za očekivanje

Dobar procjenitelj za očekivanje μ populacije je aritmetička sredina odnosno prosjek uzorka $X_1, \ldots, X_n$. To je slučajna varijabla

$$\overline{X}_n = \frac{X_1 + \ldots + X_n}{n}.$$

Uočimo:

- Za računanje varijance aritmetičke sredine uzorka važna pretpostavka je da su $X_1, X_2, X_3, ..., X_n$ **nezavisne** slučajne varijable.
- Povećanjem veličine uzorka (n) smanjuje se varijanca procjenitelja.

Procjenitelj za očekivanje

Dobar procjenitelj za očekivanje μ populacije je aritmetička sredina odnosno prosjek uzorka $X_1, \ldots, X_n$. To je slučajna varijabla

$$\overline{X}_n = \frac{X_1 + \ldots + X_n}{n}.$$

Uočimo:

- Za računanje varijance aritmetičke sredine uzorka važna pretpostavka je da su $X_1, X_2, X_3, ..., X_n$ **nezavisne** slučajne varijable.
- Povećanjem veličine uzorka (n) smanjuje se varijanca procjenitelja.
- **Standardna pogreška** je standardna devijacija procjenitelja, u ovom slučaju je

$$sem = \sqrt{\mathrm{Var}(\bar{X})} = \frac{\sigma}{\sqrt{n}}$$

Procjenitelj za očekivanje

Dobar procjenitelj za očekivanje μ populacije je aritmetička sredina odnosno prosjek uzorka $X_1, \ldots, X_n$. To je slučajna varijabla

$$\overline{X}_n = \frac{X_1 + \ldots + X_n}{n}.$$

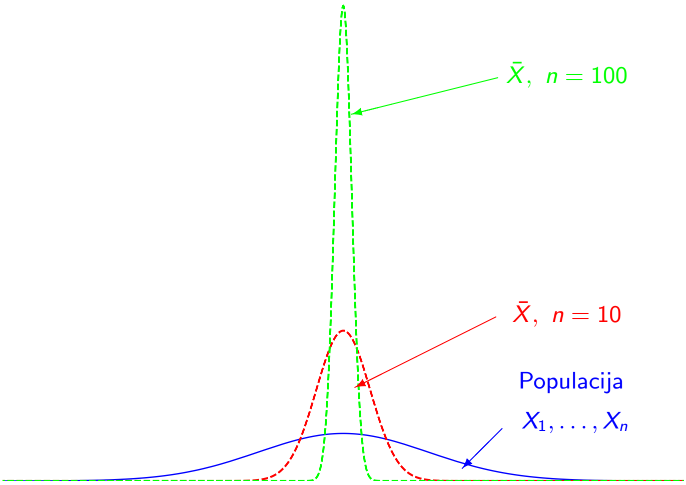
Uočimo:

- Za računanje varijance aritmetičke sredine uzorka važna pretpostavka je da su $X_1, X_2, X_3, ..., X_n$ **nezavisne** slučajne varijable.
- Povećanjem veličine uzorka (n) smanjuje se varijanca procjenitelja.
- **Standardna pogreška** je standardna devijacija procjenitelja, u ovom slučaju je

$$sem = \sqrt{\mathrm{Var}(\bar{X})} = \frac{\sigma}{\sqrt{n}}$$

- Varijanca procjenitelja i standardna pogreška su mjere točnosti procjenitelja - manja varijanca znači da je procjenitelj bolji.

Aritmetička sredina uzorka - funkcija gustoće vjerojatnosti



Procjenitelj za varijancu

Varijanca populacije σ^2 procjenjujemo **varijancom uzorka**

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

[2] S^2 je nepristrani procjenitelj varijance.
[3] S nije nepristrani procjenitelj za σ.

Procjenitelj za varijancu

Varijanca populacije σ^2 procjenjujemo **varijancom uzorka**

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

Podsjetimo se:

- U gornjem izrazu se suma dijeli s $n-1$ a ne s n! To je zato što ovako definiran procjenitelj zadovoljava
$$\mathbf{E}(S^2) = \sigma^2.$$
[2]
- Standardnu devijaciju procjenjujemo s $S = \sqrt{S^2}$.[3]

[2] S^2 je nepristrani procjenitelj varijance.
[3] S nije nepristrani procjenitelj za σ.

Procjenitelj za parametar proporcije

Porporciju nekog tipa promatranog obilježja u populaciji, odnosno parametar p iz Bernoullijeve distribucije, procjenjujemo s

$$\hat{p} = \text{udio} = \frac{\text{broj jedinki s određenim tipom obilježja}}{\text{ukupan broj jedinki}}$$

Podsjetimo se:

- Uzorak $X_1, X_2, \ldots, X_n$ određen je s $X_i = \begin{cases} 1 & \text{ukoliko se tip obilježje pojavio} \\ 0 & \text{ukoliko se tip obilježje nije pojavio} \end{cases}$
- $X_i \sim B(p)$, $\mathbf{E}(X_i) = p$, $\mathrm{Var}(X_i) = p \cdot (1-p)$
- Broj jedinki u uzorku koje imaju tip promatranog obilježja jednak je $X = \sum X_i$ i vrijedi $X \sim B(n, p)$ te

$$\mathbf{E}(X) = n \cdot p, \quad \mathrm{Var}(X) = n \cdot p \cdot (1-p).$$

- Sada je[4]

$$\mathbf{E}(\hat{p}) = \mathbf{E}\left(\frac{X}{n}\right) = \frac{1}{n}\mathbf{E}(X) = \frac{1}{n} n \cdot p = p.$$

i

$$\mathrm{Var}(\hat{p}) = \mathrm{Var}\left(\frac{X}{n}\right) = \frac{1}{n^2}\mathrm{Var}(X) = \frac{1}{n^2} n \cdot p \cdot (1-p) = \frac{p \cdot (1-p)}{n}$$

[4] Procjenitelj je nepristran.

Intervalna procjena očekivanja

Koliko je pojedina procjena očekivanja točna, pouzdana? Standardna pogreška je mjera točnosti. Umjesto točkovne procjene očekivanja odgovarajućim procjeniteljem (prosjekom uzorka) možemo promotriti **intervalnu procjenu** tako da tražimo da se naš parametar nalazi u tom intervalu s vjerojatnošću npr. 95%. [5mm] Prisjetimo se:

- Ako je razdioba obilježja u populaciji normalna, tada je

$$Y = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

- Ako je uzorak je dovoljno velik ($n \geqslant 30$) tada

$$Y = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx N(0,1).$$

- Stoga, s vjerojatnošću 95% je

$$\left| \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right| \leqslant z_{0.975}$$

odnosno

$$\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}.$$

Interval

$$\left[\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right]$$

se naziva 95%-tni **interval pouzdanosti za očekivanje.**

Interval

$$\left[\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right]$$

se naziva 95%-tni **interval pouzdanosti za očekivanje.**

Što ako nam nije poznata standardna devijacija populacije (σ)?

Interval

$$\left[\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right]$$

se naziva 95%-tni **interval pouzdanosti za očekivanje.**

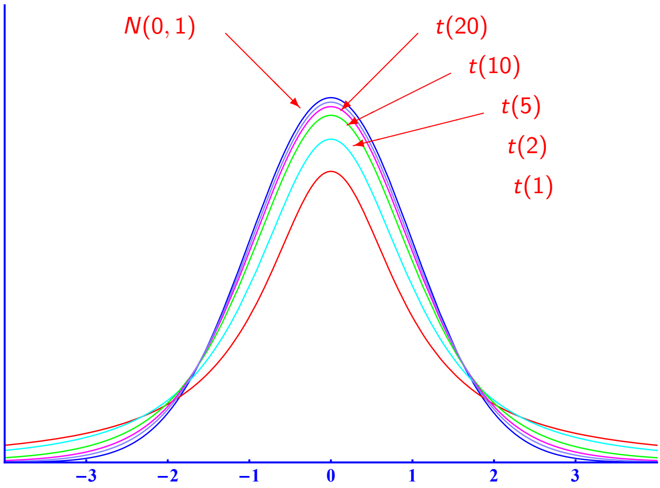
Što ako nam nije poznata standardna devijacija populacije (σ)? Umjesto σ možemo staviti procjenitelj S, odnosno standardnu devijaciju uzorka. Dobijemo interval

$$\left[\bar{X} - a \cdot \frac{S}{\sqrt{n}}, \bar{X} + a \cdot \frac{S}{\sqrt{n}}\right],$$

gdje je

- $a = t_{0.975}(n-1)$ kvanil *Studentove t-distribucije* s $n-1$ stupnjeva slobode za mali uzorak ($n < 30$);
- $a = z_{0.975} = 1.96$ za veliki uzorak ($n \geqslant 30$).

Usporedba jedinične normalne i različitih Studentovih razdioba



DZ. Odredite 90%-tni interval pouzdanosti za očekivanje (σ poznat).

DZ. Odredite 90%-tni interval pouzdanosti za očekivanje (σ poznat).

$$\left[\bar{X} - \mathbf{1.64} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + \mathbf{1.64} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Intervalna procjena proporcije

Podsjetimo se, $X = n\hat{p}$ je ukupan broj uspjeha, gdje je $\hat{p}$ procjena proporcije (vjerojatnosti) uspjeha. Vrijedi $X = n\hat{p} \sim B(n, p)$. Tu činjenicu bismo mogli iskoristiti da nađemo $100(1-\alpha)\%$ pouzdani interval za broj uspjeha ($X = n \cdot \hat{p}$):

$$[b_{n,\alpha/2}, b_{n,1-\alpha/2}]$$

gdje je

$$P(X \leqslant b_{n,p}) = p.$$

Intervalna procjena proporcije

Podsjetimo se, $X = n\hat{p}$ je ukupan broj uspjeha, gdje je $\hat{p}$ procjena proporcije (vjerojatnosti) uspjeha. Vrijedi $X = n\hat{p} \sim B(n, p)$. Tu činjenicu bismo mogli iskoristiti da nađemo $100(1-\alpha)\%$ pouzdani interval za broj uspjeha ($X = n \cdot \hat{p}$):

$$[b_{n,\alpha/2}, b_{n,1-\alpha/2}]$$

gdje je

$$P(X \leqslant b_{n,p}) = p.$$

Ali ovaj problem nema rješenje za svaki p, jer je X diskretna slučajna varijabla (poprima samo konačno mnogo vrijednosti $0, 1, \ldots, n$. Umjesto toga koristimo normalnu aproksimaciju binomne razdiobe (centralni granični teorem).

Intervalna procjena proporcije

Koristit ćemo činjenicu da za veliku duljinu uzorka n i $X_1, \ldots, X_n \sim B(p)$ vrijedi

$$\frac{\bar{X}_n - \mathbf{E}(\bar{X}_n)}{\sqrt{\mathrm{Var}(\bar{X}_n)}} = \frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$$

Intervalna procjena proporcije

Koristit ćemo činjenicu da za veliku duljinu uzorka n i $X_1, \dots, X_n \sim B(p)$ vrijedi

$$\frac{\bar{X}_n - \mathbf{E}(\bar{X}_n)}{\sqrt{\mathrm{Var}(\bar{X}_n)}} = \frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$$

Uočimo da nam se prava vrijednost proporcije p (nepoznata vrijednost) javlja i u brojniku i u nazivniku. Da bismo si olakšali procjenu intervala standardnu pogrešku za proporciju

$$\sigma_{\hat{p}} = \sqrt{\frac{p \cdot (1-p)}{n}}$$

zamijenit ćemo procjenom

$$\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}.$$

Intervalna procjena proporcije

Koristit ćemo činjenicu da za veliku duljinu uzorka n i $X_1, \ldots, X_n \sim B(p)$ vrijedi

$$\frac{\bar{X}_n - \mathbf{E}(\bar{X}_n)}{\sqrt{\mathrm{Var}(\bar{X}_n)}} = \frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$$

Uočimo da nam se prava vrijednost proporcije p (nepoznata vrijednost) javlja i u brojniku i u nazivniku. Da bismo si olakšali procjenu intervala standardnu pogrešku za proporciju

$$\sigma_{\hat{p}} = \sqrt{\frac{p \cdot (1-p)}{n}}$$

zamijenit ćemo procjenom

$$\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}.$$

Sada s vjetrojatnošću 95% znamo da vrijedi da je

$$\left| \frac{\bar{X}_n - p}{\hat{\sigma}_{\hat{p}}} \right| \leqslant z_{0.0975} = 1.96,$$

pa je 95%-tni interval pouzdanosti (Waldov interval)

$$[\hat{p} - \mathbf{1.96} \cdot \hat{\sigma}_{\hat{p}}, \hat{p} + \mathbf{1.96} \cdot \hat{\sigma}_{\hat{p}}]$$

Određivanje veličine uzorka za intervalnu procjenu očekivanja

Za danu razinu pouzdanosti $1-\alpha$ želimo odrediti veličinu uzorka potrebnu da $(1-\alpha)$%-tni pouzdani interval za očekivanje ne bude širi od od neke fiksne duljine $2d$. Uzmimo $\alpha = 0.05$ i promotrimo 95%-tni interval pouzdanosti za očekivanje (varijanca poznata):

$$\left[\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Određivanje veličine uzorka za intervalnu procjenu očekivanja

Za danu razinu pouzdanosti $1-\alpha$ želimo odrediti veličinu uzorka potrebnu da $(1-\alpha)$%-tni pouzdani interval za očekivanje ne bude širi od od neke fiksne duljine $2d$. Uzmimo $\alpha = 0.05$ i promotrimo 95%-tni interval pouzdanosti za očekivanje (varijanca poznata):

$$\left[\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Širina intervala jednaka je:

$$2 \cdot 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

pa tražimo n takav da je

$$2 \cdot 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leqslant 2d.$$

Određivanje veličine uzorka za intervalnu procjenu očekivanja

Za danu razinu pouzdanosti $1-\alpha$ želimo odrediti veličinu uzorka potrebnu da $(1-\alpha)$%-tni pouzdani interval za očekivanje ne bude širi od od neke fiksne duljine $2d$. Uzmimo $\alpha = 0.05$ i promotrimo 95%-tni interval pouzdanosti za očekivanje (varijanca poznata):

$$\left[\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Širina intervala jednaka je:

$$2 \cdot 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

pa tražimo n takav da je

$$2 \cdot 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leqslant 2d.$$

odnosno

$$n \geqslant \left(1.96 \frac{\sigma}{d}\right)^2.$$

Određivanje veličine uzorka za intervalnu procjenu proporcije

Kod procjene proporcije, $(1-\alpha)$%-tni interval pouzdanosti je dan s:

$$\left[\hat{p} - z_{1-\alpha/2} \cdot \sigma_{\hat{p}},\ \hat{p} + z_{1-\alpha/2} \cdot \sigma_{\hat{p}}\right]$$

gdje je

$$\sigma_{\hat{p}} = \frac{\sqrt{p(1-p)}}{\sqrt{n}}.$$

Određivanje veličine uzorka za intervalnu procjenu proporcije

Kod procjene proporcije, $(1-\alpha)$%-tni interval pouzdanosti je dan s:

$$\left[\hat{p} - z_{1-\alpha/2} \cdot \sigma_{\hat{p}},\ \hat{p} + z_{1-\alpha/2} \cdot \sigma_{\hat{p}}\right]$$

gdje je

$$\sigma_{\hat{p}} = \frac{\sqrt{p(1-p)}}{\sqrt{n}}.$$

Veličina uzorka bi trebala opet zadovoljavati

$$n \geqslant \left(z_{1-\alpha/2}\frac{\sigma}{d}\right)^2,$$

no kako je p nepoznat, koristimo ocjenu $p \cdot (1-p) \geqslant 0.5$ da bismo zaključili da je

$$n \geqslant \left(z_{1-\alpha/2}\frac{0.5}{d}\right)^2.$$

Određivanje veličine uzorka za intervalnu procjenu proporcije

Kod procjene proporcije, $(1-\alpha)$%-tni interval pouzdanosti je dan s:

$$\left[\hat{p} - z_{1-\alpha/2} \cdot \sigma_{\hat{p}},\ \hat{p} + z_{1-\alpha/2} \cdot \sigma_{\hat{p}}\right]$$

gdje je

$$\sigma_{\hat{p}} = \frac{\sqrt{p(1-p)}}{\sqrt{n}}.$$

Veličina uzorka bi trebala opet zadovoljavati

$$n \geqslant \left(z_{1-\alpha/2}\frac{\sigma}{d}\right)^2,$$

no kako je p nepoznat, koristimo ocjenu $p \cdot (1-p) \geqslant 0.5$ da bismo zaključili da je

$$n \geqslant \left(z_{1-\alpha/2}\frac{0.5}{d}\right)^2.$$

Ovako dobiveni n je sigurno veći od optimalnog (dobivena točnost će biti veća od tražene).

Primjer. Odredite veličinu uzorka potrebnu da širina 95%-tnog intervala pouzdanosti bude manja od 2 postotna poena u procjeni udjela ženskih riba vrste tilapia u uzgajalištu.

Primjer. Odredite veličinu uzorka potrebnu da širina 95%-tnog intervala pouzdanosti bude manja od 2 postotna poena u procjeni udjela ženskih riba vrste tilapia u uzgajalištu.

Rješenje.

$2d = 2\% = 0.02$ pa je $d = 0.01$. Očekujemo $p \approx 0.5$ pa uočite da nećemo jako povećati uzorak ako iskoristimo ocjenu $p(1-p) \geqslant 0.5$. Dobivamo

$$n \geqslant \left(z_{1-\alpha/2}\frac{\sigma}{d}\right)^2 =$$

Primjer. Odredite veličinu uzorka potrebnu da širina 95%-tnog intervala pouzdanosti bude manja od 2 postotna poena u procjeni udjela ženskih riba vrste tilapia u uzgajalištu.

Rješenje.

$2d = 2\% = 0.02$ pa je $d = 0.01$. Očekujemo $p \approx 0.5$ pa uočite da nećemo jako povećati uzorak ako iskoristimo ocjenu $p(1-p) \geqslant 0.5$. Dobivamo

$$n \geqslant \left(z_{1-\alpha/2}\frac{\sigma}{d}\right)^2 = \left(1.96 \cdot \frac{0.5}{0.01}\right)^2 = 9604.$$