

Prosečna dužina skoka udalj petnaestogodišnjaka: procena, interval pouzdanosti i test hipoteze

Lekcija za samostalno učenje (korak po korak)

1. Šta želim da znam

Želim da znam kolika je prosečna dužina skoka udalj petnaestogodišnjaka.

2. Da li mogu da znam tačan prosek bez merenja svih 15000?

Da li ću moći da znam koliki je tačan prosek ako ne izmerim svih 15000? Strogo uzevši, matematički je nemoguće dati tačan odgovor ako nisam izmerio sve. Odgovor je: NE MOGU.

3. Šta mogu: procena i pouzdanost procene

Ono što mogu je da procenim i da obezbedim da ta procena bude što pouzdanija.

Osnova statističkog zaključivanja svodi se na to da odaberemo uzorak. Recimo da uzmem $n = 100$ nasumično izabranih petnaestogodišnjaka. Izmerim skok udalj i izračunam srednju vrednost uzorka $\bar{x}$. Mogu da tvrdim da je $\bar{x}$ verovatno blizu stvarne vrednosti μ , ali ne znam koliko sam pogrešio.

4. Šta se dešava ako više puta ponovim uzorkovanje?

Ako bih više puta ponovio taj postupak (ja ili neko drugi), svaki put bih dobio novu srednju vrednost uzorka: $\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots$. Te vrednosti bi bile slične jedna drugoj, ali ne identične — malo bi se razlikovale zbog greške uzorkovanja.

Ako bih taj postupak ponovio dovoljno često i grafički prikazao učestalost dobijenih srednjih vrednosti uzorka, dobio bih krivu koja je približno normalna i koja ima najveću frekvenciju pojavljivanja oko prave srednje vrednosti μ , čak i ako rezultati svih 15000 učenika nisu normalno raspodeljeni.

5. Standardna greška sredine (SE)

Standardna devijacija raspodele srednjih vrednosti uzorka zove se standardna greška sredine (SE):

$$SE = s / \sqrt{n}$$

gde je s standardna devijacija uzorka, a n veličina uzorka.

6. Z vrednosti i 95% (ideja 'sigurnosti')

Iz onoga što smo ranije naučili o normalnoj raspodeli, znamo da se približno 95% vrednosti nalazi između -1.96 i $+1.96$ standardnih devijacija od sredine (Z vrednosti). Prevedeno na raspodelu sredina uzorka, to znači: oko 95% srednjih vrednosti uzorka bi bilo u intervalu:

$$\mu \pm 1.96 \cdot SE$$

Sa time mogu da budem zadovoljan, ali ovde imam dva problema: (1) ne znam kolika je prava μ , i (2) ovakav postupak je nepraktičan jer bih morao da uzorkujem mnogo puta da bih video tu raspodelu.

7. Rešenje: interval umesto jedne vrednosti

Zato umesto da pokušavam da 'vidim' celu raspodelu srednjih vrednosti uzorka, ja iz jednog uzorka dobijem statistiku $\bar{x}$ i svesno prihvatim da pravim grešku. Umesto jedne vrednosti definišem interval za koji očekujem da obuhvata pravu srednju vrednost μ .

Ako povećavam interval, biću sigurniji da sam u njemu obuhvatio μ , ali što je interval širi, njegov praktični značaj je manji.

8. Od čega zavisi širina intervala?

Da bih procenio optimalnu širinu intervala, polazim od tri činjenice:

- Što je uzorak heterogeniji, veći je s i interval mora biti širi.
- Što je uzorak manji, veći je $SE = s/\sqrt{n}$ i interval mora biti širi.
- Što više želim da budem siguran (90%, 95%, 99%), kritična vrednost mora biti veća i interval mora biti širi.

9. Zašto se pojavljuje t-raspodela (korekcija Z vrednosti)

Do sada smo kritičnu vrednost uzimali iz Z tabele (npr. za 95% je $Z = 1.96$). Međutim, u realnosti ne znamo populacionu SD (σ), nego je procenjujemo preko s . Ta dodatna nesigurnost zahteva korekciju kritične vrednosti u zavisnosti od veličine uzorka.

Ta korekcija daje t-raspodelu, čije kritične vrednosti zavise od stepena slobode:

$$df = n - 1$$

10. Nivo poverenja, verovatnoća greške i margina greške

10.1. Nivo poverenja (CL) i verovatnoća greške (α)

Nivo poverenja (CL) određuje koliko želim da budem siguran, a verovatnoća greške je:

$$\alpha = 1 - CL$$

Ako je interval poverenja 95%, onda je $\alpha = 0.05$. To znači da bi, kada bih postupak pravljenja intervala ponovio mnogo puta na isti način, oko 5% tako napravljenih intervala promašilo pravu μ .

10.2. Margina greške (MoE)

U praksi se najčešće koristi:

$$MoE = t^* \cdot (s / \sqrt{n}) \quad (t^* \text{ za } df = n - 1)$$

11. Interval poverenja (CI)

Interval poverenja (CI) je:

$$CI = \bar{x} \pm MoE$$

12. Konkretna primer (n = 100)

12.1. Podaci iz uzorka

U jednom uzorku od 100 učenika dobijeno je:

$$\bar{x} = 174.02 \text{ cm}, \quad s = 22.62 \text{ cm}, \quad n = 100$$

12.2. Standardna greška (SE)

Standardna greška je:

$$SE = s / \sqrt{n} = 22.62 / \sqrt{100} = 2.262$$

12.3. Kritične vrednosti t^* (df = 99)

Kritične vrednosti su približno:

- 90%: $t^* = 1.660$
- 95%: $t^* = 1.984$
- 99%: $t^* = 2.626$

12.4. Margina greške (MoE)

- $MoE_{90} = t^* \cdot SE = 1.660 \cdot 2.262 = 3.76 \text{ cm}$
- $MoE_{95} = t^* \cdot SE = 1.984 \cdot 2.262 = 4.49 \text{ cm}$
- $MoE_{99} = t^* \cdot SE = 2.626 \cdot 2.262 = 5.94 \text{ cm}$

12.5. Intervali poverenja (CI)

- $CI_{90} = 174.02 \pm 3.76 = (170.26, 177.78) \text{ cm}$
- $CI_{95} = 174.02 \pm 4.49 = (169.53, 178.51) \text{ cm}$
- $CI_{99} = 174.02 \pm 5.94 = (168.08, 179.96) \text{ cm}$

13. Testiranje statističke hipoteze (na istom primeru)

13.1. Tvrdnja (μ_0) i hipoteze

Pretpostavimo da postoji tvrdnja (norma) da je prosečna dužina skoka: $\mu_0 = 175$ cm.

$$H_0: \mu = 175 \quad H_1: \mu \neq 175$$

13.2. Nivo značajnosti (α)

Biramo nivo značajnosti:

$$\alpha = 0.05$$

13.3. t statistika i p-vrednost

Računam t statistiku:

$$t = (\bar{x} - \mu_0) / SE = (174.02 - 175) / 2.262 = -0.433, \quad df = 99$$

Dobijena p-vrednost (dvostrani test) je:

$$p = 0.666$$

13.4. Odluka

Pošto je $p > \alpha$, ne odbacujem H_0 . To znači: na osnovu ovog uzorka ne mogu da tvrdim da se prosečna vrednost razlikuje od 175 cm — razlika koju vidim je kompatibilna sa greškom uzorkovanja.

Ako bih dobio $p \leq \alpha$, tada bih odbacio H_0 : rezultat bi bio dovoljno 'ekstreman' da bih zaključio da prosečna vrednost verovatno nije 175 cm.

14. Greška tipa I i greška tipa II (praksa)

14.1. Greška tipa I (Type I) — verovatnoća = α

Odbacim H_0 , a u stvarnosti je H_0 tačna. Na ovom primeru: zaključim da prosečan skok nije 175 cm, iako je u stvarnosti $\mu = 175$.

14.2. Greška tipa II (Type II) — verovatnoća = β

Ne odbacim H_0 , a u stvarnosti je H_0 netačna. Na ovom primeru: stvarno je $\mu \neq 175$ (recimo $\mu = 178$), ali naš uzorak 'ispadne' blizu 175 i mi kažemo: 'nemamo dokaza protiv 175'.

Ova verovatnoća se označava sa:

$$P(\text{Type II}) = \beta$$

Suprotno tome je snaga testa (power):

$$\text{power} = 1 - \beta$$

14.3. Ilustracija: promena α (strožiji kriterijum)

Ako smanjim α (želim da budem stroži), teže odbacujem H_0 . To smanjuje Type I grešku, ali obično povećava Type II grešku (lakše propustim stvarnu razliku).

Ilustracija: pretpostavimo da je stvarna razlika 3 cm ($\mu = 178$ umesto 175), a $SD \approx 22.62$ cm. Za $n = 100$, dvostrani test daje sledeću snagu (power):

- $\alpha = 0.10 \rightarrow \text{power} \approx 0.37$
- $\alpha = 0.05 \rightarrow \text{power} \approx 0.26$
- $\alpha = 0.01 \rightarrow \text{power} \approx 0.10$

14.4. Ilustracija: povećanje uzorka n (smanjuje β bez menjanja α)

Ako želim da smanjim Type II grešku bez 'žrtvovanja' α , najčistiji način je da povećam n . Veće n smanjuje SE, pa test postaje osetljiviji i snaga raste.

Za isti primer razlike 3 cm, uz $\alpha = 0.05$, snaga je približno:

- $n = 100 \rightarrow \text{power} \approx 0.26$
- $n = 200 \rightarrow \text{power} \approx 0.46$
- $n = 400 \rightarrow \text{power} \approx 0.75$

14.5. Veza sa procenom veličine uzorka

Izbor α , očekivane razlike (koliko cm želimo da uočimo) i procene varijabilnosti (s) direktno određuju standardnu grešku, β i snagu testa. To je osnova za procenu potrebne veličine uzorka (sample size), o čemu će se detaljnije učiti na predavanju kada se bude sistematski obrađivao t-test.

Literatura

Weir, J. P., & Vincent, W. J. (2021). *Statistics in kinesiology* (5th ed.). Human Kinetics.