

Ocenjevanje populacijskih parametrov

- Cilji vzorčenja:
 - Opredelitev obravnavane populacije.
 - Dobiti in analizirati količinsko in stroškovno obvladljiv nabor podatkov.
 - Dobiti reprezentativen vzorec (ali več takih vzorcev).
- **Parametre** (številске karakteristike) populacije ocenimo iz parametrov vzorca (statistično sklepanje). Oceniti želimo:

Lastnosti populacije:

- populacijsko povprečje;
- populacijski delež;
- populacijski standardni odklon;
- normalnost porazdelitve.

Razlike med populacijama:

- v povprečju, deležu, standardnem odklonu, porazdelitvi;
- neodvisnost količin pri populacijah.

Problem vzorčenja

- Študij lastnosti populacije.
- Velika ali namišljena populacija: Nemogoče je ugotoviti, kako je lastnost porazdeljena na populaciji.
- Porazdelitev natančno izmerimo na delu populacije - **vzorcu**.
- Parametre populacije **ocenimo** iz parametrov vzorca.
- **Zahteva:**
Vzorčenje mora ohranjati statistične značilnosti opazovanih parametrov - vzorec mora biti **reprezentativen**.
- Vzorec se ukvarja s postopki, kako zagotoviti reprezentativen vzorec.

Reprezentativen vzorec je podoben populaciji

Na reprezentativnost vzorca vplivajo naslednji dejavniki:

- **Razpršenost pojava** (obravnavane količine) na populaciji: Večje so razlike (velik standardni odklon), slabši so vzorci.
- **Velikost vzorca:** Več je bolje.
- **Način izbora vzorca:**
 - Slučajnostni izbor (s ponavljanjem ali **brez**): žreb, računalniški program, večkratni met kovanca ali kocke, tabela naključnih števil ipd.
 - Sistematični izbor (približki za slučajnostni izbor): izbor po datumu, intervalni izbor, priložnostni izbor.

ENOSTAVNI SLUČAJNOSTNI VZOREC

- Osnova za vse statistične metode.
- Izbran slučajnostno brez ponavljanja.
- **Problem:** Nimamo vedno takega vzorca.

(Ne)odvisnost vzorcev

- Odvisna vzorca: iz iste skupine.
 - Podatki nastopajo v parih.
- Neodvisna vzorca: iz različnih skupin.
 - Podatki niso v parih.

Primer: Primerjava deleža osušenih smrek z deležem okužbe smrek z lubadarjem.

- Odvisen vzorec: Izberemo n smrek in pri vsaki preverimo, ali se suši in ali je okužena z lubadarjem, nato primerjamo deleže.
- Neodvisen vzorec: Izberemo skupini H_1 po n_1 smrek in H_2 po n_2 smrek. Pri H_1 preverimo, ali se sušijo. Pri H_2 preverimo, ali so okužene z lubadarjem. Primerjamo deleže.

Ocenjevanje populacijskih parametrov

Cenilke – namen

- Zanimajo nas populacijski parametri: povprečje, minimalna vrednost, maksimalna vrednost, modus, mediana, kvantili, delež, disperzija, standardni odklon, asimetrija porazdelitve ...
- **Statistika** - vsaka simetrična funkcija vzorčnega vektorja:
 - $U = F(X_1, \dots, X_i, X_{i+1}, \dots, X_n)$
 $= F(X_1, \dots, X_{i+1}, X_i, \dots, X_n)$.
 - Vse, kar lahko izračunamo na podlagi vzorca in je neodvisno od zaporedja statističnih enot v vzorcu.
- **Cenilka** populacijskega parametra q stat. spremenljivke X :
 - Statistika, ki iz vrednosti X na vzorcu oceni vrednost populacijskega parametra q .
 - $U_{(q)} = F(X_1, \dots, X_i, X_{i+1}, \dots, X_n)$.
 - Točkovna ali intervalna ocena populacijskega parametra.

Ocenjevanje populacijskih parametrov

Cenilke – predpostavke

- X – statistična spremenljivka na populaciji G .
- H – vzorec velikosti n .
- x_1, x_2, \dots, x_n – vrednosti X na vzorcu H .
- x_i – dejanska realizacija naključne spremenljivke X .
- (X_1, X_2, \dots, X_n) – vektorska naključna spremenljivka, katere realizacija so vrednosti na vzorcu velikosti n .
- X_i – naključna spremenljivka, porazdeljena kot X .
- X_1, X_2, \dots, X_n – paroma neodvisne naključne spremenljivke.

Ocenjevanje populacijskih parametrov

Zgledi cenilk - srednje in ekstremne vrednosti

- Populacijsko povprečje μ ocenimo z **vzorčnim povprečjem**:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Ali razumemo razliko med formulama?

- \bar{X} je točkovna cenilka za μ .
- Vrednost \bar{x} je realizacija naključne spremenljivke \bar{X} .
- Populacijska mediana, modus: vzorčna mediana, modus.
- Populacijski minimum, maksimum: vzorčni minimum, maksimum.

Ocenjevanje populacijskih parametrov

Zgledi cenilk – populacijska disperzija

- Populacijsko disperzijo σ^2 ocenimo z vzorčno disperzijo.
- Vzorčna disperzija: $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.
- Zakaj delimo z $n - 1$? Na prvi pogled bi bilo prav n !
- Disperzija vzorca: $S_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$.
- S_0^2 in S^2 sta točkovni cenilki za populacijsko disperzijo.

Ocenjevanje populacijskih parametrov

Zgledi cenilk – populacijski delež

- Oceniti želimo delež statističnih enot z določeno lastnostjo.
- Populacijski delež ocenimo z vzorčnim deležem: $\bar{p} = \frac{k}{n}$.
Pri tem je:
 - k – frekvenca enot v vzorcu, ki to lastnost imajo.
 - n – število vseh enot v vzorcu.
- Matematična utemeljitev: \bar{p} kot vzorčno povprečje indikatorskih spremenljivk!
 - $X_i = \begin{cases} 1 & ; \text{ enota ima opazovano lastnost,} \\ 0 & ; \text{ enota nima opazovane lastnosti.} \end{cases}$
 - $\bar{p} = \frac{1}{n} \sum_{i=1}^n X_i$.
 - Ista cenilka kot za povprečje.

Ocenjevanje populacijskih parametrov

Nepriustranskost in učinkovitost cenilk

Nepriustranskost cenilk:

- Pričakovana vrednost cenilke je enaka vrednosti populacijskega parametra, ki ga ocenjuje: $E(C) = q$.
- Vzorčno povprečje je nepriustranska cenilka: dokaz!
- Vzorčna disperzija in disperzija vzorca: katera je nepriustranska?
 - $E(S^2) = \sigma^2$ (S^2 je nepriustranska cenilka).
 - $E(S_0^2) = \frac{n-1}{n}\sigma^2$ (S_0^2 ni nepriustranska cenilka - v povprečju daje prenizke ocene).
- Je vzorčni delež nepriustranski?
Cenilka mora biti **učinkovita**:
 - Pri dani velikosti vzorca ima najmanjšo možno disperzijo.
 - Vzorčno povprečje, disperzija, delež – učinkovite cenilke.
- Nepriustranskost in učinkovitost na grafu porazdelitve cenilke.

Ocenjevanje populacijskih parametrov

Standardna napaka

- **Standardni odklon cenilke** se imenuje standardna napaka cenilke za parameter q .
Oznaka: $SE = SE(C)$.
To napako naredimo pri točkovnem ocenjevanju parametra.
- **Zgled 1:** Standardna napaka pri ocenjevanju populacijskega povprečja spremenljivke X s standardnim odklonom σ :
 $SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$. Dokaz.
 - Če σ ni znan, ga ocenimo iz vzorca. Potem je ocena za $SE(\bar{X})$ enaka $\frac{S}{\sqrt{n}}$.
 - Večji je n , manjša je standardna napaka.
- **Zgled 2:** Standardna napaka pri ocenjevanju populacijskega deleža: $SE(\bar{p}) = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$.

Ocenjevanje populacijskih parametrov