

UZORAK I POPULACIJA

Mirjana Kujundžić Tiljak i Davor Ivanković

Populacija (osnovni skup, univerzum) predstavlja skup svih jedinki (elemenata, članova) s određenim zajedničkim karakteristikama. Može biti konačna ili beskonačna. Jedinke promatranja u populaciji nazivaju se *entiteti*. Razlikujemo ih prema njihovim obilježjima, koje još nazivamo i *atributi*.

Proučavanje cijele populacije najčešće je vrlo skupo i predstavlja izrazito opsežan posao, a često je i potpuno nemoguće, kao kad se radi o hipotetskoj populaciji (npr. pacijenti koji bi mogli u budućnosti biti liječeni na određeni način). Stoga najčešće istražuje *uzorak* pojedinaca koji dobro reprezentiraju populaciju (engl. sample, franc. sondage, njem. die Stichprobe).

Dobivene rezultate promatranja uzorka treba generalizirati na populaciju, pri čemu koristimo *teoriju uzoraka*. Iz svojstva uzorka procjenjujemo svojstva populacije. Pri tome moramo procijeniti i veličinu odstupanja rezultata dobivenih na uzorku od točnih vrijednosti populacije (aritmetička sredina, varijanca, relativna frekvencija, proporcija).

Na temelju *statističkih veličina uzoraka* procjenjujemo *parametre populacije*.

Tablica 1. Simboli statističkih veličina uzorka i populacijskih parametara.

UZORAK statističke veličine	POPULACIJA parametri
\bar{x}	μ
s	σ
p	Π

Procjena parametara populacije na temelju statističkih veličina uzorka ovisi o *reprezentativnosti uzorka* i odabranoj *vjerojatnosti*.

Reprezentativan uzorak dobro opisuje (reprezentira populaciju). Na reprezentativnost uzorka utiču barem: (1) metoda uzorkovanja, tj. odabira uzorka, (2) veličina uzorka i (3) varijabilnost obilježja.

METODA ODABIRA UZORKA

Kada svaki element populacije ima jednaku šansu da bude izabran i svaki uzorak ima jednaku šansu da bude izabran uzorak je *slučajan*, randomiziran (engl. random sample). Najčešće korištene metode za odabir slučajnog uzorka su lutrijska metoda, ili odabir uzorka pomoću

tablice slučajnih brojeva. Danas se za odabir slučajnog uzorka pretežno koriste računala. Kada se odabir jedinki promatranja odvija po nekom sistemu (npr. svaki drugi vlasnik telefona prema telefonskom imeniku) radi se o *sistematskom uzorku*. Pri tome valja imati na umu da naš sistem može koincidirati s nepoznatim sistemom. Kod sistemskog uzorka nije moguće izračunati pogrešku vezanu uz zaključivanje o populaciji na temelju uzorka.

VELIČINA UZORKA

Zaključci izvedeni na osnovu uzorka bit će to kvalitetniji što je uzorak veći. Ovo pravilo naziva se “zakog velikih brojeva”.

Veličina uzorka ovisi o: homogenosti populacije s obzirom na ispitivano obilježje, kao i o učestalosti ispitivanog obilježja u populaciji.

EFEKT VARIJABILNOSTI

Varijabilnost uzorka istraživaču je često nepoznata.

U slučaju kada poznata velika varijabilnost uzorka ugrožava njegovu reprezentativnost uzorak bi trebalo povećati.

DISTRIBUCIJA ARITMETIČKIH SREDINA UZORAKA

Imamo jedan osnovni skup od N jedinki (članova):

$$x_1, x_2, \dots, x_n$$

Iz osnovnog skupa odaberimo niz slučajnih uzoraka od kojih svaki ima n članova:

1. uzorak	$x_{11}, x_{12}, \dots, x_{1n},$
2. uzorak	$x_{21}, x_{22}, \dots, x_{2n},$
3. uzorak	$x_{31}, x_{32}, \dots, x_{3n},$
...	
k. uzorak	$x_{k1}, x_{k2}, \dots, x_{kn},$

Aritmetičke sredine uzoraka su:

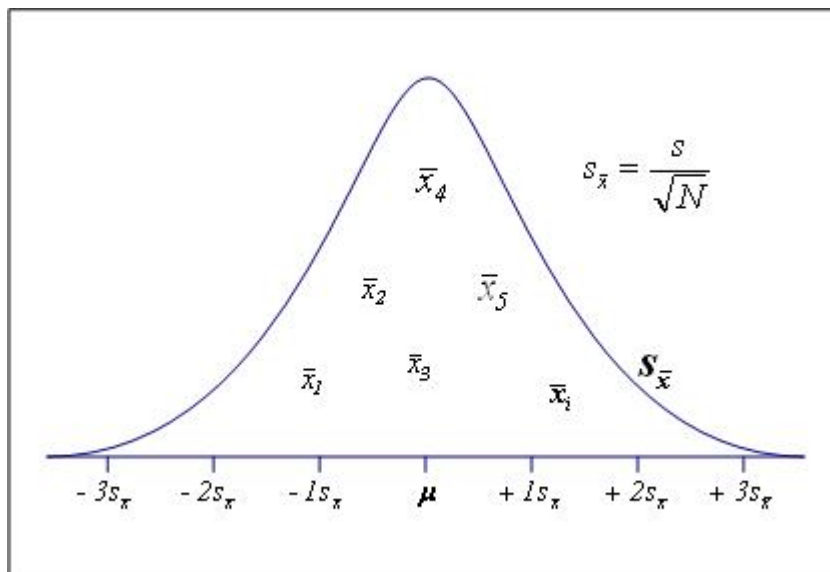
$$\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_k$$

Aritmetička sredina aritmetičkih sredina svih uzoraka je *aritmetička sredina populacije* (μ). Navedeno vrijedi samo ako načinimo sve moguće uzorke bez ponavljanja iz jednog osnovnog skupa.

Distribucija aritmetičkih sredina uzoraka iz jedne populacije bit će normalna ako je distribucija vrijednosti promatranog obilježja u populaciji normalna.

Međutim, po *centralnom graničnom teoremu* distribucija aritmetičkih sredina uzoraka iz jedne populacije bit će normalna i ako distribucija promatranog obilježja u populaciji nije normalna ukoliko su uzorci dovoljno veliki i ako je varijanca populacije (σ^2) konačan broj.

Slika 1. Centralni granični teorem



Aritmetičke sredine uzoraka iz iste populacije se grupiraju oko μ . Oočekivana vrijednost aritmetičke sredine (aritmetička sredina aritmetičkih sredina) jednaka je aritmetičkoj sredini populacije.

$$E(\bar{x}) = \mu$$

STANDARDNA POGREŠKA ARITMETIČKE SREDINE

Standardna devijacija distribucije aritmetičkih sredina uzoraka naziva se *standardna pogreška aritmetičke sredine* (engl. Standard Error of the Mean, SEM).

Računa se:

$$s_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$$

Mjera je odstupanja aritmetičkih sredina uzoraka iz jedne populacije od aritmetičke sredine populacije i kao takva predstavlja pogrešku kojoj se izlažemo zaključujući o populaciji na temelju uzorka.

Standardnu devijaciju populacije, σ u pravilu ne poznajemo jer imamo samo uzorak. Ako je uzorak slučajan i dovoljno velik može se pretpostaviti da je standardna devijacija uzorka (s) dobra procjena standardne devijacije populacije (σ), pa standardnu pogrešku aritmetičke sredine računamo:

$$s_{\bar{x}} = \frac{s}{\sqrt{N}}$$

Standardna devijacija opisuje varijabilnost podataka, a standardna pogreška aritmetičke sredine opisuje preciznost procjene aritmetičke sredine populacije na temelju aritmetičke sredine uzorka.

Velika standardna pogreška ukazuje na *nepreciznu* procjenu, dok mala standardna pogreška ukazuje na *preciznu* procjenu populacijskih parametara na temelju uzorka.

Standardna pogreška je to manja što je uzorak veći i što je varijabilnost podataka manja.

DISTRIBUCIJA PROPORCIJA UZORAKA

Proporcija jedinki u populaciji koje posjeduju određenu karakteristiku također se procijenjuje na temelju uzorka. Pri tome:

n = veličina uzorka

p = procjena proporcije populacije (π)

r = broj jedinki u uzorku koje posjeduju određenu karakteristiku

$$p = r/n$$

Distribucija proporcija uzoraka slijedi normalnu distribuciju sa srednjom vrijednosti π .

Standardna pogreška proporcije je zapravo standardna devijacija distribucije proporcija uzoraka. Računa se:

$$s_p = \sqrt{\frac{pq}{n}}$$

Mala standardna pogreška proporcije ukazuje na preciznu procjenu.

INTERVAL POUZDANOSTI

Na uzorku izračunate statističke veličine pojedinačne su procjene parametara populacije (“point estimate”).

Na temelju statističke veličine uzorka, koristeći standardnu pogrešku, možemo, uz određenu vjerojatnost procijeniti interval u kojem se nalazi parametar populacije (“interval estimate”). Procijenjeni interval u kojem se nalazi parametar populacije (aritmetička sredina ili proporcija) naziva se *interval pouzdanosti* (engl. confidence interval, CI).

Za računanje intervala pouzdanosti koristimo teorijske distribucije vjerojatnosti. Interval pouzdanosti proširuje procjenu parametra populacije na obje strane za nekoliko standardnih pogreški. *Granice pouzdanosti* (engl. confidence limits) definiraju interval, navode se između zagrada, odijeljene zarezom.

INTERVAL POUZDANOSTI ZA ARITMETIČKU SREDINU

Računa se

$$\bar{x} - z \cdot s_{\bar{x}} \leq \mu \leq \bar{x} + z \cdot s_{\bar{x}}$$

Pri čemu:

z = standardizirana vrijednost normalne raspodjele

$$z_i = \frac{(x_i - \bar{x})}{s}$$

Interval pouzdanosti uz 95% vjerojatnost iznosi:

$$\bar{x} - 1,96 \cdot s_{\bar{x}} \leq \mu \leq \bar{x} + 1,96 \cdot s_{\bar{x}}$$

Interval pouzdanosti uz 99% vjerojatnost iznosi

$$\bar{x} - 2,58 \cdot s_{\bar{x}} \leq \mu \leq \bar{x} + 2,58 \cdot s_{\bar{x}}$$

INTERVAL POUZDANOSTI ZA PROPORCIJU

Računa se:

$$p - z \cdot s_p \leq \pi \leq p + z \cdot s_p$$

Pri čemu:

z = standardizirana vrijednost normalne raspodjele

$$z_i = \frac{(x_i - \bar{x})}{s}$$

Interval pouzdanosti uz 95% vjerojatnost iznosi:

$$p - 1,96 \cdot s_p \leq \pi \leq p + 1,96 \cdot s_p$$

Interval pouzdanosti uz 99% vjerojatnost iznosi

$$p - 2,58 \cdot s_p \leq \pi \leq p + 2,58 \cdot s_p$$

PRIMJERI:

1. Kolika je prosječna dob menarhe u Zagrebu uz vjerojatnost 95%?

$$n = 2529$$

$$\bar{x} = 13,16 \text{ godina} \quad s = 1,18 \text{ godina}$$

$$s_{\bar{x}} = \frac{1,18}{\sqrt{2529}} = 0,023 \text{ godina}$$

$$\text{Interval pouzdanosti: } 13,115 \leq \mu \leq 13,205$$

2: Kolika je proporcija alergičnih reakcija u populaciji cijepljenih uz vjerojatnost 95%

$$n = 1000$$

$$p = 0,2 \quad q = 0,8$$

$$s_p = \sqrt{\frac{0,2 \cdot 0,8}{1000}} = 0,0126$$

$$\text{Interval pouzdanosti: } 0,175 \leq \pi \leq 0,225$$

Literatura:

1. *Ivanković D, i sur. Osnove statističke analize za medicinare. Zagreb: Medicinski fakultet Sveučilišta u Zagrebu, 1989.*
2. *Petrie A, Sabin C. Medical Statistics at a Glance (2nd Ed). Oxford: Blackwell Science Ltd, 2005.*
3. *Glantz. SA. Primer of Biostatistics (4th Ed). New York: McGraww-Hill: 1997.*
4. *Altman DG. Practical Statistics for Medical Research. London. Chapman & Hall, 1991.*
5. *Bland M. An Introduction to Medical Statistics (3rd Ed). Oxford: Oxford University Press, 2005.*
6. *Armitage P, Berry P. Statistical Methods in Medical Research. Oxford: Blackwell Science Ltd, 1994.*