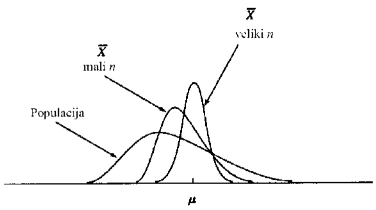
Središnji granični teorem (CLT)

Neka je $X_1, X_2,\ldots,X_n$ slučajni uzorak <u>bilo koje raspodjele</u> s očekivanjem μ i standardnom devijacijom σ. Ako je n dovoljno velik, $\overline{X}$ ima približno **<u>normalnu</u>** raspodjelu s očekivanjem

$\mu_{\overline{X}} = \mu$ i standardnom devijacijom $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$.

T_0 ima također približno normalnu raspodjelu s očekivanjem $\mu_{T_0} = n\mu$ i standardnom devijacijom $\sigma_{T_0} = \sqrt{n}\cdot\sigma$.



Procjena parametara populacije na osnovi uzorka

- Populacija koja nas zanima ima karakteristične parametre μ, σ, σ^2, α_3, itd.
- Nismo u mogućnosti odrediti ih na cijelom skupu.
- Uzimamo uzorak
- Na osnovi uzorka želimo procijeniti parametre osnovnog skupa (populacije)

Neka je θ neki takav parametar populacije koji želimo procijeniti na osnovu slučajnog uzorka $\{X_1, X_2, \ldots X_n\}$!

Definiramo <u>slučajnu varijablu</u> $\hat{\theta}$ kao funkciju slučajnih varijabli $X_1, X_2, \ldots X_n$.
$\hat{\theta}$ nazivamo **procjenitelj** parametra θ.

Za varijablu $\hat{\theta}$ kažemo da je **nepristrani procjenitelj** parametra θ ako je njezino očekivanje

$$E(\hat{\theta}) = \theta$$

Propozicija:
Ako je $\{X_1, X_2, \ldots X_n\}$ slučajni uzorak neke populacije, onda je slučajna varijabla $\overline{X}$ nepristrani procjenitelj očekivanja te populacije (μ).

Propozicija:
Ako je $\{X_1, X_2, \ldots X_n\}$ slučajni uzorak neke populacije, onda je slučajna varijabla

$$\hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}$$

nepristrani procjenitelj varijance osnovnog skupa σ^2.

Mogli smo za procjenitelja odabrati $\hat{\sigma}_u^2 = \dfrac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$

Ali ! $E(\hat{\sigma}_u^2) = \sigma^2 - \dfrac{\sigma^2}{n}$

Interval pouzdanosti prosjeka populacije

Najčešće želimo procijeniti μ populacije
(npr. pravu vrijednost mjerene veličine)

Uzimamo uzorak od n elemenata (velik n) i izračunamo njegov prosjek. Tako smo nepristrano procijenili μ.
Zanima nas koliko pouzdano, tj., u kojem intervalu se pouzdano nalazi prosjek populacije μ.

Neka je varijanca populacije σ^2.

Prema središnjem graničnom teoremu (CLT), $\overline{X}$ je normalno raspodijeljen:

$$\overline{X} \sim N(\mu, \sigma^2/n)$$

Kad odaberemo uzorak, slučajna varijabla $\overline{X}$ poprimit će vrijednost $\overline{x}$.

Za normalnu raspodjelu vrijedi:

$$P\left(-1 \le \frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \le 1\right) = 68\% \quad P\left(-1{,}96 \le \frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \le 1{,}96\right) = 95\% \quad P\left(-2{,}575 \le \frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \le 2{,}575\right) = 99\%$$

To možemo pisati:
Prava vrijednost prosjeka populacije nalazi se u intervalu

$$\mu \in \overline{x} \pm \frac{\sigma}{\sqrt{n}}$$ s pouzdanošću 68%.

```
regresija = Regress[podaciLog, {1, x}, x, RegressionReport → {BestFit, ParameterCITable}]

{BestFit → 0.688514 + 1.98287 x,

                      | Estimate  SE          CI
ParameterCITable →  1 | 0.688514  0.00912217  {0.659483, 0.717545}}
                    x | 1.98287   0.0242844   {1.90559, 2.06015}
```

Obično ne znamo varijancu osnovnog skupa pa ju moramo procijeniti.

Nepristrani procjenjitelj varijance je $$\hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

Za procjenjitelja standardne devijacije najčešće uzimamo slučajnu varijablu
$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$
koja je vrlo bliska nepristranom procjenjitelju standardne devijacije.

Interval 68% pouzdanosti prosjeka je tada $$\bar{x} \pm \frac{s}{\sqrt{n}} = \bar{x} \pm \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n(n-1)}}$$

U fizici se dogovorno svi rezultati mjerenja pišu s navođenjem iznosa 68-postotne nepouzdanosti:
$$\bar{x} \pm M$$
gdje je $M = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n(n-1)}}$ **nepouzdanost** izračunata iz n nezavisnih mjerenja.

STANDARDNA POGREŠKA

Također se upotrebljava veličina
$$m = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = s$$
koja označava **preciznost** svakog pojedinog mjerenja.

i	x_i (ms)	$(x_i - \bar{x})^2$ (ms^2)
1	584	14.5785
2	588	0.03306
3	591	10.12398
4	587	0.66942
5	590	4.76034
6	585	7.94214
7	585	7.94214
8	592	17.48762
9	588	0.03306
10	589	1.3967
11	587	0.66942

$$\sum x_i = 6466\,\text{ms}$$
$$\bar{x} = \frac{1}{11}\sum x_i = 587.8182\,\text{ms}$$
$$\sum (x_i - \bar{x})^2 = 65.6364\,\text{ms}^2$$
$$m = s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{11-1}} = 2.56196\,\text{ms}$$
$$M = \frac{m}{\sqrt{11}} = 0.7725\,\text{ms}$$

Pisanje rezultata mjerenja
$$x = (587.8 \pm 0.8)\,\text{ms}$$

relativna pogreška: $R = \frac{M}{\bar{x}} = 0.14\%$

Propagacija pogreške

Posredno mjerene veličine.

Npr.:
- Mjerimo vanjski promjer cijevi D i unutarnji promjer d, a zanima nas debljina stijenke $x=D/2-d/2$.
- Mjerimo promjer kugle d, a zanima nas volumen $V=4\pi(d/2)^3/3$.
- Mjerimo period matematičkog njihala T i duljinu niti l, a zanima nas ubrzanje sile teže $g=4\pi^2 l/T^2$.

Pretpostavke:
- Pogreške su slučajne (normalno raspodijeljene)
- i malene ($M_i << \mu_i$).

Propagacija pogreške

I. Posredna veličina je linearna kombinacija izravno mjerenih veličina
$$h(X,Y) = aX + bY$$
$$\bar{h}(X,Y) = a\bar{x} + b\bar{y} \qquad M_h = \sqrt{a^2 M_X^2 + b^2 M_Y^2}$$

II. Posredna veličina je nelinearna funkcija jedne mjerene veličine $h(X)$
$$\bar{h}(X) \approx h(\bar{x}) \qquad M_h = \left|\frac{dh}{dx}\right|_{x=\mu} \cdot M_X$$

IIa. Posredna veličina je potencija jedne mjerene veličine $h(X) = X^\alpha$
$$\Rightarrow \quad M_h = \alpha\,\bar{x}^{\alpha-1} \cdot M_X$$

relativna pogreška: $$\frac{M_h}{h} = \alpha \cdot \frac{M_X}{\bar{x}}$$

III. Najopćenitiji slučaj: Posredna veličina je nelinearna funkcija dviju ili više mjerenih veličina $h(X,Y)$

Pokazuje se da malena pogreška propagira na sljedeći način:
$$\boxed{M_h^2 = \left(\frac{\partial h}{\partial X}\right)^2_{\bar{x},\bar{y}} \cdot M_X^2 + \left(\frac{\partial h}{\partial Y}\right)^2_{\bar{x},\bar{y}} \cdot M_Y^2}$$

Općenito za n izravno mjerenih veličina imamo: $h(X_1, X_2, ..., X_n)$
$$\bar{h}(X_1, X_2, ..., X_n) = h(\bar{x}_1, \bar{x}_2, ..., \bar{x}_n)$$
$$M_h = \sqrt{\sum_i \left(\frac{\partial h}{\partial X_i}\right)^2_{\bar{x}_1,\bar{x}_2,...,\bar{x}_n} M_{X_i}^2}$$

Primjer iz praktikuma: Modul elastičnosti

Mjeri se ovisnost savijenosti šipke o sili teže utega, zatim debljina i širina šipke te udaljenost potpornja.

Iz dobivenih podataka određuje se modul elastičnosti čelika.

Teorijska formula za savijenost šipke:

$$\lambda = \frac{1}{4E}\frac{L^3}{ab^3}F$$

Izmjerene su sljedeće veličine:

$$A = \frac{\lambda}{F} = (0{,}76 \pm 0{,}01)\,\frac{\text{mm}}{\text{N}} = (7{,}6 \pm 0{,}1)\cdot 10^{-4}\,\frac{\text{m}}{\text{N}} = \overline{A} \pm M_A$$

$$a = (10{,}26 \pm 0{,}05)\,\text{mm} = (1{,}026 \pm 0{,}005)\cdot 10^{-2}\,\text{m} = \overline{a} \pm M_a$$

$$b = (1{,}53 \pm 0{,}03)\,\text{mm} = (1{,}53 \pm 0{,}03)\cdot 10^{-3}\,\text{m} = \overline{b} \pm M_b$$

$$L = (29{,}0 \pm 0{,}1)\,\text{cm} = (2{,}90 \pm 0{,}01)\cdot 10^{-1}\,\text{m} = \overline{L} \pm M_L$$

Najvjerojatnija vrijednost za modul elastičnosti čelika je:

$$\overline{E} = \frac{1}{4}\frac{\overline{L}^3}{\overline{a}\overline{b}^3}\frac{1}{\overline{A}} = 2{,}183226 \cdot 10^{11}\,\frac{\text{N}}{\text{m}^2}$$

Varijanca modula elastičnosti je:

$$M_E^2 = \left(\frac{\partial E}{\partial A}\right)^2 M_A^2 + \left(\frac{\partial E}{\partial a}\right)^2 M_a^2 + \left(\frac{\partial E}{\partial b}\right)^2 M_b^2 + \left(\frac{\partial E}{\partial L}\right)^2 M_L^2 =$$

$$= \left(\frac{\overline{L}^3}{4\overline{a}\overline{b}^3\overline{A}^2}\right)^2 M_A^2 + \left(\frac{\overline{L}^3}{4\overline{a}^2\overline{b}^3\overline{A}}\right)^2 M_a^2 + \left(\frac{3\overline{L}^3}{4\overline{a}\overline{b}^4\overline{A}}\right)^2 M_b^2 + \left(\frac{3\overline{L}^2}{4\overline{a}\overline{b}^3\overline{A}}\right)^2 M_L^2 =$$

$$= \overline{E}^2\left[\left(\frac{M_A}{\overline{A}}\right)^2 + \left(\frac{M_a}{\overline{a}}\right)^2 + \left(3\frac{M_b}{\overline{b}}\right)^2 + \left(3\frac{M_L}{\overline{L}}\right)^2\right] =$$

$$= \overline{E}^2\left[1{,}73\cdot 10^{-4} + 2{,}37\cdot 10^{-5} + 3{,}46\cdot 10^{-3} + 1{,}07\cdot 10^{-4}\right]$$

Standardna pogreška je:

$$M_E = \overline{E}\sqrt{3{,}76\cdot 10^{-3}} = 0{,}13\cdot 10^{11}\,\frac{\text{N}}{\text{m}^2}$$

Rezultat za modul elastičnosti pišemo:

$$\boxed{E = (2{,}2 \pm 0{,}1)\cdot 10^{11}\,\frac{\text{N}}{\text{m}^2}}$$

Mjerenja različitih statističkih težina

Fizikalnu veličinu mjerimo u više navrata.
Dobivamo različite rezultate.
Zanima nas koja je prava, ili barem najvjerojatnija, vrijednost te fizikalne veličine.
Mjerenja se obično izvode na različite načine i s različitim pouzdanostima pa ih ne smijemo tretirati ravnopravno.
Kažemo da mjerenja imaju *različite statističke težine*.

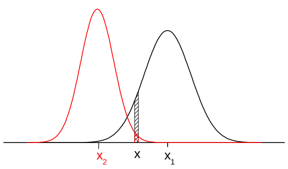
Jedna takva serija mjerenja daje rezultat u obliku

$$x_1 = \overline{x}_1 \pm M_1$$

gdje M_1 predstavlja interval 68% pouzdanosti za nalaženje prave vrijednosti x_p.

Učinimo li još jednu seriju mjerenja, njezin rezultat imat će oblik:

$$x_2 = \overline{x}_2 \pm M_2$$



Opća srednja vrijednost

Ako istu veličinu x mjerimo u raznim prigodama dobivamo različite rezultate s različitim pouzdanostima.

$$x_1 = \overline{x}_1 \pm M_1$$
$$x_2 = \overline{x}_2 \pm M_2$$
$$\vdots$$
$$x_k = \overline{x}_k \pm M_k$$

Funkcija gustoće vjerojatnosti za x_p:

$$f\left(x_p; \overline{x}_1, \overline{x}_2, \ldots, \overline{x}_k\right) = \frac{1}{M_1 M_2 \cdots M_k\left(\sqrt{2\pi}\right)^k}\, e^{-\sum_{i=1}^{k}\frac{\left(\overline{x}_i - x_p\right)^2}{2M_i^2}}$$

Tražimo onaj x_p za koji je ova f maksimalna.

Najvjerojatnija vrijednost je:

$$\boxed{x_p^* = \frac{\sum_{i=1}^{k} \frac{\bar{x}_i}{M_i^2}}{\sum_{j=1}^{k} \frac{1}{M_j^2}}}$$

Definiramo **statističku težinu:**

$$w_i = \frac{\frac{1}{M_i^2}}{\sum_{j=1}^{k} \frac{1}{M_j^2}}$$

$$\boxed{x_p^* = \sum_{i=1}^{k} w_i \bar{x}_i}$$

<u>Nepouzdanost opće srednje vrijednosti</u>

Procjenjitelj prave vrijednosti mjerene veličine jest linearna kombinacija prosjeka iz *k* mjerenja:

$$X_p^* = \sum_{i=1}^{k} w_i \bar{X}_i$$

Varijancu linearne kombinacije znademo izračunati:

$$M^2 = V\left(X_p^*\right) = \sum_{i=1}^{k} w_i^2 V\left(\bar{X}_i\right) = \sum_{i=1}^{k} w_i^2 M_i^2$$

<u>nepouzdanost opće srednje vrijednosti:</u>

$$\boxed{M^2 = \frac{1}{\sum_{i=1}^{k} \frac{1}{M_i^2}}}$$

To je bilo za konzistentna mjerenja!

Ako mjerenja nisu konzistentna, uzimamo srednje vrijednosti svake serije i tretiramo ih kao pojedinačna mjerenja.