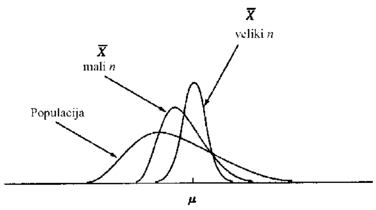
Središnji granični teorem (CLT)

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

Definiramo slučajnu varijablu $\hat{\theta}$ kao funkciju slučajnih varijabli $X_1, X_2, ... X_n$.
$\hat{\theta}$ 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 $\{X_1, X_2, ... X_n\}$ slučajni uzorak neke populacije, onda je slučajna varijabla $\overline{X}$ nepristrani procjenitelj očekivanja te populacije (μ).

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

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

nepristrani procjenitelj varijance osnovnog skupa σ^2.

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

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:

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

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

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

Obično <u>ne znamo varijancu</u> osnovnog skupa pa ju moramo procijeniti.

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

Za procjenjitelja standardne devijacije najčešće uzimamo slučajnu varijablu

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

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

```
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}
```

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)	$(x_i-\bar{x})^2$ (ms²)
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

$\sum x_i = 6466\,\mathrm{ms}$

$\bar{x} = \frac{1}{11}\sum x_i = 587.8182\,\mathrm{ms}$

$\sum (x_i - \bar{x})^2 = 65.6364\,\mathrm{ms}^2$

$m = s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{11-1}} = 2.56196\,\mathrm{ms}$

$M = \frac{m}{\sqrt{11}} = 0.7725\,\mathrm{ms}$

Pisanje rezultata mjerenja

$$x = (587.8 \pm 0.8)\,\mathrm{ms}$$

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 ($M_i << \mu_i$).