STATISTICĂ - LABORATOR 5 (partea I)

METODA INTERVALELOR DE ÎNCREDERE

Fie caracteristica X, a cărei lege de probabilitate depinde de parametrul necunoscut $\theta \in \mathbb{R}$. Se consideră o selecţie repetată de volum n. Fie $X_1, \ldots, X_n$ variabilele de selecţie.
Se definesc următoarele funcţii de selecţie (statistici):

- media de selecţie ($\boxed{\text{mean}}$):

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{1}$$

- dispersia de selecţie ($\boxed{\text{var}}$):

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

- abaterea standard de selecţie ($\boxed{\text{std}}$):

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2} \tag{3}$$

Fie numărul $\alpha \in (0,1)$ numit **probabilitate de risc**.
$1 - \alpha$ se numeşte **probabilitate de încredere**.

Definiţie. Se numeşte **interval de încredere** pentru parametrul θ, corespunzător probabilităţii de încredere $1 - \alpha$, intervalul aleator

$$\left(\widehat{\theta}_1, \widehat{\theta}_2\right) = \left(\widehat{\theta}_1\left(X_1, \ldots, X_n\right), \widehat{\theta}_2\left(X_1, \ldots, X_n\right)\right),$$

cu proprietatea că

$$P\left(\theta \in \left(\widehat{\theta}_1, \widehat{\theta}_2\right)\right) = 1 - \alpha. \tag{4}$$

Interval de încredere pentru medie când dispersia este cunoscută

Fie caracteristica $X \sim N(m, \sigma)$, cu $m = E(X)$ necunoscut şi $\sigma = \sqrt{Var(X)}$ **cunoscut**.
Statistica care se utilizează este :

$$Z = \frac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}},$$

care urmează legea normală $N(0, 1)$.
Intervalul de încredere pentru media teoretică m este:

$$(\widehat{m}_1, \widehat{m}_2) = \left(\overline{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}},\ \overline{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) \tag{5}$$

S-a notat cu $z_{1-\frac{\alpha}{2}}$ cuantila de ordin $1 - \frac{\alpha}{2}$ pentru legea $N(0, 1)$.
Cuantila $z_{1-\frac{\alpha}{2}}$ verifică relaţia:

$$\Phi\left(z_{1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2} \qquad \text{sau, echivalent,} \qquad z_{1-\frac{\alpha}{2}} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \tag{6}$$

unde Φ este funcţia de repartiţie a legii normale $N(0, 1)$.

Pentru calculul cuantilei $z_{1-\frac{\alpha}{2}}$ se foloseşte funcţia Matlab norminv.

Observaţie: Pentru $n \geq 30$, din teorema limită centrală avem că rezultatele obţinute pot fi aplicate pentru o caracteristică X ce urmează o lege de probabilitate oarecare.

Interval de încredere pentru medie când dispersia este necunoscută

Fie caracteristica $X \sim N(m, \sigma)$, cu $m = E(X)$ necunoscut şi $\sigma = \sqrt{Var(X)}$ **necunoscut**.
Deoarece abaterea standard σ e necunoscută, se utilizează abaterea standard de selecţie:

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

Se construieşte statistica:

$$T = \frac{\overline{X} - m}{\frac{s}{\sqrt{n}}},$$

care urmează legea $T(n-1)$.
Intervalul de încredere pentru media teoretică m este:

$$(\widehat{m}_1, \widehat{m}_2) = \left(\overline{X} - t_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}},\ \overline{X} + t_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}\right) \tag{7}$$

S-a notat cu $t_{1-\frac{\alpha}{2}}$ cuantila de ordin $1-\frac{\alpha}{2}$ a legii $T(n-1)$.

Cuantila $t_{1-\frac{\alpha}{2}}$ verifică relaţia:

$$F_{n-1}(t_{1-\frac{\alpha}{2}}) = 1-\frac{\alpha}{2}, \qquad \text{adică} \qquad t_{1-\frac{\alpha}{2}} = F_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right), \tag{8}$$

unde F_{n-1} este funcţia de repartiţie a legii $T(n-1)$.

Pentru calculul cuantilei $t_{1-\frac{\alpha}{2}}$ se foloseşte funcţia Matlab tinv.

Observaţie: Pentru $n \geq 30$, rezultatele obţinute pot fi aplicate pentru o caracteristică X ce urmează o lege de probabilitate oarecare.

Interval de încredere pentru dispersia legii normale

Fie caracteristica $X \sim N(m,\sigma)$, cu $\sigma = \sqrt{Var(X)} > 0$ necunoscut.

Se utilizează statistica:

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

care urmează legea $\chi^2(n-1)$.

Intervalul de încredere pentru dispersia teoretică $\sigma^2 = Var(X)$ este:

$$(d_1, d_2) = \left(\frac{(n-1)s^2}{h_{1-\frac{\alpha}{2}}}\ ,\ \frac{(n-1)s^2}{h_{\frac{\alpha}{2}}}\right) \tag{9}$$

S-au notat cu $h_{\frac{\alpha}{2}}$ şi $h_{1-\frac{\alpha}{2}}$ cuantilele de ordin $\frac{\alpha}{2}$ respectiv $1-\frac{\alpha}{2}$ pentru legea $\chi^2(n-1)$. Cuantilele verifică relaţiile:

$$F_{n-1}\left(h_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2} \qquad \text{şi} \qquad F_{n-1}\left(h_{1-\frac{\alpha}{2}}\right) = 1-\frac{\alpha}{2} \tag{10}$$

adică

$$h_{\frac{\alpha}{2}} = F_{n-1}^{-1}\left(\frac{\alpha}{2}\right) \qquad \text{şi} \qquad h_{1-\frac{\alpha}{2}} = F_{n-1}^{-1}\left(1-\frac{\alpha}{2}\right) \tag{11}$$

unde F_{n-1} este funcţia de repartiţie a legii $\chi^2(n-1)$.

Pentru calculul cuantilelor $h_{\frac{\alpha}{2}}$ şi $h_{1-\frac{\alpha}{2}}$ se foloseşte funcţia Matlab chi2inv.

Intervalul de încredere pentru abaterea standard $\sigma = \sqrt{Var(X)}$ este:

$$\left(\sqrt{d_1}, \sqrt{d_2}\right) = \left(\sqrt{\frac{(n-1)s^2}{h_{1-\frac{\alpha}{2}}}}\ ,\ \sqrt{\frac{(n-1)s^2}{h_{\frac{\alpha}{2}}}}\right) \tag{12}$$