STATISTICĂ - LABORATOR 5 (partea II)

Interval de încredere pentru raportul dispersiilor a două legi normale

Se studiază două caracteristici independente X_1 şi X_2, relative la două populaţii.
Se presupune că X_1 şi X_2 urmează fiecare legea normală:

$$X_1 \sim N(m_1, \sigma_1) \text{ şi } X_2 \sim N(m_2, \sigma_2).$$

Pentru fiecare caracteristică, se consideră câte o selecţie repetată de volum n_1, respectiv n_2. Fie

- $X_{11}, X_{12}, \ldots, X_{1n_1}$ variabilele de selecţie pentru selecţia 1;
- $X_{21}, X_{22}, \ldots, X_{2n_2}$ variabilele de selecţie pentru selecţia 2.

Se definesc următoarele funcţii de selecţie (statistici):

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

$$\overline{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i}, \qquad \overline{X}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} X_{2j} \tag{1}$$

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

$$s_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_{1i}-\overline{X}_1)^2, \qquad s_2^2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}(X_{2j}-\overline{X}_2)^2 \tag{2}$$

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

$$s_1 = \sqrt{\frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_{1i}-\overline{X}_1)^2}, \qquad s_2 = \sqrt{\frac{1}{n_2-1}\sum_{j=1}^{n_2}(X_{2j}-\overline{X}_2)^2} \tag{3}$$

Fie probabilitatea de încredere $1-\alpha$ dată, $\alpha \in (0,1)$.

Se consideră statistica

$$F = \frac{\frac{s_1^2}{\sigma_1^2}}{\frac{s_2^2}{\sigma_2^2}}, \tag{4}$$

care urmează legea $F(n_1-1, n_2-1)$.

Intervalul de încredere pentru raportul dispersiilor teoretice $\dfrac{\sigma_1^2}{\sigma_2^2}$ este:

$$\left(\frac{1}{f_{1-\frac{\alpha}{2}}} \cdot \frac{s_1^2}{s_2^2}\,, \quad \frac{1}{f_{\frac{\alpha}{2}}} \cdot \frac{s_1^2}{s_2^2} \right), \tag{5}$$

unde $f_{\frac{\alpha}{2}}$ şi $f_{1-\frac{\alpha}{2}}$ sunt cuantilele de ordin $\frac{\alpha}{2}$, respectiv $1-\frac{\alpha}{2}$ pentru legea $F(n_1-1, n_2-1)$. ($\boxed{\text{finv}}$)

Interval de încredere pentru diferenţa mediilor

Fie caracteristicile independente X_1 şi X_2, relative la două populaţii. Presupunem că:

$$X_1 \sim N\left(m_1, \sigma_1\right) \text{ şi } X_2 \sim N\left(m_2, \sigma_2\right).$$

Se consideră câte o selecţie repetată de volum n_1, respectiv n_2.

Folosind probabilitatea de încredere $1-\alpha$ dată, $\alpha \in (0,1)$, dăm un interval de încredere pentru diferenţa mediilor teoretice $m_1 - m_2$.

A) dispersiile σ_1^2 şi σ_2^2 sunt cunoscute

Se consideră statistica

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - (m_1 - m_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \tag{6}$$

care urmează legea normală $N\left(0,1\right)$.

Intervalul de încredere pentru diferenţa mediilor teoretice $m_1 - m_2$ este:

$$\left(\widehat{m}_1, \widehat{m}_2\right) = \left(\left(\overline{X}_1 - \overline{X}_2\right) - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \;\; \left(\overline{X}_1 - \overline{X}_2\right) + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right) \tag{7}$$

unde $z_{1-\frac{\alpha}{2}}$ este cuantila de ordin $1-\frac{\alpha}{2}$ pentru legea $N\left(0,1\right)$. ($\boxed{\text{norminv}}$)

B) dispersiile σ_1^2, σ_2^2 sunt necunoscute şi egale $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Se consideră statistica

$$T = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - (m_1 - m_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim T(n_1 + n_2 - 2) \tag{8}$$

unde s_p este abaterea standard de selecţie combinată şi se calculează din formula:

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}} \tag{9}$$

Intervalul de încredere pentru diferenţa mediilor teoretice $m_1 - m_2$ este:

$$(\widehat{m}_1, \widehat{m}_2) = \left(\left(\overline{X}_1 - \overline{X}_2\right) - t_{1-\frac{\alpha}{2}} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},\ \ \left(\overline{X}_1 - \overline{X}_2\right) + t_{1-\frac{\alpha}{2}} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right) \tag{10}$$

unde $t_{1-\frac{\alpha}{2}}$ este cuantila de ordin $1 - \frac{\alpha}{2}$ pentru legea $T(n_1 + n_2 - 2)$. ($\boxed{\text{tinv}}$)

<u>**C) dispersiile σ_1^2 şi σ_2^2 sunt necunoscute şi diferite**</u>

Se consideră statistica

$$T = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - (m_1 - m_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \tag{11}$$

care urmează legea $T(n)$, unde <u>n se calculează din formula:</u>

$$\frac{1}{n} = \frac{c^2}{n_1 - 1} + \frac{(1-c)^2}{n_2 - 1}, \quad \text{unde} \quad c = \frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}. \tag{12}$$

Intervalul de încredere pentru diferenţa mediilor teoretice $m_1 - m_2$ este:

$$(\widehat{m}_1, \widehat{m}_2) = \left(\left(\overline{X}_1 - \overline{X}_2\right) - t_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},\ \ \left(\overline{X}_1 - \overline{X}_2\right) + t_{1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right) \tag{13}$$

unde $t_{1-\frac{\alpha}{2}}$ este cuantila de ordin $1 - \frac{\alpha}{2}$ pentru legea $T(n)$. ($\boxed{\text{tinv}}$)

Observaţie: Rezultatele obţinute la cazurile A, B, C pot fi aplicate şi în cazul a două caracteristici independente X_1 şi X_2 care nu urmează legea normală, dacă volumul selecţiilor este mare, adică $n_1, n_2 \geq 20$.