

Testy dla dwóch zbiorowości

Wojciech Kotłowski

Statystyka i analiza danych 2019/2020

28.04.2020

Test niesparowany, duża próba ($n_1, n_2 \geq 30$)

- **Założenia:** $X_{1,1}, \dots, X_{1,n_1} \sim N(\mu_1, \sigma_1^2)$,
 $X_{2,1}, \dots, X_{2,n_2} \sim N(\mu_2, \sigma_2^2)$.

- **Układ hipotez:**

$$H_0 : \quad \mu_1 = \mu_2 \quad (\mu_1 \geq \mu_2) \quad (\mu_1 \leq \mu_2)$$

$$H_1 : \quad \mu_1 \neq \mu_2 \quad \mu_1 < \mu_2 \quad \mu_1 > \mu_2$$

- **Statystyka testowa:** ustandaryzowana różnica $\bar{X}_1 - \bar{X}_2$.

$$E[\bar{X}_1 - \bar{X}_2] = \mu_1 - \mu_2 = 0, \quad (\text{gdy } H_0 \text{ prawdziwe})$$

$$D^2[\bar{X}_1 - \bar{X}_2] = D^2[\bar{X}_1] + D^2[\bar{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Test niesparowany, duża próba ($n_1, n_2 \geq 30$)

- **Założenia:** $X_{1,1}, \dots, X_{1,n_1} \sim N(\mu_1, \sigma_1^2)$,
 $X_{2,1}, \dots, X_{2,n_2} \sim N(\mu_2, \sigma_2^2)$.

- **Układ hipotez:**

$$H_0: \quad \mu_1 = \mu_2 \quad (\mu_1 \geq \mu_2) \quad (\mu_1 \leq \mu_2)$$

$$H_1: \quad \mu_1 \neq \mu_2 \quad \mu_1 < \mu_2 \quad \mu_1 > \mu_2$$

- **Statystyka testowa:** ustandaryzowana różnica $\bar{X}_1 - \bar{X}_2$.

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1).$$

Jeśli σ_1^2 i σ_2^2 – nieznane, estymujemy je z danych (s_1^2 i s_2^2):

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \approx N(0, 1).$$

Test niesparowany, mała próba

- **Założenia:**

$$X_{1,1}, \dots, X_{1,n_1} \sim N(\mu_1, \sigma^2),$$

$$X_{2,1}, \dots, X_{2,n_2} \sim N(\mu_2, \sigma^2).$$

równe wariancje!

- Estymator **wariancji łącznej:**

$$s^2 = \frac{1}{n_1 + n_2 - 2} \left(\sum_{i=1}^{n_1} (X_{1,i} - \bar{X}_1)^2 + \sum_{i=1}^{n_2} (X_{2,i} - \bar{X}_2)^2 \right)$$

Test niesparowany, mała próba

- **Założenia:**

$$X_{1,1}, \dots, X_{1,n_1} \sim N(\mu_1, \sigma^2),$$

$$X_{2,1}, \dots, X_{2,n_2} \sim N(\mu_2, \sigma^2).$$

równe wariancje!

- Estymator **wariancji łącznej:**

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Test niesparowany, mała próba

- **Założenia:**

$$X_{1,1}, \dots, X_{1,n_1} \sim N(\mu_1, \sigma^2),$$

$$X_{2,1}, \dots, X_{2,n_2} \sim N(\mu_2, \sigma^2).$$

równe wariancje!

- Estymator **wariancji łącznej:**

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- **Statystyka testowa:** ustandaryzowana różnica $\bar{X}_1 - \bar{X}_2$.

$$E[\bar{X}_1 - \bar{X}_2] = \mu_1 - \mu_2 = 0, \quad (\text{gdy } H_0 \text{ prawdziwe})$$

$$D^2[\bar{X}_1 - \bar{X}_2] = D^2[\bar{X}_1] + D^2[\bar{X}_2] = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right).$$

Test niesparowany, mała próba

- **Założenia:**

$$X_{1,1}, \dots, X_{1,n_1} \sim N(\mu_1, \sigma^2),$$

$$X_{2,1}, \dots, X_{2,n_2} \sim N(\mu_2, \sigma^2).$$

równe wariancje!

- Estymator **wariancji łącznej:**

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- **Statystyka testowa:** ustandaryzowana różnica $\bar{X}_1 - \bar{X}_2$.

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t(n_1 + n_2 - 2).$$

Test sparowany

- **Założenia:**

- Dwie próby o tym samym rozmiarze n :

$$X_{1,1}, \dots, X_{1,n}, \\ X_{2,1}, \dots, X_{2,n}.$$

- Obserwacje parami **zależne**, tzn. $(X_{1,i}, X_{2,i})$ zależne dla każdego i .
- Testujemy **różnice** $\Delta_i = X_{1,i} - X_{2,i}$.
Efektywnie test dla jednej populacji – **populacji różnic**.

<hr/>	<hr/>		<hr/>
X_1	X_2		Δ
<hr/>	<hr/>		<hr/>
0.5	1		-0.5
-1	-2	\implies	1
0	2.5		-2.5
1.5	0.5		1
<hr/>	<hr/>		<hr/>

Test sparowany

- **Układ hipotez:**

$$H_0 : \quad \mu_{\Delta} = \mu_1 - \mu_2 = 0 \quad (\mu_{\Delta} \geq 0) \quad (\mu_{\Delta} \leq 0)$$

$$H_1 : \quad \mu_{\Delta} \neq 0 \quad \mu_{\Delta} < 0 \quad \mu_{\Delta} > 0$$

- **Statystyka testowa – ustandaryzowana średnia z różnic:**

$$T = \frac{\bar{X}_{\Delta}}{s_{\Delta}} \sqrt{n} \sim \begin{cases} N(0, 1) & n \geq 30, \\ t(n-1) & n < 30. \end{cases}$$