Testy dwóch populacji

Statystyka i analiza danych 2017/2018

Jurek Błaszczyński,
na podstawie slajdów Wojtka Kotłowskiego

6 maja 2018

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

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

- **Układ hipotez**:

H_0 :	$\mu_1 = \mu_2$	$(\mu_1 \geq \mu_2)$	$(\mu_1 \leq \mu_2)$
H_1 :	$\mu_1 \neq \mu_2$	$\mu_1 < \mu_2$	$\mu_1 > \mu_2$

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

$$E[\overline{X}_1 - \overline{X}_2] = \mu_1 - \mu_2 = 0, \qquad \text{(gdy } H_0 \text{ prawdziwe)}$$
$$D^2[\overline{X}_1 - \overline{X}_2] = D^2[\overline{X}_1] + D^2[\overline{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

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

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

- **Układ hipotez**:

$H_0:$	$\mu_1 = \mu_2$	$(\mu_1 \geq \mu_2)$	$(\mu_1 \leq \mu_2)$
$H_1:$	$\mu_1 \neq \mu_2$	$\mu_1 < \mu_2$	$\mu_1 > \mu_2$

- **Statystyka testowa**: ustandaryzowana różnica $\overline{X}_1 - \overline{X}_2$.
$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1).$$
Jeśli σ_1^2 i σ_2^2 – nieznane, estymujemy je z danych (s_1^2 i s_2^2):
$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \approx N(0, 1).$$

Test niesparowany, mała próba

- **Założenia**:

$$\left.\begin{aligned} X_{1,1}, \ldots, X_{1,n_1} &\sim N(\mu_1, \sigma^2), \\ X_{2,1}, \ldots, X_{2,n_2} &\sim N(\mu_2, \sigma^2). \end{aligned}\right\rangle \boxed{\text{równe wariancje!}}$$

- Estymator **wariancji łącznej**:

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

Test niesparowany, mała próba

- **Założenia**:

$$\left.\begin{aligned} X_{1,1}, \ldots, X_{1,n_1} &\sim N(\mu_1, \sigma^2), \\ X_{2,1}, \ldots, X_{2,n_2} &\sim N(\mu_2, \sigma^2). \end{aligned}\right\rangle \boxed{\text{równe wariancje!}}$$

- 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**:
$$\left.\begin{aligned} X_{1,1}, \ldots, X_{1,n_1} &\sim N(\mu_1, \sigma^2), \\ X_{2,1}, \ldots, X_{2,n_2} &\sim N(\mu_2, \sigma^2). \end{aligned}\right\rangle \boxed{\text{równe wariancje!}}$$

- 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 $\overline{X}_1 - \overline{X}_2$.
$$E[\overline{X}_1 - \overline{X}_2] = \mu_1 - \mu_2 = 0, \qquad \text{(gdy } H_0 \text{ prawdziwe)}$$
$$D^2[\overline{X}_1 - \overline{X}_2] = D^2[\overline{X}_1] + D^2[\overline{X}_2] = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right).$$

Test niesparowany, mała próba

- **Założenia**:

$$\left.\begin{array}{l} X_{1,1}, \ldots, X_{1,n_1} \sim N(\mu_1, \sigma^2), \\ X_{2,1}, \ldots, X_{2,n_2} \sim N(\mu_2, \sigma^2). \end{array}\right\rangle \boxed{\text{równe wariancje!}}$$

- 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 $\overline{X}_1 - \overline{X}_2$.

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

Test sparowany

- **Założenia**:
  - Dwie próby o tym samym rozmiarze n:

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

  - 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**.

X_1	X_2
0.5	1
−1	−2
0	2.5
1.5	0.5

$\Longrightarrow$

Δ
−0.5
1
−2.5
1

Test sparowany

- **Układ hipotez**:

H_0 :	$\mu_\Delta = \mu_1 - \mu_2 = 0$	$(\mu_\Delta \geq 0)$	$(\mu_\Delta \leq 0)$
H_1 :	$\mu_\Delta \neq 0$	$\mu_\Delta < 0$	$\mu_\Delta > 0$

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

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

Test F porównujący wariancję w dwóch populacjach

- **Układ hipotez**:

H_0 :	$\sigma_1^2 = \sigma_2^2$
H_1 :	$\sigma_1^2 > \sigma_2^2$

- **Statystyka testowa** – ma rozkład F (F-Snedecor'a) z k_1 i k_2 stopniami swobody:

$$F = \frac{s_1^2}{s_2^2}.$$