Wnioskowanie statystyczne

№2. 5 marca 2018

a. Testowanie wartości średniej

W poprzednim tygodniu mowa była o teście Z (ze znaną wariancją, rozkład $\mathrm{N}(0,1)$) i o teście t (bez wiedzy o wariancji, rozkład t-Studenta).

a1. Test dwu średnich, wariancje znane i równe

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sigma\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim \mathrm{N}(0,1).$$

a2. Test dwu średnich, wariancje znane ale różne

Statystyka testowa:

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \sim \mathrm{N}(0,1).$$

a3. Test dwu średnich, wariancje nieznane i równe

Niech $s_x^2 = \sum_{i=1}^{n_x}(x_i - \bar{x})^2,\ \ s_y^2 = \sum_{i=1}^{n_y}(y_i - \bar{y})^2,\ \ s = \dfrac{(n_x-1)s_x^2 + (n_y-1)s_y^2}{n_x + n_y - 2}.$
Statystyka testowa:

$$t = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{s\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim t(n_x + n_y - 2).$$

a4. Test dwu średnich, wariancje nieznane i równe

Niech $s_x^2 = \dfrac{1}{n_x - 1}\sum_{i=1}^{n_x}(x_i - \bar{x})^2,\ \ s_y^2 = \dfrac{1}{n_y - 1}\sum_{i=1}^{n_y}(y_i - \bar{y})^2,\ \ s = \dfrac{(n_x-1)s_x^2 + (n_y-1)s_y^2}{n_x + n_y - 2}.$
Statystyka testowa:

$$t = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} \sim t(n),\ \text{gdzie}$$

$$n = \frac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right)^2}{\frac{s_x^4}{n_x^2(n_x-1)} + \frac{s_y^4}{n_y^2(n_y-1)}}.$$

a5. Test średnich, dane skojarzone

Niech $d_i = x_i - y_i$, oraz $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$.
Statystyka testowa:

$$t = \frac{\bar{X} - \bar{Y}}{s}\sqrt{n} \ \sim t(n-1),\ \text{gdzie}$$

b. Testowanie wariancji

b1. Znana wartość oczekiwana

Niech $s^2 = \dfrac{1}{n}\sum_{k=1}^{n}(x_k - mu)^2$.
Statystyka testowa:

$$\chi^2 = \frac{ns^2}{\sigma^2} \ \sim \chi^2(n).$$

b2. Wartość oczekiwana nieznana

Niech $s^2 = \dfrac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2$.
Statystyka testowa:

$$\chi^2 = \frac{ns^2}{\sigma^2} \ \sim \chi^2(n-1).$$

W trzech początkowych zadaniach dane do obliczeń znajdują się w pliku **w-0201.csv**

Lista zadań №2, 19 marca 2018

1. Obliczyć:
   (a) średnią i wariancję,
   (b) przedział ufności dla wartości oczekiwanej (poziom ufności $\alpha = 0.05$.
2. Sprawdzić hipotezę o wartości oczekiwanej, $H_0 : \mu = 18$, $H_1 : \mu \neq 18$, $\alpha = 0.05$. Porównać przdział ufności z zadania poprzedniego z testowaną wartościa.
3. Obliczyć Wartość_p dla $\mu = 17$.
4. Dla danych z pliku **w-0204.csv**:
   (a) sprawdzić hipotezę $H_0 : \mu = 2750$, wobec hipotezy $H_1 : \mu > 2750$, $\alpha = 0.05$,
   (b) wyznaczyć przedział ufności dla wartości średniej,
   (c) obliczyć Wartość_p.
5. W ankiecie uzyskano 52 odpowiedzi Tak, ankietowano 215 osób.
   (a) Sprawdzić hipotezę $H_0 : p = 0.22$, wobec hipotezy $H_1 : p \neq 0.22$, $\alpha = 0.05$,

   (b) wyznaczyć przedział ufności dla częstości p,
   (c) obliczyć Wartość_p.

6. Dla danych z pliku **w-0206.csv**:

   (a) sprawdzić hipotezę $H_0 : \sigma^2 = 3.5$, wobec hipotezy $H_1 : \sigma^2 < 3.5$, $\alpha = 0.05$,
   (b) wyznaczyć przedział ufności dla wariancji σ^2,
   (c) obliczyć Wartość_p.

7. Liczebność próbki $n = 300$, średnia z próbki $\bar{x} = 715$, wariancja $s^2 = 400$. Sprawdzić hipotezę $H_0 : \mu = 730$, wobec hipotezy $H_1 : \mu < 730$, $\alpha = 0.05$.
8. Utworzyć wykresy gęstości następujących rozkładów:.

   (a) $N(0, 1)$ i $N(0, 2)$; oś X w granicach od -2 do 2,
   (b) $N(0, 1)$ i $N(1, 1)$; oś X w granicach od -4 do 4,
   (c) t_{10}, t_5 i $N(0, 1)$, na przedziale $[-4, 4]$,
   (d) χ^2_4, χ^2_6 i χ^2_8 na przedziale $[0; 20]$.

Witold Karczewski