Modele i Wnioskowanie Statystyczne
Laboratorium 4

Sprawozdanie

Igor Markiewicz

Zadanie 1
Jako rozkład badany przyjęto rozkład dyskretny z prawdopodobieństwem samoóbjstwa w ciągu danego miesiąca określonym jako :

$$p_i = d_i \cdot \frac{1}{d}$$

Gdzie :

- d_i – liczba dni w danym miesiącu
- $d = \sum_{i=1}^{12} d_i$ – liczba dni w roku

Badaną statystyką jest statystyka Pearsona :

$$T = \sum_{i=1}^{r} \frac{(v_i - np_i)^2}{np_i}$$

Gdzie :

- $r = 12$ – liczba miesięcy
- v_i – liczba samobójstw w danym miesiącu
- $n = \sum_{i=1}^{r} v_i$ – liczba wszystkich samobójstw w ciągu roku

Wtedy, statystyka T ma w przybliżeniu rozkład χ^2_{r-1} oraz

$$\begin{cases} H_0 - \text{założony rozkład :} & T \leq c \\ H_1 - \text{rozkład inny niż założony :} & T > c \end{cases}$$

Jako wyniki otrzymano :

- T $\approx 47,36528$
- $c_{0,1} = F^{-1}_{\chi^2_{12-1}}(1 - 0,9) \approx 5,577785$

Wniosek :
Dla $\alpha = 0,1$ możemy odrzucić hipotezę zerową na rzecz alternatywnej.

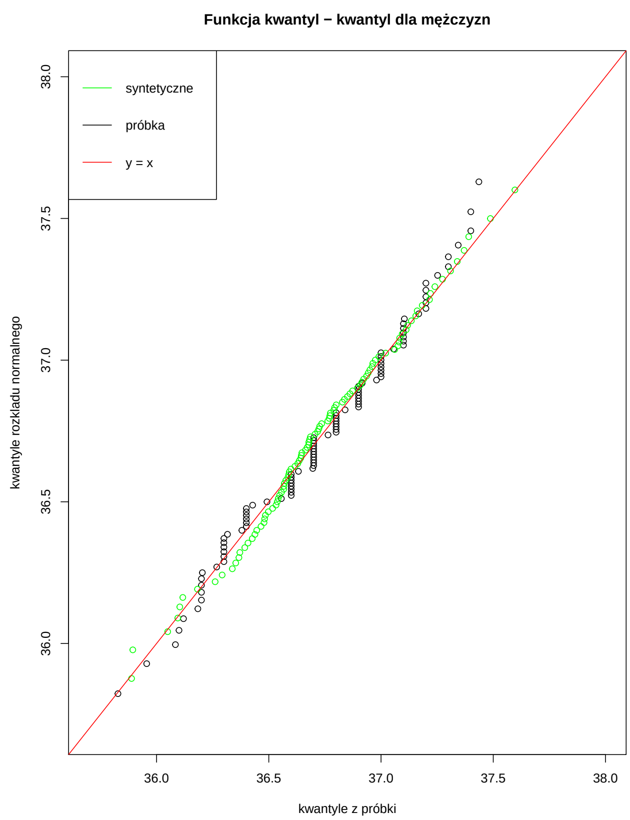
Zadanie 2
a)

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

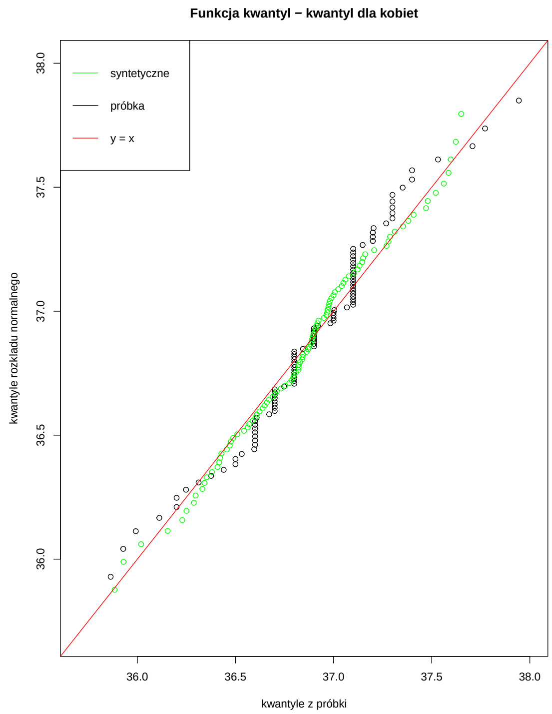
$$Var(x) = \frac{1}{n-1}\sum_{i=1}^{n}(\overline{x} - x_i)^2$$

	średnia temperatura $[°C]$	**wariancja temperatury** $[°C]$
mężczyźni	36,73	0,15
kobiety	36,89	0,17

Tab. 1: Wysetymowane parametry



Rys. 1: Wykresy kwantyl – kwantyl dla mężczyzn



Rys. 2: Wykresy kwantyl – kwantyl dla kobiet

Wniosek :
W obu przypadkach możemy stwierdzić że zarówno dla mężczyzn jak i kobiet temperatura ma w przybliżeniu rozkład normalny. Dla kobiet takie stwierdzenie może budzić pewne wątpliwości, z racji na ilość odstających pomiarów, jednak w celu bardziej precyzyjnego stwierdzenia należałoby przeprowadzić testy ilościowe.

b)
Test normalności Shapiro − Wilka

$$H_0 \ : \ \text{próba pochodzi z populacji o rozkładzie normalnym}$$
$$H_1 \ : \ \text{próba nie pochodzi z populacji o rozkładzie normalnym}$$

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2}$$

- $x_{(i)}$ – i-ta statystyka porządkowa (np: i-ta najmniejsza liczba)
- $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- $(a_1, \ldots, m_n)^T = \frac{\boldsymbol{m}^T \boldsymbol{V}^{-1}}{\sqrt{\boldsymbol{m}^T \boldsymbol{V}^{-2} \boldsymbol{m}}}$

- $\boldsymbol{m} = (m_1, \ldots, m_n)^T$
- $m_1, \ldots, m_n$ – wartości oczekiwane statystyk porządkowych typu $i.i.d$ pochodzące z rozkładu normalnego
- $\boldsymbol{V}$ – macierz kowariancji statystyk porządkowych

Wyniki :

- mężczyźni : $p - value = 0,4818$
- kobiety : $p - value = 0,03351$

Wniosek :
Dla $\alpha = 0,05$ oraz $\alpha = 0,1$ w przypadku mężczyzn nie mamy powodów do odrzucenia hipotezy zerowej, natomiast dla kobiet możemy ją odrzucić.

Test średniej

$$H_0 \ : \ \text{średnia temperatura ciała wynosi } 36,6\ ^\circ C$$
$$H_1 \ : \ \text{średnia temperatura ciała jest różna od } 36,6\ ^\circ C$$

Dla dużej liczności próbki zmienna losowa :

$$U = \frac{\overline{x} - m_0}{\frac{\sigma}{\sqrt{n}}}$$

ma asymptotycznie rozkład zbiegający do $\mathcal{N}(0,1)$ (na podstawie centralnego twierdzenia granicznego), dlatego nawet jeśli rozkład cechy w populacji nie jest normalny to możemy użyć tego testu do weryfikacji hipotezy o wartości średniej arytmetycznej, przy czym σ estymujemy jako :

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

Możemy przy tym spróbować zastosować test t – Studenta, również z racji na działanie centralnego twierdzenia granicznego.

Wyniki i wnioski :
Dla mężczyzn dwustronny $t.test$ zwraca $p - value = 0.01097$, natomiast dla kobiet $p - value = 3.985 \cdot 10^{-7}$ w związku z czym dla poziomów $\alpha = 0,05$ oraz $\alpha = 0,1$ możemy odrzucić hipotezy zerowe na rzecz altenatywnych.