

Sprawozdanie 5

Jakub Markowiak
album 255705

20 czerwca 2021

Spis treści

1	Krótki opis zagadnienia	1
2	Opis eksperymentów/analiz	1
3	Wyniki	2
3.1	Test t-Studenta dla pojedynczej próby	2
3.2	Test t-Studenta dla prób zależnych	4
3.3	Test t-Studenta dla prób niezależnych	6
3.4	Test dla wariancji rozkładu normalnego	8
4	Podsumowanie	9

1 Krótki opis zagadnienia

W tym sprawozdaniu zajmiemy się wykorzystaniem testów statystycznych do weryfikowania hipotez. Rozważymy testy t-Studenta dla pojedynczej próby, dla dwóch prób zależnych, dla dwóch prób niezależnych oraz test dla wariancji w rodzinie rozkładów normalnych.

2 Opis eksperymentów/analiz

Przeprowadzimy następujące analizy i eksperymenty:

1. test t-Studenta dla pojedynczej próby,
2. test t-Studenta dla prób zależnych,
3. test t-Studenta dla prób niezależnych,
4. test dla wariancji rozkładu normalnego.

	1	2	3	4	5	6
Wynik [h]	19.00	18.00	22.00	20.00	16.00	25.00

Tabela 1: Wyniki testu laboratoryjnego

3 Wyniki

3.1 Test t-Studenta dla pojedynczej próby

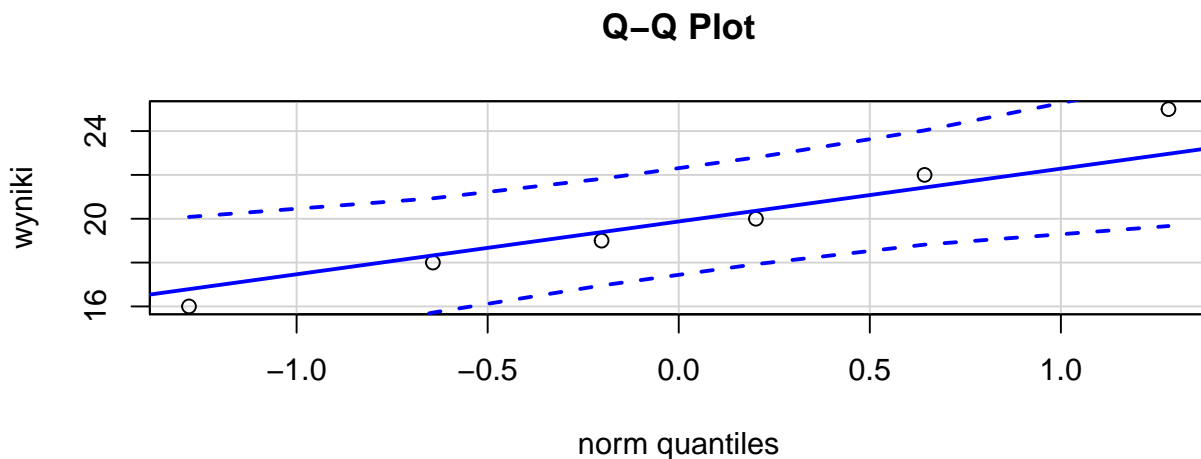
Dysponujemy wynikami testu laboratoryjnego dla 6 baterii. Zweryfikujemy hipotezę, że średni czas pracy baterii wynosi $\mu = 21.5$. Przyjęty przez nas poziom istotności to 0.01.

Zweryfikujemy najpierw, czy te dane pochodzą z rozkładu normalnego.

	Statystyka W	p-value
wartość	0.98	0.95

Tabela 2: Test Shapiro-Wilka

Widzimy, że wartość poziomu krytycznego jest wyższa niż 0.01, czyli przyjęty przez nas poziom istotności, zatem nie ma podstaw aby odrzucić hipotezę o normalności tego rozkładu.



Rysunek 1: Q-Q plot, badanie normalności rozkładu

Spoglądając na wykres Q-Q również nie mamy podstaw, aby stwierdzić, że nasze dane nie pochodzą z rozkładu normalnego.

Sformułujemy teraz hipotezę:

$$H_0 : \mu = 21.5,$$

$$H_1 : \mu \neq 21.5.$$

W celu jej weryfikacji korzystamy z testu t-Studenta dla pojedynczej próby. Wykorzystujemy statystykę testową:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}},$$

gdzie

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}.$$

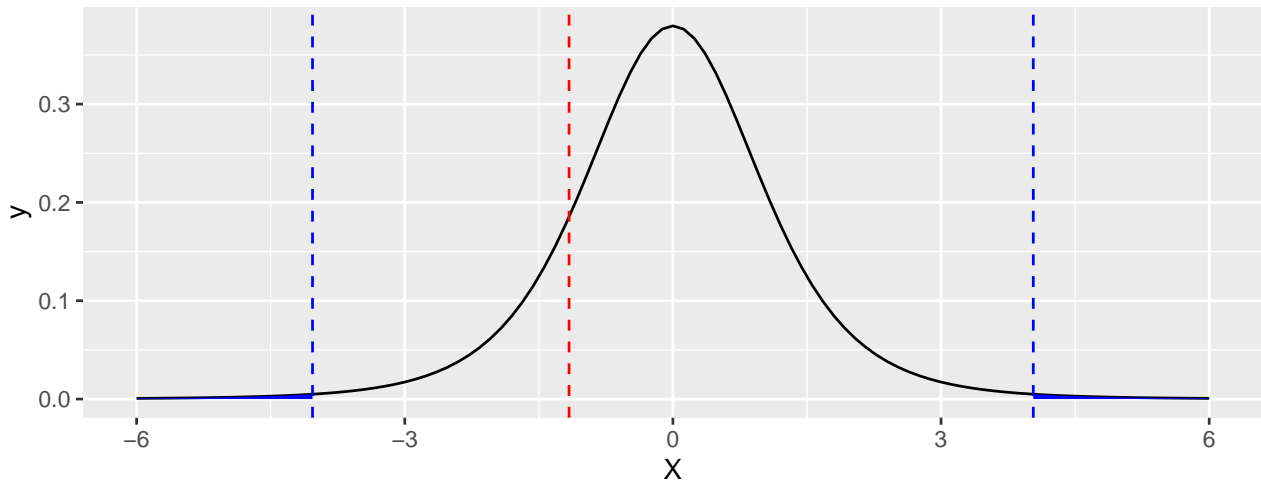
Przedział ufności na poziomie ufności $1 - \alpha$ mają postać:

$$[L, P] = \left[\bar{X} - z_{(1-\alpha)/2} \frac{S}{\sqrt{n}}, \bar{X} + z_{(1-\alpha)/2} \frac{S}{\sqrt{n}} \right].$$

	Confidence start	Confidence end	p-value	critical value	T
wartość	14.79	25.21	0.30	-4.03	-1.16

Tabela 3: Test t-studenta, $H_1 : \mu \neq 21.5$

Obszar krytyczny oraz wartosc statystyki T



Rysunek 2: Q-Q plot, badanie normalności rozkładu

Widzimy, że $\mu_0 = 21.5$ należy do przedziału ufności, a także p-value jest większe od 0.05, oraz wartość statystyki T dla naszej próby nie znalazła się w obszarze krytycznym. Stąd nie ma podstaw, aby odrzucić hipotezę zerową, więc możemy ją przyjąć na poziomie istotności 0.05.

Sformuujemy teraz drugą hipotezę:

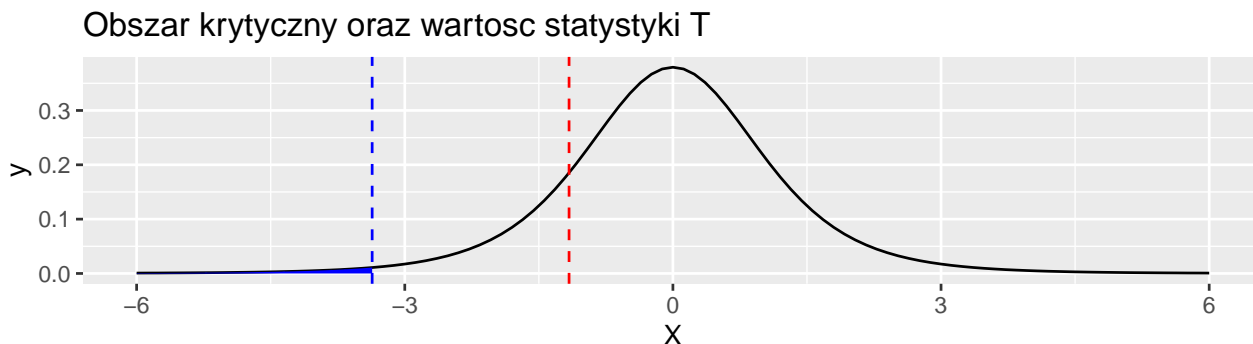
$$H_0 : \mu = 21.5,$$

$$H_1 : \mu < 21.5.$$

	Confidence start	Confidence end	p-value	critical value	T
wartość	-Inf	24.34	0.15	-3.36	-1.16

Tabela 4: Test t-studenta, $H_1 : \mu < 21.5$

Ponownie do jej weryfikacji wykorzystamy test t-Studenta.



Rysunek 3: Q-Q plot, badanie normalności rozkładu

Podobnie jak poprzednio widzimy, że $\mu_0 = 21.5$ należy do przedziału ufności, p-value jest większe od 0.05, oraz wartość statystyki T nie znalazła się w obszarze krytycznym. Stąd nie ma podstaw, aby odrzucić hipotezę zerową, więc możemy ją przyjąć na poziomie istotności 0.05.

3.2 Test t-Studenta dla prób zależnych

Dysponujemy wynikami pomiarów ciśnienia tętniczego przed leczeniem (A) i po leczeniu (B). Zweryfikujemy na poziomie istotności $\alpha = 0.05$ hipotezę, że lek ten powoduje istotny spadek ciśnienia u leczonych pacjentów.

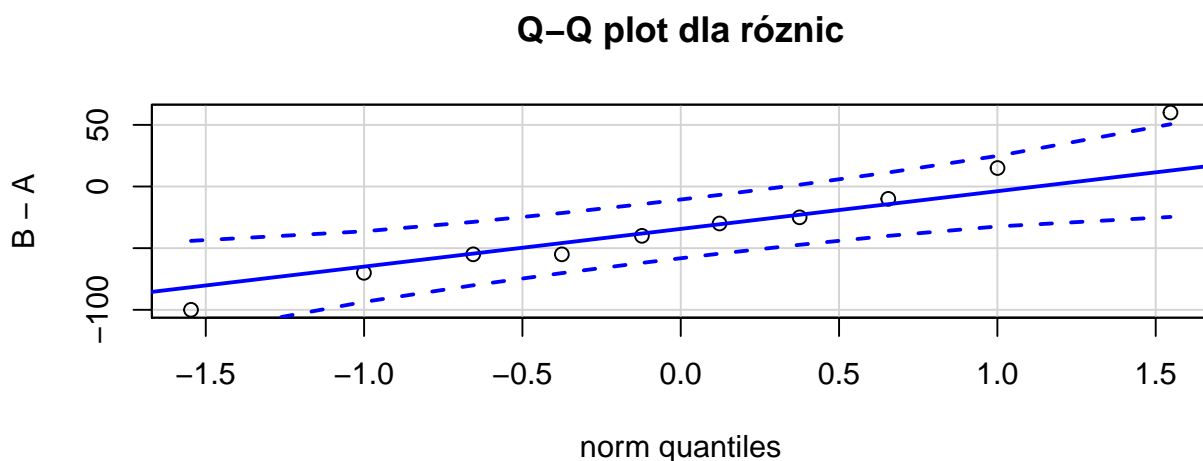
	1	2	3	4	5	6	7	8	9	10
A	220.00	185.00	270.00	285.00	200.00	295.00	200.00	190.00	225.00	230.00
B	190.00	175.00	215.00	260.00	215.00	195.00	260.00	150.00	155.00	175.00

Tabela 5: Wyniki pomiarów ciśnienia tętniczego

Sprawdzimy, czy różnice między kolejnymi obserwacjami pochodzą z rozkładu normalnego.

	Statystyka W	p-value
wartość	0.97	0.88

Tabela 6: Test Shapiro-Wilka dla różnic



Rysunek 4: Q-Q plot, badanie normalności różnic

P-value jest wyższe od przyjętego przez nas poziomu istotności, stąd można przyjąć normalność rozkładu różnic. Również wykres Q-Q potwierdza nasze przypuszczenia.

Sformułujemy teraz hipotezę:

$$H_0 : \mu_1 = \mu_2,$$

$$H_1 : \mu_1 > \mu_2.$$

W celu jej weryfikacji korzystamy z testu t-Studenta dla prób zależnych. Wykorzystujemy statystykę testową:

$$T = \frac{\bar{D}}{S_D/\sqrt{n}},$$

gdzie

$$D = (D_1 + D_2 + \dots + D_n)/n,$$

$$D_i = X_i - Y_i,$$

oraz

$$S_D^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2.$$

	Confidence start	Confidence end	p-value	T
wartość	4.80	Inf	0.03	2.17

Tabela 7: Test t-studenta, $H_1 : \mu_1 > \mu_2$

Widzimy, że p-value jest niższe od $\alpha = 0.05$, stąd odrzucamy H_0 i przyjmujemy H_1 na poziomie istotności 0.05. Z wyznaczonego przedziału ufności możemy odczytać, że na poziomie ufności 0.95 możemy stwierdzić, że wzrost ciśnienia jest większy od 4.7965285.

3.3 Test t-Studenta dla prób niezależnych

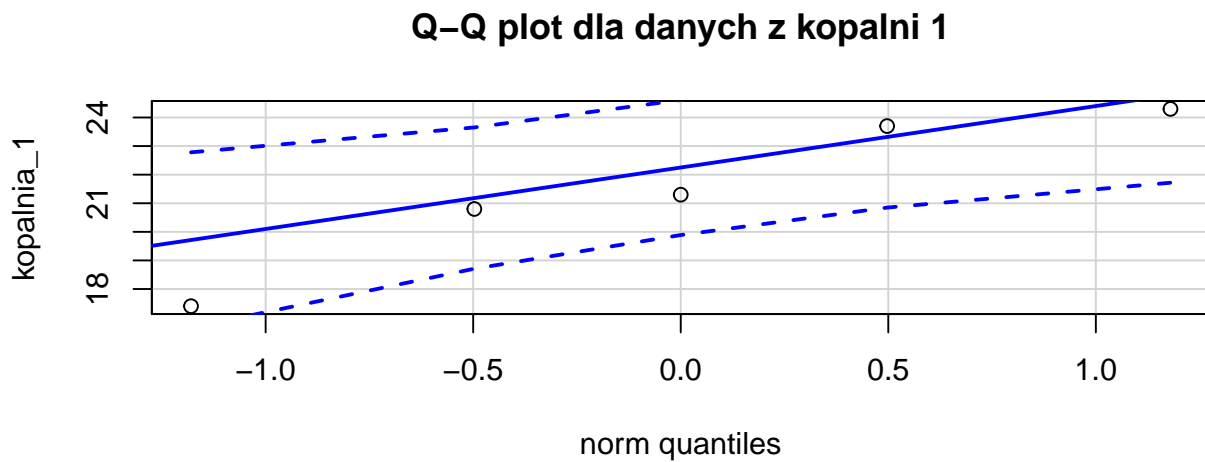
Dysponujemy danymi z dwóch kopalń, które przedstawiają ilość popiołu (w procentach), który pozostał po spaleniu węgla.

	1	2	3	4	5
Kopalnia 1	24.3	20.8	23.7	21.3	17.4
Kopalnia 2	18.2	16.9	20.2	16.7	-

Tabela 8: Wyniki pomiarów ilości popiołu

	Statystyka W	p-value
wartość	0.93	0.60

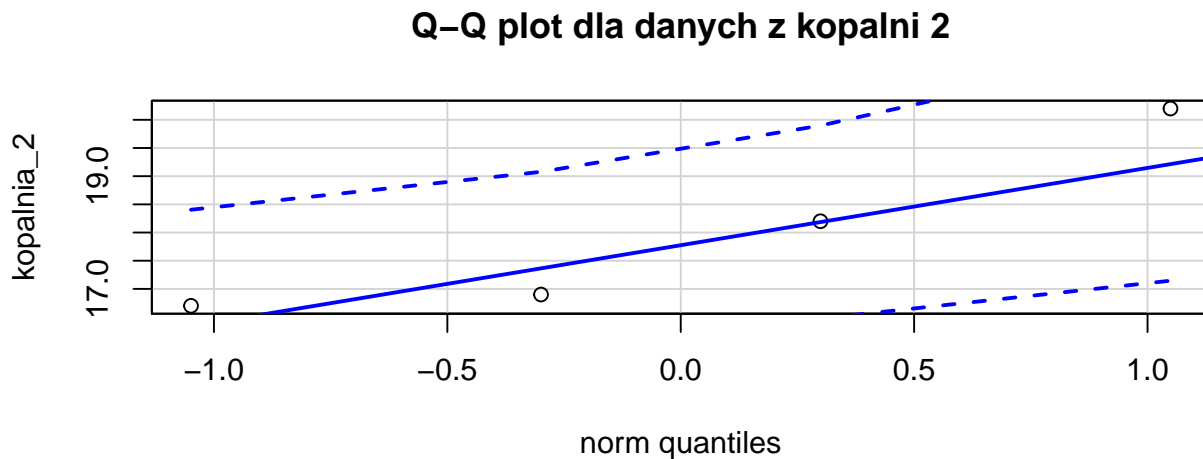
Tabela 9: Test Shapiro-Wilka dla danych z kopalni 1



Rysunek 5: Badanie normalności - Q-Q plot

	Statystyka W	p-value
wartość	0.88	0.35

Tabela 10: Test Shapiro-Wilka dla danych z kopalni 2



Rysunek 6: Badanie normalności - Q-Q plot

Ponieważ p-value jest w obu przypadkach większe od 0.05, możemy przyjąć, że nasze dane pochodzą z rozkładu normalnego. Wykresy Q-Q również nie zaprzeczyły normalności tych rozkładów. Sprawdźmy teraz, czy możemy założyć, że dla danych z kopalni 1 i kopalni 2 mamy jednakową wariancję. Rozpatrujemy hipotezę na poziomie istotności $\alpha = 0.05$:

$$H_0 : \sigma_1 = \sigma_2,$$

$$H_1 : \sigma_1 \neq \sigma_2.$$

	Confidence start	Confidence end	p-value
wartość	0.19	28.88	0.41

Tabela 11: Var Test, $H_1 : \sigma_1 \neq \sigma_2$

Ponieważ p-value jest większe od $\alpha = 0.05$, możemy przyjąć hipotezę zerową, zatem przyjmujemy, że w obu rozkładach jest taka sama wariancja.

Zweryfikujemy następującą hipotezę:

$$H_0 : \mu_1 = \mu_2,$$

$$H_1 : \mu_1 \neq \mu_2,$$

gdzie $\mu_k, k \in \{1, 2\}$ – średnia ilość popiołu, który zostaje po spaleniu węgla z kopalni k .

W celu jej weryfikacji korzystamy z testu t-Studenta dla prób niezależnych. Wykorzystujemy statystykę testową:

$$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

gdzie

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

oraz $S_i^2, i \in \{1, 2\}$ – wariancja w i -tej próbie.

	Confidence start	Confidence end	p-value
wartość	-0.19	7.19	0.06

Tabela 12: test t-Studenta, $H_1 : \mu_1 \neq \mu_2$

Ponieważ p-value jest większe niż $\alpha = 0.05$, nie ma podstaw do odrzucenia H_0 , zatem przyjmujemy H_0 . Na poziomie ufności 0.95 możemy również stwierdzić, że różnica między obiema średnimi znajduje się w przedziale $[-0.19, 7.19]$.

Zweryfikujemy teraz następującą hipotezę:

$$H_0 : \mu_1 = \mu_2,$$

$$H_1 : \mu_1 > \mu_2,$$

gdzie $\mu_k, k \in \{1, 2\}$ – średnia ilość popiołu, który zostaje po spaleniu węgla z kopalni k .

	Confidence start	Confidence end	p-value
wartość	0.55	Inf	0.03

Tabela 13: test t-Studenta, $H_1 : \mu_1 > \mu_2$

Ponieważ p-value jest mniejsza niż $\alpha = 0.05$, odrzucamy hipotezę H_0 , zatem przyjmujemy H_1 . Na poziomie ufności 0.95 możemy również stwierdzić, że różnica między obiema średnimi znajduje się w przedziale $[-0.19, 7.19]$.

3.4 Test dla wariancji rozkładu normalnego

Dysponujemy następującymi danymi, z założenia z rozkładu normalnego:

	1	2	3	4	5
wyniki	8.99	9.01	8.98	9.00	9.01

Tabela 14: Wyniki pomiarów z rozkładu normalnego

Zweryfikujemy następującą hipotezę na poziomie istotności $\alpha = 0.05$.

$$H_0 : \sigma^2 = 0,0001,$$

$$H_1 : \sigma^2 \neq 0,0001,$$

gdzie σ^2 – wariancja pomiarów.

Wykorzystujemy test dla wariancji w rodzinie rozkładów normalnych, dla którego statystyką testową jest:

$$\chi^2 = \frac{(n - 1)S^2}{\sigma_0^2}. \quad (1)$$

Przedział ufności na poziomie ufności $1 - \alpha$ ma postać:

$$\left[\frac{(n-1)S^2}{\chi^2_{(1-\alpha)/2, n-1}}, \frac{(n-1)S^2}{\chi^2_{(\alpha/2), n-1}} \right]$$

	Confidence start	Confidence end	p-value
wartość	0.0000610233	0.0014037447	0.2936847757

Tabela 15: test t-Studenta, $H_1 : \sigma^2 = 0.0001$

Ponieważ p-value jest większe od $\alpha = 0.05$ oraz $\sigma_0^2 = 0,0001$ znajduje się wewnątrz przedziału ufności, przyjmujemy hipotezę H_0 i odrzucamy hipotezę H_1 na poziomie istotności α .

4 Podsumowanie

Poniżej wypunktujemy najważniejsze wnioski, jakie można wyciągnąć z przeprowadzanych analiz:

- przed zastosowaniem testu ważne jest sprawdzenie, czy badane dane spełniają założenia,
- wykorzystanie testu Shapiro-Wilka oraz wykresu Q-Q, pozwala na zweryfikowanie, czy nasze dane pochodzą z rozkładu normalnego,
- wykorzystanie przedziałów ufności oraz p-value pozwala weryfikować hipotezy.