

# Raport zaliczeniowy

## statystyka nieparametryczna

Natalia Stefańska

2025-05-25

### Spis treści

<b>1</b>	<b>Opis danych</b>	<b>2</b>
1.1	Charakterystyka danych . . . . .	2
1.2	Wizualizacja danych . . . . .	5
<b>2</b>	<b>Weryfikowane cele badawcze</b>	<b>8</b>
2.1	Czy przechowywanie sera “A” wpływa istotnie na zmniejszenie się w nim zawartości wody? . . . . .	8
2.2	Czy przechowywanie sera “B” wpływa istotnie na zmniejszenie się w nim zawartości wody? . . . . .	10
2.3	Czy różnica w zawartości wody w terminach <code>tydzien0</code> i <code>tydzien3</code> jest istotna statystycznie? . . . . .	12
2.4	Czy różnica w zawartości wody w terminach <code>tydzien3</code> i <code>tydzien5</code> jest istotna statystycznie? . . . . .	12
2.5	Jakiego średniego procentu utraty wody można się spodziewać w terminach <code>tydzien3</code> i <code>tydzien5</code> ? . . . . .	13
2.6	Dodatkowe cele badawcze . . . . .	13
<b>3</b>	<b>Wnioski</b>	<b>16</b>

# 1 Opis danych

Dane dotyczą zawartości wody (wyrażonej w procentach) w dwóch rodzajach sera topionego: typu “A” oraz “B”, w zależności od czasu przechowywania (dojrzewania). Pomiary zostały przeprowadzone w trzech punktach czasowych - w tygodniu 0 (początek badania), tygodniu 3 oraz tygodniu 5. Dla każdego typu sera zbadano pięć próbek w każdym punkcie czasowym.

Tablica 1: Dane dla sera typu “A”

tydzien0	tydzien3	tydzien5
42.3	38.7	39.4
42.5	38.1	36.5
41.0	39.3	36.3
40.8	37.7	36.3
42.4	39.0	36.6

Tablica 2: Dane dla sera typu “B”

tydzien0	tydzien3	tydzien5
41.4	39.9	38.2
40.1	38.2	37.4
41.5	37.7	35.1
41.6	37.9	35.2
39.9	37.2	34.5

## 1.1 Charakterystyka danych

Dane składają się z dwóch prób niezależnych - próby dotyczącej sera typu “A” oraz próby dotyczącej sera typu “B”. W każdej grupie zbadano 5 próbek sera w 3 punktach czasowych - `tydzien0`, `tydzien3`, `tydzien5` - odpowiednio na początku badania, po 3 tygodniach od rozpoczęcia oraz po 5 tygodniach od rozpoczęcia.

Dane pochodzą z układu z powtarzanymi pomiarami, natomiast analizowana zmienna (*zawartość wody*) jest zmienną ilościową w skali interwałowej.

Tablica 3: Statystyki opisowe dla sera typu “A”

tydzien0	tydzien3	tydzien5
Min. :40.8	Min. :37.70	Min. :36.30
1st Qu.:41.0	1st Qu.:38.10	1st Qu.:36.30
Median :42.3	Median :38.70	Median :36.50
Mean :41.8	Mean :38.56	Mean :37.02
3rd Qu.:42.4	3rd Qu.:39.00	3rd Qu.:36.60
Max. :42.5	Max. :39.30	Max. :39.40

Tablica 4: Statystyki opisowe dla sera typu “B”

tydzien0	tydzien3	tydzien5
Min. :39.9	Min. :37.20	Min. :34.50
1st Qu.:40.1	1st Qu.:37.70	1st Qu.:35.10
Median :41.4	Median :37.90	Median :35.20
Mean :40.9	Mean :38.18	Mean :36.08
3rd Qu.:41.5	3rd Qu.:38.20	3rd Qu.:37.40
Max. :41.6	Max. :39.90	Max. :38.20

Tablica 5: Statystyki opisowe dla każdego tygodnia - bez podziału na typy serów

tydzien0	tydzien3	tydzien5
Min. :39.90	Min. :37.20	Min. :34.50
1st Qu.:40.85	1st Qu.:37.75	1st Qu.:35.48
Median :41.45	Median :38.15	Median :36.40
Mean :41.35	Mean :38.37	Mean :36.55
3rd Qu.:42.12	3rd Qu.:38.92	3rd Qu.:37.20
Max. :42.50	Max. :39.90	Max. :39.40

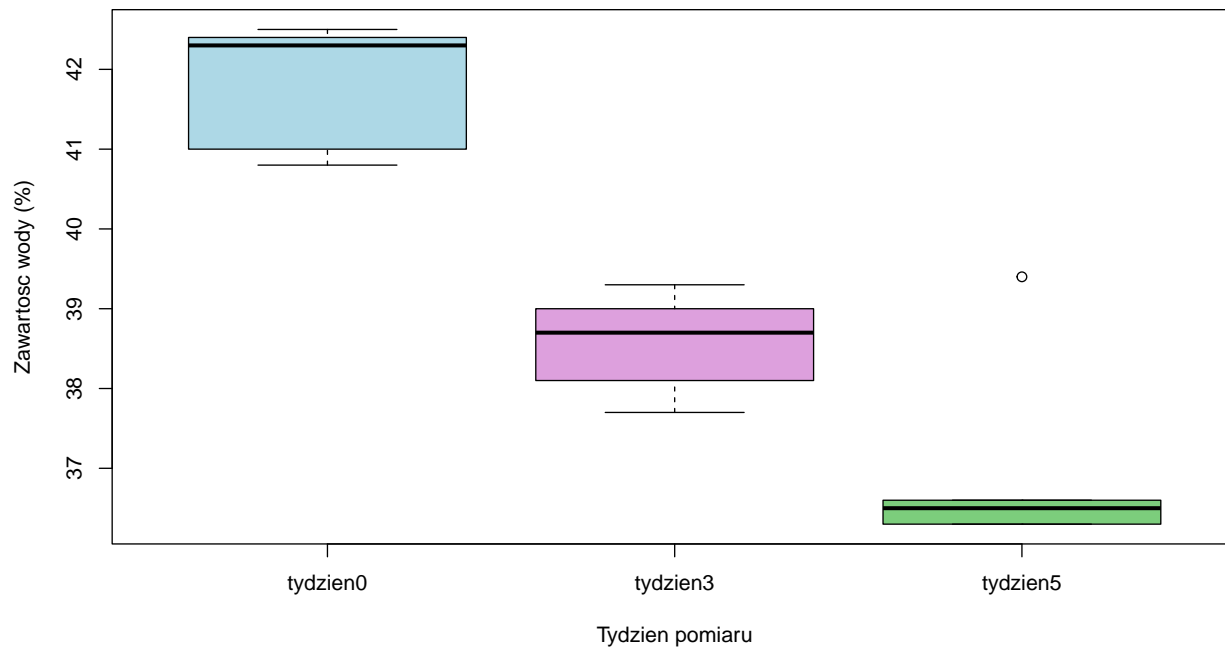
Tablica 6: Odchylenia standardowe dla każdego tygodnia -  
dla sera typu “A” oraz dla sera typu “B”

	tydzien0	tydzien3	tydzien5
sd_A	0.8276473	0.6542171	1.336787
sd_B	0.8276473	1.0281051	1.617714

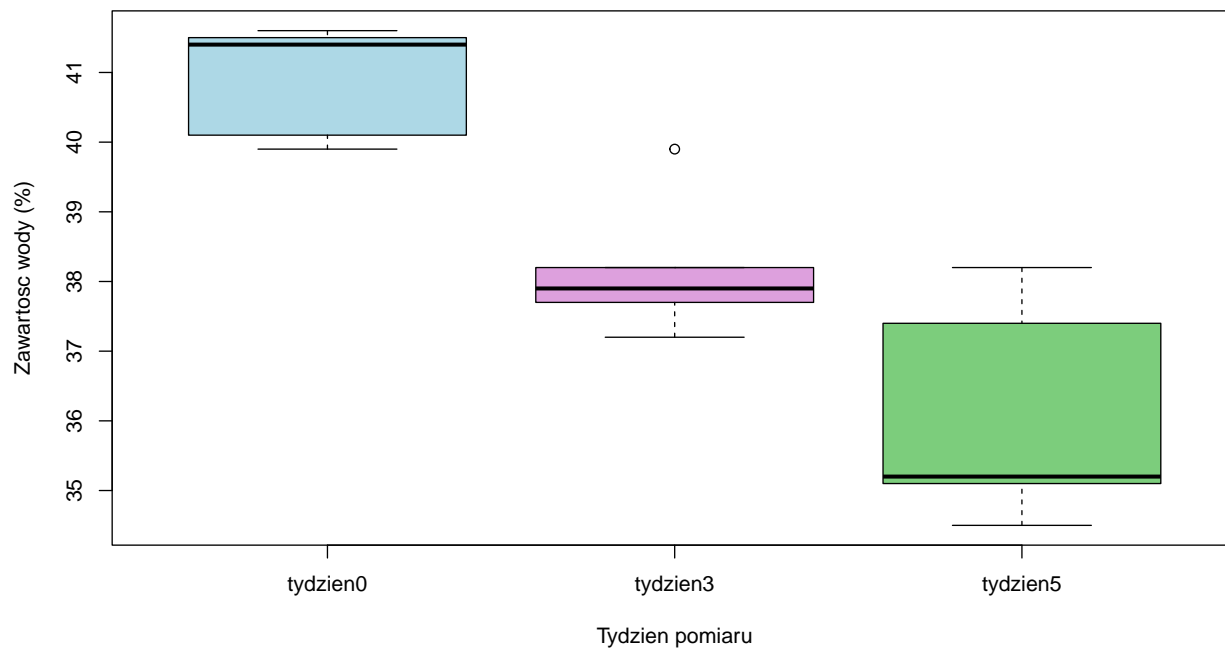
W tabelach powyżej przedstawiono statystyki opisowe - minimum, maksimum, kwartyle, średnie oraz odchylenia standardowe zawartości wody dla każdego rodzaju sera w poszczególnych tygodniach. Dane wskazują na systematyczny spadek zawartości wody w czasie, którego istotność zostanie zweryfikowana przy użyciu odpowiednich testów statystycznych.

## 1.2 Wizualizacja danych

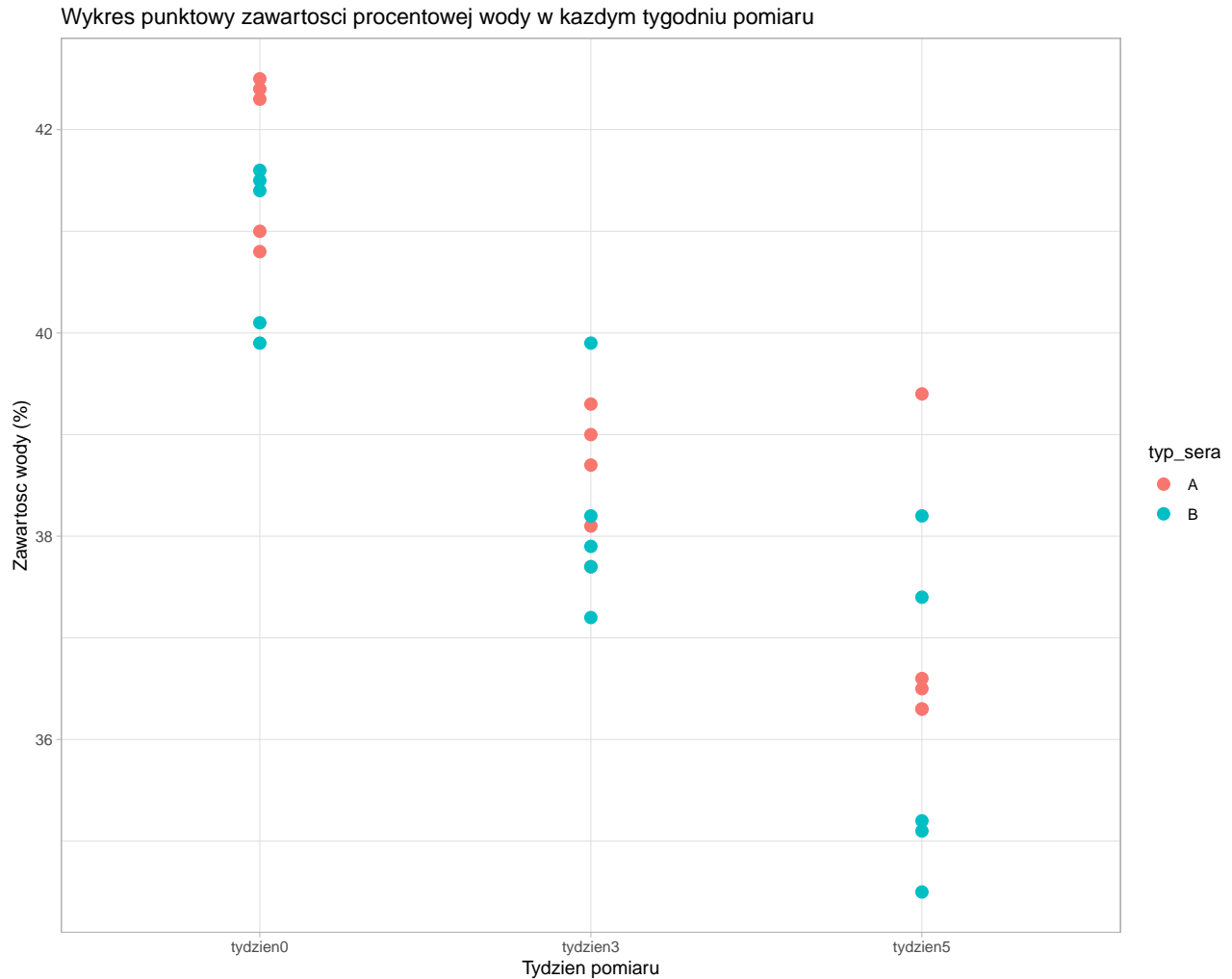
Wykres pudełkowy dla danych dotyczących sera typu "A"



Wykres pudełkowy dla danych dotyczących sera typu "B"

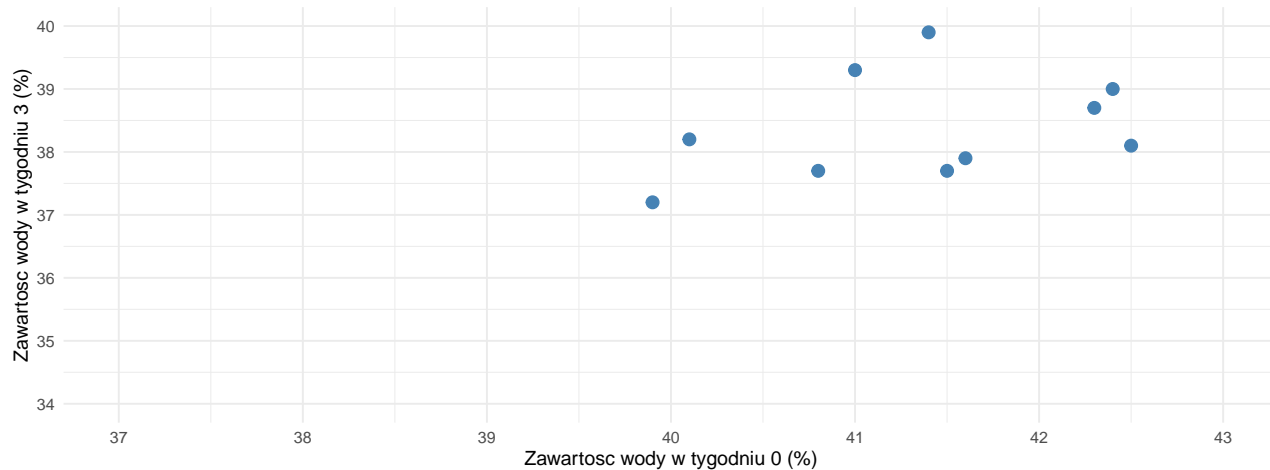


Na wykresach pudełkowych przedstawiono rozkład zawartości wody dla serów typu “A” i “B” w trzech punktach czasowych. Widoczna jest tendencja spadkowa, szczególnie wyraźna między tygodniem 0 a tygodniem 5. W obu grupach obserwuje się obniżenie median zawartości wody, co sugeruje postępujący proces odwodnienia w trakcie przechowywania.

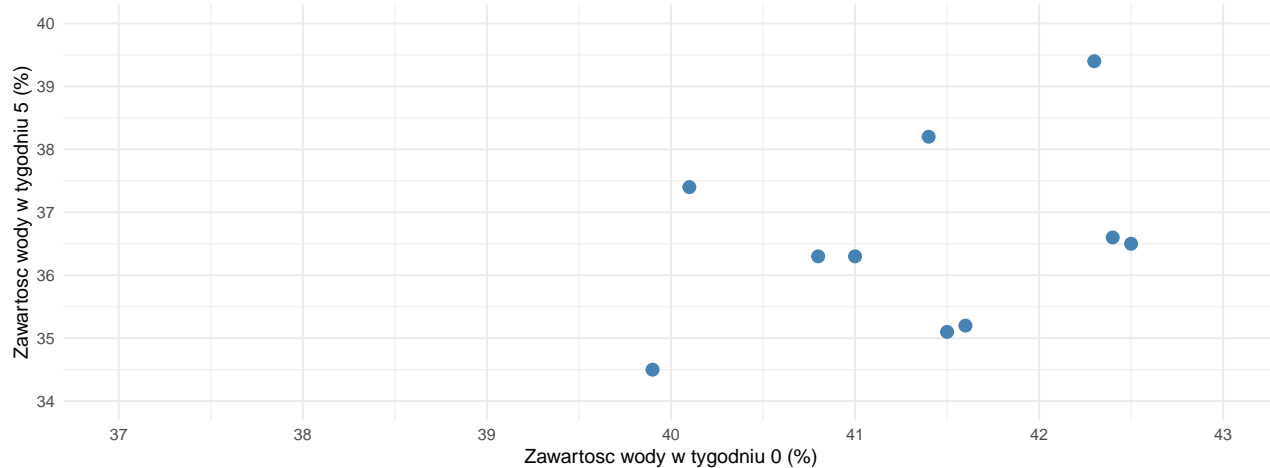


Na wspólnym wykresie punktowym przedstawiono zawartość procentową wody w serach typu “A” i “B”, przy czym kolor punktów wskazuje na typ sera.

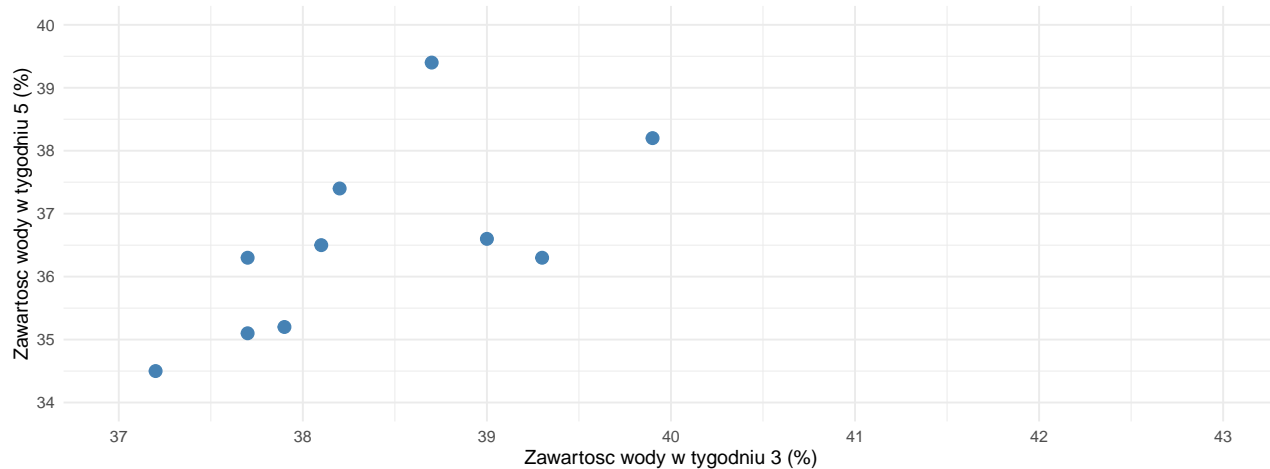
Zależność między zawartością wody w tygodniu 0 a 3



Zależność między zawartością wody w tygodniu 0 a 5



Zależność między zawartością wody w tygodniu 3 a 5



## 2 Weryfikowane cele badawcze

### 2.1 Czy przechowywanie sera “A” wpływa istotnie na zmniejszenie się w nim zawartości wody?

Pomiar zawartości wody prowadzono na tych samych pięciu próbkach sera w trzech momentach czasowych, co oznacza, że dane pochodzą z układu z powtarzanymi pomiarami (próby zależne).

#### Sformułowanie problemu w postaci hipotez

$H_0$ : rozkłady zawartości wody w tygodniach 0, 3 i 5 dla sera “A” są identyczne

$H_1$ : przynajmniej jeden z rozkładów różni się od pozostałych

#### Zapis matematyczny hipotez

$H_0: F_{tydzien0} = F_{tydzien3} = F_{tydzien5}$

$H_1: \exists_{i \neq j} F_{tydzien i} \neq F_{tydzien j}$

**Przeprowadzony test:** `friedman.test`

- otrzymana  $p$ -wartość:  $0.0149956 < 0.05$
- decyzja: odrzucenie  $H_0$
- wniosek: występują istotne różnice w rozkładach zawartości procentowej wody w serze typu “A” między tygodniami

#### Sformułowanie nowego problemu w postaci hipotez

$H_0$ : mediana różnic zawartości wody w tygodniu  $i$  oraz tygodniu  $j$  dla sera “A” jest równa 0

$H_1$ : mediana różnic zawartości wody w tygodniu  $i$  oraz tygodniu  $j$  dla sera “A” jest większa od 0

#### Zapis matematyczny hipotez

$H_0: Me(X_i - X_j) = 0$

$H_1: Me(X_i - X_j) > 0$



**Przeprowadzone testy:** Test rangowanych znaków Wilcoxona (dla prób zależnych) - `wilcox.test`

1. para między tygodniem 0 a tygodniem 3

- otrzymana  $p$ -wartość:  $0.03125 < 0.05$
- decyzja: odrzucenie  $H_0$
- wniosek: mediana różnic zawartości wody między tygodniem 0 oraz 3 jest istotnie większa od 0

2. para między tygodniem 0 a tygodniem 5

- otrzymana  $p$ -wartość:  $0.03125 < 0.05$
- decyzja: odrzucenie  $H_0$
- wniosek: mediana różnic zawartości wody między tygodniem 0 oraz 5 jest istotnie większa od 0

3. para między tygodniem 3 a tygodniem 5

- otrzymana  $p$ -wartość:  $0.0625 > 0.05$
- decyzja: nie ma podstaw do odrzucenia  $H_0$
- wniosek: mediana różnic zawartości wody między tygodniem 3 oraz 5 nie różni się istotnie od 0

**Przeprowadzony test:** Porównania parami przy użyciu testu Andersona-Darlinga dla wszystkich par - `adAllPairsTest`

Tablica 7:  $p$ -wartości dla porównań parami przy użyciu testu Andersona-Darlinga

	0	3
3	0.0110358	NA
5	0.0110358	0.045156

## 2.2 Czy przechowywanie sera “B” wpływa istotnie na zmniejszenie się w nim zawartości wody?

Pomiar zawartości wody prowadzono na tych samych pięciu próbkach sera w trzech momentach czasowych, co oznacza, że dane pochodzą z układu z powtarzanymi pomiarami (próby zależne).

### Sformułowanie problemu w postaci hipotez

$H_0$ : rozkłady zawartości wody w tygodniach 0, 3 i 5 dla sera “B” są identyczne

$H_1$ : przynajmniej jeden z rozkładów różni się od pozostałych

### Zapis matematyczny hipotez

$H_0: G_{tydzien0} = G_{tydzien3} = G_{tydzien5}$

$H_1: \exists_{i \neq j} G_{tydzien i} \neq G_{tydzien j}$

Przeprowadzony test: `friedman.test`

- otrzymana  $p$ -wartość:  $0.0067379 < 0.05$
- decyzja: odrzucenie  $H_0$
- wniosek: występują istotne różnice w rozkładach zawartości procentowej wody w serze typu “B” między tygodniami

### Sformułowanie nowego problemu w postaci hipotez

$H_0$ : mediana różnic zawartości wody w tygodniu  $i$  oraz  $j$  dla sera “B” jest równa 0

$H_1$ : mediana różnic zawartości wody w tygodniu  $i$  oraz  $j$  dla sera “B” jest większa od 0

### Zapis matematyczny hipotez

$H_0: Me(Y_i - Y_j) = 0$

$H_1: Me(Y_i - Y_j) > 0$

**Przeprowadzone testy:** Test rangowanych znaków Wilcoxona (dla prób zależnych) - `wilcox.test`

1. para między tygodniem 0 a tygodniem 3

- otrzymana  $p$ -wartość:  $0.03125 < 0.05$
- decyzja: odrzucenie  $H_0$
- wniosek: mediana różnic zawartości wody między tygodniem 0 oraz 3 jest istotnie większa od 0

2. para między tygodniem 0 a tygodniem 5

- otrzymana  $p$ -wartość:  $0.0210841 < 0.05$
- decyzja: odrzucenie  $H_0$
- wniosek: mediana różnic zawartości wody między tygodniem 0 oraz 5 jest istotnie większa od 0

3. para między tygodniem 3 a tygodniem 5

- otrzymana  $p$ -wartość:  $0.03125 < 0.05$
- decyzja: odrzucenie  $H_0$
- wniosek: mediana różnic zawartości wody między tygodniem 3 oraz 5 jest istotnie większa od 0

**Przeprowadzony test:** Porównania parami przy użyciu testu Andersona-Darlinga dla wszystkich par - `adAllPairsTest`

Tablica 8:  $p$ -wartości dla porównań parami przy użyciu testu Andersona-Darlinga

	0	3
3	0.0135597	NA
5	0.0135597	0.098776

### 2.3 Czy różnica w zawartości wody w terminach tydzień0 i tydzień3 jest istotna statystycznie?

Test został wykonany na pełnym zbiorze danych, niezależnie od typu sera. Oznacza to, że wyniki nie uwzględniają potencjalnych różnic między typami “A” i “B”.

#### Sformułowanie problemu w postaci hipotez

$H_0$ : mediana różnic zawartości wody w tygodniu 0 oraz tygodniu 3 jest równa 0

$H_1$ : mediana różnic zawartości wody w tygodniu 0 oraz tygodniu 3 jest różna od 0

#### Zapis matematyczny hipotez

$H_0: Me(X_0 - X_3) = 0$

$H_1: Me(X_0 - X_3) \neq 0$

**Przeprowadzony test:** Test rangowanych znaków Wilcoxona (dla prób zależnych) - `wilcox.test`

- otrzymana  $p$ -wartość:  $0.0019531 < 0.05$
- decyzja: odrzucenie  $H_0$
- wniosek: mediana różnic zawartości wody między tygodniem 0 oraz 3 jest istotnie różna od 0

### 2.4 Czy różnica w zawartości wody w terminach tydzień3 i tydzień5 jest istotna statystycznie?

Test został wykonany na pełnym zbiorze danych, niezależnie od typu sera. Oznacza to, że wyniki nie uwzględniają potencjalnych różnic między typami “A” i “B”.

#### Sformułowanie problemu w postaci hipotez

$H_0$ : mediana różnic zawartości wody w tygodniu 3 oraz tygodniu 5 jest równa 0

$H_1$ : mediana różnic zawartości wody w tygodniu 3 oraz tygodniu 5 jest różna od 0

#### Zapis matematyczny hipotez

$H_0: Me(X_3 - X_5) = 0$

$H_1: Me(X_3 - X_5) \neq 0$

**Przeprowadzony test:** Test rangowanych znaków Wilcoxona (dla prób zależnych) - `wilcox.test`

- otrzymana  $p$ -wartość:  $0.0039063 < 0.05$
- decyzja: odrzucenie  $H_0$
- wniosek: mediana różnic zawartości wody między tygodniem 3 oraz 5 jest istotnie różna od 0

## 2.5 Jakiego średniego procentu utraty wody można się spodziewać w terminach tydzień3 i tydzień5?

Tablica 9: Średnia procentowa utrata wody od początku do tygodnia 3 oraz do tygodnia 5

	typ "A"	typ "B"
do 3 tygodnia	7.725366	6.635823
do 5 tygodnia	11.429102	11.760554

Tablica 10: Średnia procentowa utrata wody między 3 a 5 tygodniem dojrzewania sera

typ "A"	typ "B"
3.97833	5.526704

## 2.6 Dodatkowe cele badawcze

### 2.6.1 Czy występują istotne statystycznie różnice między serem typu "A" oraz serem typu "B" w poszczególnych tygodniach?

Dane dla sera typu "A" oraz typu "B" pochodzą z dwóch niezależnych prób, co oznacza, że pomiary w jednej grupie nie są w żaden sposób powiązane z pomiarami w drugiej grupie.

### Sformułowanie problemu w postaci hipotez

$H_0$ : mediany zawartości wody w serach “A” i “B” są równe w tygodniu  $i$

$H_1$ : mediany zawartości wody w serach “A” i “B” są różne w tygodniu  $i$

### Zapis matematyczny hipotez

$H_0: Me_{A_i} = Me_{B_i}$

$H_1: Me_{A_i} \neq Me_{B_i}$

**Przeprowadzony test:** Test mediany (test Mooda) - `median_test`

1. w tygodniu 0

- otrzymana  $p$ -wartość:  $0.5485062 > 0.05$
- decyzja: nie ma podstaw do odrzucenia  $H_0$
- wniosek: mediany zawartości wody w serach typu “A” oraz “B” nie różnią się w tygodniu 0

2. w tygodniu 3

- otrzymana  $p$ -wartość:  $0.5485062 > 0.05$
- decyzja: nie ma podstaw do odrzucenia  $H_0$
- wniosek: mediany zawartości wody w serach typu “A” oraz “B” nie różnią się w tygodniu 3

3. w tygodniu 5

- otrzymana  $p$ -wartość:  $0.5485062 > 0.05$
- decyzja: nie ma podstaw do odrzucenia  $H_0$
- wniosek: mediany zawartości wody w serach typu “A” oraz “B” nie różnią się w tygodniu 5

### Sformułowanie problemu w postaci hipotez

$H_0$ : rozkłady zawartości wody w serach “A” i “B” są identyczne w tygodniu  $i$

$H_1$ : rozkłady zawartości wody w serach “A” i “B” są różne w tygodniu  $i$

### Zapis matematyczny hipotez

$H_0: F_{A_i} = G_{B_i}$

$H_1: F_{A_i} \neq G_{B_i}$

**Przeprowadzony test:** Test sumy rang Wilcoxona - `wilcox.test`

1. w tygodniu 0

- otrzymana  $p$ -wartość:  $0.2222222 > 0.05$
- decyzja: nie ma podstaw do odrzucenia  $H_0$
- wniosek: rozkłady zawartości wody w serach typu “A” oraz “B” nie różnią się w tygodniu 0

2. w tygodniu 3

- otrzymana  $p$ -wartość:  $0.4633439 > 0.05$
- decyzja: nie ma podstaw do odrzucenia  $H_0$
- wniosek: rozkłady zawartości wody w serach typu “A” oraz “B” nie różnią się w tygodniu 3

3. w tygodniu 5

- otrzymana  $p$ -wartość:  $0.4019654 > 0.05$
- decyzja: nie ma podstaw do odrzucenia  $H_0$
- wniosek: rozkłady zawartości wody w serach typu “A” oraz “B” nie różnią się w tygodniu 5

### 3 Wnioski

1. Przeprowadzona analiza statystyczna wykazała, że w obu typach sera ("A" i "B") przechowywanie produktu prowadzi do statystycznie istotnych zmian w zawartości wody w czasie. Test Friedmana potwierdził istotne różnice rozkładów zawartości wody między tygodniami 0, 3 i 5 dla obu grup, co wskazuje na zmiany poziomu uwodnienia serów podczas przechowywania. Dodatkowo przeprowadzone testy Wilcoxona pozwoliły określić, między którymi tygodniami występują istotne statystycznie różnice: dla sera typu "A" były to różnice między tygodniem 0 a tygodniem 3 oraz między tygodniem 0 a tygodniem 5, natomiast dla sera typu "B" – istotność stwierdzono we wszystkich parach pomiarowych. Przeprowadzono również porównania parami przy użyciu testu Andersona–Darlinga, które wykazały istotne różnice pomiędzy wszystkimi punktami czasowymi dla sera typu "A". W przypadku sera typu "B" różnice pomiędzy tygodniem 3 a 5 nie były istotne statystycznie.
2. Przeprowadzony test rangowanych znaków Wilcoxona dla prób zależnych (bez rozróżniania typu sera) wykazał, że zmiany zawartości wody między tygodniami 0 i 3 oraz między tygodniami 3 i 5 są istotne statystycznie, co potwierdza różnice w uwodnieniu sera w czasie.
3. Analiza średnich procentowych spadków zawartości wody wskazuje, że po 3 tygodniach ser typu "A" traci przeciętnie około 7,7% wody, a po 5 tygodniach jest to już 11,4%. Natomiast ser typu "B" po 3 tygodniach od początku badania stracił około 6,6%, a po 5 tygodniach - 11,8%. W okresie między tygodniem 3 a 5 tempo odwodnienia w serze "B" było wyższe niż w serze "A".
4. Dodatkowe porównania między typami sera w konkretnym tygodniu wykonane za pomocą testu mediany oraz testu Wilcoxona dla prób niezależnych nie wykazały istotnych różnic w zawartości wody. Oznacza to, że pomimo wyraźnej tendencji odwodnienia w czasie, nie stwierdzono istotnych różnic między typami sera w odpowiadających sobie punktach pomiarowych.