Praca_domowa_2

June 1, 2025

1 Praca domowa nr 2 „Planowanie i analiza eksperymentu’’

wykonał Paweł Jan Tłusty s95596 Zestaw zadań dla grupy 1.1

Wersja online: Notes dostępny również w serwisie GitHub (niektóre wykresy niewłaściwie dziedziczą style, zaleca zaleca się jasny motyw lub otwieranie poszczególnych wykresów jako obraz w nowej karcie)

Link:

https://github.com/Kotmin/R-Planowanie-Eksperymentu

Dla wszystkich zadań, o ile nie zaznaczono inaczej, przyjmujemy domyślny poziom istotności = **0.05**. W przypadkach, gdzie zostaje on zmieniony (np. = 0.01), informujemy o tym lokalnie przy interpretacji wyników.

1.1 Zadanie 1: Porównanie średnich wag mandarynek z dwóch plantacji

Dane:

- Waga1 = c(75, 67, 73, 78, 70, 78, 84, 75, 70, 72, 78)
- Waga2 = c(80, 75, 82, 76, 78, 82, 80, 85, 76, 72)

Cel: Sprawdzić, czy średnia waga mandarynek z dwóch plantacji jest taka sama.

Hipotezy:

- **H (hipoteza zerowa):** Średnia waga mandarynek z plantacji 1 = średnia waga z plantacji 2
- **H (hipoteza alternatywna):** Średnie są różne

[16]:
```
waga1 <- c(75, 67, 73, 78, 70, 78, 84, 75, 70, 72, 78)
waga2 <- c(80, 75, 82, 76, 78, 82, 80, 85, 76, 72)
```

[17]:
```
length(waga1); length(waga2) # Rozmiary prób
```

11

10

[18]:
```
mean(waga1); var(waga1)
mean(waga2); var(waga2)
```

74.5454545454545

23.2727272727273

78.6

15.3777777777778

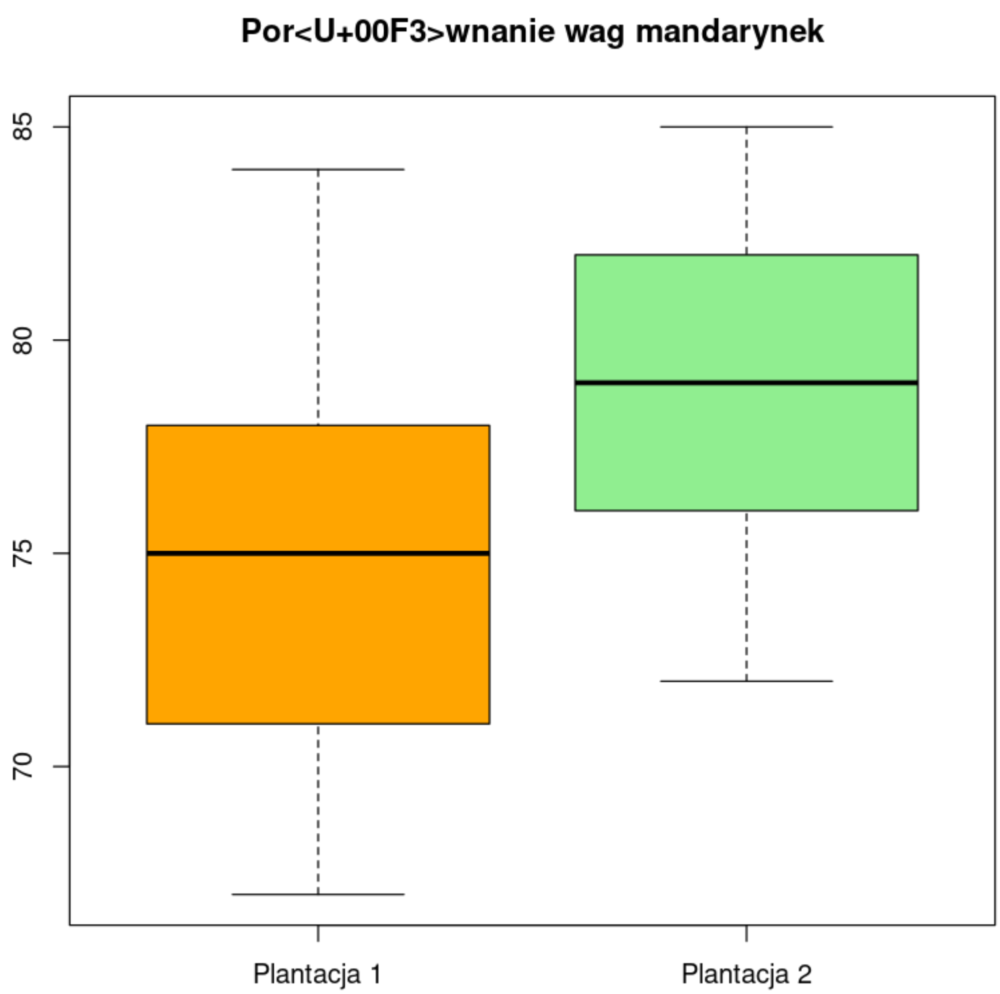
[19]:
```
waga1
```

1. 75 2. 67 3. 73 4. 78 5. 70 6. 78 7. 84 8. 75 9. 70 10. 72 11. 78

[20]:
```
waga2
```

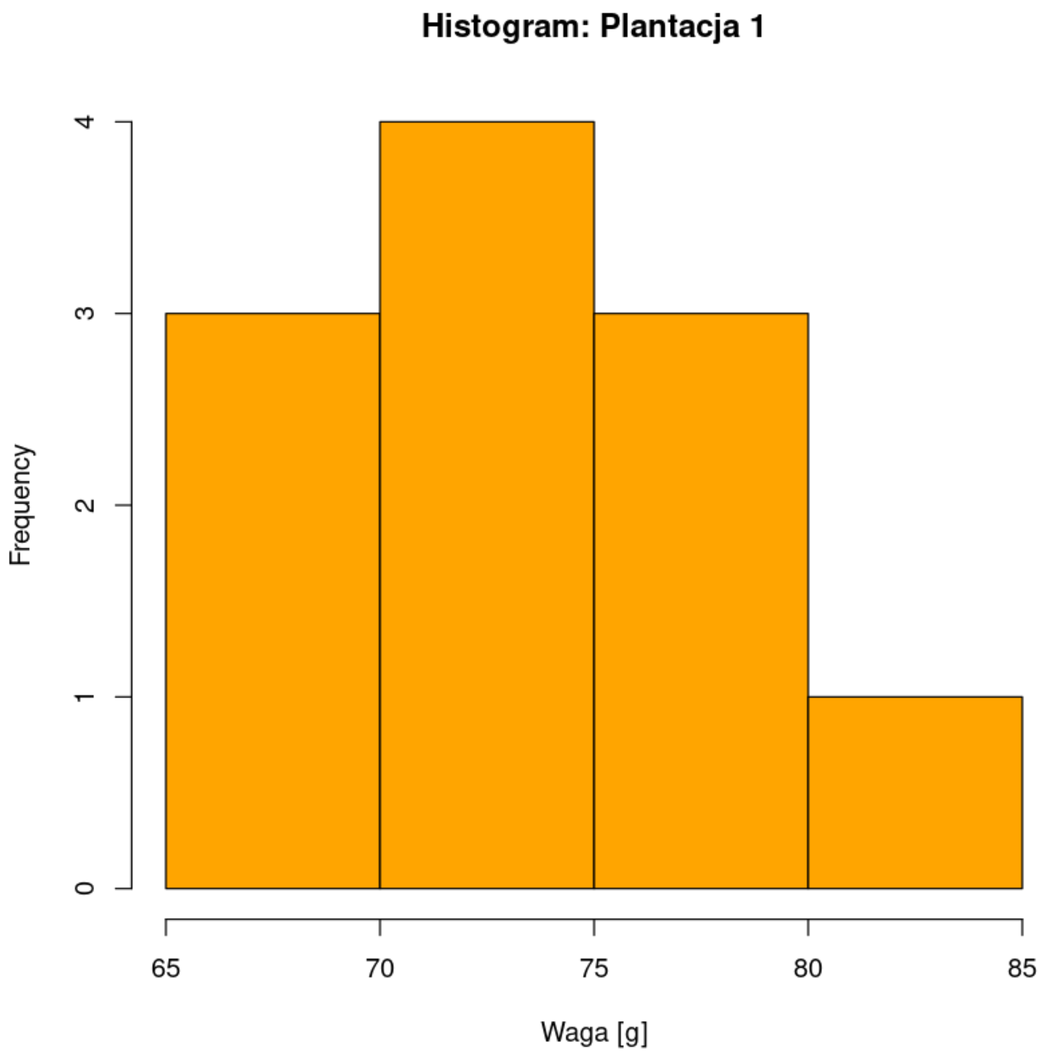
1. 80 2. 75 3. 82 4. 76 5. 78 6. 82 7. 80 8. 85 9. 76 10. 72

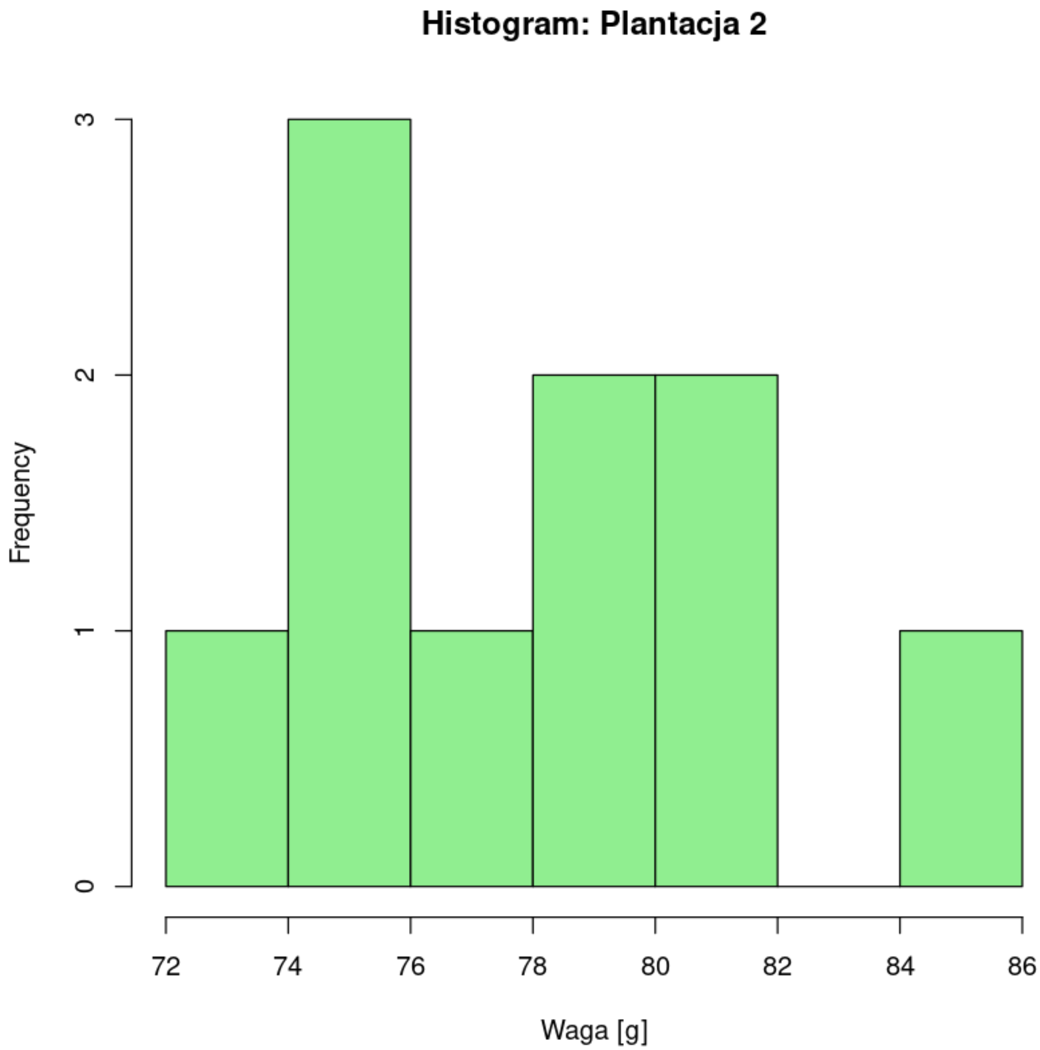
[21]:
```
boxplot(waga1, waga2, names = c("Plantacja 1", "Plantacja 2"), col =␣
↪c("orange", "lightgreen"), main = "Porównanie wag mandarynek")
```



[22]:
```
# par(mfrow = c(2, 1))
```

[23]:
```
hist(waga1, main = "Histogram: Plantacja 1", col = "orange", xlab = "Waga [g]")
hist(waga2, main = "Histogram: Plantacja 2", col = "lightgreen", xlab = "Waga␣
↪[g]")
```





[26]:
```
# sprawdźimy czy dane mają rozkład normalny
# używamy testu Shapiro-Wilka (próbki są małe)
# test SW jest skonstruowany z hipotezami:
# H0: dane pochodzą z rozkładu normalnego
# H1: dane nie pochodzą z rozkładu normalnego / dane są istotne różne od
↪rozkładu normalnego

shapiro.test(waga1)
shapiro.test(waga2)
```

```
	Shapiro-Wilk normality test
```

```
data:  waga1
W = 0.96419, p-value = 0.8227


        Shapiro-Wilk normality test

data:  waga2
W = 0.97536, p-value = 0.9356
```

[29]:
```r
# p-value większe od 0.05 --dla obu przypadków.
# nie znaleźliśmy podstaw do odrzucenia hipotezy zerowej. Dane pochodzą z␣
 ↪rozkładu normalnego
# gdyby dane nie pochodziły z r.norm. użylibyśmy zapewne testu wilcoxona dla␣
 ↪średnich, dla zgodności z rozkładem normalnym użyjemy
# dobranej wersji t-test
# w celu doboru sprawdźmy wariancję
var.test(waga1, waga2)
# h0 wariancje są równe
# h1 wariancje się różnią
```

```
        F test to compare two variances

data:  waga1 and waga2
F = 1.5134, num df = 10, denom df = 9, p-value = 0.545
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.381799 5.719082
sample estimates:
ratio of variances
            1.5134
```

[30]:
```r
# wariancje są znane, nie znaleźliśmy podstaw do odrzucenia hipotezy zerowej.␣
 ↪Wariancje nie różnią się istotnie statycznycznie
t.test(waga1, waga2, var.equal = TRUE)
```

```
        Two Sample t-test

data:  waga1 and waga2
t = -2.0996, df = 19, p-value = 0.04935
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -8.09632069 -0.01277022
sample estimates:
```

```
mean of x mean of y
 74.54545  78.60000
```

[ ]:
```
# p-value mniejsze od 0.05 - odrzucamy hipotezę zerową
```

1.1.1 Podsumowanie Z1: Porównanie średniej wagi mandarynek z dwóch plantacji

Czy można twierdzić, że średnia waga mandarynek jest taka sama dla każdej z plantacji?

– Nie.

Hipotezy:

- Hipoteza zerowa (H0): średnia waga mandarynek jest taka sama dla obu plantacji (mu1 = mu2)
- Hipoteza alternatywna (H1): średnia waga mandarynek różni się (mu1   mu2)

Założenia:

- Próbki są niezależne – pochodzą z dwóch różnych plantacji
- Dane mają rozkład normalny (sprawdzone testem Shapiro-Wilka)
- Wariancje są równe (sprawdzone testem Fishera)

Wyniki testu t:

- Statystyka t: -2.0996
- p-value: 0.04935
- Przedział ufności 95%: [-8.10, -0.01]
- Średnia (plantacja 1): 74.55 g
- Średnia (plantacja 2): 78.60 g

Wniosek: Na poziomie istotności alfa = 0.05, odrzucamy hipotezę H0 – istnieje statystycznie istotna różnica średnich wag mandarynek między plantacjami.

p-value bardzo bliskie 0.05 → wynik interpretować ostrożnie.

1.1.2 Zadanie 2:

Polecenie:

Pewna firma farmaceutyczna przetestowała działanie nowego leku na losowo wybranych pacjentach. Celem leku jest obniżenie poziomu pewnego składnika X we krwi. Wyniki badania przed podaniem i po podaniu leku są następujące (w odpowiednich jednostkach dla ustalonej objętości krwi)

Sprawdzić, czy średnio poziom składnika X spada po podaniu leku.
Czy można wnioskować, że lek obniża średnio poziom X ?

Dane:
- przed = c(160, 205, 230, 245, 180, 280, 230, 200, 170, 210)
- po   = c(150, 210, 240, 230, 170, 260, 240, 180, 190, 200)

Hipotezy: - H : _d = 0 (brak zmiany średniego poziomu X) - H : _d > 0 (lek **obniża** poziom składnika X)

Z konstukcji zadania wynika, że wartości na kolejnych pozycjach odpowiadają sobie wzajemnie przed/po.

[31]:
```r
przed <- c(160, 205, 230, 245, 180, 280, 230, 200, 170, 210)
po    <- c(150, 210, 240, 230, 170, 260, 240, 180, 190, 200)
```

[32]:
```r
przed
```

1. 160 2. 205 3. 230 4. 245 5. 180 6. 280 7. 230 8. 200 9. 170 10. 210

[33]:
```r
po
```

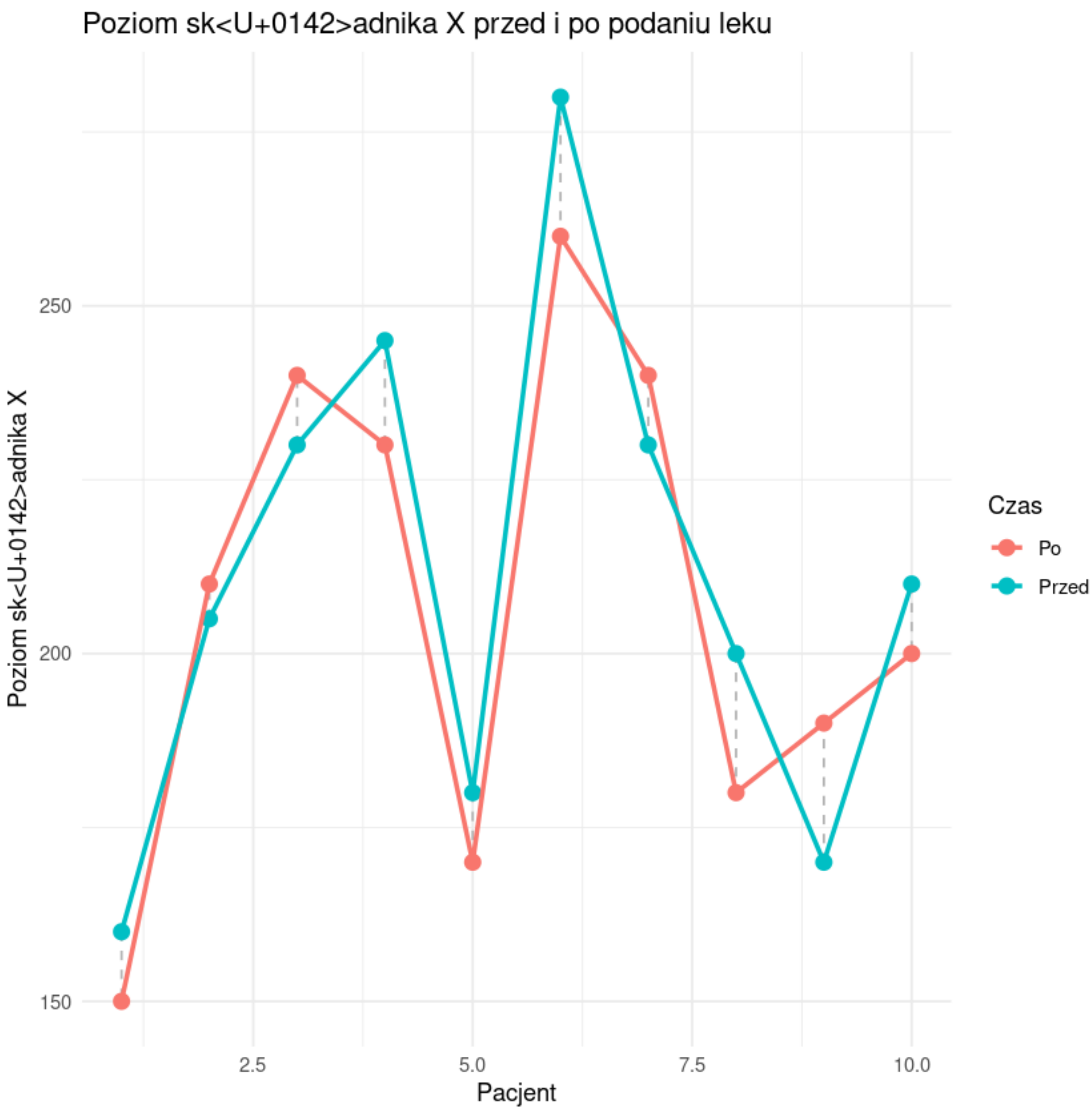
1. 150 2. 210 3. 240 4. 230 5. 170 6. 260 7. 240 8. 180 9. 190 10. 200

[36]:
```r
# chcemy zwizualicować dane
pacjenci <- 1:length(przed)
df <- data.frame(
  Pacjent = pacjenci,
  Przed = przed,
  Po = po
)

# Długa wersja danych
library(tidyr)
df_long <- pivot_longer(df, cols = c(Przed, Po), names_to = "Czas", values_to = "Wartosc")

# Wykres
library(ggplot2)
ggplot(df_long, aes(x = Pacjent, y = Wartosc, color = Czas, group = Czas)) +
  geom_line(aes(group = Pacjent), color = "gray70", linetype = "dashed") +
  geom_point(size = 3) +
  geom_line(size = 1) +
  labs(title = "Poziom składnika X przed i po podaniu leku",
       x = "Pacjent",
       y = "Poziom składnika X",
       color = "Czas") +
  theme_minimal()
```

```
Warning message:
"Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead."
```



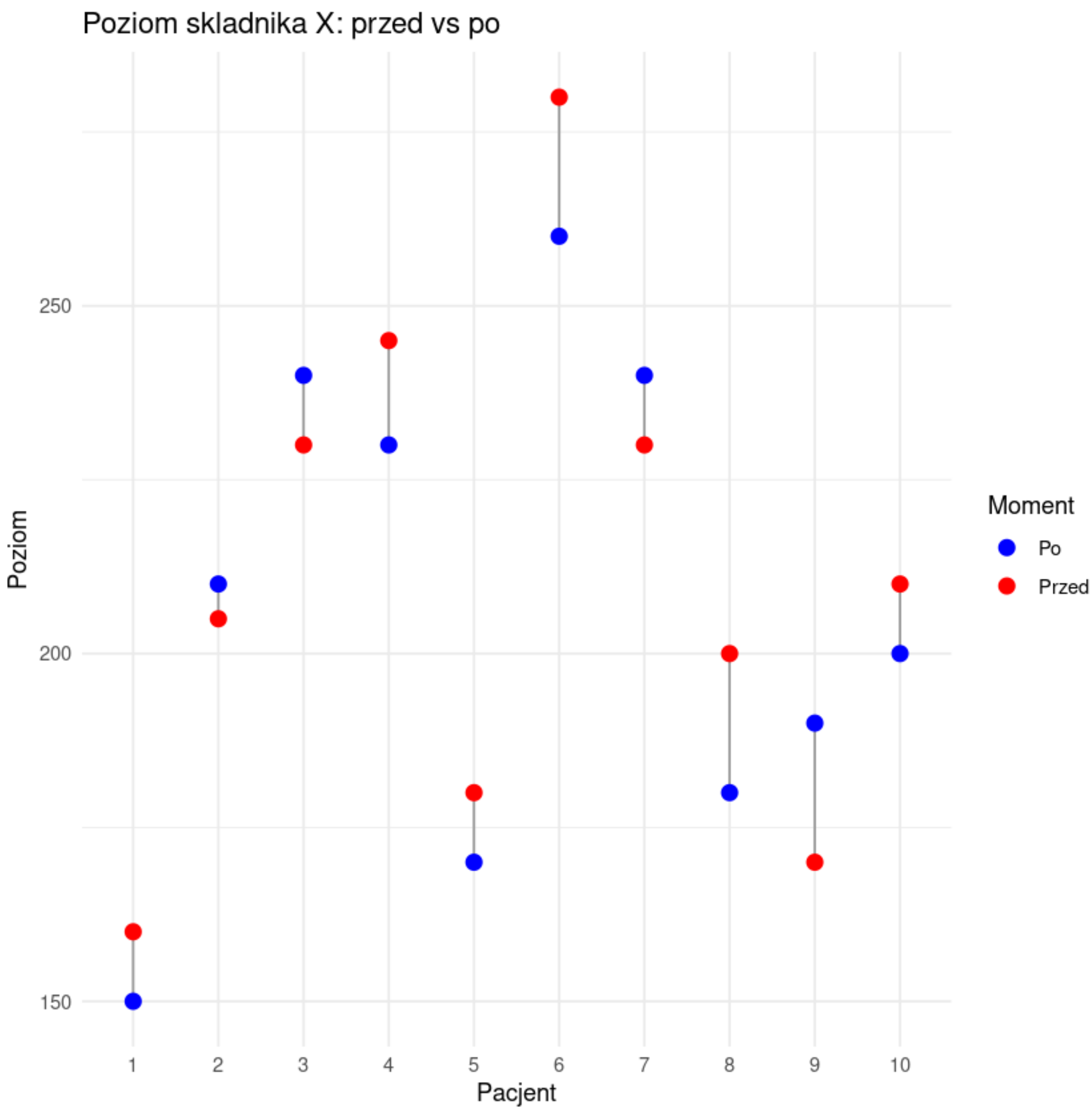
[38]:
```
# wykres liniowy zdaje się być tutaj niepoprawny. Chcieliśmy zobaczyć czy␣
↪kształt zmian będzie miał podobny przebieg.
# bardziej odpowiednia wizualizacja ponizej
```

[37]:
```
library(ggplot2)

# Ramka danych
df <- data.frame(
  Pacjent = factor(1:length(przed)),
  Przed = przed,
  Po = po
)
```

```r
# Długa wersja do wykresu
library(tidyr)
df_long <- pivot_longer(df, cols = c(Przed, Po), names_to = "Moment", values_to = "Poziom")

# Wykres z liniami łączącymi przed i po
ggplot(df_long, aes(x = Pacjent, y = Poziom, group = Pacjent)) +
  geom_line(color = "gray60") +
  geom_point(aes(color = Moment), size = 3) +
  scale_color_manual(values = c("Przed" = "red", "Po" = "blue")) +
  labs(
    title = "Poziom skladnika X: przed vs po",
    x = "Pacjent",
    y = "Poziom",
    color = "Moment"
  ) +
  theme_minimal()
```



[39]:
```
# Czy lek obniża X po zażyciu: wykres pokazuje, ze zachodzą takie przypadki.
↪Należy to jednak zweryfikować

roznice <- przed - po
```

[40]:
```
shapiro.test(roznice)
```

```
	Shapiro-Wilk normality test

data:  roznice
W = 0.89467, p-value = 0.1913
```

[41]:
```r
# poruszamy się w obszarze testów dla  dwóch średnich, w tym przypadku zależnych
# nie znaleźliśmy testem Shapiro-Wilka by różnice miały statystycznie istotną␣
↪różnicę w dystrybucji od rozkładu nromalnego
# wobec tego możemy użyć t.test, jednostronny (szukamy odpowiedzi na pytanie␣
↪przed > po)

t.test(przed, po, paired = TRUE, alternative = "greater")
```

```
        Paired t-test

data:  przed and po
t = 0.89692, df = 9, p-value = 0.1966
alternative hypothesis: true mean difference is greater than 0
95 percent confidence interval:
 -4.175127       Inf
sample estimates:
mean difference
              4
```

[ ]:
```r
# p-value jest większe od 0.05 wobec tego nie mamy podstawy do odrzucenia␣
↪hipotezy zerowej.
#  Wobec czego w mocy utrzymuje się H0: Lek nie obniża poziomu składnika X
```

1.1.3 Podsumowanie Z2: Czy lek obniża poziom składnika X

Czy można wnioskować, że lek obniża średnio poziom X?

– Nie.

Hipotezy:

- Hipoteza zerowa (H0): Średnia różnica   0 (lek nie obniża poziomu składnika X)
- Hipoteza alternatywna (H1): Średnia różnica > 0 (lek obniża poziom składnika X)

Założenia:

- Dane pochodzą od tych samych pacjentów (próby zależne)
- Różnice mają rozkład normalny (sprawdzone testem Shapiro-Wilka)

Wyniki testu t (dla prób zależnych, jednostronny):

- Statystyka t: 0.8969
- p-value: 0.1966
- Przedział ufności 95%: [–4.18, ∞)

- Średnia różnica (przed - po): 4 jednostki #### Wyniki testu t:
- Statystyka t: -2.0996
- p-value: 0.04935
- Przedział ufności 95%: [-8.10, -0.01]
- Średnia (plantacja 1): 74.55 g
- Średnia (plantacja 2): 78.60 g

Wniosek: Wnioskujemy, że nie ma statystycznie istotnych dowodów na to, że lek obniża poziom składnika X we krwi.

[ ]:
```

```

[ ]:
```

```