

Zajęcia 6

Testy istotności różnic

Zadanie 1. Zbiór danych `rat` z pakietu `UsingR` zawiera czasy przeżycia szczurów poddanych promieniowaniu. Chcemy zbadać, jaki jest średni czas życia szczurów poddawanych promieniowaniu. Przyjmując rozkład normalny badanej cechy, przetestuj hipotezę, że średni czas przeżycia wynosi 120. Jaka jest sensowna alternatywa? **Wskazówka:** `mean()`, `t.test()`

One Sample t-test

```
data: rat
t = -0.8185, df = 19, p-value = 0.2116
alternative hypothesis: true mean is less than 120
95 percent confidence interval:
 -Inf 127.2879
sample estimates:
mean of x
 113.45
```

Zadanie 2. Załaduj zbiór danych `homework` z pakietu `UsingR`, który zawiera informacje na temat ilości czasu poświęcanego na wykonywanie zadań domowych przez uczniów prywatnych oraz publicznych liceów. Naszym celem jest zbadanie czy uczniowie obu typów liceów poświęcają tyle samo czasu na wykonywania zadań domowych. Przyjmujemy rozkład normalny jako rozkład teoretyczny badanej cechy.

- (a) Zwizualizuj te dane na jednym wykresie pudełkowym. **Wskazówka:** `boxplot()` **Odpowiedź:** patrz Rysunek 1
- (b) Jaka postać testu t, o ile, jest odpowiednia dla tych danych? **Wskazówka:** `var.test()`

F test to compare two variances

```
data: homework$Private and homework$Public
F = 0.8194, num df = 14, denom df = 14, p-value = 0.7146
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.275110 2.440771
sample estimates:
ratio of variances
 0.8194392
```

- (c) Używając odpowiedniego testu zbadaj, czy rzeczywiście poświęcają tyle samo czasu na wykonywanie zadań domowych. **Wskazówka:** `t.test()`

Two Sample t-test

```
data: homework$Private and homework$Public
t = 1.7134, df = 28, p-value = 0.04884
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.01957252      Inf
sample estimates:
mean of x mean of y
 17.63333  14.90667
```

Zadanie 3. Producent proszku do prania A twierdzi, że jego produkt jest istotnie lepszy od konkurencyjnego proszku B. W celu weryfikacji tego zapewnienia CTA (Consumer Test Agency) przetestowała oba proszki do prania. W tym celu wykonano pomiary stopnia wyprania 7 wycinków tkaniny proszkiem A otrzymując wyniki (w %): 78.2, 78.5, 75.6, 78.5, 78.5, 77.4, 76.6, oraz 10 wycinków tkaniny proszkiem B otrzymując wyniki (w %): 76.1, 75.2, 75.8, 77.3, 77.3, 77.0, 74.4, 76.2, 73.5, 77.4. Jaki powinien być wniosek CTA dotyczący jakości tych proszków? Dokonaj wnioskowania na poziomie istotności 0,05.

- (a) Zilustruj różnice w stopniu wyprania tkanin w testowanych proszkach do prania przy pomocy wykresu ramkowego. **Wskazówka:** `boxplot()` **Odpowiedź:** patrz Rysunek 2
- (b) Przyjmując rozkład normalny badanej cechy, wykonaj odpowiednią procedurę testową, sprawdzając wcześniej jej założenia. **Wskazówka:** `var.test()`, `t.test()`

F test to compare two variances

```
data: dane[proszek == "A", ]$stopien and dane[proszek == "B", ]$stopien
F = 0.7397, num df = 6, denom df = 9, p-value = 0.7368
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1712288 4.0854482
sample estimates:
ratio of variances
      0.7396609
```

Two Sample t-test

```
data: dane[proszek == "A", ]$stopien and dane[proszek == "B", ]$stopien
t = 2.5735, df = 15, p-value = 0.01059
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.5082596      Inf
sample estimates:
mean of x mean of y
77.61429 76.02000
```

Zadanie 4. Zbiór danych `corn` z pakietu `UsingR` zawiera informacje na temat plonów kukurydzy. 12 poletek zostało podzielonych na dwie równe części, na jednej posadzono nową odmianę kukurydzy, a na drugiej starą. Chcemy zbadać czy nowa odmiana daje lepsze plony.

- (a) Przyjmując rozkład normalny badanej cechy, wykonaj test t dla dwóch prób niezależnych na tych danych. **Wskazówka:** `t.test()`

Two Sample t-test

```
data: corn$New and corn$Standard
t = 1.8061, df = 22, p-value = 0.08461
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-1.14919 16.64919
sample estimates:
mean of x mean of y
100.83333 93.08333
```

- (b) Czy wszystkie założenia są spełnione? **Wskazówka:** `var.test()`

F test to compare two variances

```
data: corn$New and corn$Standard
```

```
F = 1.0273, num df = 11, denom df = 11, p-value = 0.9652
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2957413 3.5685878
sample estimates:
ratio of variances
      1.027316
```

- (c) Dlaczego, w tym przypadku, właściwy wydaje się test dla prób połączonych (zależnych)? Wykonaj go. Czy wyniki się pokrywają? **Wskazówka:** `t.test()`

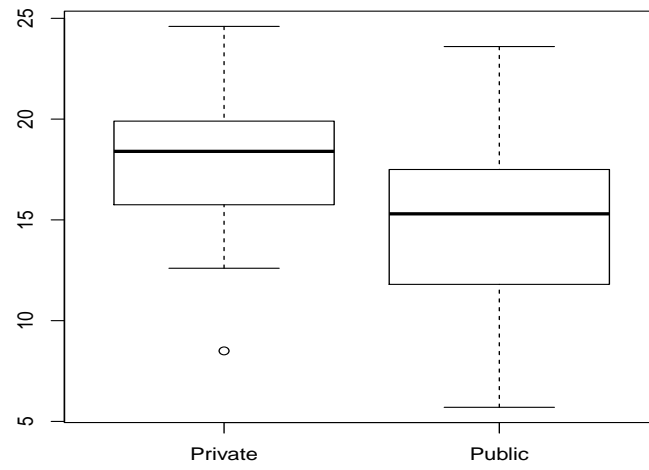
```
Paired t-test

data:  corn$New and corn$Standard
t = 3.8308, df = 11, p-value = 0.00279
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 3.297258 12.202742
sample estimates:
mean of the differences
      7.75
```

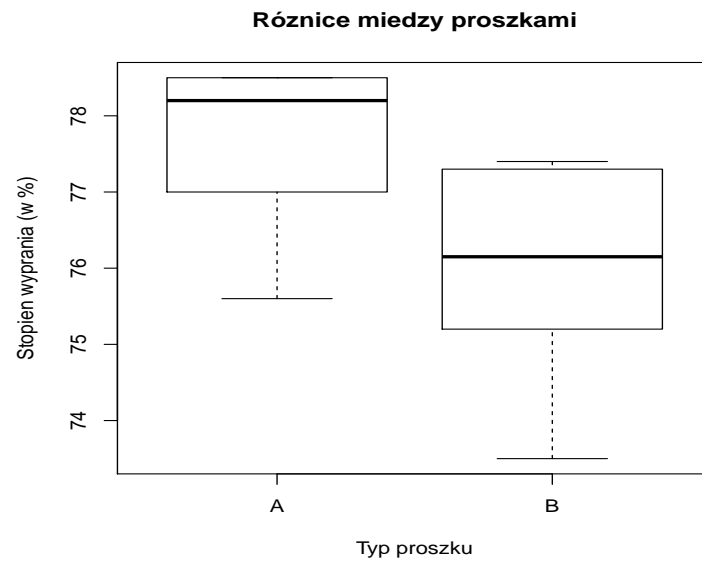
Zadanie 5. Przeprowadzono badania mające na celu zbadanie wpływu hipnozy na redukcję bólu. Poziom odczuwanego bólu wynosił: przed hipnozą 6.6, 6.5, 9.0, 10.3, 11.3, 8.1, 6.3, 11.6, po hipnozie 6.8, 2.5, 7.4, 8.5, 8.1, 6.1, 3.4, 2.0. Przyjmujemy rozkład normalny jako rozkład teoretyczny poziom odczuwanego bólu. Czy na poziomie istotności 0,05 można twierdzić, że hipnoza redukuje odczuwany ból? **Wskazówka:** `t.test()`

```
Paired t-test

data:  a and b
t = 3.0285, df = 7, p-value = 0.009577
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 1.165386      Inf
sample estimates:
mean of the differences
      3.1125
```



RYSUNEK 1. Wykres do Zadania 2 (a).



RYSUNEK 2. Wykres do Zadania 3 (a).