Sprawozdanie 2. Grupa 3. Testowanie hipotez statystycznych

Emil Olszewski, Jakub Kempa

1. Wprowadzenie

Będziemy testować hipotezy statystyczne na poziomie istotności $\alpha = 0.05$. Przestestujemy hipotezę

$$H_0 : \mu = 1$$

przeciwko

$$H_1 : \mu \neq 1$$

Wykorzystamy następujące testy:

- **test z** przy założeniu $\sigma = 2$
- **test t-Studenta**
- **test rang znakowanych Wilcoxona**

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.3     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.4.3     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.0
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

1.1 Wstęp do hipotezy zerowej

Testowanie hipotezy statystycznej $H_0 : \mu = 1$ na poziomie istotności $\alpha = 0.05$, to procedura, która ma na celu ocenę czy średnia wartość z analizowanych danych μ różni się od przyjętej wartości 1. Poziom istotności $\alpha = 0.05$ oznacza, że jesteśmy w stanie zaakceptować 5% ryzyko, że popełnimy błąd pierwszego rodzaju. Rozróżnia się dwa takie błędy:

- Błąd pierwszego rodzaju - fałszywe odrzucenie hipotezy zerowej
  Wiąże się on z nieprawidłowym odrzuceniem hipotezy mówiącej w naszym przypadku o tym, że średnia jest równa 1
- Błąd drugiego rodzaju - fałszywe zaakceptowanie hipotezy zerowej
  W tym przypadku nie odrzucamy hipotezy zerowej, mimo że powinniśmy

Badając hipotezę zerową będziemy korzystali z wyżej wymienionych testów. Procedura prezentuje się następująco:

1. Generujemy 100 wartości z zadanego rozkładu.
2. Przeprowadzamy testy statystyczne dla wygenerowanego zestawu danych.
3. Liczymy empiryczną moc testu (prawdopodobieństwo odrzucenia hipotezy zerowej).
4. Powtarzamy kroki 1. - 3. $MC = 1000$ razy.
5. Rysujemy wykres funkcji mocy w zależności od μ na zadanym przedziale.

1.2 Teoria

Moc testu statystycznego mierzy jego zdolność do wykrywania rzeczywistego efektu lub różnicy, gdy istnieje. Jest to w zasadzie prawdopodobieństwo poprawnego odrzucenia fałszywej hipotezy zerowej. Kilka czynników wpływa na moc testu:

1. **Wielkość efektu**: Wielkość różnicy lub efektu, który jest testowany. Większe efekty są łatwiejsze do wykrycia i prowadzą do większej mocy.
2. **Wielkość próby**: Liczba obserwacji lub uczestników badania. Większe próby zazwyczaj prowadzą do większej mocy, ponieważ dostarczają więcej informacji i zmniejszają efekty losowej zmienności.
3. **Poziom istotności (α)**: Próg określający istotność statystyczną. Obniżenie poziomu istotności zmniejsza prawdopodobieństwo popełnienia błędu pierwszego rodzaju (fałszywy wynik pozytywny), ale również zmniejsza moc testu.
4. **Zmiennosć lub odchylenie standardowe**: Stopień rozrzutu danych. Testy są bardziej efektywne, gdy zmienność jest mniejsza.

5. **Typ testu**: Różne testy statystyczne mają różne właściwości mocy. Wybór odpowiedniego testu dla pytania badawczego może wpłynąć na moc analizy.
6. **Założenia**: Naruszenie założeń testu statystycznego może zmniejszyć jego moc. Zapewnienie, że dane spełniają założenia wybranego testu, może pomóc maksymalizować moc.

Ogólnie rzecz biorąc, maksymalizacja mocy polega na równoważeniu tych czynników w celu uzyskania testu wystarczająco wrażliwego do wykrycia istotnych efektów, jednocześnie kontrolując błędy pierwszego i drugiego rodzaju.

2. Zadanie 1.

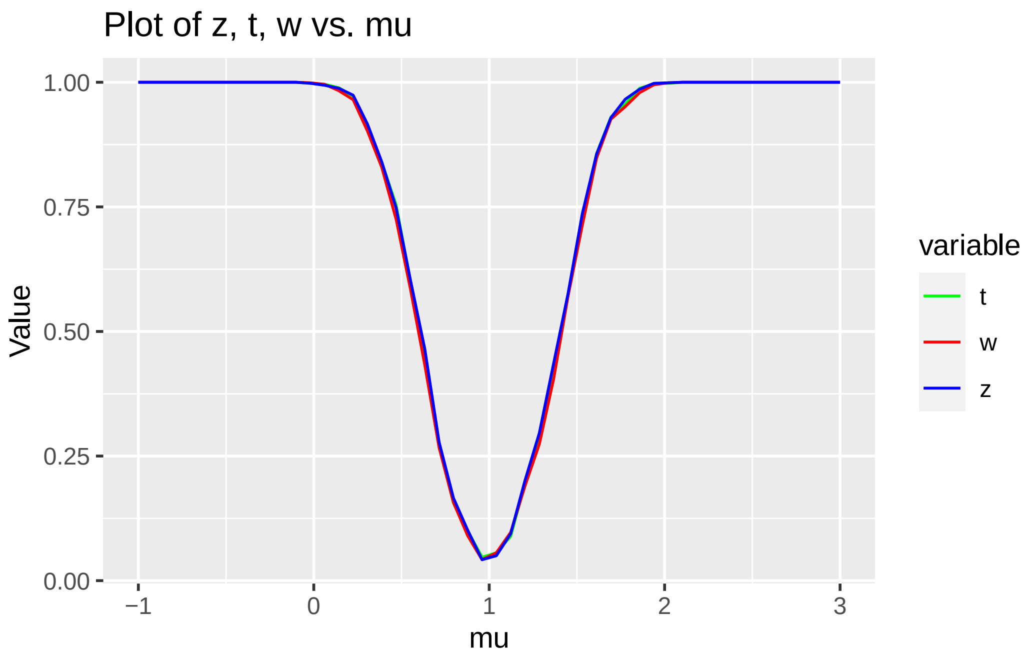
Treść zadania: *Rozważmy próbę* $(X_1, ..., X_100)$ *z rozkładu normalnego* $\mathcal{N}(\mu, 2^2)$*. Korzystając z symulacji Monte Carlo wykonaj wykres funkcji mocy w zależności od* μ *na przedziale* $(-1, 3)$ *dla wszystkich trzech testów. Czy istnieje test jednostajnie najmocniejszy spośród nich?*

Zatem, tak jak wspomniane w sekcji 1.1

1. Generujemy 100 wartości z rozkładu normalnego $\mathcal{N} \sim (\mu, 2^2)$
2. Przeprowadzamy testy statystyczne dla wygenerowanego zestawu danych.
3. Liczymy empiryczną moc testu (prawdopodobieństwo odrzucenia hipotezy zerowej).
4. Powtarzamy kroki 1. - 3. $MC = 1000$ razy.
5. Rysujemy wykres funkcji mocy w zależności od μ na przedziale $(-1, 3)$.

Wykres prezentuje się następująco:

```
plot_test_power("norm", 100, 2, seq(-1, 3, length.out=50), MCS=1000)
```



Jak widzimy wykresy wszystkich trzech testów pokrywają się ze sobą. Sugeruje to, że wszystkie trzy testy mają podobne możliwości wykrywania prawdziwego efektu lub różnicy. Może to oznaczać, że każdy z testów jest równie odpowiedni dla danego scenariusza i że ich wyniki są porównywalne pod względem mocy statystycznej. Zatem z podanych parametrów nie istnieje test jednostajnie najmocniejszy spośród nich.

3. Zadanie 2.

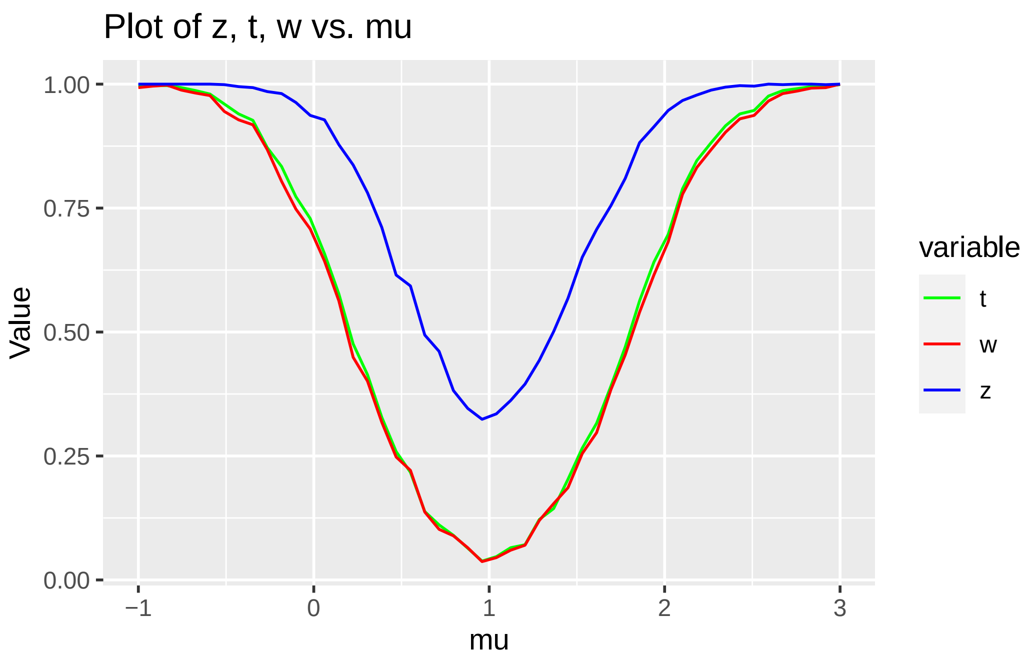
Treść zadania: *Rozważmy próbę* $(X_1, ..., X_100)$ *z rozkładu normalnego* $\mathcal{N}(\mu, 4^2)$. *Wykonaj wykres funkcji mocy na wybranym przedziale zawierającym przynajmniej po jednym punkcie z hipotezy zerowej i alternatywnej. Jak zmieniła się funkcja mocy testów? Czy w tym przypadku istnieje test jednostajnie najmocniejszy spośród nich?*

Zatem:

1. Generujemy 100 wartości z rozkładu normalnego $\mathcal{N}(\mu, 2^2)$
2. Przeprowadzamy testy statystyczne dla wygenerowanego zestawu danych.
3. Liczymy empiryczną moc testu (prawdopodobieństwo odrzucenia hipotezy zerowej).
4. Powtarzamy kroki 1. - 3. $MC = 1000$ razy.
5. Rysujemy wykres funkcji mocy w zależności od μ na wybranym przez nas przedziale (ponownie rozważamy przedział $(-1, 3)$).

Wykres prezentuje się następująco:

```
plot_test_power("norm", 100, 4, seq(-1, 3, length.out=50), MCS=1000)
```



Funkcje mocy testów zmieniły się:

- Wykresy funkcji mocy testu z oraz testu rang znakowanych Wilcoxona ponownie są do siebie bardzo zbliżone, co ponownie sugeruje podobne możliwości wykrywania prawdziwego efektu lub różnicy.
- Wykres funkcji mocy testu t-studenta odbiega od funkcji mocy pozostałych dwóch testów. Jego wygląd sugeruje, że moc testu t-studenta jest większa od pozostałych dwóch testów dla zadanych parametrów; test w większą skutecznością jest w stanie wykryć fałszywą hipotezę zerową, niż pozostałe dwa testy.

Na bazie wyglądu wykresów funkcji mocy testów możemy stwierdzić, że w tym przypadku testem jednostajnie najmocniejszym będzie test t-studenta.

4. Zadanie 3.

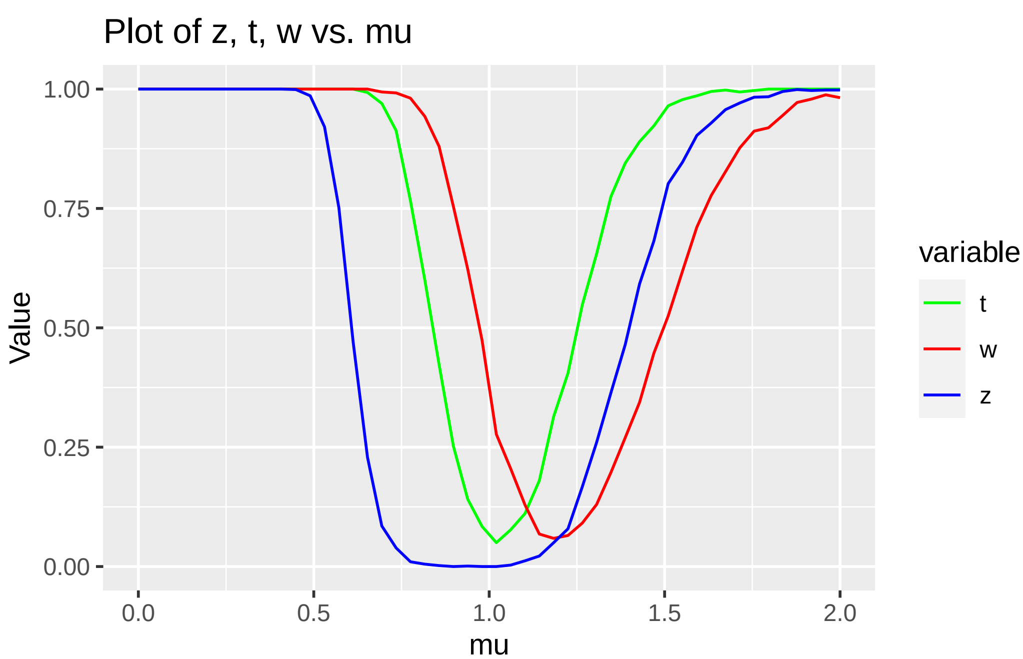
Treść zadania: *Rozważmy próbę* $(X_1, ..., X_100)$ *z rozkładu wykładniczego* $\mathcal{E}(\frac{1}{\mu})$. *Wykonaj wykres funkcji mocy na wybranym przedziale zawierającym przynajmniej po jednym punkcie z hipotezy* *zerowej i alternatywnej. Jak zmieniła się funkcja mocy testów? Czy w tym przypadku istnieje test jednostajnie najmocniejszy spośród nich?*

Zatem:

1. Generujemy 100 wartości z rozkładu normalnego $\mathcal{E}(\frac{1}{\mu})$
2. Przeprowadzamy testy statystyczne dla wygenerowanego zestawu danych.
3. Liczymy empiryczną moc testu (prawdopodobieństwo odrzucenia hipotezy zerowej).
4. Powtarzamy kroki 1. - 3. $MC = 1000$ razy.
5. Rysujemy wykres funkcji mocy w zależności od μ na przedziale $(0, 2)$.

Wykres prezentuje się następująco:

```
plot_test_power("exp", 100, NULL, seq(0, 2, length.out=50), MCS=1000)
```



Ponownie patrzymy na zachowanie funkcji dla $\mu = \mu_0$, przy czym $\mu_0 = 1$. Dla takich parametrów widzimy trzy różne wyniki:

- Najmocniejszym testem do testowania hipotezy zerowej jest test rang znakowanych Wilcoxona, gdyż dla $\mu = \mu_0$ osiąga on największą wartość. Ten sam test jest jednak narażony na duży poziom błędów I-go rodzaju.

- Test t-studenta jest słabszy, niż test rang znakowanych Wilcoxona dla testowania hipotezy zerowej, jednak ma ze wszystkich trzech testów najniższy poziom błędów I-go rodzaju.
- Test z to najsłabszy z testów dla podanych parametrów: Niska moc testu dla $\mu = \mu_0$ oraz większy, niż dla t-studenta poziom błędów I-go rodzaju sugeruje niską użyteczność dla danych o zakładanych parametrach.