

# Sprawozdanie 2

Antczak Jakub, Curkowicz Kinga

10 lutego 2024

## Spis treści

<b>1</b>	<b>Wprowadzenie</b>	<b>1</b>
<b>2</b>	<b>Opisy testów</b>	<b>1</b>
2.1	Test $t$ -studenta . . . . .	1
2.2	Test $t$ -Welcha . . . . .	2
2.3	Test sumy rang Wilcoxona . . . . .	2
2.4	Przykładowe wywołanie w R dla zadania 1 . . . . .	3
<b>3</b>	<b>Zadania</b>	<b>4</b>
3.1	Zadanie 1 . . . . .	5
3.2	Zadanie 2 . . . . .	6
3.3	Zadanie 3 . . . . .	8

## 1 Wprowadzenie

Niniejsze sprawozdanie stanowi analizę porównawczą wybranych testów statystycznych. Będą to:

- test  $t$ -studenta przy założeniu, że  $\sigma_1 = \sigma_2$ ,
- test  $t$ -Welcha,
- test sumy rang Wilcoxona.

## 2 Opisy testów

### 2.1 Test $t$ -studenta

Zakładamy, że  $X_1, \dots, X_n$  pochodzą z rozkładu  $\mathcal{N}(\mu, \sigma^2)$ . Tym razem nie zakładamy znajomości  $\sigma^2$  — jest ona nieznana. Test  $t$ -studenta opiera się na statystyce

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

gdzie  $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$ . Przy prawdziwej hipotezie zerowej  $T \sim \mathcal{T}(n-1)$  — pochodzi z rozkładu  $t$ -studenta z  $n-1$  stopniami swobody. Dla rozpatrywanej przez nas hipotezy alternatywnej  $p$ -wartość zadana jest wzorem

$$p = 1 - F_{n-1} \left( \sqrt{n} \frac{\bar{X} - \mu_0}{S} \right),$$

gdzie  $F_{n-1}$  to dystrybuanta rozkładu  $t$ -Studenta z  $n-1$  stopniami swobody.

## 2.2 Test $t$ -Welcha

Test  $t$ -Welcha zakłada, że wariancje dwóch prób, oznaczone jako  $S_X^2$  i  $S_Y^2$  dla prób  $X$  i  $Y$  odpowiednio, są nieznane i różne. Aby przybliżyć rozkład wariancji połączonej ( $S_{WS}^2$ ), można skorzystać z równania Welch-Satterthwaite'a. Równanie to jest opisane następująco:

$$S_{WS}^2 = \frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}$$

W przybliżeniu, statystyka testu  $t$ -Welcha, oznaczana jako  $t$ , ma rozkład  $t$ -Studenta z  $\nu$  stopniami swobody, gdzie

$$\nu \approx \frac{\left( \frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y} \right)^2}{\frac{\frac{S_X^4}{n_X^2}}{(n_X-1)} + \frac{\frac{S_Y^4}{n_Y^2}}{(n_Y-1)}}$$

Przy założeniu hipotezy zerowej, statystyka testu  $t$ -Welcha jest wyrażona jako:

$$t = \frac{\bar{X} - \bar{Y}}{S_{WS}}$$

gdzie  $\bar{X}$  to średnia dla próby  $X$ ,  $\bar{Y}$  to średnia dla próby  $Y$ ,  $n_X$  to liczba obserwacji w próbie  $X$ , a  $n_Y$  to liczba obserwacji w próbie  $Y$ . Test  $t$ -Welcha jest standardowym podejściem do porównywania średnich dwóch prób w sytuacji, gdy wariancje są różne.

## 2.3 Test sumy rang Wilcoxona

Dla dwóch zmiennych niezależnych używany jest test sumy rang Wilcoxona, nazywany również testem U Manna-Whitneya. Zakładamy, że zmienne losowe  $X_1, \dots, X_n$  są niezależne i pochodzą z rozkładu o dystrybuancie  $F$ , natomiast  $Y_1, \dots, Y_m$  są niezależne i pochodzą z rozkładu o dystrybuancie  $G$ . Dodatkowo zmienne  $X_i$  i  $Y_j$  są niezależne dla każdego  $i$  oraz  $j$ . Hipoteza w teście to  $H_0 : X =_{st} Y$ , przeciwko jednej z trzech hipotez alternatywnych:  $H_1 : X <_{st} Y$ ,  $H_1 : X >_{st} Y$ ,  $H_1 : \bar{X} <_{st} \bar{Y}$  lub  $\bar{X} >_{st} \bar{Y}$ .

Test opiera się na porównaniu sumy rang dla obu prób. Jeśli dystrybuanty  $F$  i  $G$  są różne, to można oczekiwać, że suma rang jednej próby będzie zazwyczaj większa niż suma rang drugiej próby.

Jeśli zmienne losowe są ciągłymi zmiennymi losowymi z rozkładu z parametrem przesunięcia i skali, czyli dystrybuanty są postaci  $F_{\mu,\sigma}(t) = F_0\left(\frac{t-\mu}{\sigma}\right)$ , hipotezy można przetłumaczyć na hipotezy o pseudomedianach. Pseudomediana to mediana zmiennej losowej  $\frac{X_1+X_2}{2}$ , gdzie  $X_1$  i  $X_2$  są niezależnymi kopiami tej samej zmiennej losowej.

Test Wilcozona dla dwóch prób dotyczy median na podstawie porównania sumy rang obu prób, co pozwala stwierdzić, czy zmienne losowe pochodzą z tych samych rozkładów.

## 2.4 Przykładowe wywołanie w R dla zadania 1

```
mu_diff_range <- seq(-2, 2, by = 0.05)
n <- 200
sigma <- 2
alpha <- 0.05
num_simulations <- 10000

# Funkcja obliczająca moc dla testu t-Studenta
power_t_student <- function(mu_diff, sigma, n, num_simulations, alpha) {
  successes <- 0

  for (i in 1:num_simulations) {
    group1 <- rnorm(n, mean = 0, sd = sigma)
    group2 <- rnorm(n, mean = mu_diff, sd = sigma)

    test_result <- t.test(group1, group2, var.equal = TRUE)

    if (test_result$p.value < alpha) {
      successes <- successes + 1
    }
  }
  return(successes / num_simulations)
}

# Funkcja obliczająca moc dla testu t-Welcha
power_t_welch <- function(mu_diff, sigma, n, num_simulations, alpha) {
  successes <- 0

  for (i in 1:num_simulations) {
    group1 <- rnorm(n, mean = 0, sd = sigma)
    group2 <- rnorm(n, mean = mu_diff, sd = sigma)

    test_result <- t.test(group1, group2)

    if (test_result$p.value < alpha) {
      successes <- successes + 1
    }
  }
  return(successes / num_simulations)
}
```

```

    }
  }
  return(successes / num_simulations)
}

# Funkcja obliczająca moc dla testu Wilcoxona
power_wilcox <- function(mu_diff, sigma, n, num_simulations, alpha) {
  successes <- 0

  for (i in 1:num_simulations) {
    group1 <- rnorm(n, mean = 0, sd = sigma)
    group2 <- rnorm(n, mean = mu_diff, sd = sigma)

    test_result <- wilcox.test(group1, group2, alternative = "greater")

    if (test_result$p.value < alpha) {
      successes <- successes + 1
    }
  }
  return(successes / num_simulations)
}

# Obliczenia funkcji mocy dla każdego testu
#power_t1 <- sapply(mu_diff_range, power_t_student, sigma, n, num_simulations, alpha)

#power_welch1 <- sapply(mu_diff_range, power_t_welch, sigma, n, num_simulations, alpha)

#power_wilcox1 <- sapply(mu_diff_range, power_wilcox, sigma, n, num_simulations, alpha)

# Tworzenie ramki danych
#df1 <- data.frame(mu_diff = mu_diff_range, power_t_student = power_t1, power_t_welch = power_welch1, power_wilcox = power_wilcox1)
#write.csv(df1, file = "zad1.csv")

```

### 3 Zadania

Przejdziemy teraz do rozwiązywania zadań, w których rozważymy testy opisane w sekcji 2. Wszystkie testy zostaną wykonane dla poziomu istotności  $\alpha = 0.05$ . Będziemy sprawdzać hipotezy  $H_0 : \mu_1 = \mu_2$ , przeciwko  $H_1 : \mu_1 > \mu_2$ . W przypadku każdego testu moc testu będziemy wyznaczać symulacyjnie przy użyciu metody Monte Carlo, a wyniki zaprezentujemy na wykresach.

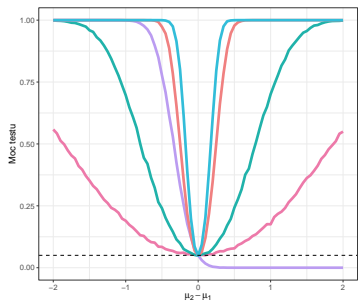
Moc testu statystycznego, znana także jako siła testu, jest prawdopodobieństwem prawidłowego odrzucenia hipotezy zerowej, gdy jest ona fałszywa. Innymi słowy, jest to zdolność testu do

wykrywania efektu, jeśli taki efekt rzeczywiście istnieje. Wartość mocy leży w zakresie od 0 do 1, gdzie wartości bliższe 1 wskazują na większą zdolność testu do wykrywania istotnych różnic lub efektów.

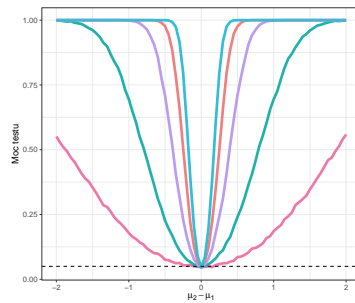
Będziemy wykorzystywać metodę Monte Carlo, która w kontekście testowania hipotez statystycznych, do estymacji mocy testu statystycznego poprzez generowanie wielu próbek z określonych rozkładów i obliczanie odsetka przypadków, w których test prawidłowo odrzuca hipotezę zerową, kiedy jest ona fałszywa.

### 3.1 Zadanie 1

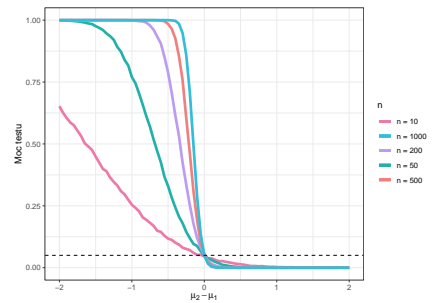
Rozważmy dwie niezależne próby losowe  $X$  i  $Y$ , które pochodzą z rozkładu normalnego o średnich  $\mu_1$  i  $\mu_2$  oraz tej samej wariancji  $\sigma^2$ . Korzystając z symulacji Monte Carlo, wykonamy wykres funkcji mocy w zależności od  $\mu_2 - \mu_1$  na przedziale  $(-2, 2)$  dla wszystkich trzech testów. Sprawdźmy, czy istnieje test jednostajnie najmocniejszy spośród rozważanych.



Rysunek 1: Funkcja mocy dla testu t-Studenta w zależności od  $n$

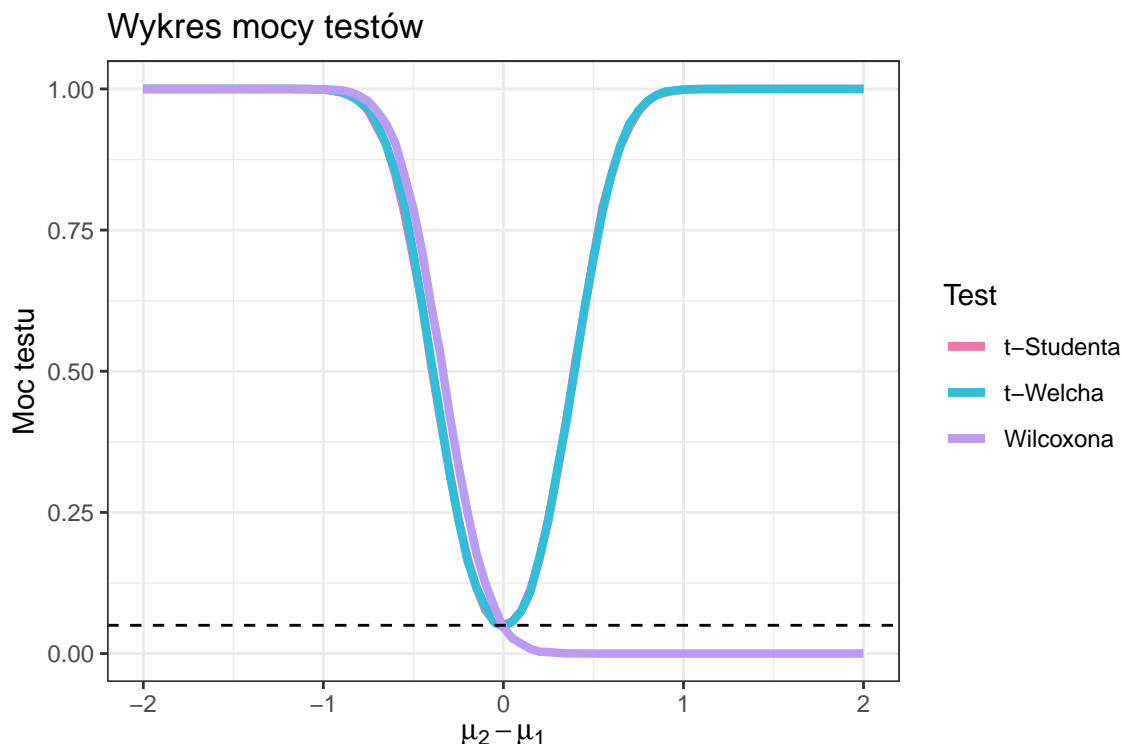


Rysunek 2: Funkcja mocy dla testu t-Welcha w zależności od  $n$



Rysunek 3: Funkcja mocy dla testu Wilcoxona w zależności od  $n$

W miejscu, gdzie różnica między średnimi  $\mu_2 - \mu_1$  wynosi zero, hipoteza zerowa jest prawdziwa, co oznacza brak rzeczywistej różnicy do wykrycia przez test. Zgodnie z definicją, moc testu powinna być niska w przypadku prawdziwej hipotezy zerowej. Wiemy, że celem testu jest minimalizacja prawdopodobieństwa błędu pierwszego rodzaju, czyli odrzucenia hipotezy zerowej pomimo jej prawdziwości. Funkcja mocy osiąga minimum w punkcie, gdzie  $\mu_2 - \mu_1 = 0$ , dla wszystkich 3 wykresów, ponieważ testy te wykrywają różnice między próbkami, a w punkcie zerowym takiej różnicy nie ma.



Rysunek 4: Funkcja mocy dla testów t-Studenta, Wilcoxona i t-Welcha

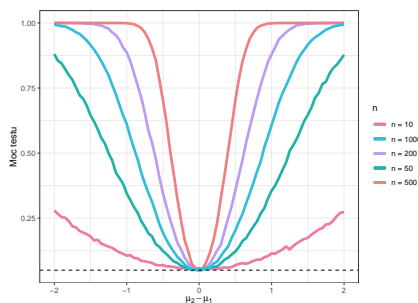
Na podstawie wykresu, można zauważyć, że test t-Studenta oraz test t-Welcha mają bardzo podobne krzywe mocy, które są najwyższe w większości zakresu wartości dla  $\mu_2 - \mu_1$  na przedziale  $(-2, 2)$ . Oba te testy mają moc bliską 1 (co jest maksymalną wartością mocy testu) dla większych różnic między średnimi, co wskazuje na to, że mają one wysoką zdolność do wykrywania różnic, kiedy takowe istnieją.

Test Wilcoxona wykazuje niższą moc w porównaniu do pozostałych dwóch testów, szczególnie w obszarze wokół punktu, gdzie nie ma różnicy między średnimi  $\mu_2 - \mu_1$ .

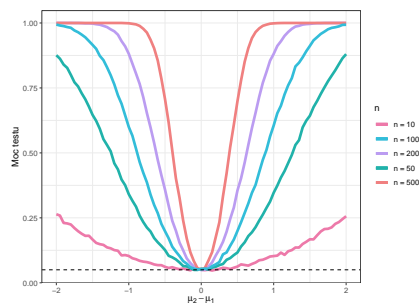
Na podstawie tego wykresu można wnioskować, że testy t-Studenta i t-Welcha są jednostajnie najmocniejsze w całym zakresie wartości różnicy średnich, które są przedstawione na wykresie. Nie można jednoznacznie stwierdzić, który z nich jest mocniejszy, ponieważ ich krzywe mocy na wykresie niemal się pokrywają.

## 3.2 Zadanie 2

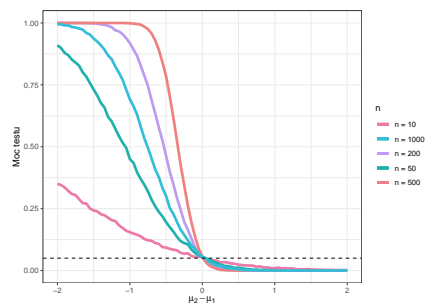
W tym zadaniu rozważymy próbę  $(X_1, \dots, X_{200})$  z rozkładu normalnego  $N(\mu_1, 2^2)$  oraz drugą próbę  $(Y_1, \dots, Y_{200})$  z rozkładu normalnego  $N(\mu_2, 4^2)$ . Wykonamy wykres funkcji mocy na przedziale  $(-2, 2)$  zawierającym przynajmniej po jednym punkcie z hipotezy zerowej i alternatywnej. Ocenimy, czy w tym przypadku istnieje test jednostajnie najmocniejszy spośród trzech rozważanych testów.



Rysunek 5: Funkcja mocy dla testu t-Studenta w zależności od  $n$

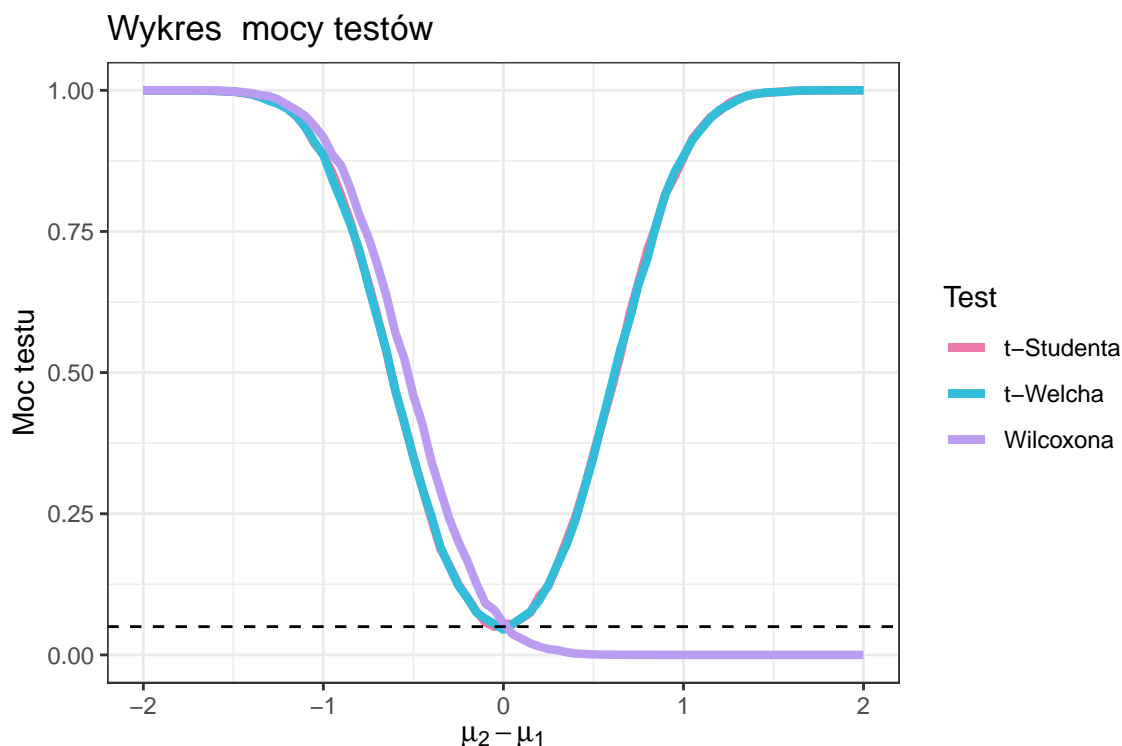


Rysunek 6: Funkcja mocy dla testu t-Welcha w zależności od  $n$



Rysunek 7: Funkcja mocy dla testu Wilcoxona w zależności od  $n$

Analogicznie do zadania 1, w miejscu gdzie różnica między średnimi  $\mu_2 - \mu_1$  wynosi zero, hipoteza zerowa jest prawdziwa, co oznacza brak rzeczywistej różnicy do wykrycia przez test. Funkcja mocy osiąga minimum w punkcie, gdzie  $\mu_2 - \mu_1 = 0$ , dla wszystkich trzech wykresów, ponieważ testy te wykrywają różnice między próbkami, a w punkcie zerowym takiej różnicy nie ma.



Rysunek 8: Wykres funkcji mocy testów w zależności od  $\mu_2 - \mu_1$  dla prób z rozkładu  $\mathcal{N} \sim (\mu, 2^2)$  i  $\mathcal{N} \sim (\mu, 4^2)$  o długości  $n = 200$  przy  $10^4$  krokach Monte Carlo dla  $H_0 : \mu_2 = \mu_1$  przeciwko  $\mu_1 > \mu_2$ .

Na nowym wykresie mocy testów widzimy, że krzywe mocy dla testu t-Studenta i testu t-Welcha nadal mają bardzo podobne kształty i są bliskie wartości 1, ale teraz test Wilcoxona wykazuje wyższą moc dla wartości różnicy średnich bliskich zero. W tym obszarze, test Wilcoxona

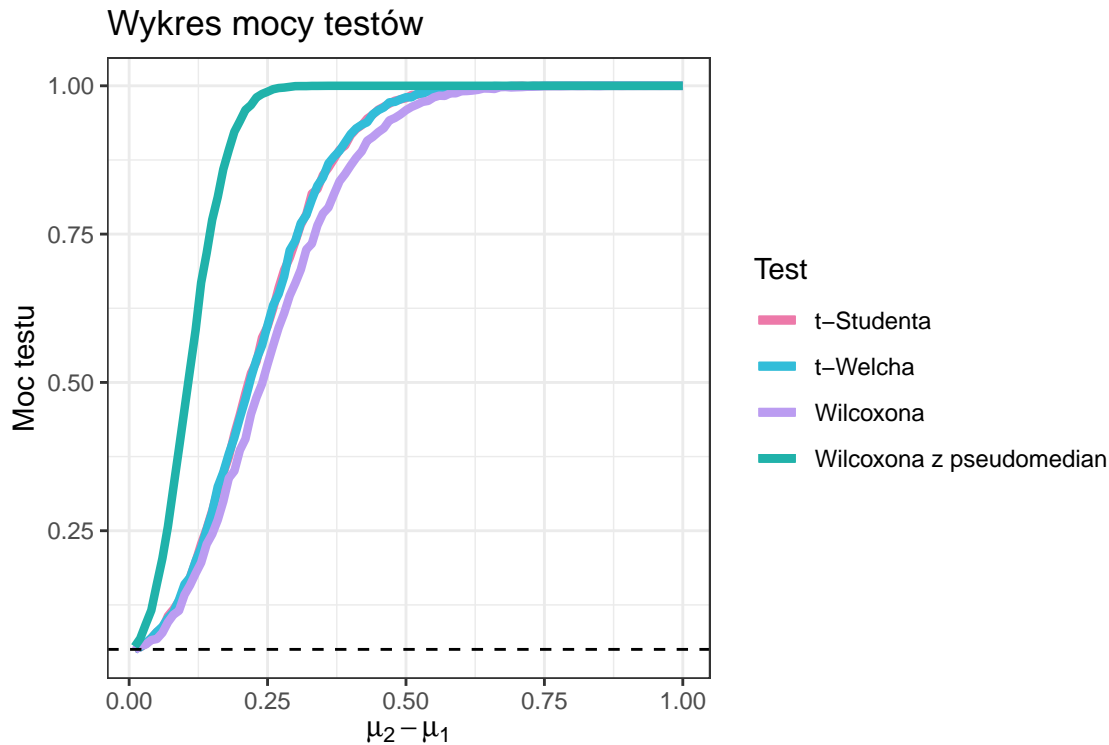
jest lepszy od pozostałych dwóch testów, co wskazuje, że ma lepszą zdolność do wykrywania różnic, kiedy średnie są bardzo podobne do siebie.

Jednakże, gdy różnica średnich zwiększa się (zarówno w kierunku dodatnim, jak i ujemnym), test Wilcoxona ma niższą moc w porównaniu do testu t-Studenta i testu t-Welcha, które to testy wykazują moc bliską 1, co oznacza bardzo wysoką zdolność do wykrywania istotnych różnic.

Na podstawie tego wykresu można stwierdzić, że nie ma testu, który byłby jednostajnie najmocniejszy na całym zakresie różnicy średnich.

### 3.3 Zadanie 3

W kolejnym zadaniu rozważymy próbę  $(X_1, \dots, X_{200})$  z rozkładu wykładniczego  $E(\frac{1}{\mu_1})$  oraz drugą próbę  $(Y_1, \dots, Y_{200})$  z rozkładu wykładniczego  $E(\frac{1}{\mu_2})$ . Wykonamy wykres funkcji mocy na przedziale  $(0, 1)$ . Sprawdzimy jak zmieniła się funkcja mocy testów i czy w tym przypadku istnieje test jednostajnie najmocniejszy spośród nich.



Rysunek 9: Wykres funkcji mocy testów w zależności od  $\mu_2 - \mu_1$  dla prób z rozkładu wykładniczego o parametrach  $\mu_1$  i  $\mu_2$  o długości  $n = 200$  przy  $10^4$  krokach Monte Carlo dla  $H_0 : \mu_2 = \mu_1$  przeciwko  $\mu_1 > \mu_2$ .

Na wykresie 9 obserwujemy wzrost mocy testu w miarę zwiększania się wartości  $\mu_2 - \mu_1$ . Jest to zgodne z naszymi oczekiwaniami, ponieważ moc testu wzrasta, gdy wzrasta różnica między parametrami, które porównujemy. Oznacza to, że zwiększenie różnicy między średnimi dwóch próbek zwiększa zdolność testu do poprawnego odrzucenia fałszywej hipotezy zerowej.



Rozkład wykładniczy, ze względu na swoją niesymetryczność, nie pozwala na interpretację testu Wilcoxona jako testu median. Tradycyjne podejście, zakładające symetrię rozkładu, nie jest adekwatne w tym przypadku. Test Wilcoxona może być jednak zinterpretowany jako test pseudomediany, która dla rozkładu wykładniczego jest jednoznacznie określona. Pseudomediana dla rozkładu wykładniczego jest równa  $q_{2,2}(0,5) \cdot \mu$ , gdzie  $q_{2,2}(0,5)$  to kwantyl rzędu 0,5 rozkładu Gamma  $G(2, 2)$ .

Na wykresie 9 przedstawiono funkcje mocy testów w zależności od różnicy parametrów dla prób z rozkładu wykładniczego o parametrach  $\mu_1 = 1$  i  $\mu_2 - \mu_1$  o długości  $n = 200$  przy  $10^4$  krokach Monte Carlo dla  $H_0 : \mu_2 = \mu_1$  przeciwko  $\mu_1 > \mu_2$ . Analizując wykres mocy testów, zauważamy, że moc testów t-Studenta i t-Welcha pozostaje wysoka dla większości wartości różnicy średnich  $\mu_2 - \mu_1$ , co sugeruje, że są one efektywne w wykrywaniu różnic między średnimi, zwłaszcza gdy różnica jest znacząca. Test Wilcoxona i test Wilcoxona z pseudomedianą wykazują niższą moc niż testy t - Welcha i t - studenta przy mniejszych różnicach średnich, ale ich moc wzrasta wraz ze wzrostem różnicy średnich. Test Wilcoxona z pseudomedianą wydaje się być lekko lepszy niż tradycyjny test Wilcoxona, co może być wynikiem lepszego dopasowania do charakterystyk danych o rozkładzie wykładniczym. Nie możemy jednak wyznaczyć jednostajnie najmocniejszego testu dla całego zakresu różnicy średnich.