

**Uniwersytet Warszawski**  
Wydział Matematyki, Informatyki i Mechaniki

**Michał Łukasik, Karol Nienałtowski**

Nr albumu: 293406, 292692

**Analiza danych o operacjach serca  
przy użyciu testów do porównywania  
dwóch rozkładów w wersjach  
klasycznej i bootstrapowej**

**Praca licencjacka  
na kierunku MATEMATYKA**

Praca wykonana pod kierunkiem  
**dra Błażeja Miasojedowa**  
Instytut Matematyki Stosowanej i Mechaniki  
Zakład Statystyki Matematycznej

Wrzesień 2013

## **Oświadczenie kierującego pracą**

Potwierdzam, że niniejsza praca została przygotowana pod moim kierunkiem i kwalifikuje się do przedstawienia jej w postępowaniu o nadanie tytułu zawodowego.

Data

Podpis kierującego pracą

## **Oświadczenie autora (autorów) pracy**

Świadom odpowiedzialności prawnej oświadczam, że niniejsza praca dyplomowa została napisana przeze mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Oświadczam również, że przedstawiona praca nie była wcześniej przedmiotem procedur związanych z uzyskaniem tytułu zawodowego w wyższej uczelni.

Oświadczam ponadto, że niniejsza wersja pracy jest identyczna z załączoną wersją elektroniczną.

Data

Podpis autora (autorów) pracy

## **Streszczenie**

W niniejszej pracy przedstawiono teorię i zastosowanie zagadnienia testowania hipotez za pomocą testów klasycznych i bootstrapowych oraz przeprowadzono analizę ich skuteczności. Opisano i pokazano przykłady dwóch metod testowania bootstrapowego : testowanie równości parametrów i testowanie równości rozkładów. W dalszej części pracy przeprowadzono analizę danych medycznych celem, której było uzyskanie rozkładów teoretycznych przybliżających dane bazowe. Rozkłady teoretyczne wraz z wyestymowanymi parametrami posłużyły do przeanalizowania testów klasycznych i bootstrapowych. Ostatecznym efektem pracy był wybór najskuteczniejszych testów i zastosowanie ich do ewaluacji danych medycznych.

## **Słowa kluczowe**

statystyka, bootstrap, testowanie hipotez, test t-studenta, test Wilcoxona, pakiet R, analiza danych, estymowanie parametrów, test Shapiro-Wilka, rozkłady skośno-normalne

## **Dziedzina pracy (kody wg programu Socrates-Erasmus)**

11.1 Matematyka

## **Klasyfikacja tematyczna**

62F40 Bootstrap, jackknife and other resampling methods

62F03 Hypothesis testing

62P10–Statistics–Applications–Applications to biology and medical sciences

## **Tytuł pracy w języku angielskim**

Analysis of heart surgery data using classical and bootstrap statistical tests for comparing two distributions



# Podział pracy

Niniejsza praca była pisana przez: Michała Łukasika i Karola Nienałtowskiego. Była ona opracowywana i konsultowana wspólnie, niemniej można wydzielić części które były pisane głównie przez jedną z osób. W tabeli przedstawiamy autora lub autorów odpowiedzialnych za kolejne rozdziały.

Rozdział/Załącznik	Autorzy
1	Michał Łukasik, Karol Nienałtowski
2.1.1	Michał Łukasik
2.1.2	Karol Nienałtowski
2.2	Michał Łukasik
3	Karol Nienałtowski
4.1.1	Michał Łukasik, Karol Nienałtowski
4.1.2	Michał Łukasik
4.1.3	Karol Nienałtowski
5	Michał Łukasik
6	Michał Łukasik, Karol Nienałtowski
A.1	Karol Nienałtowski
A.2	Karol Nienałtowski
A.3	Karol Nienałtowski
A.4	Michał Łukasik
A.5	Michał Łukasik
A.6	Karol Nienałtowski

Tablica 1: Podział pracy na autorów poszczególnych rozdziałów.



# Spis treści

<b>1. Wprowadzenie</b>	7
<b>2. Testy do porównywania dwóch rozkładów</b>	9
2.1. Testy klasyczne	9
2.1.1. Test dla porównywania średnich dwóch prób	9
2.1.2. Test Wilcoxona	11
2.2. Testy bootstrapowe	12
2.2.1. Testowanie hipotez	13
<b>3. Analiza danych</b>	15
3.1. Wstęp	15
3.2. Opis danych i wybór rozkładów teoretycznych przybliżających dane rzeczywiste	16
3.2.1. Test zgodności Shapiro-Wilka	18
3.2.2. Rozkład skośnie-normalny	18
3.3. Estymacja parametrów rozkładów teoretycznych przybliżających rozkłady bazowe	20
<b>4. Ewaluacja testów</b>	25
4.1. Porównania testów klasycznych z ich odpowiednikami bootstrapowymi	25
4.1.1. Opis eksperymentów	25
4.1.2. Test t-studenta	27
4.1.3. Test Wilcoxona	32
<b>5. Eksperymenty na danych medycznych</b>	37
<b>6. Podsumowanie</b>	39
<b>A. Implementacja eksperymentów w R</b>	41
A.1. Rozszerzona biblioteka udostępniająca rozkład skośnie-normalny	41
A.2. Implementacja estymacji parametrów	41
A.3. Analiza rozkładów poszczególnych kolumn	43
A.4. Implementacja metod bootstrapowych	45
A.5. Przykład użycia metod bootstrapowych	47
A.6. Implementacja transformacji rozkładów rangowych	49





# Rozdział 1

## Wprowadzenie

Celem niniejszej pracy jest zastosowanie metody bootstrap do przeprowadzenia analizy statystycznej danych dotyczących dwóch typów operacji serca. Dane pochodzą z Oddziału Kardiologii Innowacyjnej Warszawskiego Uniwersytetu Medycznego. Przedmiotem zainteresowania są różnice między rozkładami populacji pacjentów po dwóch typach operacji.

Wybrano następujące testy do przeprowadzenia wnioskowania: test t-studenta oraz test Wilcoxona. Głównym zagadnieniem, które należy rozpatrzyć, jest czy wersje bootstrapowe tych testów będą bardziej odpowiednie do analizy rozważanych danych. W związku z tym, niniejsza praca ma dwa cele. Pierwszym z nich jest wybranie odpowiednich testów do analizy danych medycznych. Drugim celem jest wyciągnięcie wniosków na temat danych medycznych, po zastosowaniu wybranych testów.

Pozostała część niniejszej pracy licencjackiej została podzielona na następujące części. W rozdziale 2 przeprowadzono wstęp teoretyczny do testów statystycznych. Opisano dwa klasyczne testy statystyczne: t-studenta i Wilcoxona. Ponadto, dokonano wprowadzenia do techniki bootstrap oraz przedstawiono sposoby tworzenia bootstrapowych odpowiedników klasycznych testów statystycznych. Następnie, w rozdziale 3 została opisana analiza danych medycznych, mająca na celu poznanie ich własności. Owe własności mogą pomóc w odpowiednim doborze metody ewaluacji testów statystycznych. Rozdział 4 zawiera porównania dwóch wersji testów klasycznych z ich odpowiednikami bootstrapowymi. Rozdział 5 zawiera omówienie wyników uzyskanych po zastosowaniu wybranych testów do danych medycznych.



## Rozdział 2

# Testy do porównywania dwóch rozkładów

Ważnym zagadnieniem w statystyce jest testowanie hipotez. W sytuacji, w której poszukuje się przesłanek dla odrzucenia pewnej hipotezy (zwanej **hipotezą zerową**:  $H_0$ ) na rzecz innej hipotezy (zwanej **hipotezą alternatywną**:  $H_1$ ), można użyć tego aparatu matematycznego.

W testowaniu hipotez statystyk ma do czynienia z hipotezą zerową i poszukuje przesłanek dla jej odrzucenia na rzecz hipotezy alternatywnej. W tym celu definiuje statystykę  $s$ , która przyjmuje mało prawdopodobne wartości jeśli hipoteza alternatywna jest prawdziwa ale jednocześnie bardzo prawdopodobne wartości jeśli hipoteza zerowa zachodzi. Definiowanie hipotez oraz statystyki bądź wybranie odpowiednich spośród już istniejących jest głównym problemem w testowaniu hipotez.

Po zdefiniowaniu hipotez  $H_0$  i  $H_1$  oraz statystyki  $s$ , oblicza się wartość statystyki  $s(x)$  dla zaobserwowanej próby  $x = x_1, \dots, x_n$  pochodzącej z populacji  $X$ . Następnie oblicza się prawdopodobieństwo zaobserwowania co najmniej tak ekstremalnej wartości statystyki przy założeniu hipotezy zerowej, co często sprowadza się do obliczenia wartości:  $P_{H_0}(s(X) \geq s(x))$  (gdy małe wartości statystyki są charakterystyczne dla hipotezy zerowej). Jeśli wartość statystyki jest mało prawdopodobna przy założeniu hipotezy zerowej, jest ona odrzucana. W przeciwnym przypadku brak jest przesłanek dla odrzucenia hipotezy zerowej. Nie ma tutaj ścisłej wartości granicznej od której odrzuca się hipotezę zerową. Niemniej istnieją pewne konwencje ([ET93], rozdział 15):

1.  $P_{H_0}(s(X) \geq s(x)) < 0.1$  - słaba przesłanka przeciwko  $H_0$ ,
2.  $P_{H_0}(s(X) \geq s(x)) < 0.05$  - stosunkowo mocna przesłanka przeciwko  $H_0$ ,
3.  $P_{H_0}(s(X) \geq s(x)) < 0.025$  - mocna przesłanka przeciwko  $H_0$ ,
4.  $P_{H_0}(s(X) \geq s(x)) < 0.01$  - bardzo mocna przesłanka przeciwko  $H_0$ .

W niniejszym rozdziale zostały przedstawione dwa przykłady klasycznych testów statystycznych oraz metoda testowania hipotez przy użyciu techniki bootstrap.

### 2.1. Testy klasyczne

#### 2.1.1. Test dla porównywania średnich dwóch prób

Szeroko stosowanym testem statystycznym dla porównania średnich dwóch prób jest test t-studenta. W niniejszym rozdziale opisujemy ten test i podajemy jego założenia, które w

realnych zastosowaniach mogą zostać naruszone. Opis testu poprzedzamy kilkoma istotnymi definicjami.

**Definicja 1.** *Standardowym estymatorem średniej z próby  $x = x_1, \dots, x_n$  nazywamy estymator  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ .*

**Definicja 2.** *Standardowym estymatorem wariancji z próby  $x = x_1, \dots, x_n$  nazywamy estymator  $S^2(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ .*

**Definicja 3.** *Standardowym estymatorem odchylenia standardowego z próby  $x = x_1, \dots, x_n$  nazywamy estymator  $S(x) = \sqrt{S^2(x)}$ .*

Rozważmy dwie próby losowe  $x_1, x_2, \dots, x_{n_1}$  oraz  $y_1, y_2, \dots, y_{n_2}$ , pochodzące z pewnych rozkładów o nieznanach wartościach średnich  $m_1$  i  $m_2$ . Przy założeniu, że obie próby pochodzą z rozkładów normalnych o pewnym wspólnym nieznanym odchyleniu standardowym  $\sigma$ , można przeprowadzić test statystyczny dla równości średnich dwóch prób. W przypadku różnych odchyłeń nieznan jest dokładny rozkład statystyki testowej, którą w dalszej części podamy ([KM06], rozdział 3.).

Niech  $\bar{x}$  oznacza estymator średniej z próby  $x$ , zaś  $\bar{y}$  oznacza estymator średniej z próby  $y$ . Ponadto, niech  $S_1$  i  $S_2$  oznaczają estymatory odchyłeń w próbach  $x$  i  $y$ . Wówczas estymator wspólnego dla obu prób odchylenia standardowego  $S_p$  można wyrazić wzorem ([KM06], rozdział 3.):

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}. \quad (2.1)$$

Można pokazać, że statystyka:

$$T = \frac{(\bar{x} - \bar{y}) - (m_1 - m_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (2.2)$$

ma rozkład t-studenta z  $n_1 + n_2 - 2$  stopniami swobody. Na tej podstawie można podać test statystyczny, pozwalający stwierdzić o równości bądź różności średnich z dwóch prób.

Hipoteza zerowa ma postać:

$$H_0 : m_1 = m_2. \quad (2.3)$$

W pracy zajmiemy się sytuacją symetryczną, tzn taką, gdzie hipoteza alternatywna ma postać:

$$H_1 : m_1 \neq m_2. \quad (2.4)$$

Na podstawie tego, że statystyka z równania (2.2) ma rozkład t-studenta z  $n_1 + n_2 - 2$  stopniami swobody, można podać sposób obliczania zbioru krytycznego. Mianowicie zbiór krytyczny na poziomie istotności  $1 - \alpha$  (tzn. taki, że przy założeniu hipotezy zerowej prawdopodobieństwo, że statystyka przyjmie wartość należącą do tego zbioru wynosi  $1 - \alpha$ ) można przedstawić następująco:

$$C = \{t : t \leq -t_{1-\frac{\alpha}{2}, n_1+n_2-2}\} \cup \{t : t \geq t_{1-\frac{\alpha}{2}, n_1+n_2-2}\}. \quad (2.5)$$

Warto zauważyć, jakie założenia zostały poczynione w opisywanym teście. Po pierwsze, założono że rozważane populacje pochodzą z rozkładu normalnego, co niekoniecznie musi być prawdą. W przypadku gdy próba nie pochodzi z rozkładu normalnego i jednocześnie jest małą

liczności, założenie o normalności jest bezzasadne. Drugim założeniem jest równość odchyłeń standardowych. Założenie to wydaje się być bardzo rygorystyczne i łatwe do naruszenia w realnych zastosowaniach.

### 2.1.2. Test Wilcoxona

Test Wilcoxona jest to jeden z testów kombinatorycznych (rangowych) do sprawdzenia czy wartości z dwóch niezależnych populacji są równe. Rozpatrujemy dwa niezależne ciągi obserwacji:

$$X : X_1, X_2, \dots, X_n \sim F,$$

$$Y : Y_1, Y_2, \dots, Y_m \sim G,$$

gdzie  $F$  i  $G$  są dowolnymi dystrybucjami. W szczególności nie zakładamy nic o ich kształcie i wzajemnym położeniu. Taki test nazywamy **testem nieparametrycznym**.

Hipoteza zerowa rozpatrywana przez test Wilcoxona to:

$$H_0 : F = G.$$

W pracy będziemy rozważać przypadek symetryczny, tzn. taki, w którym hipoteza alternatywna ma postać:

$$H_1 : G(t) \neq F(t).$$

W literaturze rozważane są dwie statystyki testu Wilcoxona:  $W$  i  $U$ .

**Definicja 4. Statystyką Wilcoxona  $W$**  dla hipotezy  $H_0$  przeciwko alternatywie  $H_1$  nazywamy sumę rang obserwacji  $Y_1, Y_2, \dots, Y_m$  w połączonej uporządkowanej próbie  $X_1, \dots, X_n, Y_1, \dots, Y_m$ , gdzie rangą nazywamy numer elementu w uporządkowanej próbie.

Natomiast **statystyką Wilcoxona  $U$**  przy powyższych założeniach oznacza się sumę inwersji  $Y_i$  dla  $i \in \{1, m\}$ , gdzie inwersją  $Y$ -a nazywa się liczbę  $X$ -ów poprzedzających w połączonej uporządkowanej próbie. Łatwo wywnioskować, że  $U = \sum_{x_i < y_j} 1 + \sum_{x_i = y_j} \frac{1}{2}$ .

Statystyki  $U$  i  $W$  są równoważne, a pomiędzy nimi zachodzi zależność ([Nie11]):

$$W = \frac{n(n+1)}{2} + U.$$

W ogólności wymagane jest, aby  $F$  i  $G$  były rozkładami ciągłymi, jednakże w praktyce test bywa stosowany do próbek z rozkładów dyskretnych.

Poniżej znajdują się podstawowe własności statystyki Wilcoxona (na podstawie [Nie11]):

**Fakt 1.** Jeśli  $F = G$  i dystrybucja  $F$  jest ciągła to:

- rozkład statystyki Wilcoxona  $W$  nie zależy od dystrybucji  $F$ ,
- $\mathbb{E}W = \frac{m(n+m+1)}{2}$ ,
- $\text{Var}(W) = \frac{nm(n+m+1)}{12}$ ,
- dla dowolnej liczby  $t$ :  $\mathbb{P}\left(\frac{W - \mathbb{E}W}{\sqrt{\text{Var}(W)}} \leq t\right) \rightarrow \Phi(t)$ .

**Fakt 2.** Dla próbek wielkości  $m, n > 10$  rozkład  $U$  jest aproksymacyjnie rozkładem normalnym o parametrach ([MW47]):

- $\mu_U = \frac{nm}{2},$
- $\sigma_U^2 = \frac{nm(n+m+1)}{12}.$

Stąd statystyka  $z = \frac{U - \mu_U}{\sigma_U}$  aproksymacyjnie pochodzi z rozkładu normalnego.

Test Wilcoxona jest często wykorzystywany, gdy niespełnione są założenia testu t-Studenta. Z tego względu nazywany jest alternatywą tego testu ([CI81]).

Poza opisanym wyżej zagadnieniem hipotezy zerowej  $H_0 : F = G$  za pomocą testu Wilcoxona rozważa się tzw. równość stochastyczną ([RZR05]), czyli hipotezę zerową

$$H_0 : P(X < Y) = P(Y > X),$$

przeciwko

$$H_1 : P(X < Y) \neq P(Y > X).$$

Wynika to z faktu ([RZR05]), że  $\frac{U}{nm}$  jest estymatorem  $P(X < Y) + \frac{1}{2}P(X = Y)$ , co sugeruje, że test Wilcoxona jest wrażliwy na takie różnice rozkładów  $F$  i  $G$ , które wynikają z ich nierówności stochastycznej ([RZR05]).

Na podstawie powyższych właściwości zaproponowano zastosowanie testu Welcha do rang prób  $X$  i  $Y$  do rozwiązania hipotezy równości stochastycznej. W ten sposób otrzymano rozszerzenie testu Wilcoxona.

**Definicja 5.** Niech  $r_X$  oraz  $r_Y$  oznaczają rangi odpowiednio  $X$  i  $Y$  w połączonej próbie.

Dalej niech  $\bar{r}_X$ ,  $\bar{r}_Y$  i  $s_X^2$ ,  $s_Y^2$  oznaczają odpowiednio średnią oraz wariancję  $r_X$  oraz  $r_Y$ . Wtedy *statystyką testową Welcha* nazywamy :

$$t_{RW} = \frac{\bar{r}_X - \bar{r}_Y}{\sqrt{\frac{s_X^2}{n} - \frac{s_Y^2}{m}}}.$$

## 2.2. Testy bootstrapowe

Bootstrap jest metodą wnioskowania statystycznego opartą na obliczeniach komputerowych, często omijająca potrzebę wyznaczania zależności matematycznych zadanych wzorem. Można powiedzieć, że jest to alternatywne podejście do wnioskowania statystycznego względem wcześniej omówionego klasycznego. W niniejszym rozdziale wprowadzamy metodę bootstrap, opierając się na pozycji [ET93].

Rozważmy próbę losową  $x_1, x_2, \dots, x_n$ , będącą realizacją pewnej zmiennej losowej  $X$ . Poza tym, rozważmy statystykę  $s$ . Zasada bootstrap pozwala przybliżać rozkład statystyki  $s$  w populacji na podstawie zadanej próby.

**Definicja 6.** *Próbą bootstrap* nazywamy prostą próbę losową  $x^* = (x_1^*, \dots, x_n^*)$  uzyskaną z empirycznego rozkładu zadanej próby  $x = (x_1, \dots, x_n)$ .

Poprzez wielokrotne losowanie elementów próby i każdorazowe obliczanie wartości statystyki dla realizacji próby bootstrap, uzyskuje się zbiór wartości statystyk  $s(x^*)$ , tworzący pewien rozkład. Zasada bootstrap mówi, że uzyskany rozkład empiryczny przybliża rozkład statystyki  $s(X)$  w populacji. W granicy, gdy liczba prób bootstrapowych dąży do nieskończoności, rozkład  $s(x^*)$  jest równy rozkładowi wartości statystyki w populacji wynikającej z empirycznego rozkładu zadanej próby  $x$ .

Metoda bootstrap ma wiele zastosowań. W niniejszej pracy skupiamy się na zastosowaniu jej w testowaniu hipotez.

### 2.2.1. Testowanie hipotez

Zgodnie z tym co zostało napisane we wstępie do niniejszego rozdziału, przed testowaniem statystycznym należy określić postaci: hipotezy zerowej, hipotezy alternatywnej oraz statystyki mającej pomóc w stwierdzeniu o prawdziwości bądź fałszywości hipotezy zerowej.

W procesie testowania hipotez przy pomocy metody bootstrap, w pierwszej kolejności modyfikuje się próbę wejściową  $x$  w taki sposób, aby hipoteza zerowa dla niej zachodziła. Z tak uzyskanej próby  $\hat{x}$ , losuje się  $B$  prób bootstrapowych  $\hat{x}^*$ . Dla każdej z prób oblicza się wartość statystyki  $s(\hat{x}^*)$ . Następnie dokonuje się zliczenia wartości co najmniej tak ekstremalnych jak wartość zaobserwowana dla pierwotnej próby  $x$ . W ten sposób uzyskuje się przybliżone prawdopodobieństwo uzyskania co najmniej tak ekstremalnej wartości statystyki jak  $s(x)$ , przy założeniu hipotezy zerowej:

$$P_{H_0}(s(X) \geq s(x)) = \frac{1}{B} \sum_{i=1}^B I_{s(\hat{x}_i^*) \geq s(x)}. \quad (2.6)$$

Efron w ([ET93], rozdział 16) przytacza dwa szczególne przypadki przedstawionego schematu. Poniżej je pokrótce przedstawiamy. Na nich opieramy bootstrapowe wersje testów klasycznych.

#### Testowanie równości rozkładów

W przypadku testowania hipotezy o równości rozkładów na wejściu dane są dwie realizacje:  $x$  rozmiaru  $n$  i  $y$  rozmiaru  $m$ . W pierwszym kroku procedury testowania obie próby są łączone w jedną:  $z = x + y$ . W ten sposób symuluje się zachodzenie hipotezy zerowej: jeśli rozkłady są takie same, to każda obserwacja mogła z równym prawdopodobieństwem trafić do  $x$  jak do  $y$ .

$B$  prób bootstrapowych losuje się z  $z$  w następujący sposób. Za każdym razem, losuje się ze zwracaniem  $n + m$  obserwacji:  $z^*$ . Pierwszych  $n$  tworzy próbę  $x^*$ , a ostatnich  $m$  obserwacji tworzy próbę  $y^*$ . Na każdej realizacji typu bootstrap liczona jest wartość statystyki testowej, np. w przypadku testowania równości średnich (będącym przykładem podawanym w [ET93]) postaci  $s(z^*) = x^* - y^*$ . Ostatecznie, szukane prawdopodobieństwo zaobserwowanej wartości statystyki przy założeniu hipotezy zerowej jest zadane wzorem (2.6).

Warto zauważyć, że można rozważyć inne statystyki, jak np. mediana albo maksymalna różnica rozkładów empirycznych.

#### Testowanie równości parametru

Innym typem testu statystycznego jest testowanie hipotezy o równości pewnego parametru w dwóch populacjach, reprezentowanych przez dwie próby:  $x$  rozmiaru  $n$  i  $y$  rozmiaru  $m$ . Przykładem może być testowanie równości średnich, który tutaj będziemy przytaczać dla zobrazowania metody (za Efronem, [ET93]).

W pierwszym kroku, obie próby są modyfikowane tak, aby zachodziła dla nich hipoteza zerowa. W przypadku średniej, może to być przesunięcie obu prób o pewne wartości takie, że w nowych próbach zachodzi równość średnich.

Z nowych prób:  $\hat{x}$  i  $\hat{y}$ , losowanych jest  $B$  realizacji bootstrapowych  $\hat{z}^*$  składających się z  $\hat{x}^*$  i  $\hat{y}^*$ . Próba  $\hat{x}^*$  losowana ze zwracaniem z  $\hat{x}$ , zaś próba  $\hat{y}^*$  ze zwracaniem z  $\hat{y}$ . Dla każdej realizacji  $\hat{z}^*$  liczona jest statystyka  $s(\hat{z}^*)$ , w przypadku średniej może to być studentyzowana różnica średnich. Ostatecznie, prawdopodobieństwo zaobserwowania  $s(x + y) = s(z)$  przy założeniu hipotezy zerowej jest zadane wzorem (2.6).

Tutaj również można modyfikować metodę, badając inne parametry. W przypadku parametrów innych niż średnia czy mediana wymyślenie odpowiedniej transformacji może być trudne. Efron zaznacza, że istotne jest takie zmodyfikowanie prób, aby były one tak bliskie pierwotnym próbom jak to możliwe (aby nie modyfikować zbyt mocno charakterystyki rozkładu). Przykładowo, przy badaniu każdego parametru można sobie wyobrazić taką modyfikację prób która polega na zrównaniu próby  $x$  i próby  $y$  poprzez przypisanie wszystkim obserwacjom wartości 0. Nie jest to jednak dobra metoda, gdyż gubiona jest informacja o próbach wejściowych.



## Rozdział 3

# Analiza danych

### 3.1. Wstęp

Celem pracy jest uzyskanie odpowiedzi na pytanie, który rodzaj leczenia osób chorych na serce jest odpowiedni. Do dyspozycji otrzymaliśmy dane przedstawiające stan pacjentów po przebytym zawale serca. Rozumie się przez to (na podstawie [Gaj]):

**Definicja 7. Zawał mięśnia sercowego** (łac. *infarctus myocardii*) – martwica mięśnia sercowego spowodowana jego niedokrwieniem.

*Zawał mięśnia sercowego jest postacią choroby niedokrwiennej mięśnia sercowego (obok nagłej śmierci sercowej, dławicy sercowej, przewlekłej choroby niedokrwiennej serca i tzw. kardiomiopatii niedokrwiennej).*

Pacjenci zostali poddani jednemu ze standardowych rodzajów zabiegu :

- Pomostowanie aortalno-wieńcowe (ang.bypass) - w tabeli odpowiada temu wartość 0 w kolumnie **Typ Zabiegu**,
- Angioplastyka tętnic wieńcowych (inaczej balonikowanie) - w tabeli odpowiada temu wartość 1 w kolumnie **Typ Zabiegu**,

Kolejne kolumny opisujące zabiegi zawierają wyniki niektórych badań przed i po<sup>1</sup> zastosowaniu leczenia. Są to odpowiednio:

- **AVA** - powierzchnia przekroju ujścia aorty,
- **EF** - frakcja wyrzutowa serca,
- **Mean** - średni gradient ciśnień,
- **Max** - maksymalny gradient ciśnień.

Ponadto podany jest **Wiek** każdego z pacjentów.

Do dyspozycji otrzymaliśmy 64 przypadki, jednakże ze wstępnej analizy danych wynikało (patrz tablica 3.1), iż bardzo duża ich część jest niepełna tzn. niektóre z pól były puste. Liczba pustych pozycji może znacząco wpłynąć na wyniki poszczególnych metod.

---

<sup>1</sup>Nazwy kolumn są oznaczone dopiskiem *post*

	EF	AVA	MAX	MEAN	AVApост	MAXpost	MEANpost	Wiek
Liczba	1	1	0	8	39	2	4	1
Procent	1%	1%	0%	12%	57%	3%	6%	1%

Tablica 3.1: Liczba pustych pozycji w arkuszu danych w zależności od kolumn. Z tabeli można odczytać, że zwłaszcza kolumna **AVApост** posiada dużo pustych pól.

### 3.2. Opis danych i wybór rozkładów teoretycznych przybliżających dane rzeczywiste

Celem analizy danych jest wybranie rozkładów teoretycznych, które najlepiej będą przybliżały rozkłady bazowe. Rozkłady te będą dalej wykorzystywane do oceny skuteczności działania wybranych testów klasycznych i ich bootstrapowych odpowiedników.

Analiza obejmuje zatem :

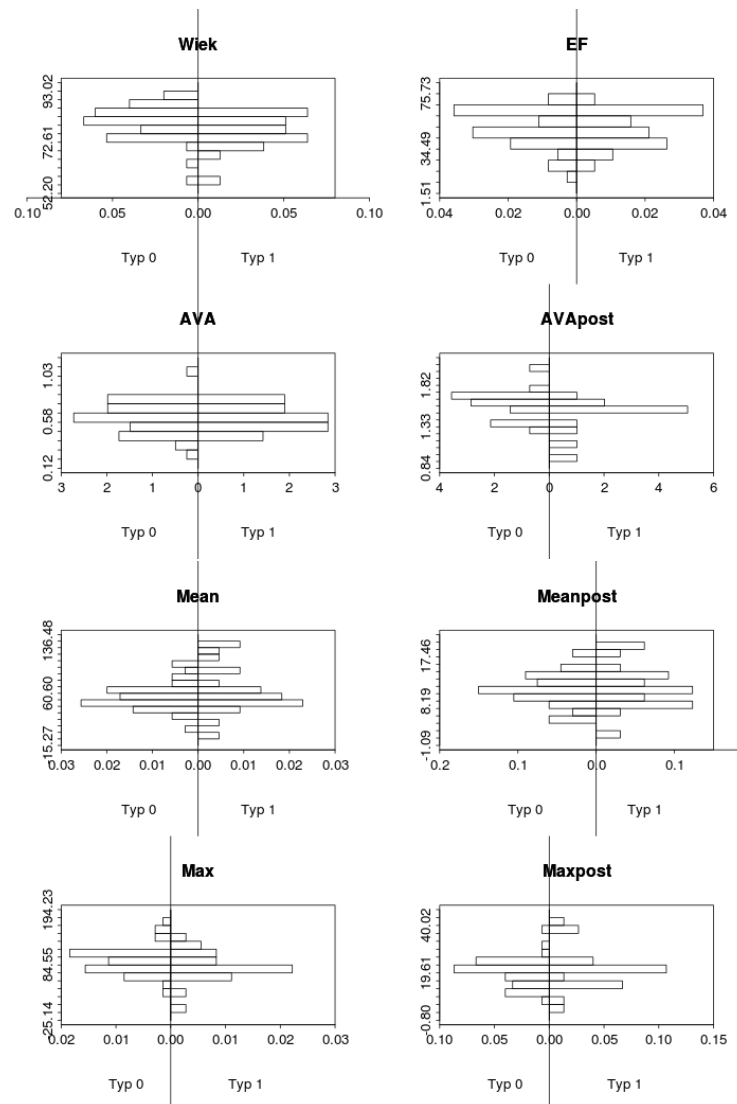
- Sprawdzenie podstawowych właściwości danych rzeczywistych takich, jak wartość średnia, mediana, itp. Wyniki tej analizy są zaprezentowane w tablicy 3.2.
- Przyjrzenie się histogramom rozkładów bazowych.
- Przeanalizowanie niezależności wartości kolejnych kolumn pomiędzy różnymi typami zabiegów.

	Min	$q_1$	Mediana $m$	Wartość średnia $\hat{\mu}$	$q_3$	Max	Skośność $\alpha$
Typ Zabiegu 0							
EF	18.0	40.00	50.0	49.4800	60.0000	75.00	-0.4918
AVA	0.3	0.50	0.6	0.6118	0.7225	1.05	0.0807
AVApост	1.3	1.50	1.6	1.6010	1.7000	2.00	0.2710
Mean	11.0	43.00	56.0	56.1600	66.0000	107.00	0.3386
Meanpost	4.0	8.75	10.0	10.4400	12.5000	19.00	0.2824
Max	23.0	72.00	92.0	92.9300	110.0000	175.00	0.4374
Maxpost	8.0	15.75	21.0	20.2000	24.0000	40.00	0.3584
Typ Zabiegu 1							
EF	20.0	36.50	50.0	48.6500	60.000	70.0	-0.3530
AVA	0.4	0.50	0.6	0.6061	0.700	0.8	0.1128
AVApост	1.0	1.33	1.5	1.4380	1.562	1.7	-0.8189
Mean	3.7	44.50	59.0	65.3800	85.000	127.0	0.4235
Meanpost	2.0	8.00	11.0	11.3300	14.000	20.0	0.3223
Max	6.2	68.50	81.0	81.9200	102.500	136.0	-0.3397
Maxpost	6.0	15.25	20.0	21.0700	23.750	42.0	0.7055

Tablica 3.2: Wartości parametrów rozkładów analizowanych dalej kolumn w zależności od przeprowadzonego typu zabiegu. Oznaczenia parametrów standardowe.  $q_1$  i  $q_3$  oznaczają odpowiednio kwantyl 25% i 75%. Skośność jest zdefiniowana wzorem :  $\alpha = 3 \frac{\hat{\mu} - \hat{m}}{\hat{\sigma}}$ , gdzie  $\hat{\sigma}$  oznacza estymator odchylenia standardowego.

Wartości w Tablicy 3.2 wskazują na duże zróżnicowanie rozkładów poszczególnych kolumn. Istotnym faktem jest, że rozkłady nie są scentrowane, a parametr skośności przyjmuje zarówno wartości dodatnie, jak i ujemne. Oznacza to, że do przybliżania prób należy rozważyć zarówno rozkłady lewoskośne, jak i prawoskośne.

Tę właściwość oraz ogólny wygląd rozkładów można dostrzec analizując histogramy danych rzeczywistych (3.1).



Rysunek 3.1: Histogramy rozkładów empirycznych poszczególnych kolumn w zależności od **Typu Zabiegu**. Wśród rozkładów można zauważyć takie, które przypominają rozkład normalny (np. **MEANpost**, bądź **MEAN**), ale także rozkłady, które są skośne (np. **Wiek** lub **EF**).

Z rysunku 3.1 wynika, iż do przybliżania danych można rozważyć rozkład normalny. Aby to sprawdzić posłużyliśmy się testem Shapiro-Wilka. Ponadto istotną właściwością prób jest skośność, którą można wyrazić przybliżając dane za pomocą rozkładu skośno-normalnego.

### 3.2.1. Test zgodności Shapiro-Wilka

Test ten służy do porównywania podobieństwa danej próby  $Z$  do rozkładu normalnego.

**Definicja 8.** Dla próby:  $Z : Z_1, Z_2, \dots, Z_n$ , test Sharipo-Wilka sprawdza, czy zachodzi hipoteza zerowa:

$$H_0 : Z \sim N(\mu, \sigma^2).$$

Z próby  $Z$  tworzymy statystykę pozycyjną

$$Z_{1:n}, Z_{2:n}, \dots, Z_{n:n}.$$

Statystyka testowa ma następującą postać :

$$W = \frac{\sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} a_{i:n} (Z_{(n-i+1):n} - Z_{i:n})^2}{Var(Z)},$$

gdzie  $a_{i:n}$  są współczynnikami zapisanymi w tablicach.

Wartość krytyczna testu Shapiro-Wilka to  $W_n(\alpha)$ .

$H_0$  odrzucamy, jeśli  $W < W_n(\alpha)$ .

W Tablicy 3.3 znajdują się p-wartości testu zgodności Shapiro-Wilka dla poszczególnych kolumn w zależności od **Typu Zabiegu**.

Przyjmując poziom istotności  $\alpha = 0.05$  nie możemy odrzucić hipotezy, iż którykolwiek z

	EF	AVA	AVApst	Mean	Meanpost	Max	Maxpost
Typ Zabiegu 0	0.05	0.41	0.50	0.85	0.34	0.24	0.07
Typ Zabiegu 1	0.25	0.08	0.11	0.20	0.41	0.61	0.08

Tablica 3.3: P-wartości testu zgodności Shapiro-Wilka.

rozkładów rzeczywistych nie pochodzi z rozkładu normalnego. W związku z tym w dalszej analizie posłużymy się tym rozkładem.

### 3.2.2. Rozkład skośno-normalny

Rozkład skośno-normalny jest naturalnym rozszerzeniem rozkładu normalnego o właściwość skośności. Jest on używany do analizy danych, które przypominają rozkład normalny, ale nie są symetryczne. Możliwość przybliżania danych rzeczywistych rozkładem normalnym (na podstawie testu Shapiro-Wilka) oraz skośność prób, która może mieć znaczący wpływ na rozróżnianie rozkładów powodują, że rozważany rozkład może nadawać się do przybliżania rozkładu bazowego.

Poniżej znajdują się definicja oraz właściwości rozkładu skośno-normalnego.

**Definicja 9.** Rozkład skośno-normalny  $SN(\mu, \sigma^2, \alpha)$  jest to rozkład ciągły na przedziale  $[-\infty, \infty]$  o gęstości wyrażającej się wzorem :

$$f(x) = 2\phi(x)\Phi(\alpha x),$$

gdzie

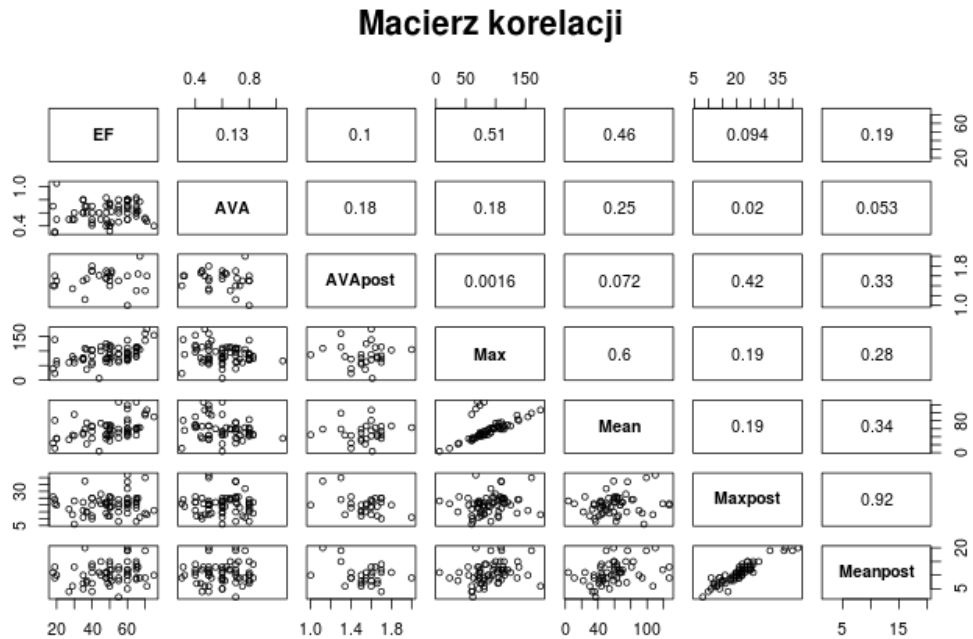
$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad \Phi(\alpha x) = \int_{-\infty}^{\alpha x} \phi(t) dt.$$

Funkcja  $\phi(x)$  to gęstość standardowego rozkładu normalnego. Parametry  $\mu, \sigma^2$  to odpowiednio wartość oczekiwana i wariancja rozkładu, natomiast parametr  $\alpha$  nazywany jest parametrem kształtu (ang. shape) i definiuje krzywiznę rozkładu skośnie-normalnego.

- Dla  $\alpha = 0$  :  $SN(\mu, \sigma^2) \sim N(\mu, \sigma^2)$ .
- $\alpha$  jest proporcjonalne do skośności  $SN$ .
- Zmiana znaku  $\alpha$  powoduje zmianę skośności na przeciwną.

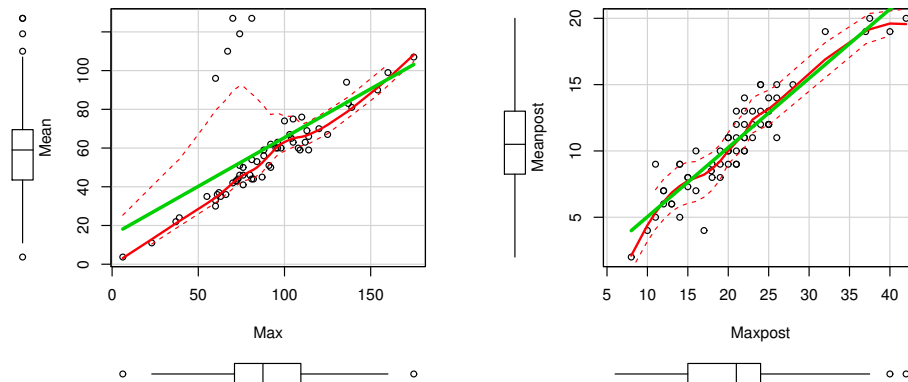
Implementacja rozkładu skośnie-normalnego znajduje się w bibliotece **VGMA**<sup>2</sup>. Ponadto w dodatku A.1 znajduje się rozszerzenie biblioteki umożliwiające przeprowadzenie analizy i estymacji parametrów.

Na rysunku 3.2 znajdują się wykresy zależności między parami cech. Wynika z nich, że mocno skorelowane są cechy: **MEAN** i **MAX** oraz **MEANpost** i **MAXpost**. Potwierdzają to wykresy z rysunku 3.3. Pozostałe cechy wydają się nie wykazywać zależności pomiędzy parami. Zależności pomiędzy wyszczególnionymi kolumnami będą miały znaczenie podczas konstruowania testów na danych rzeczywistych.



Rysunek 3.2: Wykres zależności rozkładów poszczególnych kolumn. Zgodnie z oczekiwaniami istnieje duża zależność pomiędzy kolumnami **MAX** i **MEAN** oraz pomiędzy **MAXpost** i **MEANpost**. Charakterystykę tą potwierdza rysunek 3.3. Wykres otrzymany za pomocą funkcji *pairs* z pakietu R.

<sup>2</sup>Więcej informacji na temat rozkładu oraz implementacji biblioteki można znaleźć na stronie internetowej <http://azzalini.stat.unipd.it/SN/>



Rysunek 3.3: Wykresy otrzymane za pomocą funkcji *scatterplot* z pakietu *car*. Przedstawiają zależności pomiędzy kolumnami **MAX**, a **MEAN** oraz **MAXpost**, a **MEANpost**.

### 3.3. Estymacja parametrów rozkładów teoretycznych przybliżających rozkłady bazowe

Aby znaleźć rozkłady, które optymalnie przybliżają próby należy wyestymować parametry rozważanych rozkładów teoretycznych (w tym przypadku są to rozkład normalny i skośno-normalny). Niech

$$Y : Y_1, Y_2, \dots, Y_m \sim \hat{F}$$

oznacza rozkład wybranej kolumny danych dla jednego z typów zabiegu. Niech  $F$  oznacza rozkład najlepiej przybliżający  $Y$ . Zadanie polegało zatem na wyestymowaniu parametrów rozkładu  $F$ , tzn.:

- $\mathbb{E}(F)$ ,
- $\text{Var}(F)$ ,
- $\text{Skew}(F)$  - skośność (tylko dla rozkładu skośno-normalnego).

Do estymacji parametrów rozkładu  $F$  wykorzystana została funkcja **fitdist** znajdująca się w pakiecie **fitdistrplus**. Wykorzystuje ona kilka metod dopasowywania rozkładów :

- mle - estymator największej wiarygodności,
- mge - estymator najmniejszej odległości.

Z rozkładów o wyestymowanych parametrach losowana była próba:

$$X : X_1, X_2, \dots, X_n \sim F,$$

która następnie była wykorzystywana do przeprowadzania testów.

W dalszej części użyte zostały następujące oznaczenia:

- $X_N$  - próba rozkładu normalnego,
- $X_{SN}$  - próba rozkładu skośno-normalnego.

W podrozdziale tym zostaną przedstawione metody estymacji parametrów rozkładów teoretycznych oraz opis teoretyczny zastosowanych estymatorów.

### Estymator najmniejszej odległości

**Definicja 10.** Niech  $X : X_1, X_2, \dots, X_n$  - losowa próba (iid) z rozkładu  $X \sim F(x, \theta)$ , gdzie  $\theta \in \Theta$  i  $\Theta \subseteq \mathbb{R}^k$  ( $k \geq 1$ )

Wprowadzamy oznaczenia :

- $\hat{F}_n(x)$  - dystrybuanta empiryczna,
- $\hat{\theta}$  - estymator  $\theta$ ,
- $d[.,.]$  - funkcja zwracającą odległość pomiędzy dwoma argumentami.

**Estymatorem najmniejszej odległości** parametru  $\theta$  nazywamy  $\hat{\theta} \in \Theta$  takie, że

$$d[F(x, \hat{\theta}), \hat{F}_n(x)] = \inf_{\theta \in \Theta} \{d[F(x, \theta), \hat{F}_n(x)]\}.$$

W funkcji **fitidst** można użyć jednej z trzech metod liczenia odległości pomiędzy rozkładami:

- Cramér von Mises **CvM** -  $\int [\hat{F}_n(x) - F_x(x, \theta)]^2$ ,
- Kołmogorow-Smirnow **KS** -  $\sup_{-\infty < x < \infty} |\hat{F}_n(x) - F_x(x, \theta)|$ ,
- Anderson Darling **AD** -  $\int \text{Var}(\hat{F}_n(x)) [\hat{F}_n(x) - F_x(x, \theta)]^2$ .

### Estymator największej wiarygodności

**Definicja 11.** **Wiarygodnością** nazywamy funkcję  $\Lambda : \Theta \rightarrow \mathbb{R}$ :

$$\Lambda(\theta) = f_{\theta}(x).$$

**Definicja 12.** **Estymatorem największej wiarygodności** dla parametru  $\theta$  nazywamy estymator **ENW**( $\theta$ ) =  $\hat{\theta} = \hat{\theta}(X)$ , jeśli dla dowolnego  $x$

$$f_{\hat{\theta}(x)}(x) = \sup_{\theta \in \Theta} f_{\theta}(x).$$

Równoważnie  $\text{ENW}(\theta) = \Lambda(\hat{\theta}) = \sup_{\theta \in \Theta} \Lambda(\theta)$ .

### Analiza rozkładu kolumny **EF**

Analiza **EF** ma za zadanie poza wyznaczeniem rozkładu najlepiej przybliżającego rozkład empiryczny, także wskazanie różnic otrzymywanych przy zastosowaniu różnych metod estymacji parametrów.

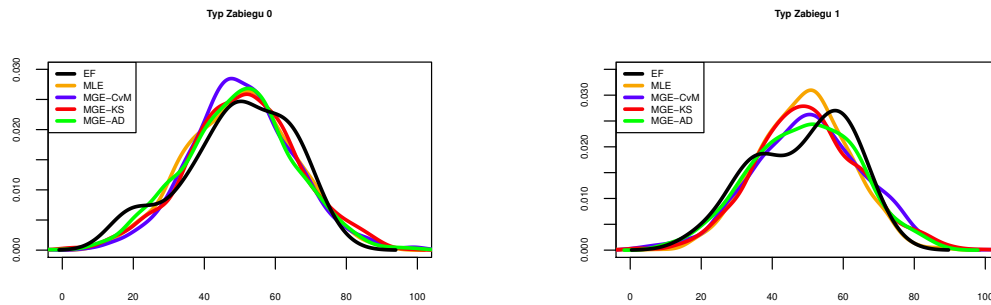
Wśród wyestymowanych za pomocą różnych metod parametrów widać znaczące różnice. Dotyczy to przede wszystkim rozkładu skośno-normalnego, gdzie różne metody wyestymowały różne kierunki skośności. W tablicy możemy dostrzec puste pozycje, które są efektem niemożliwości wyestymowania parametrów rozkładu skośno-normalnego za pomocą metody **MGE-CvM**. Wynika to z implementacji jednej z metod pakietu do generacji rozkładów skośno-normalnych.

Rysunek 3.4 przedstawia rozkłady empiryczne **EF** w zależności od **Typu Zabiegu** oraz rozkłady normalne z wyestymowanymi parametrami.

Z rysunku 3.4 możemy odczytać, że rozkłady z wyestymowanymi parametrami dobrze przybliżają rozkład **EF**, zwłaszcza wartość średnią i wariancję. Większe różnice występują

	Rozkład normalny		Rozkład skośno-normalny		
	mean	sd	location	scale	shape
Typ Zabiegu 0					
MLE	49.4773	14.7994	69.1578	24.6223	-5.2568
MGE-CvM	50.6457	15.0665	NA	NA	NA
MGE-KS	50.8320	15.4962	22.6683	37.8048	1.4290
MGE-AD	50.1604	15.5768	42.7280	17.3995	0.3238
Typ Zabiegu 1					
MLE	48.6522	13.3176	66.3319	22.1366	-5.6478
MGE-CvM	49.4546	15.3893	NA	NA	NA
MGE-KS	49.8860	14.6585	45.1644	7.5218	-0.8405
MGE-AD	49.1239	14.8472	30.8635	19.2018	1.6745

Tablica 3.4: Wyestymowane parametry rozkładów przybliżających EF.

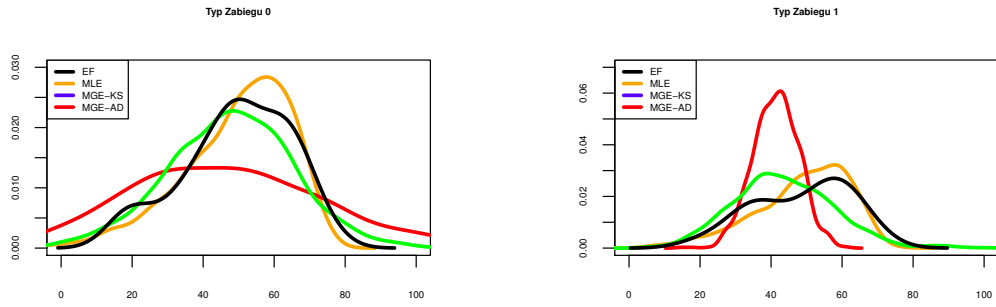


Rysunek 3.4: Wykres przedstawia rozkład rzeczywisty (kolor czarny) oraz próbki z rozkładów normalnych przybliżających **EF** z parametrami wyestymowanymi za pomocą różnych metod, odpowiednio: pomarańczowy - **MLE**, niebieski - **MGE-CvM**, czerwony - **MGE-KS**, zielony - **MGE-AD**.

jedynie w miejscach, gdzie zaburzona jest symetria (np. w rozkładzie dla Typu Zabiegu 1), co jest zgodne z naszymi przewidywaniami, ponieważ rozkład normalny jest rozkładem symetrycznym.

Z rysunku 3.5 możemy odczytać efekt opisanych wyżej różnic wartości wyestymowanych parametrów. Otrzymane rozkłady bardzo się od siebie różnią. Zauważamy, że metoda **MGE** z funkcją odległości **KS** w obu przypadkach daje najgorsze przybliżenie. W szczególności bardzo słabo estymowana jest wariancja, a skośność nie jest w ogóle uwidoczniiona. Pozostałe dwie metody lepiej przybliżają rozkład empiryczny chociaż widać znaczne różnice pomiędzy nimi. W obu przypadkach metody **MLE** i **MGE-AD** wyestymowały skośności o przeciwnych kierunkach i jedynie ta pierwsza osiągnęła kierunek zgodny z rozkładem empirycznym. Można zatem dojść do wniosku, że to metoda **MLE** estymuje parametry, przy których rozkład skośno-normalny najlepiej przybliża rozkład **EF**.





Rysunek 3.5: Wykres przedstawia rozkład rzeczywisty (kolor czarny) oraz próbki z rozkładów skośnie-normalnych przybliżających **EF** z parametrami wyestymowanymi za pomocą różnych metod, odpowiednio: pomarańczowy - **MLE**, czerwony - **MGE-KS**, zielony - **MGE-AD**.

	Wartość średnia $\mu$	Wariancja $\sigma$
EF 0	49.4773	14.7994
EF 1	48.6522	13.3176
AVA 0	0.6118	0.1611
AVA 1	0.6061	0.1249
AVApост 0	1.6006	0.1647
AVApост 1	1.4383	0.2001
MAX 0	92.9333	29.9543
MAX 1	81.9217	28.7731
MAXpost 0	20.2023	6.1747
MAXpost 1	21.0682	8.6370
MEAN 0	56.1622	20.1004
MEAN 1	65.3783	32.3570
MEANpost 0	10.4372	3.4315
MEANpost 1	11.3333	4.4863

Tablica 3.5: Parametry rozkładów normalnych przybliżających rozkłady teoretyczne wyestymowane metodą **MLE** za pomocą funkcji *estimateParameters* znajdującą się w skrypcie załączonym w dodatku A.2.

### Zbiorcza analiza rozkładów poszczególnych kolumn

Podobna analiza, jak ta znajdująca się w powyższym rozdziale powtórzona dla reszty kolumn prowadzi do analogicznych wyników. W związku z tym nie będziemy prezentować jej szczegółowo. W dodatku A.3 znajduje się kod umożliwiający przeprowadzenie analizy.

Analiza danych prowadzi do wniosku, że najbardziej uniwersalną i zarazem skuteczną metodą estymacji parametrów jest metoda **MLE**. Z tego względu w dalszych rozważaniach będziemy korzystać tylko z parametrów wyestymowanych za jej pomocą.

Są one umieszczone w tablicach 3.5 i 3.6.

	Wartość średnia $\mu$	Wariancja $\sigma$	Skośność $\alpha$
EF 0	69.1578	24.6223	-5.2568
EF 1	66.3319	22.1366	-5.6478
AVA 0	0.4084	1.0390	3.1478
AVA 1	0.3679	1.2960	7.3829
AVApост 0	1.3360	1.1857	4.3061
AVApост 1	1.0149	0.5693	-0.0033
MAX 0	64.8284	41.0680	1.6995
MAX 1	106.0696	37.5627	-1.3730
MAXpost 0	20.2330	6.1738	-14.4445
MAXpost 1	21.1160	8.6384	-30.8084
MEAN 0	37.4773	27.4439	1.6475
MEAN 1	29.2468	48.5017	2.6560
MEANpost 0	9.7185	2.5066	-1.1712
MEANpost 1	10.8000	2.5275	-0.1507

Tablica 3.6: Parametry rozkładów skośno-normalnych przybliżających rozkłady teoretyczne wyestymowane metodą **MLE** za pomocą funkcji *estimateParameters* znajdującą się w skrypcie załączonym w dodatku A.2.

## Rozdział 4

# Ewaluacja testów

### 4.1. Porównania testów klasycznych z ich odpowiednikami bootstrapowymi

#### 4.1.1. Opis eksperymentów

Celem rozdziału jest sprawdzenie działania testów klasycznych i ich odpowiedników bootstrapowych na rozkładach teoretycznych przybliżających rozkłady bazowe<sup>1</sup>. Zgodnie z wynikami analizy z rozdziału 3 rozważamy rozkłady normalne i skośnie-normalne z wyestymowanymi parametrami znajdującymi się odpowiednio w tablicach 3.5 oraz 3.6. W eksperymentach analizowano wartości błędów pierwszego rodzaju oraz drugiego rodzaju (odpowiadający dopełnieniu mocy testu).

Zastosowano dwa rodzaje doświadczeń, nazwanych przez nas odpowiednio jako: *statyczne* i *dynamiczne*. Pierwszy rodzaj eksperymentów służy do wyznaczenia wartości błędów badanych testów. Drugi typ eksperymentów ma natomiast pokazać, jak przesuwanie rozkładów (czyli manipulowanie parametrami) wpływa na moc testu oraz czy istnieje zależność pomiędzy wartością parametrów a zaistniałymi błędami. Poniżej znajduje się opis technik użytych w eksperymentach.

#### Eksperyment statyczny

Rozważamy dwa rozkłady  $F$  i  $G$  z ustalonymi parametrami. rzeprowadzamy (jeśli nie wskazano inaczej w konkretnym eksperymencie) 1000 losowań prób  $X \sim F$ ,  $Y \sim G$ , dla których realizujemy procedurę bootstrap. Wynikiem procedury jest p-wartość, którą porównujemy z oczekiwanym poziomem istotności  $\alpha$ . Testy realizujemy dla poziomów istotności

$$\alpha \in \{0.01, 0.02, 0.025, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 1\},$$

uzyskując w ten sposób wykres przedstawiający p-wartości testów. Każda z przeprowadzonych procedur bootstrap zawiera 1000 losowań realizacji prób.

---

<sup>1</sup>W dodatkach A.4 i A.5 znajduje się generyczny kod (udostępniający odpowiednio funkcje do testowania bootstrapowego oraz implementacje dwóch metod bootstrapowych) wykorzystywany w eksperymentach w niniejszym rozdziale

### **Eksperyment dynamiczny**

Niech  $\theta = \{\theta_1, \theta_2, \dots, \theta_k\}$  oznacza zbiór wartości jednego z parametrów rozkładu  $G$ . Przeprowadzamy ciąg eksperymentów. W pojedynczej instancji eksperymentu  $i \in \{1, 2, \dots, k\}$  rozważamy dwa rozkłady  $F$  i  $G_i = G(\theta_i)$ . Wykonujemy 1000 losowań prób  $X \sim F$ ,  $Y \sim G_i$ , dla których realizujemy procedurę bootstrap uzyskując p-wartość, którą porównujemy z poziomem istotności  $\alpha = 0.05$ . Każda procedura bootstrap zawiera 1000 losowań realizacji prób. Po przeprowadzeniu wszystkich eksperymentów otrzymujemy wykres przedstawiający związek między zmianą parametru, a jakością testu klasycznego i bootstrapowego.

#### 4.1.2. Test t-studenta

Niniejszy podrozdział jest poświęcony porównaniu klasycznego testu t-studenta do porównywania średnich dwóch prób z dwiema metodami bootstrapowymi do testowania hipotez, będących odpowiednikami testu t-studenta.

W sekcji 2.2.1 opisane zostały dwa podejścia do testowania hipotez metodą bootstrap. Były one jednocześnie obrazowane przykładami bazującymi na porównywaniu średnich. Owe przykłady są bootstrapowymi odpowiednikami klasycznego testu t-studenta.

W niniejszej części pracy porównujemy następujące trzy metody testowania hipotez:

1. test t-studenta,
2. metoda bootstrapowa testowania hipotez opierająca się na testowaniu równości rozkładów, używająca porównywania średnich (przykład opisany w podrozdziale 2.2.1),
3. metoda bootstrapowa testowania hipotez opierająca się na testowaniu równości parametru, używająca porównywania średnich (przykład opisany w podrozdziale 2.2.1).

#### Eksperymenty

Poniżej zostały przedstawione wyniki eksperymentów, porównujących klasyczny test t-studenta z jego bootstrapowymi odpowiednikami, wg schematu opisanego w podrozdziale 4.1.1. Za każdym razem pokazane zostały wyniki wszystkich trzech przedstawionych metod.

W trakcie przeprowadzania eksperymentów zauważyliśmy potrzebę wykonania dużej liczby iteracji, zarówno symulacji jak losowania prób bootstrapowych. W związku z tym, dla testów statycznych przyjęliśmy:

1. liczbę symulacji równą 10000,
2. liczbę losowań prób bootstrapowych równą 1000.

Z kolei dla testów dynamicznych przyjęliśmy:

1. liczbę symulacji równą 1000,
2. liczbę losowań prób bootstrapowych równą 1000.

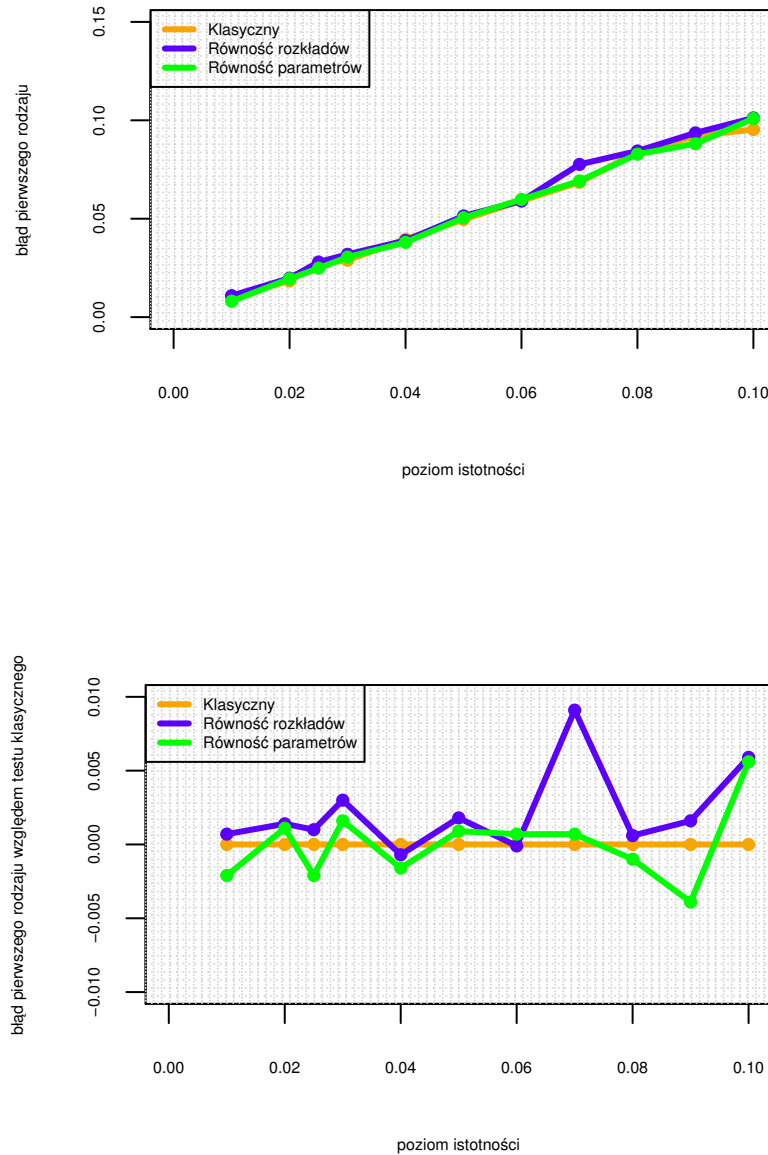
##### 1. Analiza błędów pierwszego rodzaju

Błędy pierwszego rodzaju zostały porównane przy użyciu eksperymentu statycznego (4.1.1), z parametrami:  $F \sim G = N(\mu = 1.6, \sigma = 29.95)$  Parametry rozkładu normalnego dobrane zostały na podstawie tablicy 3.5 i odpowiadają kolumnie **MEAN**.

Na podstawie wyników przedstawionych na rysunku 4.1 trudno jest stwierdzić, która metoda zachowuje się lepiej. Metody zachowują się podobnie, z wyjątkiem testu bootstrapowego metodą porównywania równości rozkładów, który zdaje się mieć większy błąd od pozostałych metod.

Zwracają uwagę niewielkie różnice pomiędzy wykresami odpowiadającymi testom:

- równość rozkładów - test klasyczny :  $\min = -0.0160$ ,  $\max = 0.0300$ ,
- równość parametrów - test klasyczny:  $\min = -0.0130$ ,  $\max = 0.0210$ ,
- równość rozkładów - równość parametrów:  $\min = -0.0220$ ,  $\max = 0.0150$ .



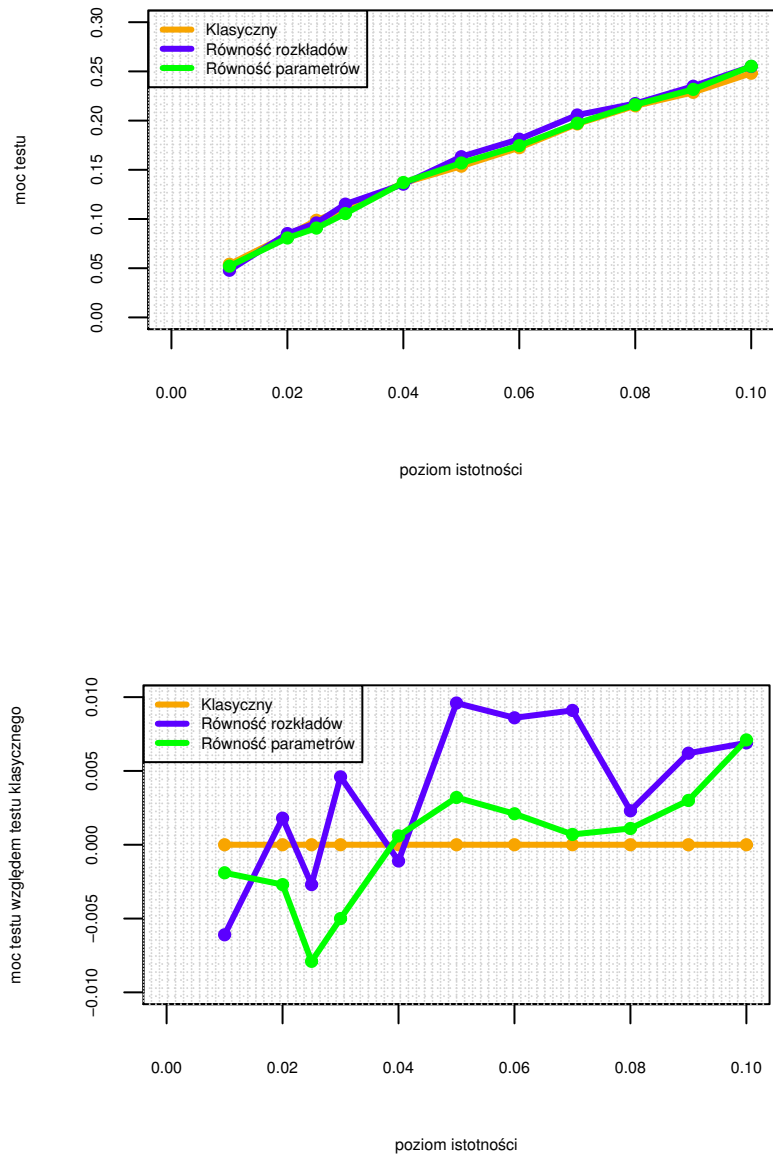
Rysunek 4.1: Wykresy przedstawiają błędy pierwszego rodzaju testów klasycznych i bootstrapowych w zależności od poziomu istotności testu zastosowanych do rozkładu normalnego dla problemu a) testowania równości rozkładów (kolor niebieski) b) testowania równości parametrów (kolor zielony), c) testowania klasycznego (kolor pomarańczowy).

## 2. Analiza mocy testu

Moce rozważanych testów zostały porównane na podstawie przeprowadzonych eksperymentów statycznych (4.1.1), porównując rozkłady  $F = N(\mu = 20.1, \sigma = 49.47)$  oraz  $G = N(\mu = 32.35, \sigma = 48.66)$ .

Parametry rozkładu normalnego dobrane zostały na podstawie tablicy 3.5 i odpowiadają

kolumnie **MAXpost** dla różnych typów zabiegu.



Rysunek 4.2: Wykresy przedstawiają moce testów klasycznych i bootstrapowych w zależności od poziomu istotności testu zastosowanych do odróżnienia rozkładów normalnych o przesuniętych parametrach dla problemu a) testowania równości rozkładów (kolor niebieski) b) testowania równości parametrów (kolor zielony), c) testowania klasycznego (kolor pomarańczowy).

Wyniki zaprezentowane na rysunku 4.2 są również trudne do zinterpretowania. Wszystkie trzy metody zachowują się podobnie, z delikatnym wskazaniem na przewagę metody bootstrapowej opierającej się na badaniu równości rozkładów.

Różnice pomiędzy testami wyglądają następująco:

- równość rozkładów - test klasyczny :  $\min = -0.0061$ ,  $\max = 0.0096$ ,
- równość parametrów - test klasyczny:  $\min = -0.0079$ ,  $\max = 0.0071$ ,
- równość rozkładów - równość parametrów:  $\min = -0.0042$ ,  $\max = 0.0096$ .

Można stwierdzić, że nie występują znaczące różnice.

### 3. Analiza szybkości rozróżniania rozkładów

Z rozdziału o analizie danych wynika, że potrzebna jest metoda, która wykrywa niewielkie różnice między próbami. W związku z tym analiza szybkości rozróżniania rozkładów oparta została o eksperymenty dynamiczne (4.1.1), które pokazują jak przy niewielkiej zmianie skośności rozkładu zmieniają się moce testów.

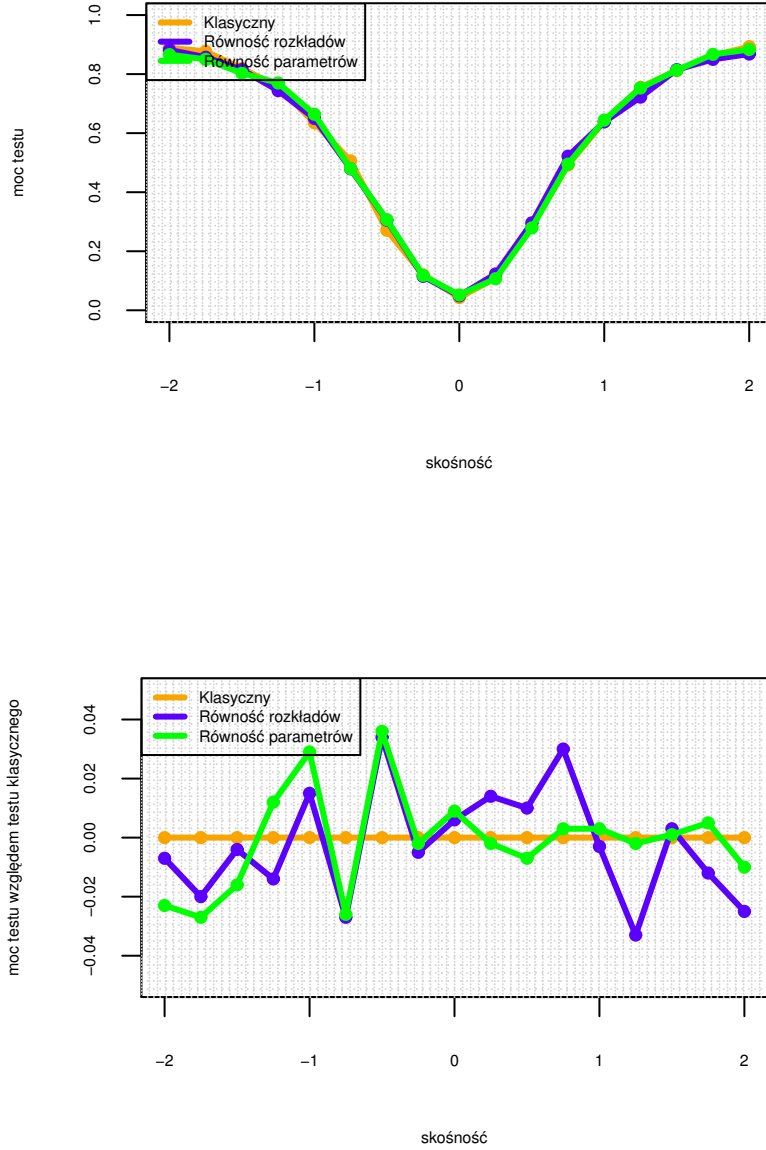
Rysunek 4.3 pokazuje, że występują nieznaczne wahania między trzema porównywanymi testami. Wydaje się jednak, że metoda polegająca na porównywaniu rozkładów dobrze sobie radzi z małymi odchyleniami. Zgodne z intuicją jest, że moce testów rosną wraz ze zmianą skośności do poziomu  $\sim 0.8$  przy zmianie skośności w zakresie  $\theta \in \{-2, 2\}$ .

Na podstawie przeprowadzonej analizy nie da się jednoznacznie określić, który z testów wypada najlepiej. Porównywane testy osiągają bardzo podobne wyniki.

Test bootstrapowy metodą porównywania rozkładów ma największy błąd pierwszego rodzaju i największą moc wg testów statycznych. Test t-studenta zachowuje się podobnie do testu bootstrapowego metodą porównywania parametrów. Z drugiej strony, bootstrapowa metoda testowania równości rozkładów najlepiej radzi sobie z rozróżnianiem rozkładów w niewielkim stopniu różniących się parametrem skośności.

Na podstawie powyższej analizy i wyciągniętych wniosków postanowiliśmy wybrać bootstrapowy test bazujący na porównywaniu rozkładów. W związku z tym, że występują niewielkie różnice między próbami, istotne jest to, żeby test był na takie różnice wrażliwy.





Rysunek 4.3: Wykresy przedstawiają moc testów klasycznych i bootstrapowych na poziomie istotności  $\alpha = 0.05$  zastosowane do rozróżnienia rozkładu normalnego  $F \sim N(0, 1)$  i skośno-normalnego przy zmieniającym się parametrze skośności  $G_i \sim SN(0, 1, \theta_i)$ , gdzie  $\theta_i \in \{-2, -1.75, \dots, 1.75, 2\}$  dla problemu a) testowania równości rozkładów (kolor niebieski) b) testowania równości parametrów (kolor zielony) c) testowania klasycznego (kolor pomarańczowy).

### 4.1.3. Test Wilcoxona

Podrozdział ten jest poświęcony porównaniu klasycznego testu rangowego Wilcoxona dla dwóch prób z zaadaptowaną do niego metodą bootstrapową.

#### Metoda bootstrapowa

Niech  $X$  i  $Y$  oznaczają losowe próby, z których oddzielnie losowane są realizacje prób bootstrapowych. W przypadku testu Wilcoxona zaadaptowano dwie różne metody testowania bootstrapowego opisane w rozdziale 2.2.1.

1. Testowanie równości rozkładów.

Realizację w  $i$ -tym kroku oznaczamy odpowiednio, jako  $X_i^*$  i  $Y_i^*$ , gdzie próby losowane były z połączonej próby  $(X, Y)$ .

Statystyka testowa w każdej realizacji próby bootstrapowej ma postać  $\omega_i = W(X_i^*, Y_i^*)$ , natomiast statystyką próby bazowej jest  $\omega = W(X, Y)$ .

P-wartość testu wynosi  $p = 2\min(p_1 + p_2)$ , gdzie:

- $p_1 = \frac{1}{R} \sum_{\omega \geq \omega_i} 1$ ,
- $p_2 = \frac{1}{R} \sum_{\omega \leq \omega_i} 1$ .

2. Testowanie równości parametru

W tym przypadku zastosowano metodę (opisaną w artykule [RZR05]) opartą o rangowy test Welcha dla hipotezy zerowej będącej równością stochastyczną rozkładów, czyli  $H_0 : P(X < Y) = P(X > Y)$ .

Zgodnie z 2.2.1 w pierwszym kroku należy zmodyfikować  $X$  i  $Y$  tak, aby zachodziła dla nich hipoteza zerowa  $H_0$ . Metoda transformacji opiera się na znalezieniu stałej normującej  $c$ , takiej aby dla  $X'$  i  $Y'$  otrzymanych w wyniku transformacji:  $X' = X$  i  $Y' = Y + c$  zachodziła  $H_0$ . W literaturze można znaleźć wiele propozycji wyboru stałej normującej<sup>2</sup>, jednakże zaleca się za każdym razem dostosować metodę do rozważanego rozkładu bazowego ([RZR05]).

Z tego względu zaimplementowaliśmy w języku  $R$  ogólny algorytm znajdowania stałej normującej polegający na binarnym wyszukaniu najlepszego  $c$ . Dokładniej  $c$  wyszukiwana jest z przedziału  $\{c_{\min} = \min(X, Y), c_{\max} = \max(X, Y)\}$ . W każdym kroku algorytmu sprawdzamy czy dla  $c = \frac{c_{\min} + c_{\max}}{2}$  zmodyfikowane próby  $X'$  i  $Y'$  spełniają  $H_0$  z dokładnością do  $\epsilon = 0.0001$ . Jeżeli hipoteza nie jest spełniona ograniczamy obszar przeszukiwań do  $\{c, c_{\max}\}$  lub  $\{c_{\min}, c\}$  w zależności czy statystyka na zmodyfikowanej próbie była odpowiednio mniejsza od zera, czy większa. Algorytm kończy działanie, jeśli hipoteza zerowa zostaje spełniona, bądź obszar przeszukiwań będzie mały tzn.  $|c_{\min} - c_{\max}| < \epsilon$ .

Kod zawierający transformację prób i wyszukiwanie stałej normującej znajduje się w załączniku A.6

Realizację w  $i$ -tym kroku oznaczamy odpowiednio, jako  $X_i^*$  i  $Y_i^*$ , gdzie próby losowane były oddzielnie z  $X'$  i  $Y'$ .

P-wartość testu jest analogiczna do przypadku testowania równości rozkładów i wynosi  $p = 2\min(p_1 + p_2)$ , gdzie:

---

<sup>2</sup>Często jako  $c$  wykorzystuje się medianę połączonej próby

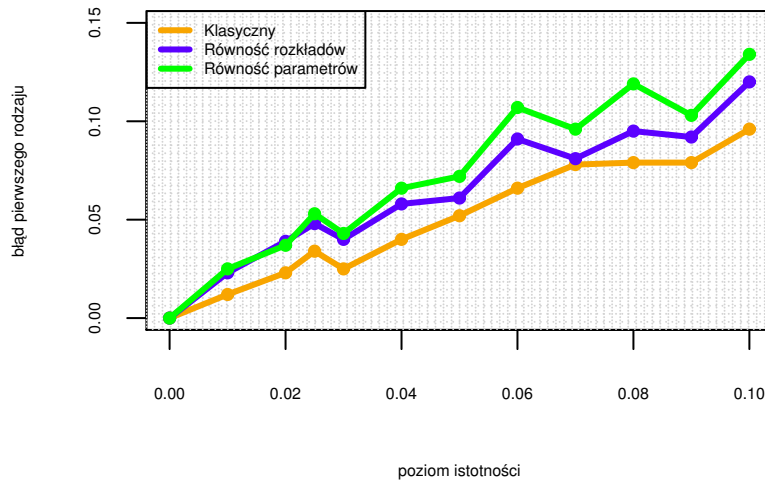
- $p_1 = \frac{1}{R} \sum_{t_{RW} \geq t_{RW}^{*(i)}} 1$ ,
- $p_2 = \frac{1}{R} \sum_{t_{RW} \leq t_{RW}^{*(i)}} 1$ .

## Eksperymenty

Poniżej znajdują się wyniki eksperymentów dla problemów testowania równości rozkładów oraz testowania równości parametrów. Miały one na celu wybór najodpowiedniejszej metody do testowania rozkładów bazowych, dlatego równoległe zaprezentowane zostały rezultaty obu metod bootstrapowych. Zastosowano eksperymenty opisane 4.1.1, które posługują się miarami mocy testu i wartościami błędów pierwszego rodzaju.

### 1. Analiza błędów pierwszego rodzaju

Na początku przyjrzymy się wartościom błędów pierwszego rodzaju testów. W tym celu wykonano eksperyment statyczny (4.1.1), gdzie  $F \sim G = N(\mu = 1.6, \sigma = 29.95)$  Parametry rozkładu normalnego dobrane zostały na podstawie tablicy 3.5 i odpowiadają kolumnie **MEAN**.



Rysunek 4.4: Wykresy przedstawiają błąd pierwszego rodzaju testów klasycznych i bootstrapowych w zależności od poziomu istotności testu zastosowanych do rozkładu normalnego dla problemu a) testowania równości rozkładów (kolor niebieski) b) testowania równości parametrów (kolor zielony).

Wyniki przedstawione na rysunku 4.4 świadczą, że obie metody bootstrapowe popełniają więcej błędów pierwszego rodzaju aniżeli klasyczny test Wilcozona. Ponadto porównując metody bootstrapowe to lepsze rezultaty osiąga testowanie równości rozkładów. Różnice pomiędzy testami są znaczące i wahają się kolejno :

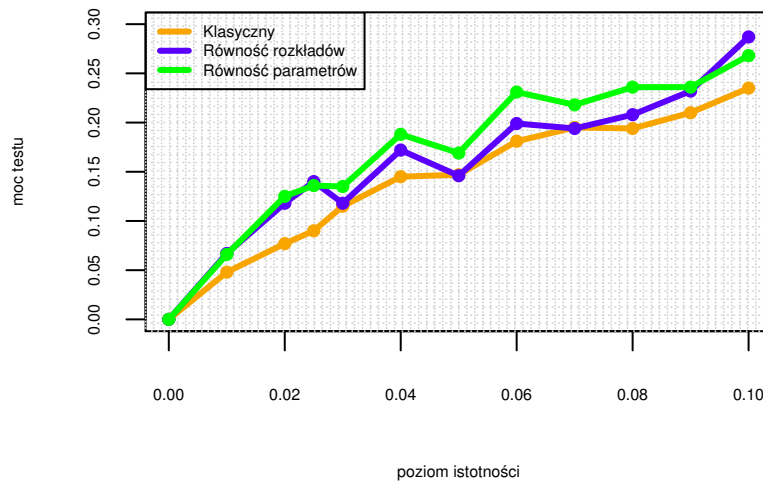
- Równość rozkładów - Test Klasyczny :  $\min = 0.009$ ,  $\max = 0.023$ ,

- Równość parametrów - Test Klasyczny:  $\min = 0.013$ ,  $\max = 0.041$ ,
- Równość parametrów - Równość rozkładów:  $\min = -0.002$ ,  $\max = 0.024$ .

## 2. Analiza mocy testu

Następnie rozważymy moc testu (dopełnienie błędu drugiego rodzaju). W tym celu wykonano eksperyment statyczny (4.1.1), gdzie  $F = N(\mu = 20.1, \sigma = 49.47)$  oraz  $G = N(\mu = 32.35, \sigma = 48.66)$

Parametry rozkładu normalnego dobrane zostały na podstawie tablicy 3.5 i odpowiadają kolumnie **MAXpost** dla różnych typów zabiegu.



Rysunek 4.5: Wykresy przedstawiają moc testów klasycznych i bootstrapowych w zależności od poziomu istotności testu zastosowanych do odróżnienia rozkładów normalnych o przesuniętych parametrach dla problemu a) testowania równości rozkładów (kolor niebieski) b) testowania równości parametrów (kolor zielony).

Wyniki zaprezentowane na rysunku 4.5 wskazują, że w przypadku mocy testu metody bootstrapowe osiągają lepsze wyniki od klasycznego testu Wilcoxona. Ponadto należy zauważyć, że różnice pomiędzy wynikami metod bootstrapowych, a testem klasycznym są znaczne i wynoszą nawet do 0.037 w przypadku testowania równości rozkładów oraz 0.05 w przypadku testowania równości parametrów. Obie metody bootstrapowe osiągają podobne moce testu i nie można jednoznacznie określić, które wyniki są lepsze.

## 3. Analiza szybkości rozróżniania rozkładów

Do tej analizy zastosowane zostaną eksperymenty dynamiczne (4.1.1). Wyniki przedstawiające moc testu w zależności od wielkości parametru skośności pokażą, który z testów najszybciej odróżnia rozkład normalny od skośnego. Jest to bardzo istotne z powodu niewielkich prób (o wielkości ok 40), aby już niewielkie różnice w rozkładach były dostrzegalne.

Z rysunku 4.6 można odczytać, że bootstrapowa metoda testowania równości parametrów w większości szybciej rozróżnia rozkład normalny od skośno-normalnego, jednakże różnice nie są bardzo duże.

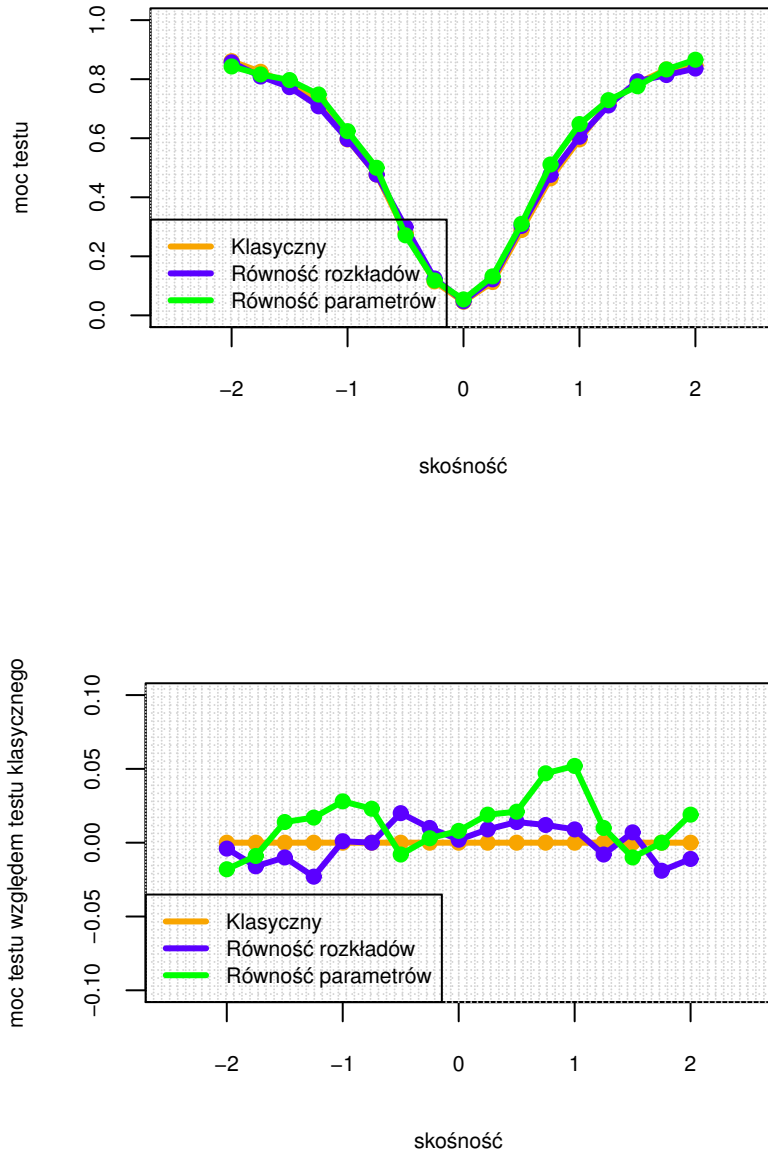
Można stwierdzić, że dla niewielkich skośności  $|\alpha| < 0.5$  testy zachowują się analogicznie. Wraz ze wzrostem parametru skośności różnice pomiędzy wynikami poszczególnych testów się zwiększają, ale nie jest to trend monotoniczny.

Na podstawie takiej analizy nie da się jednoznacznie określić, który z testów wypada najlepiej. Istotne jest, że wszystkie testy osiągają bardzo dobre wyniki, a moc testów rośnie wraz ze zmianą skośności do poziomu  $\sim 0.8$  przy skośności  $\alpha \in \{-2, 2\}$ . Jest to wynik bardzo dobry zważywszy na dosyć duże różnice skośności w rozkładach bazowych.

Podsumowując z przeprowadzonej analizy nie da się jednoznacznie określić, czy klasyczny test Wilcoxona lepiej się nadaje od jego odpowiedników bootstrapowych do testowania danych rzeczywistych.

Test klasyczny osiąga mniejsze błędy pierwszego rodzaju natomiast testy bootstrapowe uzyskują większe wartości mocy testu. Różnice w obu przypadkach są małe i wahają się w przedziale  $[-0.05, 0.05]$ . Ponadto bootstrapowa metoda testowania równości parametrów wydają się najlepiej rozpoznawać rozkłady różniące się parametrem skośności.

Na podstawie powyższej analizy i wyciągniętych wniosków uznaliśmy, że bootstrapowy odpowiednik testu Wilcoxona dla problemu testowania równości parametrów jest najwłaściwszy spośród rozważanych testów. Istotny dla nas jest przede wszystkim fakt, że rozważane próby są niewielkie (ok. 40) przez co bardzo ważne jest lepsze rozróżnianie rozkładów o niedużych różnicach parametrów, czyli bardziej znacząca w tym problemie według nas jest moc testu.



Rysunek 4.6: Wykresy przedstawiają moc testów klasycznych i bootstrapowych na poziomie istotności  $\alpha = 0.05$  zastosowane do rozróżnienia rozkładu normalnego  $F \sim N(0,1)$  i skośno-normalnego przy zmieniającym się parametrze skośności  $G_i \sim SN(0,1,\theta_i)$ , gdzie  $\theta_i \in \{-2, -1.75, \dots, 1.75, 2\}$  dla problemu a) testowania równości rozkładów (kolor niebieski) b) testowania równości parametrów (kolor zielony).

## Rozdział 5

# Eksperymenty na danych medycznych

Z analizy danych medycznych, przeprowadzonej w rozdziale 3 zostały wyciągnięte wnioski o porządkanych właściwościach testów statystycznych, które miałyby zostać użyte do wnioskowania na podstawie danych medycznych będących przedmiotem zainteresowania w niniejszej pracy. Pod tym kątem, w rozdziale 4 przeprowadzone zostały testy porównujące testy statystyczne: t-studenta i Wilcoxona z ich odpowiednikami bootstrapowymi. Różne empiryczne metody porównawcze (w sekcji 4.1.1 zostały wprowadzone metody statyczne i dynamiczne) wskazały, że najlepsze właściwości mają testy bootstrapowe: w przypadku porównywania średnich bazująca na porównywaniu rozkładów, a w przypadku testu Wilcoxona bazująca na porównywaniu parametrów (metody bootstrapowe zostały wprowadzone w rozdziale 2.2).

W niniejszym rozdziale umieszczony został opis wyników zastosowania wybranych testów do danych medycznych dotyczących operacji serca.

W rozdziale 3 zostało zauważone, że następujące pary cech są ze sobą skorelowane: Mean i Max oraz Maxpost i Meanpost. W związku z tym w następującej analizie nie rozpatrywano cech: Max i Maxpost.

W pierwszej kolejności, przeprowadzono testy statystyczne, porównujące grupy 0 i 1 względem każdej z wybranych cech. W bootstrapowym odpowiedniku testu Wilcoxona rozpatrywano hipotezę zerową:

$$H_0 : F_0 = F_1$$

przeciwko hipotezie alternatywnej

$$H_1 : F_0 \neq F_1.$$

W przypadku testu porównywania średnich:

$$H_0 : \bar{F}_0 - \bar{F}_1 = 0$$

przeciwko hipotezie alternatywnej

$$H_1 : \bar{F}_0 - \bar{F}_1 \neq 0.$$

P-wartości testów zastosowanych do porównania grup 0 i 1 względem każdej z cech zostały przedstawione w tablicach: 5.1 (dla testu porównywania średnich) i 5.2 (dla bootstrapowego odpowiednika testu Wilcoxona).

Zarówno wyniki testu porównywania średnich (tablica 5.1) jak i testu bootstrapowego odpowiednika Wilcoxona (tablica 5.2) wskazują na istotną różnicę względem cechy Avapost między grupami: 0 i 1. Wg nazewnictwa wprowadzonego w rozdziale 2, oba testy dają

Tablica 5.1: Wyniki bootstrapowego testu statystycznego do porównywania średnich bazującego na porównywaniu rozkładów, zastosowanego do danych medycznych. Hipoteza alternatywna ma postać obustronną. Zaprezentowano p-wartości.

EF	AVA	AVApost	Mean	MeanPost
0.8236	0.8794	0.0303	0.1755	0.3832

Tablica 5.2: Wyniki bootstrapowego odpowiednika testu Wilcoxona bazującego na porównywaniu parametrów, zastosowanego do danych medycznych. Hipoteza alternatywna ma postać obustronną. Zaprezentowano p-wartości.

EF	AVA	AVApost	Mean	MeanPost
0.5528	0.3784	0.0374	0.4116	0.6296

stosunkowo mocną przesłankę na różnicę między grupami. Względem każdej z pozostałych cech grupy nie różnią się statystycznie istotnie.

W związku z powyższym, przeprowadzono jednostronny test porównujący grupy 0 i 1 względem cechy AVApost. Mianowicie, hipoteza zerowa ma postać taką jak wcześniej, zaś alternatywna dla bootstrapowego odpowiednika testu Wilcoxona:

$$H_1 : F_0 > F_1,$$

oraz dla testu porównywania średnich:

$$H_1 : \bar{F}_0 - \bar{F}_1 > 0.$$

P-wartości testów zostały przedstawione w tablicy 5.3. Pokazują one mniejsze p-wartości niż wcześniej, chociaż w dalszym ciągu odpowiadające stosunkowo mocnej przesłance na różnicę między grupami. Wyciągamy z tego wniosek, że dane typu 0 w istotny statystycznie sposób mają większą wartość cechy Avapost. Oznacza to, że operacja typu 0 powoduje zwiększenie powierzchni przekroju ujścia aorty w większym stopniu niż operacja typu 1.

Tablica 5.3: Wyniki testów statystycznych z hipotezą alternatywną postaci jednostronnej, zastosowanych dla cechy Avapost. Zaprezentowano p-wartości.

porównanie średnich	bootstrapowy Wilcoxon
0.0139	0.0210



## Rozdział 6

# Podsumowanie

Podsumowując postawione przed nami zadanie polegało na ocenie skuteczności i zastosowaniu metody bootstrap do wnioskowania na temat leczenia pacjentów po zawale serca. Zostali oni poddani jednemu z dwóch typów zabiegu: pomostowaniu aortalno-wieńcowemu (*typ zabiegu 0*) lub angioplastyce tętnic wieńcowych (*typ zabiegu 1*). Należało zatem odpowiedzieć, która z metod daje lepsze rezultaty.

W pracy przedstawiliśmy teorię testowania hipotez za pomocą testów klasycznych oraz testów bootstrapowych. Opisaliśmy dwie wersje testowania bootstrapowego: testowanie równości rozkładów oraz testowanie równości parametrów. Omówiliśmy także szczegółowo dwa przykłady testów klasycznych: test dla porównywania średnich dwóch prób oraz rangowy test Wilcozona oraz dla każdego z nich zaprezentowaliśmy po dwie wersje odpowiedników bootstrapowych.

W ramach analizy opisaliśmy właściwości prób oraz zaprezentowaliśmy je w postaci tabel i histogramów. Głównym celem analizy było znalezienie rozkładów teoretycznych oraz wyestymowanie ich parametrów, tak aby jak najlepiej przybliżały one rozkłady bazowe. Za pomocą testu Shapiro-Wilka przyjęliśmy, że poszczególne kolumny można przybliżać za pomocą rozkładu normalnego. Zauważyliśmy na podstawie wyestymowanych parametrów oraz histogramów prób, że niektóre z rozkładów są skośne. Z tego względu zaproponowaliśmy do przybliżania prób rozkład skośno-normalny, jako naturalne rozszerzenie scentrowanego rozkładu normalnego o właściwość skośności. Dalej opisaliśmy różne metody estymacji parametrów, a za pomocą funkcji `fitdist` znaleźliśmy rozkłady najlepiej przybliżające rozkłady bazowe.

Przy wykorzystaniu znalezionych rozkładów testowaliśmy testy klasyczne i ich odpowiedniki bootstrapowe. Zaproponowaliśmy dwie metody testowania - nazwane przez nas metodami: statycznymi i dynamicznymi. Jako miary porównującej poszczególne testy użyliśmy błędy pierwszego rodzaju i moce testów. Na podstawie przeprowadzonych analiz stwierdziliśmy, że zarówno w przypadku testu dla porównywania średnich dwóch prób oraz testu Wilcozona ciężko jest jednoznacznie wyznaczyć najlepszy spośród rozważanych odpowiedników. Zdecydowaliśmy, jednak że w przypadku testu t-studenta najlepiej działa bootstrapowy wersja testowania równości rozkładów natomiast w przypadku testu Wilcozona najlepsze wyniki otrzymujemy za pomocą bootstrapowego testowania równości parametrów.

Następnie wybrane testy zastosowaliśmy na danych medycznych. W pierwszym kroku, przeprowadziliśmy testy statystyczne, porównujące grupy 0 i 1 względem każdej z wybranych

cech. Rezultaty obu przeprowadzonych przez nas testów nie wykazały różnicy pomiędzy wynikami badań z przed zastosowania leczenia, natomiast wskazały na istotną różnicę względem cechy AVApost między grupami: 0 i 1. W związku z powyższym przeprowadziliśmy testy jednostronne względem cechy AVApost. Uzyskane p-wartości wskazują, że dane typu 0 w istotny statystycznie sposób mają większą wartość cechy AVApost, co oznacza, że metoda 0 jest korzystniejsza dla pacjentów. Niestety z analizy danych wiemy, że aż w 57% przypadkach dane w kolumnie AVApost zawierały pustą pozycję. Możliwe zatem, że otrzymane wyniki są efektem niedokładności danych. Wskazuje na to fakt, że w przypadku cech EF, MEANpost i MAXpost nie stwierdzono dominacji żadnej z metod. Z tego względu opierając się na otrzymanych danych nie możemy jednoznacznie określić, czy któraś z metod leczenia jest korzystniejsza. Najnowsze badania medyczne wskazują, że metoda balonikowania, jako mniej inwazyjna jest równie skuteczna, co operacje na otwartym sercu<sup>1</sup>.

---

<sup>1</sup><http://test-naukawpolsce.pap.pl/aktualnosci/news,381175,balonikowanie-serca-rownie-skuteczne-jak-operacje-na-otwartym-sercu.html>

## Dodatek A

# Implementacja eksperymentów w R

### A.1. Rozszerzona biblioteka udostępniająca rozkład skośno-normalny

```
#SkosnoNormalnyBilblioteka.R
#Rozszerzona biblioteka rozkładu skośno-normalnego
library(VGAM)

R_MAX <- 10000
#Funkcje wymagane do obliczeń fitdist dla rozkładu skośno-normalnego
qsnorm <- function(p,location = 0, scale = 1, shape = 0){
  R <- R_MAX
  dist = rsnorm(R,location,scale,shape)
  quantile(dist,p)
}

psnorm <- function(q,location = 0, scale = 1, shape = 0){
  R <- R_MAX
  dist = rsnorm(R,location,scale,shape)
  sapply(q, function(qq){length(which(dist < qq))/R})
}
```

### A.2. Implementacja estymacji parametrów

Skrypt umożliwiający wyestymowanie parametrów.

```
#EstymacjaParametrowBiblioteka.R

library(fitdistrplus)
source("SkonsoNormalnyBiblioteka.R")

#Funkcja estymująca parametry tylko metodą MLE
# x - rozkład bazowy
estimateParameters <- function(x){
  x <- x[!is.na(x)]
  norm <- fitdist(
```

```

        data = x,
        distr = "norm",
        method = "mle"
    )
snorm <- fitdist(
    data = x,
    distr = "snorm",
    start = list(
        location = mean(x),
        scale = sd(x),
        shape = skewness(x)
    ),
    method = "mle"
)
list(norm=norm$estimate, snorm=snorm$estimate)
}

#Funkcje estymujące rozkład empiryczny rozkładami teoretycznymi
# x - rozkład bazowy
estimateNormParameters <- function(x){
#Estymowanie rozkładem normalnym
x <- x[!is.na(x)]
normMLE <- fitdist(data=x, distr="norm",method="mle")
normMGECvM <- fitdist(data=x, distr="norm",method="mge",gof="CvM")
normMGEKS <- fitdist(data=x, distr="norm",method="mge",gof="KS")
normMGEAD <- fitdist(data=x, distr="norm",method="mge",gof="AD")
result <- rbind(
    normMLE$estimate,
    normMGECvM$estimate,
    normMGEKS$estimate,
    normMGEAD$estimate
)
rownames(result) <- c("mle", "mge CvM", "mge KS", "mge KS")
result
}

estimateSNormParameters <- function(x){
#Estymowanie rozkładem skośno normalnym
x <- x[!is.na(x)]
snormMGEAD <- fitdist(
    data = x,
    distr = "snorm",
    start = list(
        location = mean(x),
        scale = sd(x),
        shape = skewness(x)
    ),
    method="mge",
    gof="AD"

```

```

    )
    snormMLE <- fitdist(
      data = x,
      distr = "snorm",
      start = list(
        location = mean(x),
        scale = sd(x),
        shape = skewness(x)
      ),
      method="mle"
    )
    snormMGECvM <- fitdist(
      data = x,
      distr = "snorm",
      start = list(
        location = mean(x),
        scale = sd(x),
        shape = skewness(x)
      ),
      method="mge",
      gof="CvM"
    )
    snormMGEKS <- fitdist(
      data = x,
      distr = "snorm",
      start = list(
        location = mean(x),
        scale = sd(x),
        shape = skewness(x)
      ),
      method="mge",
      gof="KS"
    )
    result <- rbind(
      snormMLE$estimate,
      snormMGECvM$estimate,
      snormMGEKS$estimate,
      snormMGEAD$estimate
    )

    rownames(result) <- c("mle", "mge CvM", "mge KS", "mge KS")
    result
  }

```

### A.3. Analiza rozkładów poszczególnych kolumn

Kod umożliwiający analizę rozkładów poszczególnych kolumn.

```

#AnalizaKolumn.R
source("EstymacjaParametrowBiblioteka.R")

methodNames <- c("MLE", "MGE-CvM", "MGE-KS", "MGE-AD")

plotDistributions <- function(X,
                             data,
                             data.seq,
                             distr,
                             n=1000,
                             title="",
                             xlim = c(0,1),
                             ylim = c(0,1),
                             xlab="",
                             ylab="",
                             names = c("Data",methodNames),
                             colours = c("black", "orange","blue", "red", "green"),
                             legend_position = "topleft",
                             filename,
                             width = 4.5,
                             height = 3.5,
                             cex = 0.5
){
  #X - próba bazowa
  #data - macierz z wyestymowanymi parametrami
  #data.seq - wiersze które wykorzystujemy w analizie np c(1,2,3)
  #distr - rozkład np rnorm
  #n=1000 - ilość losowanych próbek
  #title="" - Tytuł
  #xlim, ylim, xlab, ylab - parametry wykresów patrz : ?plot
  #names - nazwy poszczególnych rozkładów
  #colours - lista kolorów
  #legend_position - pozycja legendy na wykresie :
  # - jedna z wartości "bottomright", "bottom", "bottomleft",
  # "left", "topleft", "top", "topright", "right" and "center"
  #filename - ścieżka do pliku z rozszerzeniem eps
  #width, height - szerokość i wysokość wykresu
  #cex - wielkość liter
  X <- X[!is.na(X)]
  postscript(file=filename, encoding="ISOLatin2",
             paper="special", width = width,
             height = height, horizontal=FALSE)
  plot(NA, xlim = xlim, ylim = ylim, main=title, ylab = ylab, xlab = xlab,
       cex = cex, cex.lab=cex, cex.axis=cex, cex.main=cex, cex.sub=cex)
  sapply(data.seq,function(i){
    dens <- do.call(distr, as.list(c(n=n, data[i,])))
    lines(density(dens), col = colours[i + 1], lwd=3)
  })
  lines(density(X), col = colours[1], lwd=3)
}

```

```

legend(legend_position, # places a legend at the appropriate place
      names, # puts text in the legend
      lty=rep(1,length(data) + 1), # gives the legend appropriate symbols (lines)
      lwd=rep(3,length(data) + 1), col=colours,
      cex = cex
) # gives the legend lines the correct color
dev.off()
}

#Funkcja przeprowadzająca pełną analizę wybranej kolumny
analise_column <- function(data_, file_, title_){
  norm_ <- estimateNormParameters(data_)
  snorm_ <- estimateSNormParameters(data_)
  table_ <- round(cbind(norm_,snorm_), 4)
  rownames(table_) <- methodNames
  latex(table_, file=paste(file_, ".tex", sep=""))
  plotDistributions( data_, norm_ , title=title_, ymax =0.03, xmax = 100)
  plotSNDistributions(data_, snorm_, title=title_, ymax =0.06, xmin=-1, xmax = 100)
  table_
}

```

## A.4. Implementacja metod bootstrapowych

### 1. Testowanie równości rozkładów

```

boot_method_distribution_equality <-
function(x, y, test, R, single_sided = F){
  #Testowanie rownosci rozkładów
  #
  # Argumenty:
  #   x: pierwsza próba,
  #   y: druga próba,
  #   test: statystyka testowa,
  #   R: ile bootstrapowych losowań ma zostać wykonanych.
  #   single_sided: czy testowanie ma być jednostronne.
  #
  # Zwraca:
  #   estymowaną p-wartość.
  boot_fun <- function(data, indices, x_rows, y_rows){
    data <- data[indices]
    x_data <- data[1:x_rows]
    y_data <- data[(x_rows + 1) : (x_rows + y_rows)]
    m <- test(x_data, y_data)
    return(m)
  }
  bootobj <- boot(data=c(x, y),statistic=boot_fun,
    R=R, x_rows=length(x), y_rows=length(y))
}

```

```

if(single_sided){
  asl_boot <- sum(bootobj$t >= bootobj$t0)/length(bootobj$t)
}
else{
  asl_boot <- sum(abs(bootobj$t) >= abs(bootobj$t0))/length(bootobj$t)
}
return(asl_boot)
}

```

## 2. Testowanie równości parametrów

```

boot_method_parameters_equality <-
function(x, y, test, hipoteza, R, single_sided = F){
  #Testowanie rownosci rozkładów
  #
  # Argumenty:
  #   x: pierwsza próba, po transformacji tak aby
  #       test(x, y) było równe hipotezie H0,
  #   y: druga próba, po transformacji tak aby
  #       test(x, y) było równe hipotezie H0,
  #   test: statystyka testowa,
  #   hipoteza: wartość statystyki testowej
  #               dla pierwotnych prób (przed transformacją),
  #   R: ile bootstrapowych losowań ma zostać wykonanych.
  #   single_sided: czy testowanie ma być jednostronne.
  #
  # Zwraca:
  #   estymowaną p-wartość.
  stratum <- c(rep(1,length(x)),rep(2,length(y)))
  sampstatistic <- function(data, i, lenx){
    n <- length(i)
    part_x <- data[i[i<=lenx]]
    part_y <- data[i[i>lenx]]
    m <- test(part_x, part_y)
    return(m)
  }
  bootobj = boot(data=c(x, y), statistic=sampstatistic, strata=stratum,
                 R=R, lenx = length(x))
  if(single_sided){
    asl_boot <- sum(bootobj$t >= hipoteza)/length(bootobj$t)
  }
  else{
    asl_boot <- sum(abs(bootobj$t) >= abs(hipoteza))/length(bootobj$t)
  }
  return(asl_boot)
}

```



## A.5. Przykład użycia metod bootstrapowych

1. Funkcja ewaluująca test statystyczny.

```
get_results_classic <- function(alpha, simreps, test, genx, geny)
{
  #Testowanie testu bootstrapowego przy pomocy wielu losowań prób.
  #
  # Argumenty:
  #   alpha: poziom istotności,
  #   simreps: ile powtórzeń symulacji ma zostać wykonanych,
  #   test: statystyka testowa,
  #   genx: generator pierwszej próby,
  #   geny: generator drugiej próby.
  #
  # Zwraca:
  #   błąd pierwszego rodzaju albo moc testu, w zależności od tego,
  #   czy genx i geny zadają ten sam rozkład czy różne.
  cnt = 0
  for(j in 1:simreps){
    res = test(genx(), geny())
    if(res$p.value < alpha){
      cnt = cnt + 1
    }
  }
  return(cnt/simreps)
}
```

2. Funkcja ewaluująca metodę bootstrapową przy pomocy wielokrotnych powtórzeń.

```
get_results_boot <- function(boot_method, alpha, R, simreps, test, genx, geny)
{
  wrapped_test <- function(x, y){
    return(data.frame(p.value = boot_method(x, y, test, R)))
  }
  return(get_results_classic(alpha, simreps, wrapped_test, genx, geny))
}
```

3. Funkcja realizująca testowanie bootstrapowe metodą porównywania parametrów, z zastosowaniem transformacji dla przykładu testowania średnich.

```
boot_method_parameters_equality_means <- function(x, y, statistic,
                                                    repetitions, single_sided = F){
  hypothesis0 <- statistic(x, y)
  x2 <- x + (mean(c(x, y)) - mean(x))
  y2 <- y + (mean(c(x, y)) - mean(y))
  return(boot_method_parameters_equality(x2, y2, statistic, hypothesis0,
                                          repetitions, single_sided))
}
```

4. Przykład zastosowania funkcji bootstrapowych dla testowania średnich: przybliżenie błędu pierwszego rodzaju na podstawie porównywania prób z rozkładu normalnego o parametrach estymowanych na podstawie danych medycznych (patrz: rozdział 4.1.2).

```
simrepetitions <- 10000
bootrepetitions <- 1000
n <- 30

# statystyka dla równości rozkładów
hypothesis_rownosc_rozkladow <- function(x,y){
  return(mean(x) - mean(y))
}
# statystyka dla równości parametrów
hypothesis_rownosc_parametrow <- function(x, y){
  return((mean(x) - mean(y))/sqrt(var(x)/length(x) + var(y)/length(y)))
}

mean=1.6
sd=29.95

genx <- function(){
  return(rnorm(n, mean, sd))
}

alpha_seq <- c( 0.01, 0.02, 0.025, 0.03, 0.04, 0.05,
0.06, 0.07, 0.08, 0.09, 0.1)
result_boot1 <- c()
result_boot2 <- c()
result_classic <- c()

for(alpha in alpha_seq)
{
  resultsStatic_boot1 <- get_results_boot(boot_method_distribution_equality,
alpha, bootrepetitions, simrepetitions,
hypothesis_rownosc_rozkladow,
genx, genx)
  result_boot1 <- c(result_boot1, resultsStatic_boot1)

  resultsStatic_boot2 <- get_results_boot(boot_method_parameters_equality_means,
alpha, bootrepetitions, simrepetitions,
hypothesis_rownosc_parametrow,
genx, genx)
  result_boot2 <- c(result_boot2, resultsStatic_boot2)

  result_classic <- c(result_classic, get_results_classic(alpha, simrepetitions,
t.test, genx, genx))
}
```

## A.6. Implementacja transformacji rozkładów rangowych

Dla prób bazowych  $X$ ,  $Y$  wynikiem działania funkcji są odpowiednio próby  $X_*$ ,  $Y_*$  spełniające hipotezę zerową w testowaniu równości parametru będącym statystyką Welcha.

```
minimum_mean <- 0.001
minimum_c     <- 0.001
n1 <- length(X)
n2 <- length(Y)
transform <- function(c1, c2, X, Y){
  c      <- (c1+c2)/2
  Y_     <- Y + c
  ranks  <- rank(c(X,Y_))
  r1     <- ranks[1:n1]
  r2     <- ranks[(n1 + 1):(n1+n2)]
  m1     <- mean(r1)
  m2     <- mean(r2)
  ifelse(
    abs(m1 - m2) < minimum_mean || abs(c1- c2) < minimum_c ,
    return(list(X_ = X, Y_ = Y_)),
    ifelse(
      m1 - m2 < 0,
      return( transform(c1, c,  X, Y) ),
      return( transform(c,  c2, X, Y) )
    )
  )
}

max_c <- max(abs(X)) + max(abs(Y))
transform_result <- transform(-max_c, max_c, X, Y)
X_ <- transform_result$X_
Y_ <- transform_result$Y_
```



# Bibliografia

- [CI81] W.J. Conover and Ronald L. Iman. Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, 35:124–129, 1981.
- [ET93] Brad Efron and Rob Tibshirani. *An introduction to bootstrap*. Chapman & Hall, 1993.
- [Gaj] Piotr Gajewski. Uniwersalna definicja zawału serca. *Medycyna Praktyczna*, (1 (203)).
- [KM06] Jacek Koronacki and Jan Mielniczuk. *Statystyka*. Wydawnictwa Naukowo-Techniczne, 2006.
- [MW47] H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1):50–60, 1947.
- [Nie11] Wojciech Niemiro. *Statystyka*. Wydział Matematyki i Informatyki, Uniwersytet Mikołaja Kopernika, Toruń oraz Instytut Matematyki i Mechaniki, Uniwersytet Warszawski, 2011.
- [RZR05] Jenő Reiczigel, Ildikó Zakarias, and Lajos Rozsa. A bootstrap test of stochastic equality of two populations. *The American Statistician*, 59:156–161, 2005.