Analiza Przeżycia

Raport 3

Wiktor Niedźwiedzki (258882)
Filip Michewicz (282239)

30 grudnia 2025 Anno Domini

Spis treści

Spis wykresów

Spis tabel

1 Dane *lung*

W sprawozdaniu każda lista wykorzystuje dane *lung* z pakietu *survival* w R, które dotyczą pacjentów z zaawansowanym rakiem płuc.
Zbiór danych *lung* zawiera **228** oraz **10** cech. Liczba brakujących danych wynosi **67**, z czego przypada ona na **61** pacjentów.

Znaczenie poszczególnych cech oraz ich typ przedstawiono w Tabeli 1.

Tabela 1: Opis zmiennych w zbiorze danych lung

Zmienna	Typ	Opis
inst	numeric	Kod instytucji
time	numeric	Czas przeżycia w dniach
status	factor	Cenzura (1 - cenzura, 2 - śmierć)
age	numeric	Wiek
sex	factor	Płeć (1 - mężczyzna, 2 - kobieta)
ph.ecog	factor	Skala ECOG według lekarza
ph.karno	numeric	Skala Karnofsky'ego według lekarza
pat.karno	factor	Skala Karnofsky'ego według pacjenta,
meal.cal	numeric	Kalorie spożywane podczas posiłków
wt.loss	numeric	Utrata masy ciała w ciągu ostatnich sześciu miesięcy

- Skala (sprawności) Karnofsky'ego - skala pozwalająca określić stan ogólny i jakość życia pacjenta z chorobą nowotworową kwalifikowanego do chemioterapii bądź radioterapii. Skala ma rozpiętość od 100 do 0, gdzie 100 oznacza stan idealny, a 0 – śmierć. Skala opracowana przez Davida A. Karnofsky'ego oraz Josepha H. Burchenala w 1949 roku.
- Skala (sprawności) ECOG - skala pozwalająca określić stan ogólny i jakość życia pacjenta z chorobą nowotworową, ale stosowana też w geriatrii i psychiatrii, lub innych ciężkich i przewlekłych chorobach.

Porównanie skali ECOG do skali Karnofsky'ego:

- 0 - 100, 90,
- 1 - 80, 70,
- 2 - 60, 50,
- 3 - 40, 30,
- 4 - 20, 10,
- 5 - 0.

2 Lista 9

Lista skupia się na poznaniu modelów przyspieszonego czasu awarii oraz proporcjonalnych hazardów.

Niech $S_0(x)$ oraz $h_0(x)$ będą odpowiednio funkcjami przeżycia hazardu o znanych postaciach, które odpowiadają rozkładowi obserwowalnej zmiennej losowej X o zerowym wektorze charakterystyk. W przypadku zadań funkcje te będą odpowiadać rozkładowi Weibulla, jednakże można też przyjąć inne, co zostało uwzględnione w poniższych opisach.

2.1 Model AFT

Model, w którym funkcja przeżycia $S\left(x|z\right)$ jednostki o wektorze z charakterystyk jest postaci:

$$S\left(x|z\right) = S_0\left(\exp\left(\beta^T z\right)x\right)$$

nazywamy modelem przyspieszonego czasu awarii (AFT - accelerated failure-time). Funkcję $S_0(x)$ nazywamy bazową funkcją przeżycia, natomiast $\exp\left(\beta^T z\right)$ - czynnikiem przyspieszenia.

Model przyspieszonego czasu awarii ma następującą interpretację. Jeżeli z_1 i z_2 są dwoma wektorami charakterystyk, to prawdopodobieństwo przeżycia czasu x przez jednostkę o charakterystyce z_1 jest takie samo, jak prawdopodobieństwo przeżycia czasu $\exp\left(\beta^T\left(z_1 - z_2\right)\right)x$ przez jednostkę o charakterystyce z_2. Tym samym, jeżeli:

- $\exp\left(\beta^T\left(z_1 - z_2\right)\right) > 1$ $\left(\beta^T\left(z_1 - z_2\right) > 0\right)$ - z_1 ma w każdej chwili większe prawdopodobieństwo wystąpienia zdarzenia przed tą chwilą niż jednostka o charakterystyce z_2,
- $\exp\left(\beta^T\left(z_1 - z_2\right)\right) < 1$ $\left(\beta^T\left(z_1 - z_2\right) < 0\right)$ - z_1 ma w każdej chwili mniejsze prawdopodobieństwo wystąpienia zdarzenia przed tą chwilą niż jednostka o charakterystyce z_2,
- w szczególności $\exp\left(\beta^T\left(z_1 - z_2\right)\right) = 1$ $\left(\beta^T\left(z_1 - z_2\right) = 0\right)$ - $z_1 = z_2$.

Korzystając z faktu, że
$h(x) = -\frac{S'(x)}{S(x)}$, model AFT możemy równoważnie przedstawić w postaci:

$$h\left(x|z\right) = h_0\left(\exp\left(\beta^T z\right)x\right)\exp\left(\beta^T z\right)$$

W praktyce często za bazową funkcję przeżycia w modelu AFT przyjmuje się funkcję przeżycia odpowiadającą rozkładowi Weibulla (w szczególnym przypadku rozkładowi wykładniczemu), Gompertza, lognormalnemu, logistycznemu oraz wartości ekstremalnej.

2.2 Model PH

W (parametrycznym) modelu proporcjonalnych hazardów, w skrócie PH, funkcja hazardu $h\left(x|z\right)$ jednostki o wektorze z charakterystyk jest postaci:

$$h\left(x|z\right) = h_0(x)\exp\left(\beta^T z\right)$$

Model proporcjonalnych hazardów ma prostą interpretację, od której zawdzięcza swoją nazwę. Można zauważyć, że jeżeli $z_1 = \left(z_{11}, z_{12}, \ldots, z_{1p}\right)^T$ oraz $z_2 = \left(z_{21}, z_{22}, \ldots, z_{2p}\right)^T$ są dwoma wektorami charakterystyk, to

$$\frac{h\left(x|z_1\right)}{h\left(x|z_2\right)} = \exp\left(\beta^T\left(z_1 - z_2\right)\right),$$

czyli iloraz hazardów dwóch dowolnych jednostek jest proporcjonalny.

Korzystając z relacji między funkcją hazardu a przeżycia $S(x) = \exp\left[-\int_0^x h(u)du\right]$, model PH możemy równoważnie przedstawić w postaci:

$$S\left(x|z\right) = \exp\left[-\int_0^x h_0(u)\exp\left(\beta^T z\right)du\right] = \left[S_0(x)\right]^{\exp\left(\beta^T z\right)}.$$

W praktyce często za bazową funkcję hazardu w modelu PH przyjmuje się funkcję hazardu odpowiadającą rozkładowi Weibulla (w szczególnym przypadku rozkładowi wykładniczemu), o kawałkami stałym hazardzie, Gompertza oraz wartości ekstremalnej.

2.3 Zadanie 1

W tym zadaniu oszacowane zostały parametry modelu przyspieszonego czasu awarii (AFT). Jako zmienną zależną przyjmujemy *time*, a jako charakterystyki: *age*, *sex*, *ph.ecog* oraz *ph.karno*. W tym celu wykorzystano funkcję `survreg` z pakietu `survival`. Przy tworzeniu modelu uwzględniamy fakt, że zmienne *sex* oraz *ph.ecog* są typu `factor`, a dane *age* oraz *ph.karno* zostały zcentrowane zgodnie z zaleceniami. Po wyznaczeniu odpowiedniego podzbioru danych zostały usunięte wiersze zawierające brakujące dane w ilości **2**.

```
lung_new <- na.omit(lung[c("time", "status", "age",
                           "sex", "ph.ecog", "ph.karno")])

# Centrowanie danych
age_average <- mean(lung_new$age)
lung_new$age_new <- lung_new$age - age_average
ph.karno_average <- mean(lung_new$ph.karno)
lung_new$ph.karno_new <- lung_new$ph.karno - ph.karno_average

# Tworzenie modelu AFT
lung_new_aft <- survreg(Surv(time, status == 2) ~ age_new + as.factor(ph.ecog) +
                                                  ph.karno_new + as.factor(sex),
                        data = lung_new, dist = "weibull")
```

2.4 Zadanie 2

W tym zadaniu zinterpretowana zostały wyznaczone współczynniki dla poszczególnych zmiennych.

Tabela 2: Współczynniki modelu AFT

Charakterystyka	Wartość współczynnika	Różnica wartości bezwzględnych
ph.ecog=3	1.68724	-
ph.ecog=2	0.92611	0.76113
ph.ecog=1	0.42156	0.50455
sex=2	-0.40828	0.01328
ph.karno	0.01006	0.39822
age	0.00860	0.00146

Tabela 2. wskazuje, że największy wpływ na różnicę między poszczególnymi pacjentami ma zmienna `ph.ecog`. Zmienna `sex` ma wpływ umiarkowany, natomiast `ph.karno` oraz `age` - znikomy.

2.5 Zadanie 3

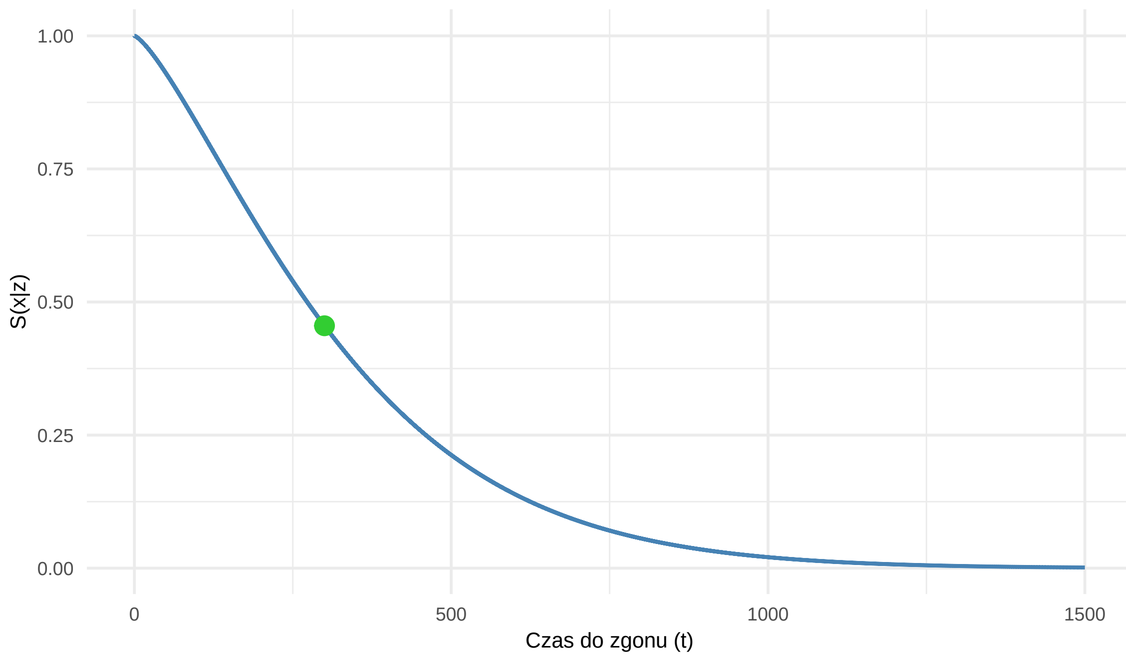
W tym zadaniu oszacowana została funkcja przeżycia (w dniach) odpowiadająca rozkładowi czasu życia 70-letniej kobiety o charakterystykach `ph.ecog=1` oraz `ph.karno=90`. Dane ciągłe zostały odpowiednio zcentrowane w oparciu o wcześniej wyznaczone średnie.

```
beta <- unname((-lung_new_aft$coefficients[-1]))
z <- c(70 - age_average, 1, 0, 0, 90 - ph.karno_average, 1)
theta_aft <- c(t(beta) %*% z)

mi <- lung_new_aft$icoef[[1]]
sigma <- exp(lung_new_aft$icoef[[2]])
alpha_aft <- 1/sigma
lambda_aft <- exp(-alpha_aft*mi)

female_time <- function(t) exp(-lambda_aft*exp(alpha_aft*theta_aft)*t^alpha_aft)
```

W oparciu o powstałą estymowaną funkcję przeżycia można wyznaczyć oszacowanie prawdopodobieństwa, że czas życia kobiety będzie większy niż 300 dni. Estymacja wynosi **0.4553973**.

2.6 Zadanie 4

W tym zadaniu narysowano wykres estymowanej funkcji przeżycia z zadania 3.



Wykres 1: Estymacja funkcji przeżycia uzyskanej metodą PH

Na wykresie 1. widać, że szacowane prawdopodobieństwo przeżycia szybko maleje do ok. 500 dni, potem zaś prędkość ta znacząco maleje. Od punktu $t = 1138$ estymowana szansa przeżycia spada poniżej 1%. Zielonym punktem zaznaczona została wcześniej wymieniona wartość estymowanego prawdopodobieństwa, że zgon kobiety nie nastąpi przed 300 dniem.

2.7 Zadanie 5

W tym zadaniu oszacowane zostały parametry modelu proporcjonalnych hazardów (PH). Podobnie jak wcześniej, za zmienną zależną przyjmujemy *time*, a jako charakterystyki: *age*, *sex*, *ph.ecog* oraz *ph.karno*. Również podobnie jak wcześniej, odpowiednie dane są traktowane jako `factor`, a ciągłe są centrowane.

```
lung_new_ph <- phreg(Surv(time, status) ~ age_new + as.factor(ph.ecog) +
                                    ph.karno_new + as.factor(sex),
                data = lung_new, dist = "weibull")
```

2.8 Zadanie 6

W tym zadaniu zinterpretowana zostały wyznaczone współczynniki dla poszczególnych zmiennych.

Tabela 3: Współczynniki modelu PH

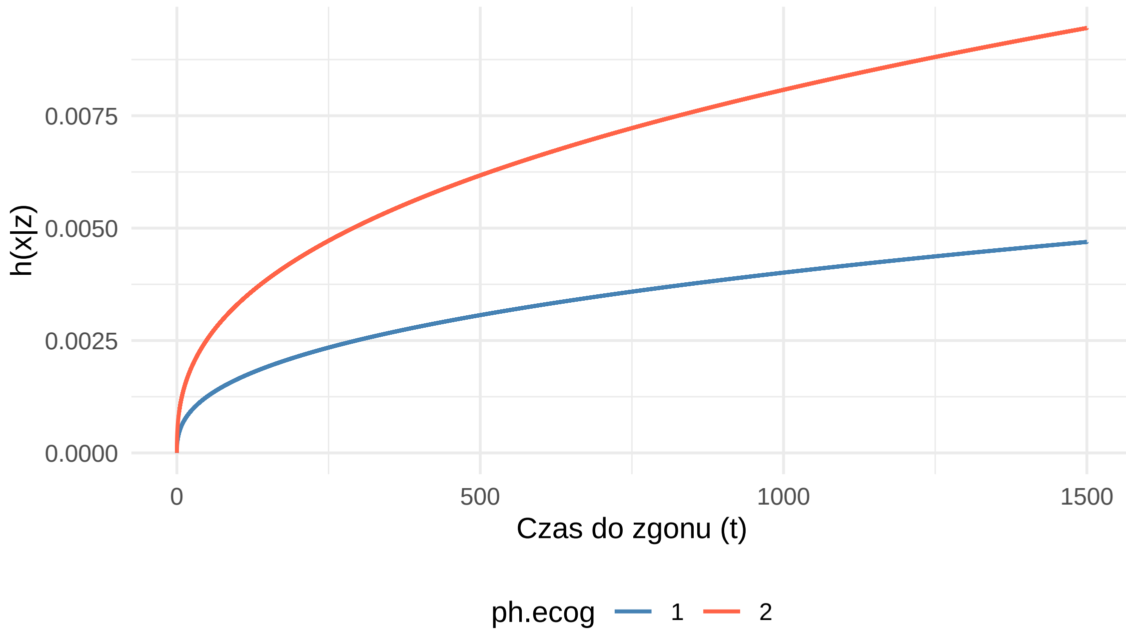
Charakterystyka	Wartość współczynnika	Różnica wartości bezwzględnych
ph.ecog=3	2.34127	-
ph.ecog=2	1.28510	1.05617
ph.ecog=1	0.58497	0.70013
sex=2	-0.56654	0.0184299999999999
ph.karno	0.01395	0.55259
age	0.01194	0.00201

Tabela 3. wskazuje na taką samą "hierarchię" wpływu poszczególnych zmiennych na estymację jak w modelu AFT. Jednakże różnica jest taka, że wartości bezwzględne są większe niż wcześniej, podobnie jak różnice.

2.9 Zadanie 7

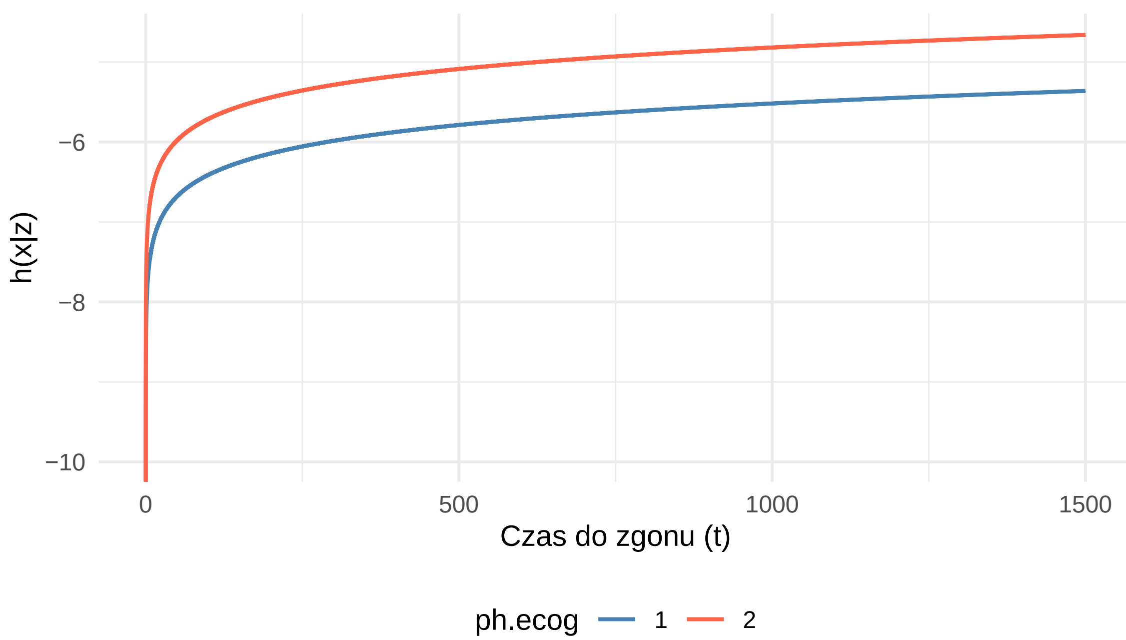
W tym zadaniu oszacowane zostały dwie funkcje hazardu (w dniach) odpowiadające rozkładowi czasu życia 70-letnich kobiet o wspólnej charakterystyce `ph.karno=90` oraz wartościach charakterystyki `ph.ecog` odpowiednio 1 i 2. Dane ciągłe zostały odpowiednio zcentrowane w oparciu o wcześniej wyznaczone średnie. Następnie porównano uzyskane estymacje.

```
beta <- unname(lung_new_ph$coefficients[-c(7,8)])
alpha_ph <- exp(lung_new_ph$coefficients[["log(shape)"]])
mu <- lung_new_ph$coefficients[["log(scale)"]]
lambda_ph <- exp(-mu*alpha_ph)

# 1 kobieta: ph.ecog=1
z <- c(70-age_average, 1, 0, 0, 90-ph.karno_average, 1)
theta_ph1 <- c(t(beta)%*%z)
female_time1 <- function(t) lambda_ph*alpha_ph*t^(alpha_ph-1)*exp(theta_ph1)

# 2 kobieta: ph.ecog=2
z <- c(70-age_average, 0, 1, 0, 90-ph.karno_average, 1)
theta_ph2 <- c(t(beta)%*%z)
female_time2 <- function(t) lambda_ph*alpha_ph*t^(alpha_ph-1)*exp(theta_ph2)
```



Wykres 2: Wykresy estymowanych funkcji hazardu dla odpowiednich charakterystyk 70-letnich kobiet, uzyskane metodą PH

Z wykresu 2. można odczytać, że funkcja hazardu dla kobiety w gorszym stanie sprawności (`ph.ecog=2`) jest zawsze wyższa niż dla tej drugiej. Oznacza to, że w jej przypadku ryzyko zgonu w każdym momencie jest wyższe.



Wykres 3: Logarytmy wykresów estymowanych funkcji hazardu dla odpowiednich charakterystyk 70-letnich kobiet, uzyskanych metodą PH

Poszczególne proste log-hazardu na wykresie 3. są w przybliżeniu równoległe, co sugeruje, że iloraz hazardów jest stały w czasie. Tym samym nie ma podstaw do kwestionowania decyzji o przyjęciu modelu proporcjonalnych hazardów.

2.10 Zadanie 8

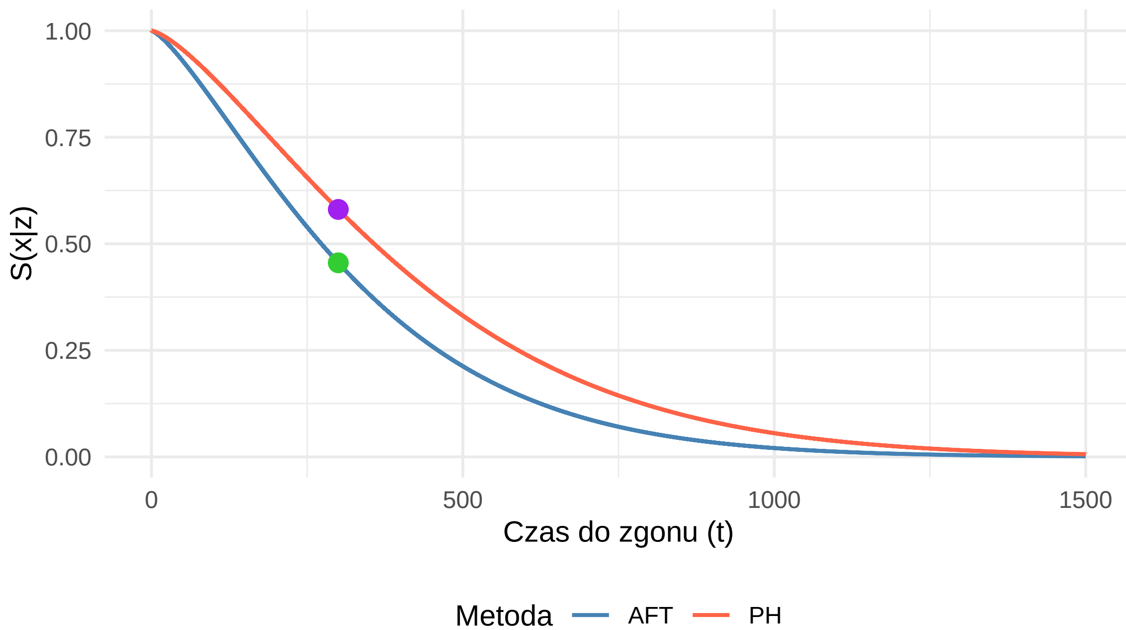
W tym zadaniu, w oparciu o wcześniej wyznaczone funkcje hazardu, wyznaczono funkcje przeżycia.

```
female_time1 <- function(t) exp(-lambda_ph*t^alpha_ph*exp(theta_ph1))
female_time2 <- function(t) exp(-lambda_ph*t^alpha_ph*exp(theta_ph2))
```

W oparciu o wyznaczone funkcje, możemy wyznaczyć estymowane prawdopodobieństwo, że kobiety przeżyją ponad 300 dni. Dla kobiety o charakterystyce `ph.karno=2` wynosi ona **0.3345465**, a dla drugiej (`ph.karno=1`) - **0.5806085**. Jest to wartość większa o **0.1252113** niż ta uzyskana metodą AFT.

2.11 Zadanie 9

W tym zadaniu zostały porównane funkcje przeżycia uzyskane metodami AFT oraz PH dla 70-letniej kobiety o charakterystykach $ph.karno = 90$ oraz $ph.ecog = 1$.



Wykres 4: Porównanie funkcji przeżycia szacowanych metodami AFT oraz PH

Z wykresu 4. widać znaczną różnicę między modelem AFT oraz PH. Estymowana funkcja przeżycia uzyskana za pomocą metody przyspieszonego czasu awarii maleje szybciej niż przy oszacowaniu metodą proporcjonalnych hazardów. Dopiero po czasie $t = 1250$ można uznać wartości wykresów za zbliżone do siebie. Maksymalna różnica między wartościami funkcji wynosi ok. **0.129286**. Zaznaczone punkty odpowiadają za estymowane wartości prawdopodobieństwa, że kobieta umrze po okresie ponad 300 dni, których konkretne wartości zostały podane przedtem.

3 Lista 10

TO DO

3.1 Zadanie 1

TO DO

3.2 Zadanie 2

TO DO

3.3 Zadanie 3

TO DO ## Zadanie 4

TO DO ## Zadanie 5

TO DO ## Zadanie 6

TO DO

4 Lista 11

TO DO

4.1 Zadanie 1

TO DO

4.2 Zadanie 2

TO DO

5 Zadanie dodatkowe

5.1 Zadanie 1

TO DO

5.2 Zadanie 2

TO DO

6 Bibliografia