

# Sprawozdanie 2

## Analiza przeżycia

Marta Stankiewicz (282244) Kacper Szmigielski (282255)

### Spis treści

<b>1</b>	<b>Zadanie 1</b>	<b>2</b>
<b>2</b>	<b>Zadanie 2</b>	<b>3</b>
<b>3</b>	<b>Zadanie 3</b>	<b>4</b>
<b>4</b>	<b>Zadanie 4</b>	<b>5</b>
<b>5</b>	<b>Zadanie 5</b>	<b>5</b>
<b>6</b>	<b>Zadanie 6</b>	<b>6</b>
<b>7</b>	<b>Zadanie 7</b>	<b>7</b>
<b>8</b>	<b>Zadanie 8</b>	<b>8</b>
<b>9</b>	<b>Zadanie 9</b>	<b>8</b>

### Spis rysunków

1	Wykres oszacowanej funkcji przeżycia - model AFT . . . . .	5
2	Wykres oszacowanej funkcji przeżycia - model PH . . . . .	8

# Spis tabel

1	Współczynniki w modelu AFT . . . . .	3
2	Współczynniki $\beta$ w modelu AFT . . . . .	3
3	Współczynniki w modelu PH . . . . .	6

## 1 Zadanie 1

Parametry zostały oszacowane zgodnie z przykładem zawartym na stronie 8 wykładu 9.

```
# parameter estimation (based on example in lecture)
model <- survreg(Surv(time, status)~age + as.factor(sex) + as.factor(ph.ecog) + ph.karno,
                   data = dane,
                   dist = "weibull")

# params assignment
beta <- -summary(model)$coefficients[-1]
mu <- model$icoef[1]
sigma <- exp(model$icoef[2])
alpha <- 1/sigma
lambda <- exp(-mu*alpha)

wsp.AFT <- data.frame(
  "alpha" = alpha,
  "beta" = beta,
  "lambda" = lambda,
  "mu" = mu,
  "sigma" = sigma
)

rownames(wsp.AFT) <- c("age",
                        "sex",
                        "ph.ecog = 1",
                        "ph.ecog = 2",
                        "ph.ecog= 3",
                        "ph.karno")
```

## 2 Zadanie 2

Tabela 1: Współczynniki w modelu AFT

Zmienne	$\alpha$	$\beta$	$\lambda$	$\mu$	$\sigma$
age	1.32569	0.00860	0.00033	6.0431	0.75432
sex	1.32569	-0.40828	0.00033	6.0431	0.75432
ph.ecog = 1	1.32569	0.42156	0.00033	6.0431	0.75432
ph.ecog = 2	1.32569	0.92611	0.00033	6.0431	0.75432
ph.ecog= 3	1.32569	1.68724	0.00033	6.0431	0.75432
ph.karno	1.32569	0.01006	0.00033	6.0431	0.75432

Dlaczego tylko współczynniki  $\beta$  się różnią?

W pakiecie R funkcja, której parametry estymujemy za pomocą survreg ma postać  $\ln(X_z) = \mu - \beta^T + \sigma W$  gdzie  $\beta^T = \beta_1 z_1 + \dots + \beta_n z_n$

Jest to postać liniowa modelu AFT a  $\alpha$  i  $\lambda$  pojawiają się dopiero po przejściu do postaci ogólnej, sa one funkcjami parametrów  $\mu$  oraz  $\sigma$

gdzie

$\mu$  - przesunięcie (poziom log-czasu gdy  $z = 0$ )

$-\beta^T$  - iloczyn skalarny

$\sigma W$  - to jest część losowa

Przy tej konwencji  $\beta > 0$  odejmujemy, więc czas się skraca

dla  $\beta < 0$  jest analogicznie, odwrotnie

Więc do interpretacji potrzebujemy jedynie współczynników  $\beta$

Tabela 2: Współczynniki  $\beta$  w modelu AFT

Zmienne	$\beta$
age	0.00860
sex	-0.40828
ph.ecog = 1	0.42156
ph.ecog = 2	0.92611
ph.ecog= 3	1.68724
ph.karno	0.01006

Z tabeli wynika, że ujemny współczynnik ma jedynie sex

sex jest zmienną typu faktor, więc ma poziomy referencyjne

Sprawdzamy poziomy referencyjne dla sex

```
## [1] "1" "2"
```

Czyli zmienna referencyjna jest 1

to oznacza, że sex = 2 wydłuża czas przeżycia w stosunku do pacjentów o wartości sex = 1

Kolejną zmienną typu factor jest ph.ecog

sprawdzamy jaki poziom jest referencyjny

```
## [1] "0" "1" "2" "3"
```

Poziom referencyjny to 0

współczynniki beta dla 1,2,3 wyszły  $> 0$ , to oznacza, że ph.ecog = 1,2,3 skraca czas przeżycia w stosunku do ph.ecog = 0

Pozostałe zmienne są typu ciągłego oraz mają współczynniki beta  $> 0$ , więc wzrost każdej z nich skraca czas przeżycia

### 3 Zadanie 3

W modelu przyspieszonego czasu życia (AFT) zakłada się, że funkcja przeżycia jednostki o wektorze charakterystyk  $z$  może zostać zapisana jako przeskalowanie czasu w bazowej funkcji przeżycia:

$$S(t | z) = S_0\left(\exp(\beta^\top z)\ t\right), \quad (1)$$

gdzie  $S_0(t)$  oznacza funkcję przeżycia jednostki bazowej, tj. odpowiadającej zerowemu wektorowi charakterystyk.

Jeżeli bazowy czas przeżycia ma rozkład Weibulla z parametrami  $\lambda > 0$  oraz  $\alpha > 0$ , to bazowa funkcja przeżycia dana jest wzorem

$$S_0(t) = \exp(-\lambda t^\alpha), \quad t \geq 0. \quad (2)$$

Podstawiając (2) do definicji (1), otrzymujemy funkcję przeżycia jednostki o wektorze charakterystyk  $z$  w postaci

$$S(t | z) = \exp\left(-\lambda \exp(\alpha \beta^\top z) t^\alpha\right). \quad (3)$$

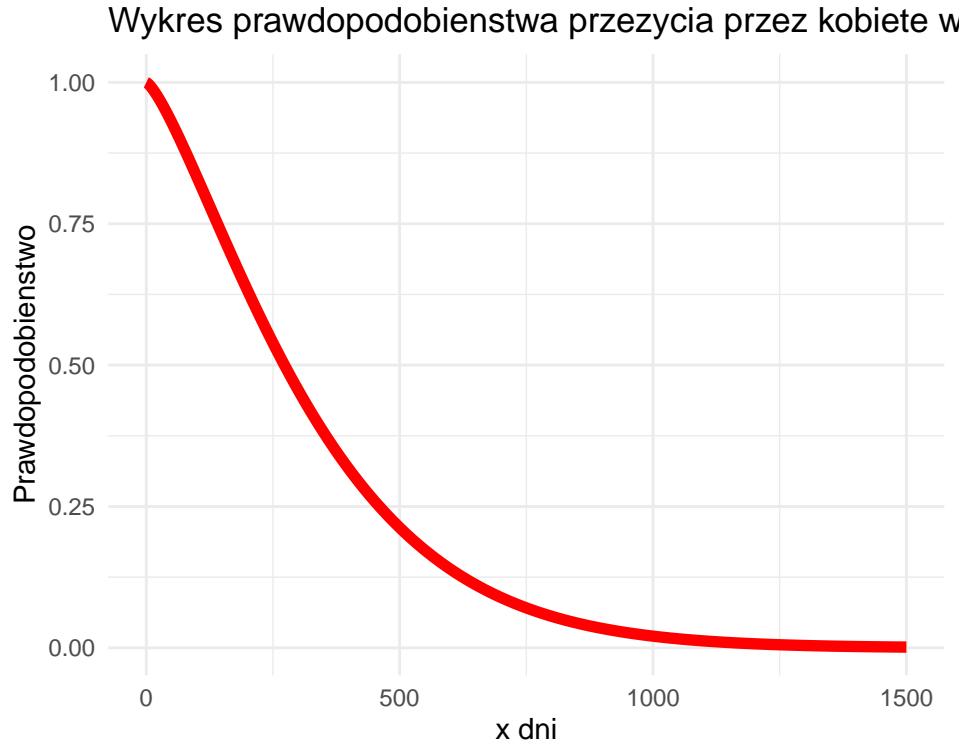
```
pacjent <- c(70 - mean_wiek, 1, 1, 0, 0, 90 - mean_ph_karno)

S_0 <- function(x) exp(-lambda*x^alpha)

S <- function (t) {
  exp(-lambda*exp(alpha*sum(beta*pacjent))*t^alpha)
}
```

Prawdopodobieństwo, że czas życia tej kobiety będzie większy niż 300 dni wynosi 0.4553973.

## 4 Zadanie 4



Rysunek 1: Wykres oszacowanej funkcji przeżycia - model AFT

## 5 Zadanie 5

**Definicja 5.1.** Niech  $h_0(x)$  będzie funkcją hazardu (o znanej postaci) obserwowalnej zmiennej losowej  $X$ , odpowiadającej jednostce o zerowym wektorze charakterystyk. Wówczas model, w którym funkcja hazardu jednostki o wektorze charakterystyk  $z$  ma postać

$$h(x | z) = h_0(x) \exp(\beta^\top z), \quad (4)$$

nazywamy (parametrycznym) modelem proporcjonalnych hazardów (PH). Funkcję  $h_0(x)$  nazywamy bazową funkcją hazardu.

Model proporcjonalnych hazardów charakteryzuje się prostą własnością interpretacyjną. Jeżeli  $z_1$  oraz  $z_2$  są dwoma wektorami charakterystyk, to na podstawie (4) otrzymujemy

$$\frac{h(x | z_1)}{h(x | z_2)} = \exp(\beta^\top (z_1 - z_2)), \quad (5)$$

co oznacza, że iloraz hazardów dwóch jednostek jest stały w czasie, a zatem hazardy są proporcjonalne.

```

model <- phreg(
  Surv(time, status)~age + as.factor(sex) + as.factor(ph.ecog) + ph.karno,
  data = dane,
  dist = "weibull"
)

beta <- model$coefficients[-c(7,8)]
mu <- model$coefficients['log(scale)']
sigma <- exp(model$coefficients['log(shape)'])
lambda <- exp(-mu*sigma)
alpha <- sigma

```

## 6 Zadanie 6

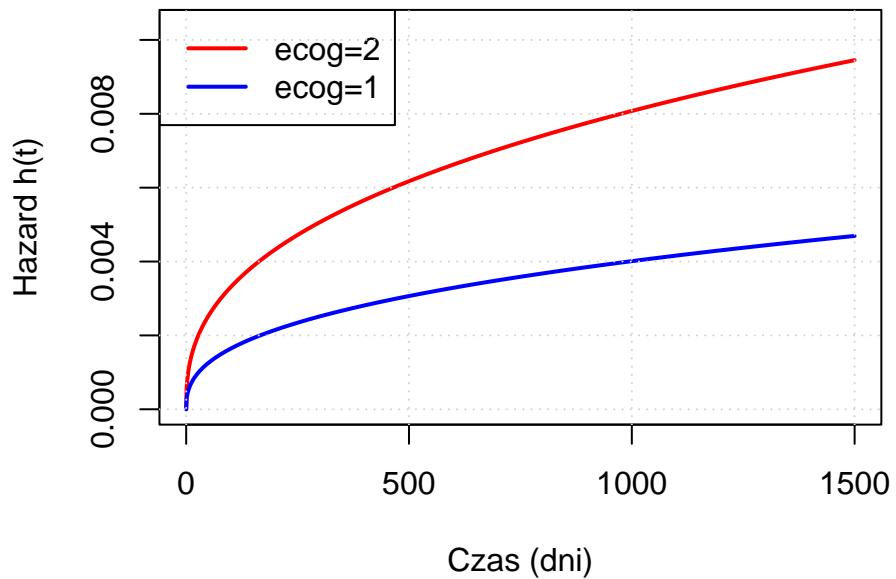
Tabela 3: Współczynniki w modelu PH

Zmienna	$\alpha$	$\beta$	$\lambda$	$\mu$	$\sigma$
age	1.38763	0.01194	0.00016	6.30118	1.38763
sex	1.38763	-0.56654	0.00016	6.30118	1.38763
ph.ecog = 1	1.38763	0.58497	0.00016	6.30118	1.38763
ph.ecog = 2	1.38763	1.28510	0.00016	6.30118	1.38763
ph.ecog= 3	1.38763	2.34127	0.00016	6.30118	1.38763
ph.karno	1.38763	0.01395	0.00016	6.30118	1.38763

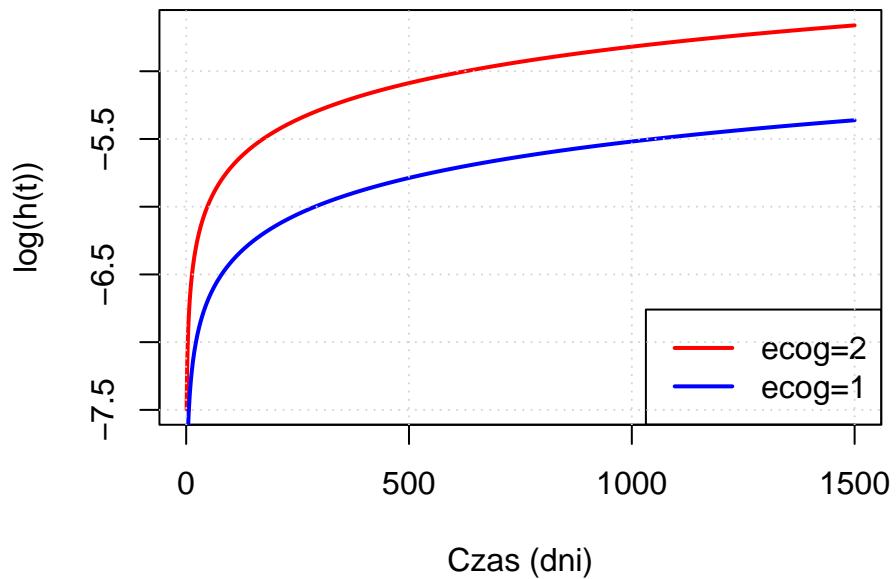
Jak interpretować współczynniki

## 7 Zadanie 7

Funkcje hazardu



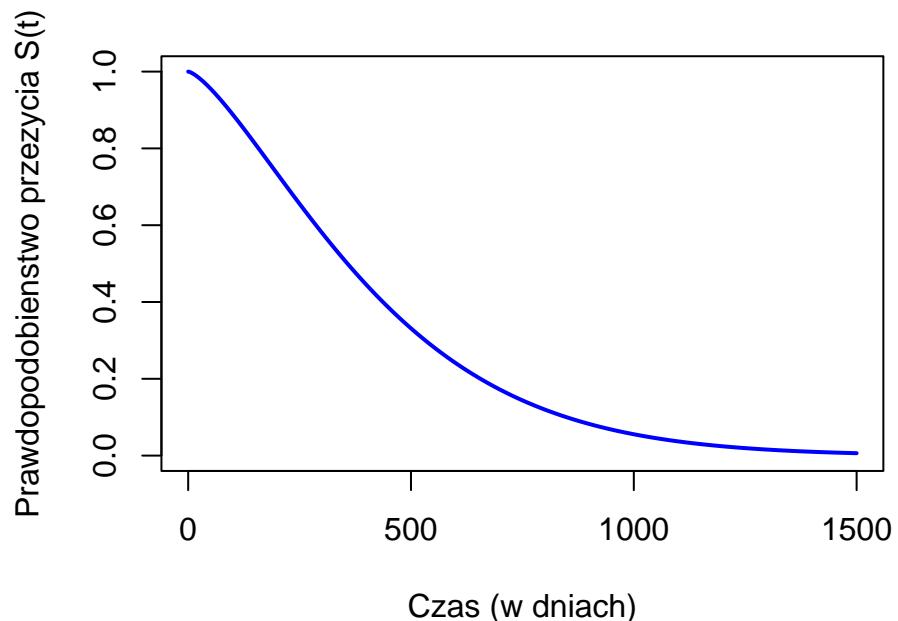
Logarytmy funkcji hazardu



## 8 Zadanie 8

Prawdopodobieństwo, że czas życia kobiety w wieku 70 lat o charakterystyce  $ph.ecog = 1$  i  $ph.karno = 90$  będzie większy niż 300 dni wynosi 0.5806085. Natomiast prawdopodobieństwo, że czas życia kobiety w wieku 70 lat o charakterystyce  $ph.ecog = 2$  i  $ph.karno = 90$  będzie większy niż 300 dni wynosi 0.3345465. Porównując wynik dla kobiety o charakterystykach opisanych w punkcie (a) z tym otrzymanym w zadaniu 3 możemy zauważać, że oszacowane prawdopodobieństwo jest większe, gdy korzystamy z modelu PH zamiast AFT.

## 9 Zadanie 9



Rysunek 2: Wykres oszacowanej funkcji przeżycia - model PH