

# Sprawozdanie 2

## Analiza przeżycia

Marta Stankiewicz (282244) Kacper Szmigielski (282255)

### Spis treści

<b>1</b>	<b>Lista 9</b>	<b>3</b>
1.1	Zadanie 1 . . . . .	3
1.2	Zadanie 2 . . . . .	4
1.3	Zadanie 3 . . . . .	5
1.4	Zadanie 4 . . . . .	6
1.5	Zadanie 5 . . . . .	7
1.6	Zadanie 6 . . . . .	8
1.7	Zadanie 7 . . . . .	8
1.8	Zadanie 8 . . . . .	10
1.9	Zadanie 9 . . . . .	11
<b>2</b>	<b>Lista 10</b>	<b>11</b>
2.1	Zadanie 1 . . . . .	11
2.2	Zadanie 2 . . . . .	12
2.3	Zadanie 3 . . . . .	13
2.4	Zadanie 4 . . . . .	15
2.5	Zadanie 5 . . . . .	16
2.6	Zadanie 6 . . . . .	17
<b>3</b>	<b>Lista 11</b>	<b>18</b>
<b>4</b>	<b>Zadanie 1</b>	<b>18</b>
4.1	Treść . . . . .	18
4.2	Rozwiązanie . . . . .	18

<b>5</b>	<b>Zadanie 2</b>	<b>18</b>
5.1	Treść . . . . .	18
5.2	Rozwiązanie . . . . .	18
<b>6</b>	<b>Zadanie 3</b>	<b>20</b>
6.1	Treść . . . . .	20
6.2	Rozwiązanie . . . . .	20
<b>7</b>	<b>Zadanie 4</b>	<b>21</b>
7.1	Treść . . . . .	21
7.2	Rozwiązanie . . . . .	21
7.3	Treść . . . . .	22
7.4	Rozwiązanie . . . . .	23
<b>8</b>	<b>Zadanie 6</b>	<b>24</b>
8.1	Treść . . . . .	24
8.2	Rozwiązanie . . . . .	24

## Spis rysunków

1	Wykres oszacowanej funkcji przeżycia - model AFT . . . . .	7
2	Porównanie estymowanych funkcji hazardu dla różnych poziomów sprawności ECOG . . . . .	9
3	Porównanie logarytmicznych funkcji hazardu dla wybranych poziomów sprawności ECOG . . . . .	10
4	Wykres oszacowanej funkcji przeżycia - model PH . . . . .	11
5	Oszacowanie bazowej skumulowanej funkcji hazardu odpowiadającej rozkładowi czasu życia . . . . .	14
6	Oszacowanie bazowej funkcji przeżycia odpowiadającej rozkładowi czasu życia . . . . .	14
7	Oszacowanie skumulowanej funkcji hazardu - model Coxa . . . . .	15
8	Logarytm skumulowanej funkcji hazardu - model Coxa . . . . .	16
9	Oszacowanie funkcji przeżycia - model Coxa . . . . .	17

# Spis tabel

1	Współczynniki w modelu AFT . . . . .	4
2	Współczynniki $\beta$ w modelu AFT . . . . .	5
3	Współczynniki w modelu PH . . . . .	8
4	Wartości oszacowanych parametrów wraz z ilorazami hazardu, przedziałami ufności oraz p-wartościami . . . . .	12

## 1 Lista 9

Sprawozdanie dotyczy analizy zbioru *lung* dostępnych w pakiecie *survival*. Dane dotyczą pacjentów z zaawansowanym rakiem płuc. Zostały one odpowiednio przygotowane, tzn. pominięto zmienne niepotrzebne dla analizy (kolumny: *inst*, *pat.karno*, *meal.cal*, *wt.loss*), usunięto wiersze zawierające wartości brakujące oraz wycentrowano zmienne ciągłe (poprzez odjęcie średniej wartości). W poniższych zadaniach przyjęto, że czas przeżycia ma rozkład Weibulla.

### 1.1 Zadanie 1

W pierwszej kolejności oszacowano (zgodnie z przykładem zawartym na stronie 8 wykładu 9) parametry modelu przyspieszonego czasu awarii, przyjmując za zmienną zależną zmienną *time*, a za charakterystyki zmienne: *age*, *sex*, *ph.ecog*, *ph.karno*.

```
# 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")

```

## 1.2 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

Analizując otrzymane współczynniki widoczne w tabeli 1 można zauważyć, że jedyne zmiany widoczne są wśród parametrów  $\beta$ . Wykorzystywana w pakiecie R funkcja *survreg* estymuje parametry modelu AFT, który w postaci liniowej wyraża się wzorem:

$$\ln(X_z) = \mu - \beta^T z + \sigma W$$

gdzie  $\beta^T z = \beta_1 z_1 + \dots + \beta_n z_n$  jest iloczynem skalarnym wektora współczynników i zmiennych objaśniających. Pozostałe elementy oznaczają:

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

$\sigma W$  - błąd losowy (postać rozkładu zmiennej  $W$  jest znana)

Warto zaznaczyć, że parametry  $\alpha$  oraz  $\lambda$ , charakterystyczne dla postaci ogólnej rozkładów, są funkcjami  $\mu$  oraz  $\sigma$  i pojawiają się dopiero po przekształceniu modelu do formy hazardu lub funkcji przeżycia. Z tego względu w bezpośredniej interpretacji wpływu zmiennych objaśniających kluczowe znaczenie mają współczynniki  $\beta$ .

Jeśli  $\beta > 0$ , to wartość tę odejmujemy w równaniu, co prowadzi do zmniejszenia logarytmu czasu, a w konsekwencji do skrócenia przewidywanego czasu przeżycia pacjenta. Jeśli  $\beta < 0$ , wpływ zmiennej jest odwrotny – prowadzi to do zwiększenia wartości  $\ln(X_z)$ , co interpretujemy jako wydłużenie oczekiwanego czasu przeżycia.

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 analizy parametrów zamieszczonych w tabeli 2 wynika, że jedyną zmienną o ujemnym współczynniku  $\beta$  jest *sex*. Ponieważ jest to zmienna typu faktor, sprawdzany jest jej poziom referencyjny.

```
levels(as.factor(lung$sex))
```

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

Kategorią odniesienia (kodowaną jako 1) są mężczyźni. Ujemna wartość współczynnika dla poziomu 2 (kobiety) oznacza, że w przyjętej konwencji modelu — gdzie dodatnie  $\beta$  skraca czas przeżycia — płeć żeńska jest czynnikiem sprzyjającym wydłużeniu oczekiwanego czasu przeżycia w stosunku do mężczyzn.

Kolejną istotną zmienną jakościową jest *ph.ecog*, określająca stan sprawności pacjenta w skali ECOG.

```
levels(as.factor(lung$ph.ecog))
```

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

W tym przypadku poziomem referencyjnym jest wartość 0 (pacjent w pełni sprawny). Dodatnie wartości współczynników  $\beta$  dla poziomów 1, 2 oraz 3 wskazują, że każdy stopień pogorszenia sprawności fizycznej wiąże się ze skróceniem przewidywanego czasu przeżycia w porównaniu do grupy o najwyższej sprawności (poziom 0).

Pozostałe zmienne w modelu mają charakter ciągły. Ich dodatnie współczynniki  $\beta$  sugerują, że wzrost wartości tych parametrów skraca czas przeżycia.

### 1.3 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:

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

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

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

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

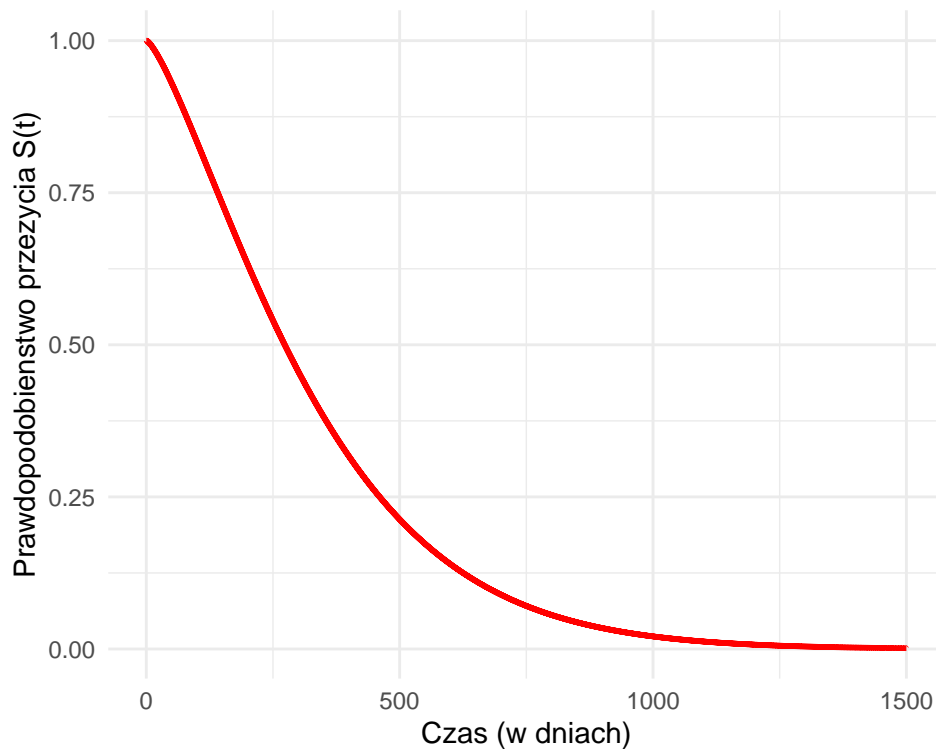
S_0 <- function(x) exp(-lambda*(x^alpha))

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

Wykorzystując wyprowadzony wzór na funkcję przeżycia, wyznaczono prawdopodobieństwo przeżycia powyżej 300 dni dla 70-letniej pacjentki o charakterystyce  $ph.ecog = 1$  oraz  $ph.karno = 90$ . Prawdopodobieństwo to wynosi 0.4554.

## 1.4 Zadanie 4

Poniżej przedstawiono wykres oszacowanej w zadaniu 3. funkcji przeżycia.



Rysunek 1: Wykres oszacowanej funkcji przeżycia - model AFT

## 1.5 Zadanie 5

**Definicja 1.1.** 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

```

## 1.6 Zadanie 6

Tabela 3: Współczynniki w modelu PH

Zmienne	$\alpha$	$\beta$	$\lambda$	$\mu$	$\sigma$
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

Obserwując wyniki w tabeli 3 jedynym zidentyfikowanym czynnikiem redukującym ryzyko jest płeć żeńska, dla której współczynnik  $\beta$  wynosi NA. Na jego podstawie można obliczyć wartość  $\exp(\beta) \approx \text{NA}$ , oznaczający ryzyko zgonu niższe o NA% w stosunku do mężczyzn. W przypadku zmiennej jakościowej *ph.ecog*, dodatnie wartości współczynników  $\beta$  dla poziomów 1, 2 oraz 3 (wynoszące odpowiednio: NA, NA oraz NA wskazują na wzrost ryzyka wraz z pogarszającym się stanem sprawności pacjenta — dla *ph.ecog* = 3 hazard jest aż NA-krotnie wyższy niż w grupie referencyjnej. Pozostałe zmienne ciągłe, *age* oraz *ph.karno*, również charakteryzują się dodatnimi wartościami  $\beta$  (odpowiednio 0.0119361 oraz 0.013955), co interpretujemy jako wzrost hazardu wraz ze zwiększaniem się wartości tych parametrów, co w konsekwencji prowadzi do skrócenia prognozowanego czasu przeżycia.

## 1.7 Zadanie 7

Na podstawie dopasowanego modelu PH wyznaczono oszacowania funkcji hazardu dla 70-letniej kobiety (*ph.karno* = 90), uwzględniając dwa alternatywne poziomy sprawności: *ph.ecog* = 1 oraz *ph.ecog* = 2.

```

z1 <- c(70 - mean_wiek, 1, 1, 0, 0, 90 - mean_ph_karno)
z2 <- c(70 - mean_wiek, 1, 0, 1, 0, 90 - mean_ph_karno)

```



```

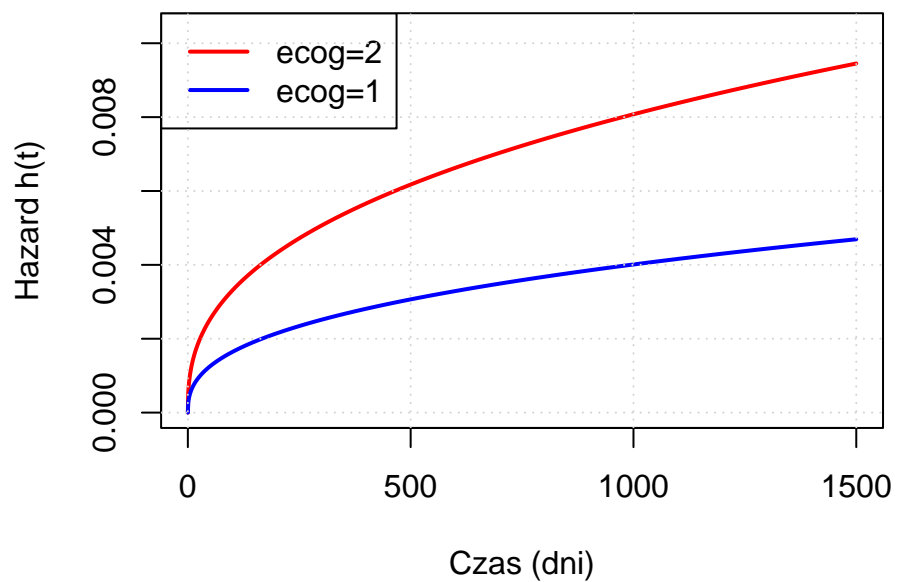
h_0 <- function(x) {
  return (lambda*alpha*x^(alpha-1))
}

h <- function(x, z) {
  h_0(x)*exp(sum(beta*z))
}

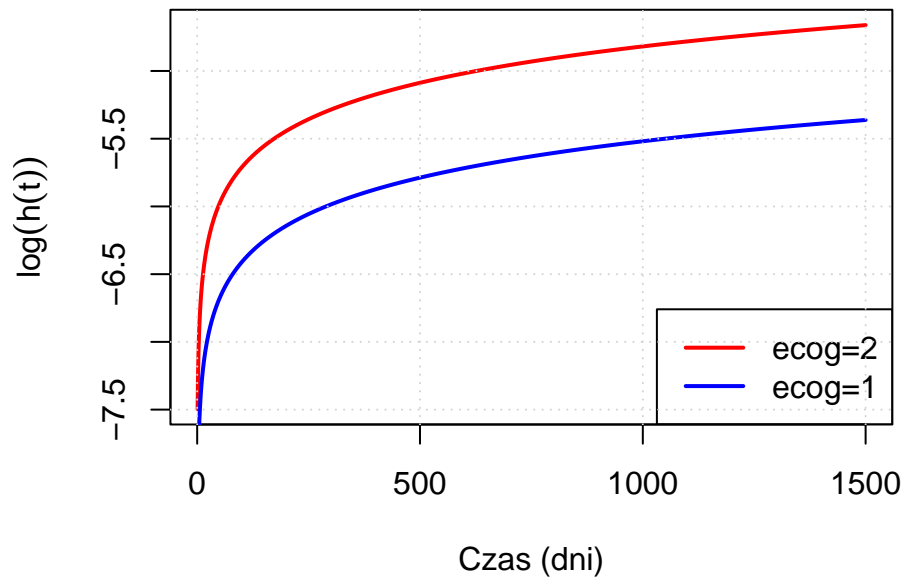
x <- 0:1500

haz1 <- sapply(x, function(t) h(t, z1))
haz2 <- sapply(x, function(t) h(t, z2))

```



Rysunek 2: Porównanie estymowanych funkcji hazardu dla różnych poziomów sprawności ECOG



Rysunek 3: Porównanie logarytmicznych funkcji hazardu dla wybranych poziomów sprawności ECOG

Weryfikacja założeń modelu, przeprowadzona poprzez analizę wyników przedstawionych na wykresach @ref(fig:hazard\_plot) i @ref(fig:log\_plot), potwierdza słuszność przyjęcia modelu proporcjonalnych hazardów. Stały odstęp między krzywymi  $\log(h(t))$  dla różnych poziomów zmiennej  $ph.ecog$  dowodzi, że iloraz hazardu pozostaje stały w czasie.

## 1.8 Zadanie 8

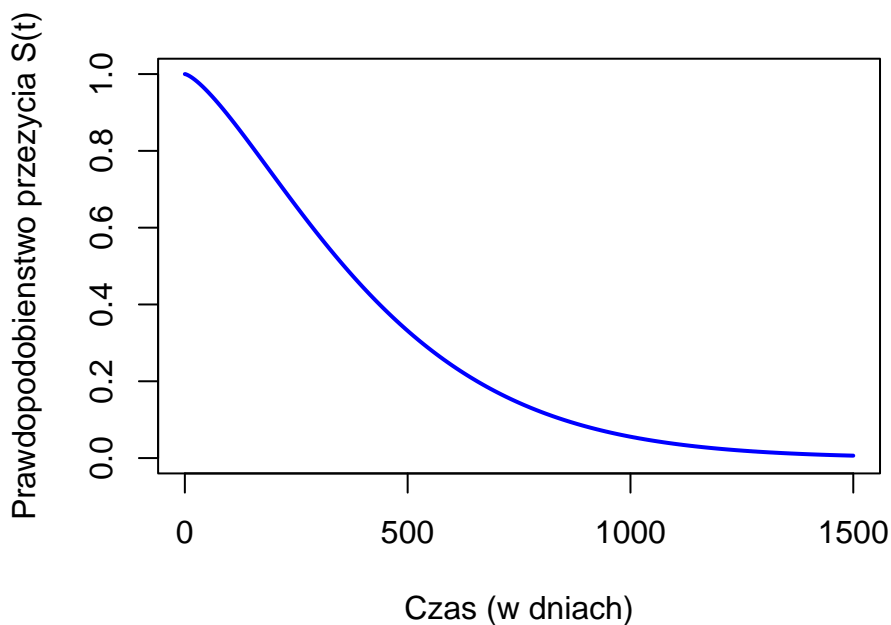
W modelu proporcjonalnych hazardów (PH) funkcja przeżycia dla jednostki o wektorze cech  $z$  jest definiowana jako bazowa funkcja przeżycia podniesiona do potęgi odpowiadającej ilorazowi hazardu. Przyjmuje ona następującą postać:

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

S <- function (t, z) {
  (S_0(t))^exp(sum(beta*z))
}
```

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 zauważyć, że oszacowane prawdopodobieństwo jest większe, gdy korzystamy z modelu PH zamiast AFT.

## 1.9 Zadanie 9



Rysunek 4: Wykres oszacowanej funkcji przeżycia - model PH

Porównując wykresy @ref(fig:wykres\_AFT) oraz @ref(fig:wykres\_PH) można zauważyć różnicę w tempie spadku oszacowanej funkcji przeżycia. Krzywa odpowiadająca modelowi proporcjonalnych hazardów (PH) maleje wolniej niż w przypadku modelu przyspieszonego czasu życia (AFT). W rezultacie dla tych samych punktów czasowych model PH generuje systematycznie wyższe prawdopodobieństwa przeżycia, co skutkuje uzyskaniem bardziej optymistycznych prognoz w porównaniu do modelu AFT.

## 2 Lista 10

W poniższych zadaniach, nie przyjęto żadnego konkretnego rozkładu czasu życia.

### 2.1 Zadanie 1

W modelu proporcjonalnych hazardów Coxa zakłada się, że rozkład czasu do wystąpienia zdarzenia jednostki o charakterystyce  $z$  ma funkcję hazardu postaci

$$h_z(t) = h_0(t)\psi(z)$$

Najczęściej przyjmuje się, że

$$\psi(z) = \exp(\beta^T z)$$

Zatem

$$h_z(t) = h_0(t) \exp(\beta^T z)$$

Poniżej oszacowano parametry modelu proporcjonalnych hazardów Coxa, przyjmując za zmienną zależną zmienną *time*, a za charakterystyki zmienne: *age*, *sex*, *ph.ecog* oraz *ph.karno*.

```
model <- coxph(Surv(time, status) ~ age + factor(sex) + factor(ph.ecog) + ph.karno,
               data = dane, ties = "efron")

s <- summary(model)

beta  <- s$coefficients[, "coef"]
pval  <- s$coefficients[, "Pr(>|z|)"]
HR    <- s$coefficients[, "exp(coef)"]

# 95% CI dla HR:
CI <- s$conf.int[, c("lower .95", "upper .95")]

wyniki <- cbind(beta=beta, HR=HR, CI, p.value=pval)
```

## 2.2 Zadanie 2

Tabela 4: Wartości oszacowanych parametrów wraz z ilorazami hazardu, przedziałami ufności oraz p-wartościami

	beta	HR	lower .95	upper .95	p.value
age	0.01256	1.01264	0.99405	1.03157	0.18390
factor(sex)2	-0.56568	0.56798	0.40723	0.79218	0.00086
factor(ph.ecog)1	0.57806	1.78257	1.12157	2.83316	0.01447
factor(ph.ecog)2	1.23990	3.45525	1.72205	6.93289	0.00048
factor(ph.ecog)3	2.39585	10.97756	1.31878	91.37750	0.02670
ph.karno	0.01242	1.01250	0.99365	1.03171	0.19512

Analizując oszacowania współczynników  $\beta$  wraz z ich ilorazami hazardu, przedziałami ufności oraz *p*-wartościami przedstawionymi w tabeli 4, można zauważyć wyraźny podział na zmienne istotne i nieistotne statystycznie. Zmienne *age* oraz *ph.karno* nie wykazują istotnego wpływu na ryzyko zgonu przy przyjętym poziomie istotności  $\alpha = 0.05$ , ponieważ ich *p*-wartości wynoszą odpowiednio 0.1839 oraz 0.1951 - są większe od przyjętego poziomu 0.05.

Pozostałe parametry modelu są istotne statystycznie. Dla zmiennej *sex* współczynnik  $\beta$  przyjmuje wartość ujemną (-0.5657), co przekłada się na iloraz hazardu  $HR = \exp(\beta) \approx 0.568$ . Oznacza to redukcję ryzyka zgonu o około 43.2% w porównaniu do mężczyzn. Z kolei w przypadku zmiennej *ph.ecog* obserwujemy dodatnie wartości współczynników, które rosną wraz z pogarszającym się stanem pacjenta. Dla najwyższego poziomu niesprawności (*ph.ecog* = 3) iloraz hazardu wynosi 10.98, co wskazuje na wielokrotny wzrost ryzyka względem grupy referencyjnej, przy czym szeroki przedział ufności dla tej kategorii sugeruje mniejszą precyzję oszacowania.

## 2.3 Zadanie 3

Poniżej wyznaczono oszacowanie bazowej skumulowanej funkcji hazardu i bazowej funkcji przeżycia odpowiadającej rozkładowi czasu życia.

Tutaj wykorzystamy wzory przedstawione w zadaniu nr.1

$$h(t | x) = h_0(t) \exp(x^\top \beta)$$

$$H(t | x) = H_0(t) \exp(x^\top \beta)$$

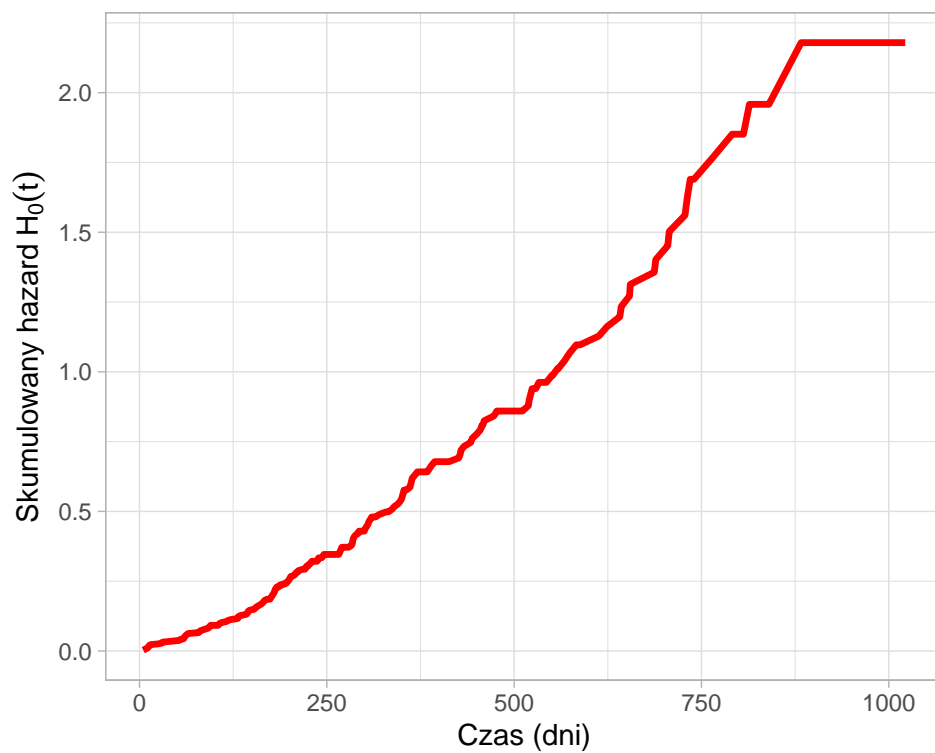
$$S(t | x) = \left( S_0(t) \right)^{\exp(x^\top \beta)}$$

$$S_0(t) = \exp\left( - H_0(t) \right)$$

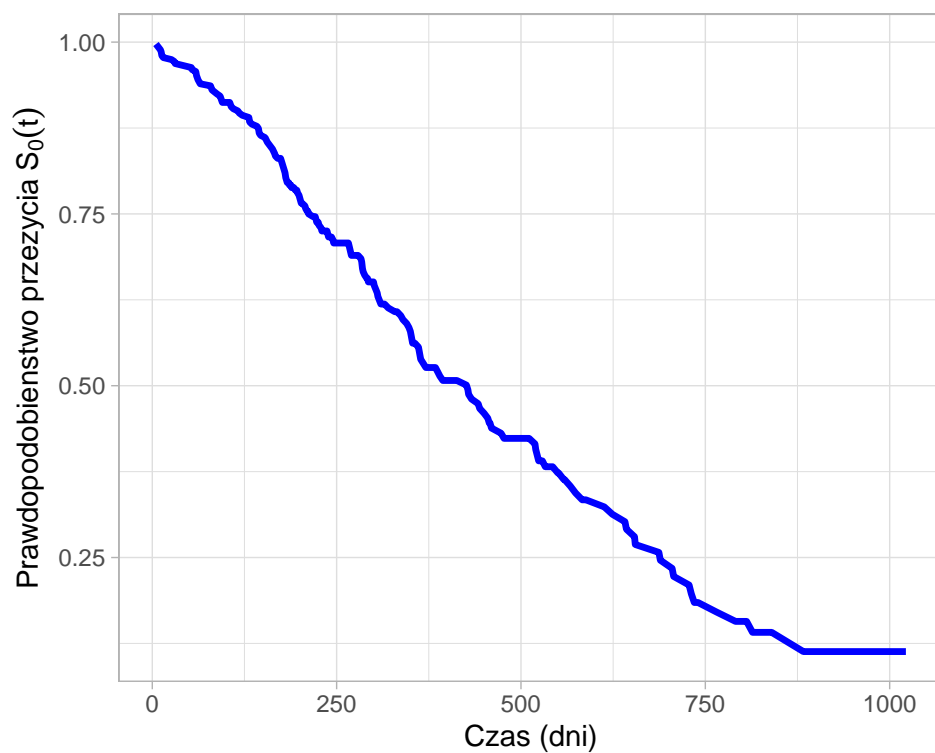
A dla bazowego Weibulla

$$H_0(t) = \left( \frac{t}{\sigma} \right)^\alpha$$

$$S_0(t) = \exp\left( - \left( \frac{t}{\sigma} \right)^\alpha \right)$$



Rysunek 5: Oszacowanie bazowej skumulowanej funkcji hazardu odpowiadającej rozkładowi czasu życia



Rysunek 6: Oszacowanie bazowej funkcji przeżycia odpowiadającej rozkładowi czasu życia

## 2.4 Zadanie 4

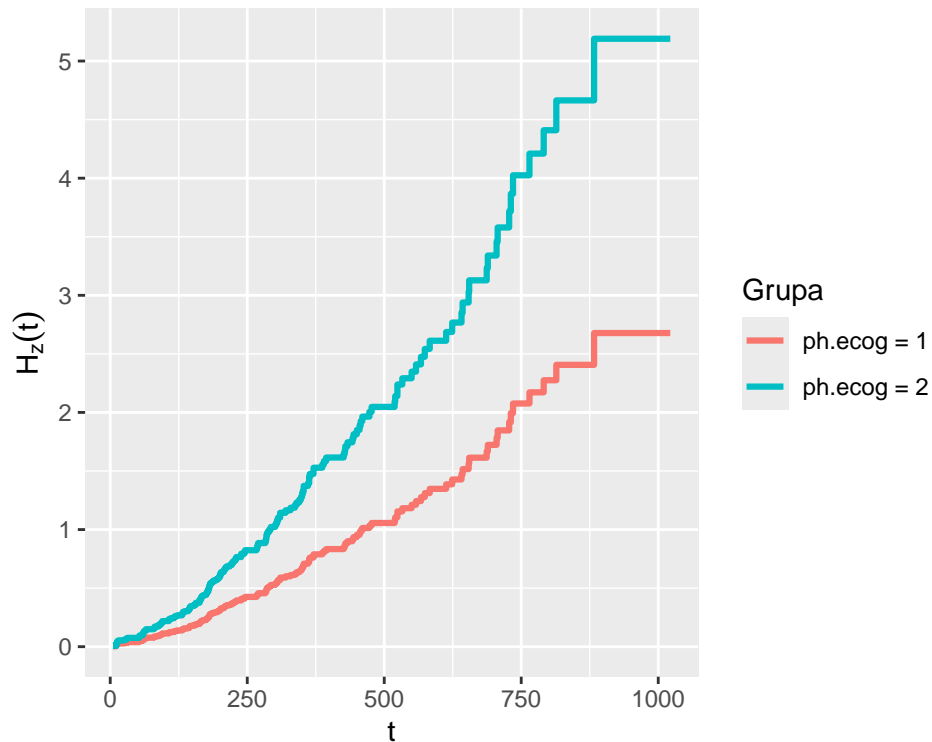
Dla modelu Coxa skumulowana funkcja hazardu dla pacjenta o wektorze cech  $z$  wyraża się wzorem:

$$H_z(t) = H_0(t) \exp(\beta^T z)$$

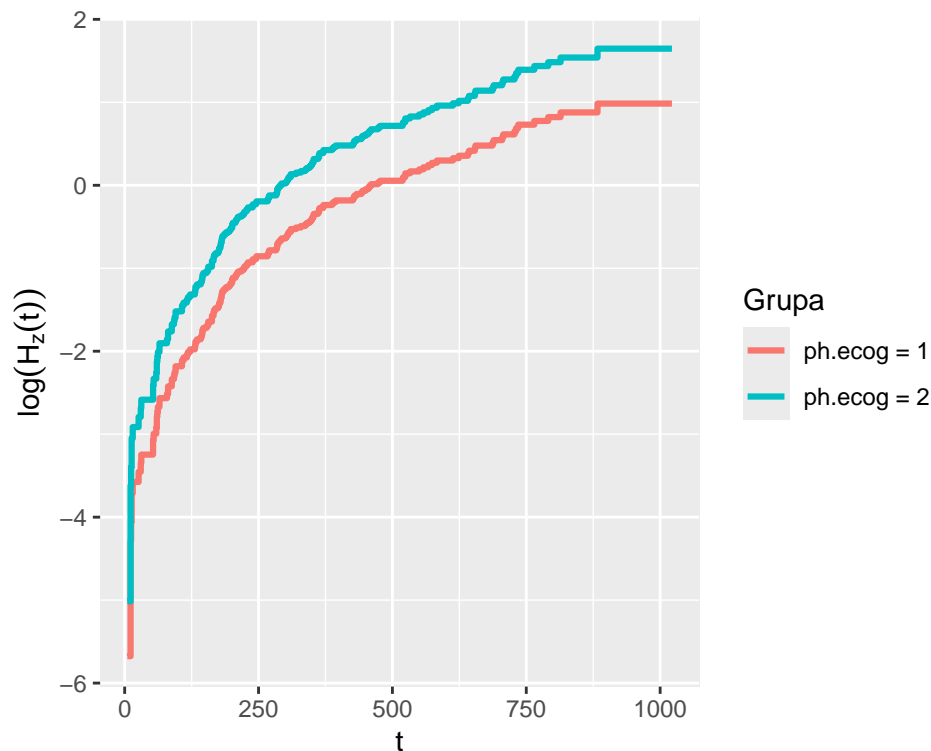
gdzie  $H_0(t)$  to bazowy skumulowany hazard, a czynnik  $\exp(\beta^T z)$  jest stały w czasie dla ustalonego pacjenta. Logarytmując powyższe równanie stronami otrzymujemy:

$$\log H_z(t) = \log H_0(t) + \beta^T z$$

Oznacza to, że na wykresie logarytmicznym krzywe dla różnych grup powinny być przesunięte względem siebie o stałą wartość, a odległość między nimi odpowiada różnicy w ryzyku  $\beta^T(z_1 - z_2)$ .



Rysunek 7: Oszacowanie skumulowanej funkcji hazardu - model Coxa



Rysunek 8: Logarytm skumulowanej funkcji hazardu - model Coxa

Analiza wykresów 7 oraz 8 potwierdza poprawność zastosowania modelu proporcjonalnych hazardów dla badanych danych. Na pierwszym wykresie widać, że krzywa dla pacjentki z gorszym rokowaniem ( $ph.ecog = 2$ ) jest proporcjonalnie przeskalowana w górę względem krzywej referencyjnej, zachowując podobny kształt przebiegu. Wniosek ten jest jeszcze wyraźniejszy na drugim wykresie przedstawiającym logarytmy funkcji. Krzywe  $\ln(H_z(t))$  są względem siebie równolegle przesunięte, zachowując stały dystans w całym analizowanym przedziale czasowym. Zatem iloraz hazardu jest stały w czasie, co oznacza, że założenia modelu Coxa są spełnione i nie ma podstaw do ich kwestionowania.

## 2.5 Zadanie 5

Wyznaczono oszacowanie funkcji przeżycia (w dniach) odpowiadającej rozkładowi czasu życia kobiet w wieku 70 lat ( $ph.karno = 90$ ), których  $ph.ecog = 1$  lub  $ph.ecog = 2$ :

```
nd <- data.frame(
  age = c(70, 70)-mean_wiek,
  sex = factor(c(2, 2), levels = levels(factor(dane$sex))),
  ph.ecog = factor(c(1, 2), levels = levels(factor(dane$ph.ecog))),
  ph.karno = c(90, 90)-mean_ph_karno
)
```



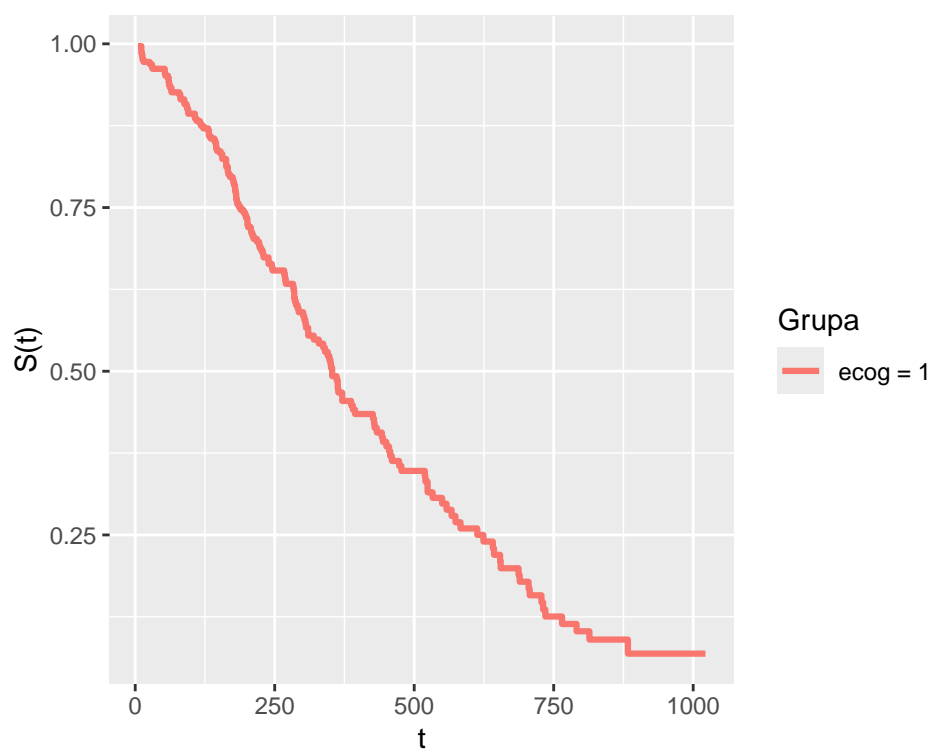
```
fit <- survfit(model, newdata = nd)

wynik <- summary(fit, times = 300)

prob_a <- wynik$surv[1]
prob_b <- wynik$surv[2]
```

Analiza uzyskanych wyników wskazuje na istotne różnice w rokowaniach zależne od stanu sprawności. Szacowane prawdopodobieństwo, że czas życia kobiety o charakterystyce  $ph.ecog = 1$  (i  $ph.karno = 90$ ) przekroczy 300 dni, wynosi  $\text{round}(\text{prob\_a}, 4)$ . W przypadku pacjentki o gorszym stanie sprawności ( $ph.ecog = 2$ ), prawdopodobieństwo to ulega wyraźnemu obniżeniu i wynosi  $\text{round}(\text{prob\_b}, 4)$ .

## 2.6 Zadanie 6



Rysunek 9: Oszacowanie funkcji przeżycia - model Coxa

Zestawienie funkcji przeżycia na wykresach 4 i 9 wykazuje różnicę w charakterze estymacji: model Coxa generuje funkcję schodkową, zmieniającą się wyłącznie w momentach wystąpienia zdarzeń, co wynika z braku założeń co do kształtu hazardu bazowego. Z kolei model parametryczny PH prezentuje gładką aproksymację tego procesu, wynikającą z przyjęcia teoretycznego rozkładu czasu życia - rozkładu Weibulla. Zbliżony przebieg obu krzywych

(trend spadkowy) świadczyłby o tym, że przyjęty rozkład parametryczny dobrze odzwierciedla rzeczywistą strukturę danych empirycznych.

### 3 Lista 11

## 4 Zadanie 1

### 4.1 Treść

Oszacować parametry modelu proporcjonalnych szans, przyjmując za zmienną zależną zmienną time, a za charakterystyki zmienne: age, sex, ph.ecog, ph.karno. Uwaga. Zadanie to można wykonać korzystając np. z funkcji prop.odds biblioteki timereg lub funkcji nltm biblioteki o tej samej nazwie pakietu R. Należy jednak zwrócić uwagę na inną definicję modelu proporcjonalnych szans niż była podana na wykładzie wykorzystywaną w tej drugiej funkcji.

### 4.2 Rozwiązanie

```
model<-prop.odds(Event(time, cause = status) ~ age + factor(sex) + factor(ph.ecog) + ph.karno,
  data = dane,n.sim=500,profile=1)
```

## 5 Zadanie 2

### 5.1 Treść

Podać interpretację współczynników modelu z zadania 1

### 5.2 Rozwiązanie

	estimate
age	0.0023530
factor(sex)2	-0.5613368
factor(ph.ecog)1	0.3199014
factor(ph.ecog)2	1.1477675
factor(ph.ecog)3	1.8507250
ph.karno	-0.0022423

Korzystając z poniższych wzorów możemy podać interpretację:

$$\theta_0(t) = \frac{1 - S_0(t)}{S_0(t)}$$

$$\ln\left(\frac{\theta_{z_1}(t)}{\theta_{z_2}(t)}\right) = \beta^T(z_1 - z_2)$$

- Zmienna **age** ma współczynnik

$$\gamma > 0,$$

co oznacza, że wraz ze wzrostem wieku rosną **szanse (odds) wystąpienia zdarzenia do czasu**

$$t$$

(przy stałych pozostałych zmiennych). Równoważnie:

$$\exp(\gamma) > 1.$$

- Zmienna **sex** (kategoryczna) ma współczynnik

$$\gamma < 0$$

dla poziomu porównywanego do poziomu referencyjnego (bazowego). Oznacza to, że w tej grupie **szanse (odds) zajścia zdarzenia do czasu**

$$t$$

są mniejsze niż w grupie bazowej, tj.

$$\exp(\gamma) < 1.$$

- Zmienna **ph.ecog** (kategoryczna) ma współczynniki

$$\gamma > 0$$

dla poszczególnych poziomów w porównaniu do poziomu referencyjnego (np.

$$1 \text{ vs } 0, 2 \text{ vs } 0, 3 \text{ vs } 0$$

). Wskazuje to, że osoby z wyższym **ph.ecog** mają **większe szanse (odds) wystąpienia zdarzenia do czasu**

$$t$$

niż osoby z poziomu bazowego, czyli

$$\exp(\gamma) > 1.$$

- Dla zmiennej ciągłej `ph.karno` otrzymano

$$\gamma < 0,$$

co oznacza, że wraz ze wzrostem `ph.karno` **maleją szanse (odds) zajścia zdarzenia do czasu**

$t$

(tj.

$$\exp(\gamma) < 1$$

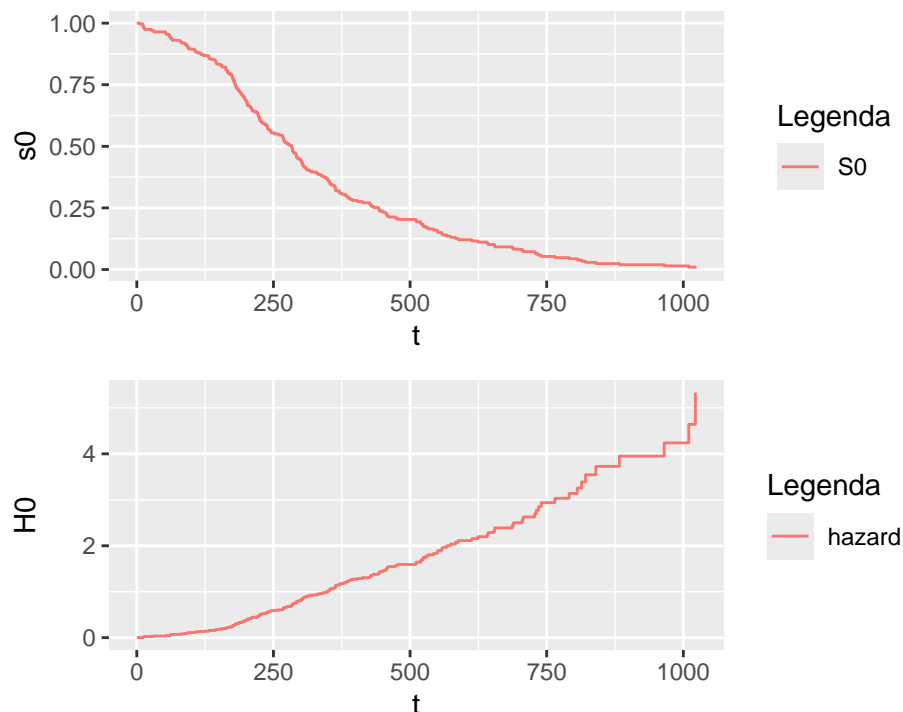
). Jeśli jednak w poprzedniej analizie (Lista 10) współczynnik dla `ph.karno` był **nieistotny statystycznie**, to również tutaj wniosek o kierunku efektu należy traktować ostrożnie: przy braku istotności znak współczynnika może zmieniać się między dopasowaniami/modelami i nie powinien być interpretowany jako stabilny efekt.

## 6 Zadanie 3

### 6.1 Treść

Wyznaczyć oszacowanie bazowej skumulowanej funkcji hazardu i bazowej funkcji przeżycia odpowiadającej rozkładowi czasu życia (przyjmując model opisany w zadaniu 1). Uwaga. To zadanie można wykonać korzystając np. z funkcji `nlm` biblioteki o tej samej nazwie.

### 6.2 Rozwiązanie

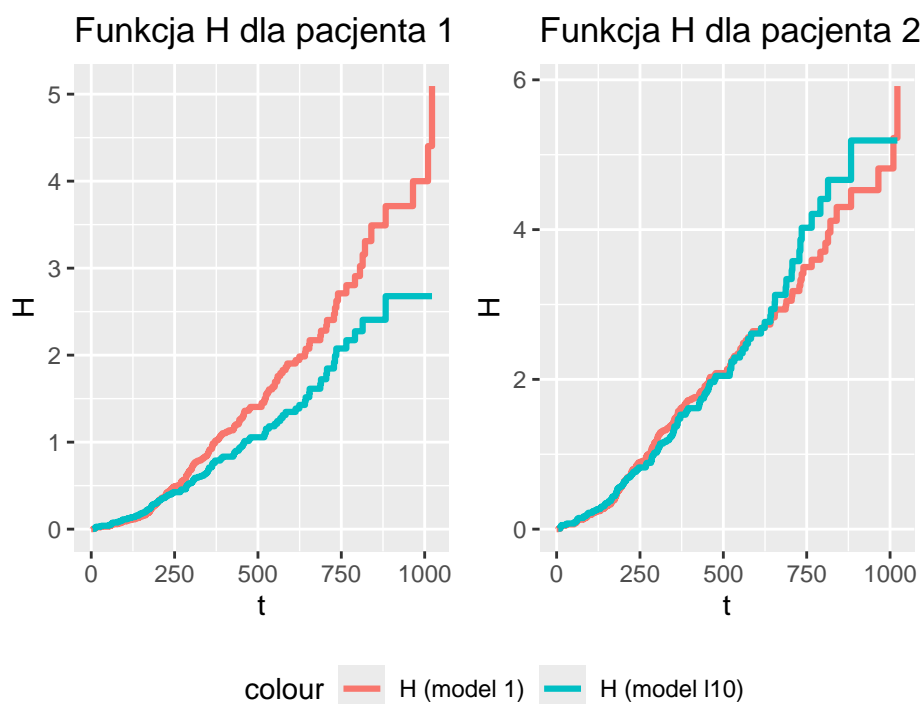


## 7 Zadanie 4

### 7.1 Treść

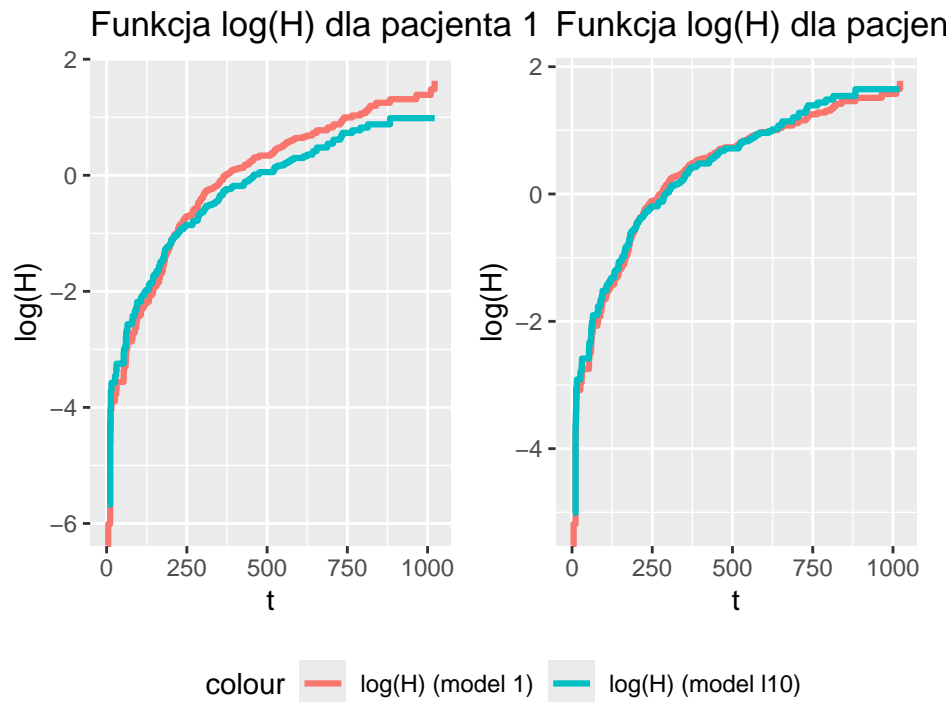
Wyznaczyć oszacowanie skumulowanej funkcji hazardu odpowiadającej rozkładowi czasu życia (przyjmując model opisany w zadaniu 1)

### 7.2 Rozwiązanie



Widzimy, że w przypadku pacjenta 1 skumulowany hazard rośnie wyraźnie szybciej, gdy użyjemy modelu proporcjonalnych szans, niż wtedy, gdy korzystamy z modelu proporcjonalnych hazardów Coxa. Oznacza to, że model proporcjonalnych szans przewiduje dla tego pacjenta gorsze przeżycie (większe skumulowane ryzyko w czasie) niż model Coxa.

W przypadku pacjenta 2 krzywe są bardzo podobne mniej więcej do 625 dnia, a później zaczynają się rozjeżdżać. W dalszej części obserwacji różnice między modelami rosną, jednak interpretację samej końcówki należy traktować ostrożnie, ponieważ w ogonie czasu często zostaje niewiele obserwacji i pojedyncze zdarzenia potrafią powodować wyraźne „skoki” krzywej.

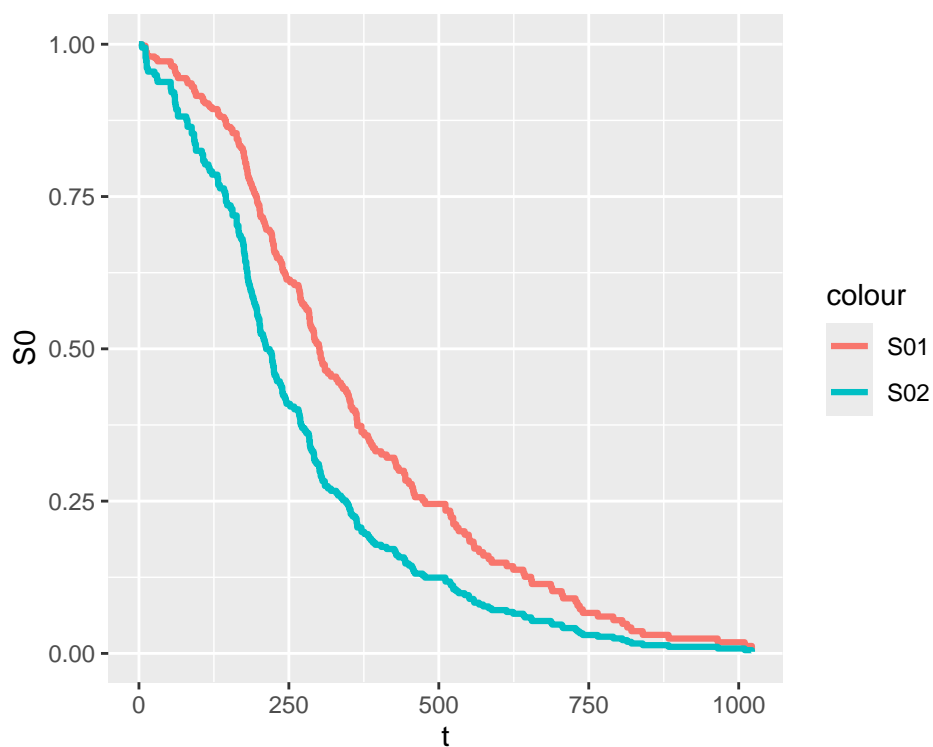


Analogiczne wnioski możemy wysnuć także na podstawie wykresów logarytmów tych funkcji, ponieważ logarytm jest funkcją ściśle rosnącą i dla dodatnich wartości zachowuje relacje. #  
Zadanie 5

### 7.3 Treść

Wyznaczyć oszacowanie funkcji przeżycia (w dniach) odpowiadającej rozkładowi czasu życia

## 7.4 Rozwiązanie



Z wykresów widać, że funkcja przeżycia dla pacjenta 2 maleje znacznie szybciej, co jest zgodne z charakterystyką danych.

Prawdopodobieństwa tego, że czas życia będzie większy od 300, to

$$S_{02}(300)$$

i

$$S_{01}(300)$$

Dla pacjenta 1 to prawdopodobieństwo wynosi:

```
## [1] 0.5033085
```

Jest to spora szansa.

A na liście 10 otrzymaliśmy:

```
## [1] 0.5901344
```

Prawdopodobieństwo dla modelu proporcjonalnych hazardów coxa jest większe.

Dla pacjenta 2 to prawdopodobieństwo wynosi:

```
## [1] 0.3069627
```

W stosunku do pacjenta pierwszego widać spadek aż o 20 punktów procentowych na liście 11.

```
## [1] 0.3597682
```

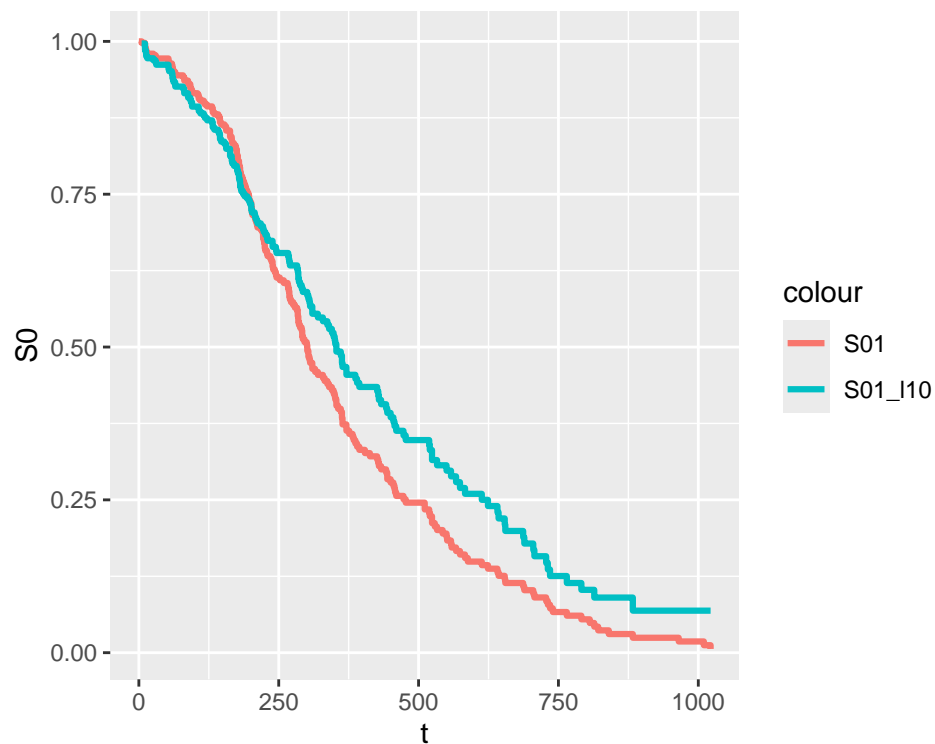
Prawdopodobieństwo dla modelu proporcjonalnych hazardów coxa jest większe.

## 8 Zadanie 6

### 8.1 Treść

Narysować wykres oszacowanej w zadaniu 5 punkt (a) funkcji przeżycia i porównać go z wykresem z zadania 6 z listy 10.

### 8.2 Rozwiązanie



Widać, że funkcja przeżycia jest większa dla modelu proporcjonalnych hazardów coxa, niż w obecnym modelu proporcjonalnych szans.