

# Analiza Przeżycia

## Raport 3

Wiktor Niedźwiedzki (258882)  
Filip Michewicz (282239)

21 stycznia 2026 Anno Domini

# Spis treści

<b>1 Dane lung</b>	<b>4</b>
<b>2 Lista 9</b>	<b>5</b>
2.1 Model przyspieszonego czasu awarii (AFT) . . . . .	5
2.2 Model proporcjonalnych hazardów (PH) . . . . .	5
2.3 Zadanie 1 . . . . .	6
2.4 Zadanie 2 . . . . .	7
2.5 Zadanie 3 . . . . .	8
2.6 Zadanie 4 . . . . .	8
2.7 Zadanie 5 . . . . .	9
2.8 Zadanie 6 . . . . .	10
2.9 Zadanie 7 . . . . .	10
2.10 Zadanie 8 . . . . .	13
2.11 Zadanie 9 . . . . .	13
<b>3 Lista 10</b>	<b>15</b>
3.1 Model proporcjonalnych hazardów Coxa . . . . .	15
3.2 Zadanie 1 . . . . .	15
3.3 Zadanie 2 . . . . .	16
3.4 Zadanie 3 . . . . .	16
3.5 Zadanie 4 . . . . .	17
3.6 Zadanie 5 . . . . .	19
3.7 Zadanie 6 . . . . .	20
<b>4 Lista 11</b>	<b>22</b>
4.1 Model proporcjonalnych szans . . . . .	22
4.2 Zadanie 1 . . . . .	22
4.3 Zadanie 2 . . . . .	23
4.4 Zadanie 3 . . . . .	23
4.5 Zadanie 4 . . . . .	23
4.6 Zadanie 5 . . . . .	26
4.7 Zadanie 6 . . . . .	26
<b>5 Lista 12</b>	<b>28</b>
5.1 Zadanie 1 . . . . .	28
5.2 Zadanie 2 . . . . .	28
5.3 Zadanie 3 . . . . .	28
5.4 Zadanie 4 . . . . .	28
<b>6 Bibliografia</b>	<b>29</b>

# Spis wykresów

1 Estymacja funkcji przeżycia - model AFT . . . . .	9
2 Estymowane funkcje hazardu dla odpowiednich charakterystyk 70-letnich kobiet - model PH .	11
3 Logarytmy estymowanych funkcji hazardu dla odpowiednich charakterystyk 70-letnich kobiet - model PH . . . . .	12
4 Różnica logarytmów estymowanych funkcji hazardu dla odpowiednich charakterystyk 70-letnich kobiet - model PH . . . . .	12
5 Porównanie estymowanych funkcji przeżycia - modele AFT oraz PH . . . . .	14
6 Estymowana skumulowana funkcja hazardu dla odpowiednich charakterystyk 70-letnich kobiet - model Cox'a . . . . .	17

7	Logarytm estymowanej skumulowanej funkcja hazardu dla odpowiednich charakterystyk 70-letnich kobiet - model Cox'a . . . . .	18
8	Różnica logarytmów estymowanej skumulowanej funkcji hazardu dla odpowiednich charakterystyk 70-letnich kobiet - model Cox'a . . . . .	19
9	Estymacja funkcji przeżycia w modelu Coxa dla różnych charakterystyk pacjentek . . . . .	20
10	Porównanie estymowanych funkcji przeżycia - modele PH oraz Coxa . . . . .	21
11	Estymowana skumulowana funkcja hazardu dla odpowiednich charakterystyk 70-letnich kobiet - modele Coxa i proporcjonalnych szans . . . . .	24
12	Logarytm estymowanej skumulowanej funkcji hazardu dla odpowiednich charakterystyk 70-letnich kobiet - model proporcjonalnych szans . . . . .	25
13	Różnica logarytmów estymowanej skumulowanej funkcji hazardu dla odpowiednich charakterystyk 70-letnich kobiet - model proporcjonalnych szans . . . . .	25
14	Różnica logarytmów estymowanej szans dla odpowiednich charakterystyk 70-letnich kobiet . . . . .	26
15	Porównanie estymowanych funkcji przeżycia - modele Coxa oraz proporcjonalnych szans . . . . .	27

## Spis tabel

1	Opis zmiennych w zbiorze danych lung . . . . .	4
2	Współczynniki modelu AFT . . . . .	7
3	Współczynniki modelu PH . . . . .	10
4	Współczynniki modelu Coxa . . . . .	16
5	Estymowana funkcja przeżycia dla czasu $t = 300$ w procentach - porównanie modeli PH oraz Coxa . . . . .	20
6	Współczynniki modelu proporcjonalnych szans . . . . .	23
7	Estymowana funkcja przeżycia dla czasu $t = 300$ w procentach - porównanie modeli Coxa oraz proporcjonalnych szans . . . . .	26

## 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 (dana kompletna))
age	numeric	Wiek w latach
sex	factor	Płeć (1 - mężczyzna, 2 - kobieta)
ph.ecog	factor	Skala ECOG według lekarza
ph.karno	numeric	Ocena pacjenta w skali Karnofsky'ego według lekarza
pat.karno	factor	Ocena pacjenta w skali 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 chemicznej 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 [1].
- 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 [2].

Porównanie skali ECOG ze skalą Karnofsky'ego:

- **ECOG 0** – 100, 90
- **ECOG 1** – 80, 70
- **ECOG 2** – 60, 50
- **ECOG 3** – 40, 30
- **ECOG 4** – 20, 10
- **ECOG 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 obserwowej 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 przyspieszonego czasu awarii (AFT)

Modelem, w którym funkcja przeżycia  $S(x|z)$  jednostki o wektorze charakterystyk  $z$  ma postać

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

nazywamy modelem przyspieszonego czasu awarii (AFT - ang. *accelerated failure time*).

Funkcję  $S_0(x)$  nazywamy bazową funkcją przeżycia, natomiast wyrażenie  $\exp(\beta^T z)$  — czynnikiem przyspieszenia (skalującym czas).

Model AFT posiada następującą interpretację. Jeżeli  $z_1$  oraz  $z_2$  są dwoma wektorami charakterystyk, to prawdopodobieństwo przeżycia czasu  $x$  przez jednostkę o charakterystyce  $z_1$  jest równe prawdopodobieństwu przeżycia czasu  $\exp(\beta^T(z_1 - z_2)) x$  przez jednostkę o charakterystyce  $z_2$ . Oznacza to, że porównanie jednostek sprowadza się do odpowiedniego przeskalowania osi czasu.

W szczególności:

- jeżeli  $\exp(\beta^T(z_1 - z_2)) > 1$  (tj.  $\beta^T(z_1 - z_2) > 0$ ), to jednostka o charakterystyce  $z_1$  ma większe prawdopodobieństwo wystąpienia zdarzenia przed chwilą  $x$  niż jednostka o charakterystyce  $z_2$ ,
- jeżeli  $\exp(\beta^T(z_1 - z_2)) < 1$  (tj.  $\beta^T(z_1 - z_2) < 0$ ), to jednostka o charakterystyce  $z_1$  ma mniejsze prawdopodobieństwo wystąpienia zdarzenia przed chwilą  $x$  niż jednostka o charakterystyce  $z_2$ ,
- w szczególności, gdy  $\exp(\beta^T(z_1 - z_2)) = 1$  (tj.  $\beta^T(z_1 - z_2) = 0$ ), mamy  $z_1 = z_2$ .

Korzystając z zależności

$$h(x) = -\frac{S'(x)}{S(x)},$$

model AFT można równoważnie zapisać w postaci

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

W praktyce, jako bazową funkcję przeżycia  $S_0(x)$  w modelu AFT często przyjmuje się funkcję odpowiadającą rozkładowi Weibulla (w szczególnym przypadku rozkładowi wykładniczemu), Gomperta, lognormalnemu, logistycznemu oraz rozkładowi wartości ekstremalnych.

### 2.2 Model proporcjonalnych hazardów (PH)

W (parametrycznym) modelu proporcjonalnych hazardów (PH – ang. *proportional hazards*) funkcja hazardu  $h(x|z)$  jednostki o wektorze charakterystyk  $z$  ma postać

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

gdzie  $h_0(x)$  oznacza bazową funkcję hazardu, a wektor parametrów  $\beta$  opisuje wpływ charakterystyk na intensywność wystąpienia zdarzenia.

Model proporcjonalnych hazardów zawsze ma nazwę bezpośrednio interpretacji. Jeżeli  $z_1 = (z_{11}, z_{12}, \dots, z_{1p})^T$  oraz  $z_2 = (z_{21}, z_{22}, \dots, z_{2p})^T$  są dwoma wektorami charakterystyk, to

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

czyli iloraz hazardów dwóch jednostek nie zależy od czasu  $x$ , a jedynie od różnicy ich charakterystyk. Oznacza to, że hazardy są proporcjonalne w całym horyzoncie obserwacji.

Korzystając z zależności pomiędzy funkcją hazardu a funkcją przeżycia

$$S(x) = \exp\left(-\int_0^x h(u) du\right),$$

model PH można równoważnie zapisać w postaci funkcji przeżycia:

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

gdzie  $S_0(x) = \exp(-\int_0^x h_0(u) du)$  jest bazową funkcją przeżycia.

W praktyce, jako bazową funkcję hazardu  $h_0(x)$  w modelu PH często przyjmuje się funkcję odpowiadającą rozkładowi Weibulla (w szczególnym przypadku rozkładowi wykładniczemu), model o kawałkami stałym hazardzie, rozkład Gompertza oraz rozkład ekstremalnych.

### 2.3 Zadanie 1

W tym zadaniu oszacowano parametry modelu przyspieszonego czasu awarii (AFT). Jako zmienną zależną przyjęto *time*, natomiast jako wektor charakterystyk: *age*, *sex*, *ph.ecog* oraz *ph.karno*. Do estymacji modelu wykorzystano funkcję *survreg* z pakietu *survival*.

Podczas konstruowania modelu uwzględniono fakt, że zmienne *sex* oraz *ph.ecog* mają charakter jakościowy (typ *factor*), natomiast zmienne *age* i *ph.karno* zostały poddane centrowaniu. Zabieg ten jest zalecany, gdyż poprawia własności numeryczne estymacji oraz interpretowalność parametrów modelu.

**UWAGA:** Po ograniczeniu zbioru danych, od tego momentu każdorazowe odwołanie do wartości w skali ECOG będzie dotyczyło zmiennej *ph.ecog*, czyli oceny nadawanej przez lekarza. Analogicznie, stopień w skali Karnofsky'ego będzie oznaczał wartość zmiennej *ph.karno*, również ustalaną przez lekarza.

Po wyznaczeniu odpowiedniego podzbioru danych usunięto obserwacje zawierające braki danych. Liczba usuniętych wierszy wyniosła **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
model.aft <- survreg(
  Surv(time, status) ~ age_new + as.factor(ph.ecog) +
    ph.karno_new + as.factor(sex),
  data = lung_new,
  dist = "weibull"
)
```

W poniższej tabeli przedstawiono współczynniki modelu. Ich interpretacja znajduje się w kolejnym zadaniu.

Tabela 2: Współczynniki modelu AFT

Charakterystyka	Wartość współczynnika $\beta$	$\exp(\beta)$
Wyraz wolny	-6.301	0.002
Wiek	0.009	1.009
ECOG 1	0.422	1.524
ECOG 2	0.926	2.525
ECOG 3	1.687	5.405
Stopień Karnofsky'ego	0.010	1.010
Płeć = 2 (kobieta)	-0.408	0.665

## 2.4 Zadanie 2

W tym zadaniu zinterpretowane zostały wyznaczone współczynniki dla poszczególnych zmiennych. Współczynniki zostały przedstawione w Tabeli 2.

Interpretacja współczynników w modelu przyspieszonego czasu awarii (AFT) opiera się na czynniku przyspieszenia czasu, czyli wartości  $\exp(\beta)$ .

Wartości  $\exp(\beta_i) > 1$  oznaczają przyspieszenie wystąpienia zdarzenia (skrócenie czasu przeżycia), natomiast  $\exp(\beta_i) < 1$  oznaczają spowolnienie zdarzenia (wydłużenie czasu przeżycia) w porównaniu z jednostką odniesienia.

Wyraz wolny odpowiada bazowej funkcji przeżycia dla jednostki referencyjnej, której zmienne objaśniające przyjmują wartości odniesienia: wiek i stopień Karnofsky'ego równy średniej w próbce (po centrowaniu), ECOG = 0 oraz płeć męska. Wtedy przewidywany czas przeżycia tej jednostki opisuje funkcja

$$S_0(\exp(\beta^T z_0) t),$$

gdzie  $z_0$  oznacza wektor zmiennych objaśniających przy wartościach odniesienia, a  $\exp(\beta^T z_0) = 0.002$  pełni rolę czynnika skalującego oś czasu względem bazowej funkcji  $S_0(t)$ .

Współczynniki modelu AFT działają multiplikatywnie na przewidywany czas przeżycia: przewidywany czas jednostki o charakterystykach  $z$  jest skalowany względem jednostki referencyjnej przez czynnik  $\exp(\beta^T z) = \exp(\beta_1 z_1) \cdot \exp(\beta_2 z_2) \cdots \exp(\beta_p z_p)$ . Oznacza to, że efekty wszystkich zmiennych łączą się poprzez mnożenie, wpływając jednocześnie na przewidywany czas przeżycia.

Po odpowiednich obliczeniach można pokazać, że wartość ilorazu prawdopodobieństw zgonu w chwili  $x$  dla charakterystyk  $z_1$  oraz  $z_2$  jest zależna od czasu, tym samym nie jest ona stała.

$$\frac{S(x|z_1)}{S(x|z_2)} = \exp \left( \left( \frac{\exp(\beta^T z_2) x}{\lambda} \right)^k - \left( \frac{\exp(\beta^T z_1) x}{\lambda} \right)^k \right)$$

$$\frac{S(x|z_1)}{S(x|z_2)} = \exp \left( \left( \left( \frac{\exp(\beta^T z_2)}{\lambda} \right)^k - \left( \frac{\exp(\beta^T z_1)}{\lambda} \right)^k \right) x^k \right)$$

Oznacza to, że nie można podać konkretnego procentu o ile zwiększa się szansa przeżycia przy zmianie charakterystyki dla dowolnego  $x$ . Można jedynie podać różnicę w wartościach  $x$  dla tej samej wartości prawdopodobieństwa zgonu.

W takim razie prawdopodobieństwo, że osoba A starsza o rok od osoby B przeżyje czas  $x$  jest równe szansie, że osoba B przeżyje czas o 0.9% dłuższy ( $1.009x$ ). Starszy wiek zwiększa w każdej chwili szansę na zgon.

Wpływ stopnia sprawności ECOG na czas przeżycia jest bardzo wyraźny. Prawdopodobieństwo, że pacjent w stanie ECOG = 1 przeżyje czas  $x$ , odpowiada szansie na przeżycie czasu o 52,4% dłuższego ( $1.524x$ ) przez pacjenta w pełni sprawnego (ECOG = 0). W przypadku stanu ECOG = 2, proces ten zachodzi jeszcze szybciej – czas  $x$  u chorego odpowiada aż 252,5% tego czasu u osoby sprawnej. Największą różnicę widać dla grupy ECOG 3, gdzie dany poziom ryzyka osiągany w czasie  $x$  jest równy poziomowi ryzyka sprawnego pacjenta w momencie ponad pięciokrotnie większym (5.405x).

Każdy punkt mniejszy w skali Karnofsky'ego względem średniej skraca przewidywany czas przeżycia o około 1% ( $\exp(-0.01) \approx 0.990$ ).

Każdy punkt mniejszy w skali Karnofsky'ego względem średniej zwiększa szacowany czas przeżycia o prawdopodobieństwie  $p$  o 1%.

Prawdopodobieństwo na zgon mężczyzn w chwili  $x$  wynosi tyle samo, co szansa na śmierć kobiety w momencie  $\frac{x}{0.665} \approx 1.504x$ .

## 2.5 Zadanie 3

W tym zadaniu oszacowano funkcję przeżycia (w dniach) dla 70-letniej kobiety o stopniu sprawności fizycznej ECOG równym 1 oraz stopniu Karnofsky'ego równym 90. Dane ciągle zostały wcześniej zcentrowane względem średnich w próbie.

```
print(df$beta)

## [1] -6.301180488  0.008601783  0.421560818  0.926113456  1.687243371
## [6]  0.010056702 -0.408280141

beta_aft <- unname((-model.aft$coefficients[-1]))
print(beta_aft)

## [1]  0.008601783  0.421560818  0.926113456  1.687243371  0.010056702
## [6] -0.408280141

z <- c(70 - age_average, 1, 0, 0, 90 - ph.karno_average, 1)

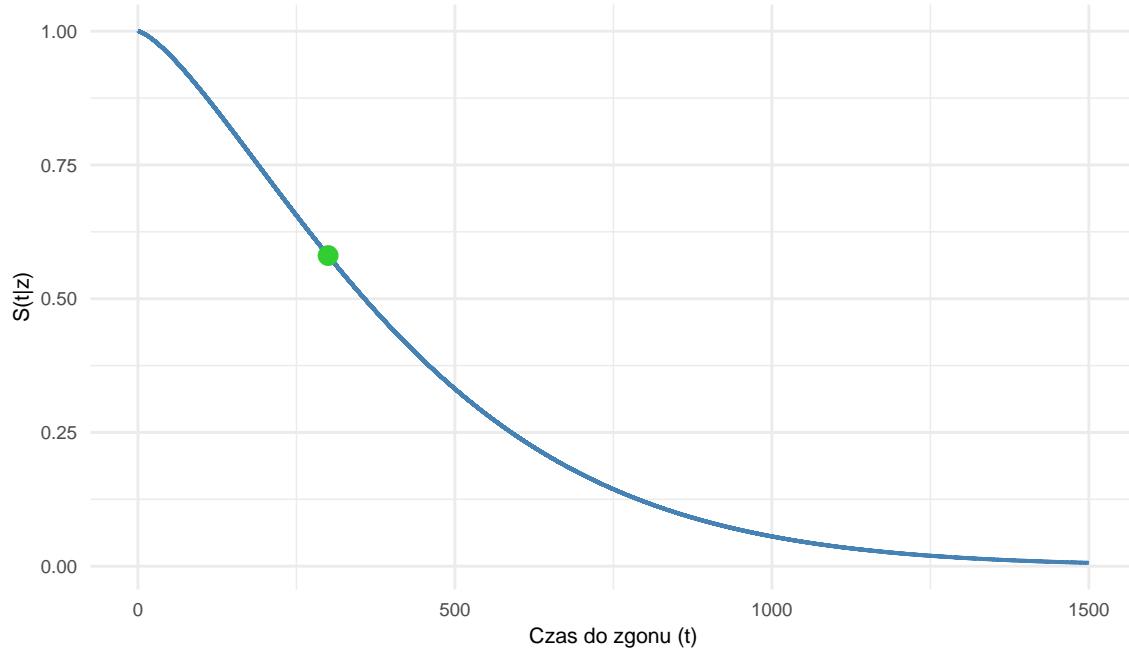
mi <- unname(model.aft$coefficients[1])
sigma <- model.aft$scale
alpha_aft <- 1 / sigma
lambda_aft <- exp(-alpha_aft * mi)

surv_fun_aft <- function(t, z) {
  theta_aft <- c(t(beta_aft) %*% z)
  exp(-lambda_aft * exp(alpha_aft * theta_aft) * t^alpha_aft)
}
```

Na podstawie oszacowanej funkcji przeżycia można obliczyć prawdopodobieństwo, że czas życia kobiety przekroczy 300 dni. Estymacja wynosi **58.06 %**.

## 2.6 Zadanie 4

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



Wykres 1: Estymacja funkcji przeżycia - model AFT

Na Wykresie 1. widać, że szacowane prawdopodobieństwo przeżycia szybko maleje do ok. 500 dni, potem zaś prędkość ta znaczowo maleje. Od czasu  $t = 1138$  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 oszacowano parametry modelu proporcjonalnych hazardów (PH). Podobnie jak wcześniej, zmienną zależną jest *time*, a zmiennymi objaśniającymi: *age*, *sex*, *ph.ecog* oraz *ph.karno*. Podobnie jak wcześniej, zmienne jakościowe traktowane są jako **factor**, natomiast zmienne ciągłe zostały zcentrowane.

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

W poniższej tabeli przedstawiono współczynniki modelu. Ich interpretacja znajduje się w kolejnym zadaniu.

Tabela 3: Współczynniki modelu PH

Charakterystyka	Wartość współczynnika $\beta$	$\exp(\beta)$
Wiek	0.012	1.012
ECOG 1	0.585	1.795
ECOG 2	1.285	3.615
ECOG 3	2.341	10.394
Stopień Karnofsky'ego	0.014	1.014
Płeć = 2 (kobieta)	-0.567	0.567
$\log(\text{scale})$	6.301	545.215
$\log(\text{shape})$	0.328	1.388

## 2.8 Zadanie 6

W tym zadaniu zinterpretowane zostały wyznaczone współczynniki dla poszczególnych zmiennych. Współczynniki zostały przedstawione w Tabeli 3.

Interpretacja współczynników w modelu PH opiera się na ilorazie hazardów, czyli wartości  $\exp(\beta)$ . Wartości  $\exp(\beta_i) > 1$  oznaczają zwiększenie ryzyka wystąpienia zdarzenia (wyższy hazard), natomiast  $\exp(\beta_i) < 1$  oznaczają zmniejszenie ryzyka (niższy hazard) w porównaniu z jednostką odniesienia.

Jednostka odniesienia to hipotetyczna jednostka, dla której wszystkie zmienne jakościowe przyjmują wartości referencyjne (ECOG = 0, mężczyzna), a zmienne ciągłe są ustawione na wartości średnie w próbie (po centrowaniu). Współczynniki  $\beta$  opisują więc zmianę hazardu jednostki w stosunku do tej jednostki odniesienia.

Każdy dodatkowy rok życia zwiększa hazard zgonu o około 1,2% ( $\exp(0.012) \approx 1.012$ ).

W porównaniu z ECOG = 0 osoby z ECOG = 1 mają hazard większy o około 79,5% ( $\exp(0.585) \approx 1.795$ ), z ECOG = 2 - o około 261,5% ( $\exp(1.285) \approx 3.615$ ), a z ECOG = 3 - o około 939,4% ( $\exp(2.341) \approx 10.394$ ). Wyższe wartości ECOG oznaczają znacznie większe ryzyko zdarzenia, czyli krótszy przewidywany czas przeżycia.

Każdy punkt wyższy w skali Karnofsky'a zwiększa hazard o około 1,4% ( $\exp(0.014) \approx 1.014$ ).

W porównaniu z mężczyznami, kobiety mają hazard mniejszy o około 43% ( $\exp(-0.567) \approx 0.567$ ), co oznacza dłuższy przewidywany czas przeżycia.

Parametry  $\log(\text{scale})$  i  $\log(\text{shape})$  określają odpowiednio skalę i kształt bazowego rozkładu Weibulla przyjętego w modelu.

## 2.9 Zadanie 7

W tym zadaniu oszacowano dwie funkcje hazardu (w dniach) odpowiadające rozkładowi czasu życia 70-letnich kobiet o tej samej wartości stopnia Karnofsky'ego równym 90 oraz zróżnicowanej wartości charakterystyki ECOG wynoszących 1 oraz 2. Dane ciągłe zostały wcześniej zcentrowane względem średnich w próbie. Następnie porównano uzyskane estymacje funkcji hazardu.

Funkcje hazardu oszacowano w modelu proporcjonalnych hazardów z rozkładem Weibulla. Współczynniki regresji odpowiadają poszczególnym cechom jednostek, natomiast parametry kształtu i skali definiują bazową funkcję hazardu Weibulla:

$$h(t | z) = h_0(t) \exp(\beta^T z),$$

gdzie bazowy hazard Weibulla jest określony wzorem:

$$h_0(t) = \lambda \alpha t^{\alpha-1}, \quad \lambda = \exp(-\log(\text{scale}) \cdot \alpha), \quad \alpha = \exp(\log(\text{shape})).$$

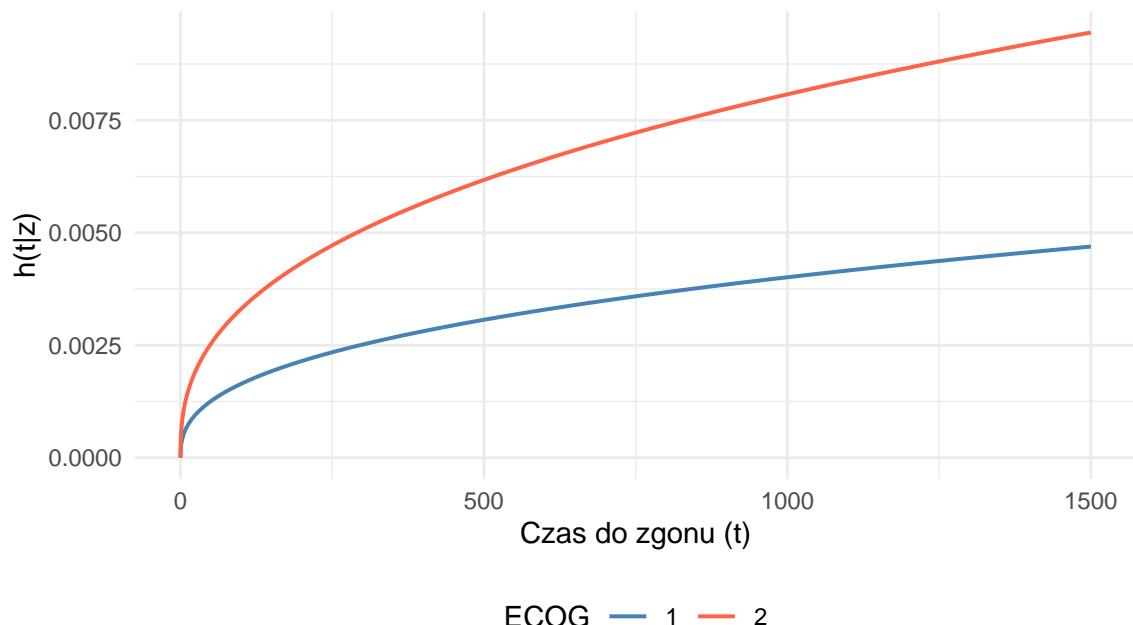
Wektor  $z$  zawiera wartości zmiennych objaśniających zcentrowanych względem średnich oraz zakodowanych zmiennych jakościowych. Bazowy hazard  $h_0(t)$  opisuje tempo ryzyka w jednostce odniesienia (dla wszystkich zmiennych jakościowych przyjmujących wartości referencyjne i zmiennych ciągłych ustawionych na średnich), natomiast współczynniki  $\beta$  skalują go dla jednostek o innych wartościach cech.

```
beta_ph <- unname(model.ph$coefficients[-c(7, 8)])
alpha_ph <- exp(model.ph$coefficients[["log(shape)"]])
mu <- model.ph$coefficients[["log(scale)"]]
lambda_ph <- exp(-mu * alpha_ph)

# 1 kobieta: ph.ecog=1
z1 <- c(70 - age_average, 1, 0, 0, 90 - ph.karno_average, 1)
theta_ph1 <- c(t(beta_ph) %*% z)

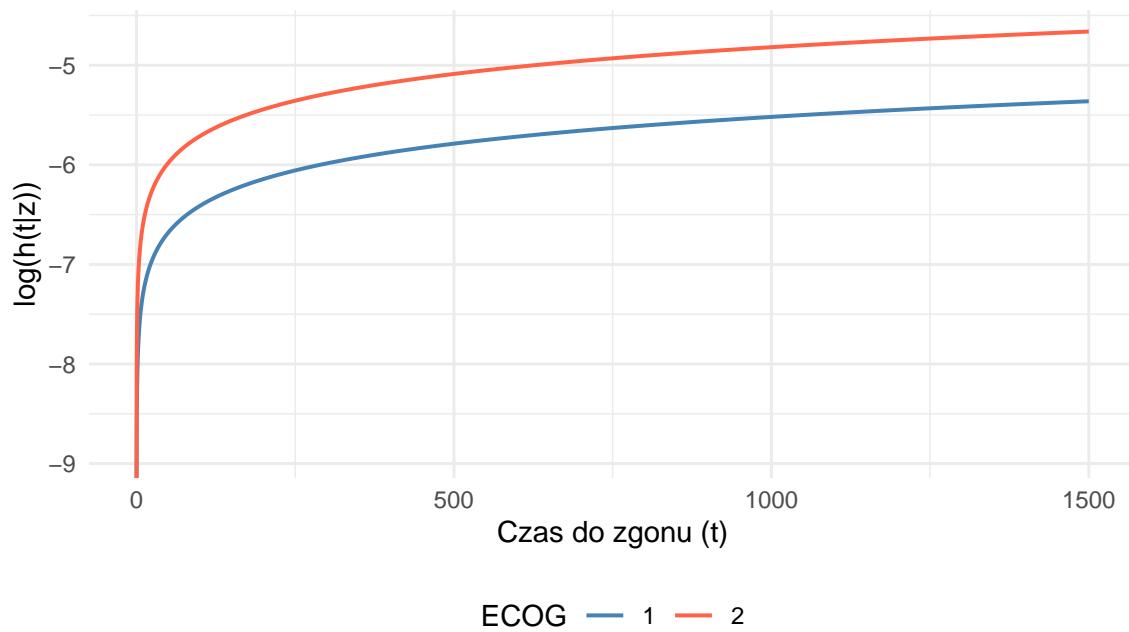
# 2 kobieta: ph.ecog=2
z2 <- c(70 - age_average, 0, 1, 0, 90 - ph.karno_average, 1)

hazard_fun_ph <- function(t, z) {
  theta_ph <- c(t(beta_ph) %*% z)
  lambda_ph * alpha_ph * t^(alpha_ph - 1) * exp(theta_ph)
}
```



Wykres 2: Estymowane funkcje hazardu dla odpowiednich charakterystyk 70-letnich kobiet - model PH

Z Wykresu 2. można odczytać, że funkcja hazardu dla kobiety o charakterystyce ECOG równej 2 jest zawsze wyższa niż dla tej drugiej. Oznacza to, że w jej przypadku ryzyko zgonu w każdym momencie jest wyższe.



Wykres 3: Logarytmy estymowanych funkcji hazardu dla odpowiednich charakterystyk 70-letnich kobiet - model PH

Poszczególne krzywe funkcji log-hazardu na Wykresie 3. wydają się równoległe, co sugeruje, że iloraz hazardów może być stały w czasie.



Wykres 4: Różnica logarytmów estymowanych funkcji hazardu dla odpowiednich charakterystyk 70-letnich kobiet - model PH

Jak pokazano na wykresie 4. różnica logarytmów funkcji hazardu jest w przybliżeniu stała w czasie, co wskazuje na stałość ilorazu hazardów. W konsekwencji brak jest przesłanek empirycznych do odrzucenia

założenia proporcjonalnych hazardów, a przyjęcie modelu proporcjonalnych hazardów w analizie można uznać za uzasadnione.

## 2.10 Zadanie 8

W tym zadaniu, w oparciu o wcześniej wyznaczone funkcje hazardu w modelu proporcjonalnych hazardów z bazowym rozkładem Weibulla, wyznaczono odpowiadające im funkcje przeżycia.

Dla rozkładu Weibulla oraz parametryzacji zastosowanej w modelu proporcjonalnych hazardów otrzymujemy bazową funkcję przeżycia

$$S_0(t) = \exp(-\lambda t^\alpha),$$

natomiast funkcja przeżycia dla jednostki o wektorze cech ( $z$ ) przyjmuje postać

$$S(t|z) = [S_0(t)]^{\exp(\beta^T z)} = \exp(-\lambda t^\alpha \exp(\beta^T z)).$$

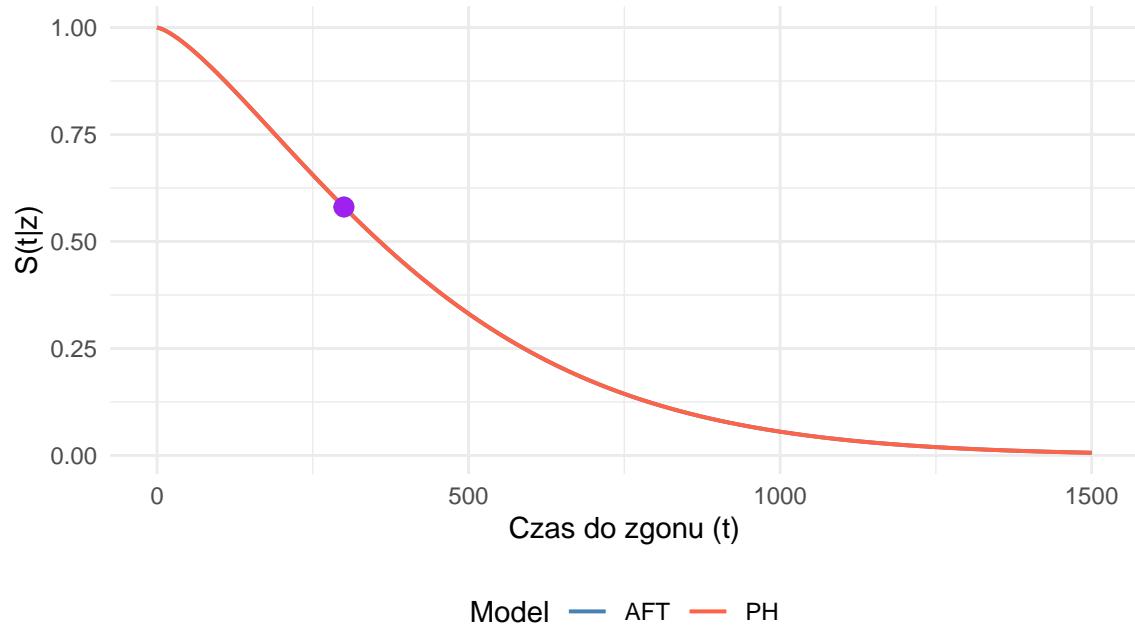
```
surv_fun_ph <- function(t, z) {  
  theta_ph <- c(t(beta_ph) %*% z)  
  exp(-lambda_ph*t^alpha_ph*exp(theta_ph))  
}
```

W oparciu o wyznaczone funkcje przeżycia obliczono estymowane prawdopodobieństwo, że kobiety przeżyją ponad 300 dni. Dla kobiety o stopniu sprawności w skali ECOG równym 2 wynosi ono **33.45%**, natomiast dla kobiety o stopniu sprawności w skali ECOG równym 1 — **58.06%**.

Otrzymana wartość dla kobiety o stopniu sprawności w skali ECOG równym 1 jest równa estymacji uzyskanej przy zastosowaniu modelu AFT. Sugeruje to, że oba modele definiują taki sam rozkład pomimo odmiennej interpretacji współczynników  $\beta$ . W przypadku modelu proporcjonalnych hazardów są one związane z multiplikatywną zmianą funkcji hazardu, a w modelu przyspieszonego czasu awarii odpowiednio przyspieszają lub opóźniają zdarzenie.

## 2.11 Zadanie 9

W tym zadaniu porównano funkcje przeżycia uzyskane za pomocą modeli AFT oraz PH dla 70-letniej kobiety o stopniu sprawności w skali Karnofsky'ego równym 90 oraz stopniu sprawności w skali ECOG równym 1.



Wykres 5: Porównanie estymowanych funkcji przeżycia - modele AFT oraz PH

Wykres 5. wskazuje idealne pokrycie rozkładów uzyskanych za pomocą modeli AFT oraz PH. To z kolei potwierdza nasze poprzednie przypuszczenia o uzyskanej matematycznej równoważności. Wynika ona bezpośrednio z faktu, że rozkład Weibulla jest jedynym rozkładem prawdopodobieństwa, który posiada jednocześnie cechę proporcjonalności hazardów oraz strukturę przyspieszonego czasu życia. Tym samym powoduje to brak zmiany kształtu funkcji przeżycia.

Ze względu na uzyskane pokrycie, w dalszej części raportu wyniki będą porównywane jedynie z modelem PH, gdyż wnioski dla modelu AFT będą identyczne.

### 3 Lista 10

Lista polega na praktycznym zastosowaniu modelu proporcjonalnych hazardów Coxa do danych o czasie przeżycia. Celem jest oszacowanie parametrów modelu, interpretacja wpływu zmiennych na ryzyko zdarzenia, wyznaczenie bazowej funkcji hazardu i funkcji przeżycia, a także ocena prawdopodobieństwa przeżycia dla wybranych profili pacjentów oraz weryfikacja założeń modelu poprzez wizualizacje.

#### 3.1 Model proporcjonalnych hazardów Coxa

Modelem, w którym funkcja hazardu jednostki o charakterystyce  $z$  ma postać

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

nazywamy modelem proporcjonalnych hazardów Coxa. Funkcja  $h_0(t)$  to bazowa funkcja hazardu, zależna od czasu, ale niezależna od charakterystyk, przyjmująca wartości nieujemne. Funkcja  $\psi(z)$  zależy wyłącznie od wektora charakterystyk  $z$ , jest nieujemna i zwykle przyjmuje postać parametryczną

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

gdzie  $\beta$  to wektor nieznanych współczynników modelu. Wówczas funkcja hazardu przyjmuje postać

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

co pozwala interpretować wpływ zmiennych objaśniających na ryzyko wystąpienia zdarzenia w sposób względny.

Ryzyko względne między jednostką o charakterystyce  $z_1$  a jednostką o charakterystyce  $z_2$  definiuje się jako

$$\frac{h_{z_1}(t)}{h_{z_2}(t)} = \exp[\beta^T(z_1 - z_2)],$$

czyli jest stałe w czasie i zależy tylko od różnicy charakterystyk. Dzięki temu nazwa modelu – proporcjonalne hazardy – odzwierciedla fakt, że hazard jednostek są proporcjonalne w całym okresie obserwacji.

Mimo że model proporcjonalnych hazardów Coxa przypomina parametryczny model proporcjonalnych hazardów (PH), podstawowa różnica polega na tym, że w modelu Coxa nie zakłada się konkretnej postaci bazowego hazardu  $h_0(t)$ . W modelu PH bazowy hazard jest parametrycznie określony, natomiast w modelu Coxa pozostaje nieznany i może mieć dowolny kształt, co sprawia, że estymacja współczynników  $\beta$  jest oparta wyłącznie na względnych hazardach i obserwowanych czasach przeżycia.

#### 3.2 Zadanie 1

Zadanie polega na oszacowaniu parametrów modelu proporcjonalnych hazardów Coxa. Jako zmienną zależną przyjęto *time*, natomiast jako wektor charakterystyk: *age*, *sex*, *ph.ecog* oraz *ph.karno*. Do estymacji modelu wykorzystano funkcję *coxph* z pakietu *survival*.

```
model.cox <- coxph(Surv(time, status) ~ age_new +
                      as.factor(ph.ecog) + ph.karno_new + as.factor(sex), data = lung_new)
```

W poniższej tabeli przedstawiono współczynniki modelu. Ich interpretacja znajduje się w kolejnym zadaniu.

Tabela 4: Współczynniki modelu Coxa

Charakterystyka	Wartość współczynnika $\beta$	$\exp(\beta)$
Wiek	0.013	1.013
ECOG 1	0.578	1.783
ECOG 2	1.240	3.455
ECOG 3	2.396	10.978
Stopień Karnofsky'ego	0.012	1.013
Płeć = 2 (kobieta)	-0.566	0.568

### 3.3 Zadanie 2

W tym zadaniu zinterpretowane zostały wyznaczone współczynniki dla poszczególnych zmiennych. Współczynniki zostały przedstawione w Tabeli 4. Interpretacja współczynników  $\beta_i$  w modelu Coxa opiera się na ryzyku względnym i hazardzie. Ryzyko wzajemne jednostki o charakterystyce przesuniętej o jeden punkt względem średniej próbki wynosi  $\exp(\beta_i)$ . Oznacza to, że funkcja hazardu tej jednostki jest równa funkcji hazardu jednostki o średnich wartościach charakterystyk pomnożonej przez  $\exp(\beta_i)$ , niezależnie od czasu.

Dla wieku każda jednostka wieku powyżej średniego wieku zwiększa hazard zgonu o około 1.3% ( $\beta = 0.01256$ ,  $\exp(\beta) \approx 1.0126$ ).

Hazard zgonu kobiety jest około 43% mniejszy niż hazard mężczyzny ( $\beta = -0.5657$ ,  $\exp(\beta) \approx 0.568$ ).

Jednostka z ECOG równym 1 ma hazard zgonu około 78% wyższy niż jednostka z ECOG równym 0 ( $\beta = 0.5781$ ,  $\exp(\beta) \approx 1.783$ ). Jednostka z ECOG równym 2 ma hazard ponad trzykrotnie większy niż jednostka z ECOG równym 0 ( $\beta = 1.2399$ ,  $\exp(\beta) \approx 3.455$ ), a jednostka z ECOG równym 3 ma hazard niemal 11-krotnie wyższy ( $\beta = 2.3959$ ,  $\exp(\beta) \approx 10.978$ ).

Dla stopnia Karnofsky'ego każdy punkt powyżej średniego wyniku zwiększa hazard zgonu o około 1.25% ( $\beta = 0.01242$ ,  $\exp(\beta) \approx 1.0125$ ).

Efekty są multiplikatywne, zmiana w jednej zmiennej mnoży bazowy hazard  $h_0(t)$  przez  $\exp(\beta_i)$ , a zmiany w różnych zmiennych działają razem poprzez iloczyn tych czynników, określając całkowity hazard jednostki.

### 3.4 Zadanie 3

Zadanie polega na oszacowaniu bazowej funkcji hazardu oraz bazowej funkcji przeżycia.

W przypadku modelu Coxa funkcja przeżycia jednostki o charakterystyce  $z$  ma postać

$$S_z(t) = [S_0(t)]^{\exp(\beta^T z)}.$$

Oszacowanie bazowej funkcji przeżycia uzyskuje się na podstawie oszacowanej bazowej funkcji hazardu

$$\hat{S}_0(t) = \exp \left[ - \int_0^t \hat{h}_0(u) du \right] = \exp \left[ -\hat{H}_0(t) \right].$$

Do oszacowania bazowej funkcji hazardu wykorzystuje się estymator Breslowa oparty na metodzie częstkowej wiarogodności

$$\hat{H}_0(t) = \sum_{\tau_i \leq t} \frac{1}{\sum_{j \in R(\tau_i)} \exp(\hat{\beta}^T z_j)},$$

gdzie  $R(\tau_i)$  jest zbiorem ryzyka, czyli jednostek znajdujących się w ryzyku w chwili  $\tau_i$ .

W pakiecie R w bibliotece `survival` estymację tę realizuje funkcja `basehaz()`.

```
basehaz_df <- basehaz(model.cox)

surv_df <- data.frame(surv = exp(-basehaz_df$hazard), time = basehaz_df$time)
```

### 3.5 Zadanie 4

Zadanie polega na oszacowaniu bazowych skumulowanych funkcji hazardu (w dniach) odpowiadających rozkładowi czasu życia dwóch 70-letnich kobiet o tej samej wartości stopnia Karnofsky'ego równym 90 oraz różnicowanej wartości charakterystyki ECOG wynoszącym 1 oraz 2. Dane ciągłe zostały wcześniej zcentrowane względem średnich w próbie. Następnie porównano uzyskane estymacje skumulowanej funkcji hazardu.

W modelu Coxa skumulowana funkcja hazardu jednostki o charakterystyce  $z$  wyraża się wzorem

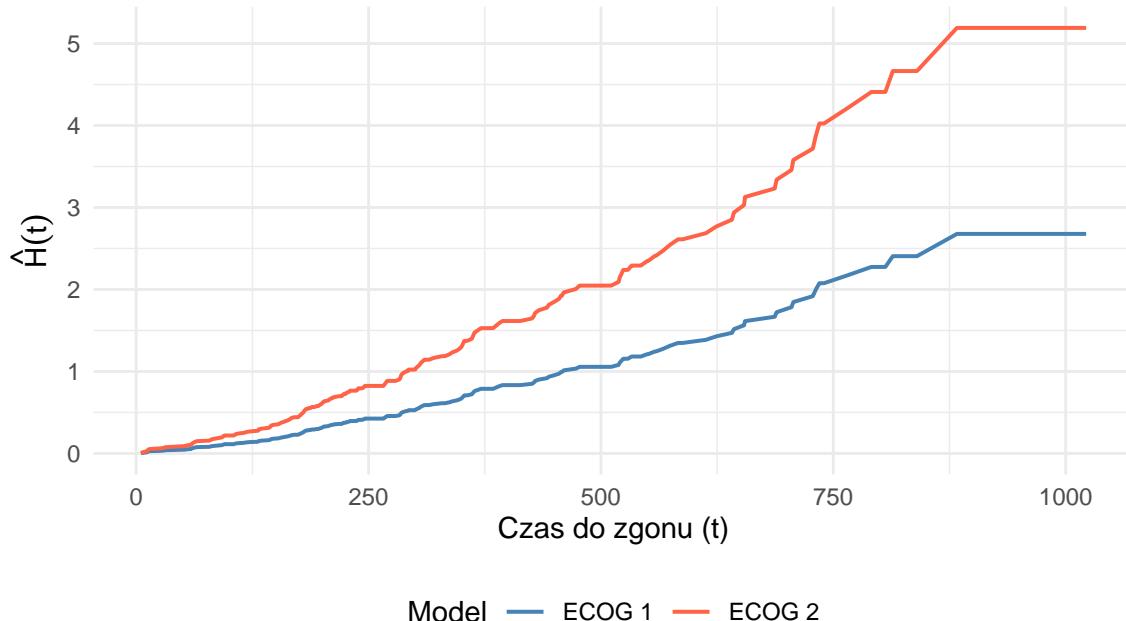
$$H_z(t) = H_0(t) \cdot \exp(\beta^T z),$$

Ten wzór wynika bezpośrednio z definicji modelu Coxa:

$$h_z(t) = h_0(t) \exp(\beta^T z) \Rightarrow H_z(t) = \int_0^t h_z(u) du = \int_0^t h_0(u) du \cdot \exp(\beta^T z) = H_0(t) \exp(\beta^T z).$$

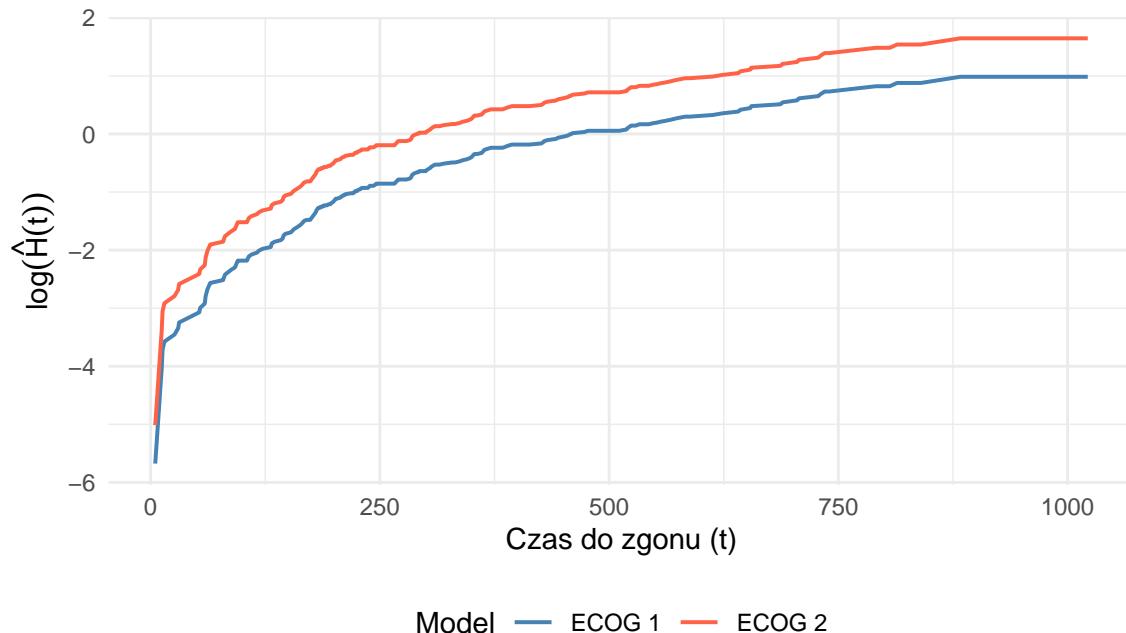
```
cum_hazard <- function(model, z) {
  basehazard <- basehaz(model)
  hazard <- basehazard$hazard * exp(sum(model$coefficients * z))
  data.frame(time = basehazard$time, hazard = hazard, loghazard = log(hazard))
}

cum_hazard_1 <- cum_hazard(model.cox, z1)
cum_hazard_2 <- cum_hazard(model.cox, z2)
```



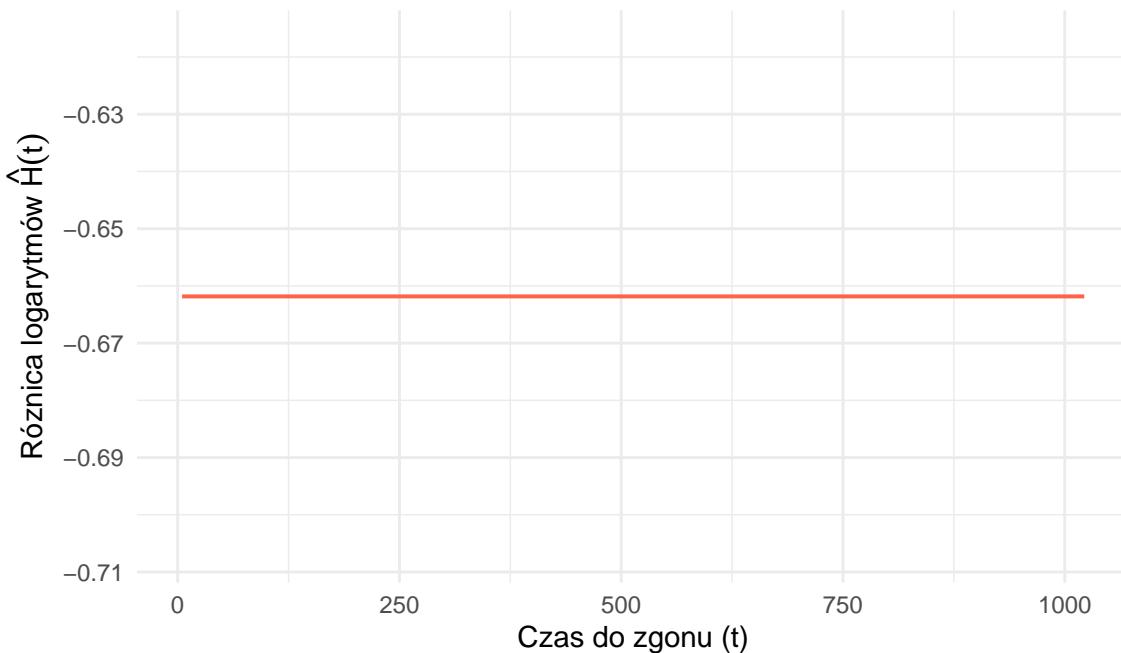
Wykres 6: Estymowana skumulowana funkcja hazardu dla odpowiednich charakterystyk 70-letnich kobiet - model Cox'a

Z Wykresu 6. można odczytać że skumulowany hazard dla kobiety o charakterystyce ECOG równej 2, jest większy niż dla kobiety o charakterystyce ECOG równej 1, co oznacza, że do danego momentu czasu kobieta z gorszym stanem ogólnym zdrowia (ECOG = 2) zgromadziła większe łączne ryzyko zgonu, a w konsekwencji odpowiadająca jej funkcja przeżycia przyjmuje mniejsze wartości w porównaniu z funkcją przeżycia kobiety o charakterystyce ECOG = 1.



Wykres 7: Logarytm estymowanej skumulowanej funkcji hazardu dla odpowiednich charakterystyk 70-letnich kobiet - model Cox'a

Na Wykresie 7. można zobaczyć, że logarytmy skumulowanej funkcji hazardu mają bardzo podobny kształt i wyglądają jakby były względem siebie przesunięte o stałą wartość w górę, co jest zgodne z założeniem modelu proporcjonalnych hazardów, zgodnie z którym różnica logarytmów skumulowanych hazardów nie zależy od czasu.



Wykres 8: Różnica logarytmów estymowanej skumulowanej funkcji hazardu dla odpowiednich charakterystyk 70-letnich kobiet - model Cox'a

Na Wykresie 8. przedstawiono różnicę logarytmów skumulowanej funkcji hazardu, która dla wszystkich argumentów pozostaje stała, oznacza to, że iloraz skumulowanych hazardów nie zależy od czasu.

Można łatwo pokazać, że w modelu Coxa zachodzi

$$\frac{H_{z_1}(t)}{H_{z_2}(t)} = \frac{\exp(\beta^T z_1) H_0(t)}{\exp(\beta^T z_2) H_0(t)} = \exp(\beta^T (z_1 - z_2)) = \frac{h_{z_1}(t)}{h_{z_2}(t)}.$$

Po zalogarytmowaniu otrzymujemy

$$\log H_{z_1}(t) - \log H_{z_2}(t) = \beta^T (z_1 - z_2) = \log h_{z_1}(t) - \log h_{z_2}(t),$$

czyli różnica logarytmów skumulowanych funkcji hazardu oraz różnica logarytmów funkcji hazardu jest stała w czasie.

W konsekwencji, na podstawie wykresów nie ma podstaw do wątpienia w poprawność przyjętego modelu proporcjonalnych hazardów, gdyż założenie proporcjonalności hazardów jest spełnione.

### 3.6 Zadanie 5

Zadanie polega na oszacowaniu funkcji przeżycia (w dniach) odpowiadającej rozkładowi czasu życia 70-letnich kobiet o takiej samej o tej samej wartości stopnia Karnofsky'ego równym 90 oraz zróżnicowanej wartości charakterystyki ECOG wykaczącym 1 oraz 2. Dane ciągle zostały wcześniej zcentrowane względem średnich w próbie.

```
surv_cox <- function(model, z){
  basehaz <- basehaz(model)
  hazard <- basehaz$hazard * exp(sum(model$coefficients * z))
  surv <- exp(-hazard)
  data.frame(time = basehaz$time, survival = surv)
```

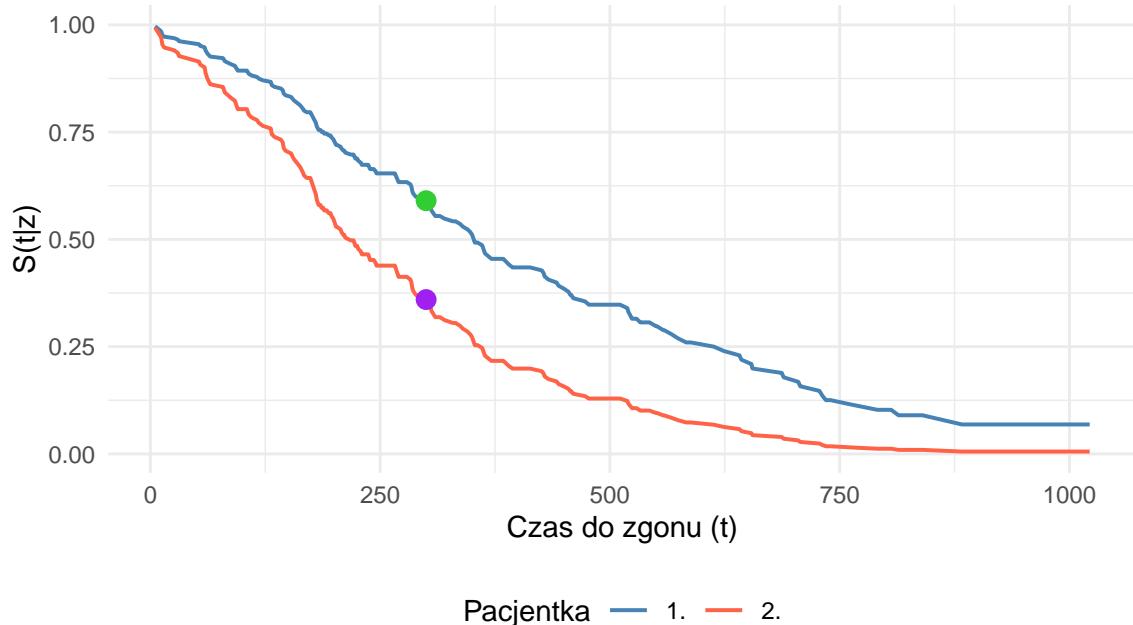
```

}

surv_at_time_cox <- function(surv_df, t) {
  idx <- which.min(abs(surv_df$time - t))
  surv_df$survival[idx]
}

df_surv_z1 <- surv_cox(model.cox, z1)
df_surv_z2 <- surv_cox(model.cox, z2)

```



Wykres 9: Estymacja funkcji przeżycia w modelu Coxa dla różnych charakterystyk pacjentek

Na Wykresie 9. widać różne funkcje przeżycia dla różnych charakterystyk pacjentek. Prawdopodobieństwo przeżycia kobiety o niższej charakterystyce ECOG jest większe dla każdego czasu.

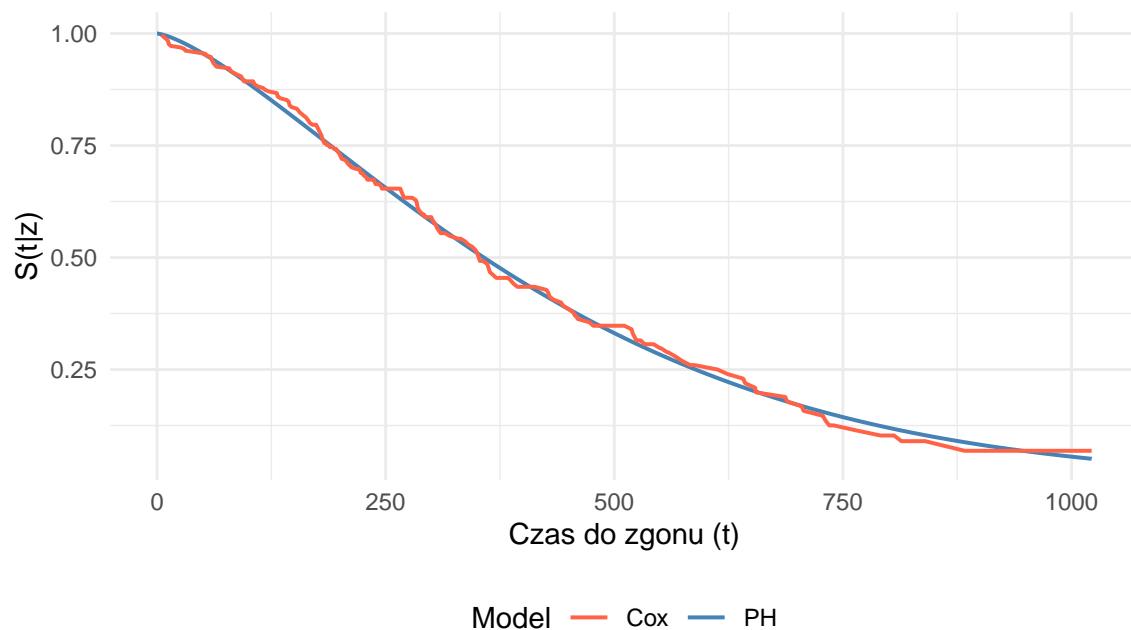
Tabela 5: Estymowana funkcja przeżycia dla czasu  $t = 300$  w procentach - porównanie modeli PH oraz Coxa

	PH	Cox
Pacjentka 1	58.06	59.01
Pacjentka 2	33.45	35.98

Porównując prawdopodobieństwa z Tabeli 5., można dojść do wniosku, że estymowane prawdopodobieństwa wartości dla modeli PH i Cox są do siebie bardzo zbliżone. To może sugerować, że nawet mimo braku założenia konkretnego rozkładu w modelu Coxa, wyniki są zbliżone do tych uzyskanych przy ustalonym z góry rozkładzie Weibulla.

### 3.7 Zadanie 6

W tym zadaniu narysowano wykresy estymowanej funkcji przeżycia dla 70-letniej kobiety o stopniu sprawności fizycznej ECOG równym 1 oraz stopniu Karnofsky równym 90 dla modeli PH oraz Coxa.



Wykres 10: Porównanie estymowanych funkcji przeżycia - modele PH oraz Coxa

Wykres 10. wskazuje zbliżoną do siebie estymację prawdopodobieństw uzyskaną modelami Coxa i PH. To z kolei sugeruje, że założenie dotyczące rozkładu Weibulla było prawidłowe.

## 4 Lista 11

Celem listy jest zastosowanie modelu proporcjonalnych szans. Analiza uwzględnia: oszacowanie parametrów modelu, interpretacja wpływu zmiennych na ryzyko zdarzenia, wyznaczenie bazowych funkcji skumulowanego hazardu i funkcji przeżycia oraz wyznaczenie wartości prawdopodobieństw dla wybranych profili pacjentów.

### 4.1 Model proporcjonalnych szans

Modelem proporcjonalnych szans nazywamy model, w którym szansę na wystąpienie zdarzenia (iloraz prawdopodobieństwa sukcesu do prawdopodobieństwa porażki) dla jednostki o charakterystyce  $z$  w chwili  $t$  określamy wzorem

$$\theta_z(t) = \frac{1 - S_z(t)}{S_z(t)}.$$

Model ten jest stosowany, kiedy założenie proporcjonalności hazardów nie jest spełnione i jest on semiparametryczny, gdyż nie zakładamy konkretnego rozkładu jednostki o zerowym wektorze charakterystyk oraz nie znamy wartości współczynników  $\beta$ . Po ich wyznaczeniu jednakże możemy określić dokładniejsze wzory na  $\theta_z$ :

$$\theta_z(t) = \theta_0(t) \exp(\beta^T z), \text{ gdzie } \theta_0(t) = \frac{1 - S_0(t)}{S_0(t)}.$$

Ze wzoru wynika też prosta interpretacja, mamy bowiem

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

czyli iloraz szans dwóch jednostek o dowolnych charakterystykach  $z_1, z_2$  jest stały i niezależny od czasu.

Przyjmując model proporcjonalnych szans, możemy dla jednostki o charakterystyce  $z$  wyznaczyć postać funkcji przeżycia

$$S_z(t) = \frac{1}{1 + \theta_z(t)} = \frac{1}{1 + \theta_0(t) \exp(\beta^T z)}$$

oraz funkcji hazardu

$$h_z(t) = -\frac{S'_z(t)}{S_z(t)} = \frac{\theta'_z(t)}{1 + \theta_z(t)} = \frac{\theta'_0(t)}{1 + \theta_0(t) \exp(\beta^T z)}$$

### 4.2 Zadanie 1

Zadanie polega na oszacowaniu parametrów modelu proporcjonalnych szans. Jako zmienną zależną przyjęto *time*, natomiast jako wektor charakterystyk: *age*, *sex*, *ph.ecog* oraz *ph.karno*. Do estymacji modelu wykorzystano funkcję *prop.odds* z pakietu *timereg*. Dane ciągle zostały zcentrowane. Dodatkowo, ze względu na specyfikację funkcji, zmieniono oznaczenia statusu danych na 0 - obserwacja cenzurowana oraz 1 - dana kompletna.

```
lung_new$status_new <- lung_new$status - 1
model.odds <- prop.odds(Event(time, status_new) ~ age_new +
                           as.factor(ph.ecog) +
                           ph.karno_new +
                           as.factor(sex),
                           data = lung_new)
```

W tabeli poniżej przedstawione są estymowane wartości współczynników.

Tabela 6: Współczynniki modelu proporcjonalnych szans

Charakterystyka	Wartość współczynnika $\beta$	$\exp(\beta)$
Wiek	0.013	1.013
ECOG 1	0.547	1.728
ECOG 2	1.447	4.249
ECOG 3	1.925	6.852
Stopień Karnofsky'ego	-0.004	0.996
Płeć = 2 (kobieta)	-0.954	0.385

### 4.3 Zadanie 2

W tym zadaniu zinterpretowane zostały wyznaczone współczynniki dla poszczególnych zmiennych znajdujące się w tabeli 6. Interpretacja opiera się na ilorazie estymowanych szans i dla jednostki o charakterystyce  $z_i$  przesuniętej o jeden punkt względem średniej próbki wynosi  $\exp(\beta_i)$ . Jest to równoznaczne z tym, że szansa tej jednostki na zdarzenie przed chwilą  $t$  jest równe szansie jednostki o zerowym wektorze charakterystyk pomnożonej przez wartość  $\exp(\beta_i)$ , niezależnie od ustalonego  $t$ .

W naszym przypadku badanym zdarzeniem jest śmierć pacjenta. Oznacza to, że większa wartość szansy w chwili  $t$  jest równoznaczna z większą szansą na zgon przed tym momentem.

W przypadku wieku każdy dodatkowy rok życia zwiększa szansę na zgon o około 1,3% ( $\exp(0.0132) \approx 1.0132$ ).

Zmiana skali ECOG z wartości 0 na wartość 1 powoduje wzrost szansy na śmierć o prawie 73% ( $\exp(0.547) \approx 1.728$ ). W przypadku wartości ECOG = 2 szansa ta wzrasta ponad czterokrotnie ( $\exp(1.4466) \approx 4.2487$ ), a ECOG = 3: prawie siedmiokrotnie ( $\exp(1.925) \approx 6.852$ ).

Wzrost skali Karnofsky'ego powoduje zmniejszenie szansy śmierci o wartość poniżej 0,4% ( $1 - \exp(-0.0037) \approx 0.369$ ).

Szansa mężczyzn na śmierć przed chwilą  $t$  jest około 2,5 raza większa niż w przypadku kobiet ( $\exp(0.9535) \approx 2.595$ ).

### 4.4 Zadanie 3

Zadanie polega na oszacowaniu wartości skumulowanej funkcji hazardu oraz przeżycia. W tym celu zastosowano funkcję `predict` z pakietu `stats`. Wykorzystano zależność  $H(t) = -\log(S(t))$ .

```
S_0 <- predict(model.odds, Z=rep(0, 6))
S <- data.frame("time"=S_0$time,
                 "S0"=c(S_0$S0),
                 "H0"=c(-log(S_0$S0)))
```

### 4.5 Zadanie 4

W tym zadaniu oszacowano wartości i narysowano wykresy skumulowanych funkcji hazardu dla 70-letnich kobiet o stopniu Karnofsky'ego równym 90 oraz stopniu sprawności fizycznej ECOG równym odpowiednio 1 lub 2. Dokonano również porównania uzyskanych wyników z estymowanymi wartościami dla modelu proporcjonalnych hazardów Coxa.

W tym celu napisano funkcję, która dla danego wektora charakterystyk  $z$  wylicza estymowane wartości funkcji przeżycia, skumulowanego hazardu oraz dodatkowo szans.

```

beta_odds <- unname(model.odds$gamma)

surv_cum_haz_fun_odds <- function(z) {
  theta_odd <- exp(sum(beta_odds*z))
  df <- data.frame("time" = S$time,
    "S" = 1 / (1 + (1-S$S0)/S$S0 * theta_odd))
  df$H <- -log(df$S)
  df$odds <- (1-df$S)/df$S
  df
}

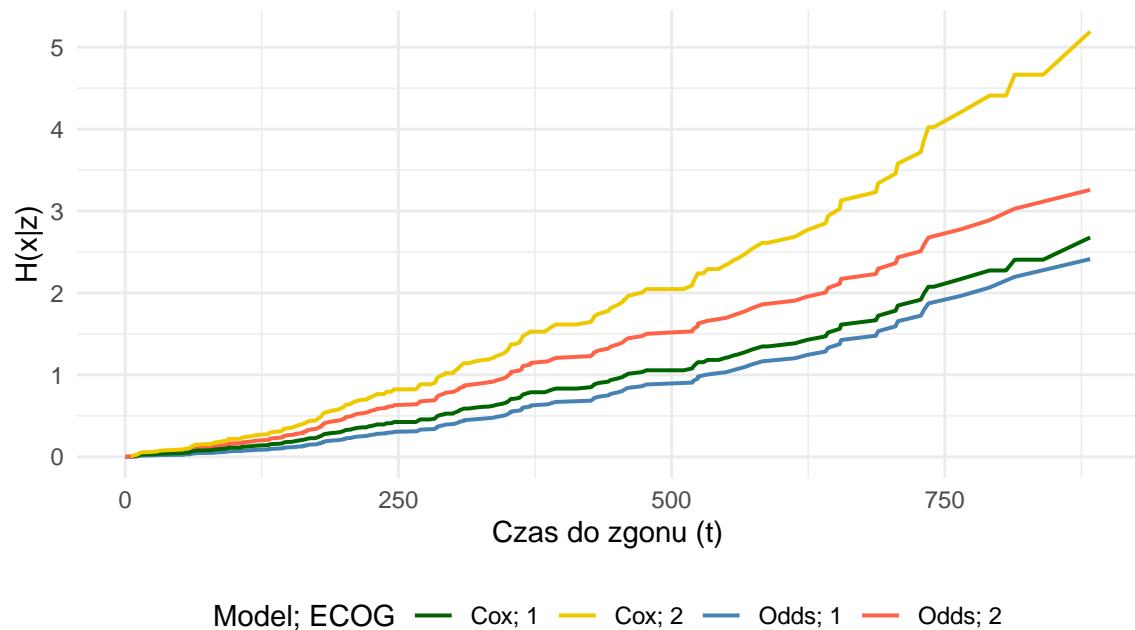
```

Tworzymy ramki danych dla odpowiednich wektorów charakterystyk  $z$ .

```

Z1 <- surv_cum_haz_fun_odds(c(70-age_average, 1, 0, 0, 90-ph.karno_average, 1))
Z2 <- surv_cum_haz_fun_odds(c(70-age_average, 0, 1, 0, 90-ph.karno_average, 1))

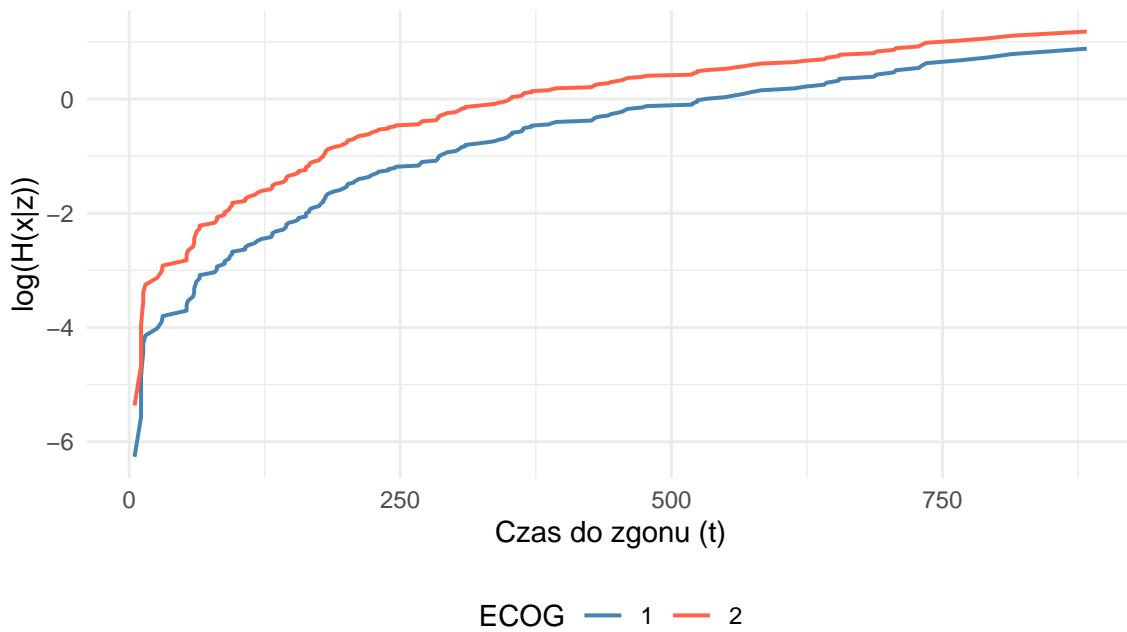
```



Wykres 11: Estymowana skumulowana funkcja hazardu dla odpowiednich charakterystyk 70-letnich kobiet - modele Coxa i proporcjonalnych szans

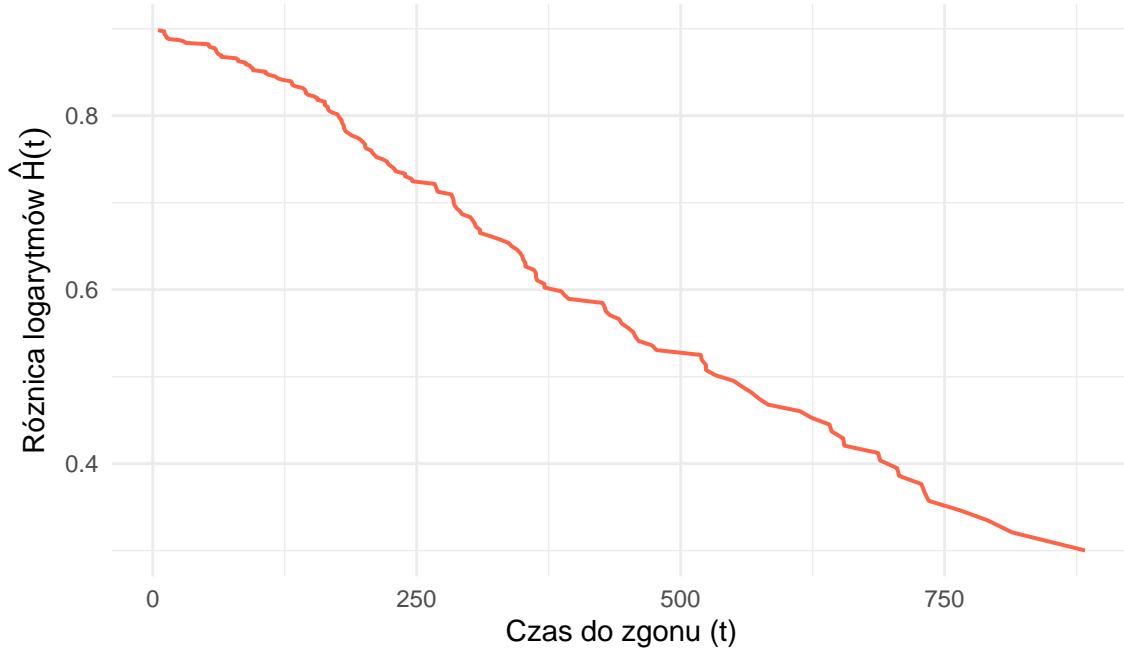
Wykres 11. pokazuje, że dla modelu proporcjonalnych szans estymowane wartości funkcji skumulowanego hazardu są mniejsze, niż te dla modelu PH Coxa. Dla kobiety o niższym stopniu ECOG = 1 różnica ta wygląda na stałą i niewielką. Jednakże przy ECOG = 2 różnica ta jest dużo bardziej wyraźna i rosnąca w czasie. Wyniki takie oznaczają, że w każdej chwili  $t$  model PO wskazuje mniejsze ryzyko zgonu niż model PH Coxa, tym samym szacowane prawdopodobieństwa przeżycia będą większe.

Ze względu na to, że nałożenie logarytmu zachowa monotoniczność oraz “ranking” grup, następny wykres pominie rysowanie funkcji związanych z modelem proporcjonalnych hazardów Coxa.



Wykres 12: Logarytm estymowanej skumulowanej funkcji hazardu dla odpowiednich charakterystyk 70-letnich kobiet - model proporcjonalnych szans

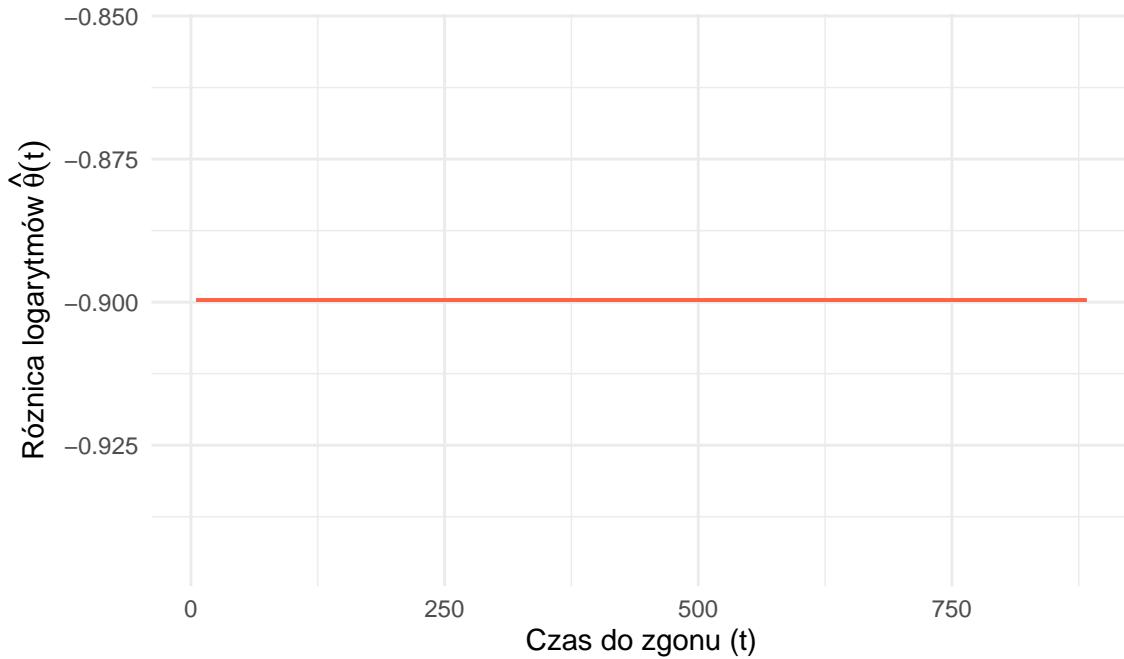
Z wykresu 12. można wyciągnąć przypuszczenia, że różnica logarytmów skumulowanych funkcji hazardu (odpowiednio: iloraz skumulowanych funkcji hazardu) nie jest stały w czasie i zależy od wartości  $t$ .



Wykres 13: Różnica logarytmów estymowanej skumulowanej funkcji hazardu dla odpowiednich charakterystyk 70-letnich kobiet - model proporcjonalnych szans

Wykres 13. potwierdza nasze poprzednie wnioski. Należy jednak pamiętać, że model proporcjonalnych szans, jak sama nazwa wskazuje, zachowuje stałą wartość ilorazu estymowanych funkcji szans, a nie skumulowanych

funkcji hazardu. Model ten stosuje się, gdy drugi przypadek nie zachodzi.



Wykres 14: Różnica logarytmów estymowanej szansy dla odpowiednich charakterystyk 70-letnich kobiet

Na podstawie wykresu 14. nie ma podstaw do wątpienia w poprawność przyjętego modelu proporcjonalnych szans. Widoczna różnica logarytmów funkcji szans pozostaje stała w całym badanym przedziale czasu, tym samym nie zależy od jego wartości.

#### 4.6 Zadanie 5

Celem zadania jest oszacować dla wcześniej wymienionych pacjentek prawdopodobieństwo, że ich czas życia będzie większy niż 300 dni, na podstawie modelu proporcjonalnych szas. Dodatkowo dokonano porównania otrzymanych wyników z tymi uzyskanymi metodą proporcjonalnych hazardów Coxa.

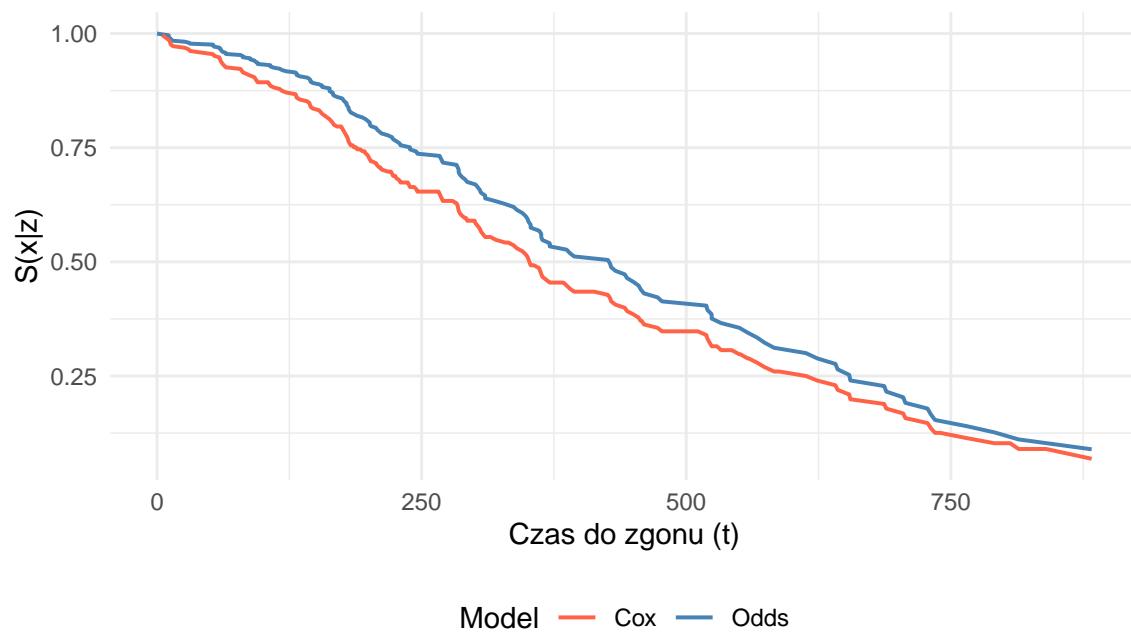
Tabela 7: Estymowana funkcja przeżycia dla czasu  $t = 300$  w procentach - porównanie modeli Coxa oraz proporcjonalnych szans

	Cox	Odds
Pacjentka 1	59.01	67.50
Pacjentka 2	35.98	45.79

Porównanie wyników z tabeli 7. prowadzi do wniosku, że estymowane wartości dla modelu PO są większe (korzystniejsze dla pacjentek) niż te dla modelu Coxa. Potwierdza to również wniosek wysnuty na podstawie wykresu 11. - mniejszy skumulowany hazard oznacza mniejsze ryzyko, a mniejsza różnica między danymi funkcjami daje również mniejszą różnicę w szacowanych prawdopodobieństwach.

#### 4.7 Zadanie 6

W zadaniu narysowano wykres estymowanej funkcji przeżycia dla pacjentki o stopniu skali ECOG = 1 dla modelu proporcjonalnych szans oraz dokonano graficznego porównania z funkcją uzyskaną za pomocą modelu PO Coxa.



Wykres 15: Porównanie estymowanych funkcji przeżycia - modele Coxa oraz proporcjonalnych szans

Wykres 15. pokazuje, że estymowane wartości prawdopodobieństwa są większe dla modelu PO niż Coxa o około 10%. Należy jednak zauważyć, że nie zmienia się ona gwałtownie wraz z czasem, lecz stopniowo.

Biorąc pod uwagę wyniki uzyskane dla wszystkich analizowanych podejść można przypuszczać, że duże rozbieżności dla modelu proporcjonalnych szans mogą wynikać ze względu na nieadekwatnego dopasowania do danych. Faktycznie, model ten stosuje się, gdy nie jest spełnione założenie proporcjonalności hazardów, które w przypadku naszych danych jednak zostało potwierdzone empirycznie.

## **5 Lista 12**

### **5.1 Zadanie 1**

TO DO

### **5.2 Zadanie 2**

TO DO

### **5.3 Zadanie 3**

TO DO

### **5.4 Zadanie 4**

TO DO

## 6 Bibliografia

- [1] *Skala Karnofsky'ego*, Wikipedia, [https://pl.wikipedia.org/wiki/Skala\\_Karnofskyego](https://pl.wikipedia.org/wiki/Skala_Karnofskyego), dostęp: 22.12.2025.
- [2] *Skala ECOG*, Wikipedia, [https://pl.wikipedia.org/wiki/Skala\\_ECOG](https://pl.wikipedia.org/wiki/Skala_ECOG), dostęp: 22.12.2025.