

Sprawozdanie 3

Mateusz Cieślak

2026-02-23

Spis treści

| | | |
|----------|---------------------|-----------|
| 1 | Lista 9 | 1 |
| 1.1 | Zadanie 1 | 1 |
| 1.2 | Zadanie 2 | 1 |
| 1.3 | Zadanie 3 | 2 |
| 1.4 | Zadanie 4 | 3 |
| 1.5 | Zadanie 5 | 4 |
| 1.6 | Zadanie 6 | 5 |
| 1.7 | Zadanie 7 | 5 |
| 1.8 | Zadanie 8 | 7 |
| 1.9 | Zadanie 9 | 8 |
| 2 | Lista 10 | 8 |
| 2.1 | Zadanie 1 | 8 |
| 2.2 | Zadanie 2 | 9 |
| 2.3 | Zadanie 3 | 9 |
| 2.4 | Zadanie 4 | 11 |
| 2.5 | Zadanie 5 | 13 |
| 2.6 | Zadanie 6 | 14 |
| 3 | Lista 11 | 14 |
| 3.1 | Zadanie 1 | 14 |
| 3.2 | Zadanie 2 | 15 |
| 3.3 | Zadanie 3 | 16 |
| 3.4 | Zadanie 4 | 17 |
| 3.5 | Zadanie 5 | 20 |
| 3.6 | Zadanie 6 | 21 |
| 4 | Lista 12 | 21 |
| 4.1 | Zadanie 1 | 21 |
| 4.1.1 | a) | 21 |
| 4.1.2 | b) | 22 |
| 4.1.3 | c) | 22 |
| 4.2 | Zadanie 2 | 22 |
| 4.2.1 | a) | 22 |
| 4.2.2 | b) | 23 |
| 4.2.3 | c) | 23 |
| 4.3 | Zadanie 3 | 23 |
| 4.3.1 | a) | 23 |
| 4.3.2 | b) | 24 |
| 4.3.3 | c) | 25 |

Spis rysunków

| | | |
|----|---|----|
| 1 | Krzywa funkcji przeżycia dla kobiety w wieku 70 lat, ph.ecog=1 i ph.karno=90 | 4 |
| 2 | Porównanie funkcji hazardu dla kobiet w wieku 70 lat, ph.karno=90 i dwóch różnych statystyk ph.ecog=1 i ph.ecog=2 (model proporcjonalnych hazardów) | 6 |
| 3 | Porównanie logarytmów funkcji hazardu dla kobiet w wieku 70 lat, ph.karno=90 i dwóch różnych statystyk ph.ecog=1 i ph.ecog=2 (model proporcjonalnych hazardów) | 7 |
| 4 | Porównanie funkcji przeżycia modelu proporcjonalnych hazardów i modelu przyspieszonego czasu awarii | 8 |
| 5 | Bazowa funkcja skumulowanego hazardu w modelu Coxa | 10 |
| 6 | Bazowa funkcja przeżycia w modelu Coxa | 11 |
| 7 | Porównanie skumulowanych funkcji hazardu dla kobiet w wieku 70 lat, ph.karno=90 i dwóch różnych statystyk ph.ecog=1 i ph.ecog=2 (model Coxa) | 12 |
| 8 | Porównanie logarytmów skumulowanych funkcji hazardu dla kobiet w wieku 70 lat, ph.karno=90 i dwóch różnych statystyk ph.ecog=1 i ph.ecog=2 (model Coxa) | 13 |
| 9 | Porównanie funkcji przeżycia modelu proporcjonalnych hazardów i modelu przyspieszonego czasu awarii do modelu proporcjonalnych hazardów Coxa | 14 |
| 10 | Bazowa funkcja skumulowanego hazardu w modelu proporcjonalnych szans | 16 |
| 11 | Bazowa funkcja przeżycia w modelu proporcjonalnych szans | 17 |
| 12 | Porównanie skumulowanych funkcji hazardu dla kobiet w wieku 70 lat, ph.karno=90 i dwóch różnych statystyk ph.ecog=1 i ph.ecog=2 (model Proporcjonalnych Szans) | 18 |
| 13 | Porównanie logarytmów skumulowanych funkcji hazardu dla kobiet w wieku 70 lat, ph.karno=90 i dwóch różnych statystyk ph.ecog=1 i ph.ecog=2 (model Proporcjonalnych Szans) | 19 |
| 14 | Porównanie hazardów modelu Coxa i modelu Proporcjonalnych Szans | 20 |
| 15 | Porównanie funkcji przeżycia modelu Coxa i modelu Proporcjonalnych Szans dla kobiety (a) . | 21 |

Spis tabel

| | | |
|---|---|----|
| 1 | Oszacowania parametrów modelu przyspieszonego czasu przeżycia (AFT) | 2 |
| 2 | Parametry modelu proporcjonalnych hazardów (PH) | 4 |
| 3 | Szacowane prawdopodobieństwo przeżycia powyżej 300 dni | 7 |
| 4 | Oszacowania parametrów modelu proporcjonalnych hazardów Coxa | 9 |
| 5 | Oszacowania parametrów modelu proporcjonalnych szans (PO) | 15 |

Wykorzystane biblioteki:

```
library(timereg)
library(ggplot2)
library(ggsurvfit)
library(survival)
library(patchwork)
library(kableExtra)
library(dplyr)
library(tidyr)
library(eha)
library(knitr)
library(broom)
```

1 Lista 9

1.1 Zadanie 1

Wczytujemy dane `lung` z pakietu `survival`. Wybieramy tylko kolumny: `time`, `age`, `sex`, `ph.ecog`, `ph.karno`. Następnie usuwamy 2 wiersze z brakującymi danymi.

Centrujemy zmienne ciągłe (wiek oraz wynik w skali Karnofsky’ego) względem ich wartości średnich, aby umożliwić bezpośrednią interpretację wyrazu wolnego modelu jako logarytmu czasu przeżycia typowego pacjenta.

```
lung.subset.complete$age_c <- scale(lung.subset.complete$age, center = TRUE, scale = FALSE)
lung.subset.complete$ph.karno_c <- scale(lung.subset.complete$ph.karno, center = TRUE, scale = FALSE)
```

Używamy funkcji `survreg` do oszacowania parametrów modelu przyspieszonego czasu awarii.

Tabela 1: Oszacowania parametrów modelu przyspieszonego czasu przeżycia (AFT)

| | β | e^β (TR) | Błąd std. | z | p-value |
|-------------------------|---------|----------------|-----------|---------|---------|
| Intercept | 6.3012 | 545.2152 | 0.1476 | 42.6968 | 0.0000 |
| Płeć (Kobieta) | 0.4083 | 1.5042 | 0.1230 | 3.3185 | 0.0009 |
| ECOG 1 | -0.4216 | 0.6560 | 0.1676 | -2.5149 | 0.0119 |
| ECOG 2 | -0.9261 | 0.3961 | 0.2514 | -3.6835 | 0.0002 |
| ECOG 3 | -1.6872 | 0.1850 | 0.7673 | -2.1990 | 0.0279 |
| Wiek (scentrowany) | -0.0086 | 0.9914 | 0.0068 | -1.2666 | 0.2053 |
| Karnofsky (scentrowany) | -0.0101 | 0.9900 | 0.0068 | -1.4836 | 0.1379 |
| Log(scale) | -0.3276 | 0.7207 | 0.0623 | -5.2613 | 0.0000 |

1.2 Zadanie 2

Interpretacja współczynników modelu. Wnioskowanie przeprowadzono na poziomie istotności $\alpha = 0.05$.

W modelu przyspieszonego czasu awarii współczynniki $\exp(\beta)$ (TR – Time Ratio) interpretowane są jako ilorazy czasów przeżycia.

Ogólna zasada interpretacji:

- TR = 1: Zmienna nie ma wpływu na czas przeżycia.
- TR > 1: Zmienna wydłuża oczekiwany czas przeżycia.
- TR < 1: Zmienna skraca czas przeżycia (przyspiesza zgon).

Szczegółowa interpretacja:

Intercept:

Wyraz wolny intercept odpowiada logarytmowi oczekiwanego czasu przeżycia mężczyzny o średnim wieku (62.5), o średnim stanie Karnofsky’ego (82), w stanie ECOG = 0 i wynosi ok. 545 dni. Ten wynik jest istotny statystycznie ($p < 0.001$).

Płeć:

Kobiety mają oczekiwany czas przeżycia około 1.5 raza dłuższy niż mężczyźni. Zależność ta jest statystycznie istotna ($p = 0.0009$).

Skala sprawności ECOG:

W porównaniu do pacjentów z ECOG = 0, pacjenci z ECOG = 1, 2 i 3 charakteryzują się odpowiednio krótszym czasem przeżycia o ok. 34%, 60% i 81.5%. P-wartości dla każdego poziomu ECOG są statystycznie istotne.

Wiek:

Wzrost scentrowanego wieku o 1 rok skraca oczekiwany czas przeżycia o ok. 0.9%. Jednakże wartość $p = 0.2053$ sugeruje, że wpływ tej zmiennej nie jest statystycznie istotny.

Skala Karnofsky'ego (scentrowana):

Wzrost scentrowanego wskaźnika Karnofsky'ego o 1 punkt skraca go o ok. 1%. Ta zmienna również nie jest statystycznie istotna ($p = 0.1379$).

Log(scale):

Oszacowany parametr `Log(scale)` wynosi -0.3276 i jest istotny statystycznie ($p < 0.001$). Odpowiada on wartości parametru skali $\sigma = \exp(-0.3276) = 0.72$, a ponieważ jest on mniejszy od 1, oznacza to, że funkcja hazardu dla analizowanego rozkładu Weibulla jest rosnąca w czasie.

1.3 Zadanie 3

W niniejszym zadaniu wyznaczymy oszacowanie funkcji przeżycia (w dniach) odpowiadającej rozkładowi czasu życia kobiety w wieku 70 lat o charakterystyce `ph.ecog=1` i `ph.karno=90` i na podstawie uzyskanego oszacowania obliczymy szacowane prawdopodobieństwo, że czas życia tej kobiety będzie większy niż 300 dni.

Do obliczeń wykorzystujemy parametry modelu przyspieszonego czasu awarii (AFT), przekształcając je na parametryzację rozkładu Weibulla. Funkcja przeżycia przyjmuje postać:

$$S(t|z) = \exp\left(-\lambda_0 \cdot e^{\beta_{PH}^T z} \cdot t^\alpha\right)$$

$$\alpha = \frac{1}{\sigma}, \quad \lambda_0 = \exp\left(-\frac{\mu}{\sigma}\right), \quad \beta_{PH} = -\frac{\beta_{AFT}}{\sigma}$$

Współczynniki μ , σ oraz β_{AFT} pochodzą z oszacowanego modelu z funkcji `survreg`.

```
sigma <- fit_a$scale
alfa0 <- 1 / sigma

mu_intercept <- fit_a$coef[1]

lambda0 <- exp(-mu_intercept / sigma)

bety_aft <- fit_a$coefficients[-1]
bety_ph <- -bety_aft / sigma

mean_age <- mean(lung.subset.complete$age)
mean_karno <- mean(lung.subset.complete$ph.karno)

z_pacjent <- c(
  1,          # sex2
  1,          # ph.ecog1
  0,          # ph.ecog2
  0,          # ph.ecog3
  70 - mean_age, # age_c
  90 - mean_karno # ph.karno_c
)
```

```

beta_T_z <- sum(bety_ph * z_pacjent)

S_t_ph <- function(t) {
  exp( -lambda0 * exp(beta_T_z) * t^alfa0 )
}

wynik <- S_t_ph(300)

```

Obliczone prawdopodobieństwo:

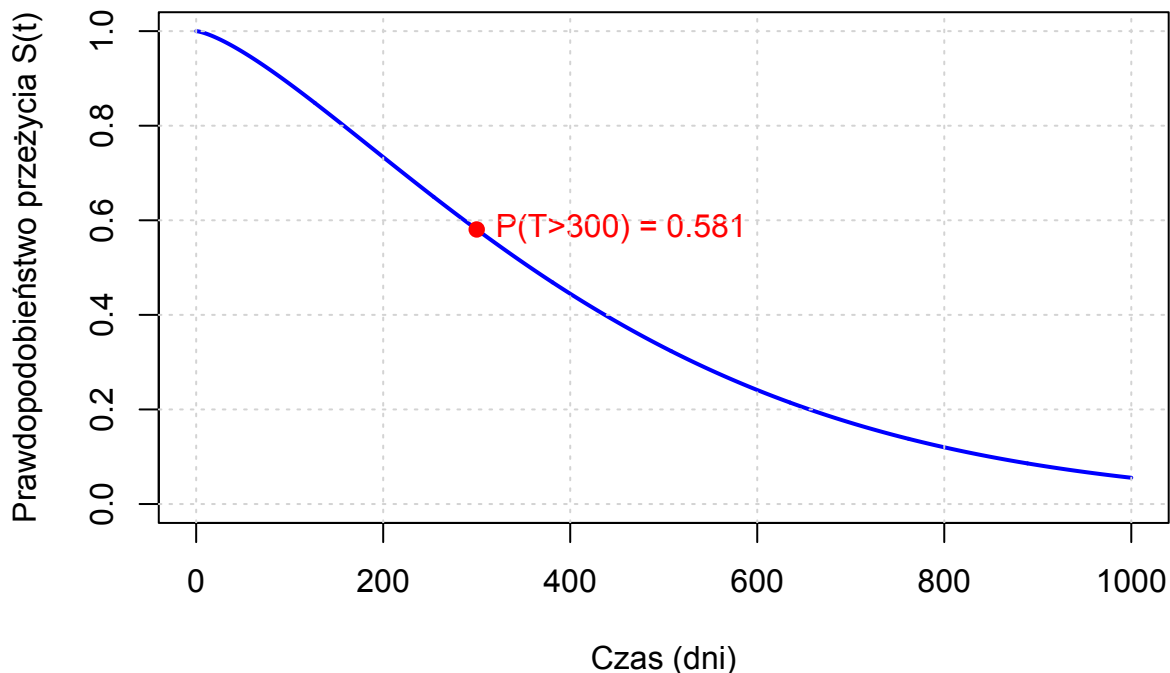
$$\hat{P}(T > 300) = S(300|z)$$

Prawdopodobieństwo wynosi $\hat{P}(T > 300)$: 0.581.

1.4 Zadanie 4

Przjrzyjmy się wykresowi oszacowanej w poprzednim zadaniu funkcji przeżycia.

Krzywa przeżycia (Kobieta, 70 lat, ECOG=1, Karno=90)



Rysunek 1: Krzywa funkcji przeżycia dla kobiety w wieku 70 lat, ph.ecog=1 i ph.karno=90

Na wykresie 1 wartość prawdopodobieństwa, że czas życia tej kobiety będzie większy niż 300 dni zaznaczono czerwoną kropką.

1.5 Zadanie 5

W tym zadaniu oszacujemy parametry modelu proporcjonalnych hazardów, przyjmując za zmienną zależną zmienną `time`, a za charakterystyki zmienne: `age`, `sex`, `ph.ecog`, `ph.karno`.

Tabela 2: Parametry modelu proporcjonalnych hazardów (PH)

| | β | e^β (Rel. Risk) | Błąd std. | z | p-value |
|---------------------|---------|-----------------------|-----------|---------|---------|
| Wiek (age) | 0.0119 | 1.0120 | 0.0094 | 1.2680 | 0.2048 |
| Płeć (Kobieta) | -0.5665 | 0.5675 | 0.1693 | -3.3460 | 0.0008 |
| ECOG 1 | 0.5850 | 1.7949 | 0.2343 | 2.4967 | 0.0125 |
| ECOG 2 | 1.2851 | 3.6150 | 0.3515 | 3.6557 | 0.0003 |
| ECOG 3 | 2.3413 | 10.3944 | 1.0758 | 2.1762 | 0.0295 |
| Skala Karnofsky’ego | 0.0140 | 1.0141 | 0.0095 | 1.4746 | 0.1403 |

Używamy funkcji `phreg` do oszacowania parametrów modelu proporcjonalnych hazardów.

1.6 Zadanie 6

Wnioskowanie o istotności statystycznej parametrów modelu przeprowadzono przy założeniu poziomu istotności $\alpha = 0.05$.

Ogólna zasada interpretacji Rel.Risk:

- Rel.Risk = 1: Brak wpływu zmiennej na ryzyko zgonu.
- Rel.Risk > 1: Czynniki zwiększa ryzyko śmierci (skraca czas przeżycia).
- Rel.Risk < 1: Czynniki zmniejsza ryzyko śmierci (wydłuża czas przeżycia, działa ochronnie).

Szczegółowa interpretacja współczynników:

Wiek:

Rel.Risk = 1.012 co oznacza, że każdy kolejny rok życia zwiększa ryzyko zgonu o 1,2%. Jednakże wartość p-value wynosi 0.2048, co oznacza, że w tym modelu wpływ wieku nie jest istotny statystycznie.

Płeć:

Poziomem odniesienia jest 1 (Mężczyźni). Dla 2 (Kobiety): Rel.Risk = 0.5675.

Interpretacja: U kobiet ryzyko zgonu stanowi 56,75% ryzyka u mężczyzn.

Wynik jest istotny statystycznie ($p = 0.0008$), przy założonym przez nas $\alpha = 0.05$.

Skala sprawności ECOG:

Poziomem odniesienia jest 0 (pacjent w pełni sprawny).

Im wyższy status ECOG, tym gorszy stan pacjenta i drastycznie wyższe ryzyko:

- ECOG 1: Ryzyko jest o 79,5% wyższe niż u pacjentów z ECOG 0.
- ECOG 2: Ryzyko jest 3,6 raza wyższe niż u pacjentów z ECOG 0.
- ECOG 3: Ryzyko jest ponad 10 razy wyższe niż u pacjentów z ECOG 0.

Dla każdego poziomu ECOG p-wartość jest poniżej założonego przez nas $\alpha = 0.05$.

Skala Karnofsky’ego:

Rel.Risk = 1.014

Interpretacja wynikająca z tabeli: Wzrost wskaźnika Karnofsky'ego o 1 punkt zwiększa ryzyko o 1,4%.

Jest to wynik sprzeczny z intuicją medyczną (wyższy Karnofsky oznacza lepsze zdrowie, więc ryzyko powinno maleć, a współczynnik powinien być < 1). Jednakże wartość p wynosi 0.1403, co oznacza, że wynik ten nie jest istotny statystycznie.

1.7 Zadanie 7

Wyznamy oszacowanie funkcji hazardu odpowiadającej rozkładowi czasu życia

- kobiety w wieku 70 lat o charakterystyce `ph.ecog=1` i `ph.karno=90`
- kobiety w wieku 70 lat o charakterystyce `ph.ecog=2` i `ph.karno=90`

Następnie przyjrzymy się wykresom tych funkcji i wykresom logarytmów tych funkcji.

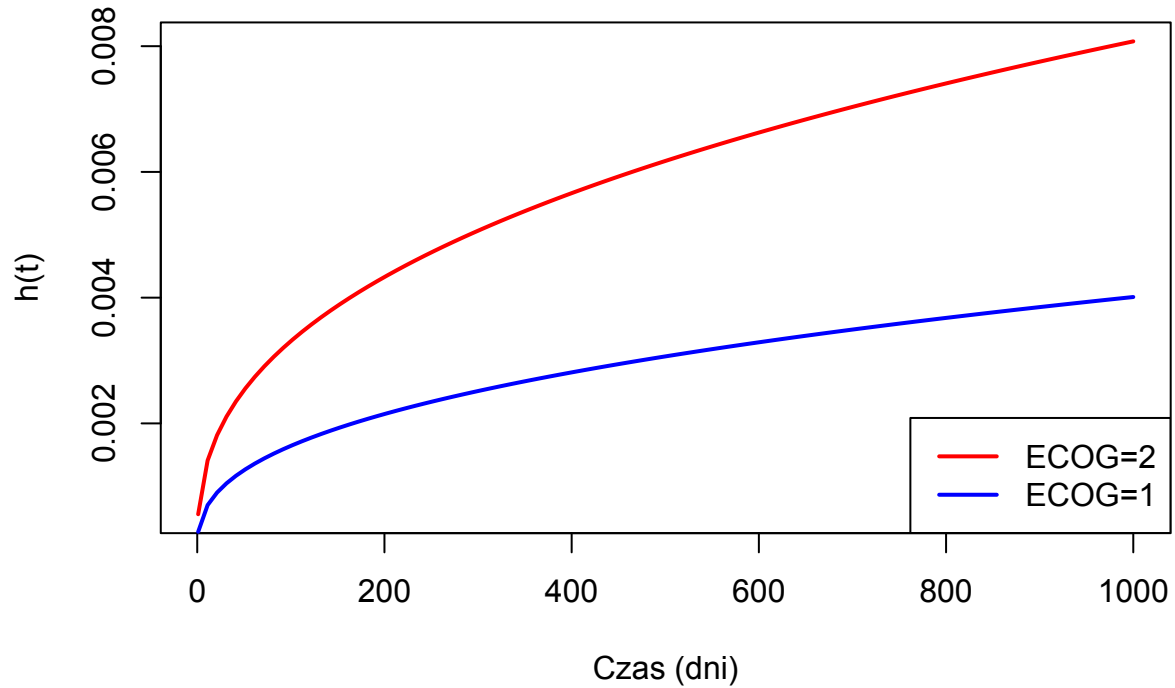
```
h0 <- function(t) alpha * lambda * t ^ (alpha - 1)
h <- function(t,z) h0(t) * exp(sum(beta * z))

mean_age <- mean(lung.subset.complete$age)
mean_karno <- mean(lung.subset.complete$ph.karno)

z_a <- c(
  70 - mean_age,    # age_c
  1,                # as.factor(sex)2 (Kobieta)
  1,                # as.factor(ph.ecog)1
  0,                # as.factor(ph.ecog)2
  0,                # as.factor(ph.ecog)3
  90 - mean_karno  # ph.karno_c
)

z_b <- c(
  70 - mean_age,    # age_c
  1,                # as.factor(sex)2
  0,                # as.factor(ph.ecog)1
  1,                # as.factor(ph.ecog)2
  0,                # as.factor(ph.ecog)3
  90 - mean_karno  # ph.karno_c
)
```

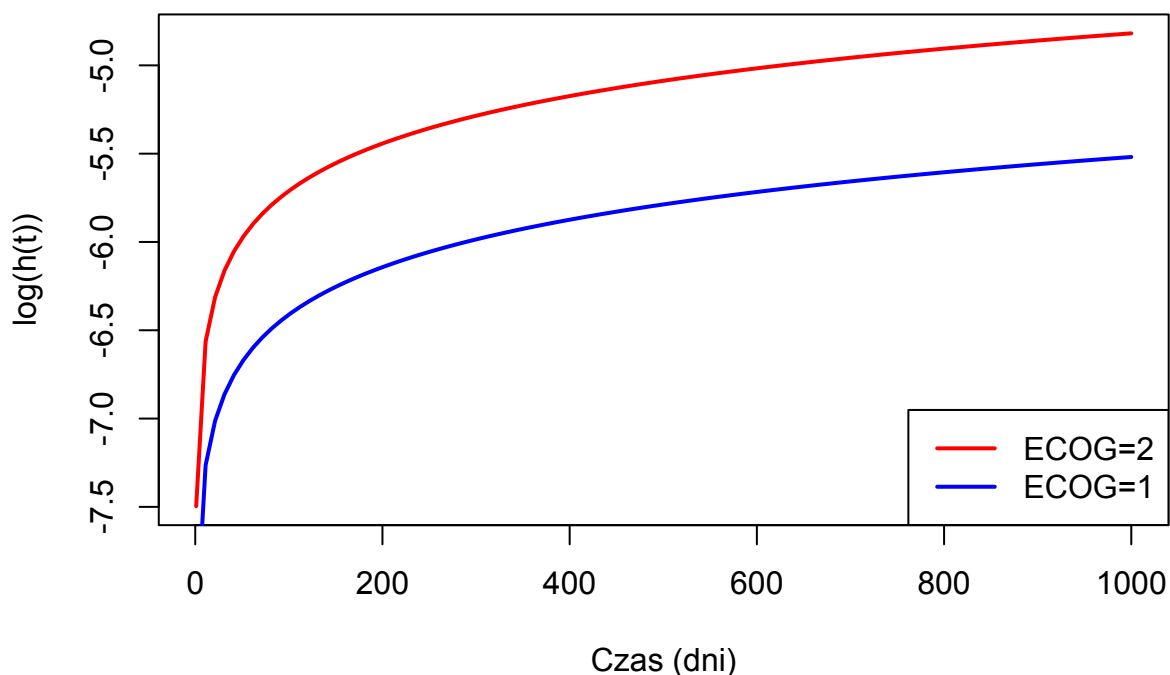
Funkcje hazardu



Rysunek 2: Porównanie funkcji hazardu dla kobiet w wieku 70 lat, $ph.karno=90$ i dwóch różnych statystyk $ph.ecog=1$ i $ph.ecog=2$ (model proporcjonalnych hazardów)

Funkcje hazardu na wykresie 2 przedstawiają, że dla profilu pacjentki z $ph.ecog=2$ (linia czerwona) funkcja hazardu przyjmuje wyższe wartości w całym badanym okresie niż dla profilu z $ph.ecog=1$ (linia niebieska). Obie funkcje są rosnące, co jest charakterystyczne dla rozkładu Weibulla z parametrem kształtu $\alpha > 1$.

Logarytm funkcji hazardu



Rysunek 3: Porównanie logarytmów funkcji hazardu dla kobiet w wieku 70 lat, $ph.karno=90$ i dwóch różnych statystyk $ph.ecog=1$ i $ph.ecog=2$ (model proporcjonalnych hazardów)

Nie mamy podstaw do tego żeby mieć wątpliwości co do przyjętego modelu hazardów. Zgodnie z oczekiwaniami krzywe na wykresie 3 są do siebie równoległe, co wskazuje, że założenie o stałym w czasie ilorazie hazardu jest spełnione.

1.8 Zadanie 8

W niniejszym zadaniu wyznaczmy oszacowanie funkcji przeżycia (w dniach) odpowiadającej rozkładowi czasu życia

- (a) kobiety w wieku 70 lat o charakterystyce $ph.ecog=1$ i $ph.karno=90$,
- (b) kobiety w wieku 70 lat o charakterystyce $ph.ecog=2$ i $ph.karno=90$

i na podstawie tego oszacowania obliczymy szacowane prawdopodobieństwo, że czas życia kobiet o powyżej podanych charakterystykach będzie większy niż 300 dni.

```
S0 <- function(t) exp(-lambda * t^alpha)
S <- function(t, z) S0(t)^exp(sum(beta*z))
```

```
p_300_a <- S(300, z_a)
p_300_b <- S(300, z_b)
```

Tabela 3: Szacowane prawdopodobieństwo przeżycia powyżej 300 dni

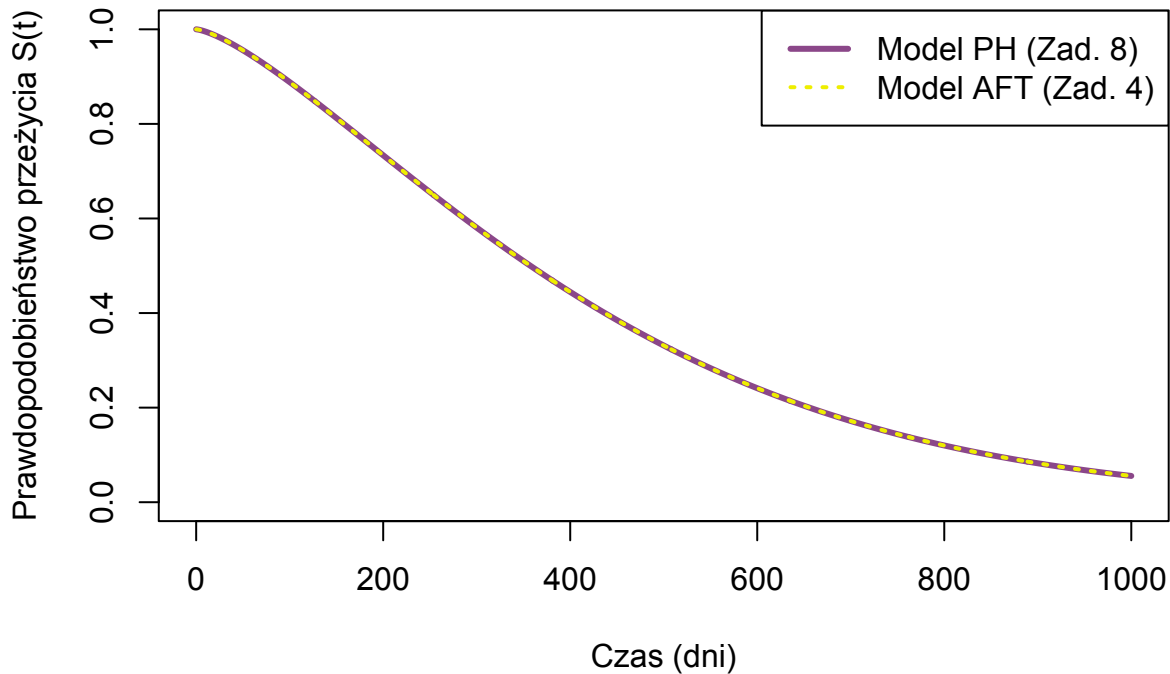
| Profil pacjentki | $\hat{P}(T > 300)$ |
|------------------|--------------------|
| Kobieta (ECOG=1) | 0.5806 |
| Kobieta (ECOG=2) | 0.3345 |

Wynik z zadania 3 (0.5806) i zadania 8 a) (0.5806) są identyczne. Jest to zgodne z oczekiwaniami, ponieważ dla rozkładu Weibulla parametryzacja PH oraz parametryzacja AFT są matematycznie równoważne. Obie metody prowadzą do wyznaczenia tej samej funkcji przeżycia dla konkretnego profilu pacjenta.

1.9 Zadanie 9

W tym zadaniu porównamy wykresy oszacowanej funkcji przeżycia z zadania 8 a) i 4.

Porównanie krzywych przeżycia (Zadanie 4 vs Zadanie 8)



Rysunek 4: Porównanie funkcji przeżycia modelu proporcjonalnych hazardów i modelu przyspieszonego czasu awarii

Zgodnie z oczekiwaniami, na wykresie 4, krzywe funkcji przeżycia nakładają się. Wynika to z faktu, że dla rozkładu Weibulla model AFT i model PH są matematycznie tożsame, a estymacja parametrów oboma metodami (`survreg` oraz `phreg`) prowadzi do identycznej postaci oszacowanej funkcji przeżycia $S(t | z)$ dla tego samego profilu pacjenta.

2 Lista 10

2.1 Zadanie 1

W niniejszym zadaniu oszacujemy parametry modelu proporcjonalnych hazardów Coxa, przyjmując za zmienną zależną zmienną `time`, a za charakterystyki zmienne: `age`, `sex`, `ph.ecog`, `ph.karno`.

Używamy funkcji `coxph` do oszacowania parametrów modelu proporcjonalnych hazardów Coxa.

Tabela 4: Oszacowania parametrów modelu proporcjonalnych hazardów Coxa

| Zmienna | β | e^β (HR) | Błąd std. | p-value |
|--------------------------------|---------|----------------|-----------|---------|
| Wiek (scentrowany) | 0.0126 | 1.0126 | 0.0095 | 0.1839 |
| Płeć (Kobieta) | -0.5657 | 0.5680 | 0.1697 | 0.0009 |
| ECOG 1 | 0.5781 | 1.7826 | 0.2364 | 0.0145 |
| ECOG 2 | 1.2399 | 3.4553 | 0.3553 | 0.0005 |
| ECOG 3 | 2.3959 | 10.9776 | 1.0812 | 0.0267 |
| Karnofsky (scentrowany) | 0.0124 | 1.0125 | 0.0096 | 0.1951 |

2.2 Zadanie 2

Wnioskowanie przeprowadzono na poziomie istotności $\alpha = 0.05$.

Ogólna zasada:

- HR = 1: Brak wpływu zmiennej na ryzyko.
- HR > 1: Zmienna zwiększa hazard (ryzyko zgonu), co skraca czas przeżycia.
- HR < 1: Zmienna zmniejsza hazard (działa ochronnie), co wydłuża czas przeżycia.

Szczegółowa interpretacja:

Wiek (scentrowany):

HR=1.0126. Każdy dodatkowy rok życia zwiększa ryzyko zgonu o około 1.26%. Wartość p=0.1839 sugeruje, że wpływ wieku w tym modelu nie jest istotny statystycznie.

Płeć (kobieta):

HR=0.5680. Interpretacja: Ryzyko zgonu u kobiet jest o ponad 43% niższe niż u mężczyzn (przy założeniu takich samych pozostałych cech). Jest to wynik istotny statystycznie (p=0.0009).

Skala sprawności ECOG: Referencyjnym poziomem jest ECOG 0 (pacjent w pełni sprawny). Każdy kolejny stopień zwiększa ryzyko:

- ECOG 1: Ryzyko zgonu jest o 78% wyższe niż w grupie ECOG 0 (p=0.0145).
- ECOG 2: Ryzyko zgonu jest ok. 3.46 razy wyższe niż w grupie ECOG 0 (p=0.0005).
- ECOG 3: Ryzyko zgonu jest prawie 11-krotnie wyższe niż w grupie ECOG 0 (p=0.0267). Wszystkie poziomy sprawności ECOG są istotnymi predyktorami ryzyka.

Skala Karnofsky'ego (scentrowana):

HR=1.0125. Wartość p=0.195 wskazuje, że zmienna ta nie jest istotna statystycznie. Podobnie jak w modelu Weibulla (Lista 9), kierunek wpływu (HR>1) jest sprzeczny z intuicją medyczną, ale brak istotności pozwala nam pominąć ten fakt.

2.3 Zadanie 3

Teraz wyznaczmy oszacowanie bazowej skumulowanej funkcji hazardu i bazowej funkcji przeżycia odpowiadającej rozkładowi czasu życia.

W modelu Coxa bazowa skumulowana funkcja hazardu jest szacowana metodą Breslowa

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

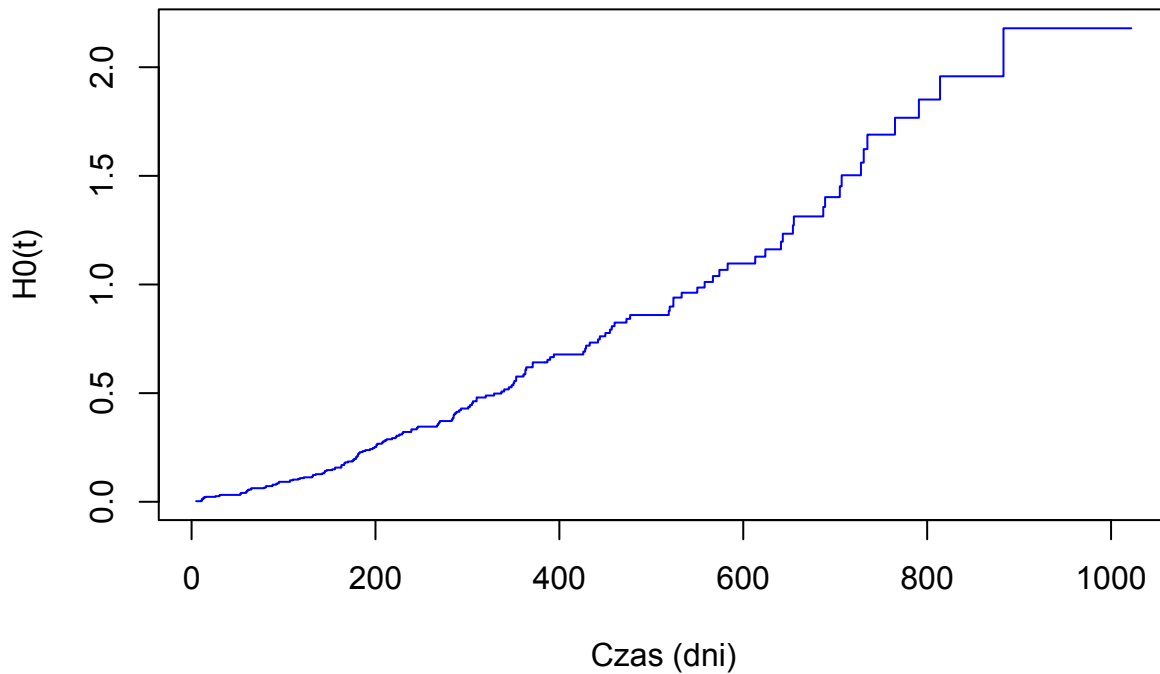
Natomiast bazowa funkcja przeżycia jest z nią związana zależnością:

$$\hat{S}_0(t) = \exp[-\hat{H}_0(t)]$$

```
bh_data <- basehaz(fit_cox, centered = FALSE)
```

```
bh_data$survival <- exp(-bh_data$hazard)
```

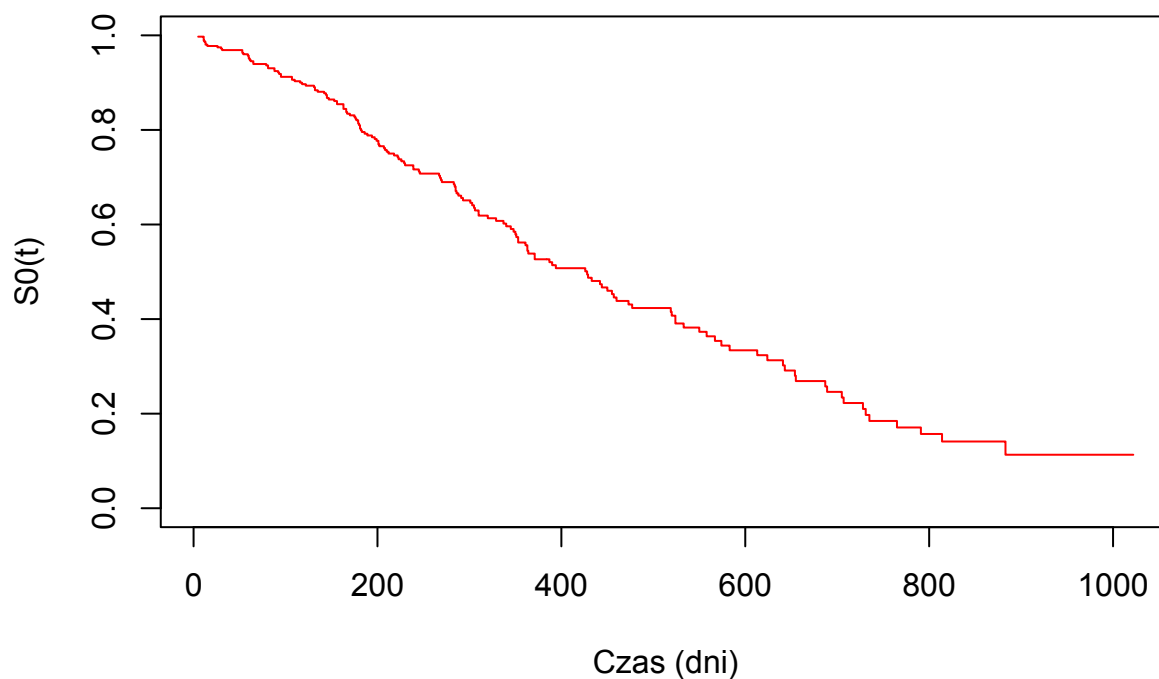
Bazowy skumulowany hazard



Rysunek 5: Bazowa funkcja skumulowanego hazardu w modelu Coxa

Wykres bazowej funkcji skumulowanego hazardu pokazuje, że łączne ryzyko zgonu wzrasta wraz z upływem czasu obserwacji.

Bazowa funkcja przeżycia



Rysunek 6: Bazowa funkcja przeżycia w modelu Coxa

Bazowa funkcja przeżycia przedstawia szacowane prawdopodobieństwo przeżycia dla pacjenta referencyjnego. Na wykresie widać, że po około 400 dniach prawdopodobieństwo przeżycia dla takiego profilu spada poniżej 0.5.

2.4 Zadanie 4

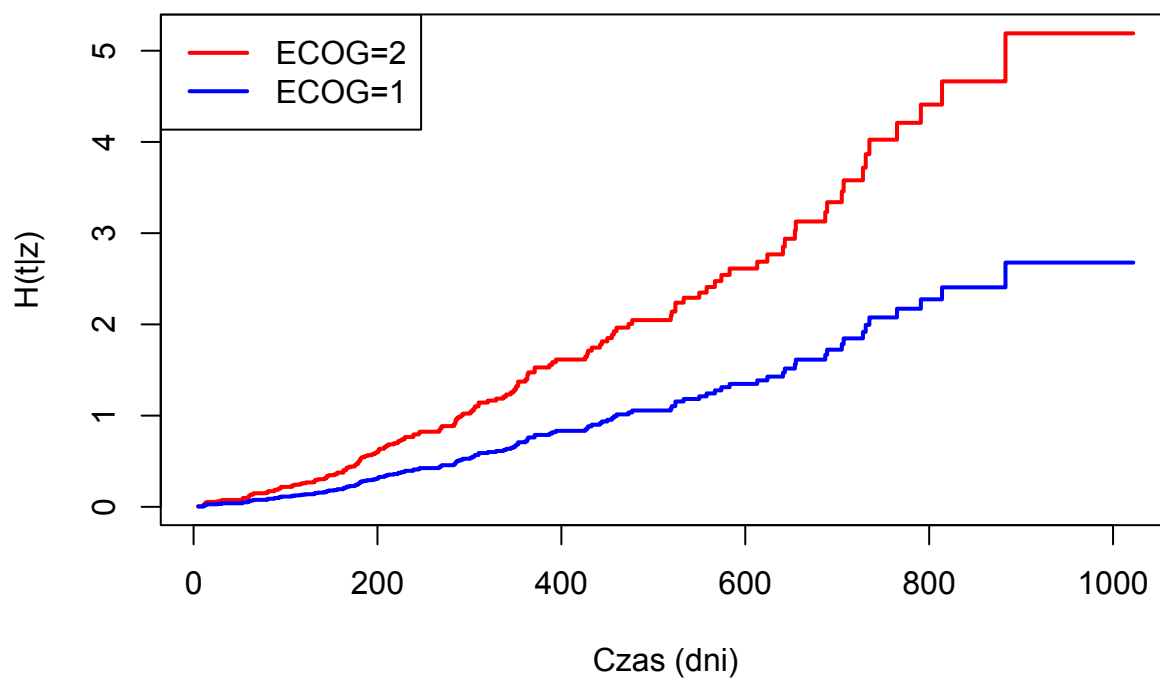
W tym zadaniu wykorzystamy oszacowany model Coxa, aby wyznaczyć funkcje hazardu dla konkretnych profili pacjentów.

W modelu Coxa skumulowany hazard dla pacjenta o cechach z dany jest wzorem:

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

```
H_a <- bh_data$hazard * exp(sum(beta * z_a))  
H_b <- bh_data$hazard * exp(sum(beta * z_b))
```

Skumulowany hazard $H(t|z)$

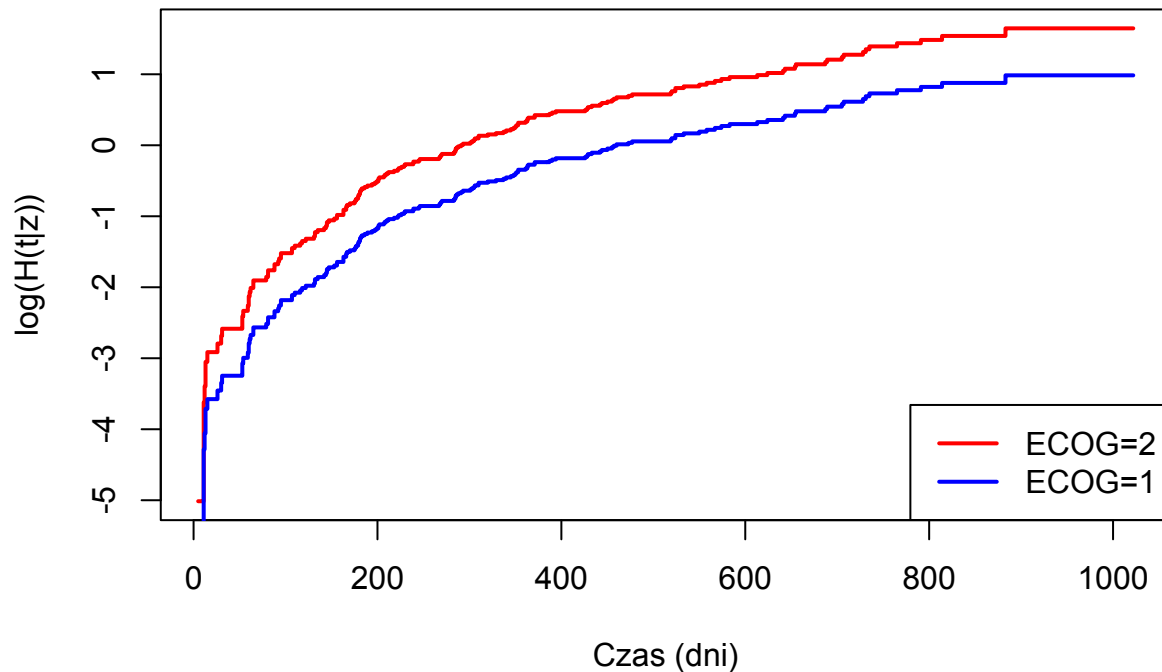


Rysunek 7: Porównanie skumulowanych funkcji hazardu dla kobiet w wieku 70 lat, $ph.karno=90$ i dwóch różnych statystyk $ph.ecog=1$ i $ph.ecog=2$ (model Coxa)

Logarytmując obustronnie, otrzymujemy:

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

Logarytm skumulowanego hazardu



Rysunek 8: Porównanie logarytmów skumulowanych funkcji hazardu dla kobiet w wieku 70 lat, ph.karno=90 i dwóch różnych statystyk ph.ecog=1 i ph.ecog=2 (model Coxa)

Na podstawie powyższych wykresów nie mamy wątpliwości co do przyjętego modelu, ponieważ krzywe na wykresie logarytmicznym są równoległe.

2.5 Zadanie 5

W tym zadaniu wyznaczmy oszacowanie funkcji przeżycia (w dniach) odpowiadającej rozkładowi czasu życia

- (a) kobiety w wieku 70 lat o charakterystyce ph.ecog=1 i ph.karno=90
- (b) kobiety w wieku 70 lat o charakterystyce ph.ecog=2 i ph.karno=90

Aby wyznaczyć funkcję przeżycia dla konkretnego pacjenta w modelu Coxa, korzystamy z relacji:

$$S(t | z) = [S_0(t)] \exp(\beta^T z)$$

gdzie:

- $S_0(t)$ to bazowa funkcja przeżycia (którą wyznaczyliśmy w Zadaniu 3).
- $\exp(\beta^T z)$ to współczynnik ryzyka dla danego pacjenta.

```
S_cox_a <- exp(-H_a)
S_cox_b <- exp(-H_b)

idx_300 <- max(which(bh_data$time <= 300))

prob_a <- S_cox_a[idx_300]
prob_b <- S_cox_b[idx_300]
```

Prawdopodobieństwo $P(T > 300)$

- (a) Kobieta (ECOG=1): 0.5901
- (b) Kobieta (ECOG=2): 0.3598

Porównanie dla kobiety z podpunktu (a)

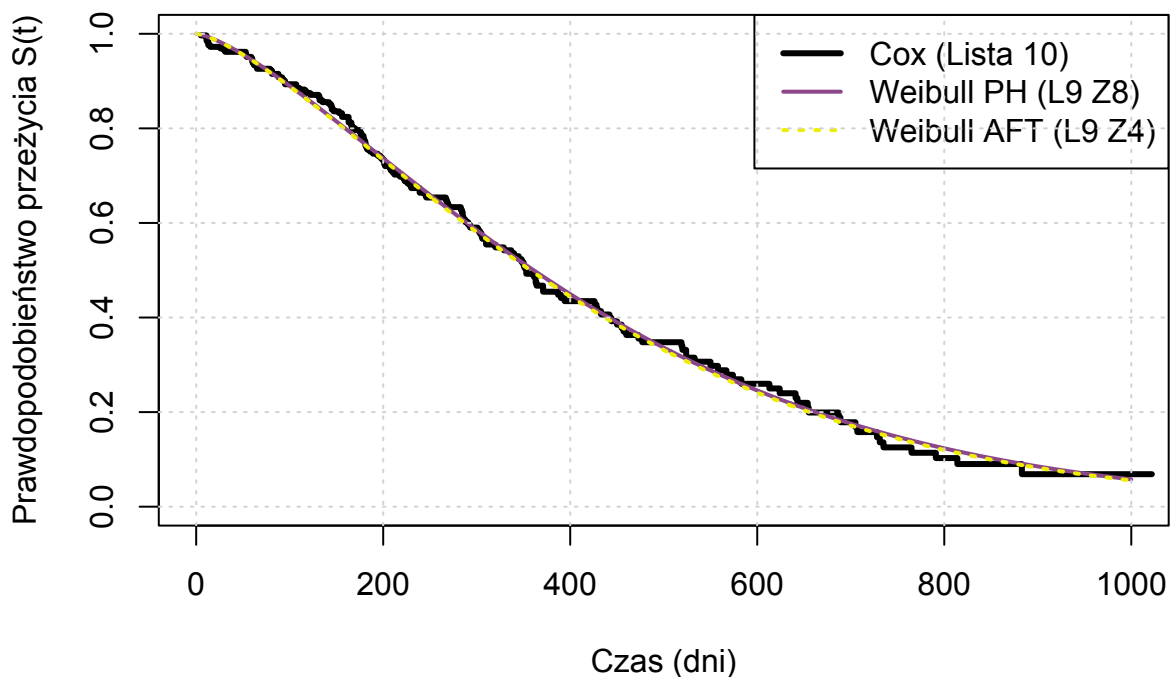
- Model Weibulla AFT (Lista 9, Zad 3): 0.5806
- Model Weibulla PH (Lista 9, Zad 8): 0.5806
- Model Coxa (Lista 10, Zad 5): 0.5901

Wynik uzyskany za pomocą semiparametrycznego modelu Coxa jest bardzo zbliżony, co świadczy o dobrej kalibracji modeli parametrycznych i poprawności przyjętych założeń dotyczących rozkładu czasu życia.

2.6 Zadanie 6

W tym zadaniu porównamy krzywe przeżycia uzyskane za pomocą modelu proporcjonalnych hazardów i modelu przyspieszonego czasu awarii do krzywej z modelu proporcjonalnych hazardów Coxa.

Porównanie: Cox vs Weibull (PH i AFT)



Rysunek 9: Porównanie funkcji przeżycia modelu proporcjonalnych hazardów i modelu przyspieszonego czasu awarii do modelu proporcjonalnych hazardów Coxa

Zgodnie z oczekiwaniami wykresy są bardzo zbliżone.

3 Lista 11

3.1 Zadanie 1

W niniejszym zadaniu oszacujemy parametry modelu Proporcjonalnych Szans, przyjmując za zmienną zależną zmienną `time`, a za charakterystyki zmienne: `age`, `sex`, `ph.ecog`, `ph.karno`.

Używamy funkcji `prop.odds` do oszacowania parametrów modelu Proporcjonalnych Szans. Ten model jest estymowany numerycznie, zatem dla stabilności wyników ustawiamy liczbę powtórzeń symulacji na 1000.

```
set.seed(123)

lung.subset.complete$status_bin <- lung.subset.complete$status - 1

fit_po <- prop.odds(Event(time, status_bin) ~ age_c + as.factor(sex) +
                    as.factor(ph.ecog) + ph.karno_c,
                    data = lung.subset.complete,
                    n.sim = 1000, profile = 1)
```

Tabela 5: Oszacowania parametrów modelu proporcjonalnych szans (PO)

| Zmienna | β | e^β (OR) |
|-------------------------|---------|----------------|
| Wiek (scentrowany) | 0.0132 | 1.0133 |
| Płeć (Kobieta) | -0.9535 | 0.3854 |
| ECOG 1 | 0.5470 | 1.7280 |
| ECOG 2 | 1.4466 | 4.2487 |
| ECOG 3 | 1.9245 | 6.8518 |
| Karnofsky (scentrowany) | -0.0037 | 0.9963 |

3.2 Zadanie 2

Wnioskowanie oparto na parametrach modelu proporcjonalnych szans. Skrót OR oznacza Odds Ratio, czyli Iloraz Szans.

Ogólna zasada:

- OR = 1: Brak wpływu zmiennej na szansę zgonu.
- OR > 1: Zmienna zwiększa szansę na wystąpienie zgonu (pogarsza rokowania).
- OR < 1: Zmienna zmniejsza szansę na wystąpienie zgonu (poprawia rokowania).

Szczegółowa interpretacja:

Wiek (scentrowany): OR = 1.0132. Każdy dodatkowy rok życia względem średniej zwiększa szansę na wystąpienie zgonu o około 1.3%.

Płeć (kobieta): OR = 0.3854. Szansa na zgon u kobiet jest o ponad 61% niższa niż u mężczyzn (przy założeniu takich samych pozostałych cech), co wskazuje na silny efekt ochronny.

Skala sprawności ECOG: Referencyjnym poziomem jest ECOG 0 (pacjent w pełni sprawny):

- ECOG 1: Szansa na zgon jest o 72.8% wyższa niż w grupie ECOG 0.
- ECOG 2: Szansa na zgon jest ok. 4.25 raza wyższa niż w grupie ECOG 0.
- ECOG 3: Szansa na zgon jest ok. 6.85 raza wyższa niż w grupie ECOG 0.

Skala Karnofsky’ego (scentrowana): OR = 0.9963. Wzrost wskaźnika sprawności o 1 punkt względem średniej wiąże się ze spadkiem szansy na zgon o ok. 0.4%. Kierunek ten jest zgodny (w końcu) z intuicją medyczną (wyższa sprawność to lepsze rokowania).

Warto zauważyć, że oszacowania dla zmiennej ECOG 3 różnią się między modelami (OR = 6.85 w modelu PO vs HR = 11 w modelu Coxa).

3.3 Zadanie 3

Teraz wyznaczmy oszacowanie bazowej skumulowanej funkcji hazardu i bazowej funkcji przeżycia odpowiadającej rozkładowi czasu życia. Następnie przyjrzymy się ich wykresom.

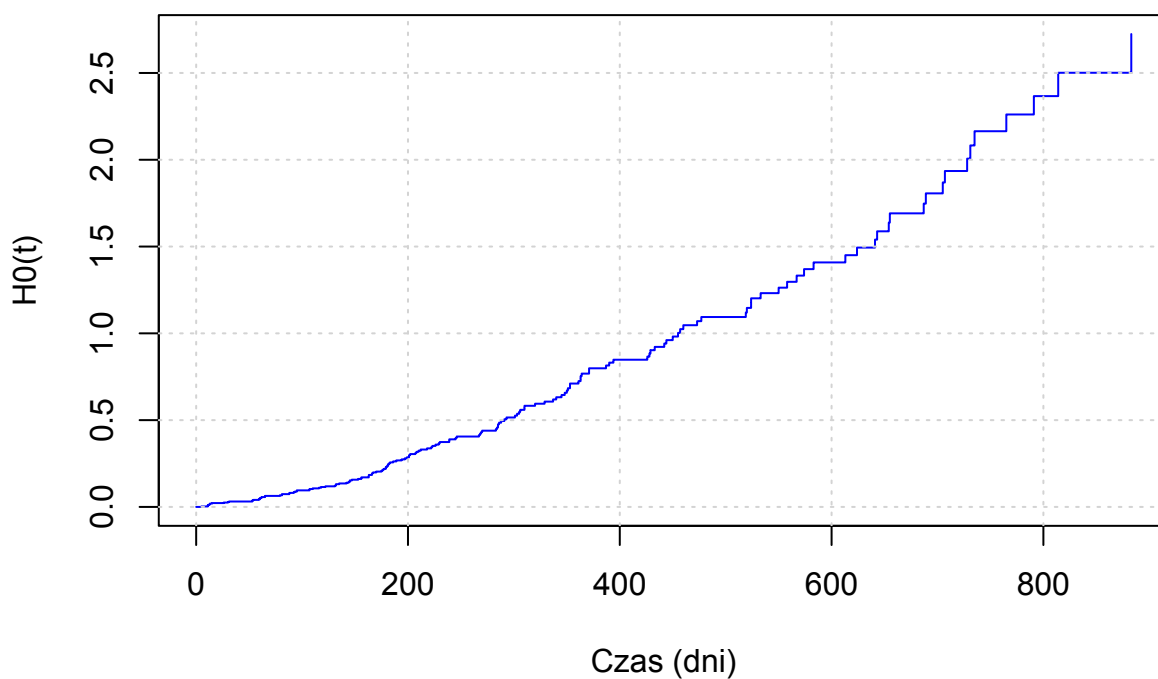
```
pred0 <- predict(fit_po, Z = rep(0, 6))
```

```
t_val <- pred0$time
```

```
S0_val <- pred0$S0
```

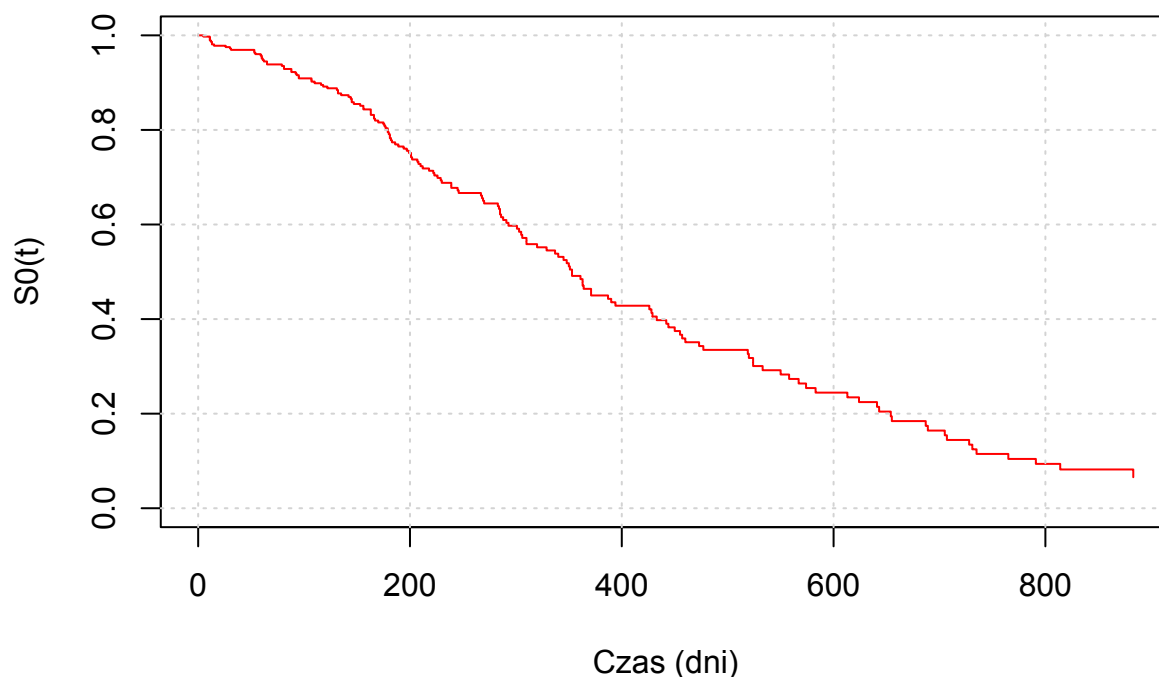
```
H0_val <- -log(S0_val)
```

Bazowy skumulowany hazard (PO)



Rysunek 10: Bazowa funkcja skumulowanego hazardu w modelu proporcjonalnych szans

Bazowa funkcja przeżycia (PO)



Rysunek 11: Bazowa funkcja przeżycia w modelu proporcjonalnych szans

Kształt bazowej funkcji przeżycia w modelu PO jest zbliżony do wyniku uzyskanego w modelu Coxa.

Powyższe wykresy przedstawiają charakterystykę bazową modelu. Dzięki scentrowaniu wieku i skali Karnofsky’ego, krzywe te opisują proces przeżycia dla pacjenta “typowego”: mężczyzny w wieku 62.5 lat, o sprawności ECOG 0 oraz średnim wyniku w skali Karnofsky’ego 82.

3.4 Zadanie 4

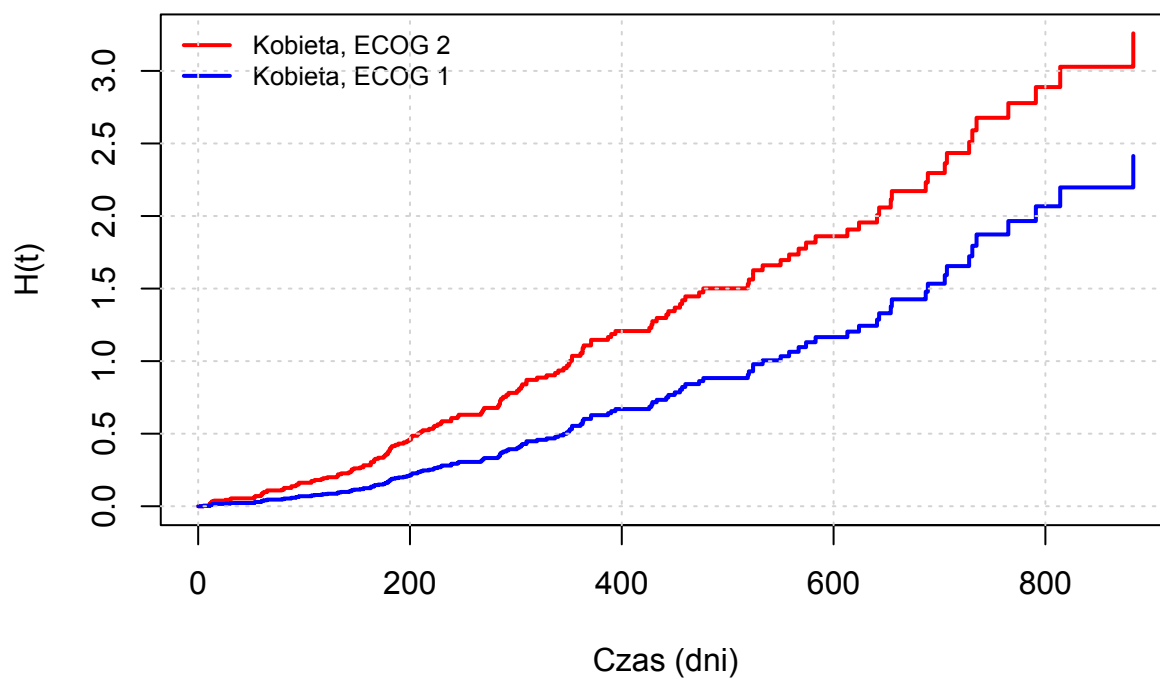
W tym zadaniu wyznaczymy oszacowanie skumulowanej funkcji hazardu odpowiadającej rozkładowi czasu życia przyjmując model Proporcjonalnych Szans.

```
z_a <- c(70 - mean_age, 1, 1, 0, 0, 90 - mean_karno)
z_b <- c(70 - mean_age, 1, 0, 1, 0, 90 - mean_karno)

pred_a <- predict(fit_po, Z = z_a)
pred_b <- predict(fit_po, Z = z_b)

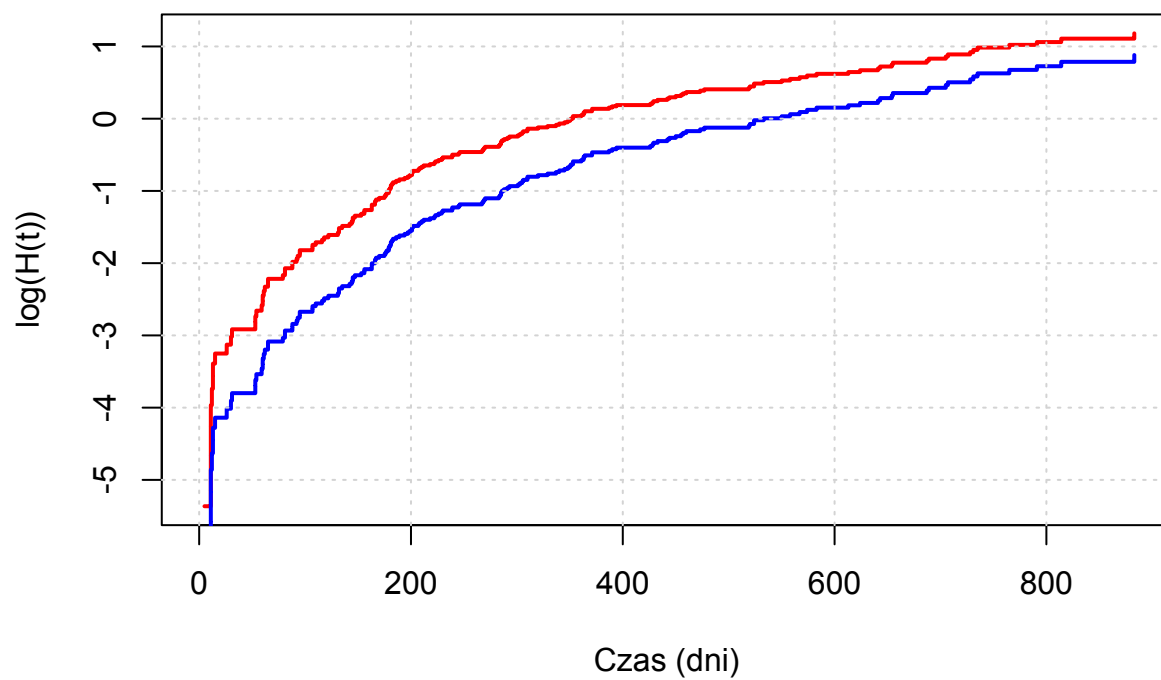
times <- pred_a$time
Ha_t <- -log(pred_a$S0)
Hb_t <- -log(pred_b$S0)
```

Skumulowany hazard $H(t)$

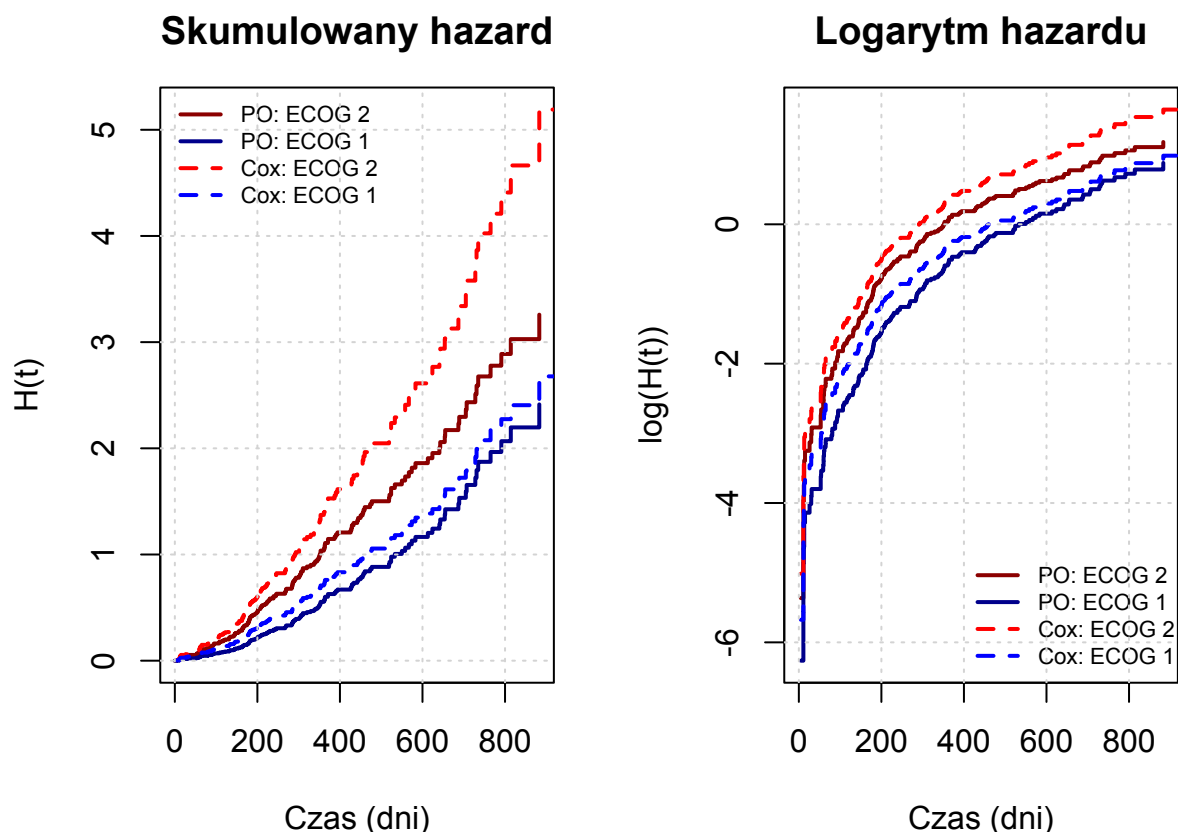


Rysunek 12: Porównanie skumulowanych funkcji hazardu dla kobiet w wieku 70 lat, $ph.karno=90$ i dwóch różnych statystyk $ph.ecog=1$ i $ph.ecog=2$ (model Proporcjonalnych Szans)

Logarytm hazardu $\log(H(t))$



Rysunek 13: Porównanie logarytmów skumulowanych funkcji hazardu dla kobiet w wieku 70 lat, ph.karno=90 i dwóch różnych statystyk ph.ecog=1 i ph.ecog=2 (model Proporcjonalnych Szans)



Rysunek 14: Porównanie hazardów modelu Coxa i modelu Proporcjonalnych Szans

- Oba modele poprawnie identyfikują, że pacjentka z ECOG 2 ma znacznie wyższy skumulowany hazard niż ta z ECOG 1.
- Model Coxa jest w tym przypadku bardziej pesymistyczny, bo przewiduje szybszy wzrost skumulowanego hazardu w porównaniu do modelu Proporcjonalnych Szans.
- Na wykresie logarytmów hazardu widać, że linie modelu Coxa są wobec siebie niemal idealnie równoległe, ponieważ log-hazardy różnią się o stałą. Linie modelu PO nie są równoległe - widać, że zbliżają się do siebie w czasie, co sugeruje, że różnica w ryzyku między grupami ECOG może maleć wraz z upływem czasu.

3.5 Zadanie 5

W niniejszym zadaniu wyznaczmy oszacowanie funkcji przeżycia (w dniach) odpowiadającej rozkładowi czasu życia

- (a) kobiety w wieku 70 lat o charakterystyce $ph.ecog=1$ i $ph.karno=90$
- (b) kobiety w wieku 70 lat o charakterystyce $ph.ecog=2$ i $ph.karno=90$

i na podstawie tego oszacowania obliczymy szacowane prawdopodobieństwo, że czas życia kobiet o powyżej podanych charakterystykach będzie większy niż 300 dni.

Prawdopodobieństwo $P(T > 300)$

- (a) Kobieta (ECOG=1): 0.675
- (b) Kobieta (ECOG=2): 0.4579

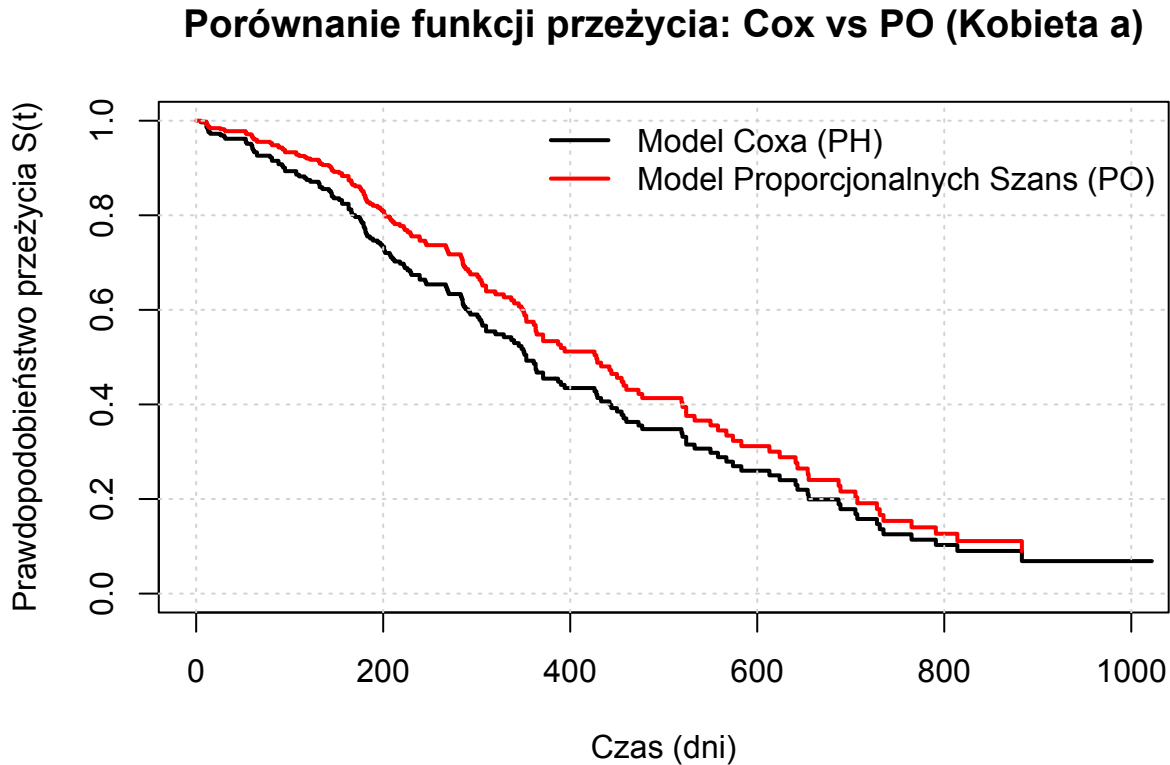
Wyniki otrzymane z modelu Coxa z listy 10:

- (a) Kobieta (ECOG=1): 0.5901

Wyższe prawdopodobieństwo przeżycia w modelu PO wynika z faktu, że w tej metodzie wpływ cech pacjenta (takich jak wiek czy ECOG) na ryzyko zgonu słabnie wraz z upływem czasu. W przeciwieństwie do modelu Coxa, gdzie różnica ryzyka między grupami jest stała, model PO zakłada, że po dłuższym czasie ryzyko różnych pacjentów zaczyna się do siebie upodabniać (iloraz hazardów dąży do 1).

3.6 Zadanie 6

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



Rysunek 15: Porównanie funkcji przeżycia modelu Coxa i modelu Proporcjonalnych Szans dla kobiety (a)

Zgodnie z przewidywaniami, model proporcjonalnych szans szacuje szanse na przeżycie pacjentki (a) optymistyczniej niż model Coxa. Różnica ta jest wyraźna na całym horyzoncie czasowym.

4 Lista 12

4.1 Zadanie 1

Przyjmując model przyspieszonego czasu awarii, w którym bazowa funkcja przeżycia odpowiada rozkładowi Weibulla, przyjmując za zmienną zależną zmienną `time`, a za charakterystyki zmienne: `age`, `sex`, `ph.ecog`, `ph.karno`, wykonamy poniższe zadania.

4.1.1 a)

Korzystając z testu Walda oraz testu IW, na poziomie istotności $\alpha = 0.05$, zweryfikujemy hipotezę, że zmienna `age` nie jest istotna w przyjętym modelu.

Na poziomie istotności $\alpha = 0.05$ przeprowadzono weryfikację hipotezy zerowej $H_0 : \beta_{age} = 0$ (zmienna **age** nie jest istotna) przeciwko hipotezie alternatywnej $H_1 : \beta_{age} \neq 0$ (zmienna **age** jest istotna).

- Test Walda: Uzyskana p-wartość wynosi 0.205
- Test Ilorazu Wiarogodności (IW): Uzyskana p-wartość wynosi 0.201

Decyzja statystyczna: Ponieważ w obu przypadkach p-wartość jest większa od przyjętego poziomu istotności nie mamy podstaw do odrzucenia hipotezy zerowej.

Interpretacja: Zmienna **age** nie jest istotna statystycznie w przyjętym modelu przyspieszonego czasu awarii (AFT) z rozkładem Weibulla. Oznacza to, że wiek pacjenta nie wpływa w sposób znaczący na czas przeżycia przy uwzględnieniu pozostałych zmiennych (płeć, stan ECOG, skala Karnofsky’ego).

4.1.2 b)

Korzystając z testu Walda oraz testu IW, na poziomie istotności $\alpha = 0.05$, zweryfikujemy hipotezę, że zmienna **sex** nie jest istotna w przyjętym modelu.

Na poziomie istotności $\alpha = 0.05$ przeprowadzono weryfikację hipotezy zerowej $H_0 : \beta_{sex} = 0$ (zmienna **sex** nie jest istotna) przeciwko hipotezie alternatywnej $H_1 : \beta_{sex} \neq 0$ (zmienna **sex** jest istotna).

- Test Walda: Uzyskana p-wartość wynosi 0.000905
- Test Ilorazu Wiarogodności (IW): Uzyskana p-wartość wynosi 0.000596

Decyzja statystyczna: Ponieważ p-wartości dla obu testów są znacznie mniejsze od przyjętego poziomu istotności, odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej.

Interpretacja: Zmienna **sex** jest istotna statystycznie w modelu. Płeć pacjenta ma znaczący wpływ na czas przeżycia; statystycznie kobiety charakteryzują się dłuższym czasem przeżycia niż mężczyźni przy zachowaniu pozostałych parametrów bez zmian.

4.1.3 c)

Korzystając z testu ilorazu wiarygodności, na poziomie istotności $\alpha = 0.05$, zweryfikujemy hipotezę, że rozkład czasu życia jest taki sam dla pacjenta o sprawności ECOG wg. lekarza na poziomie 0, 1, 2, 3

Na poziomie istotności $\alpha = 0.05$ przeprowadzono weryfikację hipotezy zerowej $H_0 : \beta_{ECOG_1} = \beta_{ECOG_2} = \beta_{ECOG_3} = 0$ (zmienna ECOG nie jest istotna) przeciwko hipotezie alternatywnej $H_1 : \exists i \beta_{ECOG_i} \neq 0$ (przynajmniej jeden poziom sprawności ECOG istotnie różnicuje czas życia).

- Test Ilorazu Wiarogodności (IW): Uzyskana p-wartość wynosi 0.00214

Decyzja statystyczna: Ponieważ p-wartość testu zbiorczego jest mniejsza od przyjętego poziomu istotności, odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej.

Interpretacja: Zmienna **ph.ecog** jest istotna statystycznie w modelu. Oznacza to, że ogólny stan sprawności pacjenta oceniany przez lekarza w skali ECOG istotnie różnicuje rozkład czasu życia pacjentów. Różne poziomy sprawności (od 0 do 3) wiążą się z różnymi prognozami dotyczącymi czasu przeżycia.

4.2 Zadanie 2

Przyjmując model proporcjonalnych hazardów Coxa ze zmienną zależną **time**, a za charakterystyki zmienne: **age**, **sex**, **ph.ecog**, **ph.karno**, wykonamy poniższe zadania.

4.2.1 a)

Korzystając z testu Walda oraz testu IW, na poziomie istotności $\alpha = 0.05$, zweryfikujemy hipotezę, że zmienna **age** nie jest istotna w przyjętym modelu.

Na poziomie istotności $\alpha = 0.05$ przeprowadzono weryfikację hipotezy zerowej $H_0 : \beta_{age} = 0$ (zmienna **age** nie jest istotna) przeciwko hipotezie alternatywnej $H_1 : \beta_{age} \neq 0$ (zmienna **age** jest istotna).

- Test Walda: Uzyskana p-wartość wynosi 0.184
- Test Ilorazu Wiarogodności (IW): Uzyskana p-wartość wynosi 0.18

Decyzja statystyczna: Ponieważ w obu przypadkach p-wartość jest większa od przyjętego poziomu istotności nie mamy podstaw do odrzucenia hipotezy zerowej.

Interpretacja: Zmienna **age** nie jest istotna statystycznie w przyjętym modelu Proporcjonalnych Hazardów Coxa. Oznacza to, że wiek pacjenta nie wpływa w sposób znaczący na funkcję hazardu przy uwzględnieniu pozostałych zmiennych (płeć, stan ECOG, skala Karnofsky’ego).

4.2.2 b)

Korzystając z testu Walda oraz testu IW, na poziomie istotności $\alpha = 0.05$, zweryfikujemy hipotezę, że zmienna **sex** nie jest istotna w przyjętym modelu.

Na poziomie istotności $\alpha = 0.05$ przeprowadzono weryfikację hipotezy zerowej $H_0 : \beta_{sex} = 0$ (zmienna **sex** nie jest istotna) przeciwko hipotezie alternatywnej $H_1 : \beta_{sex} \neq 0$ (zmienna **sex** jest istotna).

- Test Walda: Uzyskana p-wartość wynosi 0.000861
- Test Ilorazu Wiarogodności (IW): Uzyskana p-wartość wynosi 0.000629

Decyzja statystyczna: Ponieważ p-wartości dla obu testów są znacznie mniejsze od przyjętego poziomu istotności, odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej.

Interpretacja: Zmienna **sex** jest istotna statystycznie w modelu. Płeć pacjenta ma znaczący wpływ na czas przeżycia; statystycznie kobiety charakteryzują się dłuższym czasem przeżycia niż mężczyźni przy zachowaniu pozostałych parametrów bez zmian.

4.2.3 c)

Korzystając z testu ilorazu wiarygodności, na poziomie istotności $\alpha = 0.05$, zweryfikujemy hipotezę, że rozkład czasu życia jest taki sam dla pacjenta o sprawności ECOG wg. lekarza na poziomie 0, 1, 2, 3

Na poziomie istotności $\alpha = 0.05$ przeprowadzono weryfikację hipotezy zerowej $H_0 : \beta_{ECOG_1} = \beta_{ECOG_2} = \beta_{ECOG_3} = 0$ (zmienna **ecog** nie jest istotna) przeciwko hipotezie alternatywnej $H_1 : \exists i \beta_{ECOG_i} \neq 0$ (przynajmniej jeden poziom sprawności ECOG istotnie różnicuje czas życia).

- Test Ilorazu Wiarogodności (IW): Uzyskana p-wartość wynosi 0.00365

Decyzja statystyczna: Ponieważ p-wartość testu zbiorczego jest mniejsza od przyjętego poziomu istotności, odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej.

Interpretacja: Zmienna **ph.ecog** jest istotna statystycznie w modelu. Oznacza to, że ogólny stan sprawności pacjenta oceniany przez lekarza w skali ECOG istotnie różnicuje rozkład czasu życia pacjentów. Różne poziomy sprawności (od 0 do 3) wiążą się z różnymi prognozami dotyczącymi czasu przeżycia.

4.3 Zadanie 3

Przyjmując model przyspieszonego czasu awarii, w którym bazowa funkcja przeżycia odpowiada rozkładowi Weibulla, przyjmując za zmienną zależną zmienną **time**, a za charakterystyki zmienne: **age**, **sex**, **ph.ecog**, **ph.karno**, wykonamy poniższe zadania.

4.3.1 a)

Korzystając z metody eliminacji, w oparciu o test ilorazu wiarygodności, dokonamy wyboru zmiennych do modelu przyspieszonego czasu awarii.

Sprawdziliśmy już istotność `age`, `sex`, `ph.ecog`. Musimy sprawdzić jeszcze `ph.karno`.

- `age`: p-wartość wynosi 0.201
- `sex`: p-wartość wynosi 0.000596
- `ph.ecog`: p-wartość wynosi 0.00214
- `ph.karno`: p-wartość wynosi 0.133

Zaczynamy od usunięcia zmiennej `age`, dla której wartość poziomu krytycznego w teście jest największa i większa od przyjętego poziomu istotności $\alpha = 0.05$.

Teraz definiujemy model bez zmiennej `age` i ponownie sprawdzamy poziomy istotności dla pozostałych zmiennych.

- `sex`: p-wartość wynosi 0.000658
- `ph.ecog`: p-wartość wynosi 0.00204
- `ph.karno`: p-wartość wynosi 0.176

Teraz usuniemy `ph.karno`, bo dla niej p-wartość jest największa i przekracza przyjęty poziom istotności $\alpha = 0.05$.

- `sex`: p-wartość wynosi 0.00109
- `ph.ecog`: p-wartość wynosi 0.000379

W kroku trzecim wartości poziomów krytycznych dla wszystkich pozostałych w modelu zmiennych są mniejsze od przyjętego poziomu istotności $\alpha = 0.05$. Zgodnie z procedurą, nie możemy usunąć kolejnej zmiennej. Proces eliminacji zostaje zakończony.

Model końcowy (wybrany metodą eliminacji IW) przyjmuje postać: `Surv(time, status) ~ as.factor(sex) + as.factor(ph.ecog)`

4.3.2 b)

Korzystając z kryterium informacyjnego Akaike'a (AIC), dokonamy wyboru najlepszego modelu przyspieszonego czasu awarii, korzystając z funkcji `step`. Kryterium Akaike dodaje karę o wartości 2 za każdy parametr modelu.

```
## Start:  AIC=2266.6
## Surv(time, status) ~ as.factor(sex) + as.factor(ph.ecog) + age_c +
##      ph.karno_c
##
##              Df    AIC
## - age_c          1 2266.2
## <none>            2266.6
## - ph.karno_c      1 2266.8
## - as.factor(ph.ecog) 3 2275.2
## - as.factor(sex)   1 2276.4
##
## Step:  AIC=2266.23
## Surv(time, status) ~ as.factor(sex) + as.factor(ph.ecog) + ph.karno_c
##
##              Df    AIC
## - ph.karno_c      1 2266.1
## <none>            2266.2
## - as.factor(ph.ecog) 3 2275.0
## - as.factor(sex)   1 2275.8
##
```

```
## Step: AIC=2266.07
## Surv(time, status) ~ as.factor(sex) + as.factor(ph.ecog)
##
##              Df      AIC
## <none>              2266.1
## - as.factor(sex)      1 2274.7
## - as.factor(ph.ecog)  3 2278.4
##
## Call:
## survreg(formula = Surv(time, status) ~ as.factor(sex) + as.factor(ph.ecog),
##          data = lung.subset.complete, dist = "weibull")
##              Value Std. Error      z      p
## (Intercept)      6.1907      0.1244 49.77 < 2e-16
## as.factor(sex)2      0.3900      0.1240  3.14  0.0017
## as.factor(ph.ecog)1 -0.2964      0.1450 -2.04  0.0410
## as.factor(ph.ecog)2 -0.6789      0.1658 -4.09 4.2e-05
## as.factor(ph.ecog)3 -1.4200      0.7397 -1.92  0.0549
## Log(scale)      -0.3158      0.0617 -5.11 3.1e-07
##
## Scale= 0.729
##
## Weibull distribution
## Loglik(model)= -1127 Loglik(intercept only)= -1141.1
## Chisq= 28.05 on 4 degrees of freedom, p= 1.2e-05
## Number of Newton-Raphson Iterations: 5
## n= 226
```

Proces doboru modelu za pomocą funkcji step przebiegał w następujących krokach:

- Model początkowy (AIC = 2266.60): Zawierał wszystkie zmienne (`sex`, `ph.ecog`, `age_c`, `ph.karno_c`).
- Krok 1 (AIC = 2266.23): Usunięto zmienną `age_c`. Zauważono, że AIC spadło z 2266.6 na 2266.23, co oznacza poprawę modelu po uproszczeniu.
- Krok 2 (AIC = 2266.07): Usunięto zmienną `ph.karno_c`. AIC spadło z 2266.23 na 2266.07, co algorytm uznał za korzystne.
- Algorytm zatrzymał się na modelu ze zmiennymi `sex` i `ph.ecog`. Jakakolwiek dalsza próba usunięcia zmiennych spowodowałaby gwałtowny wzrost AIC (np. z 2266.1 do 2274.7).

Postać modelu końcowego: `Surv(time, status) ~ as.factor(sex) + as.factor(ph.ecog)`

Interpretacja:

- Płeć (`as.factor(sex)2 = 0.390`): oznacza to, że czas przeżycia kobiet jest dłuższy niż mężczyzn przy takim samym stanie sprawności ECOG.
- Stan sprawności (`ph.ecog`): Współczynniki są ujemne, co oznacza skrócenie czasu przeżycia wraz z pogarszaniem się stanu pacjenta względem grupy referencyjnej (ECOG 0).

4.3.3 c)

Korzystając z bayesowskiego kryterium informacyjnego (BIC), dokonamy wyboru najlepszego modelu przyspieszonego czasu awarii, korzystając z funkcji `step`. Bayesowskie kryterium informacyjne dodaje karę $\ln(n)$ za każdy parametr, gdzie n to liczba kompletnych obserwacji. W naszym przypadku $\ln(163)$ co wynosi 5.09.

```
## Start: AIC=2291.35
## Surv(time, status) ~ as.factor(sex) + as.factor(ph.ecog) + age_c +
##      ph.karno_c
```

```

##
##              Df    AIC
## - age_c      1 2287.9
## - ph.karno_c  1 2288.5
## - as.factor(ph.ecog) 3 2290.7
## <none>        2291.3
## - as.factor(sex) 1 2298.0
##
## Step: AIC=2287.89
## Surv(time, status) ~ as.factor(sex) + as.factor(ph.ecog) + ph.karno_c
##
##              Df    AIC
## - ph.karno_c  1 2284.6
## - as.factor(ph.ecog) 3 2287.4
## <none>        2287.9
## - as.factor(sex) 1 2294.4
##
## Step: AIC=2284.63
## Surv(time, status) ~ as.factor(sex) + as.factor(ph.ecog)
##
##              Df    AIC
## <none>        2284.6
## - as.factor(ph.ecog) 3 2287.7
## - as.factor(sex) 1 2290.2
##
## Call:
## survreg(formula = Surv(time, status) ~ as.factor(sex) + as.factor(ph.ecog),
## data = lung.subset.complete, dist = "weibull")
##              Value Std. Error      z      p
## (Intercept)    6.1907    0.1244 49.77 < 2e-16
## as.factor(sex)2    0.3900    0.1240  3.14  0.0017
## as.factor(ph.ecog)1 -0.2964    0.1450 -2.04  0.0410
## as.factor(ph.ecog)2 -0.6789    0.1658 -4.09 4.2e-05
## as.factor(ph.ecog)3 -1.4200    0.7397 -1.92  0.0549
## Log(scale)     -0.3158    0.0617 -5.11 3.1e-07
##
## Scale= 0.729
##
## Weibull distribution
## Loglik(model)= -1127 Loglik(intercept only)= -1141.1
## Chisq= 28.05 on 4 degrees of freedom, p= 1.2e-05
## Number of Newton-Raphson Iterations: 5
## n= 226

```

Proces doboru modelu za pomocą funkcji step przebiegał w następujących krokach:

- Model początkowy (BIC = 2293.96): Zawierał wszystkie zmienne (sex, ph.ecog, age_c, ph.karno_c).
- Krok 1 (BIC = 2290.17): Usunięto zmienną age_c.
- Krok 2 (BIC = 2286.59): Usunięto zmienną ph.karno_c.
- Algorytm zatrzymał się na modelu ze zmiennymi sex i ph.ecog.

Postać modelu końcowego: `Surv(time, status) ~ as.factor(sex) + as.factor(ph.ecog)`

Interpretacja dla podpunktów a) b) i c) jest identyczna, bo otrzymaliśmy takie same modele.