Sprawozdanie 2

Analiza przeżycia

Marta Stankiewicz (282244) Kacper Szmigielski (282255)

Spis treści

1 Zadanie 1 **2**
1.1 Treść 2
1.2 Rozwiązanie 2

2 Zadanie 2 **3**
2.1 Treść 3
2.2 Rozwiązanie 3

3 Zadanie 3 **5**
3.1 Treść 5
3.2 Rozwiązanie 5

4 Zadanie 4 **7**
4.1 Treść 7
4.2 Rozwiązanie 7

5 Zadanie 5 **9**
5.1 Treść 9
5.2 Rozwiązanie 9

6 Zadanie 6 **11**

Spis rysunków

Spis tabel

2 Prawdopodobieństwo przeżycia > 300 dni 9

1 Zadanie 1

1.1 Treść

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

1.2 Rozwiązanie

W modelu proporcjonalnych hazardów Coxa mamy

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

Najczęściej przyjmuje się, że

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

Czyli

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

```
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 Zadanie 2

2.1 Treść

Podać interpretację współczynników modelu z zadania 1.

2.2 Rozwiązanie

	beta	HR	lower .95	upper .95	p.value
age	0.0125599	1.0126391	0.9940524	1.0315733	0.1839029
factor(sex)2	-0.5656752	0.5679765	0.4072293	0.7921761	0.0008610
factor(ph.ecog)1	0.5780583	1.7825739	1.1215652	2.8331564	0.0144745
factor(ph.ecog)2	1.2398953	3.4552518	1.7220481	6.9328870	0.0004836
factor(ph.ecog)3	2.3958532	10.9775597	1.3187800	91.3774984	0.0266993
ph.karno	0.0124226	1.0125001	0.9936499	1.0317079	0.1951182

Dla age mamy wsp

$$\beta > 0$$

co wskazuje że ta zmienna powoduje wzrost hazardu. Przedział ufności wskazuje również na dokładne oszacowanie współcztynnika beta. Problem pojawia się przy p-value, ponieważ

$$p > 0.05$$

co oznacza, że uzyskany wynik nie jest statystycznie istotny

Dla zmiennej sex wsp.

$$\beta < 0$$

oraz

$$p < 0.05$$

co wskazuje, że zmienna sex ma stosunkowy wpływ na hazard w stosunku do poziomu bazowego. Zmniejsza go. Również przedział ufności jest dość wązki, co wskazuje na dokładność oszacowania.

Dla zmiennej ph.ecog widać, że w każdym przypadku wsp.

$$\beta > 0$$

co oznacza, że zwiększa ona hazard w stosunku do poziomu bazowego. W każdym przypadku również

$$p < 0.05$$

więc wyniki są statystycznie istotne. Problem pojawia się przy przedziałach Walda, które w każdym przypadku są stosunkowo szerokie co może wskazywać na niedokładność oszacowania.

Dla zmiennej ph.karno mamy wsp.

$$\beta > 0$$

oraz wązki przedział ufności, lecz niestety

$$p > 0.05$$

, więc nie możemy uznawać wyniku za statystycznie istotny. co myślisz o takich wnioskach?

3 Zadanie 3

3.1 Treść

Wyznaczyć oszacowanie bazowej skumulowanej funkcji hazardu i bazowej funkcji przeżycia odpowiadającej rozkładowi czasu życia

3.2 Rozwiązanie

Tutaj wykorzystamy wzory przedstawione w zadaniu nr.1

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

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

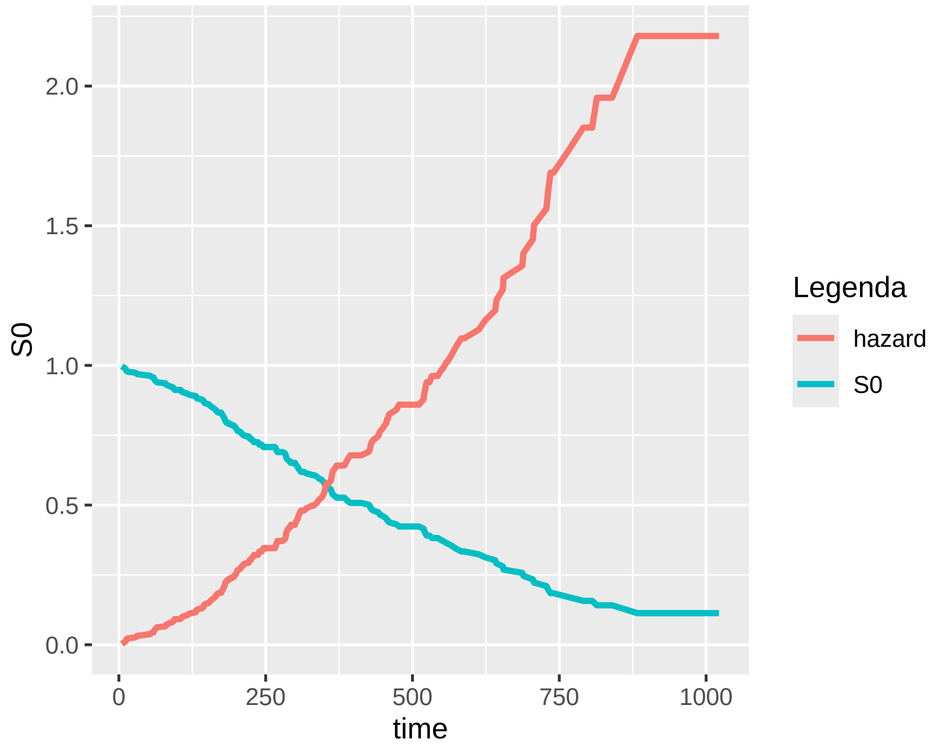
$$S(t \mid x) = \Big(S_0(t)\Big)^{\exp(x^\top \beta)}$$

$$S_0(t) = \exp\Big(- H_0(t)\Big)$$

A dla bazowego Weibulla

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

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


S0
2.0
1.5
1.0
0.5
0.0
0
250
500
750
1000
time
Legenda
hazard
S0

4 Zadanie 4

4.1 Treść

Wyznaczyć oszacowanie skumulowanej funkcji hazardu 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, narysować wykresy tych funkcji i wykresy logarytmów tych funkcji. Czy na podstawie tych wykresów możemy mieć wątpliwości co do przyjętego modelu proporcjonalnych hazardów?

4.2 Rozwiązanie

Skumulowany hazard:

$$H(t \mid x) = \int_0^t h(u \mid x)\, du$$

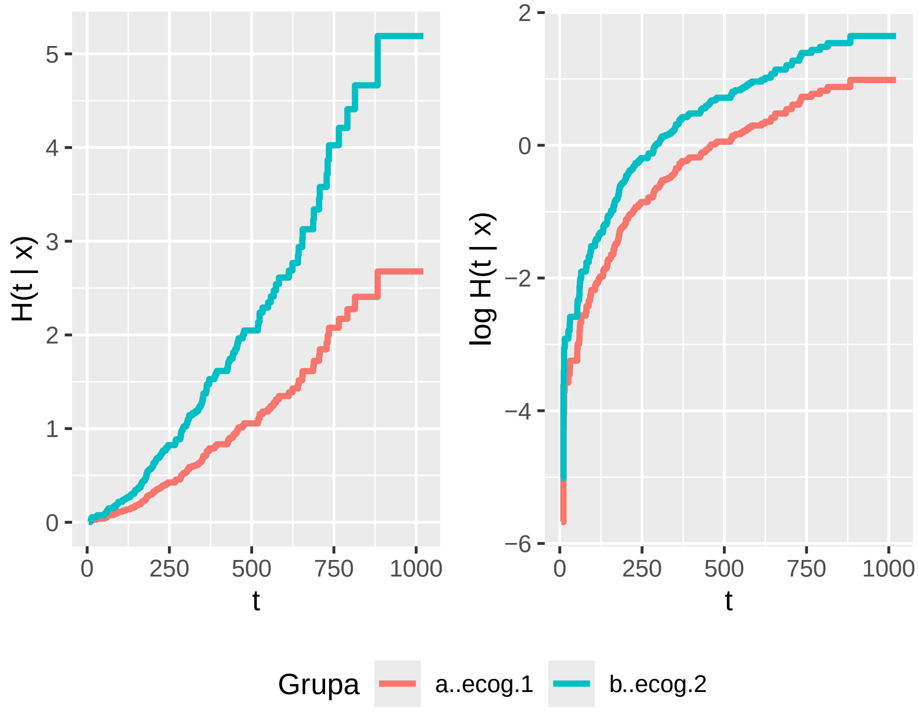
W modelu Coxa:

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

Po zalogowaniu:

$$\log H(t \mid x) = \log H_0(t) + \beta^\top x$$

Więc wykresy logarytmów powinny być równoległe i przesunięte jedynie na osi OY.



Wykresy wskazują na prawidłowośc zastosowanego modelu. Dla

$$H(t|x)$$

widać że wykresy są przeskalowane (przemnożnoe względem siebie). Natomiast dla

$$log(H(t|x))$$

wykresy są przesuniętem wzgledem siebie na osi OY. Czyli wyniki są zgodne z założeniami. Wyniki nie skłaniają ku wątpieniu w prawidłowość modelu.

5 Zadanie 5

5.1 Treść

Wyznaczyć 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 obliczyć szacowane prawdopodobieństwo, że czas życia kobiet o powyżej podanych charakterystykach będzie większy niż 300 dni. Oszacowane prawdopodobieństwo dla kobiety o charakterystykach opisanych w punkcie (a) porównać z prawdopodobieństwem uzyskanym w zadaniu 3 oraz 8 z listy 9

5.2 Rozwiązanie

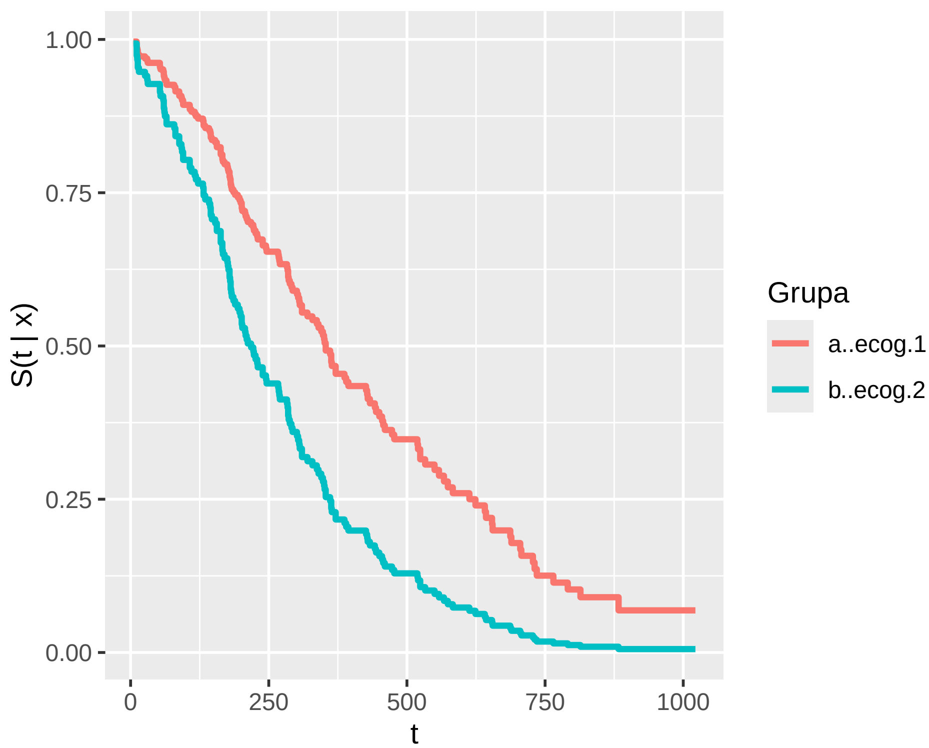


Tabela 2: Prawdopodobieństwo przeżycia > 300 dni

	ph.ecog. . . 1	ph.ecog. . . 2
Prawdopodobieństwo	0.5901344	0.3597682

Prawdopodobiestwo przeżycia przez pacjentkę z ph.ecog = 1 conajmniej 300 dni wynosi a przez pacjentkę z ph.ecog = 2 wynosi

$$0.2$$

6 Zadanie 6