

Lista 11

Analiza przeżycia

Marta Stankiewicz (282244) Kacper Szmigielski (282255)

Spis treści

1	Zadanie 1	3
1.1	a)	3
1.2	b)	3
1.3	c)	3
2	Zadanie 2	4
2.1	a)	4
2.2	b)	4
2.3	c)	4
3	Zadanie 3	5
3.1	a)	5
3.2	b)	6
3.3	c)	7
4	Zadanie 4	8
4.1	a)	8
4.2	b)	10
4.3	c)	11

Spis rysunków

Spis tabel

1	Wartości p testu IW	5
2	Wartości p testu IW	6
3	Wartości p testu IW	6
4	Wartości AIC	7
5	Wartości AIC	7
6	Wartości AIC	7
8	Wartości p testu IW	9
9	Wartości p testu IW	9
10	Wartości p testu IW	10
11	Wartości AIC	10
12	Wartości AIC	10
13	Wartości AIC	11

1 Zadanie 1

W tym zadaniu omawiamy model **przyspieszonego czasu awarii**

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

1.1 a)

Dla zmiennej *age* w teście Walda otrzymujemy $p_{value} = 0.2052959 > 0.05$, co oznacza, że przy poziomie istotności $\alpha = 0.05$ nie ma podstaw do odrzucenia hipotezy $H_0 : \beta_{age} = 0$, a więc brak jest dowodów na istotność tej zmiennej w modelu.

Dla testu Ilorazu Wiarygodności otrzymaliśmy $p_{value} = 0.2013662$, również > 0.05 , co wskazuje, że uwzględnienie zmiennej *age* nie powoduje istotnej zmiany w dopasowanym modelu.

1.2 b)

Dla zmiennej *sex* w teście Walda otrzymujemy $p_{value} = 9.0513757 \times 10^{-4} < 0.05$, co oznacza, że przy poziomie istotności $\alpha = 0.05$ odrzucamy hipotezę $H_0 : \beta_{sex} = 0$. Wskazuje to, że zmienna *sex* ma istotny wpływ na czas do zdarzenia w modelu AFT, tzn. istotnie przyspiesza lub spowalnia czas awarii względem poziomu bazowego.

Dla testu Ilorazu Wiarygodności otrzymaliśmy $p_{value} = 5.9586684 \times 10^{-4} < 0.05$, co wskazuje, że uwzględnienie zmiennej *sex* ma istotny wpływ na dopasowany model.

1.3 c)

W teście ilorazu wiarygodności otrzymano $p_{value} = 0.0021387 < 0.05$, więc przy poziomie istotności $\alpha = 0.05$ odrzucamy hipotezę zerową $H_0 : \beta_{ph.ecog=1} = \beta_{ph.ecog=2} = \beta_{ph.ecog=3} = 0$

(tj. brak wpływu poziomu ECOG na czas do zdarzenia względem poziomu bazowego, np. ECOG = 0). Oznacza to, że zmienna *ph.ecog* jest istotna w przyjętym modelu AFT, a poziom sprawności (ECOG) istotnie różnicuje czas przeżycia/czas awarii pacjentów.

2 Zadanie 2

```
model_podst <- coxph(Surv(time, status)~age +  
                      as.factor(sex) +  
                      as.factor(ph.ecog) +  
                      ph.karno,  
                      data = dane,)
```

2.1 a)

Dla zmiennej *age* w teście Walda otrzymujemy $p_{value} = 0.1839029 > 0.05$, co oznacza, że przy poziomie istotności $\alpha = 0.05$ nie ma podstaw do odrzucenia hipotezy $H_0 : \beta_{age} = 0$, a więc brak jest dowodów na istotność tej zmiennej w modelu.

Dla testu Ilorazu Wiarygodności otrzymaliśmy $p_{value} = 0.1803751$, również > 0.05 , co wskazuje, że uwzględnienie zmiennej *age* nie powoduje istotnej zmiany dopasowanego modelu.

2.2 b)

Dla zmiennej *sex* w teście Walda otrzymujemy $p_{value} = 8.609901 \times 10^{-4} < 0.05$, co oznacza, że przy poziomie istotności $\alpha = 0.05$ odrzucamy hipotezę $H_0 : \beta_{sex} = 0$. Wskazuje to, że zmienna *sex* ma istotny wpływ na czas do zdarzenia w modelu AFT, tzn. istotnie przyspiesza lub spowalnia czas awarii względem poziomu bazowego.

Dla testu Ilorazu Wiarygodności otrzymaliśmy $p_{value} = 1 < 0.05$, co wskazuje, że uwzględnienie zmiennej *sex* istotnie poprawia dopasowanie modelu w porównaniu z modelem bez tej zmiennej. Zatem *sex* jest istotnym predyktorem w modelu przyspieszonego czasu awarii, wpływając na skalę czasu przeżycia względem poziomu bazowego.

2.3 c)

W teście ilorazu wiarygodności otrzymano $p_{value} = 0.0021387 < 0.05$, więc przy poziomie istotności $\alpha = 0.05$ odrzucamy hipotezę zerową $H_0 : \beta_{ph.ecog=1} = \beta_{ph.ecog=2} = \beta_{ph.ecog=3} = 0$

(tj. brak wpływu poziomu ECOG na czas do zdarzenia względem poziomu bazowego, np. ECOG = 0). Oznacza to, że zmienna *ph.ecog* jest istotna w przyjętym modelu proporcjonalnych hazardów Coxa, a poziom sprawności (ECOG) istotnie różnicuje czas przeżycia/czas awarii pacjentów.

3 Zadanie 3

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

3.1 a)

KROK 1

```
age <- anova(update(m, . ~ . - age), m)[["Pr(>Chi)"]][2]
sex <- anova(update(m, . ~ . - as.factor(sex)), m)[["Pr(>Chi)"]][2]
ph.ecog <- anova(update(m, . ~ . - as.factor(ph.ecog)), m)[["Pr(>Chi)"]][2]
ph.karno <- anova(update(m, . ~ . - ph.karno), m)[["Pr(>Chi)"]][2]

p_values <- data.frame(
  age = age,
  sex = sex,
  ph.ecog = ph.ecog,
  ph.karno = ph.karno
)

kable(p_values, digits = 3, caption = "Wartości p testu IW")
```

Tabela 1: Wartości p testu IW

age	sex	ph.ecog	ph.karno
0.201	0.001	0.002	0.133

Usuujemy zmienną *age*, ponieważ ma ona najwyższe *p*-value spośród rozważanych predyktorów ($p = 0.2013662 > 0.05$), co oznacza brak podstaw do uznania jej wpływu za istotny statystycznie w przyjętym modelu.

KROK 2

```
m <- survreg(Surv(time, status)~as.factor(sex) +
             as.factor(ph.ecog) +
             ph.karno,
```

```

data = dane,
dist = "weibull")

sex <- anova(update(m, . ~ . - as.factor(sex)), m)[["Pr(>Chi)"]][2]
ph.ecog <- anova(update(m, . ~ . - as.factor(ph.ecog)), m)[["Pr(>Chi)"]][2]
ph.karno <- anova(update(m, . ~ . - ph.karno), m)[["Pr(>Chi)"]][2]

p_values <- data.frame(
  sex = sex,
  ph.ecog = ph.ecog,
  ph.karno = ph.karno
)

kable(p_values, digits = 3, caption = "Wartości p testu IW")

```

Tabela 2: Wartości p testu IW

sex	ph.ecog	ph.karno
0.001	0.002	0.176

Usuujemy zmienną *ph.karno*, ponieważ ma ona najwyższe *p*-value spośród rozważanych predyktorów ($p = 0.1756437 > 0.05$), co oznacza brak podstaw do uznania jej wpływu za istotny statystycznie w przyjętym modelu.

KROK 3

Tabela 3: Wartości p testu IW

sex	ph.ecog
0.001	0

Dla wszystkich pozostałych predyktorów otrzymano wartości $p < 0.05$, dlatego procedura selekcji (w oparciu o test IW) została zakończona i przyjęto model końcowy

3.2 b)

```

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

```

KROK 1

Tabela 4: Wartości AIC

wszystko	bez_age	bez_sex	bez_ph.ecoh	bez_ph.karno
2266.599	2266.231	2276.388	2275.252	2266.854

Ponieważ po usunięciu zmiennej *age* kryterium AIC przyjmuje najmniejszą wartość, uznajemy model bez *age* za lepiej dopasowany i dlatego *age* zostaje usunięta z modelu.

KROK 2

Tabela 5: Wartości AIC

wszystko	bez_sex	bez_ph.ecoh	bez_ph.karno
2266.231	2275.834	2274.984	2266.065

Ponieważ po usunięciu zmiennej *ph.karno* kryterium AIC osiąga najniższą wartość, wybieramy model bez *ph.karno* jako lepszy.

KROK 3

Tabela 6: Wartości AIC

wszystko	bez_sex	bez_ph.ecoh
2266.065	2274.735	2278.379

Usunięcie którejkolwiek z pozostałych zmiennych nie prowadzi do dalszego obniżenia wartości kryterium AIC, dlatego przyjmujemy uzyskany model jako model końcowy.

3.3 c)

Tutaj już korzystam z funkcji `step`, która automatyzuje proces przedstawiony powyżej.

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

n <- sum(dane$status)

m_bic <- step(m, direction = "backward", k = log(n), trace = 0)

kable(m_bic$anova)
```

Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
	NA	NA	218	2250.599	2291.349
- age	1	1.632438	219	2252.231	2287.887
- ph.karno	1	1.834102	220	2254.065	2284.628

Według kryterium bayesowskiego (BIC) procedura selekcji prowadzi do takiego samego wniosku jak w poprzednich metodach: w modelu końcowym pozostają zmienne *sex* oraz *ph.ecog*. Eliminacja predyktorów przebiega analogicznie jak w przypadku selekcji opartej na AIC, przy czym decyzje podejmowane są na podstawie minimalizacji wartości BIC. Najpierw został odrzucony *age* a potem *ph.karno*.

4 Zadanie 4

```
m <- coxph(Surv(time, status)~age +
  as.factor(sex) +
  as.factor(ph.ecog) +
  ph.karno,
  data = dane)
```

4.1 a)

KROK 1

```
age <- anova(update(m, . ~ . - age), m)$"Pr(>|Chi|)"[2]
sex <- anova(update(m, . ~ . - as.factor(sex)), m)$"Pr(>|Chi|)"[2]
ph.ecog <- anova(update(m, . ~ . - as.factor(ph.ecog)), m)$"Pr(>|Chi|)"[2]
ph.karno <- anova(update(m, . ~ . - ph.karno), m)$"Pr(>|Chi|)"[2]

p_values <- data.frame(
  age = age,
  sex = sex,
  ph.ecog = ph.ecog,
  ph.karno = ph.karno
)

kable(p_values,digits = 3 ,caption = "Wartości p testu IW")
```


Tabela 8: Wartości p testu IW

age	sex	ph.ecog	ph.karno
0.18	0.001	0.004	0.189

Usuujemy zmienną *age*, ponieważ ma ona najwyższe *p*-value spośród rozważanych predyktorów ($p = 0.1803751 > 0.05$), co oznacza brak podstaw do uznania jej wpływu za istotny statystycznie w przyjętym modelu.

KROK 2

```
m <- coxph(Surv(time, status)~as.factor(sex) +
           as.factor(ph.ecog) +
           ph.karno,
           data = dane)

sex <- anova(update(m, . ~ . - as.factor(sex)), m)$"Pr(>|Chi|)"[2]
ph.ecog <- anova(update(m, . ~ . - as.factor(ph.ecog)), m)$"Pr(>|Chi|)"[2]
ph.karno <- anova(update(m, . ~ . - ph.karno), m)$"Pr(>|Chi|)"[2]

p_values <- data.frame(
  sex = sex,
  ph.ecog = ph.ecog,
  ph.karno = ph.karno
)

kable(p_values,digits = 3 ,caption = "Wartości p testu IW")
```

Tabela 9: Wartości p testu IW

sex	ph.ecog	ph.karno
0.001	0.003	0.246

Usuujemy zmienną *ph.karno*, ponieważ ma ona najwyższe *p*-value spośród rozważanych predyktorów ($p = 0.2455646 > 0.05$), co oznacza brak podstaw do uznania jej wpływu za istotny statystycznie w przyjętym modelu.

KROK 3

```
## [1] 0.001032973
```

Tabela 10: Wartości p testu IW

sex	ph.ecog
0.001	0

Dla wszystkich pozostałych predyktorów otrzymano wartości $p < 0.05$, dlatego procedura selekcji (w oparciu o test IW) została zakończona i przyjęto model końcowy

4.2 b)

```
m <- coxph(Surv(time, status)~age +
            as.factor(sex) +
            as.factor(ph.ecog) +
            ph.karno,
            data = dane)
```

KROK 1

Tabela 11: Wartości AIC

wszystko	bez_age	bez_sex	bez_ph.ecog	bez_ph.karno
1458.537	1458.332	1468.226	1466.05	1458.266

Ponieważ po usunięciu zmiennej *age* kryterium AIC przyjmuje najmniejszą wartość, uznajemy model bez *age* za lepiej dopasowany i dlatego *age* zostaje usunięta z modelu.

KROK 2

Tabela 12: Wartości AIC

wszystko	bez_sex	bez_ph.ecog	bez_ph.karno
1458.332	1467.841	1465.936	1457.68

Ponieważ po usunięciu zmiennej *ph.karno* kryterium AIC osiąga najniższą wartość, wybieramy model bez *ph.karno* jako lepszy.

KROK 3

Tabela 13: Wartości AIC

wszystko	bez_sex	bez_ph.ecog
1457.68	1466.447	1470.137

Usunięcie którejkolwiek z pozostałych zmiennych nie prowadzi do dalszego obniżenia wartości kryterium AIC, dlatego przyjmujemy uzyskany model jako model końcowy.

4.3 c)

Tutaj już korzystam z funkcji `step`, która automatyzuje proces przedstawiony powyżej.

```
m <- coxph(Surv(time, status)~age + as.factor(sex) + as.factor(ph.ecog) + ph.karno,
           data = dane)

n <- sum(dane$status)

m_bic <- step(m, direction = "backward", k = log(n), trace = 0)

kable(m_bic$anova)
```

Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
	NA	NA	157	1446.537	1477.100
- ph.karno	1	1.728580	158	1448.266	1473.734
- age	1	1.414313	159	1449.680	1470.055

Według kryterium bayesowskiego (BIC) procedura selekcji prowadzi do takiego samego wniosku jak w poprzednich metodach: w modelu końcowym pozostają zmienne *sex* oraz *ph.ecog*. Eliminacja predyktorów przebiega analogicznie jak w przypadku selekcji opartej na AIC, przy czym decyzje podejmowane są na podstawie minimalizacji wartości BIC. Najpierw został odrzucony *age* a potem *ph.karno*.