

Lista 11

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	4
2.1	Treść	4
2.2	Rozwiązanie	4
3	Zadanie 3	6
3.1	Treść	6
3.2	Rozwiązanie	6
4	Zadanie 4	9
4.1	Treść	9
4.2	Rozwiązanie	9

Spis rysunków

Spis tabel

1 Zadanie 1

1.1 Treść

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, wykonać poniższe zadania.

1.2 Rozwiązanie

1.2.1 a)

1.2.1.1 Treść Korzystając z testu Walda oraz testu IW, na poziomie istotności 0.05, zweryfikować hipotezę, że zmienna age nie jest istotna w przyjętym modelu

1.2.1.2 Rozwiązanie

1.2.1.2.1 Wald P value dla testu Walda wynosi:

```
##           p
## 0.2052959
```

Widać, że $p > 0.05$

Nie odrzucamy

H_0

, więc uznajemy zmienną age za nieistotną

1.2.1.2.2 IW P value dla testu największej wiarygodności wynosi:

```
## [1] 0.2013662
```

Widać, że $p > 0.05$ Nie odrzucamy

H_0

, więc uznajemy zmienną age za nieistotną

1.2.2 b)

1.2.2.0.1 Wald P value dla testu Walda wynosi:

```
##                p
## 0.0009051376
```

Odrzucamy

$$H_0$$

, więc uznajemy zmienną sex za istotną

1.2.2.0.2 IW P value dla testu największej wiarygodności wynosi:

Widać, że $p < 0.05$ Odrzucamy

$$H_0$$

, więc uznajemy zmienną sex za istotną

1.2.3 c)

1.2.3.1 Treść

- (c) Korzystając z testu ilorazu wiarygodności, na poziomie istotności 0.05, zweryfikować hipotezę, że rozkład czasu życia jest taki sam dla pacjenta o sprawności ECOG wg. lekarza na poziomie 0, 1, 2, 3 (zmienna ph.ecog teoretycznie może przyjmować wartości 0, 1, 2, 3, 4, 5, ale w zbiorze danych mamy tylko wartości od 0 do 3)

1.2.3.2 Rozwiązanie

1.2.3.2.1 IW P value dla testu największej wiarygodności wynosi:

```
## [1] 0.002138659
```

W teście ilorazu wiarygodności otrzymano $p = 0.00214 < 0.05$, więc odrzucamy hipotezę

$$H_0$$

, że rozkład czasu życia jest taki sam dla pacjentów z ECOG 0,1,2,3. Oznacza to, że zmienna ph.ecog jest istotna w przyjętym modelu

2 Zadanie 2

2.1 Treść

Przyjmując model proporcjonalnych hazardów Coxa ze zmienną zależną time, i z charakterystykami: age, sex, ph.ecog, ph.karno, wykonać poniższe zadania.

2.2 Rozwiązanie

2.2.1 a)

2.2.1.1 Treść Korzystając z testu Walda oraz z testu IW, na poziomie istotności 0.05, zweryfikować hipotezę, że zmienna age nie jest istotna w przyjętym modelu

2.2.1.2 Rozwiąza

2.2.1.2.1 Wald P value dla testu największej wiarygodności wynosi:

```
## Pr(>|z|)
## 0.1839029
```

Widać, że $p > 0.05$ Nie odrzucamy

H_0

, więc uznajemy zmienną age za nieistotną

2.2.1.2.2 IW P value dla testu największej wiarygodności wynosi:

```
## Pr(>|Chi|)
## 1
## 2 0.1804
```

Widać, że $p > 0.05$ Nie odrzucamy

H_0

, więc uznajemy zmienną age za nieistotną

2.2.2 b)

2.2.2.1 Treść Korzystając z testu Walda oraz testu IW, na poziomie istotności 0.05, zweryfikować hipotezę, że zmienna sex nie jest istotna w przyjętym modelu

2.2.2.2 Rozwiązanie

2.2.2.2.1 Wald P value dla testu największej wiarygodności wynosi:

```
##      Pr(>|z|)
## 0.0008609901
```

Widać, że $p < 0.05$ Odrzucamy

$$H_0$$

, więc uznajemy zmienną sex za istotną

2.2.2.2.2 IW P value dla testu największej wiarygodności wynosi:

```
##      Pr(>|Chi|)
## 1
## 2 0.0006286 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Widać, że $p < 0.05$ Odrzucamy

$$H_0$$

, więc uznajemy zmienną sex za istotną

2.2.3 c)

2.2.3.1 Treść Korzystając z testu ilorazu wiarygodności, na poziomie istotności 0.05, zweryfikować hipotezę, że rozkład czasu życia jest taki sam dla pacjenta o sprawności ECOG wg. lekarza na poziomie 0, 1, 2, 3 (zmienna ph.ecog teoretycznie może przyjmować wartości 0, 1, 2, 3, 4, 5, ale w zbiorze danych mamy tylko wartości od 0 do 3)

2.2.3.2 Rozwiązanie P value dla testu największej wiarygodności wynosi:

```
##      Pr(>|Chi|)
## 1
## 2 0.003648 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

W teście ilorazu wiarygodności otrzymano $p < 0.05$, więc odrzucamy hipotezę

$$H_0$$

, że rozkład czasu życia jest taki sam dla pacjentów z ECOG 0,1,2,3. Oznacza to, że zmienna ph.ecog jest istotna w przyjętym modelu

3 Zadanie 3

3.1 Treść

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, wykonać poniższe zadania.

3.2 Rozwiązanie

3.2.1 a)

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

3.2.1.2 Rozwiązanie Korzystając z metody eliminacji wybierzemy model przyspieszonego czasu awarii

KROK 1

zmienna	p_value
age	0.2014
sex	0.0006
ph.ecog	0.0021
ph.karno	0.1332

Usuwamy age bo ma największe $p > 0.05$

KROK 2

zmienna	p_value
sex	0.0007
ph.ecog	0.0020
ph.karno	0.1756

Usuwamy zmienną ph.karno bo ma największe $p > 0.05$

KROK 3

zmienna	p_value
sex	0.0011
ph.ecog	0.0004

Dla teraz wszystkie $p < 0.05$ więc mamy już gotowy model AIC. W modelu końcowym zakłada się, że jedynymi znaczącymi zmiennymi są sex i ph.ecog

3.2.2 b)

3.2.2.1 Treść Korzystając z kryterium informacyjnego Akaike'a (AIC), dokonać wyboru najlepszego modelu przyspieszonego czasu awarii, korzystając z funkcji step.

3.2.2.2 Rozwiązanie

```
## Start:  AIC=2266.6
## Surv(time, status) ~ age + as.factor(sex) + as.factor(ph.ecog) +
##      ph.karno
##
##              Df      AIC
## - age              1 2266.2
## <none>              2266.6
## - ph.karno         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
##
##              Df      AIC
## - ph.karno         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
```

KROK 1 Z funkcji step widzimy, że wskaźnik AIC był najmniejszy dla zmiennej ph.karno, więc została ona usunięta.

KROK 2 Z funkcji step widzimy, że wskaźnik AIC był najmniejszy dla zmiennej age, więc została ona usunięta.

KROK 3 Z funkcji step widzimy, że wskaźnik AIC zwiększa się w każdym możliwym kroku, więc nie ma sensu usuwać już żadnych zmiennych.

3.2.3 c)

3.2.3.1 Treść Korzystając z bayesowskiego kryterium informacyjnego (BIC), dokonać wyboru najlepszego modelu przyspieszonego czasu awarii, korzystając z funkcji step.

3.2.3.2 Rozwiązanie

```
## Start:  AIC=2291.35
## Surv(time, status) ~ age + as.factor(sex) + as.factor(ph.ecog) +
##      ph.karno
##
##
##      Df      AIC
## - age      1 2287.9
## - ph.karno  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
##
##
##      Df      AIC
## - ph.karno      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
```

KROK 1 Z funkcji step widzimy, że wskaźnik BIC był najmniejszy dla zmiennej ph.karno, więc została ona usunięta.

KROK 2 Z funkcji step widzimy, że wskaźnik BIC był najmniejszy dla zmiennej age, więc została ona usunięta.

KROK 3 Z funkcji step widzimy, że wskaźnik BIC zwiększa się w każdym możliwym kroku, więc nie ma sensu usuwać już żadnych zmiennych.

Nasz końcowy model zawiera zmienne sex oraz ph.ecog ponieważ możemy uznać je za istotnie znaczące.

4 Zadanie 4

4.1 Treść

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

4.2 Rozwiązanie

4.2.1 a)

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

4.2.1.2 Rozwiązanie

	age	sex	ph.ecog	ph.karno
P_value	0.1803751	0.0006286	0.0036483	0.1885923

Usuujemy ph.karno bo ma największe $p > 0.05$

	age	sex	ph.ecog
P_value	0.2343418	0.0010293	0.0010004

Usuujemy zmienną age bo ma największe $p > 0.05$

	sex	ph.ecog
P_value	0.001033	0.000354

Teraz wszystkie $p < 0.05$ więc mamy już gotowy model AIC.

W modelu końcowym zakłada się, że jedynymi znaczącymi zmiennymi są sex i ph.ecog

4.2.2 b)

4.2.2.1 Treść Korzystając z kryterium informacyjnego Akaike'a (AIC), dokonać wyboru najlepszego modelu przyspieszonego czasu awarii, korzystając z funkcji step.

4.2.2.2 Rozwiązanie

```
## Start:  AIC=1458.54
## Surv(time, status) ~ age + as.factor(sex) + as.factor(ph.ecog) +
##      ph.karno
##
##              Df      AIC
## - ph.karno      1 1458.3
## - age           1 1458.3
## <none>          1458.5
## - as.factor(ph.ecog) 3 1466.0
## - as.factor(sex)     1 1468.2
##
## Step:  AIC=1458.27
## Surv(time, status) ~ age + as.factor(sex) + as.factor(ph.ecog)
##
##              Df      AIC
## - age           1 1457.7
## <none>          1458.3
## - as.factor(sex)  1 1467.0
## - as.factor(ph.ecog) 3 1468.5
##
## Step:  AIC=1457.68
## Surv(time, status) ~ as.factor(sex) + as.factor(ph.ecog)
##
##              Df      AIC
## <none>          1457.7
## - as.factor(sex)  1 1466.5
## - as.factor(ph.ecog) 3 1470.1
```

KROK 1 Z funkcji step widzimy, że wskaźnik AIC był najmniejszy dla zmiennej ph.karno, więc została ona usunięta.

KROK 2 Z funkcji step widzimy, że wskaźnik AIC był najmniejszy dla zmiennej age, więc została ona usunięta.

KROK 3 Z funkcji step widzimy, że wskaźnik AIC zwiększa się w każdym możliwym kroku, więc nie ma sensu usuwać już żadnych zmiennych. Nasz końcowy model zawiera zmienne sex oraz ph.ecog ponieważ możemy uznać je za istotnie znaczące.

4.2.3 c)

4.2.3.1 Treść Korzystając z bayesowskiego kryterium informacyjnego (BIC), dokonać wyboru najlepszego modelu przyspieszonego czasu awarii, korzystając z funkcji step.

4.2.3.2 Rozwiązanie

```
## Start:  AIC=1479.06
## Surv(time, status) ~ age + as.factor(sex) + as.factor(ph.ecog) +
##      ph.karno
##
##              Df      AIC
## - ph.karno      1 1475.4
## - age           1 1475.4
## - as.factor(ph.ecog) 3 1476.3
## <none>           1479.1
## - as.factor(sex)  1 1485.3
##
## Step:  AIC=1475.37
## Surv(time, status) ~ age + as.factor(sex) + as.factor(ph.ecog)
##
##              Df      AIC
## - age           1 1471.4
## <none>           1475.4
## - as.factor(ph.ecog) 3 1475.4
## - as.factor(sex)  1 1480.7
##
## Step:  AIC=1471.36
## Surv(time, status) ~ as.factor(sex) + as.factor(ph.ecog)
##
##              Df      AIC
## <none>           1471.4
## - as.factor(ph.ecog) 3 1473.6
## - as.factor(sex)  1 1476.7
```

KROK 1 Z funkcji step widzimy, że wskaźnik BIC był najmniejszy dla zmiennej ph.karno, więc została ona usunięta.

KROK 2 Z funkcji step widzimy, że wskaźnik BIC był najmniejszy dla zmiennej age, więc została ona usunięta.

KROK 3 Z funkcji step widzimy, że wskaźnik BIC zwiększa się w każdym możliwym kroku, więc nie ma sensu usuwać już żadnych zmiennych.

Nasz końcowy model zawiera zmienne sex oraz ph.ecog ponieważ możemy uznać je za istotnie znaczące.