Extra exercises in STK4080 2021

Solution Exercise E3.1

Time to discharge (days):

No pneumonia at admission (x = 0): 2, 3+, 6, 6, 10, 11, 12+, 23

Pneumonia at admission (x = 1): 4+, 9, 12+, 17, 24, 26+, 32

It was decided to analyze the data by a Cox regression model with the single covariate x defined above, representing the status of pneumonia at admission.

a) Write down an expression for the hazard function of a patient with pneumonia status x.

Solution: $\alpha(t|x) = \alpha_0(t)e^{\beta x}$

b) Write down the factors of the partial likelihood corresponding to the observed discharge times 2 and 6. You may as well write down the full partial likelihood, in which case you can use R or another package to plot the partial likelihood as a function of β.

Solution: The event times T_j are: 2, 6, 9, 10, 11, 17, 23, 24, 32

If we use the Breslow method (see (c) below), we get the contributions

$$\frac{1}{8+7e^{\beta}} \quad \text{and} \quad \frac{1}{6+6e^{\beta}} \cdot \frac{1}{6+6e^{\beta}}$$

from times 2 and 6, respectively.

You may load the data into R by the command

```
pneu=read.table("https://folk.ntnu.no/bo/STK4080/ex3.txt",header=T)
```

c) Estimate β using R (see below for commands).

Use the following commands:

```
library(survival)
fit.pneu=coxph(Surv(time,cens==1)~x, method="breslow", data=pneu)
summary(fit.pneu)
```

Output:

```
Call:
coxph(formula = Surv(time, cens == 1) ~ x, data = pneu, method = "breslow")

  n= 15, number of events= 10

     coef exp(coef) se(coef)      z Pr(>|z|)
x -1.5971    0.2025   0.8395 -1.902   0.0571 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

  exp(coef) exp(-coef) lower .95 upper .95
x    0.2025      4.938   0.03907      1.05

Concordance= 0.701  (se = 0.104 )
Rsquare= 0.249   (max possible= 0.913 )
Likelihood ratio test= 4.29  on 1 df,   p=0.03843
Wald test            = 3.62  on 1 df,   p=0.05713
Score (logrank) test = 4.3  on 1 df,   p=0.03801
```

d) Is there a significant difference between the discharge times for patients without and with pneumonia at admission? Formulate this problem as a testing problem regarding β, and derive the conclusion when the significance level is set to 5%.

Solution: Test $H_0: \beta = 0$ *vs.* $\beta \neq 0$. *p-values for different tests are given in the R-output. For example, the likelihood ratio test leads to rejection at 5% significance level. Note, however, that the data set is small, so asymptotic results may be inaccurate.*

Compute the estimate of the relative risk of a patient without pneumonia as compared to a patient with pneumonia (see subpoint a)). What is the practical interpretation of this number in the current situation?

Solution: The relative risk of 'withouot' compared to 'with' is $1/$***exp(coef)*** $= 1/0.2025 = 4.94$. *This means that the rate of discharge for a non-pneumonia patient is approximately 5 times the rate for a patient with pneumonia at admission.*

e) Discuss the difference between the test for $H_0 : \beta = 0$ as considered above, and a logrank test for the equality of the hazard functions of the two groups (with or without pneumonia at admission).

Solution: The logrank test is a test for the apparently more flexible hypothesis $\alpha_0(t) = \alpha_1(t)$ *in a model where the two hazard functions can vary freely. As will be seen in the exercise below, the logrank test is however the same as the score test for* $H_0 : \beta = 0$.

Perform the logrank test by hand calculation, using the given data and the setup at page 14/32 of Slides 10. [*Hint:* the computed expected number of discharges under the null hypothesis should be, respectively, 3.20 and 6.80 for the patients without and with pneumonia at admission.]

Solution: The handwritten table below shows how to calculate E_0 and E_1. See also page 14/32 in Slides 10, which is given below the table.

Time	Risk 0	Risk 1	Risk	Fail 0	Fail 1	Fail	E0	E1
2	8	7	15	1	0	1	$8 \cdot \frac{1}{15}$	$7 \cdot \frac{1}{15}$
6	6	6	12	2	0	2	$6 \cdot \frac{2}{12}$	$6 \cdot \frac{2}{12}$
9	4	6	10	0	1	1	$4 \cdot \frac{1}{10}$	$6 \cdot \frac{1}{10}$
10	4	5	9	1	0	1	$4 \cdot \frac{1}{9}$	$5 \cdot \frac{1}{9}$
11	3	5	8	1	0	1	$3 \cdot \frac{1}{8}$	$5 \cdot \frac{1}{8}$
17	1	4	5	0	1	1	$1 \cdot \frac{1}{5}$	$4 \cdot \frac{1}{5}$
23	1	3	4	1	0	1	$1 \cdot \frac{1}{4}$	$3 \cdot \frac{1}{4}$
24	0	3	3	0	1	1	0	$3 \cdot \frac{1}{3}$
32	0	1	1	0	1	1	0	$1 \cdot 1$
							3.20	6.80

Hand-calculation of log-rank test

$O_1 - E_1 = N_1(t_0) - \int_0^{t_0} \frac{Y_1(t)}{Y_\bullet(t)} dN_\bullet(t), \qquad V_{11} = \int_0^{t_0} \frac{Y_1(t)Y_2(t)}{Y_\bullet(t)^2} dN_\bullet(t)$

Go through all *failure times* $T_1, \cdots, T_r$:

	Group 1	Group 2	Total at T_j
# at risk at T_j	Y_{1j}	Y_{2j}	Y_j
Observed # fail at T_j	O_{1j}	O_{2j}	O_j
Est prob of fail under H_0			$\frac{O_j}{Y_j}$
Estim expect # failures	$E_{1j} = Y_{1j} \cdot \frac{O_j}{Y_j}$	$E_{2j} = Y_{2j} \cdot \frac{O_j}{Y_j}$	
Estimated variance			$V_j = \frac{Y_{1j} Y_{2j} O_j}{Y_j^2}$

Then sum over all failure times $T_1, \cdots, T_r$:

$$O_h = \sum_{j=1}^{r} O_{hj}, \quad E_h = \sum_{j=1}^{r} E_{hj} \text{ for } h = 1, 2, \quad \text{and } V_{11} = \sum_{j=1}^{r} V_j$$

Test statistics are then

$$\frac{(O_1 - E_1)^2}{V_{11}} \text{ or the conservative } \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

f) Perform the logrank test also by using R, for example using the command

```
survdiff(Surv(time,cens)~x, data=pneu)
```

Output:

```
Call:
survdiff(formula = Surv(time, cens) ~ x, data = pneu)

    N Observed Expected (O-E)^2/E (O-E)^2/V
x=0 8        6      3.2      2.44      4.42
x=1 7        4      6.8      1.15      4.42

 Chisq= 4.4  on 1 degrees of freedom, p= 0.0356
```

Calculate and compare the statistics

$$\frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \quad \text{and} \quad \frac{(O_1 - E_1)^2}{V_{11}}$$

Give a comment on the difference.

Solution:

$$\frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} = 2.44 + 1.15 = 3.59 \ < \ \frac{(O_1 - E_1)^2}{V_{11}} = 4.42$$

The inequality here holds in general, so that a test using the left hand side is conservative (and would not lead to rejection here at 5% significance level when using the asymptotic critical values).

How does the conclusion of the logrank test fit with the conclusion of the corresponding problem in subpoint d)? Give a comment. Note the conclusion of Exercise 4.2 in the book.

Solution: The p-value of the logrank test is close to the one of the score test in (d). This is as it should be by the next exercise.

Exercise E3.2

Do Exercise 4.2 p. 203 in the book. You may use the following hint:

Assume for simplicity that there are no ties. Consider the event times $T_1 < T_2 < \ldots$ where one of the two processes jumps. At T_j, define

$$\begin{aligned} x_j &= \text{value of } x \text{ (0 or 1) for the individual failing at } T_j \\ y_{j0} &= \text{the number of individuals in the risk set } \mathcal{R}_j \text{ having } x = 0 \\ y_{j1} &= \text{the number of individuals in the risk set } \mathcal{R}_j \text{ having } x = 1 \end{aligned}$$

Show that the partial likelihood can be written

$$L(\beta) = \prod_j \frac{e^{\beta x_j}}{y_{j0} + y_{j1}e^{\beta}}$$

Solution: This is easily seen by inspection (see also previous exercise).

Then calculate the score test statistic

$$\frac{U(0)^2}{I(0)}$$

where.

$$\begin{aligned} U(\beta) &= \frac{\partial \log L(\beta)}{\partial \beta} \\ I(\beta) &= -\frac{\partial U(\beta)}{\partial \beta} \end{aligned}$$

Show that the score statistic is exactly equal to the logrank statistic.

Solution: Let

$$\ell(\beta) = \log L(\beta) = \sum_j \{\beta x_j - \log(y_{j0} + y_{j1}e^{\beta})\}$$

Then

$$U(\beta) = \ell'(\beta) = \sum_j \left(x_j - \frac{y_{j1}e^{\beta}}{y_{j0} + y_{j1}e^{\beta}} \right)$$

Next,

$$I(\beta) = -U'(\beta) = \sum_j \frac{y_{j0}y_{j1}e^{\beta}}{(y_{j0} + y_{j1}e^{\beta})^2}$$

But now

$$U(0) = \sum_j \left(x_j - \frac{y_{j1}}{y_{j0} + y_{j1}} \right) = O_2 - E_2$$

and

$$I(0) = \sum_j \frac{y_{j0}y_{j1}}{(y_{j0} + y_{j1})^2} = V_{11}$$

(see for example the page from slides 10 displayed above). But then the test statistic $U(0)^2/I(0)$ is exactly equal to the logrank statistic.