Extra exercises in STK4080 2021

Exercise E3.1

At an intensive care unit at a hospital one is interested in whether the presence or no-presence of pneumonia for a patient at admission will have an impact on the length of hospital stay, T, i.e., time from entry to discharge from the hospital unit.

Below are times (in days) to discharge from the hospital for 8 patients *without* pneumonia at admission ($x = 0$), and 7 patients *with* pneumonia at admission ($x = 1$).

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

Here + means a right censored observation. (Right censoring occurred in the cases when a patient was still in hospital at the end of the study, or died in hospital.)

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. [*Answer:* $\alpha(t|x) = \alpha_0(t)e^{\beta x}$]

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.

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).

Note the tied events at time 6. The simplest approach here is to use Breslow's approximation. This method considers multiple event times as different events, but with the same risk set. Thus each failure at a certain tied time produces one factor in the partial likelihood, with all having the same risk set (i.e., using the same denominator in these factors). You will have to write `method="breslow"` (see below), since the default method in R is slightly different (`method="efron"`)

You may use the following commands:

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

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%.

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?

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).

Perform the logrank test (conservative version) 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.]

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

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

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.

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 ABG. *This exercise is given with hints as the next exercise below!*

Exercise E3.2

Do Exercise 4.2 p. 203 in ABG. 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

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

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.