

Project Exam for STK4080/9080

Fall semester 2012

The exam in STK4080/9080 consists of this project exam and a written exam.

The written solution to the project exam (in Norwegian or English) must be handed in or sent by email to Ørnulf Borgan (borgan@math.uio.no) no later than Thursday 29 November.

You are not allowed to collaborate with others on the project exam.

The written exam takes place Tuesday 11 December.

Details are posted on the course web-page.



Primary biliary cirrhosis (PBC) is a rare but serious liver disease of unknown origin. Between January 1974 and May 1984, 312 PBC-patients were included in a double-blind randomized study at the Mayo Clinic in the USA, comparing D-penicillamine with placebo. A number of covariates were registered for each patient at the entry to the study. In addition to treatment, we will in this assignment only consider the two demographic covariates age and sex, and the two biochemical covariates albumin (the most abundant blood plasma protein) and bilirubin. Bilirubin is a red bile pigment derived from the degradation of hemoglobin during the destruction of red blood cells. A large concentration of bilirubin may be a sign of bad liver function.

You may read the data into R by the command:

```
pbcr=
read.table("http://www.uio.no/studier/emner/matnat/math/STK4080/h12/pbc.txt",header=T)
```

The data are organized with one line for each of the 312 patients, and with the following variables in the seven columns:

- **days**: number of days from randomization to death/censoring
- **status**: indicator for death (1) or censoring (0)
- **treat**: treatment (1=D-penicillamine, 2=placebo)
- **age**: age at randomization (in years)
- **sex**: sex (0=male, 1=female)
- **bil**: concentration of bilirubin in blood plasma (in mg/dl)
- **alb**: concentration of serum albumin (in g/dl)

The purpose of the assignment is to determine the importance of the covariates treatment, age, sex, albumin and bilirubin on the mortality of PBC-patients.

The analysis consists of three main parts:

a) Simple univariate analyses:

First you perform simple univariate analyses for one covariate at a time by means of Kaplan-Meier plots and log rank tests. For the numeric covariates age, albumin, and bilirubin you have to choose a reasonable grouping.

b) Univariate regression:

Then you perform univariate Cox regression analyses for each of the covariates. In this connection you should in particular consider how the numeric covariates should be coded.

c) Multivariate regression:

Finally you perform a multivariate Cox regression analysis where the importance of the covariates is studied simultaneously. In this analysis you should among other things:

- (i) decide which covariates are of importance for mortality,
- (ii) decide whether the effect of treatment (or other covariates) depends on the values of the other covariates (interaction),
- (iii) check model assumptions.

The results of your analyses should be written up in a report. In this report you should describe the problem to be investigated and how the statistical analyses have been performed. You should also describe, interpret, and discuss the results you arrive at. In particular you should make a careful interpretation of your “final model” in part c). You should not include R code in the main body of the report. But the R code you use to arrive at your results should be given in an appendix. (Only include code for results that are presented in your report.)