DEPARTMENT OF STATISTICS

Paper 475.330 **Assignment 2** **Due: Thursday 15, April**

The following data represent a random sample of 79 patients that underwent a particular type of liver surgery. This is an observational study in which the response is the survival time of the patient. Prior to the surgery, data were obtained on four variables that were thought to be possible predictors of survival time:

clot	A score that measures blood clotting
prog	A prognostic index, that includes age
enz	An enzyme function test score
liv	A liver function test score

A data frame containing this data called liver.df has been created in Splus on the advanced lab server. To access this data simply logon to the advlab computer, start Splus and type

```
> liver.df
   clot prog enz  liv time
 1  6.7   62  81 2.59  200
 2  5.1   59  66 1.70  101
 3  7.4   57  83 2.16  204
 4  6.5   73  41 2.01  101
 5  7.8   65 115 4.30  509
 6  5.8   38  72 1.42   80
 7  5.7   46  63 1.91   80
 .   .    .   .   .    .
 .   .    .   .   .    .
 .   .    .   .   .    .
```

You are asked to investigate how the four possible predictors are related to survival time and to each other. Your ultimate goal is to identify a model that can be used to predict survival times for future patients. You should discuss the validity of your chosen model and evaluate its ability to make such predictions accurately.

As usual, your assignment should consist of two main components. The first part should be a report that is suitable to be read by a hospital administrator. It should explain your results in non-technical terms. The second part will be a statistical appendix where you explain how you analysed the data and why you chose the model you did. You should include and comment on the diagnostic procedures you used to check the validity of your fitted model.