**Final Project**

**Instructions:** In this final project, you conduct a comprehensive analysis of a survival data using statistical methodologies learned in ST525. The final project is a great way to solidify your understanding of statistical methods for survival data.

* It is recommended to work in groups of two to three people on the project.
* There will be one grade for the entire group. So everyone in your group is expected to participate in the effort.
* Each group need to submit a final report. Aim for no more than five pages of text, plus tables and figures. The report should start with a short **Introduction** stating the problem and end with a brief **Discussion** of your results. The bulk of the report should be a **Methods and Result** section in which you describe your analyses and your findings.
* Please also include your SAS code as an **Appendix**.
* The PBC3 data given below is an example of data set with expectations for statistical analyses that should be included in the final project.
* You can use the PBC3 data for your final project. But it is recommended to find your own survival data that is interesting to you and your group. And please make sure that it is a different data set from the examples we discussed in the lectures.

**An example of data set:** PBC3 was a multi-centre randomized clinical trial conducted in six European hospitals. Between 1 Jan. 1983 and 1 Jan. 1987, 349 patients with the liver disease primary biliary cirrhosis (PBC) were randomized to either treatment with Cyclosporin A (CyA, 176 patients) or placebo (173 patients). The purpose of the trial was to study the effect of treatment on the survival time. However, during the course of the trial an increased use of liver transplantation for patients with this disease made the investigators redefine the main response variable to be time to “failure of medical treatment” defined as either death or liver transplantation (Lombard et al., 1993). Patients were followed from randomization until treatment failure, drop-out or 1 Jan, 1989. A total of 61 patients died (CyA: 30, placebo: 31), another 29 were transplanted (CyA:14, placebo: 15) and 4 patients were lost to follow-up before 1 Jan. 1989. At entry, a number of clinical, biochemical and histological variables, including serum bilirubin, serum albumin, sex, age were recorded:

* ptno: patient identification
* unit: hospital (1: Hvidovre 2: London 3: Copenhagen 4: Barcelona 5: Munich 6: Lyon)
* tment: treatment (0: placebo, 1: CyA)
* sex: (1: males, 0: females)
* age: years
* stage: histological stage (1, 2, 3, 4)
* gibleed: previous gastrointestinal bleeding (1: yes, 0: no)
* crea creatinine (micromoles/L)
* alb: albumin (g/L)
* bili: bilirubin (micromoles/L)
* alkph: alkaline phosphatase (IU/L)
* asptr: aspartate transaminase (IU/L)
* weight: body weight (kg)
* days: observation time (days)
* status: status at exit (0: censored, 1: liver transplantation, 2 : dead)

**Analysis expectations:** The main goal is to evaluate the effect of treatment on time to failure of medical treatment, while controlling for patient’s other characteristics. Your analysis should include a description of the data set; univariate analyses for each covariate; construction of a “best” multiple-variable regression model to predict time to failure of medical treatment. The assumptions of the model must be suitably checked. In Discussion section, please also comment on pros and cons of your analysis. For example, the study was a multi-center study and observations in the same hospital are likely to be correlated, and your regression model may fail to account for this correlation.