VA/UW Summer Session: **Exercise Set #3** Day 3: June 25, 2003

Reading: • Kleinbaum *Survival Analysis*, Chapter 4
◦ Kleinbaum *Survival Analysis*, Chapter 5 (optional)

NOTE: Unless explicitly stated, direct computer output is **not** desired. Typically only part of the computer output is asked for (such as a confidence interval) and then proper interpretation of the statistics is requested.

DATA: The data for these exercises can be found on the class web page:
`http://faculty.washington.edu/heagerty/Courses/VA-summer-epi/` in the *Homeworks* directory (ie. click on Homeworks from the main summer course 513 page).

Survival Analysis: Model Checking

1. A subset of the Mayo PBC data is on the web as `mayo_sub.dat`. This contains the 5 variables that were used in the final Mayo model as well as the survival time and status (also included is stage, but this doesn't appear to add to the prognostic potential – you can check this if you wish). Our question is: Do these data appear to satisfy the PH assumption? Note that formal methods (tests) were not fully developed until 1994, and the reported analysis using the Cox model appeared in 1989.

(a) Fit the Mayo model and then assess whether the variables appear to satisfy the PH assumption. Specifically, test the PH assumption for each variable. Interpret these tests.

(b) For the variable (or variables) that are suggested to poorly satisfy the PH assumption divide them into 3 groups and plot $\log(-\log(\widehat{S}(t)))$ versus time. Interpret what this plot suggests about whether the PH assumption is satisfied for the variable. Turn this plot in.

(c) Fit the Cox model with the 5 Mayo model variables and plot the Schoenfeld residual versus time for each variable. Use the smooth curve to help visualize trends. Interpret these plots with respect to whether the PH assumption appears to be violated. Turn these plots in.

(optional) PBC legend has it that there is an observation which is an entry error (ie. the value is wrong!), and that it has a large influence on one coefficient estimate. Create delta-beta's and plot these influence statistics against either time and/or the predictor variable that they correspond to. Interpret these plots. Can you identify the error? (Note: see the web page to obtain code to calculate the delta-beta's).

Survival Analysis: Stratified Analysis

2. The file `addicts.dat` on the class web page contains data regarding the time that heroin addicts remain in methadone treatment. In the lecture notes we found that the variable `clinic` did not satisfy the PH assumption and we were able to make inference on other predictor variables by using `clinic` as a stratifying variable.

These data were analyzed by Caplehorn and Bell (1991) who were interested in factors associated with retaining subjects: "As methadone maintenance is of proven benefit only to those in treatment, retention in treatment is an important measure of the effectiveness of treatment programmes." and "To elucidate the reasons that programmes fail to retain patients, we have studied the relationship between the maximum daily dose and retention in a cohort of addicts." Scientific interest is in whether factors other than dose can be used to identify subjects at high risk for failing to be retained.

(a) Calculate bivariate summaries for each of the predictor variables and their association with time retained in treatment. Summarize these by creating a table of hazard ratios and 95% CI's for each variable when it is the single predictor in a Cox regression that uses clinic as a stratifying variable.

(b) Calculate a Cox regression model using all of the predictors. Summarize the results by creating a single table of regression parameters (or hazard ratios) – use the computer output as directly as possible in order to create this table. (again stratify on clinic)

(c) Describe the assumptions in your Cox regression model, in particular what it means to use clinic as a stratifying variable.

(d) Are there other variables besides `dose` that appear to be predictive of retention failure? Summarize the results of your analysis.

(e) Do the other covariates appear to satisfy the PH assumption? Justify your conclusion.