BIOST/EPI 357
SURVIVAL DATA ANALYSIS IN EPIDEMIOLOGY

Winter 2020
Instructor: Marco Carone

PROBLEM SET 4
due on March 9th, 2020 at 11:59pm

PROBLEM 1.

In this problem, you will once more revisit the dataset on methadone maintenance programs for heroin addicts you studied in Problem Sets 1, 2 and 3. Please refer to Problem Set 1 for a description of the relevant variables.

As indicated before, Caplehorn and Bell (1991) provided an analysis of these data, seeking to identify factors favoring the retention of subjects because maintenance is known to be effective only in patients who remain in treatment. They were particularly interested in determining whether daily dosage is related to the probability of retention, and whether other factors can be used to identify subjects at high risk for failing to be retained.

(a) Fit a Weibull accelerated failure time (AFT) model to estimate the association between risk of exit from maintenance and methadone dosage adjusting for both clinic and history of previous incarceration as dummy variables. Provide an estimate and 95% confidence interval for each of the resulting time ratios. Do your findings agree qualitatively with output from a corresponding Cox model? By what percentage does each time ratio estimate change if you use a generalized gamma baseline distribution instead?

(b) Based on the Weibull AFT model fitted in (a), provide a model-based estimate and 95% confidence interval of the ratio of mean time until exit from maintenance comparing individuals from clinic 2 without a history of incarceration administered dosage of 40 mg/day to individuals from clinic 1 with a history of incarceration administered a dosage of 100 mg/day. Provide a model-based estimate and 95% confidence interval of the median time until exit from maintenance for each of these two subpopulations of individuals.

(HINT: To answer the latter part, it will be useful to revisit your notes from Chapter 2 on expressing the median of a Weibull distribution in terms of the parameter values outputted by R or STATA.)

(c) Fit another Weibull AFT model to estimate the association between risk of exit from maintenance and methadone dosage adjusting for both clinic and history of previous incarceration as dummy variables and allowing this association to differ based on history of previous incarceration. Provide an estimate and 95% confidence interval for the time ratio comparing, among patients with a history of previous incarceration and from the same clinic, patients administered 80mg/day to patients administered 60mg/day. Do the same among patients without a history of previous incarceration and from the same clinic. Are these two subgroup-specific time ratios significantly different from one another?

PROBLEM 2.

Diabetic retinopathy (DR) is a complication of diabetes that affects retinal blood vessels and can lead to severe vision loss. A total of 83 high-risk adult patients (age $\geq$ 20) with DR in both eyes and visual acuity of at least 20/100 were recruited into a study at their diagnosis of DR. For each patient, an eye (left or right) was chosen at random and was subjected to laser photocoagulation therapy; the other eye served as a control. All patients were followed until onset of blindness or end of follow-up. Data from this study can be found in `diabetes.csv`.

We are interested in describing the distribution of age at blindness due to diabetic retinopathy in the absence of treatment in this high-risk patient population.

(a) Estimate nonparametrically the distribution of age at blindness due to diabetic retinopathy **using data from untreated eyes only** accounting for delayed entry at age at diagnosis. Plot the resulting survival function estimate. Provide a point estimate and 95% confidence interval for the median age at blindness due to diabetic retinopathy. Repeat the exercise but ignore any left-truncation present.

(b) Do the results in (a) appear sensible? Do you see cause for concern? How plausible is the independence assumption required for valid delayed entry adjustment?