PROBLEM SET 3
due on February 26th, 2020 at 11:59pm

In all problems below, please do not simply provide a numerical answer: interpret the estimand(s) involved and summarize your findings in language appropriate for a general scientific audience.

PROBLEM 1.

In the CCG803 study, 268 children in remission of acute lymphoblastic leukemia were recruited from a number of clinical institutions and randomized to one of two maintenance regimes. Patients in the control group were assigned to the standard of care - chemotherapy with 6MP and methotrexate - while patients in the treatment group were also given dactinomycin.

Dactinomycin administration (`rx`), age at baseline (`age`), white blood cell count (`wbc`), clinical institution (`institution`), observed follow-up time (`duration`) and relapse status at end of follow-up (`relapse`) were recorded for each patient.

(a) Describe and fit a proportional hazards model to estimate the association between risk of relapse and treatment. Provide an estimate and 95% confidence interval for the hazard ratio.

(b) Describe and fit a proportional hazards model to estimate the association between risk of relapse and treatment adjusting for white blood cell count and age. Provide an estimate and 95% confidence interval for each of the resulting hazard ratios.

(c) Describe and fit a proportional hazards model to determine whether the association between risk of relapse and treatment adjusting for white blood cell count and age differs in subpopulations of patients defined by white blood cell count being either below 10,000 (`wbc` < 100), or above or at 10,000 (`wbc` ≥ 100). Provide an estimate and 95% confidence interval for the hazard ratio corresponding to each of these subpopulations. Perform a test of the hypothesis that this association does not differ in these subpopulations.

(d) Describe and fit a proportional hazards model to estimate the association between risk of relapse and treatment adjusting for white blood cell count, age and recruitment site. Provide an estimate and 95% confidence interval for each of the resulting hazard ratios.

(e) Based on the proportional hazards model you fitted in (b), display on a single graph estimates of the relapse-free survival curves for the subpopulation of

  i.   5 year-old treated patients with `wbc` = 45;
  ii. 5 year-old control patients with `wbc` = 45;
  iii. 5 year-old treated patients with `wbc` = 210;
  iv. 5 year-old control patients with `wbc` = 210.

PROBLEM 2.

In this problem, you will once more revisit the dataset on methadone maintenance programs for heroin addicts you studied in Problem Sets 1 and 2. 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)   Describe and fit a proportional hazards model to estimate the association between risk of exit from maintenance and methadone dosage adjusting for history of incarceration and clinic as regular predictors. Provide an estimate and 95% confidence interval for each of the resulting hazard ratios.

In the following questions, use stratified proportional hazards model with clinic as stratifying variable.

(b)   Describe and fit a proportional hazards model to estimate the association between risk of exit from maintenance and methadone dosage adjusting for clinic via stratification as well as adjusting for history of previous incarceration. Provide an estimate and 95% confidence interval for each of the resulting hazard ratios. Discuss the implication of using clinic as a stratifying variable. Has the interpretation of these hazard ratios changed relative to the model you fitted in (a)?

(c)   Describe and fit a proportional hazards model to determine whether the association between methadone dosage and risk of exit from maintenance adjusting for history of incarceration and clinic is different in those with and without a history of incarceration? Provide an estimate and 95% confidence interval for any summary that allows you to answer this question. Based on this model, provide an estimate and 95% confidence interval for the hazard ratio comparing patients from a given clinic with a history of incarceration receiving 100 mg/day of methadone to patients from the same clinic without a history of incarceration receiving 40 mg/day of methadone.