BST 665 - Survival Analysis
Homework 4

Assigned: April 3, 2019
Due: April 10, 2019, by end of class

Instructions:
Review Chapter 9 of *Applied Survival Analysis*. Then complete the exercises below. Responses should be typed or printed neatly (if you have multiple pages, please staple them). Unless specified otherwise, you are encouraged to use software whenever possible to create plots and perform calculations. To receive full credit, you must show your work. It is not sufficient to circle a result on the SAS output or to simply write down a numerical answer without an explanation.

Exercises:

1. An investigator is studying infections in patients receiving dialysis. He believes that the method used to access the blood during dialysis may be associated with the risk of infection.

Download the dataset named "Dialysis.sas7bdat". The variables in this dataset are:

- ID number for each patient (ID)
- The time to infection is represented using the counting process syntax (Start, End)
- A censoring variable for infection (Censor)
    - Censor = 1 if the patient got an infection; Censor = 0 otherwise
- Infection number (Infection)
- Blood access method used (Method)
    - Catheter, Fistula, or Graft
- An indicator for whether the patient has diabetes (Diabetes)

A. Using the time to infection as the survival time, fit a Cox proportional hazards model using diabetes status and access method as covariates. Use the Prentice-Williams-Peterson conditional probability (PWP-CP) model. Create a table reporting the estimated hazard ratios (and 95% confidence intervals) for each of the effects in this model. Provide your interpretations of these results.

B. Repeat Part A, this time using the Prentice-Williams-Peterson gap time (PWP-GT) model.

C. The investigator would like you to explain the difference between the results from Parts A and B. Write a short paragraph explaining the PWP-CP and PWP-GT models. (This explanation should be written so that it can be understood by the investigator, who has not taken a survival analysis course).

D. The investigator is preparing a manuscript on this study and has asked for your help writing the methods and results sections. Write 1-2 paragraphs for the investigator's paper. Be sure to

include your interpretation of any results you include and to take the stated goal of the study into consideration.

2. You are the statistician for a large clinical trial that is being designed to evaluate a new drug for preventing stroke in high-risk patients.

The outcome of interest for this study will be the time to an occurrence of stroke. The planned study will last for a total of ten years, with patients being enrolled for the first four years of the study and then followed for an additional six years. Half of the subjects will be randomized to receive the new drug; the other half will receive the traditional drug.

A. The trial's principal investigator believes that patients taking the new drug will have a 30% reduction in their risk of stroke as compared to those receiving the standard treatment. How many strokes will the PI need to observe in order to have 90% power to detect this effect at significance level $\alpha = 0.05$?

B. Results from a prior study suggest that 12% of patients taking the traditional drug will have a stroke within the first year of being diagnosed as high-risk. Use this finding to estimate the probability that a subject enrolled in the new study will have a stroke.

C. The principal investigator needs to include a sample size justification in her grant application for this trial. Propose a sample size for this trial and write a brief paragraph explaining your choice. Include a figure showing the study power at a range of hazard ratios.

3. A nephrologist is studying the time to disease progression following nephrectomy for patients with renal cell carcinoma. While the nephrologist is primarily interested in progression, he knows that he must also take into account the fact that death is a competing risk for this population.

Download the dataset named "Renal.sas7bdat". The variables in this dataset are:

- ID number for each patient (ID)
- Patient age (Age)
- Type of nephrectomy (Type)
    - Complete or Partial
- Cancer stage (Stage)
    - Stage 1; Stage 2; Stage 3 or 4
- The length of the follow-up period, in years (Years)
- Patient outcome (Outcome) and censoring variable (Censor)
    - 1 = Progression, 2 = Death, 0 = Censored

A. Plot the cumulative incidence function for progression. What proportion of patients will have experienced disease progression at two years after nephrectomy? Provide a 95% confidence interval for this proportion.

B. Plot the cumulative incidence functions for progression and death on the same plot. Include the Kaplan-Meier estimate of survival in the plot. What conclusions can we draw from this figure?

C. Does the cumulative incidence of progression vary by cancer stage? Conduct an appropriate hypothesis test to answer this question. (Make sure you state the null and alternative hypotheses, level of significance used, test statistic, p-value and conclusion in terms of the problem).