

BST 665 - Survival Analysis Homework 3

Assigned: March 9, 2019

Due: March 20, 2019, by end of class

Instructions:

Review Chapters 6 and 7 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. A physician is studying survival following bone marrow transplant in leukemia patients. She has conducted a multicenter trial of 137 patients.

Download the “Bone_marrow.sas7bdat” dataset. The variables in this dataset include:

- Subject ID number (ID)
- The length of the follow-up period, in months (Months)
- Survival at last follow-up (Died)
 - Died = 1 if the patient died; Died = 0 otherwise
- Disease group (Group)
 - acute lymphoblastic leukemia (ALL), acute myelocytic leukemia with low risk of remission (AML Low Risk), or acute myelocytic leukemia with high risk of remission (AML High Risk)
- Patient age (Age)
- Patient sex (Sex)
- Center where patient was treated (Center)
 - Alferd, Hahnemann, Ohio State, or St. Vincent

A. Using the length of the follow-up period as the survival time and the status at last follow-up as the censoring variable, plot the log of the estimated cumulative hazard function for each treatment center. What does this figure tell us?

B. Is it reasonable to assume proportional hazards for treatment center? Why or why not?

[**Note:** In class, we discussed several ways of checking the proportional hazards assumption. For this problem, you should use at least two different methods to assess the proportional hazards assumption for treatment center].

2. The physician from Exercise 1 would like to compare the survival for the three disease groups. Using the Bone_marrow.sas7bdat dataset, fit a stratified Cox proportional hazards model using

the length of the follow-up period as the survival time and the status at last follow-up as the censoring variable. Stratify on treatment center and include effects for age, sex, and disease group in your model. Assume that the effects of each covariate are constant across strata.

A. Write the model you just fit using mathematical notation (i.e., using β coefficients). Be sure to define all variables used in the model, including any dummy variables.

B. Create a table reporting the estimated hazard ratios (and 95% confidence intervals) for each of the effects in this model. This table should include all pairwise hazard ratios comparing the three disease groups.

C. Does the effect of disease group on survival significantly differ by treatment center? Revise your model to answer this question and justify your answer using a hypothesis test. (Make sure you state the null and alternative hypotheses, level of significance used, test statistic, p-value and conclusion in terms of the problem).

D. Write a short summary (1-2 paragraphs) of the results of this study. Be sure to include your interpretation of the results and to take the stated goal of the study into consideration.

3. While recovering from bone marrow transplants, patients have lowered platelet counts and it may take some time for their platelet count to rebound. The physician is also interested in the effect of the platelet recovery process on survival and whether that effect varies by disease group. The Bone_marrow.sas7bdat dataset also contains variables related to platelet recovery:

- An indicator for whether the subject's platelet count recovered (Recovery)
- Time between transplant and platelet recovery, in months (Recovery_Time)
 - Recovery_Time is missing if the patient's platelet count did not recover.

Using the length of the follow-up period as the survival time and the status at last follow-up as the censoring variable, fit a Cox proportional hazards model to assess the effects of disease group and platelet recovery on survival. [Disease group and platelet recovery should be the only effects in your model. For this exercise, do not stratify by treatment center].

A. Report a hazard ratio (and 95% confidence interval) for the effect of platelet recovery. Give your interpretation of this hazard ratio.

B. Does the effect of platelet recovery vary by disease group? Justify your answer using a hypothesis test. (Make sure you state the null and alternative hypotheses, level of significance used, test statistic, p-value and conclusion in terms of the problem).

C. Report hazard ratios and 95% confidence intervals for the effect of platelet recovery for each disease group. Discuss these results.

D. Write a short summary (1-3 paragraphs) of the results of this study. Be sure to include your interpretation of the results and to take the stated goal of the study into consideration. The physician has also asked that, in your summary, you briefly explain the concept of a time-

varying covariate and how it relates to this data set. Include this explanation in your summary. (This explanation should be written so that it can be understood by the physician, who has not taken a survival analysis course).