

## BST 665 - Survival Analysis Homework 2

Assigned: February 13, 2019

Due: February 20, 2019, by end of class

### Instructions:

Review Chapters 3 and 4 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 researcher is interested in studying risk factors for myocardial infarction (MI). Four hundred patients identified to be potentially at risk of MI were studied.

Download the dataset named “MI.sas7bdat”. The variables in this dataset are:

- Subject ID number (ID)
- Body mass index (BMI), in  $\text{kg/m}^2$
- An indicator for whether the subject had atrial fibrillation (AFib)
- Blood pressure (BP), categorized as either “Normal” or “High”
- An indicator for whether the subject had a myocardial infarction during follow-up (MI)
  - MI = 1 if the subject suffered a myocardial infarction; MI = 0 if the subject was censored
- The length of the follow-up period, in days (Days)

Fit a Cox proportional hazards model to assess the effect of atrial fibrillation and blood pressure on time to MI. [Your model should only have main effects for atrial fibrillation and blood pressure]. Use the results of this model to answer the following questions.

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

B. Report a hazard ratio comparing a patient with high blood pressure and atrial fibrillation to a patient with normal blood pressure and no atrial fibrillation. Provide a 95% confidence interval for this hazard ratio.

C. On the same graph, plot the estimated survival functions for four hypothetical subjects:

- A patient with normal BP and no atrial fibrillation
- A patient with high BP and no atrial fibrillation
- A patient with normal BP and atrial fibrillation

- A patient with high BP and with atrial fibrillation

What conclusions can we draw from this graph?

2. Add BMI and an interaction between BMI and atrial fibrillation to the model used in Exercise 1. Use the results of this model to answer the following questions.

A. . Does this model provide a significantly better fit to these data than the model used in Exercises 1-4? Use an appropriate hypothesis test to justify your answer. (Make sure you state the null and alternative hypotheses, level of significance used, test statistic, p-value and conclusion in terms of the problem).

B. How much reduction in risk of death is associated with a 5-unit decrease in BMI for a person without atrial fibrillation? What about for a person with atrial fibrillation? Provide 95% confidence intervals for your answers.

C. Write 2-3 paragraphs summarizing the results of this study. Be sure to include an explanation of the interaction. Provide your interpretation of any results presented. Include figures as necessary.

3. A researcher is interested in identifying risk factors for multiple myeloma. He has fit a Cox proportional hazards model assessing the effects of age, protein level (g/dL), and hemoglobin on time to multiple myeloma diagnosis. Hemoglobin has been divided into three categories: low, medium, and high. Part of his SAS output is shown below.

Type 3 Tests			
Effect	DF	Wald Chi Square	Pr > ChiSq
Age	1	28.89	< 0.0001
Protein	1	4.69	0.0304
Hemoglobin	2	9.63	0.0081

Analysis of Maximum Likelihood Estimates					
Parameter		Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
Age		0.044	0.008	28.89	< 0.0001
Protein		-0.492	0.227	4.69	0.0304
Hemoglobin	High	-0.525	0.196	7.22	0.0072
Hemoglobin	Medium	-0.490	0.200	6.01	0.0142

- A. Calculate the hazard ratio and 95% confidence interval for the hazard ratio associated with a four-unit increase in protein.

B. Calculate all pairwise hazard ratios comparing subjects with high, medium, and low hemoglobin levels. Provide 95% confidence intervals all hazard ratios except the hazard ratio comparing a subject with high hemoglobin to one with medium hemoglobin.

C. Is hemoglobin significantly associated with the risk of multiple myeloma? Support 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. Help the researcher write the results section of his paper. Write 1-2 paragraphs describing the results of these analyses. Be sure to include your interpretation of the results and to take the stated goal of the study into consideration.