BIOS 845 Homework #4

Date assigned: 04/18/2018
Due Date: 05/02/2018 by 11:59 pm (Blackboard clock time);

Instructions:

1. To receive full credit, show all work. Please make your work legible.
2. Total points for this homework are 100.
3. Do not forget to write your name on the homework.
4. Insert page numbers on all pages and also total # of pages submitted.
5. Homework can be typed or hand-written. Provide SAS code wherever necessary.
6. Use the BLACKBOARD drop box to turn in the homework (preferably as pdf).

Question # 1: **30 points**

I. This problem will use the data set LEUKEMIAB.SAS7BDAT that we have previously considered in class. The data set contains the following variables:

RX = Treatment status (=0 if new treatment, =1 if standard)
LOGWBC = log white blood cell count
SEX = gender (=0 if female, =1 if male)
WEEKS = the number of weeks in the study
RELAPSE = indicator of relapse or censoring (=0 if censored, =1 if relapsed)

A. Suppose that you want to examine the effect of RX, adjusting for SEX and LOGWBC. In class, we noted that the proportional hazards assumption appears to be violated with regards to the covariate SEX. Suppose that we wish to fit a model which allows the hazard ratio for SEX to differ before and after 15 weeks. In other words, there is a constant hazard ratio associated with SEX for less than 15 weeks of follow-up and a constant hazard ratio associated with SEX for 15 weeks of follow-up or more, but these two hazard ratios are not necessarily constant. Use SAS to fit this extended Cox model. Notice that there are a large number of tied event times for this data set. For that reason, please run your model for this homework using the TIES=EXACT option. Report a 95% confidence interval for the effect of treatment on time to relapse, controlling for LOGWBC and SEX (and accounting for the fact that SEX does not satisfy the PH assumption).

B. Perform the same analysis using a Complementary log-log proportional hazards model. Recall from class that, to fit this model, you must expand the data set so that each subject has one record for each week during which that person was observed

(i.e. a person observed for 30 weeks would need to have 30 observations). For each record, you will need to create an indicator of whether or not the subject experienced the event in that interval. The advantage of this model is that it allows you to model the effect of time on the event. In the example that we discussed in class, we used a completely nonparametric approach to model the effect of time. However, since there are no relapses during weeks 9, 14, 18-21, and 24+, we can not estimate a completely nonparametric (week-specific using the CLASS statement) hazard rate for these intervals. For this problem, I want you to consider fitting a linear effect of time (i.e. list the variable on the MODEL statement alone). Take special note that you need to create a time-dependent covariate to allow the hazard ratio for SEX to differ before and after 15 weeks. This can be done in the DATA step required for putting this data set in the proper form for fitting the complementary log-log model. Report a 95% confidence interval for the effect of treatment on time to relapse, controlling for LOGWBC, time, and SEX (and accounting for the fact that SEX does not satisfy the PH assumption).

C. Compare the parameter estimates for the effect of RX, LOGWBC, and SEX on time to relapse in the models that you fit in (a) and (b). Are they similar? Why or why not?

II. Read Page # 236 – 240 (uploaded on Blackboard) from the Survival Analysis using SAS book by P D. Allison. In these pages, you will see the analysis conducted using the Logit Model for the 'Job Duration' Study using the dataset titled JOBDUR (also uploaded on Blackboard.

A. Confirm that you have understood this analysis using the Logit Model.

B. Now, conduct a similar analyses using the Complementary log-log model. Interpret your results and compare them to those obtained from Part [A] above.

Question # 2: **30 points**

I. Suppose that we are planning a study and we want to determine the required sample size needed in order to have 90% power to detect a hazard ratio of 1.38 or greater at the 0.05 significance level. Furthermore, assume that the event rate at the end of the first year is approximately 7% in the control group (i.e., approximately 93% of those receiving the control treatment will be disease-free at the end of the first year).

A. Compute the required number of deaths needed to have 90% power to detect this difference at the 0.05 significance level, assuming equal sample sizes in the two groups.

B. Assuming that the study will last for four years, with patients being accrued during the first three years, what is the total number of patients that will need to be enrolled?

C. Assuming that the study will last for four years, with patients being accrued during the first three years, use PROC POWER to produce a plot of power as a function of total sample size.

D. Repeat (a)-(b) assuming that 75% of all patients enrolled will be allocated to the treatment group and 25% of patients enrolled will be allocated to the control group.

II. For the Example #2 discussed in class (see Page #7 – 9 from Chapter 8_Part2), calculate the sample size required when the effect size corresponds to $S_{TRT}^{-1}(p)/S_{CTL}^{-1}(p) = 2$. All other parameters can be assumed to be the same as discussed in the example.

Question # 3 **25 points**

I. Solve Exercise 5.2 from your textbook using the method described on Page #4 – 6 of the lecture notes for Chapter 9_Part1. Use maximum of three iterations.

II. Use PROC ICLIFETEST to report the estimates of the survival function for this dataset.

Question # 4 **15 points**

According to the Institute of Food Science Technology (IFSI) guidelines (1993), shelf life is defined as the time during which a food product will [1] remain safe [2] retain desired sensory, chemical, physical and microbiological characteristics, and [3] comply with any label declaration of nutritional data, when stored under the recommended conditions.
For one particular study, the objective was to determine the sensory shelf life (SSL) of whole-fat, stirred yoghurt with strawberry pulp. A reversed storage design was used in which some yoghurt pots of 150 g were kept at 4^{o} C, and some of them were stored in a 42^{o} C oven for 0, 4, 8, 12, 24, 36, and 48 hours. These times were chosen because previous experiments showed that deterioration in flavor occurred quickly up to approximately 12 hours and then slowed down. After being stored at 42^{o} C, the samples were refrigerated at 4^{o} C until they were tasted. 46 adults (between 18 and 30 years) and 47 children (between 10 and 12 years) who consumed stirred yoghurt at least once a week were recruited from a town. For each of the 7 samples presented in random order, the subject tasted the sample and answered the question; "Would you normally consume this product? YES or NO?". If a subject would consume the samples up to 8 hours storage but not the samples with 12 hours storage or longer, it is known that shelf life is somewhere between 8 and 12 hours storage. The data are thus comprised of interval-censored observations. Right-censored data occur when the subject accepts all samples and left-censored data occur if the sample with the first storage time is rejected.
The main research question is to assess whether adults and children judge the shelf life of the yoghurt differently.

Help the researchers answer this question by analyzing this data using PROC ICPHREG using:
[1] Piecewise Exponential Model
[2] Cubic Spline Model

Shelf-life study data:

Adult (N=46)

$(0,4],(0,24],(0,24],(0,36],(0,36],(0,36],(4,8],(4,8],(4,8],(4,8],(4,8],$
$(4,24],(4,24],(4,24],(4,36],(4,48],(8,12],(8,12],(8,12],(8,12],(8,36],$
$(12,24],(12,24],(12,24],(12,24],(12,24],(12,48],(12,48],(12,48],(12,48],$
$(24,36],(24,36],(24,36],(24,36],(36,48],(36,48],(36,48],(36,48],(48,\infty),$
$(48,\infty),(48,\infty),(48,\infty),(48,\infty),(48,\infty),(48,\infty),(48,\infty)$

Child (N=47)

$(0,12],(0,12],(8,12],(8,36],(8,36],(8,48],(12,24],(12,48],(24,36],(24,36],$
$(24,36],(24,36],(24,36],(24,36],(36,48],(36,48],(36,48],(36,48],(36,48],$
$(36,48],(36,48],(48,\infty),(48,\infty),(48,\infty),(48,\infty),(48,\infty),(48,\infty),(48,\infty),$
$(48,\infty),(48,\infty),(48,\infty),(48,\infty),(48,\infty),(48,\infty),(48,\infty),(48,\infty),(48,\infty),$
$(48,\infty),(48,\infty),(48,\infty),(48,\infty),(48,\infty),(48,\infty),(48,\infty),(48,\infty),(48,\infty),$
$(48,\infty)$

Question # 5 (Bonus) **5 points**

Read Page # 149 – 150 of your textbook about "Nonparametric estimation of the survival function for right-truncated data". Confirm that you understood how the calculations are done.

GOOD LUCK ☺☺