BIOS - 845 Spring 2019 Take-Home Midterm Exam

Name:_____

Date assigned: 03/22/2019; Due Date: 04/01/2019 (by 11:59 pm Blackboard clock time)

Instructions:

1. To receive full credit, show all work. Please make your work readable.
2. Use $\alpha = 0.05$ unless otherwise mentioned.
3. Total points for this exam are **50**.
4. Do not forget to write your name and page numbers on the homework.
5. Do not turn in irrelevant SAS output. Mark/circle important numbers on the SAS output.
6. You cannot discuss this exam with anyone other than the instructor.

Question # 1: **(12 points)**

The Melanoma data set contains data collected on thirty melanoma patients to compare the immunotherapies BCG (Bacillus Calmette-Guerin) and CPVM (Cornyebacterium Parvum) for their abilities to prolong survival time. The age, sex, and survival time (in months) were recorded for each patient. Observations for those patients who either withdrew from the study or had not died by the end of the study are indicated as censored. This data set has been stored in the file MELANOMA.SAS7BDAT and can be downloaded from the DATASETS folder on BLACKBOARD.

A. Create the tables that yield the Kaplan-Meier estimates for the two treatment groups and then plot the corresponding curves on the same graph. Comment on how these curves compare with one another. Use PROC LIFETEST to compute and graph the Kaplan-Meier curves for survival time for the two treatment groups. Perform both a log-rank and a Wilcoxon test comparing survival times among the two treatment groups. What conclusions do you draw?

B. Repeat part (A) comparing time to remission among the two groups. Do the overall conclusions differ depending on whether we consider survival or time to remission as our outcome variable?

C. Use the TEST statement to implement a forward selection screening procedure for examining the effect of the variables TRT, SEX, and AGE on survival. What conclusions do you draw?

D. Compute and graph both the log-survival and the log-log survival plots for survival time in the two treatment groups. What do these figures tell you about the hazard function?

Question # 2: **(20 points)**

As you recall from class, one of the drawbacks to PROC LIFETEST is that it becomes difficult to look at several variables simultaneously. This is one of the advantages of PROC LIFEREG. Using the MELANOMA dataset, we would like to use parametric models to simultaneously examine the effects of treatment status, gender, and age on survival.

A. Fit a series of parametric models (all those allowable by SAS + Gompertz) to these data including TRT, SEX, and AGE in your model as covariates. In a summary table, describe the models that you fit providing any distributional assumptions that you made. Using some objective criteria, choose a 'best' parametric model (i.e., the one with the 'best' fit). Based on your choice of a 'best' model, write a short summary (NO MORE THAN TWO PAGES) of your conclusions regarding the effects of TRT, SEX, and AGE. You should write the summary under the assumption that you are creating it for inclusion in a paper to be submitted for publication in a medical journal. The summary should include two sections:

   (1) A Statistical Methods section describing how you fit the models and how you decided on a final model to report.

   (2) A Results section describing your results in such a manner that they could be easily understood by a non-statistical audience.

   Note: In completing the above mentioned tasks, you should comprehensively consider all topics taught to you in Chapter 5 of the lecture notes on PARAMETRIC METHODS. That is, look into all sections of the lecture notes and ask yourself how it can be applied to your analyses. Think about all concerns a statistical reviewer of the medical journal you are submitting to, may want you to address. Separately, provide SAS code for all analyses conducted by you and annotate it (so that it is easy for the instructor to read it).

B. From the model that you fit in part (A), compute and plot the estimated survival curves for females 40 years of age receiving the two treatments on the same plot.

Question # 3: **(6 points)**

A. Derive the expression for $E[\mathrm{T}^n]$ given in Section 5.11 of the lecture notes when T follows a Gompertz distribution.

B. Show that if the distribution is truncated, that is, we consider the distribution of T conditional on $T > v$ and study the residual lifetime, $T - v$, for some fixed v, the distribution is again Gompertz, with the same value of b, but with a replaced by ae^{bv}

Question # 4: **(12 points)**

Consider the data set ADDICTS.SAS7BDAT that can be downloaded from Blackboard. These data represent survival data on the number of days spent by 238 heroin addicts in one of two methadone clinics. The primary purpose of the study was to determine whether there are differences in the amount of time spent in the two clinics. The response variable is the number of days spent in the clinic from entry to departure. There are two covariates which are thought to be of interest:
1) whether the person had a prior prison record and
2) the maximum methadone dose used.

The data set contains the following variables:

CLINIC = Clinic (= 0 or 1)
STATUS = Status (1=Released, 0=Censored)
DOSE = Maximum methadone dosage used
LENGTH = Number of days spent in the clinic
PRISON = Prior prison record (1=Yes)

A.

i. Recall that parameter estimates for parametric survival models are obtained using maximum likelihood estimation. How many terms are involved in the likelihood function used to obtain parameter estimates for a parametric survival model using this data set?

ii. Parameter estimates for the Cox regression model are obtained using partial maximum likelihood estimation. How many terms are involved in the partial likelihood function used to obtain parameter estimates for a Cox model using this data set?

B. Consider fitting a Cox Proportional Hazards model to compare the number of days spent in the clinic for the two clinics, controlling for PRISON and DOSE (treat DOSE as a continuous variable). Write out the model, in terms of the hazard function.

C. Use SAS fit the Cox PH model described in (B). Report a 95% confidence interval for the effect of CLINIC on number of days spent in the clinic, controlling for PRISON and DOSE. Is this effect significant at the 0.05 significance level?

D. Analyze the model in part (B) treating DOSE as a categorical variable with three categories: <60, [60, 80), >=80. Are there any changes to your interpretation of HR when you treat DOSE as an ordinal variable versus a nominal variable?

E. State at least one drawback of using the Cox model for this data.

☺ GOOD LUCK ☺