STA 312s19 Assignment Ten[1]

The paper and pencil part of this assignment is not to be handed in. It is practice for Quiz 10 on March 25th. The R parts may be handed in as part of the quiz. **Bring hard copy of your printout for Questions 4 and 5 to the quiz**. Do not write anything on your printouts in advance except possibly your name and student number. *Answers to the "plain language" questions are specifically prohibited.* Do not write them, or type them, or otherwise cause them to appear on your printouts.

1. Our main concrete example of a proportional hazards regression model is Weibull regression.
   (a) What is the baseline hazard function for Weibull regression? Assume e^{β_0} is part of the baseline hazard function.
   (b) Suppose that the Weibull regression mode is the true model for a set of data. When we fit a proportional hazards regression model by maximum partial likelihood and estimate β_1, what function of the Weibull regression model parameters are we estimating?
2. Prove $S(t) = e^{-H(t)}$, where $H(t) = \int_0^t h(y)\,dy$. This is a general statement, not just for the proportional hazards model.
3. For the proportional hazards model, again assume that e^{β_0} is part of the baseline hazard function. We will always do this from now on. Prove that for the proportional hazards model, $S(t) = S_0(t)^{\exp\{\mathbf{x}^\top \boldsymbol{\beta}\}}$.

4. The classic data set `veteran` is available as part of the `survival` package. Type `help(veteran)` for details.

   (a) Based on a preliminary analysis (one that you don't have to do), I request that you fit a model with just experimental treatment, cell type and Karnofsky score. Based on this model, carry out significance tests to answer the following questions. Be able to state H_0, give the value of the test statistic (Z or chi-squared) and the p-value. Be able to state your conclusions, if any, in plain, non-statistical language. Guidelines for the plain language statements are

   - Be guided by the 0.05 significance level, but never mention it. If you do, you get a zero even if what you say is correct.
   - Any use of statistical vocabulary such as p-value, null hypothesis, significance etc. will get you a zero. Instead of saying "controlling for," say "allowing for," or "correcting for," or "taking into account." The phrase "controlling for" will not get you a zero, but please avoid it when talking to non-statisticians.
   - If a directional conclusion is posible, make it. Don't say "Survival time was related to sex." Say "Women tended to live longer."
   - If a test is not significant , do *not* say there was no effect, or no difference. Avoid accepting the null hypothesis, or implying that you accept it. Say "There was no evidence that surgery was related survival time," or "These results do not provide evidence of a connection between marital status and time required to graduate," or something like that.
   - For any explanatory variable that was *not* randomly assigned, avoid language that suggests influence, or causal connection. Say "Patients with a health club memberships were at less risk for heart attack," not "Exercise prevented heart attacks."

   Now here are the questions.

   i. Controlling for cell type and Karnofsky score, does treatment appear to affect survival time?

   ii. Allowing for experimental treatment and cell type, does Karnofsky score help predict survival? In spite of the word "predict," you are beng asked for a significance test.

   iii. Correcting for experimental treatment and Karnofsky score, do patents with different types of cancer (cell type) differ in their hazard of dying? Do a partial likelihood ratio test.

   iv. Follow up the last question by carrying out tests for all pairwise comparisons of cancer types. Some of the comparisons you want are Z-tests on the `summary` output. Use Wald tests for the other comparisons. Directional conclusions are possible for all the tests that are statistically significant, including the Wald tests.

(b) Now we are interested in whether there could be an effect of experimental treatment that depends on the type of cancer. Fit another model with experimental treatment, cell type and Karnofsky score – except this one also allows for interaction between treatment and cell type. Use a different dummy variable coding. Display `summary` and carry out a partial likelihood ratio test for the interaction. Are you able to conclude that the effect of treatment depends on type of cancer?

(c) Returning to the model of Question 4a (which has no interaction of treatment with cell type), test whether there might be an effect of treatment that depends on Karnofsky score. Just look at the Z-test. In plain language, what do you conclude?

5. This question uses the same old `cancer` data set you have been analyzing in the past two assignments. Even though you may be getting tired of it, there is an interesting technical question we have not explored. *Please print the output for this question on a separate sheet.*

(a) Fit a proportional hazards model using the same explanatory variables you did for Weibull regression and log-normal regression: `sex` and `ph.ecog`. Be able to state the conclusions in plain, non-statistical language. See Question 4a for guidelines.

(b) Now do a table of `ph.ecog`. Also do `summary`. You can see that even though it's technically a 6-point scale, in practice the physicians are using just a few categories. It makes you wonder whether we should be treating `ph.ecog` as quantitative or categorical.

(c) Fit a model with `sex` and `ph.ecog`, in which `ph.ecog` is represented by dummy variables. Before you do this, make `ph.ecog = 3` into `NA`, since there's only one patient.

(d) If `ph.ecog` is quantitative, the proportional hazards model says the hazard is multiplied by e^{β_2} when we increase by one unit (whether it's from 0 to 1 or from 1 to 2). Express this as a null hypothesis about the parameters of your model from Question 5c.

(e) Test the null hypothesis with a Wald test. What do you conclude? (This is not a plain language question.) Does it seem okay to treat `ph.ecog` as quantitative?

Please bring **both** printouts to the quiz. Your printout should show *all* R input and output,and *only* R input and output. Do not write anything on your printouts in advance except your name and student number. The rule is that you may not put anything on your printout that you could not have known before seeing the results. So question numbers are okay. You may even copy-paste the entire questions (for the computer parts) into comment statements if you wish. But results, conclusions and interpretation are not allowed.

In particular, do not write answers to "plain language" questions on your printout, put them in comment statements, or otherwise cause them to appear on your printout. If you do, it's an unauthorized aid and you will be charged with an academic offence, whether or not that particular question was asked.