

## STA 312s19 Assignment Eleven<sup>1</sup>

The paper and pencil part of this assignment is not to be handed in. It is practice for Quiz 11 on April 1st. The R parts may be handed in as part of the quiz. **Bring hard copy of your printout for Questions 2 and 3 to the quiz.** Separate printouts would be a good idea. 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 the answers to the plain language questions, or type them, or otherwise cause them to appear on your printouts.

1. In a study of whether vitamin C can help protect against the common cold, volunteer subjects were randomly assigned to either an observation only condition (no pill), a placebo condition (sugar pill), a 100 mg pill, a 500 mg pill or a 2,000 mg pill. The pills all had the same size and appearance. Time until the first reported cold was recorded. Subjects were followed for 6 months. At that point, the data for any subject who had not caught a cold was censored. In addition to experimental condition, age was recorded for each subject. Naturally, a lot more data about the subjects would be recorded in a real study.
  - (a) These data will be analyzed using proportional hazards regression. Write the hazard function, including the baseline hazard. Denote age by  $x$ , and put it first. There are no interactions yet. You do not have to say how your dummy variables are defined. You will do that in the next part.
  - (b) In the table below, make columns showing how your dummy variables are defined. Make placebo the reference category. In the last column, write the hazard function. If *symbols* for your dummy variables appear in the last column, the answer is wrong.

	Hazard Function	
Observation Only		
Placebo		
100 mg		
500 mg		
2,000 mg		

---

<sup>1</sup>This assignment was prepared by [Jerry Brunner](#), Department of Mathematical and Computational Sciences, University of Toronto. It is licensed under a [Creative Commons Attribution - ShareAlike 3.0 Unported License](#). Use any part of it as you like and share the result freely. The L<sup>A</sup>T<sub>E</sub>X source code is available from the course website: <http://www.utstat.toronto.edu/~brunner/oldclass/312s19>

- (c) You want to test whether, controlling for age, experimental condition has any effect on the risk (technically, the hazard) of getting a cold. Experimental condition includes Observation Only.
- Write the null hypothesis in scalar form, using the notation of your answer to Question 1a.
  - Write the null hypothesis in matrix form as  $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$ .
- (d) Placebo effects are interesting. It is well documented that sometimes, people can get better by taking a pill that is medically inert, just because they *believe* it will make them better. It is particularly likely in this study, because we have data only on time to the first *reported* cold. You want to test for a placebo effect.
- Write the null hypothesis in scalar form, using the notation of your answer to Question 1a.
  - Write the null hypothesis in matrix form as  $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$ .
- (e) You want to test whether, controlling for age, dosage level has any effect on the risk of getting a cold. This test includes the placebo, which is a dosage level of zero, but excludes Observation Only.
- Write the null hypothesis in scalar form, using the notation of your answer to Question 1a.
  - Write the null hypothesis in matrix form as  $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$ .
- (f) If that last test is statistically significant, you definitely want to follow up with all pairwise comparisons of dosage levels, to see where the overall difference comes from. How many pairwise comparisons are there? The answer is a number.
- (g) One of the pairwise comparisons is between the placebo and 2,000 mg.
- Write the null hypothesis in scalar form, using the notation of your answer to Question 1a.
  - Write the null hypothesis in matrix form as  $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$ .
- (h) Another pairwise comparison is between 100 mg and 500 mg.
- Write the null hypothesis in scalar form, using the notation of your answer to Question 1a.
  - Write the null hypothesis in matrix form as  $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$ .

- (i) Now we want to check whether the magnitude of the effects might depend on the age of the subject. Write the hazard function for proportional hazards regression, including the baseline hazard. Dummy variable definitions are the same as in Question 1b.
- (j) Fill in the table for the model of Question 1i above.

Hazard Function

Obs. Only		
Placebo		
100 mg		
500 mg		
2,000 mg		

- (k) For a test of the interaction of age by experimental condition (including Observation Only),
- i. Write the null hypothesis in scalar form, using the notation of your answer to Question 1i.
  - ii. Write the null hypothesis in matrix form as  $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$ .
- (l) Maybe older people have more faith in the medical system. Test whether the magnitude of the placebo effect depends on age.
- i. Write the null hypothesis in scalar form, using the notation of your answer to Question 1i.
  - ii. Write the null hypothesis in matrix form as  $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$ .
- (m) Test whether the effect of dosage depends on age. This excludes Observation Only.
- i. Write the null hypothesis in scalar form, using the notation of your answer to Question 1i.
  - ii. Write the null hypothesis in matrix form as  $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$ .

2. The `channing` data set in the `KMsurv` package has data on death in nursing homes as a function of age and sex. You will need to install the `KMsurv` package. Information about the data is provided by `help channing`. For some reason you need `data(channing)` before you can use it.

Please make gender a binary variable with 1=Female and 0=Male. Also, create a centered version of age, by subtracting off the sample mean for the entire sample. This variable should be in years, not months. When a question refers to a resident of “average age,” it refers to sample mean age. When I say ‘age,’ do I mean the variable `ageentry` or the variable `age` in the data file? Why?

- (a) In any data analysis, you should start with some basic descriptive statistics to see what the data are like. Please answer these questions. The answers are numbers. A summary of the data frame will help.
- What was the median age at which participants entered the nursing home, in years?
  - What was the age of the youngest person to enter the nursing home, in years?
  - What was the age of the oldest person to enter the nursing home, in years?
  - What proportion of the nursing home residents were women?
  - What was the longest length of stay at the nursing home, in years?
  - What proportion of the observations were censored?
- (b) Fit a proportional hazards model with just age at entry, and gender. There is no interaction yet.
- Interpret both  $Z$ -tests in plain, non-statistical language. This is one sentence each.
  - Correcting for age, the estimated hazard of death is \_\_\_\_\_% as great for women.
  - Give a 95% confidence interval for that last answer. Allowing for age, we estimate the hazard to be between \_\_\_\_\_% and \_\_\_\_\_% as great for women.
  - Controlling for gender, if age at entry is increased by 10 years, we estimate the hazard of death to be multiplied by \_\_\_\_\_.
  - Give a 95% confidence interval for that last number.
  - Test the proportional hazards assumption. What do you conclude?
- (c) Now fit a model that allows you to test whether the sex difference in risk of death depends on the age of the nursing home resident.
- Do it two ways, with age at entry centered and age at entry uncentered. Which way do you like more? What is the evidence that the fit of the two models is the same?
  - State your conclusion (about the interaction) in plain, non-statistical language.

3. In the liver disease data, patients were randomly assigned to one of two drugs, or to a placebo. The data file includes age and sex (1=F). Blood platelet count was recorded for each patient in each time period. The data are available at

<http://www.utstat.toronto.edu/~brunner/data/legal/liver.data.txt>

- (a) What was the failure time for Patient 1? Patient 2? Patient 3?
- (b) For a data file in this format, it can be challenging to get basic descriptive statistics. Let's see if you can do it. See `help(subset)`.
- How many patients took part in the study?
  - What was their mean age?
  - What percent were male?
- (c) Fit a proportional hazards model with no interactions.
- Controlling for sex, experimental condition and platelet count, is there evidence that the risk of death depends on age?
    - Answer the question Yes or No.
    - If the answer is Yes, state the conclusion in plain, non-statistical language. If the answer is No, that's all you need to say.
    - What test statistic value supports your conclusion? The answer is a number on your printout.
    - What  $p$ -value supports your conclusion? The answer is a number on your printout.
    - State the null hypothesis in symbols.
    - Do you reject  $H_0$ ? Answer Yes or No.
    - Are the results statistically significant? Answer Yes or No.
    - All other things being equal, if age is increased by one year, the estimated hazard of death is multiplied by \_\_\_\_\_.
    - Give a 95% confidence interval to go with that last estimate.
    - All other things being equal, if age is increased by ten years, the estimated hazard of death is multiplied by \_\_\_\_\_. You could get this from the output of `summary` with a calculator, but please use R and display the result on your printout.
    - Give a 95% confidence interval to go with that last estimate. Again, this is an easy calculation based on the output of `summary`.
  - Controlling for age, experimental condition and platelet count, is there evidence that the risk of death depends on sex of patient?
    - Answer the question Yes or No.
    - If the answer is Yes, state the conclusion in plain, non-statistical language. If the answer is No, that's all you need to say.

- C. What test statistic value supports your conclusion? The answer is a number on your printout.
  - D. What  $p$ -value supports your conclusion? The answer is a number on your printout.
  - E. State the null hypothesis in symbols.
  - F. Do you reject  $H_0$ ? Answer Yes or No.
  - G. Are the results statistically significant? Answer Yes or No.
- iii. Controlling for age, sex and platelet count, is there evidence that experimental treatment (including the placebo condition) affects the chances of survival? Answer the question with a Wald test.
- A. Answer the question Yes or No.
  - B. What test statistic value supports your conclusion? The answer is a number on your printout.
  - C. What  $p$ -value supports your conclusion? The answer is a number on your printout.
  - D. State the null hypothesis in symbols.
  - E. Do you reject  $H_0$ ? Answer Yes or No.
  - F. Are the results statistically significant? Answer Yes or No.
  - G. If the results are statistically significant, follow up with all pairwise comparisons. State your conclusions in plain, non-statistical language.
- (d) Now fit a model that allows you to test whether the effect of experimental treatment depends on the patient's blood platelet count. Test the null hypothesis with a partial likelihood ratio test. What do you conclude?
- (e) Test the proportional hazards assumption. What do you conclude?

---

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 on the quiz.