**Take-Home Exam 1**

**Psychology 610**

**Due: 13 October 2017, 9:00am (100 points total)**

**Instructions**

Welcome to your first take-home exam! Remember, you are allowed and encouraged to refer to prior lab scripts, homework assignments, books/articles, and the internet. **You may *not* consult with other students.** Use the same general format as you do for homework assignments (see template document on the website). A few things to note:

* **Submit your files anonymously.** Instead of using the normal Lastname\_HWXX format, use the **last 4 digits of your student ID number** (e.g., “3180\_THE1.R”). *Include no identifying information in your files!!*
* Label your answers in your Word doc according to the question numbers given here (with a Reading Questions section and a Data Analysis section).
* Follow course norms for naming variables and models (or use names that will be crystal clear to us).
* You don’t need to include the question text next to the question number in your R script; the number is plenty for us.
* You can email us asking questions, but we won’t be as instructive as we would be on a homework assignment. This is an exam, after all.

**Reading Questions**

Answer the following questions in your Word document.

1. Which summary measure of error do we typically use in our analysis? Why? (2 pts)

2. One assumption about our models of ERRORS is that the errors are normally distributed. Name three other assumptions about the errors that the book makes very explicit. Briefly explain each assumption. (3 pts)

3. Imagine that we have a compact model C and an augmented model A with one additional parameter (b) not contained in model C. We fit both of our models to our data, and find that the total error for C equals the total error for A (using our standard method for measuring error from Question 1).  What is the best estimate for b that we found for model A? Briefly explain your answer. (3 pts)

4. Does the F statistic for a given model comparison increase or decrease for each of the following scenarios, holding all else constant? (Respond with either “increase” or “decrease”). (4 pts)

a. Sample size increases

b. Residuals (errors) for the augmented model increase

c. The number of parameters estimated increases

d. The difference between the SSE for the compact and augmented model increases

**Data analysis**

Some researchers at UW-Madison hope to reduce student loneliness with a new social media app called “Look@ME”. Look@ME was designed to be similar to Instagram except that it only allows pictures of the user doing interesting things (as determined by research team review). The options for reacting to pictures posted to Look@ME are a thumbs up, a heart, a larger heart, or a kiss.

The researchers conducted a study in which 80 heavy social media using student participants were randomly assigned to use the new Look@ME app every day for a month, while another 100 heavy social media users were assigned to simply continue using whatever apps they normally would for the month of the study. Before assignment to conditions, participants filled out a loneliness questionnaire for which they rated how lonely they generally felt on a scale from 1 (not at all lonely) to 7 (extremely lonely).

After one month, researchers asked the students to fill out the loneliness scale again. The researchers were also able to record how many hours a day participants in the Look@ME app condition actually used the app.

The researchers expect their new Look@ME app to reduce self-reported loneliness for the heavy social media using students.

The codebook for the data is given below. **Unless otherwise specified, answer the following questions in your R script.**

|  |  |  |
| --- | --- | --- |
| Var. name | Description | Range in sample |
| BaseLoneli | Self-reported loneliness at baseline | 1-5 |
| AppCondition | Whether the student was given the new app (No = 0) | 0 or 1 |
| AppHours | How many hours a day participants used new app | 0-7.02 |
| LaterLoneli | Self-reported loneliness one month later | 2-6 |

1. Read in and inspect the data. Run as many commands as you need to in order to thoroughly understand the data. (2 pts)
2. The researchers notice that their assistant left two demographic variables in the data frame that they have no interest in, at least for the purposes of this study (Height, Weight). Remove these variables from the data frame. (2 pts)
3. Obtain the means and SDs of LaterLoneli for participants in the two different conditions separately to get a general sense of whether the outcome variables are affected by the treatment (do not run any models). Do not write comments about what you learn; this is just so you can become familiar with the data. (2 pts)
4. The researchers want to see whether participants in the sample are lonely or not, on average. In other words, they want to test if scores on the loneliness scale are different than 0. It will be easier to interpret this test if the loneliness variables themselves have a potential floor (i.e., low anchor point) of 0 (instead of 1). Create two new variables for which you shift the possible *scale* of the BaseLoneli and LaterLoneli variables from 1-7 to 0-6. Name these two new variables “BaseLoneli06” and “LaterLoneli06”. Check the range of each new variable to make sure you did this correctly. Use these new variables in all analysis below. You may remove the old variables from the data frame if you like. (3 pts)
5. Create a very simple histogram of all participants’ baseline loneliness scores. Label each axis so that you can give this histogram to someone else and have them understand what is displayed. Paste this figure in your word doc and label it “Figure 1.” (2 pts)
6. Use an R command to determine the values used to code AppCondition. In your script, write a comment describing how AppCondition is coded, and what this coding scheme is called. (2 pts)
7. Create a zero-centered version of the AppCondition variable and call it AppCondition0C. Write a sentence in your R script describing what a 0 on this zero-centered variable represents. (4 pts)
8. Using your original dummy-coded condition variable, determine if the two experimental groups (new app vs no new app) differ in their baseline loneliness. Run appropriate analyses to see. In your R script, write 1-2 sentences answering this question (reporting all relevant stats including effect size). Be sure to indicate the size of the difference between the two groups (in loneliness scale units), and whether it is significant. Finally, explain why checking for baseline differences is important. (8 pts)
9. What is the interpretation of b0 in the model you used in question 8? (3 pts)
10. Estimate a model where b0 is interpreted as the unweighted mean for later loneliness, controlling for their app condition. In your R script, write 1-2 sentences describing the test of b0 (reporting all relevant stats including effect size) and explain why you may want to assess b0 in your data in this design. (5 pts)
11. Is the app effective at reducing loneliness? Run the relevant model(s) using loneliness one month later as the outcome variable. Report all relevant stats including effect size. Describe the relationship in one sentence in your R script. (5 pts)
12. Does your answer to question 11 depend on whether you used AppCondition or AppCondition0C in your model? Why or why not? (3 pts)
13. Create a publication-quality barplot that displays later loneliness scores as a function of AppCondition. Include error bars, raw data points (jittered), and informative labels. Paste this plot into your Word doc and label it “Figure 2.” (8 pts)
14. One researcher suddenly remembers that the research team also collected the number of daily hours people used the new app. Run a new model, using raw AppHours, testing whether loneliness changes as a function of number of hours spent on the app. Be sure to include the data from both app conditions in your analysis. Interpret the parameter estimates for both b0 and b1 and report all relevant statistics. (5 pts)
15. Create a new variable that represents mean-centered AppHours. Run the model from question 14 again, using this new variable. Visually compare the results from this model with the results from the model with raw AppHours. What changed and what stayed the same? Why? (4 pts)
16. *Using the effects package,* create a quick and dirty plot visualizing the effect of AppHours using the Look@ME app on loneliness scores at one month. Paste this graph into your Word doc and label it “Figure 3”. (2 pts)
17. Now create a *publication-quality* scatterplot that displays loneliness at one month as a function of number of hours using the new app. Include a regression line, appropriate error bands, raw data points, and informative labels. Paste this graph into your Word doc and label it “Figure 4.” (8 pts)
18. What is different about the results obtained in question 11 (using the dichotomous predictor) and question 14 (using the continuous predictor)? Based on the results of the two models alone, which of these two predictors appears to be more powerful? Why? Answer in your script. (4pts)
19. In your Word document, write a results section summarizing the background of the study, the hypotheses you tested, the statistical results of your tests using both the dichotomous and the continuous predictor (in APA format), and the practical interpretation of these results. Indicate whether or not the researchers’ hypothesis was supported. Write for an audience that has some statistical knowledge, but little knowledge about this research, and reference the figures you made. In your last sentence, make a conclusive statement about the new app’s effects on student loneliness. Aim for 250-350 words. (15 pts)
20. At the end of your word, type how long it took to complete the assignment (1 pt).