**Comprehensive Assignment 3**

**Psychology 610**

**Due date: Wednesday, December 20th 2023, 1:30pm (90 total points)**

Welcome to your third comprehensive assignment! 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, including more advanced students in your program.** Use RMarkdown as you do for homework assignments. A few things to note:

**Submit your files anonymously.** Submit a .Rmd and .PDF file.Instead of using the normal Lastname\_HWXX format, use the **last 4 digits of your student ID number** (e.g., “3180\_comp\_03.Rmd”). *Include no identifying information in your files!!!*

You can email us asking questions, but we will not be as instructive as we would be on a regular homework assignment. Do not post questions to Slack channels where other students can see.

The number of points each question is worth appears in parentheses next to the question, for a total of 100 points. Three points are awarded for the correct file submission and naming. To obtain these points you must submit both an R Markdown and HTML file.

**Formatting** (Five points are awarded for proper formatting. To obtain these points you must adhere to the following instructions):

Label your answers in RMarkdown according to the sections and question numbers given here (e.g., with a Reading Questions section and a Data Analysis section). Use RMarkdown heading level 1 (#) for section headings (e.g., “Reading questions”, “Study 1” etc.). Use RMarkdown heading level 2 (##) for question numbers. Use RMarkdown heading level 3 (###) if there are question letters. Always start your answer on a new line following any question number, so your response is not formatted like a heading.

Do not repeat the question prompt next to the question number in your RMarkdown script; the number is plenty for us (and repeating the question adds more clutter we have to sort through for grading!).

When reporting *b* values, *p*, *F*, and ηp2, use the standard conventions for the correct number of decimal places. These conventions are listed in lab 5.

Remember to create a chunk in your RMarkdown file that is specifically for loading in any packages needed to complete this assignment, as well as sourcing in functions 610710\_functions.R.

Follow the course style guide and norms for naming variables and models.

When you’re done, check your work & the formatting in your HTML file for mistakes!

You got this!

**Part 1: Reading and Conceptual Questions, 7 points**

1. **Hypothesis 1**: A researcher hypothesizes that the severity of ADHD symptoms (*adhd*) will predict performance on a basic cognitive task (*perf*).   
   **Hypothesis 2**:The researcher further hypothesizes that the relationship between ADHD severity and performance will be moderated by condition (*condition*), an experimentally manipulated predictor, in which participants are assigned to either view (or to not view) brightly colored, animated icons on the screen of the computer being used for the task.  
   **Control Variable**: When testing these two hypotheses she decides to control for processing speed (*speed*).
   1. Write out the correct (augmented) model that would allow the researcher to test both of the hypotheses and control for processing speed. When writing your model, use the *italicized* variable names provided in parentheses. (2)
   2. Write the compact model for question 1a which tests whether the main effect of ADHD symptom severity is significant. (1)
   3. Write the compact model for question 1a which tests whether the interaction between ADHD symptom severity and condition is significant. (1)
2. According to Judd et al., 2014, what complicates the testing of interactions with measured predictors, specifically with regard to obtaining adequate statistical power? (2)
3. Explain the *additive assumption* described by Judd et al., 2017. What does the “reasonableness” of this assumption depend on? (2)

**Part 2: Data Analysis Questions**

**Study 1 (24 points)**

Researchers at a medical school want to determine how their students’ confidence changes throughout medical school and if they can help increase their students’ confidence. The researchers want to determine if students will enter medical school with high confidence, then experience a drop in confidence, and finally have a rise in confidence by the end of medical school. The researchers also decide to implement a confidence-boosting workshop, which they randomly assign students to either complete or to not complete. Following the completion of the confidence-boosting workshop, all students (both those who completed and who did not complete the workshop) responded to a survey assessing their confidence. The researchers also record how long the student has spent in medical school (measured in academic quarters).

The researchers have shared their data with you. They have asked you to test whether confidence varies not linearly, but rather curvilinearly as a function of time spent in medical school. The researchers also ask you to test whether the confidence-boosting workshop improves students’ confidence. (Note that the researchers made no hypotheses concerning interactive effects between time spent in medical school and attending the workshop, therefore you do not need to test for interactions).

|  |  |
| --- | --- |
| med\_school.csv |  |
| ID | Student ID |
| Time | The number of academic quarters students have spent in medical school ranging from 1 to 20 |
| Workshop | 1 = completed the confidence-boosting workshop, 0 = did not complete the confidence-boosting workshop |
| Confidence | Students’ confidence measured from 0 (not at all confident) to 10 (very confident) |

1. Import and examine your data file (“med\_school.csv”). Include 1 to 2 sentences noting strong/weak relationships between predictors and any odd or unusual stats or distributions. If you notice any observations that must be removed (e.g., observations with impossibly large values), remove them and explain why you did so in one sentence. (3)
2. Fit an additive linear regression model predicting confidence from time, including workshop as a covariate. Interpret the model fully, including the *b*’s, *F* values, *p* values, and ηp2. (3)
3. Fit a regression model to test for the quadratic effect of time, statistically controlling for whether students completed the workshop. Interpret the quadratic effect fully, including the *b*’s, *F* values, *p* values, and ηp2. (3)
4. What is the linear effect of time on confidence at the mean time spent in medical school, controlling for completion of the workshop and the quadratic effect of time? (2)
5. From the quadratic model that includes the covariate, what is the predicted confidence for students prototypically early in medical school (1 SD below the mean on time) who have completed the confidence-boosting workshop? What is the linear effect of time for these same individuals? (2)
6. Make a scatterplot that shows confidence as a function of time. Add a (curved) line that represents the model predictions of the model you estimated in question #3. The model predictions should be plotted for (hypothetical) participants who are half way between having completed and having not completed the confidence-boosting workshop. For full points, your graph must include a regression line (or lines), raw data points, confidence intervals representing the standard error of the point estimate, a title, clear labels on the x and y axis, and have a blank white background. (6)
7. Write a short results section in your Rmarkdown file. Very briefly set up the study and report all statistics that are relevant to the researchers’ hypothesis (report the *b*’s, *F* values, *p* values, and ηp2). Provide a one-sentence summary or conclusion at the end. (5)

**Study 2 (21 points)**

Affective forecasting is an individual’s ability to predict how one will feel in the future. A group of clinical psychologists want to examine if a person’s affective forecasting ability predict depression. The psychologists recruited participants to complete a survey assessing each participant’s affective forecasting ability (using a 4-item scale) and their depression symptoms (using a 6-item scale).

|  |  |
| --- | --- |
| forecasting.csv |  |
| ID | Participant ID |
| Forecast\_1 Forecast\_2 Forecast\_3 Forecast\_4 | Participants’ affective forecasting scores. Scores on each item range from 0 to 10 with higher score indicating better affective forecasting abilities. |
| Dep\_a  Dep\_d  Dep\_c  Dep\_d  Dep\_e  Dep\_f | Depression scores 6 depression items. Scores on each item range from 1 to 7 with higher scores indicating greater depression. |

1. Import and examine your data file (“forecasting.csv”). Include 1 to 2 sentences noting strong/weak relationships between predictors and any odd or unusual stats or distributions. If you notice any observations that must be removed (e.g., observations with impossibly large values), remove them and explain why you did so in one sentence. (3)
2. Create composite (average) scores for depression and affective forecasting. For each of these composite scores, check the reliability of the scale. (3)
3. Fit a model predicting depression from active forecasting. Provide a 1- to 2-sentence summary of your results, include the *b* estimate of the focal predictor, *F* values, *p* values, and ηp2. (2)
4. Create a scatterplot in which you plot a participant’s average depression score on the y axis and their average affective forecasting score on the x axis. Does this relationship appear linear? (2)
5. Determine if your data include any observations that should be removed because they are regression outliers or have high influence on the model. (2)
6. Determine if the model assumptions of normality, constant variance, and linearity have been violated. Write 2 to 3 sentences describing what you find. (2)  
     
   Note: R Mardown does not include the graphs from the function used to test model assumptions in the knitted PDF or HTML file. If your graphs do not appear when you knit your file, you will still receive full credit for this question.
7. Regardless of your answers to questions 5 and 6, use the same model from question 3 and determine a recommended transformation that you could apply to these data to ensure that your model better fits the linearity assumption. (2)
8. Based on your answer to question 7 apply a transformation to your data. Be sure to pick a transformation that you can interpret. (2)
9. Re-fit a model predicting depression scores from the affective forecasting scores, using the transformed data. Determine if your transformed data corrected the violations of our model assumptions. (2)
10. Provide a 1- to 2-sentence summary of the model you fit in question 9. Including the *b* estimate of the focal predictor, *F* values, *p* values, and ηp2. Make sure that your summary includes a conceptual interpretation of your *b1* estimate, which could be understood by someone who is not an expert in statistics. (3)

**Study 3 (30 points)**

For many students, recognizing that a subject is important or connected to their personal values is a strong prediction of success in school. For this reason, researchers have used utility-value interventions, which emphasize the personal relevance and importance of course content, to improve academic outcomes.

Suppose researchers wanted to determine if the effectiveness of the utility-value intervention differed between STEM and non-STEM courses. The researchers assigned students in either STEM or non-STEM courses to take part in a utility-value task or a task unrelated to the course. They asked students about their previous experience with the course content (e.g., previous biology courses prior to taking intro to biology in college). Finally, the researchers obtained the students’ GPA at graduation.

utility\_value.csv

|  |  |
| --- | --- |
| intervention | 1 = took part in the utility-value intervention  0 = did not take part in the utility-value intervention |
| stem | 1 = intervention was administered in a STEM class  0 = intervention was administered in a non-STEM class |
| exp | Experience with the content from the course. Scores range from 0 (no experience) to 5 (lots of experience). |
| gpa | GPA measured from 0 to 4 |

1. Import and examine your data file (“utility\_value.csv”). Include 1-2 sentences noting strong/weak relationships between predictors and any odd or unusual stats or distributions. If you notice any observations that must be removed (e.g., observations with impossibly large values), remove them and explain why you did so in one sentence. (3)
2. Fit a model to test if being in a STEM (versus) non-STEM course interacts with whether students took part in the utility-value intervention to predict students’ GPA at graduation. Be sure to center any variables that should be centered for easier interpretation of your results. (3)
3. Provide a *full* interpretation of each of the parameter estimates from the model you fit in question 2. A full interpretation of a parameter includes a practical explanation for the *b* value and states which other effects (if any) were controlled for when calculating that *b* value. (4)
4. What is the simple effect of the utility-value intervention for participants in a non-STEM course. (2)
5. What is the predicted GPA for each of the following groups:
   1. Students who took part in the utility-value intervention and are in a non-STEM course? (.5)
   2. Students who took part in the utility-value intervention and are in a STEM course? (.5)
   3. Students who did not take part in the utility-value intervention and are in a non-STEM course? (.5)
   4. Students who did not take part in the utility-value intervention and are in a STEM course? (.5)
6. Fit a model to test if being in a STEM (versus) non-STEM course interacts with experience to predict students’ GPA at graduation. Be sure to center any variables that should be centered for easier interpretation of your results. (3)
7. Refit the model from question 6 using uncentered predictors. Make a publication quality graph depicting the interaction from this recentered model. Experience should be graphed on the x-axis. Include a dashed line, representing participants who are hypothetically half way between being in a STEM versus non-STEM course. For full points, your graph must include a regression line (or lines), raw data points, confidence intervals representing the standard error of the point estimate, a title, clear labels on the x and y axis, and have a blank white background. (6)
8. Identify where on the graph you observe each of the parameter estimates from the model you fit in question 6. (4)  
     
   Note: if you run into issues graphing in question 7, draw (either by hand, using excel, or a graphic illustrating program) what you think the graph should look like. If you include that image in your R markdown file and identify all points on the graph representing the parameter estimates from your model, you will still receive full points for question 8.
9. Provide a 2- to 3-sentence summary of the results from the model you fit in question 6. For full points, explain what variables you controlled for and why; include the *b*’s, *F* values, *p* values, and ηp2; and include a conceptual interpretation of your *b3* estimate, which could be understood by someone who is not an expert. (3)

How long did this assignment take to complete? (Report time in hours)