**Psych 610/710**

**HW 4, Due Wed October 4th, 2023, at 1:30 PM.**

Answer questions in an R Markdown file.

Submit both an R Markdown file and a knitted pdf or html file.

Follow the class style guide for variable naming, spacing, etc.

**Setting Up R Markdown**

1. Set your working directory, load the appropriate libraries, and source any functions you will use for your analyses.
2. (Bonus) include code in your R Markdown file to suppress warning messages when knitting your file.

**Reading questions**

1. Sampling procedures: What core assumption is required for the mean to be the most efficient estimator for a given sample?
2. *F* statistic: What are the two reasons for calculating *F*, according to Judd et al.’s *Data Analysis*?
3. Statistical terms: What special name is given to the square root of the mean-squared-error from the mean-only model?
4. Confidence Intervals: According to Judd and colleagues, what information does one need in order to calculate the 95% confidence intervals for the parameter of a simple linear regression that represents the slope?

**Data analysis**

A group of researchers is interested in determining why people believe conspiracy theories. Specifically, the researchers want to see if individuals who have high anxiety scores are more likely to also believe conspiracy theories. The researchers want to further test if individuals who use social media are more likely to believe conspiracy theories than individuals who do not use social media.

The file hw\_04\_data.xlsx contains data from 1,974 (fake) participants who completed a survey assessing their (1) belief in conspiracy theories, (2) anxiety, (3) use of social media, and (4) demographic characteristics for age, gender, and political affiliation. Further information about each variable is written below.

**Variables:**

* *Id*: a unique id assigned to each participant ranging from 1 to 1974
* *cons1 – cons15*: 15 items assessing the extent to which individuals believe conspiracy theories. All fifteen conspiracy items are forward coded. Response options range from 1 to 5 with greater scores indicating stronger belief in conspiracy theories.
* *SocialMedia*: Indicates whether or not an individual uses social media. A score of 1 indicates a person does not use social media. A score of 2 indicates a person does use social media.
* *AnnoyingVariable*: this is a useless variable that your TAs included so that you can practice removing it.
* *Anx1 – Anx5*: 5 items assessing an individual’s anxiety. Items Anx4 and Anx5 are reverse coded. Response options range from 1 to 5 with greater scores indicating higher anxiety.
* *Age*: the participant’s age
* *Gender*: the participant’s gender identity. Gender was identified as 1 (Female), 2 (Male), 3 (Another Gender Identity), 4 (Prefer not to say)
* Political Affiliation was measured using the American National Election Studies 7-Point Political Party Affiliation Scale. This scale includes 4 separate items of which participants only answer 2. The four items contained within the ANES scale are described as follows:
  + *anes\_0:* In this scale participants are first asked if they are a (1) Republican, (2) Democrat, or (3) an Independent (i.e., someone who is neither a Republican nor Democrat). Responses to this first item are included in the *anes\_0* column.
  + *anes\_rep:* If participants answer “Republican” to *anes\_0*, they complete question *anes\_rep*. In the *anes\_rep* column, a score of 1 indicates the participant is a “Strong Republican” and a score of 2 indicates the participant is a “Weak Republican.”
  + *anes\_dem:* If participants answer “Democrat” to *anes\_0*, they complete question *anes\_dem*. In the *anes\_dem* column, a score of 7 indicates the participant is a “Strong Democrat” and a score of 6 indicates the participant is a “Weak Democrat.”
  + *anes\_ind:* If participants answer “Independent” to *anes\_0*, they complete question *anes\_ind* (for independents). In the *anes\_ind* column, a score of 3 indicates the participant is a “Republican Leaning Independent,” a score of 4 indicates the participant is an “Independent,” and a score of 5 indicates the participant is a “Democratic Leaning Independent.”
  + To create a composite ANES score, columns *anes\_rep*, *anes\_dem*, and *anes\_ind* should be combined. The resulting variable should include the following numeric values: 1 (Strong Republican), 2 (Weak Republican), 3 (Republican Leaning Independent), 4 (Independent), 5 (Democratic Leaning Independent), 6 (Weak Democrat), and 7 (Strong Democrat).

1. Read in the data, convert variable names to snake case, and get summary statistics for each of the columns included in the dataframe.
2. Notice that there is a variable that we do not need for these analyses (labeled AnnoyingVariable). Using bracket notation, remove this variable while also putting the remaining variables (columns) in an order that makes sense to you. Run a command that shows the names of the variables that are left.
3. Prepare your two predictor variables and your outcome variable for data analysis. To do so, create composite scores for participant’s belief in conspiracy theory scores and anxiety scores. Do not include any participant in a composite score if they have missed more than 1/3 of the response items for a variable. While it will not affect your interpretation of b1, later in the homework we will ask you to interpret b0. For the sake of easier interpretation of b0, recode social media use age -0.5 (does not use social media) and 0.5 (uses social media).
4. Make two quick and dirty plots exploring our main questions (i.e., do anxiety or social media predict the belief in conspiracy theories?).
5. Fit a linear model to determine if anxiety predicts the belief in conspiracy theories. Report the corresponding (b1) parameter estimate, *F-*statistic, *df,* *p*-value, and h*p2*. Provide a 95% confidence interval for the (b1) parameter estimate.
6. Create a publication-quality graph that depicts the model that you fit in Question 5. Make sure you label each axis and include the data points.
7. Fit a linear model to determine if social media use predicts the belief in conspiracy theories. Report the corresponding (b1) parameter estimate*, F*-statistic, *df*, and *p*-value. Provide a 95% confidence interval for the (b1) parameter estimate.
8. Report h*p2* (PRE) along with its interpretation in a sentence to describe the effect of social media use on the belief in conspiracy theories.
9. From the output of the linear model from Question 7, interpret the “intercept” or b0 coefficient. What does it mean in this sample? What does its corresponding *p* value mean?
10. Create a publication-quality graph that depicts the model you fit in Question 7.
11. Write a concise (a few sentences) description of the tests you conducted in Questions 5 and 7. Include the statistical results of your tests, and the practical interpretation of the results. Write your response so that someone who does not know anything about these data could understand your results.
12. Create a composite score for political affiliation. (Hint: you may need to consult early labs form the course or use external sources like Chat GPT to complete this question)
13. Provide statistics for your demographic variables. For age, report a mean and standard deviation. For gender and political affiliation, identify the number and percentage of participants in each demographic category (i.e., the number of Strong Republicans, Weak Republicans, etc.). The table does not need to be publication quality.
14. (Bonus) When using some functions (such as describeBy) in R Markdown, the output often appears in multiple windows (see the picture below for an example from an unrelated dataframe). Create a table in R which gives the average age for each gender group. Make sure the table appears in a single window.



1. At the bottom of your script write the number of hours you spent working on this homework assignment.