Homework 2

PSC-103B

YOUR NAME HERE

2025-01-16

Instructions

Please use R/RStudio to complete the following questions. You will submit your filled-out version of this document **as a PDF** on Canvas. Make sure your PDF looks as expected before submitting. Please **include the code you used to generate your answer for each question (when relevant) as well as the final answer and/or relevant output**.

You may consult with your classmates while working on the assignment, but **you must do all the work yourself – everything you turn in must be your own code and words**. Academic dishonesty will not be tolerated.

Please submit **a pdf version** of this document with your answers on Canvas by **2:00pm on Thursday, January 16th (before lab)**.

Data

For this homework assignment, we will be using a subset of data from the Nerdy Personality Attributes Scale (NPAS). This is a real scale, and if you're interested, you can take the quiz yourself: https://openpsychometrics.org/tests/NPAS/

The original dataset had over 25,000 responses from all over the world. In order to make things a bit simpler, I chose 200 responses randomly from each continent (in order to prevent the responses from being too skewed towards one continent, such as the Americas). There are a bunch of variables in this dataset, but the ones we'll be interested in for this assignment are:

- `nerdy_scale`: Participant's average "nerdiness" score from the NPAS (1-5 scale)
- `TIPI1`: How extraverted or enthusiastic the participant rates themselves (1-7 scale)
- `TIPI3`: How dependable or self-disciplined the participant rates themselves (1-7 scale)
- `nerdy_selfreport`: Do you see yourself as someone who is nerdy? (1-7 scale)

1. Please read the NPAS dataset (NPAS.csv) into R using `read.csv()`. Create a dataset that only contains the variables listed above (recall: how do you subset columns from a dataframe?) (.5 pt). Show the first 6 rows (.5 pt) and last 6 rows (.5 pt) of the new dataset. Show your code.

2. Show the correlations among all the variables in the new dataset (1pt). Show your code.

3. Test whether the correlation between the participants' nerdiness scores on the NPAS (`nerdy_scale`) their level of extraversion is significant (1 pt). Report this correlation using APA format (.5 pt). Show your code.

4. Create a scatterplot to reveal the relationship between participants' level of self-discipline and their nerdiness score on the NPAS (make sure to display their level of self-discipline on the x-axis and nerdiness score on the y-axis, and include appropriate axis labels and a title) (1 pt). Use a few words to describe the general trend revealed from the plot (1 pt). Show your code.

Note: This plot might look a little weird (e.g., a bunch of straight lines), and that is because the level of self-discipline is an ordinal variable, so it cannot take on all possible values between 1 and 7, only the integers.

5. Build up a linear model to test if level of self-discipline can significantly predict the nerdiness score (1 pt). Show your code and the summary of the model.

6. Interpret the intercept (.5 pt) and slope (.5 pt) of your model.

7. Report and interpret the (unadjusted) R^2 value of your model (1 pt).

The R^2 value is 0.0004, which means that approximately 0.04% of the variation in nerdiness scores is explained by participants' self-discipline scores.

8. Report the results of this linear regression in APA format (.5 pt).

9. Assuming a person gives themselves a self-discipline rating of 4, what is their expected nerdiness score on the NPAS based on the linear model from Q5? (.5 pt).