

Programming summer school project instructions

pyschtech office

Well done!

If you've made it this far, you've survived the first 3 days of the programming summer school. You've been exposed to a lot of new things, and this is where you start putting things into practice.

Project overview

You and your colleagues will be working with a real data set to write a methods and results section for a research paper. You'll include a table of descriptive statistics, a scatter plot, a bar chart and a histogram.

You'll be checking to see if there is a relationship between 2 scale variables, and if there is a difference between 3 conditions on a scale variable.

Data set

The data set is on brightspace and it's called `project_data.csv`.

It is a messy data set, and you'll need to clean it up before you can start your analysis.

Cleaning steps

- Remove all the columns you don't need
- Remove all the rows with missing data
- Check for rows that should be removed for other reasons
- Rename the columns to something more meaningful
- Calculate the total score for both the scale variables and save them as new columns
- Make a single column that represents which condition the participant was in (combine the 3 'FL_12_DO' columns into one)
- Save the cleaned data set as a new csv file called `cleaned_project_data.csv`

Analysis

Descriptive statistics

- You'll need to generate descriptive statistics for the 2 scale variables
- You'll need to generate the mode for the scale variable both overall and for each condition
- You'll need to check if the data is normally distributed (comparing the mean and median, and using a histogram)

Scatter plot

- You'll need to generate a scatter plot of the 2 scale variables

Bar chart

- You'll need to generate a bar chart of the mean for the scale variable for each condition

Inferential statistics

- You'll need to run a correlation test to see if there is a relationship between the 2 scale variables
- You'll need to run an ANOVA to see if there is a difference between the 3 conditions on the scale variable

Report

- Using the apaquarto extension you can render the document in APA format
- You'll need to include a reference list at the end of the document (don't worry about the veracity of the references)

Over view of the data

- The data set is a survey of 100(ish) participants
- It was conducted on qualtrics
- They were randomly assigned to 1 of 3 conditions (positive, negative, neutral)
- They were then asked to complete 2 5 item scales
 - The first scale was about their mood
 - The second scale was about their self esteem
 - Both scales produce a total score between 5 and 25
- The data set also includes some demographic information about the participants
 - age
 - Gender

The raw variable names are as follows:

- 'age'
- Q_1 (gender item)
- Q_2 (mood scale item 1)
- Q_3 (mood scale item 2)
- Q_4 (mood scale item 3)
- Q_5 (mood scale item 4)
- Q_6 (mood scale item 5)
- Q_7 (self esteem scale item 1)
- Q_8 (self esteem scale item 2)
- Q_9 (self esteem scale item 3)
- Q_10 (self esteem scale item 4)
- Q_11 (self esteem scale item 5)
- FL_12_DO_positive (condition 1)
- FL_12_DO_negative (condition 2)
- FL_12_DO_neutral (control condition)

Work together!

- You can each take on a different part of the project, and then combine your work at the end
- For example, one person could set up the quarto document, install the apaquarto extension, and write the Yaml front matter
- Some people could look up how to make a new column that is the sum of other columns
- Some people could look up how to make a new column that says which condition the participant was in (combining the 3 columns)
- Some people could look into how to run an Anova on a pandas data frame

You have almost everything you need

- For almost all the operations you have examples in the previous notebooks and slide decks
- But you will need to look up how to do some things
- You can use the internet, the pandas documentation, and you can ask the instructors for help

Good luck!