LING 3040
Independent Project: Method & descriptive statistics, due Feb 25

Part 1: Assessing preliminary validities (20pts)

First, restate the research question that this faculty member was asking with their experimental study (please check for my feedback in the previous assignment to make sure you have the strongest statement of the research question).

- (300-400 words, 10pts) Assess the construct validity of the experiment. Use the following questions to prompt your assessment:
    - How is each variable in the research question operationalized? Is this a good operationalization?
    - What type of measurement is provided for each variable?
    - Not all variables that you see in the dataset are part of the research question. For each variable that you have not discussed yet, what is it operationalizing and how is it measured?
- (300-400 words, 10pts) Assess the external validity of the experiment. Use the following questions to prompt your assessment:
    - What was the target population of the study?
    - Who was in the sample of study? What was the sample size?
    - Will the results be generalizable from the sample to the target population? Are there other factors that may have been overlooked that would limit the opportunity to generalize the results?

Part 2: Descriptive / exploratory statistics (25pts)

Concepts / skills we have learned in the statistics portion of the course so far:

- Measures of central tendency
- Measures of spread
- Identifying & trimming outliers
- Visualization of data

For this portion of the project, you can pick whichever variables interest you the most, and you can subset them however you like (ex. maybe you are interested in comparing the reaction times of some subgroup of the participants to the other). These need not align with the research question stated in Part 1!

Write up a report in which you present the following:

- Select two continuous variables (if your dataset only has one continuous variable, consider subsetting it so that you have two subsets of the continuous variable).
    - Calculate measures of central tendency & spread (4pts)
    - Explain why you chose mean&SD or median&IQR. (2pts)
    - Are there outliers? Show your work in identifying outliers. (4pts)

    - What method did you use to identify and (if applicable) trim? (2pts)
    - What are the new measures of central tendency & spread after trimming? (4pts)
- Present histograms of your two continuous variables. (2pts)
- Provide at least two other plots that show the relationship between two variables (ex. one of the variables from the above and one of the group variables). (4pts)
    - You may choose from bargraphs, boxplots, scatterplots, and line plots, among others, but make sure you justify your choice. For each graph, why is that the most appropriate graph for the given relationship? (2pts)
    - Based on each graph, would you recommend investigating further for a possible significant difference? (1pt)

Be sure to submit both a report **and the R code as a .R file**.

If you do not have variables that conform to the type of analysis I expect from you here, then please discuss with me what you can do instead.