# Research Paper Submission 2 Instructions

For this submission, you need to perform a general analysis of your data in R. You need to report the essential numerical characteristics of your variables and do some visual analysis. The objective here is to obtain a general understanding of the patterns in the data.

Note that the first submission gave you the building blocks to write the Introduction of your Research Paper. This assignment gives you the building block to write the Descriptive Analysis part of your paper.

*Tasks*

You need to use what we have learned so far to describe the variables of interest in your dataset. The variables of interest are the dependent variable, the primary independent variable, and two control variables. These are the variables you mentioned in Submission 1 (unless you were required to change them).

1. Mention and verbally describe the dataset you are using, your dependent variable, primary independent variable, and control variables. Your variables should be either continuous or binary. If your variable is categorical and has more than two categories, transform it into a meaningful binary variable using 1 and 0 as values. Remember to mention how the variables are named in your dataset. Mention if you constructed new variables or deleted observations. Also, note the number of observations in your data (the number of observations must be greater than 100).
2. Present summary statistics for all your variables in a table. Comment on the averages of your variables and anything else you think is interesting or important. Be careful in properly describing the averages of your binary variables.
3. Present a graph showing the relationship between your dependent and primary independent variable. You should present a scatter plot if both of your variables are continuous. If any of the two variables is binary, use a bar plot. Summarize what you learn.
4. Confirm what you found in (3) by finding and reporting the correlation between your dependent and primary independent variable.
5. In one or more tables, report the correlation between your dependent variable and control variables. Comment on what you find. (Feel free to combine 5 and 6.)
6. In one or more tables, report the correlation between your primary independent variable and control variables. Comment on what you find.
7. Given the correlations you find, do you think not accounting for your control variables while studying the relationship between your dependent and primary independent variable might bias your results? If yes, what would be the direction of this bias? Comment separately for each of your control variables.

*How to insert tables and figures in your write-up?*

You should prepare your submission in Microsoft Word. You need to manually export all your tables and figures from R. For tables, you can copy from the output and format them as a table in Word (use a fixed-width font like Courier New). You need to export the figures from R and insert them in word. Finally, save your word document as a PDF.

*What needs to be submitted?*

- A 2–4-page essay in PDF format that outlines the analysis you performed and summarizes your results. You should include all tables and figures in the body of your essay. (All tables and figures should have a name and number. Also, the write-up should be in essay format, with no bullet points.)
- Your R code that reproduces the analysis you performed.
- The dataset used in the analysis if you made significant changes to the data or are using an external dataset (in excel, csv, or rdata format)

Submissions need to be made on Canvas. Any member of the group can submit on your group's behalf.

Here is an example submission with its accompanying code to get you started.

Click here to download the above instructions in PDF format.