Final Course Project: Checkpoint 2
Stat 420/ASRM 450, Fall 2023

Our second project checkpoint is due on Tuesday, November 7th with late submissions accepted until Thursday, November 9th. This checkpoint allows you to start working on your analyses for the project, to identify any potential issues with your dataset of interest, and to reach out to the instructional team for support or with questions.

For this project checkpoint, you will create your own RMarkdown document for analyses & interpretations, knit to pdf, and submit the pdf document to Gradescope. If you are working with a group, only one member of the group needs to submit this document. Be sure that the submitting member includes all team members when submitting to Gradescope.

Instructions for the Project Checkpoint & Analysis:

For this project checkpoint, you will begin analyzing your data. At a minimum, the following tasks should be included in your checkpoint:

1. Load your data into R. Print the first few rows of the dataset and the size (dimensions) of the dataset.
2. Generate & interpret a histogram for your response variable of interest.
3. Generate summary statistics for your response variable. Be sure that the standard deviation is included in this output.
    a. This is the variability that you will be trying to explain with your linear model.
4. Create a scatterplot matrix for your response variable and the quantitative predictor variables of interest. Interpret the scatterplot matrix, including any potential concerns or encouraging features for fitting a linear model.
5. Create the first model for your project, which includes all first order terms for your five predictors (two quantitative, one categorical, + two of any type).
6. Report the fitted model & the R^2 for this model.
7. Generate the default plots in R used to assess the assumptions of the model fit. No interpretation is needed for this assignment, although you may start to think about if these plots indicate that your modeling assumptions are met.

Note that after completing your initial analyses, you may choose to adjust your project goals or possibly your dataset. If you do make these changes, you should note them at the end of your project checkpoint. You are welcome and encouraged to change, adjust, and pivot as needed and as you learn more information about your variables of interest. Creating and fitting a linear model should be an iterative process that uses different techniques to adjust your model using additional information until it is optimal.

Also note that you will be graded on your statistical decisions and the quality of your explanations and presentation rather than on the quality of your fitted models.

Grading for Checkpoint 2:

Each of the 7 tasks in the instructions above will be worth 2 points. The final 2 points for this assignment will be earned for neat & clear formatting of the markdown document.

Instructions for creating the RMarkdown document:

To record your data analysis steps (code & output), you will use an RMarkdown document. Follow these instructions to format this document appropriately.

1. Create a new RMarkdown Document by opening RStudio and selecting ”File > New File > R Markdown…” from the program menu.
2. Change the heading of the document to include an appropriate title, your name(s), the due date of the checkpoint, and indicate that you’d like pdf_document for your output.
3. Delete the default body of the RMarkdown document (starting at line 8 for most).
4. Add your dataset to the folder where your RMarkdown document is stored and read the data into your document.
5. Be sure to keep your code and analysis organized in your document. It may be helpful to include multiple code chunks, headers (using # in the plain text area of the document), or comments (using # in the code chunk of the document). Your final document should be *neatly* formatted.
6. Check to make sure that your code is readable in your final document. That is, make sure that your code isn’t printing off the page. If your code cannot be printed in one line, add a new line within a function/line of code.

Please, please, please do not hesitate to reach out if you have any questions throughout your analysis! We are here to help! Post to campuswire, send an email, or pop into office hours.