Project for STAT 513

You will analyze a data set of your choosing, using (1) a regression method like the type we have discussed in this class, or (2) some suitable method that you have learned in your 500-level STAT classes here at UofSC. If you choose option (2), you will have to have a solid knowledge of the method you choose, in order to apply it. The specific guidelines about the data analysis will mainly relate to option (1).

The data set could be one that is of personal or academic interest to you and that fits the types of regression analysis studied in STAT 513. It should be a real data set which you have not analyzed before and which has not been analyzed in a textbook. You may work individually or in groups of 2 or 3 people. If you work in a group, each person must contribute substantially and submit a separate one-page document (see below).

Project Part I– Data Set Proposal/Description:

You should write a one- to two-page typed description of the data set you propose to study. You should include details about the response/dependent variable, about independent variable(s), and about the number of observations. If there are issues such as obvious potential model violations, comment on these. Discuss the source of the data set, and whether the data come from an experimental or observational study. In addition, please include a printout of the data set (or if it is quite large, a selected part of the data set).

You should also include a general proposal for what sort of analysis (simple linear regression? multiple linear regression? quadratic regression? something else?) you plan to do with this data set. (You could alter your model later if necessary.) Think of one or more hypotheses/research questions that are of interest from the beginning, and mention those. These should be questions that could be answered with the types of inference we studied in Chapter 11.

This part is due on or before Friday, November 4, 2022.

Project Part II – Written Report:

You should write a concise report summarizing your analysis. The report should be no longer than four (typed) pages, not counting any R output, graphs, etc., which you may wish to include as support or illustration for your analysis. For those working in groups, each group member should submit separately a one-page typed document describing his or her contributions to the project and what he or she specifically learned while doing the project.

The style of the report is up to you, but the best reports will address many of the questions and details studied in class when we discussed regression analysis and in the course-webpage examples.

Some things to include (depending on the data set and choice of model) might be:

- An introduction and discussion of the data set itself
- If your data set has several potential independent variables, you may wish to perform some automated model selection before fitting the model (see the example R code for ideas about this)
- A statement of the fitted model (i.e., with estimated parameters) with any relevant interpretations
- Summary results for any relevant hypothesis tests or confidence intervals
- A summary of model assumptions
- A discussion of any diagnostics you examined to look for model violations, along with your findings
- Any remedial action you took to fix such violations, if necessary
- Your overall conclusions about the data, based on your analysis

Part II of the project (the final project report) will be due ***on or before*** **Friday, December 2, 2022**. It will count for 6% of your final grade (1% for Part I, 5% for Part II).