STAT 477/577 - Homework Assignment 1

Due Tuesday, January 25, by 11:59 PM

General homework guidelines: All homework assignments should be submitted using Canvas. Please submit your answers separately, for each of the problems, as set-up in the Canvas submission portal. You are allowed to either type in your answers, as well as submit graphs directly, within each submission item. You are also allowed to submit a scanned copy of your answer, as long as the answers are submitted separately for each question, as instructed. You can either scan your answers (ask if you don't have access to a scanner or do not know how to use your phone to do so), or just submit a picture of your answer. Please note that if we can't read your answer, we won't be able to award any (partial) credit.

You have one attempt to submit your answers. If technical issues appear and your submission portal has closed for some reason, please email Prof. Caragea explaining the situation and requesting permission to resubmit. Please note that such requests must be made before the deadline.

For full credit, please make sure you submit your HW by the deadline. A late submission is possible, with a 20% penalty, as long as it is turned in before the end of the day on Thursday following the due day (for this HW, it is January 27-th). No submissions will be accepted past this date.

Homework problems.

1. **[5 pts.]** Give five examples of categorical random variables (at least two should be ordinal, and your examples should be different from the ones discussed in class). For each variable, specify whether it is Nominal or Ordinal. Please make sure you state clearly what the categories (or levels) are for each of the variables. For the ordinal variables, justify briefly the ordering.

2. **[20 pts.]** In the file **SPRING2022STAT_477_577.csv** you will find information on the first Major, College and Classification year for each of the students in our class, as of Friday, January 14-th. From this data set, select the variables major and classification year and write a summary of the distribution for each variable. Make sure to use R and include a **summary table**, a **bar graph** and a **pie chart** for each of these two variables in your summary. Include a brief commentary of what you have learned about the profile of the students in the class from creating these summaries, and the shortcomings of the pie chart for representing these data.

   To produce the required graphs, please use the R script provided in Unit 1, Section 1 for a list of the commands you should be able to master. You are required to use ggplot in R to produce the barchart, but you can choose to use commands in base R to produce the pie charts (sample commands are provided within the R script).

   The goal for this problem is not only to have you think on how to present and describe a variable in a data set, but also to get you familiar with R and ggplot, which we will be using extensively in this class.