STAT 6012: Linear Models for Data Science
Class Activity 3

Due date: Wednesday, July 18 by 10:50 am Via Canvas.

Complete the following questions an R Markdown file and submit your compiled HTML file. If you are working in a group, list the names (last, first) of the group members in alphabetical order of last names.

Consider the same dataset from Class Activity 2: A fast-food chain plans to add a new item to its menu. However, they are still undecided between three possible marketing campaigns for promoting the new product. In order to determine which promotion has the greatest effect on sales, the new item is introduced at locations in several randomly selected markets. A different promotion is used at each location, and the weekly sales of the new item are recorded for the first four weeks.

The attached dataset contains information on the data collected for the marketing campaign. Read more about this dataset and the variable descriptions:

https://www.kaggle.com/datasets/chebotinaa/fast-food-marketing-campaign-ab-test?resource=download

The fast-food chain is particularly interested in know if there is a difference in the average sales amount across all the three promotions.

1. [3] As a data scientist, the first step you want to take is explore the data with the research question (is the average sales amount is the same across all three promotions?) in mind. Make an appropriate graph for the research question.

2. [1] What analysis should you conduct to investigate if the research question.

3. [2] For the research question, state the null and alternative hypothesis in symbols **and also in words**.

4. [3] Conduct the analysis in R. State the p-value and write the conclusion of your investigation, in context. Test at $\alpha = 0.05$.

5. [1] Based on your conclusion from Question 4, would you need to do a post hoc test to investigate which promotion yields the most sales, on average? Explain briefly.