DSA 8010 - Final exam (25 points)

Part 1 - data analysis (10 points)

Part 1 of the exam is a written portion similar to mini-project 3. Please submit your response to Part 1 through Canvas by 11:59 pm on Tuesday, December 10. You may, but are not required to, work in a group of up to 4 students. Each student must turn in the report document.

For each of the following questions, give your answer as a short report that answers the key question and supports your answer using the data. Descriptive and inferential methods should be used. Make sure that every statistic, plot, or table that is included is also described in the text.

1. (5 points) An agricultural research team in Idaho investigated the effect of Fuji apple trees' branch configuration on the fruit produced by the trees. They planted 24 trees in an experimental orchard in three rows. Each row was treated to promote one of three branch configurations: "overlapping arm," "tipped arm," and "tall spindle." Each of the treatments was applied to eight trees.

In the 2012 growing season, the trees' fruit was harvested and the average weight per apple (grams) was recorded for each tree. The data are found in the file `fuji_apples.csv`.

Key question: Do the data suggest that the branch configuration affects the fruit weights?

2. (5 points) The dataset `crash_maryland.csv` contains records of vehicle crashes in Montgomery County, Maryland for which police reports were filed between 2015 and 2023. The posted data reflects a subset of all crashes: those that occurred during rain, snow, or sleet.

A team of government officials is looking to better understand conditions under which serious traffic accidents occur. One question they are interested in is whether the type of precipitation is related to increased chance of injury. This question is related to the variables `Weather` and `Injury.Severity`.

Key question: Do the data suggest that the weather condition is associated with injury severity? Under which condition(s) are crashes more likely to occur?

Rubric for Q1 and Q2:

Give an answer of 3-4 paragraphs. Make sure to include a check of the model assumptions (e.g. Normality, adequate sample size in each group, expected cell count high enough, etc.) as well as consideration of data quality.

Category	Points
The results are presented clearly in a readable, narrative form. The writing is clear and concise. All quantitative results, including plots, tables, and summary statistics are referenced and interpreted in the text. Captions, graph titles, and axis labels are included where appropriate.	1
A brief and informative descriptive summary is given. The summary statistics are correct and their relevance is described in the text.	1
An appropriate inferential procedure is chosen and implemented correctly. An accurate conclusion is given based on the results.	2
Assumptions of the statistical model are checked for each analysis. A thoughtful and accurate verbal conclusion is given. It summarizes findings with respect to the key question. Limitations or complications of analysis are discussed.	1

Part 2 (15 points)

There will be about 20 quiz questions posted in Canvas. This portion is cumulative. You will have two attempts and 120 minutes for each attempt. Corrections will not be accepted for the Canvas portion. The quiz will be open in Canvas between 12pm on Friday, Dec 6 until 11:59pm on Wednesday, Dec 11.