Taniya Adhikari

DSC 520

Assignment 3 Analysis

1. **What are the observational units in this study?**

In Statistics, measured characteristics are called variables of subjects called observational units. For every study, numerical values assigned to the variables are called observations. In this study, professor is measuring Test Scores for each section, in this case **Two Types of Section, regular and sports** is the Observational Unit.

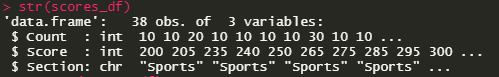
However, one can have many observational units in a study, if the professor was measuring Test Scores for each student then students will be observational unit.

1. **Identify the variables mentioned in the narrative paragraph and determine which are categorical and quantitative?**



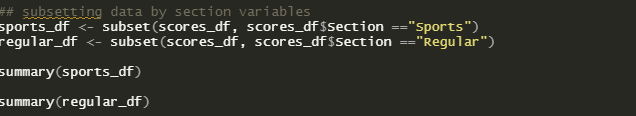
Using above code, Quantitative variable are Count and Scores because it is numerical values and Categorical variable is Section because it’s string variable

**Output**

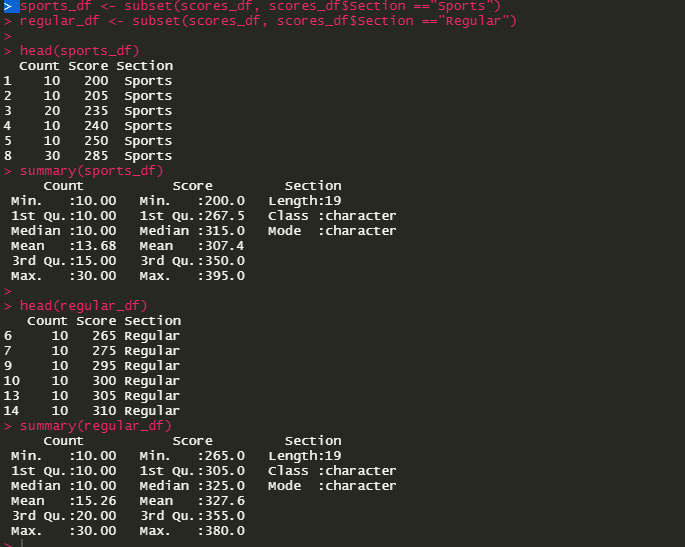


**3. Create one variable to hold a subset of your data set that contains only the Regular Section and one variable for the Sports Section.**

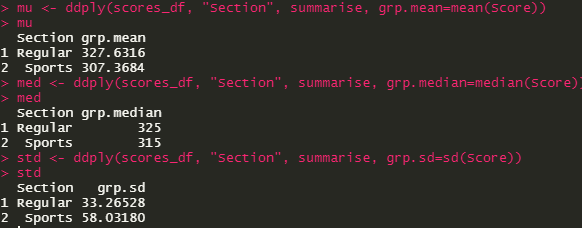
**R Code**



**Output**



**4. Use the Plot function to plot each Sections scores and the number of students achieving that score. Use additional Plot Arguments to label the graph and give each axis an appropriate label. Once you have produced your Plots answer the following questions:**



**4a. Comparing and contrasting the point distributions between the two section, looking at both tendency and consistency: Can you say that one section tended to score more points than the other? Justify and explain your answer.**

A screenshot of a cell phone

Description automatically generated

A screenshot of text

Description automatically generated

Using the Histogram comparison of two section, On average students in regular Section seem to consistently score more points compared to sports section. The mean (327) and median (325) score for regular section is also higher than sports section mean (307.4) and median (315). However, Sports section has higher standard deviation (58.03) compared to regular section standard deviation (33.26), this means sports section didn’t consistently scored higher points, and this section has larger spread. Box plot also shows the larger spread in sports section,

b. Did every student in one section score more points than every student in the other section? If not, explain what a statistical tendency means in this context.

A close up of a map

Description automatically generated

On average, regular section has outperformed the sports section, but not every student scored low in sports section and not every student scored high in regular section. The lowest score for regular section was 265 and highest score for sports section was 395. By looking at the density plot, both section’s distribution is slightly skewed to the left, showing that students in both section tend to score higher, however, it shows that the Regular Section has more students who scored higher points compared to Sports section.

c. What could be one additional variable that was not mentioned in the narrative that could be influencing the point distributions between the two sections?

Scores are usually distributed in multiple categories, such as assignments, exams, quizzes and projects. Total scores of a section is a sum of these categories. Therefore, further distributing the scores into additional categorical variable will show a better breakdown of a score. For example, one student may have high score in exams, but low score in project, likewise one student may turn in all assignments but had low test scores. This will show how each student performed in assignments, exams, etc. Breaking down score distribution will show how students performed in each category of score in each section and comparing those result by section will give better insights on performance. Thus, more data is needed to make any conclusions.