MAT 243 Project One Summary Report

Alyssa Bailey
alyssa.bailey1@snhu.edu
Southern New Hampshire University

1. **Introduction: Problem Statement**

Discuss the statement of the problem in terms of the statistical analyses that are being performed. In your response, you should address the following questions:

- *What is the problem you are going to solve?*
- *What data set are you using?*
- *What statistical methods will you be using to do the analysis for this project?*

The problem that I am attempting to solve is to assist the coach and management of an NBA team in making future decisions. This will improve overall team performance. I am doing this by analyzing descriptive statistics and data visualization. The data sets I will use are the team's points per game and the points allowed by their opponents. I will analyze the data by calculating a team's mean, median, standard deviation, and variance over a specified time period using statistical methods.

2. **Introduction: Your Team and the Assigned Team**

In this project, you picked a team and you were assigned a team to do comparative analysis.

See Steps 1 and 2 in the Python script to address the following items:

- *What team did you pick and what years were picked to do the analysis?*
- *What team and range of years were you assigned for the comparative study? (Hint: This is called the assigned team in the Python script.) Present this information in a formatted table as shown below.*

Table 1. Information on the Teams

	Name of Team	Assigned Years
1. Yours	Lakers	2013 - 2015
2. Assigned	Bulls	1996-1998

The team I picked was my favorite NBA team, the Lakers. The years assigned to do the analysis were 2013-2015. I was initially assigned the Bulls for the comparative study.

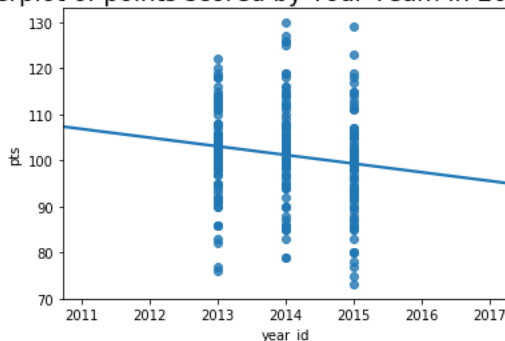3. **Data Visualization: Points Scored by Your Team**

In the Python script, you created a visualization for the distribution of points scored by your team.

See Step 3 in the Python script to address the following items in a paragraph response:

- *In general, how is data visualization used to study data distributions and trends?*
- *In this activity, you were asked to pick one of the two plots that best describes the data distribution of the variable for your team. Include a screenshot of this plot in your report.*
- *Why did you pick this plot? Explain.*
- *What can you say about the distribution of the variable by visually inspecting this plot? What does this signify?*

Data visualization aims to study data distributions and trends by giving the viewer an overall picture of how things are going, for example, the NBA team. Charts and graphs can convey a lot of information. This is especially true when the information contains a large number of averages spanning over time or occurrences. Following is a scatterplot that I chose to represent the data best. Based on this scatterplot, I can easily see how many points the Lakers scored per game over the 2013 - 2015 NBA season. The graph shows how many points were scored per game and in which years on the x-axis and y-axis. There was an average game score of just over 100, with most points scored between 90 and 110.

Scatterplot of points scored by Your Team in 2013 to 2015

4. Data Visualization: Points Scored by the Assigned Team

In the Python script, you created a visualization for the distribution of points scored by the assigned team.
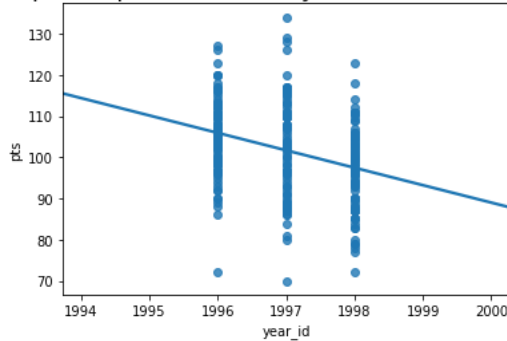
See Step 4 in the Python script to address the following items in a paragraph response:

- *In this activity, you were asked to pick one of the two plots that best describes the data distribution of the variable for the assigned team. Include this plot in your report.*
- *Why did you pick this plot? Explain.*
- *What can you say about the distribution of the variable by visually inspecting this plot? What does this signify?*

To demonstrate how much time was spent scoring, I used a scatterplot to show the Chicago Bulls' difference in total scores during the 1996-1998 NBA season. An accurate representation of an NBA team's scoring performance can be obtained with this type of visual. Based on the

scoring scatterplot graph of the Bulls' 1996-1998 season, it is evident that despite a few outliers, the team's scoring performance was nearly average. However, the team scored fewer points per game each year.

Scatterplot of points scored by the Bulls in 1996 to 1998

5. Data Visualization: Comparing the Two Teams

In the Python script, you created a visualization for the difference in the distributions of points scored by your team and the assigned team.

See Step 5 in the Python script to address the following items in a paragraph response:

- *In general, how is data visualization used to compare two different data distributions?*
- *In this activity, you were asked to pick one of the two plots that best compares the data distributions of your team with the assigned team. Include a screenshot of this plot in your report.*
- *Why did you pick this plot? Explain.*
- *How do the two distributions compare to each other?*

By comparing two different data distributions side by side, data visualizations can provide insight into the differences between their performance. I used a boxplot to compare the 1996-1998 Bulls with the 2013-2015 Lakers. In years measured with a higher mean score per game, the Lakers scored slightly higher than the Bulls. From the graph, we can see that the Lakers' box and the line across the middle are higher than the Bulls' box

Boxplot to compare points distribution
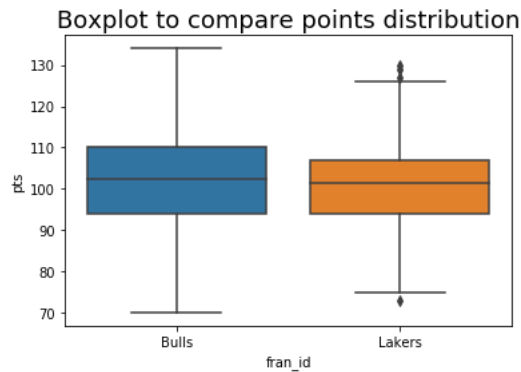
6. Descriptive Statistics: Points Scored By Your Team in Home Games

In the Python script, you calculated descriptive statistics on the points scored by your team in games played at home venue. These included the mean, median, variance, and standard deviation for the relative skill of your team.

See Step 6 in the Python script to address the following items:

- *Summarize **all** statistics in a formatted table as shown below. Use one row for each statistic. You will need to add rows to the table in order to include all of your statistics.*

Table 2. Descriptive Statistics for Points Scored by Your Team in Home Games

-	Value
Mean	101.7
Median	102.0
Variance	149.18
Standard Deviation	12.21

- *In general, how are the measures of central tendency and variability used to analyze a data distribution?*
- *Interpret each statistic in detail and explain what it represents in this scenario.*
- *Use the mean and the median to describe the distribution of points scored by your team in home games.*
  - *Describe the skew: Is it left, right, or bell-shaped?*
  - *Explain which measure of central tendency is best to use to represent the center of the distribution based on its skew.*

When analyzing data distribution, distribution refers to the frequencies of different responses while central tendency measures provide the average for each response. The mean represents the sum of all numbers in the given dataset divided by the number of numbers in the set. The median is the middle number in the set. The variance is the average of the squared differences

from the mean while the standard deviation is the square root of the variance. Regarding the measure of central tendency, the skew is represented because of its bell-shape.

7. Descriptive Statistics: Points Scored By Your Team in Away Games

In the Python script, you calculated descriptive statistics on the points scored by your team in games played at opponent's venue (away). These included the mean, median, variance, and standard deviation for the relative skill of the assigned team.

See Step 7 in the Python script to address the following items:

- *Summarize **all** statistics in a formatted table as shown below. Use one row for each statistic. You will need to add rows to the table in order to include all of your statistics.*

Table 3. Descriptive Statistics for Points Scored by Your Team in Away Games

Statistic Name	Value
Mean	100.71
Median	101.0
Variance	88.16

- *Interpret each statistic in detail and explain what it represents in this scenario.*
- *Use the mean and the median to describe the distribution of points scored by your team in away games.*
    - a. *Describe the skew: Is it left, right, or bell-shaped?*
    - b. *Explain which measure of central tendency is best to use to represent the center of the distribution based on its skew.*
- *Is your team performing better in games played at home than those played away? Use the mean and the standard deviation to answer this question. What can be deduced by comparing the standard deviation of points scored in home games and points scored in away games?*

When analyzing data distribution, distribution refers to the frequencies of different responses while central tendency measures provide the average for each response. The mean represents the sum of all numbers in the given dataset divided by the number of numbers in the set. The median is the middle number in the set. The variance is the average of the squared differences from the mean while the standard deviation is the square root of the variance. Regarding the measure of central tendency, the skew is represented because of its bell-shape.

The skew for both teams appears bell-shaped. My team has a smaller variance and standard deviation than the 1996-1998 Bulls. According to the data, the Lakers have a lower mean and median and were more consistent with scoring that more closely matched their mean score. This implies the Lakers have more consistent skill than the Chicago Bulls.

8. Confidence Intervals for the Average Relative Skill of All Teams in Your Team's Years

In the Python script, you calculated a 95% confidence interval for the average relative skill of all teams in the league during the years of your team. Additionally, you calculated the probability that a given team in the league has a relative skill level less than that of the team that you picked.

See Step 8 in the Python script to address the following items:

- *Report the confidence interval in a formatted table as shown below.*

Table 4. Confidence Interval for Average Relative Skill of Teams in Your Team's Years

Confidence Level (%)	Confidence Interval
95%	(1502.02, 1507.18)

- *Describe how confidence intervals are generally used in estimating the measures of central tendency for a population.*
- *Provide a detailed interpretation of the confidence interval in terms of the average relative skill of teams in the range of years that you picked.*
- *How would your interval be different if you had used a different confidence level?*
- *What is the probability that a given team in the league has a relative skill level less than that of the team that you picked? Is it unusual that a team has a skill level less than your team?*

When estimating the measures of central tendency for population confidence levels are used to quantify the uncertainty by providing a lower limit and upper limit. These limits represent a range of values that represent the true population parameter. It is given that a 95% confidence interval for the average relative skill of all teams in the NBA during the 2013-2015 seasons is 1502.02 as the lower limit and 1507.18 as the upper limit. We are 95% confident that the average relative skill of all teams in the league is less than that of the Lakers. It is not unusual that a team has a skill level less than due to our consistent average of high scores between the 2013-2015 season.

9. Confidence Intervals for the Average Relative Skill of All Teams in the Assigned Team's Years

In the Python script, you calculated a 95% confidence interval for the average relative skill of all teams in the league during the years of the assigned team. Additionally, you calculated the probability that a given team in the league has a relative skill level less than that of the assigned team.

See Step 9 in the Python script to address the following items:

- *Report the confidence interval in a formatted table as shown below.*

Table 5. Confidence Interval for Average Relative Skill of Teams in Assigned Team's Years

Confidence Level (%)	Confidence Interval
95%	(1487.66, 1493.65)

- *Provide a detailed interpretation of the confidence interval in terms of the average relative skill of teams in the assigned team's range of years.*
- *Discuss how your interval would be different if you had used a different confidence level.*
- *How does this confidence interval compare with the previous one? What does this signify in terms of the average relative skill of teams in the range of years that you picked versus the average relative skill of teams in the assigned team's range of years?*

As the level of confidence increases, the width increases. If we consider a confidence level of 99%, the difference between the upper and lower limits will be higher than that of a confidence level of 95%.

10. Conclusion

Describe the results of your statistical analyses clearly, using proper descriptions of statistical terms and concepts.

- *What is the practical importance of the analyses that were performed?*
- *Describe what these results mean for the scenario.*

The practical importance of the analyses of the calculations of standard deviation, variance and confidence level, as well as intervals, enables the ability to test and prove each team's success. Using graphs and calculations allowed us to compare each team's mean and median of points scored. The results concluded the Lakers performed better than the Bulls.