# MAT 243 Project One Summary Report

Nicholas Cleveland

[nicholas.cleveland1@snhu.edu](mailto:nicholas.cleveland1@snhu.edu)

Southern New Hampshire University

## Introduction: Problem Statement

We are a basketball data analyst and have been assigned to analyze performance patterns of basketball teams. This will be done using descriptive statistics and data visualization techniques in order to study the distributions of key performance metrics of the basketball teams.

For the data set, we have been given data with the name of the team, points scored by each team in a game, a measure of the relative skill level of the team in the league, and the year when the team played the games.

The statistical methods that I’ll be using to do an analysis will be calculating the mean the median, the variance, the standard deviation, and the confidence intervals.

## Introduction: Your Team and the Assigned Team

The team that I was chose was the **Cleveland Cavaliers** from 2013 to 2015.

We generated the following data from our Python program. We calculated the 2013-2015 Cleveland Cavaliers observations and displayed the first 5 out of 246:

Graphical user interface, text

Description automatically generated

For the comparative study, I was assigned the **Chicago Bulls** from 1996 to 1998.

We generated the following data from our Python program. We calculated the 1996-1998 Chicago Bulls observations and displayed the first 5 out of 246:

Table

Description automatically generated with medium confidence

|  | **Name of Team** | **Assigned Years** |
| --- | --- | --- |
| 1. Yours | Cleveland Cavaliers | 2013 - 2015 |
| 2. Assigned | Chicago Bulls | 1996 - 1998 |

Table 1. Information on the Teams

## Data Visualization: Points Scored by Your Team

Data visualization is used to show a graphical representation of generated data so that it can be read and understood in a more comprehensive way to not just the analyst, but to the general readers of the report.

For my team, the Cleveland Cavaliers, I chose a **Histogram** as this graphical representation depicts the data of the team’s scores in a more comprehensible way. We are not yet comparing data from another team, so this graph makes the most sense in providing the scores among one team to show the distribution of points and their frequencies.

By visually inspecting this plot, the distribution is centered around 90 to 100 points, which indicates the team may score in this range most of the time. This can be further analyzed with statistical methods, but this is what we can intuitively tell from just looking at the Histogram.

Chart, histogram

Description automatically generated

## Data Visualization: Points Scored by the Assigned Team

I picked this plot as it shows the distribution of points scored by the team that I was assigned, the Chicago Bulls in 1996 to 1998.

* *What can you say about the distribution of the variable by visually inspecting this plot? What does this signify?*

From visually inspecting the distribution of the scores on this plot, it shows that there was a distributed frequency of the number of points that the Bulls scored in 1996 to 1998 where they had a higher frequency in the range from about 105 to 110 points. This

Chart, histogram

Description automatically generated

## Data Visualization: Comparing the Two Teams

For the comparison of both teams, I decided to choose a Box Plot as it shows the ranges of points, the mean and the outliers. It gives more information about each data than just a mean and dispersion of points made.

Chart, box and whisker chart

Description automatically generated

Comparing both box plots, we can see that the Bulls had a higher mean than the Cavaliers. The Bulls also had a larger number of outliers in the data that scored higher and lower than the outliers found in the Cavaliers plot.

## Descriptive Statistics: Relative Skill of Your Team

| **Statistic Name** | **Value** |
| --- | --- |
| Mean | 1427.57 |
| Median | 1390.92 |
| Variance | 9939.24 |
| Standard Deviation | 96.64 |

Table 2.

Descriptive Statistics for Relative Skill of Your Team (Cleveland Cavaliers)

*In general, how are the measures of central tendency and variability used to analyze a data distribution?*

In general, measures of central tendency and variability are usually manifested in the forms of statistical function such as the mean, median, variance, and standard deviation. When these components are used together to describe the data frame in question, it shows various patterns that can lead to conclusions with strong evidence.

Looking at the relative skill value of my team, the Cleveland Cavaliers, it should be noted that there is a significant difference between the mean and median of the values. By subtracting the median from the mean (1427.57 – 1390.92) we get approximately a difference of 36.65 relative skill value points. Due to this, outcome, we can say that this is a **positive skew**, which indicates a skew to the **right**. The relative skill of my team had a large variance in the data, which tells us that there is a large spread in the relative skill points.

* + *Explain which measure of central tendency is best to use to represent the center of the distribution based on its skew.*

Based on this right skew,

## Descriptive Statistics: Relative Skill of the Assigned Team (Chicago Bulls)

| **Statistic Name** | **Value** |
| --- | --- |
| Mean | 1739.80 |
| Median | 1751.23 |
| Variance | 2651.55 |
| Standard Deviation | 51.49 |

Table 3.

Descriptive Statistics for Relative Skill of the Assigned Team (Chicago Bulls)

* *Interpret each statistic in detail and explain what it represents in this scenario.*
* *Use the mean and the median to describe the distribution of relative skill of the assigned team.*
  1. *Describe the skew: Is it left, right, or bell-shaped?*
  2. *Explain which measure of central tendency is best to use to represent the center of the distribution based on its skew.*
* *Use the variance and the standard deviation to compare the distributions of relative skill of your team and relative skill of the assigned team. Which of the two teams has a more consistent skill? How do you know?*

## Confidence Intervals for the Average Relative Skill of All Teams in Cavaliers’ Years (My Team)

| **Confidence Level (%)** | **Confidence Interval** |
| --- | --- |
| 95% | [1502.02, 1507.18] |

Table 4.

Confidence Interval for Average Relative Skill of Teams in Cleveland Cavaliers’ Years

Confidence intervals are generally used to estimate the measures of central tendency by

* *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.*

If we had chosen a different confidence interval, this would affect both our confidence range and our amount of margin error. If we chose a smaller confidence level, liked 90%, the range would have become smaller as the amount of error would have become smaller. If we chose a higher confidence level, such as 99%, our range would have increased and therefore our margin for the error would have also increased.

* *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?*

The probability that a given team in the league has a relative skill level less than that of the Cleveland Cavaliers is 0.9732.

This would not be unusual as

## Confidence Intervals for the Average Relative Skill of All Teams in the Chicago Bulls’ Years (Assigned Team)

| **Confidence Level (%)** | **Confidence Interval** |
| --- | --- |
| 95% | [1487.66, 1493.65] |

Table 5.

Confidence Interval for Average Relative Skill of Teams in Assigned Team’s Years (Chicago Bulls)

* *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.*

The confidence interval calculated from the data in our Python program as well as our confidence level shows the

If we had chosen a different confidence interval, this would affect both our confidence range and our amount of margin error. If we chose a smaller confidence level, liked 90%, the range would have become smaller as the amount of error would have become smaller. If we chose a higher confidence level, such as 99%, our range would have increased and therefore our margin for the error would have also increased.

* *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?*

The confidence interval for the average relative skill of all teams in the Cleveland Cavaliers’ years, which was **2013 – 2015**, was **[1502.02, 1507.18]**. The confidence interval for the average relative skills of all teams in the Chicago Bulls’ years, which was **1996 – 1998** was **[1487.66, 1493.65]**. We can see that there was an increase in ranges from the 1996-1998 sample versus the more recent 2013 – 2015 sample.

If we look at the difference of each of these intervals, we can see that the teams in 2013-2015 had a 95% confidence interval difference of **5.16** while the teams from 1996-1998 had a 95% confidence interval difference of **5.99**.

## 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?*

*In conclusion,*

* *Describe what these results mean for the scenario.*

## Citations

*You were* ***not*** *required to use external resources for this report. If you did not use any resources, you should remove this entire section. However, if you did use any resources to help you with your interpretation, you* ***must*** *cite them. Use proper APA format for citations.*

Insert references here in the following format:

Author's Last Name, First Initial. Middle Initial. (Year of Publication). Title of book: Subtitle of book, edition. Place of Publication: Publisher.