# 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 (Cleveland Cavaliers)

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. From visually inspecting the distribution of the scores on this plot, it shows there is a trend to the right after some median value in the data. This trend is decreasing and is resembling a skew right pattern.

.

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. A histogram will show prominent values in the graph such as the median where we can measure central tendency. In this histogram, the number of points made is compared to the frequency of each. There is prominent mean and median show as the graph depicts a bell-curve shape like pattern.

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 (Cleveland Cavaliers)

| **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)

As shown in the table above, the Cleveland Cavaliers had a mean of 1427.57, a median of 1390.92, a variance of 9939.24, and a standard deviation 96.64. 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, as shown. 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. We can see this pattern depicted in the Histogram attached above.

Based on this right skew, the median would be best chosen to represent the center of the distribution as we can that after the median, the graph will start skewing to the right and the frequency of scores will decrease.

## 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)

The relative skill that was calculated during the 1996-1998 period for the Chicago Bulls had a mean of 1739.80. a median of 1751.23, a variance of 2651.55 and a standard deviation of 51.49. Since the mean is 1739.80 and the median is 1751.23, this leaves the median with a higher score of a difference of merely 11.43. This graph takes on a **bell-shaped curve** pattern as the mean and median are so close in proximity.

We can also compare these statistics with the team that I chose, the Cavaliers. The Cavaliers had a mean of 1427.57, a median of 1390.92, a variance of 9939.24, and a standard deviation 96.64. The variance and standard deviation in this case are much larger than the ones found in the table above for the Bulls. This tells us that the Cavaliers had a larger spread and less central tendency than the Bulls did. The Bulls are more consistent in this case as their standard deviation was almost half of the Cavaliers (51.49).

## 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 calculating a **mean**, **median**, or **mode**. The confidence level that was used for the Cavaliers years of 2013-2015 was 95% and the program calculated a confidence interval of [1502.02, 1507.18]. This means that the skill level with falls into this range with a 95% confidence level.

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.

The probability that a given team in the league has a relative skill level less than that of the Cleveland Cavaliers is **0.2478**. This is only about a ~25% probability. This would be unusual for another team to have a probability near this, due to the contrast of this being approximately **a ~75.22% probability of having a higher** skill level than the Cavaliers. This shows a low relative skill level for the Cleveland Cavaliers in the 2013-2015 period.

## 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)

The confidence interval calculated from the data in our Python program as well as our confidence level shows the range that we predict to be 95% accurate. We calculated a probability that a team has average relative skill less than the average relative skill of the Chicago Bulls in 1996-1998 to be **0.9732**. This means that there is a 97.32% probability that the other teams during this season will score less than the Bulls. This shows a great level of skill in the Chicago Bulls team during this period.

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.

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

In conclusion, after gathering the data on each of our basketball teams, we can conclude that each team’s data was able to successfully draw results to calculate measures of central tendency, variance, and confidence intervals. The experiment behind this study was to find how a team that I chose, compared against the assigned team. The Chicago Bulls during the period of 1996 – 1998 had a higher average score than the Cleveland Cavaliers during the period of 2013-2015. In comparison to the other teams’ scores that played during the 1996-1998 period and 2013-2015 period, there was only a 2.68% chance that another team would pass the Chicago Bulls in skill level in 1998 and over a 75.22% chance that another team would pass the Cleveland Cavaliers in skill level in 2013-2015.

## Citations

*2.6 Skewness and the mean, median, and mode - introductory business statistics*. OpenStax. (n.d.). Retrieved November 14, 2021, from https://openstax.org/books/introductory-business-statistics/pages/2-6-skewness-and-the-mean-median-and-mode.

Bhandari, P. (2020, October 26). *Central tendency: Mean, median and Mode*. Scribbr. Retrieved November 14, 2021, from <https://www.scribbr.com/statistics/central-tendency/>.

Gong, D. (2020, November 21). *How to choose the most appropriate chart?* Medium. Retrieved November 14, 2021, from https://towardsdatascience.com/which-chart-to-choose-4b21929539eb.