# MAT 243 Project One Summary Report

Sarah Steinbaum

sarah.steinbaum@snhu.edu

Southern New Hampshire University

## Introduction: Problem Statement

The coach and management team have asked me to conduct an analysis to further improve team performance. I will be using data-driven analytics to help management make decisions that will advance the team’s performance. I will use a large set of historical data from the National Basketball Association, NBA, to conduct my analysis. The statistical methods I will be using for this analysis are data visualization, descriptive statistics, and confidence intervals. I will also be utilizing the programming language, Python. The problem statement I will be addressing is: How do we improve team performance?

## Introduction: Your Team and the Assigned Team

The team I picked for this analysis is the Atlanta Hawks from the years 2013 – 2015. The assigned team for the comparative study is the Chicago Bulls from 1996 – 1998.

Table 1. Information on the Teams

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

## Data Visualization: Points Scored by Your Team

Data visualization is the visual representation of data that provides us with an overview of the data in a graph or chart which makes the data easier to understand and analyze patterns or trends in data. The plot I chose to display the points scored by the Atlanta Hawks in 2013 to 2015 is the scatterplot. I chose the scatterplot over the histogram because it illustrates data on the number of points for each game the Hawks scored in 2013, 2014, and 2015. The scatterplot also shows outliers such as one low scoring game that occurred in 2013. It also directly demonstrates an upward trend showing the Hawks scored more points with each consecutive year between 2013 and 2015. See below. Although the histogram displayed data in a clear, organized manor, the scatterplot seems to represent the data more appropriately.

*Chart, icon

Description automatically generated*

## Data Visualization: Points Scored by the Assigned Team

I chose the scatterplot for similar reasons as to why I chose the scatterplot for the Hawks graph. The scatterplot seems to best represent the data for the Chicago Bulls from 1996 to 1998 because it displays the number of points scored per game in each for 1996, 1997, and 1998. It also shows outliers in the data such as when the Bulls scored very low or high during a game for each season. The trendline on the scatterplot for the Chicago Bulls in 1996 to 1998 showed a downward trend, meaning they scored less points with each consecutive year. See below.

*Chart

Description automatically generated*

## Data Visualization: Comparing the Two Teams

Data visualization coverts data into a visual form which provides a better overview of the data. This can be helpful when comparing two different data distributions because it allows the viewer to analyze data side by side within the same chart or graph. A boxplot will be the best way to show this data. The boxplot below provides us with a lot of information to compare the datasets from the Chicago Bulls and the Atlanta Hawks. The Bulls had a higher range than the Hawks. They both have a similar mean. Overall, the Chicago Bulls have a slightly higher performance than the Hawks.

Chart, box and whisker chart

Description automatically generated

## Descriptive Statistics: Points Scored By Your Team in Home Games

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

| **Statistic Name** | **Value** |
| --- | --- |
| Mean | 101.19 pts |
| Median | 102.0 pts |
| Variance | 138.47 pts |
| Standard Deviation | 11.77 |

The measures of central tendency demonstrates the central location of a distribution and the mean, median, and mode. Variability is the variance and standard deviation while also indicating the spread of the data. The mean is just the average of the data. In this scenario the mean is 101.19 which means that this is the average number of points scored with home games the Hawks scored between 2013 to 2015. The median is the middle value of a dataset. In this circumstance, the median is 102.0 which means this is the mid-value score of the team. The variance is the average of the square difference from the mean, which in this case, is 138.47. Standard deviation is the square root of the variance and is 11.77 in this scenario. The standard deviation explains how the data is spread out from the average value.

## Descriptive Statistics: Points Scored By Your Team in Away Games

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

| **Statistic Name** | **Value** |
| --- | --- |
| Mean | 99.74 pts |
| Median | 101.0 pts |
| Variance | 130.77 pts |
| Standard Deviation | 11.44 |

The mean in this scenario is 99.74 which means the this is the average number of points the Hawks scored in away games between 2013 to 2015. The median is the mid-point value the Hawks scored in away games which is 101.0. The variance in this circumstance is 130.77 which represents the average of the square difference from the mean value. The standard deviation represents how the data is spread from the mean which is 11.44 in this scenario. The median is 101.0 which is roughly similar to the mean, 99.74. Since the measures are close to each other, the distribution is considered bell-shaped. Since the median is less effected by outliers, it is the best measurement of central tendency to use for this distribution. The Hawks had a mean score of 101.19 for home games and 99.74 for away games. The standard deviation was 11.77 for home games and 11.44 for away games. Both means and standard deviations are close. However, the lower a standard deviation is, the more reliable it becomes. Therefore, since the standard deviation for the away games is less than the standard deviation for home games, the Hawks played slightly better at away games than they did home games.

## Confidence Intervals for the Average Relative Skill of All Teams in Your Team’s Years

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

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

Confidence intervals are a range of values between an upper and lower limit around the sample mean, with a certain probability. The confidence intervals are generally used in estimating the measures of central tendency for a population by creating a range of likely distributions. The confidence interval in terms of the average relative skill of teams between 2013 to 2015 is (1502.02, 1507.18) with a 95% confidence level, which is the level that is used to determine the interval. The confidence interval is calculated by subtracting the margin of error from the point estimate and then adding the margin of error to the point estimate. If we had used a different confidence level, the confidence interval would have also changed. If the confidence level decreases, so will the confidence interval. Likewise, if the confidence level increases, the confidence interval does as well. The probability that a given team has a skill level less than the Hawks is 0.6203. This is considered Choice 2 from the python script which used the function norm.cdf. This is not considered unusual because it is greater than a probability of 0.5.

## Confidence Intervals for the Average Relative Skill of All Teams in the Assigned Team’s Years

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

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

The confidence interval in terms of the average relative skill of teams between 1996 to 1998 is (1487.66, 1493.65) with a 95% confidence level, which is the level that is used to determine the interval. The confidence interval is calculated by subtracting the margin of error from the point estimate and then adding the margin of error to the point estimate. If I had used a different confidence level, the confidence interval would have also changed. If the confidence level decreases, so will the confidence interval. Likewise, if the confidence level increases, the confidence interval does as well. Both confidence intervals used a 95% confidence level. This confidence interval (1487.66, 1493.65) is lower than the other confidence interval (1502.02, 1507.18). This signifies that the average level of skill of all teams was higher between the years of 2013 to 2015 compared the years 1996 to 1998.

## Conclusion

Analyzing data incredibly important for businesses to make well informed decisions. Accompanied with data visualization, it helps demonstrate the numbers and information in a method that is easier to understand. In this scenario, I used data visualization, descriptive statistics including central tendency and variability, and confidence intervals to conduct my analysis. The data visualizations included scatterplots and a boxplot to demonstrate the number of points scored in each game the Hawks played between 2013 to 2015 as well as the number of points scored each game that the Bulls played between 1996 to 1998. The boxplot presented the overall data between the two teams and their respective years played for the dataset. The descriptive statistics used mean, median, variance, and the standard deviation to determine if the Hawks played better at home or away games as well as the overall relative skill and performance level. The confidence intervals provided a range of probability. Overall, the Chicago Bulls had a decrease in their performance between the years of 1996 to 1998 and the Atlanta Hawks had an increase in their performance between the years of 2013 to 2015.