# MAT 243 Project One Summary Report

Kenneth Fernandez

Kenneth.fernandez1@snhu.edu

Southern New Hampshire University

## Introduction: Problem Statement

In this project we will be analyzing data sets from the National Basketball Association (NBA) for two teams. We will be using descriptive statistics and data visualization to study distributions of key variables that are associated with the performance of these teams. The data set we will be using is a two-year period of data from our team and will be compared against a two-year period of an assigned team. Statistical methods such as measures of central tendency and variability and confidence intervals will be used to interpret this data. By using data-driven analytics, we will be able to observe trends in performance and compare that with the performance of other teams in the franchise. This will hopefully lead to better decision-making by management in terms of offensive and defensive strategies and impact our future trading philosophies for better on-court performance.

## Introduction: Your Team and the Assigned Team

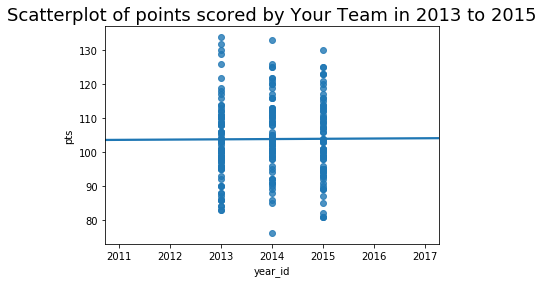
My team is the San Antonio Spurs, and this analysis will be observing the team for the years ranging from 2013 – 2015. The assigned team is the Chicago Bulls, and this comparative study will be covering this team for the years ranging from 1996 – 1998.

Table 1. Information on the Teams

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

## Data Visualization: Points Scored by San Antonio Spurs

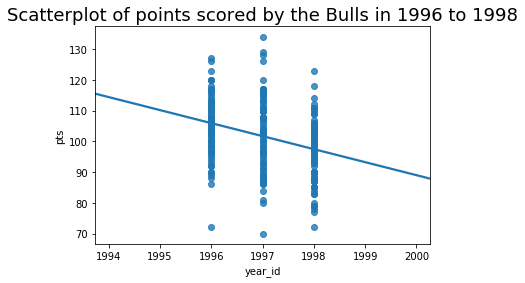
Data visualization is the graphical representation of information and data, often using charts, graphs, and maps. By depicting data graphically, data visualization allows the observer to understand trends, outliers, and patterns in a clear and easy to understand way. The following scatter plot is being provided to show the trend between two variables (points scored and year) for the San Antonio Spurs. The scatter plot was chosen because it best describes the data distribution for the variable being observed here as it provides us with an easy way to observe the trend in points scored over time.



By visually inspecting this plot, we can see that the distribution of this variable is around the same for the years 2013 and 2015 with only a slight decrease in max points scored for the year 2015. However, we can also see that 2014 was our most erratic year with outliers being reported at 77 points and 134 points. Based on this data, we can also see that 2013 was our most successful year in terms of points scored.

## Data Visualization: Points Scored by the Chicago Bulls

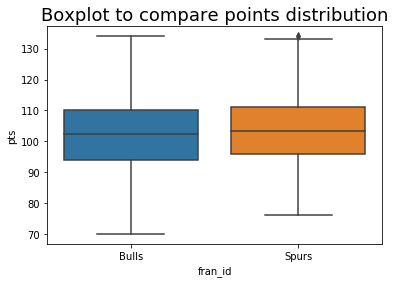
The following scatter plot is a graphical representation of the points scored by the Chicago Bulls in 1996 – 1998. In this example, we can see that 1996 and 1997 were both years in which extreme outliers were recorded at around 70 points and 134 points respectively. We can also say that 1996 was the most consistent year in terms of scoring despite the one outlier, as their data points were grouped much closer together.

**

This method of data visualization was chosen because it shows an interesting trend in the linear regression curve. A linear regression curve, or line of best fit, is a straight line that represents the relationship between two variables in a dataset (zyBooks, 2019). By visually inspecting this plot, we can see that the linear regression curve is sloped downwards. This indicates that as each year progressed, the trend in points went down. Based on this data we can safely assume that the trend will continue downwards for the subsequent year.

## Data Visualization: Comparing the Two Teams

In general, data visualization can play an important role in comparing two different data distributions. By allowing the observer to visually explore and understand its characteristics, data visualization helps to identify patterns and measure differences amongst key variables. A good way to visually approach a data distribution is by using a box plot. Box plots summarize the distribution by showing the median, quartiles, and potential outliers. This is useful for identifying skewness and spread as well as comparing interquartile ranges (Courthoud, 2022). The following box plot is a comparison of the points distribution for the Chicago Bulls and the San Antonio Spurs.



The box plot was chosen because it easily identifies the median score for each team as well as allowing us to quickly identify the range of scores for each team. Furthermore, we can identify outliers and interquartile ranges as well. The data distributions can be visually observed as having a relatively close median between both teams at around 104 points. We can also visually observe that while both teams had a relatively equal max score of around 134 points, the Spurs had a few outliers above that. Additionally, we can see that the Chicago Bulls had a larger Q1 range than the San Antonio Spurs did indicating that they had a lower minimum score. Finally, we observe that both boxes span nearly the same area, meaning that 50% of the data provided by these teams is in the same location. Box plots are also useful because they provide a five-number summary: minimum, median, maximum, Q1, and Q3.

Table 2. Chicago Bulls Five Number Summary

| **Category** | **Value** |
| --- | --- |
| Minimum | 70 |
| Q1 | 95 |
| Median | 103 |
| Q3 | 110 |
| Max | 134 |

Table 3. San Antonio Spurs Five Number Summary

| **Category** | **Value** |
| --- | --- |
| Minimum | 77 |
| Q1 | 97 |
| Median | 104 |
| Q3 | 111 |
| Max | 134 |

## Descriptive Statistics: Points Scored By San Antonio Spurs in Home Games

In general, measures of central tendency help identify three characteristics of a data set: mean, mode, and median. These characteristics are useful as they tell us the most frequent (mode) value, the middle value (median) of the data when sorted in order, and the average (mean) of the values (Bhandari, 2020). This helps to provide insights into where data points cluster within a distribution. To provide context beyond central tendency, measures of variability can be observed such as variance, range, and standard deviation. The range can be used to calculate the difference between the maximum and minimum values in a data set, giving a quick assessment of the overall spread. The variance measures the average squared difference of each value from the mean and a higher variance indicates that data points are more spread out from the mean (Whatley, 2022). Likewise, a lower variance indicates the opposite. Variance is helpful in identifying whether data is dispersed far from the average. Finally, the standard deviation is the square root of the variance and is useful because it makes interpreting the variance easier since it is not measured in a squared unit.

Table 4. Descriptive Statistics for Points Scored by San Antonio Spurs in Home Games

| **Statistic** | **Value** |
| --- | --- |
| Mean | 106.12 |
| Median | 105.0 |
| Variance | 107.16 |
| Standard Deviation | 10.35 |

Based on these figures, we can see that the San Antonio Spurs scored an average of 106.12 points in home games and the median or middle value of those games was 105 points. By comparing the mean and median we can see that the distribution of these points is skewed to the right since the mean is greater than the median. Since this is a positively skewed distribution of data, the median is more resistant to outliers and can better represent the center of the data here. The data also showed the variance of 107.16 and standard deviation of 10.35. This indicates that data points were somewhat (around 10 points) from the mean for home games.

## Descriptive Statistics: Points Scored By San Antonio Spurs in Away Games

Descriptive statistics were also recorded for points scored by the San Antonio Spurs in away games. Measures such as mean, median, variance, and standard deviation were calculated to obtain a clearer picture as to whether the San Antonio Spurs were performing better at home or on the road.

Table 5. Descriptive Statistics for Points Scored by San Antonio Spurs in Away Games

| **Statistic** | **Value** |
| --- | --- |
| Mean | 101.59 |
| Median | 101.0 |
| Variance | 131.31 |
| Standard Deviation | 11.46 |

We can see from the table that the San Antonio Spurs had a lower mean score in away games (101.59) than they did at home (106.12). The median was reported lower as well at 101 points. Unsurprisingly, we saw a higher variance (131.31) and standard deviation (11.46) indicating that points scored were more likely to differ from the mean on the road than when the team played at home. These figures still show a right-shaped, or positively skewed, distribution of data and as such, the median is the best measure of central tendency which can represent the center of this distribution.

On average, the San Antonio Spurs scored five points more per game at home than when the team was traveling. With a higher variance, we can see that the San Antonio Spurs were not as likely to remain consistent with their points scored as well. This would lead to a higher standard deviation of points per game indicating that their points scored would be more unpredictable. Based on this information, we can say that the San Antonio Spurs are performing better at home than on the road.

## Confidence Intervals for the Average Relative Skill of All Teams in San Antonio Spurs’ Years

Confidence intervals were also used to estimate the average relative skill (ELO) of all teams in the San Antonio Spurs’ years. Confidence intervals can be used to estimate measures of central tendency by accounting for the uncertainty associated with sample variability. For instance, in constructing a 95% confidence interval for a mean, by repeatedly taking samples, approximately 95% out of 100 intervals would contain the true population mean.

Table 6. Confidence Interval for Average Relative Skill of Teams in San Antonio Spurs’ Years

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

By using a 95% confidence interval here, we have calculated that 95% of teams will have an ELO between 1502.02 and 1507.18 for these years. This means that out of 100 intervals, 95 of them produced an ELO for these teams between these ranges. If this interval had used a different confidence level, the interval would change. A higher confidence level would be used to capture the true population parameters. A wider interval can provide more confidence, but the interval would not be as precise. A lower level would allow us to see a more precise number but with less certainty that our figure is accurate.

The probability that a given team in the league has an ELO less than that of the San Antonio Spurs is 95%. Because the San Antonio Spurs won the NBA Finals in 2014 and had the highest Win/Loss record in 2015, it is not unusual for a team to have a skill level less than the Spurs and the probability accurately depicts that.

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

Confidence intervals were also used to study the average relative skill of all teams in the Chicago Bulls’ years. This was done to identify and compare the ELO of teams in the franchise against the Bulls and contrast the difference between them and the Spurs.

Table 7. Confidence Interval for Average Relative Skill of Teams in Chicago Bulls’ Years

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

The 95% confidence level provided us with a confidence interval ranging between 1487.66 and 1493.65. This shows that out of 100 intervals, 95 of them provided teams within this range. As stated previously, using a different confidence level would have been able to either provide us with a more accurate representation or a more precise one depending on whether we chose to use a higher or lower level. The probability that a team had a lower ELO than that of the Bulls in these years was 97%.

Comparing this data to the previous table shows me that while the Bulls were more likely to have a higher ELO than other teams and as such be a better team, this was true because the ELO of other teams in the franchise was significantly lower than for the Spurs’ years. This signifies that teams are getting more skilled, and the NBA is playing at a higher caliber than they were from 1996 – 1998.

## Conclusion

Based on these findings, we can say that players in the NBA are performing at a higher standard than they were before. We can also analyze that while the Spurs may not have been scoring as high of an average relative skill of the Bulls, they were performing on par with them, if not better in some areas. These results showed me that two teams at the top of their league can differ in skill level and provide meaningful statistics such as measures of central tendency and variability. By using these statistics, a practical analysis of their scores was observed and a confidence interval was calculated to understand the probability of a team performing at a lesser skill level than either team.

Descriptive statistics including measures of central tendency (mean, median, and mode) and variability (variance and standard deviation) were used to provide insights on scoring patterns and variability of performance. Confidence intervals were used to estimate ranges within the average relative skill of teams for specified time periods. Finally, comparative analysis was conducted between the performance for home and away games.

This analysis was useful to improve team performance by informing management of strategic decisions that could be made in the future. Based on these scoring patterns, variability, and other methods used to conduct this review, coaches can make better choices for future team development. The practical implications of this study are inherent in the data it provides to management. The NBA is an evolving league with franchise teams always trading and seeking the best players or free agents. By understanding this type of information, management can adapt and transform strategies for a competitive edge.

## Citations

Bhandari, P. (2020, July 30). *Central Tendency | Understanding the mean, median & mode*. Scribbr. https://www.scribbr.com/statistics/central-tendency/

Courthoud, M. (2022, June 22). *How to compare two or more distributions.* Matteo Courthoud. <https://matteocourthoud.github.io/post/distr/>

Whatley, M. (2022). Measures of Variability. In: Introduction to Quantitative Analysis for International Educators. Springer Texts in Education. Springer, Cham. https://doi.org/10.1007/978-3-030-93831-4\_3

zyBooks. (2019). MAT 243: Applied Statistics I for Science, Technology, Engineering, and Math.