# MAT 243 Project One Summary Report

Nicholas Kreuziger

nicholas.kreuziger@snhu.edu

Southern New Hampshire University

## Introduction: Problem Statement

The problem at hand is determining a comparison of the 2013-2015 San Antonio Spurs against the 1996-1998 Chicago Bulls to assist management with improving their team’s performance. This analysis will begin by cross analyzing the performance of various NBA teams using historical Elo rating systems. The Elo rating system is a statistical predictor for the outcomes of competitive matches originally created by Hungarian American physics professor Arpad Elo. This rating system was implemented by the United States Chess Federation in 1960. The difference in the ratings between two players serves as a predictor of the outcome of a match. Two players with equal ratings who play against each other are expected to score an equal number of wins. A player whose rating is 100 points greater than their opponent's is expected to score 64%; if the difference is 200 points, then the expected score for the stronger player is 76% (Wikipedia, 2019). We will use a dataset of historical Elo ratings of the NBA League. With this dataset we will be calculating points scored, points distribution and Relative Skill between the Bulls (1996-1998) and Spurs (2013 – 2015). Additionally, we will use the Average Relative Skill of All Teams in the time frames of 2013-2015 and 1996-1998 to calculate, within a confidence interval, the likelihood of our team’s scoring better than the two teams being examined.

## Introduction: Your Team and the Assigned Team

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 Your Team

Data visualization is used to make sense of data. It is often that raw data does not provide the full story. A solution to this is to use visualizations to highlight data distribution and trends. The maxim “Seeing is believing” exists for a reason; if yourself or the audience cannot visualize the data and message, they will not comprehend it.   
 A Histogram and Scatterplot were the choices for the visualization. The scatterplot was one-dimensional, the x-axis showed the years, and the y-axis showed the points scored per game. An inherent flaw to this approach is that duplicate coordinates will overlap and not visualize the entire story. The histogram was two-dimensional, the y-axis showed probability of the value occurring and the x-axis plotted the sum of points scored in a game. I chose the histogram for two reasons. Firstly, the histogram represents all the data in a visually pleasing manner. Secondly, the histogram is adding additional dimension by representing the probability of a value occurring.

Chart, histogram

Description automatically generated

The above Histogram has a unimodal distribution, there is one prevalent peak. Visually you can tell the data is normally distributed, which means 68.3% of all data is 1 standard deviation from the average, 95.5% of the data is within 2 standard deviations, and 99.7% of the data is within 3 standard deviations.

## Data Visualization: Points Scored by the Assigned Team

Data visualization is used to make sense of data. It is often that raw data does not provide the full story. A solution to this is to use visualizations to highlight data distribution and trends. The maxim “Seeing is believing” exists for a reason; if yourself or the audience cannot visualize the data and message, they will not comprehend it.   
 A Histogram and Scatterplot were the choices for the visualization. The scatterplot was one-dimensional, the x-axis showed the years, and the y-axis showed the points scored per game. An inherent flaw to this approach is that duplicate coordinates will overlap and not visualize the entire story. The histogram was two-dimensional, the y-axis showed probability of the value occurring and the x-axis plotted the sum of points scored in a game. I chose the histogram for two reasons. Firstly, the histogram represents all the data in a visually pleasing manner. Secondly, the histogram is adding additional dimension by representing the probability of a value occurring.

Chart, histogram

Description automatically generated

The above histogram has a unimodal distribution, there is one prevalent peak. Visually you can tell the data is normally distributed, which means 68.3% of all data is 1 standard deviation from the average, 95.5% of the data is within 2 standard deviations, and 99.7% of the data is within 3 standard deviations.

## Data Visualization: Comparing the Two Teams

Data visualization is used to make sense of data. It is often that raw data does not provide the full story. A solution to this is to use visualizations to highlight data distribution and trends. The maxim “Seeing is believing” exists for a reason; if yourself or the audience cannot visualize the data and message, they will not comprehend it.

The plots available in this exercise were a Histogram or Boxplot. This exercise specifically is asking for a visualization that best encapsulates the points distribution between the 2013-2015 Spurs and the 1996-1998 Bulls. A histogram overlay would show the probability of both teams scoring a specified number of points per game. A Box and Whisker plot shows the range of points distributions for each team in the respective sampling years in four quartiles while also plotting outliers from the quartiles. The Box and Whisker Plot is a superior choice for answering the question “What is the difference in points distribution between the Spurs (2013-2015) and the Bulls (1996-1998)?”.

Chart, box and whisker chart

Description automatically generated

From this Box and Whisker Plot a few things can be determined about the Spurs (2013-2015) and the Bulls (1996-1998). The “Box” of the data represents the lower boundary of Q1 and the upper boundary of Q3, which is the middle 50% of the data. The Bulls have a wider spread of the middle 50% of their data, whereas the Spurs have a smaller spread with higher values; This indicates the Spurs 50% of the time are slightly more likely to score better. The Spurs also have a tighter spread of minimum and maximum, around 135 for their maximum and around 78 minimum. The Bulls have a higher maximum of 135 and a lower minimum of 71. Outliers are also intriguing, showing that the Spurs were the only team of the two with any values outside of 1.5 times the length of the box from either end of the box. The composition of these outliers also indicates “When the Spurs played a game that resulted in an outlier they always scored above the maximum within the dataset”.

## Descriptive Statistics: Relative Skill of Your Team

Measures of central tendency and variability are used to peer into the overall composition of a dataset. Using these cursory measures, you can instantly know some important features of your dataset. With the Mean you know what on average a value of the dataset will be. With the Median you will know the center-most value of the dataset, which can help to determine if the Mean is being skewed by variable data inputs. The Variance will tell you how much the data points are spread around the Mean and how dependably the Mean will be output. The Standard Deviation will tell you on average how far away from the Mean a data point will be. In short, measures of central tendency and variability give you a better idea of what the dataset looks like without reviewing every individual record.

Table 2. Descriptive Statistics for Relative Skill of Spurs (2013-2015)

| **Statistic Name** | **Value** |
| --- | --- |
| Mean | 1685.42 |
| Median | 1689.17 |
| Variance | 1038.61 |
| Standard Deviation | 32.23 |

The Mean of Relative Skill is 1685.42, which indicates on average in the dataset the Spurs will have a Relative Skill of 1685.42. The Median of Relative Skill is 1689.17, which indicates that the middle-most value of the entire dataset is 1689.17. This tells us that the Mean is not being influenced by any extreme outliers. The Variance is 1038.61, this tells us that the Relative Skill of the Spurs has varied between 1038.61. Variances are unintuitive and are a better metric as compared to a second sample. The Standard Deviation is 32.23, this tells us that the Relative Skill has on average varied 32.23 from the mean of 1685.42.

The Mean and the Median tell us that the data is bell-shaped. The Mean and Median are only varied by around 4 units, which indicates to us that there is higher likelihood of any given Relative Skill is centered around the Mean. With this dataset the Mean is best to use to represent the center of the distribution, this is because there are no extremely large values or outliers present. If there were larger outliers or variation the Median would have to be considered.

## Descriptive Statistics: Relative Skill of the Assigned Team

Measures of central tendency and variability are used to peer into the overall composition of a dataset. Using these cursory measures, you can instantly know some important features of your dataset. With the Mean you know what on average a value of the dataset will be. With the Median you will know the center-most value of the dataset, which can help to determine if the Mean is being skewed by variable data inputs. The Variance will tell you how much the data points are spread around the Mean and how dependably the Mean will be output. The Standard Deviation will tell you on average how far away from the Mean a data point will be. In short, measures of central tendency and variability give you a better idea of what the dataset looks like without reviewing every individual record.

Table 3. Descriptive Statistics for Relative Skill of Bulls (1996-1998)

| **Statistic Name** | **Value** |
| --- | --- |
| Mean | 1739.8 |
| Median | 1751.23 |
| Variance | 2651.55 |
| Standard Deviation | 51.49 |

The Mean of Relative Skill is 1739.8, which indicates on average in the dataset the Bulls will have a Relative Skill of 1739.8. The Median of Relative Skill is 1751.23, which indicates that the middle-most value of the entire dataset is 1751.23. This tells us that the Mean is not being influenced by any extreme outliers. The Variance is 2651.55, this tells us that the Relative Skill of the Bulls has varied between 2651.55. This is significantly higher than the variance of the Spurs. This tells us that the Bulls will not as dependably have the same Relative Skill between any given game. This is likely due to the Elo system more heavily penalizing losing a Winning Streak than maintaining a winning streak.

The Standard Deviation is 51.49, this tells us that the Relative Skill has on average varied 51.49 from the mean of 1739.8.

The Mean and the Median tell us that the data is bell-shaped. The Mean and Median are varied by around 11 units. With this dataset the Mean is best to use to represent the center of the distribution, this is because there are no extremely large values or outliers present. If there were larger outliers or variation the Median would have to be considered.

## 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 2013-2015

| **Confidence Level (%)** | **Confidence Interval** | **Range** |
| --- | --- | --- |
| 95% | (1502.02, 1507.18) | 5.16 |
| 80% | (1502.92, 1506.29) | 3.37 |
| 60% | (1503.5, 1505.71) | 2.21 |

Confidence intervals are used to estimate measures of central tendency for a population. The confidence interval for the average relative skill of an NBA team for the years 2013-2015 is 1502.02 to 1507.18 with a 95% confidence interval. This means that if this test was redone, there’s a 95% confidence the value would fall between 1502.02 and 1507.18. Changing the confidence level will affect the overall error bound. If a larger confidence level is used the interval is wider. If a smaller confidence level is used, the error bound is smaller. In example, using the above table you can see the Range Decrease as the Confidence Interval Increases.

The probability that a team in the league will have a relative Skill Level less than the Spurs between 2013-2015 is 94.52%. This makes sense considering the Spurs were in the NBA Championship finals in 2013 and won the finals in 2014 (Wikipedia Contributors, 2019).

## 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 1996-1998

| **Confidence Level (%)** | **Confidence Interval** | **Range** |
| --- | --- | --- |
| 95% | (1487.66, 1493.65) | 5.99 |
| 80% | (1488.69, 1492.61) | 3.92 |
| 60% | (1489.37, 1491.94) | 2.57 |

Confidence intervals are used to estimate measures of central tendency for a population. The confidence interval for the average relative skill of an NBA team for the years 1996-1998 is 1487.66 to 1493.65 with a 95% confidence interval. This means that if this test was redone, there’s a 95% confidence the value would fall between 1487.66 and 1493.65. Changing the confidence level will affect the overall error bound. If a larger confidence level is used the interval is wider. If a smaller confidence level is used, the error bound is smaller. In example, using the above table you can see the Range Decrease as the Confidence Interval Increases.

The probability that a team in the league will have a relative Skill Level less than the Bulls between 1996-1998 is 97.32%. This makes sense considering the Bulls in the years 1996-1998 are considered one of the best NBA teams of all time. They won the Championship every year in that date range (Wikipedia Contributors, 2019).

## Conclusion

As the table below demonstrates, the Relative Skills broken down for the Teams in their respective years are statistically similar.

Table 6. Statistics of Assigned Team and Your Team

|  |  |  |
| --- | --- | --- |
| Statistic Name | Bulls 1996-1998 | Spurs 2013-2015 |
| Mean | 1739.8 | 1685.42 |
| Median | 1751.23 | 1689.17 |
| Variance | 2651.55 | 1038.61 |
| Standard Deviation | 51.49 | 32.23 |
| Probability of Lower Rel. Skill | 2.68% | 5.48% |

The Bulls scored higher in their Mean and Median. Their Relative Skill was dependably higher than the Spurs. That takeaway indicates that The Bulls were a better team than The Spurs. The Spurs did have a lower Variance and Standard Deviation, which indicates they were more consistent than the Bulls when it came to their performance. The Bulls were 2.8% more likely to have a higher Relative Skill than The Spurs.

## Citations

Wikipedia Contributors. (2019, April 2). *Elo rating system*. Wikipedia; Wikimedia Foundation. <https://en.wikipedia.org/wiki/Elo_rating_system>

Wikipedia Contributors. (2019, October 30). *List of NBA champions*. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/List\_of\_NBA\_champions

‌

‌