# MAT 243 Project One Summary Report

[Luna Saccoia]

[Luna.Saccoia@snhu.edu]

Southern New Hampshire University

Note: Replace the bracketed text on page one (the cover page) with your personal information.

## Introduction: Problem Statement

We are looking at performance data from basketball teams during certain years. We will be looking at a specific team’s performance during the assigned years. We will be using descriptive statistics and visualizing it so it is easy to see where the team’s performance was the best during the given time.

## Introduction: Your Team and the Assigned Team

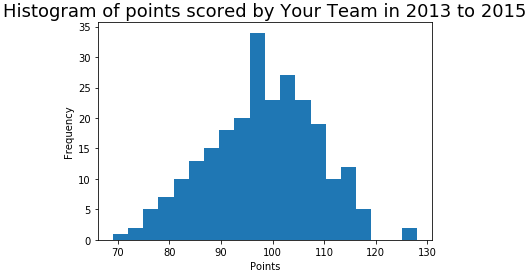
Table 1. Information on the Teams

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

I chose the Pistons as my team, and 2013-2015 were the years assigned. The bulls were the team assigned and the years assigned were 1996-1998 for comparative study.

## Data Visualization: Points Scored by Your Team

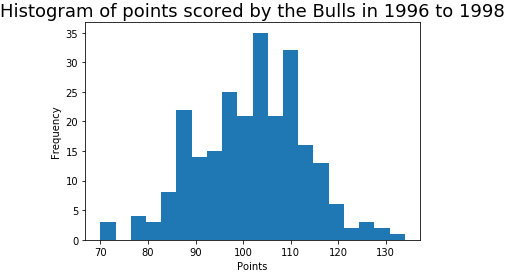
In general, data visualization is the representing of data and information using graphs and charts to do so. This makes it easy to see and understand the information and data that is being presented. Trends, patterns and outliers are easy to see and understand.



The Histogram is much easier to see where the highest frequency was for points scored during this time. This allows us to see that the team’s points were mostly in the 95-110 point range, with 98 being the score that had the highest frequency.

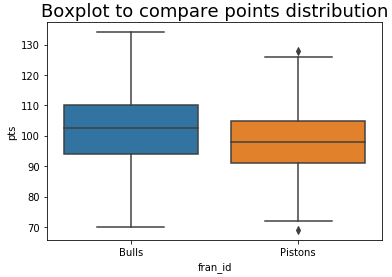
## Data Visualization: Points Scored by the Assigned Team

I chose the histogram as the plot I think best visualizes this data set. It gives us the frequency that they scored various points over this period. From this histogram, there is more outliers than the one for my team, but it shows that the number of points scored with the highest frequency is 104. I feel like the histogram is easier to visualize, where the scatter plot is unable to show each of the points clearly.



## Data Visualization: Comparing the Two Teams

In general, data visualization is the representing of data and information using graphs and charts to do so. This makes it easy to see and understand the information and data that is being presented. Trends, patterns and outliers are easy to see and understand.



The box plot shows the comparison of these two data distributions more clearly than the overlaid histogram. With the box plot we can clearly see where the average scores were for each team and that the Bulls performed better than the Pistons.

## Descriptive Statistics: Relative Skill of Your Team

Table 2. Descriptive Statistics for Relative Skill of Your Team

| **Statistic Name** | **Value** |
| --- | --- |
| Statistic  *(for example, Mean)* | X.XX  *\*Round off to 2 decimal places.* |
| Mean | 1406.66 |
| Median | 1413.78 |
| Variance | 1754.65 |
| Standard Deviation | 41.89 |

In general, the measures of central tendency are the mean, median and mode of a data set. The standard deviation and variability are how far the data set deviates from the mean.

Honestly I am confused by the numbers it is giving. None of them are the mean of points scored, the numbers are much too high. So I am confused on what they actually show, since it is clearly not variability of points scored. From reading the step, it seems to be showing the skill of the teams but does not show what the data means, such as a mean score of 1406.66 is confusing without knowing what the upper and lower bounds are. If it does show that, I am not seeing it.

It’s skew is bell shaped. The median and mean being approximately the same gives it a bell shaped skew and normal distribution.

## Descriptive Statistics: Relative Skill of the Assigned Team

Table 3. Descriptive Statistics for Relative Skill of the Assigned Team

| **Statistic Name** | **Value** |
| --- | --- |
| Statistic  *(for example, Mean)* | X.XX  *\*Round off to 2 decimal places.* |
| Mean | 1739.80 |
| Median | 1751.23 |
| Variance | 2651.55 |
| Standard Deviation | 51.49 |

The mean is the average of the data, median is the middle number of the data set when arranged in ascending or descending order, the variance and standard deviation show how far some of the points are from the mean. This is to show outliers in the data.

The skew is bell shaped. This is because the mean and median are very close to each other, enough that it does not skew right or left.

The variance and standard deviation are higher in the assigned team than in my team, the fact that there is less variance and a lower standard deviation from the mean in my team, shows that they are the team with the most consistent skill.

## 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 used to estimate the measures of central tendency of a dataset by using the standard deviation to be, in this case, 95% sure that the range of values will contain the true mean of the dataset. The true mean of the relative skill for the team I chose, during the chosen years, is 1406.66. This actually falls just outside the 95% confidence intervals calculation. The probability that a given team has a relative skill level less than that of the team I picked is 0.1931 (Choice 2 in the notebook step 8 question). It is not unusual that a team would have a skill less than mine, same is true for the fact that there are teams that would have greater skill than mine. My chosen team is not the top team, not the bottom team in the league. So there will be teams with greater skill and some with lower skill. The probability of a team having a greater skill than mine is 0.8069 (choice 1 in the notebook step 8 question.)

## 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) |

My interpretation of the confidence interval in terms of the average relative skill of teams in the assigned team’s range of years is that the calculations show the Bull’s had a higher relative skill during the years of 1996-1998 than the Piston’s did during the years of 2013-2015. The probability of a team having LESS relative skill than the Bull’s during this time was 0.9732. Calculated to a percentage this would mean that the probability of a team having a higher relative skill than the Bull’s would be approximately 3%. This means that 97% of teams had LESS skill than the Bull’s, during this time. Comparatively, the probability of a team having LESS skill than the Piston’s during their given timeframe, is 0.1931 or approximately 19%. This would show that the Bull’s were very close to the top of their league at 97% during the years of 1996-1998 and the Piston’s were much lower at around 20% during the years of 2013-2015.

## Conclusion

If I understand it correctly, the importance of the analysis we performed was to gauge the skill level of each team during a given block of time. Taking their performance scores, finding the central tendency, and using that to find the confidence interval allowed us to find the probability that they performed better or worse than other teams in the league during the same timeframe. The results showed us that during their given years (1996-1998) the assigned team (Bull’s) were performing very well, and were at the top of their league, whereas the team I chose (Piston’s) did not perform nearly as well during their given years (2013-2015)

## Citations

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