

MAT 243 Project One Summary Report

Walker Martin
walker.martin@snhu.edu
Southern New Hampshire University

1. Introduction: Problem Statement

This statistical project is analyzing the relative skill level of two NBA teams from different years. One team is the 1996-1998 Bulls and the other is the 2013-2015 Wizards. The data being used is from the teams performances over the three years each having 246 relative skill data points. The project begins with gathering the data, then we get into descriptive statistics, then we finish up with the confidence intervals and comparing teams relative skills in the years they were active.

2. Introduction: Your Team and the Assigned Team

For this project the team I picked is the Wizards, and the years were preselected to be 2013-2015. The preassigned team is the Bulls during 1996 -1998, assumingly because their skill was so high and they were so dominant that there should be a clear disparity between then and most teams during 2013 - 2015. Out of the five rows that were printed in steps one and two the bulls have five wins while the Wizards have five losses.

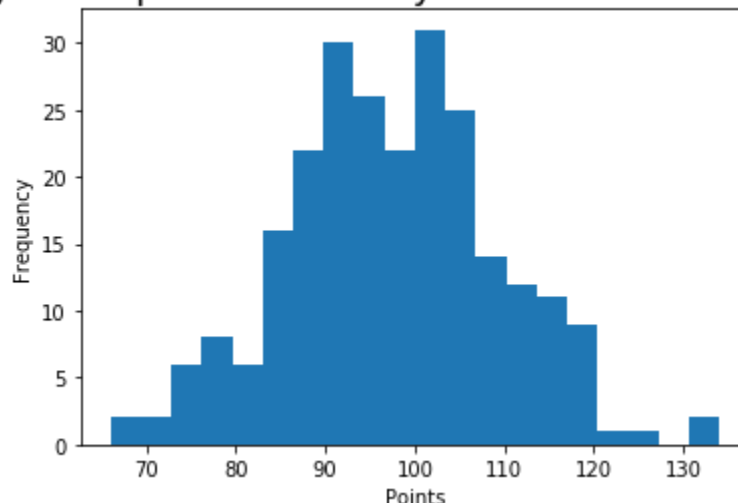
Table 1. Information on the Teams

	Name of Team	Assigned Years
1. Yours	Wizards	2013 - 2015
2. Assigned	Bulls	1996 - 1998

3. Data Visualization: Points Scored by Your Team

Data visualization is used for large and small scale data to organize it in a manner that people understand. It's much more comprehensive to view a chart with 250 data plots than it is to view a csv file with 250 rows. I selected the histograms because it's better to represent the data distribution than the scatter plot. That is due to the x axis being points and y being frequency as opposed to the scatter that uses years as the X. Scatter plot would be better to find the trend of points scored across the years but not the distribution across the actual points scored. Based on the histogram it's clear that the data resembles a bell curve therefore is near normal distribution. The density in the middle buckets leads me to believe that the median and mean will lie around 100 points.

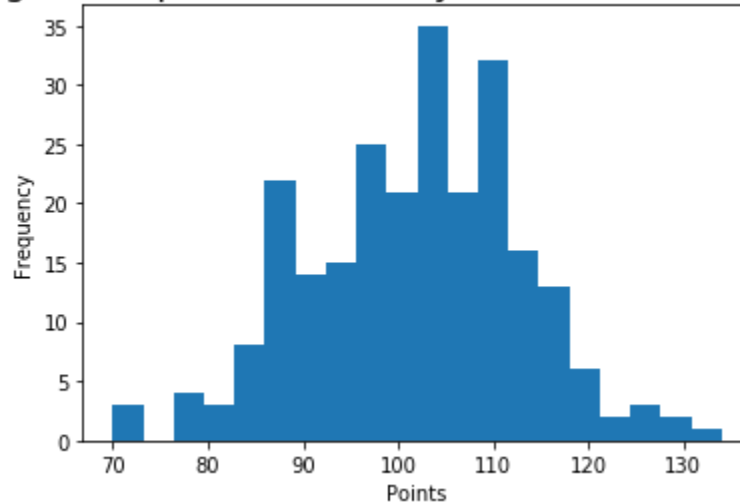
Histogram of points scored by Your Team in 2013 to 2015



4. Data Visualization: Points Scored by the Assigned Team

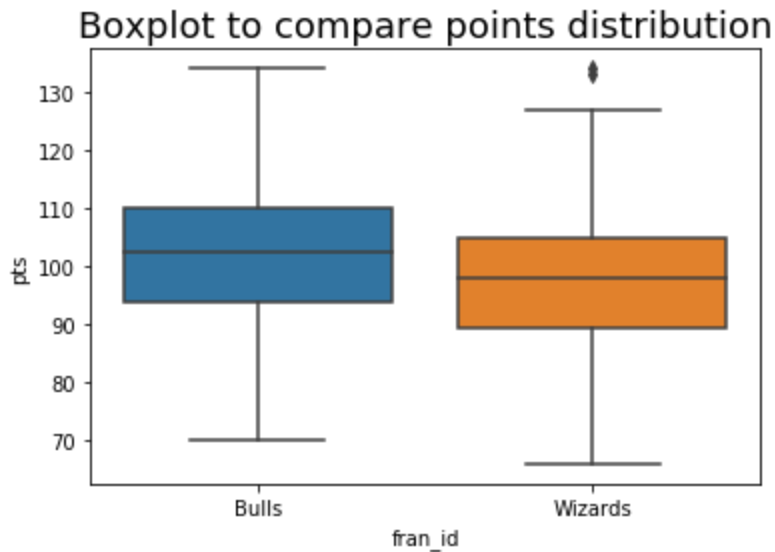
Similarly to the visualization of my chosen teams data I choose to represent this with the histogram rather than scatter plot. Which makes sense since this and the previous data visualization both depict frequency of points scored. The scatter plot would show the correlation or trends from points scored across the three years. Also like the chosen teams visualization this one also presents a normal distribution. The assigned team has a higher max frequency though the distributions appear similar.

Histogram of points scored by the Bulls in 1996 to 1998



5. Data Visualization: Comparing the Two Teams

The data visualizations here are intended to help compare the two teams' relative skill by analyzing their points scored throughout the three respective seasons. I chose the boxplot as my preferred data visualization because it shows a clear median, interquartile length and outliers. Also the previous visualizations are histograms so I could just compare the two together without the overlay. However because I chose the boxplot I know that my chosen team has a median of about 100 while the bulls have a median of 102. My first impression that the distributions are similar appears to be correct since their inter quartile lengths are both approximately 15 points. Though the bulls have a more balanced center distribution, the wizards interquartile length is slightly heavy towards the lower quartile.



6. Descriptive Statistics: Points Scored By Your Team in Home Games

Home Games	Value
Mean	100.18
Median	101.0
Variance	116.07
Standard Deviation	10.77

Central tendency is how we get a feel for the datasets ranges and the most common occurrences. Variability is the measurement of how spread or distributed the data is from the central tendency. Which is why they go hand and hand. The central tendencies mean of 100.18 and median of 101.0 depict the average number of points scores for the wizards from 2013-2015. The variance of 116.07 and standard deviation of 10.77 tells how far and often the team drift away from their average. The home game statistics of mean and median appear to line up with the box plot which holds all games, which means the Home games data describes a normal distribution of bell shaped curve. Therefore we can use the mean as the measurement of central tendency. Since the median is used when the data is skewed or there are more outliers present.

7. Descriptive Statistics: Points Scored By Your Team in Away Games

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

Away Games	Value
Mean	96.76
Median	95.0
Variance	147.98
Standard Deviation	12.16

This dataset represents the home field advantage phenomenon in sports. The average points scored per game shifted from 100.18 at home to 96.76 away. The variance also increased pretty significantly, indicating inconsistent points scored per game. The mean and median in away games have a larger difference in away games than home. For home games the mean is less than the median while games away the mean is larger. This represents a slightly skewed data in the away games, the median is 95 therefore the data is heavier in the upper quartile since the mean of 96.76 lies there. This provides a reason to choose median over mean for the central tendency to represent the dataset, but I think the difference is small enough to validate using the mean.

8. 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)

The confidence interval is a method of estimating an accurate population mean. The confidence interval is the range in which the population mean likely lies therefore it's directly related to central tendency. The confidence interval for years 2013-2015 is very stable with a 5.16 point difference between high and low interval bounds. This means that on average the teams during these years played at a similar skill level to one another. The confidence level impacts the intervals, as the level increases the confidence intervals increase in width too. Therefore a 99% confidence level leads to a larger interval while a 90% would decrease the width. The probability that a team has a relative skill less than my team is 42%, I used the cdf function rather than the sf since cdf refers to less or equal to while cdf is greater than.

9. 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%	(1456.06, 1525.24)

The confidence intervals for the assigned teams years are (1456.06, 1525.24). The confidence level is the same as my chosen teams, however, this interval width is significantly larger. Therefore the relative skill of teams during 1996-1998 must vary more so than teams during 2013-2015. The assigned teams confidence intervals upper bound is about 18 greater than the previous though the lower bound is around 46 points lower. Therefore some teams during the 1996-1998 had a higher relative skill than teams during 2013-2015 however the majority were significantly lower in skill. The CDF probability for teams in 1996-1998 being worse than the bulls is 57%, which seems relatively low for comparing teams to the Bulls. Especially since the first step printed five rows where their relative skill is greater than 1600, but 1600 is well over the upper confidence interval for 1996-1998. Therefore the five games where Bulls has a relative skill of around 1600 must have been slight outliers.

10. Conclusion

In conclusion there are a few ways to wrap up the analysis. From comparing the wizards home vs away games we can tell that the team plays better at home and is more consistent. Based on the confidence intervals for all teams for both year sections we know that the teams during the 2013-2015 years played much more consistently and overall better than teams in 1996-1998. However, the max ceiling for relative skill of a team was higher in 1996-1998. The Bulls were dominant during the given years, therefore, the probability that a team is worse than them seems low at 57%. There may be an error in the calculations here since the Bulls are the team that should have the highest relative skill out of the 30 teams leading to a probability closer to 95% that one of the other teams have a lower relative skill.