# MAT 243 Project One Summary Report

William Paddock

william.paddock@shnu.edu

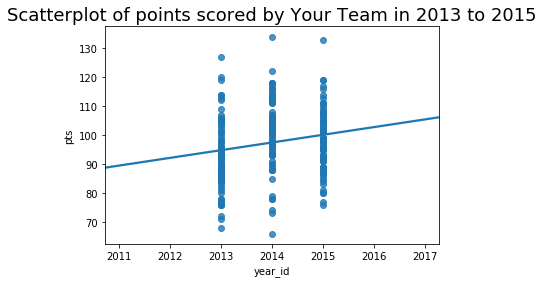
Southern New Hampshire University

The issue is to derive meaningful information from a large set of historical data on two different basketball teams finding patterns in the information. Data driven review of key variables will then help the team's management make decisions going forward.

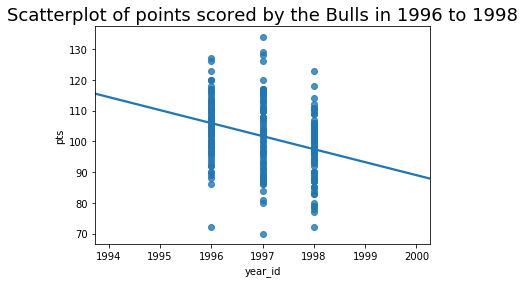
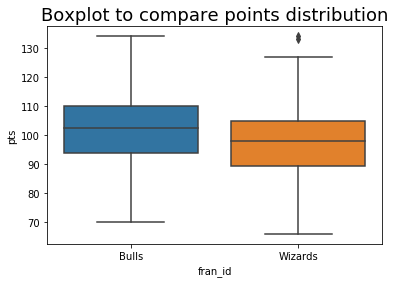
The two teams being compared will be the Wizards and the Chicago Bulls. These teams will be analyzed via the years 2013-2015. The report will compare the teams scores per game as well as the Average Relative Skills. The above points will be compared via the mean, and median, then used to extract the confidence.

|  | **Name of Team** | **Assigned Years** |
| --- | --- | --- |
| 1. Yours | Wizards | 2013-2015 |
| 2. Assigned | Bulls | 2013-2015 |

Data visualization is used to study the distribution of trends via visuals such as graphing. Graphing shows various differences in data via visual acuity. Which means that graphing the data in question reveals abnormalities in data that tables may miss.



This Scatterplot shows the points scored each year for the Wizard’s team. This makes it easy to see when in each year the score was drastically higher or lower than normal. The viewer can then analyze the distribution of scores through the years of the scatterplot. Thus showing a significant increase in points earned each year. The highest in 2014 with over 130 points. Furthermore, Scatterplot’s also show the consistency of the Wizard‘s points revealing outliers.

The scatterplot for the Bull’s shows the date for the years between 2013-2015. This association shows outliers in the data that can be ignored as well as the expected team performance going forwards. Via visual inspection, there are several outliers for the Bull’s over the years, though the overall points scored does decrease, thus it can be surmised that the following year will also trend further downward.

Generally visualizing data distributions will show the differences in the data, revealing points of interest for comparison. Furthermore, diagrams also show data in a way that is generally understandable to large numbers of people to quickly follow. The Boxplot here shows the average as well as the spread of the data distribution of the two teams point values for each team.

This Boxplot has been picked for easily representing the points scored between the Bulls, and the Wizards, showing the Bulls averaging higher than the wizards in the given timeframe. The Wizard’s also show a few outliers for data comparison. Since the Bulls have a higher score annually; the graphic confirms the Bull’s generally then score more points..

| **-** | **Value** |
| --- | --- |
| Mean | 100.18 |
| Median | 101.00 |
| Variance | 116.07 |
| Standard Deviation | 10.77 |

The central tendency of a statistic is the measure of the central value for a probability distribution. A central tendency is generally measured via a mean, median and the mode of the given data. This is contrasted with the variability of the distribution for which the data in the set shows range from the mean, or distance from the center of a distribution. Per the statistics above, the Mean for home games that the Wizards played is 100.18 points. This is calculated as the middle of the data collected for all home games. A similar statistic is Median, though this shows the exact middle score of the scores given. This then leaves variance and the standard deviation, variance is the calculated difference from the middle of the data at 116.07, the average distance from the mean showing the spread of data or scores in this case. Standard deviation is then calculated from the variance to give the average distances between each score of the individual games played at home. The data itself is generally a bell shape , as the mean is 100.18 which is similar to the median at 101. Thus the suggested way to measure central tendency for this bell shape is with the mean.

| **Statistic Name** | **Value** |
| --- | --- |
| Mean | 94.76 |
| Median | 95.00 |
| Variance | 147.98 |
| Standard Deviation | 12.16 |

The mean scores for the away games are 94.76; this is the central score of all scores taken for away games related to wizards. The median is 95.00 in which this would be the calculated score of the middle of the distribution for scores off the games played away from home. Thus we then have the variance of the away games, in which this is at 147.98; meaning the scores are very spread out from the center or mean of the scores. The standard deviation shows the scores are spread apart by an average of 12.16 points each. This means that the Skew of the away games is still in a bell curve but the curve is spread out, thus the mean is still the best statistic to use with these scores.

Generally the wizards play better at home games than at away games, as the mean and median of the home games are higher; 100.18 vs 94.78 and 101.00 vs 95.00; as the variance of the away games is higher, the scores are also noticeably spread out here.

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

The confidence interval of data is used to explain the expected range of values in the central tendency for this population. Thus the Average Relative Skill for all teams within the years of 2013-2015 should be within 1502.02 and 1507.18. This range will shrink with a confidence interval of 90%. Though it would have stayed roughly the same should it have been higher. The probability of other teams having a skill level less than the Wizard’s is at 17%. Thus it would be unusual for a team to have a lower level of skill.

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

The confidence interval of 95% shows that the expected team Wizards skill level will likely be between 1502.02 and 1507.15 between the years of 2013-2015. While this does represent the considerations of the interval for the 95% confidence interval, using a 90% confidence interval will shrink this interval down to ( 1502.44 , 1506.77 ); while a 99% confidence interval will leave the interval almost the same ( 1501.21 , 1507.99 ). This interval is very similar to the overall Average Relative Skill in the years 2013-2015.

Conclusively it can then be said the statistical differences between teams can be found via the use of numpies statistical library. This same library can also be used extensively to build graphable information which can then be passed over matplot. When done, both show the statistical considerations of the data in easy to view means, such as comparing the confidence interval. Along with the consideration for finding the expected mean of the associated data. These results then show that the Wizards average overall is similar in comparison to the overall teams average for scores for the years of 2013-2015. While the bulls scored higher in this regard to the mean averages. This same consideration is true for the Average relative skills, where the wizards again scored similar to the overall average for the years of 2013-2015.