# MAT 243 Project One Summary Report

Austin Niwa

Austin.niwa@snhu.edu

Southern New Hampshire University

I have been asked to study variables associated with different basketball teams using statistics and data visualization techniques. The data I’m analyzing will be used to improve my team’s performance in comparison to the Chicago Bulls in 1996-1998. For my assignment I will compare averages between my team, the Atlanta Hawks with the Chicago Bulls. I will be using the data sets for the Chicago Bulls in 1996-1998 and the Atlanta Hawks from 2013-2015.

|  | **Name of Team** | **Assigned Years** |
| --- | --- | --- |
| 1. Yours | Team (Atlanta Hawks) | (2013 - 2015) |
| 2. Assigned | Team (Chicago Bulls) | (1996 - 1998) |

A picture containing pixel, silhouette, design

Description automatically generated

Data visualization such as tables and charts are used to quickly interpret and compare data sets. A histogram shows the frequency and points in an easy to see chart but it’s harder to identify when those points were scored as there is no time axis. I picked this plot because it allows me to get a good estimate of how many points were scored by my chosen team, the Atlanta Hawks. Histograms also spread the data out over a larger area compared to the scatterplot which centers all the data above the chosen years and makes it difficult to see all the data. The histogram shows that the data is a bit left-skewed considering there are not many low points scored by my team. Most of the data points start in the low 70’s and go up, peaking in the mid 100-110s range. This signifies that the Hawks scored around 100-110 points most frequently with a low range in the mid 70’s. There was also an outlier in the high 50’s/low 60’s.

* A picture containing pixel, design, silhouette

  Description automatically generated

In this data plot I can see that the Bulls most frequently scored in the high 90’s to low 110’s. The average is likely to be somewhere in the 100-110 area considering that is the height of the bell curve here. I chose this graph for the Bulls as well because it gives a better visualization of the points the Bulls got over the period of 1996-1998. Most of the outliers are on the right side this signifies that the Bulls had more games where they scored a high number of points such as those in the 120‘s and the 130’s than lower in the 70’s and 80’s.

A picture containing rectangle, screenshot, diagram, line

Description automatically generated

Data visualization is used to compare distributions by showing the differences in the averages and variables between two sets of data. By comparing averages using the measures of central tendencies, I can get insight into the similarities and differences in the data distributions between the points scored by the Chicago Bulls and the Atlanta Hawks. In this activity I picked the box plot to compare the points between my assigned team and my chosen team. I picked this plot because despite the double histogram showing both teams’ points overlayed over each other it doesn’t show me the averages between the two teams. With the box plot I can see the general average, compare the point distributions through the upper and lower quartiles, and see how low and high the outliers go. Through the box plot I can see that the Chicago Bulls have higher average points scored than the Atlanta Hawks and higher quartiles, both upper and lower. They also have a wider variation in their outliers than the Hawks do.

| **Statistic name** | **Value** |
| --- | --- |
| Mean | 101.19 |
| Median | 102.00 |
| Variance | 138.47 |
| Standard Deviation | 11.77 |

The measures of central tendency can be used to analyze and summarize values in a given data set. These measures of central tendency all have important roles in summarizing the data between these two basketball teams scores. The mean shows the average in the data set. The median is the middle number in the data set. The variance shows the difference between the data points and the mean. Standard deviation shows the dispersion of the data points from the mean. The points scored by my team in home games is a bell curve since my mean is very close to my median, only being 1.19 off. This makes the best measure to use the mean since it is the average.

| **Statistic Name** | **Value** |
| --- | --- |
| Mean | 99.74 |
| Median | 101.00 |
| Variance | 130.77 |
| Standard deviation | 11.44 |

The mean shows the average in the data set. The median is the middle number in the data set. The variance shows the difference between the data points and the mean. Standard deviation shows the dispersion of the data points from the mean. The mean and the median are close to each other in my team’s away data set, being only 1.26 away from each other. This makes my data bell-shaped. It’s best to use the mean to represent bell-shaped data. My team performs better in home games than in away games as the mean and standard deviation are higher on the home data set. By comparing the deviation of points from the home and the away games I can see that there is a higher variance between the points scored in home games than away games.

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

Confidence intervals are used to estimate the range of certainty of a data set. For example, if my teams relative skill is taken between the years of 2013 and 2015 then there is a 95% chance it will be within the range of the confidence interval of (1502.02, 1507.18). The probability of a team being better than my team, the Atlanta Hawks during the years of 2013 and 2015 was 37.98% and the chances of them being worse than my team was 62.02%. It would not be unusual for a team to be worse than the Hawks at a 62% chance. If a different confidence level was used, then the confidence interval would be further apart if it was higher and closer together if it was lower.

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

The confidence interval for the Chicago Bulls between 1996 and 1998 was 1487.66, 1493.65 at a confidence level of 95%. If data was taken from the Bulls ELO between 1996 and 1998 there’s a 95% chance it would be within 1487.66 and 1493.65. If a different confidence level was used, then the confidence interval would be further apart if it was higher and closer together if it was lower. The confidence interval of the Bulls is lower than the Hawks confidence interval. The probabilities of a team having an average skill lower than the Bulls in 1996 to 1998 is 97.32% whereas the chance of a team having a lower skill level than the Hawks in 2013 to 2015 is 62.03%. This means at the time the bulls had a higher relative elo than the Hawks. This also signifies that the average skill level of teams has gone up from the years of 1996 to 1998 and 2013 to 2015.

Describe the results of your statistical analyses clearly, using proper descriptions of statistical terms and concepts.

The results of the data that I analyzed were that the Chicago Bulls had a higher relative skill level from the years 1996 to 1998 than the Atlanta Hawks in 2013 to 2015. The Bulls also had a higher average points per game, but they also had a higher variance of points per game. The importance of the analyses that I just performed allow me to see how the measures of central tendency can be used to summarize and compare statistics. What this means for the scenario is that the Hawks should analyze the Bulls old plays from 1996 to 1998 to try to improve their performance.

REFERENCES:

FiveThirtyEight. (April 26, 2019). FiveThirtyEight NBA Elo dataset. Kaggle. Retrieved from https://www.kaggle.com/fivethirtyeight/fivethirtyeight-nba-elo-dataset/