**MAT 243 Project One Summary Report**

Mason Cabirac

mason.cabirac@snhu.edu

Southern New Hampshire University

* **Introduction: Problem Statement**

Using key variables associated with the performance of different basketball teams, I will solve the problem of comparing two of them: the Chicago Bulls and the Houston Rockets, thereby using data-driven analytics to help make decisions to improve the Houston Rocket's performance. While the dataset includes the points scored in each game by all teams between the following years for each of the two respective teams, the analyses heavily involve that by the Chicago Bulls, from 1996 to 1998 and the Houston Rockets, from 2013 to 2015. I will be running a data visualization of points scored for each team during these respective years, in addition to one comparing the two teams, descriptive statistics, confidence intervals for the average relative skill of all teams in both ranges of years. I will also be determining the probability that a team has an average skill level less than that of the Chicago Bulls and the Houston Rockets for their respective years.

* **Introduction: Your Team and the Assigned Team**

Table 1. Information on the Teams

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

* **Data Visualization: Points Scored by Your Team**

Data visualization is used to help make decisions by clarifying data sets. For instance, a pie chart may show some alternative materials used to produce a good by an industry. It may then follow to determine the reasons certain materials are chosen over others and recommend an option that aligns with a company's standards.



I picked this histogram because it revealed something the scatterplot didn't clearly show: the relatively extreme peak of frequency between the scored points of about 95 and 115. That is, there is clearly an extreme effort and ability to break a score of 95, but not as much to go above 115.

* **Data Visualization: Points Scored by the Assigned Team**

*In the Python script, you created a visualization for the distribution of points scored by the assigned team.*

*See Step 4 in the Python script to address the following items in a paragraph response:*

* *In this activity, you were asked to pick one of the two plots that best describes the data distribution of the variable for the assigned team. Include this plot in your report.*
* *Why did you pick this plot? Explain.*
* *What can you say about the distribution of the variable by visually inspecting this plot? What does this signify?*



It was difficult this time to choose one plot over the other. In this case, the decreasing points scored by the years imply also that the frequency of higher scores decreased, leading to a wide variance of their performance. Therefore, the histogram, which does not show the decline by years, was not as telling.

* **Data Visualization: Comparing the Two Teams**

Data visualization is used to compare two different data distributions in a way that may shed light on the relationship between the meeting of a goal and that which is left to be desired. For instance, a company may hire a skilled worker with competency A over competency B, given that a line chart shows that the mean of the value of the former is consistently greater than the latter.



I chose this boxplot because it clarified something extant, but not as visible in the histogram: The teams are not terribly far apart in skill. The maximums are close, as are the medians and quartiles. The largest difference is in the minimums and the presence of outliers in the Houston Rockets.

* **Descriptive Statistics: Relative Skill of Your Team**

*See Step 6 in the Python script to address the following items:*

* *Summarize* ***all*** *statistics in a formatted table as shown below. Use one row for each statistic. You will need to add rows to the table in order to include all of your statistics.*

Table 2. Descriptive Statistics for Relative Skill of Your Team

|  |  |
| --- | --- |
| **Statistic Name** | **Value** |
| Mean | 1596.29 |
| Median | 1602.14 |
| Variance | 1633.70 |
| Standard Deviation | 40.79 |

* *In general, how are the measures of central tendency and variability used to analyze a data distribution?*
* *Interpret each statistic in detail and explain what it represents in this scenario.*
* *Use the mean and the median to describe the distribution of points scored by your team in home games.*
* *Describe the skew: Is it left, right, or bell-shaped?*
* *Explain which measure of central tendency is best to use to represent the center of the distribution based on its skew.*

Central tendency tries to determine a value in a dataset which most of the other values are distributed around. Variability attempts to show how far numbers in a data set are from each-other.

A mean is an average, where the data set is added together and divided by the number of numbers within it. A median is the middle value in a dataset. A lower mean or median might be interpreted, in this case, as a team having relatively less skill, while a higher mean or median might be interpreted as a team having relatively more skill. The variance is the spread between numbers in a data set. A higher variance may show a less inconsistent skill level, while a lower variance may show a more consistent skill level.

The mean of the distribution of points scored by the Houston Rockets from 2013 to 2015 is about 105.85. The median of the same distribution is 106.0. The skew is bell-shaped. Considering that only, the mean may be used as a representative of central tendency.

* **Descriptive Statistics: Relative Skill of Assigned Team**

*In the Python script, you calculated descriptive statistics on the points scored by your team in games played at opponent’s venue (away). These included the mean, median, variance, and standard deviation for the relative skill of the assigned team.*

*See Step 7 in the Python script to address the following items:*

* *Summarize* ***all*** *statistics in a formatted table as shown below. Use one row for each statistic. You will need to add rows to the table in order to include all of your statistics.*

Table 3. Descriptive Statistics for Relative Skill of Assigned Team

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

* *Interpret each statistic in detail and explain what it represents in this scenario.*
* *Use the mean and the median to describe the distribution of points scored by your team in away games.*
* *Describe the skew: Is it left, right, or bell-shaped?*
* *Explain which measure of central tendency is best to use to represent the center of the distribution based on its skew.*
* *Is your team performing better in games played at home than those played away? Use the mean and the standard deviation to answer this question. What can be deduced by comparing the standard deviation of points scored in home games and points scored in away games?*

The mean of the points scored by the Bulls from 1996 to 1998 is 101.68, while the median is 102.5. The data, when visualized, represents a bell shape, so the mean may be used as a representative of the distribution’s central tendency.

For each team during their respective years, the relative skill is higher for the Chicago Bulls than the Houston Rockets for their respective ranges of years. The standard deviation for the skill of the Bulls is greater than the Rockets, however, so the Bulls are more varied in their scores.

* **Confidence Intervals for the Average Relative Skill of All Teams in Your Team’s Years**

*In the Python script, you calculated a 95% confidence interval for the average relative skill of all teams in the league during the years of your team. Additionally, you calculated the probability that a given team in the league has a relative skill level less than that of the team that you picked.*

*See Step 8 in the Python script to address the following items:*

* *Report the confidence interval in a formatted table as shown below.*

Table 4. Confidence Interval for Average Relative Skill of Teams in Your Team’s Years

|  |  |
| --- | --- |
| **Confidence Level (%)** | **Confidence Interval** |
| 95% | ( 1502.02 , 1507.18 )  *\*Round off to 2 decimal places.* |

* *Describe how confidence intervals are generally used in estimating the measures of central tendency for a population.*
* *Provide a detailed interpretation of the confidence interval in terms of the average relative skill of teams in the range of years that you picked.*
* *How would your interval be different if you had used a different confidence level?*
* *What is the probability that a given team in the league has a relative skill level less than that of the team that you picked? Is it unusual that a team has a skill level less than your team?*

Confidence intervals are generally used to show that the parameter being estimated, in this case the average relative skill of basketball teams in the data set, is likely to lie between the first and second values of the interval. Regarding a 95% confidence level, the average skill of teams between 2013 to 2015 is likely to lie between 1502.02 and 1507.18.

With a greater confidence level, the confidence interval widens and there is a greater chance that the parameter is between the first and second values.

The probability that a given team in the league has a relative skill level less than that of the Houston Rockets is 79.14%. Therefore, it is unusual that a team had a skill level less than the Houston Rockets during between 2013 to 2015.

* **Confidence Intervals for the Average Relative Skill of All Teams in the Assigned Team’s Years**

*In the Python script, you calculated a 95% confidence interval for the average relative skill of all teams in the league during the years of the assigned team. Additionally, you calculated the probability that a given team in the league has a relative skill level less than that of the assigned team.*

*See Step 9 in the Python script to address the following items:*

* *Report the confidence interval in a formatted table as shown below.*

Table 5. Confidence Interval for Average Relative Skill of Teams in Assigned Team’s Years

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

* *Provide a detailed interpretation of the confidence interval in terms of the average relative skill of teams in the assigned team’s range of years.*
* *Discuss how your interval would be different if you had used a different confidence level.*
* *How does this confidence interval compare with the previous one? What does this signify in terms of the average relative skill of teams in the range of years that you picked versus the average relative skill of teams in the assigned team’s range of years?*

The confidence interval of the average skill level of teams between 1996 to 1998 is wider in range than during that of 2013 to 2015, the former being between 1487.66 and 1493.65. It’s maximmum value is less than the minimum value for 2013 to 2015, showing that the average skill of the those respective years were slightly greater than in 1996 to 1998.

The probability that a team would have an average skill level less than the Bulls was 97.32%, so it can be inferred that the Bulls had better scores during those years than most other teams did and that the Rockets, while seemingly matching or surpassing them in 2013 to 2015, had fiercer competition during their own years.

* **Conclusion**

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

* *What is the practical importance of the analyses that were performed?*
* *Describe what these results mean for the scenario.*

The main importance of the analyses that were performed is that the Bulls and the Rockets were about neck and neck for the years of 1996 to 1998 and 2013 to 2015, respectively, however with the Bulls being more skilled than other teams during their own years. This shows that the Bulls, during their years, did not face as great of competition as the Rockets, during their years. I advise management to carefully observe high-performing teams in the present time and manage the Rockets after the fashion of the foremost achieving.

Another interesting observation is that the skill level varied more for the Bulls than the Rockets. I want to suppose there was likely a certain level of trust that the Bulls would outperform many teams, resulting in laissez-faire management, resulting in this variance, despite an excellent average result. If this is true, and it could be gotten to the root of how it occurred, something like this may be emulated in another team, even the currently middling Houston Rockets.

* **Citations**

*You were* ***not*** *required to use external resources for this report. If you did not use any resources, you should remove this entire section. However, if you did use any resources to help you with your interpretation, you* ***must*** *cite them. Use proper APA format for citations.*

Insert references here in the following format:

Author's Last Name, First Initial. Middle Initial. (Year of Publication). Title of book: Subtitle of book, edition. Place of Publication: Publisher.