# MAT 243 Project One Summary Report

[Full Name]

[SNHU Email]

Southern New Hampshire University

**Notes:**

* Replace the bracketed text on page one (the cover page) with your personal information.
* You will use your selected team for all three projects

## Introduction: Problem Statement

*Discuss the statement of the problem in terms of the statistical analyses that are being performed. In your response, you should address the following questions:*

* What is the problem you are going to solve?
* What data set are you using?
* What statistical methods will you be using to do the analysis for this project?

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include Python code in your report.*

|  | **game\_id** | **year\_id** | **fran\_id** | **pts** | **opp\_pts** | **elo\_n** | **opp\_elo\_n** | **game\_location** | **game\_result** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 199511030CHI | 1996 | Bulls | 105 | 91 | 1598.2924 | 1531.7449 | H | W |
| **1** | 199511040CHI | 1996 | Bulls | 107 | 85 | 1604.3940 | 1458.6415 | H | W |
| **2** | 199511070CHI | 1996 | Bulls | 117 | 108 | 1605.7983 | 1310.9349 | H | W |
| **3** | 199511090CLE | 1996 | Bulls | 106 | 88 | 1618.8701 | 1452.8268 | A | W |
| **4** | 199511110CHI | 1996 | Bulls | 110 | 106 | 1621.1591 | 1490.2861 | H | W |

printed only the first five observations...

Number of rows in the data set = 246

## Introduction: Your Team and the Assigned Team

*In this project, you picked a team and you were assigned a team to do comparative analysis.*

*See Steps 1 and 2 in the Python script to address the following items:*

* What team did you pick and what years were picked to do the analysis?
* What team and range of years were you assigned for the comparative study? (Hint: This is called the assigned team in the Python script.) Present this information in a formatted table as shown below.

Table 1. Information on the Teams

|  | **Name of Team** | **Assigned Years** |
| --- | --- | --- |
| 1. Yours | Team (e.g., Heat) | XXXX-YYYY (e.g. 2013 - 2015) |
| 2. Assigned | Team (e.g. Bulls) | XXXX-YYYY (e.g. 2013 - 2015) |

* Answer the questions in a paragraph response. Remove all questions and this note (but not the table) before submitting! Do not include Python code in your report.*

|  | **game\_id** | **year\_id** | **fran\_id** | **pts** | **opp\_pts** | **elo\_n** | **opp\_elo\_n** | **game\_location** | **game\_result** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 199511030CHI | 1996 | Bulls | 105 | 91 | 1598.2924 | 1531.7449 | H | W |
| **1** | 199511040CHI | 1996 | Bulls | 107 | 85 | 1604.3940 | 1458.6415 | H | W |
| **2** | 199511070CHI | 1996 | Bulls | 117 | 108 | 1605.7983 | 1310.9349 | H | W |
| **3** | 199511090CLE | 1996 | Bulls | 106 | 88 | 1618.8701 | 1452.8268 | A | W |
| **4** | 199511110CHI | 1996 | Bulls | 110 | 106 | 1621.1591 | 1490.2861 | H | W |

printed only the first five observations...

Number of rows in the data set = 246

## Data Visualization: Points Scored by Your Team

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

*See Step 3 in the Python script to address the following items in a paragraph response:*

* In general, how is data visualization used to study data distributions and trends?
* In this activity, you were asked to pick one of the two plots that best describes the data distribution of the variable for your team. Include a screenshot of 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?

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include Python code in your report.*

| **game\_id** | **year\_id** | **fran\_id** | **pts** | **opp\_pts** | **elo\_n** | **opp\_elo\_n** | **game\_location** | **game\_result** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 201210300MIA | 2013 | Heat | 120 | 107 | 1666.3193 | 1586.1121 | H | W |
| **1** | 201211020NYK | 2013 | Heat | 84 | 104 | 1647.6675 | 1548.2699 | A | L |
| **2** | 201211030MIA | 2013 | Heat | 119 | 116 | 1650.0934 | 1554.4674 | H | W |
| **3** | 201211050MIA | 2013 | Heat | 124 | 99 | 1656.5652 | 1504.0280 | H | W |
| **4** | 201211070MIA | 2013 | Heat | 103 | 73 | 1659.7239 | 1361.5804 | H | W |

printed only the first five observations...

Number of rows in the data set = 246

## 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?
* *Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include Python code in your report.*
* Chart, histogram

  Description automatically generated
* Chart, line chart

  Description automatically generated

## Data Visualization: Comparing the Two Teams

*In the Python script, you created a visualization for the difference in the distributions of points scored by your team and the assigned team.*

*See Step 5 in the Python script to address the following items in a paragraph response:*

* In general, how is data visualization used to compare two different data distributions?
* In this activity, you were asked to pick one of the two plots that best compares the data distributions of your team with the assigned team. Include a screenshot of this plot in your report.
* Why did you pick this plot? Explain.
* How do the two distributions compare to each other?

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include Python code in your report.*

*Chart, box and whisker chart

Description automatically generated*

*Chart, histogram

Description automatically generated*

## Descriptive Statistics: Points Scored By Your Team in Home Games

*In the Python script, you calculated descriptive statistics on the points scored by your team in games played at home venue. These included the mean, median, variance, and standard deviation for the 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 Points Scored by Your Team in Home Games

| **-** | **Value** |
| --- | --- |
| Statistic  *(for example, Mean)* | X.XX  *\*Round off to 2 decimal places.* |

* 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.

* Answer the questions in a paragraph response. Remove all questions and this note (but not the table) before submitting! Do not include Python code in your report.*

Points Scored by Your Team in Home Games (2013 to 2015)

-------------------------------------------------------

Mean = 102.56

Median = 103.0

Variance = 126.05

Standard Deviation = 11.23

## Descriptive Statistics: Points Scored By Your Team in Away Games

*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 Points Scored by Your Team in Away Games

| **Statistic Name** | **Value** |
| --- | --- |
| Statistic  *(for example, Mean)* | X.XX  *\*Round off to 2 decimal places.* |

* 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.
  1. Describe the skew: Is it left, right, or bell-shaped?
  2. 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?
* *Answer the questions in a paragraph response. Remove all questions and this note (but not the table) before submitting! Do not include Python code in your report.*

Points Scored by Your Team in Away Games (2013 to 2015)

-------------------------------------------------------

Mean = 97.28

Median = 97.0

Variance = 101.4

Standard Deviation = 10.07

## 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** |
| --- | --- |
| XX% (for example, 95%) | (X.XX, X.XX)  *\*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?
* *Answer the questions in a paragraph response. Remove all questions and this note (but not the table) before submitting! Do not include Python code in your report.*

Confidence Interval for Average Relative Skill in the years 2013 to 2015

------------------------------------------------------------------------------------------------------------

95% confidence interval (unrounded) for Average Relative Skill (ELO) in the years 2013 to 2015 = (1502.0236894390478, 1507.1824625533618)

95% confidence interval (rounded) for Average Relative Skill (ELO) in the years 2013 to 2015 = ( 1502.02 , 1507.18 )

Probability a team has Average Relative Skill LESS than the Average Relative Skill (ELO) of your team in the years 2013 to 2015

----------------------------------------------------------------------------------------------------------------------------------------------------------

Which of the two choices is correct?

Choice 1 = 0.159

Choice 2 = 0.841

Choice #2 is correct

## 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** |
| --- | --- |
| XX% (for example, 95%) | (X.XX, X.XX)  *\*Round off to 2 decimal places.* |

* 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?

* Answer the questions in a paragraph response. Remove all questions and this note (but not the table) before submitting! Do not include Python code in your report.*

Confidence Interval for Average Relative Skill in the years 1996 to 1998

------------------------------------------------------------------------------------------------------------

95% confidence interval (unrounded) for Average Relative Skill (ELO) in the years 2013 to 2015 = (1487.7065262875703, 1493.596609849239)

95% confidence interval (rounded) for Average Relative Skill (ELO) in the years 2013 to 2015 = ( 1487.71 , 1493.6 )

Probability a team has Average Relative Skill LESS than the Average Relative Skill (ELO) of your team in the years 2013 to 2015

----------------------------------------------------------------------------------------------------------------------------------------------------------

Which of the two choices is correct?

Choice 1 = 0.1629

Choice 2 = 0.8371

Number 2 is correct

## 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.

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include Python code in your report.*

## 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.