# MAT 243 Project Three Summary Report

Tyler Ellis

Tyler.ellis1@snhu.edu

Southern New Hampshire University

## 1. Introduction

* Understanding and anticipating a team's performance is essential for making informed choices and development in the professional world of basketball. In the current research, we look closely into our basketball team's past performance data to find trends and factors which significantly impact the number of regular-season victories.

The dataset under analysis includes past performance benchmarks for our team through various seasons. This dataset covers a variety of factors like player scoring averages, assists, rebounds, and turnovers, amongst others. The main objective is to determine how each of these separate and combined factors correspond with the number of regular-season wins.

This findings will be useful for a number of purposes. We allow tactical decision-making for specific training and greater in game strategies by recognizing performance factors that greatly affect wins. These insights allow the coaching staff to develop customized player training programs that ensure optimal contributions to team victories. This clarity provides coaches with the necessary planning needed for things like player trades, drafts, and other roster changes all of which are based on the predictive power of certain statistics.

To get a complete outlook of the data, we will use a multitude of analytical tools. The methods will include:

1. Scatterplots: These visual representations will throw light on the links between certain performance indicators and the number of wins, assisting with early visual interpretations.
2. Correlation Coefficient: We may prioritize which measures are most significant by evaluating the degree and direction of the relationship between performance metrics and wins.
3. multiple Regression: This complex approach analyzes multiple performance metrics at the same time, evaluating their combined impact on the number of wins and detecting possible relationships or similarities between metrics.
4. Simple Linear Regression: We will look at the predictive value of specific performance metrics on the number of wins, offering insight into metric outcomes.

## 2. Data Preparation

Avg\_pts\_differential:

The variable avg\_pts\_differential is the average points differential between each team and an opposite team in a regular season.

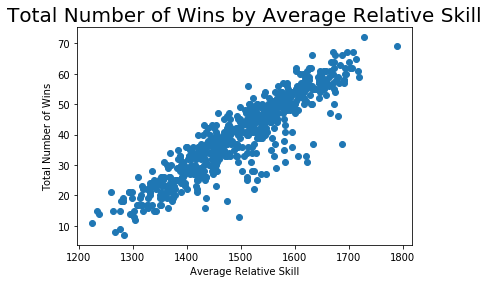
In other words, think about two rival teams facing each other. At the end of the game, find the difference of the 2 scores. This is called the "the points differential." If we take the points differential for all the games in the season, and find the average, this is called "average points differential." The averate points differential is a good indicator of how much better or worse a team's scoring percentage is compared to their opponents over the course of the season.

Avg\_elo\_n:

The avg\_elo\_n variable is the average relative skill of each team in a regular season.  
  
In other words, the average elo rating is a method for evaluating players or teams based on their skill levels. A score is issued to each team that displays their skill level. Whenever a lower scored team beats a higher scored team, the team with the lower score significantly increases in rank, while the higher scored team drops in score. In the NFL, this is how they determine who goes to the Superbowl. The average Elo rating is the total of the scores for 1 season. It represents a team's relative skill level or strength compared to others.

## 3. Simple Linear Regression: Scatterplot and Correlation for the Total Number of Wins and Average Relative Skill

Analysts may visually analyze the link between two variables utilizing data visualization techniques such as scatterplots. Individual data points for each paired set of values may be plotted to identify patterns, trends, clusters, or even outliers in the data. This visual analysis may offer a basic understanding of the type of the association between the two variables.  
 The correlation coefficient calculates the strength of the relationship between two variables. It has a value between -1 and 1.A correlation of one or close to one denotes a good association. A -1 or near -1 indicates a negative association. There is no association if the value is 0.The coefficient's absolute value indicates the strength of the link. The higher the number, the stronger the association.



We can determine the following from the scatterplot and Pearson correlation coefficient:

* The total number of victories and the average relative skill have a positive linear relationship. As a result, teams with greater relative skill levels win more games.
* The correlation coefficient number 0.9072 indicates the exact strength of this linear relationship.
* The coefficient is 0.9072, which is nearly one. This demonstrates a very strong positive linear relationship between total number of wins and average relative skills. In essence, teams with greater relative skill levels are more likely to win a greater percentage of games. This positive relationship shows that increasing a team's relative skill level is likely to improve the overall number of wins they may earn.
* After rounding, the p value calculated is essentially 0.0, which is significantly lower than the 0.01 significance limit. Thus, at a 1% level of significance, the correlation coefficient of 0.9072 is statistically significant. This suggests that the observed higher positive relationship between total number of wins and average relative skill is not attributable simply to chance, instead due to a real relationship in the data.

## 4. Simple Linear Regression: Predicting the Total Number of Wins using Average Relative Skill

* The simple linear regression model is used to compare the correlation between two variables. One variable,the predictor or independent variable, is used to predict the value of the other, the response or dependent variable. The method of linear regression determines the best fitting linear path, or regression line, to demonstrate this relationship.
* What is the equation for your model?   
  total\_wins = -128.2475 + 0.1121 × avg\_elo\_n
* What are the results of the overall F-test? Summarize all important steps of this hypothesis test. This includes:
* **Null Hypothesis** (H₀): β=0 The slope of the relationship between average relative skill and total wins is zero.
* **Alternative Hypothesis**(H₁): β ≠ 0 The slope of the relationship between average relative skill and total wins is not zero.
* Level of Significance   
  α=0.01 or 1%

Table 1: Hypothesis Test for the Overall F-Test

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | X.XX  \*Round off to 2 decimal places. |
| P-value | X.XXXX  \*Round off to 4 decimal places. |

* 1. Since the p value is less than the level of significance = 0.01, we reject the null hypothesis. This suggests there is a significant linear relationship between the average relative skill and total wins.
* Based on the results of the overall F-test, can average relative skill predict the total number of wins in the regular season?
* What is the predicted total number of wins in a regular season for a team that has an average relative skill of 1550? Round your answer down to the nearest integer.
* What is the predicted number of wins in a regular season for a team that has an average relative skill of 1450? Round your answer down to the nearest integer.

** 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.

**5. Multiple Regression: Scatterplot and Correlation for the Total Number of Wins and Average Points Scored**

You will now add a second predictor to the model from section 3 and create a multiple regression model for the total number of wins.

You constructed a scatterplot of total number of wins and average points scored. You also calculated the Pearson correlation coefficient along with its P-value.

See Step 4 in the Python script to answer the following questions:

* In this activity, you generated a scatterplot of the total number of wins and average points scored. Include a screenshot of this plot in your report.
* What do the scatterplot and the Pearson correlation coefficient tell you about the association between total number of wins and average points scored?
* Is the correlation coefficient statistically significant based on the P-value? Use a 1% level of significance.

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

## 6. Multiple Regression: Predicting the Total Number of Wins using Average Points Scored and Average Relative Skill

You created a multiple regression model with the total number of wins as the response variable, with average points scored and average relative skill as predictor variables.

See Step 5 in the Python script to answer the following questions:

* In general, how is a multiple linear regression model used to predict the response variable using predictor variables?
* What is the equation for your model?
* What are the results of the overall F-test? Summarize all important steps of this hypothesis test. This includes:
  1. Null Hypothesis (statistical notation and its description in words)
  2. Alternative Hypothesis (statistical notation and its description in words)
  3. Level of Significance
  4. Report the test statistic and the P-value in a formatted table as shown below:

Table 2: Hypothesis Test for the Overall F-Test

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | X.XX  \*Round off to 2 decimal places. |
| P-value | X.XXXX  \*Round off to 4 decimal places. |

* 1. Conclusion of the hypothesis test and its interpretation based on the P-value
* Based on the results of the overall F-test, is at least one of the predictors statistically significant in predicting the total number of wins in the season?
* What are the results of individual t-tests for the parameters of each predictor variable? Is each of the predictor variables statistically significant based on its P-value? Use a 1% level of significance.
* Report and interpret the coefficient of determination.
* What is the predicted total number of wins in a regular season for a team that is averaging 75 points per game with a relative skill level of 1350?
* What is the predicted total number of wins in a regular season for a team that is averaging 100 points per game with an average relative skill level of 1600?

** 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.

## 7. Multiple Regression: Predicting the Total Number of Wins using Average Points Scored, Average Relative Skill, Average Points Differential, and Average Relative Skill Differential

You created a multiple regression model with the total number of wins as the response variable, with average points scored, average relative skill, average points differential, and average relative skill differential as predictor variables.

See Step 6 in the Python script to answer the following questions:

* In general, how is a multiple linear regression model used to predict the response variable using predictor variables?
* What is the equation for your model?
* What are the results of the overall F-test? Summarize all important steps of this hypothesis test. This includes:
  1. Null Hypothesis (statistical notation and its description in words)
  2. Alternative Hypothesis (statistical notation and its description in words)
  3. Level of Significance
  4. Report the test statistic and the P-value in a formatted table as shown below:

Table 3: Hypothesis Test for Overall F-Test

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | X.XX  \*Round off to 2 decimal places. |
| P-value | X.XXXX  \*Round off to 4 decimal places. |

* 1. Conclusion of the hypothesis test and its interpretation based on the P-value
* Based on the results of the overall F-test, is at least one of the predictors statistically significant in predicting the number of wins in the season?
* What are the results of individual t-tests for the parameters of each predictor variable? Is each of the predictor variables statistically significant based on its P-value? Use a 1% level of significance.
* Report and interpret the coefficient of determination.
* What is the predicted total number of wins in a regular season for a team that is averaging 75 points per game with a relative skill level of 1350, average point differential of -5 and average relative skill differential of -30?
* What is the predicted total number of wins in a regular season for a team that is averaging 100 points per game with a relative skill level of 1600, average point differential of +5 and average relative skill differential of +95?

** 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.

## 8. Conclusion

Describe the results of the statistical analyses clearly, using proper descriptions of statistical terms and concepts. Fully describe what these results mean for your scenario.

* Briefly summarize your findings in plain language.
* What is the practical importance of the analyses that were performed?

## 9. 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.