# 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.

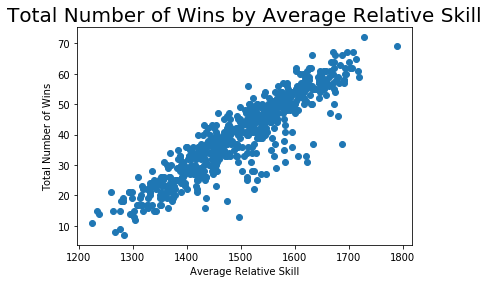
In other words, think about two rival teams facing each other. At the end of the game, find the difference of the 2 scores. This is called the "the points differential." If we take the points differential for all the games in the season, and find the average, this is called "average points differential." 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 | 53.52 |
| P-value | 0.00000 |

* 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.
* Yes, due to the f test results, average relative skill is able to predict the overall number of wins in the regular season. The predicted total number of wins is 43. The predicted number of wins in a regular season is 34.

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

As the amount of scores increase, the total number of wins increases as well. This indicates that this is a positive relationship. This positive linear connection is supported by the coefficient value of 0.4777. While it indicates an acceptable relationship, it is not as significant as with value closer to 1.  
  
 Based on the p value, the coefficient is significant. The p value is 0.0. This is less than the level of significance, 0.01. We reject the null hypothesis. There is no linear correlation between average points scored and the total wins.

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

A multiple linear regression model is used by data analysts and scientists to explain the relationship between one response value/variable and two or more predictor values/variables. Multiple regression is used to understand more about the connection between a number of independent factors and a dependent variable.  
  
total\_wins=−152.5736+0.3497×avg\_pts+0.1055×avg\_elo\_n

* Null Hypothesis (H0): β1 =β2 =0 There is no relationship between the predictors and the response.
* Alternative Hypothesis (H1): At least one βx (where x = 1, 2) is not equal to 0. There is a relationship between at least one predictor and the response.
* Level of Significance: We can assume a significance level of 0.05 or 5%.

Table 2: Hypothesis Test for the Overall F-Test

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | 1580.0 |
| P-value | 0.00001 |

* 1. For the f test, the p value is 0.0, which is less than the 0.01 level of significance. The null hypothesis is rejected. In the model, at least one predictor is significant.

Based on the results of the f test, at least one predictor is significant. Both predictors have p values of 0.000. Both are significant at a 0.01 level of significance.

The coefficient of determination is 0.837. The model with the predictor variables avg\_pts and avg\_elo\_n can explain approximately 83.7% of the variability in total wins.  
  
A team with an average of 75 points per game and a relative skill level of 1350 has approximately 62 wins in a regular season.  
   
predicted\_wins=−152.5736+0.3497(75)+0.1055(1350) predicted\_wins≈62.44predicted\_wins≈62.44

A team with an average of 100 points per game and a relative skill level of 1600 has approximately 143 wins in a regular season.

predicted\_wins=−152.5736+0.3497(100)+0.1055(1600) predicted\_wins≈143.73predicted\_wins≈143.73

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

* 1. Multiple linear regression is a statistical method that calculates the value of one dependent variable based on the values of two or more independent variables. The main goal is to identify the best fit straight line that accurately determines the output values within a given range.  
       
     total\_wins=34.5753+0.2597(avg\_pts)−0.0134(avg\_elo\_n)+ 1.6206(avg\_pts\_differential)+ 0.0525(avg\_elo\_differential)
* Null Hypothesis :H0 : βavg\_pts =βavg\_elo\_n = βavg\_pts\_differential = βavg\_elo\_differential =0  
   This means that none of the predictor variables are significant in predicting the response variable.
* Alternative Hypothesis : Ha : At least one β=0  
   This means that at least one predictor variable is statistically significant in predicting the response variable.
* Level of Significance: Not given, but usually is 0.05 or 0.01

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. |

The p value is relatively close to zero. The null hypothesis is rejected. There is considerable evidence that at least one of the predictors can accurately forecast the total number of wins.  
  
At least one of the predictors is significant in predicting total season wins.  
  
The coefficient of determination which is also called R squared is 0.878. This means that about 87.8% of the total number of wins can be predicted by the model.  
  
“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”  
  
We can find this answer with the equation below:   
  
total\_wins = 34.5753 + 75 \* 0.2597 + 1350 \* (-0.0134) + (-5) \* 1.6206 + (-30) \* 0.0525  
  
The total number of wins in a regular season for the team is approximately 26.3, or 26.  
  
“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”  
  
We can find this answer with the equation below:  
  
total\_wins = 34.5753 + 100 \* 0.2597 + 1600 \* (-0.0134) + 5 \* 1.6206 + 95 \* 0.0525  
  
The total number of wins in a regular season for the second team is about 52.2, or 52.

## 8. Conclusion We implemented a multiple regression model in our analysis to determine how various metrics like average points scored and relative skill levels determine a team's wins in a single season. The metrics explain about 87.8% of the variance in the total amount of wins. The most significant metric was the average points differential. Essentially, the more a team outscores the other team, the more wins it will likely get. The practicality of this analysis stems from the fact that it can improve team and coach's strategies towards practice. Knowing which metrics strongly influence wins and losses will assist coaches in focusing their efforts in the right direction.

## 9. Citations

zyBooks. (2023). Module 1: escriptive Statistics and Probability. MAT-243. <https://learn.zybooks.com/zybook/MAT-243-J1254-OL-TRAD-UG.23EW1/chapter/1/>   
zyBooks. (2023). Module 2: Probability Distributions. MAT-243. <https://learn.zybooks.com/zybook/MAT-243-J1254-OL-TRAD-UG.23EW1/chapter/2/>   
zyBooks. (2023). Module 3: Confidence Intervals and Hypothesis Testing. MAT-243. <https://learn.zybooks.com/zybook/MAT-243-J1254-OL-TRAD-UG.23EW1/chapter/3/>

zyBooks. (2023). Module 4: Hypothesis Testing with Two Populations. MAT-243. [https://learn.zybooks.com/zybook/MAT-243-J1254-OL-TRAD-UG.23EW1/chapter/4/](https://learn.zybooks.com/zybook/MAT-243-J1254-OL-TRAD-UG.23EW1/chapter/1/)   
zyBooks. (2023). Module 5: Linear Regression. MAT-243. [https://learn.zybooks.com/zybook/MAT-243-J1254-OL-TRAD-UG.23EW1/chapter/5/](https://learn.zybooks.com/zybook/MAT-243-J1254-OL-TRAD-UG.23EW1/chapter/2/)zyBooks. (2023). Module 6: Multiple Regression. MAT-243. [https://learn.zybooks.com/zybook/MAT-243-J1254-OL-TRAD-UG.23EW1/chapter/6/](https://learn.zybooks.com/zybook/MAT-243-J1254-OL-TRAD-UG.23EW1/chapter/1/)   
zyBooks. (2023). Module 7: Interpreting Multiple Regression. MAT-243. [https://learn.zybooks.com/zybook/MAT-243-J1254-OL-TRAD-UG.23EW1/chapter/7/](https://learn.zybooks.com/zybook/MAT-243-J1254-OL-TRAD-UG.23EW1/chapter/2/)