# MAT 243 Project Three Summary Report

Armon Wilson

Armon.wilson@snhu.edu

Southern New Hampshire University

## 1. Introduction

The dataset utilizes basketball team data from 1995 to 2015, looking at how stats like average points, relative skill, and score differentials relate to the total number of wins in a season. The goal is to create models that predict how many wins a team might get based on these stats. The results obtained from statistical analyses, including correlation, simple linear regression, and multiple regression, will serve as a predictive model for the team. These analyses will unveil relationships between performance metrics and wins, providing insights into which factors, such as average points scored and relative skill differentials, influence the team's success. The results will help the team's coach and management make educated decisions to improve and win more matches.

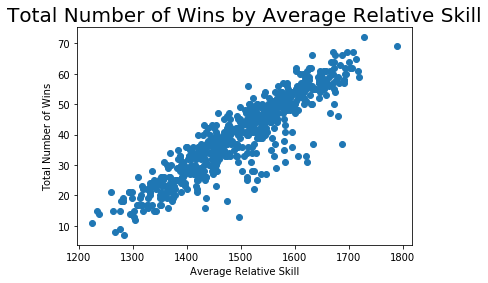
## 2. Data Preparation

The variable avg\_pts\_differential reflects the average difference in points scored between a team and their opponents in a regular basketball season. The metric shows, on average, how many more points a team scores compared to their opponents per game throughout a season. A positive value indicates the team, on average, scores more than their opponents, while a negative value means the opponents outscore them.

The variable avg\_elo\_n represents the average relative skill of each team in a regular season, calculated using the ELO rating system. It as a measure that gauges a team's average level of skill throughout a season. The higher the avg\_elo\_n value, the higher the perceived skill level of the team.

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

Visuals like scatterplots show a clear picture of how two things relate. In this case, the scatterplot showed how the total wins and skill levels of basketball teams connect. The Pearson correlation coefficient, on the other hand, quantifies the strength and direction of this relationship. For this specific scatterplot, a positive correlation coefficient of 0.9072 suggests a strong positive association between total wins and average relative skill. The significance of the correlation coefficient is determined by the associated p-value. With a p-value of 0.0 (assuming a 1% level of significance), it indicates a highly significant correlation between these variables. The scatterplot visualizes this strong positive relationship, while the p-value emphasizes the statistical significance of this association.



## 

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

The linear regression model predicts a response variable using a single predictor variable.

Total Wins = -128.2475 + 0.1121 \* Average Relative Skill

The overall F-test helps determine if the model's coefficients are significant in predicting the response variable. For this test:

1. Null Hypothesis (Ho):
   1. All coefficients are equal to zero, implying the model has no predictive power.
   2. (Ho): β1 = 0
2. Alternative Hypothesis (Ha):
   1. At least one coefficient is not equal to zero, indicating the model has some predictive power.
   2. (Ha): β1 ≠ 0

c. Level of Significance: 0.05.

Table 1: Hypothesis Test for the Overall F-Test

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | 2865.00 |
| P-value | 0.0000 |

With the p-value less than 0.05, we reject the null hypothesis. This implies that at least one coefficient in the model is not zero, suggesting the model has predictive power.

For a team with an average relative skill of 1550, the predicted total number of wins in a regular season would be approximately 25. For a team with an average relative skill of 1450, the predicted wins would be approximately 9.

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

A blue dots on a white background

Description automatically generated

The scatterplot visually illustrates the relationship between the total number of wins and average points scored. The data points form a pattern where higher average points scored correspond to more total wins, it suggests a positive association. The Pearson correlation coefficient, which measures the strength and direction of this relationship, further solidifies this understanding. A correlation coefficient of 0.4777 indicates a moderate positive relationship between the total number of wins and average points scored in the dataset. This value suggests that as the average points scored increases, there's a tendency for the total number of wins to also increase, but not as strongly as in a relationship with a coefficient closer to 1.

At a 1% level of significance, a p-value of 0.0 signifies that the observed correlation between total wins and average points scored is statistically significant. This suggests that the likelihood of this relationship occurring by random chance is extremely low, reinforcing the idea of a strong and meaningful association between scoring more points on average and achieving more wins in a regular basketball season.

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

A multiple linear regression model assesses the relationship between a dependent variable (in this case, total number of wins) and multiple independent variables (predictors such as average points scored and average relative skill). It determines the impact of each predictor on the response variable while considering the effects of other predictors.

The equation for the model can be expressed as:

Total Wins = -152.5736 + 0.3497 \* Avg\_Points + 0.1055 \* Avg\_Elo\_N

Null Hypothesis (Ho):

1. The coefficients for the predictors are zero
2. (Ho): β1 = β2 = 0

Alternative Hypothesis (Ha):

1. At least one of the coefficients for the predictors are not zero
2. (Ha): β1 or β2 ≠ 0

Level of Significance: 0.05

Table 2: Hypothesis Test for the Overall F-Test

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | 1580.00 |
| P-value | 0.0000 |

Based on the overall F-test with a p-value of 0.0000, it's evident that at least one of the predictors, either the average points scored or the average relative skill, significantly influences the total number of wins in a basketball season. Individually, the t-tests for each predictor variable's parameter indicate their statistical significance. With a 1% level of significance, both average points scored and average relative skill are statistically significant based on their respective p-values in the model.

The coefficient of determination (R-squared) at 0.837 suggests that roughly 83.7% of the variability in total wins can be explained by the included predictors, emphasizing the strong predictive capacity of the model.

For a team averaging 75 points with a relative skill level of 1350:

Total Wins = -152.5736 + 0.3497 \* 75 + 0.1055 \* 1350

Total Wins = -152.5736 + 26.2275 + 142.275

Total Wins = 15.929 ≈16

For a team averaging 100 points with a relative skill level of 1600:

Total Wins = -152.5736 + 0.3497 \* 100 + 0.1055 \* 1600

Total Wins = -152.5736 + 34.97 + 168.8

Total Wins = 51.196≈ 51

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

A multiple linear regression model is utilized to predict a response variable based on multiple predictor variables. It aims to establish a relationship between the response variable and several predictors simultaneously, allowing us to estimate how changes in the predictors correspond to changes in the response. The model assumes a linear relationship between the response and each predictor variable, collectively capturing their combined influence on the response.

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):

1. None of the included predictor variables have a significant impact on the total number of wins.
2. H₀: β₁ = β₂ = β₃ = β₄ = 0

Alternative Hypothesis (H₁):

1. At least one of the included predictor variables has a significant impact on the total number of wins.
2. H₁: At least one β ≠ 0

Level of Significance: 0.01

Table 3: Hypothesis Test for Overall F-Test

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | 1102.00 |
| P-value | 0.0000 |

The F-statistic here is high, indicating that the model's overall fit is statistically significant. The associated P-value is close to zero (3.07e-278), indicating a very low likelihood that this F-statistic occurred by chance. Therefore, at least one of the predictors is statistically significant in predicting the number of wins in the season.

In the presented model, avg\_pts and avg\_pts\_differential appear to be statistically significant with p-values close to zero (0.000). Avg\_elo\_n and avg\_elo\_differential exhibit p-values above the 1% threshold (0.442 and 0.004, respectively), indicating they might not be statistically significant at this level.

The coefficient of determination for this model is approximately 0.878, implying that around 87.8% of the variability in the total number of wins is explained by the predictors included in the model.

Team averaging 75 points per game, with a relative skill level of 1350, an average point differential of -5, and an average relative skill differential of -30:

Total Wins = 34.5753 + (0.2597 \* 75) − (0.0134 \* 1350) + (1.6206 \* (−5)) + (0.0525 \* (−30))

Total Wins = 34.5753 + 19.4755 − 18.09 − 8.103 − 1.575

Total Wins = 26.2728 ≈ 26

Team averaging 100 points per game, with a relative skill level of 1600, an average point differential of +5, and an average relative skill differential of +95:

Total Wins = 34.5753 + (0.2597 \* 100) − (0.0134 \* 1600) + (1.6206 \* 5) + (0.0525 \* 95)

Total Wins = 34.5753 + 25.97 − 21.6 + 8.103 + 4.9875

Total Wins = 52.0368 ≈ 52

## 8. Conclusion

The statistical analyses conducted on the basketball team's historical data from 1995 to 2015 revealed insights into the relationship between various performance metrics and the total number of wins in a regular season. The analyses used correlation, simple linear regression, and multiple regression to understand how factors like average points scored, relative skill, points differentials, and skill differentials impact a team's success.

The correlation analyses indicated strong positive relationships between total wins and average relative skill, as well as a positive association between wins and average points scored. These findings suggest that maintaining higher skill levels and scoring more points on average positively correlates with a team's success in winning games during a season.

The regression models provided predictive equations to estimate the number of wins based on different performance metrics. The simple linear regression model demonstrated that for every increase in average relative skill, there was an increase in predicted wins, emphasizing its influence on the total wins. The multiple regression models expanded on this by considering multiple factors simultaneously. They confirmed that both average points scored, and average relative skill significantly contribute to predicting total wins. However, the inclusion of additional metrics like points differentials and skill differentials in the more complex regression model highlighted their mixed influence, with some variables showing stronger significance than others in predicting wins.

Practically, these analyses offer actionable insights for the team's management and coach. They provide a predictive model that can forecast the team's potential wins based on specific performance metrics. This understanding enables strategic decisions to enhance the team's performance, such as focusing on improving skill levels, scoring more points, and managing point differentials effectively against opponents.