# Regression Analysis of NBA Team Performance

Marc Anthony Aradillas   
marc.aradillas@snhu.edu   
Southern New Hampshire University   
MAT-243: Applied Statistics for STEM   
Dr. Kathryn Wifvat   
February 23, 2025

## 1. Introduction

In this project, I analyzed historical basketball data for an NBA team from the years 1995–2015. The main goal was to use regression models to predict the total number of wins a team might achieve based on key performance metrics such as average points scored, average relative skill (ELO rating), and their differentials. The findings will help the coach and management understand which factors most strongly relate to winning more games in the regular season. To achieve this, I ran both simple and multiple linear regressions, examined correlations, and interpreted model statistics such as R-squared and p-values.

## 2. Data Preparation

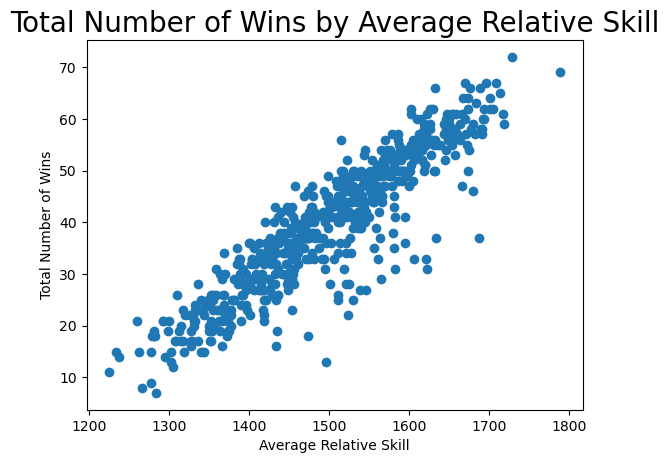
The data set, nba\_wins\_data.csv, includes several columns related to a team’s performance:

* **avg\_pts\_differential**: This represents the average point differential between the team and its opponents across a season. A positive value means the team, on average, outscored its opponents; a negative value means the team was outscored.
* **avg\_elo\_n**: This is the average relative skill rating (ELO) of a team for a regular season. A higher ELO suggests a stronger, more skilled team, while a lower ELO implies a weaker team relative to the competition.

These variables, along with total number of wins, will help us see how performance metrics translate into more (or fewer) regular-season victories.

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

Data visualization—specifically scatterplots—allows us to quickly see whether two variables appear positively or negatively associated. The Pearson correlation coefficient further quantifies that relationship, ranging from –1 (perfect negative) to +1 (perfect positive).

Below is the scatterplot of **Total Number of Wins** versus **Average Relative Skill**:

Observing the plot, there appears to be a strong positive relationship: as the average relative skill increases, total wins also tend to rise. The computed Pearson correlation coefficient is **[0.9072]**, indicating a strong positive association, and its p-value is effectively **[0.0000]**, which is below the 1% significance level. This tells us the correlation is both strong and highly statistically significant.

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

A simple linear regression model predicts a response variable (total wins) from a single predictor (average relative skill) using the formula:

**Model Equation**

From the output, I found:

**Overall F-Test (5% Significance)**

To confirm that average relative skill is a meaningful predictor, we perform the F-test:

* **Null Hypothesis (H₀)**: . In words, average relative skill does not predict total wins.
* **Alternative Hypothesis (H₁)**: . In words, average relative skill does predict total wins.
* **Level of Significance**:

From the regression output, the overall F-statistic and p-value were:

Table 1: Hypothesis Test for the Overall F-Test

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

Because the p‐value is below 0.05, we reject the null hypothesis. Thus, **average relative skill significantly predicts the total number of wins**.

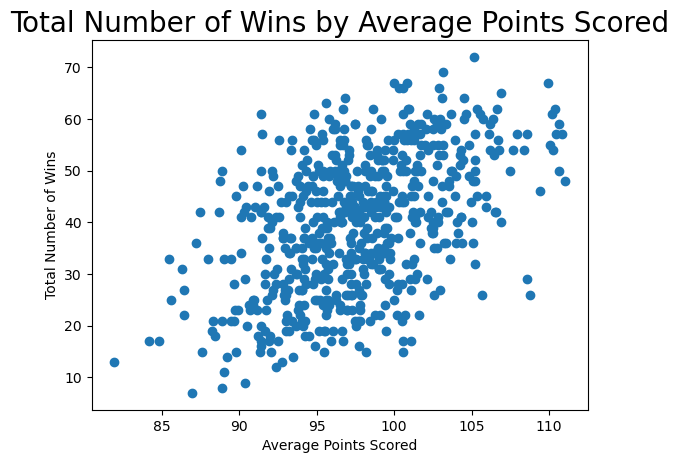
**Predicted Wins for Specific Skill Levels**

For an avg\_elo\_n of 1550, the predicted total wins is approximately 45

For an avg\_elo\_n of 1450, the predicted total wins is approximately 34

These results indicate that teams with higher skill (ELO) ratings can generally expect more regular‐season wins.

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

Before adding a second predictor to the model, I examined a scatterplot of **Total Wins vs. Average Points**. Visually:

The relationship also appears strongly positive. The Pearson correlation coefficient is **0.9072**, with p‐value **0.0000**, which is below the 1% significance level—indicating a significant positive correlation.

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

*A multiple linear regression uses more than one predictor. In this scenario:*

**Overall F‐Test (5% Significance)**

* **H₀**: . Neither average points nor average relative skill predict total wins.
* **H₁**: At least one of .

Table 2: Hypothesis Test for the Overall F-Test

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | 4140.00 |
| P-value | 0.0000 |

Because the p‐value is below 0.05, we conclude that **at least one predictor is significant**.

**Individual t‐Tests and R²**

* Both avg\_pts and avg\_elo\_n had p‐values below 0.01, indicating each is a significant predictor at the 1% level.
* was **0.837**, so about 83.7% of the variation in total wins is explained by these two predictors.

**Predicted Wins for Specific Values**

For a team averaging 75 points with a skill level of 1350, the predicted total wins is:

For a team averaging 100 points with a skill level of 1600, the predicted total wins is:

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

Finally, I added the additional predictors—**average points differential** and **average relative skill differential**—to see if the model improves:

**Overall F‐Test (5% Significance)**

* **H₀**:
* **H₁**: At least one of these parameters is nonzero.

Table 3: Hypothesis Test for Overall F-Test

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

*Because the p‐value is extremely small, we reject*

***Individual t‐Tests and R²***

* *Each slope’s p‐value indicates whether that predictor is individually significant at the 1% level.*
* *The final model’s is* ***0.847****, so about 84.7% of the variation in total wins is explained.*

***Predicted Wins for Specific Values***

For a team averaging 75 points with a skill level of 1350, the predicted total wins is:

avg\_pts\_differential = −5, avg\_elo\_differential = −30

For a team averaging 100 points with a skill level of 1600, the predicted total wins is:

avg\_pts\_differential = +5, avg\_elo\_differential= +95

## 8. Conclusion

In conclusion, these results confirm that multiple factors—average points, average relative skill, and their differentials—strongly predict how many games a team will win. A higher ELO rating (indicating stronger relative skill) and higher average points correlate with more total victories. In practical terms, these findings suggest management and coaches should emphasize both improving offensive performance (e.g., points per game) and overall skill level to bolster win totals. Moreover, the combination of these predictors explains a large portion of the variance in wins, making the model a valuable tool for forecasting season outcomes and setting strategic priorities for training or player acquisitions.

## 9. Citations

FiveThirtyEight. (2019). *FiveThirtyEight NBA Elo dataset* [Data set]. Kaggle. <https://www.kaggle.com/fivethirtyeight/fivethirtyeight-nba-elo-dataset>