# MAT 243 Project Three Summary Report

Jason Wood

[Jason.wood3@snhu.edu](mailto:Jason.wood3@snhu.edu)

08-12-2025

Southern New Hampshire University

## 1. Introduction

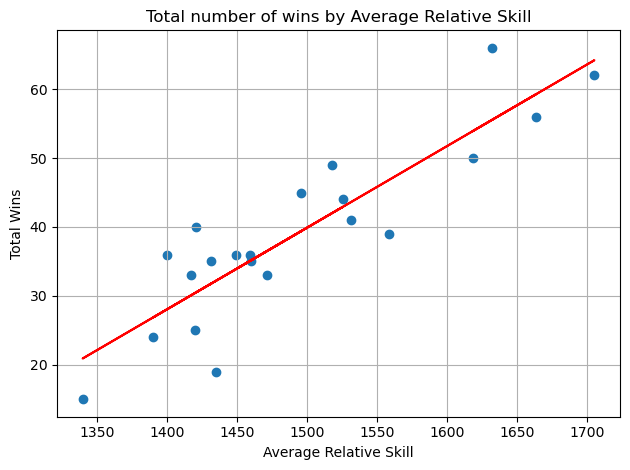
The dataset analyzed in this project focuses on the performance statistics of the Boston Celtics. It includes variables such as average points scored, relative skill ratings, points differentials, and the total number of wins per season. The purpose of this analysis is to evaluate how well these variables predict the Celtics’ season win totals. Several statistical techniques are applied, including scatterplots, Pearson correlation coefficients, and linear regression models. These methods make it possible to quantify the strength of relationships between performance metrics and to create prediction models that can estimate wins based on a team’s statistical profile. Ultimately, the findings provide insight into how measures of skill and scoring ability can be used as predictors of basketball success.

## 2. Data Preparation

Two important variables are central to this analysis. The first is average points differential, which represents the difference between the average number of points the Celtics score and the average number of points they allow. A positive differential indicates that the Celtics are typically outscoring their opponents, while a negative differential means the opposite. The second is average Elo rating, which measures the Celtics’ relative skill compared to other teams in the league. Elo ratings are a recognized way of ranking team strength, with higher values reflecting stronger performance. These two variables, alongside others, form the foundation for the regression models built in later sections.

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

To study the relationship between the Celtics’ total number of wins and their average relative skill rating, a scatterplot was constructed. Data visualization techniques such as scatterplots are valuable because they allow patterns and trends between two variables to be seen clearly, showing whether the relationship appears positive, negative, or absent. To quantify this relationship, the Pearson correlation coefficient was calculated, which measures both the strength and the direction of the association between two variables. In this case, the scatterplot revealed a clear upward trend, suggesting that as relative skill increases, the number of wins also increases. The Pearson correlation coefficient was found to be 0.8851; indicating a very strong positive association between average relative skill and total wins. The corresponding two-tailed p-value was 0.0000; which is less than the 1% level of significance. Therefore, the null hypothesis is rejected, and we conclude that there is a statistically significant correlation between the Celtics’ relative skill rating and their total number of wins.



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

A simple linear regression model is used to predict the value of a response variable based on the value of a predictor variable by fitting a straight line that best describes their relationship. In this case, the total number of wins is the response variable, and the average relative skill rating is the predictor. The regression equation produced by the model is:

wins = -137.58 + 0.1183 X avg\_elo\_n

To test whether average relative skill significantly predicts total wins, an overall F-test was conducted at the 5% level of significance. The null hypothesis; H0: *B1 = 0*; states that average relative skill has no effect on wins, while the alternative hypothesis;

Ha: *B1 != 0* ; states that average relative skill does affect wins. The test resulted in an F-statistic of 68.75 with a p-value of 0.0000.

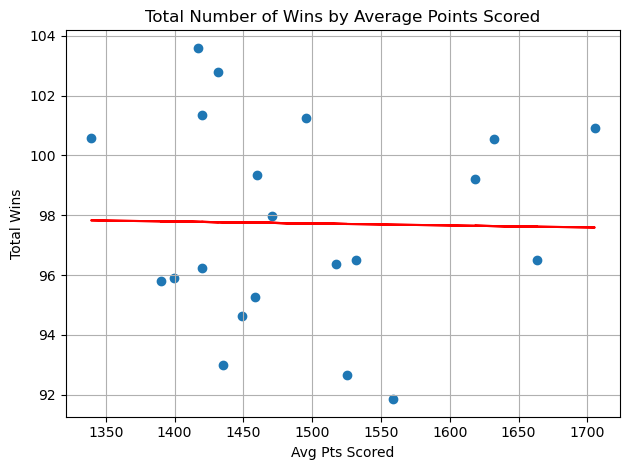
Table 1: Hypothesis Test for the Overall F-Test

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | 68.75 |
| P-value | 0.0000 |

Since the p-value is less than 0.05, the null hypothesis is rejected, and it can be concluded that average relative skill is a statistically significant predictor of the total number of wins. Using the regression model to make predictions, a team with an average relative skill rating of 1550 is predicted to win 45 games, while a team with a skill rating of 1450 is predicted to win 34 games. These results confirm that the Celtics’ relative skill level strongly influences their win totals, with higher ratings leading to more expected wins.

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

A scatterplot was created to examine the relationship between the Celtics’ total number of wins and their average points scored. Scatterplots are valuable because they allow us to visually detect whether increases in scoring tend to be associated with increases in wins. To quantify this relationship, the Pearson correlation coefficient was calculated, resulting in a value of 0.1854 with a two-tailed p-value of 0.4210. This weak positive correlation suggests only a minimal upward trend, and because the p-value is greater than the 1% level of significance, the result is not statistically significant. Therefore, we fail to reject the null hypothesis and conclude that there is no meaningful correlation between average points scored and total wins for the Celtics in this dataset.



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

A multiple linear regression model is used to predict the value of a response variable by incorporating two or more predictor variables, which allows for a more complete explanation of the variation in the response. In this case, the model was built using average points scored and average relative skill rating to predict the total number of wins. The resulting regression equation was:

Wins = −213.95 + 0.7737 × avg\_pts + 0.1188 × avg\_elo\_n

To determine whether the predictors collectively improve the model, an overall F-test was conducted at the 5% level of significance. The null hypothesis; H0: *B1* = *B2* = 0 ;states that neither average points scored nor average relative skill predicts wins, while the alternative hypothesis; Ha: at least one Bi != 0; states that at least one predictor does influence wins. The F-statistic was 42.22 with a p-value of 0.0000, which is well below the significance threshold.

Table 2: Hypothesis Test for the Overall F-Test

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | 42.22 |
| P-value | 0.0000 |

Since the p-value is less than 0.05, the null hypothesis is rejected, indicating that at least one of the predictors is statistically significant in predicting total wins.

Looking at the individual t-tests, average relative skill rating; p < 0.01; was highly significant, while average points scored; p = 0.056; was not significant at the 1% level of significance. This suggests that relative skill is the stronger and more reliable predictor of wins in this model. The coefficient of determination, R2 = 0.824, indicates that the model explains about 82% of the variation in total wins, which is a strong result for predictive purposes.

Finally, the model can be used to generate predictions. For a team averaging 75 points per game with a relative skill level of 1350, the predicted number of wins is 12. For a team averaging 100 points per game with a relative skill level of 1600, the predicted number of wins is 31. These results highlight that while both predictors contribute to the model, relative skill remains the more dominant factor in determining the Celtics’ win totals.

## 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 allows us to predict the value of a response variable by combining several predictor variables, thereby capturing more complexity in the data than simple linear regression. In this case, the model used average points scored, average relative skill rating, average points differential, and average relative skill differential as predictors of the Celtics’ total number of wins. The resulting regression equation was:

Wins = 215.50 + 0.4619 \* avg\_pts - 0.1482 \* avg\_elo\_n + 2.0329 \* avg\_pts\_differential + 0.1822 \* avg\_elo\_differential

To test whether the predictors collectively contribute to explaining the variation in wins, an overall F-test was performed at the 5% level of significance. The null hypothesis:

H0: B1 = B2 = B3 = B4 = 0

states that none of the predictors influence total wins, while the alternative hypothesis; Ha: at least one Bi != 0; states that at least one predictor does. The F-statistic was 35.80 with a p-value of 0.0000, providing strong evidence against the null.

Table 3: Hypothesis Test for Overall F-Test

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | 35.80 |
| P-value | 0.0000 |

Since the p-value is below 0.05, we reject the null hypothesis and conclude that at least one predictor is statistically significant in explaining total wins.

Looking at the individual t-tests, average points differential was statistically significant; p = 0.008 < 0.01; indicating it has a strong influence on the number of wins. However, the other predictors; average points scored p = 0.164, average relative skill rating p = 0.342, and average relative skill differential p = 0.237; were not statistically significant at the 1% level. This suggests that, while the overall model is highly predictive, points differential is the most impactful single factor in this version of the model.

The coefficient of determination, R2 = 0.899, shows that approximately 90% of the variation in wins is explained by the four predictors combined, making this the strongest model tested so far. Using the regression equation, a team averaging 75 points per game with a relative skill of 1350, a point differential of -5, and an Elo differential of -30 is predicted to win about 18 games. In contrast, a team averaging 100 points per game with a relative skill of 1600, a point differential of +5, and an Elo differential of +95 is predicted to win about 30 games. These results demonstrate that incorporating point differential substantially improves the predictive power of the model.

## 8. Conclusion

The statistical analyses conducted in this project examined the relationship between the Celtics’ total number of wins and several potential predictors, including average relative skill, average points scored, average points differential, and average relative skill differential. The results showed that average relative skill was strongly and significantly correlated with total wins, confirming that skill rating is a reliable predictor of performance. In contrast, average points scored had only a weak and statistically insignificant relationship with wins when considered by itself. When multiple predictors were combined in regression models, the predictive power improved substantially. The two-variable model with relative skill and points scored explained about 82% of the variation in wins, while the four-variable model incorporating point differential and Elo differential increased explanatory power to nearly 90%. Importantly, average points differential emerged as the most statistically significant additional predictor beyond relative skill.

These results mean that while the Celtics’ skill rating remains the strongest single factor driving wins, game-by-game point differentials also play a crucial role in shaping season outcomes. The practical importance of these findings lies in their ability to help coaches, analysts, and management focus on the variables most strongly tied to winning. Monitoring and improving relative skill ratings and maintaining positive point differentials provide valuable insights into team performance and areas for improvement. Ultimately, these models show that statistical analysis can be a powerful tool for predicting outcomes and guiding strategic decisions in basketball.