# MAT 243 Project Three Summary Report

Brandon Taylor

brandon.taylor6@snhu.edu

Southern New Hampshire University

**Notes:**

* Replace the bracketed text on page one (the cover page) with your personal information.
* You will use your selected team for all three projects

## 1. Introduction

*Basically, a multiple linear regression model helps us figure out how a bunch of different factors (our "predictor variables") work together to influence something we're trying to predict (our "response variable"). It tries to draw a straight line (or a plane, or a hyperplane in higher dimensions, but let's stick to the idea of a line for simplicity) through our data points to show the relationship. Each predictor variable gets its own "weight" or coefficient, which tells us how much that factor changes our predicted outcome when everything else stays the same. So, if we know the values of these different factors, the model gives us a pretty good guess at what the response variable will be.*

## 2. Data Preparation

*In this project, several key variables are utilized to analyze basketball team performance. The avg\_pts\_differential variable represents the average difference between the points a team scores and the points their opponents score in a regular season. To put it simply, it tells us, on average, how much a team wins or loses by each game; a positive value means they typically outscore their opponents, while a negative value means they're usually outscored. The avg\_elo\_n variable, on the other hand, stands for a team's average relative skill level throughout a regular season. This "relative skill" is like a ranking system that measures how good a team is compared to others, taking into account game scores, location, and outcomes. A higher avg\_elo\_n number indicates a stronger, more skilled team.*

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

*Data visualization techniques, such as scatterplots, are incredibly useful tools for initially exploring the relationship between two variables. They allow us to visually inspect if there's a trend, pattern, or any outliers in the data, giving us an intuitive sense of how variables might be related. For instance, we can see if as one variable increases, the other tends to increase (positive relationship), decrease (negative relationship), or show no clear pattern at all. The Pearson correlation coefficient, often denoted as 'r', quantifies this observed relationship by providing both its strength and direction. The strength is indicated by the absolute value of 'r' (closer to 1 means a stronger relationship, closer to 0 means weaker), while the direction is shown by the sign (+ for positive correlation, - for negative correlation).*

*In this activity, a scatterplot was generated to visualize the association between the total number of wins and a team's average relative skill.*

A graph of blue dots

AI-generated content may be incorrect.

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

*For the overall F-test, we're basically asking: "Are any of these predictor variables actually useful for predicting total wins, or are they all just random noise?"*

* ***Null Hypothesis (H0​):*** *β1​=β2​=β3​=β4​=0. This means that none of the predictor variables (average points scored, average relative skill, average points differential, and average relative skill differential) are helpful in predicting the total number of wins in a regular season.*
* ***Alternative Hypothesis (Ha​):*** *At least one βi​=0. This means that at least one of our predictor variables is useful for predicting the total number of wins.*
* ***Level of Significance (α):*** *We're using 0.05 (or 5%) for this test, which is a common threshold.*

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

Since our P-value (0.0000) is super tiny – way smaller than our α of 0.05 – we can confidently reject the null hypothesis. This tells us that our model isn't just guessing; at least one of these factors affects the outcome.

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

*For this section, I made another scatterplot, this time looking at the relationship between the total number of wins and the average points a team scored. This plot helps us visualize if there's any kind of pattern between scoring more points and winning more games.*

*The scatterplot for Total Number of Wins by Average Points Scored is included right here:*

*A diagram of a number of points

AI-generated content may be incorrect.*

*When I looked at the scatterplot, it seemed like there was a positive, but maybe not super strong, connection between how many points a team scored and how many games they won. As teams scored more, they generally won more, but it wasn't as tight a relationship as with the relative skill. To get a clearer picture, I calculated the Pearson correlation coefficient, which came out to be about 0.4777. This positive number confirms that there's a moderate positive linear relationship – so, more points generally mean more wins. To be sure this wasn't just a fluke, I checked the P-value, which was 0.0. Since this P-value is much, much smaller than our 1% significance level (α=0.01), we can say that this correlation is statistically significant. That means it's a real relationship, not just random chance!*

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

*A multiple linear regression model is a statistical method employed to predict the value of a dependent (response) variable by utilizing the values of two or more independent (predictor) variables. This model establishes a linear relationship, allowing for the estimation of how each predictor influences the response, while simultaneously accounting for the effects of other predictors. This capability facilitates comprehensive prediction and a deeper understanding of how multiple factors collectively impact an outcome.*

*The equation for this multiple regression model (model2) is derived from the estimated coefficients provided in the output: Total Wins=−152.5736+(0.3497×avg\_pts)+(0.1055×avg\_elo\_n)*

*To assess the overall significance of this model, an F-test is conducted. The hypothesis test steps are as follows:*

* ***Null Hypothesis (H0​):*** *β1​=β2​=0. This hypothesis posits that neither average points scored nor average relative skill has a statistically significant linear relationship with the total number of wins in the regular season.*
* ***Alternative Hypothesis (Ha​):*** *At least one βi​=0. This hypothesis suggests that at least one of the predictor variables (average points scored or average relative skill) possesses a statistically significant linear relationship with the total number of wins.*
* ***Level of Significance (α):*** *0.05 (or 5%).*

*Table 2: Hypothesis Test for the Overall F-Test*

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

*With a P-value of 0.0000, which is considerably less than the chosen level of significance (α=0.05), the null hypothesis is rejected. This indicates strong statistical evidence to conclude that at least one of the predictor variables (average points scored or average relative skill) significantly contributes to predicting the total number of wins in a regular season. Therefore, based on the results of the overall F-test, at least one of the predictors is statistically significant in predicting the total number of wins in the season.*

*Moving to the individual t-tests for each predictor variable, we evaluate their statistical significance using a 1% level of significance (α=0.01). For avg\_pts (Average Points Scored), the P-value is 0.000. Since 0.000 is less than 0.01, avg\_pts is statistically significant. Similarly, for avg\_elo\_n (Average Relative Skill), the P-value is also 0.000. As 0.000 is less than 0.01, avg\_elo\_n is likewise statistically significant. Consequently, both avg\_pts and avg\_elo\_n are statistically significant predictors of the total number of wins in this model at the 1% level of significance.*

*The coefficient of determination, or R-squared, for this model is 0.837. This implies that approximately 83.7% of the variability observed in the total number of wins can be explained by this multiple linear regression model, which incorporates average points scored and average relative skill as predictor variables. This high R-squared value suggests that the model accounts for a substantial proportion of the variance in wins.*

*Using the model equation, predictions can be made for various scenarios. For a team averaging 75 points per game with a relative skill level of 1350, the predicted total number of wins is calculated as: Predicted Wins=−152.5736+(0.3497×75)+(0.1055×1350)=16.0789. Rounding down, the predicted total number of wins is 16.*

*For a team averaging 100 points per game with an average relative skill level of 1600, the predicted total number of wins is: Predicted Wins=−152.5736+(0.3497×100)+(0.1055×1600)=51.1964. Rounding down, the predicted total number of wins is 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 a powerful tool that helps us predict an outcome (like total wins) by looking at how several different factors (our predictor variables) influence it all at once. This model creates a formula that shows the combined effect of these factors on the response variable. For this specific model, we're predicting Total Wins using average points scored, average relative skill, average points differential, and average relative skill differential.*

*The equation we found is: Total Wins=34.5753+(0.2597×avg\_pts)+(−0.0134×avg\_elo\_n)+(1.6206×avg\_pts\_differential)+(0.0525×avg\_elo\_differential).*

*To start, we perform an overall F-test to see if the entire model is statistically significant in predicting total wins. Our null hypothesis is that none of the predictors are significant (β1​=β2​=β3​=β4​=0), and our alternative hypothesis is that at least one of them is significant. Using a 5% level of significance, our F-test yielded a statistic of 1102.00 with a P-value of 0.0000. Since this P-value is well below 0.05, we reject the null hypothesis, meaning that at least one of our predictors is indeed statistically significant in predicting total wins.*

*Delving into the individual t-tests (using a 1% level of significance), we found that average points scored (P-value = 0.000), average points differential (P-value = 0.000), and average relative skill differential (P-value = 0.004) are all statistically significant predictors. However, average relative skill (P-value = 0.442) was not statistically significant in this model, suggesting its individual contribution might be less pronounced when other factors are considered.*

*The model's R-squared value is 0.878, which is quite high. This means that about 87.8% of the variability in the total number of wins can be explained by the four predictor variables in our model, indicating a strong fit to the data.*

*Finally, we can use our equation for predictions: for a team averaging 75 points per game with a relative skill of 1350, an average point differential of -5, and an average relative skill differential of -30, the model predicts about 26 total wins. For a team averaging 100 points per game with a relative skill of 1600, an average point differential of +5, and an average relative skill differential of +95, the model predicts around 52 total wins.*

***8. Conclusion***

*Overall, our statistical analyses consistently demonstrate a strong relationship between various team performance metrics and the total number of wins in a regular basketball season. Both simple and multiple linear regression models proved highly effective in predicting wins, with the final multiple regression model (model3) being particularly robust, explaining approximately 87.8% of the variability in total wins (as indicated by its R-squared value). This means that factors like average points scored, average points differential, and average relative skill differential are very strong indicators of a team's success, as confirmed by their statistically significant P-values in the individual t-tests. While average relative skill (avg\_elo\_n) was a significant predictor on its own, its individual significance decreased when combined with other strong predictors in the more complex model, suggesting that its influence might be captured by or correlated with the other variables. In plain language, this project shows that if a team scores more points, has a better point differential (scoring more than opponents), and maintains a higher skill advantage over opponents, they are much more likely to win more games. The practical importance of these analyses for the coach and management is significant: these models provide data-driven insights into which performance areas most directly impact wins. By focusing on improving metrics like point differential and overall scoring, the team can set clear, quantifiable goals to enhance their competitive performance and increase their chances of winning more games in future seasons.*

## 9. Citations

*N/A*