# MAT 243 Project Three Summary Report

Cadden Buist

caddenbuist@snhu.edu

Southern New Hampshire University

## 1. Introduction

We are trying to predict our total wins during the regular season based on key performance metrics using regression models. These key metrics are total wins, average points, average relative skill, average point differential and average relative skill differential. These results will help not only predict our win total at the end of the season, but we will also use this information to help make key decisions. We will be using simple linear regression and multiple regression to analyze our data in this project.

## 2. Data Preparation

Before diving too deep into this report lets go over some of the terms that may be less common terminology outside of this field. First, we have avg\_pts\_differential or average point differential which is the average points difference between us, and our opponents based off the season stats. Now let’s get to avg\_elo\_n which stands for average relative skill level of each team during the regular season. These terms are going to be important towards the end of this report which you’ll find out at the end.

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

Data visualization techniques make it easier for stockholders to understand and for analyst to be able to study the data much easier and to understand it much faster. The correlation coefficient is used to determine strength and direction for a scatterplot like in step 2 we see the scatterplot has a correlation coefficient is a strong positive one. A picture containing screenshot

Description automatically generated

Based on the scatterplot and the Pearson correlation coefficient is a very strong and it shows a strong linear relationship between our 2 variables total wins and average relative skill. The correlation coefficient is not significant as we can say that 0.00>0.01 is false.

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

*You created a simple linear regression model for the total number of wins in a regular season using the average relative skill as the predictor variable.*

*See Step 3 in the Python script to address the following items:*

A simple linear regression model uses the predictor variable to predict the response variable by finding the best-fitting straight line through the data points. The equation for the linear regression model is Y=β0+β1X where β0 is our y intercept, β1 is our is the measurement unit and x is our predictor variable. Our null hypothesis is 0 also known to look like H0=β1=β2=0 while the alternative hypothesis is anything but 0 or Ha: at least one Bi≠0 for i=1. Our significance level is 5% which would make our alpha level 0.05 or α=0.05.

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| Test Statistic | 53.52 |
| P-value | 0.0000 |

To start things off here our null hypothesis is rejected because our p-value is 0.0000 which makes 0.0000>0.05 false. Based off the F-test there is a significant linear relationship that exists because the null hypothesis was rejected which means the average relative skill can predict the number of wins in the regular season. If our team had a relative skill level of 1550, we would have 45 wins during the regular season using the equation Y=-128.2475+1550(0.1121). If our team had a relative skill level of 1450, we would win 34 games losing 11 more games than if we had a relative skill level of 1550 with the equation being Y=-128.2475+1450(0.1121).

**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 with low confidence

The scatterplot and the Pearson correlation coefficient tell me that this scatterplot is not very strong it’s a moderate correlation in a positive direction. It has a moderate linear relationship between the 2 variables average points scored and total number of wins. With the p-value being 0.0 making 0.0>0.01 false the coefficient correlation is not statistically significant.

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

*You created a multiple regression model with the total number of wins as the response variable, with average points scored and average relative skill as predictor variables.*

*See Step 5 in the Python script to answer the following questions:*

A multiple linear regression model uses multiple predictor variables to predict response variable by adding a slope coefficient to every predictor variable in our equation. The equation for our model is going to be Y=β0+β1X1+β2X2 with B0 being the y-intercept, B1 being the first predictor variable, B2 being the second predictor variable and X1/X2 being the slopes for their respected variables. Our null hypothesis is going to be 0 or H0=β1=β2=0 and our alternative hypothesis is anything but 0 or Ha: at least one Bi≠0 for i=1. The level of significance is 5% or our alpha level equals 0.05 which would look like α=0.05.

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| Test Statistic | 1580 |
| P-value | 4.41e-243 |

We are going to start this conclusion by stating that the null hypothesis was rejected with the p-value being less than the alpha value 4.41e-243<0.05. Looking at the p-value for the individual predictor variables are both at 0.000 which means they are both less than the alpha level making them both statistically significant. The coefficient of determination is 0.837 is the percentage of the response variable explained by the predictor variable which would be the accuracy which is 83.7 percent. If we scored an average of 75 points per game and had an average relative skill level of 1350 our equation would look like Y=-152.5736+75(0.3497)+1350(0.1055) which tells us we would only win 16 games all season long. Finally, if we had an average of 100 points per game and an average of 1600 relative skill level we would get an equation that is Y=-152.5736+100(0.3497)+1600(0.1055) which would give us a win total of 51 games which would give us a high spot in the playoffs.

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

*You created a multiple regression model with the total number of wins as the response variable, with average points scored, average relative skill, average points differential, and average relative skill differential as predictor variables.*

*See Step 6 in the Python script to answer the following questions:*

A multiple linear regression model uses multiple predictor variables to predict response variable by adding a slope coefficient to every predictor variable in our equation. The equation for our model is going to be Y=β0+β1X1+β2X2+β3X3+β4X4 with B0 being the y-intercept, B1 being the first predictor variable, B2 being the second predictor variable, B3 and B4 being the 3rd and 4th predictor variable and with X1, X2, X3, X4 being the slopes for their respected variables. Our null hypothesis is going to be 0 or H0=β1=β2=0 and our alternative hypothesis is anything but 0 or Ha: at least one Bi≠0 for i=1. The level of significance is 5% or our alpha level equals 0.05 which would look like α=0.05.

|  |  |
| --- | --- |
| **Statistic** | **Value** |
| Test Statistic | 1102 |
| P-value | 3.07e-278 |

Looking at our Null hypothesis for the group we can reject it because the p-value 3.07e-278 is less than the alpha level 0.05 which makes 3.07e-278>0.05 false. We can see from the individual p-values the only predictor variable that is not statistically significant which is the average relative skill level which fails to reject the null hypothesis because 0.442>0.01 is true. Our coefficient of determination is 0.878 or 87.8% making the response variable to predictor variable very accurate. If we had 75 points per game, a relative skill level of 1350, a point differential -5 and a relative skill level differential of -30 we get the equation of Y=34.5753+75(0.344)+1350(-0.0134)+-5(1.6206)+-30(0.0525) which means we would have a win total of 32 games in the regular season. If we had 100 points per game, a relative skill level of 1600, a point differential +5 and a relative skill level differential of +95 we get the equation of Y=34.5753+100(0.344)+1600(-0.0134)+5 (1.6206)+95(0.0525) which means we would have a win total of 60 games in the regular season.

## 

## 8. Conclusion

From my findings if a team were to have a high points per game average and a high relative skill level always ends with a large number of wins. When only one is large and the other is smaller there is a 50/50 shot the team gets many wins with the same chance of the team having many losses. Having a positive differential with points and skill will give any team a better chance of getting wins during the regular season. The importance of this analysis lets us know what we need to work on with the players or if we need to go out and get more star power for this team. If a relative skill is high than we need to work on shooting baskets with the team getting that muscle memory with shots the team is missing the most. If our points per game is high but we have a lower skill level us as the management need to go out and get a free agent or trade for a player with a high relative skill level to help the team win.