# MAT 243 Project Three Summary Report

Jared Baker

Jared.baker@snhu.edu

Southern New Hampshire University

## 1. Introduction

The data set that I am exploring includes the total wins, average points, average relative skill, average point differential, and average relative skill differential of NBA teams from the years 1995-2015. The results of the tests will be used to compare the Bulls and the Nuggets during this time period. I will pe performing simple linear regression models and multiple regression models to make predictions about the teams.

## 2. Data Preparation

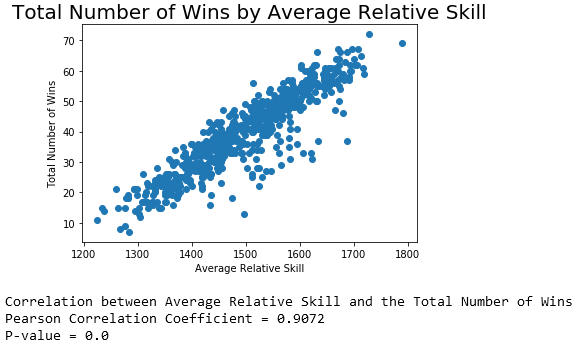
The avg\_pts\_differential variable represents the average difference between two team’s average points scored per game. A higher value shows a greater difference in average points scored.

The avg\_elo\_n variable represents the average relative skill level for a team. A higher value means a team is more skilled. A team with a higher value than another team is statistically better than the other team.

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

Data visualization techniques are used to visibly represent response and predictor varaibles in a way that clearly shows correlations and trends. Plots easily show how different variables are alike and different.

The correlation coefficient represents the strength and direction of two variable’s relationship by sign and value. A value above .8 (80%) indicates a strong correlation for example. If that value is negative (-0.8), then that coefficient represents a strong negative correlation. The closer to 0, the weaker the trend.



This plot and coefficient (0.9072) show a strong positive correlation between number of wins and average relative skill.

The correlation is statistically significant as shown by the P-value of 0 being less than 0.01 (1% level of significance).

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

Generally, a simple linear regression model shows the relationship between two variables. In this case it compares total wins and average relative skill.

The equation for my model is: total\_wins = -128.2475 + 0.1121 \* avg\_elo\_n.

The null hypothesis for this test is that teams that maintain a higher relative skill will not have more wins ().

The alternative hypothesis is that teams that maintain a higher relative skill will have more wins ().

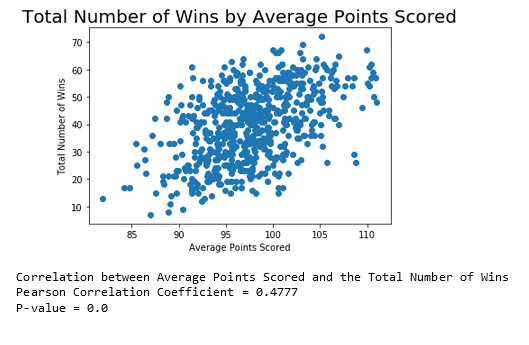
The l;evel of significance is 5%.

Table 1: Hypothesis Test for the Overall F-Test

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | 2865.00 |
| P-value | 0.000 (8.06e-234) |

* 1. The conclusion of the test is to reject the null hypothesis based on the P-value of zero. Average relaive skill can predict the total number of wins in the regular season.
  2. The predicted total number of wins in a regual season for a team with an average relative skill of 1550 is 45 games.
  3. The predicted total number of wins in a regular season for a team with an average relative skill of 1450 is 34 games.

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



The scatterplot and the Pearson correlation coefficient show a moderate positive correlation between average points scored and total number of wins.

Using a 1% level of significance, the correlation is shown to be statistically significant due to the P-value being 0.0 (< 0.01).

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

Generally, a multiple regression model is used to predict the response variable by assessing its relationship between multiple predictor variables. An F-test is performed to see if ANY of the predictor variables have a statistically significant relationship to the response variable. Individiual t-tests can be found within the OLS Regression results to look at individual variables.

The equation for my model is: total\_wins = -152.5736 + 0.1055 \* avg\_elo\_n + 0.3497 \* avg-pts.

The null hypothesis is that average relative skill and average points do not affect total wins in a season for a team ().

The alternative hypothesis is that average relative skill and average points affect total wins in a season for a team ().

The level of significane is 5%.

Table 2: Hypothesis Test for the Overall F-Test

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | 1580 |
| P-value | 0.000 (4.41e-243) |

The conclusion of the test is to reject the null hypothesis. The F-test shows that at least one of the predictors are statistically significant in predicting the total number of wins.

The results of the individual t-tests for the paramaters of each predictor variable all show that they are statistically significant using a 1% level of significance (all P-values are zero).

The coefficient of determination is 0.837, meaning that 83.7% of the variation in total wins can be explained by average points scored and average relative skill.

The predicted total number of wins in a regular season for a team that is averaging 75 points per game with a relative skill level is 1350 is 16.

The predicted total number of wins in a regular season for a team that is averaging 100 points per game with a relative skill level is 1600 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

Generally, a multiple regression model is used to predict the response variable by assessing its relationship between multiple predictor variables. An F-test is performed to see if ANY of the predictor variables have a statistically significant relationship to the response variable. Individiual t-tests can be found within the OLS Regression results to look at individual variables.

The equation for my model is: total\_wins = 34.5753 -0.0134 \* avg\_elo\_n + 0.2597 \* avg\_pts + 1.6206 \* avg\_pts\_differential + 0.0525 \* avg\_elo\_differential.

The null hypothesis is that there is no relationship between total wins and any of the predictor variables ().

The alternative hypothesis is that there is a relationship between total wins and any of the predictor variables ().

The level of significance is 5%.

Table 3: Hypothesis Test for Overall F-Test

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | 1102 |
| P-value | 0.0000 (3.07e-278) |

The conclusion of the hypothesis test is to reject the null hypothesis.

The overall F-test shows that at least one of hte predictors is statistically significant in predicting hte number of wins in the season.

The results of the individual t-tests for each predictor varible, using a 1% level of significance, show that only average relative skill does not have a statistically significant relationship to total wins (P-value 0.442). Average points, average point differential, and average elo differential do have a statistically significant relationship to total wins (P-values < 0.01).

The coeficcient of determination is 0.878, meaning that 87.8% of total wins can be explained by the predictor variables.

The predicted number of wins in a regular season for a team that is averaging 75 points per game with a relative skill level of 1350, average point differential of –5, and average relative skill differential of –30 is 26.

The predicted number of wins in a regular season for a team that is averaging 100 points per game with a relative skill level of 1600, average point differential of +5, and average relative skill differential of +95 is 52.

## 8. Conclusion

The results of the statiscal analyses show that there are clear and statistically significant relationships between an NBA team’s total wins and their average points scored, average point differential to the opposing teams, and average relative skill level to the opposing teams.

Using scatterplots, this data shows that you can visualize and predict how many games a team wil win based on these factors with relatively strong accuracy. The trend between wins and relative skill is strong, and the trend between wins and points scored is moderate.

The practical importance of these findings is that coaches can gauge how well their team is doing as well as how other teams are doing in a season. With this information, they can make predictions about outcomes and tailor their training against certain teams. They can also identify outlying performances to determine the cause of why they did exceptionally well or poor versus what was expected of them.