

MAT 243 Project Three Summary Report

Sean Toon
sean.toon@snhu.edu
Southern New Hampshire University

1. Introduction

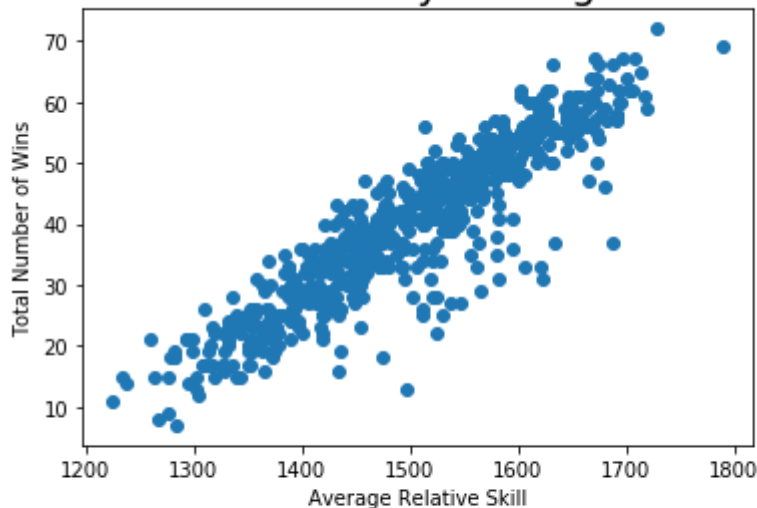
The data set that I will be exploring is historical data of performance patterns from NBA basketball teams. I will analyze the total wins, avg points scored, average relative skill, average point differential between the team and their opponents, and average relative skill differential between the team and their opponent in a regular season. My goal is to come up with regression models that help predict the total number of wins for a team in the regular season based on these key performance metrics.

2. Data Preparation

Some important variables to understand in this project are the average points differential, and the average relative skill. The average points differential represents the difference between the average number of points a team scores per game and the average number of points their opponents score per game. The average relative skill level is measured by the final score of a game, the game location and the outcome of the game relative to the probability of that outcome, also called an ELO rating.

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

Total Number of Wins by Average Relative Skill



Correlation between Average Relative Skill and the Total Number of Wins
Pearson Correlation Coefficient = 0.9072
P-value = 0.0

For this analysis the scatterplot worked particularly well in representing the data. It worked well because each dot on the scatterplot represents a pair of values from the two variables being analyzed. The scatterplot also makes identifying trends and patterns easily.

The Pearson correlation coefficient is valuable in this case because it shows the strength and direction of the association between the two variables. If the correlation coefficient is positive that means that there is a visible upward trend in the data, which in this case there is. If the correlation coefficient is greater

than 0.80 that indicates that there is a strong correlation between the two variables, and in this case there is.

Because there is a strong correlation between the two variables, the plot shows that as the average relative skill of a team increases, so does the total number of wins and vice versa.

The correlation coefficient is statistically significant based on the p-value at a 1% level of significance because the p-value is less than 0.01.

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

A simple linear regression model is used to predict the response variable using the predictor value. By assuming a linear relationship between the two variables being used, the parameters that best fit the data can be estimated.

The equation for this model is:

$$\text{total wins} = -128.2475 + 0.1121 \times \text{average relative skill}$$

The null hypothesis in statistical notation is:

$$H_0 : B_1 = 0$$

The null hypothesis in words is:

There is no linear relationship between the average relative skill and the total number of wins.

The alternative hypothesis is:

$$H_a: B_1 \neq 0$$

The alternative hypothesis in words is:

There is a linear relationship between the average relative skill and the total number of wins.

The level of significance used for the test is 5%

Table 1: Hypothesis Test for the Overall F-Test

Statistic	Value
Test Statistic	2865.00
P-value	8.06e-234

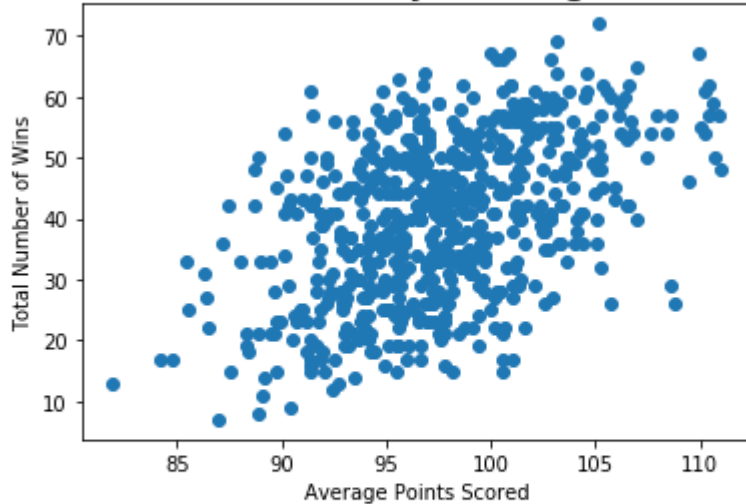
Based on the results of the overall F-test, the average relative skill can predict the total number of wins in the regular season.

An example using the equation is, the predicted total number of wins in a regular season for a team that has an average relative skill of 1550 is 46.

Another example is if a team has an average relative skill of 1450, the equation predicts that the number of wins in a regular season is likely to be 34!

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

Total Number of Wins by Average Points Scored



Correlation between Average Points Scored and the Total Number of Wins
Pearson Correlation Coefficient = 0.4777
P-value = 0.0

The scatterplot and the Pearson correlation coefficient indicate that there is a moderate correlation between the total number of wins and the average points scored. The correlation is moderate because it is smaller than 0.80 but larger than 0.3. Additionally, the p-value of 0.0 indicates that the correlation coefficient is statistically significant because it is less than 0.01.

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.

A multiple linear regression model is used to predict the response variable using predictor variables by first assuming there is a linear relationship and then the prediction is made by fitting a line that minimizes the sum of squared errors between the actual data points and predicted values.

The equation for the model produced is:

$$\text{Total wins} = -152.5736 + 0.3497 \times \text{average points} + 0.1055 \times \text{average relative skill}$$

The null hypothesis in statistical notation is:

$$H_0: B_1 = B_2 = 0$$

The null hypothesis in words is:

There is no linear relationship between the predictors and the total number of wins.

The alternative hypothesis in statistical notation is:

$H_a: B_1 \neq 0 \text{ or } B_2 \neq 0$

The alternative hypothesis in words is:

At least one of the predictors is statistically significant in predicting the total number of wins.

The level of significance used is 5%.

Table 2: Hypothesis Test for the Overall F-Test

Statistic	Value
Test Statistic	1580.00
P-value	4.41e-243

The p-value indicates that at least one of the predictors is statistically significant in predicting the total number of wins in the season because it is much lower than the level of significance.

The results of the average points and the average relative skill individual t-tests have p-values less than the significance level. This means that both predictors are statistically significant in predicting the total number of wins in the season.

The coefficient of determination is 0.837. This means that approximately 83.7% of the variability in total wins can be explained by the predictor variables.

An example using the equation is, 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 of 1350 is 16 wins.

Another example is, if a team is averaging 100 points per game with an average skill level of 1600, the predicted total number of wins in a regular season for that team 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

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.

A multiple linear regression model is used to predict the response variable by considering multiple predictor variables. In this case, four predictors are used.

The equation for the model is:

Total wins = $34.5753 + 0.2597 \times \text{average points} - 0.0134 \times \text{average relative skill} + 1.6206 \times \text{average points differential} + 0.0525 \times \text{average relative skill differential}$

The null hypothesis in statistical notation is:

$H_0: B_1 = B_2 = B_3 = B_4 = 0$

The null hypothesis in words is:

None of the predictor variables are statistically significant in predicting total wins.

The alternative hypothesis is:

Ha: At least one of $B_1, B_2, B_3, B_4 \neq 0$

The alternative hypothesis in words is:

At least one of the predictor variables is statistically significant in predicting total wins.

The level of significance is 5%

Statistic	Value
Test Statistic	1102.00
P-value	3.07e-278

Since the p-value is larger than the significance level, the null hypothesis is rejected and there is enough evidence to support that at least one of the predictors is statistically significant in predicting the number of wins in the season.

Intercept p-value: 0.182 (not statistically significant)

Average points p-value: 0.000 (statistically significant)

Average relative skill p-value: 0.442 (not statistically significant)

Average points differential p-value: 0.000 (statistically significant)

Average relative skill p-value: 0.004 (statistically significant)

The coefficient of determination is 0.878. This means 87.8% of the variance in total wins is explained by the model.

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 of 1350, average point differential of -5 and average relative skill differential of -30 is 26 wins.

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 of 1600, average point differential of +5 and average relative skill differential of +95 is 52 wins.

8. Conclusion

Describe the results of the statistical analyses clearly, using proper descriptions of statistical terms and concepts. Fully describe what these results mean for your scenario.

In this project I have been able to develop useful models that NBA management teams can use to predict the total number of wins based on key performance metrics. I produced a simple linear regression model that accurately showed that average relative skill is a significant predictor of the total number of wins. I produced a multiple regression model that proved average points scored and average relative

skill are significant predictors. I also produced a multiple regression model. As I added more predictors, I was able to increase the accuracy of the models.

The model can be used by NBA teams and management to analyze what performance metrics are the most important to focus on if the total number of wins of games an important goal is.