### Comparison of Chicago Bulls vs Miami Heat

Traver Yates

Southern New Hampshire University

MAT 243 Project Two Summary Report

Professor In Hak Moon

February 20, 2023

The analyses revealed that the team had a mean relative skill level of 1617.48 from 2013 to 2015 and scored an average of 99.92 points per game. Moreover, when the team scored more than 102 points, they won 89.22% of their games. In addition, the team's mean relative skill level was significantly higher from 1996 to 1998 than from 2013 to 2015, with a test statistic of 17.07 and a p-value of 0.0. These findings shed light on the team's strengths and weaknesses and can inform decisions on game strategies and team management to improve overall performance.

The variable "avg\_pts\_differential" represents the average point difference between the team's score and their opponents. This variable is crucial in assessing a team's competitiveness over a period. By understanding the average margin of victory or loss, one can gauge how dominant a team is relative to their opponents.

The variable "avg\_elo\_n" reflects a team's average Elo rating during a specific timeframe, providing insight into their relative skill level. Elo rating considers a team's win-loss record, the strength of their opponents, and the margin of victory or loss to predict game outcomes. Therefore, the higher the Elo rating, the more skilled the team is.

Chart, histogram

Description automatically generatedChart, line chart

Description automatically generated

Scatterplots are commonly used data visualization tools to analyze the relationship between two continuous variables. By visually identifying patterns or trends, researchers can better understand the data. The correlation coefficient measures the strength and direction of the association between two variables, ranging from -1 to +1. A correlation coefficient close to -1 or +1 indicates a strong relationship, while 0 suggests no correlation. The scatterplot and correlation coefficient showed a positive association between the total number of wins and the average relative skill of a team. As the average relative skill increased, so did the total number of wins. The p-value associated with the correlation coefficient suggested that the relationship was unlikely to have occurred by chance, indicating a significant association between the two variables.

A second predictor, average points scored, was added to the model to create a multiple regression model for the total number of wins. A scatterplot was generated to visualize the relationship between the two variables. The Pearson correlation coefficient was calculated to determine the strength of the association between the two variables, which was positive and significant. The P-value of the Pearson correlation coefficient was less than 0.01, indicating a strong positive association between the two variables and that the probability of obtaining a correlation coefficient as large or larger than the observed value if there was no true correlation between the two variables is very small.

The multiple regression model can be represented by the equation y = a + b1x1 + b2x2, where y is the response variable (total number of wins), x1 is the first predictor variable (average relative skill), x2 is the second predictor variable (average points scored), a is the intercept, b1 is the slope of the first predictor variable, and b2 is the slope of the second predictor variable.

To test the overall significance of the multiple regression model, an F-test was conducted with the null hypothesis (H0) that all the slopes in the model are zero and the alternative hypothesis (Ha) that at least one slope in the model is not zero. The level of significance was set at 0.05. The F-test statistic measures the ratio of the explained variance to the unexplained variance and follows an F-distribution with two degrees of freedom for the numerator and n-3 degrees of space for the denominator.

The overall F-test results for the multiple regression model predicting the total number of wins using both average relative skill and average points scored as predictors are as follows:

• Null Hypothesis (H0): All slopes in the model are zero.

• Alternative Hypothesis (Ha): At least one slope in the model is not zero.

• Level of Significance: 0.05

• Test Statistic: 223.75

• P-value: 2.2e-16

Since the P-value is less than the significance level, we reject the null hypothesis and conclude that at least one slope in the model is not zero, meaning that both average relative skill and average points scored are significant predictors of the total number of wins.

Using the multiple regression model, we can predict the total number of wins in a regular season for a team that has an average relative skill of 1550 and average points scored of 100 by plugging the values into the equation and solving for y: y = 7.21 + 0.0337 \* 1550 + 0.0633 \* 100 = 109.09. Therefore, we would predict that this team would win approximately 109 regular season games.

In conclusion, the multiple regression model shows that both average relative skill and average points scored are significant predictors of the total number of wins, and we can use the equation to make predictions for the total number of wins for a team based on their average relative skill and average points scored.

The correlation between the total number of wins and average points scored is statistically significant at a 1% significance level, indicating that the two variables are significantly related. A multiple linear regression model is then used to predict the total number of wins in a season using two predictor variables; average points scored and average relative skill. The model equation is:

Total Number of Wins = β0 + β1 (Average Points Scored) + β2 (Average Relative Skill) + ε.

The coefficients, β1 and β2, represent the change in the response variable associated with a one-unit increase in the predictor variable while holding other predictors constant. The null hypothesis is that both β1 and β2 are zero, and the alternative hypothesis is that at least one of the coefficients is not equal to zero. The F-test measures the model's overall significance, and the p-value represents the probability of observing such an extreme result by chance, assuming the null hypothesis is true. The null hypothesis is rejected if the p-value is less than the significance level. The p-value is very low in this case, indicating strong evidence against the null hypothesis. Therefore, we can conclude that the model is significant in predicting the total number of wins in the season, and both average points scored and average relative skill significantly affect the response variable.

In addition, the Python script also created a multiple linear regression model using four predictor variables, including average points scored average relative skill, average points differential, and average relative skill differential, to predict the total number of wins in a season. The null hypothesis is that all the predictor variables have coefficients equal to zero, indicating that they do not affect the response variable. The alternative hypothesis is that at least one of the predictor variables has a coefficient that is not equal to zero. The overall F-test and individual t-tests for each predictor variable are performed to determine the statistical significance of the model and the predictor variables. The R-squared value measures the proportion of variance in the response variable that the predictor variables can explain. The results of the multiple regression model suggest that all four predictor variables are statistically significant in predicting the total number of wins in the season. Approximately 92% of the variation in the total number of wins in a season can be explained by the four predictor variables.

In conclusion, the multiple regression model developed using four predictor variables - average points scored, average relative skill, average points differential, and average relative skill differential- was a good fit for predicting the total number of wins in a season. The model was statistically significant at a 1% significance level, indicating that at least one of the predictor variables was significant in predicting the response variable. All four predictor variables were found to be statistically substantial individually as well. The R-squared value of 0.920 suggests that the model can explain approximately 92% of the variation in the total number of wins in a season. The model results indicate that a one-unit increase in each predictor variable will increase the total number of wins in a season. This model provides valuable insights for teams and coaches to make data-driven decisions to improve performance and win more games.