# MAT 243 Project Three Summary Report

Nicholas Cleveland

nicholas.cleveland1@snhu.edu

Southern New Hampshire University

## 1. Introduction

*Discuss the statement of the problem in terms of the statistical analyses that are being performed. Be sure to address the following:*

* *What is the data set that you are exploring?*

The data set that I will be exploring in this report is the correlation between predictor and response variables used in given basketball game statistics from the NBA. I will be looking at the correlation specifically between the average number of points to the number of wins, the average relative skill to the number of wins, and both predictors, the average number of points and skill level to the number of wins. I will finally add a third predictor variable of the average points differential along with the first two predictor variables compared to the response variable of the number if game wins.

* *How will your results be used?*
* *What type of analyses will you be running in this project?*

The types of analyses that will be conducted will be…

## 2. Data Preparation

*There are some important variables that are used in this project. Identify and explain these variables.*

*See the introductory section and Step 1 of the Python script to address the following:*

* *What does the variable* ***avg\_pts\_differential*** *represent? How would you explain it to someone who does not understand the data?*

The average points differential represents the average difference of points between a team and their opponents during the season*.*

* *What does the variable* ***avg\_elo\_n*** *represent? How would you explain it to someone who does not understand the data?*

The average relative skill level variable represents the relative skill of each team in a regular season.

## 3. Scatterplot and Correlation for the Total Number of Wins and Average Points Scored

*You constructed a scatterplot of the total number of wins and the average points scored to study their correlation. You also calculated the Pearson correlation coefficient along with its P-value.*

*See Step 2 in the Python script to address the following items:*

* *In general, how are data visualization techniques used to study relationship trends between two variables?*

In general, data visualization techniques can be used to show relationships between two variables more easily than another representation such as a table or similar. The data visualization manifested as a scatter plot can show a clear relationship between a predictor variable and a response variable to the point where the trend can be identified. Most of the time trends will show representations such as a positive or negative trend.

* *How is the correlation coefficient used to get the strength and direction of the association between two variables?*
* *In this activity, you generated a scatterplot of the total number of wins and the average points scored. Include a screenshot of this plot in your report.*

Chart, scatter chart

Description automatically generated

* *What do the scatterplot and the Pearson correlation coefficient tell you about the association between total number of wins and average points scored?*
* *Is the correlation coefficient statistically significant based on the P-value? Use a 1% level of significance.*

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include Python code in your report.*

## 4. Simple Linear Regression: Predicting the Total Number of Wins using Average Points Scored

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

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

* *In general, how is a simple linear regression model used to predict the response variable using the predictor variable?*
* *What is the equation for your model?*
* *What are the results of the overall F-test? Summarize all important steps of this hypothesis test. This includes:*
  1. *Null Hypothesis (statistical notation and its description in words)*
  2. *Alternative Hypothesis (statistical notation and its description in words)*
  3. *Level of Significance*
  4. *Report the test statistic and the P-value in a formatted table as shown below:*

Table 1: Hypothesis Test for the Overall F-Test

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | X.XX  *\*Round off to 2 decimal places.* |
| P-value | X.XXXX  *\*Round off to 4 decimal places.* |

* 1. *Conclusion of the hypothesis test and its interpretation based on the P-value*
* *Based on the results of the overall F-test, can average points scored predict the total number of wins in the regular season?*
* *What is the predicted total number of wins in a regular season for a team that is averaging 75 points per game? Round your answer down to the nearest integer.*
* *What is the predicted number of wins in a regular season for a team that is averaging 90 points per game? Round your answer down to the nearest integer.*

* Answer the questions in a paragraph response. Remove all questions and this note (but not the table) before submitting! Do not include Python code in your report.*

**5. Scatterplot and Correlation for the Total Number of Wins and Average Relative Skill**

*You constructed a scatterplot of total number of wins and average relative skill. You also calculated the Pearson correlation coefficient along with its P-value.*

*See Step 4 in the Python script to answer the following questions:*

* In this activity, you generated a scatterplot of the total number of wins and average relative skill. Include a screenshot of this plot in your report.

Chart, scatter chart

Description automatically generated

* *What do the scatterplot and the Pearson correlation coefficient tell you about the association between total number of wins and average relative skill?*
* *Is the correlation coefficient statistically significant based on the P-value? Use a 1% level of significance.*

* Answer the questions in a paragraph response. Remove all questions and this note before submitting! Do not include Python code in your report.*

## 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:*

* *In general, how is a multiple linear regression model used to predict the response variable using predictor variables?*
* *What is the equation for your model?*
* *What are the results of the overall F-test? Summarize all important steps of this hypothesis test. This includes:*
  1. *Null Hypothesis (statistical notation and its description in words)*
  2. *Alternative Hypothesis (statistical notation and its description in words)*
  3. *Level of Significance*
  4. *Report the test statistic and the P-value in a formatted table as shown below:*

Table 2: Hypothesis Test for the Overall F-Test

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | X.XX  *\*Round off to 2 decimal places.* |
| P-value | X.XXXX  *\*Round off to 4 decimal places.* |

* 1. *Conclusion of the hypothesis test and its interpretation based on the P-value*
* *Based on the results of the overall F-test, is at least one of the predictors statistically significant in predicting the total number of wins in the season?*
* *What are the results of individual t-tests for the parameters of each predictor variable?*

*Is each of the predictor variables statistically significant based on its P-value? Use a 1% level of significance.*

* *Report and interpret the coefficient of determination.*
* *What 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?*
* *What is the predicted total number of wins in a regular season for a team that is averaging 100 points per game with an average relative skill level of 1600?*

** *Answer the questions in a paragraph response. Remove all questions and this note (but not the table) before submitting! Do not include Python code in your report.*

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

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

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

* *In general, how is a multiple linear regression model used to predict the response variable using predictor variables?*
* *What is the equation for your model?*
* *What are the results of the overall F-test? Summarize all important steps of this hypothesis test. This includes:*
  1. *Null Hypothesis (statistical notation and its description in words)*
  2. *Alternative Hypothesis (statistical notation and its description in words)*
  3. *Level of Significance*
  4. *Report the test statistic and the P-value in a formatted table as shown below:*

Table 3: Hypothesis Test for Overall F-Test

| **Statistic** | **Value** |
| --- | --- |
| Test Statistic | X.XX  *\*Round off to 2 decimal places.* |
| P-value | X.XXXX  *\*Round off to 4 decimal places.* |

* 1. *Conclusion of the hypothesis test and its interpretation based on the P-value*
* *Based on the results of the overall F-test, is at least one of the predictors statistically significant in predicting the number of wins in the season?*
* *What are the results of individual t-tests for the parameters of each predictor variable?*

*Is each of the predictor variables statistically significant based on its P-value? Use a 1% level of significance.*

* *Report and interpret the coefficient of determination.*
* *What 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 and average point differential of -5?*
* *What is 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 and average point differential of +5?*

* Answer the questions in a paragraph response. Remove all questions and this note (but not the table) before submitting! Do not include Python code in your report.*

## 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.*

* *Briefly summarize your findings in plain language.*
* *What is the practical importance of the analyses that were performed?*

## 9. Citations

*You were* ***not*** *required to use external resources for this report. If you did not use any resources, you should remove this entire section. However, if you did use any resources to help you with your interpretation, you* ***must*** *cite them. Use proper APA format for citations.*

Insert references here in the following format:

Author's Last Name, First Initial. Middle Initial. (Year of Publication). Title of book: Subtitle of book, edition. Place of Publication: Publisher.