**Statistical Analysis of NBA Champions:**

**A Guide to Creating an Ideal NBA Team**

**(This project was designed by a group of 7 professionals)**

**Statistical Analysis of NBA Champions**

**Part One: Project Introduction**

The National Basketball Association (NBA) provides entertainment for viewers after a long day. However, data from basketball games can assist in future decision-making. The importance of this review allows for the necessary data to build an ideal team. This goal guided Group 7 into conducting a statistical analysis on NBA champions over the last 50 years. The group has chosen 7 variables to develop the dataset. They include four conscious variables, two categorical, and one binary. This data will be analyzed through six methods including: descriptive statistics, test of hypothesis for one mean and one proportion, test of hypothesis for the difference between two means, test of independence, linear regression, and multiple regression.

The variables are all related to the champion won for that given year. NBA basketball seasons begin mid-fall, October, and continue until late spring, April, of the following year with the championship running from April to June (Basketball Word, 2021). This category displays the beginning and end years of that season. Additionally, the playoffs follow a “best-of-seven” structure where the team scoring the highest in the first four games is declared the winner of the series (Dictionary.com, 2021). The continuous variables are as follows, points per game (PPG), opponent’s points per game (Opp PPG), winning percentage, and average age of team. The categorical variables are the winning team of that game and the division of competition. NBA divisions are organized by team location for a total of six. Lastly, the group is using whether the champion had home court advantage as the binary variable. Home court advantage refers to a team playing in the venue at or near their practice stadium (Hoops Addict, 2021). These factors, among many others, all contribute to the managerial decision-making process when developing a team.

**Part Two: Managerial Summary**

Sports fanatics have been known to make assumptions that fall into a subjective or an objective category. Some people take their passion for sports and invest in their favorite team. In a professional setting, businesses have been seen financially supporting teams or sponsoring individual players. Oftentimes, these avenues influence a return on product marketing. Before investing in either business or personal avenues, it is important to view the current information and understand NBA outcomes from a quantitative perspective. The proper data analysis promotes capitalization on creating a winning championship team.

**Understanding the Numerical Values:**

To understand the NBA data, a summarized version of the information was created. The first four factors observed describe popular elements that many analyze or discuss and, as alluded to earlier, are classified as continuous variables. The first two variables in the continuous category relate to points per game scored and points per game allowed. In basketball, teams serve as either defense or offense depending on which player has the ball. The team acting as defense protects the net whereas the team acting as offense travels and passes toward said net (Breakthrough Basketball, LLC, 2021). Both defensive and offensive prowess fall in the descriptive category since both factors displayed relevance in previous champions.

The next two variables included in the description are the winning percentage and the average age of the championship team. Not only do these categories explain the values presented, but an additional relationship can be observed. All values represented in the numerical spreadsheet shows that championship teams come in all shapes and sizes. Some teams have better offensive skills but weaker defenses and vice versa. Usually, a team's performance is based on previous data such as average age of team and/or previous winning percentages. However, it is hard to prove these categories provide criteria for a winning team without viewing statistical data.

**Testing Relevant Assumptions:**

Sometimes, it is difficult to separate a fact and an opinion especially in terms of basketball. Testing an assumption using numerical values allows for a more accurate answer. Quantitative values also allow for a stronger argument against verbal ideas mentioned in the decision-making process. As mentioned previously, winning NBA championship teams could have similar factors. Understanding these areas with the use of historical values can help predict which team will win. Additionally, a championship NBA team could be created after reviewing similarities amongst winning teams.

One assumption pertains to one team having leverage over the other depending on where the championship game is held, often referred to as home court advantage. The scores obtained by opposing teams playing on that did not have this support were compared against teams that did. The goal was to see if teams that played on their home court scored higher. When looking at the outcome, the information tested did not support the home court advantage. It is important to note that the NBA follows a “best-of-seven” structure to reduce this bias. Additionally, the NBA could still begin their investment in building neutral courts (Cacciola, 2020). Even though the data tested is not in support, it does not mean that the advantage does not exist.

When looking at NBA championship games, points per game for both teams has been reported to provide essential information. Again, it is assumed that teams entering the championship series are highly skilled and experienced. One can guess that teams entering the playoffs will score above average, however, that might not actually be true. To see if the chances of winning are greater for these teams, the average value of 70% was used. The results supported this guess with each team having over a 70% chance of winning once they begin championship games.

Along with the two relationships discussed, the average of points between PPG and Opp PPG was observed. The outcome showed that there was not enough significant evidence to prove that there was a difference between the two averages. The calculated conclusion supports a common expectation. The outcomes reported are important to know when recruiting, training, and investing in a potential “championship” team. The NBA officially recognized the significance of data analysis as it pertains to basketball. The organization promotes tracking basketball and business analytics during the annual Hackathon event (NBA Media Ventures, LLC, 2019).

**Connection Between Two Variables:**

The significance of having the league’s best player on a team is a source of debate. While having a very talented individual on a roster is beneficial, basketball is a team sport. It requires a group of players to work together towards the goal of winning a championship. An important metric is an association between a team having the regular season Most Valuable Player (MVP) on their roster and earning home court advantage during the NBA finals. The regular season MVP is voted based on their performance up to the playoffs. Home court advantage during the finals is awarded to the team with the best record during the regular season.

To determine if there is a relationship, we viewed data over the last 50 NBA seasons. Our assumption going into these tests was that there is no relationship between these two factors with the alternative being there is a relationship. Through statistical analysis it was determined there is not an association between having the regular season MVP on a roster and achieving home court advantage. This analysis can help guide team management to make informed choices when building a roster. Creating a well-balanced team provides an organization a better opportunity to win a championship instead of seeking out a single great player.

**Relationships amongst One or More Factors:**

Linear regression is a form of linear algebra that was allegedly invented by Carl Friedrich Gauss (1777–1855) but was first published in a scientific paper by Adrien-Marie Legendre (1752–1833). Gauss used the least squares method to guess when and where the asteroid Ceres would appear in the night sky (The Discovery of Statistical Regression, 2015). This was not a hobby project, this was a well-funded research project for the purpose of oceanic navigation, a highly competitive field that was sensitive to technological disruption. (Peterson 2020). It Is now one of the best obtainable methods to a data scientist or a statistician.

Linear regression is a simple method of comparing two variables. The method consists of trying to come up with a prediction of one variable based on another variable believed to have more significance. The factor believed to have more influence is referred to as independent or x while the factor that is being observed is considered dependent or y.

It is not clear whether there is significance in the “independent” variable which is the main reason for an analysis. When creating an ideal NBA team, the average age of the team was assumed to be independent, therefore, influencing team wins. Although it is assumed that basketball skills decrease as a team ages, the results observed state the opposite. In the values organized, the older average age of a team was associated with more team wins. The two factors affect each other positively from a numerical perspective since the increase in one increases the results of the other. This outcome is relevant based on the values reviewed; however, different numerical values could represent a different outcome.

Sometimes, one factor shows enough of an influence to end the statistical review. However, in this situation, a few other factors were reported to affect the winning percentage. The ideal calculation resulted in both points per game and average age of team reflecting the percentage of wins. Additionally, opponent points per game were also added to the model to see if the results changed. To create a winning NBA team, it is important to know how each factor benefits the end goal. Unfortunately, the calculations displayed some consistency in wins, but not a large amount. If a larger result is requested, more factors should be observed and added into the current model.

**Part Three: Statistical Output**

**Descriptive Statistics:**

In the last fifty years, the average NBA champion scored 106.69 points per game, while only giving up 99.87 points per game for an average margin of victory of 6.82 points per game. Unsurprisingly, the ability to score and stop the opponent from scoring is important to winning. Does this mean that every champion was built the same way? Absolutely not. The minimum points per game a NBA champion scored was 90.1, which is 16.59 points per game lower than the average champion. On the defensive side, the most points a NBA champion surrendered was 110.9 points per game, 11 points per game higher than the average champion.

***Chart, box and whisker chart

Description automatically generated*Table I**

***Boxplot of PPG***

***Chart, box and whisker chart

Description automatically generated*Table II**

***Boxplot of Opp PPG***

In the next two continuous variables, the winning percentage and average team age illustrates the wide-ranging differences between champions in any given year. The mean value of winning percentage was 73.13% and the median value was 73.21%, implying normal distribution. The average age was 28.128 with a median age of 28.2 so age also appears to be normally distributed. The minimum winning percentage was 53.66% and the maximum was 87.8% for a range of 34.15%. Meanwhile, the youngest NBA champion was, on average, 24.8 years old, while the oldest average age was 31.2 years old for a range of 6.4 years. It is extremely rare for a team with a win percentage below 65% and/or an average age below 27 to win a championship, but it is possible.

Chart, box and whisker chart

Description automatically generated**Table III**

***Boxplot of Win %***

**Table IV**

***Boxplot of Aver Age of Team***

Chart, box and whisker chart

Description automatically generated

The last factors we looked at were three categorical variables: home-court advantage, division, and team. 33 of the last 50 champions or 66% had home court advantage. 58% of the champions came from the Pacific or Midwest division. The division with the fewest championships was the central division with 1. The three teams that have won the most are the Lakers, Celtics, and Bulls who all happen to reside in major markets.

**Testing Hypothesis:**

When creating a hypothesis, two ideas often associated with championship wins were tested. The first one examined home court advantage. Many believe that teams that are competing in the same stadium where they practice, have an advantage over teams who traveled. Even though the final games strive to make the competition fair, a hypothesis was tested to see if away opponents lacked this benefit. The baseline is that there is no advantage, therefore, the null hypothesis is equal to 0 whereas the alternative hypothesis is not equal to 0. The hypothesis being tested is best reflected in a Two-Sample T Test. Opp PPG compared to home court Advantage was tested using Minitab. Before beginning this model, the variances of the data set were observed to see if Opp PPG and Homecourt advantage had equal variances. An alpha of 0.05 with a 95% confidence interval was used as well. In the outcome, the two variances did not overlap with one another as the multiple comparison P-Value was 0.006 and the Levene’s test P-value was 0.481.

**Chart, box and whisker chart

Description automatically generated**It is important to note that unequal variances, otherwise known as heteroscedasticity, could influence a Type I error. However, a two-sample T- Test was carried out because the sample data is over 30 and the standard deviation is unknown. When running the hypothesis test, the P-value resulted in 0.332 with the two variances not assumed as equal. Because the P-value is greater than the 0.05 alpha, we fail to reject the null hypothesis. There is not enough evidence to support that points per game scored by the away opponent team is affected by home court advantage.

**Table V**

***Opp PPG vs Homecourt Advantage Equal Variance Test***

**Chart, box and whisker chart

Description automatically generatedTable VI**

***Opp PPG vs Homecourt Advantage Equal Variance Test***

Additionally, another hypothesis looking at the proportion of wins was examined. For teams in the NBA finals, is the chance of winning greater than 70%? The null hypothesis would equal 0.70 while the alternative would be greater than 0.70. A one-sample T-Test was used in Minitab displaying a P-Value of 0.002. Since this value is less than the 0.05 alpha value, the null hypothesis is rejected. Each team has over a 70% chance of winning. This outcome makes sense logically since each team entering the finals succeeded throughout the competition season. The teams have been shown to work together well while carrying out exceptional basketball techniques. Additionally, these results support the decision to invest monetarily when teams enter the playoffs since there is an above average chance of them winning according to previous results.

Testing for the difference between two means was also observed. The two factors considered were PPG and Opp PPG to see if there was a difference between the two averages. The null hypothesis in this calculation refers to (PPG - Opp PPG = 0) with the alternative being (PPG - Opp PPG not = 0). Prior to the calculation, the test of equal variances was calculated. The results physically shown in the summary plot, individual value plot, and boxplot as well as the P-values displayed variances overlapping. In the Two-Sample T Test, the variances were assumed equal resulting in a P value of 0.790 which was higher than the 0.05 alpha. This resulted in a fail to reject null hypothesis conclusion as there was not enough evidence to show a significant difference.

When testing hypothesis and confidence intervals, a Chi-Square test would not have been valid due to the data and the analysis purpose. It is also important to note that different data could result in different outcomes. For example, home court advantage has assisted in teams winning about 61% of games in the finals. The NBA historical values show that Game 1 and Game 7 in the series are highly influenced by home court advantage. The results of these tests show variables to observe when choosing a winning team. Even though these outcomes could be reported differently in other models, the results were constructed through the data samples with the use of Minitab (Karim, 2021).

**Test of Independence:**

When diving deeper into the statistical outputs, there could be an association between having the regular seasons, Most Valuable Player (MVP) on a championship team compared to earning home court advantage in the NBA finals. This analysis was conducted through a test of independence, otherwise referred to as the Chi-Square Test for Association. The null hypothesis assumes there is no association between the regular season MVP and home court advantage while the alternative hypothesis tries to prove that there is an association.

**Table VII**

***Test of Independence on Minitab***

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Chi-Square** | **DF** | **P-Value** |
| Pearson | 1.666 | 1 | 0.197 |
| Likelihood Ratio | 1.686 | 1 | 0.194 |

**Table VIII**

***Common Average Ages compared to Number of Times Won***

|  |  |  |
| --- | --- | --- |
| Age |  | Times Won |
| 26 | X | 9 |
| 27 | X | 11 |
| 28 | X | 18 |
| 29 | X | 6 |
| 30 | X | 4 |

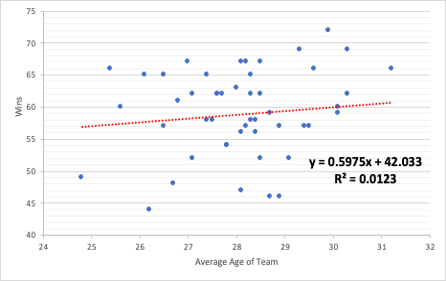
In the Chi-Square output, the Pearson’s value was 1.666 with the P-value of 0.197. It is important to note that the P-value is greater than the alpha of 0.05. This outcome results in failing to reject the null hypothesis. There is not enough significant evidence to determine a relationship between having the regular season MVP on a team’s roster and earning home court advantage in the NBA finals.

**Linear Regression:**

For our linear regression analysis, the two variables that were used consist of the average age of the teams (x) and the win percentage (y). The win percentage was calculated by taking the number of games won divided by the total of games played. With this data a scatter plot was created modeling the relationship between the two variables by fitting a linear equation to our observed data. From looking at the thread line on the scatter plot a positive slope 0.5975 was determined, identifying that there is a positive relationship between the two variables. With this information we can observe that the older the average age of the team, the more wins can be predicted.

**Table VIIII**

***Linear Regression Analysis***

****

**Multiple Regression:**

As seen in the dataset, the average age of NBA championship teams was reported. This variable is important when creating an ideal basketball team since age is believed to affect physical performance. Many believe that the peak age of a basketball player is reached at age 26. However, the gathered data displayed a combination of ages creating an average between age 27 and 28 resulted in 40% of the wins.

Continuing with the linear regression model, points per game and average age of team were calculated. The total R-Squared reported in this summary resulted in about 8.42%. Since this was a lower percentage, an additional variable, opponent points per game, was added. The three variables increased the R-Squared to 9.33%. Again, this value shows some significance on winning percentages, but not as much as one would want. On a positive note, multicollinearity was not reported as an issue since the variance inflation factor (VIF) was less than 5. Observations that are considered unusual can be seen under Fits and Diagnostics for Unusual Observations chart.

**Table X**

A picture containing chart

Description automatically generated***Regression Analysis for Two Independent Variables***

**Table XI**

***Regression Analysis for Three Independent Variables***

Graphical user interface

Description automatically generated with medium confidence

**Table XII**

***Fits and Diagnostics for Unusual Observations***

Table

Description automatically generated

After constructing the multiple regression model, a stepwise analysis was performed in addition to adding a categorical variable through recoding. The stepwise process allowed for each independent variable to be added as well as subtracted from the analysis. However, since this model first focused on two independent variables followed by three, the R-Squared adjusted value would provide the most appropriate conclusion. The R-Squared adjusted value considers how many independent variables, two or three, are in the model. In the regression observing points per game and average age as independent, the R-Squared adjusted is about 6.52% which remains low. However, the P-value of 0.041 is less than the 0.05 alpha concluding in a rejection of the null hypothesis. The residual plots regarding the winning percentage values in relation to two independent variables was also recorded. Overall, it is important to note that a combination of skills make up a team. There are many other variables that

Additional variables to consider would be positions assigned (defense or offense), physique (height, weight, BMI, muscle mass).

**Table XIII**

***Stepwise for Two Independent Variables***

**Table

Description automatically generated**

**Table XIIII**

***Residual Plots for Two Independent Variables*Chart

Description automatically generated**

**Conclusion**

By analyzing multiple variables, we can determine numerous factors related to developing an NBA championship team. These outcomes can be leveraged by team leadership to make managerial decisions to optimize team success. Of significance, we determined that the average age of a team can help predict how many games a team will win. Additionally, it was also identified that there is no association with a team having the regular season MVP on their roster and earning home-court advantage in the finals. These factors support developing an experienced team that has well-balanced talent. The players making up a championship team should have individual skills that complement each other, supported by their experience playing the game.

**References:**

Breakthrough Basketball, LLC. (2021). Basketball basics for new players and coaches—Learn

the basic rules, concepts, court layout, and player positions. Retrieved from

https://www.breakthroughbasketball.com/basics/basics.html

Basketball Word. (2021). Is basketball a winter or fall sport? Quick answer! Retrieved from

https://basketballword.com/is-basketball-a-winter-or-fall-sport/

Cacciola, S. (2020). On neutral courts, nba bubble ‘home’ teams still have an edge. Retrieved

from <https://www.nytimes.com/2020/08/16/sports/basketball/nba-homecourt-advantage>

wnba.html

Dictionary.com. (2021). Historical & current events dictionary: Nba finals. Retrieved from

https://www.dictionary.com/e/historical-current-events/nba-finals/

Hoops Addict. (2021). Home court advantage in the nba playoffs: Does it matter? Retrieved from

https://www.hoopsaddict.com/home-court-advantage-in-the-nba-playoffs/

Kalbrosky, B. (2018). What is the peak age in the nba? Probably 27 years old. Retrieved from

<https://hoopshype.com/2018/12/31/nba-aging-curve-father-time-prime-lebron-james>

decline/

Karim, S. A. (2021). Nba finals: Why home court advantage is an important factor. Retrieved

from <https://www.mediareferee.com/2021/06/27/nba-finals-why-home-court>

advantageis-animportant-factor/

NBA Media Ventures, LLC. (2019). Nba hackathon. Retrieved from https://hackathon.nba.com/

NBA Media Ventures, LLC. (2021). Nba season recaps: A look back at every season since 1946.

Retrieved from <https://www.nba.com/history/season-recap-index>

Peterson, R. (2020) Linear regression by hand. Retrieved from <https://towardsdatascience.com/linear-regression-by-hand-ee7fe5a751bf>

Tokarz, D. (2018). Per game team statistics in the nba. Retrieved from

https://sports.sites.yale.edu/game-team-statistics-nba

Sportradar. (n.d.). Basketball statistics and history. Retrieved from https://www.basketball-reference.com/