Taylor Spence

Intro to Probability and Stats

Dr. Jacob

January 15, 2023

2022-2023 NBA Basketball Players Statistics

**Project Part 1:**

The data I collected is from the website [List of all the NBA and ABA Players | Basketball-Reference.com](https://www.basketball-reference.com/players/) by using all the teams in the NBA, selecting a player from each team but making sure each position is represented equally, and using their personal statistics for the 2022-2023 season to fill out the rest of the information. This data set is about professional basketball players playing various positions and for various teams. I chose this data set because I grew up playing basketball and always watching it on television, even to this day. Also, a lot of my close friends play fantasy basketball, so, I figured I would use this opportunity to learn more about the players and the numerous factors of how that works.

In my analysis, I am using the variables of the different teams each player plays for, their position, height, games played, and how many team wins their respective team has. The players’ names do not count as variables since they are case names. However, the players' teams and positions they play are considered categorical variables. Whereas the players’ height, games played, and team wins are quantitative variables since these can be counted.

All the variables mentioned above are some that I feel are important to establish a relationship that can be useful when watching basketball or playing fantasy. Through this data set, I am hoping to establish a relationship between the height and games played by certain players to the success of their team. I am using height specifically because basketball is a sport where success can be driven from the height of their players. Height also helps determine the position the player will take on, but I want to show that this along with how much they are playing contributes to how many wins their team has.

By establishing this relationship between players’ height, games played, and team wins it can help my friends choose more wisely when selecting their teams for fantasy basketball. It will also change the way that I watch basketball by giving me a better understanding of how height in certain positions can lead to success. Included below is the data set that contains the different variables I am using.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Player** | **Team** | **Position** | **Height** | **Games Played** | **Team Wins** |
| John Collins | Atlanta Hawks | Power Forward | 6'9" | 35 | 21 |
| Jaylen Brown | Boston Celtics | Small Forward | 6'6" | 40 | 32 |
| Kyrie Irving | Brooklyn Nets | Point Guard | 6'2" | 31 | 27 |
| Terry Rozier | Charlotte Hornets | Shooting Guard | 6'1" | 34 | 11 |
| Patrick Williams | Chicago Bulls | Power Forward | 6'7" | 43 | 19 |
| Jarrett Allen | Cleveland Cavaliers | Center | 6'11" | 37 | 27 |
| Isaiah Stewart | Detroit Pistons | Center | 6'8" | 36 | 12 |
| Tyrese Haliburton | Indiana Pacers | Point Guard | 6'5" | 40 | 23 |
| Jimmy Butler | Miami Heat | Small Forward | 6'7" | 30 | 24 |
| Giannis Antetokounmpo | Milwaukee Bucks | Power Forward | 7'0" | 35 | 27 |
| RJ Barrett | New York Knicks | Shooting Guard | 6'6" | 37 | 24 |
| Wendell Carter Jr. | Orlando Magic | Center | 6'10" | 24 | 16 |
| Tobias Harris | Philadelphia 76ers | Power Forward | 6'8" | 38 | 26 |
| OG Anunoby | Toronto Raptors | Small Forward | 6'7" | 39 | 19 |
| Kristaps Prozingis | Washington Wizards | Center | 7'3" | 38 | 18 |
| Luka Doncic | Dallas Mavericks | Point Guard | 6'7" | 40 | 24 |
| Kentavious Caldwell-Pope | Denver Nuggets | Shooting Guard | 6'5" | 41 | 29 |
| Stephen Curry | Golden State Warriors | Point Guard | 6'2" | 28 | 21 |
| Jalen Green | Houston Rockets | Shooting Guard | 6'4" | 42 | 10 |
| Paul George | Los Angeles Clippers | Small Forward | 6'8" | 30 | 22 |
| Russell Westbrook | Los Angeles Lakers | Point Guard | 6'3" | 39 | 19 |
| Dillon Brooks | Memphis Grizzlies | Small Forward | 6'7" | 38 | 29 |
| Rudy Gobert | Minnesota Timberwolves | Center | 7'1" | 38 | 22 |
| Zion Williamson | New Orleans Pelicans | Power Forward | 6'6" | 29 | 26 |
| Josh Giddey | Oklahoma City Thunder | Shooting Guard | 6'8" | 38 | 20 |
| Chris Paul | Pheonix Suns | Point Guard | 6'0" | 26 | 21 |
| Jerami Grant | Portland Trail Blazers | Power Forward | 6'8" | 40 | 20 |
| Domantas Sabonis | Sacramento Kings | Center | 6'11" | 40 | 23 |
| Keldon Johnson | San Antonio Spurs | Small Forward | 6'5" | 37 | 13 |
| Jordan Clarkson | Utah Jazz | Shooting Guard | 6'4" | 45 | 22 |

[StatsProject\_Spence](https://eastcentraluniversity-my.sharepoint.com/:x:/r/personal/tayspe_email_ecok_edu/Documents/StatsProject_Spence.xlsx?d=wf21187c9998f439eb209047be7bff957&csf=1&web=1&e=zXDNan)

For the second part of my project, I chose to examine the frequency and relative frequency of the position these players play. Refer to the table below that shows this data:

|  |  |  |
| --- | --- | --- |
| **Players' Positions** | **Frequency** | **Relative Frequency** |
| Point Guard | 6 | 0.2 |
| Small Forward | 6 | 0.2 |
| Shooting Guard | 6 | 0.2 |
| Power Forward | 6 | 0.2 |
| Center | 6 | 0.2 |
| **Total** | **30** | **1** |

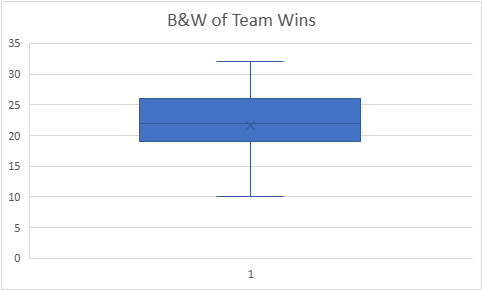
This section of the data may seem a little odd, but that’s because I chose to represent each position equally since there are 30 teams in the league and 5 positions that could be played. This made all this data equal, since there are 6 players for each position, making the relative frequency 0.2 for each.

Next, I examined the players’ position to the team they play for in the table below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Point Guard** | **Small Forward** | **Shooting Guard** | **Power Forward** | **Center** |
| **Atlanta Hawks** | 0 | 0 | 0 | 1 | 0 |
| **Boston Celtics** | 0 | 1 | 0 | 0 | 0 |
| **Brooklyn Nets** | 1 | 0 | 0 | 0 | 0 |
| **Charlotte Hornets** | 0 | 0 | 1 | 0 | 0 |
| **Chicago Bulls** | 0 | 0 | 0 | 1 | 0 |
| **Cleveland Cavaliers** | 0 | 0 | 0 | 0 | 1 |
| **Detroit Pistons** | 0 | 0 | 0 | 0 | 1 |
| **Indiana Pacers** | 1 | 0 | 0 | 0 | 0 |
| **Miami Heat** | 0 | 1 | 0 | 0 | 0 |
| **Milwaukee Bucks** | 0 | 0 | 0 | 1 | 0 |
| **New York Knicks** | 0 | 0 | 1 | 0 | 0 |
| **Orlando Magic** | 0 | 0 | 0 | 0 | 1 |
| **Philadelphia 76ers** | 0 | 0 | 0 | 1 | 0 |
| **Toronto Raptors** | 0 | 1 | 0 | 0 | 0 |
| **Washington Wizards** | 0 | 0 | 0 | 0 | 1 |
| **Dallas Mavericks** | 1 | 0 | 0 | 0 | 0 |
| **Denver Nuggets** | 0 | 0 | 1 | 0 | 0 |
| **Golden State Warriors** | 1 | 0 | 0 | 0 | 0 |
| **Houston Rockets** | 0 | 0 | 1 | 0 | 0 |
| **Los Angeles Clippers** | 0 | 1 | 0 | 0 | 0 |
| **Los Angeles Lakers** | 1 | 0 | 0 | 0 | 0 |
| **Memphis Grizzlies** | 0 | 1 | 0 | 0 | 0 |
| **Minnesota Timberwolves** | 0 | 0 | 0 | 0 | 1 |
| **New Orleans Pelicans** | 0 | 0 | 0 | 1 | 0 |
| **Oklahoma City Thunder** | 0 | 0 | 1 | 0 | 0 |
| **Pheonix Suns** | 1 | 0 | 0 | 0 | 0 |
| **Portland Trail Blazers** | 0 | 0 | 0 | 1 | 0 |
| **Sacramento Kings** | 0 | 0 | 0 | 0 | 1 |
| **San Antonio Spurs** | 0 | 1 | 0 | 0 | 0 |
| **Utah Jazz** | 0 | 0 | 1 | 0 | 0 |

Once again, this data may seem odd because I chose to represent each team with one player and each position with 6 players. Due to this there are not really many relationships to establish besides the one I already have, which is to represent each team and position equally. I do believe though that more valuable relationships can/will be established by examining the quantitative variables.

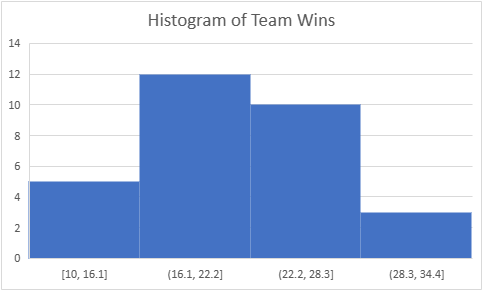
For the third part of the project, I have created a Box-and-Whisker Plot and a Histogram Chart to show the data of team wins, which is a quantitative variable.



This graph is a box-and-whisker plot that shows many different important data points. Some being the five number summary, mean, and standard deviation which I have listed below:

* Five Number Summary:
  + Minimum: 10
  + 1st Quartile: 19
  + Median: 22
  + 3rd Quartile: 26
  + Maximum: 32
* Mean: 21.56667
* Standard Deviation: 5.4118855999667

Another graph that shows important data points/descriptions is a histogram, as seen below:



In this histogram, we can confirm that there are no outliers in this data set. It can also be seen that the histogram is distributed symmetrically. When one analyzes this histogram. They can see that the majority of the players’ wins fall in the 16.1-22.2 category, but the 22.2-28.3 category is not far behind. I thought it was interesting to see that there were more players in the lowest wins category (10-16.1) than there were in the highest wins category (28.3-34.4). This can be expected though, because there are six divisions and only one leader in each of these divisions.

In the fourth part of the project, I created a hypothesis test for a quantitative and categorical variable. The quantitative variable that I used for the hypothesis test was team wins. Since this data was taken before the preseason was over, I had to shift the data just a bit. Normally, there are 82 games played by each team in the preseason so winning over 41 games would be considered successful. In my case, though, I am using 24 wins to consider the team successful. My null hypothesis is that the player’s team wins are equal to 25, but my alternative hypothesis is that the wins are over 25 games. I chose this for the hypothesis test to test how many players are having a successful season in terms of the team.

H0: µ = 25

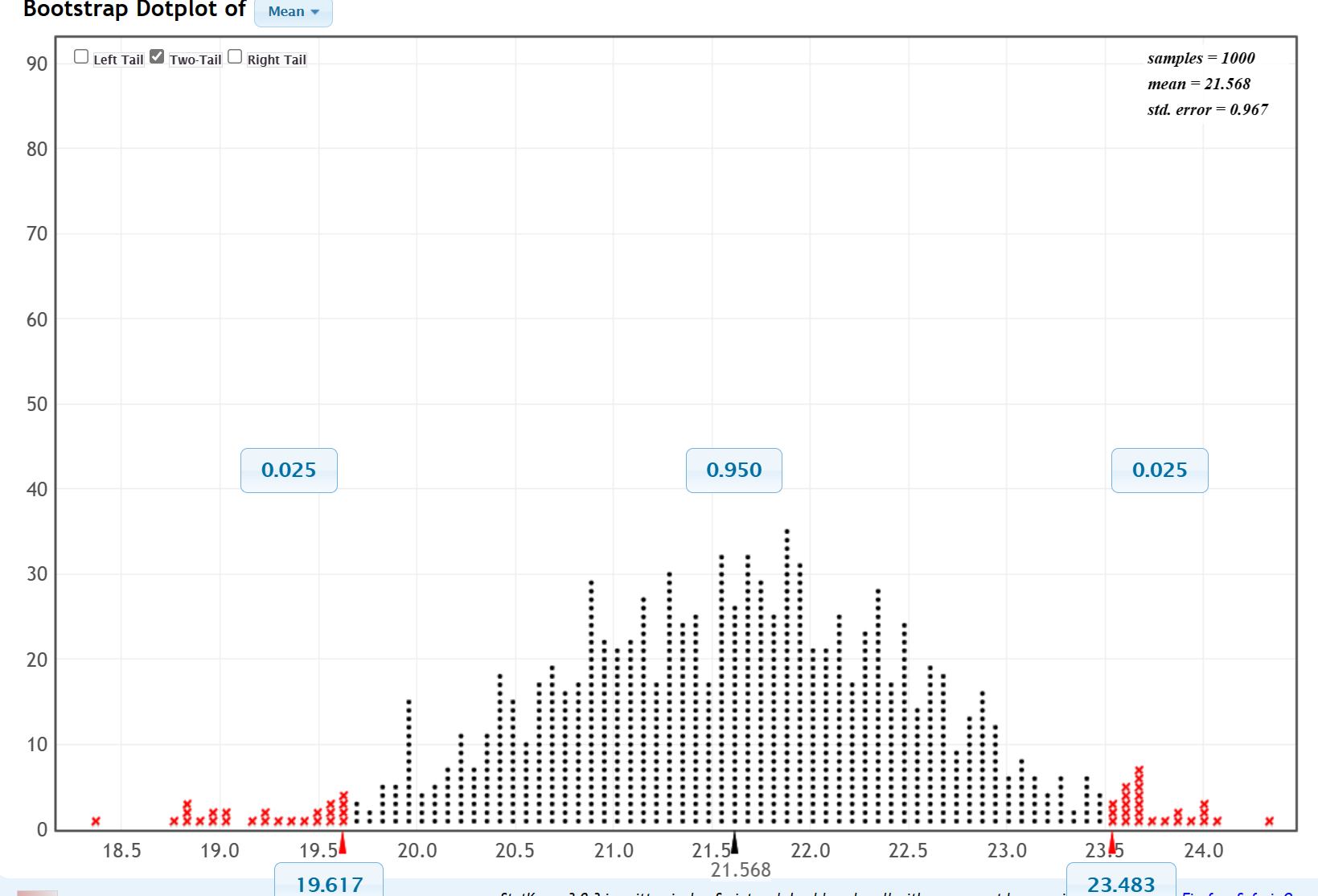
Ha: µ > 25

The next hypothesis test is for the categorical variable, player’s position. For this test, I chose to use the position, shooting guard, out of the five different positions because generally this position gets the most opportunity to score in the game because of their ability to shoot long and short-range shots. The “long-range shot” is called a “3-pointer” and given its name; it is worth 3 points whereas other shots are worth 2. So, this position has the chance to score the most due to ability and the number of points certain shots that they shoot are. So, in fantasy basketball it is smart to include at least one of these players on a person’s roster. So, I decided to create a null and alternative hypothesis to see if shooting guards are 1/5 of the players on a fantasy roster or greater than that. One of my friends, has 2 shooting guards on his team, so the alternative hypothesis is true in his case.

H0: PSG = 0.2

Ha: PSG > 0.2

For the fifth part of the project, I created a histogram below to show the bootstrap distribution for my quantitative variable, team wins.

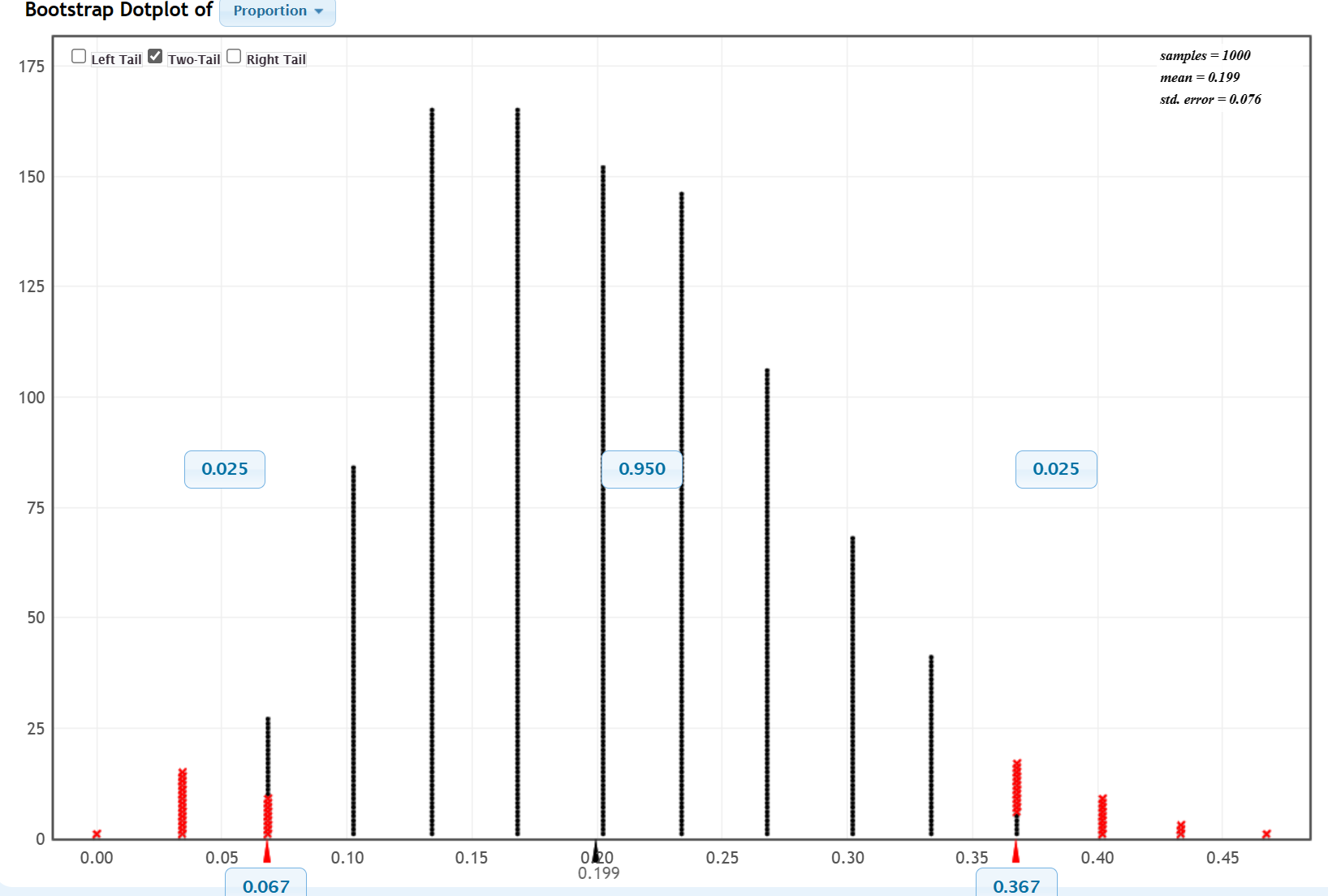


The standard error for this data set is 0.967.

The 95% confidence interval is between 19.617-23.483.

Due to this information, I can reject my null hypothesis that teams’ wins are equal to 25 since the highest part of the confidence interval is under 25.

Below is the histogram showing my categorical variable, player’s position, specifically focused on the shooting guard.



The standard error for this data set is 0.076.

The 95% confidence interval is between 0.067-0.367.

Due to this information, I fail to reject my null hypothesis that shooting guards make up 0.2 of a fantasy team since 0.2 falls within the confidence interval listed above.

For the sixth part of my project, I am repeating my hypothesis test for my categorical variable, player’s positions, and compare this to my results of the bootstrapping.

H0: PSG = 0.2

Ha: PSG > 0.2

Basketball teams have 5 different positions, and I am considering the shooting guard the most valuable since they have the most chances to score the most, so, I am testing to see if shooting guards make up more than 20% of the teams on fantasy teams.

To start, I will be finding my standard error (SE), z-value (z), followed by the 95% confidence interval. Below are my starting values and equations I used to find these values:

p=0.2 p(hat)=0.2 n=30

SE= SE==0.073

Z value= Z value==0

CI(95%)=P(hat) +/- Z\*SE CI(LOW)=0.2-(0)(0.073)=0.2

CI(HIGH)=0.2+(0)(0.073)=0.2

Based on these findings, I still can fail to reject my null hypothesis.

On to the seventh part of my project, which is repeating the hypothesis test for my quantitative variable, team wins! Below is my null hypothesis, which states that the wins will be equal to 25 wins, and my alternative hypothesis, which states that the wins will be greater than 25 wins.

H0: µ = 25

Ha: µ > 25

To repeat this hypothesis test, I used Excel so below is a chart of the variables I am looking for, the formulas I used, and the results I obtained.

|  |  |  |
| --- | --- | --- |
| MU | Hypothesis | 25 |
| X-Bar | Average=(sum/n) | 21.56667 |
| Standard Deviation | =STDEV.S(column of numbers) | 5.411886 |
| Sample Size (n) | Entire size of group | 30 |
| Standard Error (SE) | =STDEV.S/(SQRT(n)) | 0.988071 |
| T-score | =(Xbar-MU)/SE | -3.474780658 |
| Alpha | 95%CI (1-0.95) | 0.05 |
| T-critical or T\* | =T.INV(alpha/2,n-1) | -2.04523 |
| Low 95% | =(Xbar-Tcritical)\*SE | 19.52144 |
| High 95% | =(Xbar+Tcritical)\*SE | 23.6119 |

Next, I have included another table to compare this to my bootstrap results.

|  |  |  |
| --- | --- | --- |
|  | Bootstrap Results | Repeat Test Results |
| Mean | 21.568 | 21.56667 |
| Standard Error | 0.967 | 0.988071 |
| 95% Confidence Interval | 19.617-23.483 | 19.521-23.612 |

So, as you can see all my results from the bootstrap and repeated test results are remarkably similar. The only differences are by a few decimals. Given this, I can still reject my null hypothesis that teams must win games to be successful since it does not fit into either confidence interval. Instead, teams need to win more than 20 games to be considered to have a successful season.

For the final part of the project, I created two conditional probabilities. The data I used comes from a two-way table that I created in the second part of this project. This table compares a player’s position to their respective team.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Point Guard** | **Small Forward** | **Shooting Guard** | **Power Forward** | **Center** |
| **Atlanta Hawks** | 0 | 0 | 0 | 1 | 0 |
| **Boston Celtics** | 0 | 1 | 0 | 0 | 0 |
| **Brooklyn Nets** | 1 | 0 | 0 | 0 | 0 |
| **Charlotte Hornets** | 0 | 0 | 1 | 0 | 0 |
| **Chicago Bulls** | 0 | 0 | 0 | 1 | 0 |
| **Cleveland Cavaliers** | 0 | 0 | 0 | 0 | 1 |
| **Detroit Pistons** | 0 | 0 | 0 | 0 | 1 |
| **Indiana Pacers** | 1 | 0 | 0 | 0 | 0 |
| **Miami Heat** | 0 | 1 | 0 | 0 | 0 |
| **Milwaukee Bucks** | 0 | 0 | 0 | 1 | 0 |
| **New York Knicks** | 0 | 0 | 1 | 0 | 0 |
| **Orlando Magic** | 0 | 0 | 0 | 0 | 1 |
| **Philadelphia 76ers** | 0 | 0 | 0 | 1 | 0 |
| **Toronto Raptors** | 0 | 1 | 0 | 0 | 0 |
| **Washington Wizards** | 0 | 0 | 0 | 0 | 1 |
| **Dallas Mavericks** | 1 | 0 | 0 | 0 | 0 |
| **Denver Nuggets** | 0 | 0 | 1 | 0 | 0 |
| **Golden State Warriors** | 1 | 0 | 0 | 0 | 0 |
| **Houston Rockets** | 0 | 0 | 1 | 0 | 0 |
| **Los Angeles Clippers** | 0 | 1 | 0 | 0 | 0 |
| **Los Angeles Lakers** | 1 | 0 | 0 | 0 | 0 |
| **Memphis Grizzlies** | 0 | 1 | 0 | 0 | 0 |
| **Minnesota Timberwolves** | 0 | 0 | 0 | 0 | 1 |
| **New Orleans Pelicans** | 0 | 0 | 0 | 1 | 0 |
| **Oklahoma City Thunder** | 0 | 0 | 1 | 0 | 0 |
| **Pheonix Suns** | 1 | 0 | 0 | 0 | 0 |
| **Portland Trail Blazers** | 0 | 0 | 0 | 1 | 0 |
| **Sacramento Kings** | 0 | 0 | 0 | 0 | 1 |
| **San Antonio Spurs** | 0 | 1 | 0 | 0 | 0 |
| **Utah Jazz** | 0 | 0 | 1 | 0 | 0 |

I will use the formula below to calculate the conditional probabilities:

P(A|B) =

Here A refers to the position of the players while B refers to their respective teams.

So, for example, let’s consider that A is the shooting guards and B is the Dallas Mavericks. This allows for P(A) = 1/5 because there are 5 different positions and the probability of falling into one is 1/5. P(B) is 1/30 because there are 30 different teams in the NBA and the probability of falling into one is 1/30. For (AB) is the intersection of A and B with the specific categories of the position shooting guard and playing for the Dallas Mavericks which is equal to 1/150. So, to find the conditional probability of P(A|B) compute (1/150)/ (1/30) so this gives the answer 1/5. To find the conditional probability of P(B|A) compute (1/150)/ (1/5) so this gives the answer 1/30.