Final Project

Dan Brooks

April 23, 2016

Introduction

In the NBA (National Basketball Association) we hear about players earing huge contracts. We hear the top players in the league making 10's of millions of dollars to play basketball. The better the player, the more they get paid. Most players fresh out of college make the league minimum while players im their prime get the huge pay days. The questions that comes to mind is, does paying these players the large amount of money really lead to more wins? Does a team really have to spend an outrageous amoutn of money on their players to get more wins? If they don't, then they can tak ethe moeny they would be spending on the salaries and invest in other parts of the organization.

Data

Collection

The data was pulled for this project has two parts:

1. Number of wins for each team from 2007-2015
    - found at http://www.basketball-reference.com/leagues/NBA_wins.html
2. Total salary for each team from 2007-2015
    - found at https://www.eskimo.com/~pbender/misc/salaries07.txt The data was pullaed from each of these sources into R Studio. From there, the data was then transformed into a tidy data set (One case per role) and analyzed from there.

Cases/Variables

Each case in the data consist of a team, a year, a number of wins and a salary. Here is an example of what the data looks like:

```
head(data_bball)
```

```
##   Teams Year Wins Salaries
## 1   ATL 2015   60 58328957
## 2   BOS 2015   40 60403234
## 3   BRK 2015   38 88424372
## 4   CHI 2015   50 65680052
## 5   CHO 2015   33 67432537
## 6   CLE 2015   53 81377634
```

Type of Study

This is an observational study. There is not experiment that is taking place. We are simply observing the number of wins that each team had over the past 9 years comparing it to the teams salaries that they had over the same time period.

Scope of Interest-Generalizability

THe population of interest in the entire NBA, Since I am taking information from all of the teams in the entire league, this can be generalized for the entire population. A potential source is the amount of money the team has to offer. Not all of the teams have the same cap space (amount of moeny to spend of players). Some teams have a lot of money some teams do not. The teams witha lot of moeny can affore to pay their players more (they do not have to be good players). They could be making more money on a team, jsut because they have the cap space. If they would have been on any other team, then they would havce made a lot less money.

Scope of Inference - Causality

I would say that you cannot link the number of wins to the total salaries given to the players. There are so many things that result in the winning of a basketball game. There is the injuries to important players, how well the players play together on the court, the coaching decisions, referees, noe of those have to do with the amount of money a player gets paid. I would not say that the salary the players make will be a cause for the number of wins.

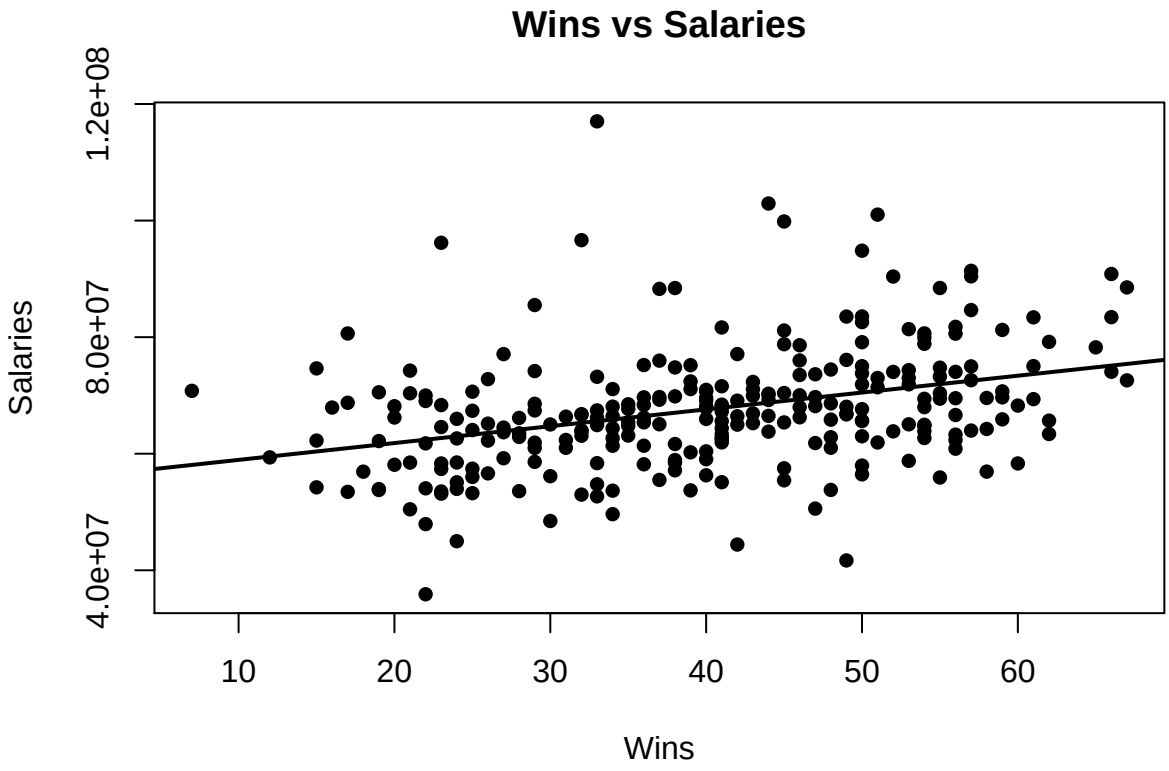
Exploratory Data Analysis

Graph

Here is a plot that shows that compares the number of wins by each NBA team and the total salaries for that team.

```
plot(x = data_bball$Wins, data_bball$Salaries, xlab = "Wins", ylab = "Salaries", pch = 16, main = "Wins

regression <- lm(Salaries ~ Wins, data = data_bball)
abline(regression, lwd = 2)
```



We can see that there is a positive slope to the graph. Than means that as the number of wins increase, the higher the salary. That is showing that the higher the salary, the more wins the team had for that year. Which means higher paid players give teams more wins.

Summary Stats

```
summary(regression)
```

```
##
## Call:
## lm(formula = Salaries ~ Wins, data = data_bball)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -28515494  -5924709   -644682   3964714  51442382
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 56047192    1983250  28.260  < 2e-16 ***
## Wins          288928      47095   6.135 3.05e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9933000 on 268 degrees of freedom
```

```
## Multiple R-squared:  0.1231, Adjusted R-squared:  0.1199
## F-statistic: 37.64 on 1 and 268 DF,  p-value: 3.046e-09
```

We can see the sumamry statsistics of the plot that is created above. The summary gives us the equation of the line as wellas the R squared. We can see that the coefficients of the line are very significant. Haveing a very small p-value. The equation of the line is as follows:

$$\hat{y} = 56,047,192 + 288,928 * Wins$$

This is telling us that with each win that a team gets, their salary should increase by $288,928 dollars. It also tells us, if a team has zero wins in a season, they will have to play their players a total of $56,047,192 dollars. That is not a bad salary for not winning a single game thorughout the season.

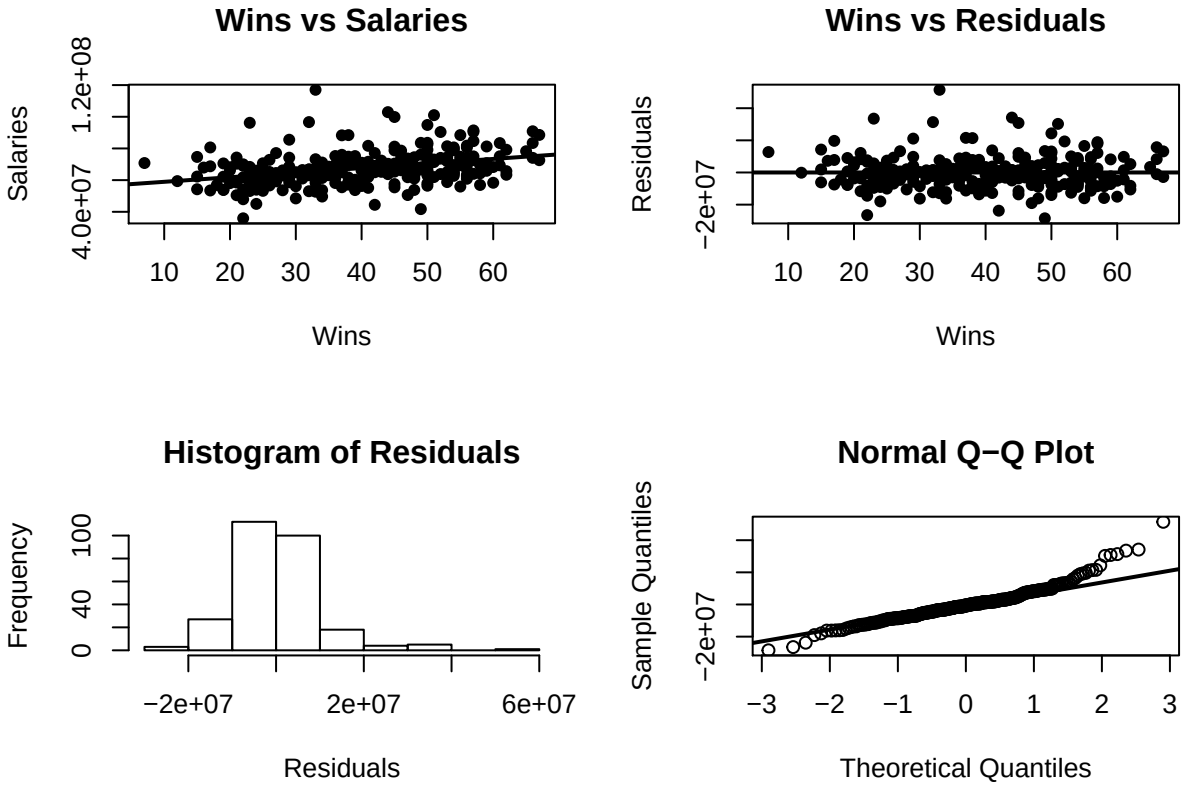
Reliability of Linear Model

```
par(mfrow=c(2,2))

plot(x = data_bball$Wins, data_bball$Salaries, xlab = "Wins", ylab = "Salaries", pch = 16, main = "Wins
abline(regression, lwd = 2)

plot(regression$residuals ~ data_bball$Wins, xlab = "Wins", ylab = "Residuals", pch = 16, main = "Wins
abline(h = 0, lwd = 2)

hist(regression$residuals, xlab = "Residuals", main = "Histogram of Residuals")
qqnorm(regression$residuals)
qqline(regression$residuals, lwd = 2)
```



There are a few conditions that have to be met in order for a graph linear graph to be reliable

1. We need to check for linearity

- We can see from the plot above (top left corner), the graph appears to be linear. There are no curves or bends in the graph. It appears to be a relative straight line

2. Nearly normal residuals

- The histogram (lower left corner) looks normal. There are a few outliers on the left of the graph, but most of the graph is contained around zero in the middle of the graph.

3. Constant variability

- The other two graphs (lower and upper right) are ploting the residuals. The top right graph is showing the residuals compared to the wins. We can see that the residuals are pretty evenly spread out around the line y = 0. They are not skewed to one side or the other. The lower right graph also shows that the residuals are rather constant. They stick to the line pretty well. There are a few outliers towards the ends of the graph but most of it contained on the line.

Statistical Inference

I am a huge Cavs fan. I grew up in Cleveland and I watch them whenever I can. I am interested in seeing how the Cavs salary from the 2015 season compares to the rest of the league. The Cavs have LeBron and

Kyrie and others that bring in hige paydays, but I want to see if they are paying hteir players more then the rest of the league. I believe that they are paying their players more then the rest of the league. We will see what the data says.

Conditions

In order for me to make this comparison, there has to be a few conditions that are met:

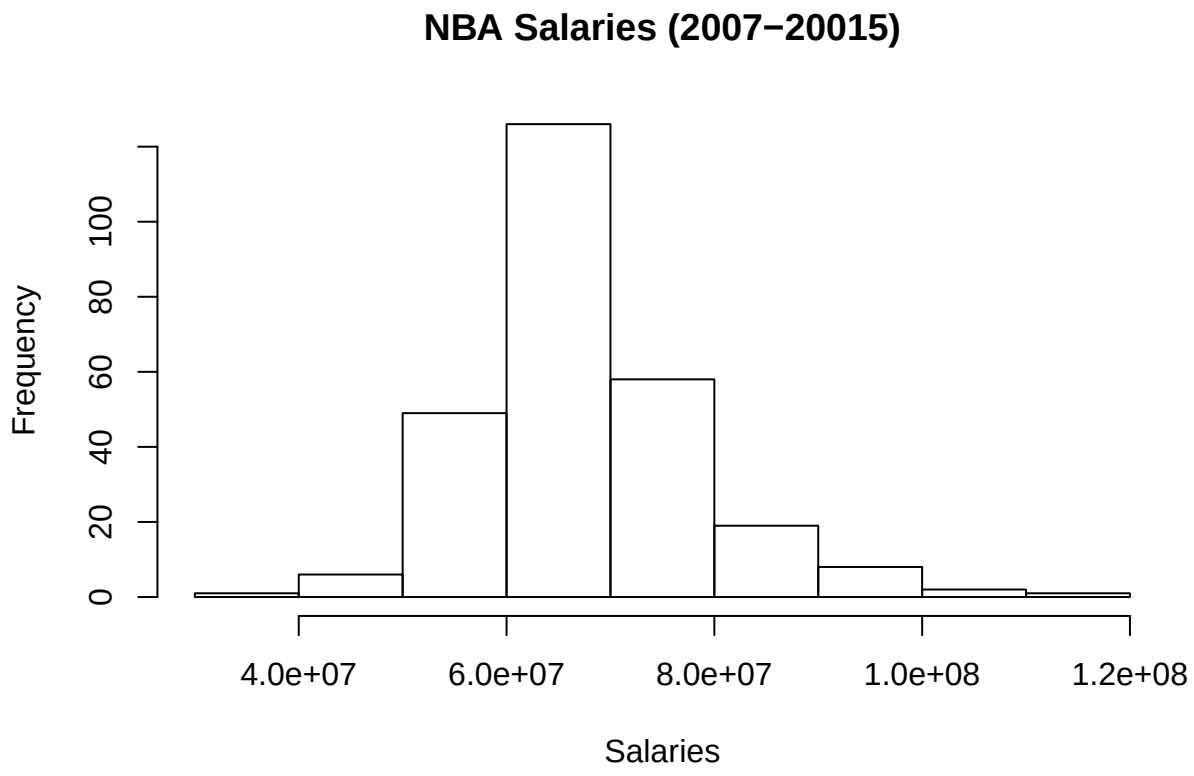
Random

The variables have to be from a random sample. These variable are picked from the teams of the NBA from random years.

Normal

The population has to be roughly normal

```
hist(tidy.salaries$Salaries, xlab  = "Salaries", main = "NBA Salaries (2007-20015)")
```



We can see by the above plot that the distributions of the salaries are approx. normal.

Independent

The population has to be large compared to the samples that are being chose. There at 270 total observations in this population and we are choosing on year out of that sample. The general rule is, the population has to be 20 times larger then the sample that is chosen. Since we are picking one team out of the 9 years of data, this condition is met.

Confidence Interval

A confidence interval allows us to estimate the population mean of a sample.

$$\bar{x} \pm 1.96 * \frac{s}{\sqrt{n}}$$

- x bar is the sample mean
- 1.96 is the z-value associated with a 95% confidence interval
- s the sample standard deviation
- n is the number of observations in the sample

```
mean <- mean(tidy.salaries$Salaries)
sd <- sd(tidy.salaries$Salaries)
n <- NROW(tidy.salaries)

posCI <- mean + 1.96 * (sd/sqrt(n))
negCI <- mean - 1.96 * (sd/sqrt(n))
```

I can be 95 % confident that the average salary of the NBA between the years of 2007-2015 is between (66372348.40, 68898327.16).

Cavs Salary

I want to see if the Cavs Salary is higher then the rest of the teams for the 2015 season. We can use a t-score to see if this is a true statement of not.

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$$H_0 : \mu_Cavs = The same as the rest of the NBA teams salaries$$

$$H_A : \mu_Cavs > The rest of the NBA teams salaries$$

```
average <- mean(subset(tidy.salaries$Salaries, tidy.salaries$Year==2015))
std <- sd(subset(tidy.salaries$Salaries, tidy.salaries$Year==2015))
n <- NROW(subset(tidy.salaries$Salaries, tidy.salaries$Year==2015))

t <- (subset(tidy.salaries$Salaries, tidy.salaries$Year==2015&tidy.salaries$Teams=="CLE")-average)/(std/
```

The t-value that is 39.50. That means that the p-vlaue is pretty much 0. We want that value to be less than .05. 0 is less than .05. That means we can reject H not and accept our alternative hypothesis. That means that the Cavs do have a higer salary then the league average for the 2015 season.

Conclusion

We can see from the graphs and information above, that the salary of the teams is a pretty good indication of the wins a team will get. It is probably not a good way to pick the exact number of wins a team will get, but it is a good indication of the trend. The more a team pays their players, the more wins they should get per season. That is probably due to the fact that more seasoned players, players who have been playing for a while in the league, will be paied more then incoming rookies. More seasoned players give better skills that cater to the NBA, which turns into more wins.

I have also found that the Cavs salary for the 2015 season is higher then that over the rest of the NBA. They pay their players very well. They have Lebron James, Kyrie Erving and Kevin Love. All very good players. They also had the 3rd best record in the NBA that year. That follows with the trend that was stated earlier, that the more you pay the more wins you get. Money does produce results in the NBA.