STEPHEN CURRY

A FINAL PROJECT REPORT SUBMITTED
IN FULFILMENT OF THE REQUIREMENTS FOR COURSE STAT 291 –
STATISTICAL COMPUTING I
DEPARTMENT OF STATISTICS OF
MIDDLE EAST TECHNICAL UNIVERSITY

GROUP 12

BY

Serhat Habil Çelik 231202
Murat Çelebi 2361186
Halil Ege Evran 2218121
Kazım Tuğşad Yar 2361590

January 2022

1-Introduction

For our project, we have chosen Stephen Curry, a worldwide famous NBA basketball player. Our choice of him is because of the fact that we are all his fans and we all know his splendid performances during NBA. In other words, it is because we are all impressed by his playing that we have chosen for our project. We want to start by introducing Stephen Curry to you.

His full name is Wardell Stephen Curry II; however, everyone calls him Stephen Curry. He started his basketball career at his high school years. As time elapsed, he developed himself and he was selected by Golden State Warriors to play at NBA. Since then he has achieved great performances in his matches, and now he is a popular NBA Star. NBA is an abbreviation of National Basketball Association, established in 1945. He has been a player of Golden State Warriors since 2009. He has had numerous achievements at Golden State Warriors. Between 2014 and 2015, he was selected the best player in that season, and he brought championship to his team by beating Cleveland. He is still a very good player and helps his team win.

In this project, we want to find out his contribution to his team. Thanks to the data available to us, we have tried to answer some questions about Stephen Curry and our target is to determine how effectively Stephen Curry plays in the matches. One of these questions concerns the relationship of the total points that Stephen Curry scored with the matches won and lost. Firstly, we have used lots of pie charts, bar charts and summary tables to show some percentages regarding Stephen Curry. Because our data consists of many continuous variable, pie charts are so efficient to visualize data variables. We have also used a scatter plot to show the correlation between some variables that will be explained in the following pages of our report and the box plot that we will use to represent statistics. Secondly, we have applied some t-tests and simple linear regression and multiple linear regression to show relationships between some variables.

1.1 Data Description

Link of the data studied is https://data.world/datatouille/stephen-curry-stats. Tristan Malherbe prepared this data 4 years ago. This data consists of 879 rows and 21 columns, four of which are categorical variables. They are the dates of the matches in which Curry either played or

did not between 2009 and 2018; his opponents; the type of the matches with the names of NBA Finals, Regular Season and Conference; and results of the matches. The rest is continuous variables in this data. We are going to work on it as a group.

1.2 Research Questions

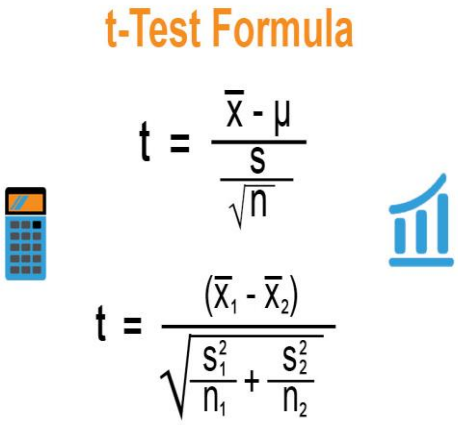
─ What is the summary of variables?
─ What is the percentage of matches won and matches lost?
─ What is the percentage of successful free throws and unsuccessful free throws?
─ What is the percentage of successful and unsuccessful shots?
─ Is there any relationships between assist numbers and minutes he played?
─ Is there any relationships between assists and points?
─ How much he contributed to his team at the matches more than 35 minutes against Cleveland?
─ Is there any relationships between total points, total 3 points and total free throws?
─ What is the relationships between total points between the matches won and matches lost?
─ Is there any relationships between time-out, minutes and personal foul?

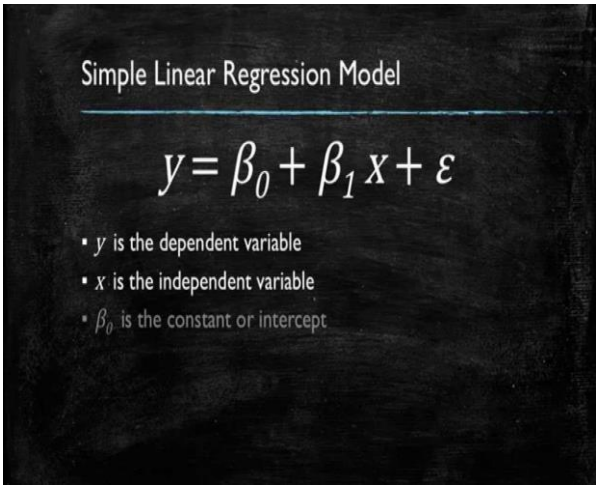
1.3 Aim of the study

Our main purpose in this study is that is how effective Stephen Curry played at the matches and whether he has an important role of making his team win or not. We actually find out the latter. If he has a big role for his team, what is he doing for that? We wanted to see some results about winning percentage, free throw percentage etc… By looking at these result, we could make some inferences on Stephen Curry.

2. Methodology/Analysis

We applied one t-test, two multiple regressions and one simple linear regression so as to analyze our data. Their formulas will be given below.



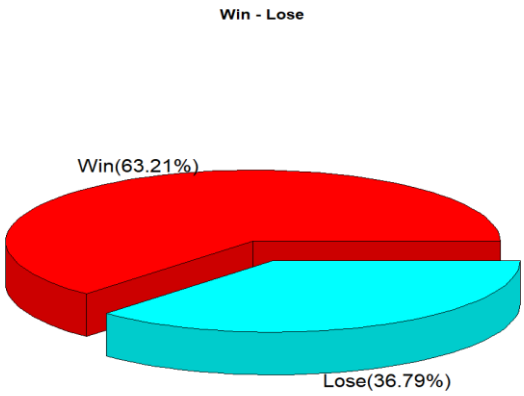


Moreover, we used lots of pie charts, summary table, box plot and scatter plot. Thanks to them, we easily analyzed our values in data studied.

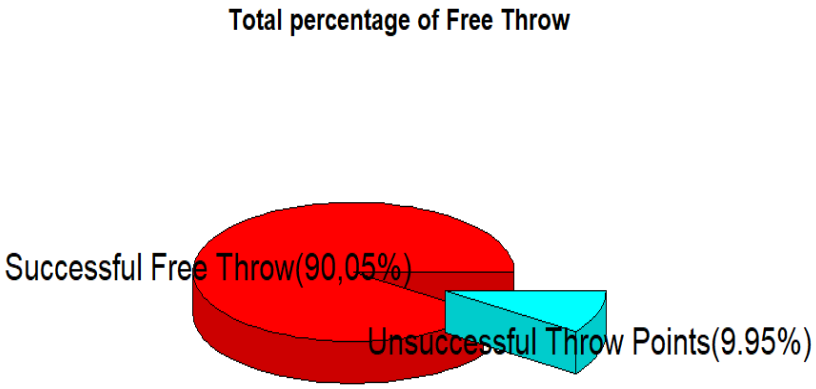
3.Result and Findings

STEPHEN CURRY	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
Minutes	0.0	28.0	34.0	29.9	38.0	58.0
Successful Shoots	0.0	4.250	7.000	7.006	10.000	20.000
Total Shoots	0.0	11.00	16.00	14.74	20.00	35.00
3 Points Successful	0.0	1.00	3.00	3.06	5.00	13.00
Total 3 Points	0.0	4.000	7.000	7.059	10.000	19.000
Successful Free Throw	0.0	0.000	3.000	3.233	5.000	16.000
Total Free Throw	0.0	0.000	3.000	3.591	6.000	16.000
Rebound	0.0	2.000	.4.000	3.902	6.000	14.000
Assists	0.0	4.00	6.00	5.78	8.00	16.00
Block	0.0	0.000	0.000	0.1913	0.000	3.000
Steal	0.0	0.000	1.000	1.525	2.000	7.000
Personal Faul	0.0	1.000	2.000	2.148	3.000	6.000
Turnover	0.0	1.000	3.000	2.759	4.000	11.000
Points	0.0	13.00	22.00	20.31	28.00	54.00
Home Score	70.0	99.0	108.0	108.5	117.0	196.0
Opponent Score	63	95	104	104	112	182

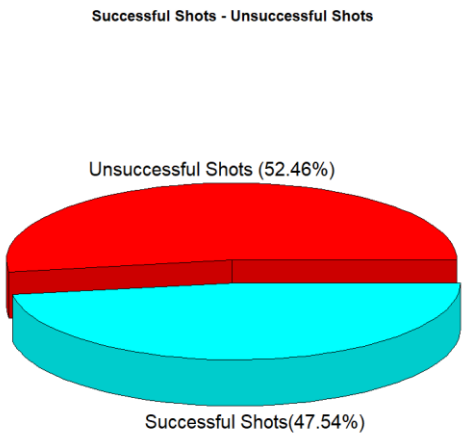
This table shows that summary of continuous variables in our data. We can interpret minimum-maximum values of our variables in this table.



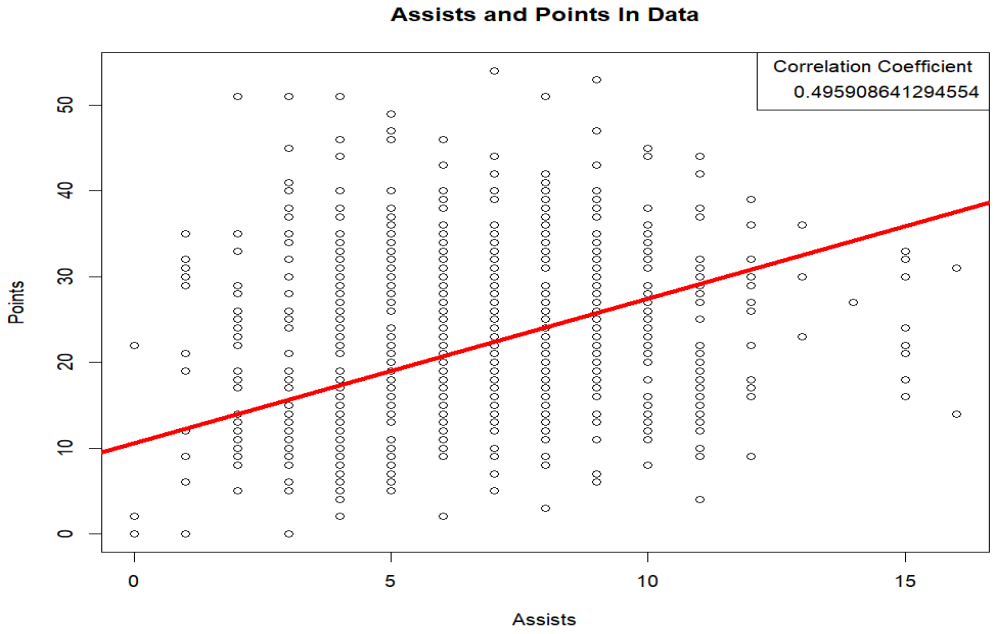
This pie chart shows win percentage of matches of Curry's total match. According to this pie chart, Stephen Curry won 63.21% of his matches played and lost 36.79% of his matches played. We can see that there is a significant difference between matches lost and matches won. He won 555 matches and he lost 323 matches.



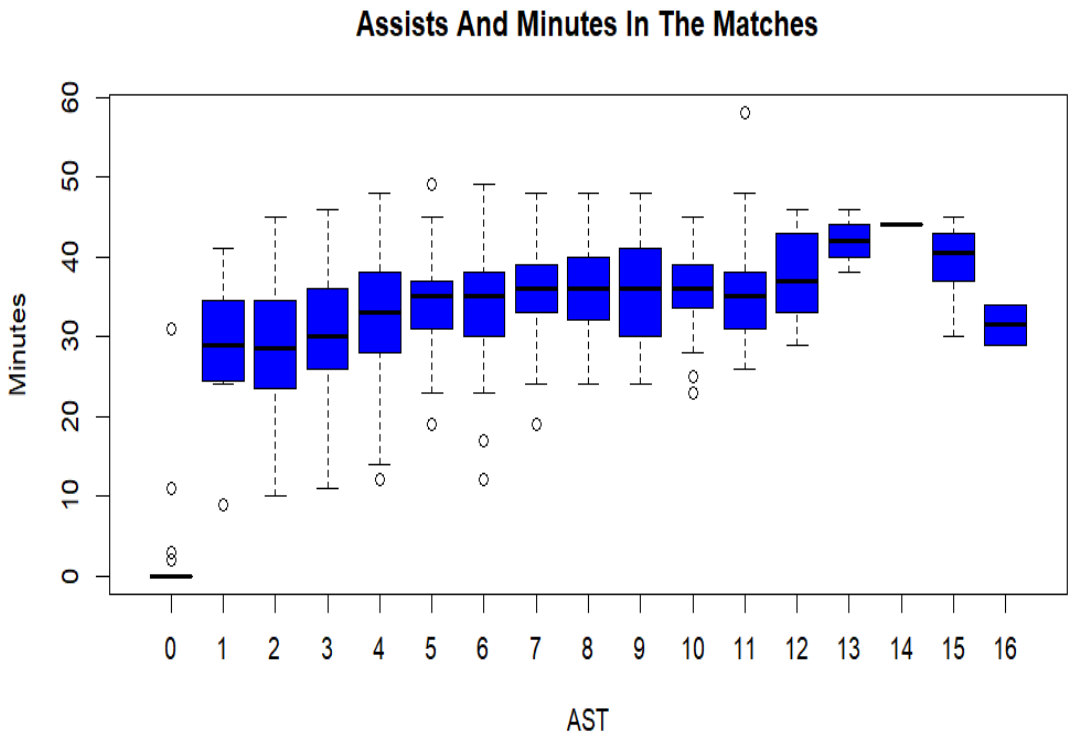
When we look at the pie chart above, we can see that 90.05% of his free throw is successful. He just missed 9.95 percentage of his free throw. We can understand that Stephen Curry helps his team in terms of points.



Curry's total shots is 12941. This pie chart shows that the percentage of unsuccessful shots is 52.48%. We can conclude that there is no significant difference between successful and unsuccessful shots Curry used.

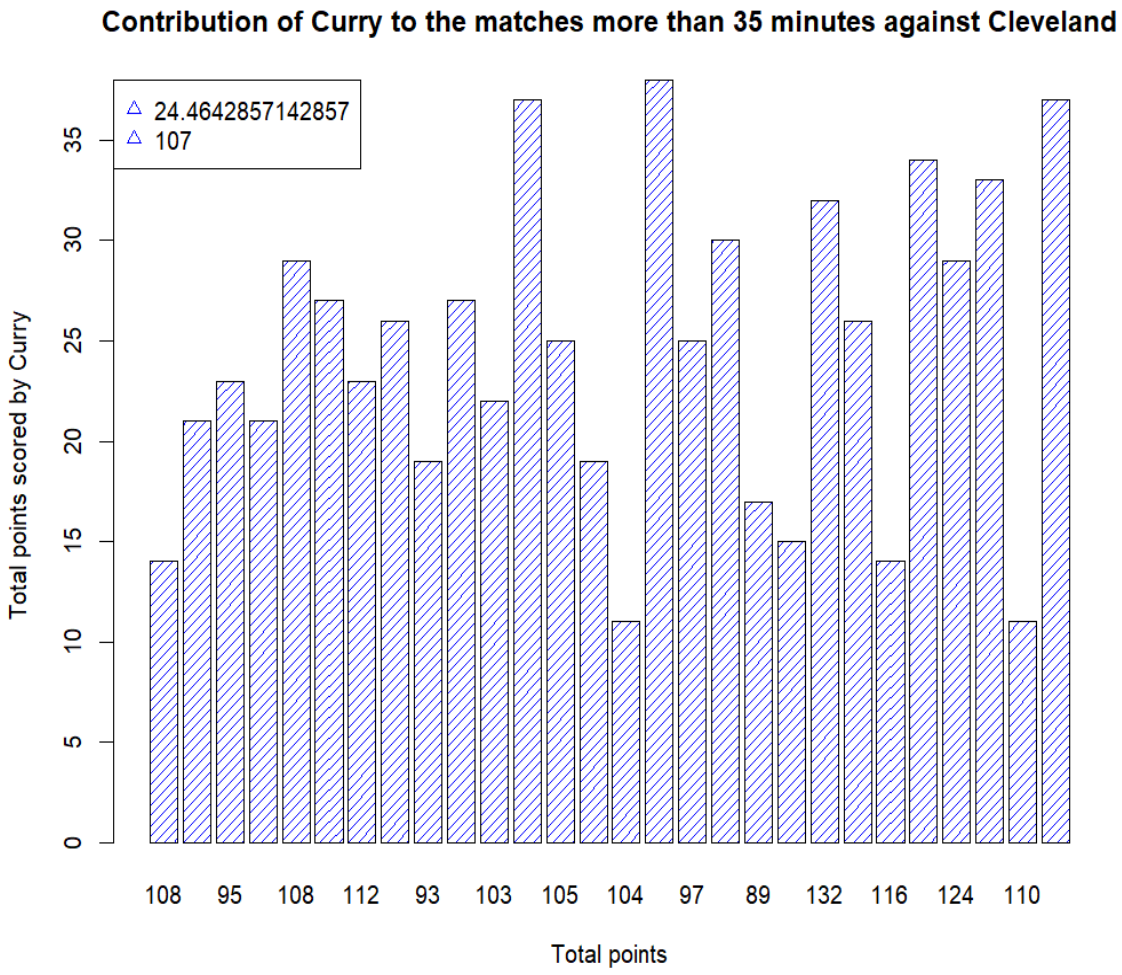


The Correlation Coefficient is 0.495 between assists and points in data studied. We can conclude that there is a positive linear relationship between them. As the number of assists increases, number of points increases.



When the correlation coefficient is calculated by R –Studio between assists and minutes in the matches, it is found as 0.68

So there is a strong positive relationship between these variables. And when we look at our boxplot, we can say that mean of Minutes that Curry played is spread between 25-30. When we think total match duration is 48 minutes in NBA, Stephen Curry almost play 30 minutes. We can conclude that the team cannot give up Stephen Curry.



In this bar charts above, we can see the contribution of Stephen Curry to the matches more than 35 minutes against Cleveland. While Golden State Warriors scored an average of 107 points in these games, Stephen Curry scored an average of 24.46. He scored also more than 30 points in 6 games. It means that almost 22 % of total points comes from Stephen Curry. Actually it could be concluded that without Stephen Curry, Golden State Warriors is nothing.

AGAINST CLEVELAND

Result	L	W	Sum
PTS			
0	0	3	3
11	0	2	2
14	2	0	2
15	1	1	2
17	1	0	1
18	1	1	2
19	2	1	3
20	0	1	1
21	1	1	2
22	0	1	1
23	0	3	3
25	1	1	2
26	0	2	2
27	2	0	2
28	0	1	1
29	0	2	2
30	1	0	1
32	0	1	1
33	0	1	1
34	0	1	1
35	0	1	1
37	0	2	2
38	0	1	1
Sum	12	27	39

AGAINST UTAH

Result	L	W	Sum
PTS			
0	6	1	7
2	0	1	1
4	0	1	1
8	0	1	1
10	1	0	1
12	0	1	1
13	1	0	1
14	1	0	1
15	1	1	2
16	0	1	1
17	1	0	1
20	0	1	1
22	1	2	3
23	0	3	3
24	0	1	1
25	0	1	1
26	0	2	2
27	0	2	2
28	1	0	1
29	1	1	2
30	0	1	1
31	0	3	3
32	1	0	1
44	0	1	1
Sum	15	25	40

When we compare Cleveland and Utah crosstab table made by R, we can conclude that Curry wins more matches against Cleveland. This is because when we 27 divide 39, it is 0.69, on the other hand, when we 25 divide 40, it is equal to 0.625. Stephen Curry must be playing more careful against Cleveland.

```
        Two Sample t-test

data:  W and L
t = 4.1305, df = 398, p-value = 4.413e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.59927 7.32073
sample estimates:
mean of x mean of y
   21.465    16.505
```

H0: μ W – μ L <= 0

HA: μ W – μ L > 0

In this t-test, as our alpha value is smaller than our p- value, we reject null hypothesis. Consequently, Stephen Curry scored in the matches won more than in the matches lost.

```
Call:
lm(formula = PTS ~ Total.3.Points + Total.FT, data = h)

Residuals:
     Min       1Q   Median       3Q      Max
-21.0910  -3.1045  -0.8806   3.7363  19.1928

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.10453    0.38757    8.01 3.64e-15 ***
Total.3.Points  1.83461    0.05117   35.86  < 2e-16 ***
Total.FT        1.18341    0.06752   17.53  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.844 on 875 degrees of freedom
Multiple R-squared:  0.7541,    Adjusted R-squared:  0.7535
F-statistic:  1342 on 2 and 875 DF,  p-value: < 2.2e-16
```

B0^= The estimated mean of points 3.10 when total 3 points and total free throw is 0

B1^= We expect the points to increase by 1.83461 for a unit increase in total 3 points when Total free throw is held constant.

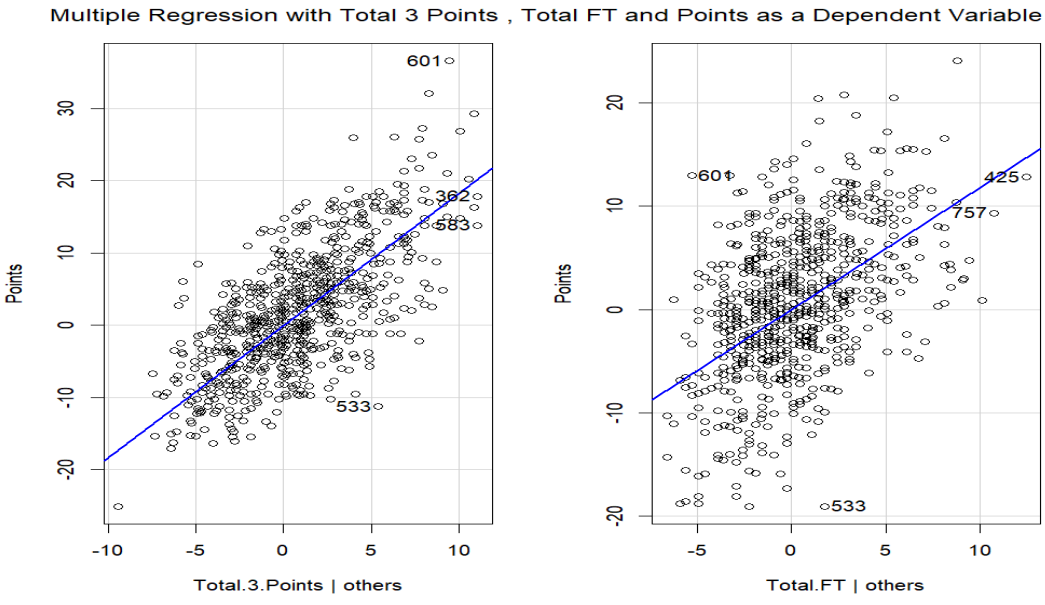
B2^= The expected increase in points is 1.18 corresponding to one-unit increase in the Total free throw for a fixed value of total three points.

P value =2.2e-16 < 0.05 So, at least one of the regressors is significant at 5% significance level.

For testing H_0:B_1=0 vs H_1:B_1 is not equal to 0, the p-value is 2e-16 < 0.05 So, total three points has significant contribution to the model.

Also, for testing H_0:B_2=0 vs H_1:$B_2$is not equal to 0, the p-value is 2e-16 < 0.05 So, total free throws have significant contribution to the model.

R^2adj =0.7541 So, almost 75% of the variation in points can be explained by the model with total three points and total free throw.



```
Call:
lm(formula = TO ~ Minutes + PF, data = h)

Residuals:
    Min      1Q  Median      3Q     Max
-3.7174 -1.1691 -0.1085  0.8736  7.4979

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.108534   0.140816   0.771   0.4411
Minutes     0.081682   0.005026  16.253   <2e-16 ***
PF          0.096530   0.041194   2.343   0.0193 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.668 on 875 degrees of freedom
Multiple R-squared:  0.3253,    Adjusted R-squared:  0.3238
F-statistic:   211 on 2 and 875 DF,  p-value: < 2.2e-16
```

B0^= The estimated mean of turn over 0.108534 when minutes and personal foul is 0

B1^= We expect the turn over increase by 0.081682 for a unit increase in minutes' points when Personal faul is held constant.

B2^= The expected increase in turnover is 0.096530 corresponding to one-unit increase in the Personal faul for a fixed value of minutes.

P value =2.2e-16 < 0.05 So, at least one of the regressors is significant at 5% significance level.

For testing H_0:B_1=0 vs H_1:B_1 is not equal to 0, the p-value is 2e-16 < 0.05 So, minutes has significant contribution to the model.

Also, for testing H_0:B_2=0 vs H_1:B_2 is not equal to 0, the p-value is 2e-16 < 0.05 So, personal faul has significant contribution to the model.

R^2adj =0.3238 So, almost 32% of the variation in points can be explained by the model with total three points and total free throw.

4.Discussion/Conclusion

Stephen Curry is a very outstanding player in NBA. He has lots of championships and medals. When we analyzed Stephen Curry, we were all excited. First, we started to think about what we could learn about Stephen Curry. We realized that there are lots of things to learn about Curry. We first looked at the percentages of win and lose. He won almost 64% percentage of his matches. It is an unbelievable statistic. After that, we decided to use a pie-chart in our report. It is a very good tool for visualization. Then, we decided to test some values in our data studied by us. Accordingly, we applied t-test and linear regression models to find out relationships between our variables. After doing this, we realized that Stephen Curry is a very valuable player in Golden State Warrior. For instance, he just scored 44 points against UTAH. Moreover, there are lots of positive linear relationships between some variables such as the minutes and points of the assists. To conclude, Stephen Curry has contributed to his team very much by looking at the above statistics. He has helped his team win the matches. Moreover, he proved to be the MVP in most of the matches in which he played.

REFERENCES

Malherbe, T. (2018, October 31). Stephen Curry Stats. Retrieved From

https://data.world/datatouille/stephen-curry-stats

APPENDICES

Data

```
h <- read.csv("Stephen Curry Stats.csv")
```

3D Pie Charts

```
install.packages("plotrix")

library("plotrix")

sum(h$Result=="L")

sum(h$Result=="W")
```

```
a <- c(555,323)

lbl <- c("Win(63.21%)","Lose(36.79%)")

pie3D(a,labels = lbl,explode = 0.1, main="Win - Lose")

sum(h$Successful.Shots)

sum(h$Total.Shots)

x <- c(6790,6151)

lbl5 <- c("Unsuccessful Shots (52.46%)","Successful Shots(47.54%)")

pie3D(x,labels = lbl5,explode = 0.1, main="Successful Shots – Unsuccessful
      Shots")sum(h$Total.FT) # 3153

sum(h$Successful.FT) #2839

d <- c(2839,314)

lbl4 <- c("Successful Free Throw(90,05%)","Unsuccessful Throw Points(9.95%)")

pie3D(d,labels = lbl4,explode = 0.3,main="Total percentage of Free Throw")
```

T-Test

```
#Hypothesis testing #1:

# H0: μW μL <= 0

# HA: μW – μL > 0

qW <- subset(h,select = c(PTS),subset = Result=="W")
```

```
qL <- subset(h,select = c(PTS),subset = Result=="L")

list=as.list(qW)

list2=as.list(qL)

vec=unlist(list)

vec1=unlist(list2)

W <- sample(vec,200,replace = F)

L <- sample(vec1,200,replace = F)

t_test0 <- t.test(W,L,var.equal = T,alternative = "two.sided")

t_test0
```

Linear Regressions

```
MLR2 <- lm(TO ~ Minutes + PF,data=h)

summary(MLR2)

MLR1 <- lm(PTS~ Total.3.Points+ Total.FT,data=h)

summary(MLR1)
```

Bar Plot

```
A <- subset(h,select = c(Opponent,PTS,Score.GS),subset = Minutes>=35)

B <- subset(A,select=c(PTS),subset=Opponent=="CLE")

E <- subset(A,select=c(Score.GS),subset=Opponent=="CLE")
```

```
BB <- sum(B)/length(unlist(B))

CC <- sum(E)/length(unlist(E))

barplot(unlist(B),ylab = "Total points scored by Curry",xlab = "Total points",names.arg =
unlist(E),col = "Blue",

    main="Contribution of Curry to the matches more than 35 minutes against
Cleveland",density = 20)

legend("topleft",legend=c(BB,CC),pch=c(2,2),col=c("Blue","Blue"))
```

Box Plot

```
vvv <- subset(h,select = c(Minutes,AST))
cor(vvv)
boxplot(Minutes~AST,data = vvv,col="blue",main="Assists And Minutes In The Matches")
```

Cross Tab

```
source("http://pcwww.liv.ac.uk/~william/R/crosstab.r")
aaa <- subset(h,select = c(PTS,AST,Result),subset =Opponent=="UTAH" )
bbb <- subset(h,select = c(PTS,AST,Result),subset =Opponent=="CLE" )
crosstab(aaa,row.vars = "PTS",col.vars = "Result",type = "f")
crosstab(bbb,row.vars = "PTS",col.vars = "Result",type = "f")
```

Correlation Graph

```
cor1 <- cor(h$AST,h$PTS)
hh <- lm(h$PTS~h$AST,data = h)
plot(h$AST,h$PTS,
    xlim = c(min(h$AST), max(h$AST)),
    ylim = c(min(h$PTS), max(h$PTS)),
    main = "Assists and Points In Data", xlab = "Assists", ylab = "Points")
abline(hh,lwd=4,lty=1,col = "Red")
legend("topright",legend=cor1,title="Correlation Coefficient")
```