Patient Times

Table of Contents

# Step:1

We have a data set that doesn’t give too much information ;- we have the average\_time which is the average time spent by the patients in the department this is measured in minutes. We have the patients\_2hrs this is the percentage of patients for the day whom spend less than 2 hrs in the department.

We created extra two columns ; - average\_time\_selection which is the average time less than 120minutes is given 1 and average time more than 120 minutes is represented by 0. We also have patients\_2hrs\_selection which is the percentage number of patients less than 50 who spend 2hrs in th department is given 1 and 0 for otherwise.

Read and inspect the data set. Provide a descriptive analysis for each of the variables in the data set.

* **Anwser**

# Load packages  
library("plotly")

## Loading required package: ggplot2

##   
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':  
##   
## last\_plot

## The following object is masked from 'package:stats':  
##   
## filter

## The following object is masked from 'package:graphics':  
##   
## layout

library("tidyverse")

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v tibble 3.1.6 v dplyr 1.0.8  
## v tidyr 1.1.4 v stringr 1.4.0  
## v readr 2.1.1 v forcats 0.5.1  
## v purrr 0.3.4

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks plotly::filter(), stats::filter()  
## x dplyr::lag() masks stats::lag()

library("data.table")

##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':  
##   
## between, first, last

## The following object is masked from 'package:purrr':  
##   
## transpose

library("dplyr")  
library("gridExtra")

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

library("knitr")  
library("scales")

##   
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':  
##   
## discard

## The following object is masked from 'package:readr':  
##   
## col\_factor

# Load Cloud Data  
df.al <- readxl::read\_xlsx("DATA Times in Hospital (1).xlsx")  
df.al <- df.al[,-1]  
df.al$patients\_2hrs\_selection <- as.factor(df.al$patients\_2hrs\_selection)  
df.al$average\_time\_selection <- as.factor(df.al$average\_time\_selection)  
# diplay data  
head(df.al,26)

## # A tibble: 25 x 4  
## average\_time patients\_2hrs average\_time\_selection patients\_2hrs\_selection  
## <dbl> <dbl> <fct> <fct>   
## 1 132. 53.3 0 0   
## 2 143. 45.9 0 1   
## 3 111. 57.9 1 0   
## 4 128. 51.7 0 0   
## 5 124. 54.5 0 0   
## 6 123. 56.7 0 0   
## 7 146. 38.5 0 1   
## 8 137. 49.0 0 1   
## 9 134. 49.7 0 1   
## 10 144. 47.0 0 1   
## # ... with 15 more rows

# reveal cloud.csv data analysis  
str(df.al)

## tibble [25 x 4] (S3: tbl\_df/tbl/data.frame)  
## $ average\_time : num [1:25] 132 143 111 128 124 ...  
## $ patients\_2hrs : num [1:25] 53.3 45.9 57.9 51.7 54.5 ...  
## $ average\_time\_selection : Factor w/ 2 levels "0","1": 1 1 2 1 1 1 1 1 1 1 ...  
## $ patients\_2hrs\_selection: Factor w/ 2 levels "0","1": 1 2 1 1 1 1 2 2 2 2 ...

# descriptive analysis of cloud.csv  
print("Summary of average time spent by each patients in a day at the department in minutes")

## [1] "Summary of average time spent by each patients in a day at the department in minutes"

summary(df.al$average\_time)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 103.2 123.6 130.4 129.4 137.1 146.0

print("Summary of percentage number of Patients whom spend less than 2 hours in a day")

## [1] "Summary of percentage number of Patients whom spend less than 2 hours in a day"

summary(df.al$patients\_2hrs)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 38.48 49.24 51.64 51.90 54.05 62.47

The Maximum average time spent by patients in the hospital Outpatients deptartment is 146 minutes , the minimum is 103 minutes , while the median and mode are 130 minutes and 129 minutes respectively. This indicates that most patients spend more than two hours on an average day in the hopital department.

Again the median and mode of patients that spend less than 2 hours in day was as high as 62% for some days and as low as 38% for some days the mode and median indicates that most patients or half of the patients visiting the hospital department spend most spend more than 2 hours at the hospital facility.

# Step:2

We need to fit a model that will keep the time patients spend in hospital outpatients department to less than 2hours.

We use the Linear Model functions to test both variables against easch other , since we have very little information and no outline dependent variable we can troubleshoot to find a model that fits the information given best.

print("Model fit of percentage number of Patients whom spend less than 2 hours in a day")

## [1] "Model fit of percentage number of Patients whom spend less than 2 hours in a day"

fit.1 <- lm(patients\_2hrs\_selection == "1" ~ average\_time , data = df.al )  
summary(fit.1)

##   
## Call:  
## lm(formula = patients\_2hrs\_selection == "1" ~ average\_time, data = df.al)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.50598 -0.26351 -0.00958 0.27868 0.55870   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -4.339672 0.835071 -5.197 2.87e-05 \*\*\*  
## average\_time 0.036629 0.006434 5.693 8.50e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.329 on 23 degrees of freedom  
## Multiple R-squared: 0.585, Adjusted R-squared: 0.5669   
## F-statistic: 32.42 on 1 and 23 DF, p-value: 8.497e-06

print("Model fit for Average time spent by Patients in a day at the Outpaitents Hospital Department")

## [1] "Model fit for Average time spent by Patients in a day at the Outpaitents Hospital Department"

fit.2 <- lm(average\_time\_selection == "1" ~ patients\_2hrs , data = df.al)  
summary(fit.2)

##   
## Call:  
## lm(formula = average\_time\_selection == "1" ~ patients\_2hrs, data = df.al)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.32655 -0.18169 -0.03261 0.02173 0.62275   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.11979 0.57375 -3.695 0.001197 \*\*   
## patients\_2hrs 0.04315 0.01101 3.920 0.000686 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2623 on 23 degrees of freedom  
## Multiple R-squared: 0.4005, Adjusted R-squared: 0.3745   
## F-statistic: 15.37 on 1 and 23 DF, p-value: 0.0006858

fit.1 model has an Adjusted R-Squared of 58.5%, suggesting that the model can explain the data well enough. We might also interested in seeing how good the model will be on the testing dataset.

fit.2 model has an Adjusted R-Squared of 40.5%, suggesting that the model can’t explain the data well enough as fit.1.

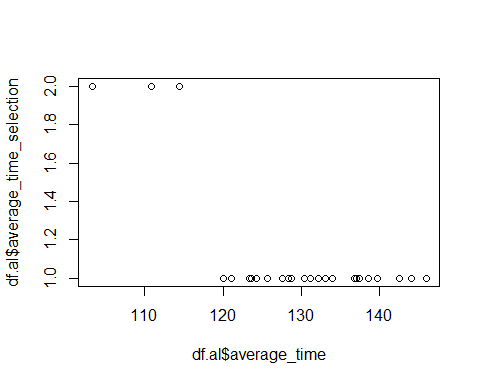
# Step:3

plot models

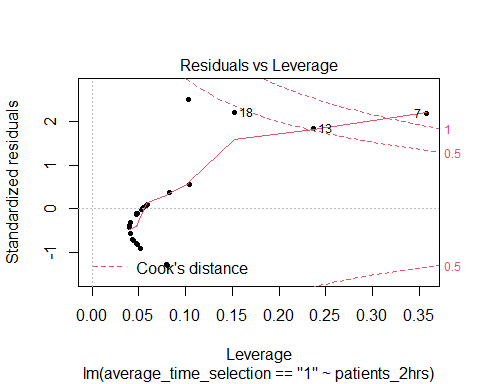
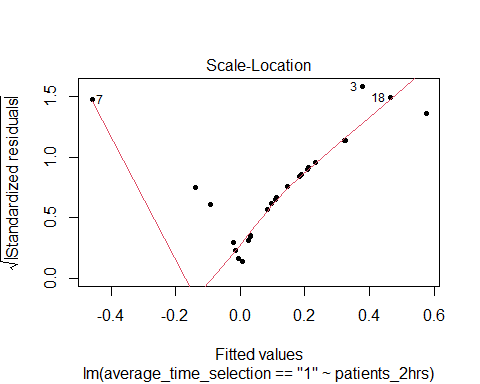
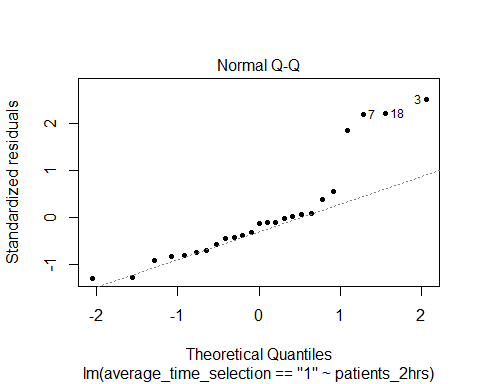
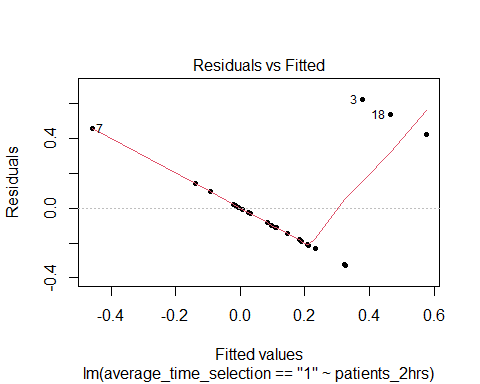
print("Plots of Model fit for Average time spent by Patients in a day at the Outpaitents Hospital Department")

## [1] "Plots of Model fit for Average time spent by Patients in a day at the Outpaitents Hospital Department"

plot(x = df.al$average\_time , y = df.al$average\_time\_selection)



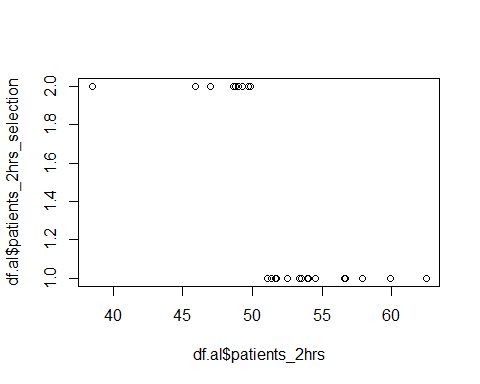
plot(fit.2,pch=20)



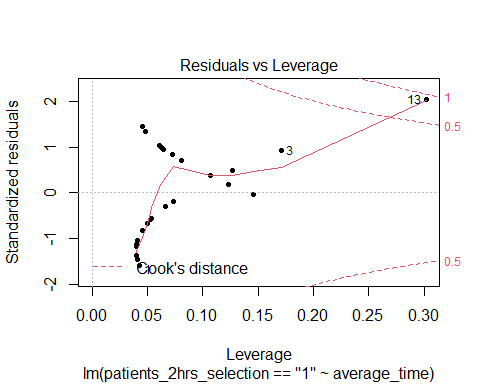
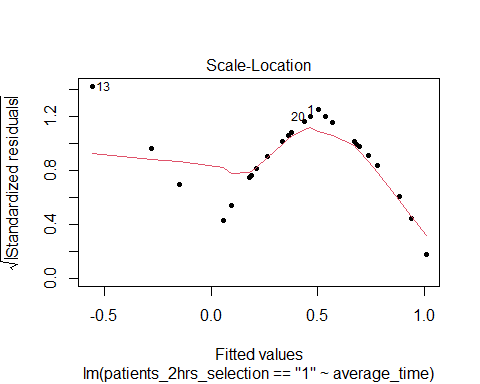
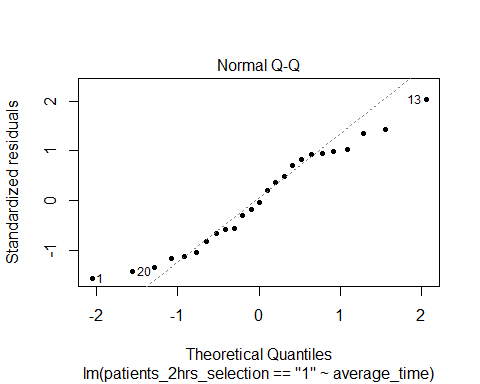
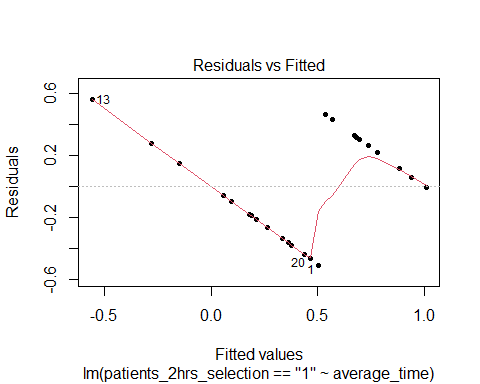
print("Plots of Model fit for percentage number of Patients whom spend less than 2 hours in a day")

## [1] "Plots of Model fit for percentage number of Patients whom spend less than 2 hours in a day"

plot(x = df.al$patients\_2hrs , y = df.al$patients\_2hrs\_selection)



plot(fit.1, pch=20)



# Step:4

Perform white test using fit.2

#load lmtest library  
library(lmtest)

## Warning: package 'lmtest' was built under R version 4.1.3

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 4.1.3

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

#perform White's test  
bptest(fit.2, ~ patients\_2hrs + I(patients\_2hrs^2), data = df.al)

##   
## studentized Breusch-Pagan test  
##   
## data: fit.2  
## BP = 16.145, df = 2, p-value = 0.000312

Here is how to interpret the output: The test statistic is X2 = 16.145. The degrees of freedom is 2. The corresponding p-value is 0.000312.

Null (H0): Homoscedasticity is present. Alternative (HA): Heteroscedasticity is present.

Since the p-value is less than 0.05, we reject the null hypothesis. We have sufficient evidence to accept the Alternative that there is Heteroscedasticity present. When this assumption is violated, we say that heteroscedasticity is present in the residuals. When this occurs, the results of the regression become unreliable

# Step:5

Perform white test using fit.1

#load lmtest library  
library(lmtest)  
  
#perform White's test  
bptest(fit.1, ~ average\_time + I(average\_time^2), data = df.al)

##   
## studentized Breusch-Pagan test  
##   
## data: fit.1  
## BP = 1.1694, df = 2, p-value = 0.5573

Here is how to interpret the output: The test statistic is X2 = 1.1694. The degrees of freedom is 2. The corresponding p-value is 0.5573.

Null (H0): Homoscedasticity is present. Alternative (HA): Heteroscedasticity is present.

Since the p-value is not less than 0.05, we fail to reject the null hypothesis. We do not have sufficient evidence to say that heteroscedasticity is present in the regression model.