ALY 6015 M2 Report - Thota, Sunil Raj.R

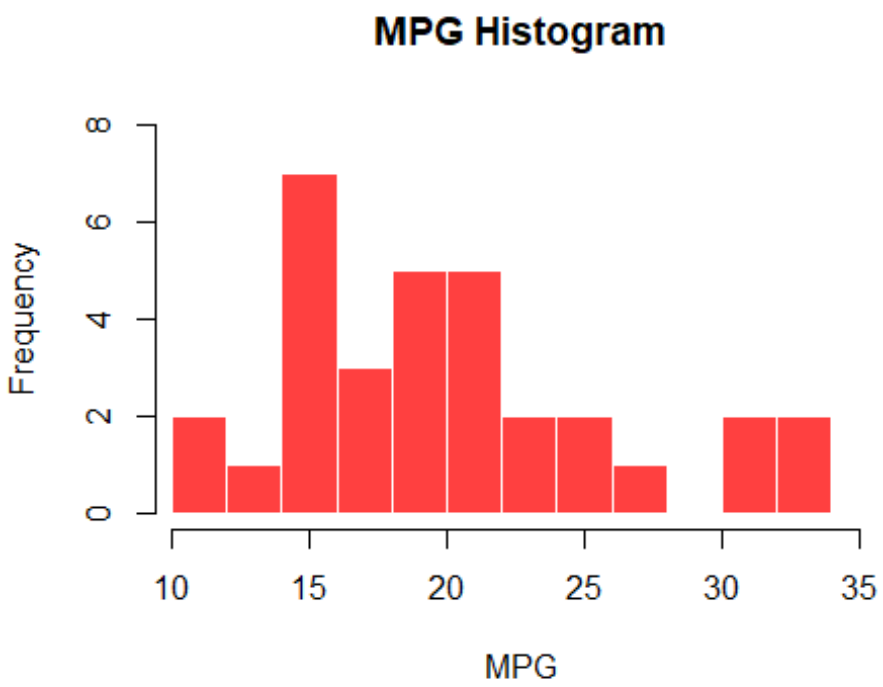
```
# Intermediate Analytics
# ALY 6015
# Module 1 - Hypothesis Testing with R
# 01/30/2021
# Sunil Raj Thota
# NUID: 001099670

# Get and set the working directories
getwd()

## [1] "G:/NEU/Coursework/2021 Q1 Winter/ALY 6015 IA/Discussions &
Assignments"

setwd('G:/NEU/Coursework/2021 Q1 Winter/ALY 6015 IA/Discussions &
Assignments')
getwd()

## [1] "G:/NEU/Coursework/2021 Q1 Winter/ALY 6015 IA/Discussions &
Assignments"

# Installed the above packages into the work space

install.packages("datasets")
install.packages("plyr")
install.packages("dplyr")
install.packages("tidyr")
install.packages("tidyverse")
install.packages("ggplot2")
install.packages("ggcorrplot")

# Loaded the below libraries into the workspace

library(plyr)
library(dplyr)
library(tidyr)
library(tidyverse)
library(ggplot2)
library(ggcorrplot)
require(grDevices)
require(datasets)
```

```
data(mtcars) # Load the mtcars Data set into the Environment
View(mtcars) # To View the mtcars Data set
head(mtcars) # It shows first few rows in the Data set
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1

tail(mtcars) # It shows last few rows in the Data set

##                 mpg cyl  disp  hp drat    wt qsec vs am gear carb
## Porsche 914-2  26.0   4 120.3  91 4.43 2.140 16.7  0  1    5    2
## Lotus Europa   30.4   4  95.1 113 3.77 1.513 16.9  1  1    5    2
## Ford Pantera L 15.8   8 351.0 264 4.22 3.170 14.5  0  1    5    4
## Ferrari Dino   19.7   6 145.0 175 3.62 2.770 15.5  0  1    5    6
## Maserati Bora  15.0   8 301.0 335 3.54 3.570 14.6  0  1    5    8
## Volvo 142E     21.4   4 121.0 109 4.11 2.780 18.6  1  1    4    2

summary(mtcars) # Provides the Descriptive Stats of the mtcars Data set

##       mpg             cyl             disp             hp
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
##       drat             wt             qsec             vs
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
##        am              gear            carb
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
##  Median :0.0000   Median :4.000   Median :2.000
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000

mtcars$am <- factor(mtcars$am)
mtcars$am

##  [1] 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1
## Levels: 0 1
```

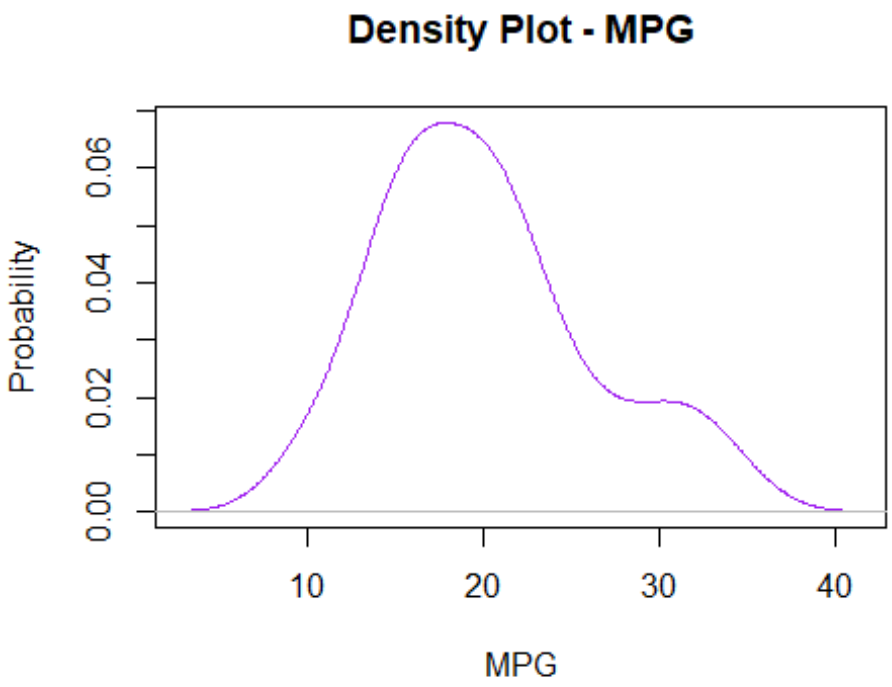
```r
str(mtcars) # To observe the structure of the Data set
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : Factor w/ 2 levels "0","1": 2 2 2 1 1 1 1 1 1 1 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

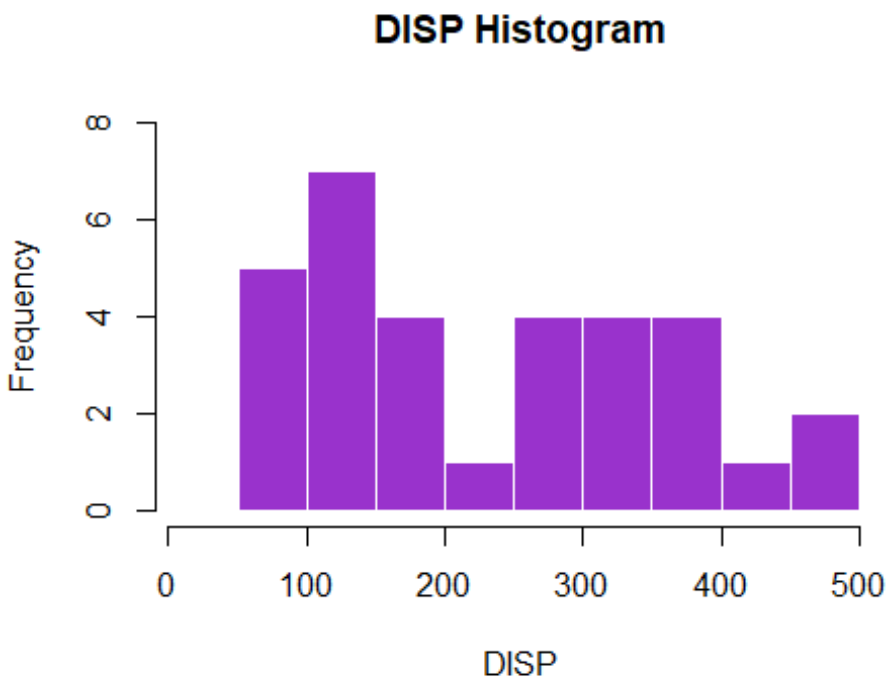
```r
hist(
  mtcars$mpg,
  ylab = "Frequency",
  xlab = "MPG",
  breaks = 10,
  xlim = c(min(mtcars$mpg), 35),
  main = "MPG Histogram",
  ylim = c(0, 8),
  col = "brown1",
  border = FALSE
) # Histogram of mpg
```



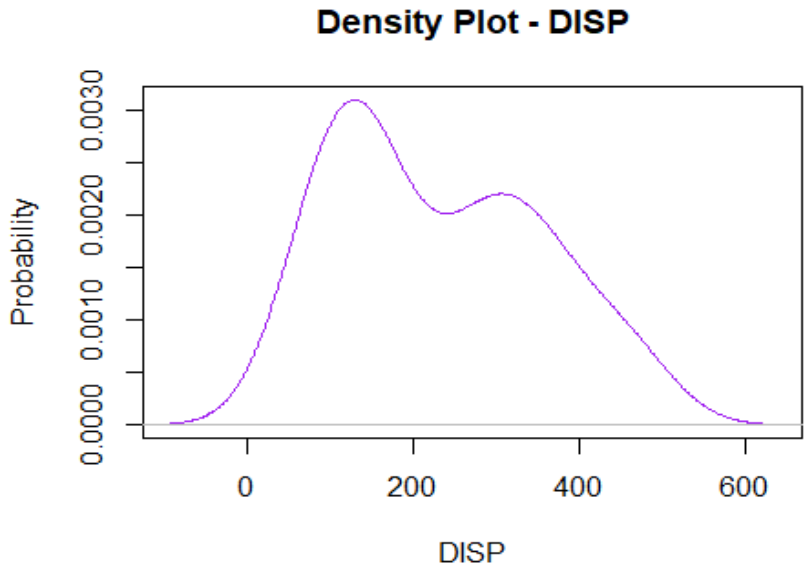
```r
plot(
  density(mtcars$mpg),
  main = "Density Plot - MPG",
  xlab = "MPG",
  ylab = "Probability",
  col = "purple"
) # Density Plot of mpg
```



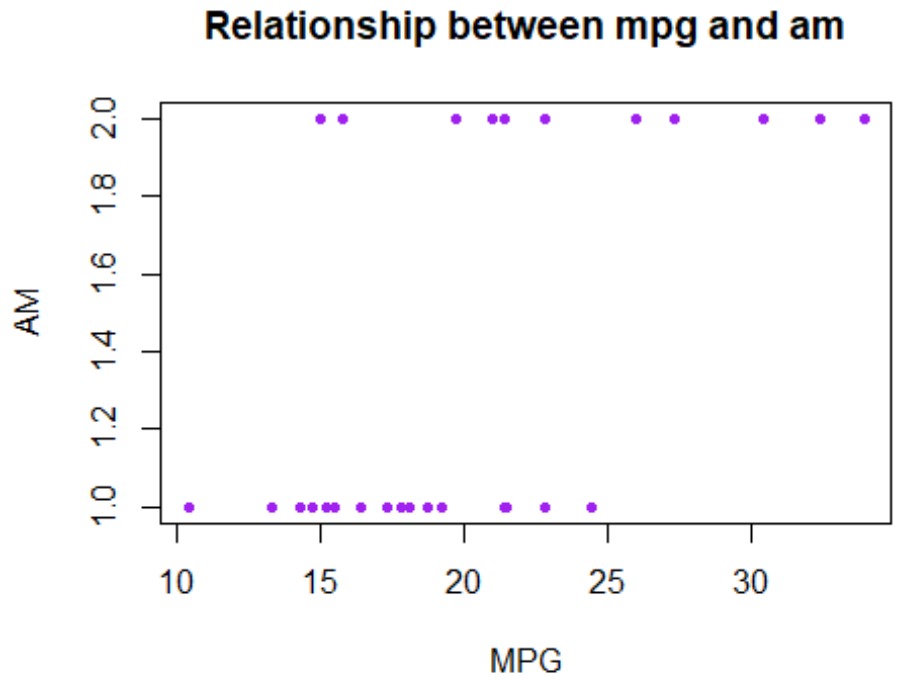

```r
hist(
  mtcars$disp,
  ylab = "Frequency",
  xlab = "DISP",
  breaks = 9,
  xlim = c(min(mtcars$mpg), 500),
  main = "DISP Histogram",
  ylim = c(0, 8),
  col = "darkorchid",
  border = FALSE
) # Histogram of disp
```



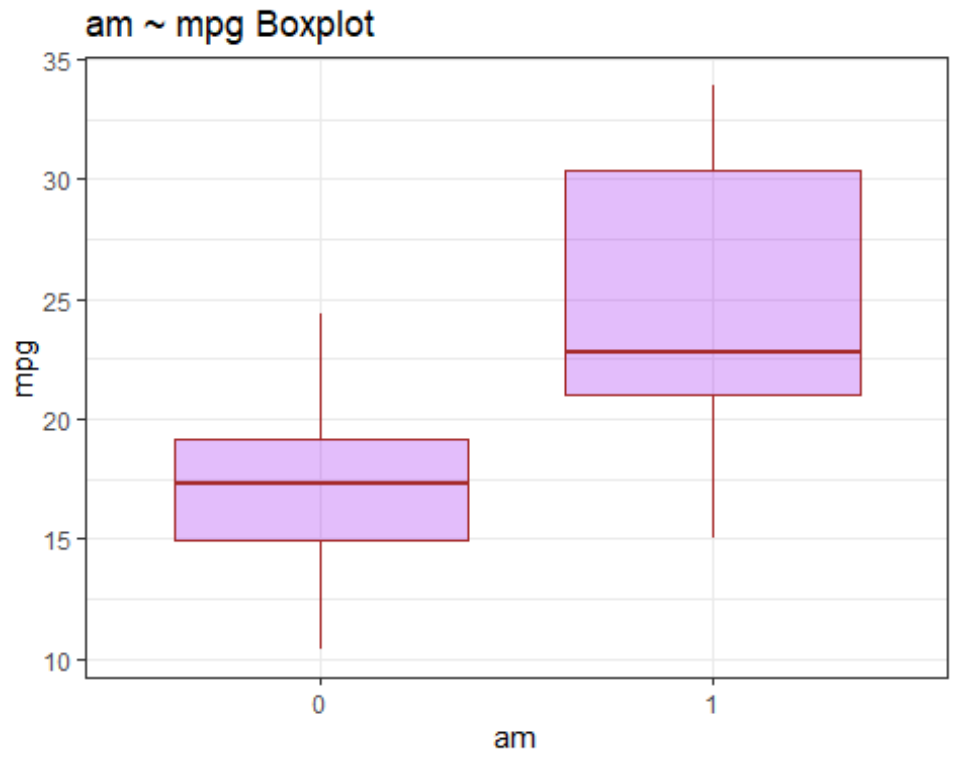

```r
plot(
  density(mtcars$disp),
  main = "Density Plot - DISP",
  xlab = "DISP",
  ylab = "Probability",
  col = "purple"
) # Density Plot of disp
```



```r
plot(
  x = mtcars$mpg,
  y = mtcars$am,
  xlab = "MPG",
  ylab = "AM",
  main = "Relationship between mpg and am",
  col = "purple",
  pch = 20,
  xlim = c(min(mtcars$mpg), max(mtcars$mpg))
) # Scatter Plot is used to depict the relationship between the MPG and AM
```




```r
boxPltMPGandAM <-
  data.frame(mtcars$mpg, mtcars$am) # Boxplot of mpg and am
ggplot(boxPltMPGandAM,
       aes(
         x = mtcars$am,
         y = mtcars$mpg,
         fill = mtcars$am
       )) + geom_boxplot(color = "brown",
                         alpha = 0.3,
                         fill = "purple") + ggtitle("am ~ mpg Boxplot") +
ylab("mpg") + xlab("am") + theme_bw() # GGplot is used to plot the Boxplot
```

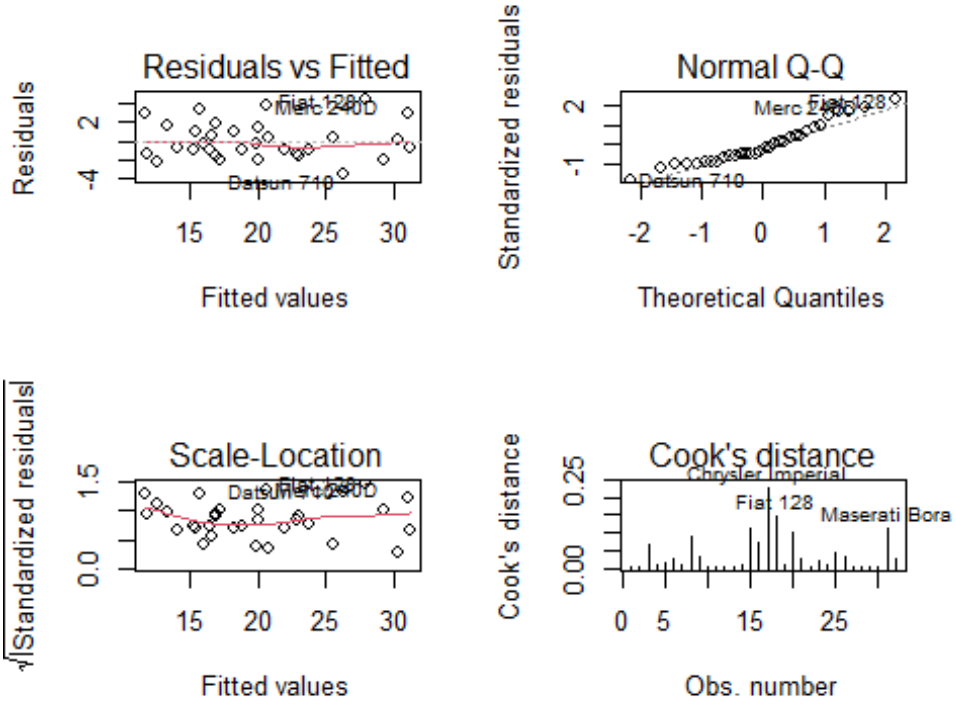



```
lmFit <- lm(mpg ~ am, data = mtcars)
lmFit # Linear Model relationship between mpg and am in mtcars dataset

##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Coefficients:
## (Intercept)          am1
##      17.147        7.245

summary(lmFit) # Summary Stats of the linear model

##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## am1            7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285

rSquared <- summary(lmFit)$r.squared
rSquared # r squared value

## [1] 0.3597989

lmFitAll <- lm(mpg ~ wt + am + wt:am + qsec, data = mtcars)
lmFitAll # Linear Model relationship between mpg and all other parameters in
mtcars dataset

##
## Call:
## lm(formula = mpg ~ wt + am + wt:am + qsec, data = mtcars)
##
## Coefficients:
## (Intercept)           wt          am1         qsec       wt:am1
##       9.723       -2.937       14.079        1.017       -4.141

summary(lmFitAll) # Summary Stats of the linear model

##
## Call:
## lm(formula = mpg ~ wt + am + wt:am + qsec, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5076 -1.3801 -0.5588  1.0630  4.3684
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.723      5.899   1.648 0.110893
## wt            -2.937      0.666  -4.409 0.000149 ***
## am1           14.079      3.435   4.099 0.000341 ***
## qsec           1.017      0.252   4.035 0.000403 ***
## wt:am1        -4.141      1.197  -3.460 0.001809 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.084 on 27 degrees of freedom
## Multiple R-squared:  0.8959, Adjusted R-squared:  0.8804
## F-statistic: 58.06 on 4 and 27 DF,  p-value: 7.168e-13

rSquaredAll <- summary(lmFitAll)$r.squared
rSquaredAll # r squared value

## [1] 0.8958514

confint(lmFitAll) # confidence interval of the Linear Model
```

```
##                   2.5 %    97.5 %
## (Intercept) -2.3807791 21.826884
## wt          -4.3031019 -1.569960
## am1          7.0308746 21.127981
## qsec         0.4998811  1.534066
## wt:am1      -6.5970316 -1.685721
```

```
par(mfrow = c(2, 2)) # To form the Plots in a 2X2 matrix
plot(lmFitAll, which = 1:4)# Plotted the residuals vs fitted, normal QQ,
Scale-location, and Cooks distance plots
```




```
tTestMPG <- t.test(mtcars$mpg, mu = 20)
tTestMPG # Performed One-sample T Test
```

```
##  One Sample t-test
##
## data:  mtcars$mpg
## t = 0.08506, df = 31, p-value = 0.9328
## alternative hypothesis: true mean is not equal to 20
## 95 percent confidence interval:
##  17.91768 22.26357
## sample estimates:
## mean of x
##  20.09062
```

```r
twoSampleTTest <-
  t.test(
    mpg ~ am,
    data = mtcars,
    var.equal = FALSE,
    paired = FALSE ,
    conf.level = .95
  )
twoSampleTTest # Performed Two-sample T Test

##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group 0 mean in group 1
##        17.14737        24.39231

tTestPaired <-
  t.test(mtcars$mpg,
         mtcars$disp,
         paired = TRUE,
         alternative = "less")
tTestPaired # Performed Paired T Test

##
##  Paired t-test
##
## data:  mtcars$mpg and mtcars$disp
## t = -9.2303, df = 31, p-value = 1.047e-10
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -171.9404
## sample estimates:
## mean of the differences
##               -210.6312

tableDataSet <- table(mtcars$mpg, mtcars$am)
propTestMtcars <- prop.test(tableDataSet,
                            conf.level = 0.95,
                            alternative = "two.sided")

## Warning in prop.test(tableDataSet, conf.level = 0.95, alternative =
## "two.sided"): Chi-squared approximation may be incorrect

propTestMtcars # Performed Prop Test
```

```
##
##  25-sample test for equality of proportions without continuity
##  correction
##
## data:  tableDataSet
## X-squared = 27.854, df = 24, p-value = 0.2662
## alternative hypothesis: two.sided
## sample estimates:
##  prop 1  prop 2  prop 3  prop 4  prop 5  prop 6  prop 7  prop 8  prop 9
prop 10
##     1.0     1.0     1.0     1.0     0.0     1.0     1.0     0.0     1.0
1.0
## prop 11 prop 12 prop 13 prop 14 prop 15 prop 16 prop 17 prop 18 prop 19
prop 20
##     1.0     1.0     1.0     1.0     0.0     0.0     0.5     1.0     0.5
1.0
## prop 21 prop 22 prop 23 prop 24 prop 25
##     0.0     0.0     0.0     0.0     0.0
```

```
varTestMtcars <-
  var.test(mtcars$mpg, mtcars$disp)
varTestMtcars # Performed Var Test
```

```
##
##  F test to compare two variances
##
## data:  mtcars$mpg and mtcars$disp
## F = 0.0023647, num df = 31, denom df = 31, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.001154324 0.004844338
## sample estimates:
## ratio of variances
##        0.002364727
```