Cars

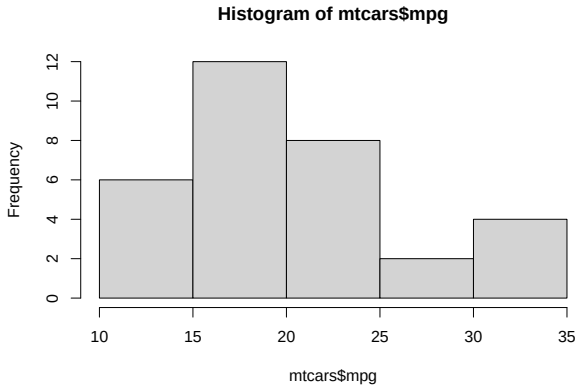
Vasileios Papadopoulos Georgios Papadopoulos

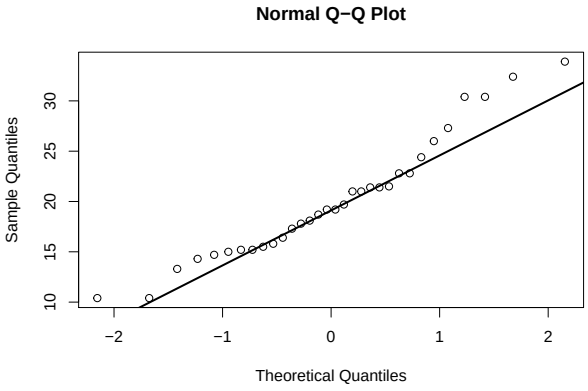
```
head(mtcars)
```

```
##                   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
```

Check whether *mpg* comes from a Gaussian distribution with Q-Q plot.

```
hist(mtcars$mpg)
qqnorm(mtcars$mpg, pch = 1)
qqline(mtcars$mpg, lwd = 2)
```





Binary feature *am* represents *automatic* with 0 and *manual* with 1. For convenience we convert it into a categorical variable.

```
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c('at','mt')
head(mtcars)
```

```
##                   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 mt    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0 mt    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1 mt    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1 at    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0 at    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1 at    3    1
```

We split the dataset into two subsets *Automatic(AT)* and *Manual (MT)* transmission. We want to find confidence intervals for 0.95 confidence for the mean *MPG* for both categories. In order to do this we run *t-test*

```
mpg.automatic <- mtcars[mtcars$am == "at",]$mpg
mpg.manual <- mtcars[mtcars$am == "mt",]$mpg
t.test(mpg.automatic, mpg.manual)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg.automatic and mpg.manual
## 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 of x mean of y
##  17.14737  24.39231
```

P-value is *0.001374* which is below the 5% confidence, so alternative hypothesis is accepted. The 95% confidence interval of the difference in mean fuel consumption between *Automatic* and *Manual* transmission is between *3.2* and *11.2*