

# Melwin Poovakottu Homework Week 1

*Melwin Poovakottu*

*Sunday, January 15, 2017*

## Statistics for Data Science

### Unit 1 Homework

#### Exercise

Load the dataset found in the file, cars.csv.

```
cars = read.table("C:/Users/Melwin/Desktop/Data Science files/UC Berkeley/W203 Stats/W203_Assignments/W
```

1. What are the variables in the file?

-> Colnames function displays all the variables in the file

```
colnames(cars)
```

```
## [1] "mpg" "cyl" "disp" "hp" "drat" "wt" "qsec" "vs" "am" "gear"  
## [11] "carb"
```

2. Find the mean, median, minimum, maximum, 1st quartile and 3rd quartile for the mpg variable.

-> Here the summary function is used to calculate the mean, median, minimum, maximum, 1st quartile and 3rd quartile for mpg variable

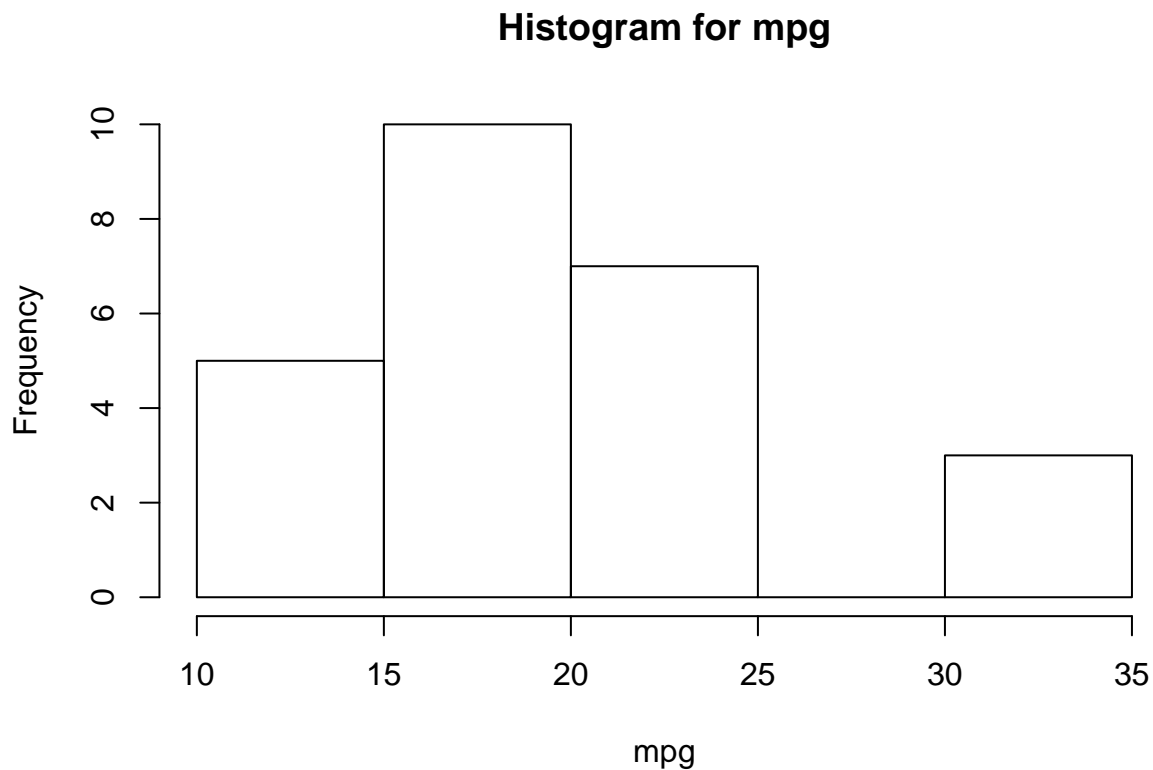
```
summary(cars$mpg)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##   10.40   15.20   18.70   19.49   21.50   33.90
```

3. Create a histogram of the mpg variable.

-> The hist function is used to create the histogram. The chart is named as “Histogram for mpg” and the x-axis is named as “mpg”

```
hist(cars$mpg, xlab="mpg", main= 'Histogram for mpg')
```



4. What is the standard deviation of mpg variable?

-> To determine the standard deviation the sd command is used

```
sd(cars$mpg)
```

```
## [1] 6.047446
```

5. What is the variance of mpg variable?

-> The variance of the mpg variable is determined by using the var command

```
var(cars$mpg)
```

```
## [1] 36.5716
```

6. What is the relationship of the standard deviation to the variance? Why does the standard deviation and variance of the mpg variable differ?

-> the Standard Deviation is the square root of variance. i.e

```
Standard deviation = squareroot(variance)
```

That is why the standard deviation and the variance of the mpg variable differs. Standard deviation has the same unit as the mean(or median) of the variable whereas the variance has square units of the variable.

7. How many data points are there for the cyl variable?

-> There 23 data points in the cyl variable, excluding the na points

```
length(cars$cyl[!is.na(cars$cyl)])
```

```
## [1] 23
```

8. What is the mean of the cyl variable?

```
mean(cars$cyl, na.rm=TRUE)
```

```
## [1] 6.26087
```