

# DATA ANALYTICS WITH R, EXCEL AND TABLA EU

## ASSIGNMENT 7.1 ANSWERS

By ASHISH S SHANBHAG

ashishshanbhag108@gmail.com

### Question no:

5)

1) Histogram for all variables in a dataset mtcars. Write a program to create histograms for all columns.

**Ans**

```
library(tidyr)
library(ggplot2)
mtcars %>% gather() %>% head()
ggplot(gather(mtcars), aes(value)) +
  geom_histogram(bins = 10) +
  facet_wrap(~key, scales = 'free_x')
for (col in 2:ncol(df))
{
  hist(df[,col])
}
hist(df[,col], breaks=10)
```

2) Check the probability distribution of all variables in mtcars.

**Ans**

```
y = mtcars$mpg
group = mtcars$am
y; group

# baseline group means and difference
baselineMeans = tapply(mtcars$mpg, mtcars$am, mean)
baselineMeansDiff = baselineMeans[2] - baselineMeans[1]

tStat = function(w, g) mean(w[g == 1]) - mean(w[g == 0])
observedDiff = tStat(y, group)

# check if function works - should be 0:
baselineMeansDiff - observedDiff

# execute shuffle:
permutations = sapply(1:100000, function(i) tStat(y, sample(group)))
```

**3)** Write a program to create boxplot for all variables.

**Ans:**

```
input <- mtcars[,c('mpg','cyl')]
print(head(input))
```

When we execute above code, it produces following result –

	mpg	cyl
Mazda RX4	21.0	6
Mazda RX4 Wag	21.0	6
Datsun 710	22.8	4
Hornet 4 Drive	21.4	6
Hornet Sportabout	18.7	8
Valiant	18.1	6

Creating the Boxplot

# Give the chart file a name.

```
png(file = "boxplot.png")
```

# Plot the chart.

```
boxplot(mpg ~ cyl, data = mtcars, xlab = "Number of Cylinders",
        ylab = "Miles Per Gallon", main = "Mileage Data")
```

# Save the file.

```
dev.off()
```

**Mileage Data**

