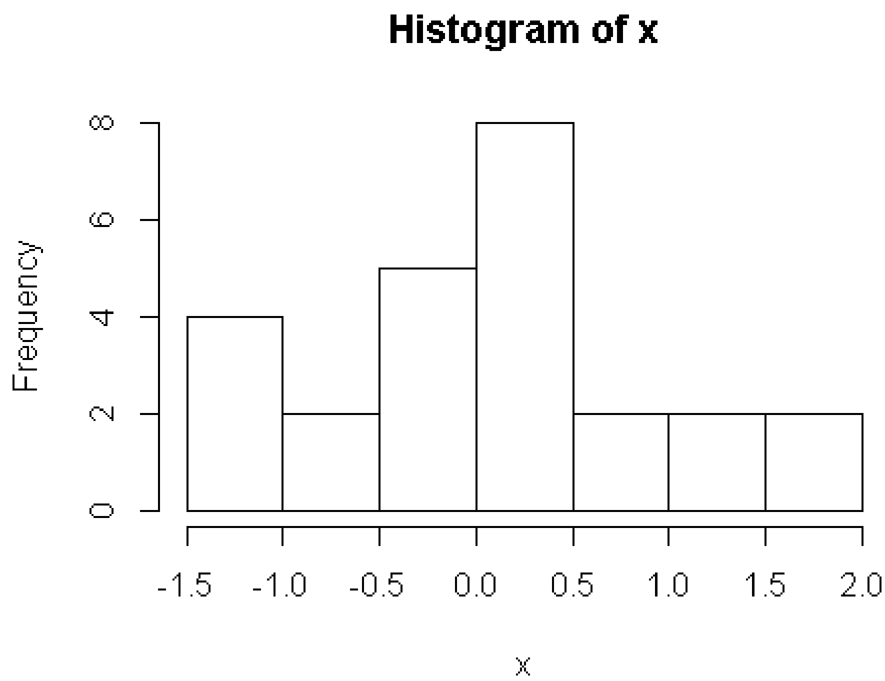
Homework 1

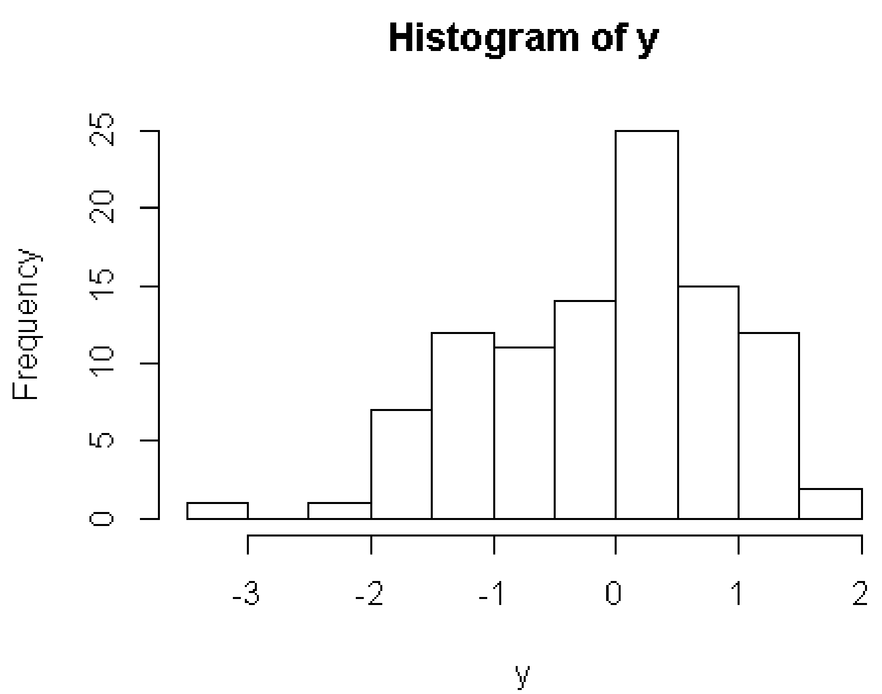
Abhishek Reddy Palle

11 September 2018

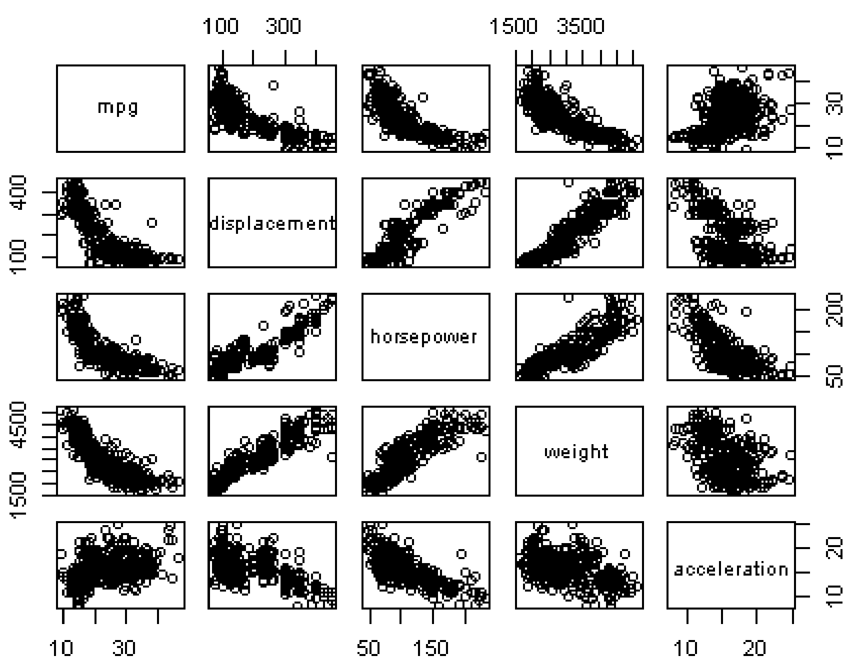
```
# Generate a vector with 25 elements and each element independently follows a
normal distribution (with mean =0 and sd=1)
x <- rnorm(25,0,1)
x

##  [1]  0.86721196  0.41126351 -0.17586801 -0.12964846 -1.48799041
##  [6]  0.56020730 -0.11099003 -0.80286027 -0.82116933  0.09124826
## [11] -1.16271406  0.31221629  0.17867202 -1.30920205  1.12244170
## [16]  0.27789255  1.23186626  1.87503837 -1.23088156  0.29473630
## [21]  0.31979631 -0.02844221 -0.34055262  1.59024610  0.04029159

# Reshape this vector into a 5 by 5 matrix in two ways (arranged by row and
column)
a <- matrix(x, nrow = 5)
a

##            [,1]        [,2]       [,3]       [,4]        [,5]
## [1,]  0.8672120  0.56020730 -1.1627141  0.2778925  0.31979631
## [2,]  0.4112635 -0.11099003  0.3122163  1.2318663 -0.02844221
## [3,] -0.1758680 -0.80286027  0.1786720  1.8750384 -0.34055262
## [4,] -0.1296485 -0.82116933 -1.3092021 -1.2308816  1.59024610
## [5,] -1.4879904  0.09124826  1.1224417  0.2947363  0.04029159

# Similarly, generate another vector with 100 elements and plot its
histogram.
b <- matrix(x, nrow = 5, byrow = T)
b

##            [,1]        [,2]       [,3]       [,4]        [,5]
## [1,]  0.8672120  0.41126351 -0.1758680 -0.1296485 -1.48799041
## [2,]  0.5602073 -0.11099003 -0.8028603 -0.8211693  0.09124826
## [3,] -1.1627141  0.31221629  0.1786720 -1.3092021  1.12244170
## [4,]  0.2778925  1.23186626  1.8750384 -1.2308816  0.29473630
## [5,]  0.3197963 -0.02844221 -0.3405526  1.5902461  0.04029159

# Provide screenshots of the R code used for the above questions as well as
the plots in the report. Explain the plots in your own words.

hist(x)
```



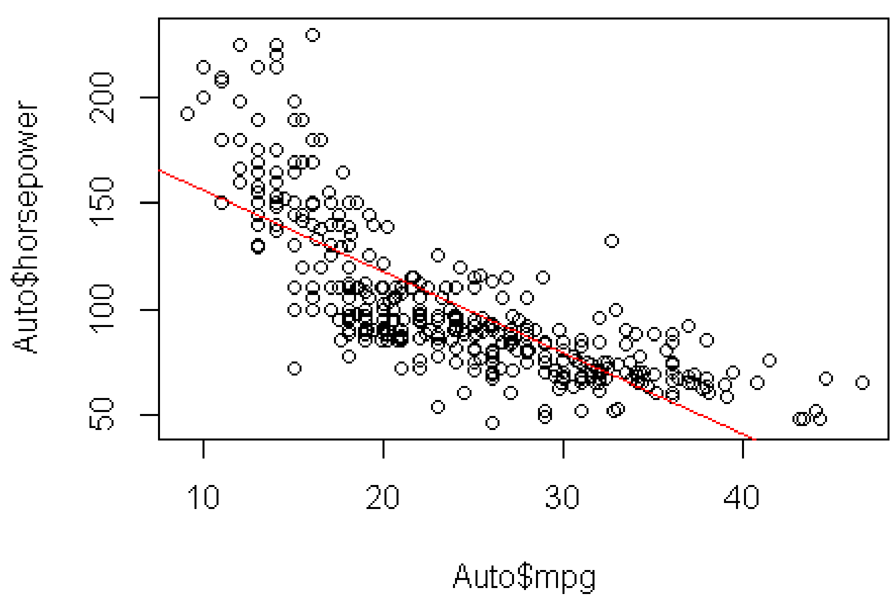

```r
y <- rnorm(100,0,1)
hist(y)
```



```r
# Ans : The plot with 100 elements is a better representation of normal
distribution compared to the one with 25 elements. As the number of samples
increases, the more closely the sample probability distribution resembles the
theoretical normal probability distribution.

# Upload the Auto data set, which is in the ISLR library. Understand
information about this data set by either ways we introduced in class (like
“?Auto” and names(Auto))

library(ISLR)
?Auto

## starting httpd help server ... done

names(Auto)

## [1] "mpg"          "cylinders"    "displacement" "horsepower"
## [5] "weight"       "acceleration" "year"         "origin"
## [9] "name"

# Make a scatterplot between every pair of the following variables (try to
plot all scatterplots in one figure; hint: use pairs() command): “mpg”,
“displacement”, “horsepower”, “weight”, “acceleration”. By observing the
plots, do you think the two variables in each scatterplot are correlated? If
so, how?

pairs(Auto[,c(1,3:6)])
```



```r
cor(Auto[,c(1,3:6)])

##                     mpg displacement horsepower     weight acceleration
## mpg           1.0000000   -0.8051269 -0.7784268 -0.8322442    0.4233285
## displacement -0.8051269    1.0000000  0.8972570  0.9329944   -0.5438005
## horsepower   -0.7784268    0.8972570  1.0000000  0.8645377   -0.6891955
## weight       -0.8322442    0.9329944  0.8645377  1.0000000   -0.4168392
## acceleration  0.4233285   -0.5438005 -0.6891955 -0.4168392    1.0000000

#Positive Correlation means the x axis increases as the y increase
approximately. In the plots displacement Vs horsepower, displacement Vs
weight, horsepower vs weight are positively correlated. Negative Correlation
is when one axis decreases when the other increases. In the plots mpg vs
displacement, mpg Vs horsepower, mpg Vs weight are negatively correlated.

# Draw a line on the scatterplot of mpg vs. horsepower to represent
relationship between the two variables.

plot(Auto$mpg, Auto$horsepower)
abline(lm(horsepower ~ mpg, Auto), col = "red")
```




```r
lm(horsepower ~ mpg, Auto)

##
## Call:
## lm(formula = horsepower ~ mpg, data = Auto)
##
## Coefficients:
```

```
## (Intercept)          mpg
##     194.476       -3.839

# Is there a better way to represent their relationship rather than the
linear model you just drew? (No need to use mathematical formula. Just draw
something on the figure)

plot(Auto$mpg, Auto$horsepower)
lines(lowess(Auto$mpg,Auto$horsepower), col ="red")
```