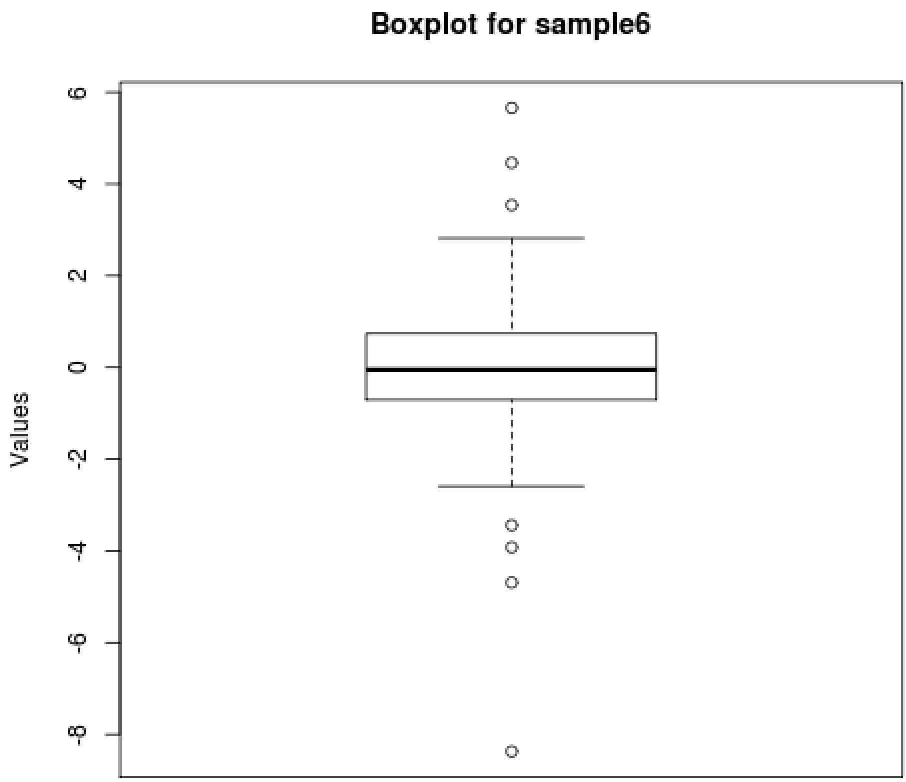
Homework 1 – Rais Mense
rme360 / 2510240

Exercise 1.1

a.



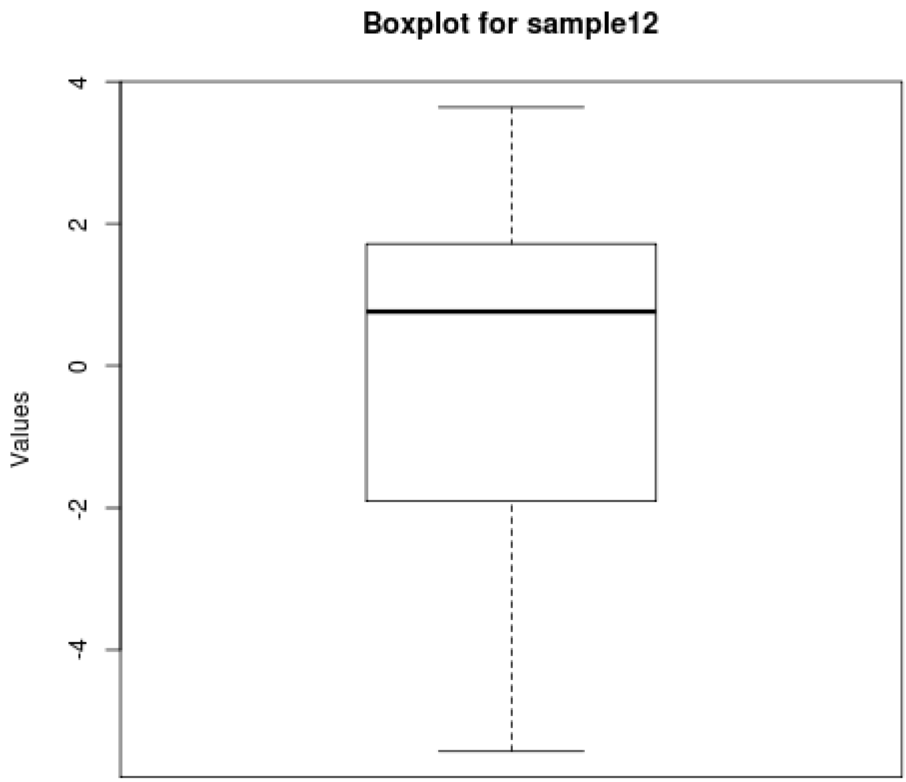
b.

Min: -8.37000
Max: 5.66000
1st Qu.: -0.70500
2rd Qu.: -0.05
3rd Qu.: 0.74000
Median: -0.05000
Mean: -0.01415
Std dev.: 1.593163

c.

Most of the data points are concentrated around the zero, roughly between -1.5 and 1.5.

d.



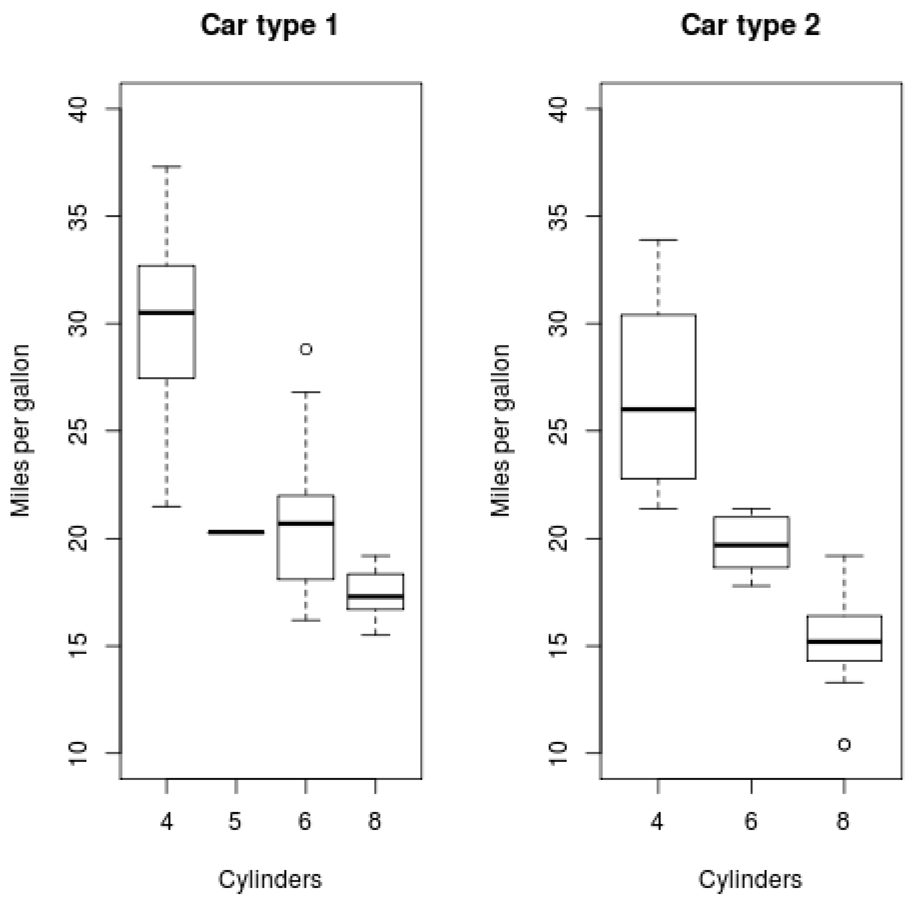
Min: -5.43000
Max: 3.64000
1st Qu.: -1.89800
2rd Qu.: 0.76
3rd Qu.: 1.71200
Median: 0.76000
Mean: 0.02661
Std dev.: 2.140151

The data is split in two, roughly centered around -3 and 1.

e.

The two data sets do not seem to be from the same population. Their distribution is vastly different and their ranges are quite dissimilar.

Exercise 1.2



Type 1:

Four Cylinders:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	StDev.
21.50	27.45	30.50	30.02	32.70	37.30	4.182447

Data points in category: 19

Five Cylinders:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
20.3	20.3	20.3	20.3	20.3	20.3

Data points in category: 1

Six Cylinders:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	StDev.
16.20	18.23	20.70	21.08	21.98	28.80	4.077526

Data points in category: 10

Eight Cylinders:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	StDev.
15.50	16.80	17.30	17.42	18.27	19.20	1.192536

Data points in category: 8

Type 2:

Four Cylinders:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	StDev.
21.40	22.80	26.00	26.66	30.40	33.90	4.509828

Data points in category: 11

Six Cylinders:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	StDev.
17.80	18.65	19.70	19.74	21.00	21.40	1.453567

Data points in category: 7

Eight Cylinders:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	StDev.
10.40	14.40	15.20	15.10	16.25	19.20	2.560048

Data points in category: 14

In both types of cars the four cylinder versions seem to get the most miles to the gallon on average, but also show the greatest fluctuation, in some cases falling well below six cylinder models. Cars of type 1 appear to get worse mpg overall than their type 2 counterparts.
The single five cylinder data point for type 1 looks like it may be roughly consistent with the drop of in mpg for this type of car with the number of cylinders, but since it is just a single data point, it cannot be validate one way or the other.

Exercise 1.3

a.

Yes, based on the correlation coefficient of 0.9962575, it is very likely that this sample is from population with a normal distribution.

b.

N(12.22,0.5285562) appears to be appropriate for this distribution.

c.

i) 0.3547935

ii) 0.06431387

iii) 0.9239101

d.

i) 0.33

ii) 0.06

iii) 0.92

The values obtained from counting the sample data matches the predicted values from the normal distribution about as closely as possible with this amount of sample data.

e.

The correlation between the data from c and d shows that it is highly probable that the earlier assumption that the sample data is from a population with a normal distribution is correct.

Exercise 1.4

a.

μ_L = 792.458

m_{1879} = 852.4

s_{1879} = 79.01055

b.

The chance for a single measurement to be smaller than m_{1878} is 0.5, or 50%.

c.

This chance is also 0.5.

d.

These probabilities seem to suggest that based on the extracted values (m_{1879} and s_{1879}) that the errors in Michelson's measurements were made in a manner that followed a normal distribution, since the mean and standard deviation. However, further investigation is needed if this is to be determined with greater certainty, as the mean and standard deviation may also hold for other distributions and simply be a coincidence in this case.

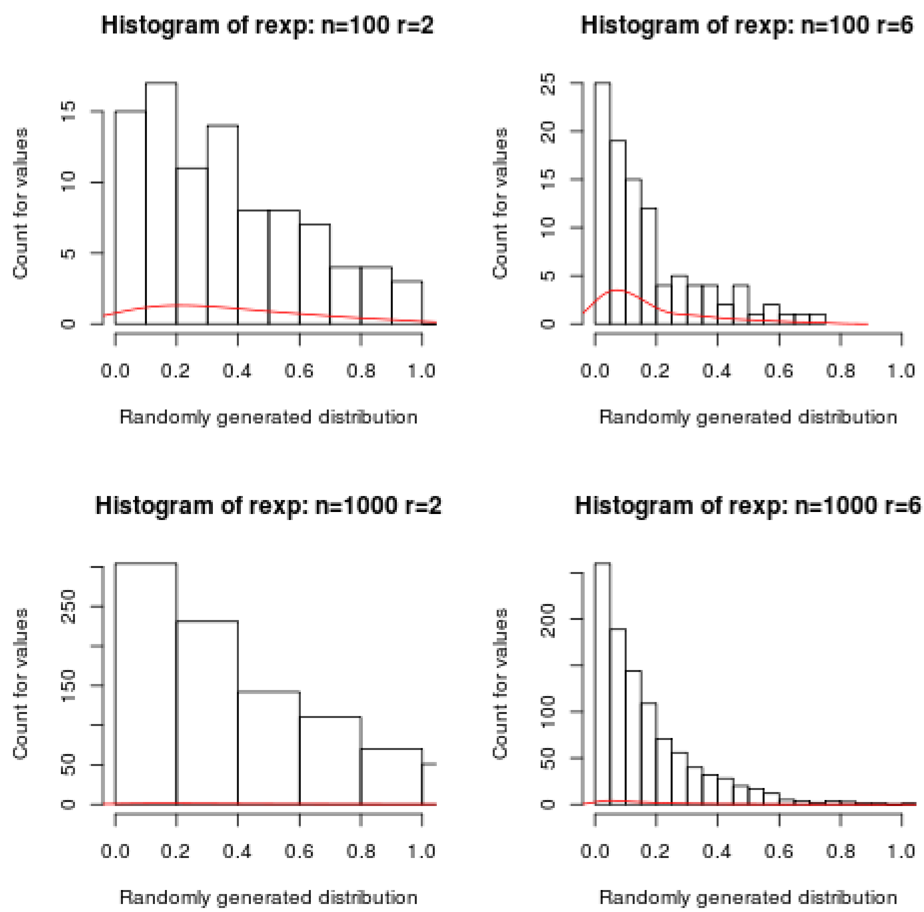
Exercise 1.5

a.

```
f <- function(n, r) {
     dist = rexp(n, r);
     main=paste("Histogram of rexp: n=", n, " r=", r, sep="");
     hist(dist, xlim=range(0,1), xlab="Randomly generated
distribution", ylab="Count for values", main=main, breaks=20);
     lines(density(dist), col="red");
}
```

b.

The parameter r has the most influence on how well the histogram follows the density. As the sample size increases, the distribution becomes smoother.



Appendix

Exercise 1.1

a.

```
data = scan("sample6.txt")
boxplot(data, main="Boxplot for sample6", ylab="Values")
```

b.

```
summary(data)
quantile(data,0.5)
sd(data)
```

c.

-

d.

Same as a & b. (With the exception of the filename.)

e.

-

Exercise 1.2

```
source("mileage.txt")
par(mfrow=c(1,2))
boxplot(mpg1~cyl1, data=mileage[1:2], ylim=c(10,40), xlab="Cylinders",
ylab="Miles per gallon", main="Car type 1")
boxplot(mpg2~cyl2, data=mileage[3:4], ylim=c(10,40), xlab="Cylinders",
ylab="Miles per gallon", main="Car type 2")

type1=matrix(c(mileage$mpg1, mileage$cyl1), 38, 2)
type2=matrix(c(mileage$cyl2, mileage$mpg2), 32, 2)
type1=type1[order(type1[, 1]),]
type2=type2[order(type2[, 1]),]
"type_1" <- list(
      "four" = c(type1[1:19,2]),
      "five" = c(type1[20,2]),
      "six" = c(type1[21:30,2]),
      "eight" = c(type1[31:38,2])
)
"type_2" <- list(
      "four" = c(type2[1:11,2]),
      "six" = c(type2[12:18,2]),
      "eight" = c(type2[19:32,2])
)

summary(type_1$four)
sd(type_1$four)
length(type_1$four)
```

```
summary(type_1$five)
length(type_1$five)
summary(type_1$six)
sd(type_1$six)
length(type_1$six)
summary(type_1$eight)
sd(type_1$eight)
length(type_1$eight)
summary(type_2$four)
sd(type_2$four)
length(type_2$four)
summary(type_2$six)
sd(type_2$six)
length(type_2$six)
summary(type_2$eight)
sd(type_2$eight)
length(type_2$eight)
```

Exercise 1.3

a.

```
dell = scan("logdell.txt")
V=qqnorm(dell, plot=FALSE)
cor(V$x, V$y)
```

b.

```
mean(dell)
sd(dell)
qqplot(dell, rnorm(n=300, m=12.22, sd=0.5285562));abline(0,1, col="red");
```

c.

```
i)   pnorm(12, mean=mean(dell), sd=sd(dell))
ii)  1 - pnorm(13, mean=mean(dell), sd=sd(dell))
iii) pnorm(13, mean=mean(dell), sd=sd(dell)) - pnorm(11, mean=mean(dell),
sd=sd(dell))
```

d.

```
CountLower <- function(x, max) {
      count = 0
      for(i in 1:length(x)) {
            if(x[i] <= max) {
                  count = count + 1
            }
      }
      return(count)
}

CountHigher <- function(x, max) {
      count = 0
```

```
    for(i in 1:length(x)) {
        if(x[i] >= max) {
            count = count + 1
        }
    }
    return(count)
}
CountLower(dell, 12)/length(dell)
CountHigher(dell, 13)/length(dell)
(CountLower(dell, 13) - CountLower(dell, 11)) / length(dell)
```

e.
-

Exercise 1.4
a.

```
muL=299792.458-299000
source("light.txt")
m1879=mean(light$`1879`)
s1879=sd(light$`1879`)
```

b.

```
pnorm(m1879, m1879, s1879)
```

c.

```
(100*0.5)/100
```

d.
-

Exercise 1.5
a.
Code for this exercise is in the main document.

b.

```
par(mfrow=c(2,2));
source("1.5.a.R");
f(100,2);
f(100,6);
f(1000,2);
f(1000,6);
```