Assignment 2, EDDA 2017

Fabio Curi Paixao (2592802) Arash Parnia (2591051) - Group 22

19 April 2017

Introduction

In the present document, the results for the second and third assignment of the EDDA course are presented.

Assignment 2

Exercise 1

1

The following R code will check whether the 'telephone.txt' dataset stems from the exponential distribution exp(0.035).

```r
telephone = read.table('telephone.txt')
d <- as.numeric(levels(telephone$V1))[telephone$V1]
```

```
## Warning: NAs introduced by coercion
```

```r
d <- d[!is.na(d)]

B=1000
Tstar=numeric(B)
tall = numeric(B)
for(i in 1:B){
  Xstar=sample(d,replace=TRUE)
  Tstar[i]=median(Xstar)
  tall[i] = median(rexp(n = B, rate = 0.035))
}

# Tstar25=quantile(Tstar,0.025)
# Tstar975=quantile(Tstar,0.975)
# sum(Tstar<Tstar25)
# T1 = mean(d)
# c(2*T1-Tstar975,2*T1-Tstar25)


t = median(tall)

pl=sum(Tstar<t)/B
pr=sum(Tstar>t)/B
p=2*min(pl,pr)
p
```
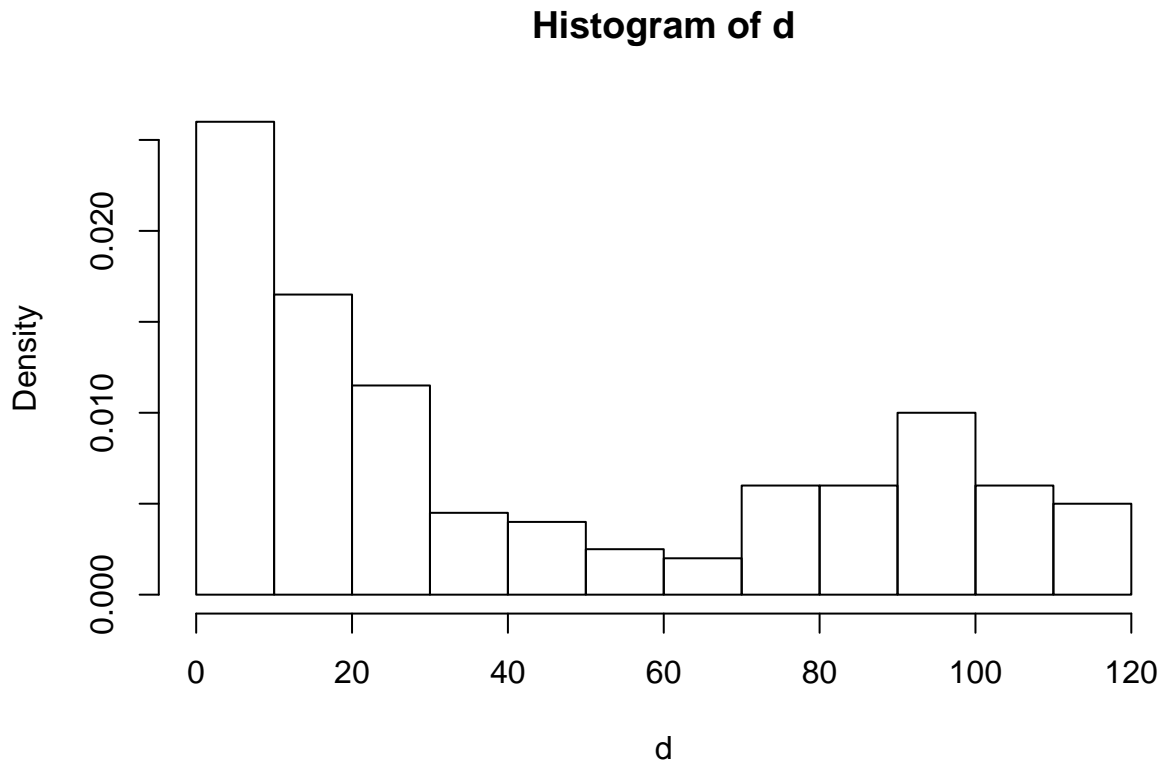
```
## [1] 0.024
```

Since the p-value in lower than 0.05, we reject H0 and we conclude that it is not likely that the sample comes from the standard exponential distribution exp(0.035).

```r
hist(d,prob=TRUE)
```

Histogram of d

An appropriate plot for this data is the histogram shown here above, as it indicates the amplitudes in which the clients lie with their first month bills. As the histogram indicates, the highest portion of the sample lies in lower intervals. However, a considerable portion of the same sample lies in the end of the histogram, which indicates that their bills have been rather expensive. Thus, as consultant, we could advise this company to propose to their clients a new offer in which their bills would have a fixed price, however with limited amount of calls. Thus, we could expect the beginning of the histogram to remain the same, an averaging in the middle, and a shortening in the end.

Exercise 2

First of all, we will make the adjusments necessary for the three databases to be able to be compared. These adjustments are needed since the measurement procedures between both scientists were different.

```r
light1879 = scan(file="light1879.txt");
light1882 = scan(file="light1882.txt");
light = scan(file="light.txt");

light = light/1000  + 24.8

light = 7.442 * 1000 /light
light = light * 1000


light1879 = light1879 + 299000
light1882 = light1882 + 299000

light = light[light > 0]
```
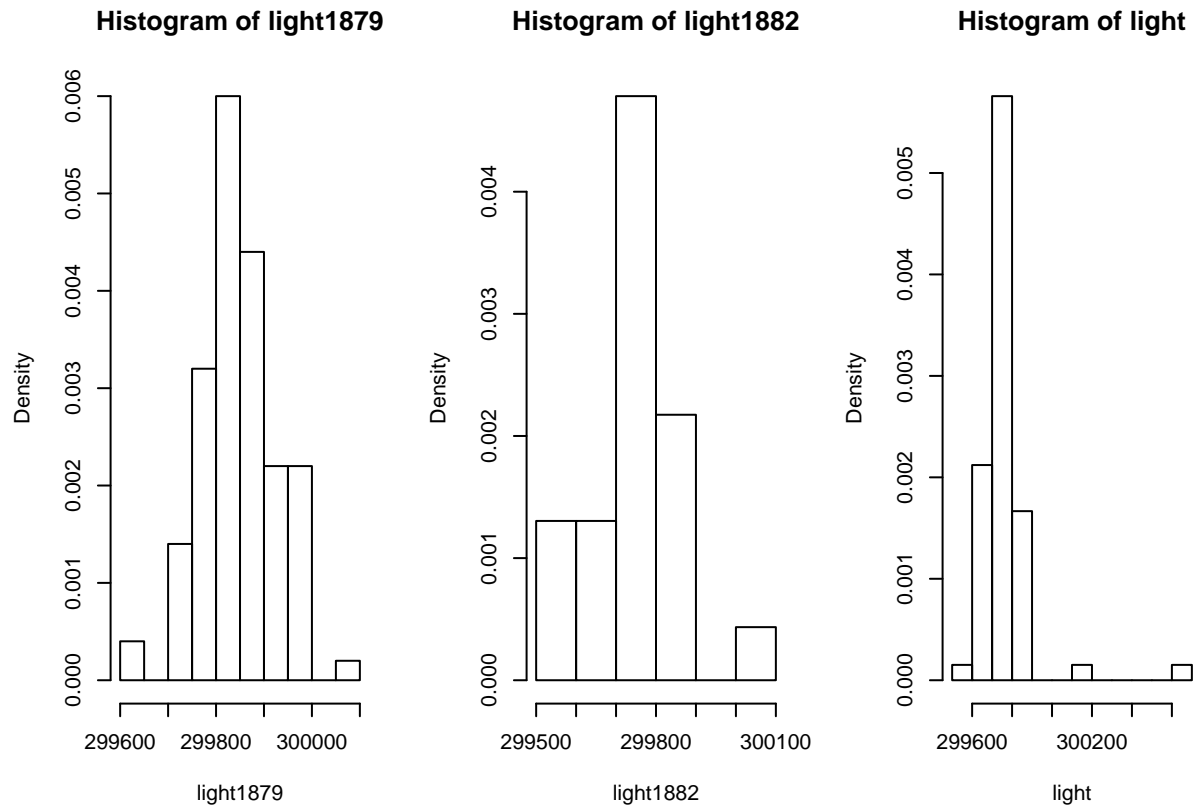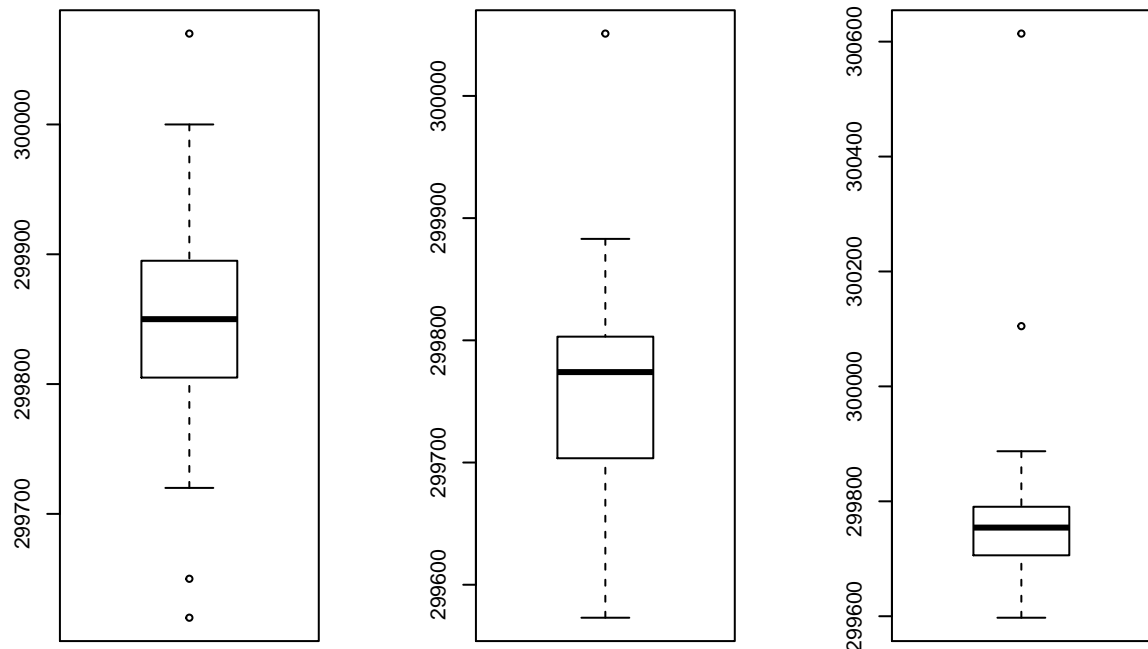
1

Here below are the histograms and the boxplots for the three datasets.

```
par(mfrow=c(1,3))
hist(light1879,prob = T)
hist(light1882,prob = T)
hist(light,prob = T)
```

Histogram of light1879   **Histogram of light1882**   **Histogram of light**


```
par(mfrow=c(1,3))
boxplot(light1879)
boxplot(light1882)
boxplot(light)
```

As it can be seen from the boxplots, there is a significant difference between the measurements made in 1879 by Michelson in 1879 and 1882 and from Newcomb's measurements from 1882. The median values of the boxplots are considerably different between the former and these two last. Furthermore, between the two measurements from 1882, Newcomb managed to obtain his results in a fairly narrower amplitude, which is maybe thanks to his methodology and/or equipments used.

2

Here, we will make use of both mean and median values of each dataset. We have created a function to take into account this choice. The results are given right after.

```
# for means

B=1000
T1 = mean(light1879)
d = light1879
Tstar=numeric(B)
for(i in 1:B){
Xstar=sample(d,replace=TRUE)
Tstar[i]=mean(Xstar)
}

Tstar25=quantile(Tstar,0.025)
Tstar975=quantile(Tstar,0.975)
# sum(Tstar<Tstar25)
T1 = mean(d)
c(2*T1-Tstar975,2*T1-Tstar25)
```

```
##     97.5%     2.5%
## 299837.1 299867.8
```

```
mean(light1882)
```

```
## [1] 299756.2
```

```r
B=1000
T1 = mean(light1882)
d = light1882
Tstar=numeric(B)
for(i in 1:B){
Xstar=sample(d,replace=TRUE)
Tstar[i]=mean(Xstar)
}

Tstar25=quantile(Tstar,0.025)
Tstar975=quantile(Tstar,0.975)
# sum(Tstar<Tstar25)
T1 = mean(d)
c(2*T1-Tstar975,2*T1-Tstar25)
```

```
##     97.5%      2.5%
## 299715.5 299795.9
```

```r
mean(light)
```

```
## [1] 299763.9
```

```r
B=1000
T1 = mean(light)
d = light
Tstar=numeric(B)
for(i in 1:B){
Xstar=sample(d,replace=TRUE)
Tstar[i]=mean(Xstar)
}

Tstar25=quantile(Tstar,0.025)
Tstar975=quantile(Tstar,0.975)
# sum(Tstar<Tstar25)
T1 = mean(d)
c(2*T1-Tstar975,2*T1-Tstar25)
```

```
##     97.5%      2.5%
## 299729.0 299790.1
```

```r
# for median

median(light1879)
```

```
## [1] 299850
```

```r
B=1000
T1 = median(light1879)
d = light1879
Tstar=numeric(B)
for(i in 1:B){
Xstar=sample(d,replace=TRUE)
Tstar[i]=median(Xstar)
}

Tstar25=quantile(Tstar,0.025)
Tstar975=quantile(Tstar,0.975)
```

```r
# sum(Tstar<Tstar25)
T1 = median(d)
c(2*T1-Tstar975,2*T1-Tstar25)
```

```
##  97.5%   2.5%
## 299830 299860
```

```r
median(light1882)
```

```
## [1] 299774
```

```r
B=1000
T1 = median(light1882)
d = light1882
Tstar=numeric(B)
for(i in 1:B){
Xstar=sample(d,replace=TRUE)
Tstar[i]=median(Xstar)
}

Tstar25=quantile(Tstar,0.025)
Tstar975=quantile(Tstar,0.975)
# sum(Tstar<Tstar25)
T1 = median(d)
c(2*T1-Tstar975,2*T1-Tstar25)
```

```
##  97.5%   2.5%
## 299751 299825
```

```r
median(light)
```

```
## [1] 299754.3
```

```r
B=1000
T1 = median(light)
d = light
Tstar=numeric(B)
for(i in 1:B){
Xstar=sample(d,replace=TRUE)
Tstar[i]=median(Xstar)
}

Tstar25=quantile(Tstar,0.025)
Tstar975=quantile(Tstar,0.975)
# sum(Tstar<Tstar25)
T1 = median(d)
c(2*T1-Tstar975,2*T1-Tstar25)
```

```
##    97.5%     2.5%
## 299742.2 299772.4
```

3

Intervals show that all measurements are closely ranged around 299000.

4

Speed of light is 299792458 m / s and 299792.458 kps, and it closely ressembles the measurements of the French physicists Fizeau and Foucault.

Exercise 3

1

Now, we will explore the file "klm.txt". We will now test the null hypothesis that the median duration is smaller or equal to 32 days against the alternative hypothesis that this median is greater than 32 days.

```r
klm = read.table('klm.txt')
klm = c(klm$V1,klm$V2,klm$V3,klm$V4,klm$V5)
par(mfrow=c(1,2))
hist(klm)
boxplot(klm)
summary(klm)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    6.00   29.00   42.00   57.73   70.00  303.00
```

```r
library(BSDA)
```

```
## Loading required package: e1071
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'BSDA'
```
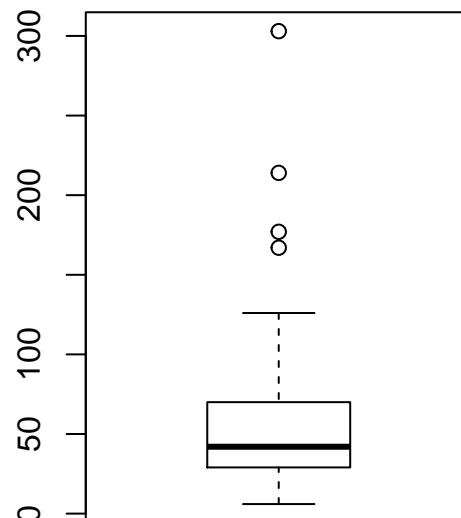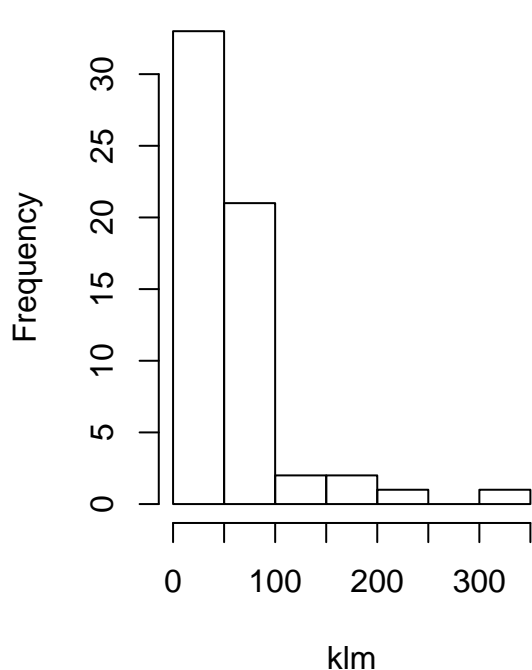
```
## The following object is masked from 'package:datasets':
##
##     Orange
```

Histogram of klm

```
SIGN.test(klm , md = 32,alternative = "greater")
```

```
##
##  One-sample Sign-Test
##
## data:  klm
## s = 37, p-value = 0.01657
## alternative hypothesis: true median is greater than 32
## 95 percent confidence interval:
##  33.36164      Inf
## sample estimates:
## median of x
##          42

##                 Conf.Level  L.E.pt U.E.pt
## Lower Achieved CI   0.9225 36.0000    Inf
## Interpolated CI     0.9500 33.3616    Inf
## Upper Achieved CI   0.9538 33.0000    Inf
```

The sign test concludes that the true median is not equal to 32, therefore inducing to reject the null hypothesis that the median is smaller or equal to 32 days.

2

```
klm-70
```

```
## [1]    0 -22 -41 -18  97 -24 -38  56 -44   7 -64 -39 -11  14   2 -30   0
## [18]   5 -49  33 -26  -5 -38 -47 -42 -50 -33 107 -15   4 -48 233 -38   7
## [35] -33 -14 -11  -6 -40 -16   6 -13 -52 -63  10 -34 -45 -41 -37 -47 -40
## [52] 144 -41 -14 -31 -32 -41  -5 -40 -43
```

```
rank(abs(klm-70))
```

```
## [1]   1.5 22.0 41.5 21.0 57.0 23.0 34.0 54.0 46.0 10.5 56.0 36.0 13.5 17.0
## [15]   3.0 25.0  1.5  6.0 51.0 29.0 24.0  6.0 34.0 48.5 44.0 52.0 29.0 58.0
## [29] 19.0  4.0 50.0 60.0 34.0 10.5 29.0 17.0 13.5  8.5 38.0 20.0  8.5 15.0
## [43] 53.0 55.0 12.0 31.0 47.0 41.5 32.0 48.5 38.0 59.0 41.5 17.0 26.0 27.0
## [57] 41.5  6.0 38.0 45.0
```

```
rank(abs(klm-70))[klm-70>0]
```

```
## [1] 57.0 54.0 10.5 17.0  3.0  6.0 29.0 58.0  4.0 60.0 10.5  8.5 12.0 59.0
```

```
sum(rank(abs(klm-70))[klm-70>0])
```

```
## [1] 388.5
```

```
wilcox.test(klm,mu=70,alternative = "less")
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  klm
## V = 360.5, p-value = 6.431e-05
## alternative hypothesis: true location is less than 70
```

The Wilcox test shows that the null hypothesis indicating that the delivery period is maximum of 70 days can not be rejected. Therefore it is plausible to think that parts arrive after the maximum.

Exercise 4

In this subsection, we will explore the file "clouds.txt".

```
clouds = read.table('clouds.txt',header = FALSE)
a=as.numeric(levels(clouds$V1))[clouds$V1]
```

```
## Warning: NAs introduced by coercion
```

```
a=a[2:end(a)]
```

```
## Warning in 2:end(a): numerical expression has 2 elements: only the first
## used
```

```
b=as.numeric(levels(clouds$V2))[clouds$V2]
```

```
## Warning: NAs introduced by coercion
```

```
b=b[2:end(b)]
```

```
## Warning in 2:end(b): numerical expression has 2 elements: only the first
## used
```

1

To see if there was an improvement, deterioration, or if the means of times have remained substantially the same (hypothesis H0), we need to make a Student's t-test for paired samples:

```
t.test(a,b, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  a and b
## t = 2.1204, df = 25, p-value = 0.04407
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##    7.961957 546.883428
## sample estimates:
## mean of the differences
##                277.4227
```

Here, we asked R to check whether the mean of the values contained in the vector a is equal to the mean of the values contained in the vector b, thus checking whether there has not been any improvement/nor worsening in the amount of rainfall after seeding (H0). The p-value is right below 0.05, so we reject the null hypothesis.

Then, we perform Mann-Whitney test and the Kolmogorov-Smirnov test.

```
wilcox.test(a,b,paired=TRUE)
```

```
##
##  Wilcoxon signed rank test
##
## data:  a and b
## V = 281, p-value = 0.006131
## alternative hypothesis: true location shift is not equal to 0
```

The null hypothesis is that the rainfall of the two sample unseeded and seeded clouds are identical populations. As the p-value turns out to be 0.006131, and is less than the .05 significance level, we again reject the null

hypothesis.

```
ks.test(a,b,paired=TRUE)
```

```
## Warning in ks.test(a, b, paired = TRUE): cannot compute exact p-value with
## ties
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  a and b
## D = 0.42308, p-value = 0.01905
## alternative hypothesis: two-sided
```

Here, the same conclusion can be made, as the p-value for is below 0.05. In the present case, applying these three tests might not be applicable, as it could be proved in practice that seeding clouds would be an improvement for the amount of rainfall. What is lacking here is maybe the inclusion of other variables important for such rise of rainfall, such as local temperature, wind, among others. Or simply, make use of other procedures to calculate an eventual evidence.

2

Now, we will take the squared root of the values in the file.

```
a_2=sqrt(a)
b_2=sqrt(b)
```

```
t.test(a_2,b_2, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  a_2 and b_2
## t = 2.682, df = 25, p-value = 0.01278
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   1.656326 12.617061
## sample estimates:
## mean of the differences
##                7.136693
```

```
wilcox.test(a_2,b_2,paired=TRUE)
```

```
##
##  Wilcoxon signed rank test
##
## data:  a_2 and b_2
## V = 276, p-value = 0.009383
## alternative hypothesis: true location shift is not equal to 0
```

```
ks.test(a_2,b_2,paired=TRUE)
```

```
## Warning in ks.test(a_2, b_2, paired = TRUE): cannot compute exact p-value
## with ties
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
```

```
## data:  a_2 and b_2
## D = 0.42308, p-value = 0.01905
## alternative hypothesis: two-sided
```

The values for the t.test dropped from 0.04407 to 0.01278; for the Wilcoxon, it increased from 0.006131 to 0.009383, and for the Kolmogorov-Smirnov, it remained the same.

3

Finally, the squared root of the squared root of the values in the file are considered.

```
a_3=sqrt(a_2)
b_3=sqrt(b_2)
```

```
t.test(a_3,b_3, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  a_3 and b_3
## t = 2.8413, df = 25, p-value = 0.008811
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.2673422 1.6759523
## sample estimates:
## mean of the differences
##                0.9716472
```

```
wilcox.test(a_3,b_3,paired=TRUE)
```

```
##
##  Wilcoxon signed rank test
##
## data:  a_3 and b_3
## V = 279, p-value = 0.007291
## alternative hypothesis: true location shift is not equal to 0
```

```
ks.test(a_3,b_3,paired=TRUE)
```

```
## Warning in ks.test(a_3, b_3, paired = TRUE): cannot compute exact p-value
## with ties
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  a_3 and b_3
## D = 0.42308, p-value = 0.01905
## alternative hypothesis: two-sided
```

Now, the values for the t.test dropped from 0.01278 to 0.008811; for the Wilcoxon, it decreased from 0.009383 to 0.007291, and for the Kolmogorov-Smirnov, it still remained the same.
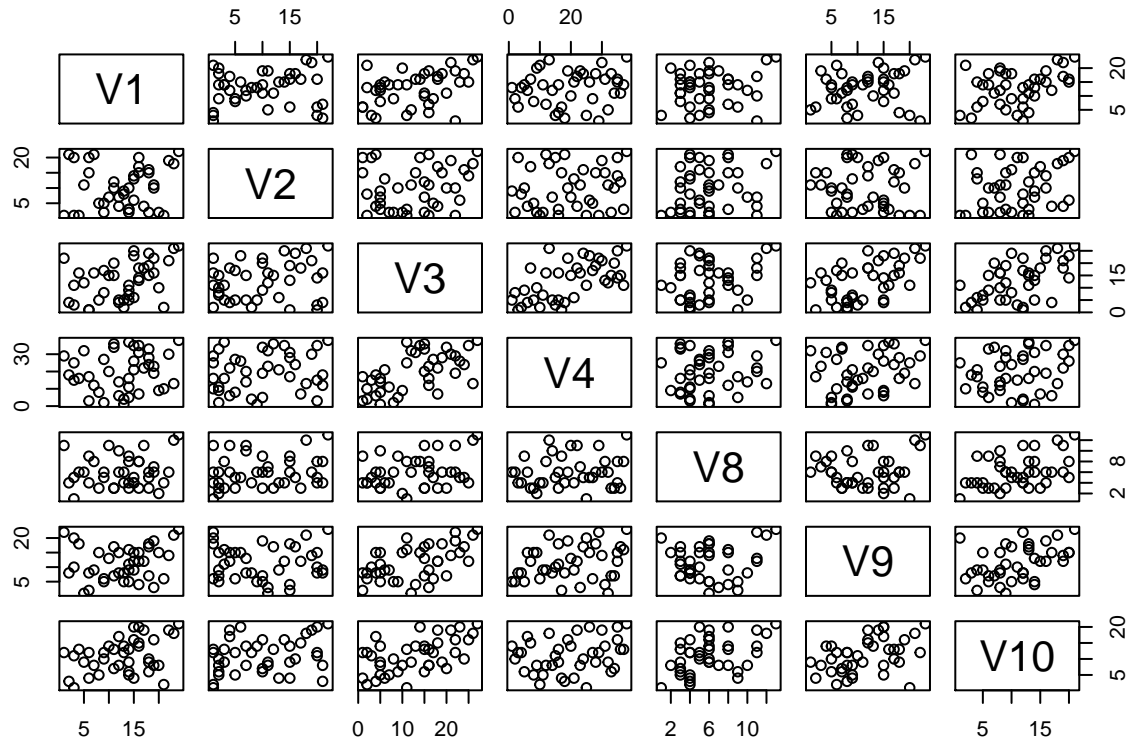
Assignment 3

Exercise 1

1

In this subsection, we will explore the file "peruvians.txt" with the data of Peruvian men after migrating to a modern society.

```
peruvians = read.table('peruvians.txt')
peruvians = peruvians[,-c(5,6,7)]
pairs(peruvians)
```


We obtain the pairs plot for the correlation between variables age (V1), weight (V3), length (V4), wrist (V8), systolic (V9) and diastolic (V10) with migration (V2). The area of interest is concentrated on the second row, and also second column. From this plot, we decided that only the variable "age" would be correlated to "migration", as it is the only which seemly respects an uprising behaviour, as in the oldest the person is, the longest he/she would be expected to having been under the status of migrant.

2

Now, we will study the correlation of these variables with "migration".

```
age=as.numeric(levels(peruvians$V1))[peruvians$V1]
```

```
## Warning: NAs introduced by coercion
```

```
age=age[2:end(age)]
```

```
## Warning in 2:end(age): numerical expression has 2 elements: only the first
## used
```

```
migration=as.numeric(levels(peruvians$V2))[peruvians$V2]
```

```
## Warning: NAs introduced by coercion
```

```
migration=migration[2:end(migration)]
```

```
## Warning in 2:end(migration): numerical expression has 2 elements: only the
## first used
```

```
weight=as.numeric(levels(peruvians$V3))[peruvians$V3]
```

Warning: NAs introduced by coercion
```
weight=weight[2:end(weight)]
```

Warning in 2:end(weight): numerical expression has 2 elements: only the
first used
```
length=as.numeric(levels(peruvians$V4))[peruvians$V4]
```

Warning: NAs introduced by coercion
```
length=length[2:end(length)]
```

Warning in 2:end(length): numerical expression has 2 elements: only the
first used
```
wrist=as.numeric(levels(peruvians$V8))[peruvians$V8]
```

Warning: NAs introduced by coercion
```
wrist=wrist[2:end(wrist)]
```

Warning in 2:end(wrist): numerical expression has 2 elements: only the
first used
```
systolic=as.numeric(levels(peruvians$V9))[peruvians$V9]
```

Warning: NAs introduced by coercion
```
systolic=systolic[2:end(systolic)]
```

Warning in 2:end(systolic): numerical expression has 2 elements: only the
first used
```
diastolic=as.numeric(levels(peruvians$V10))[peruvians$V10]
```

Warning: NAs introduced by coercion
```
diastolic=diastolic[2:end(diastolic)]
```

Warning in 2:end(diastolic): numerical expression has 2 elements: only the
first used
```
cor(age,migration)
```

[1] 0.5882125
```
cor(weight,migration)
```

[1] 0.4811534
```
cor(length,migration)
```

[1] 0.07259415
```
cor(wrist,migration)
```

[1] 0.2369046
```
cor(systolic,migration)
```

[1] -0.08748046

```
cor(diastolic,migration)
```

[1] 0.07579214

By the numbers here above, we can confirm what was before stated that the variable which is most likely to be correlated to migration is age. The other variables had a very low or even negative values. The variable "weight" had correlation of 0.4811, which is not as low as the others, however we can't consider here it being correlated to "migration" as this value is still low.
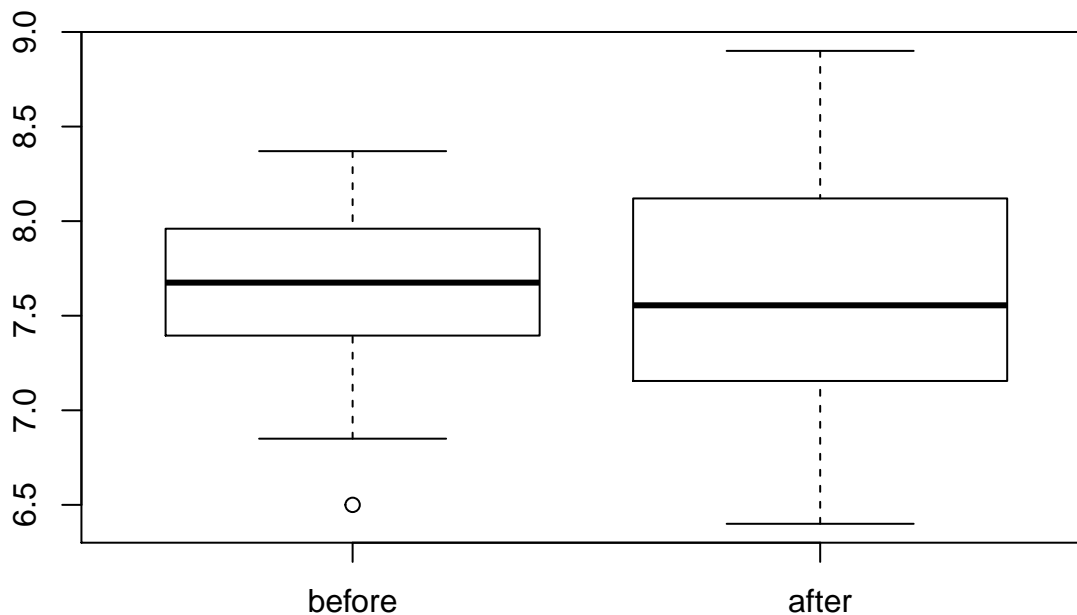
Exercise 2

The best way to represent this data is through separating it into two different datasets: one for "lemo" and another for "energy".

```
run = read.table('run.txt')
lemo = subset(run,drink=="lemo")[,-c(3)]
energy = subset(run,drink=="energy")[,-c(3)]
```
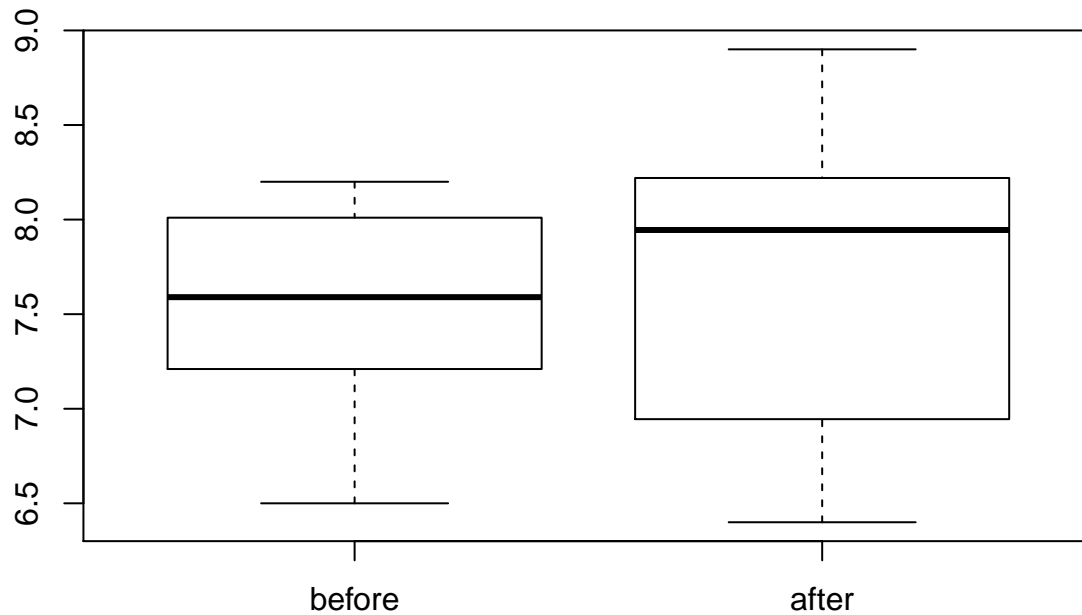
1

A graphical representation of the data is here shown:
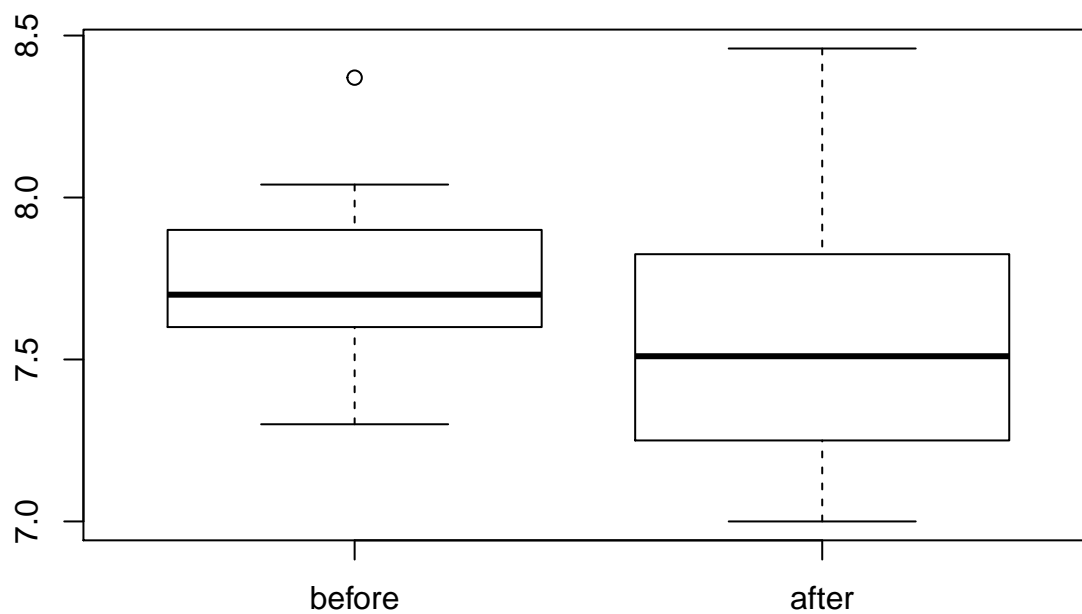
```
boxplot(run[,-c(3)])
```



```
par(mfrow=c(1,1))
boxplot(lemo)
```

14

```
boxplot(energy)
```


In general, the median of the boxplot after having drunk the lemon soda was higher, while for the energy drink it was lower. However, the upper and lower limits are limited in a wider range for both cases after having drunk the beverages.

2

```
t.test(lemo$before, lemo$after);
```

```
##
##   Welch Two Sample t-test
##
## data:  lemo$before and lemo$after
## t = -0.5036, df = 18.989, p-value = 0.6203
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.7476599  0.4576599
## sample estimates:
## mean of x mean of y
##  7.554167  7.699167
```

```
t.test(energy$before, energy$after);
```

```
##
##  Welch Two Sample t-test
##
## data:  energy$before and energy$after
## t = 0.97741, df = 18.854, p-value = 0.3407
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1761393  0.4844726
## sample estimates:
## mean of x mean of y
##  7.732500  7.578333
```

The results of the t.test show that, for those having drunk "lemo", the hypothesis H0 which states that the mean values before and after drinking the beverage are equal, is maintained as the p-value of 0.6203 is clearly higher than 0.05. For the beverage "energy", the p-value of 0.3407 maintains the hypothesis H0, however with lower value, which could suggest that, for instance, the energy could have made a difference in the overall performance.

3

```
run$difference = run$after - run$before
difference_lemo = subset(run,drink=="lemo")$difference
difference_energy = subset(run,drink=="energy")$difference
```

```
t.test(difference_lemo, difference_energy)
```

```
##
##  Welch Two Sample t-test
##
## data:  difference_lemo and difference_energy
## t = 1.4764, df = 16.509, p-value = 0.1586
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1293076  0.7276409
## sample estimates:
##  mean of x  mean of y
##   0.1450000 -0.1541667
```

Effectively, even though the hypothesis H0 is not rejected here, the p-value is still low and the means of the differences for lemo and energy shown in the report state that there is a significant difference, specially for the latter one being negative. Thus, the time differences can be considered to be affected by the beverages.

4

One objection would be, since it is expected that the energy drink takes some time to take effect in the pupils' metabolism, the data collected could have also considered that the same amount of beverage can be absorbed

differently by the pupils and thus, the results might end misleading or even not leading to any conclusion whatsoever.

5

For the same reasons, the metabolism of the pupils react differently. For instance, regardless of what beverage was consumed, a very slim and short person would not present a difference proportional to a chubbier person with more height, under normal circumstances. Thus, the height and the weight of the pupils could have been taken into account.

6

The QQ-plots for the differences are given here after.

```
par(mfrow=c(1,2))
qqnorm(difference_lemo)
qqnorm(difference_energy)
```


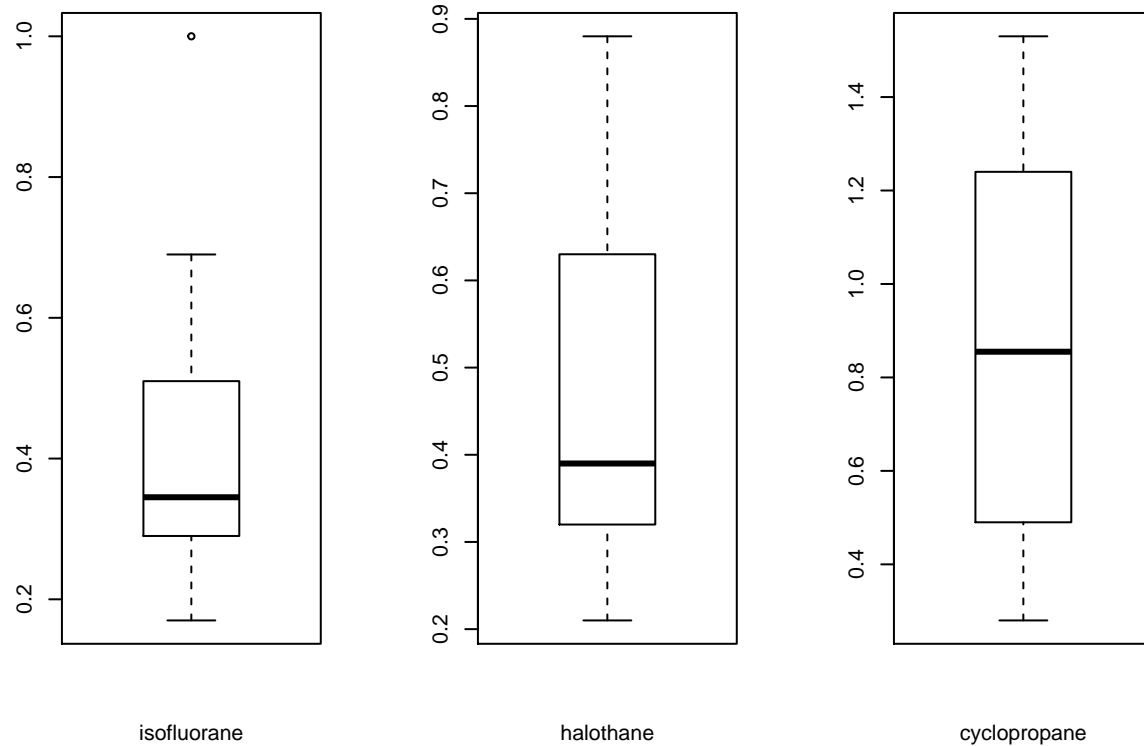**Normal Q–Q Plot**
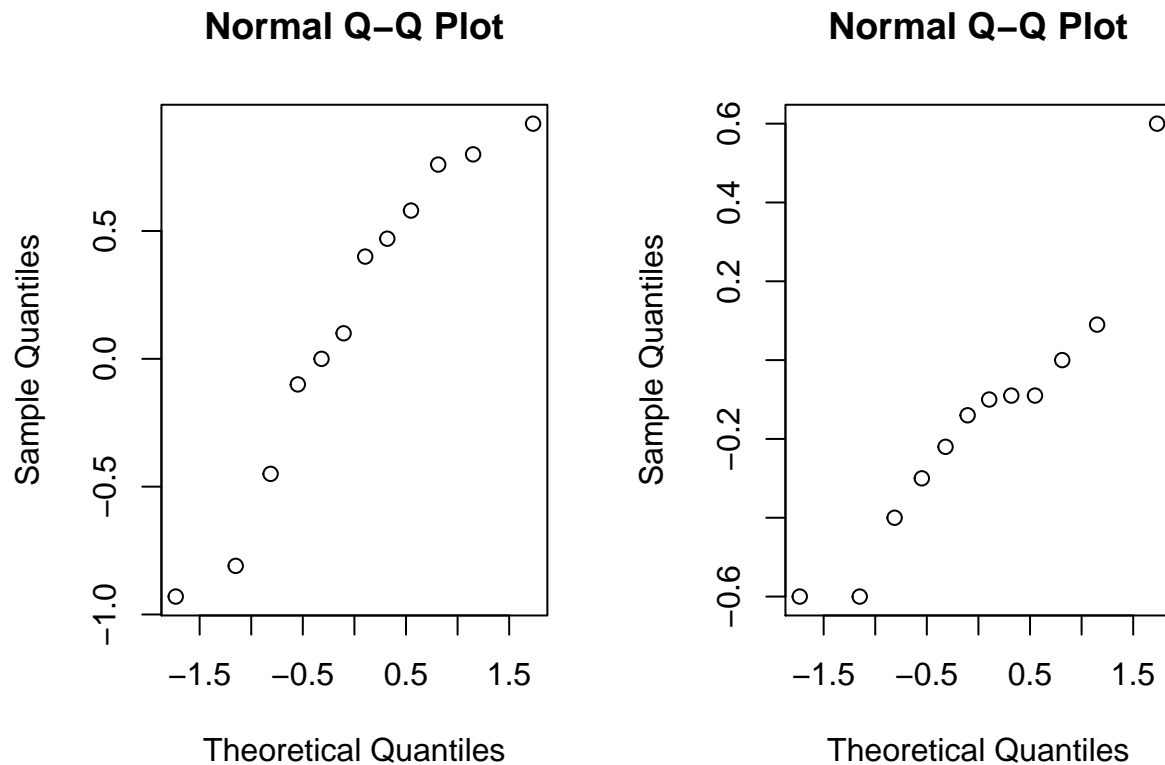
Normal Q–Q Plot

Exercise 3

```
dogs <- read.table("dogs(2).txt", header=TRUE, quote="\"")
```

1

Please find here after the boxplots and the QQ-plots of the samples, as well as the QQ-plots of the normal distributions with the same sizes of the respective samples.

```
par(mfrow=c(1,3))

boxplot(dogs$isofluorane, xlab="isofluorane")
boxplot(dogs$halothane, xlab="halothane")
boxplot(dogs$cyclopropane, xlab="cyclopropane")
```
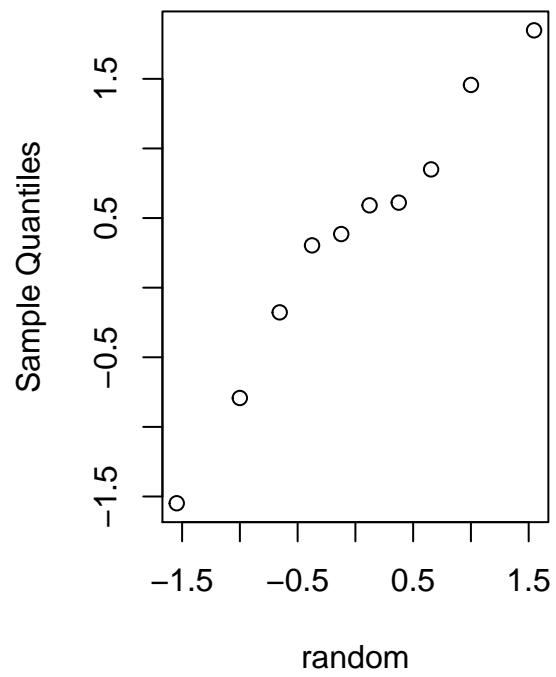


```
par(mfrow=c(1,2))

# Create QQ-plots of Isofluorane and normal distribution
qqnorm(dogs$isofluorane, xlab="sample", main="QQ-plot Isofluorane")
qqnorm(rnorm(length(dogs$isofluorane)), xlab="random", main="QQ-plot Isofluorane")
```
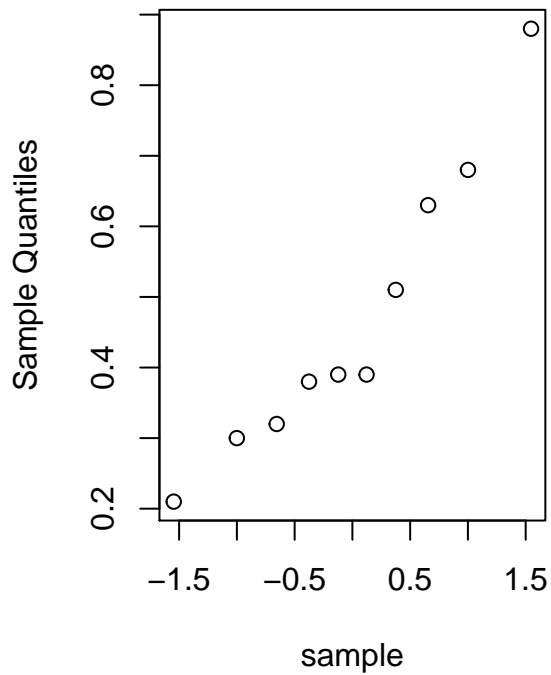
QQ–plot Isofluorane
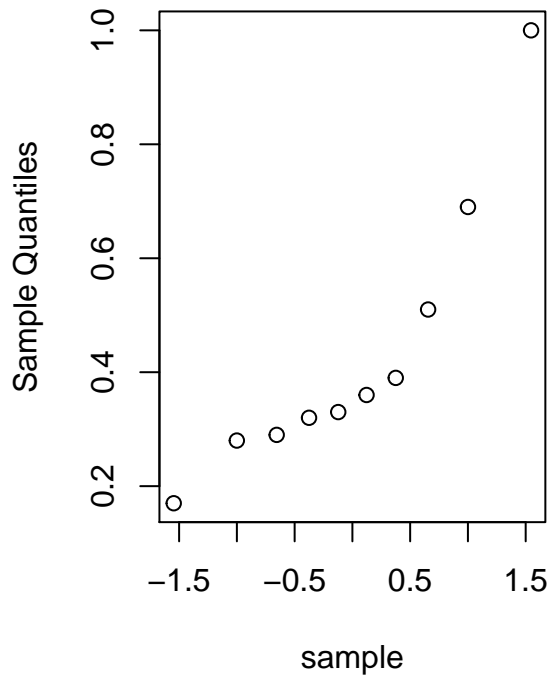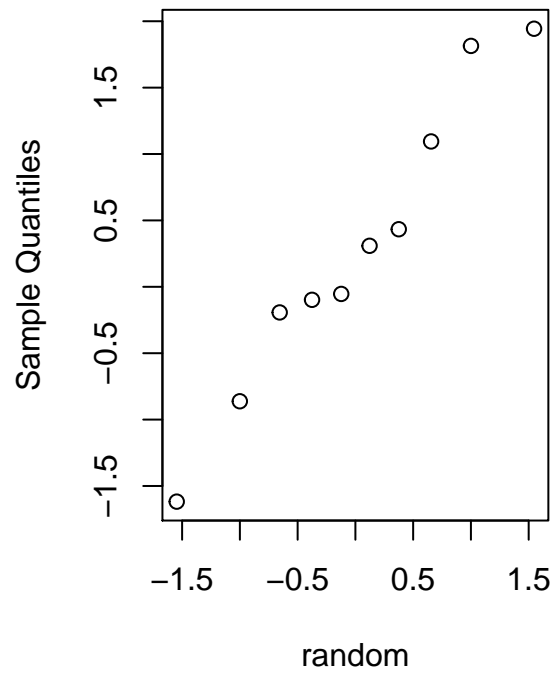

QQ–plot Isofluorane


```r
# Create QQ-plots of Halothane and normal distribution
qqnorm(dogs$halothane, xlab="sample", main="QQ-plot Halothane")
qqnorm(rnorm(length(dogs$halothane)), xlab="random", main="QQ-plot Halothane")
```
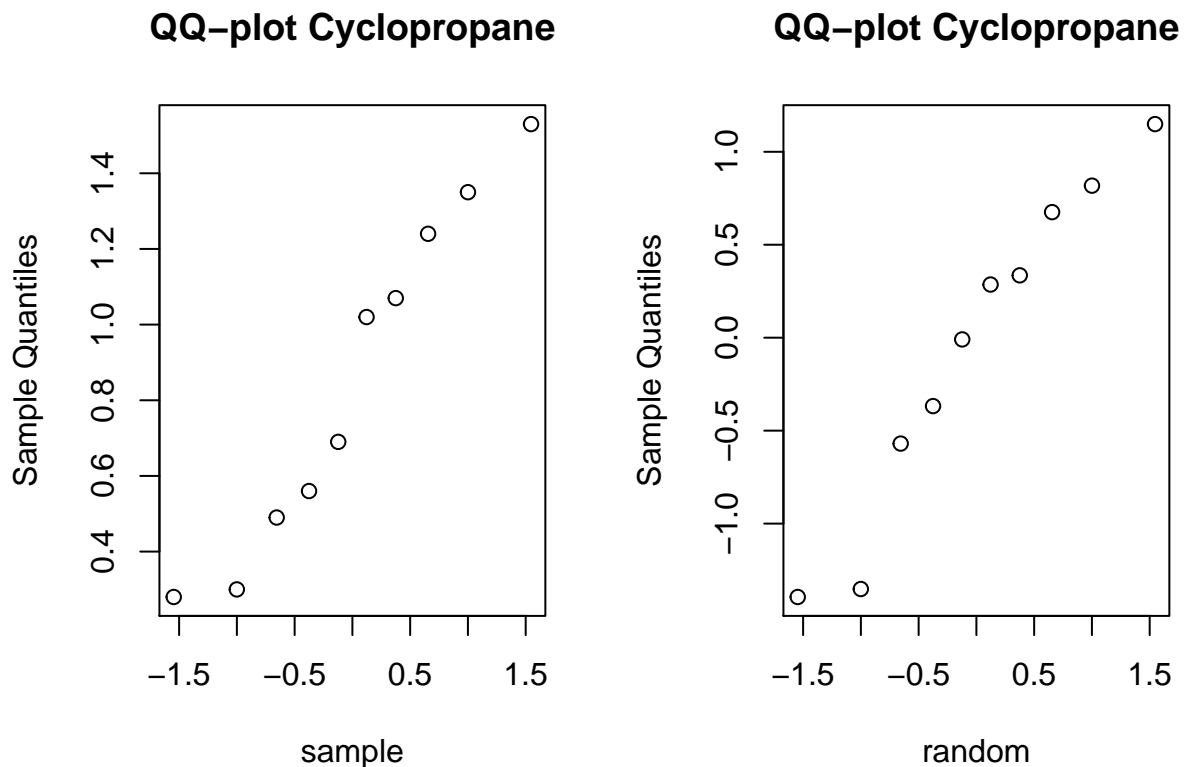
QQ–plot Halothane

QQ–plot Halothane

```
# Create QQ-plots of Cycloprane and normal distribution
qqnorm(dogs$cyclopropane, xlab="sample", main="QQ-plot Cyclopropane")
qqnorm(rnorm(length(dogs$cyclopropane)), xlab="random", main="QQ-plot Cyclopropane")
```


QQ–plot Cyclopropane

QQ–plot Cyclopropane

After analysing the three comparisons between QQ-plots of the drugs and the normal distributions, it is reasonable to state that isufluorane values do not follow a normal distribution, while the other two do. This is explained by the fact that the former one does not really respect a behavior of an uprising line in the QQ-plot graph.

2

Now, we will try to check the hypothesis that is the same under the different drugs.

```
t.test(dogs$isofluorane,dogs$halothane)
```

```
##
##  Welch Two Sample t-test
##
## data:  dogs$isofluorane and dogs$halothane
## t = -0.34687, df = 17.482, p-value = 0.7328
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2474409  0.1774409
## sample estimates:
## mean of x mean of y
##     0.434     0.469
```

```
t.test(dogs$isofluorane,dogs$cyclopropane)
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  dogs$isofluorane and dogs$cyclopropane
## t = -2.5955, df = 13.912, p-value = 0.02125
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.76544404 -0.07255596
## sample estimates:
## mean of x mean of y
##     0.434     0.853
```

```r
t.test(dogs$cyclopropane,dogs$halothane)
```

```
##
##  Welch Two Sample t-test
##
## data:  dogs$cyclopropane and dogs$halothane
## t = 2.463, df = 12.616, p-value = 0.02901
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.04613337 0.72186663
## sample estimates:
## mean of x mean of y
##     0.853     0.469
```

Thus, the results show that the p-value for assuming that the means of concentrations of plasma epinephrine under isofluorane and halothane is 0.7328, which is proof that this hypothesis can not be rejected. For the other two t.tests, the p-values are lower than 0.05, which unables us to accept the hypothesis that the respective pairwise instances in the t.tests have the same mean. So as to better picture this fact, find here after the estimated concentrations of plasma epinephrine for each scenario, which confirms the results so-far obtained:

```r
mean_iso=mean(dogs$isofluorane)
mean_halo=mean(dogs$halothane)
mean_cyclo=mean(dogs$cyclopropane)
mean_iso
```

```
## [1] 0.434
```

```r
mean_halo
```

```
## [1] 0.469
```

```r
mean_cyclo
```

```
## [1] 0.853
```

3

First, we create data frames as numeric and factor.

```r
dogsframeiso = dogs[,-c(2,3)] # subsetting only iso
isofluorane=as.vector(as.matrix(dogs))
group1=as.factor(as.numeric(rep(1:3,each=10)))
dogsframeiso = data.frame(isofluorane, group1)
# dogsframeiso = dogsframeiso[1:10,]

dogsframehalo = dogs[,-c(1,3)] # subsetting only halo
halothane=as.vector(as.matrix(dogs))
group2=as.factor(as.numeric(rep(1:3,each=10)))
```

```
dogsframehalo = data.frame(halothane,group2)
# dogsframehalo = dogsframehalo[12:21,]


dogsframecyclo = dogs[,-c(1,2)] # subsetting only cyclo
cyclopropane=as.vector(as.matrix(dogs))
group3=as.factor(as.numeric(rep(1:3,each=10)))
dogsframecyclo = data.frame(cyclopropane,group3)
# dogsframecyclo = dogsframecyclo[22:30,]

kruskal.test(dogsframeiso,group1)
```

```
## Warning in kruskal.test.default(dogsframeiso, group1): 'x' is a list, so
## ignoring argument 'g'
```

```
## Warning in kruskal.test.default(dogsframeiso, group1): some elements of 'x'
## are not numeric and will be coerced to numeric
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  dogsframeiso
## Kruskal-Wallis chi-squared = 34.642, df = 1, p-value = 3.963e-09
```

```
kruskal.test(halothane,group2)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  halothane and group2
## Kruskal-Wallis chi-squared = 5.6442, df = 2, p-value = 0.05948
```

```
kruskal.test(cyclopropane,group3)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  cyclopropane and group3
## Kruskal-Wallis chi-squared = 5.6442, df = 2, p-value = 0.05948
```

We conclude that the first Kruskal-Wallis test rejects the null hypothesis that the concentration is the same under the different drugs. The two others do not reject this hypothesis. This is different from the previous conclusion because the previous test assumed normality while this test did not.