

Assignment 3, EDDA 2017

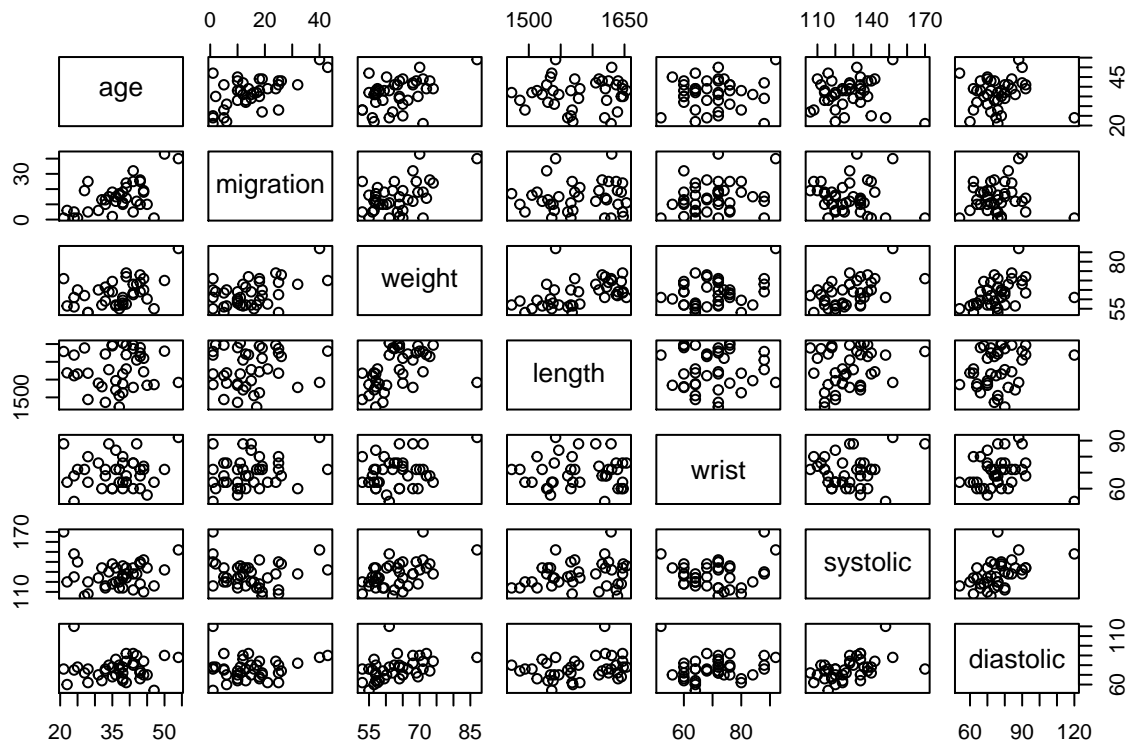
Martin de la Riva(11403799) and Kieran O'Driscoll(11426438), Group 23

24 April 2017

Exercise 1

Q1

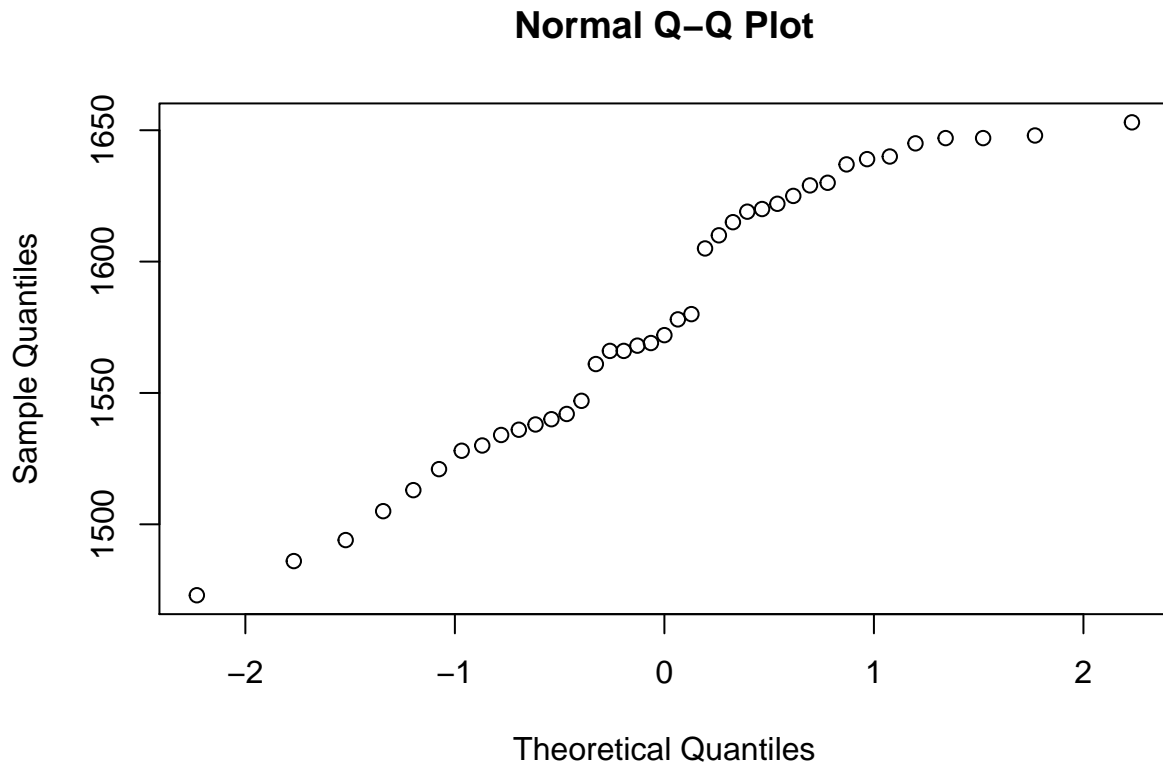
```
data = read.table("data/peruvians.txt", header=TRUE)
pairs(data[, -c(5, 6, 7)])
```



Based on the diagram above, age, weight and perhaps a case can be made for diastolic. These were chosen because their scatter plots show a cluster that shows the values are in proportion with migration whereas the other graphs have a scatter plot that don't show any connection with migration.

Q2

```
qqnorm(data[['length']])
```



```
attach(data)
```

Peruvians is not drawn from a normal distribution so therefore will use Spearman correlation test to test for correlation.

```
cor.test(age,migration,method="spearman")
```

```
## Warning in cor.test.default(age, migration, method = "spearman"): Cannot
## compute exact p-value with ties

##
## Spearman's rank correlation rho
##
## data: age and migration
## S = 5176.6, p-value = 0.002189
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.4760575
```

The p-value received is 0.002189 which falls under the 0.05 significant level. Therefore the value is correlated with migration. Since the value is quite low, it can be said there is a high positive correlation as the correlation increases as values get bigger.

```
cor.test(length,migration,method="spearman")
```

```
## Warning in cor.test.default(length, migration, method = "spearman"): Cannot
## compute exact p-value with ties
```

```
##
## Spearman's rank correlation rho
##
## data: length and migration
## S = 9044.3, p-value = 0.6087
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.08458432
```

With a P-value of 0.6087 it's clear that length and migration do not have a correlation.

```
cor.test(wrist,migration,method="spearman")
```

```
## Warning in cor.test.default(wrist, migration, method = "spearman"): Cannot
## compute exact p-value with ties
```

```
##
## Spearman's rank correlation rho
##
## data: wrist and migration
## S = 7712.8, p-value = 0.1797
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.2193498
```

Wrist gets a p-value of 0.1797 which means there is no correlation.

```
cor.test(systolic,migration,method="spearman")
```

```
## Warning in cor.test.default(systolic, migration, method = "spearman"):
## Cannot compute exact p-value with ties
```

```
##
## Spearman's rank correlation rho
##
## data: systolic and migration
## S = 11544, p-value = 0.3054
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.1684286
```

Systolic has no correlation with migration, p-value= 0.3054

```
cor.test(diastolic,migration,method="spearman")
```

```
## Warning in cor.test.default(diastolic, migration, method = "spearman"):  
## Cannot compute exact p-value with ties
```

```
##  
## Spearman's rank correlation rho  
##  
## data: diastolic and migration  
## S = 9137.6, p-value = 0.6494  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho  
## 0.07514098
```

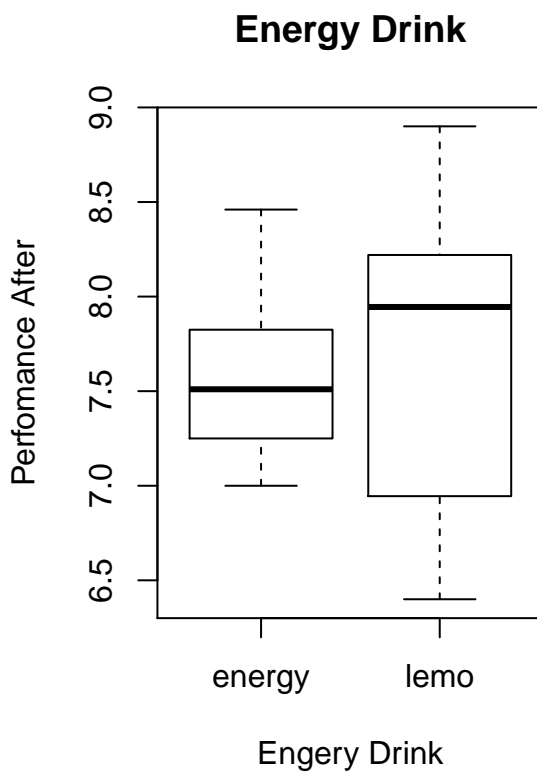
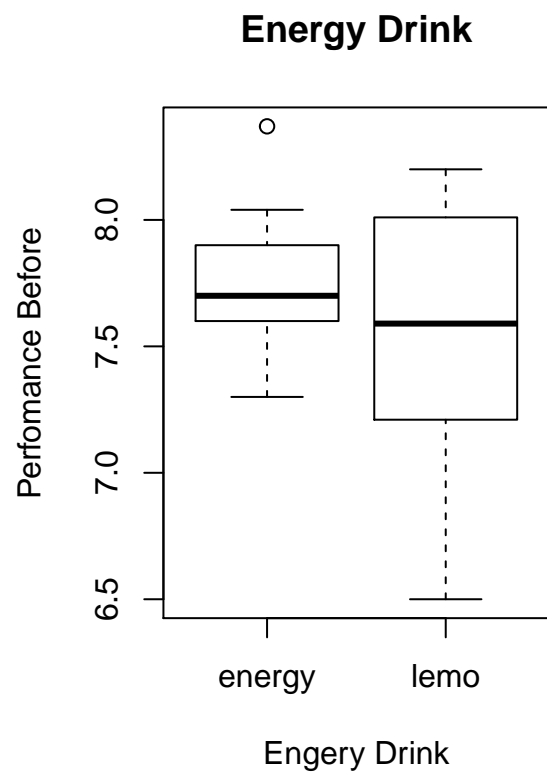
Diastolic has no correlation, the p-value was 0.6494

```
cor.test(weight,migration,method="spearman")
```

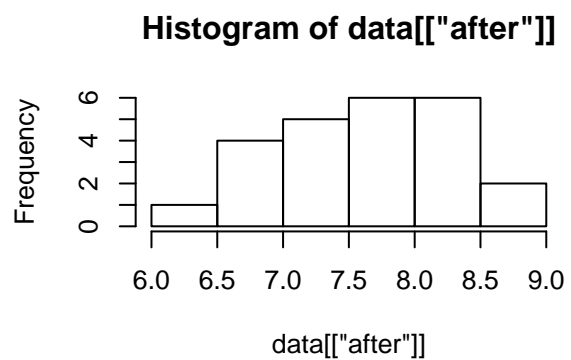
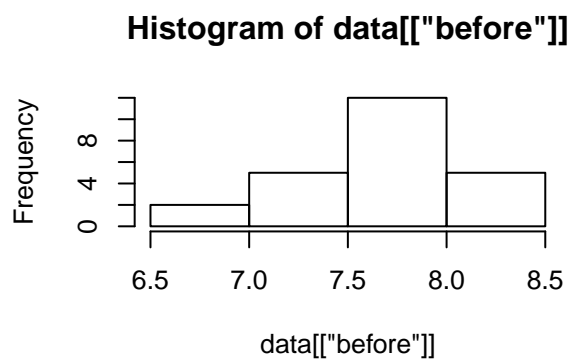
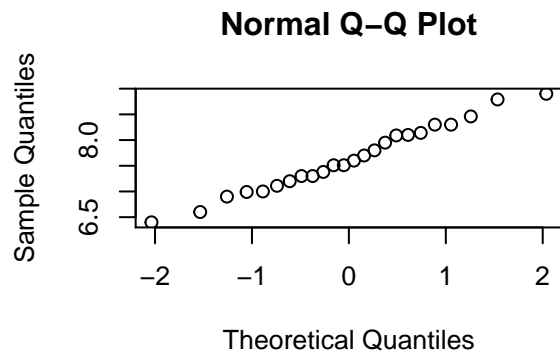
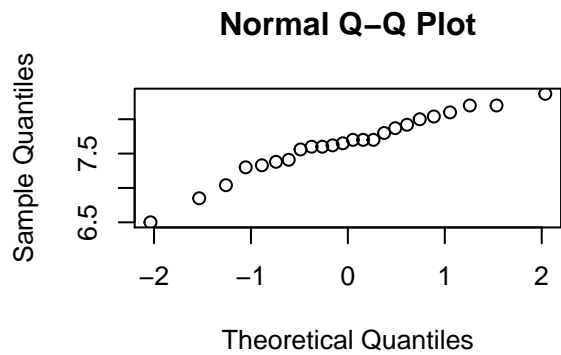
```
## Warning in cor.test.default(weight, migration, method = "spearman"): Cannot  
## compute exact p-value with ties
```

```
##  
## Spearman's rank correlation rho  
##  
## data: weight and migration  
## S = 6415.1, p-value = 0.02861  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho  
## 0.3506956
```

Lastly, weight is correlated with migration having a p-value of 0.02861. Two of the tree values predicted in question one were correlated, bith weight and age. ##Exercise 1 ##Q1

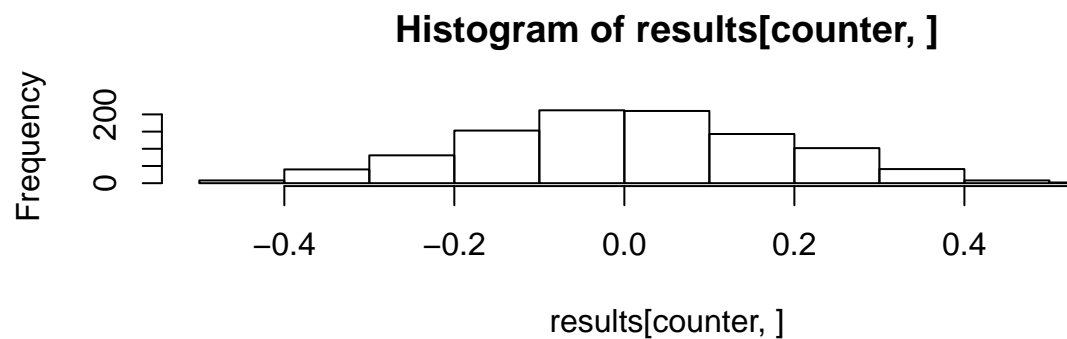
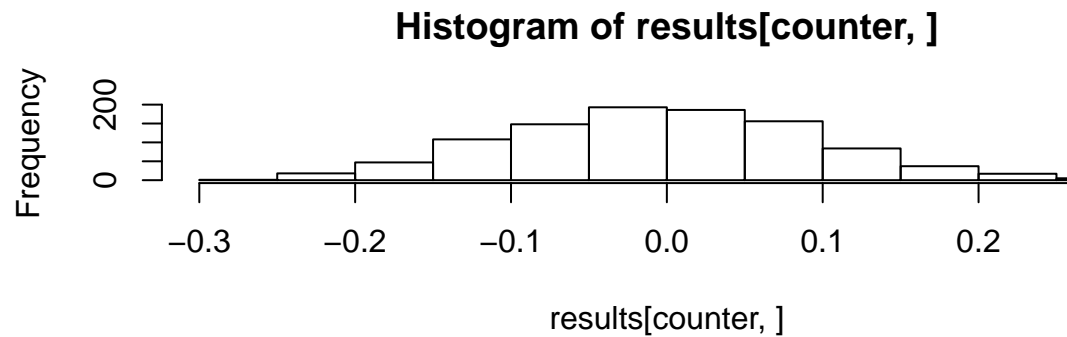


From the boxplots it indicates that The engery actaully makes students worse after 60 mins and the lemonade makes students better.



Data looks like it was drawn from a normal distribution. With the exception of the histogram for “after” but there are only 12 entries per drink type which is a small amount to guage whether this is drawn from a normal distribution

Q2



Using a two paired sample test

```
print("Energy")
```

```
## [1] "Energy"
```

```
print(value[1])
```

```
## [1] 0.1541667
```

```
p1=sum(results[1,<value[1]])/B  
pr=sum(results[1,>value[1]])/B  
p=2*min(p1,pr)  
print(p)
```

```
## [1] 0.108
```

```
print("Lemo")
```

```
## [1] "Lemo"
```

```
print(value[2])
```

```
## [1] -0.145
```

```
p1=sum(results[2,<value[2]])/B  
pr=sum(results[2,>value[2]])/B
```

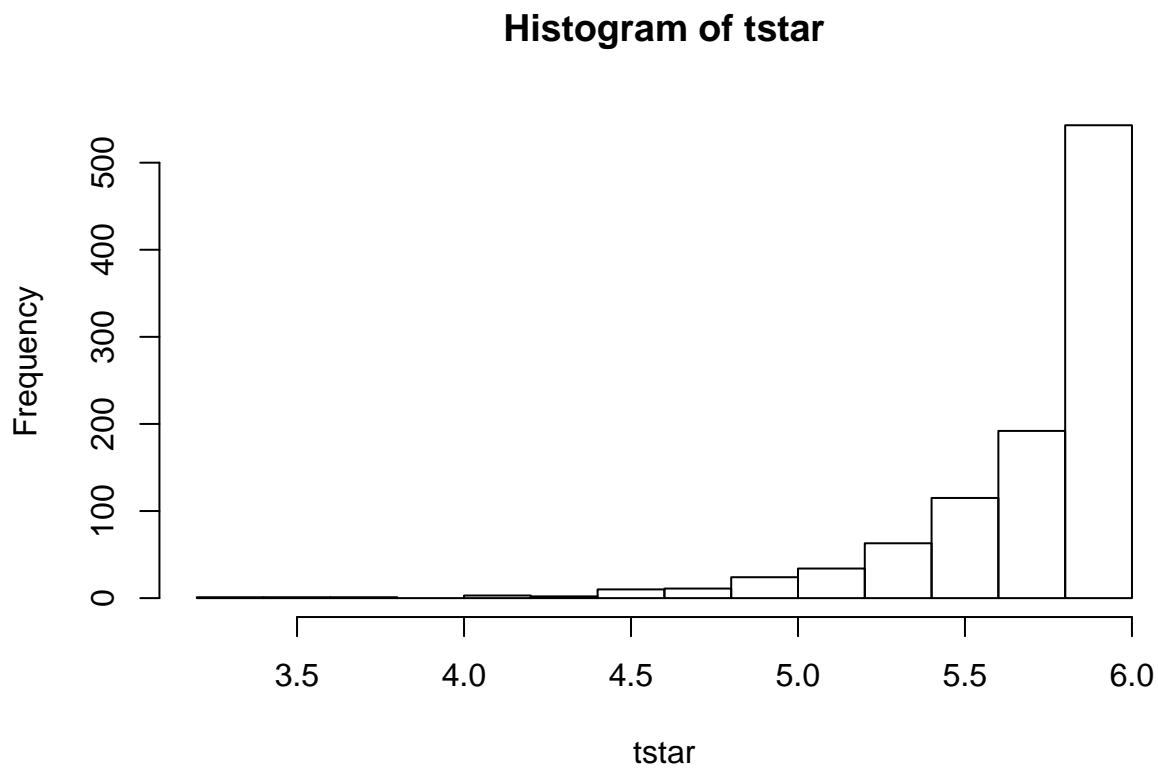
```
p=2*min(pl,pr)
print(p)
```

```
## [1] 0.402
```

For both drinks there does not seem to be a significant difference between “before” and “after”. The p-values obtained are 0.128 and 0.4 for “energy” and “lemo” respectively.

Q3

Using permutation test permutation for independent samples.



```
## [1] 5.419592
```

```
## [1] 0.162
```

The drink type does not affect the time difference from “before” to “after”.

Q4

There is only a small sample size, each drink gets allocated 12 people, it could be the case that one group were faster on average therefore would be faster before and after whereas some people might

need more than 30 minutes to recover. Also, perhaps a bigger margin then 30 minutes for the gap between before and after.

Q5

It would have affected the time difference if people were still tired after the first run.

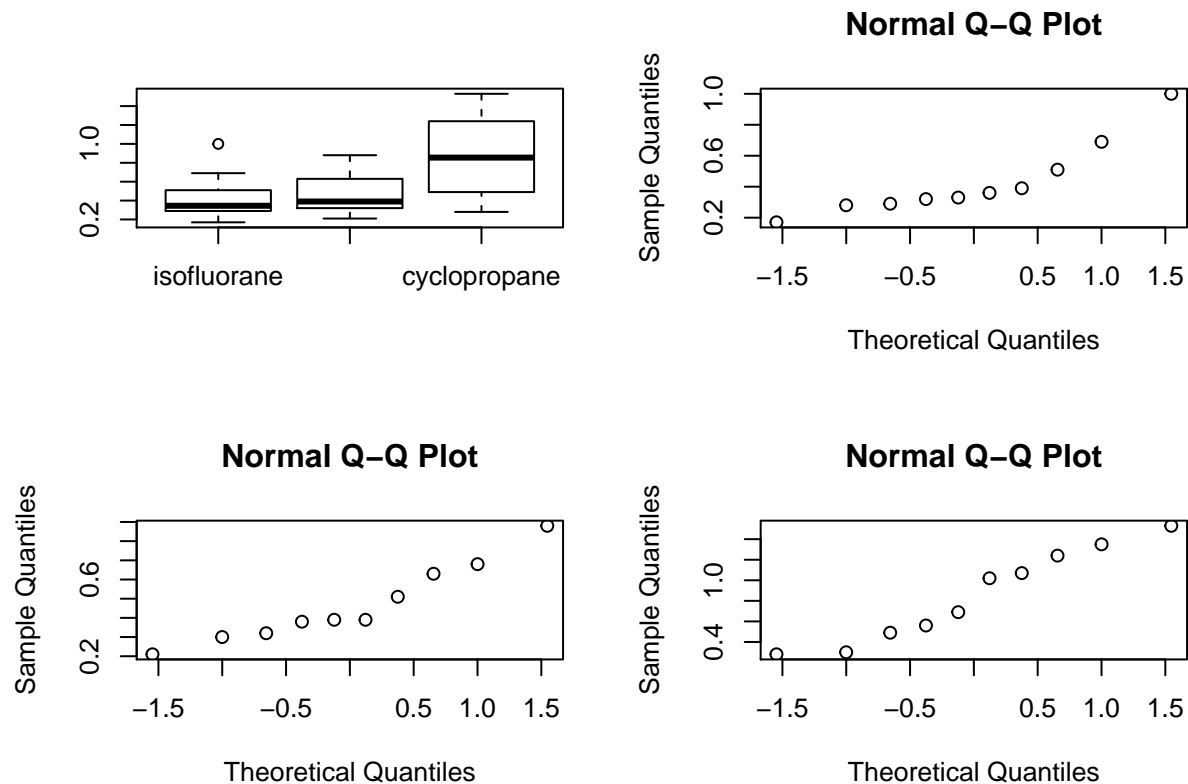
Q6

?? Used paiered sample which doesn't need normal distribution.

Exercise 3

Q1

```
data = read.table("data/dogs.txt", header=TRUE)
par(mfrow=c(2,2))
boxplot(data,data=data)
qqnorm(data[['isofluorane']])
qqnorm(data[['halothane']])
qqnorm(data[['cyclopropane']])
```



Each drug type has 10 examples. All qqplots are close to a normal distribution with the exception of isofluorane which is displayed in top right. It can be presumed that by adding more data that the qqplots would more closely resembles a qqplot of a normal distribution. In this case it is reasonable to assume sample were taken from normal distribution.

Q2

Using anova.

```
attach(data)
pvcaov=lm(isofluorane~halothane*cyclopropane,data=data)
anova(pvcaov)
```

Analysis of Variance Table

##

Response: isofluorane

##

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
## halothane	1	0.01346	0.013456	0.1605	0.7026
## cyclopropane	1	0.00819	0.008193	0.0977	0.7652
## halothane:cyclopropane	1	0.01233	0.012333	0.1471	0.7145
## Residuals	6	0.50306	0.083843		

Q3

```
kruskal.test(isofluorane,halothane)

##
##  Kruskal-Wallis rank sum test
##
## data:  isofluorane and halothane
## Kruskal-Wallis chi-squared = 8.9455, df = 8, p-value = 0.3469

kruskal.test(isofluorane,cyclopropane)

##
##  Kruskal-Wallis rank sum test
##
## data:  isofluorane and cyclopropane
## Kruskal-Wallis chi-squared = 9, df = 9, p-value = 0.4373

kruskal.test(cyclopropane,halothane)

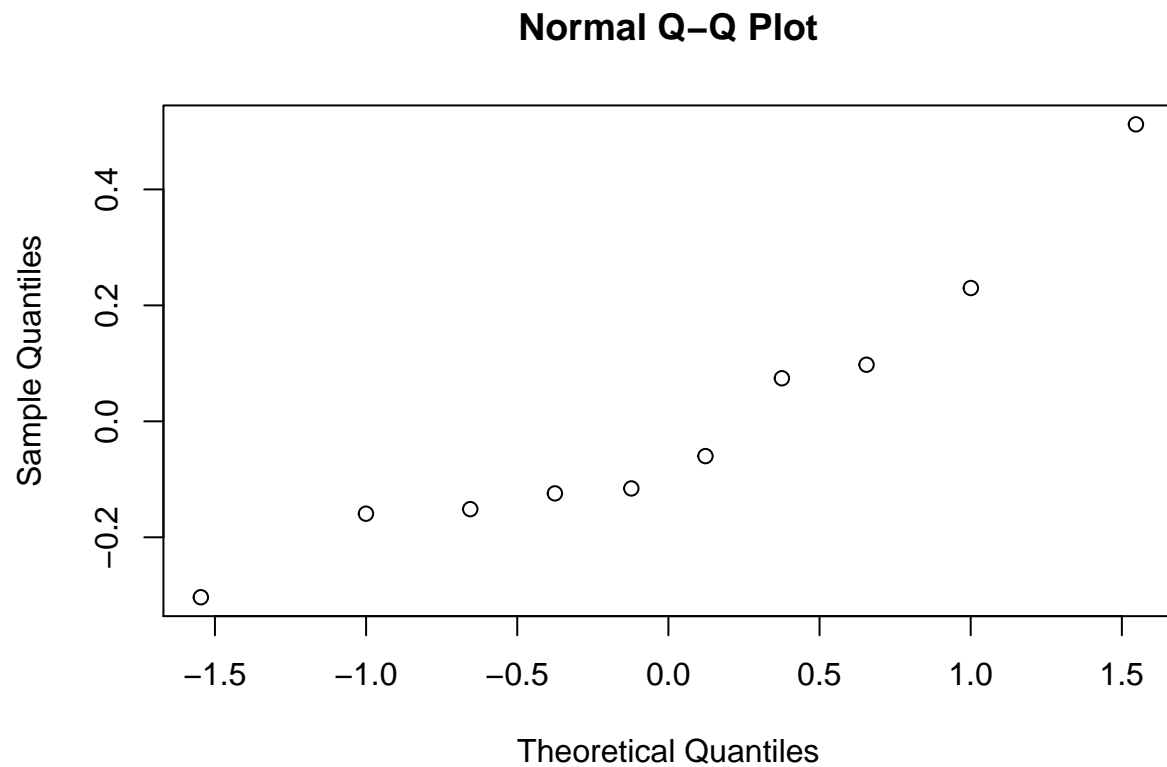
##
##  Kruskal-Wallis rank sum test
##
## data:  cyclopropane and halothane
## Kruskal-Wallis chi-squared = 7.6364, df = 8, p-value = 0.4698

pvcaov=lm(isofluorane~halothane*cyclopropane,data=data)
anova(pvcaov)

## Analysis of Variance Table
##
## Response: isofluorane
##
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
## halothane	1	0.01346	0.013456	0.1605	0.7026
## cyclopropane	1	0.00819	0.008193	0.0977	0.7652
## halothane:cyclopropane	1	0.01233	0.012333	0.1471	0.7145
## Residuals	6	0.50306	0.083843		

```
qqnorm(pvcaov$residuals)
```



The alnova ANOVA produce significantly different results than the Kruskal-Wallis. The qqnorm of the residuals deviates from the normal distrubution. This woukd account for the signifant difference in the results so in this case it would be better to use Kruskal-Wallis.