

# Assignment 1

Group 67: Jamie Faber, Laurens Prast, Amal Salman

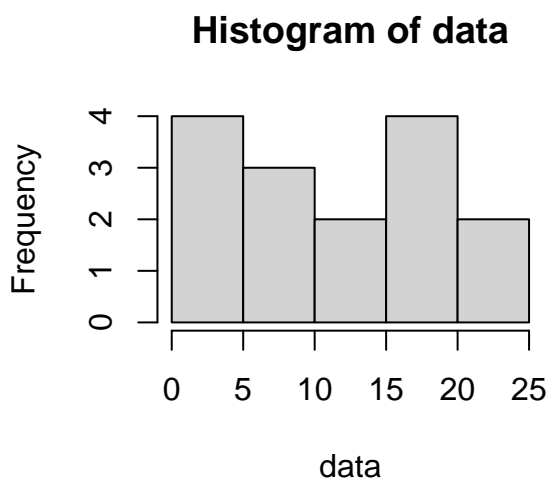
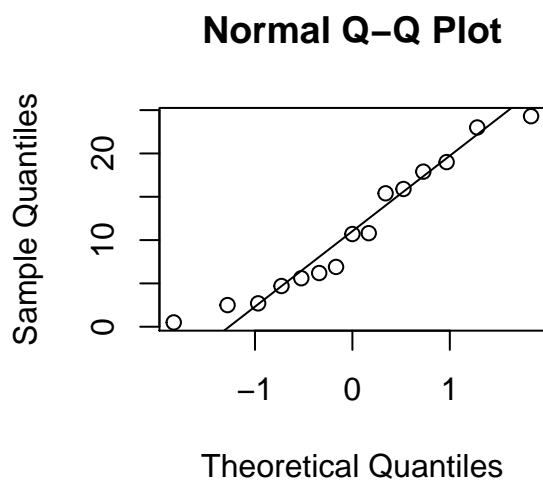
27 February 2022

## Imports

### Exercise 1

a) Checking normality of the data:

```
par(mfrow=c(1,2))
qqnorm(data); qqline(data); hist(data)
```



```
shapiro.test(data);
```

```
## W = 0.9, p-value = 0.3
```

Due to small sample size the plots vary more, however they don't show a clear non normal distribution nor does Shapiro's test find significance. P-value larger than confidence level: 0.05, therefore can state that there exists no significant departure from a normal distribution in the data set.

**Constructing a 97%-CI for  $\mu$ :**

```
mean_data = mean(data)
std_data = sd(data)
n = length(data)

error = qnorm(0.985)*std_data/sqrt(n)
left = mean_data-error
right = mean_data+error
```

```
c(left, mean_data, right)      #mean and confidence interval
```

```
## [1]  6.74 11.07 15.40
```

0.985 is used since 2-sided and confidence interval of 0.97. Normal distribution is used since we can assume normality.

**Evaluating the sample size needed to provide that the length of the 97%-CI is at most 2:**

```
E = 1
min_sample_size = (qnorm(0.985)^2*std_data^2)/ E^2
min_sample_size
```

```
## [1] 281
```

$2E=2$  so  $E=1$ . And normal distribution is used since we can assume normality. Therefore, the sample size needed is estimated to be 281.

**Computing a bootstrap 97%-CI for  $\mu$ :**

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

Tstar15=quantile(Tstar,0.015)
Tstar985=quantile(Tstar,0.985)
```

If we have a (small) sample from an unknown distribution and the distribution of  $X^-$  is not close to normal, we cannot rely on the above (asympt.) normal CI. and we can use a bootstrap method.

**Comparing t-test and bootstrap method:**

```
c(left, mean_data, right)
```

```
## [1]  6.74 11.07 15.40
```

```
c(2*mean_data-Tstar985,mean(Tstar),2*mean_data-Tstar15)
```

```
## 98.5%      1.5%
```

```
##  7.07 11.08 15.13
```

The bootstrap method has a smaller 97%-confidence interval compared to the method were we use the normal distribution.

**b) T-test to verify the claim that the mean waiting time is less than 15 minutes**

Assuming normality we can do the following t-test. We can state that there exists no significant departure from normality according to the Shapiro test from part (a).

```
mu0 = 15
t.test(data,mu=mu0,alt="less")
```

```
## t = -2, df = 14, p-value = 0.03
## alternative hypothesis: true mean is less than 15
## 95 percent confidence interval:
##  -Inf 14.6
## sample estimates:
## mean of x
##      11.1
```

Since  $p\text{-value} < 0.05$  we reject  $H_0$ , which was that  $H_0$  is equal to or larger than 15, and therefore accept  $H_1$  (that  $\text{mean} < 15$ ). We can see that the 95% confidence interval for the population mean is  $[-\infty; 14.59]$  which means that at a significance level  $\alpha = 0.05$  we reject null hypothesis as long as the hypothesized value  $\mu_0$  is below 14.59, otherwise the null hypothesis cannot be rejected.

### Sign tests:

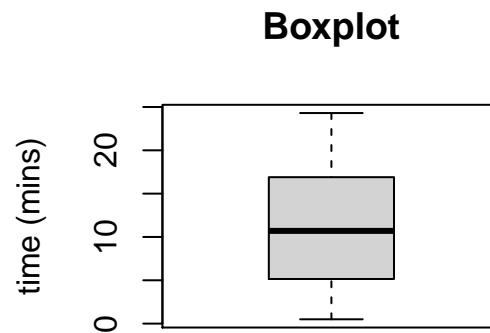
Binomial sign test:

```
larger_than_mu = sum(data>mu0)
samples = length(data)
binom.test(larger_than_mu,samples,p=0.5, alt='less')

## number of successes = 6, number of trials = 15, p-value = 0.3
## alternative hypothesis: true probability of success is less than 0.5
## 95 percent confidence interval:
##  0.00 0.64
## sample estimates:
## probability of success
##                0.4
```

Since  $p\text{-value} > 0.05$  we do not reject  $H_0$ . We do not have sufficient evidence to say that the patients are less likely to wait less than 15 min than more than 15 min with the binomial sign test.

```
boxplot(data, main="Boxplot",
        ylab="time (mins)")
```



While the histogram in a) does not seem symmetric, the above boxplot does. Also since a normal distribution is symmetric, which was found in a), the Wilcoxon signed rank test can be used. Although since the assumption of normality is not violated it makes more sense to use a parametric test.

### c) Computing the powers of the t-test and sign test from b) at $\mu = 14$ and $\mu = 13$ :

Power test for sign test and t-test with  $\mu_0=14$ :

```
mu0_1= 13
psign_13=numeric(B) ## will contain p-values of sign test
pttest_13=numeric(B) ## will contain p-values of t-test

for(i in 1:B) {
```

```

x=sample(data,replace=TRUE)
pttest_13[i]=t.test(x,mu=mu0_1,alt="less")[[3]] ## extract p-value
psign_13[i]=binom.test(sum(x>mu0),length(data),p=0.5, alt="less")[[3]]} ## extract p-value

#percentage of t test/sign tests that are lower than 0.05. T tests have a lot
#larger percentage of tests that reject the H0.
sum(psign_13<0.05)/B

```

```
## [1] 0.08
```

```
sum(pttest_13<0.05)/B
```

```
## [1] 0.244
```

```

#power test for sign test and t-test with mu0=14
mu0_2 = 14
psign_14=numeric(B) ## will contain p-values of sign test
pttest_14=numeric(B) ## will contain p-values of t-test

for(i in 1:B) {
  x=sample(data,replace=TRUE)
  pttest_14[i]=t.test(x,mu=mu0_2,alt="less")[[3]] ## extract p-value
  psign_14[i]=binom.test(sum(x>mu0),length(data),p=0.5, alt="less")[[3]]} ## extract p-value

#percentage of t test/sign tests that are lower than 0.05. T tests have a lot
#larger percentage of test than reject the H0.
sum(psign_14<0.05)/B

```

```
## [1] 0.092
```

```
sum(pttest_14<0.05)/B
```

```
## [1] 0.405
```

The t-test rejects the H0 more than the sign test for both  $\mu = 13$  and  $\mu = 14$ , which means that the wit the t-test the H1 is accepted more which was that mean  $< 15$ . What is also clearly seen for the t-test is the H1 is accepted more when  $\mu = 14$  compared to  $\mu = 13$ . This is not the case for the sign test were H1 is accepted roughly the same amount.

#### d) Recovering the whole confidence interval and its confidence level:

```

s = sd(data)
n = length(data)

# z-score is probability
pr = 0.53 # right end confidence interval
x <- sum(data>15.5)/n # sum of values > 15.5 / total
m <- pr-x # cause pr = x+m, m is error margin
pl <- x-m # left end confidence interval
cv <- (sqrt(n)*m)/s # cause m = cv*s/sqrt(n), is critical value
area <- 0.5398 # for cv 0.10 read from z-score table
alpha <- (1-(area/2)) # don't actually need alpha, cl is area/2
cl <- (1-alpha)*100 # confidence level
cbind(pl,pr,cl)

```

```

##          pl    pr cl
## [1,] 0.137 0.53 27

```

Due to asymptotic normality the Z-score is used. Thus, the confidence interval is [0.137, 0.53] and its confidence level is 26.99%.

e) Verifying the claim that the waiting time is different for men and women using the Fisher exact test:

```
tab = matrix(c(3,2,4,6), ncol=2, byrow=TRUE)
colnames(tab) = c('men', 'women')
rownames(tab) = c('more than 15.5', 'less than 15.5')
tab <- as.table(tab)
tab
```

```
##           men women
## more than 15.5    3    2
## less than 15.5    4    6
```

```
fisher.test(tab)
```

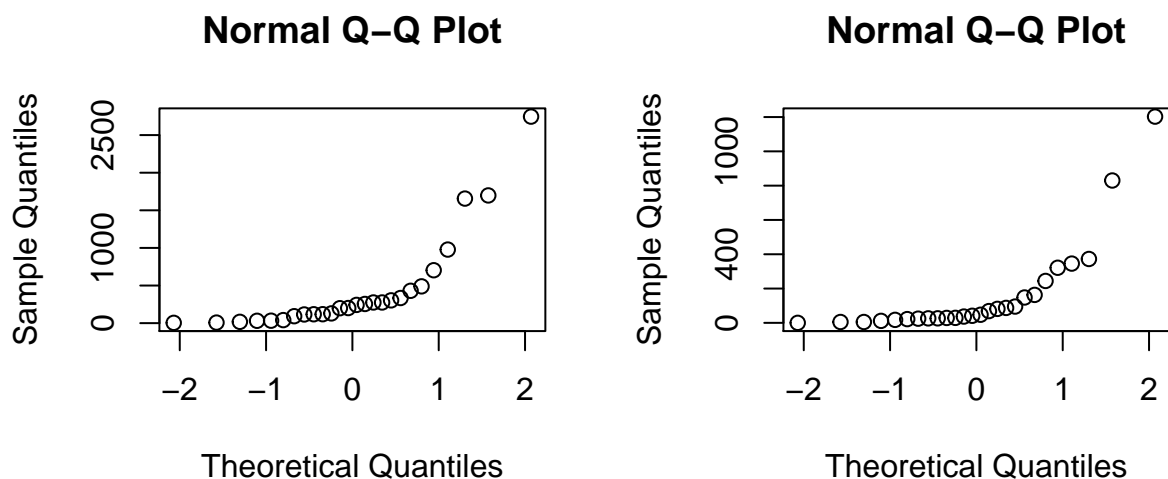
```
## p-value = 0.6
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.16 37.20
## sample estimates:
## odds ratio
##      2.13
```

Due to the small sample size Fisher's test is used. As the p-value 0.6084 is greater than the .05 significance level, we do not reject the null hypothesis that the waiting time is below or above 15.5 min is independent of the fact that if you are a man or woman. So the claim of the researcher that the waiting time is different for men and women is not proven based on the Fisher's test.

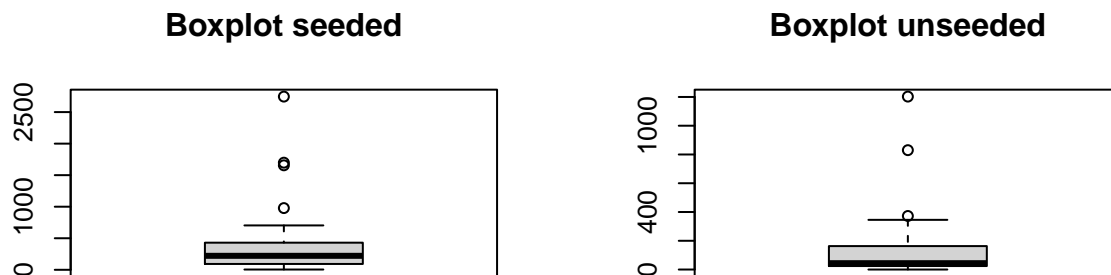
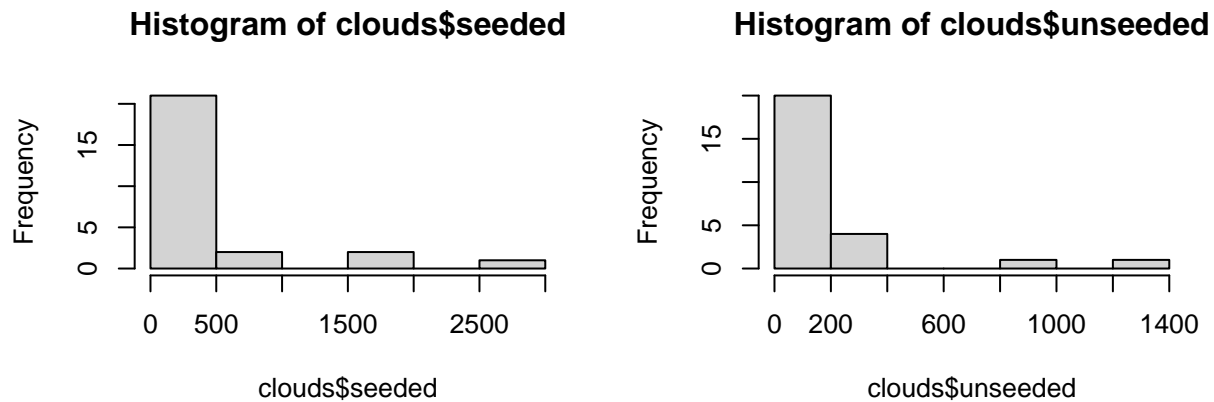
## Exercise 2

a) Testing whether silver nitrate has an effect: Plots:

```
qqnorm(clouds$seeded);qqnorm(clouds$unseeded)
```



```
hist(clouds$seeded);hist(clouds$unseeded)
boxplot(clouds$seeded,main='Boxplot seeded');boxplot(clouds$unseeded,main='Boxplot unseeded')
```



Normality test:

```
shapiro.test(clouds$seeded)
```

```
## W = 0.7, p-value = 1e-06
```

```
shapiro.test(clouds$unseeded)
```

```
## W = 0.6, p-value = 3e-07
```

Plots show a very not normal data distribution which is confirmed with Shapiro's test. Box plots show the data consists of very large outliers.

The seeded and unseeded clouds data set is not paired data. In the experiment design it is described as out of a sample of 52 clouds 26 are chosen, this already implies the clouds are separate 'individuals'. Not only that, clouds are unique and when it rains the cloud disappears (at least changes), which makes it impossible to use the same cloud twice. So we use an unpaired two samples t-test. Since we are investigating whether the rainfall increases with silver nitrate we test one sided.

```
t.test(clouds$seeded,clouds$unseeded,alternative = "greater")
```

```
## t = 2, df = 34, p-value = 0.03
```

```
## alternative hypothesis: true difference in means is greater than 0
```

Mann-Whitney test:

```
wilcox.test(clouds$seeded,clouds$unseeded,alternative = "greater")
```

```
## W = 473, p-value = 0.007
## alternative hypothesis: true location shift is greater than 0
```

Kolmogorov-Smirnov test:

```
ks.test(clouds$unseeded, clouds$seeded, alternative = "greater")
```

```
## D+ = 0.4, p-value = 0.01
## alternative hypothesis: the CDF of x lies above that of y
```

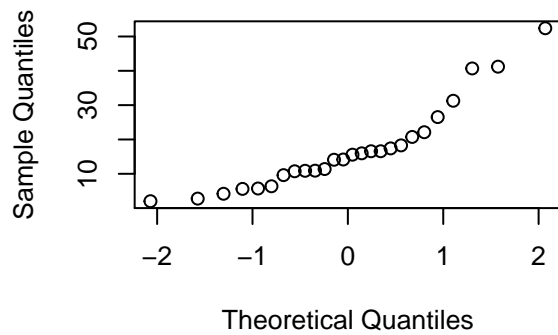
Are tests applicable? Mann-Whitney and Kolmogorov need the samples to be independent. Due to the unique nature of clouds and nothing in the experiment design suggests this assumption can't be withheld. T-test isn't applicable due to assumption of normality being violated.

Non parametric tests are more significant due to outliers for which these tests aren't as sensitive for. Mann-Whitney is a rank test (median) and Kolmogorov checks the differences of the verticality of the histograms, since outliers are infrequent, verticality of histogram for the outlier is small.

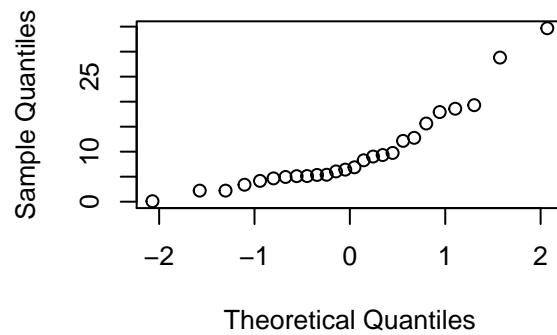
**b) Repeat (a) with square root transformation:**

```
sqrtseeded <- sqrt(clouds$seeded); sqrtunseeded <- sqrt(clouds$unseeded)
# Plots:
qqnorm(sqrtseeded); qqnorm(sqrtunseeded)
hist(sqrtseeded); hist(sqrtunseeded)
```

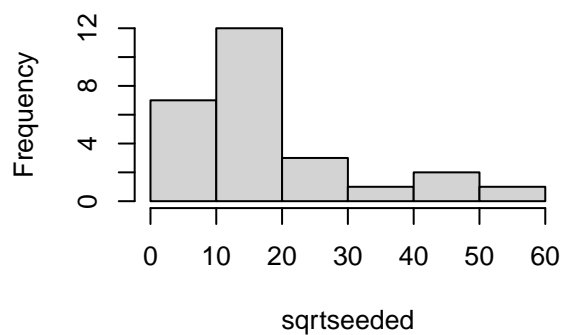
**Normal Q-Q Plot**



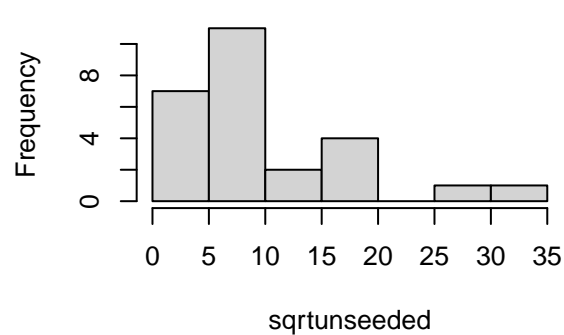
**Normal Q-Q Plot**



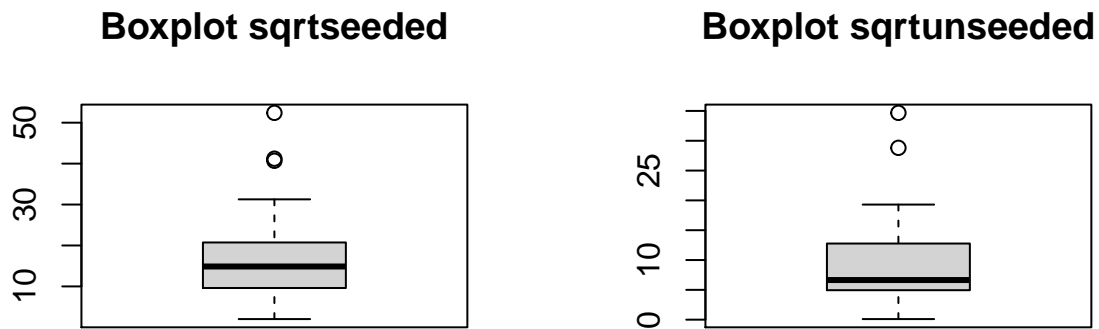
**Histogram of sqrtseeded**



**Histogram of sqrtunseeded**



```
par(mfrow=c(1,2))
boxplot(sqrtseeded, main='Boxplot sqrtseeded'); boxplot(sqrtunseeded, main='Boxplot sqrtunseeded')
```



Normality test:

```
shapiro.test(sqrtseeded)
```

```
## W = 0.9, p-value = 0.004
```

```
shapiro.test(sqrtunseeded)
```

```
## W = 0.8, p-value = 8e-04
```

An improvement can be seen in the plots and Shapiro test. However the data still does not have a normal distribution and still consists of outliers.

Two samples t-test:

```
t.test(sqrtseeded,sqrtunseeded,alternative = "greater")
```

```
## t = 2, df = 43, p-value = 0.01
```

```
## alternative hypothesis: true difference in means is greater than 0
```

Mann-Whitney test:

```
wilcox.test(sqrtseeded,sqrtunseeded,alternative = "greater")
```

```
## W = 473, p-value = 0.007
```

```
## alternative hypothesis: true location shift is greater than 0
```

Kolmogorov-Smirnov test:

```
ks.test(sqrtunseeded,sqrtseeded,alternative = "greater")
```

```
## D+ = 0.4, p-value = 0.01
```

```
## alternative hypothesis: the CDF of x lies above that of y
```

While the data set still contains outliers they are less large. T-test assumptions are still violated even though the resulting p-value is more significant. The transformation doesn't affect the significance of the ranked tests nor Kolmogorov Smirnov test since the rank and frequency of the data points is still the same.

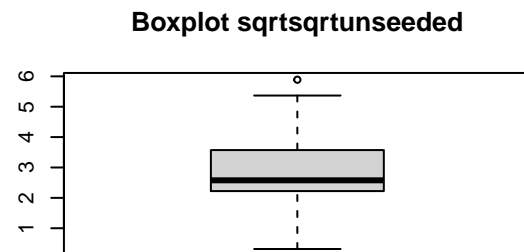
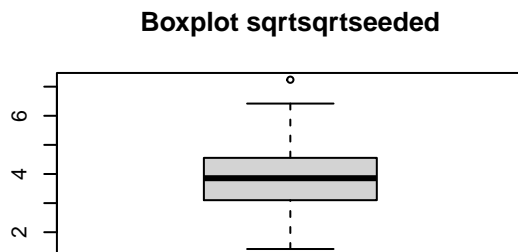
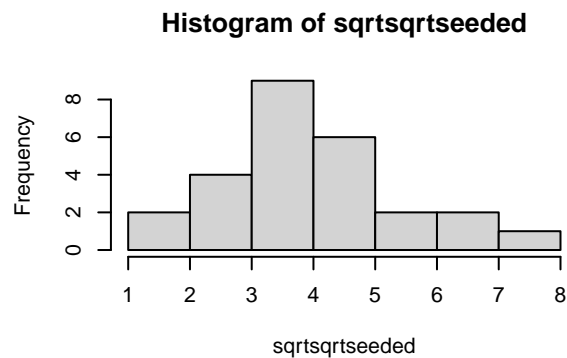
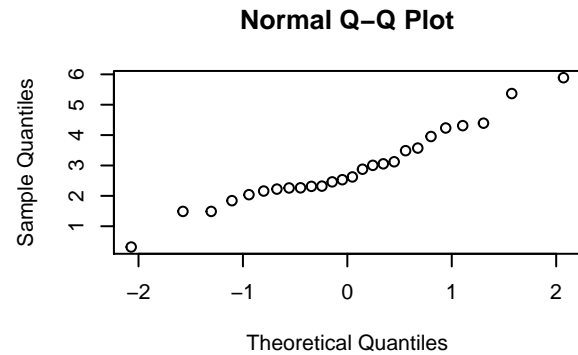
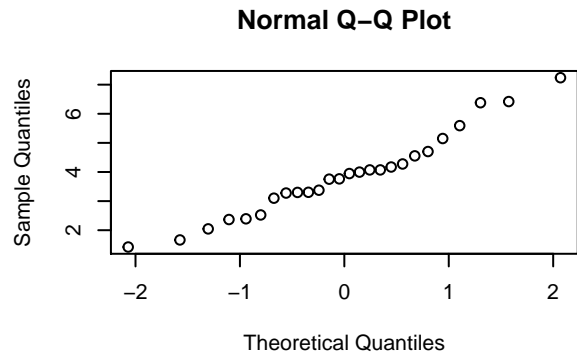
**Repeat (a) with square root of square root transformation:**



```

sqrtsqrtseeded <- sqrt(sqrtseeded); sqrtsqrtunseeded <- sqrt(sqrtunseeded)
# Plots:
qqnorm(sqrtsqrtseeded);qqnorm(sqrtsqrtunseeded)
hist(sqrtsqrtseeded);hist(sqrtsqrtunseeded)
boxplot(sqrtsqrtseeded,main='Boxplot sqrtsqrtseeded');boxplot(sqrtsqrtunseeded,main='Boxplot sqrtsqrtunseeded')

```



Normality test:

```
shapiro.test(sqrtsqrtseeded)
```

```
## W = 1, p-value = 0.5
```

```
shapiro.test(sqrtsqrtunseeded)
```

```
## W = 1, p-value = 0.3
```

The data set is now normally distributed. Since the assumption of normality is not violated any more for the t-test, we also check to see if the variances are equal. To do this we use the F-test:

```
var.test(sqrtsqrtseeded, sqrtsqrtunseeded, alternative = "two.sided")
```

```
## F = 1, num df = 25, denom df = 25, p-value = 0.4  
## alternative hypothesis: true ratio of variances is not equal to 1
```

The variance between groups is insignificant and can be considered equal.

Two samples t-test:

```
t.test(sqrtsqrtseeded, sqrtsqrtunseeded, var.equal = TRUE, alternative = "greater")
```

```
## t = 3, df = 50, p-value = 0.006  
## alternative hypothesis: true difference in means is greater than 0
```

Mann-Whitney test:

```
wilcox.test(sqrtsqrtseeded, sqrtsqrtunseeded, alternative = "greater")
```

```
## W = 473, p-value = 0.007  
## alternative hypothesis: true location shift is greater than 0
```

Kolmogorov-Smirnov test:

```
ks.test(sqrtsqrtunseeded, sqrtsqrtseeded, alternative = "greater")
```

```
## D+ = 0.4, p-value = 0.01  
## alternative hypothesis: the CDF of x lies above that of y
```

Are tests now applicable with transformations? Mann-Whitney and Kolmogorov were applicable the whole time and now since only 1 outlier is left and the data is normal distributed the t-test assumptions are not violated. P-values are similar, all around 0,01. The t-test is slightly more significant, which is to be expected due the power of the t-test. With out the remaining outlier it would most likely be even more significant compared to the non parametric tests.

### c) Finding an estimate of $\lambda$ and constructing a 95%-CI for $\lambda$ :

```
c1 <- clouds$seeded  
lambda_est <- 1/mean(c1)  
n <- length(clouds$seeded)  
sd <- 1/lambda_est  
# CI  
z <- qnorm(1-0.025)  
CI_L <- (-z+sqrt(n))/(mean(c1)*sqrt(n))  
CI_R <- (z+sqrt(n))/(mean(c1)*sqrt(n))  
cbind(CI_L, CI_R)
```

```
##          CI_L    CI_R  
## [1,] 0.00139 0.00313
```

Estimate of lambda = 0.00226 CI around lambda: [0.00139,0.00313]

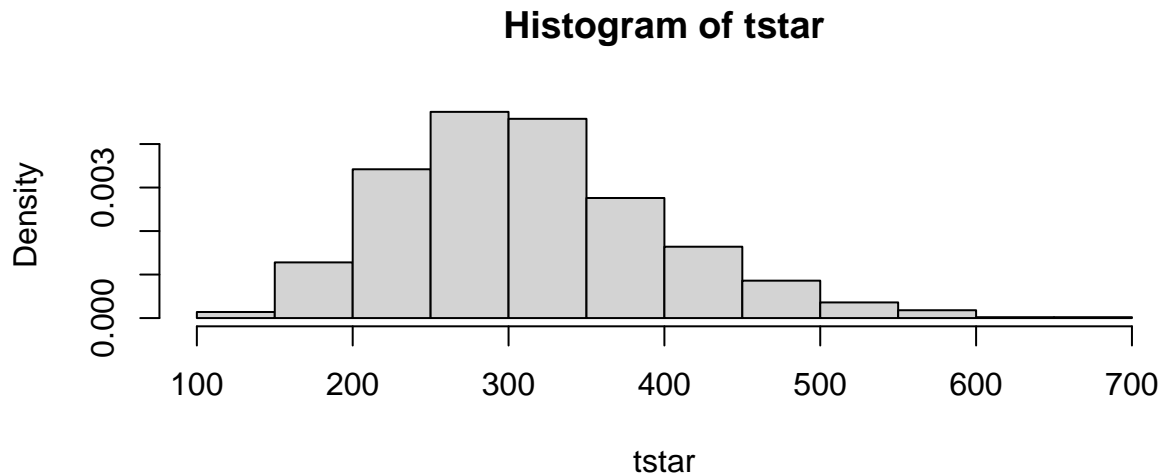
Bootstrap test:

```
t=median(c1);t
```

```
## [1] 222
```

```
B=1000  
tstar=numeric(B)  
n=length(c1)
```

```
for (i in 1:B){
  xstar=rexp(n,lambda_est)
  tstar[i]=median(xstar)}
hist(tstar,prob=T)
```



```
pl=sum(tstar<t)/B; pr=sum(tstar>t)/B; p=2*min(pl,pr)
cbind(pl,pr,p)
```

```
##      pl    pr    p
## [1,] 0.137 0.863 0.274
```

**Kolmogorov-Smirnov test:**

```
ks.test(clouds$seeded,"pexp", lambda_est)
```

```
## D = 0.2, p-value = 0.2
## alternative hypothesis: two-sided
```

For the bootstrap the test statistic median is used. The median of our sample is 222 and as can be seen in the histogram differs from the median of the simulated data. To test whether the data follows an  $\exp(\lambda_{est})$  distribution is significantly different: the  $H_0$  (that the data distribution is the same) cannot be rejected (both bootstrap test and Kolmogorov are insignificant). Thus we conclude that the distribution of the data follows a  $\exp(\lambda_{est})$  distribution. **d) Verifying whether the median precipitation for seeded clouds is less than 300**

In a) we showed that the cloud data is not symmetric nor normal distributed. So we use a sign test.

Binomial test:

```
x <- sum(clouds$seeded<300)
n <- length(clouds$seeded)
binom.test(x,n,p=0.5,alt='less')
```

```
## number of successes = 17, number of trials = 26, p-value = 1
## alternative hypothesis: true probability of success is less than 0.5
```

Not significant (P-value = 0.962) so  $H_0$  not rejected: which is medians are the same. Thus cannot conclude median precipitation is less than 300

Checking whether the fraction of the seeded clouds with precipitation less than 30 is at most 25%:

Binomial test:

```
x <- sum(clouds$seeded<30)
binom.test(x,n,p=0.25,alt='less')
```

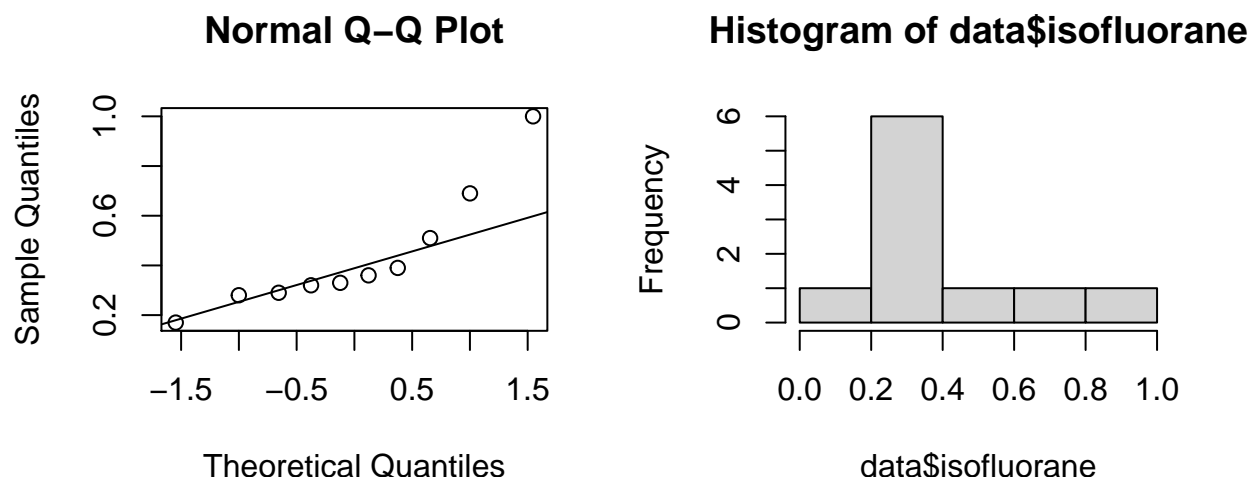
```
## number of successes = 3, number of trials = 26, p-value = 0.08
## alternative hypothesis: true probability of success is less than 0.25
```

Not significant (P-value = 0.0802)  $H_0$  not rejected: which is medians are the same. Thus cannot conclude the fraction of precipitation under 30 is at most 25%.

### Exercise 3

a) We cannot assume that the data was taken from normal populations because we have a very small sample size (only 10) and that is not enough to invoke the central limit theorem because we need a sample size of at least 30 to assume that. Furthermore, the distribution of the sample data is not normal as can be seen by the plot and normality test below.

```
qqnorm(data$isofluorane);qqline(data$isofluorane); hist(data$isofluorane)
```



```
shapiro.test(data$isofluorane)
```

```
## W = 0.8, p-value = 0.03
```

b) Checking whether the columns isofluorane and halothane are correlated:

Using Spearman's rank correlation test because it does not assume normality.

```
cor.test(data$isofluorane, data$halothane, method = "spearman")
```

```
## S = 129, p-value = 0.5
## alternative hypothesis: true rho is not equal to 0
```

p-value > 0.05 so the correlation is not significant.

Verifying whether the distributions of the isofluorane and halothane columns are different:

The permutation test is applicable here because it doesn't assume normality and the sample size is small.

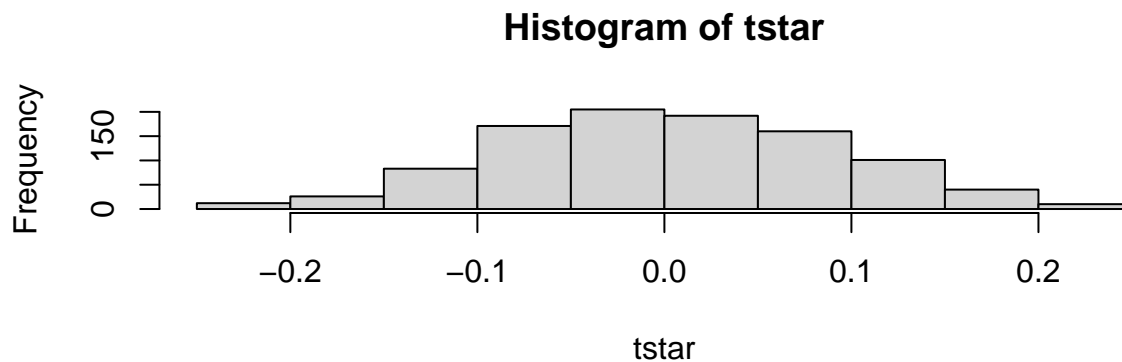
```

mystat=function(x,y) {mean(x-y)}
B=1000; tstar=numeric(B)
for (i in 1:B) {
  datastar=t(apply(cbind(data[,1],data[,2]),1,sample))
  tstar[i]=mystat(datastar[,1],datastar[,2]) }
myt=mystat(data[,1],data[,2])
myt

```

```
## [1] -0.035
```

```
hist(tstar)
```



```

pl=sum(tstar<myt)/B
pr=sum(tstar>myt)/B
p=2*min(pl,pr)
p

```

```
## [1] 0.694
```

p-value > 0.05 so no significant difference in the distributions of the isoflurane and halothane columns.

### c) One-way ANOVA test:

```

dataframe=data.frame(concentration=as.vector(as.matrix(data)),
                      variety=factor(rep(1:3,each=10)))
aov=lm(concentration~variety,data=dataframe)
anova(aov)

```

```

##           Df Sum Sq Mean Sq F value Pr(>F)
## variety    2   1.08   0.540    5.35 0.011 *

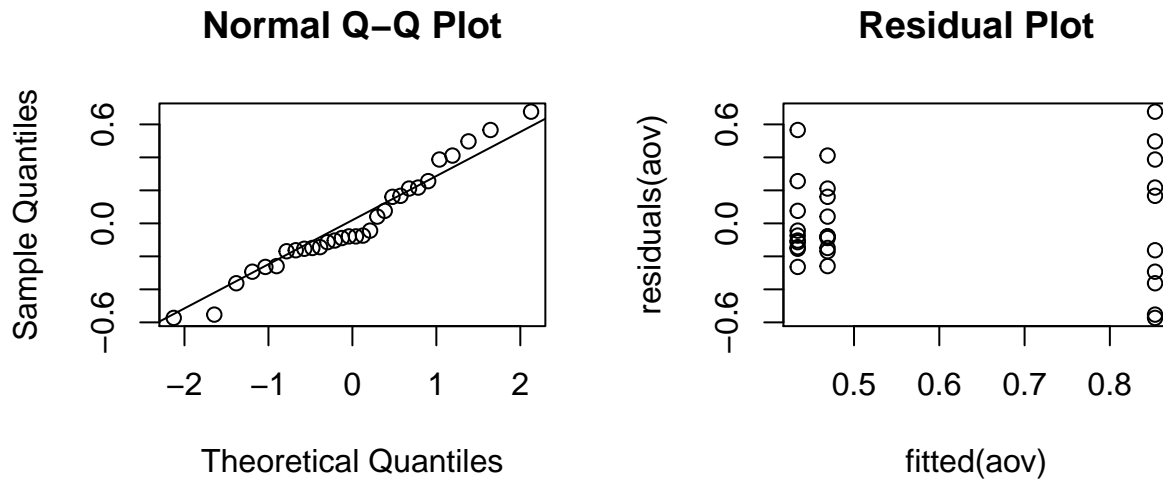
```

P-value < 0.05 so factor variety is significant. So yes, the type of drug has an effect on the concentration of plasma epinephrine.

```

par(mfrow=c(1,2))
qqnorm(residuals(aov)); qqline(residuals(aov))
plot(fitted(aov),residuals(aov), main = "Residual Plot")

```



```
leveneTest(concentration~variety,data=dataframe)
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The plot of the residuals shows the assumption of normality can withhold, however the spread in the residuals suggest that there is no equal variance in the residuals. To confirm this a Levenes test was performed ( $P < 0.05$ ). Thus the equal variances assumption of the one-way ANOVA test is not met.

**The estimated concentrations for each of the three anesthesia drugs:**

```
summary(aov)
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.434      0.101    4.32 0.00019 ***
## variety2      0.035      0.142    0.25 0.80727
## variety3      0.419      0.142    2.95 0.00650 **
```

The estimated concentrations are 0.43 for isoflurane, 0.04 for halothane, and 0.42 for cyclopropane. It's worth noting that the p-value for halothane is the only one greater than 0.05.

**d) Kruskal-Wallis test:**

```
attach(dataframe); kruskal.test(concentration,variety)
```

```
## Kruskal-Wallis chi-squared = 6, df = 2, p-value = 0.06
```

$p\text{-value} > 0.05$  so factor variety is not significant. So no, the two tests don't arrive at the same conclusion. While the one-way ANOVA test has more power than the Kruskal-Wallis test, it is less accurate in this situation because the assumption of equal variances is not met. The nonparametric Kruskal-Wallis test has less strict assumptions which are met and is thus more accurate.

## Exercise 4

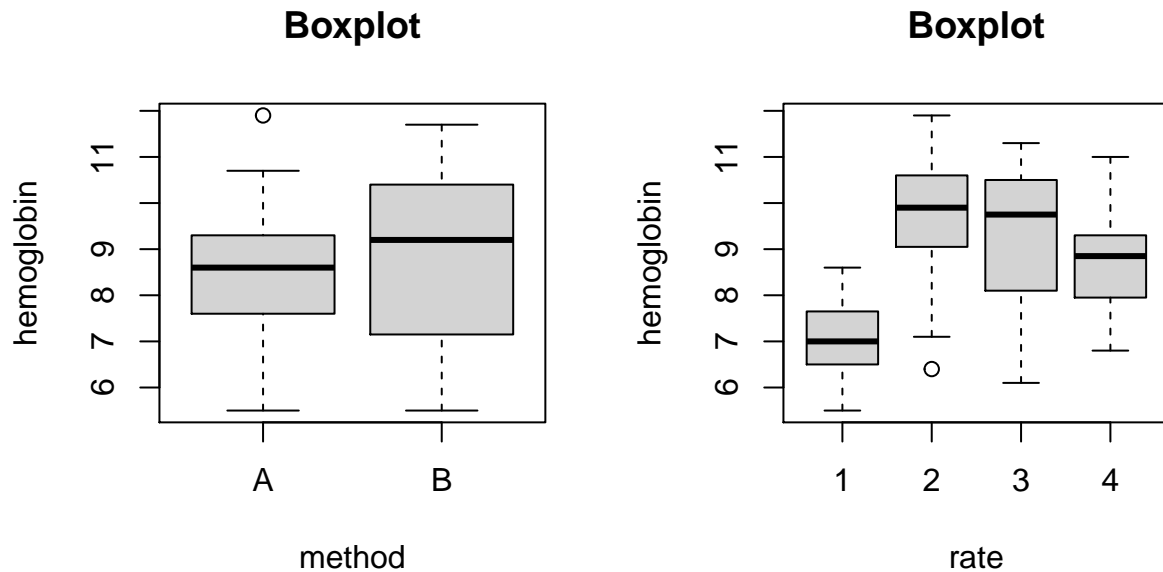
**a) Randomization process to distribute 80 fishes:**

```
I=4; J=2; N=10
rbind(rep(1:I,each=N*J),rep(1:J,N*I),sample(1:(N*I*J)))
```

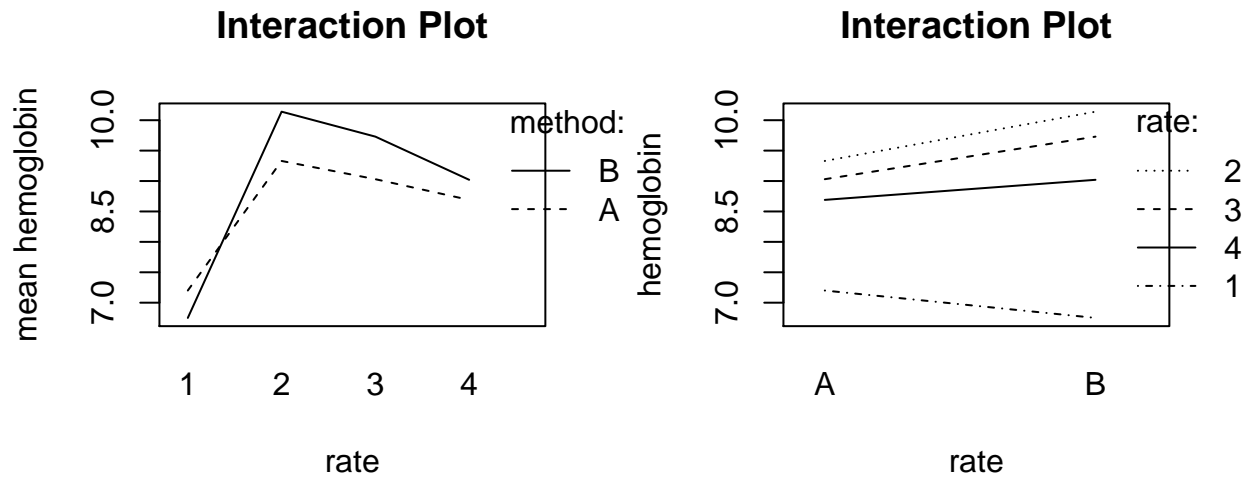
4 rate factors 2 method factors and each specific combination is repeated for 10 fishes. With this code the 80 fishes are distributed randomly over all combinations of levels of factors rate and method.

**b) Two-way ANOVA test:**

```
#necessary other wise rate is numerical and should be categorical
data$rate=as.factor(data$rate)
boxplot(hemoglobin ~ method, data = data,main = "Boxplot") ;
boxplot(hemoglobin ~ rate, data = data,main = "Boxplot")
```



```
interaction.plot(x.factor = data$rate, trace.factor = data$method, response = data$hemoglobin, main="In
                 xlab="rate",
                 ylab="mean hemoglobin",trace.label = "method:");
interaction.plot(x.factor = data$method, trace.factor = data$rate, response = data$hemoglobin, main="In
                 xlab="rate",
                 ylab="hemoglobin",trace.label = "rate:")
```



```
hemoglobinaov=lm(hemoglobin ~ method * rate, data = data)
anova(hemoglobinaov)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## method      1    2.4    2.42    1.56    0.22
## rate        3   90.6   30.19   19.47 2.4e-09 ***
## method:rate  3    4.9    1.62    1.05    0.38
```

The p-value for the interaction between method and the rate is larger than 0.05. So, there is no evidence for interaction between the two factors. Therefore we will create a new model without the interaction called the additive model.

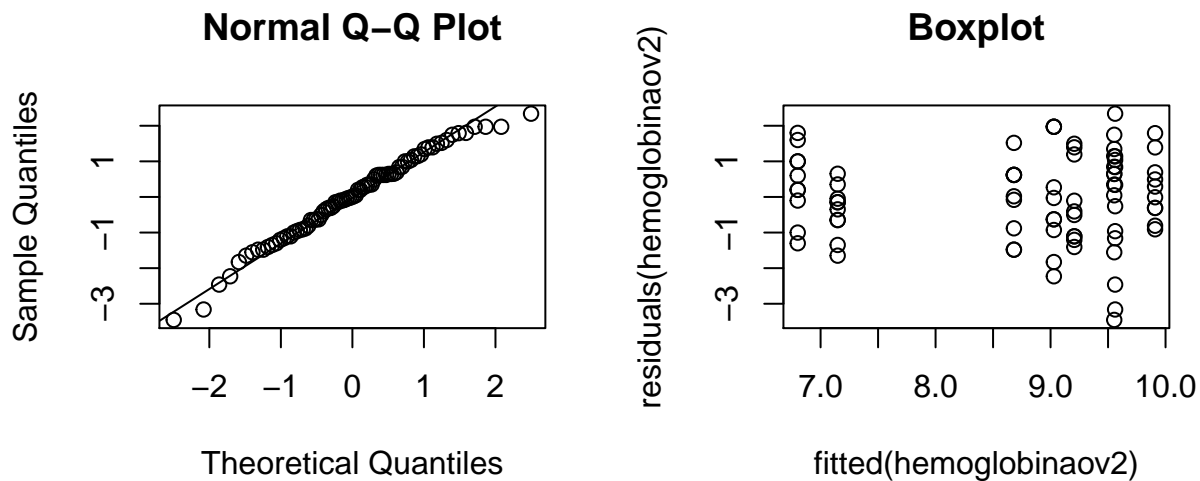
```
hemoglobinaov2 =lm(hemoglobin ~ method + rate, data = data)
anova(hemoglobinaov2)
```

```
## method      1    2.4    2.42    1.55    0.22
## rate        3   90.6   30.19   19.43 2e-09 ***
```

With the two way anova additive model we can see that the method factor had a p-value > 0.05 and therefore has no main effect for the outcome of this research. The rate factor has a p-value < 0.05 which means that the effect of different amounts of sulfamerazine that is given to the brown trout gives significantly different amounts of hemoglobin in their blood.

```
par(mfrow=c(1,2))
qqnorm(residuals(hemoglobinaov2))
qqline(residuals(hemoglobinaov2))
plot(fitted(hemoglobinaov2),residuals(hemoglobinaov2), main = "Boxplot")
```





```
hemoglobinaov2
```

```
##
## Call:
## lm(formula = hemoglobin ~ method + rate, data = data)
##
## Coefficients:
## (Intercept)      methodB        rate2        rate3        rate4
##      6.801         0.348         2.760         2.405         1.880
```

```
shapiro.test(residuals(hemoglobinaov2))
```

```
## W = 1, p-value = 0.3
```

QQ plot looks like the residuals are normally distributed. For the Shapiro's test the p-value  $> 0.05$ , therefore passed the test and can state that there exists no significant departure from normality. The spread in the residuals seems to be in the similar range for all the fitted values. So, we can assume equal variance in the residuals. Because the requirements of normality and the assumption of equal variances are met, we can use the two-way ANOVA test.

### c) Factor with greater influence:

```
group_by(data, method, rate) %>%
  summarise(
    count = n(),
    mean = mean(hemoglobin, na.rm = TRUE),
    sd = sd(hemoglobin, na.rm = TRUE)
  )
```

```
##   <chr> <fct> <int> <dbl> <dbl>
## 1 A    1     10  7.2  1.02
## 2 A    2     10  9.33 1.72
## 3 A    3     10  9.03 1.14
## 4 A    4     10  8.69 1.00
## 5 B    1     10  6.75 0.707
## 6 B    2     10 10.1  0.887
## 7 B    3     10  9.73 1.56
## 8 B    4     10  9.02 1.55
```

The rate factor has a significant main effect on the the amount of hemoglobin, while the method factor does not have a significant main effect on the amount of hemoglobin. So the rate factor has the greatest influence of the two. This is a good question, because one factor is significant and the other is not if both were insignificant than it would not be a good question.

**Combination of rate and method that yields the highest hemoglobin:** Rate 2 and method B leads to a mean of 10.1, which is the highest hemoglobin yield when both factors are considered. When rate 3 and method A was used then the mean was estimated to be 9.03.

**Rate that leads to the highest mean hemoglobin:**

```
group_by(data, rate) %>%
  summarise(
    count = n(),
    mean = mean(hemoglobin, na.rm = TRUE),
    sd = sd(hemoglobin, na.rm = TRUE)
  )
```

```
##   <fct> <int> <dbl> <dbl>
## 1 1      20  6.98 0.884
## 2 2      20  9.74 1.39
## 3 3      20  9.38 1.38
## 4 4      20  8.86 1.28
```

When only the factor rate is considered then rate 2 gives the highest mean rate with a mean of 9.74 hemoglobin yield.

**d) One-way ANOVA test:**

```
hemoglobinaov3 =lm(hemoglobin ~ rate, data = data)
anova(hemoglobinaov3)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## rate        3   90.6    30.19    19.3 2.1e-09 ***
## Residuals  76  118.9     1.56
```

With the one-way anova we see that the p-value  $< 0.05$  and therefore the rate factor gives significant different means for the hemoglobin. We can use the one-way anova test on this data set because the factor method is insignificant and the interaction between the rate factor and the method factor is also insignificant.

## Exercise 5

**a) Repeated Anova without interactions:**

```
cream$batch <- factor(cream$batch)
cream$position <- factor(cream$position)
cream$starter <- factor(cream$starter)
aovcream <- lm(acidity~batch+position+starter,data=cream); anova(aovcream)
```

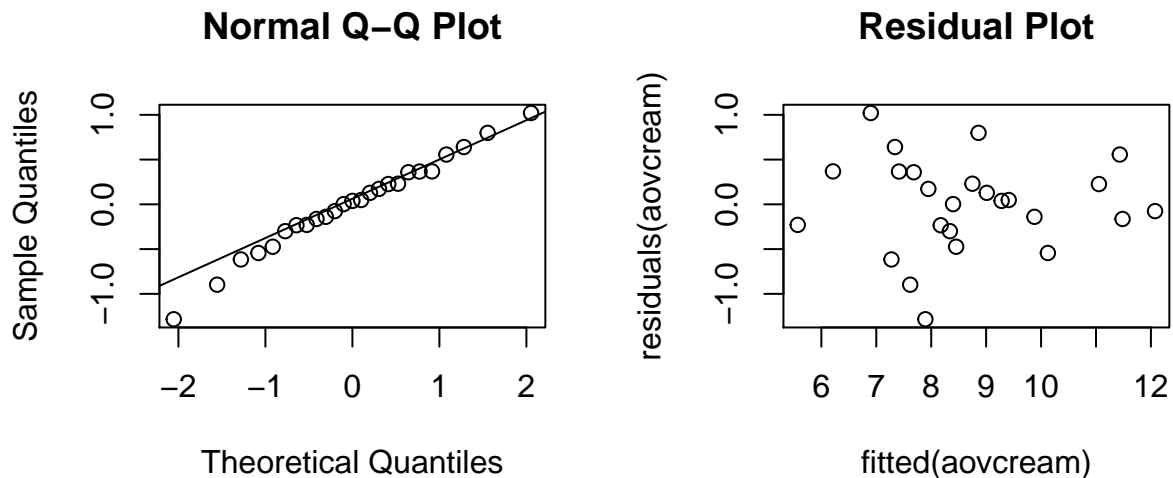
```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## batch        4   18.8     4.69     8.60 0.0016 **
## position      4    2.3     0.59     1.08 0.4112
## starter       4   44.1    11.03    20.21 2.9e-05 ***
```

```
summary(aovcream)
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.662     0.533   16.26 1.5e-09 ***
## batch2       -1.348     0.467   -2.88  0.014 *
## batch3        0.276     0.467    0.59  0.566
```

```
## batch4      1.368      0.467      2.93      0.013 *
## batch5      0.200      0.467      0.43      0.676
## position2   -0.618      0.467     -1.32      0.211
## position3   -0.038      0.467     -0.08      0.937
## position4   -0.764      0.467     -1.63      0.128
## position5   -0.264      0.467     -0.56      0.583
## starter2    -0.150      0.467     -0.32      0.754
## starter3    -0.980      0.467     -2.10      0.058 .
## starter4     2.810      0.467      6.01     6.1e-05 ***
## starter5    -0.484      0.467     -1.04      0.321
```

```
qqnorm(residuals(aovcream));qqline(residuals(aovcream));plot(fitted(aovcream),residuals(aovcream),main=
```



```
shapiro.test(residuals(aovcream));
```

```
## W = 1, p-value = 1
```

Three factors are relevant to look at so a three way Anova is used. Since the same starter combination is being tested multiple times in different locations a repeated measures test is used. So the test used is a three way repeated measure Anova. Position no significant effect ( $P = 0.411$ ). However batch does have significant effect ( $P = 0.00163$ ) (not expected in experiment design) and starter also has a significant effect ( $P = 2.904e-05$ ). From summary: p value starter2 not significant ( $p = 0.754$ ), this is compared to starter1. Thus starter 1 and 2 do not significantly differ. The plots and normality tests show the assumption of normality is not violated and looking at the residuals plot the assumption of homogeneity of variances is also not violated. So the chosen Anova test is applicable.

b)

```
aovcream=lm(acidity~batch+starter,data=cream); anova(aovcream)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## batch      4   18.8    4.69    8.44 0.00073 ***
## starter    4   44.1   11.03   19.84 4.8e-06 ***
## Residuals 16    8.9    0.56
```

```
summary(aovcream)
```

```
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)      8.325      0.447    18.60  2.9e-12 ***
## batch2          -1.348      0.472    -2.86   0.011 *
## batch3           0.276      0.472     0.59   0.567
## batch4           1.368      0.472     2.90   0.010 *
## batch5           0.200      0.472     0.42   0.677
## starter2        -0.150      0.472    -0.32   0.755
## starter3        -0.980      0.472    -2.08   0.054 .
## starter4         2.810      0.472     5.96  2.0e-05 ***
## starter5        -0.484      0.472    -1.03   0.320
```

Batch is not an insignificant block variable ( $P = 0.000735$ ), even though in the experiment design the batch was meant to be identical. However the position has no influence so can be considered an insignificant block element ( $P = 4.816e-06$ ) and be removed. The now two-way repeated measures Anova: Out of all the starters, only starter 4 is significant ( $P = 6.10e-05$ )

### c) Friedman

```
friedman.test(acidity~batch | starter,data=cream)
```

```
## Friedman chi-squared = 13, df = 4, p-value = 0.01
```

Friedman test is a non parametric version of the repeated measures Anova. Since the the assumptions for the repeated measure Anova are met, it makes more sense to use that test, no assumptions for Friedman's test (data is independent and data can be ranked) are violated. so it is possible to use it with acidity as response values, starter as the group and batch as block variables. However to see which starter of the starters causes the significant effect, a post hoc test would need to be applied.

### d)

```
creamlmer=lmer(acidity~starter+(1|batch),REML=FALSE,data=cream)
summary(creamlmer)
```

```
## Random effects:
## Groups   Name      Variance Std.Dev.
## batch    (Intercept) 0.662    0.814
## Residual                0.445    0.667
## Number of obs: 25, groups:  batch, 5
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)    8.424      0.471    17.90
## starter2       -0.150      0.422     -0.36
## starter3       -0.980      0.422     -2.32
## starter4        2.810      0.422     6.66
## starter5       -0.484      0.422     -1.15
```

```
creamlmer1=lmer(acidity~(1|batch),data=cream,REML=FALSE)
anova(creamlmer,creamlmer1)
```

```
##              npar    AIC    BIC logLik deviance Chisq Df Pr(>Chisq)
## creamlmer1      3 103.1 106.7  -48.5    97.1
## creamlmer       7  75.4  83.9  -30.7    61.4  35.7  4    3.3e-07 ***
```

Using the same model as in b, however now the block model batch is modeled with a random effect. Since lmer doesn't give p-values, it is compared to the same model without starters. The variance between these models gives us the significance of the starter factor. Comparing the model with random effects with the fixed effect model in b), Starter 4 has the most significant effect on acidity in both models and slightly larger t-value in lmer ( $T = 6.660$ ) compared to the fixed effect model in b) ( $T = 5.957$ ). Furthermore the estimates are slightly larger in the random effects model. So similar results are found, except for the difference that

no estimates were given for the batch variable, since this was modeled as a random effect. Since in the experimental design the batch was not meant to have an effect, we argue that modelling the batch as a random block variable is an improvement.