Homework 2

Palash Sharma

February 17, 2018

Ques No 1(a)

```
setwd("D:/KUMC COURSE/SPRING 2018/Design of Experiment/HOMEWORK/hw 2")

choles<-read.csv("Cholesterol.csv")
library(pwr)
```

```
## Warning: package 'pwr' was built under R version 3.3.3
```

```
delta=10 # difference
sigma=var(choles$Response) # variance
f=sqrt(10^2/(2*sigma))
pwr.anova.test(k=2,n=20,f,sig.level = 0.05)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##               k = 2
##               n = 20
##               f = 1.359946
##       sig.level = 0.05
##           power = 1
##
## NOTE: n is number in each group
```

Our effect size for power calculation is

$$f = \sqrt{\frac{\delta^2}{2\sigma^2}}$$

.

And using the default R package we can be able to detect 80 percent power with 0.05 type I error rate.

Ques No 1(b):

```
setwd("D:/KUMC COURSE/SPRING 2018/Design of Experiment/HOMEWORK/hw 2")

data<-read.csv("Cholesterol.csv")

fit<-lm(data$Response~factor(data$Treatment))
#summary(fit)
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: data$Response
##                          Df  Sum Sq Mean Sq F value   Pr(>F)
## factor(data$Treatment)  4  346.23  86.558  3.5288 0.009915 **
```

```
## Residuals              95 2330.24  24.529
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Our null hypothesis is that there is no mean difference among all treatment groups. And the alternative Hypothesis is there is a differnece among treatment groups. Lets say,

$$H_0 : \mu_1 = \mu_2 = ... = \mu_5 = 0$$

$$H_a : \mu_i \neq 0$$

Since the P values is less than 0.05, We can reject the null hypothesis in favour of alternative hypothesis. So there is a mean difference in cholesterol between any of the five treatment groups following the 16 week treatment peiod.

Ques No 1(c):

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.3.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
treat.1<-filter(data,Treatment==1)
```

```
## Warning: package 'bindrcpp' was built under R version 3.3.3
```

```
treat.2<-filter(data,Treatment==2)
treat.3<-filter(data,Treatment==3)
treat.4<-filter(data,Treatment==4)
treat.5<-filter(data,Treatment==5)

t_value<-qt((.05/6),95,lower.tail=FALSE)
mse<-anova(fit)[["Mean Sq"]][2]
sum_c<-((1+(1/4)^2+(1/4)^2+(1/4)^2+(1/4)^2)/20)
mean_y<-mean(treat.1$Response)-.25*mean(treat.2$Response)-.25*mean(treat.3$Response)-.25*mean(treat.4$R

# First preplanned contrast -0.5231651---5.511921
contrast.1.lower<-mean_y-t_value*sqrt(mse*sum_c)
contrast.1.upper<-mean_y+t_value*sqrt(mse*sum_c)


# Second contrast
sum_c2<-((.5+.5)/20)
mean_y2<-.5*mean(treat.3$Response)+.5*mean(treat.5$Response)-.5*mean(treat.2$Response)-.5*mean(treat.4$

# second preplanned contrast -5.601793 .....   -0.203848
contrast.1.lower2<-mean_y2-t_value*sqrt(mse*sum_c2)
```

```
contrast.1.upper2<-mean_y2+t_value*sqrt(mse*sum_c2)


# Third contrast
sum_c3<-((.5+.5)/20)
mean_y3<-.5*mean(treat.4$Response)+.5*mean(treat.5$Response)-.5*mean(treat.2$Response)-.5*mean(treat.3$

# third preplanned contrast -4.476069 .....  0.9218768
contrast.1.lower3<-mean_y3-t_value*sqrt(mse*sum_c3)
contrast.1.upper3<-mean_y3+t_value*sqrt(mse*sum_c3)
```

```
# 95% confidence interval for first preplanned contrast
c(contrast.1.lower,contrast.1.upper)
```

```
## [1] -0.5231651  5.5119214
```

```
# 95% confidence interval for second preplanned contrast
c(contrast.1.lower2,contrast.1.upper2)
```

```
## [1] -5.601793 -0.203848
```

```
# 95% confidence interval for Third preplanned contrast
c(contrast.1.lower3,contrast.1.upper3)
```

```
## [1] -4.4760686  0.9218768
```

We are using Bonferroni method for 95% simultaneous confidence interval for the three prelanned contrasts. Since our m=3 is small, This method gives shorter confidence interval than the other method. Also it can be used for any design but can not be used for data snopping.

For the first confidence interval, our null hypothesis is there is no mean differnce between experimental and standard drug.

$$H_o : \lambda_1 - 1/4(\lambda_2 + \lambda_3 + \lambda_4 + \lambda_5) = 0$$

$$H_a : \lambda_1 - 1/4(\lambda_2 + \lambda_3 + \lambda_4 + \lambda_5) \neq 0$$

Using Bonferroni method, our confidence interval is $(-0.5231651, 5.511921)$. Since this confidence interval includes 0 we would fail to reject null hypothesis that there is no differnece between standard vs. experimental drug. Using significant level $\alpha = 0.05$, we are 95% confident that there is no mean differnce between these two groups.

For the second confidence interval, our null hypothesis is there is a mean differnce between Drug with niacin vs. non niacin drugs among only the experimental treatments.

$$H_o : 1/2(\lambda_3 + \lambda_5) - 1/2(\lambda_2 + \lambda_4) = 0$$

$$H_a : 1/2(\lambda_3 + \lambda_5) - 1/2(\lambda_2 + \lambda_4) \neq 0$$

Using Bonferroni method, our confidence interval is $(-5.601793 - 0.203848)$. Since this confidence interval does not includes 0 we would reject the null hypothesis that there is no differnece between between Drug with niacin vs. non niacin drugs among only the experimental treatments. Using significant level $\alpha = 0.05$, we are 95 percent confident that there is a mean differnce between Drug with niacin vs. non niacin drugs among only the experimental treatments.

For the third confidence interval, our null hypothesis is there is a mean differnce between drug with simvastatin vs. non simvastatin contaings drugs among only the experimental treatments.

$$H_0 : 1/2(\lambda_4 + \lambda_5) - 1/2(\lambda_2 + \lambda_3) = 0$$

$$H_a : 1/2(\lambda_4 + \lambda_5) - 1/2(\lambda_2 + \lambda_3) \neq 0$$

Using Bonferroni method, our confidence interval is $(-4.47606860.9218768)$. Since this confidence interval includes 0 we would fail to reject null hypothesis that there is no differnece between drug with simvastatin vs. non simvastatin contaings drugs among only the experimental treatments. Using significant level $\alpha = 0.05$, we are 95% confident that there is no mean differnce between drug with simvastatin vs. non simvastatin contaings drugs among only the experimental treatments.

Ques No 1(d):

If we are interested in 95 percent simultaneous confidence intervals for pairwise contrats, we can use tukey's method. It is best for all pairwise comparison. It can also used for completely randomized design, randomized block design and balanced incomplete block design.

Critical Coefficient

The critical coefficient value for this method is

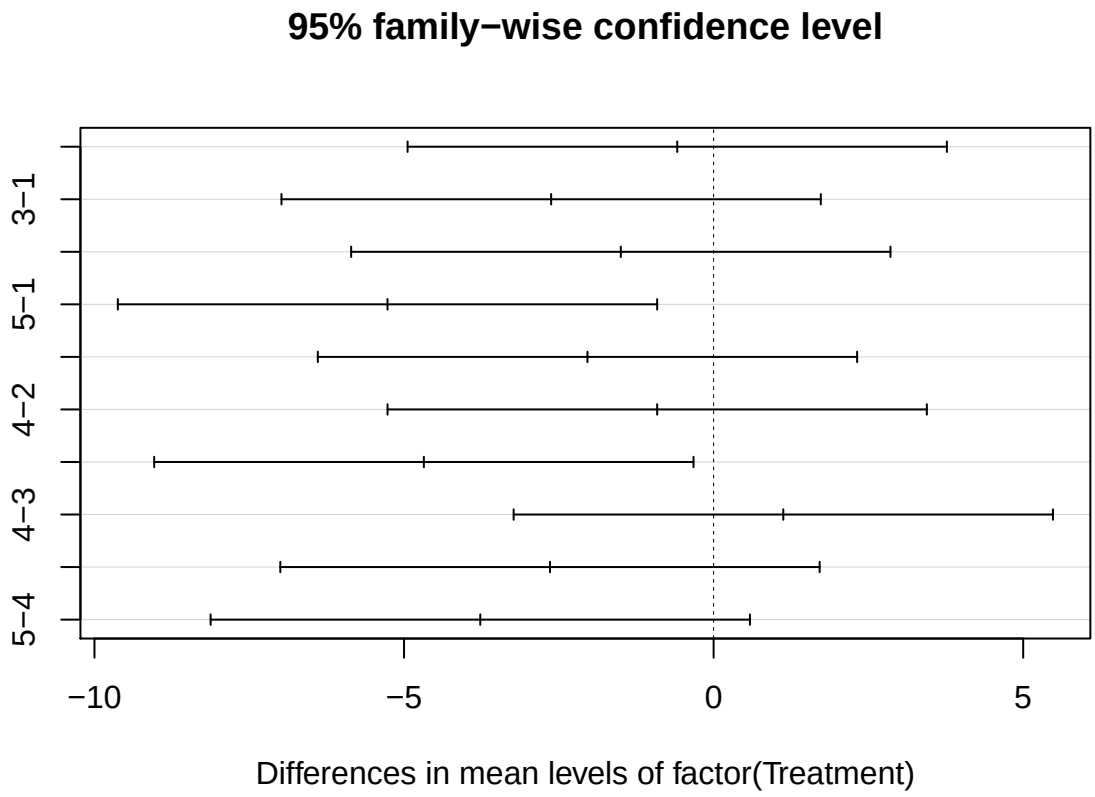
```
qtukey(.95,5,95)/sqrt(2)
```

```
## [1] 2.780865
```

```
# Tukey's test for all possible pairs
fitted<-aov(Response~factor(Treatment),data=data)
TukeyHSD(fitted,conf.level=0.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Response ~ factor(Treatment), data = data)
##
## $`factor(Treatment)`
##           diff       lwr        upr     p adj
## 2-1 -0.5873037 -4.942604  3.7679969 0.9957262
## 3-1 -2.6243567 -6.979657  1.7309439 0.4538105
## 4-1 -1.4986318 -5.853932  2.8566687 0.8735481
## 5-1 -5.2672204 -9.622521 -0.9119198 0.0095896
## 3-2 -2.0370530 -6.392354  2.3182476 0.6914471
## 4-2 -0.9113281 -5.266629  3.4439724 0.9774512
## 5-2 -4.6799167 -9.035217 -0.3246161 0.0287327
## 4-3  1.1257249 -3.229576  5.4810254 0.9517003
## 5-3 -2.6428637 -6.998164  1.7124369 0.4465489
## 5-4 -3.7685885 -8.123889  0.5867120 0.1225584
```

```
plot(TukeyHSD(fitted,conf.level=0.95))
```



Ques No 1(e):

Overall experiment wise error rate for this study across the two sets of contrast (part c and d) if we want to maintain an overall experimental error rate 0.05 is as follows:

$$\frac{\alpha}{4m} + \frac{\alpha}{2} = \frac{0.05}{12} + \frac{0.05}{2} = 0.02916$$

Ques No 2:

We know

$$Y_i = \beta_0 + \beta_1 x_i$$

$$\begin{aligned} L &= -3(\beta_0 + 24\beta_1) - 1(\beta_0 + 28\beta_1) + 1(\beta_0 + 32\beta_1) + 3(\beta_0 + 36\beta_1 \\ &= -72\beta_1 - 28\beta_1 + 32\beta_1 + 108\beta_1 \\ &= 40\beta_1 \end{aligned}$$

Since L is only a function of β_1 and a test of L=0 is equivalent of a test of $\beta_1 = 0$.

Ques No 3:

```
library(dplyr)

sample_size=function(alpha,delta,power,sigma){
   nu1=3
   nu2=1000
   f_value=qf(1-alpha,nu1,nu2) # critical value
   lambda<-seq(0,20,0.01) # non central parameter
   ff=qf(1-alpha,nu1,nu2,ncp=lambda)  # noncentral f distribution
   power.seq=1-pf(f_value,nu1,nu2,ff)  # power for f distribution
   actual.data=data.frame(lambda,ff,power.seq) # create a data frame
   # filtering data with power .80
   new.lambda<-filter(actual.data,power.seq>power)
   actual.lambda=new.lambda[1,1] # extract the first value of lambda for which power is greater than .8
   phi=actual.lambda/(nu1+1)  # estimate phi
   sample.size=(2*(nu1+1)*sigma*phi)/delta # estimate sample size
   whole.sample=ceiling(sample.size)  # rounding up the sample
   return(whole.sample)

 }

 sample_size(0.05,4,0.80,.5)
```

```
## [1] 4
```

Compare this function with build in pwr.anova.test function in the pwr package

```
library(pwr)
power.anova.test(groups = 4,sig.level = 0.05,between.var=.5,within.var=.5,power = 0.80)
```

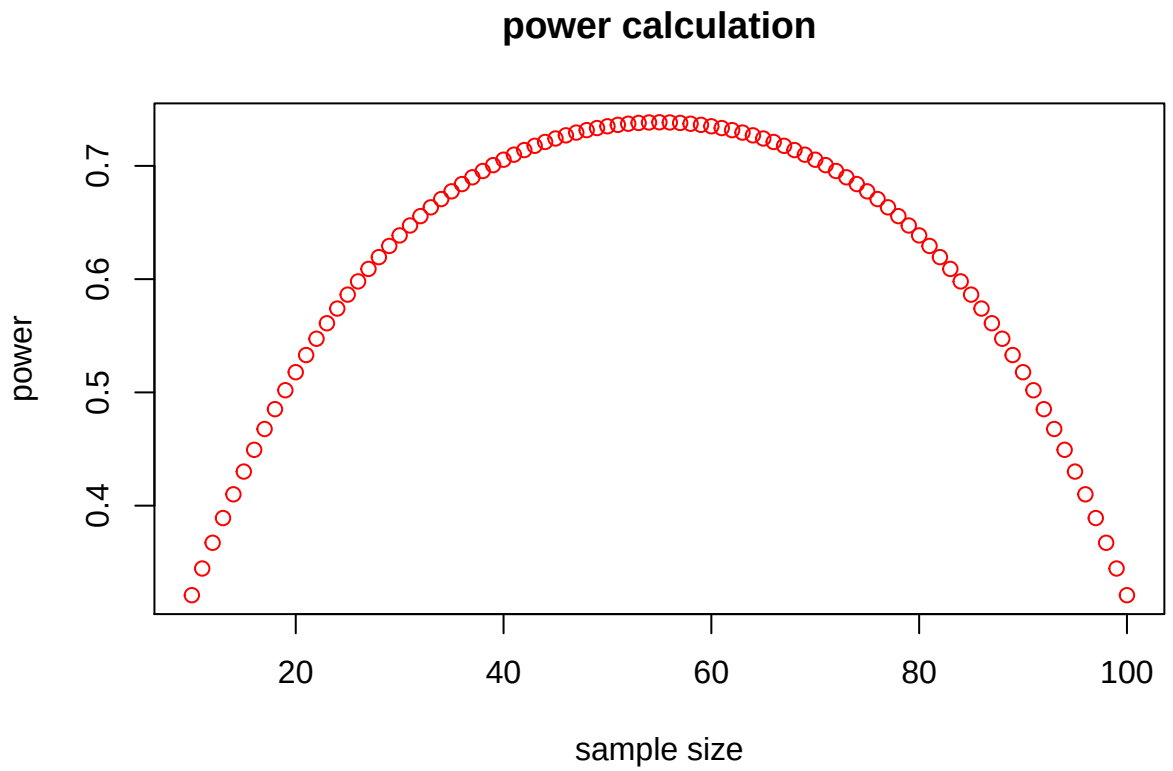
```
##
##      Balanced one-way analysis of variance power calculation
##
##          groups = 4
##               n = 4.734463
##     between.var = 0.5
##      within.var = 0.5
##       sig.level = 0.05
##           power = 0.8
##
## NOTE: n is number in each group
```

So my function is consistant with the build in R package.

Ques No 4:

```
n1=seq(10,100) # generate a sequence
n2=seq(100,10)  # generate a reverse sequence
pow<-rep(NA,length(n1)) # create a null vector
d<-rep(0.5,length(n1)) # fixed our effect size
sig.level=rep(0.05,length(n1))
power.data<-data.frame(n1,n2,d,sig.level) # create a data frame
pow<-transform(power.data,pow.value=pwr.t2n.test(power.data$n1,power.data$n2,power.data$d,power.data$sig
```

```r
plot(n1,pow$pow.value,xlab="sample size",ylab="power",main="power calculation",col="red")
```



We can see statistical power is maximized in a balanced complete randomized design and power decreases as a function of increasing unbalancedness.

Ques No 5(chapter 4 question NO 7)

Part(a):

```r
soap.origin<-read.table("soap_origin.txt",header = TRUE)
fit_soap<-aov(WtLoss~factor(Soap),data=soap.origin)
summary(fit_soap)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## factor(Soap)  2 16.122   8.061   104.5 5.91e-07 ***
## Residuals     9  0.695   0.077
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
soap1<-soap.origin%>%
  filter(Soap=="1")
soap2<-soap.origin%>%
  filter(Soap=="2")
soap3<-soap.origin%>%
  filter(Soap=="3")
```

```r
t_val<-qt((.05/2),9,lower.tail=FALSE)
mse_soap<-anova(fit_soap)[["Mean Sq"]][2]
sum_c<-(1+(1/2)^2+(1/2)^2)/4 # r is individual group
mean_y<-mean(soap1[,5])-.50*mean(soap2[,5])-.50*mean(soap3[,5])

# First preplanned contrast -0.5231651---5.511921
contrast.1.lower<-mean_y-t_val*sqrt(mse_soap*sum_c)
contrast.1.upper<-mean_y+t_val*sqrt(mse_soap*sum_c)
```

```r
# Confidence Interval for this single contrast is
c(contrast.1.lower,contrast.1.upper)
```

```
## [1] -2.766087 -1.996413
```

part(b)

Null Hypothesis Vs. Alternative Hypothesis

$$H_o : \tau_1 - 1/2(\tau_2 + \tau_3) = 0$$

$$H_a : \tau_1 - 1/2(\tau_2 + \tau_3) \neq 0$$

```r
test_statistics<-abs(mean_y/(sqrt(mse_soap*sum_c)))
test_statistics
```

```
## [1] 13.99752
```

```r
t_val<-qt((.05/2),9,lower.tail=FALSE)
t_val
```

```
## [1] 2.262157
```

Using Part (a) Since 0 in not between this interval we can reject the null hupotheis and conclude that mean weight loss of regular soap is not the same as the average weight for the other two soaps.

Using the formula 4.3.13, we can say that since the test statistics is greater than critical t- value , we can reject the null hypothesis and conclude that mean weight loss of regular soap is not the same as the average weight for the other two soaps.

Using The anova table, p value is very small and less than 0..05. So we can reject the null hypothesis too and and conclude that mean weight loss of regular soap is not the same as the average weight for the other two soaps.

All of the above test give us the same conclusion.

part(c)

We are going to estimate two contrast

1. $\tau_2 - \tau_1$ and 2. $\tau_3 - \tau_1$ Using the Duntee's method first confidence interval is (2.022,3.448) Using the Duntee's method second confidence interval is (1.314,2.741)

Using Bonferroni method for first preplanned contrast

```r
t_val<-qt((.01/4),9,lower.tail=FALSE)
mse_soap<-anova(fit_soap)[["Mean Sq"]][2]
sum_c<-(1+1)/4 # r is individual group
mean_y<-mean(soap2[,5])-mean(soap1[,5])

# First preplanned contrast -0.5231651---5.511921
contrast.1.lower<-mean_y-t_val*sqrt(mse_soap*sum_c)

contrast.1.upper<-mean_y+t_val*sqrt(mse_soap*sum_c)
c(contrast.1.lower,contrast.1.upper)
```

```
## [1] 2.010214 3.459786
```

Another preplanned interval

```r
t_val<-qt((.01/4),9,lower.tail=FALSE)
mse_soap<-anova(fit_soap)[["Mean Sq"]][2]
sum_c<-(1+1)/4 # r is individual group
mean_y<-mean(soap3[,5])-mean(soap1[,5])

# First preplanned contrast -0.5231651---5.511921
contrast.2.lower<-mean_y-t_val*sqrt(mse_soap*sum_c)

contrast.2.upper<-mean_y+t_val*sqrt(mse_soap*sum_c)

c(contrast.2.lower,contrast.2.upper)
```

```
## [1] 1.302714 2.752286
```

We can see that Bonferroni method gives a little wider confidence interval for two preplanned treatment vs. control contrast for simultaneous 99 percent confidence interval.

Tukey's method

```r
t_val<-sqrt(qt(0.01,3,9,lower.tail=FALSE))
mse_soap<-anova(fit_soap)[["Mean Sq"]][2]
sum_c<-(1+1)/4 # r is individual group
mean_y<-mean(soap2[,5])-mean(soap1[,5])

# First preplanned contrast -0.5231651---5.511921
contrast.1.lower<-mean_y-t_val*sqrt(mse_soap*sum_c)
contrast.1.upper<-mean_y+t_val*sqrt(mse_soap*sum_c)

c(contrast.1.lower,contrast.1.upper)
```

```
## [1] 1.395036 4.074964
```

Another Interval

```
t_val<-sqrt(qt(0.01,3,9,lower.tail=FALSE))
mse_soap<-anova(fit_soap)[["Mean Sq"]][2]
sum_c<-(1+1)/4 # r is individual group
mean_y<-mean(soap3[,5])-mean(soap1[,5])

# First preplanned contrast -0.5231651---5.511921
contrast.2.lower<-mean_y-t_val*sqrt(mse_soap*sum_c)

contrast.2.upper<-mean_y+t_val*sqrt(mse_soap*sum_c)


c(contrast.2.lower,contrast.2.upper)
```

```
## [1] 0.6875359 3.3674641
```

We can see that Tukey's method gives a little wider confidence interval for two preplanned treatment vs. control contrast for simultaneous 99 percent confidence interval.

Tukey's method for part (d)

For all pairwise differnce Tukey's method is the best. Since we are takeing into consideration for all pairwise differnce tukey's method in this case gives longer interval than in part (c).

```
TukeyHSD(fit_soap,conf.level=0.99)
```

```
##   Tukey multiple comparisons of means
##     99% family-wise confidence level
##
## Fit: aov(formula = WtLoss ~ factor(Soap), data = soap.origin)
##
## $`factor(Soap)`
##        diff       lwr       upr     p adj
## 2-1  2.7350  1.981034 3.4889657 0.0000006
## 3-1  2.0275  1.273534 2.7814657 0.0000074
## 3-2 -0.7075 -1.461466 0.0464657 0.0142876
```

Ques No 6

(chapter 5 question NO 2):

Checking our assumptions

Most statistical tests results rest on meeting certain assumptions when we run the test. A Oneway ANOVA is no exception. We have assumed 5 things; Check of the model, outlier detection, independence, homogeneity of variance (homoscedasticity) and normality. the fit of the model is checked by plotting the standardized residuals versus the levels of each independent variable (treatment factor, block factor, or covariate) included in the model. We didn't see any significant pattern from the fitted vs. standard residual graph.

Outliers are easy to detect from a plot of the standardized residuals versus the levels of the treatment factors. Plots reveal that there might be one or two outliers detected. But we need to find out if those outliers are influential or not.

The independence assumption is checked by plotting the standardized residuals against the order in which the corresponding observations were collected and against any spatial arrangement of the corresponding experimental units. We can seefrom the graph that soap data holds independence assumption.

The most common pattern of nonconstant variance is that in which the error variance increases as the mean response increases.

The assumption that the error variables have a normal distribution is checked using a normal probability plot, which is a plot of the standardized residuals against their normal scores. Also we can use the shapiro normality test to see the data are normal or not. From the qqplot we can see that data are approximatley normaly distributed except few points. Although our normality test indicated that data are not normal. In this case may be we need to transformation to make the data normal.

```
library(dplyr)
soap<-read.table("soap.txt",header = TRUE) %>%
  select(Soap,WtLoss)
fit_model<-aov(WtLoss~factor(Soap),data=soap)
summary(fit_model)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## factor(Soap)  2 16.122   8.061   105.7 5.62e-07 ***
## Residuals     9  0.686   0.076
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
par(mfrow=c(2,2))
plot(fit_model)

## Test for normality
res=fit_model$residuals
shapiro.test(res)
```

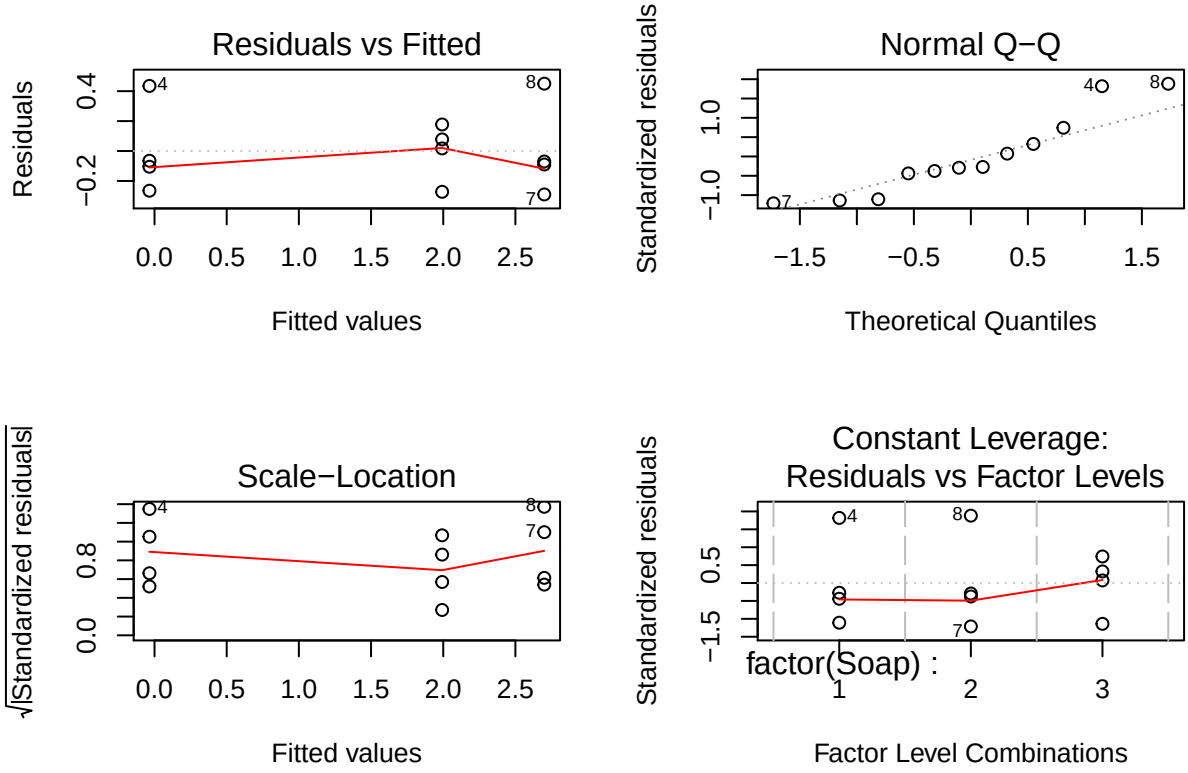
```
##
##  Shapiro-Wilk normality test
##
## data:  res
## W = 0.89383, p-value = 0.132
```

```
## Test for independence
## Durbon watson test
#install.packages("cars")
library(car)
```

```
## Warning: package 'car' was built under R version 3.3.3
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```




```
#dwtest(fit_model,alternative="two.sided")
durbin.watson(fit_model,alternative="two.sided")
```

```
## Warning: 'durbin.watson' is deprecated.
## Use 'durbinWatsonTest' instead.
## See help("Deprecated") and help("car-deprecated").
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1      -0.4036514      2.704544   0.464
##  Alternative hypothesis: rho != 0
```

```
## Test for equal variance:  Equal variances between treatments
#Bartlett Test in R:
bartlett.test(soap$WtLoss~soap$Soap)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  soap$WtLoss by soap$Soap
## Bartlett's K-squared = 0.68233, df = 2, p-value = 0.7109
```

```
oneway.test(soap$WtLoss~ soap$Soap)
```

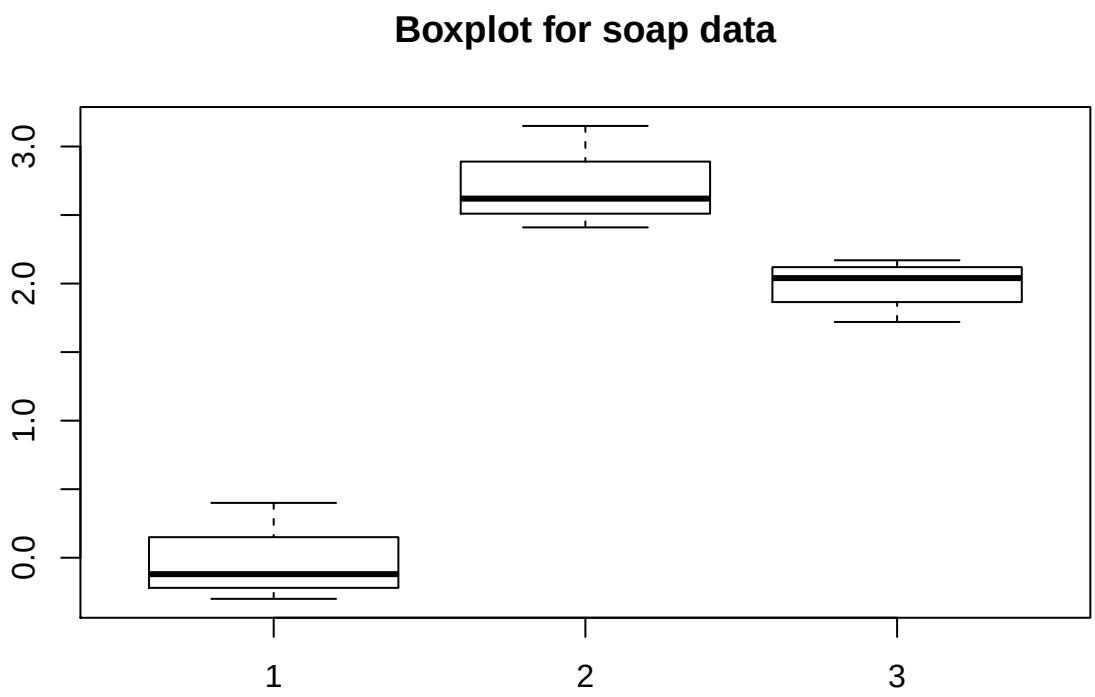
```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  soap$WtLoss and soap$Soap
## F = 80.802, num df = 2.0000, denom df = 5.6683, p-value =
```

```
## 6.822e-05
```

```
# versus this which is what we've done so far
oneway.test(soap$WtLoss~ soap$Soap, var.equal = TRUE)
```

```
##
##  One-way analysis of means
##
## data:  soap$WtLoss and soap$Soap
## F = 105.7, num df = 2, denom df = 9, p-value = 5.619e-07
```

```
# For continuous variable (convert to categorical if needed.)
par(mfrow=c(1,1))
boxplot(soap$WtLoss~ soap$Soap, data=soap, main="Boxplot for soap data")
```



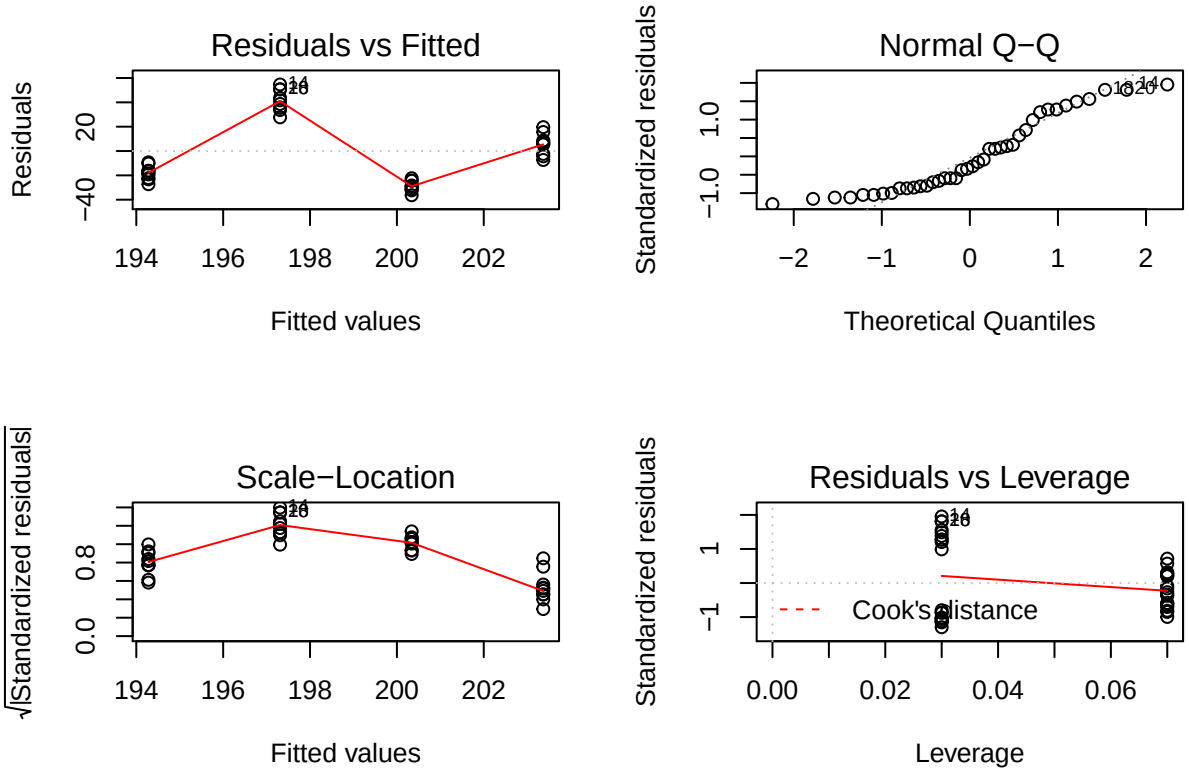
Ques No 7(chapter 5 question NO 3):

part(a)

```
margarine<-read.csv("margarine.csv")
y_bar=c(176.4,238.6,171.4,208.90)
s_bar=c(5.56,8.66,4.27,8.45)

8.66/4.27
```

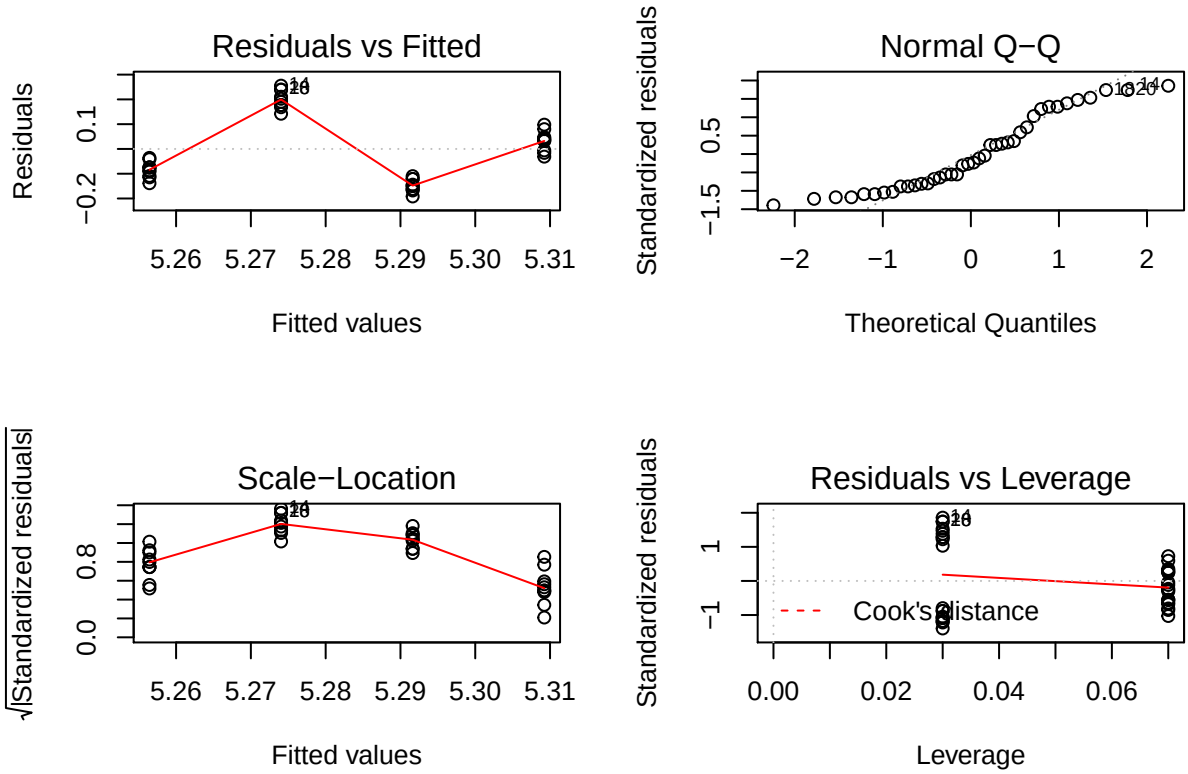
```
## [1] 2.028103
```

```r
margarin_old<-c(margarine$margarine.1,margarine$margarine.2,margarine$margarine.3,margarine$butter)
marga2<-c(rep(1,length(margarine$margarine.1)),rep(2,length(margarine$margarine.1)),rep(3,length(margari

new.merge<-data.frame(marga2,margarin_old)
fit_merge<-aov(margarin_old~marga2,data = new.merge)
par(mfrow=c(2,2))
plot(fit_merge)
```



Since the values is less than 3 we dont need any transformation. But the student residual vs fitted plot suggest that unequal variance exists.

```r
# After Transformation : Use Log transformation

margarin<-log(margarin_old)
marga2<-c(rep(1,length(margarine$margarine.1)),rep(2,length(margarine$margarine.1)),rep(3,length(margari
new.merge<-data.frame(marga2,margarin)

# Anova model fit
fit_merge<-aov(margarin~marga2,data = new.merge)
par(mfrow=c(2,2))
plot(fit_merge)
```



part(b)

```
## or merging for ready
#margarin_old<-c(margarine$margarine.1,margarine$margarine.2,margarine$margarine.3,margarine$butter)


t_val<-qt((.05/2),40,lower.tail=FALSE)
mse_soap<-anova(fit_merge)[["Mean Sq"]][2]
sum_c<-(1+(1/3)^2+(1/3)^2+(1/3)^2)/10 # r is individual group
mean_y<-1/3*mean(margarine$margarine.1)+(1/3)*mean(margarine$margarine.2)+(1/3)*mean(margarine$margarin

# First preplanned contrast -0.5231651---5.511921
contrast.1.lower<-mean_y-t_val*sqrt(mse_soap*sum_c)
contrast.1.upper<-mean_y+t_val*sqrt(mse_soap*sum_c)
c(contrast.1.lower,contrast.1.upper)
```

```
## [1] -13.53642 -13.33024
```

part (c)

```
margarin_old<-c(margarine$margarine.1,margarine$margarine.2,margarine$margarine.3,margarine$butter)
new.merge<-data.frame(marga2,margarin_old)
## We ned to use satterthwise approximation
```

```
sum_c<-((1*sd(margarine$butter)+(1/3)^2*sd(margarine$margarine.1)+(1/3)^2*sd(margarine$margarine.2)+(1/3

df=(sum_c)^2/((sum_c)^2/9)

t_val<-qt((.05/2),df,lower.tail=FALSE)

mean_y<-1/3*mean(margarine$margarine.1)+(1/3)*mean(margarine$margarine.2)+(1/3)*mean(margarine$margarine

# First preplanned contrast -0.5231651---5.511921
contrast.1.lower<-mean_y-t_val*sqrt(mse_soap*sum_c)
contrast.1.upper<-mean_y+t_val*sqrt(mse_soap*sum_c)
c(contrast.1.lower,contrast.1.upper)
```

```
## [1] -13.75724 -13.10943
```

part(d)

I would prefer transformed data.

Ques No 8(chapter 5 question NO 8):

Part (a)

```
wildflower<-read.csv("wildflower.csv")
head(wildflower)
```

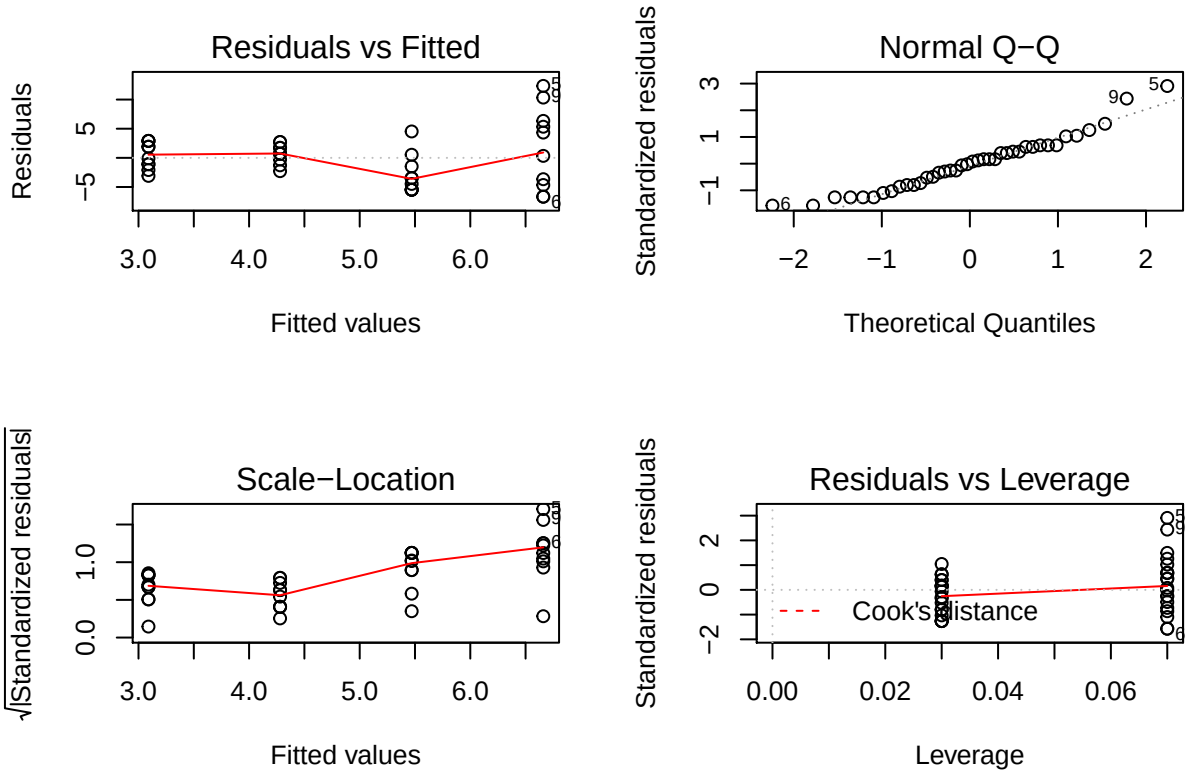
```
##   sp_stra sp_unstra summer_stra summer_unstra
## 1      12         6           6             0
## 2      13         2           4             6
## 3       2         0           5             2
## 4       7         2           7             5
## 5      19         4           6             1
## 6       0         1           5             5
```

```
# find the groupwise mean and variance

sp_stra_sd<-sd(wildflower$sp_stra)
sp_stra_mean<-mean(wildflower$sp_stra)

sp_unstra_sd<-sd(wildflower$sp_unstra)
sp_unstra_mean<-mean(wildflower$sp_unstra)

summer_stra_sd<-sd(wildflower$summer_stra)
summer_stra_mean<-mean(wildflower$summer_stra)

summer_unstra_sd<-sd(wildflower$summer_unstra)
summer_unstra_mean<-mean(wildflower$summer_unstra)

all_sd<-c(sp_stra_sd,sp_unstra_sd,summer_stra_sd,summer_unstra_sd)
all_mean<-c(sp_stra_mean,sp_unstra_mean,summer_stra_mean,summer_unstra_mean)
```

```r
sp_stra_sd/summer_stra_sd
```

```
## [1] 4.28369
```

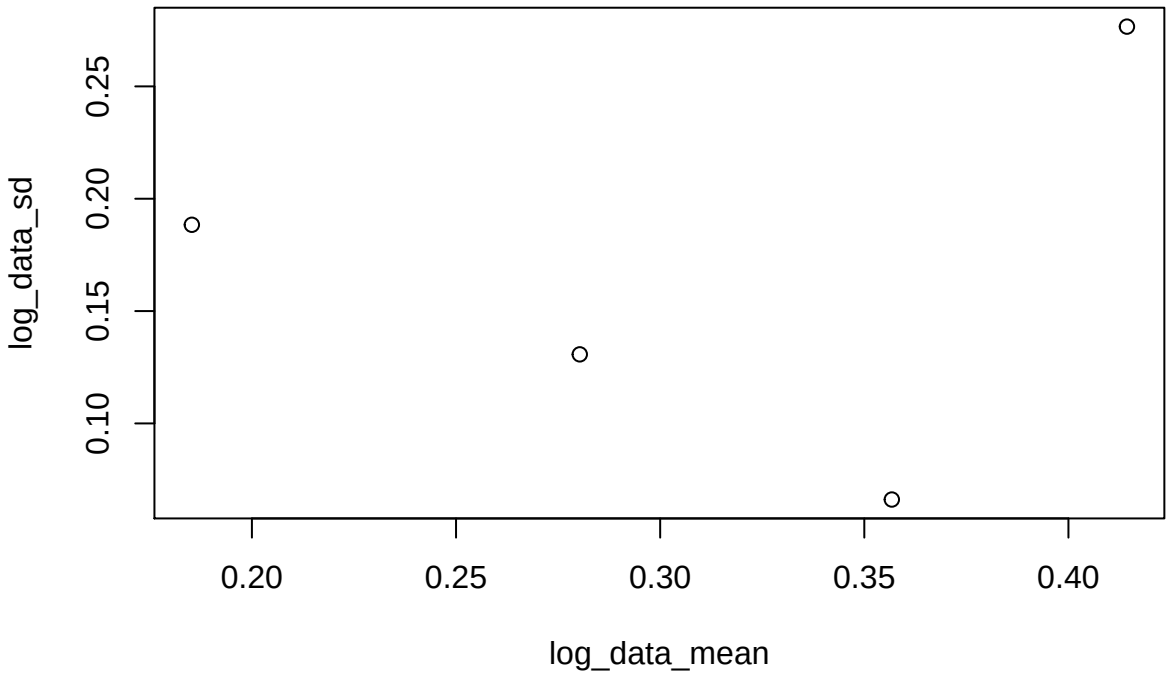
```r
#4.28369 and since it is greater than 3 so we need to perform variance stabilation technique suggested

## graphically assumption analysis
## Anova formation
wild_fl<-c(wildflower$sp_stra,wildflower$sp_unstra,wildflower$summer_stra,wildflower$summer_unstra)
tr_t<-c(rep(1,10),rep(2,10),rep(3,10),rep(4,10))
wild_new<-data.frame(tr_t,wild_fl)
fit_wild<-aov(wild_fl~tr_t,data=wild_new)
par(mfrow=c(2,2))
plot(fit_wild)
```



Part (B)

```r
## (B)
## data transformation technique

transformed_wild1<-asin(sqrt(wildflower$sp_stra/40))
transformed_wild2<-asin(sqrt(wildflower$sp_unstra/40))
transformed_wild3<-asin(sqrt(wildflower$summer_stra/40))
transformed_wild4<-asin(sqrt(wildflower$summer_unstra/40))
```

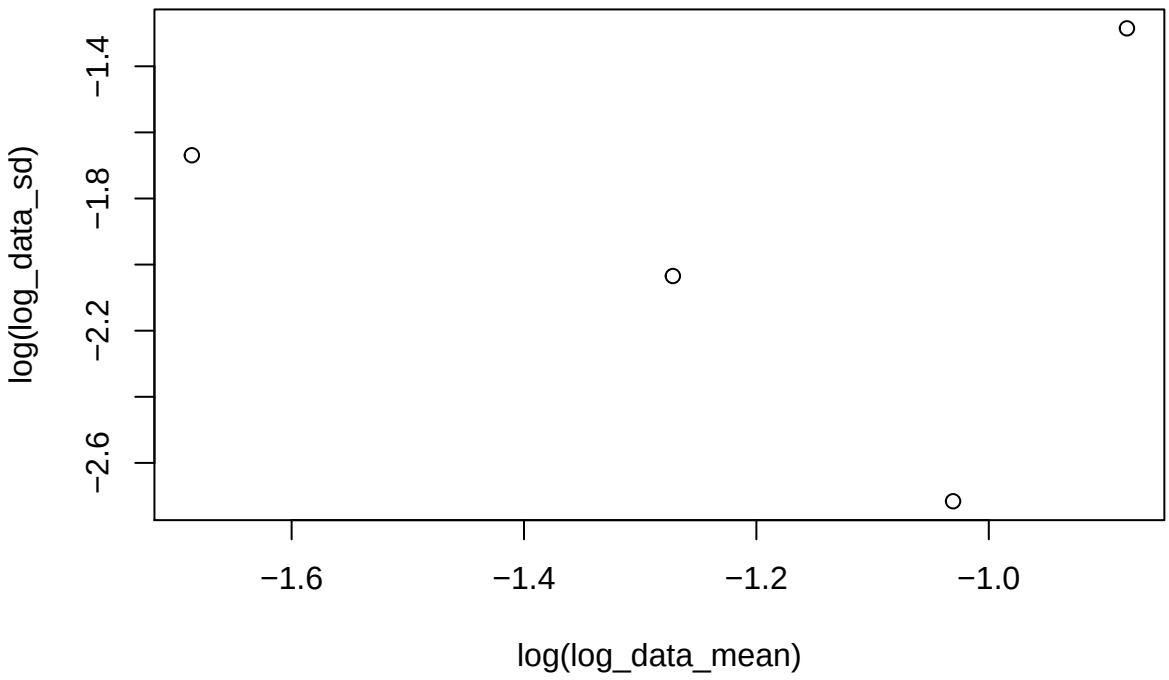
```r
log_data_mean<-c(mean(transformed_wild1),mean(transformed_wild2),mean(transformed_wild3),mean(transform
log_data_sd<-c(sd(transformed_wild1),sd(transformed_wild2),sd(transformed_wild3),sd(transformed_wild4))
plot(log_data_mean,log_data_sd)
```




```r
# does not improve
```

part (c)

```r
plot(log(log_data_mean),log(log_data_sd))
```



It does not change the variance

part (d)