H4

Mingming Xu

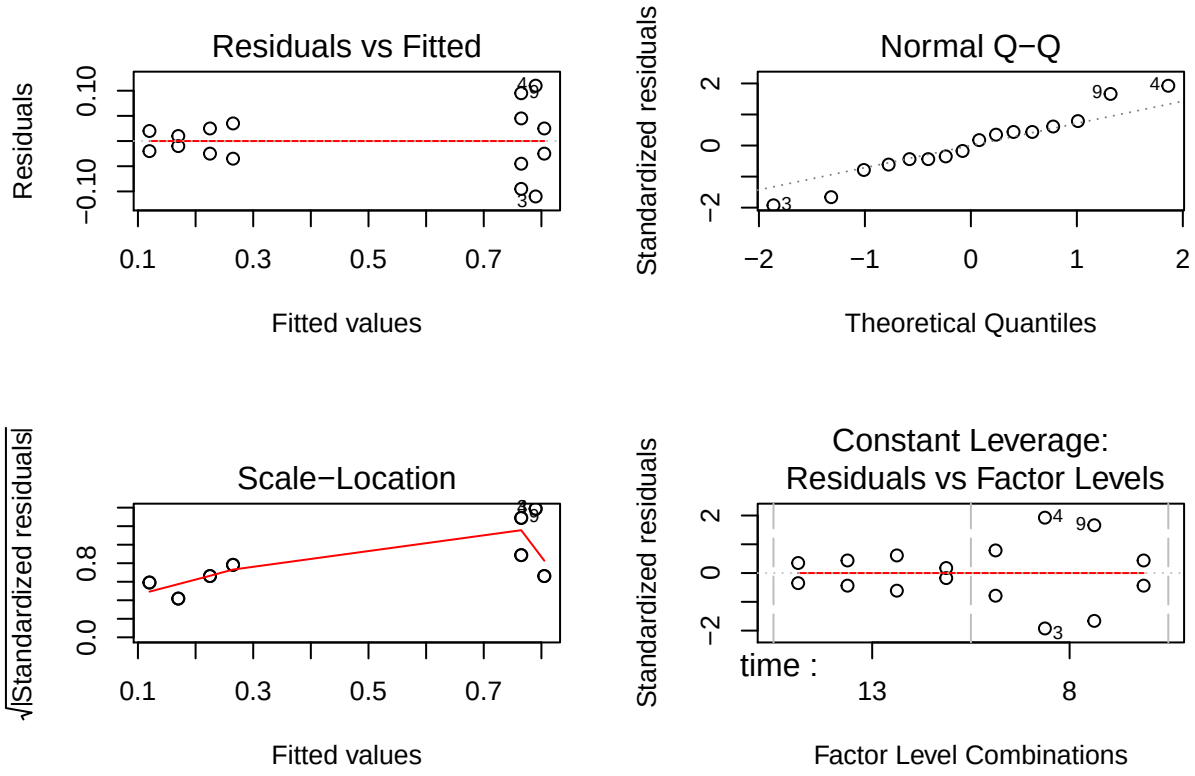
2019/10/17

Oehlert Problem 8.4.

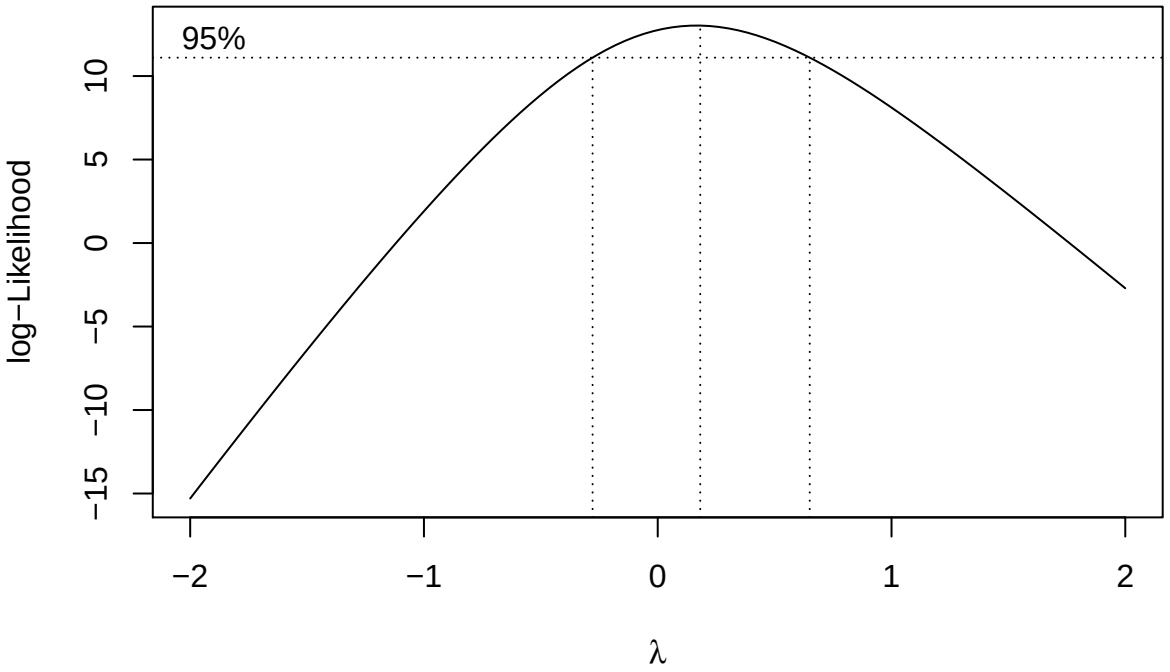
```
c1<-c(.83,.78)
c2<-c(.68,.90)
c3<-c(.18,.16)
c4<-c(.25,.20)
c5<-c(.86,.67)
c6<-c(.72,.81)
c7<-c(.30,.23)
c8<-c(.10,.14)

delamination<-data.frame(substrates=c(c1,c2,c3,c4,c5,c6,c7,c8),
                  time=c(rep("8",4),rep("13",4),rep("8",4),rep("13",4)),
                  airflow=c( rep(c(rep("low",2),rep("high",2)),2)),
                  laser=c(rep("old",8),rep("new",8)) )

attach(delamination)
model_1=aov(substrates~time*airflow*laser,data = delamination)
summary(model_1)
```

```
##                    Df Sum Sq Mean Sq F value   Pr(>F)
## time                1 1.3748  1.3748 210.489 4.99e-07 ***
## airflow             1 0.0028  0.0028   0.422    0.534
## laser               1 0.0014  0.0014   0.215    0.655
## time:airflow        1 0.0014  0.0014   0.215    0.655
## time:laser          1 0.0008  0.0008   0.116    0.742
## airflow:laser       1 0.0086  0.0086   1.310    0.285
## time:airflow:laser  1 0.0116  0.0116   1.769    0.220
## Residuals           8 0.0523  0.0065
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##examine the statistical assumptions:
##these observations and groups are independent
par(mfrow=c(2,2))
plot(model_1)
```




```
par(mfrow=c(1,1))
##In the residuals vs fitted values plot,it is not a ideal pattern.So, the assumption of cantant varian
library(MASS)
par(mfrow=c(1,1))
b<-boxcox(model_1)
```



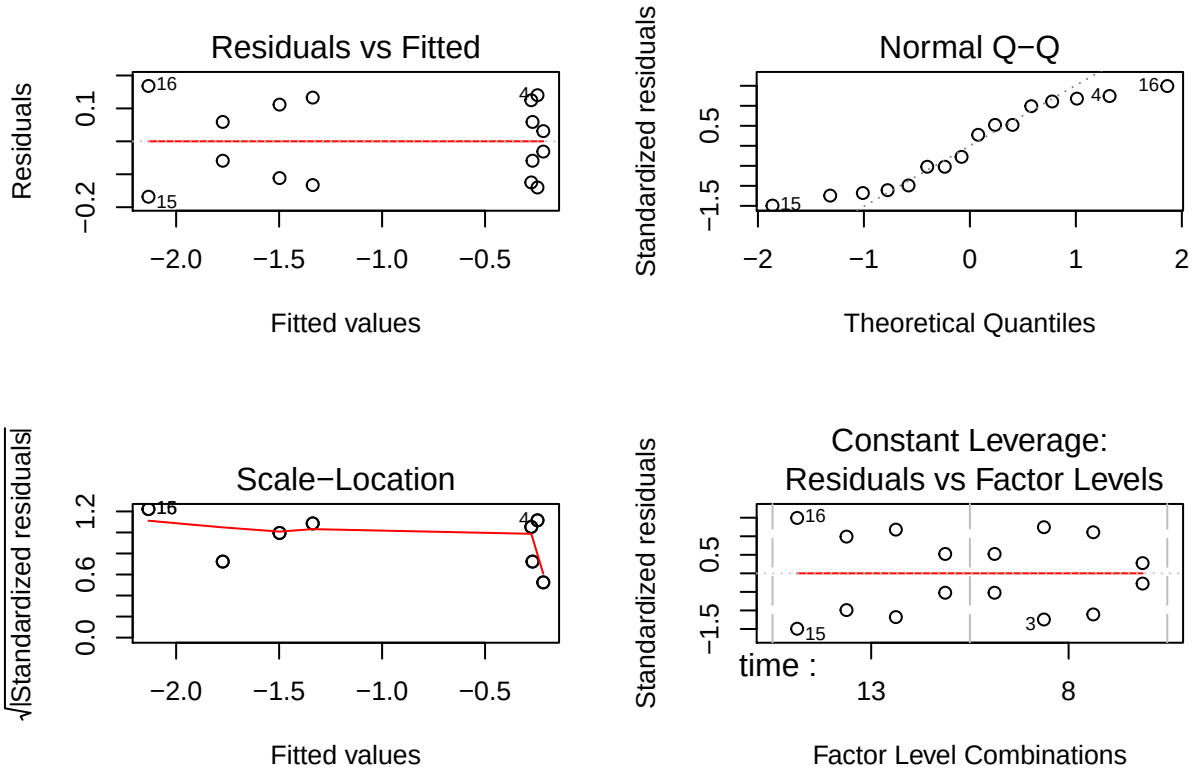

```
lamb<-b$x[which.max(b$y)]
lamb
```

```
## [1] 0.1818182
```

```
##Take lambda is 0
model_11=aov(log(substrates)~time*airflow*laser,data = delamination)
summary(model_11)
```

```
##                     Df Sum Sq Mean Sq F value   Pr(>F)
## time                 1  8.221   8.221 323.906 9.32e-08 ***
## airflow              1  0.074   0.074   2.912  0.12628
## laser                1  0.020   0.020   0.783  0.40200
## time:airflow         1  0.062   0.062   2.459  0.15551
## time:laser           1  0.003   0.003   0.135  0.72241
## airflow:laser        1  0.270   0.270  10.637  0.01150 *
## time:airflow:laser   1  0.307   0.307  12.081  0.00837 **
## Residuals            8  0.203   0.025
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##examine the statistical assumptions:
##these observations and groups are independent
par(mfrow=c(2,2))
plot(model_11)
```



Before boxcox transformation, based on the p-value, the term that has significant diffrence is time.

After boxcox transformation, based on the p-value,the terms that have significanr diffrence are time,airflow:laser and time:airflow:laser. So,the treatments time,airflow and laser exist significant diffrence.

Hothorn and Everitt Ex. 5.2

```
library(HSAUR3)
```

```
## Loading required package: tools
```

```r
data(schooldays,package="HSAUR3")
head(schooldays)
```

```
##         race gender school learner absent
## 1 aboriginal   male     F0    slow      2
## 2 aboriginal   male     F0    slow     11
## 3 aboriginal   male     F0    slow     14
## 4 aboriginal   male     F0 average      5
## 5 aboriginal   male     F0 average      5
## 6 aboriginal   male     F0 average     13
```

```r
attach(schooldays)
```

a.

```r
model2_1=aov(absent~race*gender*school*learner,data = schooldays)
summary(model2_1)
```

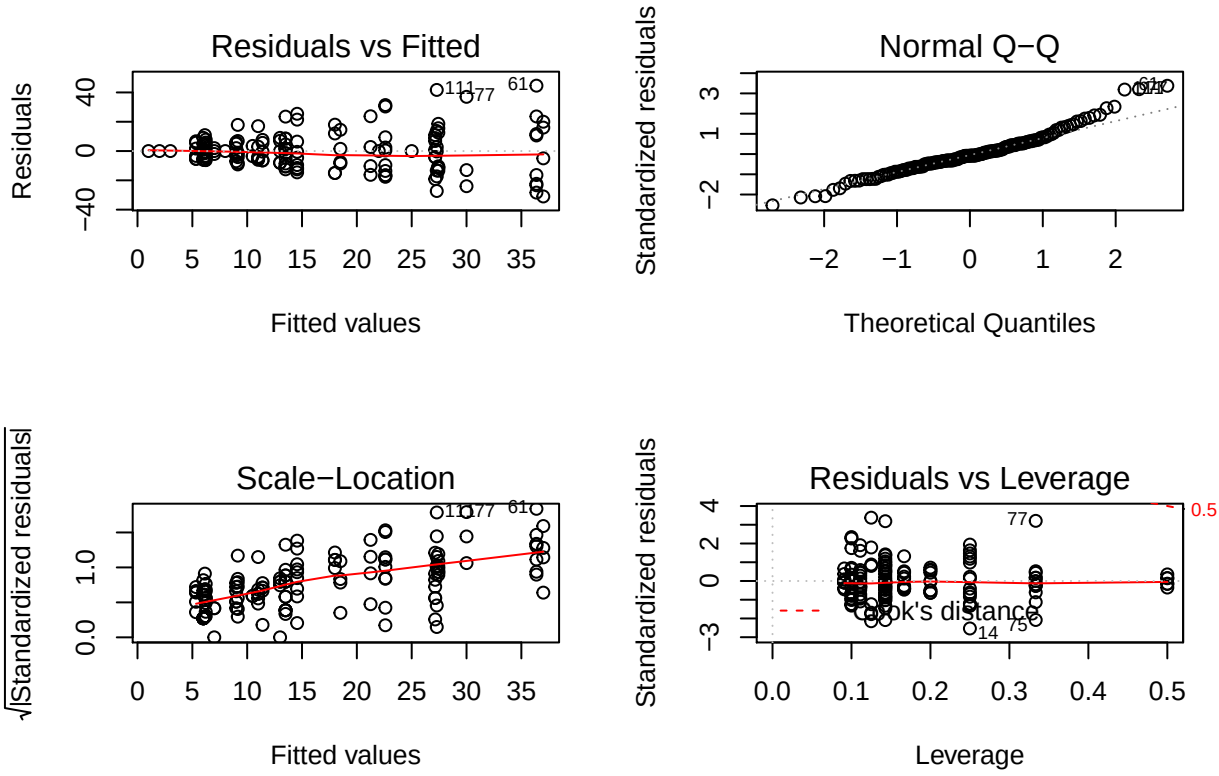
```
##                              Df Sum Sq Mean Sq F value   Pr(>F)
## race                          1   2646  2645.7  13.258 0.000400 ***
## gender                        1    339   338.9   1.698 0.194985
## school                        3   1222   407.3   2.041 0.111672
## learner                       1     17    17.3   0.087 0.769087
## race:gender                   1    173   173.3   0.868 0.353295
## race:school                   3   3628  1209.2   6.059 0.000702 ***
## gender:school                 3   1504   501.3   2.512 0.061747 .
## race:learner                  1     67    67.4   0.338 0.562291
## gender:learner                1      9     9.4   0.047 0.828522
## school:learner                3   1931   643.5   3.225 0.025011 *
## race:gender:school            3    206    68.7   0.344 0.793404
## race:gender:learner           1    388   387.6   1.942 0.165957
## race:school:learner           3   1419   473.0   2.370 0.073836 .
## gender:school:learner         3    609   203.1   1.018 0.387326
## race:gender:school:learner    3    202    67.4   0.338 0.798014
## Residuals                   122  24346   199.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
##examine the statistical assumptions:
##these observations and groups are independent
par(mfrow=c(2,2))
plot(model2_1)
```

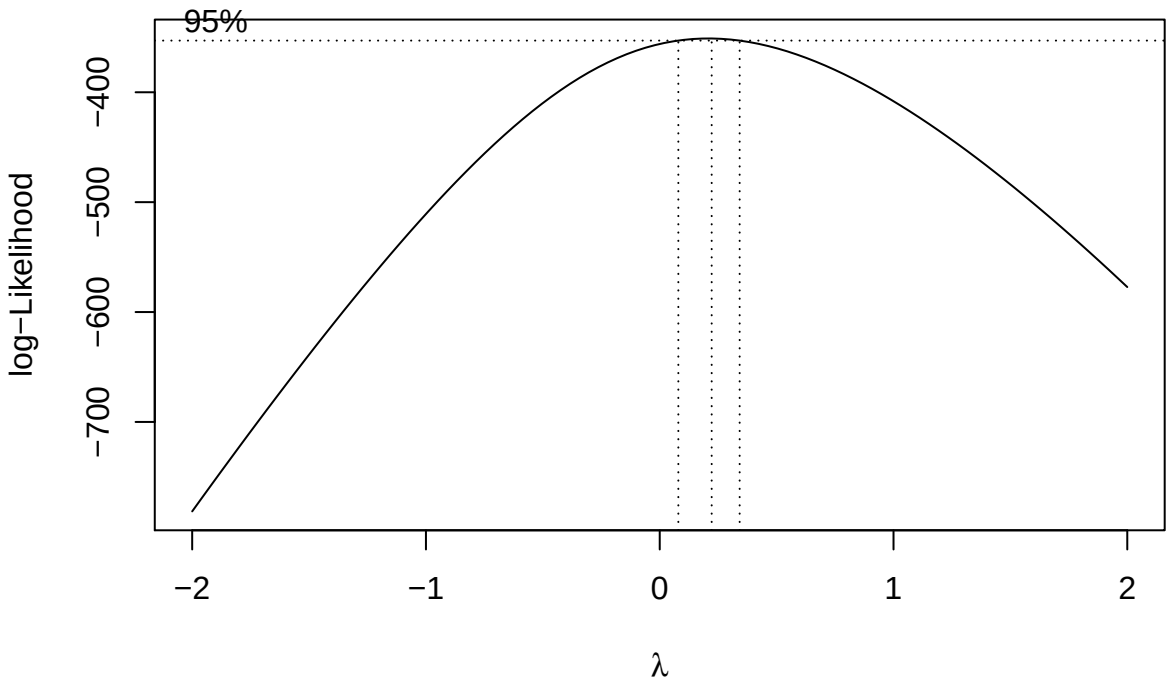
```
## Warning: not plotting observations with leverage one:
##   34, 62, 93, 112, 143, 144
```



```
## Warning: not plotting observations with leverage one:
##   34, 62, 93, 112, 143, 144
```




```
par(mfrow=c(1,1))
##In the residuals vs fitted values plot,these points have a tendency to gradually spread along the x-a
schooldays$absent=schooldays$absent+1
library(MASS)
par(mfrow=c(1,1))
b<-boxcox(model2_1)
```



```r
lam<-b$x[which.max(b$y)]
lam
```

```
## [1] 0.2222222
```

```r
##Take lambda=0
model2_11=aov(log(absent)~race*gender*school*learner,data = schooldays)
summary(model2_11)
```

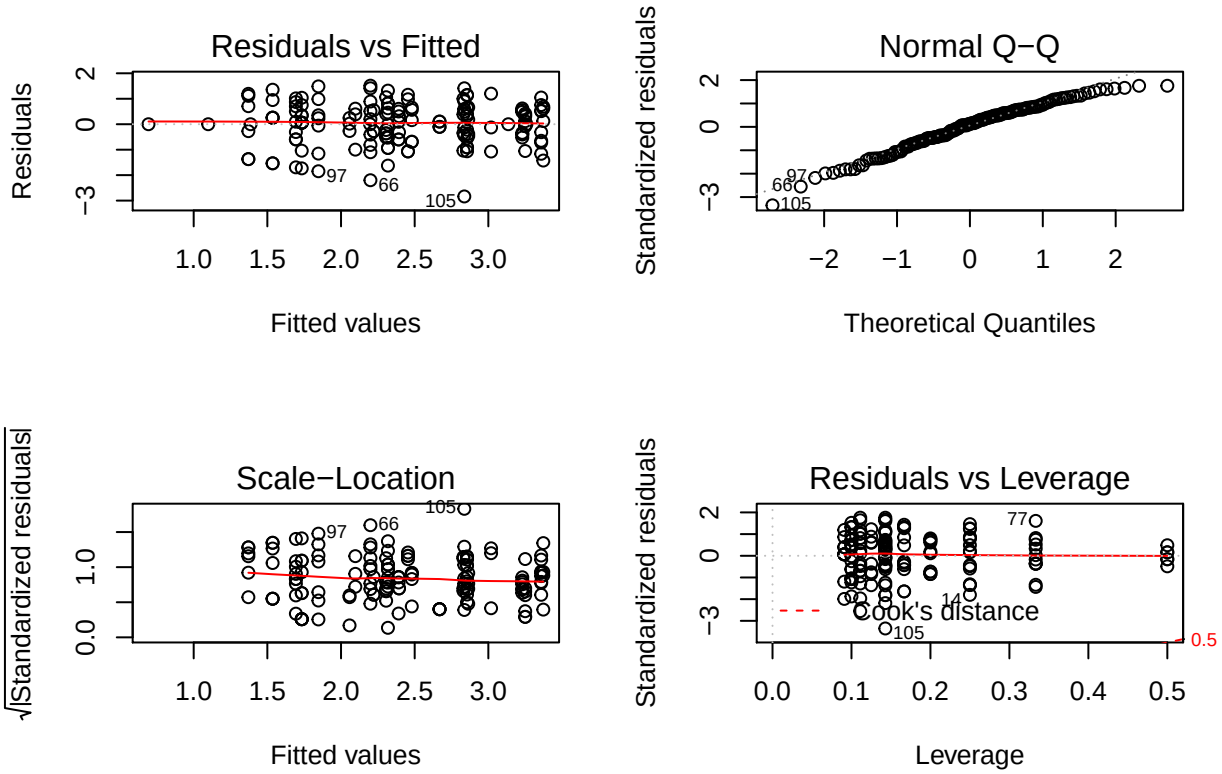
```
##                              Df Sum Sq Mean Sq F value  Pr(>F)
## race                          1  14.89  14.889  17.820 4.7e-05 ***
## gender                        1   0.90   0.902   1.080 0.30072
## school                        3   2.27   0.758   0.907 0.44003
## learner                       1   0.00   0.004   0.005 0.94349
## race:gender                   1   0.21   0.205   0.245 0.62119
## race:school                   3  11.07   3.691   4.417 0.00551 **
## gender:school                 3   8.33   2.778   3.325 0.02203 *
## race:learner                  1   0.07   0.073   0.087 0.76813
## gender:learner                1   0.01   0.012   0.014 0.90634
## school:learner                3   5.70   1.900   2.274 0.08342 .
## race:gender:school            3   2.36   0.788   0.943 0.42193
## race:gender:learner           1   2.62   2.617   3.132 0.07928 .
## race:school:learner           3   6.14   2.048   2.451 0.06670 .
## gender:school:learner         3   3.21   1.070   1.281 0.28406
## race:gender:school:learner    3   0.97   0.324   0.387 0.76224
## Residuals                   122 101.93   0.836
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
##examine the statistical assumptions:
##these observations and groups are independent
par(mfrow=c(2,2))
plot(model2_11)
```

```
## Warning: not plotting observations with leverage one:
##   34, 62, 93, 112, 143, 144
```



```
## Warning: not plotting observations with leverage one:
##   34, 62, 93, 112, 143, 144
```




```
par(mfrow=c(1,1))
##In the residuals vs fitted values plot, the most of the dots are scattered around the red line which i
##In the Q-Q plot, the most of points are on the line. So, the assumption of nomarlity distributed has b
```

Based on the p-value,the items which have sinificant diffence are race,race:school,school:leaner

So,the factors that affect absenteeism are race ,school and learner.

b.

```
model2_2=aov(log(absent)~gender*school*race*learner,data = schooldays)
summary(model2_2)
```

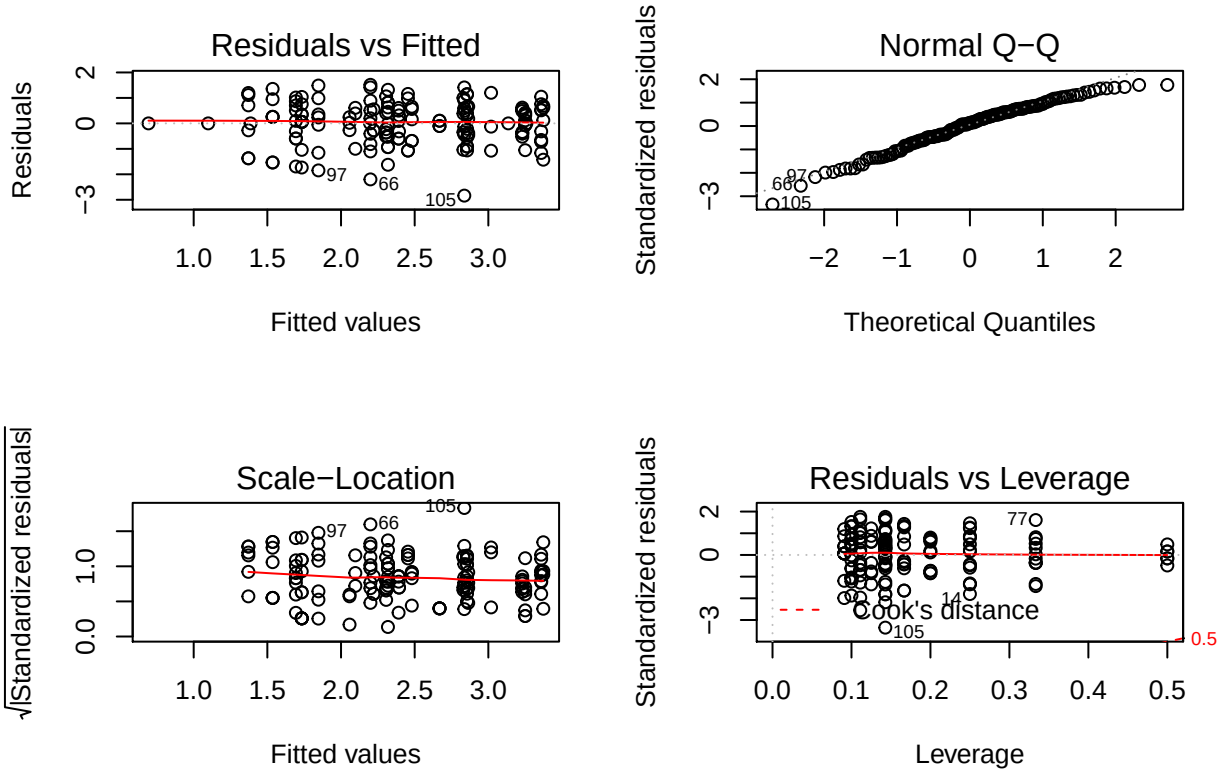
```
##                              Df Sum Sq Mean Sq F value   Pr(>F)
## gender                        1   0.78   0.785   0.939  0.33444
## school                        3   3.38   1.127   1.349  0.26187
## race                          1  13.90  13.899  16.635 8.11e-05 ***
## learner                       1   0.00   0.004   0.005  0.94349
## gender:school                 3   8.98   2.993   3.582  0.01589 *
## gender:race                   1   0.06   0.061   0.073  0.78744
## school:race                   3  10.57   3.524   4.218  0.00709 **
## gender:learner                1   0.01   0.014   0.016  0.89896
## school:learner                3   5.77   1.923   2.301  0.08056 .
## race:learner                  1   0.00   0.002   0.002  0.96561
## gender:school:race            3   2.36   0.788   0.943  0.42193
## gender:school:learner         3   3.60   1.201   1.437  0.23532
## gender:race:learner           1   2.60   2.596   3.107  0.08045 .
## school:race:learner           3   5.77   1.924   2.303  0.08039 .
## gender:school:race:learner    3   0.97   0.324   0.387  0.76224
## Residuals                   122 101.93   0.836
## ---
```

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

```
par(mfrow=c(2,2))
plot(model2_2)
```

```
## Warning: not plotting observations with leverage one:
##   34, 62, 93, 112, 143, 144
```

```
## Warning: not plotting observations with leverage one:
##   34, 62, 93, 112, 143, 144
```




```
par(mfrow=c(1,1))## the assumptions have been met
```

Each term's p-value has changed but gender:school:race:learner p-value is same (a).

c.

```
library(car)
```

```
## Loading required package: carData
```

```
Anova(lm(log(absent)~race*gender*school*learner,data = schooldays),type=2)
```

```
## Anova Table (Type II tests)
##
## Response: log(absent)
##                          Sum Sq  Df F value    Pr(>F)
## race                     12.070   1 14.4461 0.0002265 ***
## gender                    0.977   1  1.1693 0.2816804
## school                    2.007   3  0.8007 0.4958067
## learner                   0.062   1  0.0747 0.7851147
## race:gender               0.296   1  0.3542 0.5528227
## race:school               4.840   3  1.9311 0.1280931
```

```
## gender:school               11.000   3  4.3887 0.0057087 **
## race:learner                 0.086   1  0.1034 0.7483367
## gender:learner               0.357   1  0.4267 0.5148267
## school:learner               7.359   3  2.9358 0.0361005 *
## race:gender:school           1.457   3  0.5814 0.6283416
## race:gender:learner          0.676   1  0.8086 0.3703125
## race:school:learner          5.773   3  2.3030 0.0803907 .
## gender:school:learner        3.210   3  1.2808 0.2840598
## race:gender:school:learner   0.971   3  0.3874 0.7622400
## Residuals                  101.933 122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(aov(log(absent)~race*gender*school*learner,data = schooldays))
```

```
##                              Df Sum Sq Mean Sq F value  Pr(>F)
## race                          1  14.89  14.889  17.820 4.7e-05 ***
## gender                        1   0.90   0.902   1.080 0.30072
## school                        3   2.27   0.758   0.907 0.44003
## learner                       1   0.00   0.004   0.005 0.94349
## race:gender                   1   0.21   0.205   0.245 0.62119
## race:school                   3  11.07   3.691   4.417 0.00551 **
## gender:school                 3   8.33   2.778   3.325 0.02203 *
## race:learner                  1   0.07   0.073   0.087 0.76813
## gender:learner                1   0.01   0.012   0.014 0.90634
## school:learner                3   5.70   1.900   2.274 0.08342 .
## race:gender:school            3   2.36   0.788   0.943 0.42193
## race:gender:learner           1   2.62   2.617   3.132 0.07928 .
## race:school:learner           3   6.14   2.048   2.451 0.06670 .
## gender:school:learner         3   3.21   1.070   1.281 0.28406
## race:gender:school:learner    3   0.97   0.324   0.387 0.76224
## Residuals                   122 101.93   0.836
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In these two models, the Residuals' Sum Sq and df keep the same but other terms' Sum Sq changed.

Beside, in the model using Type I sum of squares, the terms that have sinificant diffence are race, race:school,gender:school. In the model using Type II sum of squares, the terms that have sinificant diffence are race,gender:school and school:learner.

d.

For factor race, if I put race into model first

Type I sum of squares: SS(race)

Type II sum of squares: SS(race|1,gender,school,learner,gender:school,gender:learner,school:learner,gender:school:learner)

e.

For race:gender:school:learner

Type I sum of squares: SS(race:gender:school:learner|race,gender,school,learnerr,race:gender,race:school,gender:school,race:learn

Type II sum of squares:SS(race:gender:school:learner|1,race,gender,school,learnerr,race:gender,race:school,gender:school,race:lea

When we calculate Type I sum of squares of race:gender:school:learner, it is actually an incremental improvment given other items added to the modle. So, they are same.

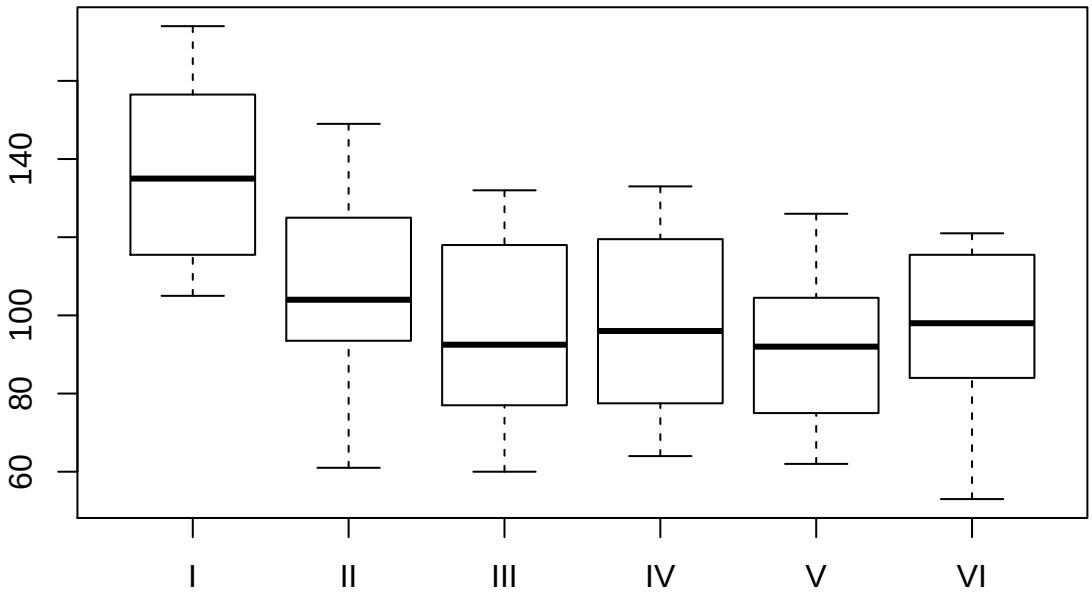
Oat Problem

```
library (MASS)
data(oats, package="MASS")
attach(oats)
head(oats)
```

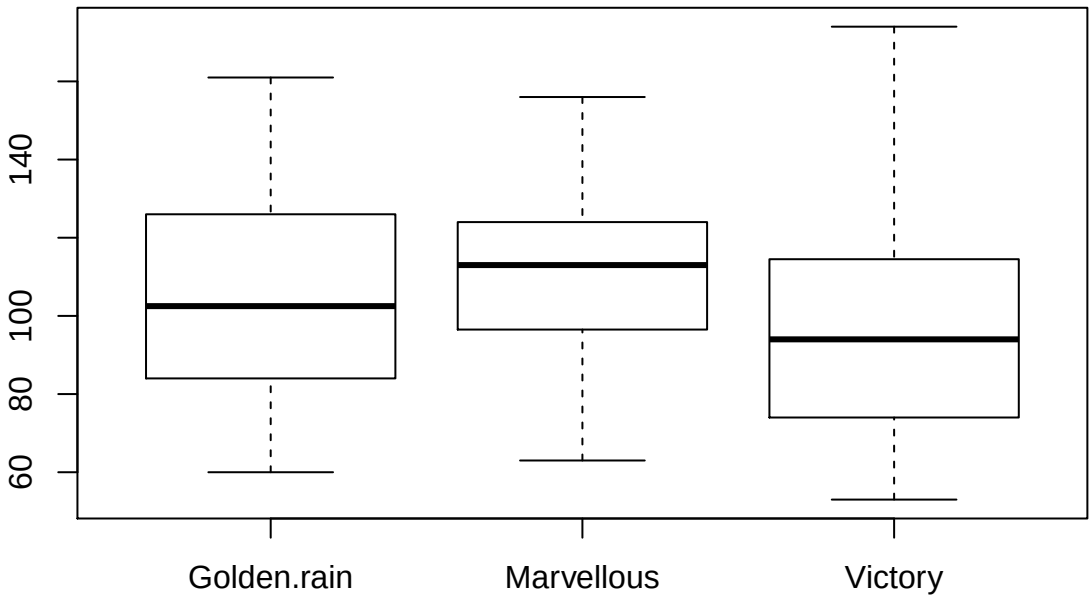
```
##   B           V      N   Y
## 1 I     Victory 0.0cwt 111
## 2 I     Victory 0.2cwt 130
## 3 I     Victory 0.4cwt 157
## 4 I     Victory 0.6cwt 174
## 5 I Golden.rain 0.0cwt 117
## 6 I Golden.rain 0.2cwt 114
```

a.

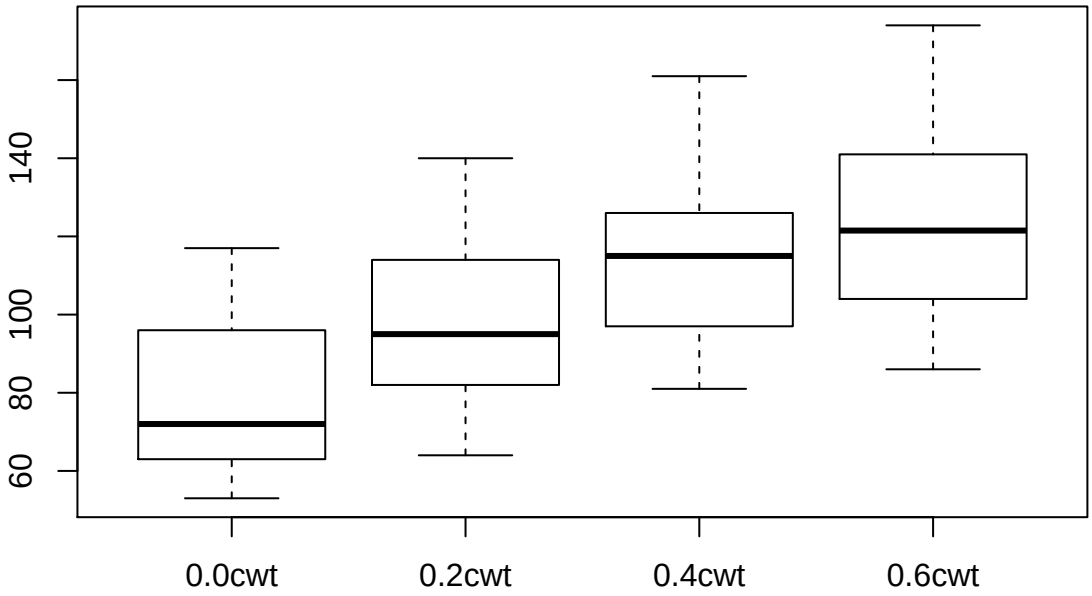
```
boxplot(Y~B)
```




```
boxplot(Y~V)
```




```
boxplot(Y~N)
```



From the plot, the treatment Nitrogen concentration appears diffrence.And,the block appears diffence.

b.

Hypothesis test:

Nitrogen concentration:

H_0: All nitrogen concentrations are the same H_a: At least one nitrogen concentration is diffrent

Variety of oats:

H_0: All variety of oats are the same H_a:At least one kind of oats is diffrent

Nitrogen concentration by Variety of oats interaction:

H_0: All interactions are the same H_a: All interactions are the same

```r
model3_1=aov(Y~V*N,data=oats)
summary(model3_1)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## V            2   1786     893   1.795    0.175
## N            3  20020    6673  13.411 8.37e-07 ***
## V:N          6    322      54   0.108    0.995
## Residuals   60  29857     498
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Because the p-value of term N is less than 0.05, we reject the null hypothesis.So, there is significant diffence between nitrogen concentrations.

Because the p-value of term V and term V:N is greater than 0.05 or 0.1, we fail to reject the null hypothesis.So, there is not significant diffence between variety of oats, and also nitrogen concentration by variety of oats interaction.

c.

```r
model3_2=aov(Y~V+N+V:N+B,data=oats)
summary(model3_2)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## V            2   1786     893   3.513   0.0366 *
## N            3  20020    6673  26.251 1.13e-10 ***
## B            5  15875    3175  12.489 4.09e-08 ***
## V:N          6    322      54   0.211   0.9719
## Residuals   55  13982     254
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Because the p-value of term V and term N is less than 0.05 , we reject the null hypothesis.So, there is significant diffrence between variety of oats, and also nitrogen concentration. Also, the p-value of block is less than 0.05, the block has significanr diffrence.

Because the p-value of term V:N is greater than 0.05 or 0.1, we fail to reject the null hypothesis.So, there is not significant diffence between nitrogen concentration by variety of oats interaction.

By including block in my model, it causes the reduction of sum squares ofresiduals ,loss of df for residuals and the reduction of MSE. Because of the reduction of MSE, F value becomes larger and p-value becomes samller of the terms N and V. Hence, we get that variety of oats have significant diffence which is opposite to part(b).

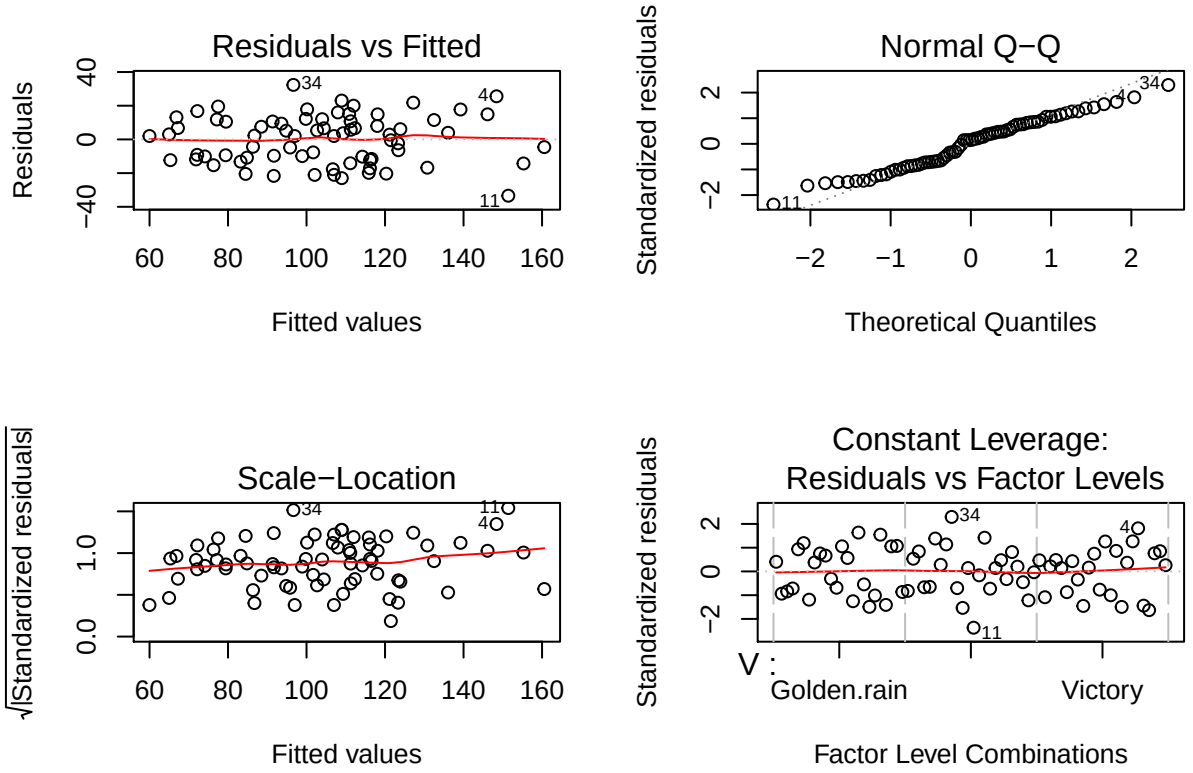
d.

```r
##my final model: from part(c)
model3_3=aov(Y~V+N+B,data=oats)
summary(model3_3)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## V            2   1786     893   3.809   0.0276 *
## N            3  20020    6673  28.460 1.24e-11 ***
## B            5  15875    3175  13.540 6.91e-09 ***
## Residuals   61  14304     234
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
##examine the statistical assumptions:
##The observations and groups are independent
```

```
par(mfrow=c(2,2))
plot(model3_3)
```




```
##In the residuals vs fitted values plot, the most of the dots are scattered around the red line which i
##In the Q-Q plot, the most of points are on the line. So, the assumption of nomarlity distributed has b
```

e.

```
TukeyHSD(aov(Y~V+N,data=oats))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Y ~ V + N, data = oats)
##
## $V
##                               diff        lwr       upr     p adj
## Marvellous-Golden.rain    5.291667  -9.509153 20.092486 0.6689436
## Victory-Golden.rain      -6.875000 -21.675819  7.925819 0.5090858
## Victory-Marvellous      -12.166667 -26.967486  2.634153 0.1274623
##
## $N
##                     diff       lwr      upr     p adj
## 0.2cwt-0.0cwt 19.500000  0.712970 38.28703 0.0389760
## 0.4cwt-0.0cwt 34.833333 16.046303 53.62036 0.0000400
## 0.6cwt-0.0cwt 44.000000 25.212970 62.78703 0.0000003
## 0.4cwt-0.2cwt 15.333333 -3.453697 34.12036 0.1479244
## 0.6cwt-0.2cwt 24.500000  5.712970 43.28703 0.0055113
## 0.6cwt-0.4cwt  9.166667 -9.620363 27.95370 0.5749975
```

By Tukey's test without block, we could see that the treatment nitrogen concentrations have sinificant diffrence from the p-value.