STAT5353HW4

Mingming Xu

10/9/2020

E8.3

Let two factors be A and B, each at two levels.

Four factor/level combinations.

N=20 wafers and balanced data.

```
##A skeleton ANOVA table
Source<-c("A","B","AB","Error")
df<-c("1","1","1","16")
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 3.6.2
```

```
kable(cbind(Source,df))
```

Source	df
A	1
B	1
AB	1
Error	16

P8.4

```
library(cfcdae)
```

```
## Warning: package 'cfcdae' was built under R version 4.0.2
```

```
## Registered S3 method overwritten by 'DoE.base':
##   method          from
##   factorize.factor conf.design
```

```
data("PineOleoresin")
head(PineOleoresin)
```
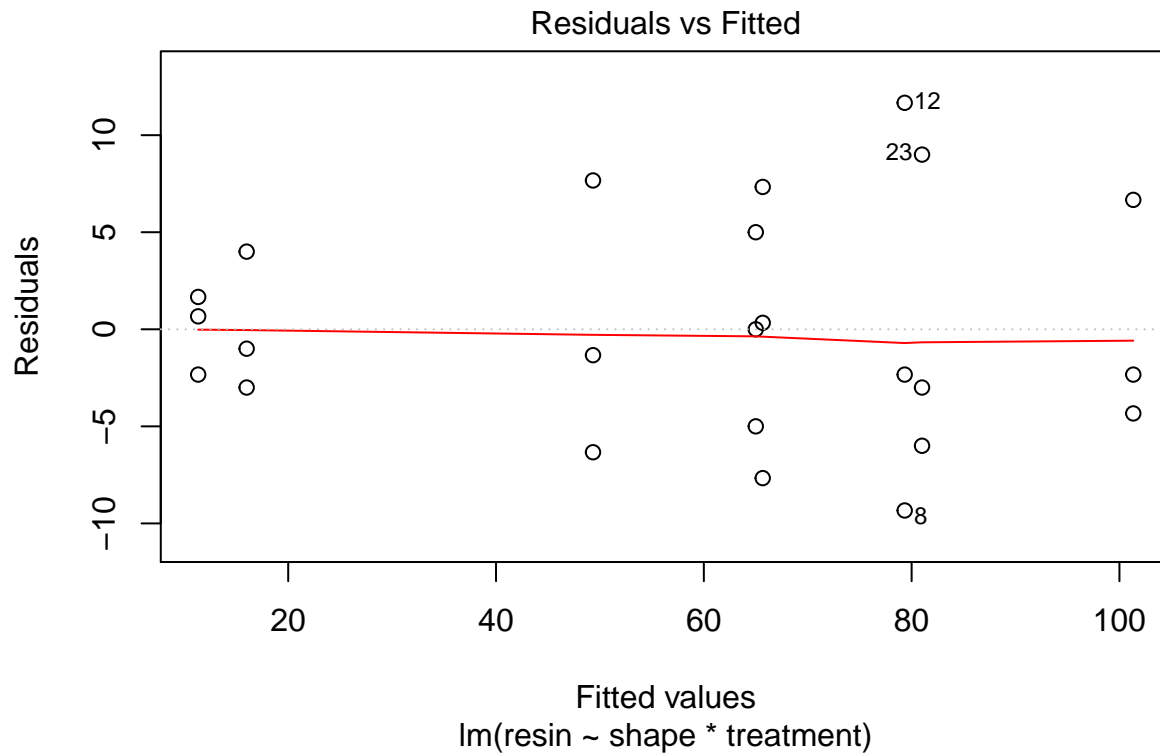
```
##          shape treatment resin
## 1     circular   control     9
## 2     diagonal   control    43
## 3        check   control    60
## 4  rectangular   control    77
## 5     circular   control    13
## 6     diagonal   control    48
```

```
pinemod_1<-lm(resin~shape*treatment,data=PineOleoresin)
##check assumptions:
```

```r
plot(pinemod_1,which=1)
```

Residuals vs Fitted


Fitted values
lm(resin ~ shape * treatment)

```r
##The residuals vs fitted plot shows we have non-constant variance.It does not meet the assumpitons.
car::boxCox(pinemod_1)
```

```
## Registered S3 methods overwritten by 'car':
##   method                             from
##   influence.merMod                   lme4
##   cooks.distance.influence.merMod    lme4
##   dfbeta.influence.merMod            lme4
##   dfbetas.influence.merMod           lme4
```
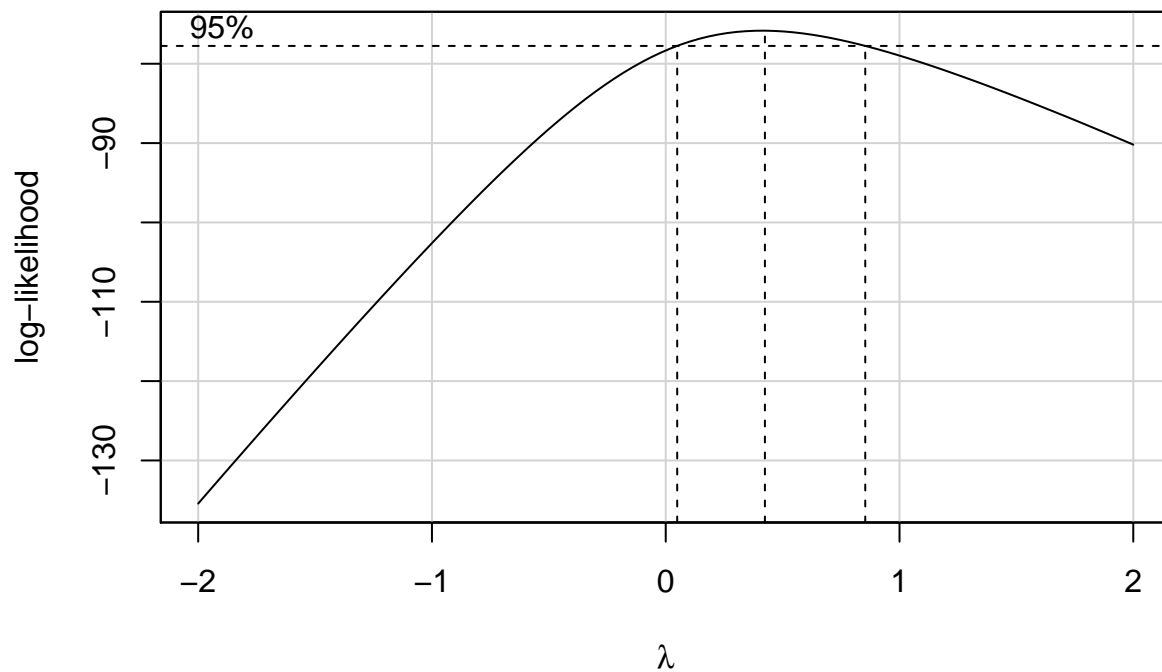
```
##box-Cox suggests a square root
pinemod_2<-lm(sqrt(resin)~shape*treatment,data=PineOleoresin)
##check assumptions:
plot(pinemod_2,which=1)
```

Residuals vs Fitted

lm(sqrt(resin) ~ shape * treatment)
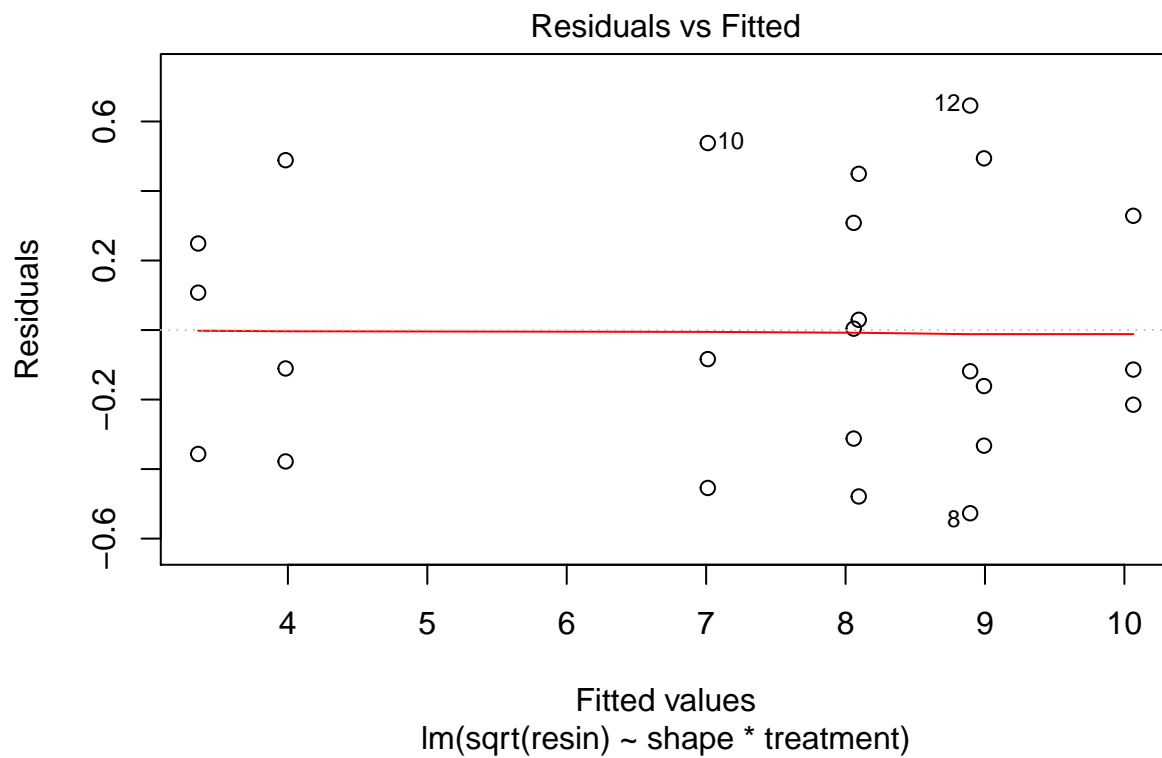
```
##The assumptions are good.
anova(pinemod_2)

## Analysis of Variance Table
```

```
## 
## Response: sqrt(resin)
##                 Df  Sum Sq Mean Sq  F value     Pr(>F)
## shape            3 116.935  38.978 209.3714 4.786e-13 ***
## treatment        1   5.456   5.456  29.3088 5.739e-05 ***
## shape:treatment  3   0.256   0.085   0.4581    0.7154
## Residuals       16   2.979   0.186
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(pinemod_2)
```

```
## 
## Call:
## lm(formula = sqrt(resin) ~ shape * treatment, data = PineOleoresin)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5271 -0.3174 -0.0971  0.3134  0.6457
## 
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       7.306887   0.088074  82.963  < 2e-16 ***
## shape1           -3.636833   0.152548 -23.841 6.28e-14 ***
## shape2            0.246329   0.152548   1.615    0.126
## shape3            1.218725   0.152548   7.989 5.64e-07 ***
## treatment1       -0.476811   0.088074  -5.414 5.74e-05 ***
## shape1:treatment1 0.163308   0.152548   1.071    0.300
## shape2:treatment1 -0.064579  0.152548  -0.423    0.678
## shape3:treatment1 0.009474   0.152548   0.062    0.951
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.4315 on 16 degrees of freedom
## Multiple R-squared:  0.9763, Adjusted R-squared:  0.9659
## F-statistic: 94.11 on 7 and 16 DF,  p-value: 8.412e-12
```

```
with(PineOleoresin,interactplot(shape,treatment,resin))
```
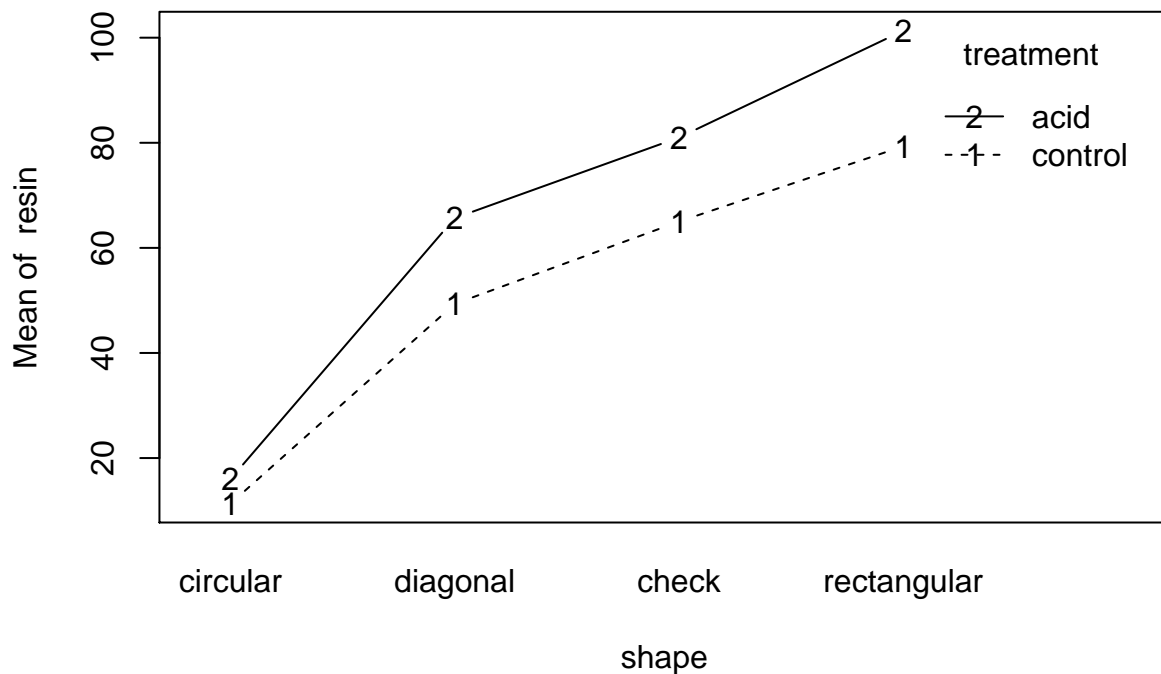
After transforming,form the anova table, we can see that the main effects shape and treatment both have significant difference but there is no evidence showing interaction is significant.

By the interaction plot, we can see that the mean of resin when treatment is acid is aparently higher than control treatment.And when shape is rectangular and treatment is acid, it has the largerst value if mean of resin.

P8.8

```r
library(cfcdae)
data("PopcornRatios")
head(PopcornRatios)
```

```
##      pType pAmt   oType oAmt ratio
## 1 generic  1/8  canola    1  24.5
## 2 generic  1/8 popcorn    1  28.0
## 3 generic  1/8  canola    2  21.5
## 4 generic  1/8 popcorn    2  22.5
## 5 generic  1/4  canola    1  21.5
## 6 generic  1/4 popcorn    1  23.0
```

```r
attach(PopcornRatios)
pAmt<-as.factor(pAmt);oAmt<-as.factor(oAmt)
pop_1<-lm(ratio~pType*pAmt*oType*pAmt)
##check assumptions:
plot(pop_1,which=1)
```

Residuals vs Fitted

Fitted values
lm(ratio ~ pType * pAmt * oType * pAmt)
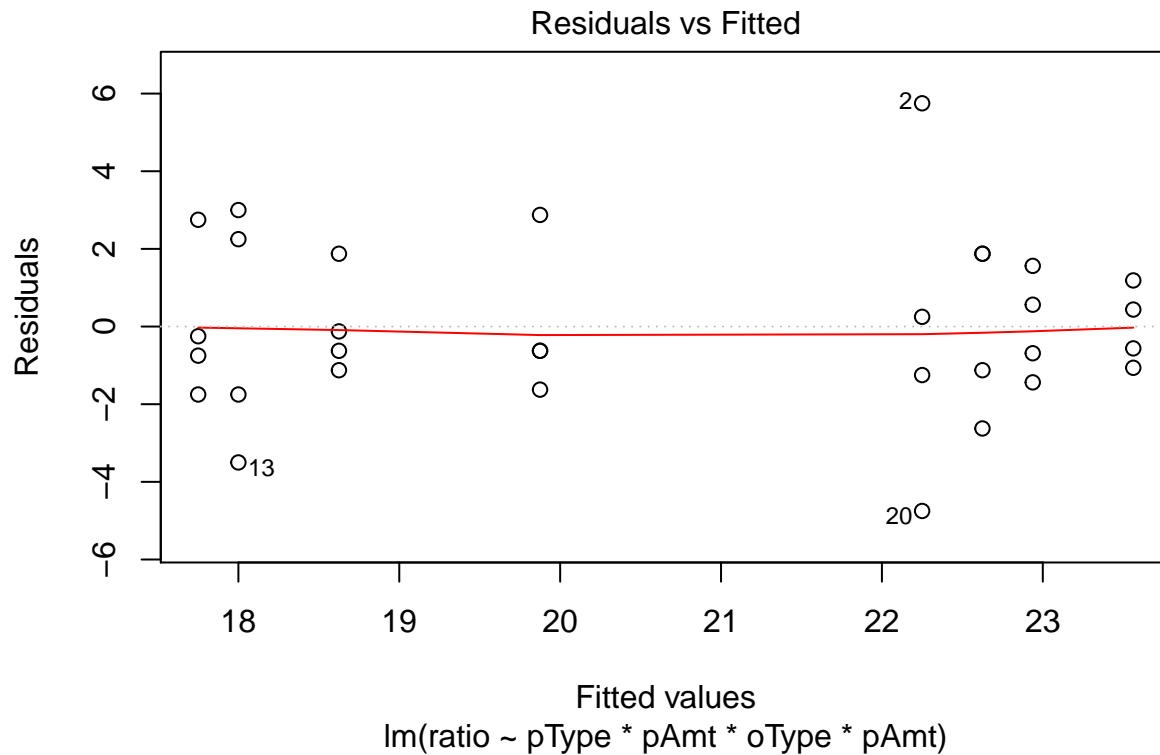
```
##The assumptions look good.
anova(pop_1)

## Analysis of Variance Table
##
## Response: ratio
##                 Df  Sum Sq Mean Sq F value    Pr(>F)
## pType            1 146.633 146.633 25.4495 3.711e-05 ***
## pAmt             1   4.883   4.883  0.8475    0.3664
## oType            1   0.781   0.781  0.1356    0.7159
## pType:pAmt       1   0.008   0.008  0.0014    0.9709
## pType:oType      1   0.281   0.281  0.0488    0.8270
## pAmt:oType       1   7.031   7.031  1.2203    0.2802
## pType:pAmt:oType 1   1.531   1.531  0.2658    0.6109
## Residuals       24 138.281   5.762
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##form the anova table, we can see that there is significant difference between pType but there is no e
summary(pop_1)

##
## Call:
## lm(formula = ratio ~ pType * pAmt * oType * pAmt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7500 -1.1562 -0.5938  1.6406  5.7500
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)              20.70313    0.42433  48.790  < 2e-16 ***
## pType1                     2.14062    0.42433   5.045 3.71e-05 ***
## pAmt1                      0.39062    0.42433   0.921    0.366
## oType1                    -0.15625    0.42433  -0.368    0.716
## pType1:pAmt1               0.01563    0.42433   0.037    0.971
## pType1:oType1              0.09375    0.42433   0.221    0.827
## pAmt1:oType1              -0.46875    0.42433  -1.105    0.280
## pType1:pAmt1:oType1        0.21875    0.42433   0.516    0.611
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.4 on 24 degrees of freedom
## Multiple R-squared:  0.5382, Adjusted R-squared:  0.4035
## F-statistic: 3.996 on 7 and 24 DF,  p-value: 0.004966
```

P8.2

The residuals vs fitted plot shows we have non-constant variance before transforming data. After transforming, the residuals vs fitted plot shows log-transformed data meet our assumptions.From the anova table of log-transformed data,we can see that the main effects r and c both have significant difference because their p-values are so small but there is no enough evidence showing interaction is significant because the p-value 0.5224 is larger than 0.05.

P8.7

```
library(cfcdae)
data("UnpoppedKernels")
head(UnpoppedKernels)
```

```
##   brand watts.z watts unpopped
## 1     P    1000  1000       19
## 2     P    1000  1000       20
## 3     P    1000  1000       22
## 4     A    1000  1000       40
## 5     A    1000  1000       33
## 6     A    1000  1000       35
```

```
attach(UnpoppedKernels)
watts<-as.factor(watts)
upop_1<-lm(unpopped~brand*watts)
##check assumptions:
plot(upop_1,which=1)
```
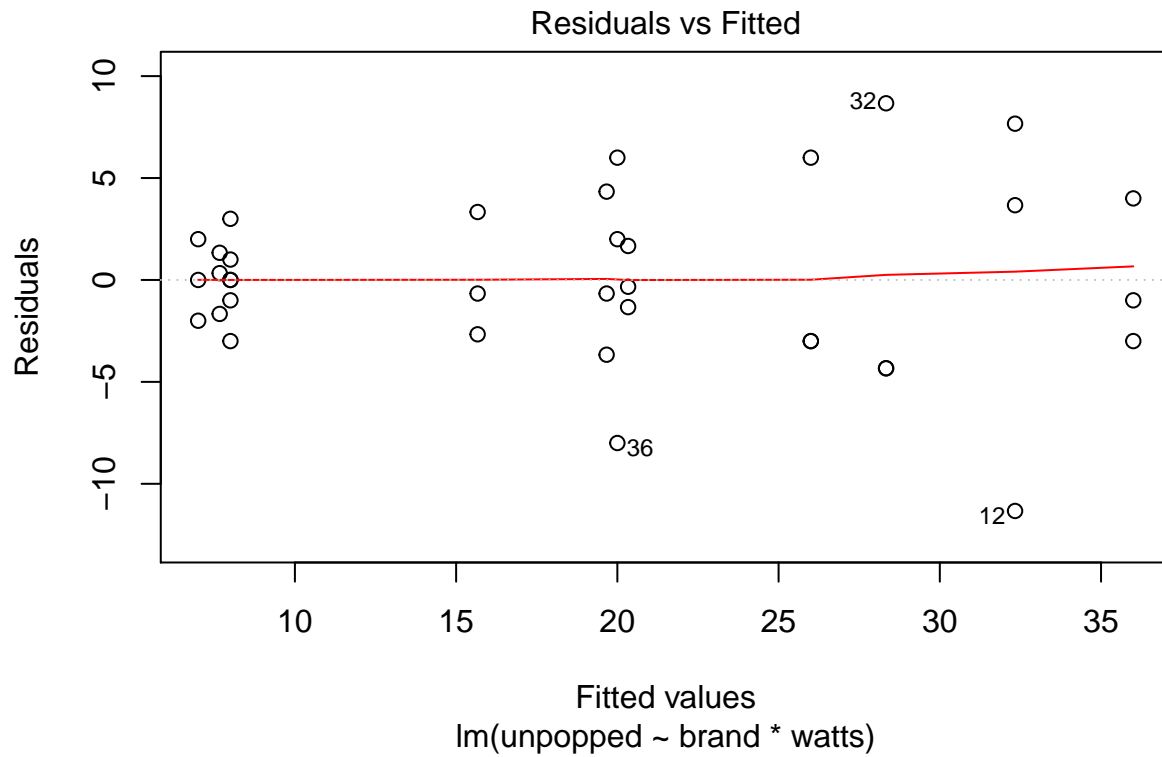
Residuals vs Fitted

Fitted values
lm(unpopped ~ brand * watts)

```
##The residuals vs fitted plot shows we have non-constant variance.It does not meet the assumpitons.
car::boxCox(upop_1)
```
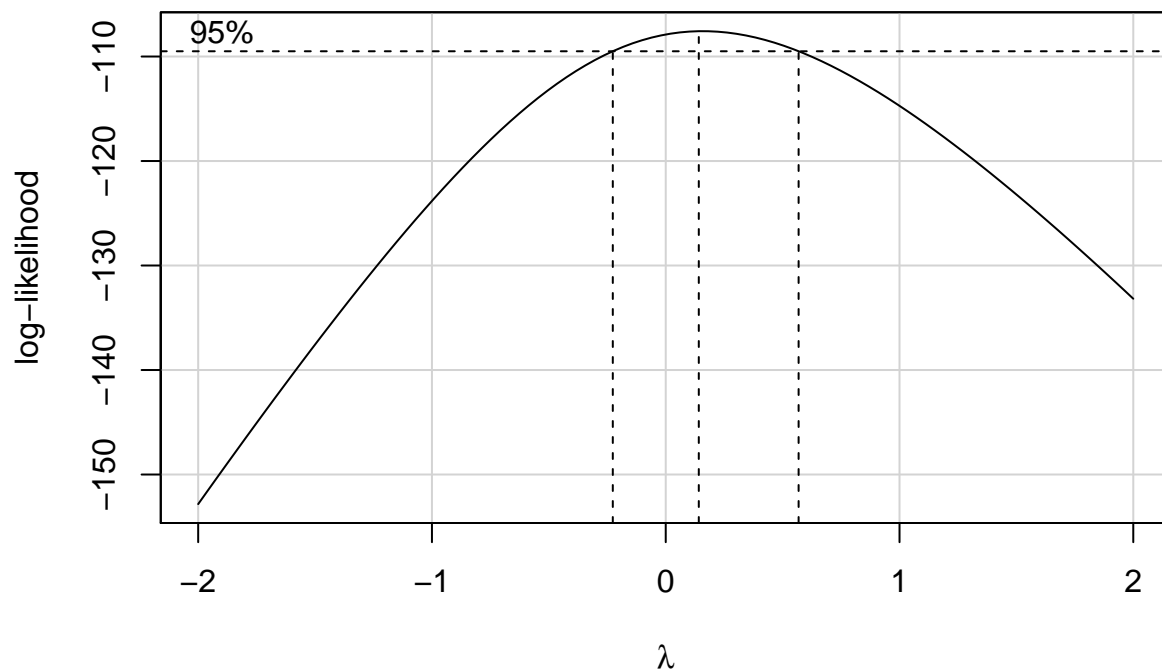


```
##  take lamta=0
upop_2<-lm(log(unpopped)~brand*watts)
##check assumptions:
plot(upop_2,which=1)
```

Residuals vs Fitted

Residuals

Fitted values
lm(log(unpopped) ~ brand * watts)
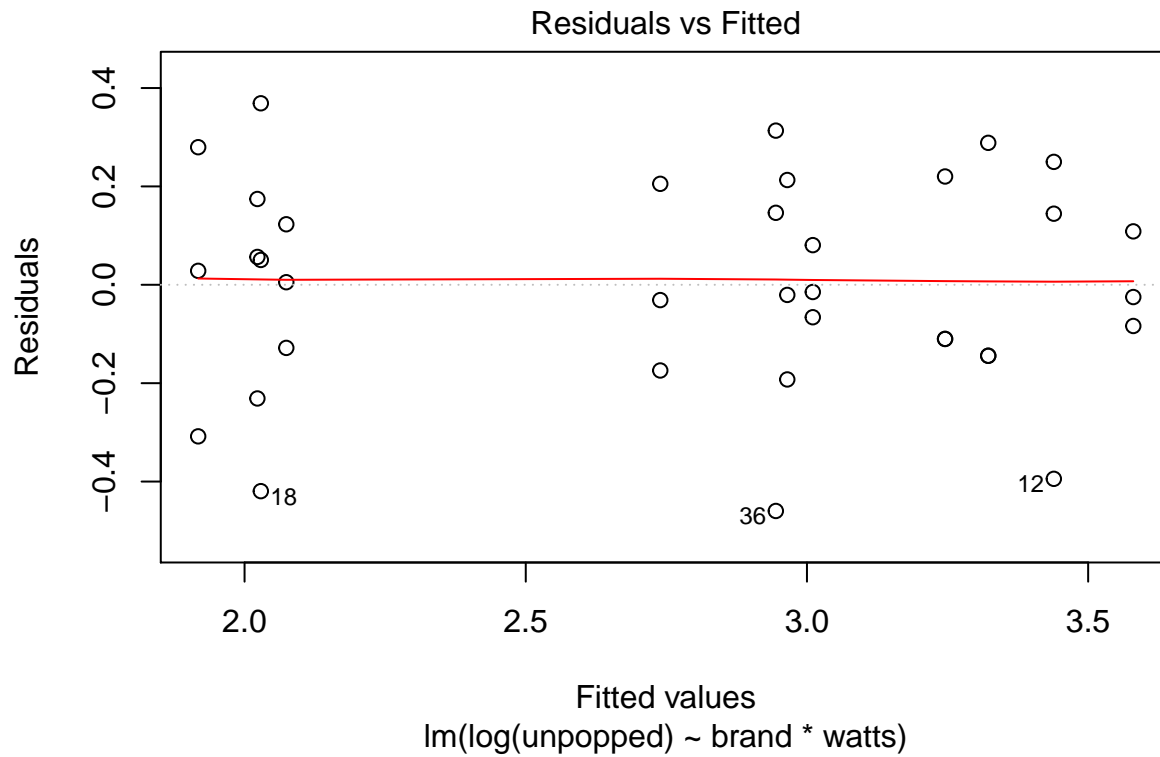
```
##The assumptions look good.
anova(upop_2)

## Analysis of Variance Table
##
## Response: log(unpopped)
##             Df  Sum Sq Mean Sq F value      Pr(>F)
## brand        3  0.5529  0.1843  2.8325     0.05965 .
## watts        2 10.6940  5.3470 82.1749 1.832e-11 ***
## brand:watts  6  0.9952  0.1659  2.5491     0.04713 *
## Residuals   24  1.5616  0.0651
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
with(UnpoppedKernels,interactplot(brand,watts,unpopped))
```
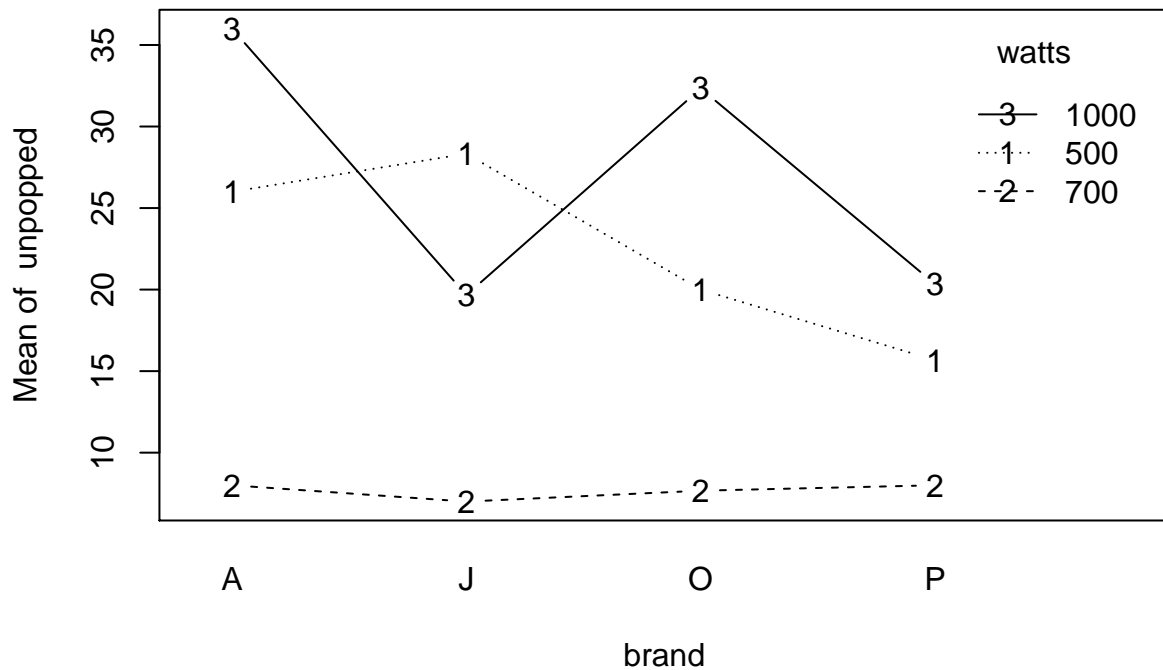
From the anova tabel, we see that the effect of watts have significant difference and the interaction is also significant but there is no evidence showing the effect of brand is significant.

P8.11

```r
library(cfcdae)
data("SewageBacteria")
head(SewageBacteria)
```

```
##   medium origin antibiotic  count
## 1     LB sludge       amox 760000
## 2     LB sludge       amox 440000
## 3     LB sludge       amox 330000
## 4     LB sludge      tetra  17000
## 5     LB sludge      tetra  11000
## 6     LB sludge      tetra  21000
```

```r
attach(SewageBacteria)
se_1<-lm(count~medium*origin*antibiotic)
##check assumptions:
plot(se_1,which=1)
```
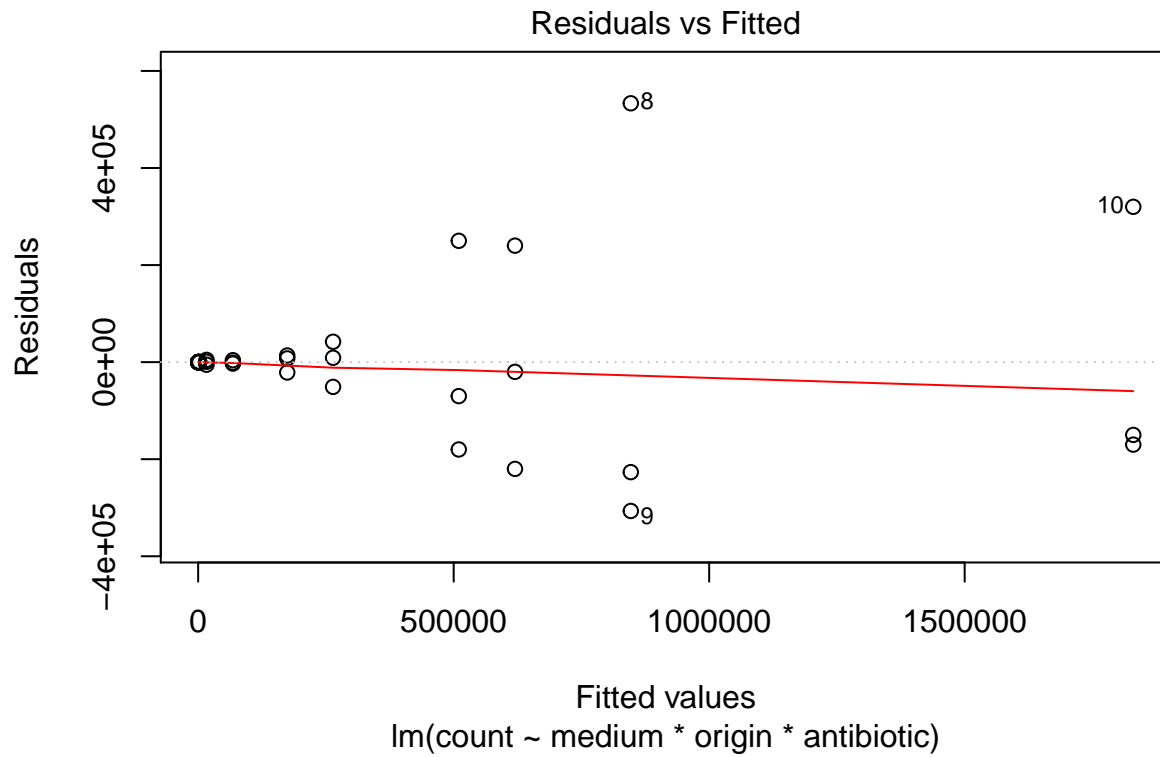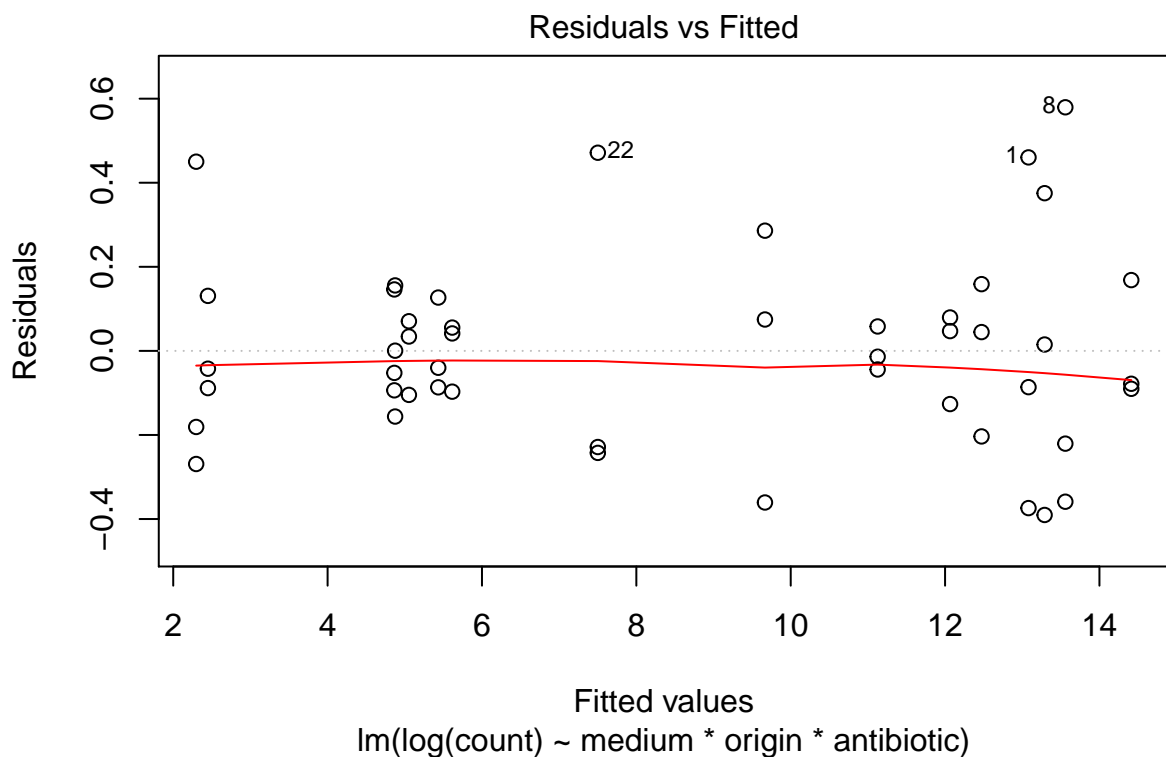
Residuals vs Fitted

Fitted values
lm(count ~ medium * origin * antibiotic)

```
##The residuals vs fitted plot shows we have non-constant variance.It does not meet the assumpitons.
car::boxCox(se_1)
```
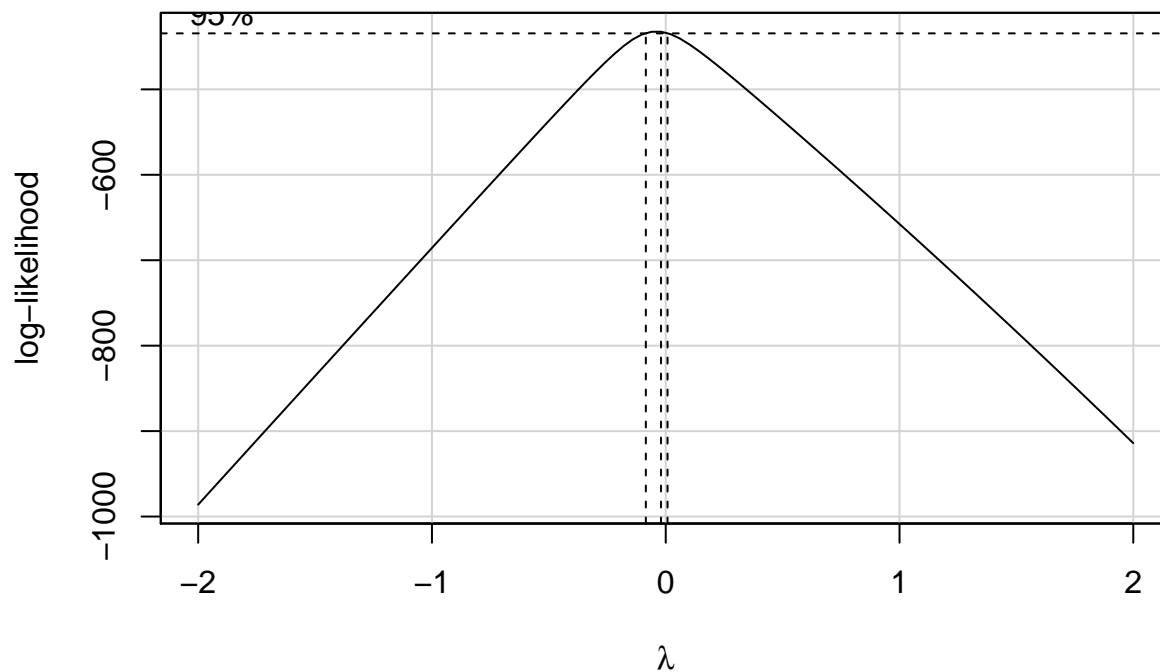


```
se_2<-lm(log(count)~medium*origin*antibiotic)
##check assumptions:
plot(se_2,which=1)
```

Residuals vs Fitted

Fitted values
lm(log(count) ~ medium * origin * antibiotic)

```
##The assumptions look good.
anova(se_2)

## Analysis of Variance Table
##
## Response: log(count)
##                         Df Sum Sq Mean Sq   F value    Pr(>F)
## medium                   1   3.38    3.38   46.4460 1.049e-07 ***
## origin                   1 711.07  711.07 9764.4800 < 2.2e-16 ***
## antibiotic               3  88.43   29.48  404.7898 < 2.2e-16 ***
## medium:origin            1   0.10    0.10    1.3270 0.2578711
## medium:antibiotic        3  11.15    3.72   51.0480 2.680e-12 ***
## origin:antibiotic        3   3.30    1.10   15.1159 2.658e-06 ***
## medium:origin:antibiotic 3   1.76    0.59    8.0787 0.0003801 ***
## Residuals               32   2.33    0.07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(se_2)

##
## Call:
## lm(formula = log(count) ~ medium * origin * antibiotic)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.39031 -0.11014 -0.02703  0.09133  0.57946
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             8.60997    0.03895 221.050  < 2e-16 ***
```

```
## medium1                       -0.26545    0.03895  -6.815 1.05e-07 ***
## origin1                        -3.84890    0.03895 -98.815  < 2e-16 ***
## antibiotic1                     0.15590    0.06746   2.311 0.027436 *
## antibiotic2                     1.39019    0.06746  20.606  < 2e-16 ***
## antibiotic3                    -2.22517    0.06746 -32.983  < 2e-16 ***
## medium1:origin1                -0.04487    0.03895  -1.152 0.257871
## medium1:antibiotic1            -0.03581    0.06746  -0.531 0.599244
## medium1:antibiotic2            -0.69121    0.06746 -10.246 1.25e-11 ***
## medium1:antibiotic3             0.66849    0.06746   9.909 2.84e-11 ***
## origin1:antibiotic1             0.04200    0.06746   0.623 0.538002
## origin1:antibiotic2             0.40651    0.06746   6.026 1.01e-06 ***
## origin1:antibiotic3            -0.16262    0.06746  -2.410 0.021851 *
## medium1:origin1:antibiotic1  0.25166    0.06746   3.730 0.000742 ***
## medium1:origin1:antibiotic2  0.05821    0.06746   0.863 0.394681
## medium1:origin1:antibiotic3 -0.28211    0.06746  -4.182 0.000210 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2699 on 32 degrees of freedom
## Multiple R-squared:  0.9972, Adjusted R-squared:  0.9958
## F-statistic:    750 on 15 and 32 DF,  p-value: < 2.2e-16
```

(a) From the anova table, we can see that except medium:origin,all effects is significant. From the summary table, we can see that the items,medium1:origin1 ,medium1:antibiotic1,origin1:antibiotic1 and medium1:origin1:antibiotic2 do not show significant.

(b) The values of count vary widely.