STATHW5

Mingming Xu

10/16/2020

E9.3

```
Variables<-c("c","r","c.r","ERROR")
DF<-c(3,3,9,14)
SS<-c(116.25,0.0255,0.48787,0.8223)
MS<-c(38.479,0.0085,0.054207,0.058736)
F_Value<-c(38.479/0.058736,0.0085/0.058736,0.054207/0.058736, NA)
P_Value<-c(1-pf(38.479/0.058736,3,14),1-pf(0.0085/0.058736,3,14),1-pf(0.054207/0.058736,9,14),NA)
library(knitr)
```

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

```
kable(cbind(Variables,DF,SS,MS,F_Value,P_Value))
```

Variables	DF	SS	MS	F_Value	P_Value
c	3	116.25	38.479	655.11781530918	2.77555756156289e-15
r	3	0.0255	0.0085	0.144715336420594	0.931363023225667
c.r	9	0.48787	0.054207	0.922892263688368	0.533949854387978
ERROR	14	0.8223	0.058736	NA	NA

```
##Only c has the statistical sinificant diffrence.
```

E9.4

(a)$MS_{AB} = SS_{AB}/df_{AB}$

(b)$MS_{ABC} = SS_{ABC}/df_{ABC}$

(c)$MS_E = SS_E/df_E$

Since df_{AB}, df_{ABC} and df_E are same whatever Type II or Type III ANOVAs, I will use SS to distinguish them.

  (a) Type II:$SS_{AB} = SS(AB|A, B, C, AC, BC)$.

Type III:$SS_{AB} = SS(AB|A, B, C, AC, BC, ABC)$.

Not same.

  (b) Type II:$SS_{ABC} = SS(AB|A, B, C, AC, BC, AB)$.

TypeIII:$SS_{ABC} = SS(AB|A, B, C, AC, BC, AB)$

Same.

  (c) SS_E is the same whatever Type II or Type III ANOVAs because it just the error that the model does not explain.

MS_{ABC} and MS_E will be the same in the two tables.

P9.1

```
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
```
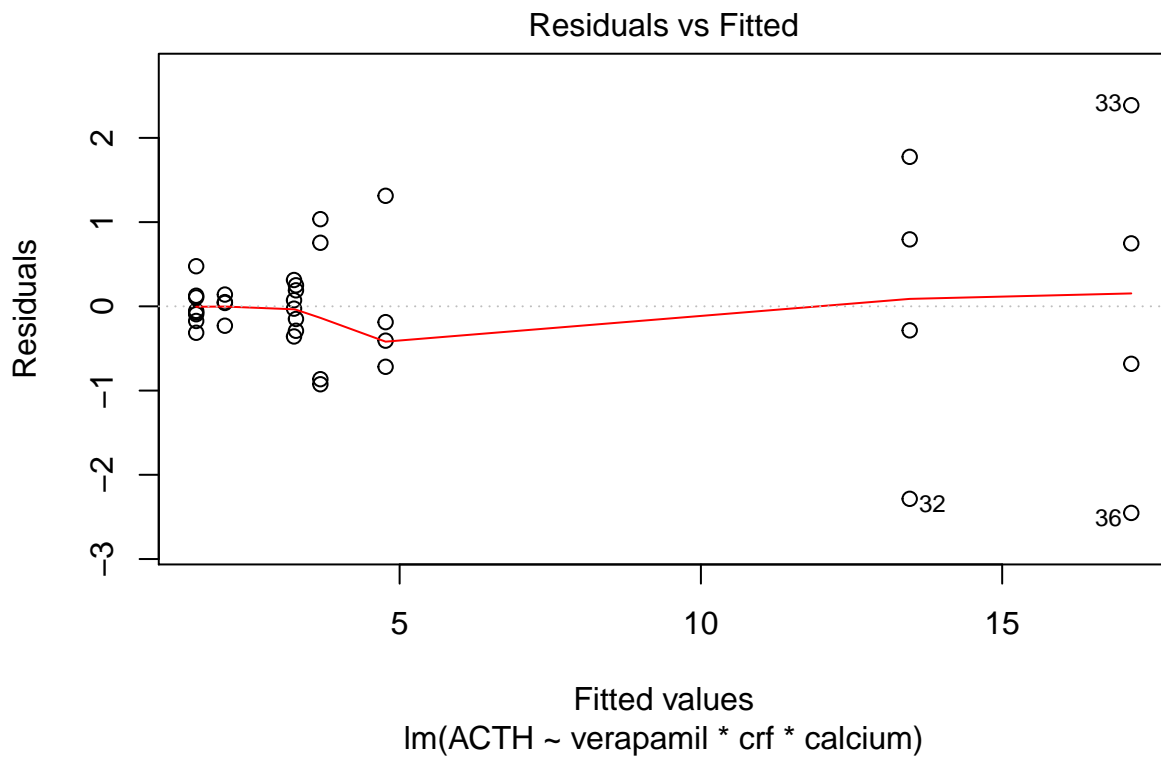
```
data("Verapamil")
attach(Verapamil)

mod1<-lm(ACTH~verapamil*crf*calcium)
##check assumptions:
plot(mod1,which=1)
```


Residuals vs Fitted

Fitted values
lm(ACTH ~ verapamil * crf * calcium)

```
##The residuals vs fitted plot looks like we do not have constant variance.It does not meet the assumpt
car::boxCox(mod1)
```

```
## Registered S3 methods overwritten by 'car':
##    method                            from
##    influence.merMod                  lme4
##    cooks.distance.influence.merMod   lme4
##    dfbeta.influence.merMod           lme4
##    dfbetas.influence.merMod          lme4
```
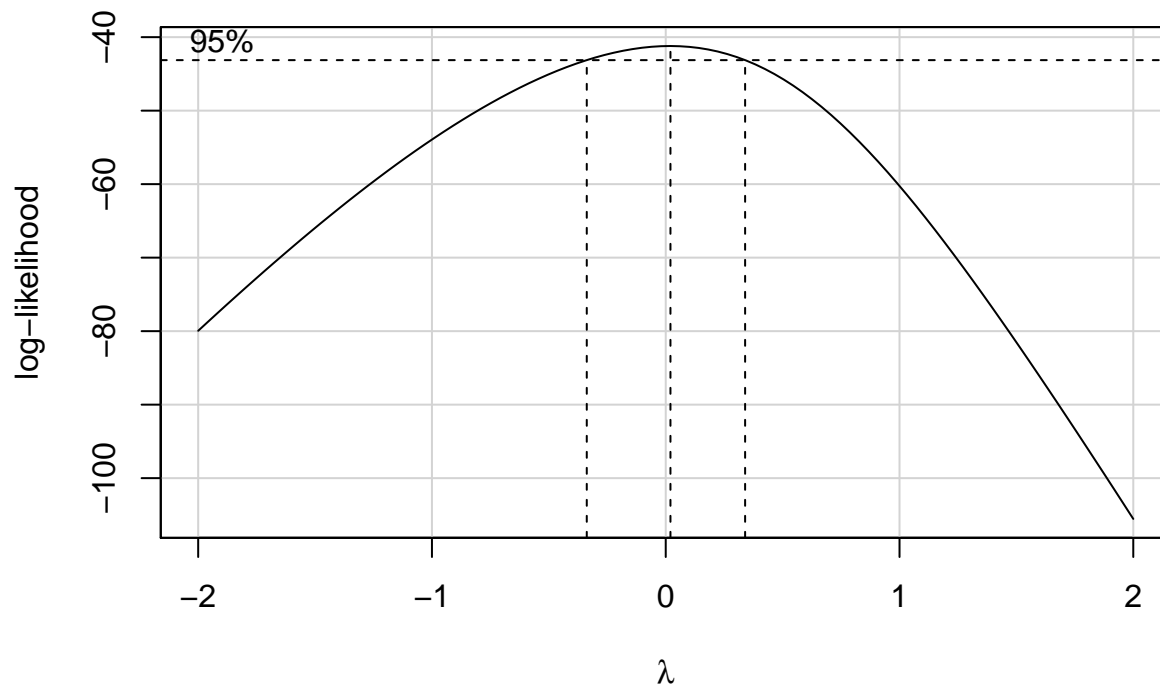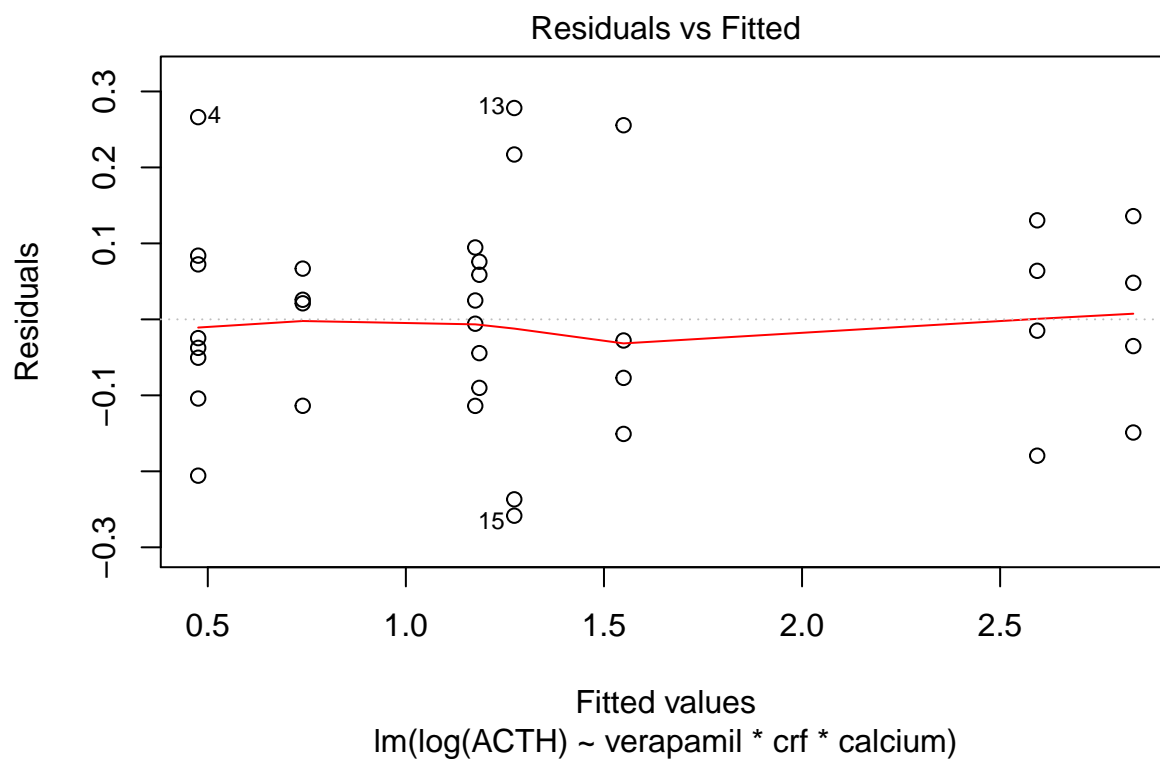
```
##choose lamta=0
mod1_1<-lm(log(ACTH)~verapamil*crf*calcium)
plot(mod1_1,which=1)####The assumptions are good
```

Residuals vs Fitted

Fitted values
lm(log(ACTH) ~ verapamil * crf * calcium)

```
##
anova(mod1_1)##the resulting by using Type I anova

## Analysis of Variance Table
##
```

```
## Response: log(ACTH)
##                       Df  Sum Sq Mean Sq  F value     Pr(>F)
## verapamil             1  1.2451  1.2451   54.3852 4.964e-08 ***
## crf                   1 13.2636 13.2636  579.3476 < 2.2e-16 ***
## calcium               1  7.2809  7.2809  318.0252 < 2.2e-16 ***
## verapamil:crf         1  0.0608  0.0608    2.6554    0.1144
## verapamil:calcium     1  0.0262  0.0262    1.1431    0.2941
## crf:calcium           1  1.1191  1.1191   48.8832 1.329e-07 ***
## verapamil:crf:calcium 1  0.0310  0.0310    1.3524    0.2547
## Residuals            28  0.6410  0.0229
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
car::Anova(mod1_1,type=2)##since this is unblanced data so I will try Type II anova.
```

```
## Anova Table (Type II tests)
##
## Response: log(ACTH)
##                    Sum Sq Df  F value     Pr(>F)
## verapamil          0.3413  1  14.9086  0.000609 ***
## crf               11.2756  1 492.5136 < 2.2e-16 ***
## calcium            7.3413  1 320.6616 < 2.2e-16 ***
## verapamil:crf      0.0331  1   1.4477  0.238972
## verapamil:calcium  0.0561  1   2.4518  0.128625
## crf:calcium        1.1191  1  48.8832 1.329e-07 ***
## verapamil:crf:calcium 0.0310 1 1.3524  0.254674
## Residuals          0.6410 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##And we can see the two anova table results are diffrent.
##From the Anova Table (Type II tests),the p-values show that the predictors verapamil,crf,
##calcium and crf:calcium have significant diffrence
mod1_2<-lm(log(ACTH)~verapamil+crf+calcium+crf:calcium)
anova(mod1_2)
```

```
## Analysis of Variance Table
##
## Response: log(ACTH)
##             Df  Sum Sq Mean Sq F value     Pr(>F)
## verapamil    1  1.2451  1.2451  51.087 4.931e-08 ***
## crf          1 13.2636 13.2636 544.213 < 2.2e-16 ***
## calcium      1  7.2809  7.2809 298.739 < 2.2e-16 ***
## crf:calcium  1  1.1226  1.1226  46.059 1.338e-07 ***
## Residuals   31  0.7555  0.0244
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

P9.3

```
library(cfcdae)
data("IceCream")
attach(IceCream)
boxplot(rating~gum);boxplot(rating~protein);boxplot(rating~gum:protein)
```
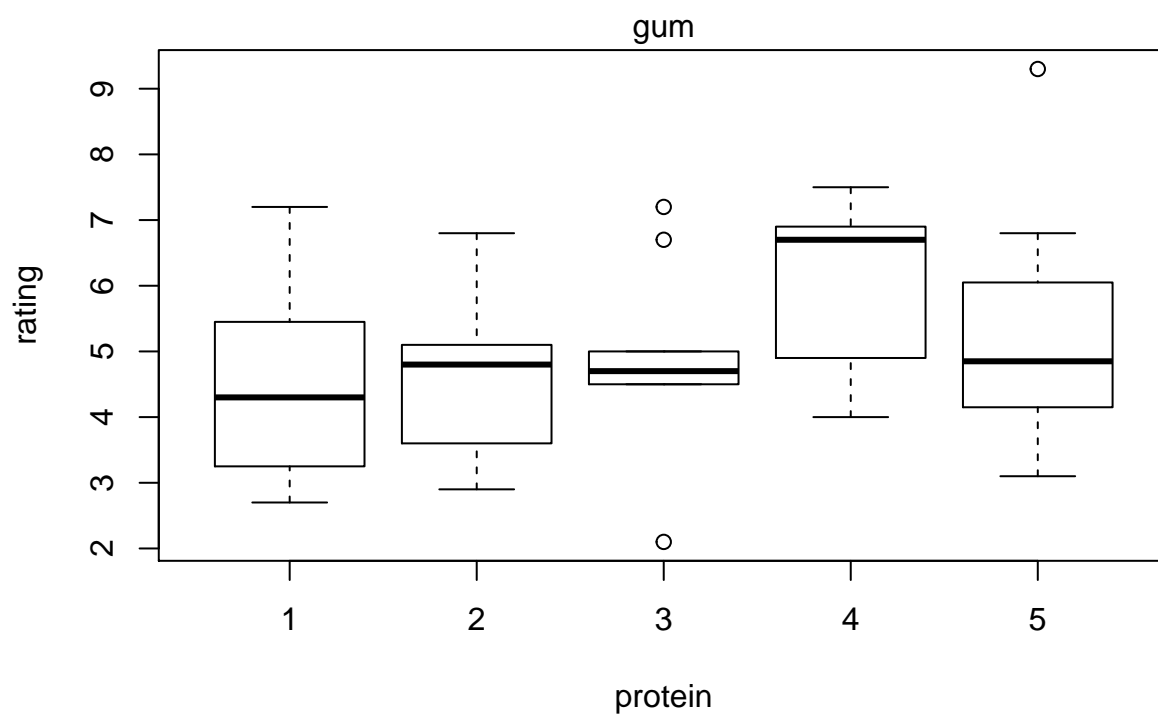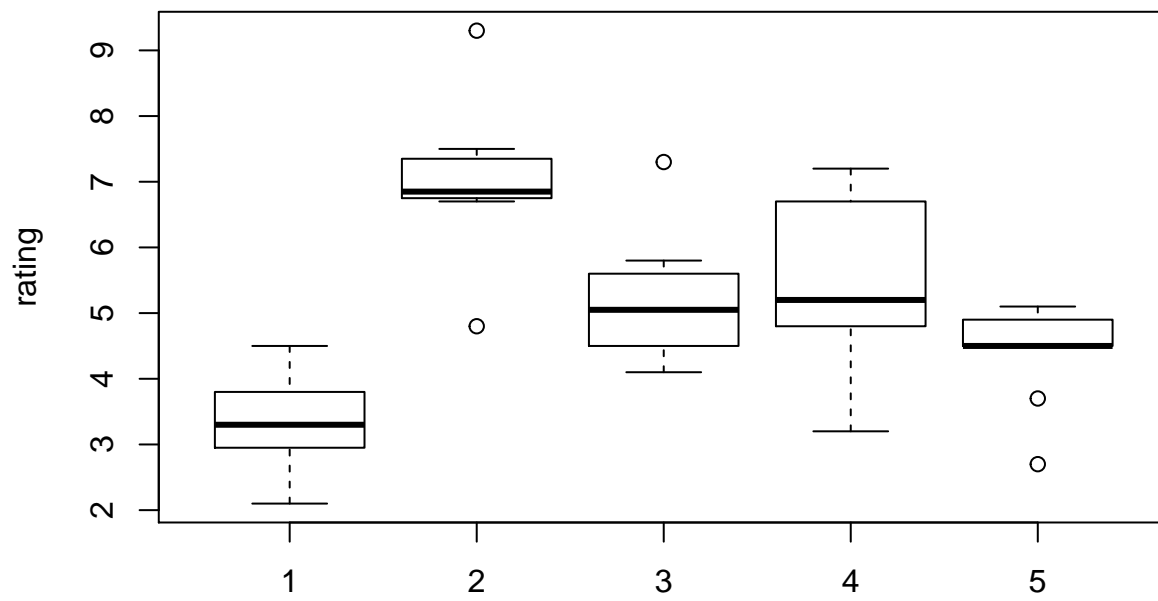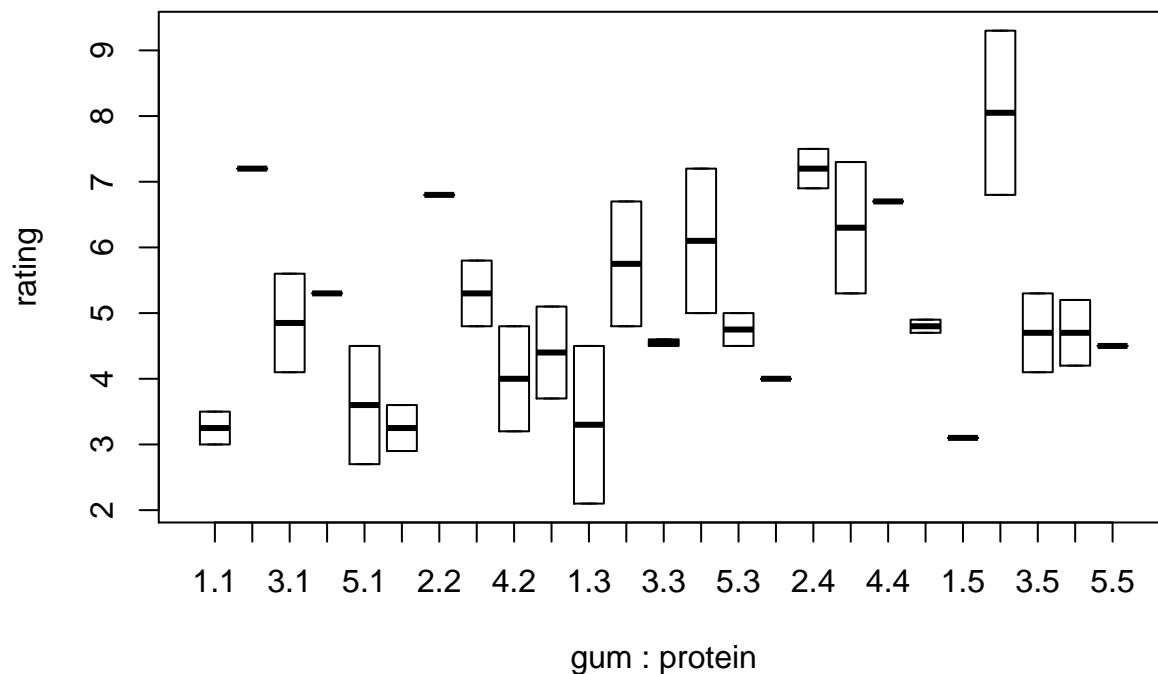
gum : protein

```
##The plots show that there is significance in gum but it is difficult to see if there is any
##diffrence in protein and gum:protein.
mod2<-lm(rating~gum*protein)
##check assumptions:
plot(mod2,which=1)
```

Residuals vs Fitted

Fitted values
lm(rating ~ gum * protein)
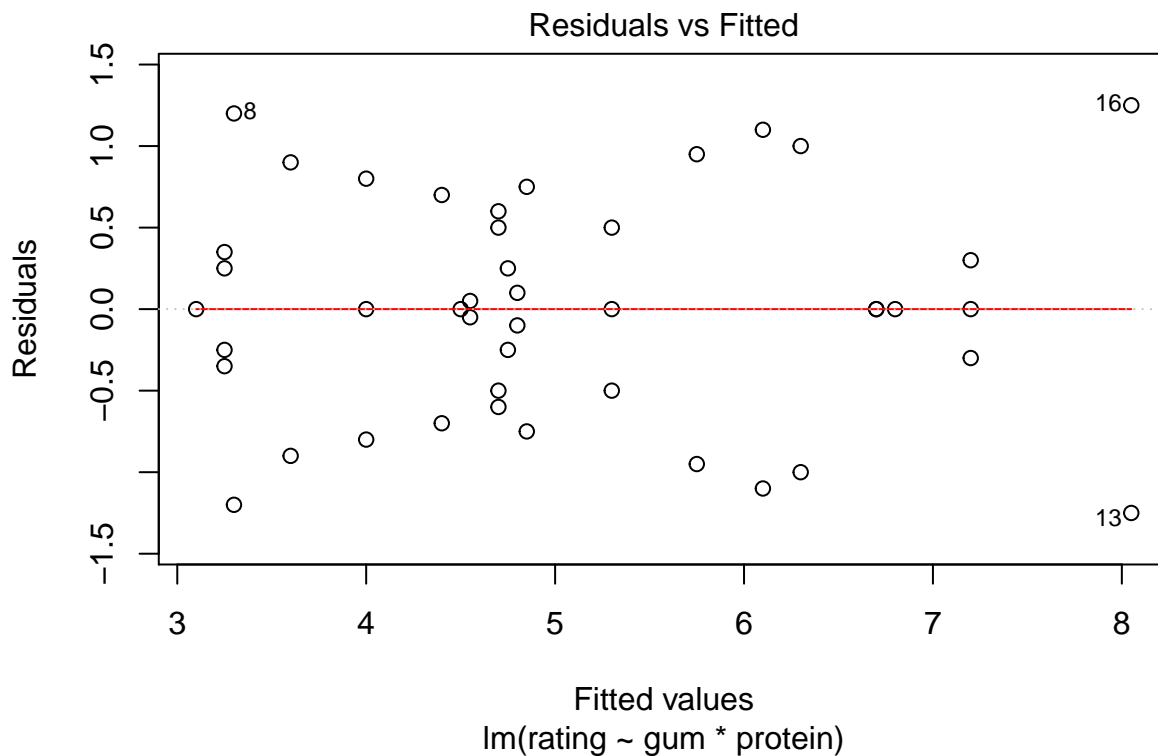
```
##The residuals vs fitted plot looks like we do have constant variance. They meet our assumptions.
##Becausethis is unblanced data ,I will try Type II anova test to analyze it
```

```r
car::Anova(mod2,type=2)
```

```
## Anova Table (Type II tests)
##
## Response: rating
##             Sum Sq Df F value    Pr(>F)
## gum         51.775  4 12.5123 3.741e-05 ***
## protein      7.075  4  1.7099    0.1894
## gum:protein 14.053 16  0.8490    0.6261
## Residuals   19.655 19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
##From the Anova Table (Type II tests), we can see that only gum has significant diffrence.
##So,only gum has effect on the sensory rating.
mod2_1<-lm(rating~gum)
model.effects(mod2_1,"gum")
```

```
##          1          2          3          4          5
## -1.71133333  1.95116667  0.09116667  0.31783333 -0.64883333
```

```r
##From this model, when gum is type 2,the fitted sensory rating is the highest.
##And when gum is type 1,the fitted sensory rating is the lowest.
```

P9.6

```r
library(cfcdae)
data("AirCells")
attach(AirCells)
var<-as.factor(var);loc<-as.factor(loc);temp<-as.factor(temp)
##The plots show that there is significance in gum but it is difficult to see if there is any diffrence
mod<-lm(area~var*loc*temp,data=AirCells)
##check assumptions:
plot(mod,which=1)
```
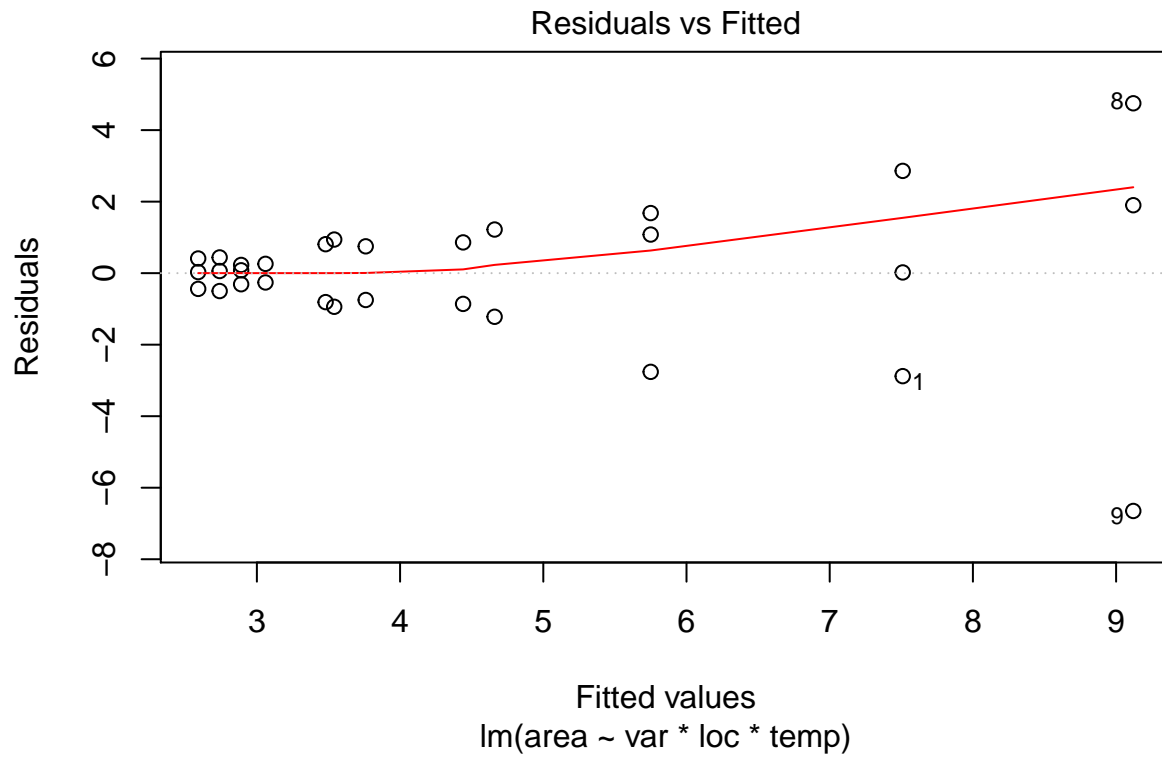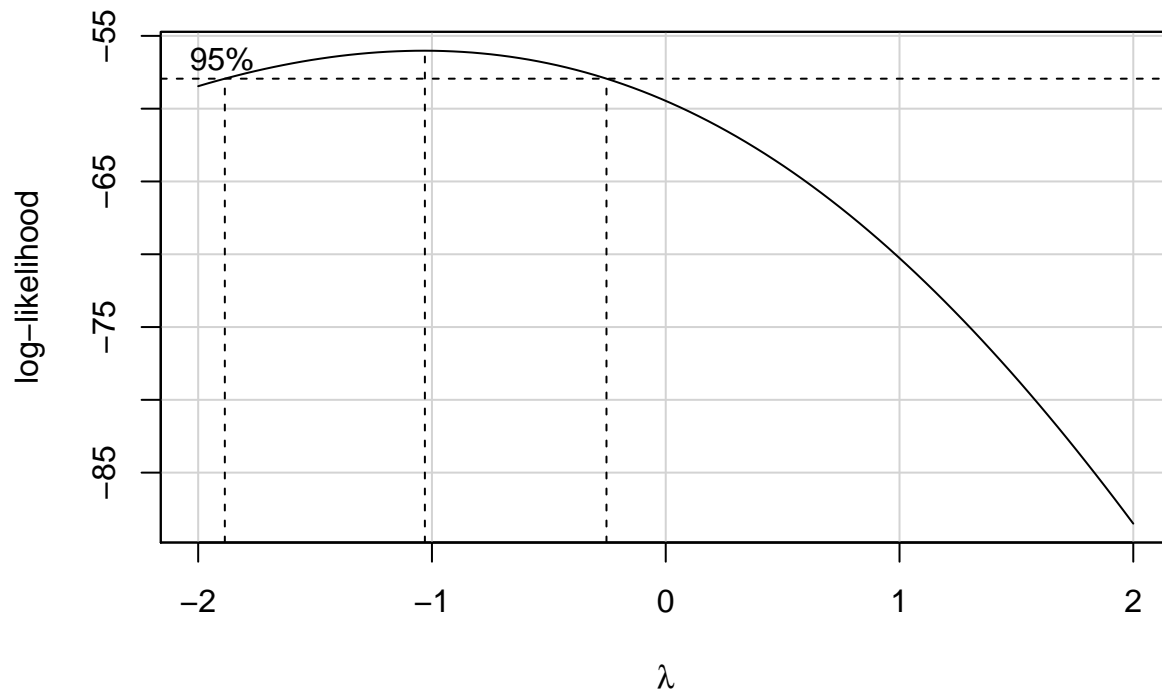
Residuals vs Fitted
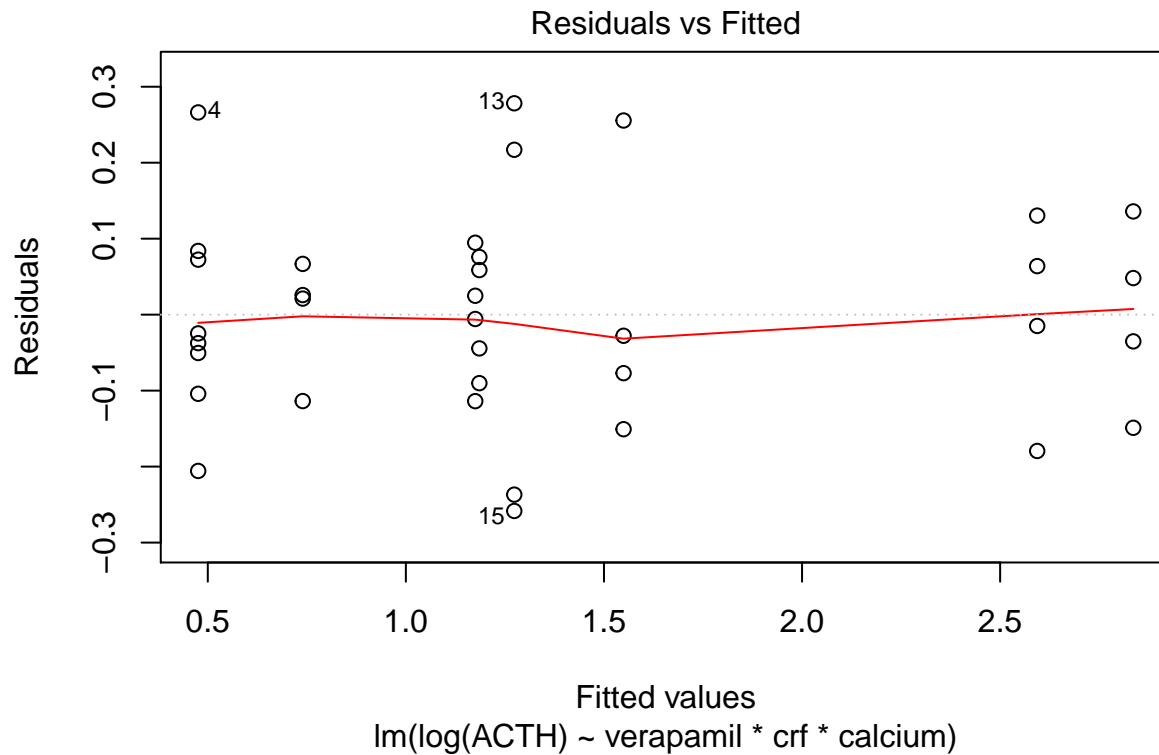

lm(area ~ var * loc * temp)

```
##The residuals vs fitted plot looks like we do not have constant variance.
##It does not meet the assumpitons.
car::boxCox(mod)
```



```
##choose lamta=-1
mod_1<-lm(log((area)^(-1))~var*loc*temp,data=AirCells)
plot(mod1_1,which=1)##The assumptions are good
```

Residuals vs Fitted

Fitted values
lm(log(ACTH) ~ verapamil * crf * calcium)

```
##
##Because this is unblanced data ,I will try Type II anova test to analyze it
car::Anova(mod_1,type=2)
```

```
## Anova Table (Type II tests)
##
## Response: log((area)^(-1))
##               Sum Sq Df F value  Pr(>F)
## var           0.1292  2  0.3515 0.70837
## loc           1.3081  1  7.1194 0.01567 *
## temp          1.4333  1  7.8011 0.01201 *
## var:loc       0.1555  2  0.4232 0.66128
## var:temp      0.1469  2  0.3997 0.67635
## loc:temp      0.1194  1  0.6501 0.43060
## var:loc:temp  0.0311  2  0.0846 0.91924
## Residuals     3.3072 18
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##From the Anova Table (Type II tests), we can see that only loc and temp have significant diffrence.
##Our model does not has enough evidence to support that variety has any effect on the size of air cells

mod_2<-lm(log((area)^(-1))~loc+temp,data=AirCells)
car::Anova(mod_2,type=2)
```

```
## Anova Table (Type II tests)
##
## Response: log((area)^(-1))
##       Sum Sq Df F value   Pr(>F)
## loc   1.3081  1  9.1613 0.005380 **
## temp  1.4333  1 10.0385 0.003788 **
```

```
## Residuals 3.8552 27
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model.effects(mod_2,"loc");model.effects(mod_2,"temp")
```
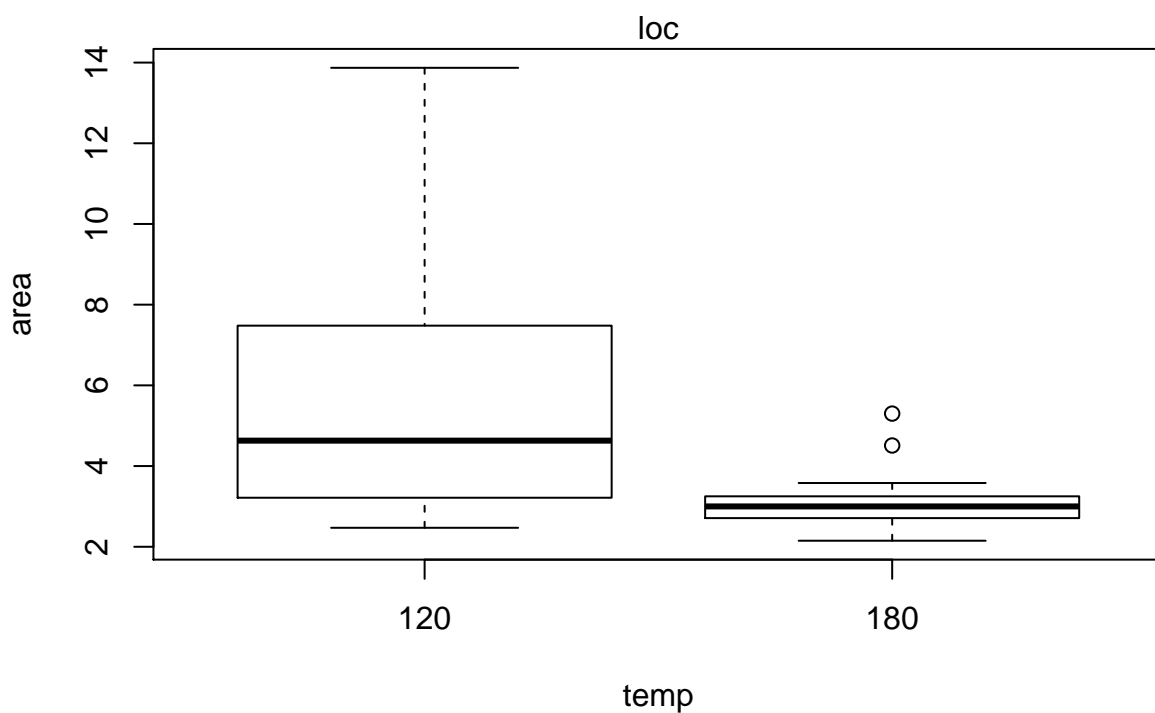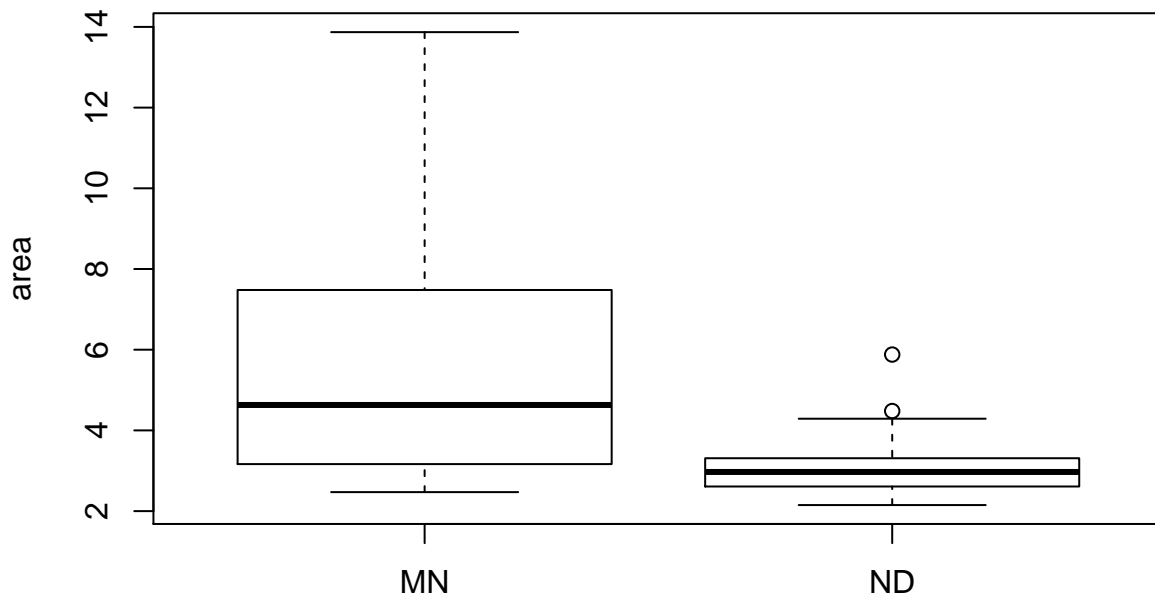
```
##         MN          ND
## -0.2131189  0.2131189

##        120         180
## -0.2230886  0.2230886
```

```
boxplot(area~loc);boxplot(area~temp)
```

```
##From the model results and plots,
##we can see that the size is larger when the growth location is MN than ND.
##Also,the size is larger when temperature of the extrusion is 180 C than 120 C.
```

P9.10

```
library(cfcdae)
data("PlasmaLeucine")
attach(PlasmaLeucine)
percent<-as.factor(percent)
fit<-lm(leucine~source*percent,data=PlasmaLeucine)
##check assumptions:
plot(fit,which=1)
```

Residuals vs Fitted

Fitted values
lm(leucine ~ source * percent)
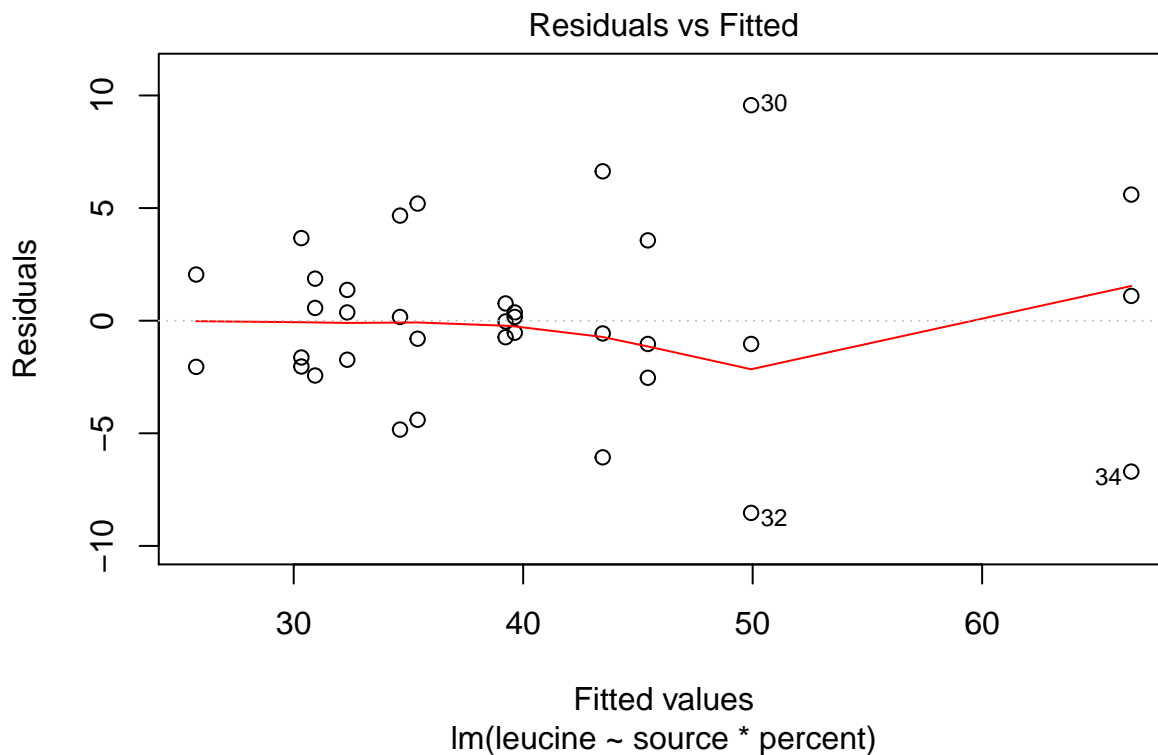
```
##examine it:
car::boxCox(fit)
```
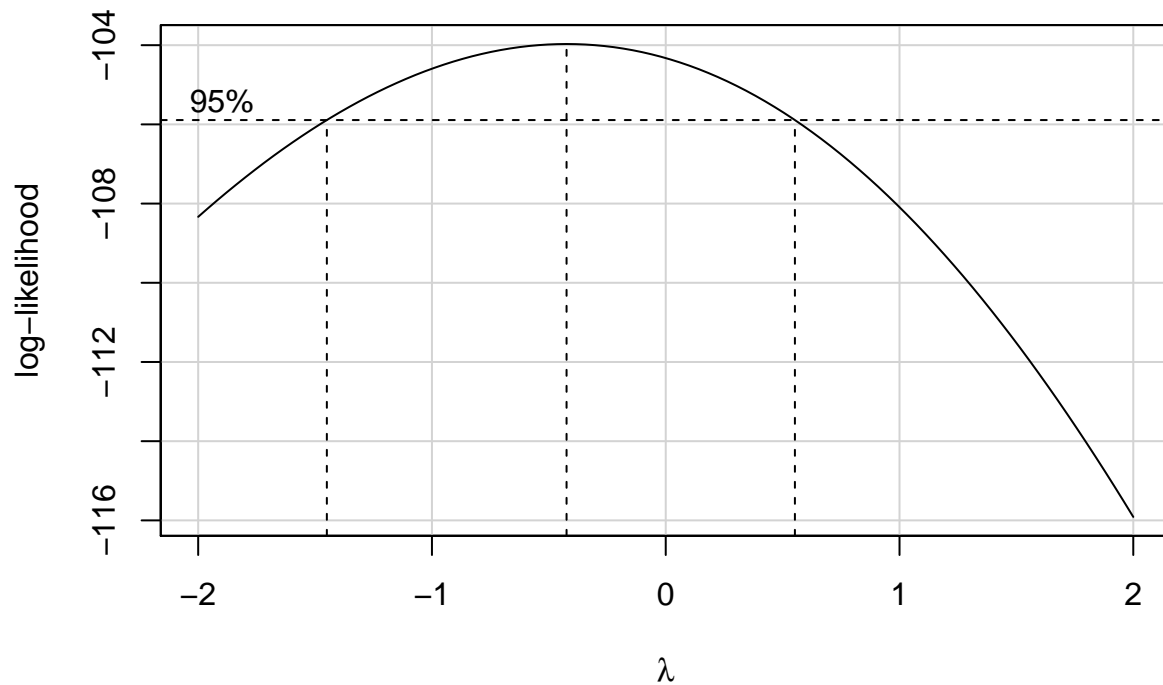
```
##choose lamta=-0.5
fit_1<-lm((leucine^(-0.5))~source*percent,data=PlasmaLeucine)
plot(fit_1,which=1)##The assumptions are good
```

Residuals vs Fitted

Fitted values
lm((leucine^(−0.5)) ~ source * percent)
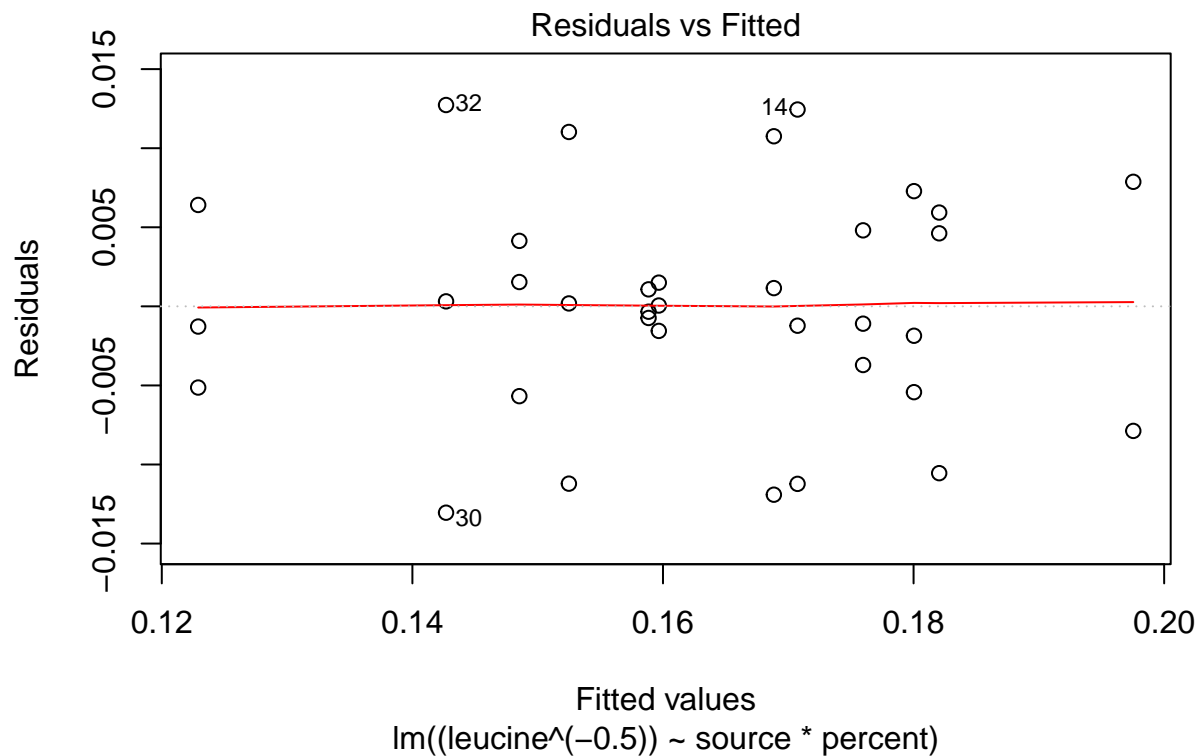
```
##Because this is unblanced data ,I will try Type II anova test to analyze it
car::Anova(fit_1,type=2)
```

```
## Anova Table (Type II tests)
##
```

```
## Response: (leucine^(-0.5))
##                 Sum Sq Df F value    Pr(>F)
## source        0.0081823  2  56.548 1.320e-09 ***
## percent       0.0038701  3  17.831 3.346e-06 ***
## source:percent 0.0007892  6   1.818    0.1398
## Residuals     0.0016640 23
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##From the Anova Table (Type II tests),
##we can see that source and percent have significant diffrence
##but the p-value does not show the interaction term source:percent has any effect on the level of leuc
fit_2<-lm((leucine^(-0.5))~source+percent,data=PlasmaLeucine)
car::Anova(fit_2,type=2)
```

```
## Anova Table (Type II tests)
##
## Response: (leucine^(-0.5))
##              Sum Sq Df F value    Pr(>F)
## source    0.0081823  2  48.362 5.796e-10 ***
## percent   0.0038701  3  15.250 3.834e-06 ***
## Residuals 0.0024532 29
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
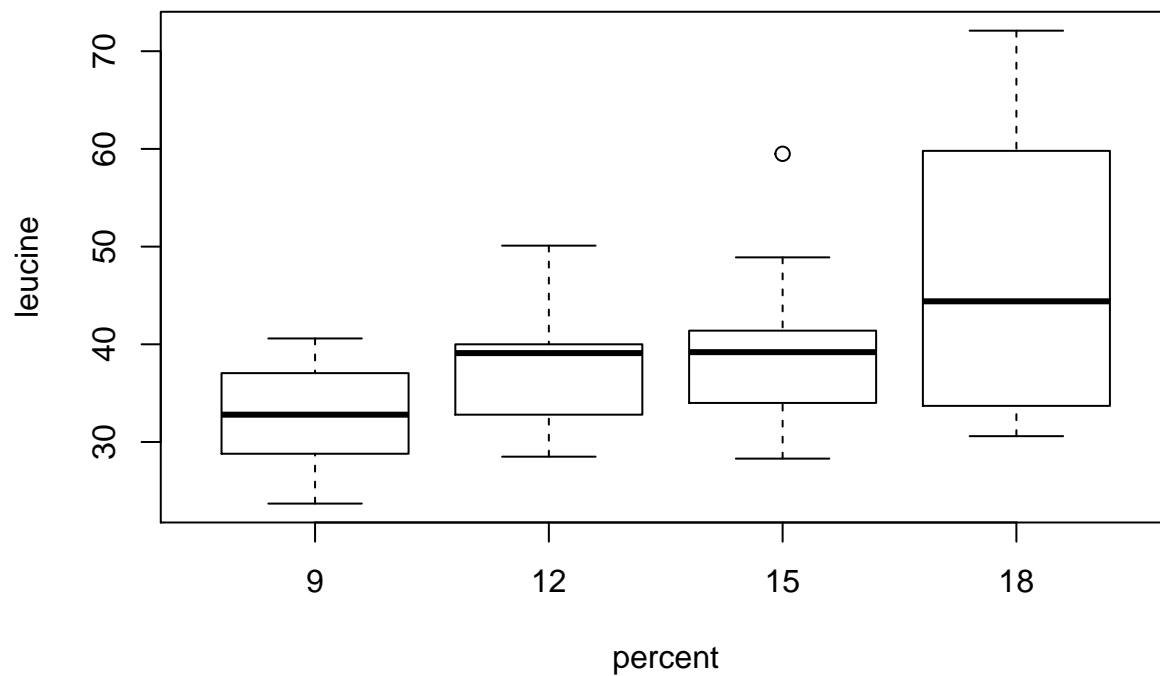
```
##
model.effects(fit_2,"source");model.effects(fit_2,"percent")
```

```
##        fish         soy        skim
##   0.020673624 -0.003979565 -0.016694060
```

```
##            9          12          15          18
##   0.0158872699  0.0003639949 -0.0019606419 -0.0142906229
```

```
boxplot(leucine~source);boxplot(leucine~percent)
```
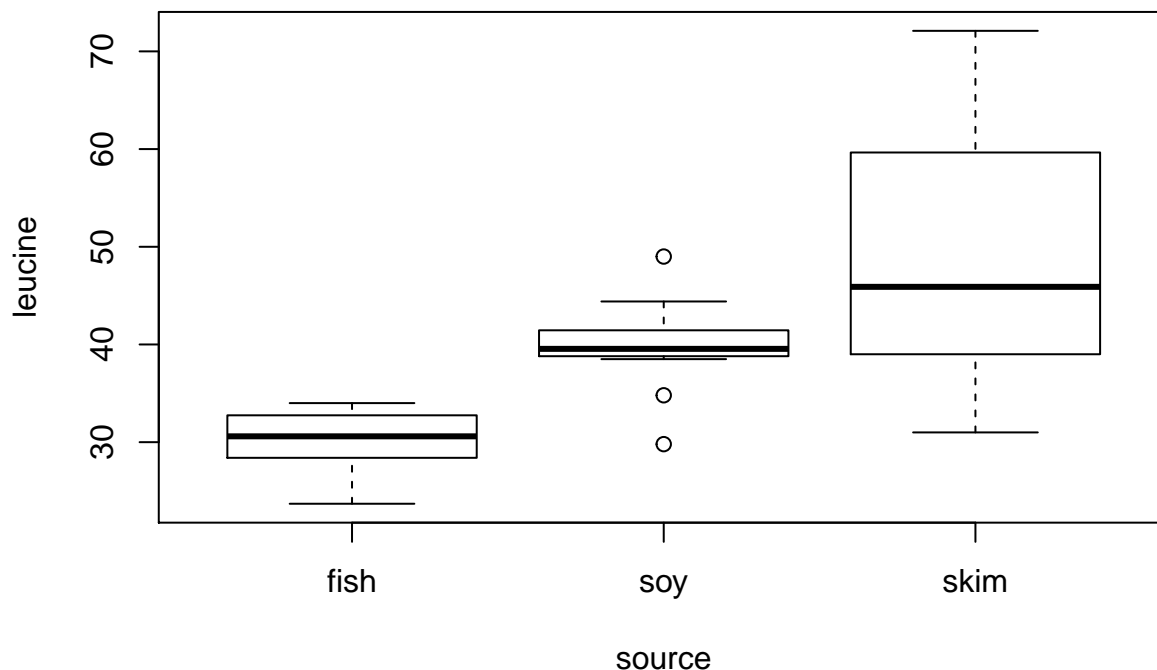
```
##The results show the higher protein concentration in the diet,the lower the level of leucine
##The results also show that fish meal can raise the level of leucine
##but soy and skim meal will lower the level of leucine.
```