

# stat4051hw5

*Mingming Xu*

*2019/11/3*

Short Answers:

- a. Fixed effect: If the researchers are interested in comparing the students reading performance of these 6 specific schools they chose for the study, we have a fixed effect.

Random effect: Instead the researchers are actually interested in a studying the variation in reading scores of students, the results of which would be applicable to all schools in Minnesota. In this case, because there are many schools in Minnesota, they may have randomly chosen a sample of 6 many schools in Minnesota and test them to characterize the variability in reading scores of students in Minnesota.

b.

Fixed effect: If the researchers are interested in comparing the effectiveness of these specific therapist they chose for the study, we have a fixed effect.

Random effect: Instead the researchers are actually interested in a studying the variation in change in score on depression test after one year of therapy for patients, the results of which would be applicable to all therapist. In this case, because there are many therapist, they may have randomly chosen a sample of some therapist and test them to characterize the variability in effectiveness of all therapist.

c.

Correlated: The data  $Y_{ij}$  and  $Y_{i'j'}$  are correlated if  $i = i'$ .

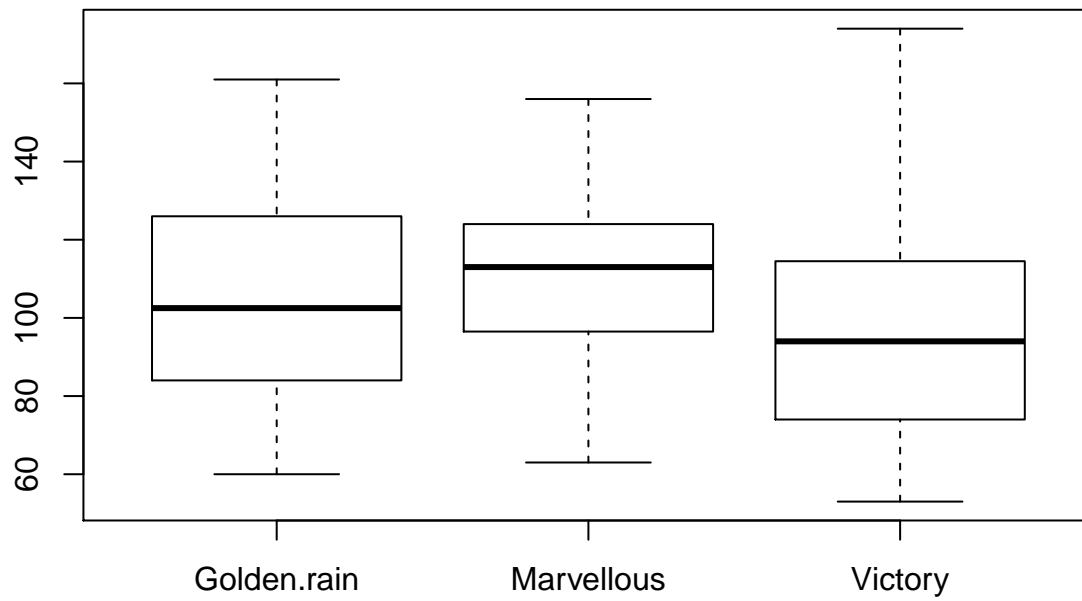
Independent: The data  $Y_{ij}$  and  $Y_{i'j'}$  are independent if  $i \neq i'$ .

Oat Problem

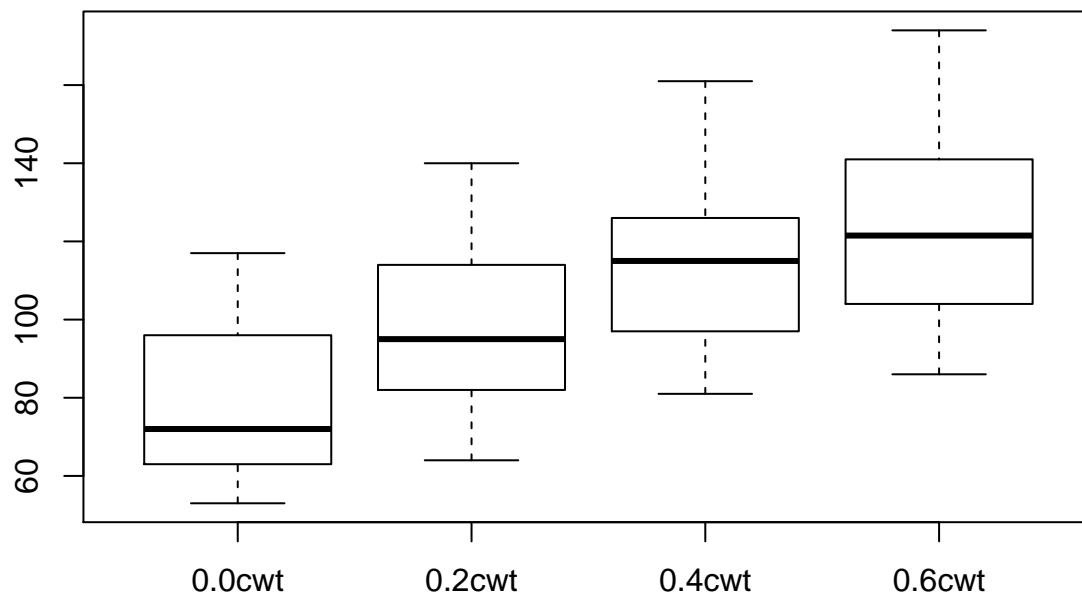
```
library(MASS)
#Read in the oats dataset
data(oats, package="MASS")
attach(oats)
```

a.

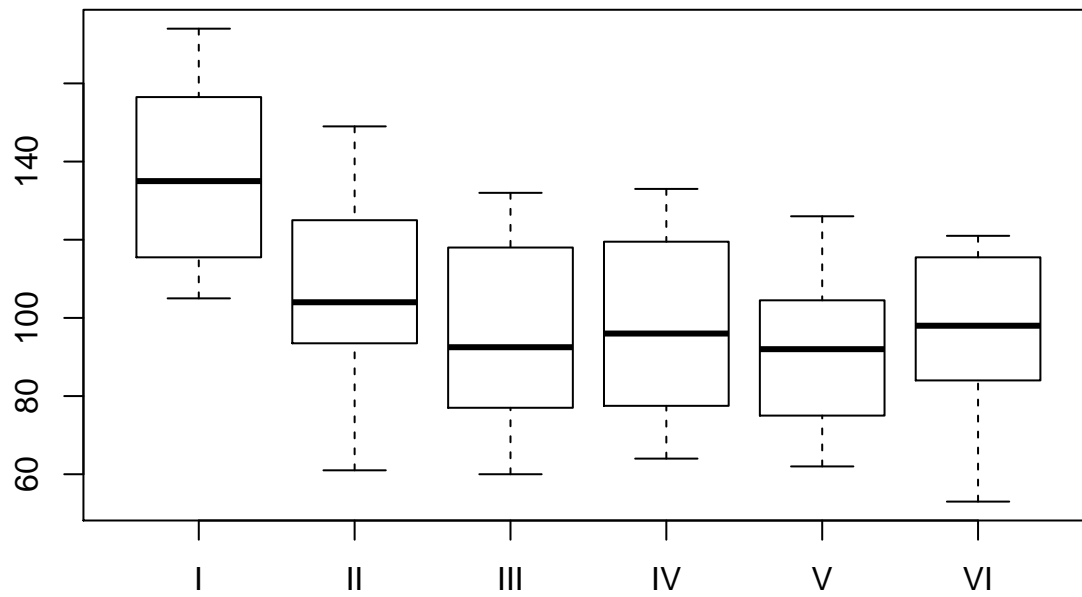
```
boxplot(Y~V)
```



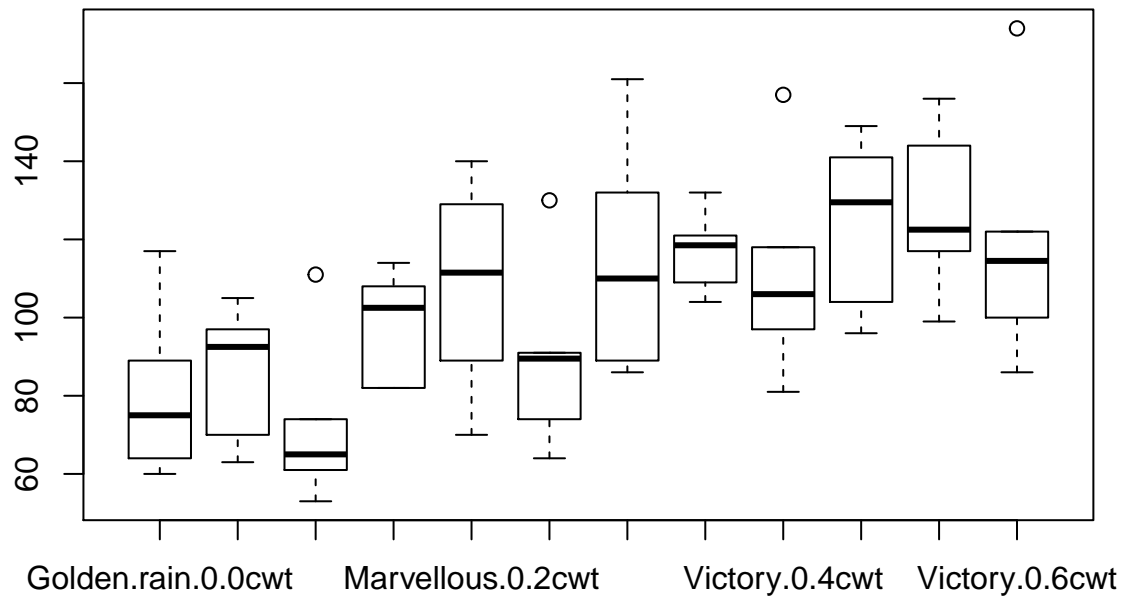
```
boxplot(Y~N)
```



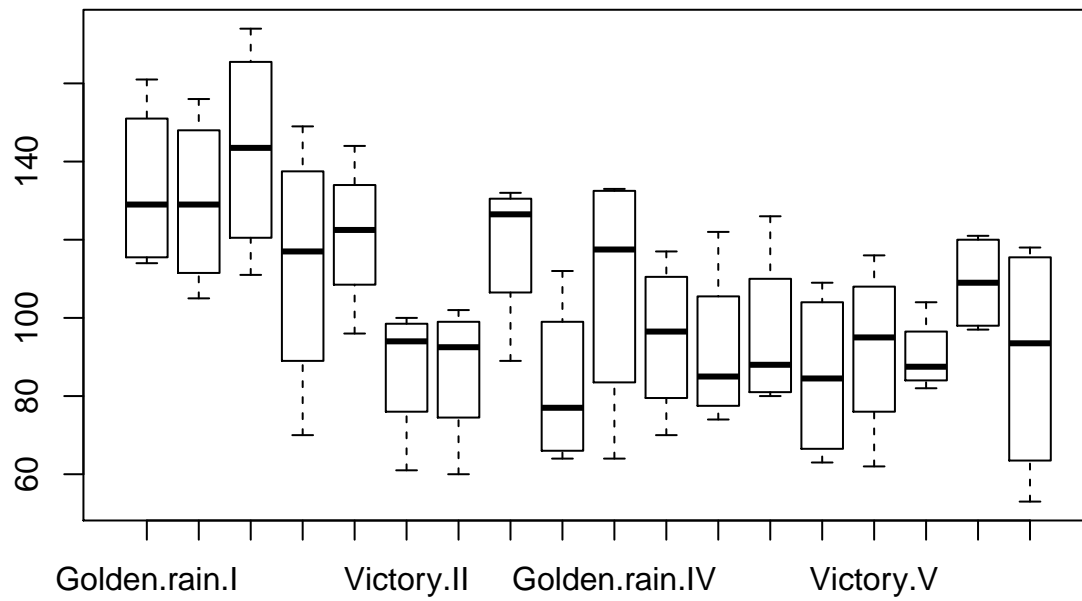
```
boxplot(Y~B)
```



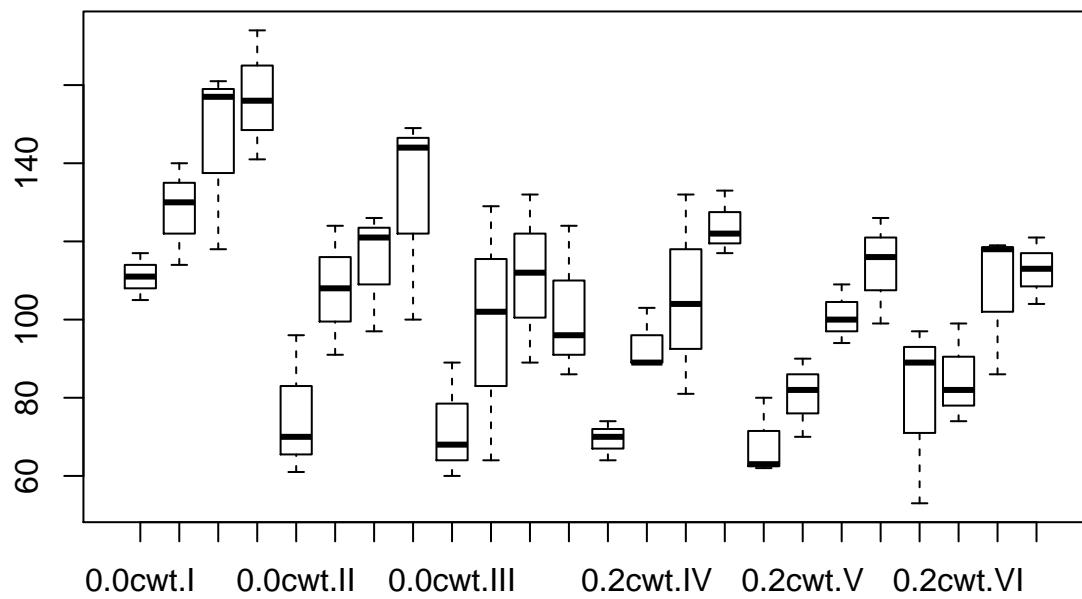
```
boxplot(Y~V*N)
```



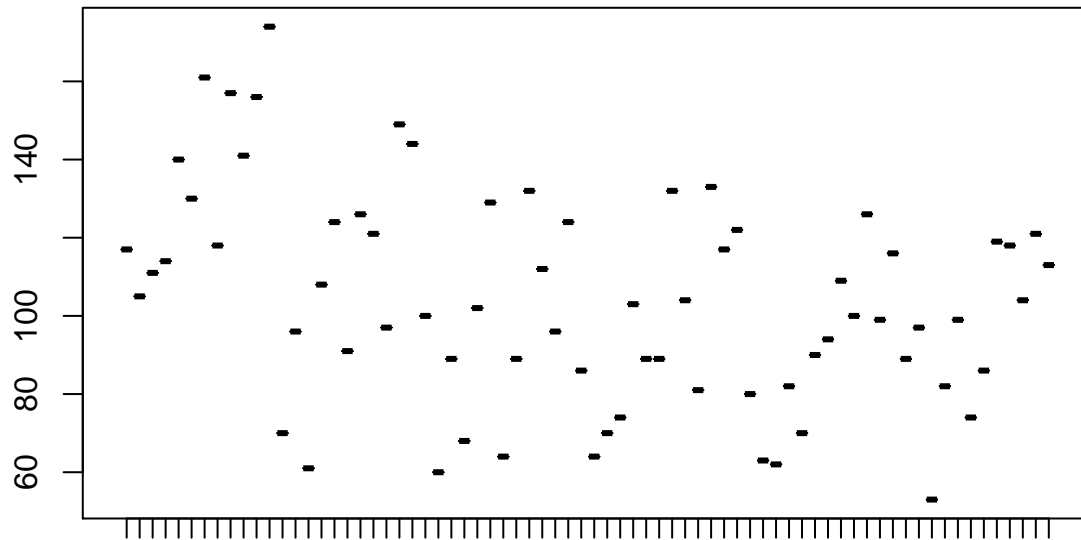
```
boxplot(Y~V*B)
```



```
boxplot(Y~N*B)
```



```
boxplot(Y~V*N*B)
```



Golden.rain.0.0cwt.I   Victory.0.6cwt.II   Victory.0.4cwt.IV   Victory.0.2cwt.VI

It is harder to tell if Varieties are different. If so, then Victory may be different from Marvellous.

There is a strong increasing trend with increasing Nitrogen concentration.

For the block, it is gradually decreasing and tending toward stability. And, II, III, IV, V and VI do not have evident difference but I is different to others.

Reanalyze this problem treating block as a random effect in model:

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
model0.2<-lmer(Y~V+N+V:N+(1|B),data=oats)
```

```
summary(model0.2)
```

```
## Linear mixed model fit by REML ['lmerMod']
```

```
## Formula: Y ~ V + N + V:N + (1 | B)
```

```
## Data: oats
```

```
##
```

```
## REML criterion at convergence: 536.7
```

```
##
```

```
## Scaled residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -1.7572 -0.7457  0.1290  0.6720  1.7505
```

```
##
```

```
## Random effects:
```

```
## Groups   Name      Variance Std.Dev.
```

```
## B        (Intercept) 243.4    15.60
```

```
## Residual                254.2    15.94
```

```
## Number of obs: 72, groups: B, 6
```

```
##
```

```
## Fixed effects:
```

```
##              Estimate Std. Error t value
```

```
## (Intercept)      80.0000     9.1070   8.784
```

```
## VMarvellous       6.6667     9.2054   0.724
```

```
## VVictory          -8.5000     9.2054  -0.923
```

```
## N0.2cwt           18.5000     9.2054   2.010
```

```

## NO.4cwt          34.6667      9.2054   3.766
## NO.6cwt          44.8333      9.2054   4.870
## VMarvellous:NO.2cwt  3.3333    13.0184   0.256
## VVictory:NO.2cwt   -0.3333    13.0184  -0.026
## VMarvellous:NO.4cwt -4.1667    13.0184  -0.320
## VVictory:NO.4cwt    4.6667    13.0184   0.358
## VMarvellous:NO.6cwt -4.6667    13.0184  -0.358
## VVictory:NO.6cwt    2.1667    13.0184   0.166
##
## Correlation of Fixed Effects:
##          (Intr) VMrvll VVctry NO.2cw NO.4cw NO.6cw VM:NO.2 VV:NO.2
## VMarvellous -0.505
## VVictory    -0.505  0.500
## NO.2cwt     -0.505  0.500  0.500
## NO.4cwt     -0.505  0.500  0.500  0.500
## NO.6cwt     -0.505  0.500  0.500  0.500  0.500
## VMrvll:NO.2  0.357 -0.707 -0.354 -0.707 -0.354 -0.354
## VVctry:NO.2  0.357 -0.354 -0.707 -0.707 -0.354 -0.354  0.500
## VMrvll:NO.4  0.357 -0.707 -0.354 -0.354 -0.707 -0.354  0.500  0.250
## VVctry:NO.4  0.357 -0.354 -0.707 -0.354 -0.707 -0.354  0.250  0.500
## VMrvll:NO.6  0.357 -0.707 -0.354 -0.354 -0.354 -0.707  0.500  0.250
## VVctry:NO.6  0.357 -0.354 -0.707 -0.354 -0.354 -0.707  0.250  0.500
##          VM:NO.4 VV:NO.4 VM:NO.6
## VMarvellous
## VVictory
## NO.2cwt
## NO.4cwt
## NO.6cwt
## VMrvll:NO.2
## VVctry:NO.2
## VMrvll:NO.4
## VVctry:NO.4  0.500
## VMrvll:NO.6  0.500  0.250
## VVctry:NO.6  0.250  0.500  0.500

```

```
library(lmerTest)
```

```

##
## Attaching package: 'lmerTest'

## The following object is masked from 'package:lme4':
##
##      lmer

## The following object is masked from 'package:stats':
##
##      step

```

```
anova(model0.2)
```

```

## Analysis of Variance Table
##      Df  Sum Sq Mean Sq F value
## V      2  1786.4   893.2   3.5134
## N      3 20020.5  6673.5  26.2510
## V:N     6   321.8    53.6   0.2109

```

```
##In hw4
model.hw4<-aov(Y~V*N+B)
summary(model.hw4)
```

```
##           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
```

The values of residuals are so similiar.The values of Sum Sq,Mean Sq, F value and Pr(>F) of V,N and V:N are so similiar between these two results.Both of them shows that Nitrogen is highly statistically significantly different and variety is statistically different.

## Denim Jeans Problem

a.

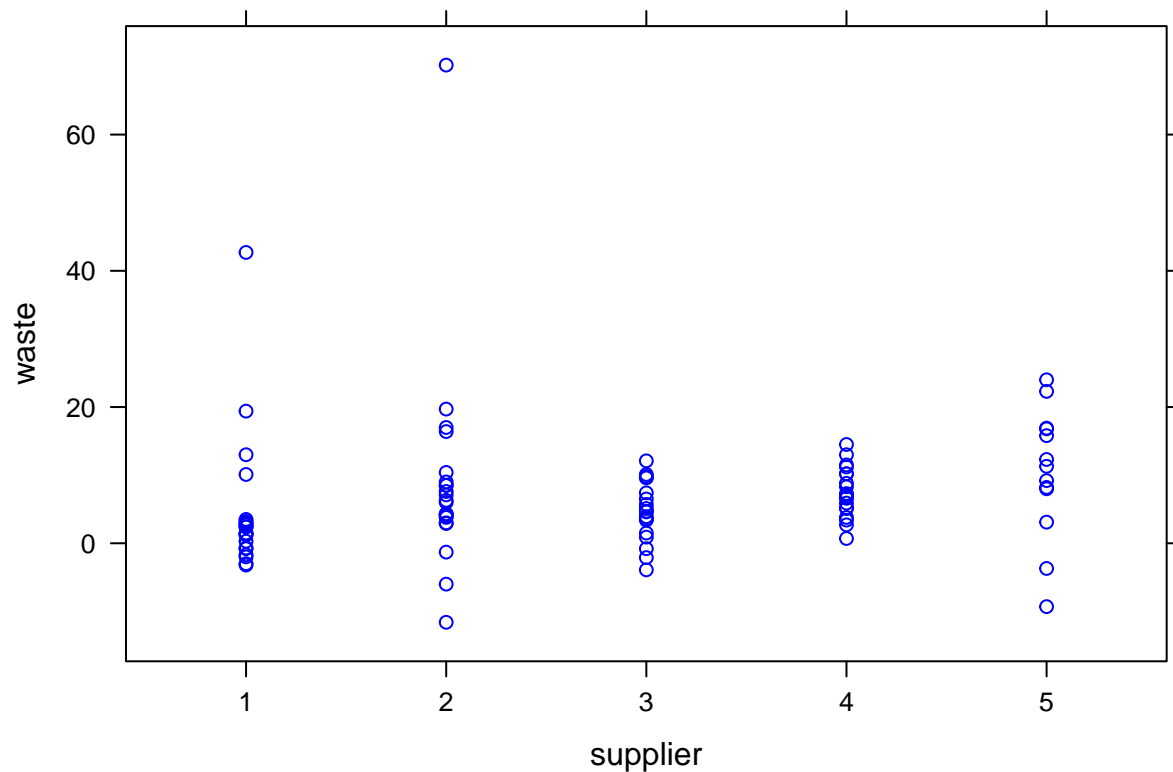
```
library(faraway)
data(denim, package = "faraway")
head(denim)
```

```
## waste supplier
## 1  1.2      1
## 2 16.4      2
## 3 12.1      3
## 4 11.5      4
## 5 24.0      5
## 6 10.1      1
```

```
attach(denim)
```

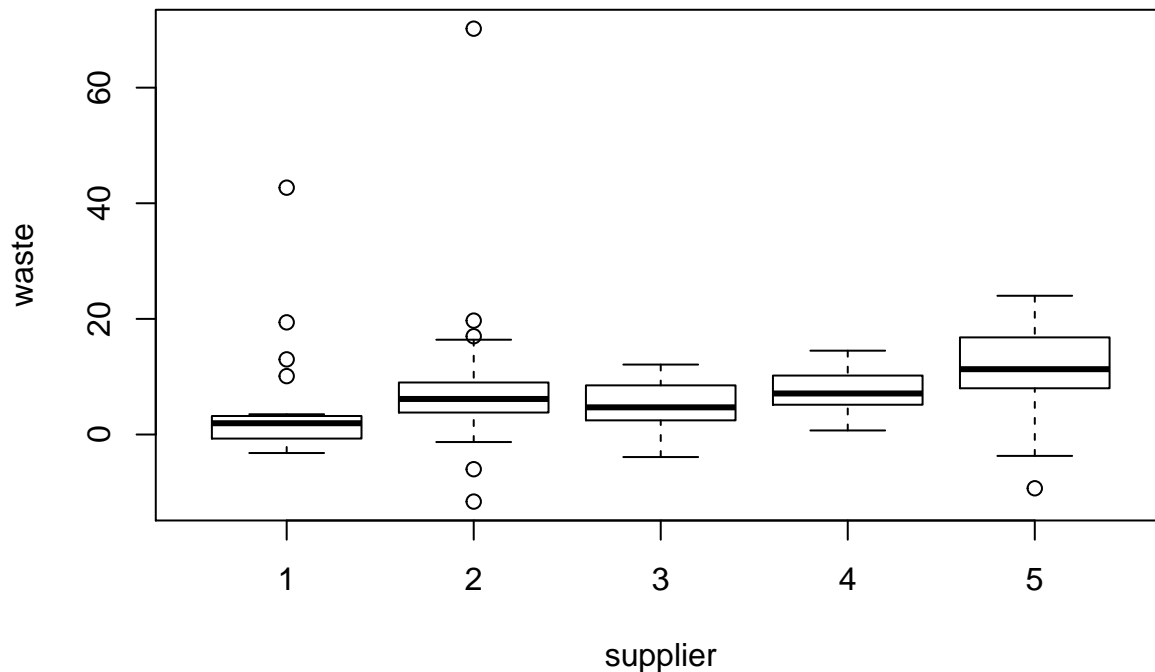
```
library(lattice)
```

```
##
## Attaching package: 'lattice'
## The following object is masked from 'package:faraway':
##
##      melanoma
plot1<-xyplot(waste~supplier,pch=21,col="blue",data=denim)
print(plot1)
```



```
boxplot(waste~supplier,xlab="supplier",ylab="waste",data=denim)
```





From the xyplot, we could predict that the variability with-in supplier is small. From the boxplot, it is harder to see that there are some difference between them.

b.

```
denim$supplier=as.factor(denim$supplier)
library(lme4)

## Loading required package: Matrix
model.1=lmer(waste~1+(1|supplier),data=denim)
summary(model.1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: waste ~ 1 + (1 | supplier)
## Data: denim
##
## REML criterion at convergence: 702.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.9095 -0.4363 -0.1669  0.3142  6.3817
##
## Random effects:
##  Groups   Name                Variance Std.Dev.
##  supplier (Intercept)  0.6711  0.8192
##  Residual                97.3350  9.8658
## Number of obs: 95, groups:  supplier, 5
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)    6.997      1.078    6.49
```

The estimate of  $\sigma_{supplier}^2$  is 0.6711 and the estimate of  $\sigma^2$  is 97.3350.

c.

The classic approach:

```
nlevels(denim$supplier)
```

```
## [1] 5
```

```
length(waste[denim$supplier==1]);length(waste[denim$supplier==2]);length(waste[denim$supplier==3]);length(waste[denim$supplier==4]);length(waste[denim$supplier==5])
```

```
## [1] 22
```

```
## [1] 22
```

```
## [1] 19
```

```
## [1] 19
```

```
## [1] 13
```

Because the data is not balanced, we can not use the classic approach.

$$e. \frac{\sigma_{supplier}^2}{\sigma_{supplier}^2 + \sigma^2} = 0.6711 / (0.6711 + 97.3350) = 0.006847533$$

f.

```
library(RLRsim)
exactRLRT(model.1)
```

```
##
```

```
## simulated finite sample distribution of RLRT.
```

```
##
```

```
## (p-value based on 10000 simulated values)
```

```
##
```

```
## data:
```

```
## RLRT = 0.029383, p-value = 0.3442
```

Because the p-value is 0.3395 which is larger than 0.05, the supplier does not have a statistically significant difference.

g.

```
confint(model.1, level = 0.95)
```

```
## Computing profile confidence intervals ...
```

```
## Warning in nextpar(mat, cc, i, delta, lowcut, upcut): Last two rows have
## identical or NA .zeta values: using minstep
```

```
## Warning in nextpar(mat, cc, i, delta, lowcut, upcut): Last two rows have
## identical or NA .zeta values: using minstep
```

```
## Warning in FUN(X[[i]], ...): non-monotonic profile for .sig01
```

```
## Warning in confint.thpr(pp, level = level, zeta = zeta): bad spline fit
## for .sig01: falling back to linear interpolation
```

```
##           2.5 %    97.5 %
## .sig01      0.000000  4.206374
## .sigma      8.591567 11.424996
## (Intercept) 4.977790  8.975895
```

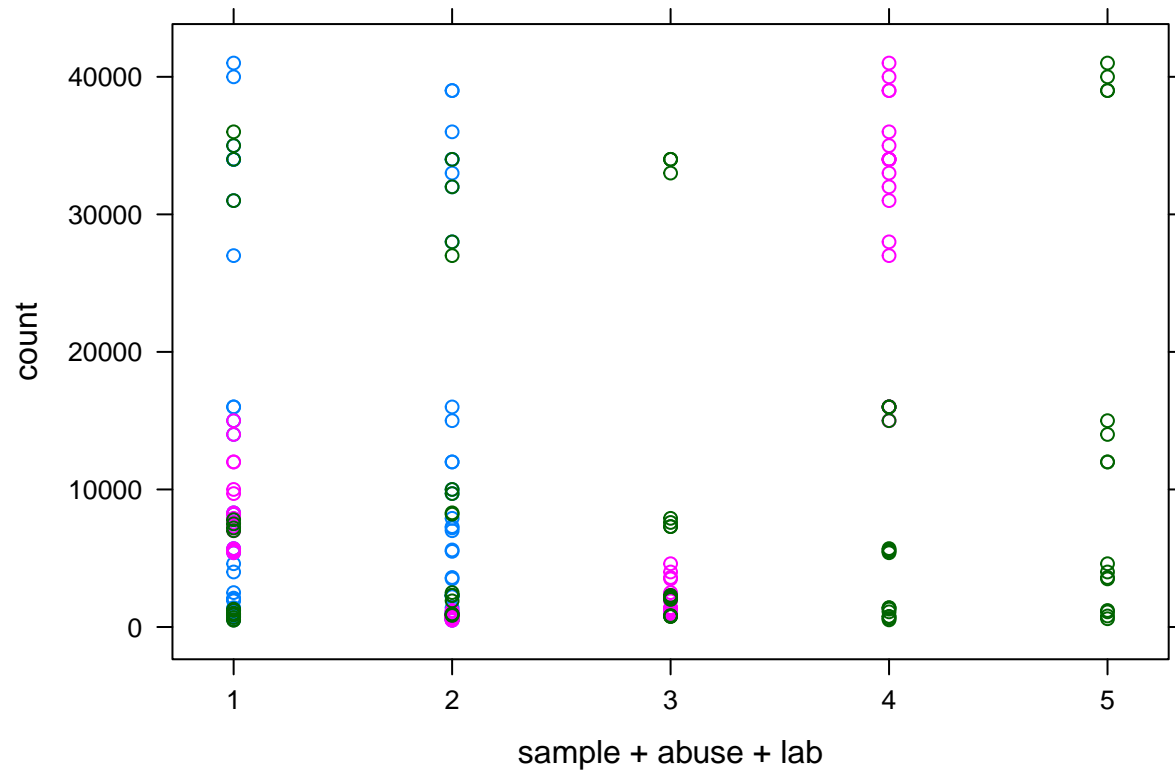
Bonus:

```
prob<-read.table("http://www.stat.umn.edu/~gary/book/fcdae.data/pr12.6",header = TRUE)
head(prob)
```

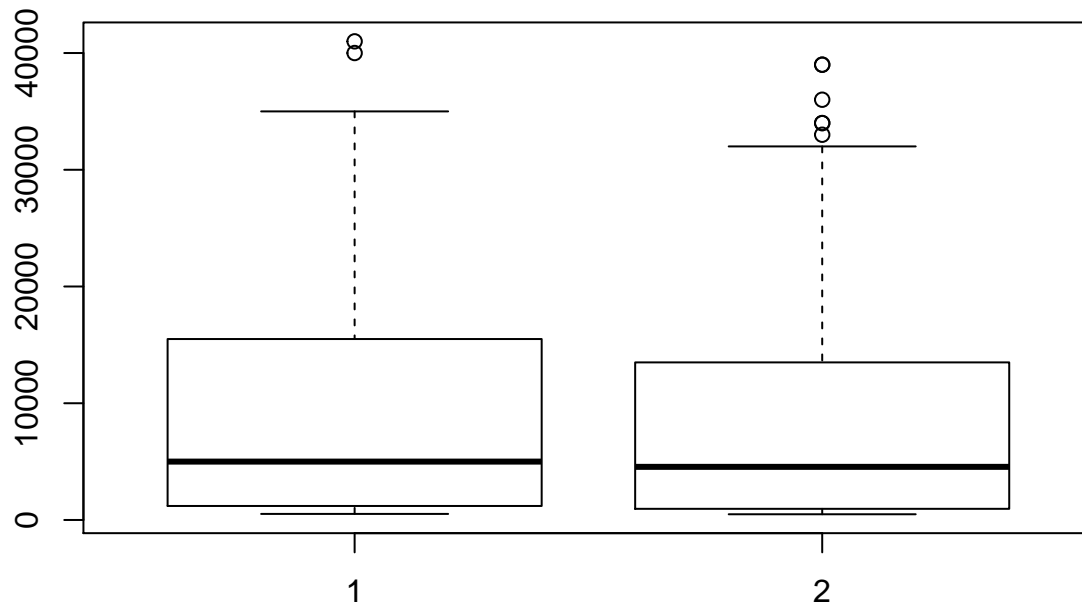
```
##      sample abuse lab count
## 1      1      1   1  7800
## 2      2      1   1  7000
## 3      1      2   1   870
## 4      2      2   1   490
## 5      1      3   1  1300
## 6      2      3   1  1000
```

a.

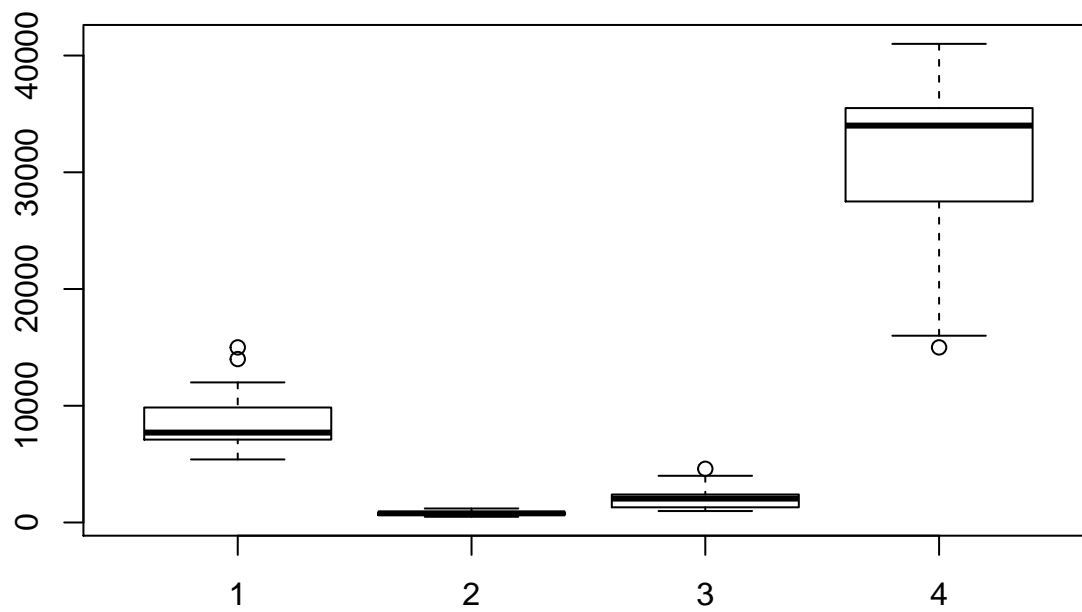
```
library(lattice)
print(xyplot(count~sample+abuse+lab,data = prob))
```



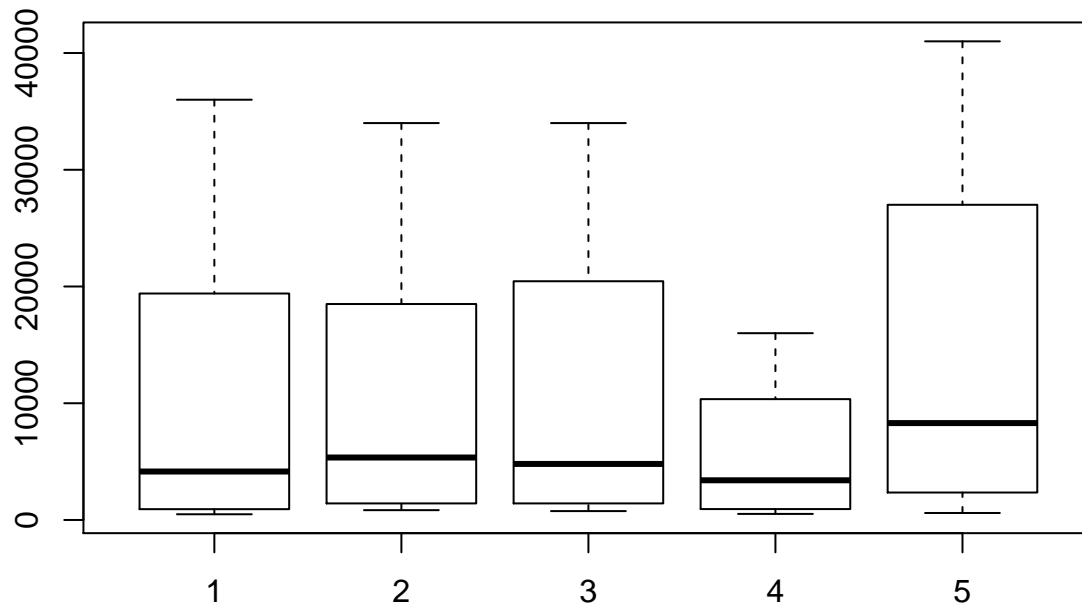
```
boxplot(count~sample,data = prob)
```



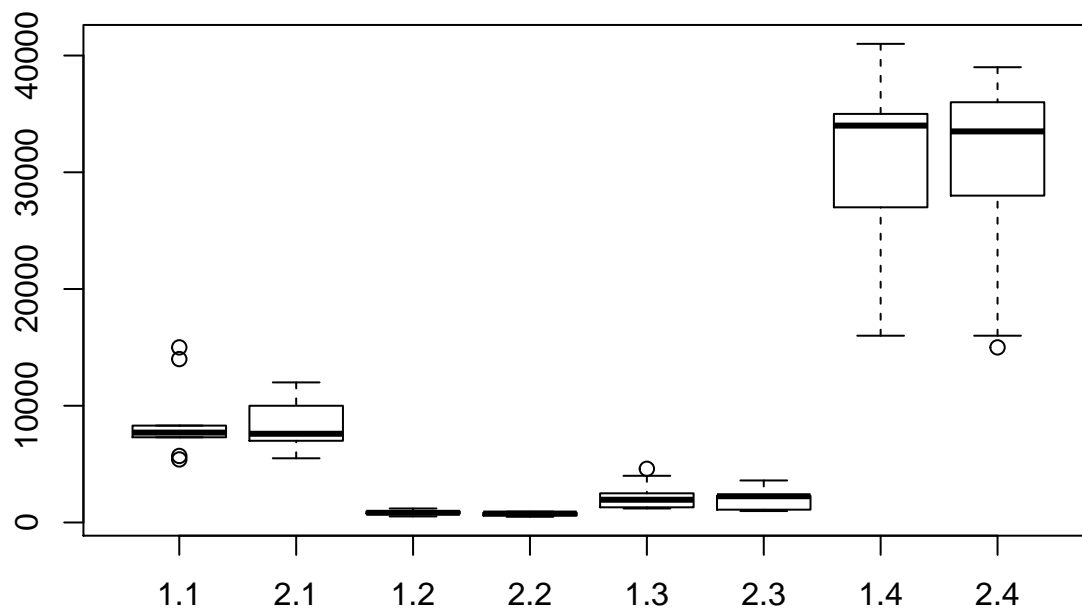
```
boxplot(count~abuse,data = prob)
```



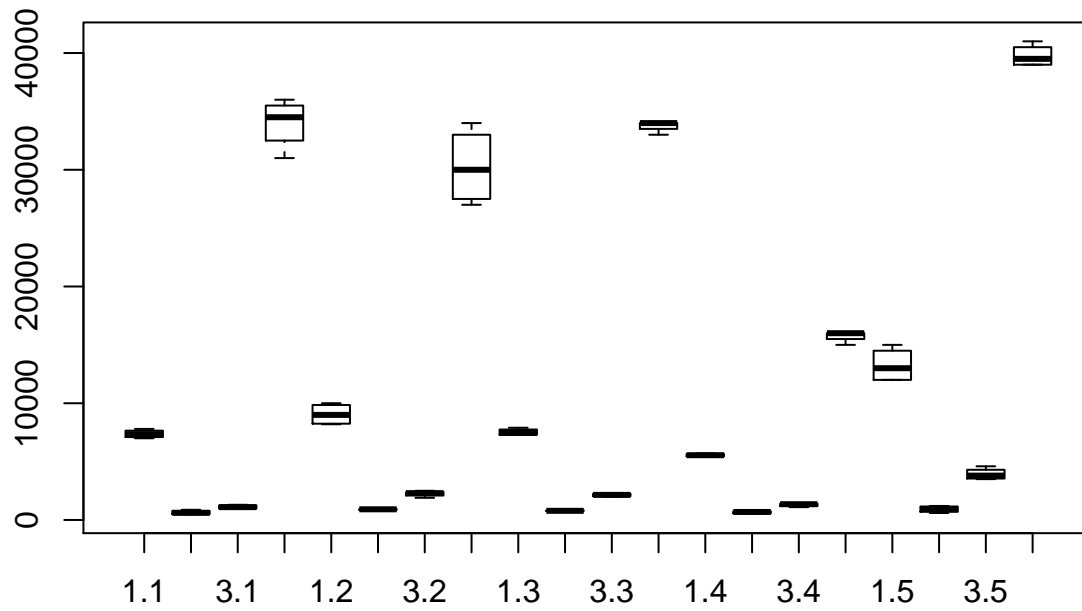
```
boxplot(count~lab,data = prob)
```



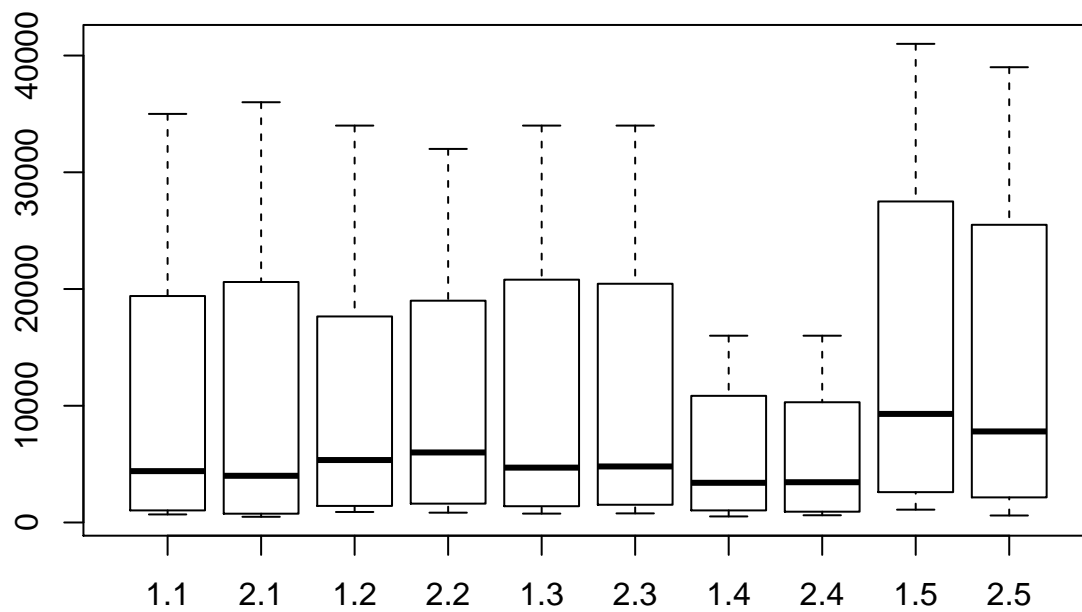
```
boxplot(count~sample*abuse,data = prob)
```



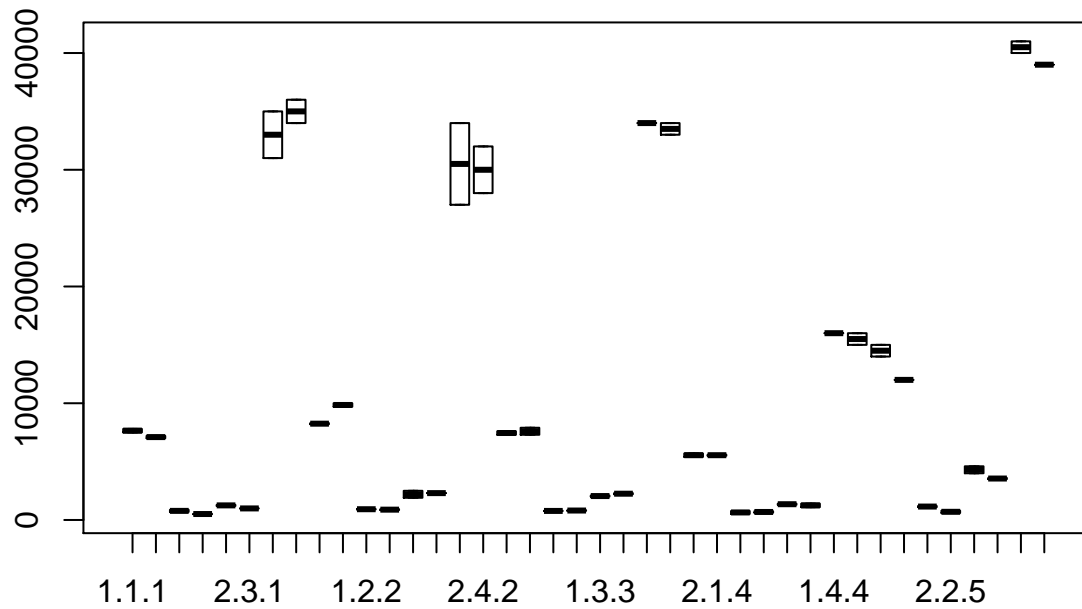
```
boxplot(count~abuse*lab,data = prob)
```



```
boxplot(count~sample*lab,data = prob)
```

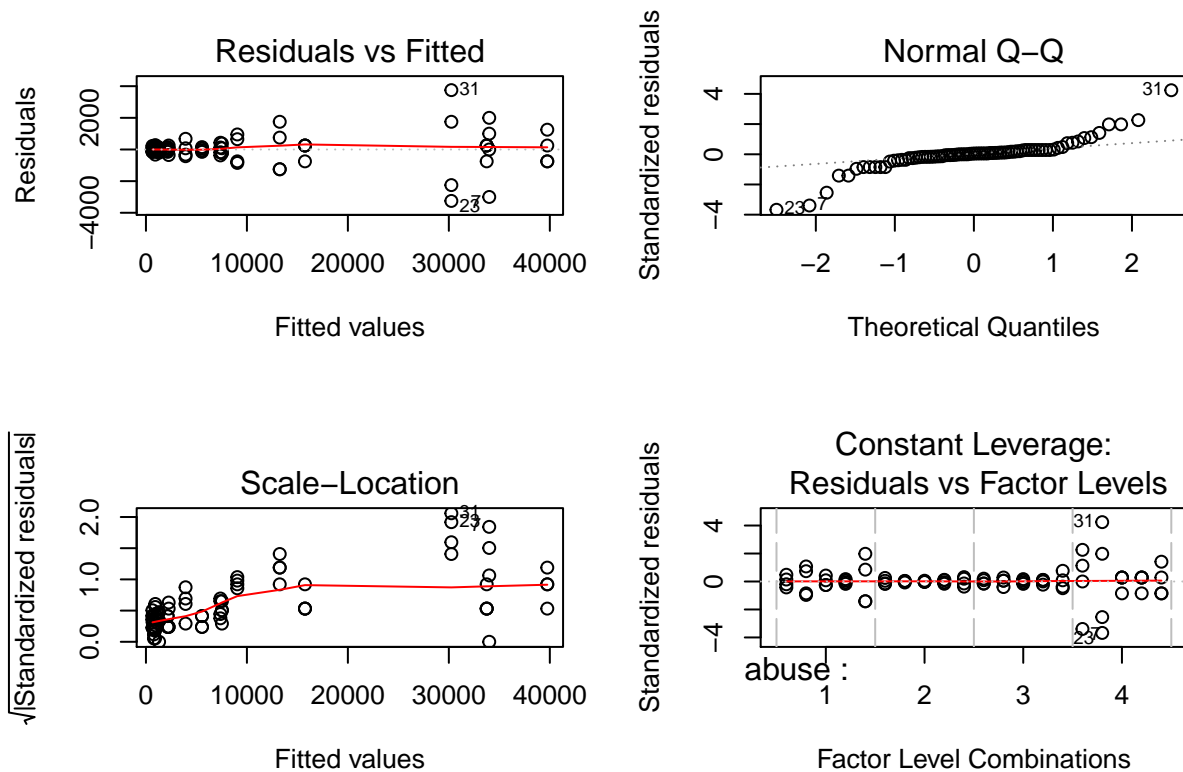


```
boxplot(count~sample*abuse*lab,data = prob)
```



b. Classic approach:

```
prob$abuse=as.factor(prob$abuse)
prob$lab=as.factor(prob$lab)
prob$sample=as.factor(prob$sample)
model_1=aov(count~abuse*lab,data=prob)
##check assumptions:
par(mfrow=c(2,2))
plot(model_1)
```



```

library(MASS)
box<-boxcox(model_1,lambda=seq(-3,3))
lam<-box$x[which.max(box$y)]
lam

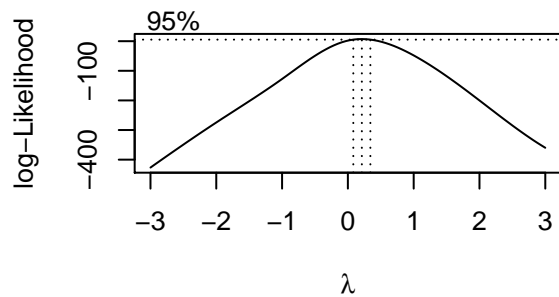
## [1] 0.2121212

##new model
model_1_new=aov((count~0.25)~abuse*lab,data=prob)
summary(model_1_new)

##              Df Sum Sq Mean Sq F value Pr(>F)
## abuse          3  723.0   240.99  5327.29 <2e-16 ***
## lab            4   28.4     7.09   156.82 <2e-16 ***
## abuse:lab      12   13.6     1.13    25.08 <2e-16 ***
## Residuals     60    2.7     0.05
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##check assumptiona:
par(mfrow=c(2,2))

```

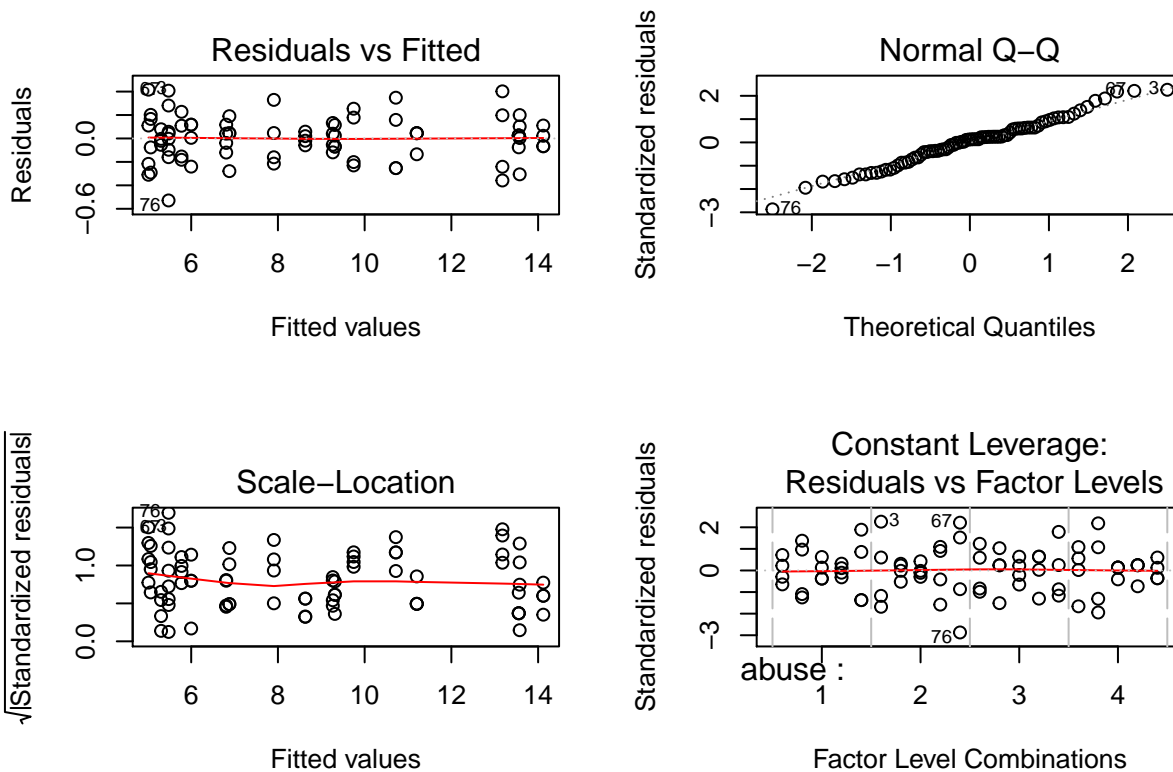


```

plot(model_1_new)

```





```
par(mfrow=c(1,1))
```

```
n=nrow(prob)
b_lab=nlevels(prob$lab)
a_abuse=nlevels(prob$abuse)
sigma_sq= 0.01
sigma_al_beta_sq=(0.28-sigma_sq)/n
##abuse-to-abuse variability:
(64.20-sigma_sq-n*sigma_al_beta_sq)/(n*b_lab)
```

```
## [1] 0.1598
```

```
##lab-to-lab variability:
(1.87-sigma_sq-n*sigma_al_beta_sq)/(n*a_abuse)
```

```
## [1] 0.00496875
```

New approach:

```
library(lme4)
model_2=lmer((count~0.25)~1+(1|abuse)+(1|lab)+(1|abuse:lab),data=prob)
summary(model_2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: (count~0.25) ~ 1 + (1 | abuse) + (1 | lab) + (1 | abuse:lab)
## Data: prob
##
## REML criterion at convergence: 68.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.5926 -0.5601  0.0499  0.5623  1.9444
```

```
##
## Random effects:
## Groups      Name      Variance Std.Dev.
## abuse:lab (Intercept)  0.27219 0.5217
## lab        (Intercept)  0.37228 0.6101
## abuse      (Intercept) 11.99414 3.4633
## Residual                0.04524 0.2127
## Number of obs: 80, groups: abuse:lab, 20; lab, 5; abuse, 4
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)    8.651      1.757    4.924
```

The abuse-to-abuse variability is 3.19716 and the lab-to-lab variability is 0.09939.

c.

Classic approach:

```
## Testing abuse
F <- 3.19716/0.06654 # F_A=MS_A/MS_AB
pf(F,df1=3,df2=12,lower.tail = FALSE) # upper tail, df1=df_A, df2=df_AB

## [1] 5.839207e-07

## Testing lab
F <- 0.09939/0.06654 # F_B=MS_B/MS_AB
pf(F,df1=4,df2=12,lower.tail=FALSE) # upper tail, df1=df_B, df2=df_AB

## [1] 0.2651045

## Testing interaction
F <- 0.06654/0.01296 # F_AB=MS_AB/MS_E
pf(F,df1=12,df2=60,lower.tail = FALSE) # upper tail, df1=df_AB, df2=df_E

## [1] 7.967097e-06
```

New approach:

```
##Test of Random Effect:
library(RLRsim)
##Test of abuse:
no_abuse<-lmer((count~0.25)~1+(1|lab)+(1|abuse:lab),data=prob)

## boundary (singular) fit: see ?isSingular
abuse<-lmer((count~0.25)~1+(1|abuse),data=prob)
exactRLRT(abuse,model_2,no_abuse)

##
## simulated finite sample distribution of RLRT.
##
## (p-value based on 10000 simulated values)
##
## data:
## RLRT = 44.407, p-value < 2.2e-16

##Test of lab:
no_lab<-lmer((count~0.25)~1+(1|abuse)+(1|abuse:lab),data=prob)
lab<-lmer((count~0.25)~1+(1|lab),data=prob)
```

```

## boundary (singular) fit: see ?isSingular
exactRLRT(lab,model_2,no_lab)

##
## simulated finite sample distribution of RLRT.
##
## (p-value based on 10000 simulated values)
##
## data:
## RLRT = 6.0862, p-value = 0.0046
##Test of abuse by lab interaction:
no_int<-lmer((count~0.25)+1+(1|abuse)+(1|lab),data=prob)
int<-lmer((count~0.25)+1+(1|abuse:lab),data=prob)
exactRLRT(int,model_2,no_int)

##
## simulated finite sample distribution of RLRT.
##
## (p-value based on 10000 simulated values)
##
## data:
## RLRT = 77.404, p-value < 2.2e-16
d.
##the percent variation due to abuse:
3.19716/(3.19716+0.01296 )

## [1] 0.9959628
##the percent variation due to lab:
0.09939/(0.09939+0.01296 )

## [1] 0.8846462

```