

# Stat461 HW7

Enbo Cao

Oct 29, 2020

## Question 1

#a

$$Y_{ijt} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijt}, \quad \epsilon_{ijt} \stackrel{iid}{\sim} N(0, \sigma^2)$$
$$i = control, f1, f2, f3 \quad j = SppA, SppB \quad t_i = 1, 2, \dots, 12 \quad t_j = 1, 2, \dots, 6$$

```
Fert<-c(rep("control", 12), rep("f1", 12),
rep("f2", 12), rep("f3", 12))
Species<-c(rep(c(rep("SppA", 6), rep("SppB", 6)),4))
Height<-c(21.0, 19.5, 22.5, 21.5, 20.5, 21.0,
23.7, 23.8, 23.8, 23.7, 22.8, 24.4,
32.0, 30.5, 25.0, 27.5, 28.0, 28.6,
30.1, 28.9, 30.9, 34.4, 32.7, 32.7,
22.5, 26.0, 28.0, 27.0, 26.5, 25.2,
30.6, 31.1, 28.1, 34.9, 30.1, 25.5,
28.0, 27.5, 31.0, 29.5, 30.0, 29.2,
36.1, 36.6, 38.7, 37.1, 36.8, 37.1)
df<-data.frame(Fert=Fert, Species=Species, Height=Height)
```

#b

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.0.3
```

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

```
## Warning: package 'carData' was built under R version 4.0.3
```

```
modell1 = aov(Height~Fert+Species+Fert:Species,data = df)
Anova(modell1, type = "III")
```

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

```
##
```

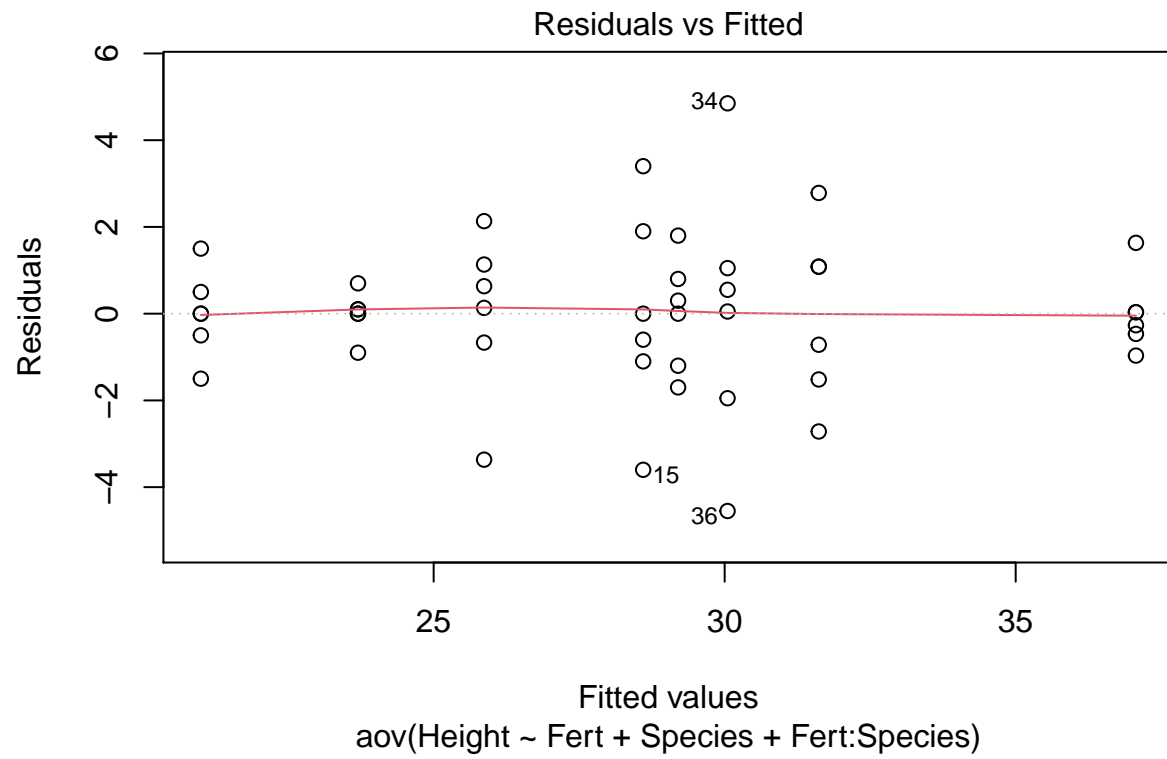
```
## Response: Height
```

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	2646.00	1	778.4070	< 2.2e-16 ***
Fert	251.44	3	24.6564	3.373e-09 ***
Species	21.87	1	6.4338	0.015204 *
Fert:Species	50.58	3	4.9603	0.005081 **
Residuals	135.97	40		

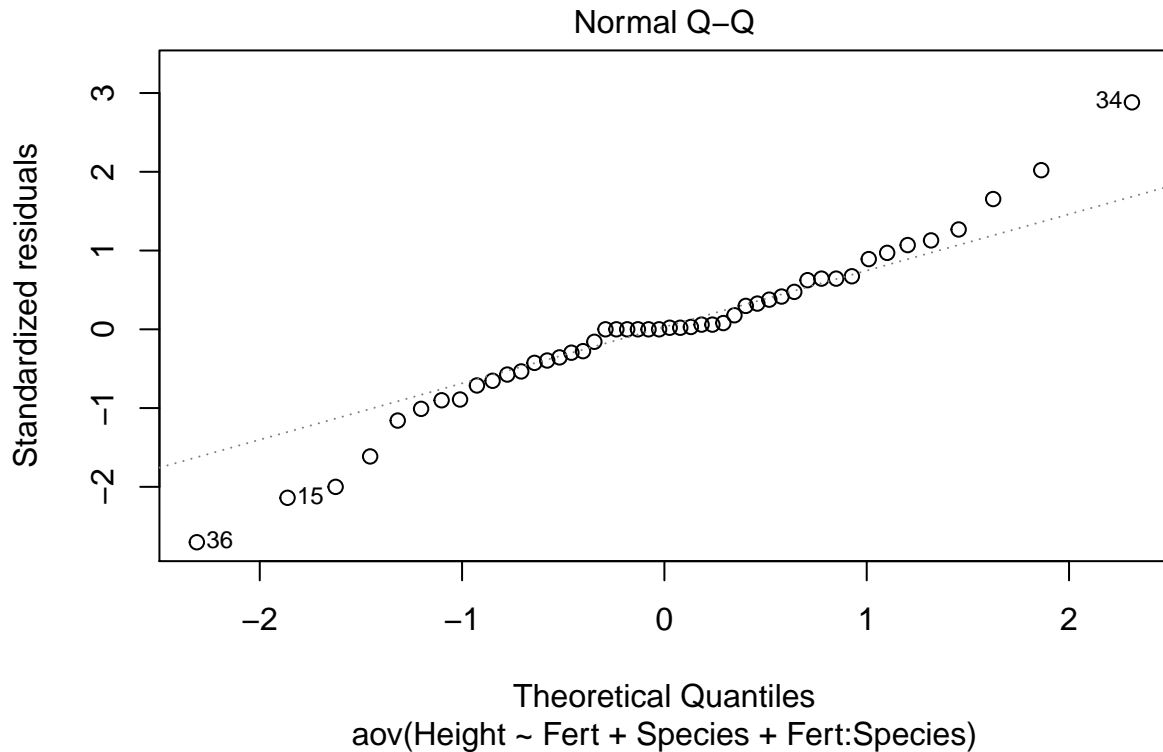
```
## ---
```

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

```
plot(model1, which = 1)
```



```
plot(model1, which = 2)
```



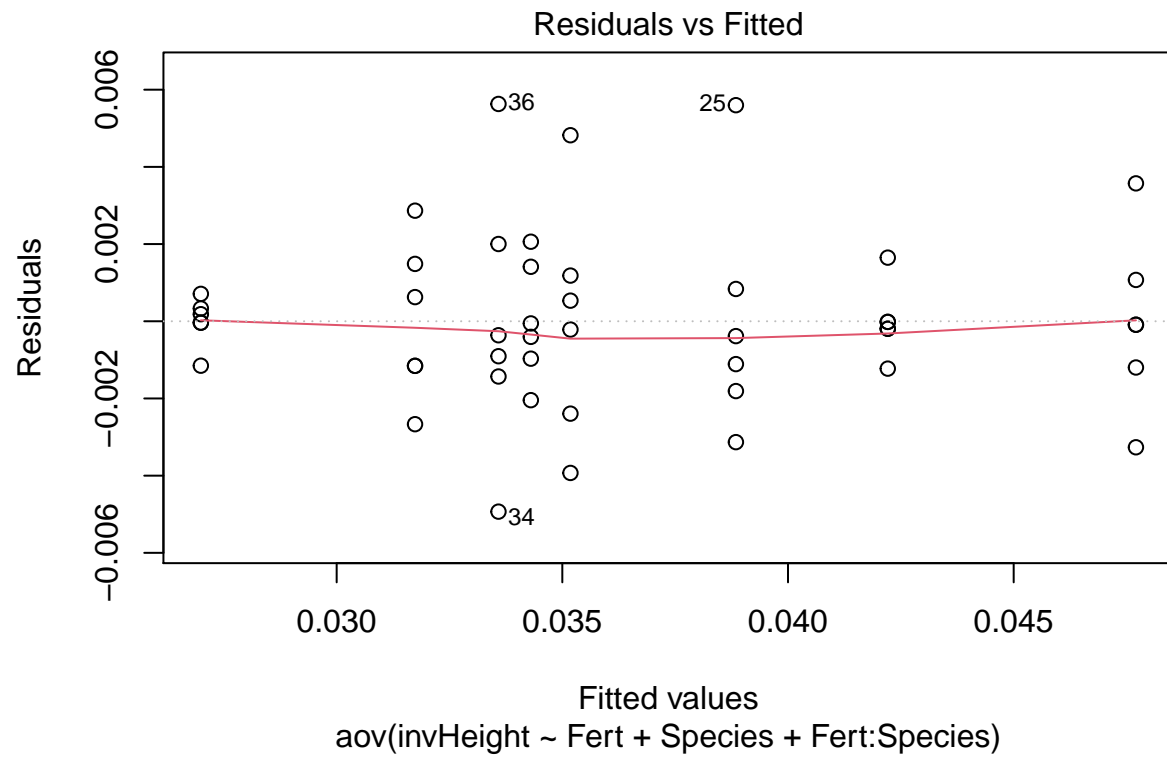
```
df$invHeight = 1/(df$Height)
invmodel = aov(invHeight~Fert+Species+Fert:Species,data = df)
Anova(invmodel, type = "III")

## Anova Table (Type III tests)
##
## Response: invHeight
##
```

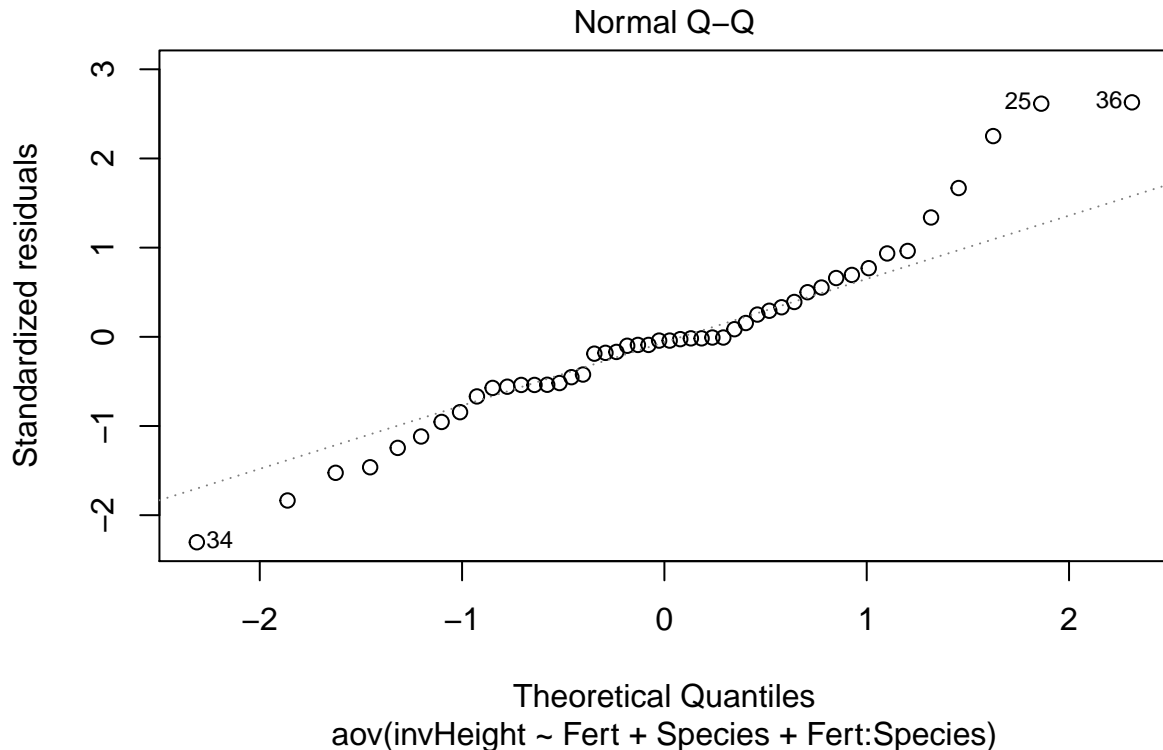
	Sum Sq	Df	F value	Pr(>F)
(Intercept)	0.0136572	1	2482.0991	< 2.2e-16 ***
Fert	0.0006752	3	40.9062	2.921e-12 ***
Species	0.0000907	1	16.4849	0.0002217 ***
Fert:Species	0.0000225	3	1.3647	0.2673800
Residuals	0.0002201	40		

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

plot(invmodel, which = 1)
```



```
plot(invmodel, which = 2)
```



```
# After trying several transformation for the model, I found that all transformations
# are close to each other. I think the inverse transformation is the best.

# The residuals are not approximately normal because the QQ line isn't a straight line.

# The assumption of constant error variance among treatments is justified,
# since the residuals are in a square shape.
modelnoInter = aov(invHeight~Fert+Species,data = df)
anova(modelnoInter)
```

```
## Analysis of Variance Table
##
## Response: invHeight
##          Df      Sum Sq   Mean Sq F value    Pr(>F)
## Fert       3 0.00138048 0.00046016  81.556 < 2.2e-16 ***
## Species    1 0.00034714 0.00034714   61.525 7.899e-10 ***
## Residuals 43 0.00024262 0.00000564
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

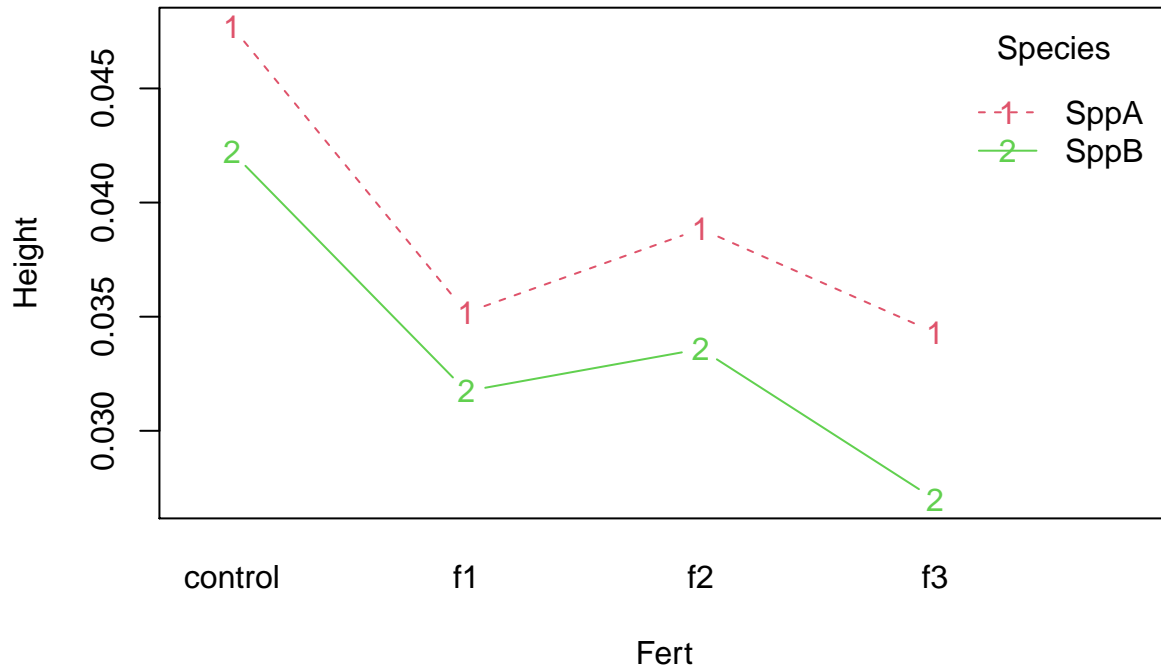
#c
```

$H_0 : (\alpha\beta)_{ij} = 0$  for all  $i, j$   $H_a$  : at least one treatment is different.

Since Fert:Species is 0.26 and larger than  $\alpha = 0.05$ , we are fail to reject  $H_0$ . There are not significant interactions.

Both Fert and Species row in the ANOVA table indicates that there are significant differences height to their removal because both P-value is smaller than 0.05.

```
interaction.plot(x.factor= df$Fert, trace.factor = df$Species, response = df$invHeight, type = "b", col
library(lsmeans)
```



```
lsmfert = lsmeans(invmodel, ~ Fert)
contrast(lsmfert, method = "pairwise")
```

```
## contrast      estimate      SE df t.ratio p.value
## control - f1  0.01150 0.000958 40 12.011  <.0001
## control - f2  0.00874 0.000958 40  9.131  <.0001
## control - f3  0.01431 0.000958 40 14.947  <.0001
## f1 - f2      -0.00276 0.000958 40 -2.879  0.0311
## f1 - f3       0.00281 0.000958 40  2.936  0.0270
## f2 - f3       0.00557 0.000958 40  5.816  <.0001
##
## Results are averaged over the levels of: Species
## P value adjustment: tukey method for comparing a family of 4 estimates
```

```
lsmsspecies = lsmeans(invmodel, ~ Species)
contrast(lsmsspecies, method = "pairwise")
```

```
## contrast      estimate      SE df t.ratio p.value
## SppA - SppB  0.00538 0.000677 40  7.943  <.0001
##
## Results are averaged over the levels of: Fert
```

```
#all species and fert are likely different.
library(multcompView)
```

```
library(multcomp)
lsminster = lsmeans(invmodel, ~ Fert:Species)
CLD(lsminster)
```

##	Fert	Species	lsmean	SE	df	lower.CL	upper.CL	.group
##	f3	SppB	0.0270	0.000958	40	0.0251	0.0289	1
##	f1	SppB	0.0317	0.000958	40	0.0298	0.0337	2
##	f2	SppB	0.0336	0.000958	40	0.0317	0.0355	2
##	f3	SppA	0.0343	0.000958	40	0.0324	0.0362	2
##	f1	SppA	0.0352	0.000958	40	0.0332	0.0371	23
##	f2	SppA	0.0388	0.000958	40	0.0369	0.0408	34
##	control	SppB	0.0422	0.000958	40	0.0403	0.0441	4
##	control	SppA	0.0477	0.000958	40	0.0458	0.0496	5

```
##
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 8 estimates
## significance level used: alpha = 0.05
```

## Question 2

#a

$$Y_{ijt} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijt}, \quad \epsilon_{ijt} \stackrel{iid}{\sim} N(0, \sigma^2)$$

$$i = 1, 2 \quad j = 1, 2 \quad t_i = 1, 2, \dots, 10 \quad t_j = 1, 2, \dots, 10$$

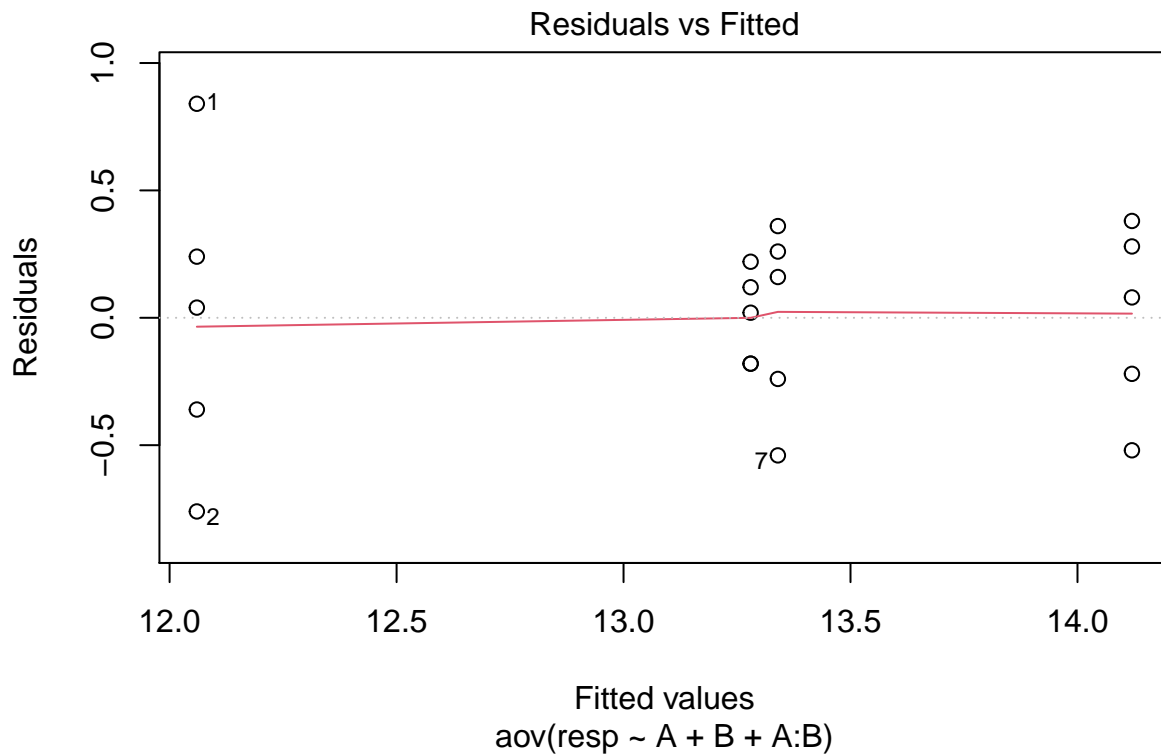
```
A<-c(rep(1, 10), rep(2, 10))
B<-rep(c(c(rep(1, 5), rep(2, 5))), 2)
resp<-c(12.9, 11.3, 11.7, 12.1, 12.3,
13.7, 12.8, 13.6, 13.1, 13.5,
14.2, 14.5, 13.9, 13.6, 14.4,
13.5, 13.1, 13.3, 13.1, 13.4)
df2<-data.frame(A=A, B=B, resp=resp)
df2
```

```
##      A B resp
## 1   1 1 12.9
## 2   1 1 11.3
## 3   1 1 11.7
## 4   1 1 12.1
## 5   1 1 12.3
## 6   1 2 13.7
## 7   1 2 12.8
## 8   1 2 13.6
## 9   1 2 13.1
## 10  1 2 13.5
## 11  2 1 14.2
## 12  2 1 14.5
## 13  2 1 13.9
## 14  2 1 13.6
## 15  2 1 14.4
## 16  2 2 13.5
## 17  2 2 13.1
## 18  2 2 13.3
## 19  2 2 13.1
```

```
## 20 2 2 13.4
#b
model2 = aov(resp~A+B+A:B,data = df2)
Anova(model2, type = "III")

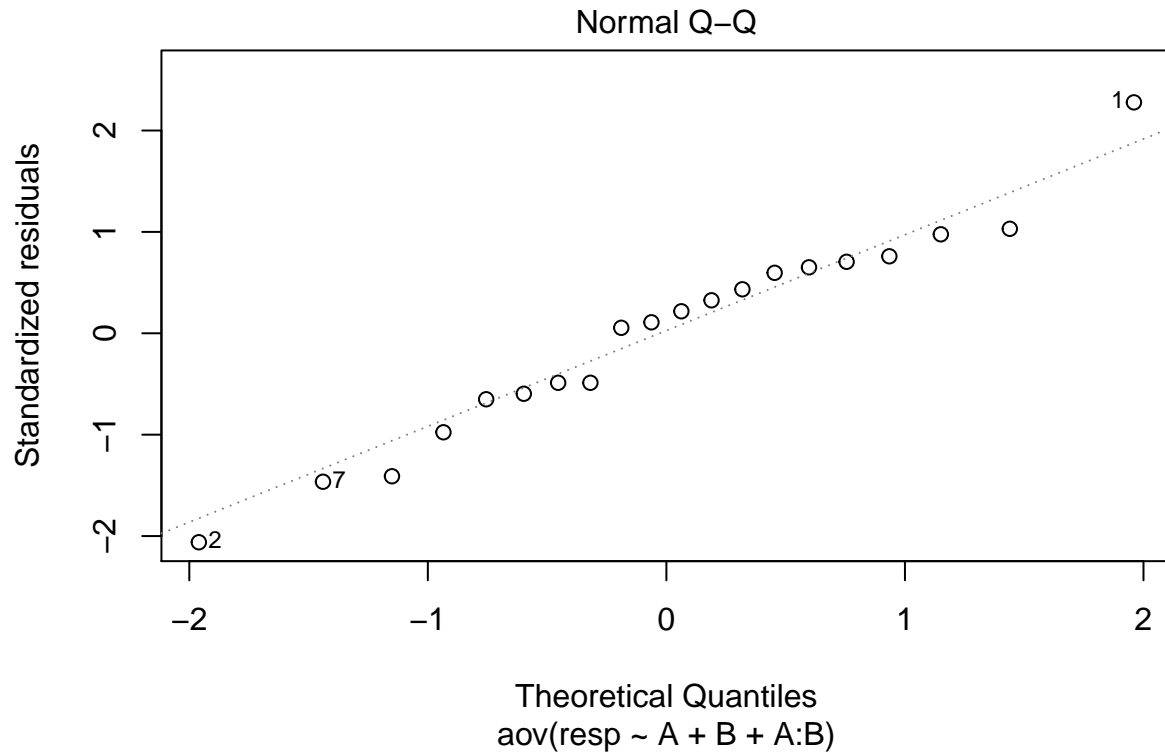
## Anova Table (Type III tests)
##
## Response: resp
##          Sum Sq Df F value    Pr(>F)
## (Intercept)  8.7120  1  51.247 2.272e-06 ***
## A            8.7362  1  51.389 2.233e-06 ***
## B            5.7800  1  34.000 2.554e-05 ***
## A:B          5.6180  1  33.047 2.991e-05 ***
## Residuals    2.7200 16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

plot(model2, which = 1)
```



```
plot(model2, which = 2)
```

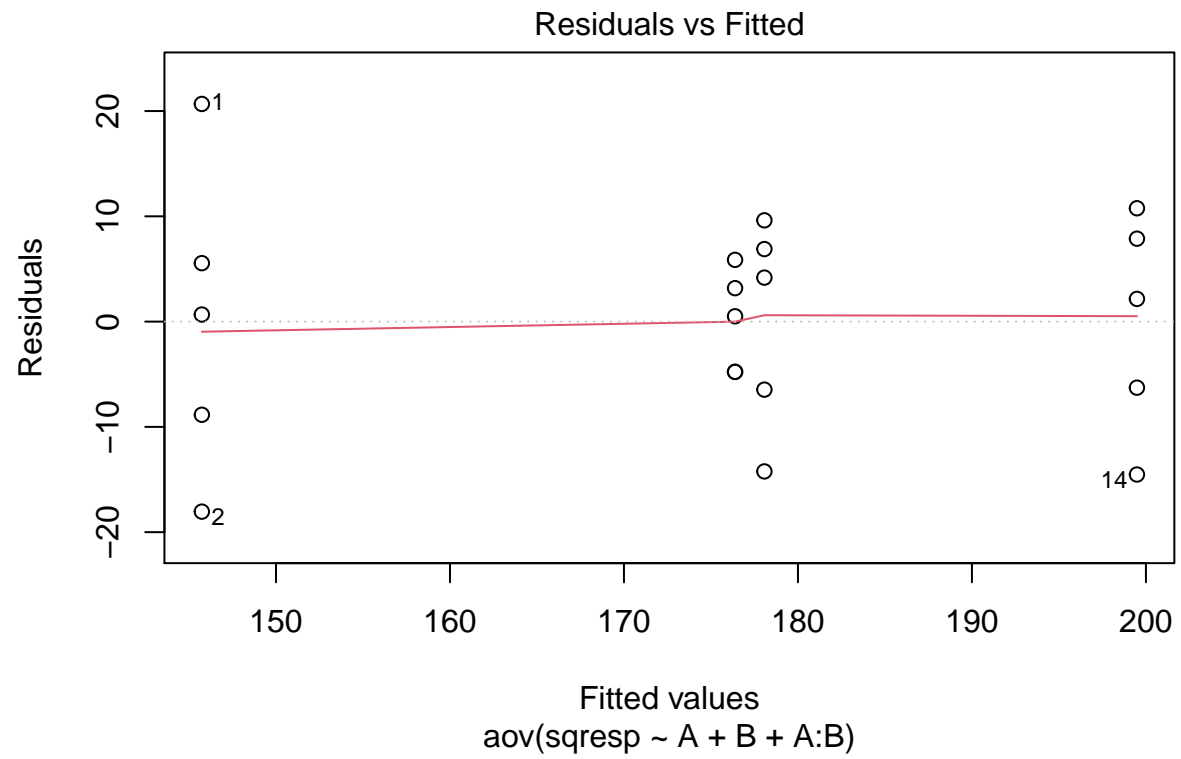




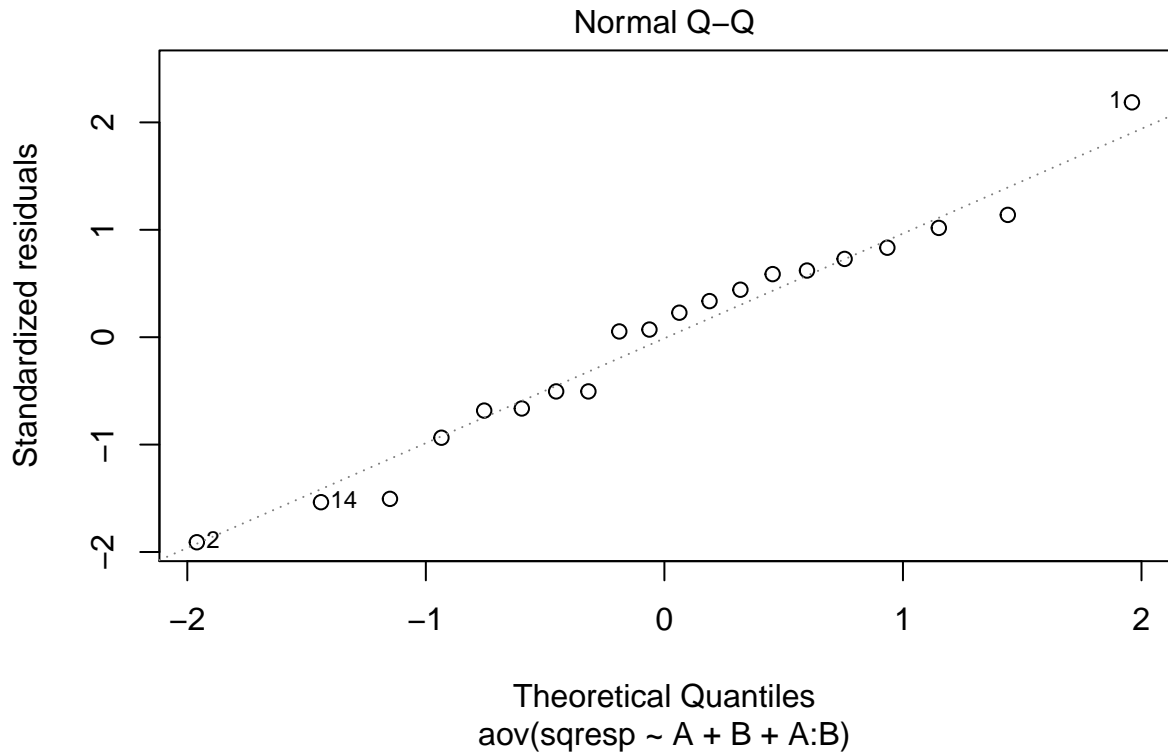
```
df2$sqresp = (df2$resp)^2
sqmodel = aov(sqresp~A+B+A:B,data = df2)
Anova(sqmodel, type = "III")
```

```
## Anova Table (Type III tests)
##
## Response: sqresp
##          Sum Sq Df F value    Pr(>F)
## (Intercept)   3.6  1   0.032    0.8603
## A             5959.9  1  53.345 1.770e-06 ***
## B             3851.3  1  34.471 2.365e-05 ***
## A:B           3840.9  1  34.379 2.401e-05 ***
## Residuals    1787.6 16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

plot(sqmodel, which = 1)
```



```
plot(sqmodel, which = 2)
```



*# After trying several transformation for the model, I found that all transformations are close to each other. I think the square transformation is the best.*

*# The residuals are approximately normal because the QQ line is a straight line.*

*# The assumption of constant error variance among treatments is justified, since the residuals are in a square shape.*

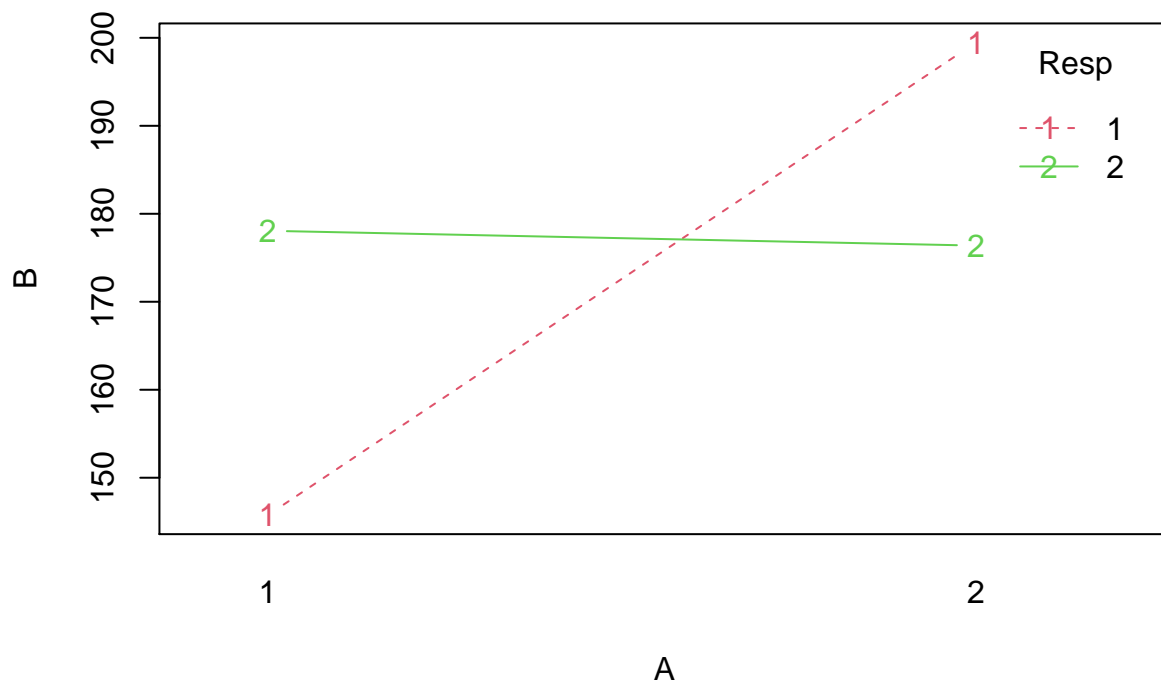
#c

$H_0 : (\alpha\beta)_{ij} = 0$  for all  $i, j$   $H_a$  :at least one treatment is different.

Since A:B is smaller than  $\alpha = 0.05$ , we reject  $H_0$ . There are significant interactions.

Both A and B row in the ANOVA table indicates that there are significant differences height to their removal because both P-value is smaller than 0.05.

```
interaction.plot(x.factor= df2$A, trace.factor = df2$B, response = df2$sqresp, type = "b", col = 2:3,xlab = "A", ylab = "sqresp", main = "Interaction Plot")
```



```
library(lsmeans)
lsmA = lsmeans(sqmodel, ~ A)
contrast(lsmA, method = "pairwise")

## contrast estimate SE df t.ratio p.value
## 1 - 2 -26 4.73 16 -5.507 <.0001
##
## Results are averaged over the levels of: B

lsmB = lsmeans(sqmodel, ~ B)
contrast(lsmB, method = "pairwise")

## contrast estimate SE df t.ratio p.value
## 1 - 2 -4.62 4.73 16 -0.977 0.3433
##
## Results are averaged over the levels of: A

#A is likely different.
library(multcompView)
library(multcomp)
lsminter2 = lsmeans(sqmodel, ~ A:B)
CLD(lsminter2)

## A B lsmean SE df lower.CL upper.CL .group
## 1 1 146 4.73 16 136 156 1
## 2 2 176 4.73 16 166 186 2
## 1 2 178 4.73 16 168 188 2
## 2 1 199 4.73 16 189 210 3
```

```
##
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 4 estimates
## significance level used: alpha = 0.05
```

### Question 3

$$Y_{it} = \mu + \tau_i + \epsilon_{it}, \quad i = A, B \quad t = 1, 2, \dots, r_i \quad r_A = 2$$

$$\epsilon_{it} \stackrel{iid}{\sim} N(0, \sigma^2)$$