

STAT461 HW8

Xiangyu Ren

10/30/2020

Problem 1. Greenhouse. Consider an experiment to study the effect of three types of fertilizer (F1, F2, and F3) on the growth of two species of plant (SppA and SppB). The data are as follows:

```
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)
```

(a) Write out the 2-way complete model for this experiment.

The 2-way complete model is:

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

and $i = \text{control}, f_1, f_2, f_3, j = \text{Spp}_A, \text{Spp}_B, t_i = 1, 2, \dots, 12, t_j = 1, 2, \dots, 6$

(b) Fit the model using R and examine the residuals. Transform the response if needed to address any problems with normality or constant error variance. If you transform the response, clearly show the residuals from the un-transformed response, and your best transformation, and describe why you chose the transformation you did.

```
library(car)

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

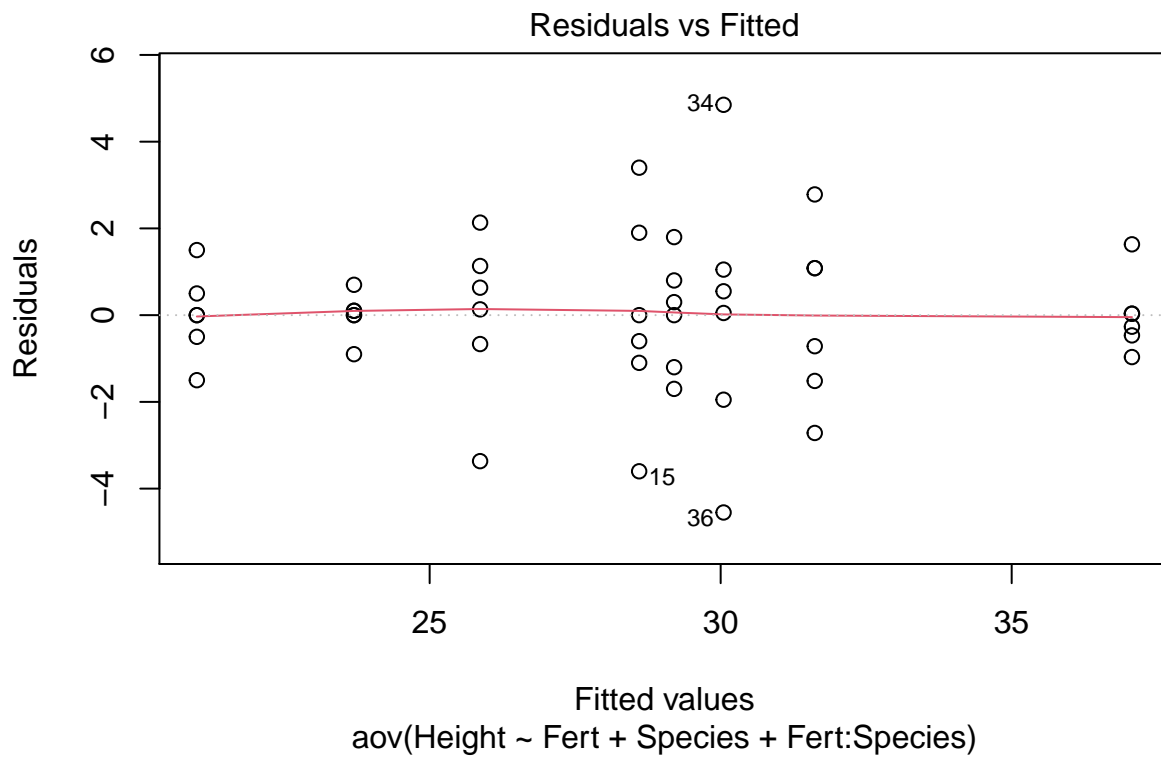
## Loading required package: carData

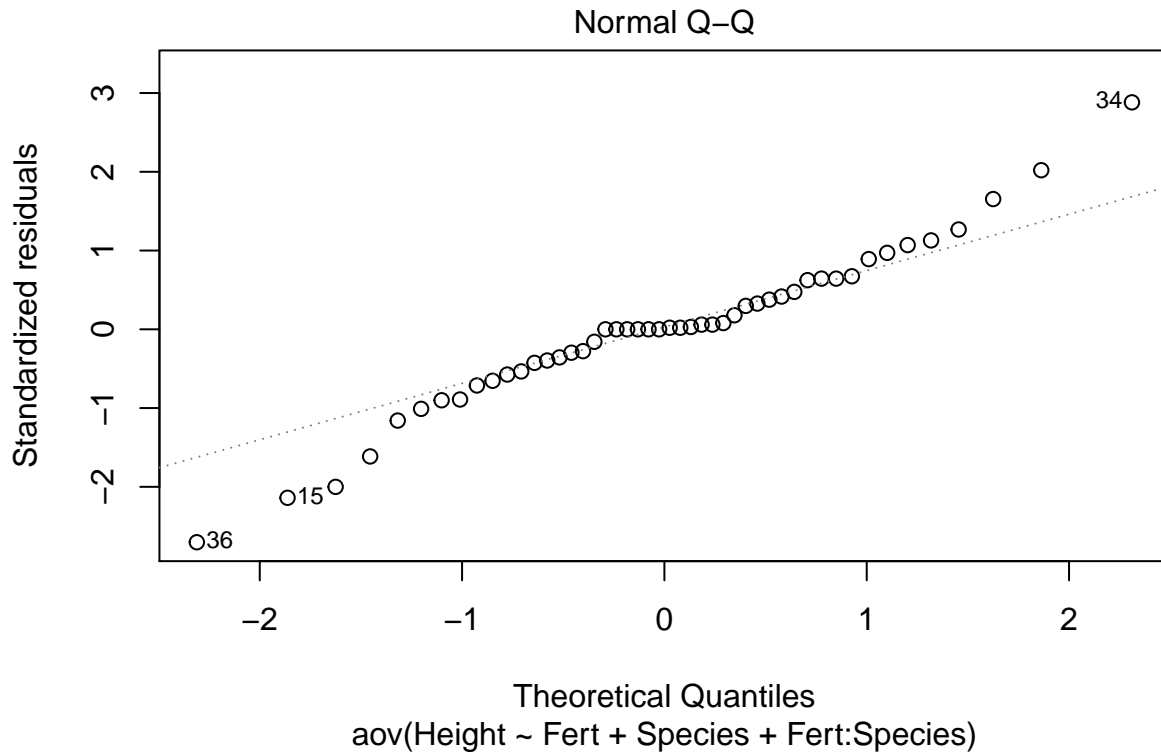
## Warning: package 'carData' was built under R version 4.0.3

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 = c(1, 2))
```

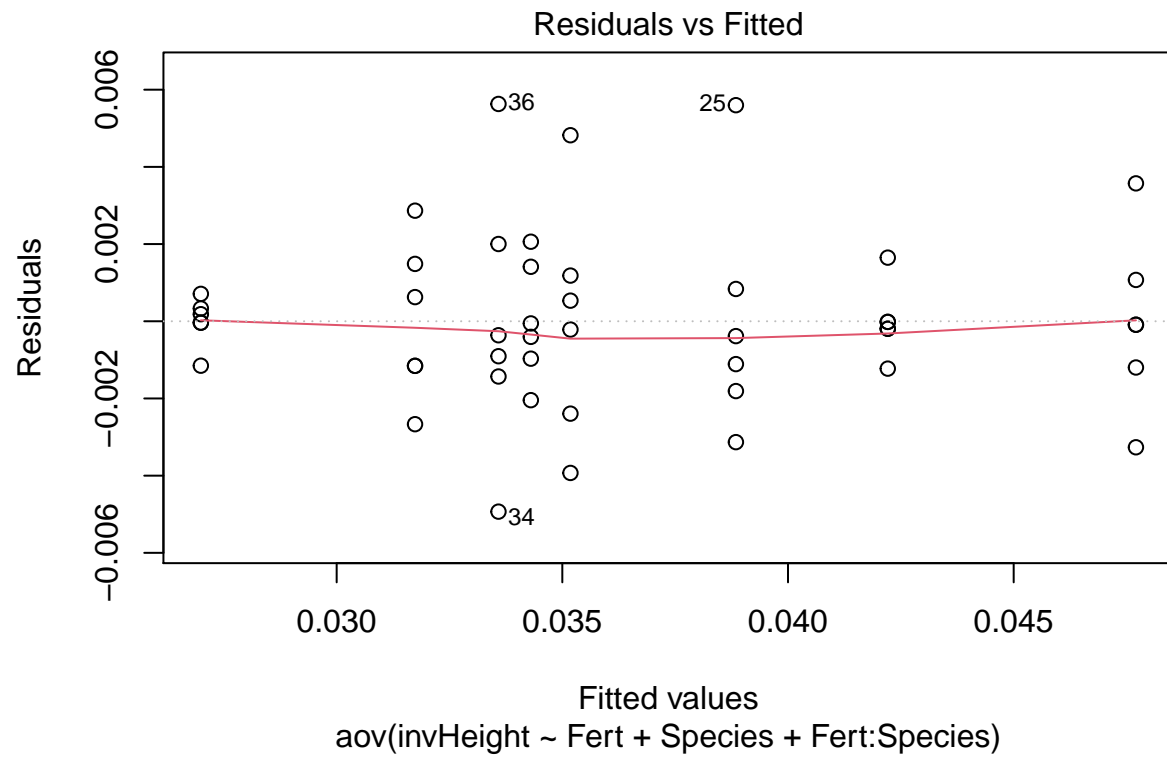


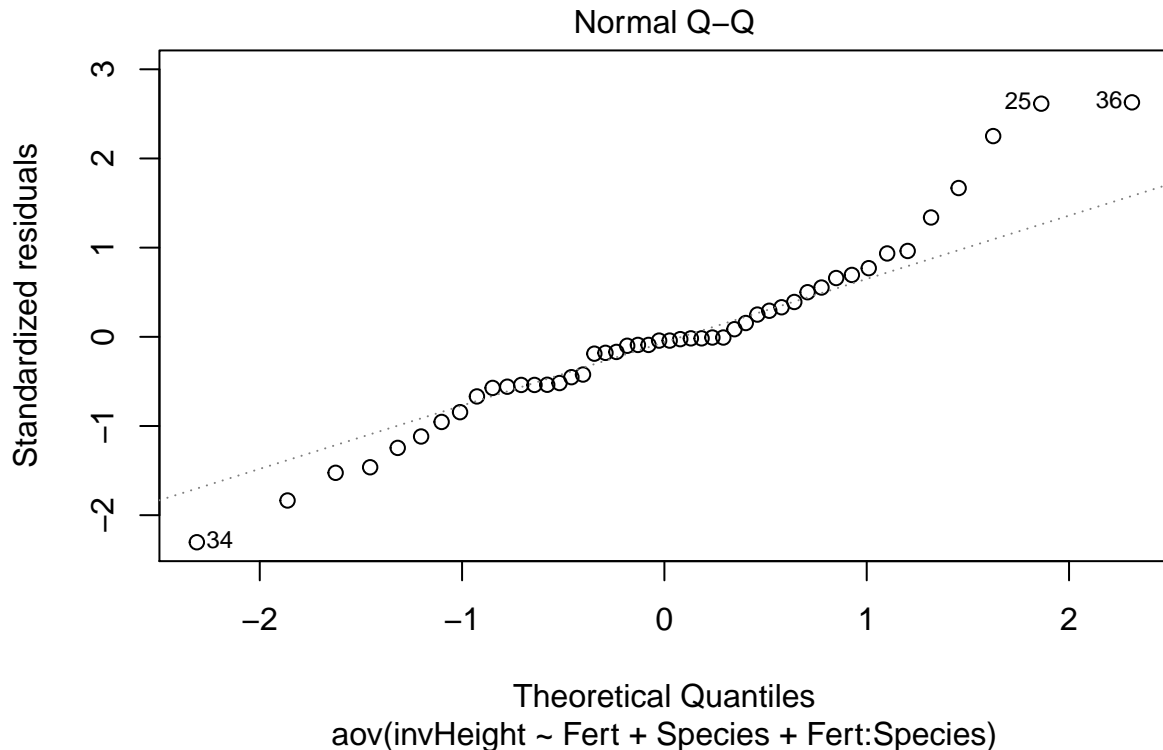


```
df$invHeight = 1 / (df$Height)
model2 = aov(invHeight ~ Fert + Species + Fert:Species, data = df)
Anova(model2, type = "III")
```

```
## 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(model2, which = c(1, 2))
```





All the transformation looks similar, but I think the inverse transformation is the best fit for the model. According to the QQ-plot the residuals are not approximately normal because it seems there isn't a straight line in the QQ-plot. The assumption of constant error variance among treatments is justified, since the residuals are in a square shape.

(c) Describe the effect of species and fertilizer on mean height. This description should use the results of hypothesis tests and p-values as described in class. Discuss any relevant interaction effects, main effects and pairwise differences between treatment means. Provide a plot that shows the means for all combinations of factor levels. Provide R code and output that supports your results.

We first give a hypothesis test with the following $\alpha = 0.05$

$$H_0 : (\alpha\beta)_{ij} = 0, \quad H_a : \text{at least one } (\alpha\beta) \text{ is different}$$

```
model3 = aov(invHeight ~ Fert + Species, data = df)
anova(model3)
```

```
## 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
```

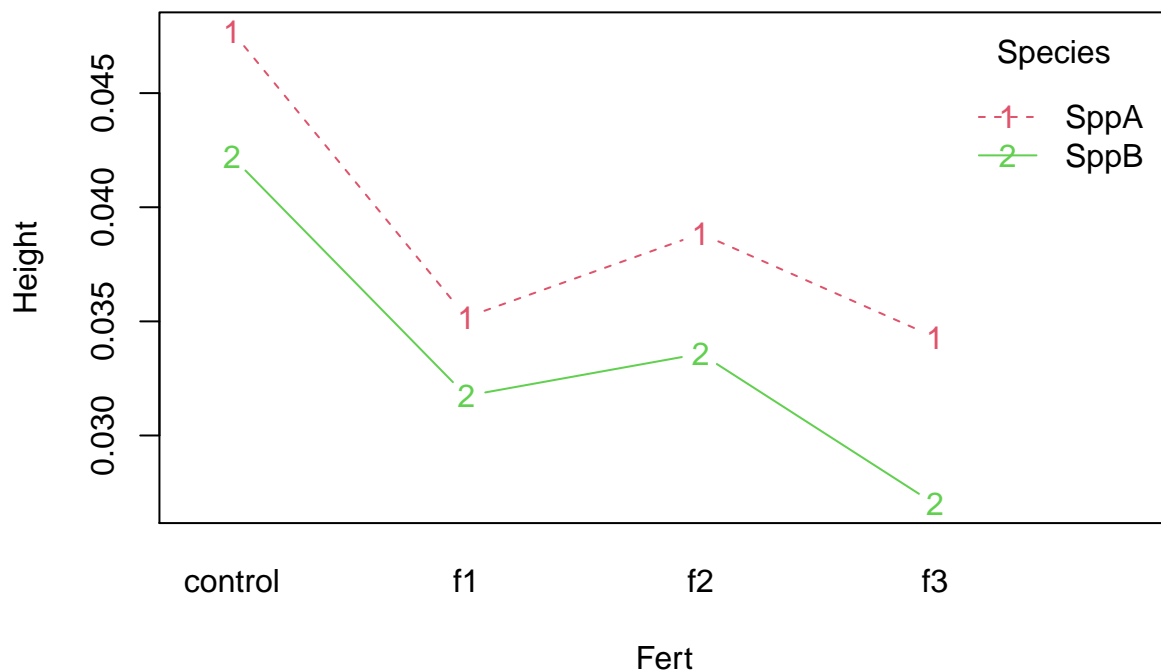
According to the inverse transformation table, we can see the Fert:Species has p-value larger than α , so we fail to reject the null hypothesis, therefore we can conclude that there are no significant interactions.

Also, from the ANOVA table, both variables: Fert & Species has p-value much smaller than 0.05, hence there are significant differences between heights and their removals.

```
interaction.plot(x.factor= df$Fert, trace.factor = df$Species, response = df$invHeight,
                 type = "b", col = 2:3, xlab = "Fert", ylab = "Height",
                 trace.label = "Species" )
library(lsmeans)
```

```
## Loading required package: emmeans
```

```
## The 'lsmeans' package is now basically a front end for 'emmeans'.
## Users are encouraged to switch the rest of the way.
## See help('transition') for more information, including how to
## convert old 'lsmeans' objects and scripts to work with 'emmeans'.
```



```
lsm.fert = lsmeans(model2, ~ Fert)
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
contrast(lsm.fert, 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
```

```
lsm.species = lsmeans(model2, ~ Species)
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
contrast(lsm.species, 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
```

With the output above, we can say that all species and fert seems to be different.

```
library(multcompView)
```

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

```
library(multcomp)
```

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

```
## Loading required package: mvtnorm
```

```
## Loading required package: survival
```

```
## Loading required package: TH.data
```

```
## Warning: package 'TH.data' was built under R version 4.0.3
```

```
## Loading required package: MASS
```

```
##
```

```
## Attaching package: 'TH.data'
```

```
## The following object is masked from 'package:MASS':
```

```
##
```

```
##      geyser
```

```
lsm.inter = lsmeans(model2, ~ Fert:Species)
CLD(lsm.inter)
```

```
## Warning: 'CLD' will be deprecated. Its use is discouraged.
## See '?cld.emmGrid' for an explanation. Use 'pwpp' or 'multcomp::cld' instead.
```

```
## 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
```

Problem 2. Consider the following data, the result of a 2-factor factorial experiment with 5 replications for each combination of Factor A and Factor B. Treatment combinations were assigned at random to the 20 experimental units.

```
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)
df<-data.frame(A=A, B=B, resp=resp)
```

(a) Write out the 2-way complete model for this experiment.

The 2-way complete model is:

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

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

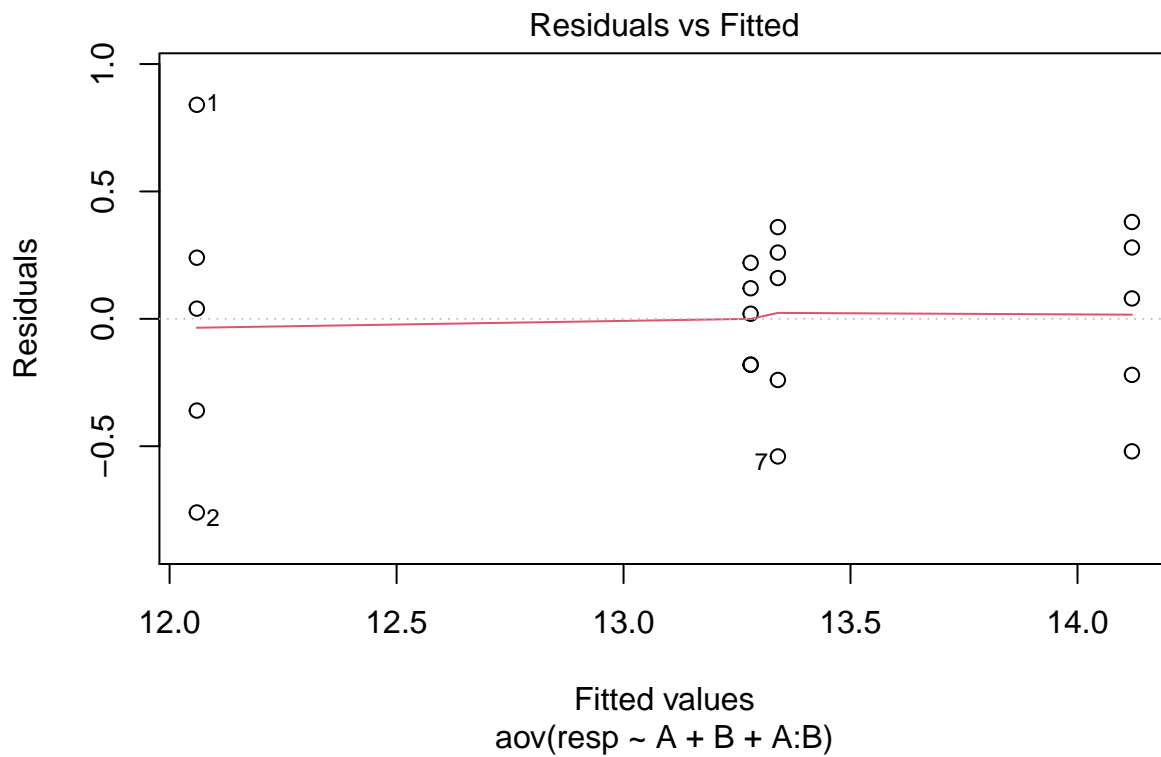
(b) Fit the model using R and examine the residuals. Transform the response if needed to address any problems with normality or constant error variance. If you transform the response, clearly show the residuals from the un-transformed response, and your best transformation, and describe why you chose the transformation you did.

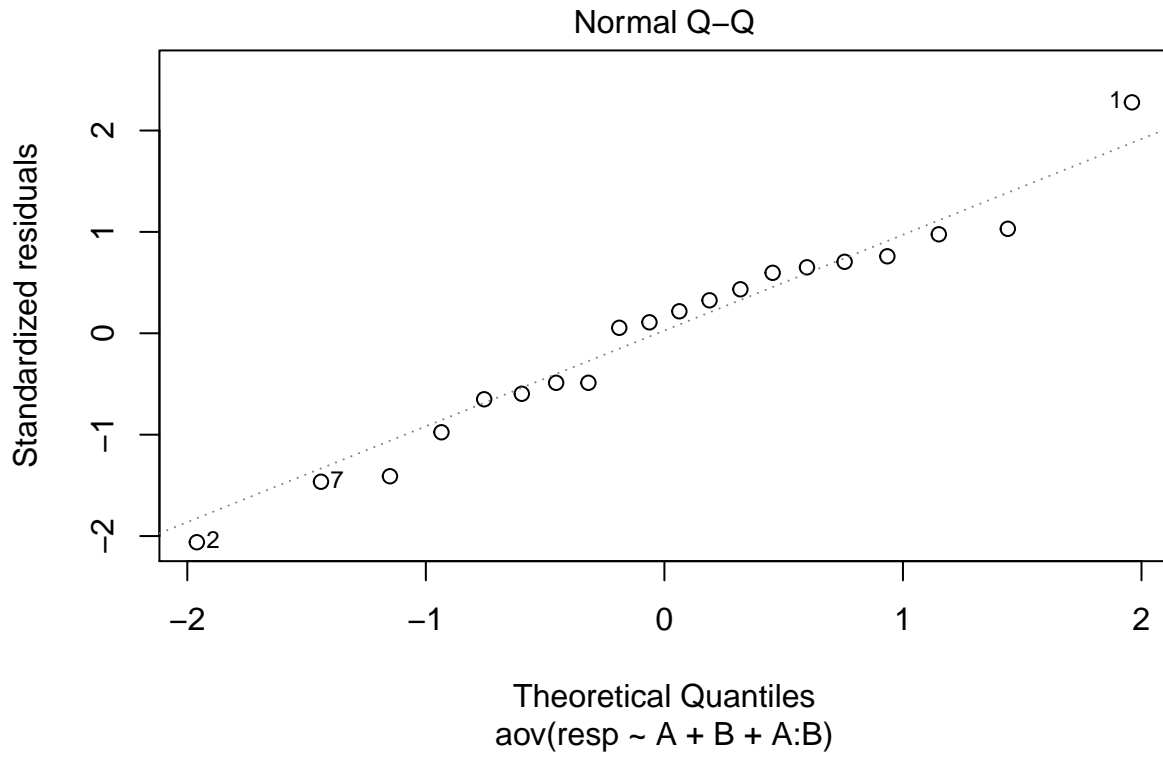
```
model4 = aov(resp ~ A + B + A:B, data = df)
Anova(model4, type = "III")
```



```
## 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(model4, which = c(1, 2))
```

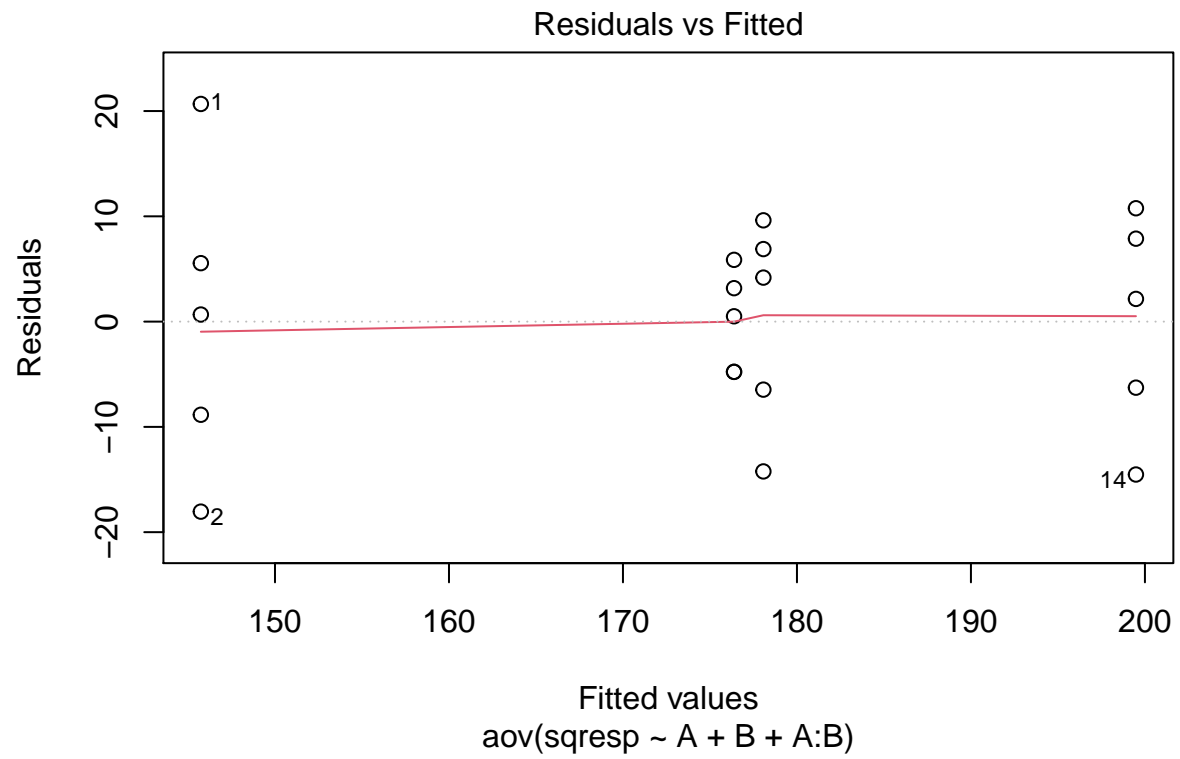


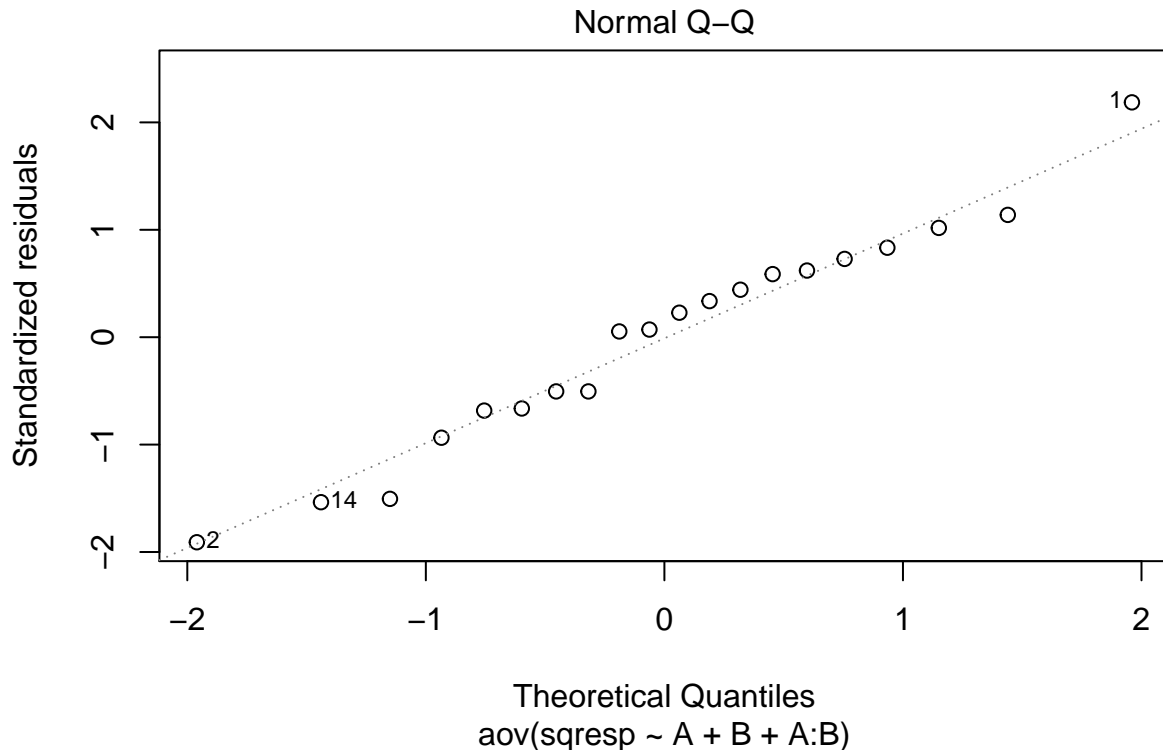


```
df$sqresp = (df$resp)^2
model5 = aov(sqresp ~ A + B + A:B, data = df)
Anova(model5, type = "III")
```

```
## 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(model5, which = c(1, 2))
```





All the transformation looks similar, but I think the square transformation is the best fit for the model. According to the QQ-plot the residuals are approximately normal because it seems there is a straight line in the QQ-plot. The assumption of constant error variance among treatments is justified, since the residuals are in a square shape.

(c) Describe the effect of Factors A and B on mean response. This description should use the results of hypothesis tests and p-values as described in class. Discuss any relevant interaction effects, main effects and pairwise differences between treatment means. Provide a plot that shows the means for all combinations of factor levels. Provide R code and output that supports your results.

We first give a hypothesis test with the following $\alpha = 0.05$

$$H_0 : (\alpha\beta)_{ij} = 0, \quad H_a : \text{at least one } (\alpha\beta) \text{ is different}$$

```
model6 = aov(sqresp ~ A + B, data = df)
anova(model6)
```

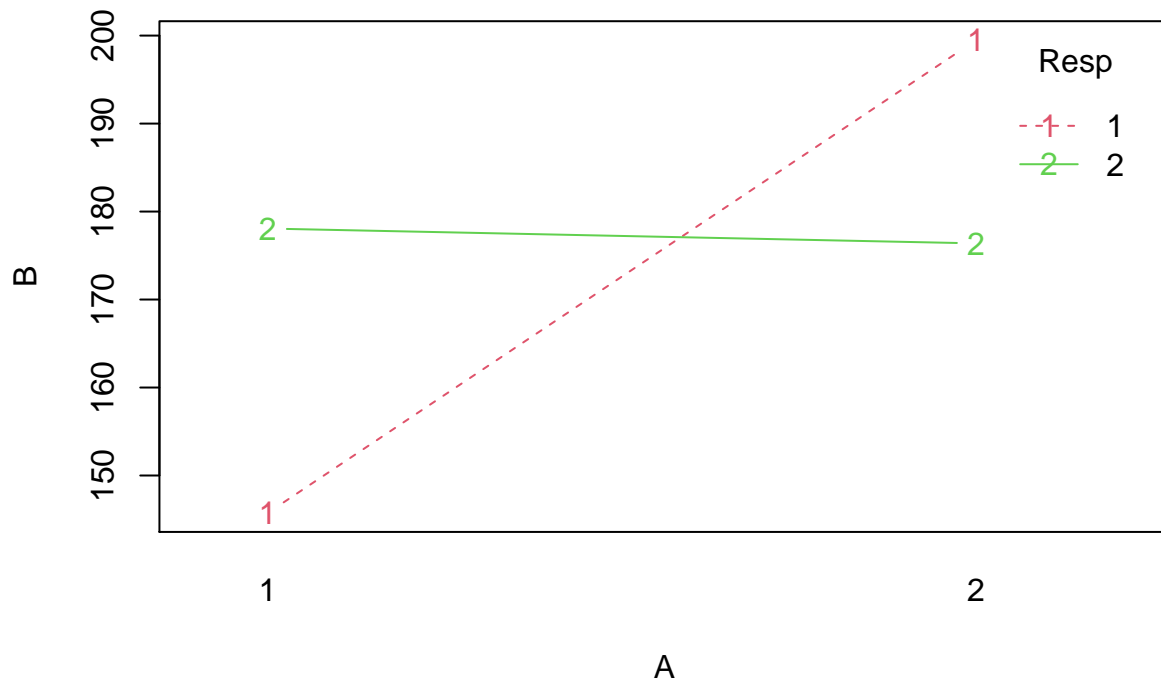
```
## Analysis of Variance Table
##
## Response: sqresp
##          Df Sum Sq Mean Sq F value    Pr(>F)
## A           1 3387.8  3387.8 10.2324 0.005261 **
## B           1  106.5   106.5  0.3218 0.577953
## Residuals 17 5628.5    331.1
```

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

According to the previous table, we see that the p-value of A:B is smaller than α , thus we can reject the null hypothesis, and conclude that there are significant interactions.

Also, from the ANOVA table, A has p-value much smaller than 0.05, hence there is significant differences between heights and its removals.

```
interaction.plot(x.factor= df$A, trace.factor = df$B, response = df$sqresp,
                 type = "b", col = 2:3, xlab = "A", ylab = "B", trace.label = "Resp" )
```



```
lsmA = lsmeans(model5, ~ A)
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
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(model5, ~ B)
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
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
```

With the output above, we can say that A seems to be different.

```
lsm.inter = lsmeans(model5, ~ A:B)
CLD(lsm.inter)
```

```
## Warning: 'CLD' will be deprecated. Its use is discouraged.
## See '?cld.emmGrid' for an explanation. Use 'pwpp' or 'multcomp::cld' instead.
```

```
## 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
```

Problem 3. Under the 2-way complete model, consider the case where Factor A takes values 1 and 2 and Factor B takes the value 1, 2, ..., b. we wish to consider testing the differences between all treatments with A = 1 and all treatments with A = 2. Show that if the interaction $(\alpha\beta)_{ij}$ is significant for all $i = 1, 2$ and all $j = 1, 2, \dots, b$ that $\alpha_1 - \alpha_2$ is not identifiable.

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

and $i = 1, 2, \dots, a$, $j = 1, 2, \dots, b$, $t = 1, 2, \dots, r_{ij}$.

if there is a significant interaction for all $i = 1, 2$ and $j = 1, 2, \dots, b$ then $(\alpha\beta)_{ij} \neq 0$

rewrite the model,

$$Y_{ijt} = \mu + \tau_{ij} + \epsilon_{ijt}$$

$$b_1(\mu + \tau_{11}) + \dots + b_n(\mu + \tau_{1b}) - b_{n1}(\mu + \tau_{21}) - \dots - b_{nn}(\mu + \tau_{2b}) = a_1 = a_2$$

Therefore, we conclude that we can't perform the test for the 2-way complete model for this experiment, since there will always exist $(\alpha\beta)_{ij}$.