

Homework 7

Jiaqi Li

1.-----

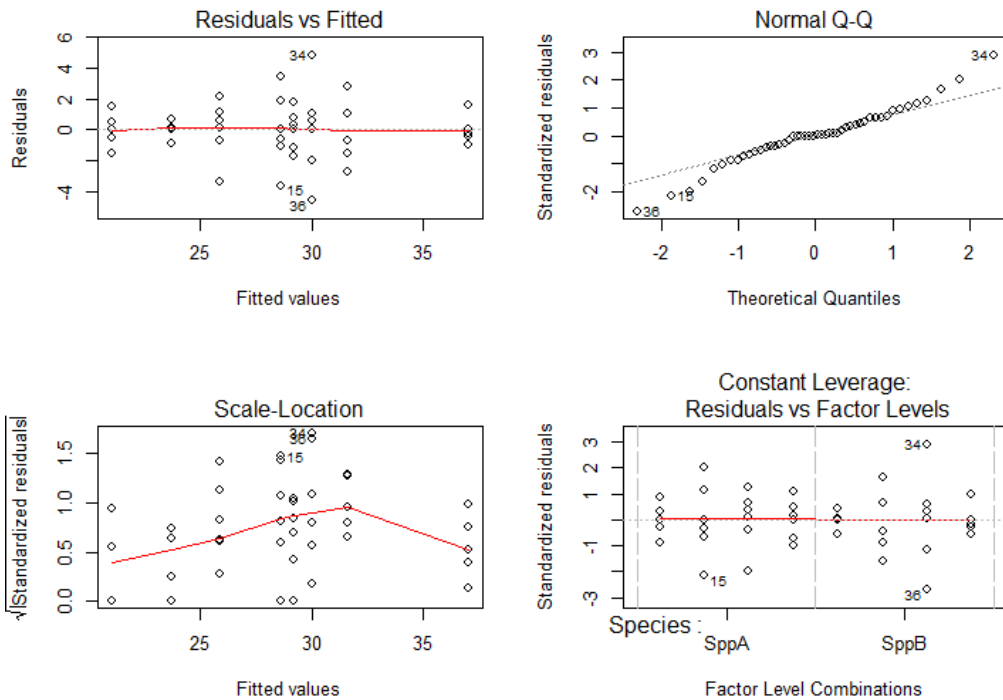
(a)

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

$$\tau_{ij} = \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

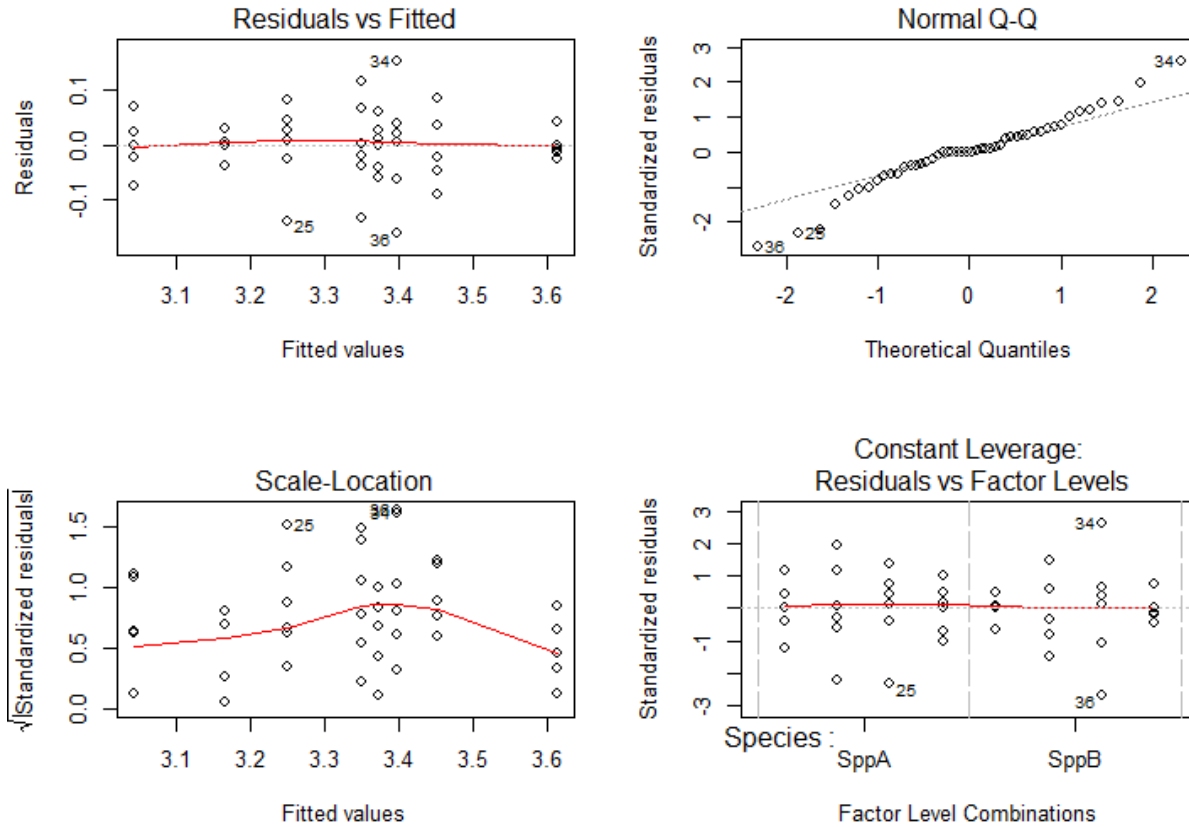
i=control, f1, f2, f3 j=SppA, SppB t=1, 2, 3, 4, 5, 6

(b)



By observing the scatter plot, we conclude that residuals do not have constant variance. By observing the Q-Q plot, we conclude that some points in the plot do not fit the line well, which indicates the violation of normality. Thus, we choose to transform the response.

```
> modelF3<-aov(LogH~Fert+Species+Fert:Species, data=df)
> par(mfrow=c(2,2))
> plot(modelF3)
```



We use the new model (coded as modelF3 in R):

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

$$\tau_{ij} = \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

$$i=\text{control, f1, f2, f3} \quad j=\text{SppA, SppB} \quad t=1, 2, 3, 4, 5, 6$$

In the model, we observe that residuals still have non-constant variance and data do not fit the line well enough in the Q-Q plot, which indicates violation of normality. However, this transformation is better than the un-transformed model. Thus, we use this transformation as our model to make further study and we need to be careful with the p-value our hypothesis tests calculating for us because assumptions for error terms are not satisfied.

(c)

```
> anova(modelF3)
Analysis of Variance Table

Response: LogH
      Df Sum Sq Mean Sq F value    Pr(>F)
Species 1  0.27920  0.27920  66.6636 4.733e-10 ***
Fert     3  0.99025  0.33008  78.8131 < 2.2e-16 ***
Species:Fert 3  0.03319  0.01106   2.6417  0.06241 .
Residuals 40  0.16753  0.00419
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

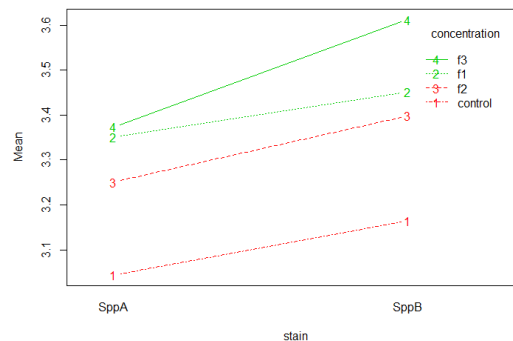
For interaction:

H_0 : there is no interaction effect on the mean height.

H_1 : there are some interaction effects on the mean height.

$p\text{-value} = 0.06241 > \alpha = 0.05$

Thus, we fail to reject null hypothesis and conclude that there is no interaction effect on the mean height.



Since we suspect the accuracy of the p-value calculated by this model because assumptions for error terms are not satisfied, we also want to see how the information about the interaction term that the original model shows us:

```
> anova(modelF1)
Analysis of Variance Table

Response: Height
      Df Sum Sq Mean Sq F value    Pr(>F)
Fert     3  745.44  248.479  73.0982 2.766e-16 ***
Species 1  236.74  236.741  69.6450 2.707e-10 ***
Fert:Species 3  50.58   16.861   4.9603  0.005081 **
Residuals 40  135.97    3.399
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

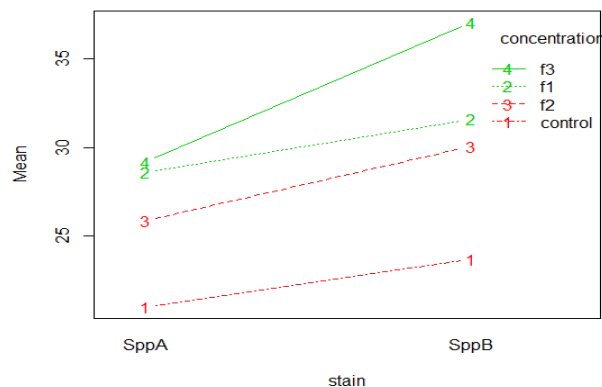
The original model shows us that p-value of the interaction term is $0.005081 < \alpha = 0.05$. Thus we have the following hypothesis test:

H_0 : there is no interaction effect on the mean height.

H_1 : there are some interaction effects on the mean height.

p-value = $0.005081 < \alpha = 0.05$

Thus, we reject null hypothesis and conclude that there is some interaction effects on the mean height.



```
> contrast(lsmInter, method="pairwise")
contrast      estimate      SE df t.ratio p.value
control,SppA effect -7.387500 0.7040767 40 -10.492 <.0001
f1,SppA effect      0.212500 0.7040767 40  0.302 0.7644
f2,SppA effect     -2.520833 0.7040767 40 -3.580 0.0015
f3,SppA effect      0.812500 0.7040767 40  1.154 0.2918
control,SppB effect -4.687500 0.7040767 40 -6.658 <.0001
f1,SppB effect      3.229167 0.7040767 40  4.586 0.0001
f2,SppB effect      1.662500 0.7040767 40  2.361 0.0309
f3,SppB effect      8.679167 0.7040767 40 12.327 <.0001
```

P value adjustment: fdr method for 8 tests

To observe easier, we do the following:

```
> cld(lsmInter,alpha=0.05)
Fert  Species  lsmean      SE df lower.CL upper.CL .group
control SppA    21.00000 0.7526896 40 19.47876 22.52124    1
control SppB    23.70000 0.7526896 40 22.17876 25.22124   12
f2      SppA    25.86667 0.7526896 40 24.34542 27.38791   23
f1      SppA    28.60000 0.7526896 40 27.07876 30.12124   34
f3      SppA    29.20000 0.7526896 40 27.67876 30.72124   34
f2      SppB    30.05000 0.7526896 40 28.52876 31.57124    4
f1      SppB    31.61667 0.7526896 40 30.09542 33.13791    4
f3      SppB    37.06667 0.7526896 40 35.54542 38.58791    5
```

Confidence level used: 0.95

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

From the table above, we see that all p-values of the contrast are smaller than $\alpha = 0.05$. Thus we conclude that (f=fertilizer, Spp=species of plants):

- Control group with SppA has less effect on the mean height than f1 with SppA or SppB, f2 with SppA or SppB, f3 with SppA or SppB have.
- Control group with SppB has less effect on the mean height than f1 with SppA or SppB, f2 with SppB, f3 with SppA or SppB have.
- f2 with SppA has less effect on the mean height than f1 with SppB, f2 with SppB, f3 with SppB have.
- f1 with SppA has less effect on the mean height than f3 with SppB has.
- f3 with SppA has less effect on the mean height than f3 with SppB has.
- f2 with SppB has less effect on the mean height than f3 with SppB has.
- f1 with SppA has less effect on the mean height than f3 with SppB has.

No other significant pairwise differences in mean height with respect to different fertilizers used by different species of plants.

For "Species":

H_0 : species has no effect on mean height.

H_1 : species has effects on mean height.

$$p\text{-value} = 4.733 \times 10^{-10} < \alpha = 0.05$$

Thus, we reject null hypothesis and conclude that species has effects on the mean height.

```
> contrast(lsms, method="pairwise")
contrast      estimate      SE df t.ratio p.value
SppA - SppB -0.1525338 0.01868192 40  -8.165  <.0001
```

Results are averaged over the levels of: Fert

$$p\text{-value} < 0.0001 < \alpha = 0.05$$

Thus, species SppB has more effect on the mean height than SppA has.

For "Fert":

H_0 : fertilizers have no effect on mean response.

H_1 : at least one fertilizer has effect on mean height.

$$p\text{-value} < 2.2 \times 10^{-16} < \alpha = 0.05$$

Thus, we reject null hypothesis and conclude that at least one fertilizer has effect on the mean height.

```
> contrast(lsmF, method="pairwise")
contrast      estimate      SE df t.ratio p.value
control - f1 -0.29674566 0.02642023 40 -11.232  <.0001
control - f2 -0.22001603 0.02642023 40  -8.328  <.0001
control - f3 -0.38849410 0.02642023 40 -14.704  <.0001
f1 - f2      0.07672963 0.02642023 40   2.904  0.0292
f1 - f3     -0.09174844 0.02642023 40  -3.473  0.0066
f2 - f3     -0.16847806 0.02642023 40  -6.377  <.0001
```

Results are averaged over the levels of: Species

P value adjustment: tukey method for comparing a family of 4 estimates

To observe easier, we can do the following:

```
> cld(lsmF, alpha=0.05)
```

Fert	lsmean	SE	df	lower.CL	upper.CL	.group
control	3.104426	0.01868192	40	3.066669	3.142184	1
f2	3.324442	0.01868192	40	3.286685	3.362200	2
f1	3.401172	0.01868192	40	3.363415	3.438930	3
f3	3.492921	0.01868192	40	3.455163	3.530678	4

From the table above, we see that all p-values of the contrast are smaller than $\alpha = 0.05$. Thus we conclude that:

- f1 has more effect on the mean height than the control group has.
- f2 has more effect on the mean height than the control group has.
- f3 has more effect on the mean height than the control group has.
- f1 has more effect on the mean height than f2 has.
- f3 has more effect on the mean height than f1 has.
- f3 has more effect on the mean height than f2 has.
- No other significant pairwise differences in mean height with respect to different fertilizers.

2.-----

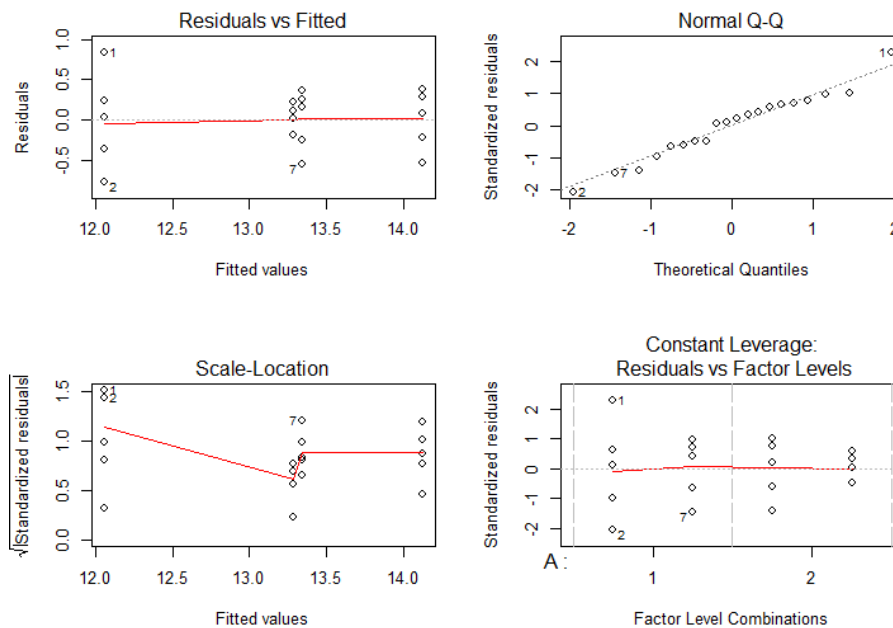
(a)

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

$\tau_{ij} = \alpha_i + \beta_j + (\alpha\beta)_{ij}$

$i=1, 2 \quad j=1, 2 \quad t=1, 2, 3, 4, 5$

(b)



In the model, we observe that residuals do not have constant variance, but data fit the line well in the Q-Q plot, which indicates normality. In this model, assumptions of normality is satisfied but the constant variance is not satisfied. Then, we try to use transformation to satisfy error terms' assumption of constant variance. However, none of the transformations makes any improvements. Thus, we use the un-transformed model for further study and we need to be careful with the p-value our hypothesis tests calculating for us because assumptions for error terms are not all satisfied. (Attempts of trying different transformations are included in R code at the last few pages and the outcomes are not shown in this report.)

(c)

```
> anova(model11)
Analysis of Variance Table

Response: resp
      Df Sum Sq Mean Sq F value    Pr(>F)
A       1  5.000   5.000 29.4118 5.632e-05 ***
B       1  0.242   0.242  1.4235  0.2502
A:B     1  5.618   5.618 33.0471 2.991e-05 ***
Residuals 16  2.720   0.170
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

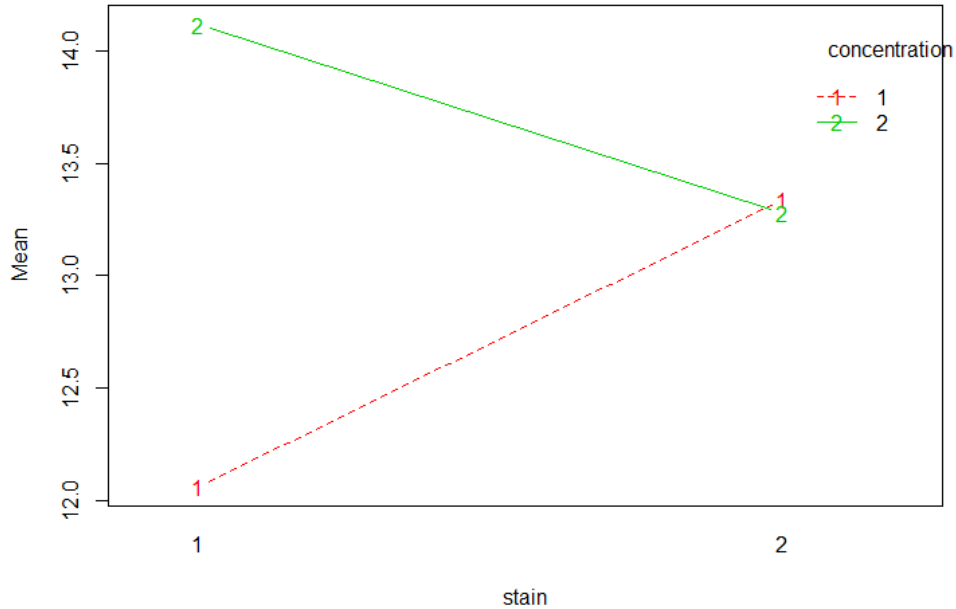
For intersection:

H_0 : there is no interaction effect on the mean response.

H_1 : there are some interaction effects on the mean response.

p-value = $2.991 \times 10^{-5} < \alpha = 0.05$

Thus, we reject null hypothesis and conclude that there is some interaction effects on the mean response.



```
> lsmAB=lsmeans(model1, ~ A:B )
> contrast(lsmAB,method="pairwise")
```

contrast	estimate	SE	df	t.ratio	p.value
1,1 - 2,1	-2.06	0.2607681	16	-7.900	<.0001
1,1 - 1,2	-1.28	0.2607681	16	-4.909	0.0008
1,1 - 2,2	-1.22	0.2607681	16	-4.678	0.0013
2,1 - 1,2	0.78	0.2607681	16	2.991	0.0389
2,1 - 2,2	0.84	0.2607681	16	3.221	0.0247
1,2 - 2,2	0.06	0.2607681	16	0.230	0.9955

P value adjustment: tukey method for comparing a family of 4 estimates

To observe easier, we can do the following:

```
> cld(lsmAB, alpha=0.05)
```

A	B	lsmean	SE	df	lower.CL	upper.CL	.group
1	1	12.06	0.1843909	16	11.66911	12.45089	1
2	2	13.28	0.1843909	16	12.88911	13.67089	2
1	2	13.34	0.1843909	16	12.94911	13.73089	2
2	1	14.12	0.1843909	16	13.72911	14.51089	3

Confidence level used: 0.95

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

We can conclude that:

- "A1 B1" has significantly lower mean than "A2 B2", "A1 B2", "A2 B1".
- "A2 B2" has significantly lower mean than "A2 B1".

- “A1 B2” has significantly lower mean than “A2 B1”.
- No other comparisons are significantly different than zero.

For A:

H_0 : A has no effect on mean response. H_1 : A has effects on mean response.

p-value = $5.632 \times 10^{-5} < \alpha = 0.05$

Thus, we reject null hypothesis and conclude that A has effects on the mean response.

```
> contrast(lsmA, method="pairwise")
contrast estimate      SE df t.ratio p.value
1 - 2          -1 0.1843909 16  -5.423  0.0001
```

Results are averaged over the levels of: B

p-value = 0.0001 < $\alpha = 0.05$

Thus, species A2 has more effect on the mean response than A1 has.

For B:

H_0 : B has no effect on mean response. H_1 : B has effects on mean response.

p-value = 0.2502 > $\alpha = 0.05$

Thus, we fail to reject null hypothesis and conclude that B does not have any effect on the mean response.

R code:

```
install.packages("lsmeans")
install.packages("multcompView")
install.packages("car")

library(lsmeans)
library(multcompView)
library(car)
options(contrasts = c("contr.sum", "contr.poly"))

#1-----
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)

modelF1<-aov(Height~Fert+Species+Fert:Species, data=df)
par(mfrow=c(2,2))
plot(modelF1)
anova(modelF1)

df$LogH<-log(Height)
df$SqrtH<-sqrt(Height)
df$InvH<-1/Height

modelF2<-aov(SqrtH~Fert+Species+Fert:Species, data=df)
par(mfrow=c(2,2))
plot(modelF2)

modelF3<-aov(LogH~Fert+Species+Fert:Species, data=df)
par(mfrow=c(2,2))
plot(modelF3)

modelF4<-aov(InvH~Fert+Species+Fert:Species, data=df)
par(mfrow=c(2,2))
plot(modelF4)
```

```
anova(modelF3)
```

```
interaction.plot(x.factor = df$Species, trace.factor = df$Fert,  
  response = df$LogH, type = "b", col = 2:3,  
  xlab = "stain", ylab = "Mean", trace.label = "concentration")
```

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

```
anova(modelF1)  
lsmInter=lsmeans(modelF1, ~Fert:Species)  
contrast(lsmInter, method="pairwise")  
cld(lsmInter,alpha=0.05)
```

```
lsmS=lsmeans(modelF3, ~Species)  
contrast(lsmS, method="pairwise")
```

```
lsmF=lsmeans(modelF3, ~Fert)  
contrast(lsmF, method="pairwise")  
cld(lsmF, alpha=0.05)
```

```
#2-----
```

```
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)
```

```
model1<-aov(resp~A+B+A:B, data=df2)  
par(mfrow=c(2,2))  
plot(model1)
```

```
df2$LogR<-log(resp)  
df2$SqrtR<-sqrt(resp)  
df2$InvR<-1/resp
```

```
model2<-aov(SqrtR~A+B+A:B, data=df2)  
par(mfrow=c(2,2))  
plot(model2)
```

```

model3<-aov(LogR~A+B+A:B, data=df2)
par(mfrow=c(2,2))
plot(model3)

model4<-aov(InvR~A+B+A:B, data=df2)
par(mfrow=c(2,2))
plot(model4)

anova(model1)

interaction.plot(x.factor = df2$B, trace.factor = df2$A,
                 response = df2$resp, type = "b", col = 2:3,
                 xlab = "stain", ylab = "Mean", trace.label = "concentration")

lsmAB=lsmeans(model1, ~ A:B )
contrast(lsmAB,method="pairwise")
cld(lsmAB, alpha=0.05)

lsmA=lsmeans(model1, ~A)
contrast(lsmA, method="pairwise")

```