

Kamala Dadashova

Problem 1: Find the missing values in the following ANOVA table.

Solution: In this problem,

1) we are given

$$MSA = 50, \quad df_A = 1$$

. Using class note,

$$MSA = \frac{SSA}{df_A}$$

where $df_A = I - 1$, we find that

$$SSA = 50 \times 1 = 50.$$

2) we have

$$SSA = 50, \quad SSB = 80, \quad SSAB = 30, \quad SST = 172.$$

Using the identity, $SST = SSA + SSB + SSAB + SSE$, we find that $SSE = 172 - 50 - 80 - 30 = 12$.

3) we have given

$$df_A = I - 1 = 1, \quad df_{AB} = (I - 1)(J - 1) = 2$$

implies that $df_B = J - 1 = 2$.

4) we know that $df_A = I - 1$, $df_B = J - 1$, $df_{AB} = (I - 1)(J - 1)$, $df_E = IJ(K - 1)$. If we add them $df_A + df_B + df_{AB} + df_E = I - 1 + J - 1 + IJ - I - J + 1 + IJK - IJ = IJK - 1 = df_T$. Using this identity, we find that $df_E = 17 - 1 - 2 - 2 = 12$.

5) using $MSB = \frac{SSB}{df_B} = 80/2 = 40$, $MSAB = \frac{SSAB}{df_{AB}} = 30/2 = 15$, $MSE = \frac{SSE}{df_E} = 12/12 = 1$.

6) using class note $F_A = \frac{MSA}{MSE} = 50/1 = 50$, $F_B = \frac{MSB}{MSE} = 40/1 = 40$ and $F_{AB} = \frac{MSAB}{MSE} = 15/1 = 15$.

The final table is :

Source	SS	df	MS	F
A	50	1	50	50
B	80	2	40	40
AB	30	2	15	15
ERROR	12	12	1	
TOTAL	172	17		

Table 1: Anova Table

Problem 2: Researchers investigated the effects of cyclic loading and environmental conditions on fatigue crack growth at 22 MPa stress for a particular material. The crack growth rate was recorded for three loading frequencies (10, 1, and 0.1 Hz) in three different environments (Air, Water, and Salt Water). The data from the experiment is provided in growth.csv

- a) Fit an ANOVA model using Growth as the response and Environment and Frequency as fixed main effects; ignore the interaction.

Solution: We have fixed effects model,

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J$$

where we want $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$. Our hypotheses are:

$$H_{0A} : \alpha_1 = \alpha_2 = \alpha_3 = 0, \quad H_{aA} : \text{at least one of the } \alpha_i \text{ s is different than 0.}$$

$$H_{0B} : \beta_1 = \beta_2 = \beta_3 = 0, \quad H_{aB} : \text{at least one of the } \beta_j \text{ s is different than 0.}$$

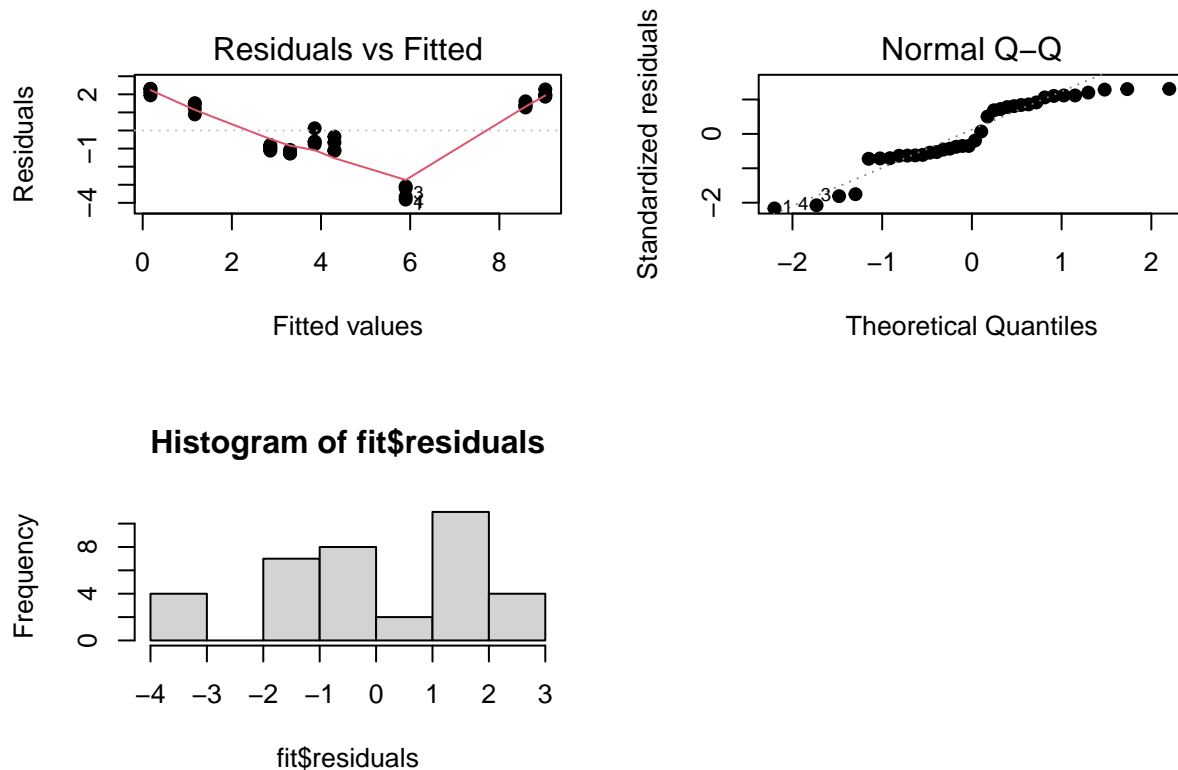
where A represents environment factor and B represents frequency factor. We firstly call the excel file **growth.csv** and then fit an Anova model using Growth as the response and Environment and Frequency as fixed main effects.

```
ex2=read.csv("growth.csv")
fit <- aov(Growth~as.factor(Frequency)+as.factor(Environment),data=`ex2`)
```

- b) Plot the residual diagnostics and identify issues with the residual plots if any.

Solution:

```
par(mfrow=c(2,2))
plot(fit,1:2,pch=19)
hist(fit$residuals)
```



We observe that model residuals demonstrate a nonlinear relationship with fitted values, thus perhaps adding interaction term could improve fitting. Regarding QQ plot, we see some major deviations from normality. We also provide a histogram of residuals and it clearly shows that residuals are not normally distributed.

- c) Use the model results to display an ANOVA table and coefficient estimates for the fit model. Is there a significant difference in any of the mean values for the different Environments or Frequencies? If so, conduct the appropriate test to determine which means are different.

Solution:

```
summary(fit) # ANOVA table
```

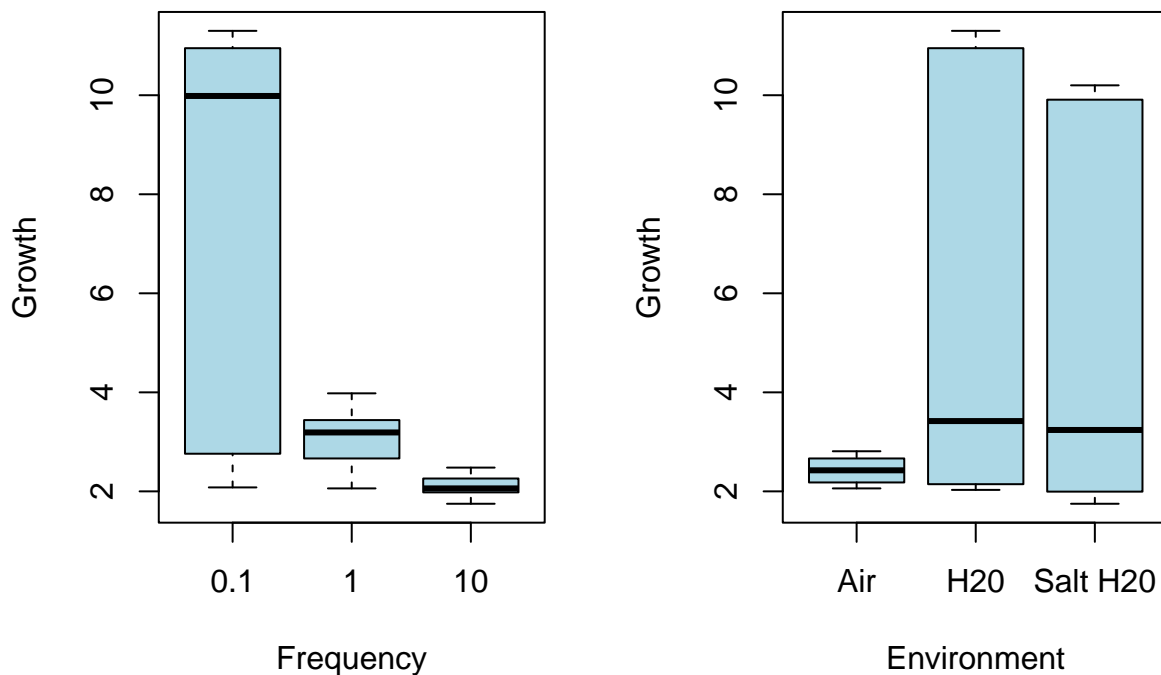
```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## as.factor(Frequency)    2  224.55   112.28   31.142 3.84e-08 ***
## as.factor(Environment)  2   68.92    34.46    9.559 0.000584 ***
## Residuals              31  111.76     3.61
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fit$coefficients # means
```

```
##              (Intercept)          as.factor(Frequency)1
##              5.899167              -4.730000
##      as.factor(Frequency)10      as.factor(Environment)H20
##              -5.725000              3.131667
## as.factor(Environment)Salt H20
##              2.688333
```

For both environment and frequency, there is a strong evidence (p values, $3.84 \times 10^{-8} < 0.05$, $0.000584 < 0.05$) that the null hypotheses defined above for all $\alpha_i = 0$ and all $\beta_j = 0$ is not true and conclude that at least one mean is different than others. At the level 0.05 significance level, all main effects are statistically significant. We would like to know which means are exactly different. To do that we initially provide a comparison box-plot which is followed by applying Tukey's test,

```
attach(ex2)
par(mfrow=c(1,2)) # box plot to view difference in means
boxplot(Growth~Frequency,col="light blue",xlab="Frequency",ylab="Growth")
boxplot(Growth~Environment,col="light blue",xlab="Environment",ylab="Growth")
```



```
TukeyHSD(fit,ordered=T)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
## factor levels have been ordered
##
## Fit: aov(formula = Growth ~ as.factor(Frequency) + as.factor(Environment), data = ex2)
##
## $`as.factor(Frequency)`
##      diff      lwr      upr      p adj
## 1-10    0.995 -0.9128324 2.902832 0.4148590
## 0.1-10  5.725  3.8171676 7.632832 0.0000001
## 0.1-1    4.730  2.8221676 6.637832 0.0000027
##
## $`as.factor(Environment)`
##      diff      lwr      upr      p adj
## Salt H20-Air 2.6883333 0.7805009 4.596166 0.0043294
## H20-Air      3.1316667 1.2238342 5.039499 0.0009288
## H20-Salt H20 0.4433333 -1.4644991 2.351166 0.8359677
```

According to Tukey's procedure for multiple comparison, we say that there is a significant difference between mean of 0.1 and others, 1,10. Because $p = 0.0000001 < 0.05$ for the pair 0.1-10 and $p = 0.0000027 < 0.05$ for the pair 0.1-1. Nonetheless, there is no significant difference between the mean of 1 and 10 since $p = 0.4148590 > 0.05$. Regarding the second factor, we say that there is a significant difference between mean of Air and mean of other two elements, Salt H20 and H20, since $p = 0.0043294 < 0.05$ and $p = 0.0009288 < 0.05$. However, there is no significant difference between the means of the H20 and Salt H20 because $p = 0.8359677 > 0.05$.

We can observe the same result by looking at box-plot of both factors.

d) Repeat part (a), but this time include the interaction.

Solution: Now, we have additional interaction term and our model is

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, K > 1,$$

where we want $\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$. Our hypotheses are:

$$H_{0A} : \alpha_1 = \alpha_2 = \alpha_3 = 0, \quad H_{aA} : \text{at least one of the } \alpha_i \text{ s is different than 0.}$$

$$H_{0B} : \beta_1 = \beta_2 = \beta_3 = 0, \quad H_{aB} : \text{at least one of the } \beta_j \text{ s is different than 0.}$$

$$H_{0AB} : \gamma_{11} = \dots = \gamma_{33} = 0, \quad H_{0AB} : \text{at least one is not equal to 0.}$$

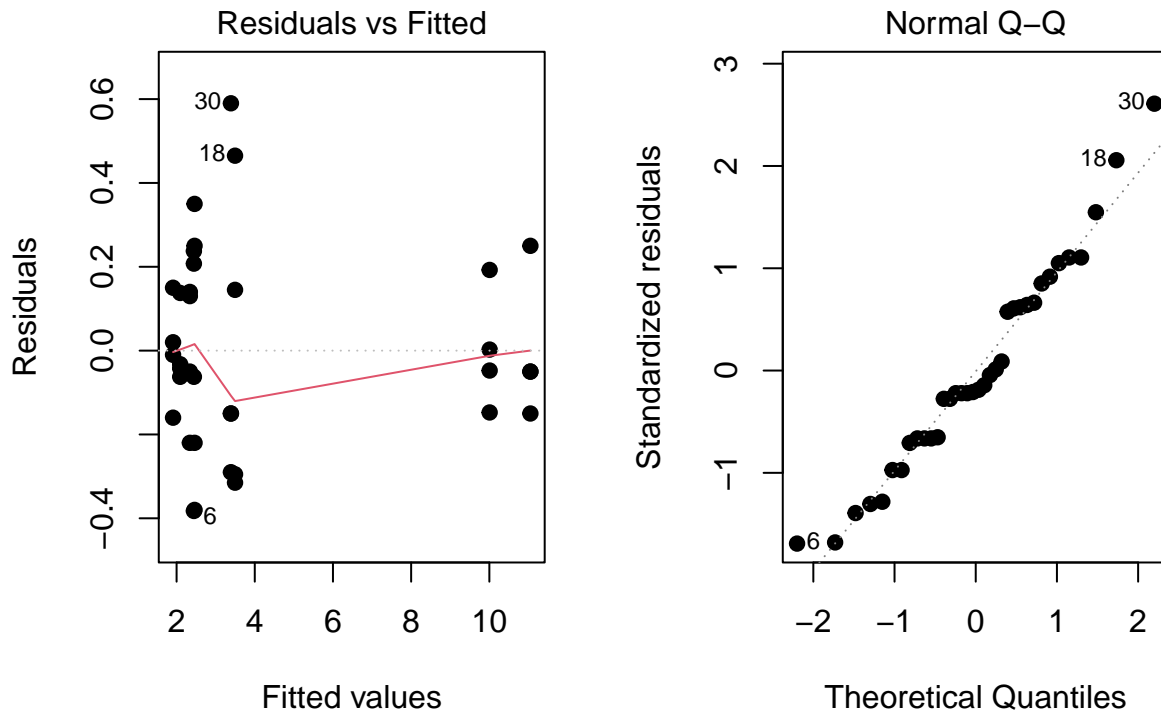
Next, we fit an ANOVA model considering interaction.

```
fit <- aov(Growth~as.factor(Frequency)*as.factor(Environment),data=ex2)
```

e) Plot the residual diagnostics and identify issues with the residual plots if any. Based on the diagnostic plots, would you say the model with interaction is better than the one without? Why or why not?

Solution:

```
par(mfrow=c(1,2))
plot(fit,1:2,pch=19) # residual diagnostics
```



In QQ plot, the pattern is sufficiently linear that there should be no concern about lack of normality. The plot of residuals against fitted values shows residuals have relatively constant spread across treatment levels. Assumptions for the error term in model hold.

We prefer the model involving interaction because the assumption for the error term in the model holds true. However, we do not see the correctness of these assumptions if we ignore interaction term as we saw in part (b).

- f) Use the model results to display an ANOVA table and coefficient estimates for the fit model. Is there a significant interaction? If so, conduct the appropriate test to determine which means are different. Should you include the differences in means for the main effects? Why or why not?

Solution:

```
anova(fit)

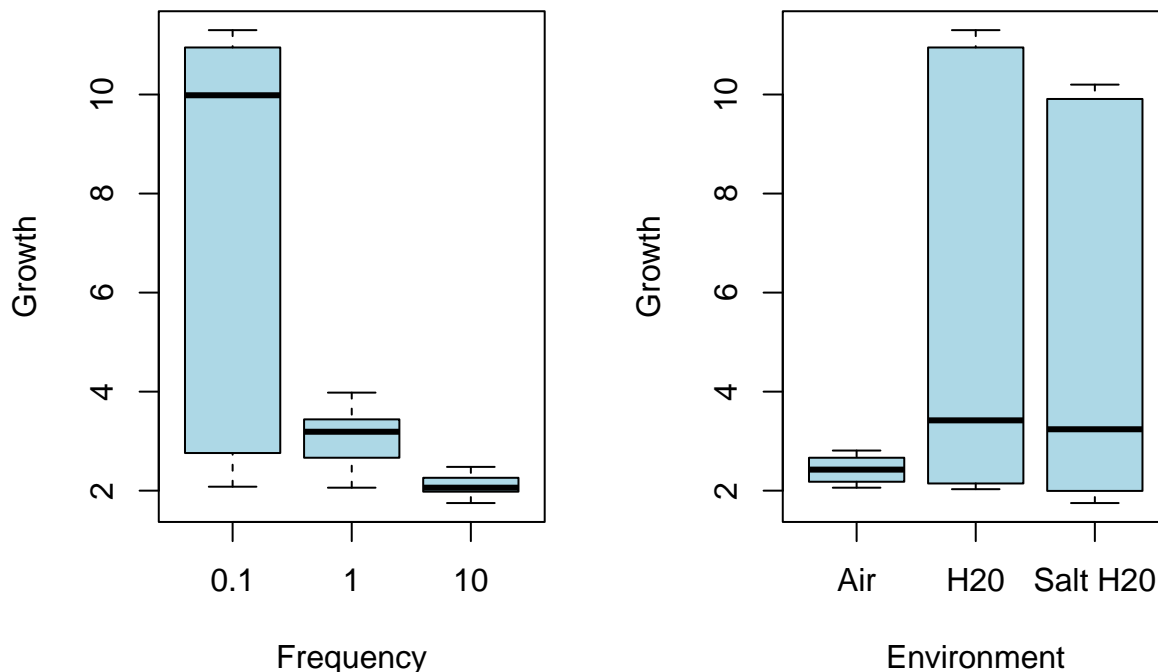
## Analysis of Variance Table
##
## Response: Growth
##
##          Df Sum Sq Mean Sq F value
## as.factor(Frequency)      2  224.554   112.277  1646.72
## as.factor(Environment)     2   68.924    34.462   505.44
## as.factor(Frequency):as.factor(Environment)  4  109.923    27.481   403.05
## Residuals                  27    1.841     0.068
##
##          Pr(>F)
## as.factor(Frequency) < 2.2e-16 ***
## as.factor(Environment) < 2.2e-16 ***
## as.factor(Frequency):as.factor(Environment) < 2.2e-16 ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

fit$coefficients
```

```
##                                (Intercept)
##                                2.4600
##                                as.factor(Frequency)1
##                                -0.0175
##                                as.factor(Frequency)10
##                                -0.1200
##                                as.factor(Environment)H2O
##                                8.5900
##                                as.factor(Environment)Salt H2O
##                                7.5475
##                                as.factor(Frequency)1:as.factor(Environment)H2O
##                                -7.5375
##                                as.factor(Frequency)10:as.factor(Environment)H2O
##                                -8.8375
##                                as.factor(Frequency)1:as.factor(Environment)Salt H2O
##                                -6.6000
##                                as.factor(Frequency)10:as.factor(Environment)Salt H2O
##                                -7.9775
```

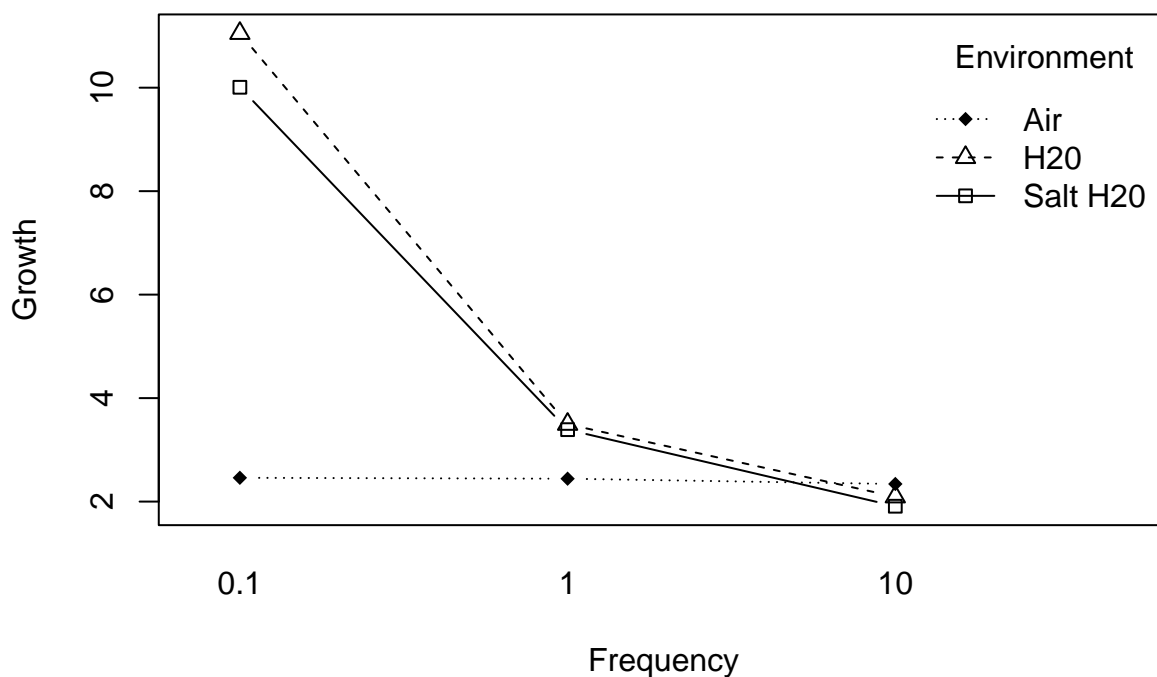
Based on ANOVA table, we see that for the p value of both factors Frequency and Environment is $2.2 \times 10^{-16} < 0.05$, so we should reject the null hypotheses H_{0A} and H_{0B} . Regarding interaction, we also see that $p = 2.2 \times 10^{-16} < 0.05$ implying we should reject the null hypothesis H_{0AB} . At the level 0.05 significance level, all main and interaction effects are statistically significant. To verify our outcomes for both main effects and interactions, we apply the following tests.

```
par(mfrow=c(1,2)) # box plot to view difference in means
boxplot(Growth~Frequency,col="light blue",xlab="Frequency",ylab="Growth")
boxplot(Growth~Environment,col="light blue",xlab="Environment",ylab="Growth")
```



```
par(mfrow=c(1,1)) # interaction plot
with(ex2, (interaction.plot(Frequency, Environment, Growth, type = "b", pch = c(18,24,22),leg.bty = "n"
```

Interaction Plot of Frequency with Environment



```
## NULL
```

```
TukeyHSD(fit,ordered=T)
```

```
## Tukey multiple comparisons of means
```

```
## 95% family-wise confidence level
```

```
## factor levels have been ordered
```

```
##
```

```
## Fit: aov(formula = Growth ~ as.factor(Frequency) * as.factor(Environment), data = ex2)
```

```
##
```

```
## $`as.factor(Frequency)`
```

```
##      diff      lwr      upr p adj
```

```
## 1-10  0.995 0.730692 1.259308    0
```

```
## 0.1-10 5.725 5.460692 5.989308    0
```

```
## 0.1-1  4.730 4.465692 4.994308    0
```

```
##
```

```
## $`as.factor(Environment)`
```

```
##      diff      lwr      upr      p adj
```

```
## Salt H2O-Air 2.6883333 2.4240254 2.9526413 0.0000000
```

```
## H2O-Air      3.1316667 2.8673587 3.3959746 0.0000000
```

```
## H2O-Salt H2O 0.4433333 0.1790254 0.7076413 0.0008227
```

```
##
```

```
## $`as.factor(Frequency):as.factor(Environment)`
```

```
##      diff      lwr      upr      p adj
```

```
## 10:H2O-10:Salt H2O 0.1825 -0.43875107 0.8037511 0.9838526
```

```
## 10:Air-10:Salt H2O 0.4300 -0.19125107 1.0512511 0.3606528
```

```
## 1:Air-10:Salt H2O 0.5325 -0.08875107 1.1537511 0.1374309
```

```
## 0.1:Air-10:Salt H2O 0.5500 -0.07125107 1.1712511 0.1138210
```

```
## 1:Salt H2O-10:Salt H2O 1.4800 0.85874893 2.1012511 0.0000004
```

```
## 1:H2O-10:Salt H2O 1.5850 0.96374893 2.2062511 0.0000001
```

```
## 0.1:Salt H2O-10:Salt H2O 8.0975 7.47624893 8.7187511 0.0000000
## 0.1:H2O-10:Salt H2O 9.1400 8.51874893 9.7612511 0.0000000
## 10:Air-10:H2O 0.2475 -0.37375107 0.8687511 0.9096918
## 1:Air-10:H2O 0.3500 -0.27125107 0.9712511 0.6222106
## 0.1:Air-10:H2O 0.3675 -0.25375107 0.9887511 0.5622686
## 1:Salt H2O-10:H2O 1.2975 0.67624893 1.9187511 0.0000048
## 1:H2O-10:H2O 1.4025 0.78124893 2.0237511 0.0000012
## 0.1:Salt H2O-10:H2O 7.9150 7.29374893 8.5362511 0.0000000
## 0.1:H2O-10:H2O 8.9575 8.33624893 9.5787511 0.0000000
## 1:Air-10:Air 0.1025 -0.51875107 0.7237511 0.9996954
## 0.1:Air-10:Air 0.1200 -0.50125107 0.7412511 0.9990412
## 1:Salt H2O-10:Air 1.0500 0.42874893 1.6712511 0.0001477
## 1:H2O-10:Air 1.1550 0.53374893 1.7762511 0.0000339
## 0.1:Salt H2O-10:Air 7.6675 7.04624893 8.2887511 0.0000000
## 0.1:H2O-10:Air 8.7100 8.08874893 9.3312511 0.0000000
## 0.1:Air-1:Air 0.0175 -0.60375107 0.6387511 1.0000000
## 1:Salt H2O-1:Air 0.9475 0.32624893 1.5687511 0.0006261
## 1:H2O-1:Air 1.0525 0.43124893 1.6737511 0.0001426
## 0.1:Salt H2O-1:Air 7.5650 6.94374893 8.1862511 0.0000000
## 0.1:H2O-1:Air 8.6075 7.98624893 9.2287511 0.0000000
## 1:Salt H2O-0.1:Air 0.9300 0.30874893 1.5512511 0.0008010
## 1:H2O-0.1:Air 1.0350 0.41374893 1.6562511 0.0001824
## 0.1:Salt H2O-0.1:Air 7.5475 6.92624893 8.1687511 0.0000000
## 0.1:H2O-0.1:Air 8.5900 7.96874893 9.2112511 0.0000000
## 1:H2O-1:Salt H2O 0.1050 -0.51625107 0.7262511 0.9996363
## 0.1:Salt H2O-1:Salt H2O 6.6175 5.99624893 7.2387511 0.0000000
## 0.1:H2O-1:Salt H2O 7.6600 7.03874893 8.2812511 0.0000000
## 0.1:Salt H2O-1:H2O 6.5125 5.89124893 7.1337511 0.0000000
## 0.1:H2O-1:H2O 7.5550 6.93374893 8.1762511 0.0000000
## 0.1:H2O-0.1:Salt H2O 1.0425 0.42124893 1.6637511 0.0001641
```

Now, we see that for both main factors, frequency and environment, p values are less than 0.05. Thus, we reject the null hypotheses H_{0A} , H_{0B} and conclude that ALL mean values for the different Environments or Frequencies are different. The box-plot also illustrates it. When it comes to interaction term, we see that while a few of p values are greater than 0.05, the most of them are less than 0.05. Hence, we reject the null hypotheses H_{0AB} and conclude that there is a significant difference between interactions. If there were no interactions between the factors, we would expect all the lines to follow the same general curve. However, in the provided interaction plot in which we also observe that the effect of Air on mean response is constant with 0.1, 1, 10, while the pattern for H2O and Salt H2O is very different. Thus, it verifies Tukey's result. If there is a significant interaction, then we should ignore the two sets of hypotheses for the main effects. A significant interaction tells us that the change in the true average response for a level of Factor A depends on the level of Factor B. The effect of simultaneous changes cannot be determined by examining the main effects separately. So, we should include them.

```
detach(ex2)
```

Problem 3: The file weather.csv contains 77 observations of relative humidity (RH) and air temperature in deg C (Temp) collected at a local weather station.

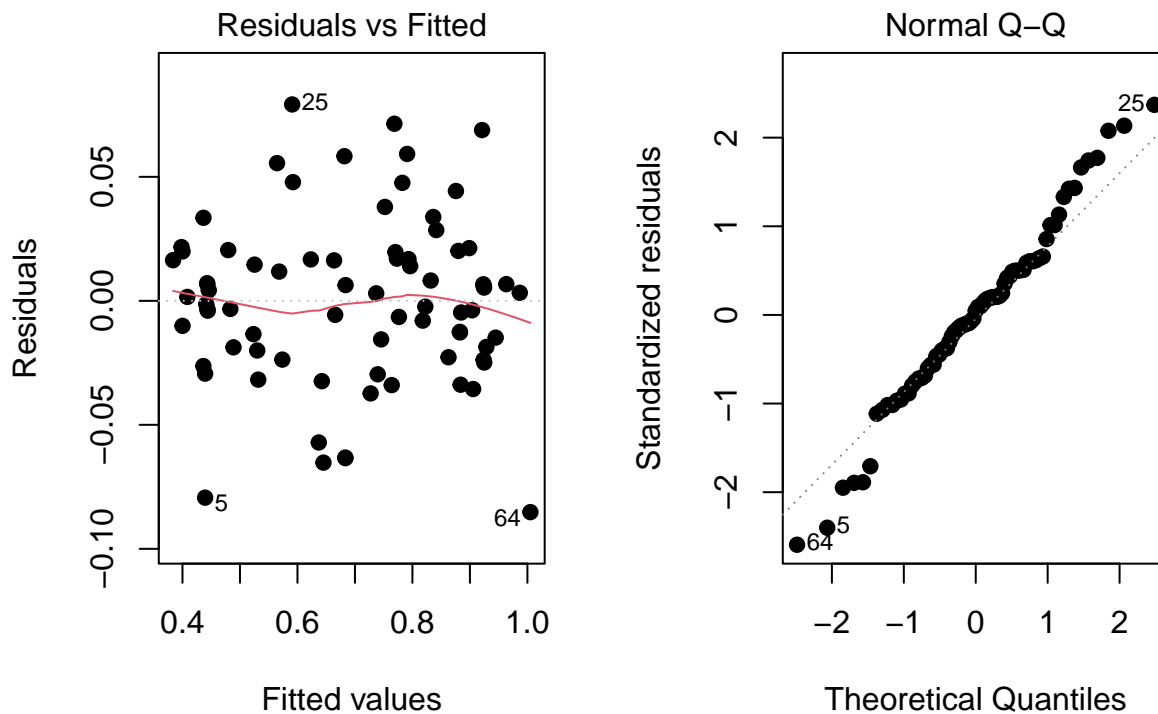
- Fit a model using RH as the response and Temp as the predictor; plot the diagnostics. Identify any issues with the residual plots. From the summary, would you conclude there a significant relationship between RH and Temp (justify your answer using output from the fit summary)?

Solution:

```
ex3=read.csv("weather.csv")
```



```
attach(ex3)
fit <- lm(Humidity~ Temperature..C.,ex3)
par(mfrow=c(1,2))
plot(fit,1:2,pch=19) # residual diagnostics
```



```
summary(fit) # fit summary
```

```
##
## Call:
## lm(formula = Humidity ~ Temperature..C., data = ex3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.085199 -0.020013  0.001582  0.017017  0.079190
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.6693619  0.0205807   81.11  <2e-16 ***
## Temperature..C. -0.0395074  0.0008203  -48.16  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0337 on 75 degrees of freedom
## Multiple R-squared:  0.9687, Adjusted R-squared:  0.9683
## F-statistic: 2319 on 1 and 75 DF,  p-value: < 2.2e-16
cor(Humidity,Temperature..C.)

## [1] -0.9842145
```

Residuals have constant variance, but have short tail. Since consequences of short tailed distribution are not serious, will proceed with current model. From the summary, we see that $p = 2.2e - 16$, so there is a

significant linear relationship. We also find correlation coefficients which is equal to -0.9842145, so there is highly negative linear relationship.

- b) Identify the estimates for intercept and slope; is there sufficient evidence to conclude they are both different from zero (justify your answer)? Compute 95% confidence intervals for both estimates. How much of the variability in RH can be explained by Temp? What is the interpretation of the slope estimate?

Solution:

```
fit$coefficients
```

```
##      (Intercept) Temperature..C.
##      1.6693619      -0.0395074
```

So, estimates for intercept is 1.6693619 and for slope is -0.0395074. From summary table, we see p values for both estimates is $2e - 16 < 0.05$. Thus, we conclude that there is no strong evidence that mean values of estimates are equal to zero.

```
confint(fit)
```

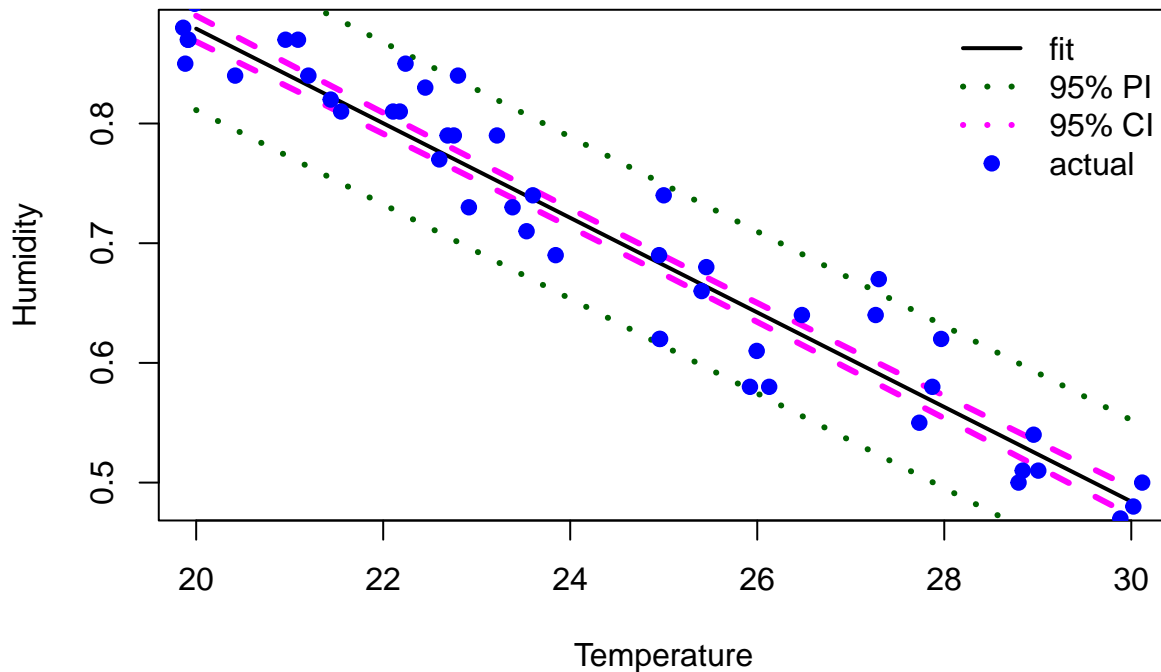
```
##              2.5 %      97.5 %
## (Intercept)    1.62836307  1.71036074
## Temperature..C. -0.04114155 -0.03787324
```

Since zero is not included in both confidence intervals, [1.62836307, 1.71036074] for intercept and [-0.04114155, -0.03787324], we say there is a sufficient evidence to conclude they are both different from zero. $R^2 = 0.9687$, so the Temp explains 96.87% of the variation in RH. The relationship is negative because the coefficient estimate is negative. If the temperature increases by 1 degree, we predict the level of the RH will decrease by approximately 0.0395074 units.

- c) Create a grid of potential Temp values ranging from 20 to 30 deg C in increments of at most 0.01 deg C and predict values of RH for those values of Temp; also calculate values for the 95% confidence interval and the 95% prediction interval. Plot the predicted values versus Temp. Overlay the confidence and prediction limits using dashed and dotted lines, respectively; also overlay the actual RH values as points. Does the 95% prediction interval seem reasonable?

Solution:

```
grid <- data.frame(seq(20, 30, by = 0.01))
colnames(grid)="Temperature..C."
muRh <- predict(fit,new=grid,interval="confidence")
Ypred <- predict(fit,new=grid,interval="prediction")
plot(grid$Temperature..C.,Ypred[,1],type="l",lwd=2,xlab="Temperature",
      ylab="Humidity")
lines(grid$Temperature..C.,Ypred[,2],lty=3, col = "darkgreen", lwd = 3)
lines(grid$Temperature..C.,Ypred[,3],lty=3, col = "darkgreen", lwd = 3)
lines(grid$Temperature..C.,muRh[,2],lty=2, col="magenta", lwd = 3)
lines(grid$Temperature..C.,muRh[,3],lty=2, col = "magenta", lwd = 3)
points(ex3$Temperature..C.,ex3$Humidity,pch=19,col="blue")
legend("topright",legend=c("fit", "95% PI", "95% CI", "actual"),
      lwd=c(2,3,3,NA),pch=c(NA,NA,NA, 19),lty=c(1,3,3, NA),
      col=c("black", "darkgreen", "magenta", "blue"),bty="n")
```

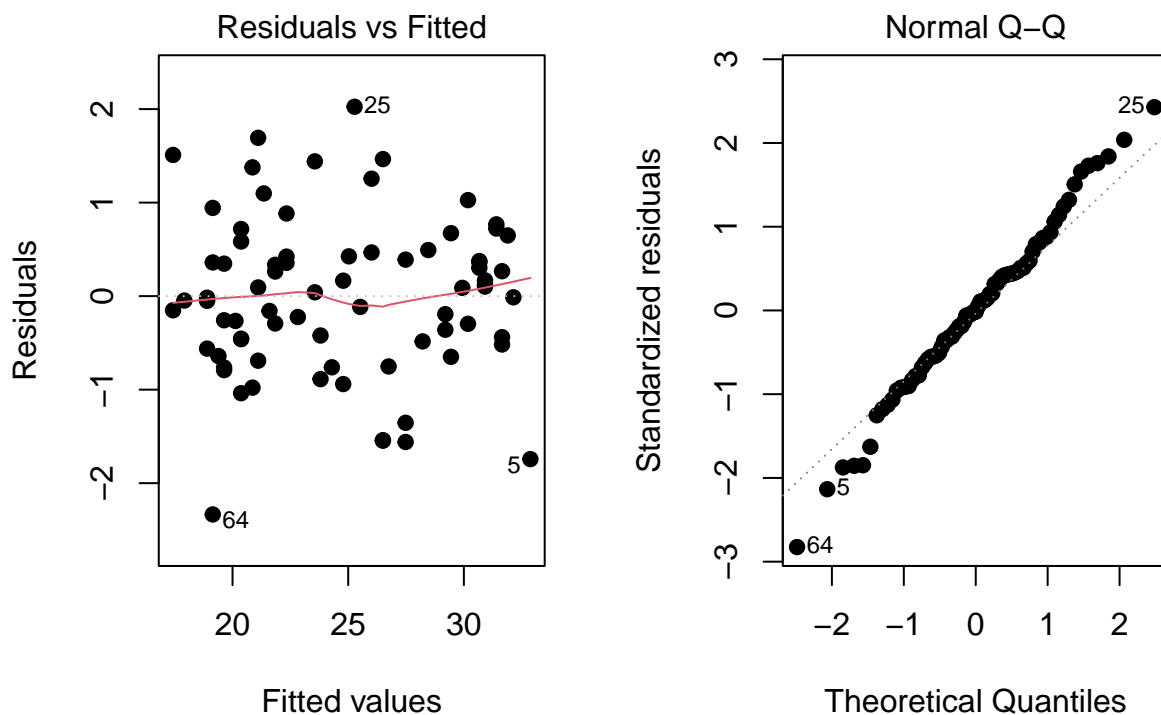


Since 95% prediction interval shown in plot covers almost all of the actual data and wider than confidence interval, it seems reasonable.

- d) Fit a second model to predict Temp from RH. Compare the summary to that of the model you fit in part (a). What elements of the output are the same? Why?

Solution:

```
fit <- lm(Temperature..C.~Humidity,ex3)
par(mfrow=c(1,2))
plot(fit,1:2,pch=19) # residual diagnostics
```



```
summary(fit) # fit summary

##
## Call:
## lm(formula = Temperature..C. ~ Humidity, data = ex3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.3344 -0.4842 -0.0128  0.4255  2.0247
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  41.7029     0.3668  113.69  <2e-16 ***
## Humidity    -24.5189     0.5091  -48.16  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8396 on 75 degrees of freedom
## Multiple R-squared:  0.9687, Adjusted R-squared:  0.9683
## F-statistic: 2319 on 1 and 75 DF,  p-value: < 2.2e-16

cor(Humidity, Temperature..C.)

## [1] -0.9842145

detach(ex3)
```

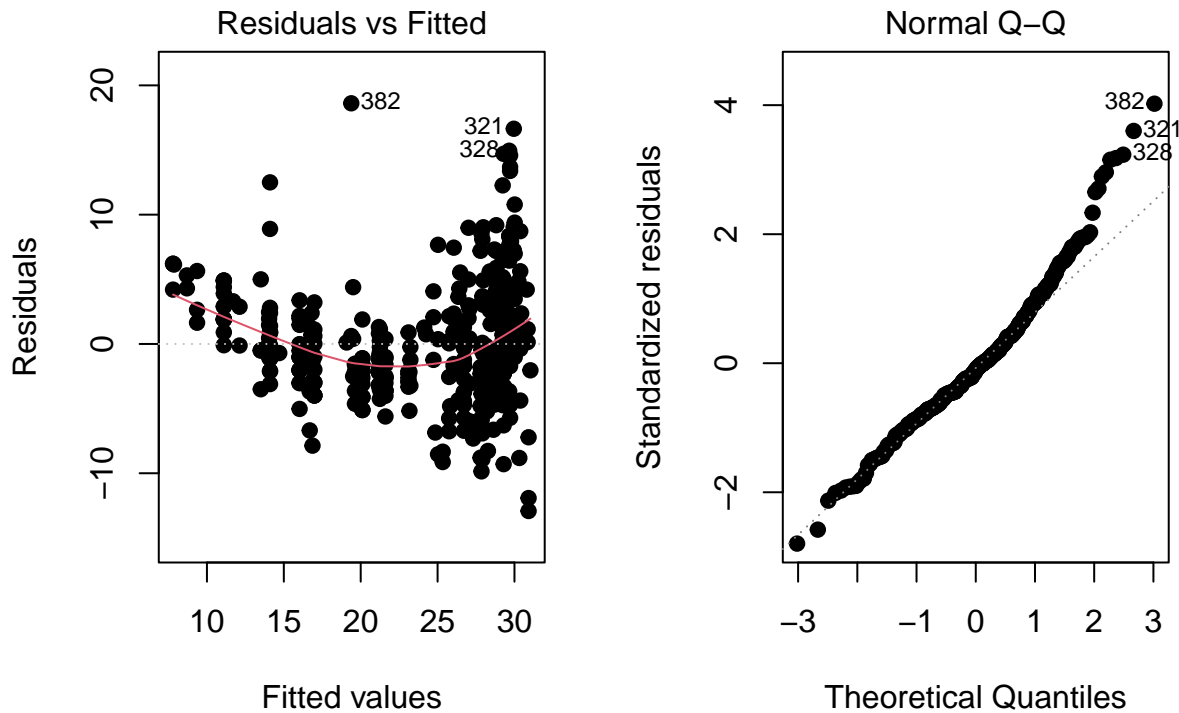
We observe that R^2 , F-statistic, p-value, p value for estimates, adjusted R^2 , degree of freedoms are still same value but other values in summary table is different than the respective ones in summary table in part(a). The reason why they are still same is that the aim of linear regression is to find a linear relationship between Y and X. Hence, $Y = \beta_0 + \beta_1 X + \varepsilon$ is equivalent to $X = \alpha_0 + \alpha_1 Y + \epsilon = -\frac{\beta_0}{\beta_1} + \frac{1}{\beta_1} Y - \frac{1}{\beta_1} \varepsilon$.

Problem 4: In this problem, we examine the relationship between the response mpg and the predictor displacement in the auto.csv data set.

- a) Fit the model to predict mpg with displacement. Print the fit summary and plot the diagnostics. From the summary, is there a significant linear relationship between displacement and mpg? How much of the variation in mpg can be explained by displacement?

Solution:

```
ex4=read.csv("auto.csv")
attach(ex4)
fit <- lm(mpg~ displacement,ex4)
par(mfrow=c(1,2))
plot(fit,1:2,pch=19) # residual diagnostics
```



```
summary(fit) # fit summary
```

```
##
## Call:
## lm(formula = mpg ~ displacement, data = ex4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.9170  -3.0243  -0.5021   2.3512  18.6128
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  35.12064    0.49443   71.03  <2e-16 ***
## displacement -0.06005    0.00224  -26.81  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.635 on 390 degrees of freedom
## Multiple R-squared:  0.6482, Adjusted R-squared:  0.6473
## F-statistic: 718.7 on 1 and 390 DF, p-value: < 2.2e-16
cor(mpg,displacement)
```

```
## [1] -0.8051269
```

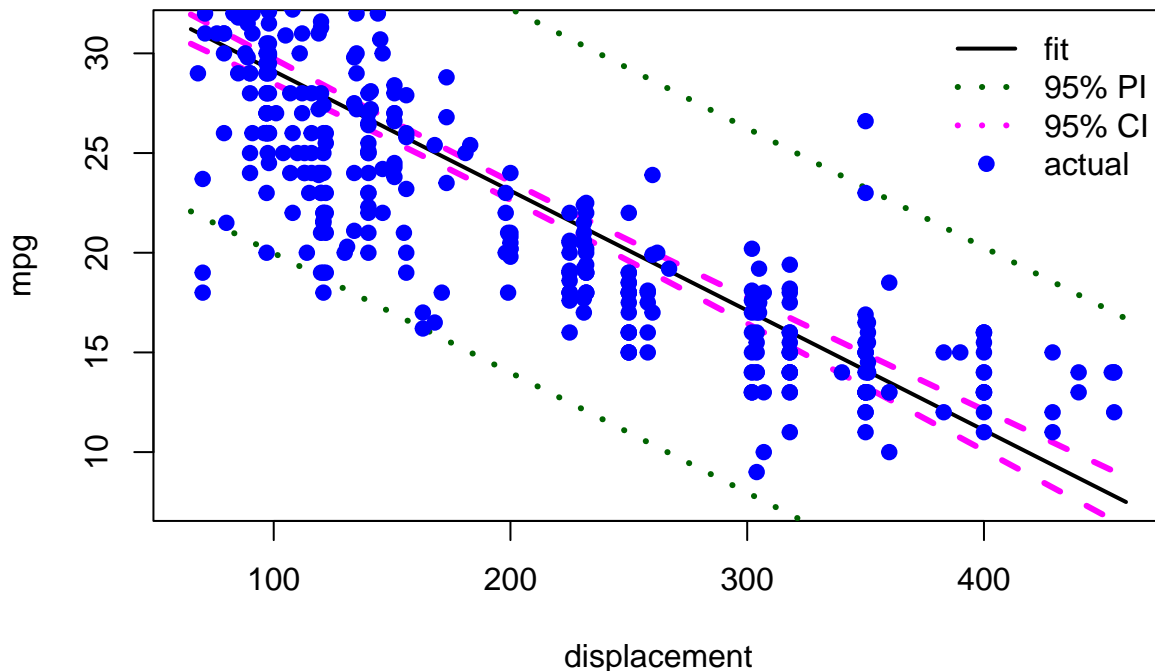
From diagnostics plots, we observe that non-constant variance, heteroscedasticity, and well as a slight right skewness to the residual distribution. P-value is $2.2e-16$, so there is a significant linear relationship. $R^2 = 0.6482$, so the displacement explains 64.82% of the variation in mpg. The relationship is negative because the coefficient estimate is negative; we would expect on average mpg to decrease by 0.06005 units for every 1 unit increase in displacement.

- b) Create a grid of potential displacement values ranging from 65 to 460 in increments of at most 0.5 and predict values of mpg for those values of displacement; also calculate values for the 95% confidence

interval and the 95% prediction interval. Plot the predicted values versus displacement. Overlay the confidence and prediction limits using dashed and dotted lines, respectively; also overlay the actual mpg values as points.

Solution:

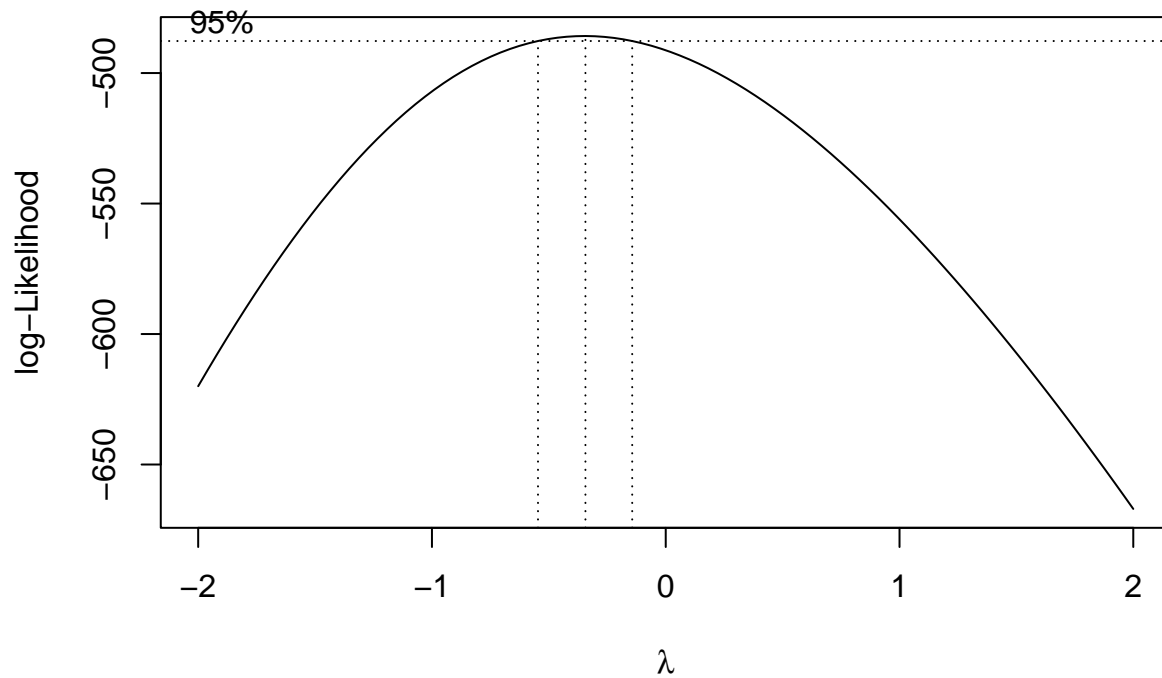
```
grid <- data.frame(seq(65, 460, by = 0.5))
colnames(grid)="displacement"
muMpg <- predict(fit,new=grid,interval="confidence")
Ypred <- predict(fit,new=grid,interval="prediction")
plot(grid$displacement,Ypred[,1],type="l",lwd=2,xlab="displacement",
      ylab="mpg")
lines(grid$displacement,Ypred[,2],lty=3, col = "darkgreen", lwd = 3)
lines(grid$displacement,Ypred[,3],lty=3, col = "darkgreen", lwd = 3)
lines(grid$displacement,muMpg[,2],lty=2, col="magenta", lwd = 3)
lines(grid$displacement,muMpg[,3],lty=2, col = "magenta", lwd = 3)
points(ex4$displacement,ex4$mpg,pch=19,col="blue")
legend("topright",legend=c("fit","95% PI", "95% CI","actual"),
      lwd=c(2,3,3,NA),pch=c(NA,NA,NA, 19),lty=c(1,3,3, NA),
      col=c("black","darkgreen","magenta","blue"),bty="n")
```



- c) From the diagnostics in part (a) you should see some heteroscedasticity and well as a slight right skewness to the residual distribution. Find a good transformation to correct these problems. Try to fit models for several different transformed mpg values until you find one that reasonably corrects the two problems mentioned above. Plot the diagnostics for the fit to the transformed response you select to show that the problems are corrected. From the summary, is there a significant linear relationship between displacement and the transformed mpg? How much of the variation in the transformed mpg can be explained by displacement?

Solution:

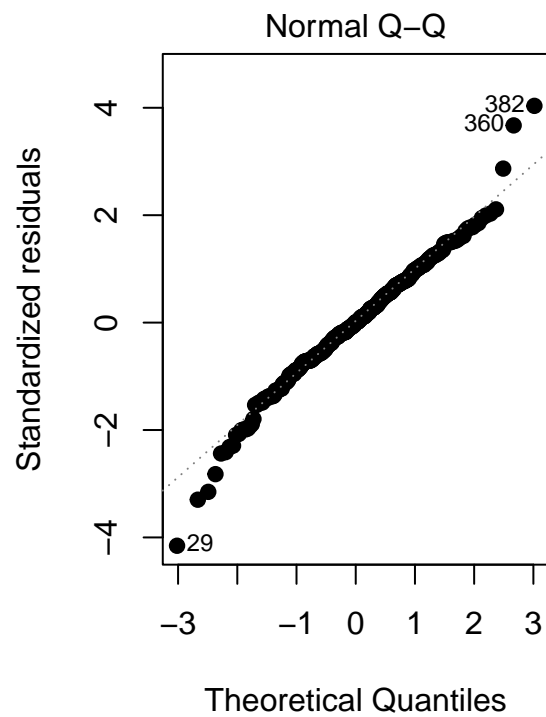
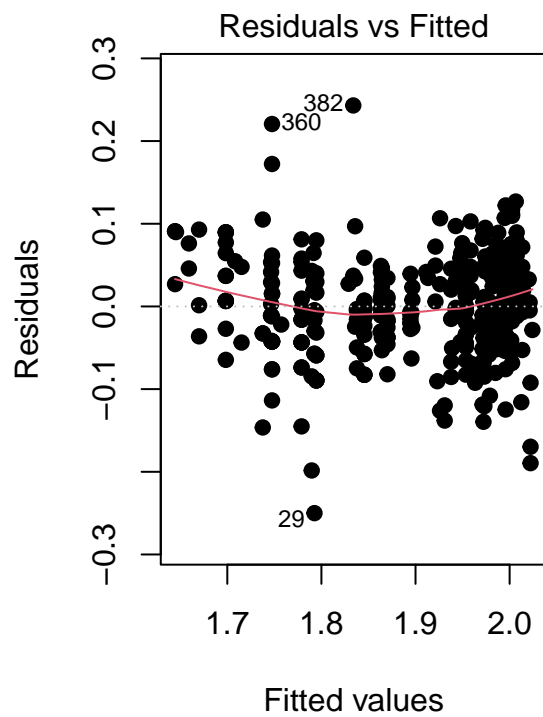
```
# Box-Cox
library(MASS)
par(mfrow=c(1,1))
box <- boxcox(fit)
```



```
lambda <- box$x[which.max(box$y)]
lambda
```

```
## [1] -0.3434343
```

```
mpg.box <- (mpg^lambda-1)/lambda
fit <- lm(mpg.box~displacement)
par(mfrow=c(1,2))
plot(fit,1:2,pch=19)
```

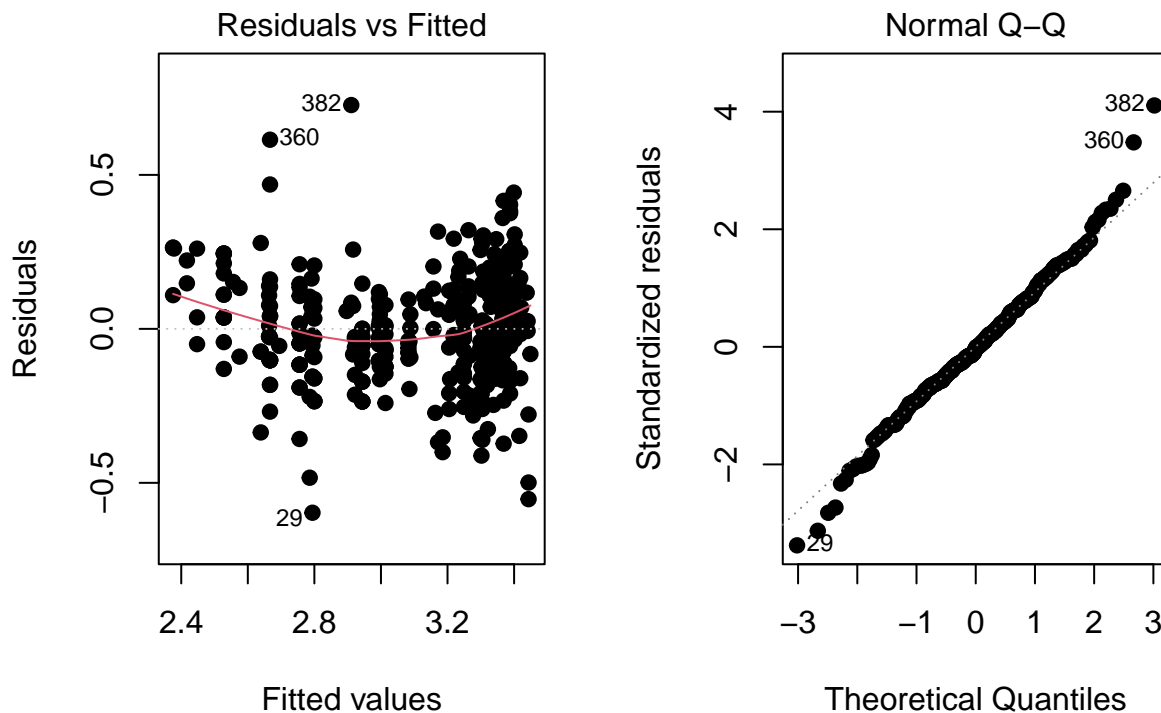


```
summary(fit)
```

```
##
## Call:
## lm(formula = mpg.box ~ displacement)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.249993 -0.037477  0.000033  0.041341  0.243061
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.091e+00  6.436e-03  324.88  <2e-16 ***
## displacement -9.812e-04  2.916e-05  -33.65  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06034 on 390 degrees of freedom
## Multiple R-squared:  0.7438, Adjusted R-squared:  0.7431
## F-statistic: 1132 on 1 and 390 DF,  p-value: < 2.2e-16
```

In order to get rid of two issues, we try different transformations. The first one we tried is Box-Cox transform. We originally have $R^2 = 0.6482$. Now, we have $R^2 = 0.7438$, so we observe some improvements. It does a quite good job of resolving the right skew issue with the Q-Q plot; however, didn't seem to help with the heteroscedasticity in the residuals plot.

```
fit <- lm(log(mpg)~ displacement,ex4)
par(mfrow=c(1,2))
plot(fit,1:2,pch=19) # residual diagnostics
```



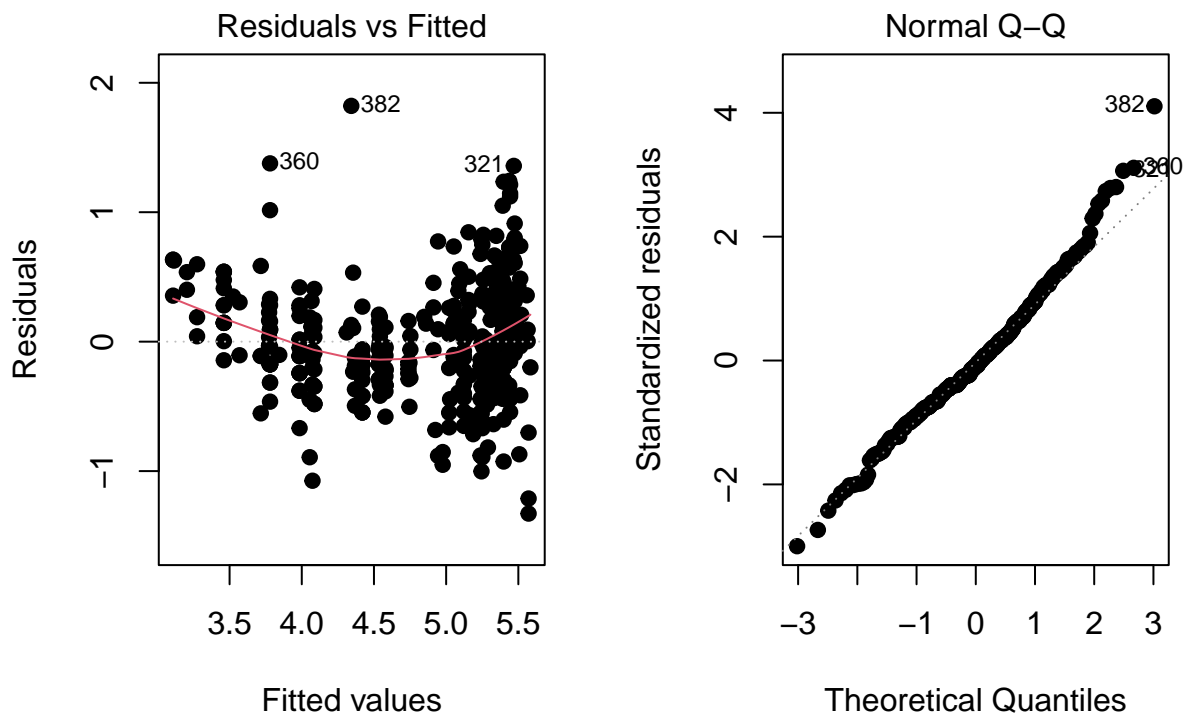
```
summary(fit) # fit summary
```



```
##
## Call:
## lm(formula = log(mpg) ~ displacement, data = ex4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59709 -0.11044 -0.00549  0.11112  0.72677
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.638e+00  1.891e-02  192.32  <2e-16 ***
## displacement -2.774e-03  8.569e-05  -32.37  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1773 on 390 degrees of freedom
## Multiple R-squared:  0.7288, Adjusted R-squared:  0.7281
## F-statistic: 1048 on 1 and 390 DF, p-value: < 2.2e-16
```

Next, we use log transformation. Now, we have $R^2 = 0.7288$ even if it shows some improvements comparing to original R^2 , it has less value comparing to Box-Cox transform. It does a quite good job of resolving the right skew issue with the Q-Q plot; however, didn't seem to help with the heteroscedasticity in the residuals plot. Thus, we need to continue to seek a correct transform.

```
fit <- lm(sqrt(mpg) ~ displacement, ex4)
par(mfrow=c(1,2))
plot(fit,1:2,pch=19) # residual diagnostics
```



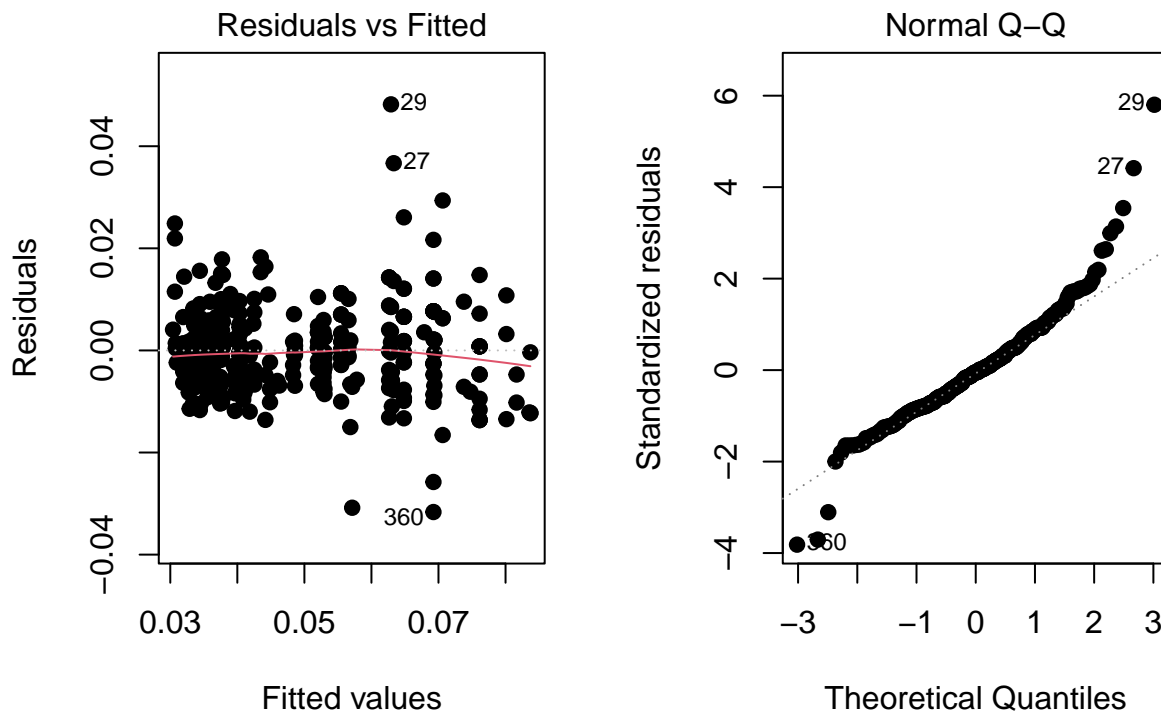
```
summary(fit) # fit summary
```

```
##
## Call:
## lm(formula = sqrt(mpg) ~ displacement, data = ex4)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.32820 -0.29233 -0.04063  0.26681  1.82144
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.0185034  0.0474141  126.94  <2e-16 ***
## displacement -0.0063951  0.0002148  -29.77  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4445 on 390 degrees of freedom
## Multiple R-squared:  0.6944, Adjusted R-squared:  0.6936
## F-statistic: 886.3 on 1 and 390 DF,  p-value: < 2.2e-16
```

Next, we use square root transformation. Now, we have $R^2 = 0.6944$, it shows a little improvement comparing to two transforms we used above. It does not resolve the right skew issue with the Q-Q plot and didn't seem to help with the heteroscedasticity in the residuals plot.

```
fit <- lm(1/mpg~ displacement,ex4)
par(mfrow=c(1,2))
plot(fit,1:2,pch=19) # residual diagnostics
```



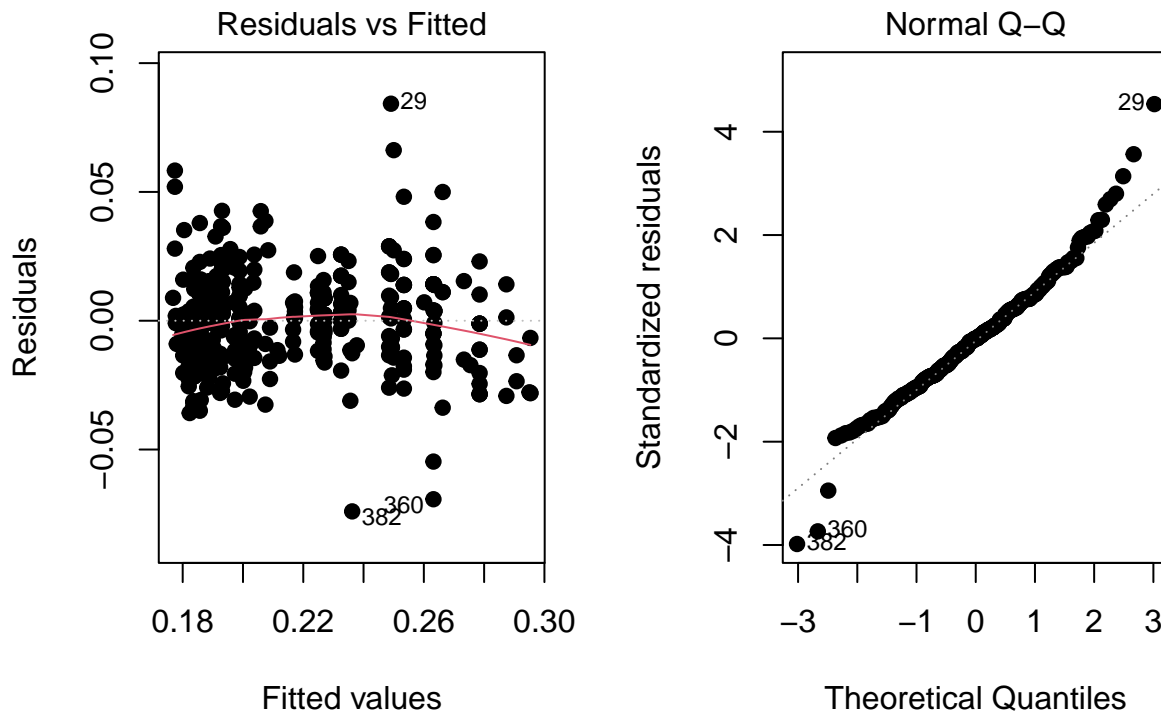
```
summary(fit) # fit summary
```

```
##
## Call:
## lm(formula = 1/mpg ~ displacement, data = ex4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.031655 -0.005309 -0.000394  0.004117  0.048197
```

```
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.105e-02  8.885e-04   23.69  <2e-16 ***
## displacement 1.377e-04  4.026e-06   34.21  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00833 on 390 degrees of freedom
## Multiple R-squared:  0.75, Adjusted R-squared:  0.7494
## F-statistic: 1170 on 1 and 390 DF, p-value: < 2.2e-16
```

Next, we use $1/Y$ (reciprocal) transformation. Now, we have $R^2 = 0.75$, which demonstrates the best R^2 value so far meaning that we have very good improvements. It does not do a good job to resolve the right skew issue with the Q-Q plot; however, it seems to help with resolving the heteroscedasticity in the residuals plot. We still want to continue to check other transforms.

```
fit <- lm(1/sqrt(mpg) ~ displacement, ex4)
par(mfrow=c(1,2))
plot(fit, 1:2, pch=19) # residual diagnostics
```



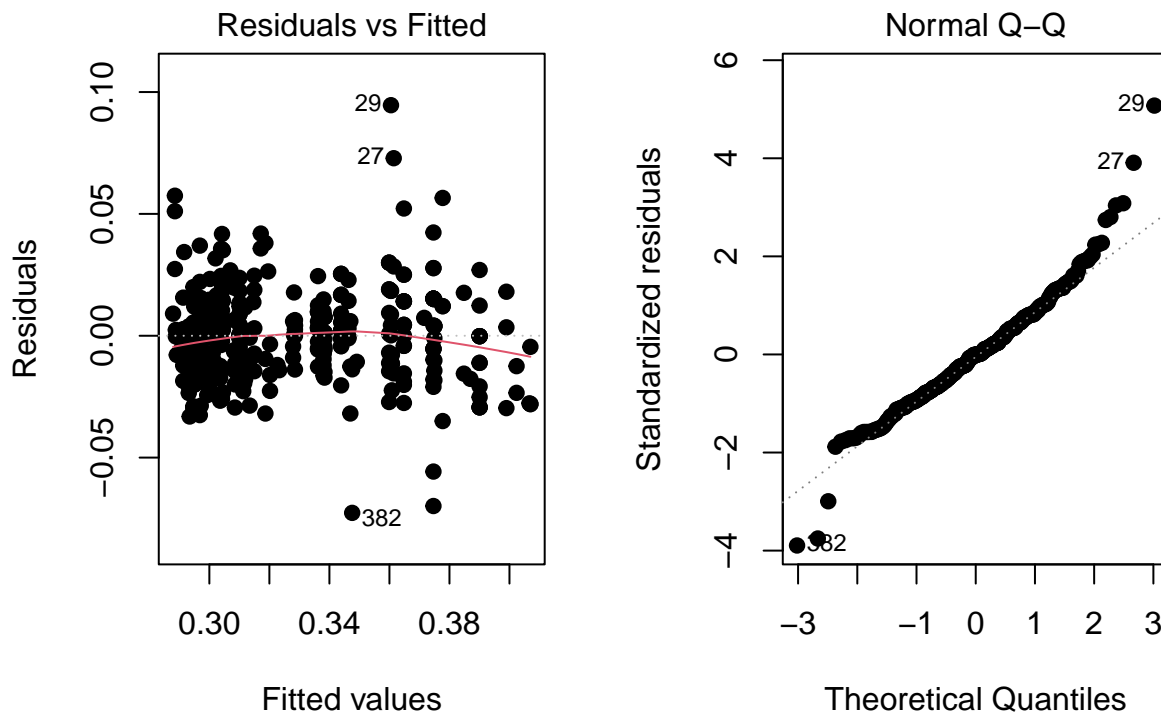
```
summary(fit) # fit summary
```

```
##
## Call:
## lm(formula = 1/sqrt(mpg) ~ displacement, data = ex4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.074009 -0.012827 -0.000057  0.010987  0.084238
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 1.560e-01 1.987e-03 78.51 <2e-16 ***
## displacement 3.063e-04 9.001e-06 34.03 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01863 on 390 degrees of freedom
## Multiple R-squared:  0.748, Adjusted R-squared:  0.7474
## F-statistic: 1158 on 1 and 390 DF, p-value: < 2.2e-16
```

Next, we use $1/\sqrt{Y}$ transformation. Now, we have $R^2 = 0.748$, we see a good improvements comparing to original R^2 . Th diagnostics plots are very similar to the one $1/Y$ transform generates. However, $R^2 = 0.75$ of $1/Y$ is greater than this transform's R^2 .

```
fit <- lm((1/log(mpg))~ displacement,ex4)
par(mfrow=c(1,2))
plot(fit,1:2,pch=19) # residual diagnostics
```



```
summary(fit) # fit summary
```

```
##
## Call:
## lm(formula = (1/log(mpg)) ~ displacement, data = ex4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.072686 -0.012393 -0.000124  0.010522  0.094605
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.670e-01  1.994e-03  133.90  <2e-16 ***
## displacement  3.077e-04  9.033e-06   34.06  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

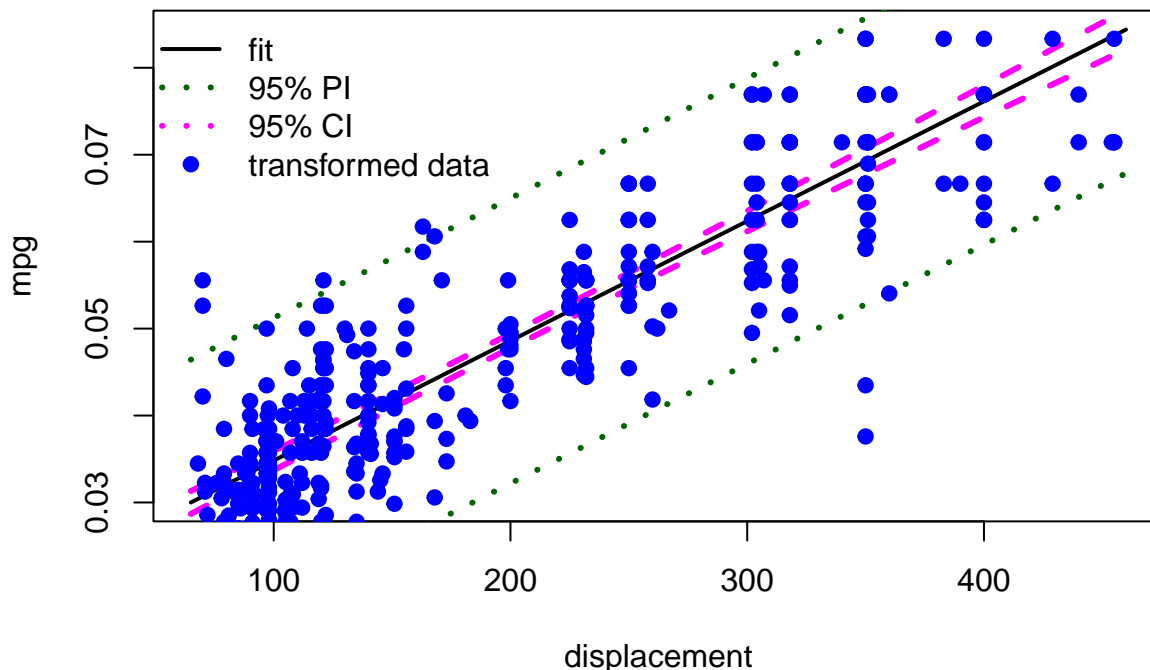
```
##
## Residual standard error: 0.01869 on 390 degrees of freedom
## Multiple R-squared:  0.7484, Adjusted R-squared:  0.7477
## F-statistic: 1160 on 1 and 390 DF,  p-value: < 2.2e-16
```

Lastly, we use $1/\log(Y)$ transformation. Now, we have $R^2 = 0.7484$, we see a good improvements comparing to original R^2 . Th diagnostics plots again are very similar to the one $1/Y$ transform generates. However, $R^2 = 0.75$ of $1/Y$ is greater than this transform's R^2 .

Consequently, $1/Y$ transform didn't completely remove the issue of right skewed data but of the transforms we tried this came the closest to resolving both the skew and the heteroscedasticity with big R^2 value. Thus, we decide to choose this transform for later analysis. We see from this summary that the R-Squared value increased from 0.64 to 0.75, meaning that this fit accounts for 75% of the variation in mpg while the original one only accounted for 64%. Since the P-value remained the same there is still significant linear relationship.

- d) Predict values of the transformed mpg for the values of displacement in your grid; also calculate values for the 95% confidence interval and the 95% prediction interval. Plot the predicted values versus displacement. Overlay the confidence and prediction limits using dashed and dotted lines, respectively.

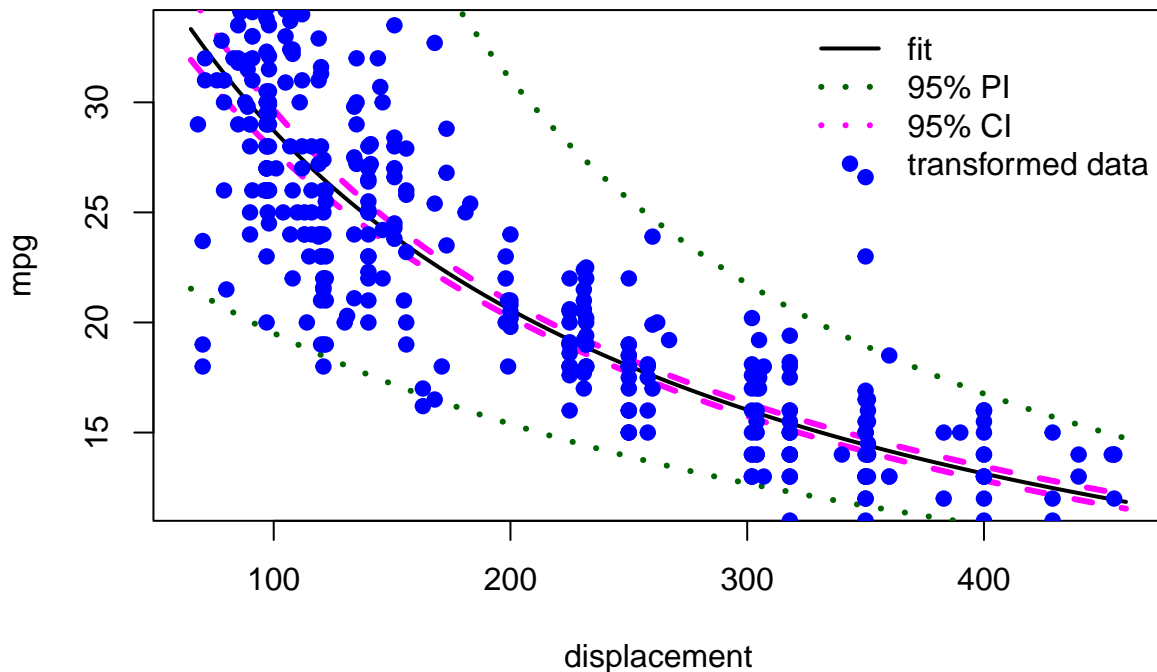
```
fit <- lm(1/mpg~ displacement,ex4)
grid <- data.frame(seq(65, 460, by = 0.5))
colnames(grid)="displacement"
muMpg <- predict(fit,new=grid,interval="confidence")
Ypred <- predict(fit,new=grid,interval="prediction")
plot(grid$displacement,Ypred[,1],type="l",lwd=2,xlab="displacement",
      ylab="mpg")
lines(grid$displacement,Ypred[,2],lty=3, col = "darkgreen", lwd = 3)
lines(grid$displacement,Ypred[,3],lty=3, col = "darkgreen", lwd = 3)
lines(grid$displacement,muMpg[,2],lty=2, col="magenta", lwd = 3)
lines(grid$displacement,muMpg[,3],lty=2, col = "magenta", lwd = 3)
points(ex4$displacement,1/ex4$mpg,pch=19,col="blue")
legend("topleft",legend=c("fit","95% PI", "95% CI","transformed data"),
      lwd=c(2,3,3,NA),pch=c(NA,NA,NA, 19),lty=c(1,3,3, NA),
      col=c("black","darkgreen","magenta","blue"),bty="n")
```



- e) Perform a reverse transform of the predicted values as well as the confidence and prediction limits to convert them back to mile per gallon. Plot the predicted values of mpg versus displacement values in the grid. Overlay the confidence and prediction limits using dashed and dotted lines respectively; also overlay the actual mpg values as points.

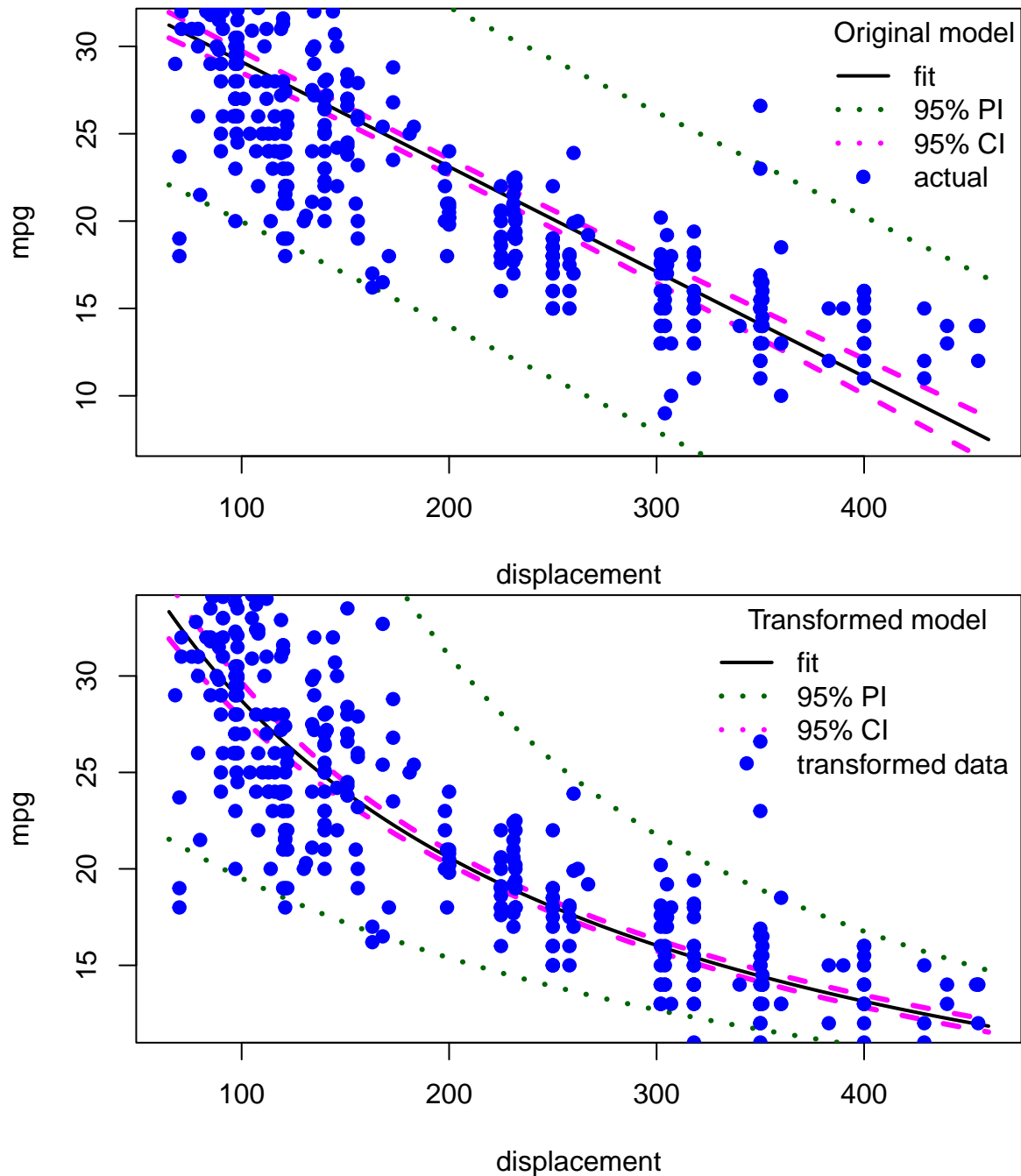
Solution:

```
grid <- data.frame(seq(65, 460, by = 0.5))
colnames(grid)="displacement"
muMpg <- predict(fit,new=grid,interval="confidence")
Ypred <- predict(fit,new=grid,interval="prediction")
plot(grid$displacement,1/Ypred[,1],type="l",lwd=2,xlab="displacement",
      ylab="mpg")
lines(grid$displacement,1/Ypred[,2],lty=3, col = "darkgreen", lwd = 3)
lines(grid$displacement,1/Ypred[,3],lty=3, col = "darkgreen", lwd = 3)
lines(grid$displacement,1/muMpg[,2],lty=2, col="magenta", lwd = 3)
lines(grid$displacement,1/muMpg[,3],lty=2, col = "magenta", lwd = 3)
points(ex4$displacement,ex4$mpg,pch=19,col="blue")
legend("topright",legend=c("fit","95% PI", "95% CI","transformed data"),
      lwd=c(2,3,3,NA),pch=c(NA,NA,NA, 19),lty=c(1,3,3, NA),
      col=c("black","darkgreen","magenta","blue"),bty="n")
```



- f) Do you prefer the model using mpg as the response or the model using the transformed response? Comment on the reasons for your choice.

Solution:



Fitting the model to a modified data set and then executing an inverse transform on the model tends to result in a better fit line than just running a linear regression on an unbalanced data set. The transformed model appears to be closer to several of the data points at the end that the original model missed. By comparing the R-Squared values from both summaries, we can definitely infer that the transformed model accounts for more of the variance in mpg, with the transformed model accounting for roughly 11 percent more variation than the original model.

Problem 5: In this problem, we evaluate an experiment to understand certain factors within the manufacturing process of a silver powder used in silicon wafer manufacturing. Specifically, we want to know the impact of Stir Rate (in RPM), Ammonium Concentration (in %), and Temperature (in deg C) on the density (in g/cm³) of the powder. The data from the experiment is contained in silver.csv.

- a) Analyze these data with a view to which factors and interactions have highly significant effects, and which levels within these factors are different. Report on all steps necessary for a proper ANOVA.

Solution:

Our fixed model has a form of

$$X_{ijkl} = \mu + \alpha_i + \beta_j + \delta_k + \gamma_{ij}^{AB} + \gamma_{ik}^{AC} + \gamma_{jk}^{BC} + \gamma_{ij} + \varepsilon_{ijkl}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J, \quad 1 \leq k \leq K$$

where we want $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Consider factor A is ammonium, factor B is stir rate, factor C is temperature, hypotheses are:

$$H_{0A} : \alpha_1 = \alpha_2 = 0, \quad H_{aA} : \text{at least one of the } \alpha_i \text{ s is different than 0.}$$

$$H_{0B} : \beta_1 = \beta_2 = 0, \quad H_{aB} : \text{at least one of the } \beta_i \text{ s is different than 0.}$$

$$H_{0C} : \delta_1 = \delta_2 = 0, \quad H_{aC} : \text{at least one of the } \delta_i \text{ s is different than 0.}$$

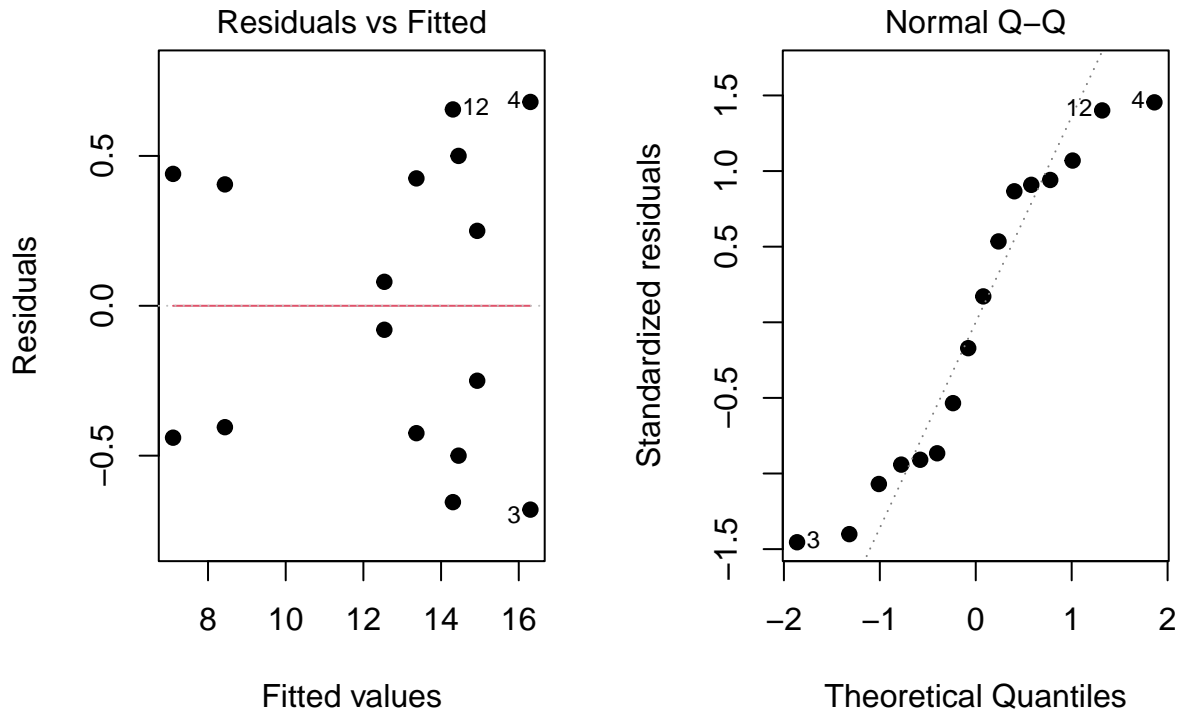
$$H_{0AB} : \gamma_{11}^{AB} = \dots = \gamma_{22}^{AB} = 0, \quad H_{aAB} : \text{at least one is not equal to 0.}$$

$$H_{0AC} : \gamma_{11}^{AC} = \dots = \gamma_{22}^{AC} = 0, \quad H_{aAC} : \text{at least one is not equal to 0.}$$

$$H_{0BC} : \gamma_{11}^{BC} = \dots = \gamma_{22}^{BC} = 0, \quad H_{aBC} : \text{at least one is not equal to 0.}$$

$$H_{0ABC} : \gamma_{111} = \dots = \gamma_{222} = 0, \quad H_{aABC} : \text{at least one is not equal to 0.}$$

```
ex5=read.csv("silver.csv")
attach(ex5)
fit <- aov(Density~as.factor(Ammonium)* as.factor(Stir_Rate)* as.factor(Temperature),data=ex5)
par(mfrow=c(1,2))
plot(fit,1:2,pch=19)
```



According to diagnostics plot, we say residuals have relatively constant spread across treatment levels, and residuals are close to normally distributed. Thus, assumptions for error term in model hold.

```
summary(fit)
```

```
##                                     Df Sum Sq
## as.factor(Ammonium)                 1  33.61
## as.factor(Stir_Rate)                 1  85.98
## as.factor(Temperature)               1   0.02
## as.factor(Ammonium):as.factor(Stir_Rate) 1  20.91
## as.factor(Ammonium):as.factor(Temperature) 1   1.03
## as.factor(Stir_Rate):as.factor(Temperature) 1   5.37
## as.factor(Ammonium):as.factor(Stir_Rate):as.factor(Temperature) 1   0.25
## Residuals                           8   3.50
##                                     Mean Sq F value
## as.factor(Ammonium)                 33.61  76.888
## as.factor(Stir_Rate)                 85.98 196.684
## as.factor(Temperature)               0.02   0.057
## as.factor(Ammonium):as.factor(Stir_Rate) 20.91  47.828
## as.factor(Ammonium):as.factor(Temperature) 1.03   2.345
## as.factor(Stir_Rate):as.factor(Temperature) 5.37  12.286
## as.factor(Ammonium):as.factor(Stir_Rate):as.factor(Temperature) 0.25   0.578
## Residuals                           0.44
##                                     Pr(>F)
## as.factor(Ammonium)                 2.24e-05 ***
## as.factor(Stir_Rate)                 6.48e-07 ***
## as.factor(Temperature)               0.817703
## as.factor(Ammonium):as.factor(Stir_Rate) 0.000123 ***
## as.factor(Ammonium):as.factor(Temperature) 0.164210
## as.factor(Stir_Rate):as.factor(Temperature) 0.008018 **
## as.factor(Ammonium):as.factor(Stir_Rate):as.factor(Temperature) 0.469047
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fit$coefficients
```

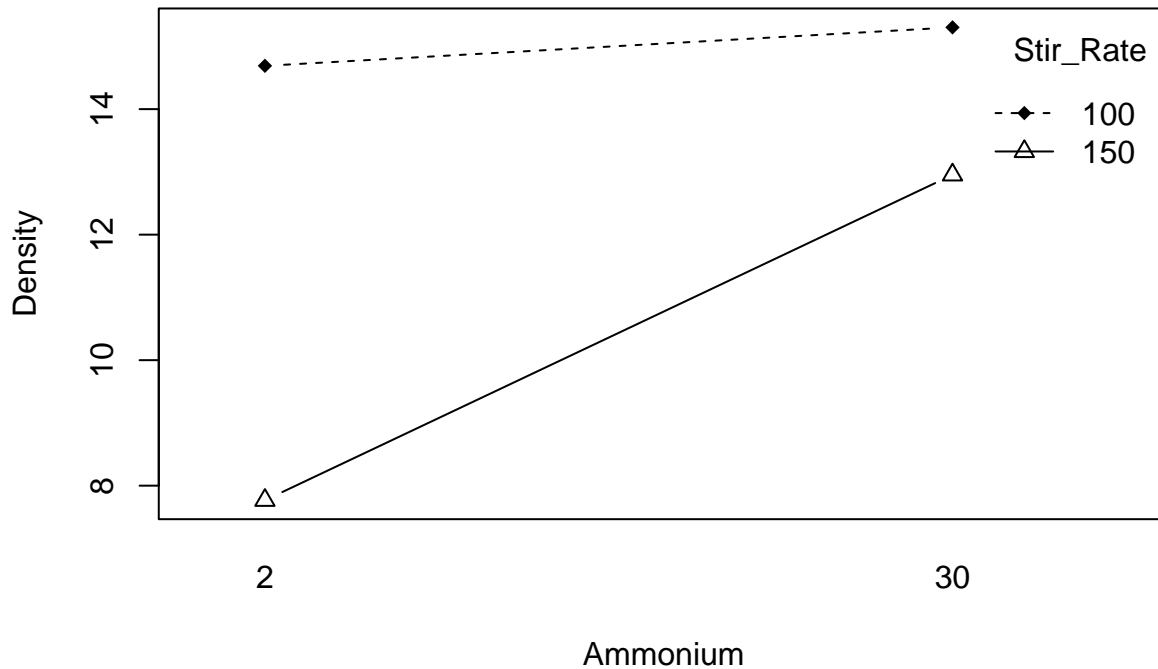
```
##                                     (Intercept)
##                                     14.930
##                                     as.factor(Ammonium)30
##                                     1.370
##                                     as.factor(Stir_Rate)150
##                                     -7.830
##                                     as.factor(Temperature)40
##                                     -0.480
##                                     as.factor(Ammonium)30:as.factor(Stir_Rate)150
##                                     4.070
##                                     as.factor(Ammonium)30:as.factor(Temperature)40
##                                     -1.515
##                                     as.factor(Stir_Rate)150:as.factor(Temperature)40
##                                     1.815
## as.factor(Ammonium)30:as.factor(Stir_Rate)150:as.factor(Temperature)40
##                                     1.005
```

Based on ANOVA table, there is strong evidence ($p = 2.24e - 05 < 0.05$, $p = 6.48e - 07 < 0.05$) that the null hypotheses for Ammonium and Stir Rate H_{0A} and H_{0B} are not true. So, they have significant effect. However,

there is strong evidence ($p = 0.817703 > 0.05$) that the null hypotheses for temperature H_{0C} is true, so it does not have any significant effect. Regarding interactions, there is strong evidence ($p = 0.000123 < 0.05$, $p = 0.008018 < 0.05$) that the null hypotheses for interactions AB and BC H_{0AB} and H_{0BC} are not true. So, they have significant effect. Nonetheless, there is strong evidence ($p = 0.164210 > 0.05$, $p = 0.469047 > 0.05$) that the null hypothesis for interactions AC and ABC, H_{0AC} and H_{0ABC} , are true.

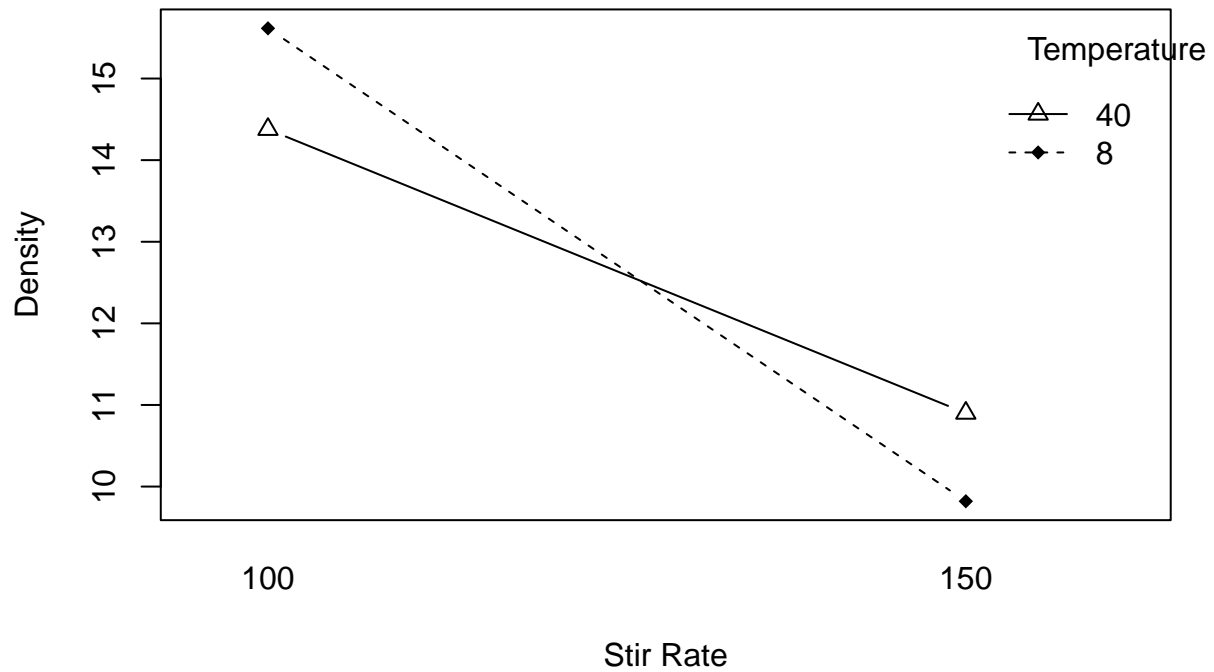
```
# interaction plot
with(ex5, (interaction.plot(Ammonium, Stir_Rate, Density, type = "b",
pch = c(18,24,22), leg.bty = "n", main = "Interaction Plot of Ammonium with Stir_Rate",
xlab = "Ammonium", ylab = "Density"))))
```

Interaction Plot of Ammonium with Stir_Rate



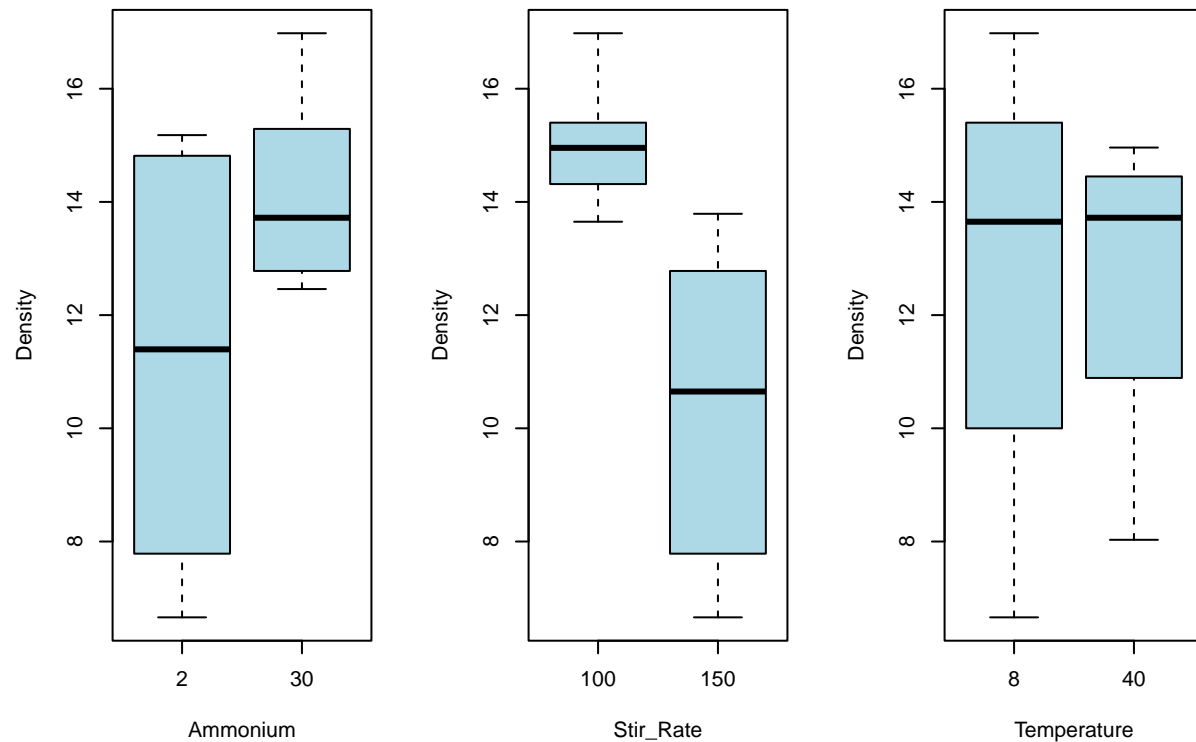
```
## NULL
with(ex5, (interaction.plot(Stir_Rate, Temperature, Density, type = "b",
pch = c(18,24,22), leg.bty = "n", main = "Interaction Plot of Stir Rate with Temperature",
xlab = "Stir Rate", ylab = "Density"))))
```

Interaction Plot of Stir Rate with Temperature



NULL

```
par(mfrow=c(1,3))
boxplot(Density~Ammonium,data=ex5,col="light blue",xlab="Ammonium",ylab="Density")
boxplot(Density~Stir_Rate,data=ex5,col="light blue",xlab="Stir_Rate",ylab="Density")
boxplot(Density~Temperature,data=ex5,col="light blue",xlab="Temperature",ylab="Density")
```



```

tuk=TukeyHSD(fit)
tuk[1:4]

## $`as.factor(Ammonium)`
##           diff      lwr      upr      p adj
## 30-2  2.89875  2.136421  3.661079  2.243707e-05
##
## $`as.factor(Stir_Rate)`
##           diff      lwr      upr      p adj
## 150-100 -4.63625 -5.398579 -3.873921  6.543709e-07
##
## $`as.factor(Temperature)`
##           diff      lwr      upr      p adj
## 40-8 -0.07875 -0.8410786  0.6835786  0.8177028
##
## $`as.factor(Ammonium):as.factor(Stir_Rate)`
##           diff      lwr      upr      p adj
## 30:100-2:100  0.6125 -0.8846531  2.1096531  5.821218e-01
## 2:150-2:100 -6.9225 -8.4196531 -5.4253469  2.057741e-06
## 30:150-2:100 -1.7375 -3.2346531 -0.2403469  2.446087e-02
## 2:150-30:100 -7.5350 -9.0321531 -6.0378469  1.063804e-06
## 30:150-30:100 -2.3500 -3.8471531 -0.8528469  4.477387e-03
## 30:150-2:150  5.1850  3.6878469  6.6821531  1.825023e-05
tuk[6]

## $`as.factor(Stir_Rate):as.factor(Temperature)`
##           diff      lwr      upr      p adj
## 150:8-100:8 -5.7950 -7.2921531 -4.2978469  7.891822e-06
## 100:40-100:8 -1.2375 -2.7346531  0.2596531  1.094458e-01
## 150:40-100:8 -4.7150 -6.2121531 -3.2178469  3.713972e-05
## 100:40-150:8  4.5575  3.0603469  6.0546531  4.778227e-05
## 150:40-150:8  1.0800 -0.4171531  2.5771531  1.746778e-01
## 150:40-100:40 -3.4775 -4.9746531 -1.9803469  3.368099e-04

```

Based on Tukey's procedure, p value for Ammonium and Stir_Rate is $2.24e-05 < .05$ and $7e-07 < 0.05$, so we reject the null hypotheses H_{0A} and H_{0B} . But, p value for Temperature is $0.8177028 > 0.05$, so we accept the null hypothesis H_{0C} . It can be seen in box plot, too.

Subsequently, there is a significant difference for interactions for pairs in Ammonium and Stir_Rate, which are 2:150-2:100 ($0.0000021 < 0.05$), 30:150-2:100 ($0.0244609 < 0.05$), 2:150-30:100 ($0.0000011 < 0.05$), 30:150-30:100 ($0.0044774 < 0.05$), 30:150-2:150 ($0.0000183 < 0.05$). We also see a significant difference for interactions for pairs in Stir_Rate and Temperature, which are 150:8-100:8 ($0.0000079 < 0.04$), 150:40-100:8 ($0.0000371 < 0.04$), 100:40-150:8 ($0.0000478 < 0.05$), 150:40-100:40 ($0.0003368 < 0.05$). We also provide interaction plots of Ammonium with Stir_Rate and Stir Rate with Temperature, respectively. A look at the first graph shows that the effect of ammonium is different for Stir rate 100 than it is for Stir rate 150. We observe the same thing in the second interaction plot, which is the effect of stir rate is different for temperature 40 than it is for temperature 8. In short, since we do not have interactions, we do not have same patterns or slopes in both plots.

- b) Based on the outcome of this experiment, identify specific operating conditions for the process that will maintain a silver power density less than $14g/cm^3$.

Solution:

```

library(dplyr)
Ypred <- predict(fit,new = ex5,interval="prediction")

```

```
New_data = cbind(ex5,Ypred)
filter(New_data, upr<=14)
```

##	X	Stir_Rate	Ammonium	Temperature	Density	fit	lwr	upr
## 5	5	150	2	8	7.54	7.100	5.232684	8.967316
## 6	6	150	2	8	6.66	7.100	5.232684	8.967316
## 13	13	150	2	40	8.03	8.435	6.567684	10.302316
## 14	14	150	2	40	8.84	8.435	6.567684	10.302316

We observe that specific operating conditions for the process that will maintain a silver power density less than $14g/cm^3$ happen when Stir Rate has a value of 150, ammonium has a value of 2, temperature has values 8 and 40.