STAT 522 —— Assignment 4

Subasish Das (sxd1684)

25th March 2014

1 Exercise 5.1

(a)

```
## Given
df_tot = 17
df_err = 12
df_a = 1
ss_a = 0.322
ss_b = 80.554
ss_err = 105.327
ss_tot = 231.551
ms_b = 40.2771
ms_err = 9.773
F_b = 4.59

ms_a = ss_a/df_a
df_b = round(ss_b/ms_b, 0)
df_ab = df_tot - df_err - df_b - df_a
ss_ab = ss_tot - ss_err - ss_a - ss_b
ms_ab = ss_ab/df_ab
F_a = ms_a/ms_err
F_ab = ms_ab/ms_err
p_a = 1 - pf(F_a, df_a, df_err)
p_b = 1 - pf(F_b, df_b, df_err)
p_ab = 1 - pf(F_ab, df_ab, df_err)

rbind(df_b, df_ab)

##       [,1]
## df_b     2
## df_ab    2

rbind(ms_a, ms_ab)

##         [,1]
## ms_a   0.322
## ms_ab 22.674

rbind(F_a, F_ab)

##         [,1]
## F_a  0.03295
## F_ab 2.32007

rbind(p_a, p_b, p_ab)

##         [,1]
## p_a  0.85899
## p_b  0.03308
## p_ab 0.14065
```

The ANOVA Table:

	DF	SS	MS	F	P
A	1	0.322	0.322	0.0367	0.859
B	2	80.554	40.2771	4.59	0.033
Interaction	2	45.348	22.674	2.5832	0.14
Error	12	105.327	8.7773		
Total	17	231.551			

(b) 3
(c) 3
(d) From the p-values, it's found that only B is significant.

2 Exercise 5.3

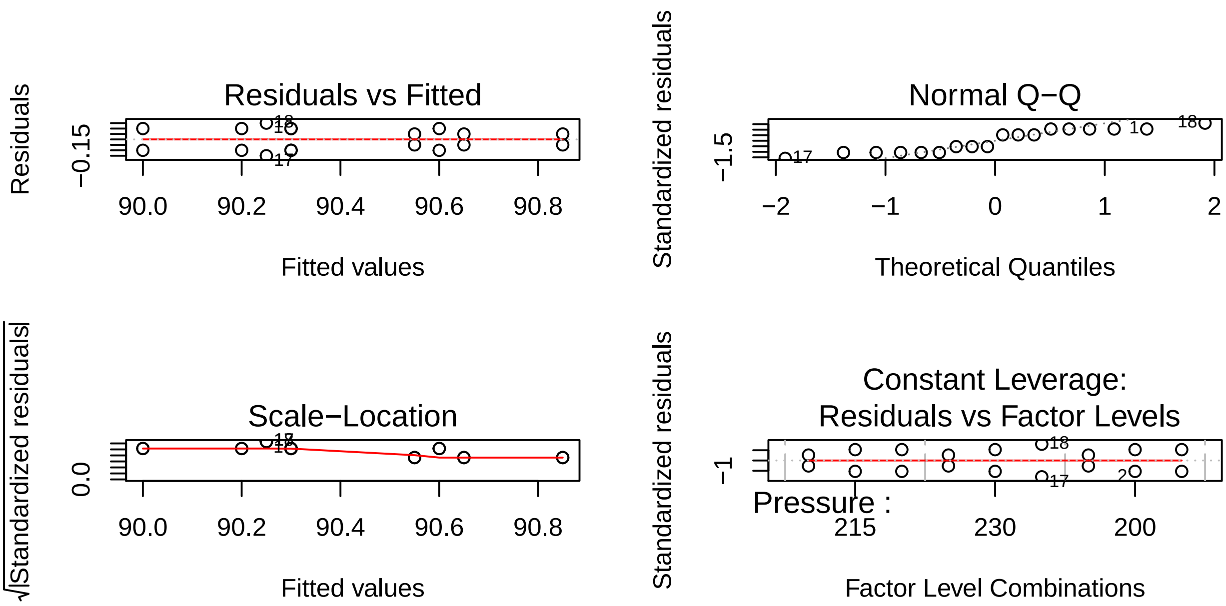
(a) Analyze the data and draw conclusions. Use $\alpha = 0.05$

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW")
yield <- read.csv("5.3.csv")
yield$Pressure <- as.factor(yield$Pressure)
yield$Temperature <- as.factor(yield$Temperature)
yield.aov <- aov(Yield ~ Pressure * Temperature, data = yield)
summary(yield.aov)

##                      Df  Sum Sq Mean Sq F value  Pr(>F)
## Pressure              2  0.768   0.384   21.59 0.00037 ***
## Temperature           2  0.301   0.151    8.47 0.00854 **
## Pressure:Temperature  4  0.069   0.017    0.97 0.47001
## Residuals             9  0.160   0.018
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

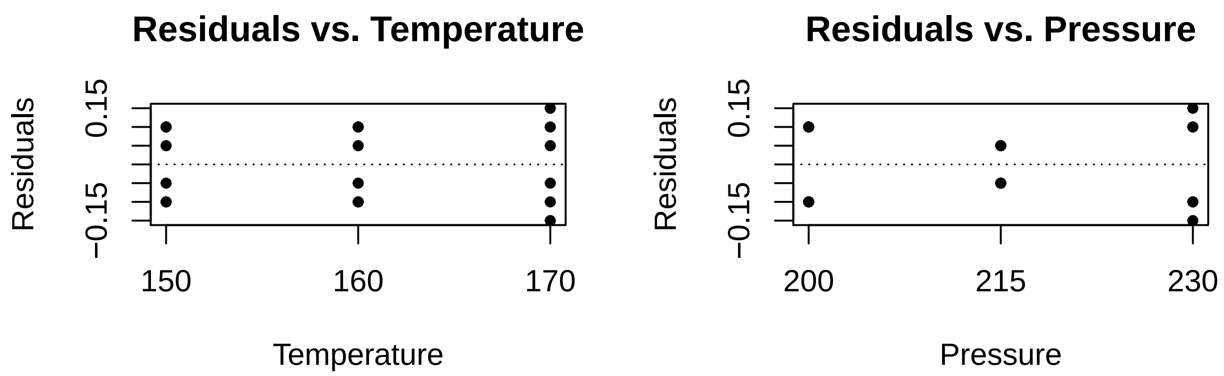
From the R code values, the F-value for pressure is 21.594 with a corresponding p-value of 0.00037 and the F-value for temperature is 8.469 with a corresponding p-value of 0.0085. From these values, we understand that both pressure and temperature are significant. The interaction of pressure and temperature is not significant (because of its higher p-value.)

(b) Prepare appropriate residual plots and comment on the model's adequacy.

```
par(mfrow = c(2, 2))
plot(yield.aov)
```



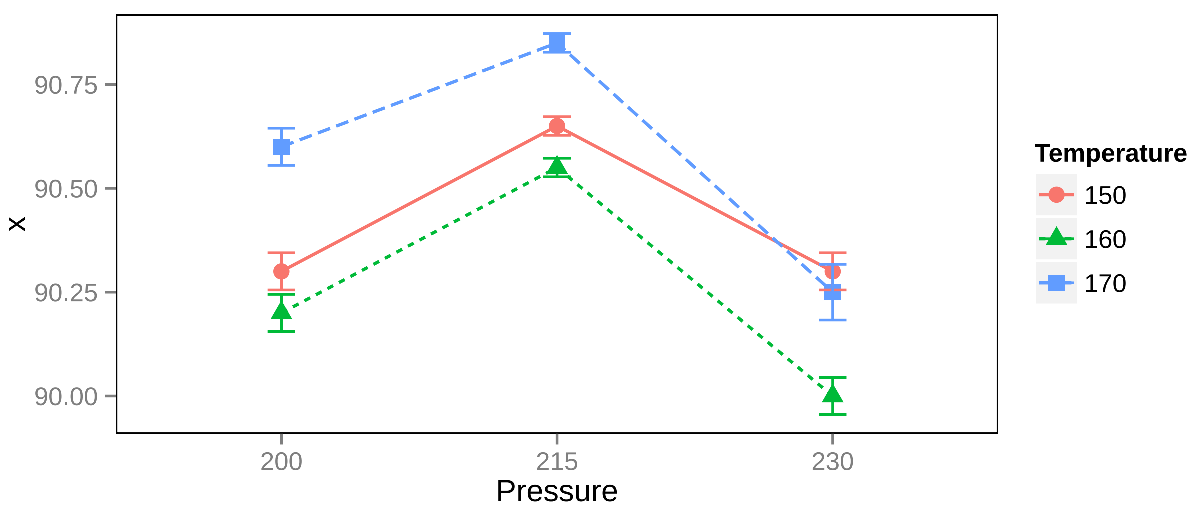
```r
par(mfrow = c(1, 2))
stripchart(residuals(yield.aov) ~ yield$Temperature, vertical = TRUE, jitter = 0,
    xlab = "Temperature", ylab = "Residuals", cex = 1, pch = 20, main = "Residuals vs. Temperature")
abline(h = 0, col = "black", lty = 3)
stripchart(residuals(yield.aov) ~ yield$Pressure, vertical = TRUE, jitter = 0,
    xlab = "Pressure", ylab = "Residuals", cex = 1, pch = 20, main = "Residuals vs. Pressure")
abline(h = 0, col = "black", lty = 3)
```



By observing the residual plots, no significant deviations from the observations are noticed.

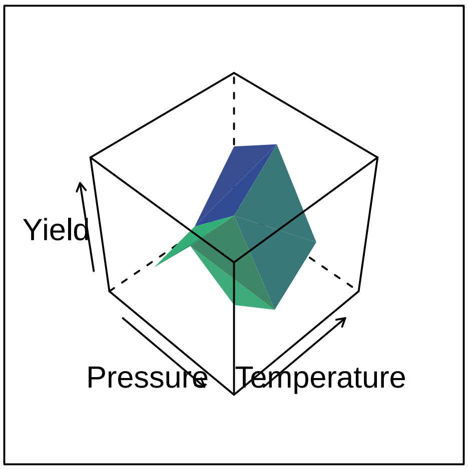
(c) Under what conditions would you operate this process?

```r
library(ggplot2)
df <- with(yield, aggregate(Yield, list(Temperature = Temperature, Pressure = Pressure),
    mean))
df$se <- with(yield, aggregate(Yield, list(Temperature = Temperature, Pressure = Pressure),
    function(x) sd(x)/sqrt(10)))[, 3]

opar <- theme_update(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
    panel.background = element_rect(colour = "black"))
gp <- ggplot(df, aes(x = Pressure, y = x, colour = Temperature, group = Temperature))
gp + geom_line(aes(linetype = Temperature), size = 0.6) + geom_point(aes(shape = Temperature),
    size = 3) + geom_errorbar(aes(ymax = x + se, ymin = x - se), width = 0.1)
```




```r
theme_set(opar)

#### Response Surface Plot
library(lattice)
```

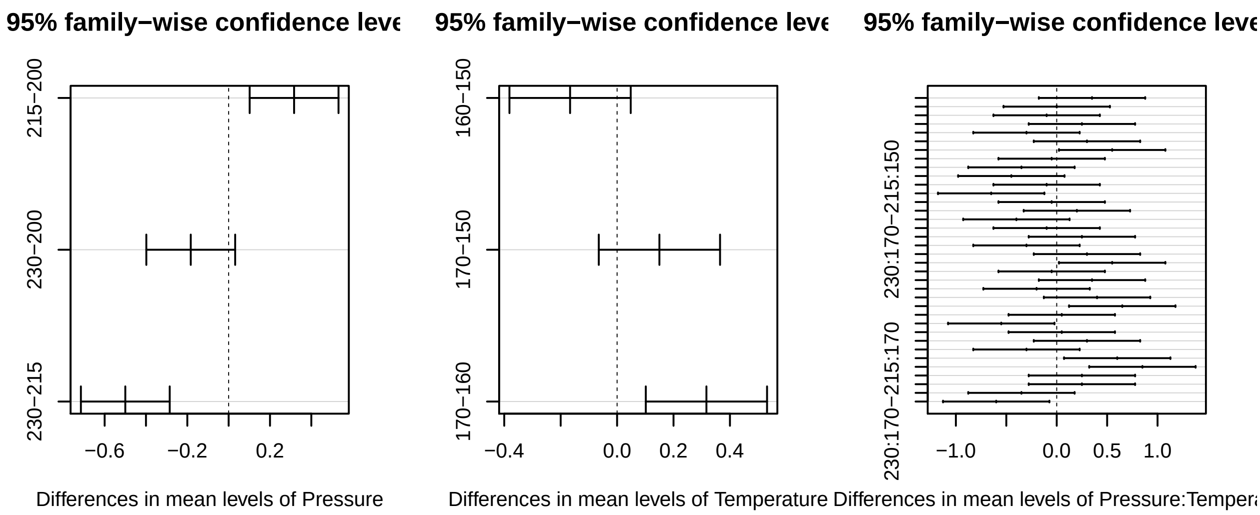
```r
wireframe(Yield ~ Pressure * Temperature, data = yield, screen = list(z = -45,
    x = -45), colorkey = FALSE, shade = TRUE, light.source = c(0, 10, 10), zlim = range(seq(90,
    91, by = 0.05)))
```



From the interaction plot, it's visible that temperature at 170 and pressure at 215 gives the highest yield.

3 Exercise 5.12

```r
TukeyHSD(yield.aov, which = "Pressure", conf.level = 0.95)

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Yield ~ Pressure * Temperature, data = yield)
##
## $Pressure
##              diff      lwr     upr  p adj
## 215-200  0.3167  0.1017  0.5316 0.0067
## 230-200 -0.1833 -0.3983  0.0316 0.0945
## 230-215 -0.5000 -0.7149 -0.2851 0.0003

par(mfrow = c(1, 3))
plot(TukeyHSD(yield.aov), cex = 0.8)
```



Three different ranges of pressure are compared. Pressure range of 230 to 200 is insignificant because of its higher p-value (greater than 0.05). The difference in pressure range (230-215) seems more than the other two pressure range.

4 Exercise 5.15

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW")
exp <- read.csv("5.15.csv")
exp$Row <- as.factor(exp$Row)
exp$Column <- as.factor(exp$Column)
exp.aov <- aov(Data ~ Row * Column, data = exp)
anova(exp.aov)

## Warning:  ANOVA F-tests on an essentially perfect fit are unreliable

## Analysis of Variance Table
##
## Response: Data
##            Df Sum Sq Mean Sq F value Pr(>F)
## Row         2    581   290.3
## Column      3     29     9.6
## Row:Column  6     29     4.8
## Residuals   0      0


## As F-values can't be determined in this test, we condider anova without
## any interaction.
exp.aov_noint <- aov(Data ~ Row + Column, data = exp)
anova(exp.aov_noint)

## Analysis of Variance Table
##
## Response: Data
##           Df Sum Sq Mean Sq F value  Pr(>F)
## Row        2    581   290.3   60.40 0.00011 ***
## Column     3     29     9.6    2.01 0.21472
## Residuals  6     29     4.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

By considering interaction, the F-value and p-values are not generated. When we consider there's no interaction, the p-values for row is very low (less than 0.05) which indicates significance.

```
library(asbio)

## Warning:  package 'asbio' was built under R version 3.0.3
## Loading required package:  tcltk

setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW")
exp <- read.csv("5.15.csv")
tukey.add.test(exp$Data, exp$Row, exp$Column)

##
## Tukey's one df test for additivity
## F = 0.6999   Denom df = 5    p-value = 0.441

### From the test
F_nonadd = 0.6999
df_nonadd = 1
ss_nonadd = 3.54
```

```
ms_nonadd = ss_nonadd/df_nonadd
ss_err = 28.83 - ss_nonadd
df_err = 6 - df_nonadd
ms_err = ss_err/df_err
F_row = 580.5
F_col = 28.92
df_row = 2
df_col = 3
p_nonadd = 1 - pf(F_nonadd, df_nonadd, df_err)
p_row = 1 - pf(F_row, df_col, df_err)
p_col = 1 - pf(F_col, df_col, df_err)
rbind(p_row, p_col, p_nonadd)

##               [,1]
## p_row    8.925e-07
## p_col    1.384e-03
## p_nonadd 4.410e-01
```

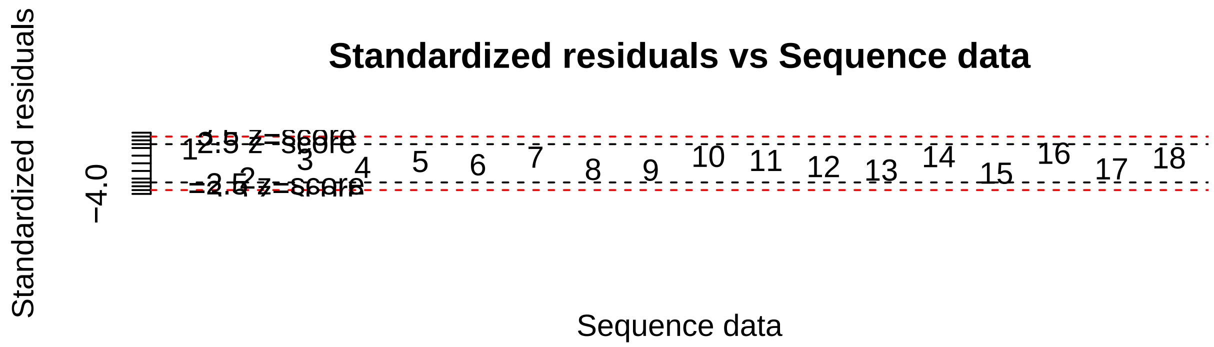
The ANOVA Table (including Nonadditivity row):

	SS	DF	MS	F	P
Row	580.25	2	290.125	57.360	8.92E-07
Column	28.92	3	9.64	1.906	1.38E-03
Nonadd	3.54	1	3.54	0.700	4.41E-01
Error	25.29	5	5.058		
Total	638.25	11			

From Tukey's one df test for additivity, the p-value is high enough (greater than 0.05). So, we can't reject the null hypothesis.

5 Exercise 5.21

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW")
yield2 <- read.csv("5.21a.csv")
head(yield2, n = 2)

##   Pressure Temperature Day Yield
## 1      250         Low   1  86.3
## 2      250         Low   2  86.1

library(easyanova)  ## 'easyanova' package is utilized for convenience.

## Loading required package:  car
## Loading required package:  MASS
## Loading required package:  nnet
## Loading required package:  nlme

r1 = ea2(yield2, design = 2)
```




```
r1[[1]]

##                    df type III SS mean square F value    p>F
## factor_1            2        5.508      2.7539  5.1838  0.036
## factor_2            2       99.854     49.9272 93.9807 <0.001
## blocks              1       13.005     13.0050   24.48 0.0011
## factor_1:factor_2   4        4.452      1.1131  2.0952 0.1733
## Residuals           8        4.250      0.5312       -      -
```

From the ANOVA table values, it is easily interpreted that both temperature and pressure have significant effects. The usage of blocks is also significant. The interaction of temperature and pressure is not significant (greater than 0.05) in this experiment.

6 Exercise 5.24

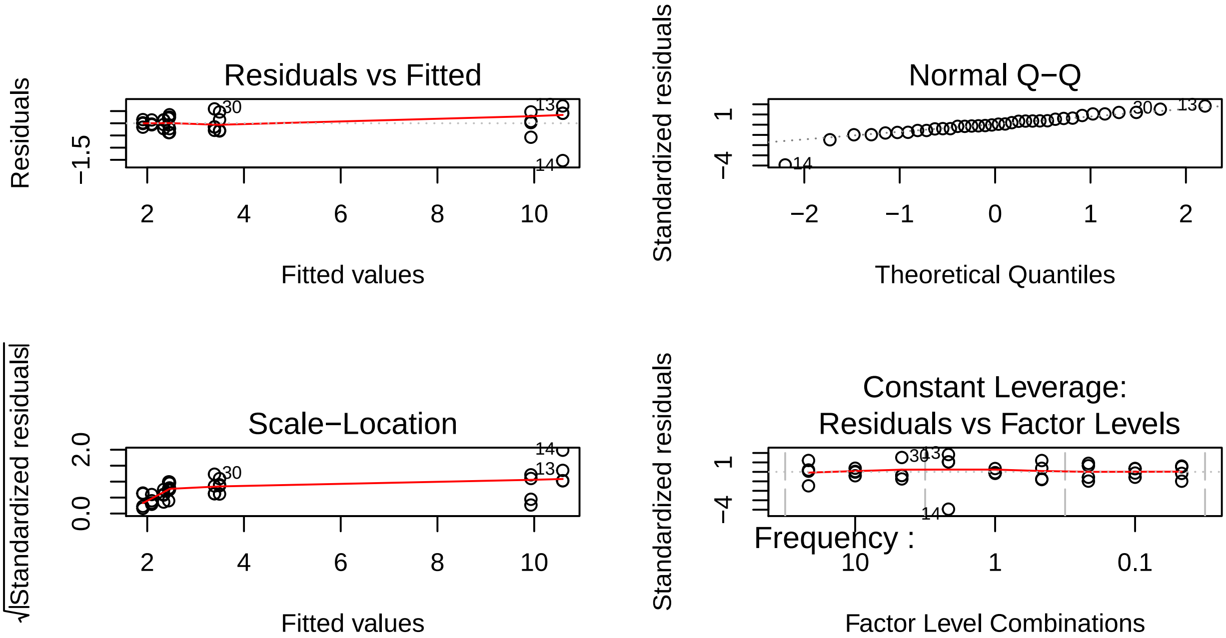
(a) Analyze the data from this experiment. Use $\alpha = 0.05$

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW")
environ <- read.csv("5.24.csv")
environ$Frequency <- as.factor(environ$Frequency)
environ.aov <- aov(Crack.Growth ~ Frequency * Environment, data = environ)
summary(environ.aov)

##                       Df Sum Sq Mean Sq F value  Pr(>F)
## Frequency              2  209.9   104.9     522 < 2e-16 ***
## Environment            2   64.3    32.1     160 1.1e-15 ***
## Frequency:Environment  4  102.0    25.5     127 < 2e-16 ***
## Residuals             27    5.4     0.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

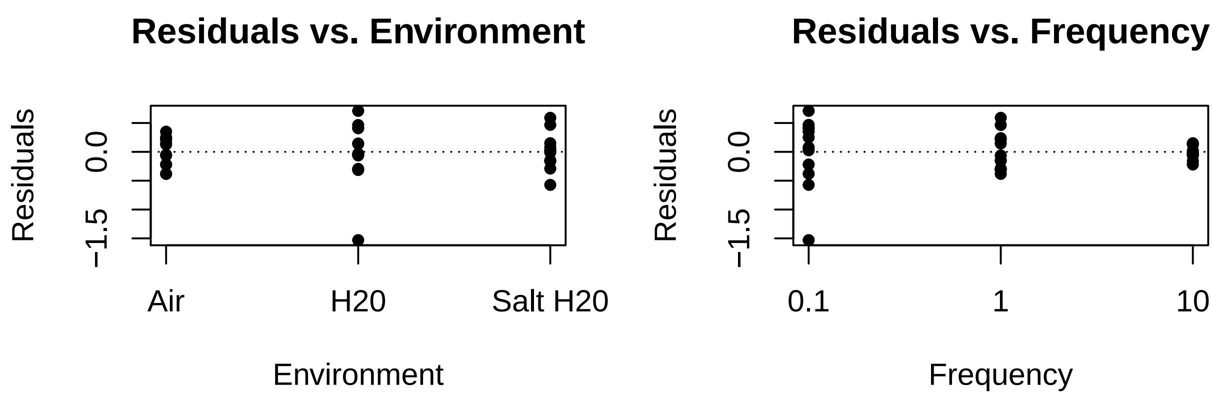
Both frequency and environment as well as the interaction between them are significant because of their low p-values (less than 0.05).

(b) Analyze the residuals.

```
par(mfrow = c(2, 2))
plot(environ.aov)
```




```r
par(mfrow = c(1, 2))
stripchart(residuals(environ.aov) ~ environ$Environment, vertical = TRUE, jitter = 0,
    xlab = "Environment", ylab = "Residuals", cex = 1.1, pch = 20, main = "Residuals vs. Environment")
abline(h = 0, col = "black", lty = 3)
stripchart(residuals(environ.aov) ~ environ$Frequency, vertical = TRUE, jitter = 0,
    xlab = "Frequency", ylab = "Residuals", cex = 1.1, pch = 20, main = "Residuals vs. Frequency")
abline(h = 0, col = "black", lty = 3)
```



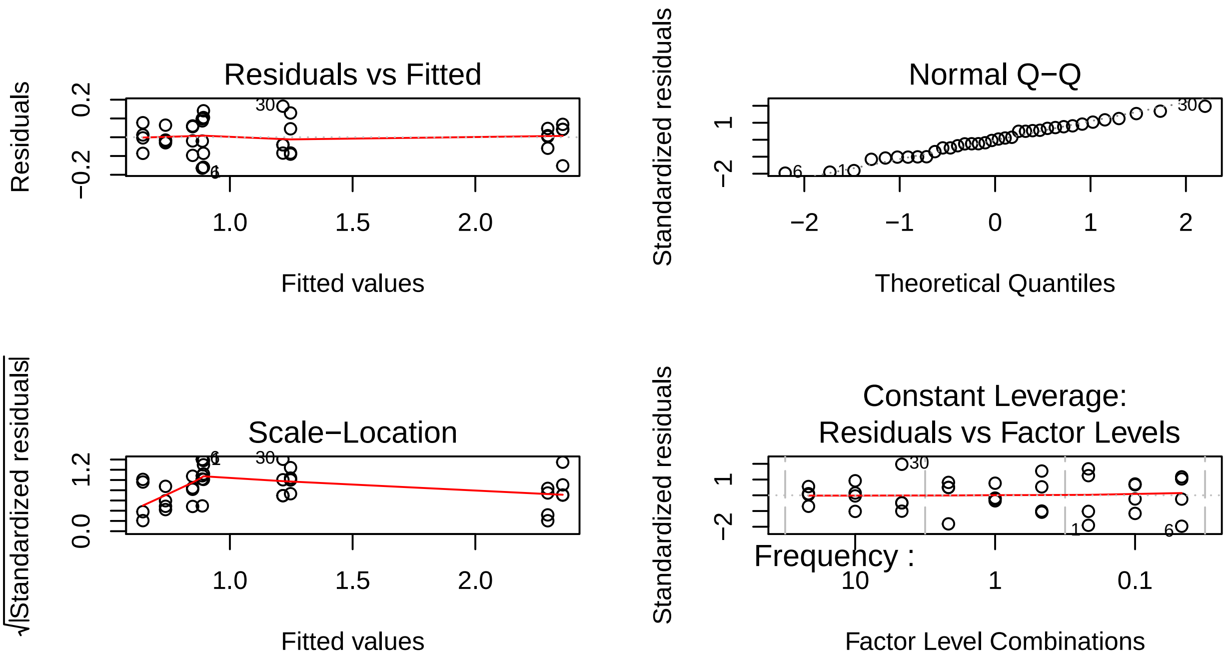
The inequality of variance seems problematic from the observation of the residuals plots. This is visible on the plots of residuals versus predicted response and the plots of residuals versus frequency.

(c) Repeat the analyses from parts (a) and (b) using ln(y) as the response. Comment on the results.

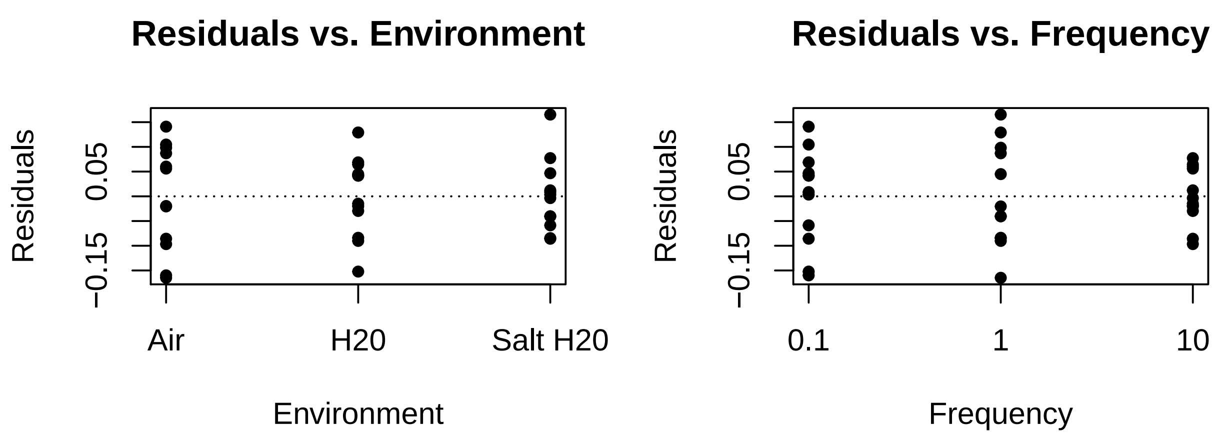
```r
environ1.aov <- aov(log(Crack.Growth) ~ Frequency * Environment, data = environ)
summary(environ1.aov)
```

```
##                       Df Sum Sq Mean Sq F value   Pr(>F)
## Frequency              2   7.57    3.79   404.1 < 2e-16 ***
## Environment            2   2.36    1.18   125.8 2.1e-14 ***
## Frequency:Environment  4   3.53    0.88    94.2 1.9e-15 ***
## Residuals             27   0.25    0.01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
par(mfrow = c(2, 2))
plot(environ1.aov)
```




```r
par(mfrow = c(1, 2))
stripchart(residuals(environ1.aov) ~ environ$Environment, vertical = TRUE, jitter = 0,
    xlab = "Environment", ylab = "Residuals", cex = 1.1, pch = 20, main = "Residuals vs. Environment")
abline(h = 0, col = "black", lty = 3)
abline(h = 0, col = "black", lty = 3)
stripchart(residuals(environ1.aov) ~ environ$Frequency, vertical = TRUE, jitter = 0,
    xlab = "Frequency", ylab = "Residuals", cex = 1.1, pch = 20, main = "Residuals vs. Frequency")
abline(h = 0, col = "black", lty = 3)
abline(h = 0, col = "black", lty = 3)
```



From the transformed data, both frequency and environment as well as their interaction are significant. The residual plots on the transformed data seems better than the original data.

7 Exercise 5.26

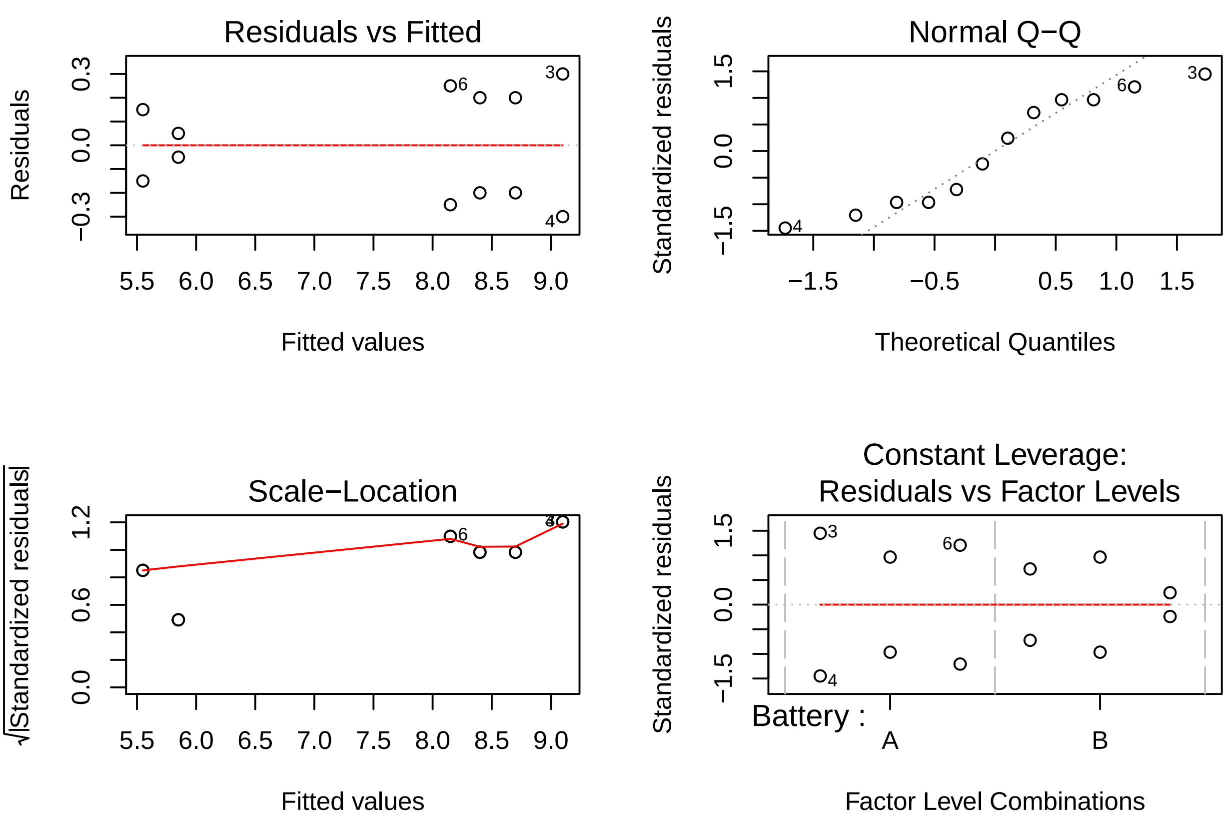
(a) Analyze the data and draw conclusions. Use $\alpha = 0.05$

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW")
battaries <- read.csv("5.26.csv")
battaries.aov <- aov(Life.Hours ~ Battery * Device, data = battaries)
summary(battaries.aov)

##                Df Sum Sq Mean Sq F value  Pr(>F)
## Battery         1   0.80    0.80    9.33   0.022 *
## Device          2  22.45   11.22  130.75 1.1e-05 ***
## Battery:Device  2   0.08    0.04    0.48   0.643
## Residuals       6   0.52    0.09
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

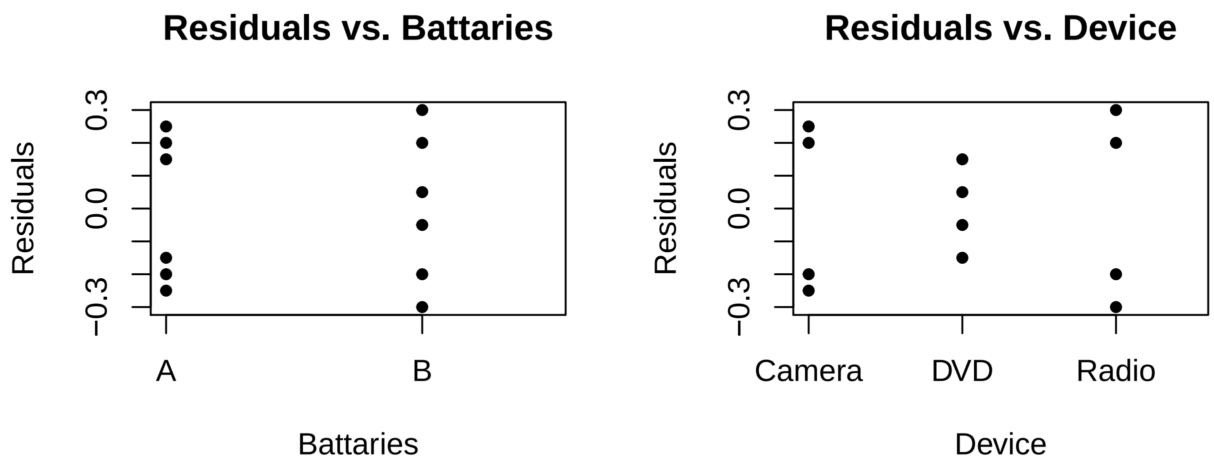
From the p-values, we can say that both the battery and device are significant (p-values are less than 0.05), but the interaction between them is not significant (p value is greater than 0.05).

(b) Investigate model adequacy by plotting the residuals.

```
par(mfrow = c(2, 2))
plot(battaries.aov)
```



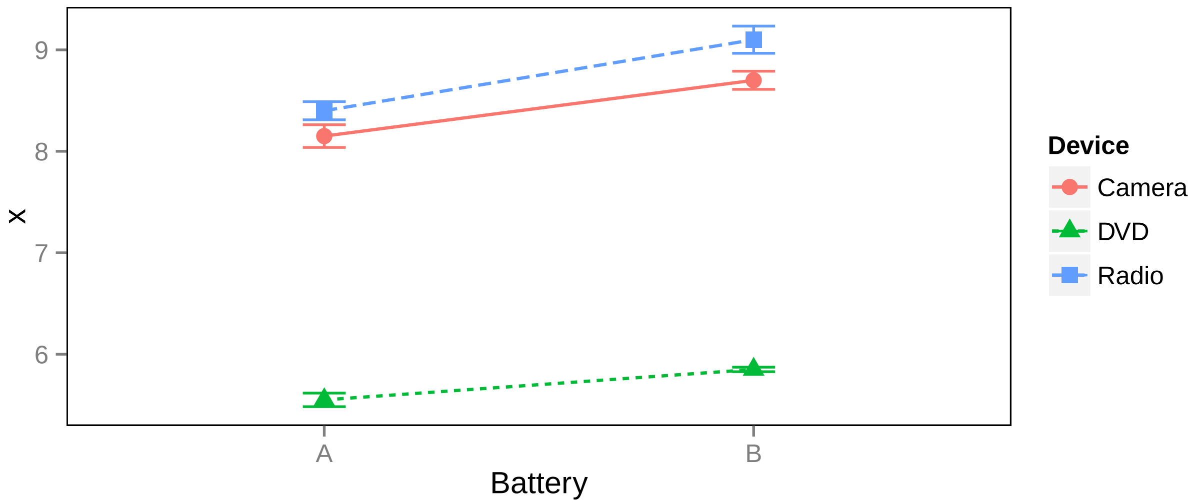

```
par(mfrow = c(1, 2))
stripchart(residuals(battaries.aov) ~ battaries$Battery, method = "stack", vertical = TRUE,
    jitter = 0, xlab = "Battaries", ylab = "Residuals", cex = 1.1, pch = 20,
    main = "Residuals vs. Battaries")

stripchart(residuals(battaries.aov) ~ battaries$Device, method = "stack", vertical = TRUE,
    jitter = 0, xlab = "Device", ylab = "Residuals", cex = 1.1, pch = 20, main = "Residuals vs. Device")
```



By observing the residual plots, no significant variation from the assumption is visible.

(c) Which brand of batteries would you recommend?

```
df <- with(battaries, aggregate(Life.Hours, list(Device = Device, Battery = Battery),
    mean))
df$se <- with(battaries, aggregate(Life.Hours, list(Device = Device, Battery = Battery),
    function(x) sd(x)/sqrt(10)))[, 3]

opar <- theme_update(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
    panel.background = element_rect(colour = "black"))
gp <- ggplot(df, aes(x = Battery, y = x, colour = Device, group = Device))
gp + geom_line(aes(linetype = Device), size = 0.6) + geom_point(aes(shape = Device),
    size = 3) + geom_errorbar(aes(ymax = x + se, ymin = x - se), width = 0.1)
```



From the interaction, we can recommend battery brand B.

8 Exercise 5.43

```
## Given
ss_tot = 1000
ss_err = 150
ss_a = 350
ss_b = 300
```

```r
ss_ab = 200
df_a = 2
df_err = 18
ms_b = 150
ms_ab = 50

ms_a = ss_a/df_a
df_b = ss_b/ms_b
df_ab = ss_ab/ms_ab
df_tot = df_a + df_b + df_ab + df_err
ms_err = ss_err/df_err
F_a = ms_a/ms_err
F_b = ms_b/ms_err
F_ab = ms_ab/ms_err
p_a = 1 - pf(F_a, df_a, df_err)
p_b = 1 - pf(F_b, df_b, df_err)
p_ab = 1 - pf(F_ab, df_ab, df_err)

rbind(df_b, df_ab, df_tot)

##        [,1]
## df_b      2
## df_ab     4
## df_tot   26

rbind(ms_a, ms_err)

##           [,1]
## ms_a   175.000
## ms_err   8.333

rbind(F_a, F_b, F_ab)

##      [,1]
## F_a    21
## F_b    18
## F_ab    6

rbind(p_a, p_b, p_ab)

##           [,1]
## p_a  1.968e-05
## p_b  5.081e-05
## p_ab 2.996e-03


## standard deviation of response variable
sd = sqrt(ms_err)
sd

## [1] 2.887
```

The ANOVA Table:

	SS	DF	MS	F	P
A	350	2	175	21	1.97E-05
B	300	2	150	18	5.08E-05
AB	200	4	50	6	0.00299
Error	150	18	8.333		
Total	1000	26			

(a) 3

(b) 4
(c) 8.333
(d) 175
(e) 3
(f) The p-values for A, B and the interaction are very low (less than 0.05). So, the effect of A, B and interaction are significant.
(g) 2.886
(h) 2

9 Exercise 5.44

```
## Given
ss_tot = 1000
ss_block = 60
ss_a = 350
ss_b = 300
ss_ab = 200
n = 3

ss_err = ss_tot - ss_block - ss_a - ss_b - ss_ab
df_block = n - 1
ms_block = ss_block/df_block

## From 5.43
df_a = 2
ms_a = 175
df_b = 2
ms_b = 150
df_ab = 4
ms_ab = 50
df_tot = 26

df_err = df_tot - df_a - df_b - df_ab - df_block
ms_err = ss_err/df_err
F_a = ms_a/ms_err
F_b = ms_b/ms_err
F_ab = ms_ab/ms_err
F_block = ms_block/ms_err
p_a = 1 - pf(F_a, df_a, df_err)
p_b = 1 - pf(F_b, df_b, df_err)
p_block = 1 - pf(F_block, df_block, df_err)

ss_err

## [1] 90

rbind(df_block, df_err, df_tot)

##          [,1]
## df_block    2
## df_err     16
## df_tot     26

rbind(ms_block, ms_err)

##            [,1]
## ms_block 30.000
## ms_err    5.625

F_block
```

```
## [1] 5.333

rbind(p_block, p_a, p_b)

##              [,1]
## p_block 1.680e-02
## p_a     3.064e-06
## p_b     8.043e-06
```

The ANOVA Table:

	SS	DF	MS	F	P
Blocks	60	2	30	5.333	1.68E-02
A	350	2	175	31.111	3.06E-06
B	300	2	150	26.667	8.04E-06
AB	200	4	50	8.889	
Error	90	16	5.625		
Total	1000	26			