STAT 522 —— Assignment 5

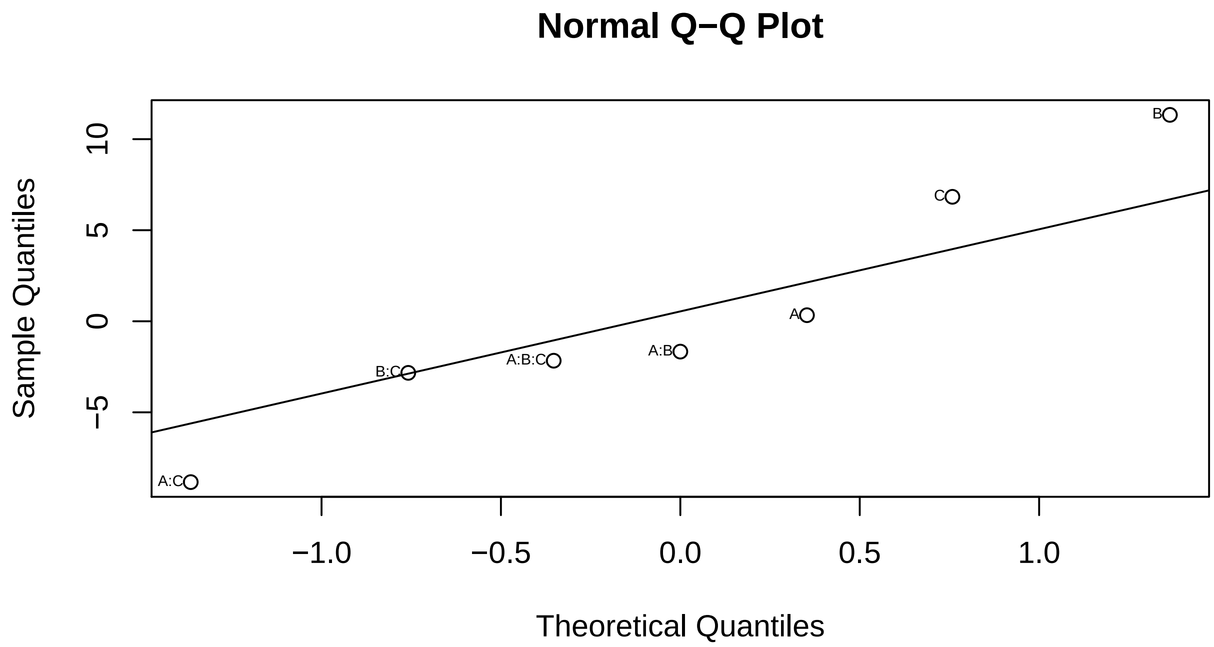
Subasish Das (sxd1684)

10th April 2014

1 Exercise 6.1

(a)

```
## Given
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5/effects")
tool <- read.csv("6.1a.csv")
mydata.lm = lm(Life.Hours ~ A * B * C, tool)
n = 3
k = 3
effects = coefficients(mydata.lm)[-c(1)] * 2
SS = effects^2 * n * 2^(k - 2)
percentage = SS/sum(SS) * 100
tem = qqnorm(effects)
qqline(effects)
text(tem$x, tem$y, names(effects), pos = 2, offset = 0.2, cex = 0.5)
```




```
cbind(effects, SS, percentage)

##        effects       SS percentage
## A       0.3333   0.6667    0.04134
## B      11.3333 770.6667   47.78834
## C       6.8333 280.1667   17.37288
## A:B    -1.6667  16.6667    1.03348
## A:C    -8.8333 468.1667   29.03059
## B:C    -2.8333  48.1667    2.98677
## A:B:C  -2.1667  28.1667    1.74659
```

From the effect values and normality effect plot, it's observed that factors B, C and interaction AC are seemed to be significant.

(b)

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5")
tool <- read.csv("6.1.csv")
hours <- tool$Life.Hours

tool[, 1:3] <- lapply(tool[, 1:3], factor)
tool1.aov <- aov(Life.Hours ~ A.Cutting.Speed * B.Tool.Geometry * C.Cutting.Angle,
    data = tool)
summary(tool1.aov)

##                                              Df Sum Sq Mean Sq F value
## A.Cutting.Speed                               1      1       1    0.02
## B.Tool.Geometry                               1    771     771   25.55
## C.Cutting.Angle                               1    280     280    9.29
## A.Cutting.Speed:B.Tool.Geometry               1     17      17    0.55
## A.Cutting.Speed:C.Cutting.Angle               1    468     468   15.52
## B.Tool.Geometry:C.Cutting.Angle               1     48      48    1.60
## A.Cutting.Speed:B.Tool.Geometry:C.Cutting.Angle  1     28      28    0.93
## Residuals                                    16    483      30
##                                               Pr(>F)
## A.Cutting.Speed                              0.88368
## B.Tool.Geometry                              0.00012 ***
## C.Cutting.Angle                              0.00768 **
## A.Cutting.Speed:B.Tool.Geometry              0.46808
## A.Cutting.Speed:C.Cutting.Angle              0.00117 **
## B.Tool.Geometry:C.Cutting.Angle              0.22448
## A.Cutting.Speed:B.Tool.Geometry:C.Cutting.Angle 0.34828
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA table confirms the significance of factors B, C and interaction AC.

The ANOVA of the reduced model is performed below:

```
tool2.aov <- aov(Life.Hours ~ A.Cutting.Speed + B.Tool.Geometry + C.Cutting.Angle +
    A.Cutting.Speed * C.Cutting.Angle, data = tool)
summary(tool2.aov)

##                                 Df Sum Sq Mean Sq F value  Pr(>F)
## A.Cutting.Speed                  1      1       1    0.02  0.8836
## B.Tool.Geometry                  1    771     771   25.44 7.2e-05 ***
## C.Cutting.Angle                  1    280     280    9.25  0.0067 **
## A.Cutting.Speed:C.Cutting.Angle  1    468     468   15.45  0.0009 ***
## Residuals                       19    576      30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Factor A is kept here to maintain the hierachy. The factors B, C and interaction AC are significant at 0.01 level.

(c)

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5")
tool_num <- read.csv("6.1.csv")
tool_num.lm <- lm(Life.Hours ~ A.Cutting.Speed + B.Tool.Geometry + C.Cutting.Angle +
    A.Cutting.Speed * C.Cutting.Angle, data = tool_num)
summary(tool_num.lm)
```

```
##
## Call:
## lm(formula = Life.Hours ~ A.Cutting.Speed + B.Tool.Geometry +
##     C.Cutting.Angle + A.Cutting.Speed * C.Cutting.Angle, data = tool_num)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.333 -4.375 -0.417  2.958 11.500
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       40.833      1.124   36.34  < 2e-16 ***
## A.Cutting.Speed                    0.167      1.124    0.15   0.8836
## B.Tool.Geometry                    5.667      1.124    5.04  7.2e-05 ***
## C.Cutting.Angle                    3.417      1.124    3.04   0.0067 **
## A.Cutting.Speed:C.Cutting.Angle   -4.417      1.124   -3.93   0.0009 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.5 on 19 degrees of freedom
## Multiple R-squared:  0.725,Adjusted R-squared:  0.667
## F-statistic: 12.5 on 4 and 19 DF,  p-value: 3.69e-05
```

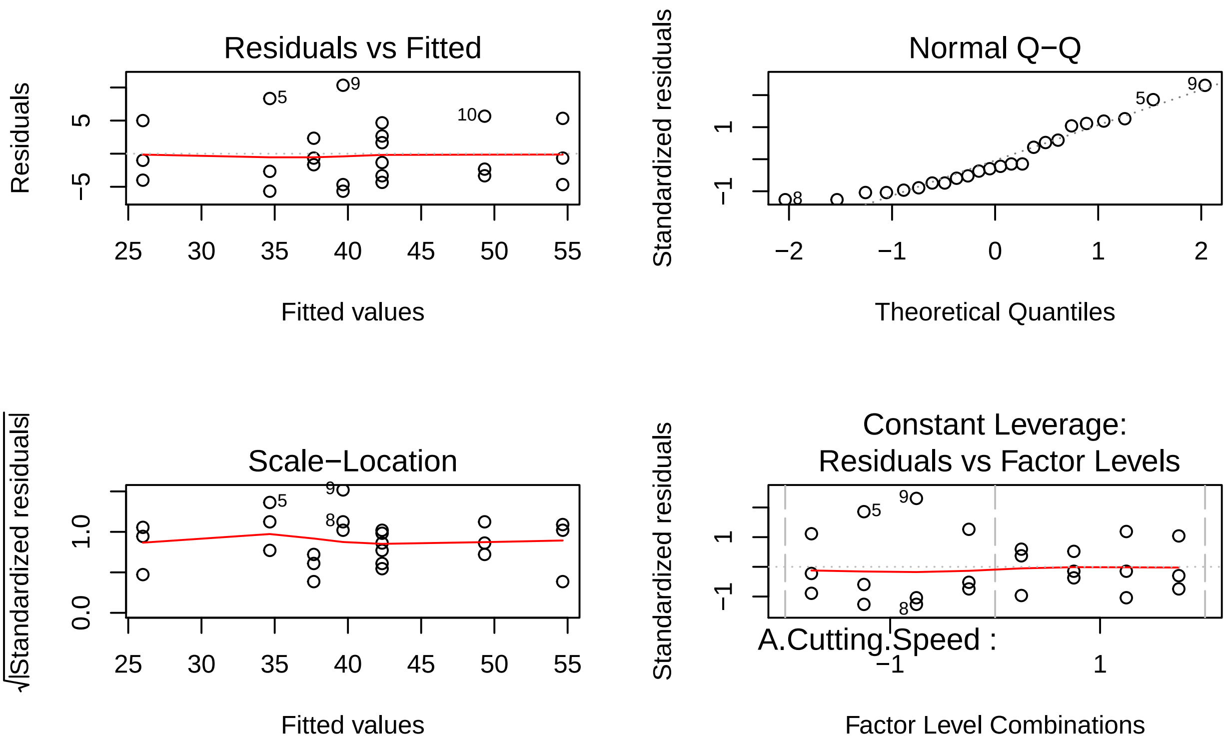
The regression model:

y_{ijk}= 40.8333+ 0.1667x_A + 5.667x_B + 3.417x_C -4.4167$x_A x_C$

The regression model is based on the significant factors B (tool geometry), C (cutting angle) and interaction of AC.

(d)

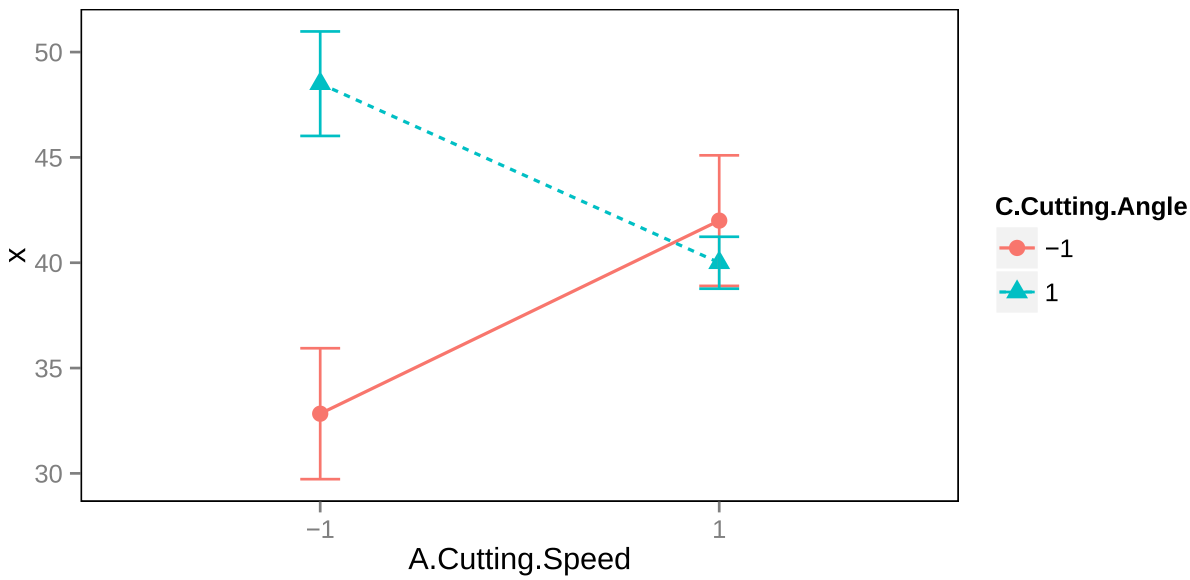
```
par(mfrow = c(2, 2))
plot(tool1.aov)
```



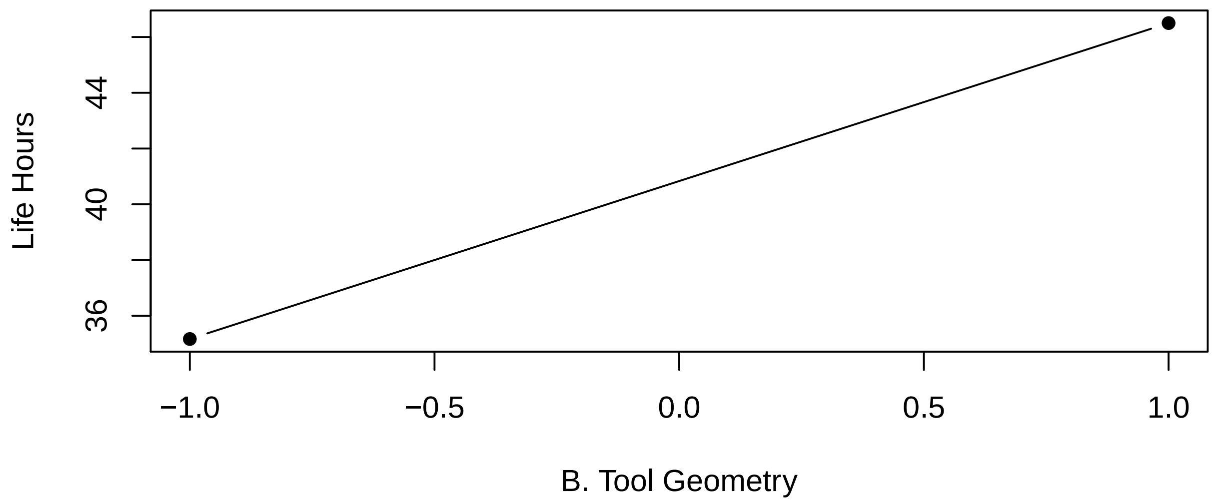
Nothing unusual is visible regarding the residual plots.

(e)

```
## Interaction Plot
tool[, 1:3] <- lapply(tool[, 1:3], factor)
library(ggplot2)
df <- with(tool, aggregate(Life.Hours, list(C.Cutting.Angle = C.Cutting.Angle,
    A.Cutting.Speed = A.Cutting.Speed), mean))
df$se <- with(tool, aggregate(Life.Hours, list(C.Cutting.Angle = C.Cutting.Angle,
    A.Cutting.Speed = A.Cutting.Speed), 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 = A.Cutting.Speed, y = x, colour = C.Cutting.Angle, group = C.Cutting.Angle))
gp + geom_line(aes(linetype = C.Cutting.Angle), size = 0.6) + geom_point(aes(shape = C.Cutting.Angle),
    size = 3) + geom_errorbar(aes(ymax = x + se, ymin = x - se), width = 0.1)
```



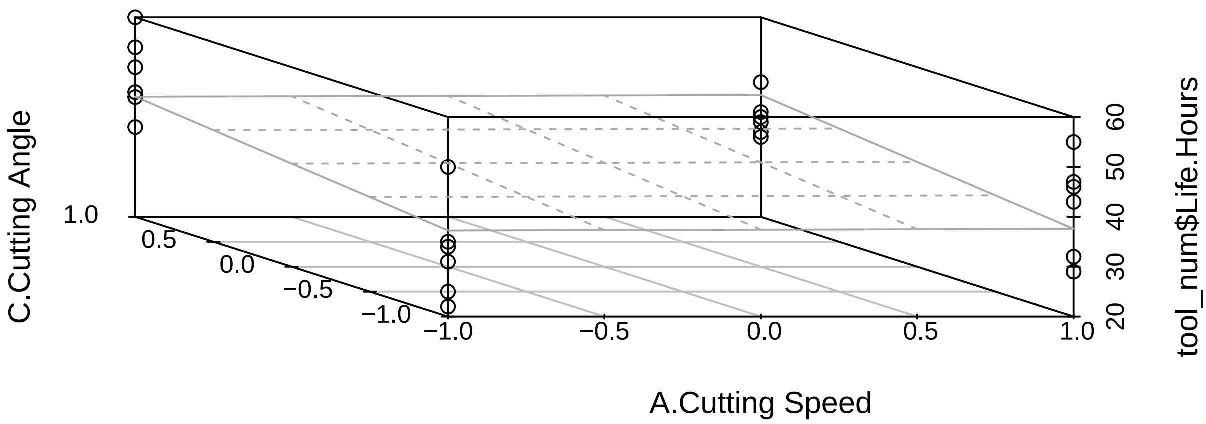

```
## One Factor Plot
with(tool, plot(c(-1, 1), tapply(Life.Hours, B.Tool.Geometry, mean), type = "b",
    pch = 16, xlab = "B. Tool Geometry", ylab = "Life Hours"))
```



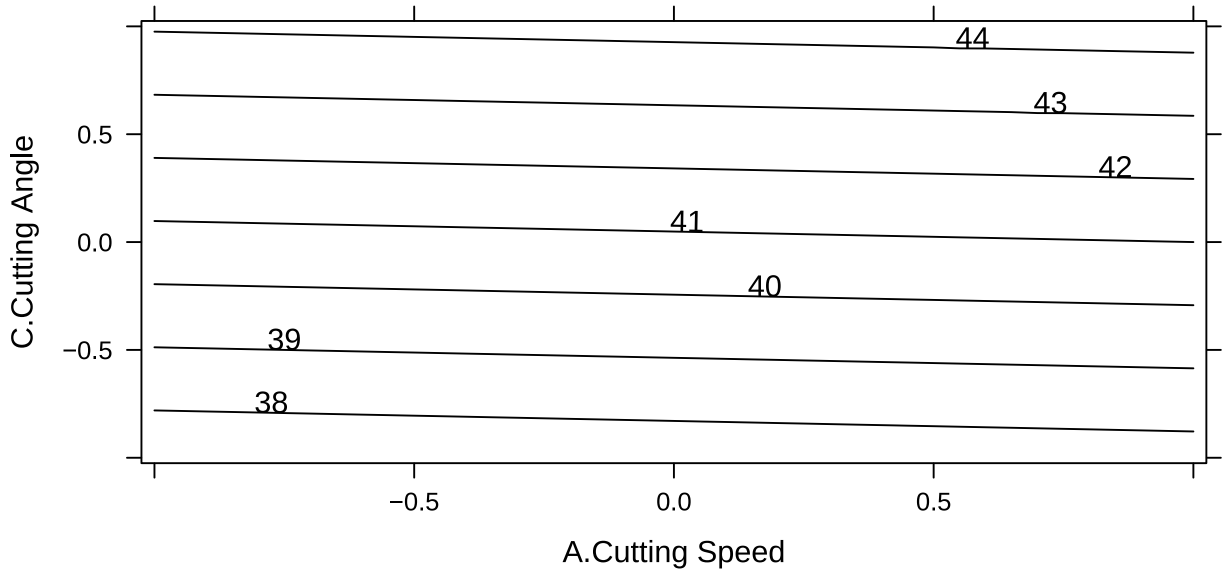
As B has positive effect, we can set B at a high level to increase the life hours. The interaction plot of AC indicates that life hours will be maximum at higher level of C and lower level of A.

2 Exercise 6.2

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5")
tool_num <- read.csv("6.1.csv")
tool_num.lm <- lm(Life.Hours ~ A.Cutting.Speed + C.Cutting.Angle, data = tool_num)


library(scatterplot3d)
s3d <- scatterplot3d(tool_num$A.Cutting.Speed, tool_num$C.Cutting.Angle, tool_num$Life.Hours,
    type = "p", angle = 135, scale.y = 1, xlab = "A.Cutting Speed", ylab = "C.Cutting Angle")
s3d$plane3d(tool_num.lm, lty.box = "solid", col = "darkgray")
```




```
tmp <- list(A.Cutting.Speed = seq(-1, 1, by = 0.05), C.Cutting.Angle = seq(-1,
    1, by = 0.05))
new.data <- expand.grid(tmp)
new.data$fit <- predict(tool_num.lm, new.data)
library(lme4)

## Loading required package:  lattice
## Loading required package:  Matrix
##
## Attaching package:  'lme4'
## The following object is masked from 'package:ggplot2':
##
##    fortify

contourplot(fit ~ A.Cutting.Speed * C.Cutting.Angle, new.data, xlab = "A.Cutting Speed",
    ylab = "C.Cutting Angle")
```



The response surface plot and the contour plot are generated by using the regression model in problem 6.1 part (c). The curvature is visible due to the interaction of AC. Yes, these plots provide insight regarding the desirable operating conditions for this process.

3 Exercise 6.3

```
## Standard Error
n = 3
k = 3
S2 = 30.17
SE_effect = sqrt(S2/(n * 2^(k - 2)))
SE_effect

## [1] 2.242


## 95% Confidence Limit
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5")
tool <- read.csv("6.1.csv")
tool[, 1:3] <- lapply(tool[, 1:3], factor)
aov <- aov(Life.Hours ~ A.Cutting.Speed * B.Tool.Geometry * C.Cutting.Angle,
    data = tool)
confint(aov)

##                                                      2.5 %  97.5 %
## (Intercept)                                        19.2777 32.7223
## A.Cutting.Speed1                                   -0.8401 18.1735
## B.Tool.Geometry1                                    4.1599 23.1735
## C.Cutting.Angle1                                    6.8265 25.8401
## A.Cutting.Speed1:B.Tool.Geometry1                 -12.4446 14.4446
## A.Cutting.Speed1:C.Cutting.Angle1                 -26.7780  0.1113
## B.Tool.Geometry1:C.Cutting.Angle1                 -14.7780 12.1113
## A.Cutting.Speed1:B.Tool.Geometry1:C.Cutting.Angle1 -27.6803 10.3469
```

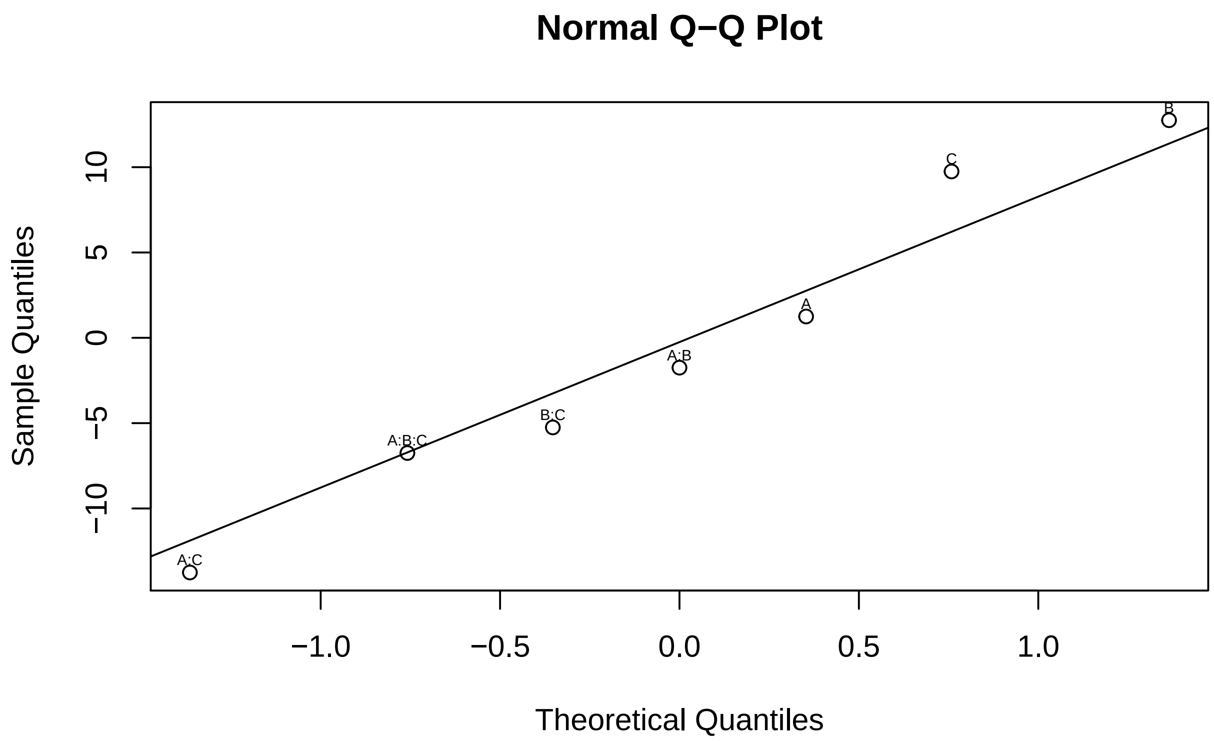
The standard error is 2.24.

The 95 percent confidence interval of the factors B and C and the interation AC don't contain zero. This completely aggrees of the varince approach.

4 Exercise 6.6

(a)

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5/effects")
tool <- read.csv("6.6a.csv")
mydata.lm = lm(Life.Hours ~ A * B * C, tool)
n = 1
k = 3
effects = coefficients(mydata.lm)[-c(1)] * 2
SS = effects^2 * n * 2^(k - 2)
percentage = SS/sum(SS) * 100
tem = qqnorm(effects)
qqline(effects)
text(tem$x, tem$y, names(effects), pos = 3, offset = 0.2, cex = 0.5)
```




```
cbind(effects, SS, percentage)

##       effects      SS percentage
## A        1.25   3.125     0.2979
## B       12.75 325.125    30.9975
## C        9.75 190.125    18.1266
## A:B     -1.75   6.125     0.5840
## A:C    -13.75 378.125    36.0505
## B:C     -5.25  55.125     5.2556
## A:B:C   -6.75  91.125     8.6879
```

From the normality plot for effects, factors B, C and interaction AC have larger effects.

(b)

```
## ANOVA including a check for pure quadratic curvature.
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5")
toola <- read.csv("6.6.csv")
```

```
tool1a.aov <- aov(Life.Hours ~ A.Cutting.Speed * B.Tool.Geometry * C.Cutting.Angle +
    I((A.Cutting.Speed)^2), data = toola)
summary(tool1a.aov)

##                                                 Df Sum Sq Mean Sq F value
## A.Cutting.Speed                                  1      3       3    0.20
## B.Tool.Geometry                                  1    325     325   21.20
## C.Cutting.Angle                                  1    190     190   12.40
## I((A.Cutting.Speed)^2)                           1      0       0    0.00
## A.Cutting.Speed:B.Tool.Geometry                  1      6       6    0.40
## A.Cutting.Speed:C.Cutting.Angle                  1    378     378   24.66
## B.Tool.Geometry:C.Cutting.Angle                  1     55      55    3.60
## A.Cutting.Speed:B.Tool.Geometry:C.Cutting.Angle  1     91      91    5.94
## Residuals                                        3     46      15
##                                                 Pr(>F)
## A.Cutting.Speed                                  0.682
## B.Tool.Geometry                                  0.019 *
## C.Cutting.Angle                                  0.039 *
## I((A.Cutting.Speed)^2)                           0.962
## A.Cutting.Speed:B.Tool.Geometry                  0.572
## A.Cutting.Speed:C.Cutting.Angle                  0.016 *
## B.Tool.Geometry:C.Cutting.Angle                  0.154
## A.Cutting.Speed:B.Tool.Geometry:C.Cutting.Angle  0.093 .
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Inclusion of a pure quadratic curvature shows no significance. The ANOVA table confirms that the effects of the factors B, C and intertion AC are larger.

The ANOVA of the reduced model is performed below:

```
tool2a.aov <- aov(Life.Hours ~ A.Cutting.Speed + B.Tool.Geometry + C.Cutting.Angle +
    A.Cutting.Speed * C.Cutting.Angle, data = toola)
summary(tool2a.aov)

##                                 Df Sum Sq Mean Sq F value Pr(>F)
## A.Cutting.Speed                  1      3       3    0.11 0.7496
## B.Tool.Geometry                  1    325     325   11.47 0.0117 *
## C.Cutting.Angle                  1    190     190    6.71 0.0360 *
## A.Cutting.Speed:C.Cutting.Angle  1    378     378   13.34 0.0082 **
## Residuals                        7    198      28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Factor A is kept here to maintain the hierachy. The factors B, C and interaction AC are significant at 0.05 level.

(c)

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5")
tool_numa <- read.csv("6.6.csv")
tool_numa.aov <- lm(Life.Hours ~ A.Cutting.Speed + B.Tool.Geometry + C.Cutting.Angle +
    A.Cutting.Speed * C.Cutting.Angle, data = tool_numa)
summary(tool_numa.aov)

##
## Call:
## lm(formula = Life.Hours ~ A.Cutting.Speed + B.Tool.Geometry +
##     C.Cutting.Angle + A.Cutting.Speed * C.Cutting.Angle, data = tool_numa)
##
## Residuals:
```

```
##    Min      1Q Median      3Q     Max
## -6.917 -2.479 -0.042  2.583  6.833
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          40.917      1.537   26.62  2.7e-08 ***
## A.Cutting.Speed                       0.625      1.882    0.33   0.7496
## B.Tool.Geometry                       6.375      1.882    3.39   0.0117 *
## C.Cutting.Angle                       4.875      1.882    2.59   0.0360 *
## A.Cutting.Speed:C.Cutting.Angle   -6.875      1.882   -3.65   0.0082 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.32 on 7 degrees of freedom
## Multiple R-squared:  0.819,Adjusted R-squared:  0.715
## F-statistic: 7.91 on 4 and 7 DF,  p-value: 0.00979
```

The regression model:

y_{ijk}= 40.917+ 0.625x_A + 6.375x_B + 4.875x_C -6.875$x_A x_C$

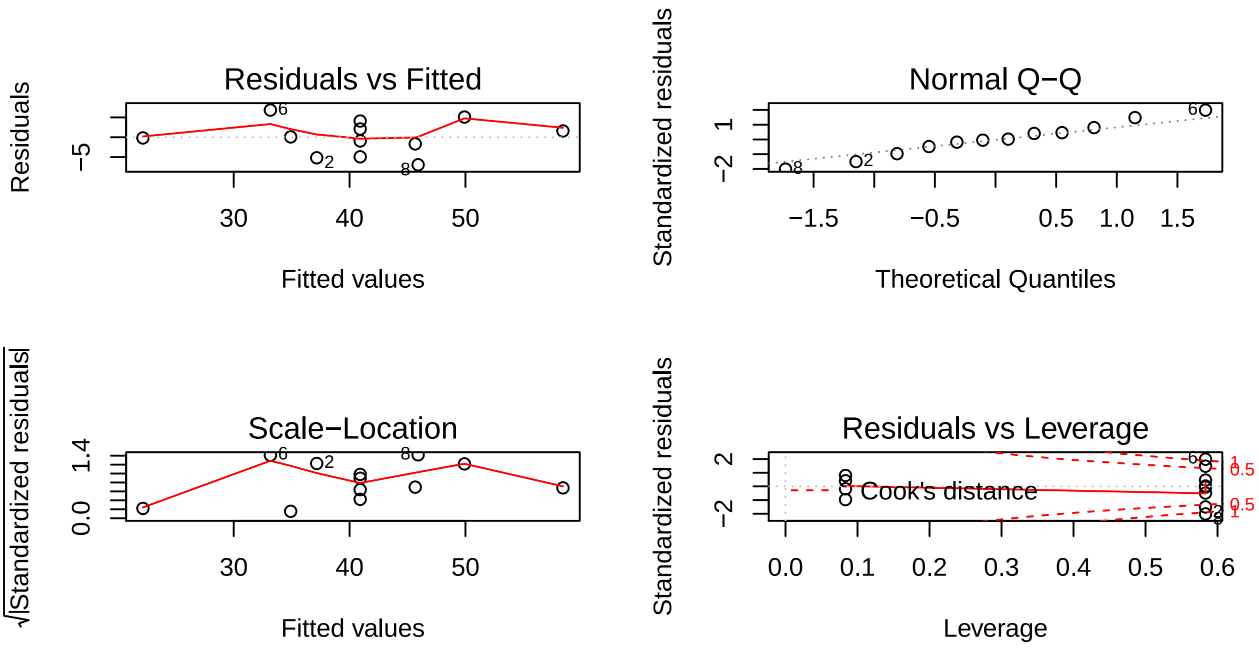
The regression model from 6.1 part (c):

y_{ijk}= 40.8333+ 0.1667x_A + 5.667x_B + 3.417x_C -4.4167$x_A x_C$

The regression model is based on the significant factors B,C and interaction of AC. The current model is not subtantially different from the regression model of problem 6.1 part (c)

(d)

```
par(mfrow = c(2, 2))
plot(tool2a.aov)
```

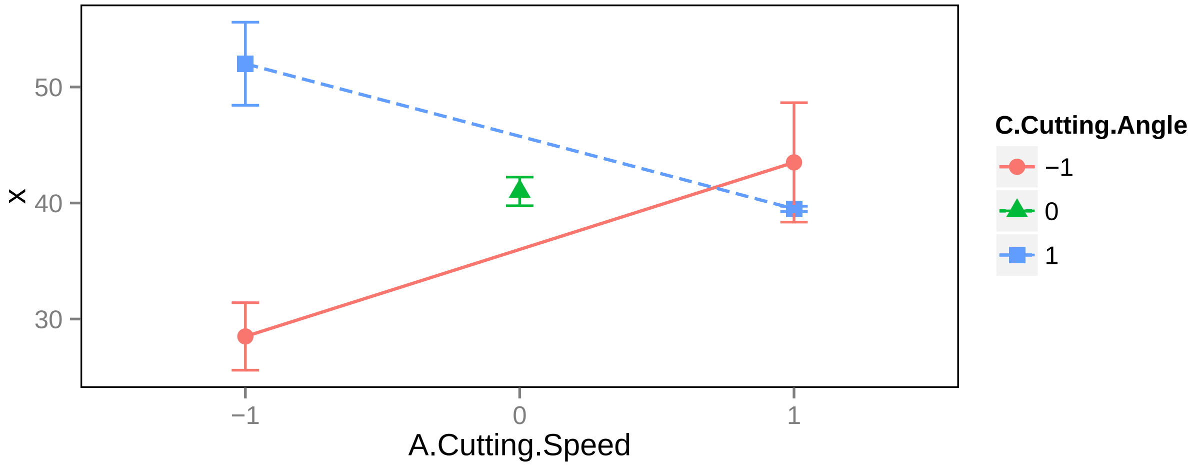


Nothing unusual is drastically visible regarding the residual plots.

(e)

```
## Interaction Graph
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5")
toola <- read.csv("6.6.csv")
toola[, 1:3] <- lapply(toola[, 1:3], factor)
library(ggplot2)
```

```r
df <- with(toola, aggregate(Life.Hours, list(C.Cutting.Angle = C.Cutting.Angle,
    A.Cutting.Speed = A.Cutting.Speed), mean))
df$se <- with(toola, aggregate(Life.Hours, list(C.Cutting.Angle = C.Cutting.Angle,
    A.Cutting.Speed = A.Cutting.Speed), 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 = A.Cutting.Speed, y = x, colour = C.Cutting.Angle, group = C.Cutting.Angle))
gp + geom_line(aes(linetype = C.Cutting.Angle), size = 0.6) + geom_point(aes(shape = C.Cutting.Angle),
    size = 3) + geom_errorbar(aes(ymax = x + se, ymin = x - se), width = 0.1)
```

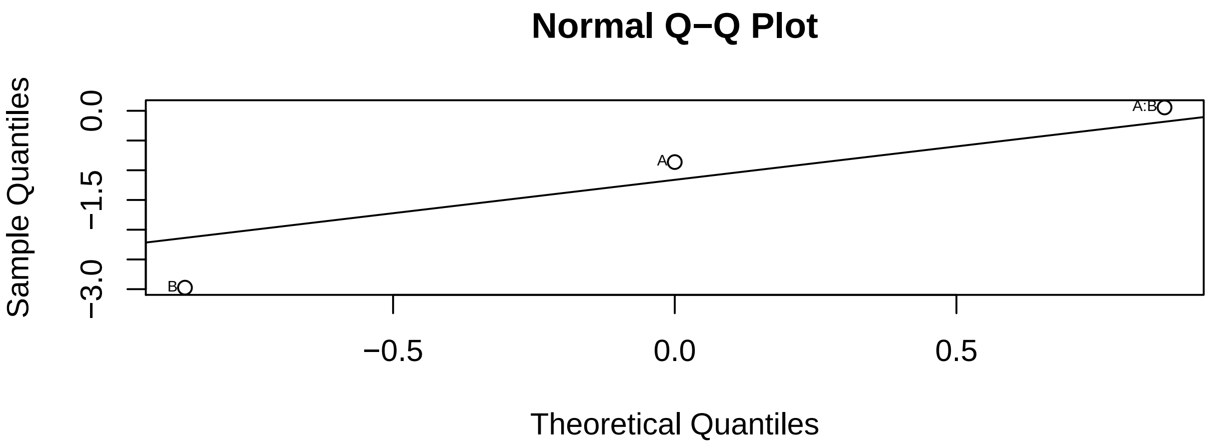


As B has positive effect, we can set at a high level to increase the life hours. The interaction plot of AC life hours will be maximum at higher level of C and lower level of A.

5 Exercise 6.12

(a)

```r
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5/effects")
circuit <- read.csv("6.12a.csv")
mydata.lm = lm(Thickness ~ A * B, circuit)
n = 2
k = 4
effects = coefficients(mydata.lm)[-c(1)] * 2
SS = effects^2 * n * 2^(k - 2)
percentage = SS/sum(SS) * 100
cbind(effects, SS, percentage)

##     effects       SS percentage
## A   -0.8624  5.94953    7.75025
## B   -2.9747 70.79072   92.21672
## A:B  0.0563  0.02536    0.03303

tem = qqnorm(effects)
qqline(effects)
text(tem$x, tem$y, names(effects), pos = 2, offset = 0.2, cex = 0.5)
```



From the factor effect values and the normality effect plot, we find that factor B (Deposition time) has significant effect.

(b)

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5")
circuit <- read.csv("6.12.csv")
circuit.aov <- aov(Thickness ~ A.Flow.Rate * B.Dep.Time, data = circuit)
summary(circuit.aov)

##                        Df Sum Sq Mean Sq F value Pr(>F)
## A.Flow.Rate             1   0.40   0.403    1.26   0.28
## B.Dep.Time              1   1.37   1.374    4.31   0.06 .
## A.Flow.Rate:B.Dep.Time  1   0.32   0.317    0.99   0.34
## Residuals              12   3.83   0.319
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the values of the ANOVA table, we find that factor B (Deposition time) has significant effect at 0.1 level.

(c)

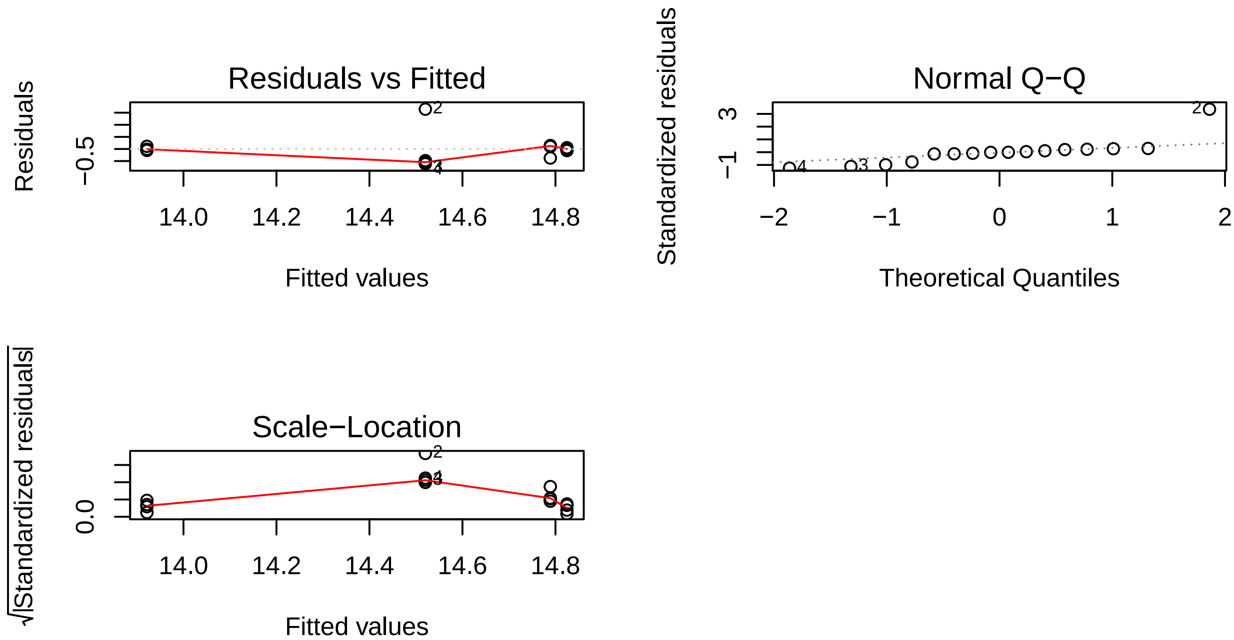
```
circuit.lm <- lm(Thickness ~ A.Flow.Rate * B.Dep.Time, data = circuit)
summary(circuit.lm)

##
## Call:
## lm(formula = Thickness ~ A.Flow.Rate * B.Dep.Time, data = circuit)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6133 -0.1443 -0.0056  0.1019  1.6447
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             37.6266    20.5334    1.83    0.092 .
## A.Flow.Rate             -0.4312     0.3600   -1.20    0.254
## B.Dep.Time              -1.4874     1.6108   -0.92    0.374
## A.Flow.Rate:B.Dep.Time   0.0282     0.0282    1.00    0.339
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.565 on 12 degrees of freedom
## Multiple R-squared:  0.353,Adjusted R-squared:  0.192
## F-statistic: 2.19 on 3 and 12 DF,  p-value: 0.142
```

The regression equation:

y= 37.626- 0.432x_A - 1.487x_B + 0.028$x_A x_B$

(d)

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

## hat values (leverages) are all = 0.25
##  and there are no factor predictors; no plot no.  5
```

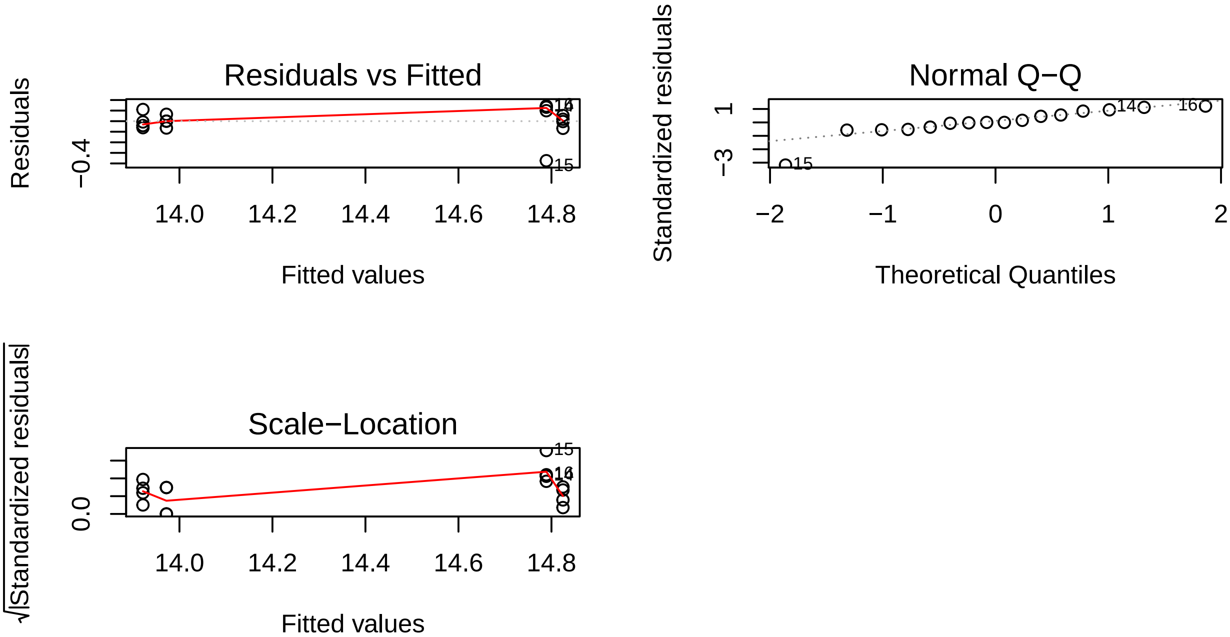


From the residual plots, it's found that observaion no. 2 (16.165) falls outside the groupings in the normal probability plot. This observation is also visible outide the groupings in residual versus predicted plot.

(e)
We can deal with the potential outlier by replacing that observation with the avergae of the observations from that particular cell. Another way is to validate that particular point. If validation produces finds out any error, it should be corrected.

Obsevation 2 is replaced by the average of the remaining three other runs in the cell which is 13.972.

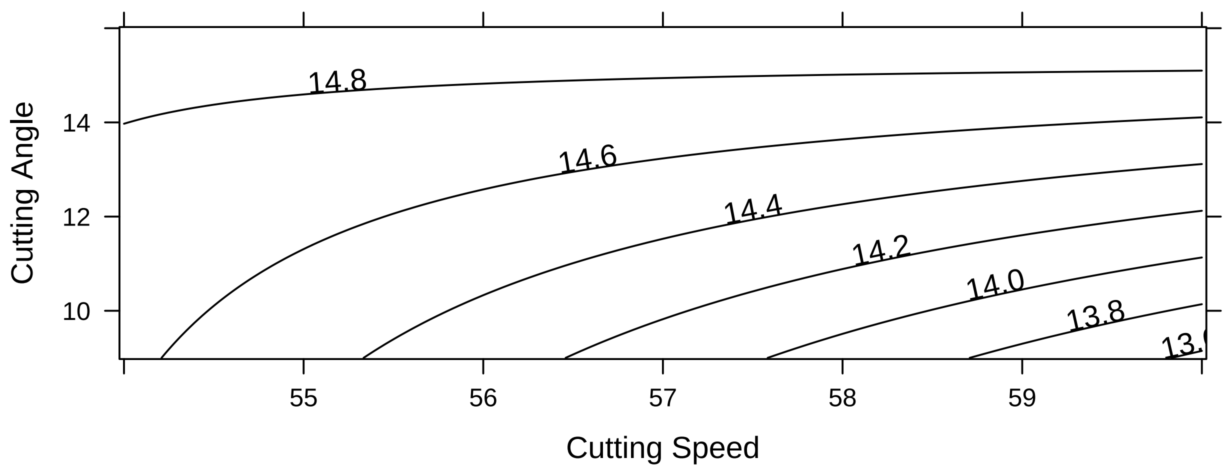
```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5")
circuit_1 <- read.csv("6.12_1.csv")
circuit_1.aov <- aov(Thickness ~ A.Flow.Rate * B.Dep.Time, data = circuit_1)
summary(circuit_1.aov)

##                         Df Sum Sq Mean Sq F value  Pr(>F)
## A.Flow.Rate              1  0.007   0.007    0.40    0.54
## B.Dep.Time               1  2.959   2.959  160.29 2.7e-08 ***
## A.Flow.Rate:B.Dep.Time   1  0.000   0.000    0.01    0.92
## Residuals               12  0.222   0.018
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

par(mfrow = c(2, 2))
plot(circuit_1.aov)

## hat values (leverages) are all = 0.25
##  and there are no factor predictors; no plot no.  5
```



Another outlier is visible again by checking the normality plot and residual versus predicted plot. It's required to be taken care of again.

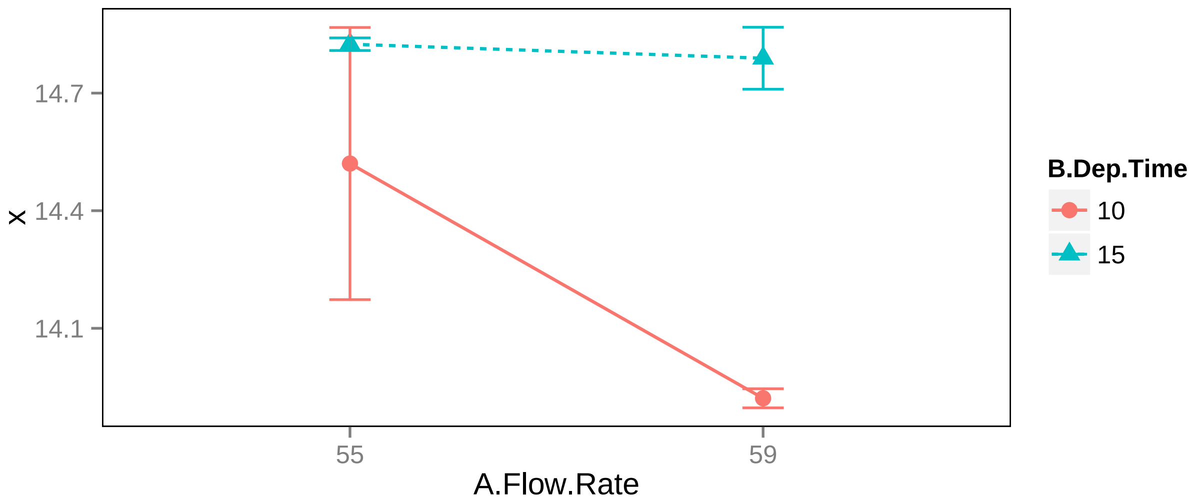
6 Exercise 6.13

```
tmp <- list(A.Flow.Rate = seq(54, 60, by = 0.05), B.Dep.Time = seq(9, 16, by = 0.05))
new.data <- expand.grid(tmp)
new.data$fit <- predict(circuit.lm, new.data)
contourplot(fit ~ A.Flow.Rate + B.Dep.Time, new.data, xlab = "Cutting Speed",
    ylab = "Cutting Angle")
```




```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5")
circuit <- read.csv("6.12.csv")
circuit[, 1:2] <- lapply(circuit[, 1:2], factor)
library(ggplot2)
df <- with(circuit, aggregate(Thickness, list(B.Dep.Time = B.Dep.Time, A.Flow.Rate = A.Flow.Rate),
    mean))
df$se <- with(circuit, aggregate(Thickness, list(B.Dep.Time = B.Dep.Time, A.Flow.Rate = A.Flow.Rate),
```

```
    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 = A.Flow.Rate, y = x, colour = B.Dep.Time, group = B.Dep.Time))
gp + geom_line(aes(linetype = B.Dep.Time), size = 0.6) + geom_point(aes(shape = B.Dep.Time),
    size = 3) + geom_errorbar(aes(ymax = x + se, ymin = x - se), width = 0.1)
```



By observing the contour plot and interaction plot, the deposition time can be recommended at 12.4 minutes (to obtain the required layer thickness). On the other hand the arsenic flow can be set at any of the experiment levels.

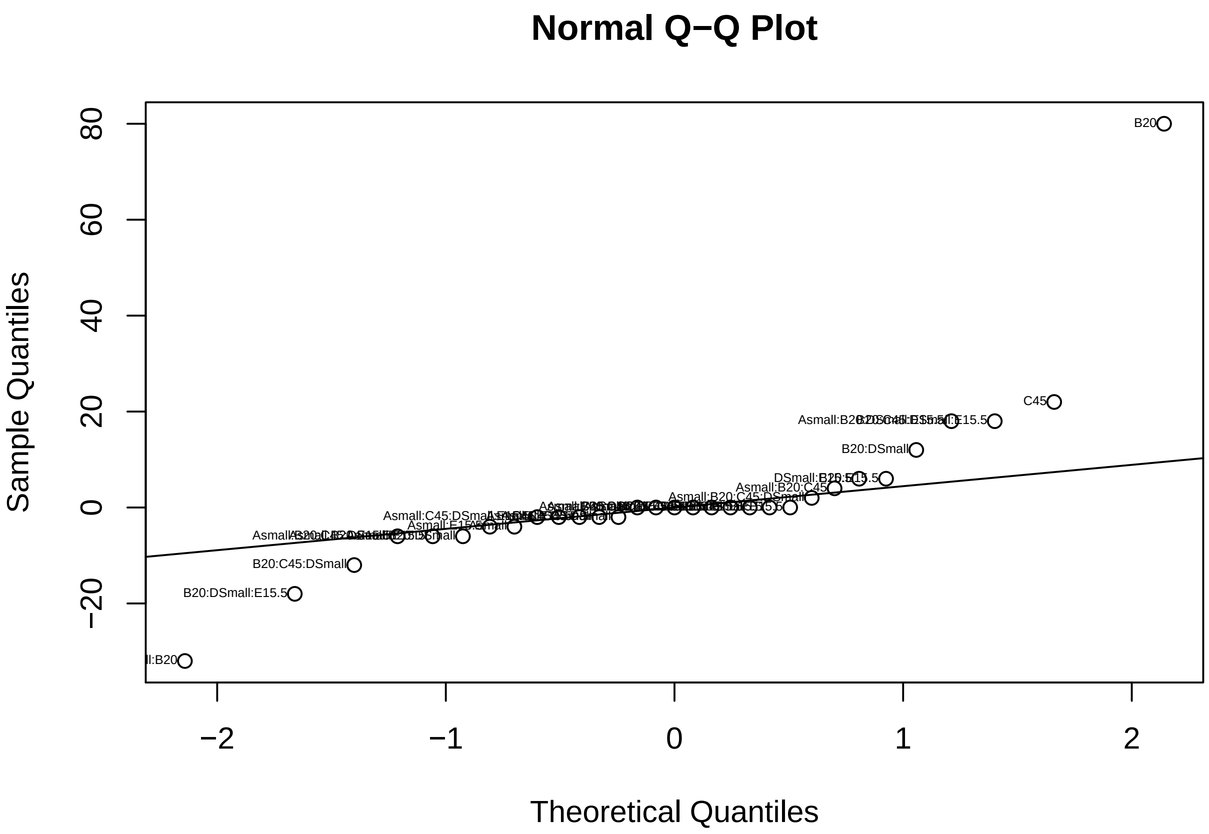
7 Exercise 6.14

From the countour plot and interaction plot of Problem 6.13, we observe that when the proccess would be run at a higher level of deposition time, there would be no change in the thickness value with the change of arsenic flow rate.

8 Exercise 6.26

(a)

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5/effects")
semicon <- read.csv("6.26a.csv")
semicon[, 1:5] <- lapply(semicon[, 1:5], factor)
mydata.lm = lm(Yield ~ A * B * C * D * E, data = semicon)
n = 1
k = 5
effects = coefficients(mydata.lm)[-c(1)] * 2
SS = effects^2 * n * 2^(k - 2)
percentage = SS/sum(SS) * 100
tem = qqnorm(effects)
qqline(effects)
text(tem$x, tem$y, names(effects), pos = 2, offset = 0.2, cex = 0.45)
```




```r
cbind(effects, SS, percentage)
```

```
##                              effects        SS percentage
## Asmall                    -4.000e+00 1.280e+02  1.699e-01
## B20                        8.000e+01 5.120e+04  6.794e+01
## C45                        2.200e+01 3.872e+03  5.138e+00
## DSmall                    -2.000e+00 3.200e+01  4.246e-02
## E15.5                      8.478e-15 5.751e-28  7.631e-31
## Asmall:B20                -3.200e+01 8.192e+03  1.087e+01
## Asmall:C45                -2.000e+00 3.200e+01  4.246e-02
## B20:C45                    7.262e-15 4.219e-28  5.598e-31
## Asmall:DSmall             -1.479e-13 1.750e-25  2.322e-28
## B20:DSmall                 1.200e+01 1.152e+03  1.529e+00
## C45:DSmall                -4.145e-14 1.375e-26  1.824e-29
## Asmall:E15.5              -4.000e+00 1.280e+02  1.699e-01
## B20:E15.5                  6.000e+00 2.880e+02  3.822e-01
## C45:E15.5                 -2.000e+00 3.200e+01  4.246e-02
## DSmall:E15.5               6.000e+00 2.880e+02  3.822e-01
## Asmall:B20:C45             4.000e+00 1.280e+02  1.699e-01
## Asmall:B20:DSmall         -6.000e+00 2.880e+02  3.822e-01
## Asmall:C45:DSmall         -2.000e+00 3.200e+01  4.246e-02
## B20:C45:DSmall            -1.200e+01 1.152e+03  1.529e+00
## Asmall:B20:E15.5          -6.000e+00 2.880e+02  3.822e-01
## Asmall:C45:E15.5          -7.154e-15 4.094e-28  5.433e-31
## B20:C45:E15.5             -3.219e-15 8.288e-29  1.100e-31
## Asmall:DSmall:E15.5        1.548e-13 1.917e-25  2.544e-28
## B20:DSmall:E15.5          -1.800e+01 2.592e+03  3.439e+00
## C45:DSmall:E15.5           4.496e-14 1.617e-26  2.146e-29
## Asmall:B20:C45:DSmall      2.000e+00 3.200e+01  4.246e-02
## Asmall:B20:C45:E15.5      -7.693e-15 4.735e-28  6.283e-31
## Asmall:B20:DSmall:E15.5    1.800e+01 2.592e+03  3.439e+00
```

```
## Asmall:C45:DSmall:E15.5     -2.000e+00 3.200e+01  4.246e-02
## B20:C45:DSmall:E15.5         1.800e+01 2.592e+03  3.439e+00
## Asmall:B20:C45:DSmall:E15.5 -6.000e+00 2.880e+02  3.822e-01
```

From the normality plot for effects, factors B, C, A and interaction AB have larger effects.

(b)

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5")
semicon <- read.csv("6.26.csv")
semicon.aov <- aov(Yield ~ A.Aperture + B.Exposure.Time + C.Develop.Time + A.Aperture *
    B.Exposure.Time, data = semicon)
summary(semicon.aov)

##                             Df Sum Sq Mean Sq F value  Pr(>F)
## A.Aperture                   1   1116    1116     382 < 2e-16 ***
## B.Exposure.Time              1   9214    9214    3155 < 2e-16 ***
## C.Develop.Time               1    751     751     257 2.5e-15 ***
## A.Aperture:B.Exposure.Time   1    504     504     173 3.0e-13 ***
## Residuals                   27     79       3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA table confirms the findings from part (a).

(c)

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5")
semicon <- read.csv("6.26.csv")
semicon_small <- subset(semicon, A.Aperture == "small")
semicon_large <- subset(semicon, A.Aperture == "large")
semicon_small.lm <- lm(Yield ~ B.Exposure.Time + C.Develop.Time, data = semicon_small)
summary(semicon_small.lm)

##
## Call:
## lm(formula = Yield ~ B.Exposure.Time + C.Develop.Time, data = semicon_small)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.000 -1.250 -0.125  1.188  2.750
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.5000     2.2776    0.66     0.52
## B.Exposure.Time   0.6500     0.0223   29.10  3.2e-13 ***
## C.Develop.Time    0.6167     0.0596   10.35  1.2e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.79 on 13 degrees of freedom
## Multiple R-squared:  0.987,Adjusted R-squared:  0.984
## F-statistic:  477 on 2 and 13 DF,  p-value: 6.83e-13


semicon_large.lm <- lm(Yield ~ B.Exposure.Time + C.Develop.Time, data = semicon_large)
summary(semicon_large.lm)

##
## Call:
## lm(formula = Yield ~ B.Exposure.Time + C.Develop.Time, data = semicon_large)
##
```

```
## Residuals:
##    Min     1Q Median     3Q    Max 
## -2.438 -0.781 -0.375  0.875  2.688 
## 
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)    
## (Intercept)      11.1250     2.1158    5.26  0.00015 ***
## B.Exposure.Time   1.0469     0.0207   50.46  2.7e-16 ***
## C.Develop.Time    0.6750     0.0553   12.20  1.7e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.66 on 13 degrees of freedom
## Multiple R-squared:  0.995,Adjusted R-squared:  0.994 
## F-statistic: 1.35e+03 on 2 and 13 DF,  p-value: 8.48e-16
```

The regression equation (Aperture=Small):

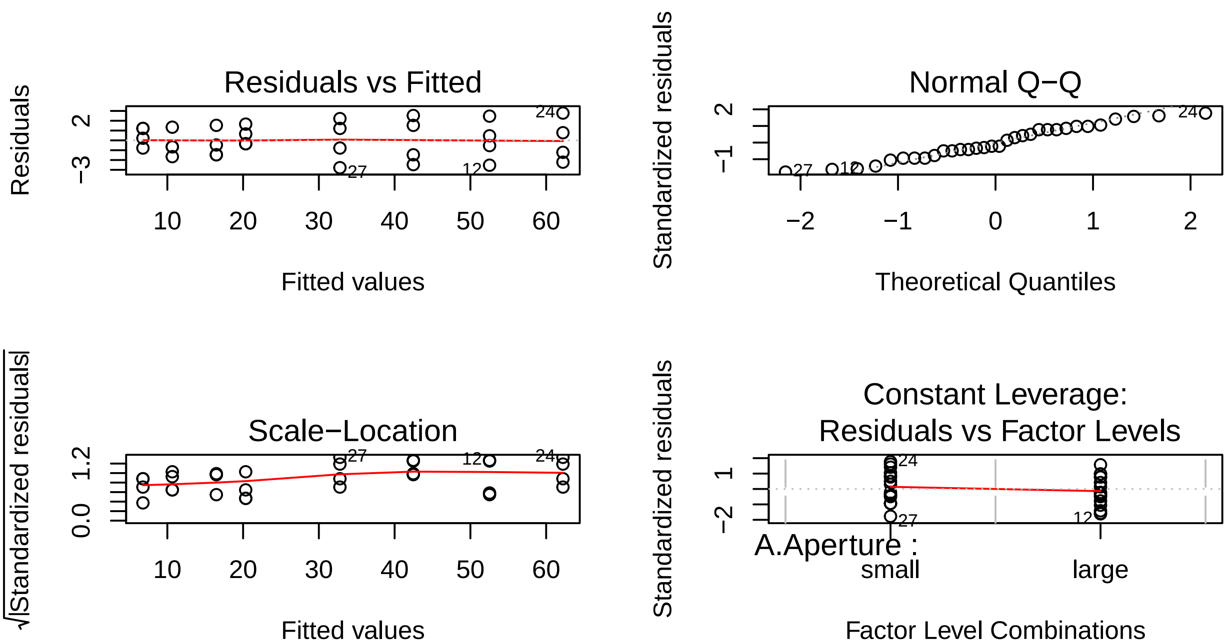
y= 1.5+ 0.650x_A + 0.645x_B

The regression equation (Aperture=Large):

y= 11.125+ 1.046x_A + 0.675x_B

(d)

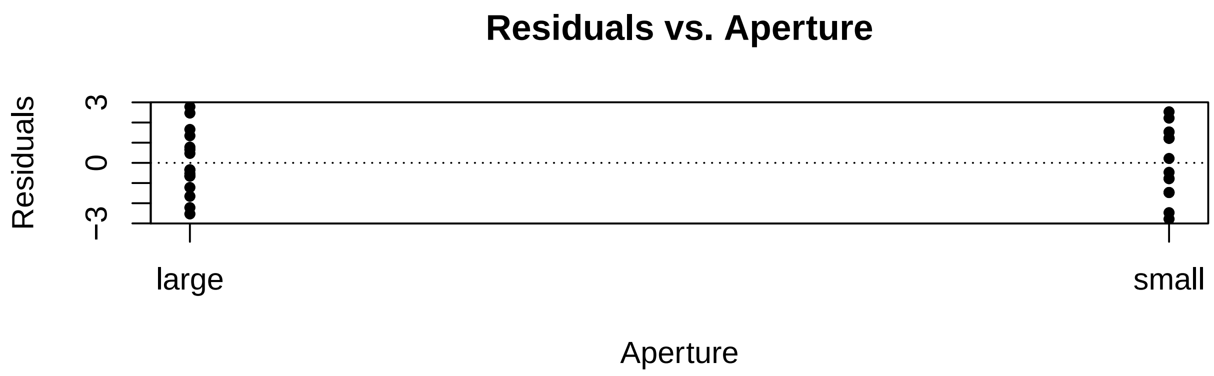
```r
par(mfrow = c(2, 2))
plot(semicon.aov)
```



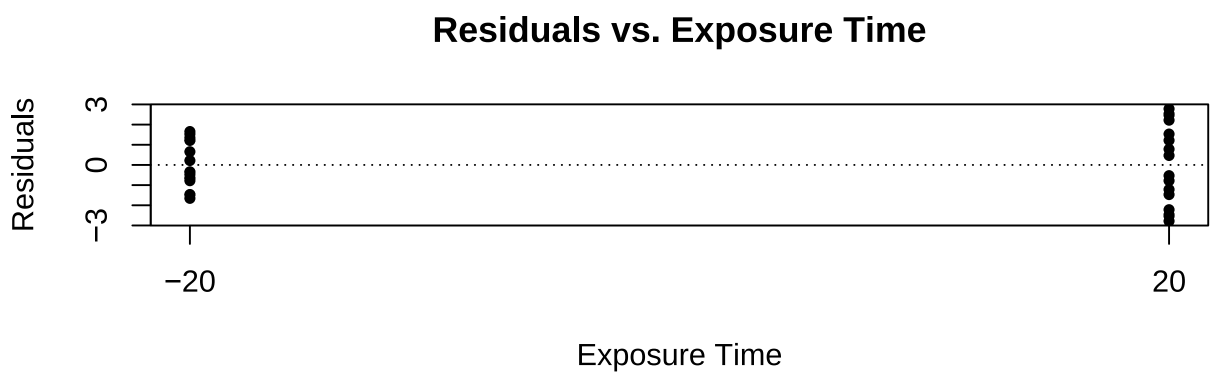
Nothing unusual is visible from the normality plot.

(e)

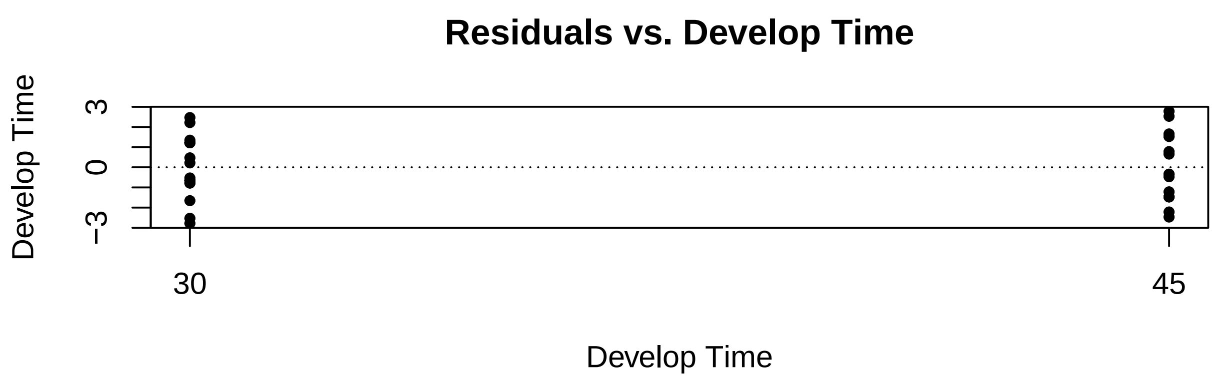
```r
stripchart(residuals(semicon.aov) ~ semicon$A.Aperture, vertical = TRUE, jitter = 0,
    xlab = "Aperture", ylab = "Residuals", cex = 1, pch = 20, main = "Residuals vs. Aperture")
abline(h = 0, col = "black", lty = 3)
```



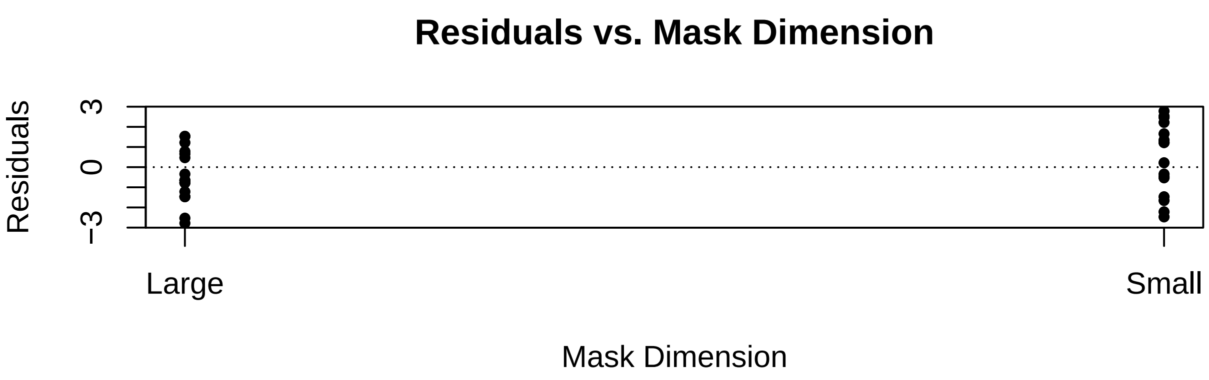

```
stripchart(residuals(semicon.aov) ~ semicon$B.Exposure.Time, vertical = TRUE,
    jitter = 0, xlab = "Exposure Time", ylab = "Residuals", cex = 1, pch = 20,
    main = "Residuals vs. Exposure Time")
abline(h = 0, col = "black", lty = 3)
```



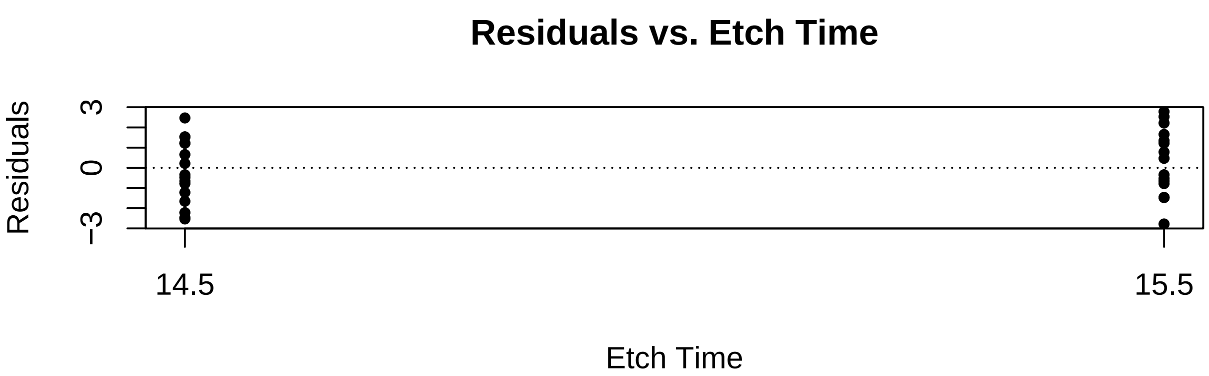

```
stripchart(residuals(semicon.aov) ~ semicon$C.Develop.Time, vertical = TRUE,
    jitter = 0, xlab = "Develop Time", ylab = "Develop Time", cex = 1, pch = 20,
    main = "Residuals vs. Develop Time")
abline(h = 0, col = "black", lty = 3)
```




```
stripchart(residuals(semicon.aov) ~ semicon$D.Mask.Dimension, vertical = TRUE,
    jitter = 0, xlab = "Mask Dimension", ylab = "Residuals", cex = 1, pch = 20,
    main = "Residuals vs. Mask Dimension")
abline(h = 0, col = "black", lty = 3)
```



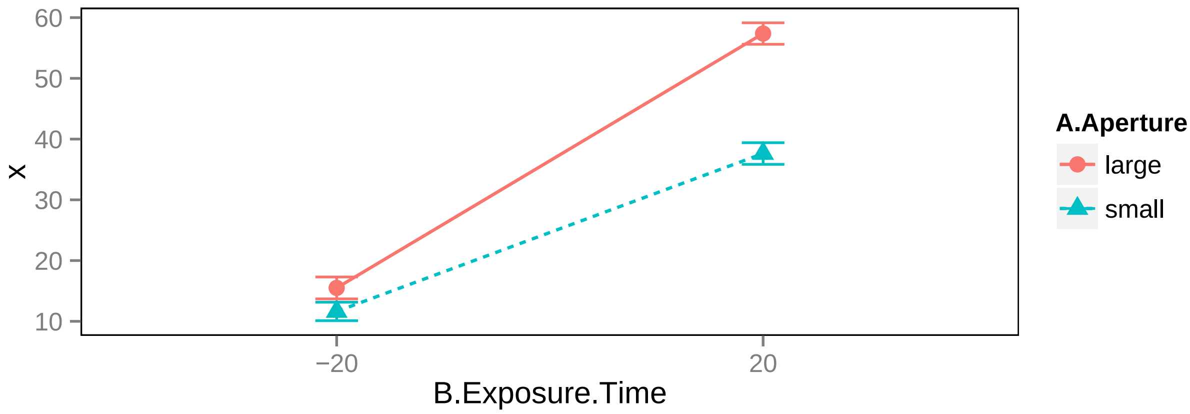

```
stripchart(residuals(semicon.aov) ~ semicon$E.Etch.Time, vertical = TRUE, jitter = 0,
    xlab = "Etch Time", ylab = "Residuals", cex = 1, pch = 20, main = "Residuals vs. Etch Time")
abline(h = 0, col = "black", lty = 3)
```



The residual versus predicted plot for exposure time show very slight amount of inequality of variance. By observing all five plots no significant problem is visible.

(f)

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5")
semicon <- read.csv("6.26.csv")
semicon[, 1:5] <- lapply(semicon[, 1:5], factor)

library(ggplot2)
df <- with(semicon, aggregate(Yield, list(A.Aperture = A.Aperture, B.Exposure.Time = B.Exposure.Time),
    mean))
df$se <- with(semicon, aggregate(Yield, list(A.Aperture = A.Aperture, B.Exposure.Time = B.Exposure.Time
    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 = B.Exposure.Time, y = x, colour = A.Aperture, group = A.Aperture))
gp + geom_line(aes(linetype = A.Aperture), size = 0.6) + geom_point(aes(shape = A.Aperture),
    size = 3) + geom_errorbar(aes(ymax = x + se, ymin = x - se), width = 0.1)
```



From the interaction plot, we find that Factor A doesn't have much effect when B is at low level. When B is at higher level, Factor A has very large effect.

(g)
For getting higher yield, we need to run B and A at a higher level by keeping C at higher level.

9 Exercise 6.27

(a)

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5")
semicon1 <- read.csv("6.27.csv")
semicon1a.aov <- aov(Yield ~ A.Aperture + B.Exposure.Time + C.Develop.Time +
    I(B.Exposure.Time^2) + A.Aperture * B.Exposure.Time, data = semicon1)
summary(semicon1a.aov)
```

```
##                            Df Sum Sq Mean Sq F value  Pr(>F)
## A.Aperture                  1    992     992   122.6 4.1e-12 ***
## B.Exposure.Time             1   9214    9214  1138.1 < 2e-16 ***
## C.Develop.Time              1    751     751    92.7 1.1e-10 ***
## I(B.Exposure.Time^2)        1   6114    6114   755.2 < 2e-16 ***
## A.Aperture:B.Exposure.Time  1    504     504    62.3 8.3e-09 ***
## Residuals                  30    243       8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the ANOVA table, we find that the factor A, B, C, interaction AB and the curvature all have significant effects.

(b)

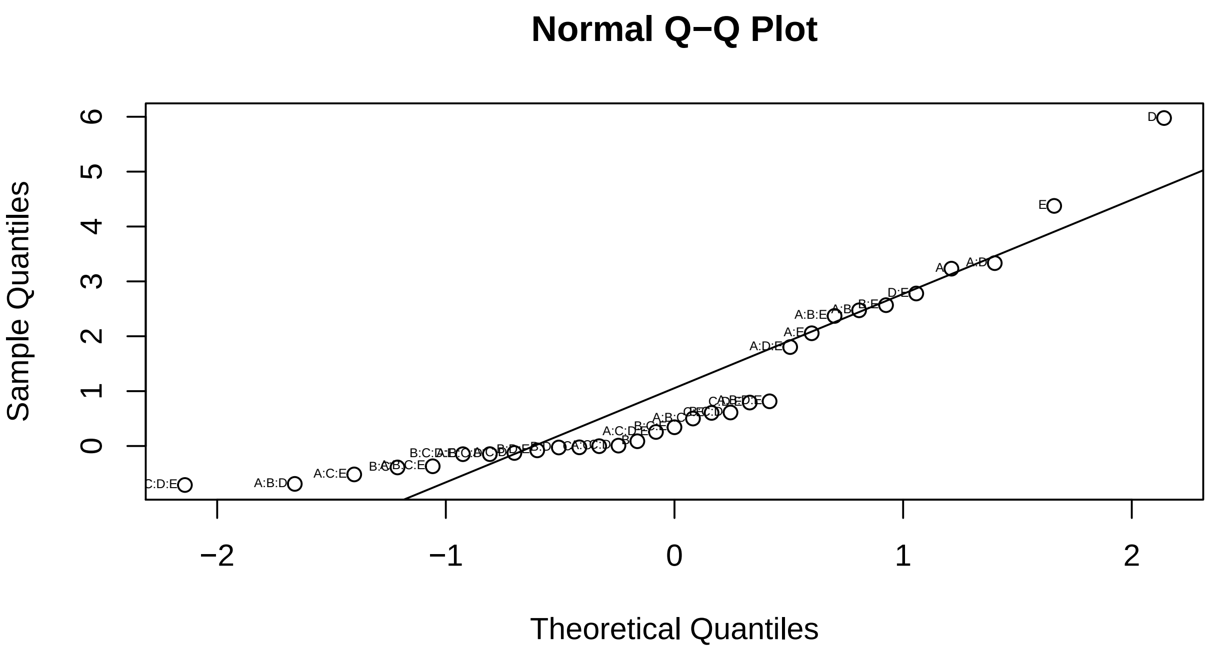
The possible next step would be adding axial points and try to fit a second-order model.

10 Exercise 6.39

(a)

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5")
exp <- read.csv("6.39.csv")
# exp[,1:5] <- lapply(exp[,1:5],factor)
mydata.lm = lm(y ~ A * B * C * D * E, data = exp)
n = 1
k = 5
effects = coefficients(mydata.lm)[-c(1)] * 2
SS = effects^2 * n * 2^(k - 2)
```

```
percentage = SS/sum(SS) * 100
tem = qqnorm(effects)
qqline(effects)
text(tem$x, tem$y, names(effects), pos = 2, offset = 0.2, cex = 0.45)
```




```
cbind(effects, SS, percentage)

##              effects        SS percentage
## A           3.231875 8.356e+01  9.161e+00
## B           0.086875 6.038e-02  6.619e-03
## C          -0.024375 4.753e-03  5.211e-04
## D           5.976875 2.858e+02  3.133e+01
## E           4.375625 1.532e+02  1.679e+01
## A:B         2.473125 4.893e+01  5.364e+00
## A:C        -0.003125 7.813e-05  8.565e-06
## B:C        -0.390625 1.221e+00  1.338e-01
## A:D         3.333125 8.888e+01  9.744e+00
## B:D        -0.026875 5.778e-03  6.335e-04
## C:D         0.006875 3.781e-04  4.145e-05
## A:E         2.054375 3.376e+01  3.701e+00
## B:E         2.566875 5.271e+01  5.779e+00
## C:E         0.603125 2.910e+00  3.190e-01
## D:E         2.779375 6.180e+01  6.775e+00
## A:B:C       0.500625 2.005e+00  2.198e-01
## A:B:D      -0.690625 3.816e+00  4.183e-01
## A:C:D      -0.126875 1.288e-01  1.412e-02
## B:C:D       0.610625 2.983e+00  3.270e-01
## A:B:E       2.370625 4.496e+01  4.929e+00
## A:C:E      -0.518125 2.148e+00  2.354e-01
## B:C:E       0.341875 9.350e-01  1.025e-01
## A:D:E       1.803125 2.601e+01  2.851e+00
## B:D:E      -0.079375 5.040e-02  5.526e-03
## C:D:E       0.791875 5.017e+00  5.500e-01
## A:B:C:D    -0.148125 1.755e-01  1.924e-02
## A:B:C:E    -0.369375 1.092e+00  1.197e-01
## A:B:D:E     0.814375 5.306e+00  5.817e-01
## A:C:D:E     0.255625 5.228e-01  5.731e-02
```

```
## B:C:D:E   -0.149375 1.785e-01  1.957e-02
## A:B:C:D:E -0.710625 4.040e+00  4.429e-01
```

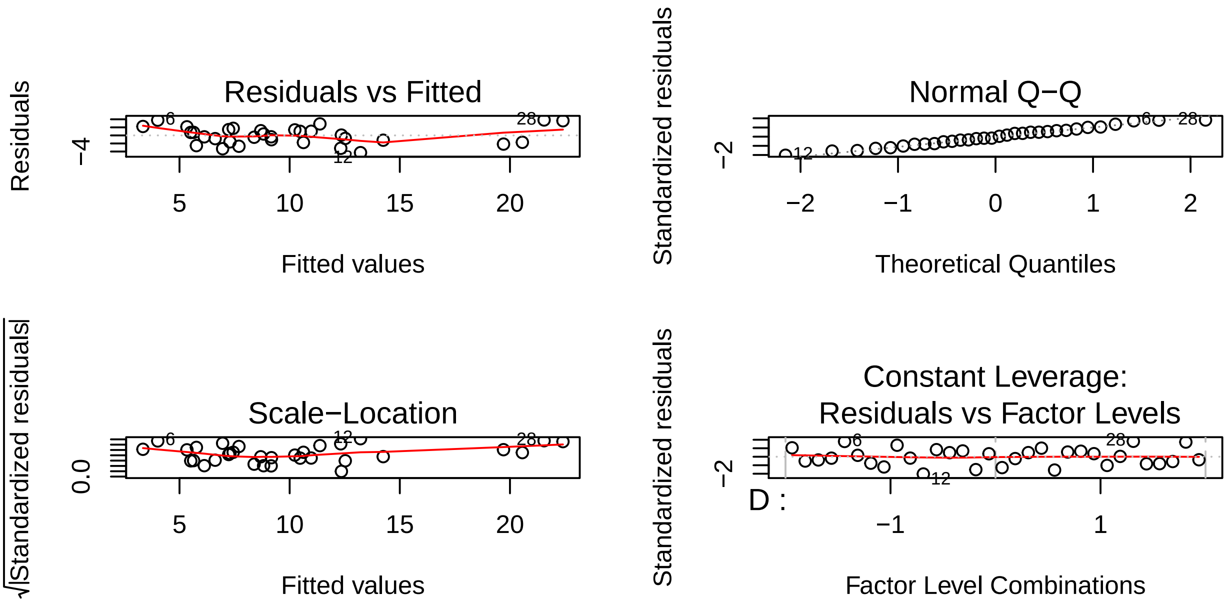
From the normality effect plot, we find that factor D, E and interaction ABD, ACE and CDE have higher effect. No single factor have individual higher effect.

The tentative ANOVA table is shown below:

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5")
exp <- read.csv("6.39.csv")
exp[, 1:5] <- lapply(exp[, 1:5], factor)
mydata.aov = aov(y ~ D + E + A * B * D + C * D * E + A * C * E, data = exp)
summary(mydata.aov)

##             Df Sum Sq Mean Sq F value  Pr(>F)
## D            1  285.8   285.8   30.12 6.2e-05 ***
## E            1  153.2   153.2   16.14  0.0011 **
## A            1   83.6    83.6    8.81  0.0096 **
## B            1    0.1     0.1    0.01  0.9375
## C            1    0.0     0.0    0.00  0.9824
## A:B          1   48.9    48.9    5.16  0.0383 *
## D:A          1   88.9    88.9    9.37  0.0079 **
## D:B          1    0.0     0.0    0.00  0.9806
## D:C          1    0.0     0.0    0.00  0.9950
## E:C          1    2.9     2.9    0.31  0.5879
## D:E          1   61.8    61.8    6.51  0.0221 *
## A:C          1    0.0     0.0    0.00  0.9977
## E:A          1   33.8    33.8    3.56  0.0787 .
## D:A:B        1    3.8     3.8    0.40  0.5355
## D:E:C        1    5.0     5.0    0.53  0.4783
## E:A:C        1    2.1     2.1    0.23  0.6411
## Residuals   15  142.3     9.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(b)

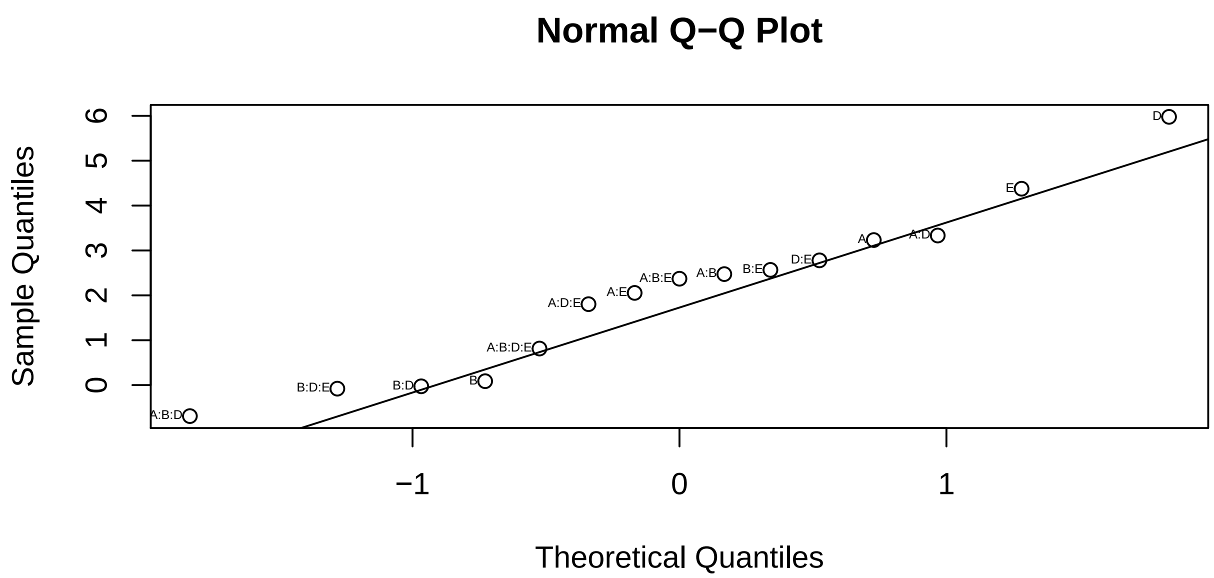
```
par(mfrow = c(2, 2))
plot(mydata.aov)
```



From the normality plot and residual versus predicted plot, we find that observations ane 28 and 32 are ourliers. This is slight indication of model inadequacy.

(c)
From the ANOVA table in part (a), we see that factor C is lees important. By dropping this factor we perform the factorial design again.

```
setwd("C:/Users/Subasish/Dropbox/A Spring 2014/Dr Novelo/HW/HW5")
exp_1 <- read.csv("6.39_1.csv")
# exp[,1:5] <- lapply(exp[,1:5],factor)
mydata.lm = lm(y ~ A * B * D * E, data = exp_1)
n = 1
k = 4
effects = coefficients(mydata.lm)[-c(1)] * 2
SS = effects^2 * n * 2^(k - 2)
percentage = SS/sum(SS) * 100
tem = qqnorm(effects)
qqline(effects)
text(tem$x, tem$y, names(effects), pos = 2, offset = 0.2, cex = 0.45)
```




```
cbind(effects, SS, percentage)

##           effects        SS percentage
## A         3.23188 4.178e+01  9.401e+00
## B         0.08688 3.019e-02  6.793e-03
## D         5.97687 1.429e+02  3.215e+01
## E         4.37562 7.658e+01  1.723e+01
## A:B       2.47312 2.447e+01  5.505e+00
## A:D       3.33313 4.444e+01  1.000e+01
## B:D      -0.02688 2.889e-03  6.501e-04
## A:E       2.05437 1.688e+01  3.799e+00
## B:E       2.56688 2.636e+01  5.931e+00
## D:E       2.77938 3.090e+01  6.953e+00
## A:B:D    -0.69062 1.908e+00  4.293e-01
## A:B:E     2.37063 2.248e+01  5.058e+00
## A:D:E     1.80312 1.301e+01  2.926e+00
## B:D:E    -0.07937 2.520e-02  5.671e-03
## A:B:D:E   0.81437 2.653e+00  5.969e-01
```

```r
### ANOVA Table
mydata.aov2 = aov(y ~ D + E + A * B * D, data = exp)
summary(mydata.aov2)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## D            1  285.8   285.8   26.51 3.2e-05 ***
## E            1  153.2   153.2   14.21  0.0010 ***
## A            1   83.6    83.6    7.75  0.0105 *
## B            1    0.1     0.1    0.01  0.9410
## A:B          1   48.9    48.9    4.54  0.0441 *
## D:A          1   88.9    88.9    8.24  0.0086 **
## D:B          1    0.0     0.0    0.00  0.9817
## D:A:B        1    3.8     3.8    0.35  0.5577
## Residuals   23  248.0    10.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(d)
The settings of the active factrs will be consisted of A, D and E to find the value of y maximum.