STAT430 Assignment4

Jiajun Zhang

November 15, 2019

```r
library(gplots)
library(BsMD)
```

Question 6.5

```r
#2^3 factorial design: 8 runs
A <- c(-1, 1, -1, 1, -1, 1, -1, 1)
B <- c(-1, -1, 1, 1, -1, -1, 1, 1)
C <- c(-1, -1, -1, -1, 1, 1, 1, 1)
#3 replicates each
R1 <- c(22, 32, 35, 55, 44, 40, 60, 39)
R2 <- c(31, 43, 34, 47, 45, 37, 50, 41)
R3 <- c(25, 29, 50, 46, 38, 36, 54, 47)


machine_data <- data.frame(Life=c(R1, R2, R3), A=rep(A,3), B=rep(B,3), C=rep(C,3))
machine_data[,2:4] <- lapply(machine_data[,2:4], factor)
```

(a)

```r
Replicate <- cbind(R1, R2, R3)
cbind(c("(1)", "a", "b", "ab", "c", "ac", "bc", "abc"), apply(Replicate, 1, sum))

##      [,1]  [,2]
## [1,] "(1)" "78"
## [2,] "a"   "104"
## [3,] "b"   "119"
```

```
## [4,] "ab"  "148"
## [5,] "c"   "127"
## [6,] "ac"  "113"
## [7,] "bc"  "164"
## [8,] "abc" "127"

#n=3
```

$$A = \frac{1}{4n}(a - (1) + ab - b + ac - c + abc - bc)$$

$$A = \frac{1}{12}(104 - 78 + 148 - 119 + 113 - 127 + 127 - 164) = 0.33$$

$$B = \frac{1}{4n}(b + ab + bc + abc - (1) - a - c - ac)$$

$$B = \frac{1}{12}(119 + 148 + 164 + 127 - 78 - 104 - 127 - 113) = 11.33$$

$$C = \frac{1}{12}(127 + 113 + 164 + 127 - 78 - 104 - 119 - 148) = 6.83$$

$$AB = \frac{1}{12}(127 - 164 + 148 - 119 - 113 + 127 - 104 + 78) = -1.67$$

$$AC = \frac{1}{12}(78 - 104 + 119 - 148 - 127 + 113 - 164 + 127) = -8.83$$

$$BC = \frac{1}{12}(78 + 104 - 119 - 148 - 127 - 113 + 164 + 127) = -2.83$$

$$ABC = \frac{1}{12}(127 - 164 - 113 + 127 - 148 + 119 + 104 - 78) = -2.167$$

From above, we can see that the effects of B(tool geometry), C(cutting angle), and the interaction between AC appear to be large, where A is the cutting speed.

(b)

```
anova(lm(Life~A*B*C, data=machine_data))

## Analysis of Variance Table
##
## Response: Life
##           Df Sum Sq Mean Sq F value    Pr(>F)
## A          1   0.67    0.67  0.0221 0.8836803
## B          1 770.67  770.67 25.5470 0.0001173 ***
## C          1 280.17  280.17  9.2873 0.0076787 **
```

```
## A:B        1   16.67    16.67  0.5525 0.4680784
## A:C        1  468.17   468.17 15.5193 0.0011722 **
## B:C        1   48.17    48.17  1.5967 0.2244753
## A:B:C      1   28.17    28.17  0.9337 0.3482825
## Residuals 16  482.67    30.17
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the ANOVA table above, we can conclude that the effects of B, C, and interaction between A and C are highly significant, and the other effects are not since their p-values are greater than the significant level of 0.05. This says, we obatined the same results from part a.

(c)

```
machine_data$A <- as.numeric(as.character(machine_data$A))
machine_data$B <- as.numeric(as.character(machine_data$B))
machine_data$C <- as.numeric(as.character(machine_data$C))
# str(machine_data)
summary(lm(Life~B+C+A*C, data=machine_data))

##
## Call:
## lm(formula = Life ~ B + C + A * C, data = machine_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.3333 -4.3750 -0.4167  2.9583 11.5000
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  40.8333     1.1236  36.342  < 2e-16 ***
## B             5.6667     1.1236   5.043 7.22e-05 ***
## C             3.4167     1.1236   3.041 0.006724 **
## A             0.1667     1.1236   0.148 0.883641
## C:A          -4.4167     1.1236  -3.931 0.000897 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 5.504 on 19 degrees of freedom
## Multiple R-squared:  0.7253,Adjusted R-squared:  0.6674
## F-statistic: 12.54 on 4 and 19 DF,  p-value: 3.688e-05
```
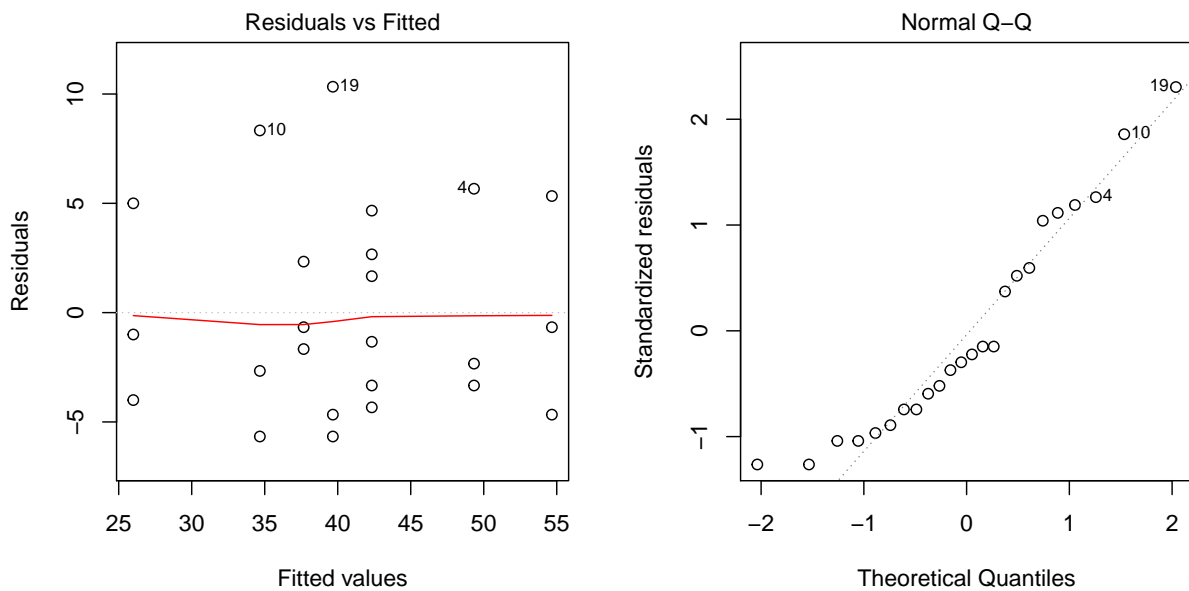
The regression model can be written as

$$\hat{y} = 40.83 + 0.167A + 5.67B + 3.412C - 4.42AC$$

where $\hat{y}$ is the estimated life of a machine tool in hours.

(d)

```
par(mfrow = c(1,2))
res <- aov(Life~A*B*C, data=machine_data)
plot(res,1)
plot(res,2)
```


Based on the two graphs above, we can conclude that there is no obvious problems with the residuals.

(e)

```
#Main effects of B, C significant, and interaction b/w A and C
par(mfrow = c(1,3))
plotmeans(Life ~ B, data = machine_data,
          xlab="Factor B", ylab="Avg. Life of a Machine Tool(in hours)",
          p=.68, main = "Main effect Plot", barcol = "black")
plotmeans(Life ~ C , data = machine_data,
          xlab="Factor C", ylab="Avg. Life of a Machine Tool(in hours)",
          p=.68, main = "Main effect Plot", barcol = "black")
with(machine_data, interaction.plot(x.factor = A,
        trace.factor = C, response = Life, fun = mean, type = "b",
        legend = TRUE, pch = c(1,2,3),
        main = "Interaction Plot(Cutting Speed-Cutting Angle)",
        xlab = "Cutting Speed(A)", ylab = "Avg. Life of a Machine Tool(in hours)"))
```
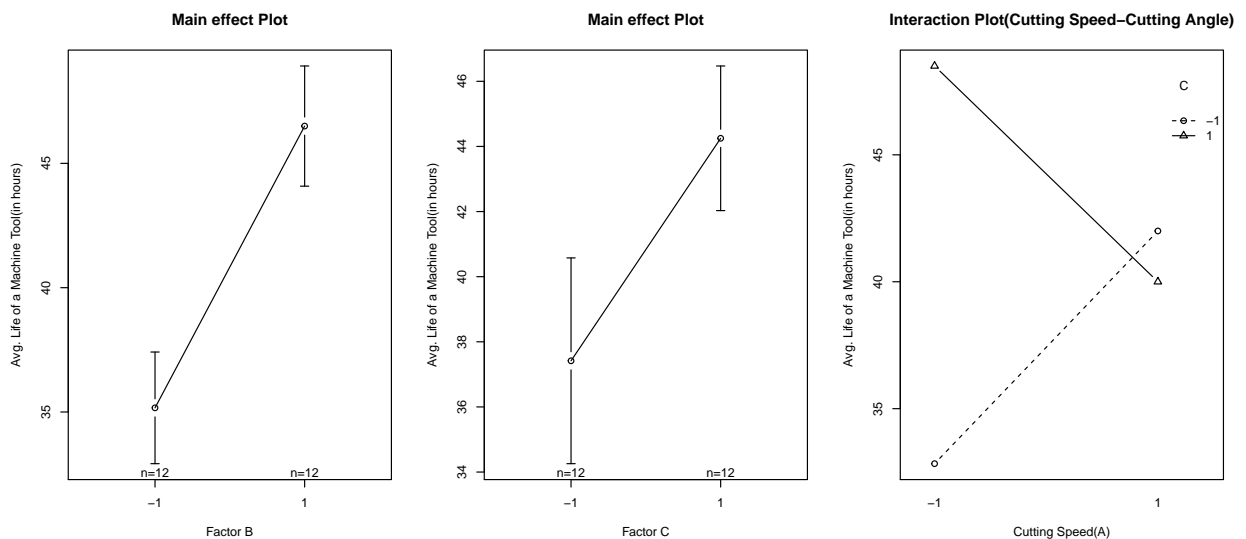

Based on the two main effect plots, we can recommend that using higher level of B and C maximizes the machine tool life time. And from the interacation plot, we suggest that the machine tool life time will be maximized when we use lower level of A and higher level of C.

Question 6.9

```
#2^2 factorial desgin; 4 runs
A <- c(-1, 1, -1, 1)
B <- c(-1, -1, 1, 1)
#4 Replicates each
R1 <- c(18.2, 27.2, 15.9, 41.0)
R2 <- c(18.9, 24.0, 14.5, 43.9)
R3 <- c(12.9, 22.4, 15.1, 36.3)
R4 <- c(14.4, 22.5, 14.2, 39.9)


CBoard.data <- data.frame(Vibration=c(R1, R2, R3, R4), A=rep(A, 4), B=rep(B,4))
CBoard.data[,2:3] <- lapply(CBoard.data[,2:3], factor)
```

(a)

```
anova(lm(Vibration~A*B, data=CBoard.data))

## Analysis of Variance Table
##
## Response: Vibration
##            Df  Sum Sq Mean Sq F value     Pr(>F)
## A           1 1107.23 1107.23 185.252 1.175e-08 ***
## B           1  227.26  227.26  38.023 4.826e-05 ***
## A:B         1  303.63  303.63  50.801 1.201e-05 ***
## Residuals 12   71.72    5.98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
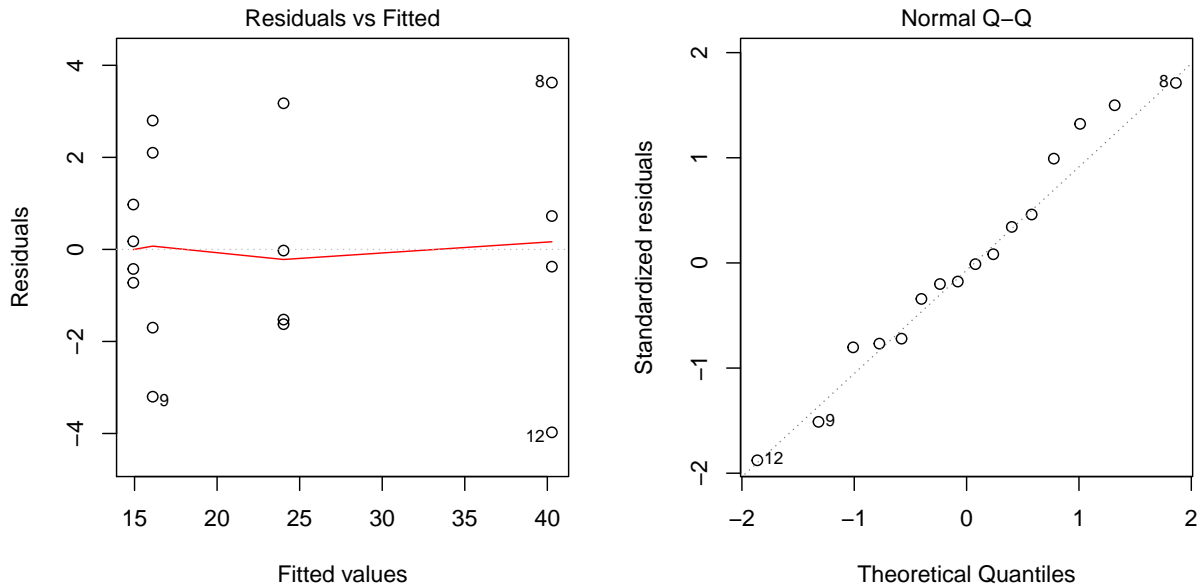
From the ANOVA table, we can see that all main effects of A and B and their interaction are significant since all p-values are less than 0.05.

(b)

```
# Model adequacy checking
par(mfrow = c(1,2))
res <- aov(Vibration~A*B, data=CBoard.data)
```

```
plot(res,1)
plot(res,2)
```


Based on the residuals vs fitted values and normality plots, we can conclude that there is nothing unusual with the residuals.

(c)

```
par(mfrow=c(1,1))
CBoard.data[,2:3] <- lapply(CBoard.data[,2:3], as.numeric)
CBoard.data$A[CBoard.data$A==1] <- 1/16
CBoard.data$A[CBoard.data$A==2] <- 1/8


CBoard.data$B[CBoard.data$B==1] <- 40
CBoard.data$B[CBoard.data$B==2] <- 90


with(CBoard.data, interaction.plot(x.factor=A, trace.factor=B,
          response=Vibration, fun=mean, type="b", legend=T,
          pch=c(1,2,3), xlab="Factor A (Bit Size in inch)",
          ylab="Avg. Vibration", main="Interaction Plot (Bit Size-Cutting Speed)"))
```

Interaction Plot (Bit Size–Cutting Speed)
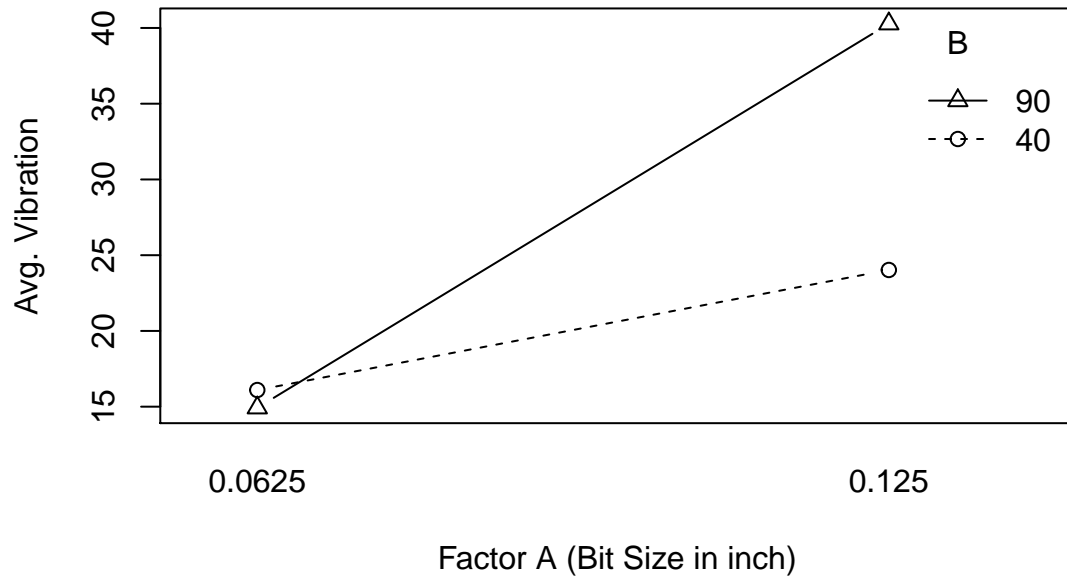
Since we are interested in minimizing the vibration, then based on the interaction plot, we would recommend using the bit size of $\frac{1}{16}$ in. and cutting speed of 90 rpm.

Question 6.16

```
#2^2 factorial desgin; 4 runs
A <- c(-1, 1, -1, 1)
B <- c(-1, -1, 1, 1)
#4 Replicates each
R1 <- c(14.037, 13.880, 14.821, 14.888)
R2 <- c(16.165, 13.860, 14.757, 14.921)
R3 <- c(13.972, 14.032, 14.843, 14.415)
R4 <- c(13.907, 13.914, 14.878, 14.932)


Layer.data <- data.frame(Layer=c(R1, R2, R3, R4), A=rep(A, 4), B=rep(B,4))
Layer.data[,2:3] <- lapply(Layer.data[,2:3], factor)
```

(a)

```
Replicate <- cbind(R1, R2, R3, R4)
cbind(c("(1)", "a", "b", "ab"), apply(Replicate, 1, sum))

##       [,1]   [,2]
## [1,] "(1)"  "58.081"
## [2,] "a"    "55.686"
## [3,] "b"    "59.299"
## [4,] "ab"   "59.156"

#n=4
```

$$A = \frac{1}{2n}[ab + a - b - (1)]$$

$$B = \frac{1}{2n}[ab + b - a - (1)]$$

$$AB = \frac{1}{2n}[ab + (1) - a - b]$$

$$A = \frac{1}{8}[59.156 + 55.686 - 59.299 - 58.081] = -0.317$$

$$B = \frac{1}{8}[59.156 + 59.299 - 55.686 - 58.081] = 0.586$$

$$AB = \frac{1}{8}[59.156 + 58.081 - 55.686 - 59.299] = 0.281$$

(b)

```
anova(lm(Layer~A*B, data=Layer.data))

## Analysis of Variance Table
##
## Response: Layer
##           Df Sum Sq Mean Sq F value  Pr(>F)
## A          1 0.4026 0.40259  1.2619 0.28327
## B          1 1.3736 1.37358  4.3054 0.06016 .
## A:B        1 0.3170 0.31697  0.9935 0.33856
## Residuals 12 3.8285 0.31904
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the ANOVA table above, we can see that the main effects of A, B, and their interaction are all not significant since their p-values are greater than 0.05.

(c)

```
#Full model
Layer.data[,2:3] <- lapply(Layer.data[,2:3], as.numeric)
Layer.data$A[Layer.data$A==1] <- 0.55
Layer.data$A[Layer.data$A==2] <- 0.59


Layer.data$B[Layer.data$B==1] <- 10
Layer.data$B[Layer.data$B==2] <- 15
str(Layer.data)

## 'data.frame': 16 obs. of  3 variables:
##  $ Layer: num  14 13.9 14.8 14.9 16.2 ...
##  $ A    : num  0.55 0.59 0.55 0.59 0.55 0.59 0.55 0.59 0.55 0.59 ...
##  $ B    : num  10 10 15 15 10 10 15 15 10 10 ...

summary(lm(Layer~A*B, data=Layer.data))

##
## Call:
## lm(formula = Layer ~ A * B, data = Layer.data)
```

```
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -0.61325 -0.14431 -0.00562  0.10187  1.64475
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   37.627     20.533   1.832   0.0918 .
## A            -43.119     36.001  -1.198   0.2542
## B             -1.487      1.611  -0.923   0.3740
## A:B            2.815      2.824   0.997   0.3386
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5648 on 12 degrees of freedom
## Multiple R-squared:  0.3535,Adjusted R-squared:  0.1918
## F-statistic: 2.187 on 3 and 12 DF,  p-value: 0.1425
```
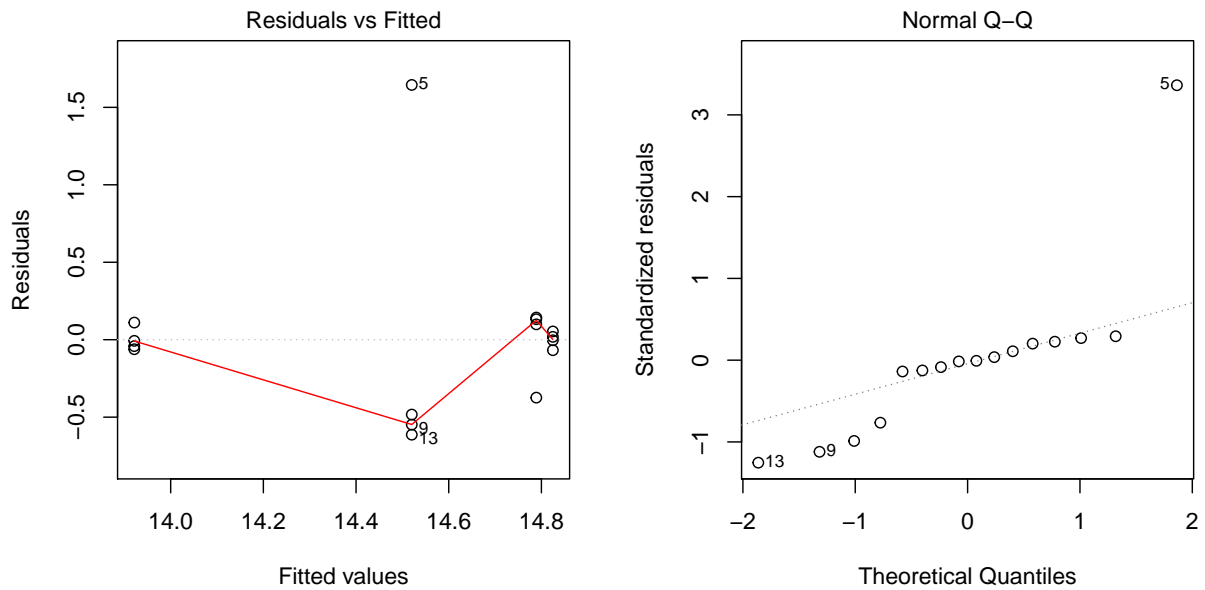
Based on the summary result above, we can conclude that the regression model with all effects can be written as

$$\hat{y} = 37.627 - 43.119A - 1.487B + 2.815AB$$

where $\hat{y}$ is the estimated epitaxial layer thickness, A is the arsenic flow rate and B is the deposition time.

(d)

```
par(mfrow = c(1,2))
res <- aov(Layer~A*B, data=Layer.data)
plot(res,1)
plot(res,2)
```

11

Residuals vs Fitted   Normal Q–Q

From the residuals plots above, we can see that there are some outliers we should concern such as the points with the labels 5, 9, and 13.

(e) For the observation of 16.165, we consider this as the outlier. We can fix this by replacing it with the average of the other three observations in the same row.

```
(14.037 + 13.972 + 13.907)/3
```

```
## [1] 13.972
```

Therefore, we should replace the outlier with the value of 13.972.

Question 6.19

```
#2^4; 16 runs;
A <- rep(c(-1, 1), 8)
B <- rep(c(-1, -1, 1, 1), 4)
C <- rep(c(rep(-1,4), rep(1,4)),2)
D <- c(rep(-1,8), rep(1,8))
# two replicates each
R1 <- c(7.037, 14.707, 11.635, 17.273, 10.403, 4.368, 9.360, 13.440, 8.561,
        16.867, 13.876, 19.824, 11.846, 6.125, 11.190, 15.653)
R2 <- c(6.376, 15.219, 12.089, 17.815, 10.151, 4.098, 9.253, 12.923, 8.951,
        17.052, 13.658, 19.639, 12.337, 5.904, 10.935, 15.053)


titanium.data <- data.frame(Crack.len=c(R1, R2), A=A, B=B, C=C, D=D)
```

(a)

```
Replicate <- cbind(R1, R2)
cbind(c("(1)", "a", "b", "ab", "c", "ac", "bc", "abc", "d", "ad",
        "bd", "abd", "cd", "acd", "bcd", "abcd"), apply(Replicate, 1, sum))

##         [,1]    [,2]
##  [1,] "(1)"  "13.413"
##  [2,] "a"    "29.926"
##  [3,] "b"    "23.724"
##  [4,] "ab"   "35.088"
##  [5,] "c"    "20.554"
##  [6,] "ac"   "8.466"
##  [7,] "bc"   "18.613"
##  [8,] "abc"  "26.363"
##  [9,] "d"    "17.512"
## [10,] "ad"   "33.919"
## [11,] "bd"   "27.534"
## [12,] "abd"  "39.463"
## [13,] "cd"   "24.183"
## [14,] "acd"  "12.029"
```

#n=2

$$A = \frac{1}{16}[-(1) + a - b + ab - c + ac - bc + abc - d + ad - bd + abd - cd + acd - bcd + abcd] = 3.019$$

$$B = \frac{1}{16}[-(1) - a + b + ab - c - ac + bc + abc - d - ad + bd + abd - cd - acd + bcd + abcd] = 3.976$$

$$AB = \frac{1}{16}[(1) - a - b + ab + c - ac - bc + abc + d - ad - bd + abd + cd - acd - bcd + abcd] = 1.934$$

Contrast Constants for the 2^4 Design

	A	B	AB	C	AC	BC	ABC	D	AD	BD	ABD	CD	ACD	BCD	ABCD
(1)	−	−	+	−	+	+	−	−	+	+	−	+	−	−	+
a	+	−	−	−	−	+	+	−	−	+	+	+	+	−	−
b	−	+	−	−	+	−	+	−	+	−	+	+	−	+	−
ab	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+
c	−	−	+	+	−	−	+	−	+	+	−	−	+	+	−
ac	+	−	−	+	+	−	−	−	−	+	+	−	−	+	+
bc	−	+	−	+	−	+	−	−	+	−	+	−	+	−	+
abc	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−
d	−	−	+	−	+	+	−	+	−	−	+	−	+	+	−
ad	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+
bd	−	+	−	−	+	−	+	+	−	+	−	−	+	−	+
abd	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−
cd	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+
acd	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−
bcd	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−
abcd	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Followed by the table above, we will have C = -3.596, AC = -4.008, BC = 0.096, ABC = 3.138, D = 1.958, AD = 0.077, BD = 0.047, ABD = 0.098, CD = -0.077, ACD = 0.019, BCD = 0.036, ABCD = 0.014.

Therefore, the factor A, B, C, AC, ABC tend to have large effects as their absolute values are greater than 3.

(b)

```
titanium.data[,2:5] <- lapply(titanium.data[,2:5], factor)
# str(titanium.data)
anova(lm(Crack.len~A*B*C*D, data=titanium.data))

## Analysis of Variance Table
```

14

```
##
## Response: Crack.len
##             Df  Sum Sq Mean Sq   F value    Pr(>F)
## A            1  72.909  72.909  898.3389 1.740e-15 ***
## B            1 126.461 126.461 1558.1720 < 2.2e-16 ***
## C            1 103.464 103.464 1274.8225 < 2.2e-16 ***
## D            1  30.662  30.662  377.8021 1.485e-12 ***
## A:B          1  29.927  29.927  368.7390 1.790e-12 ***
## A:C          1 128.496 128.496 1583.2562 < 2.2e-16 ***
## B:C          1   0.074   0.074    0.9084    0.3547
## A:D          1   0.047   0.047    0.5769    0.4586
## B:D          1   0.018   0.018    0.2201    0.6453
## C:D          1   0.047   0.047    0.5825    0.4564
## A:B:C        1  78.751  78.751  970.3255 9.485e-16 ***
## A:B:D        1   0.077   0.077    0.9467    0.3450
## A:C:D        1   0.003   0.003    0.0361    0.8518
## B:C:D        1   0.010   0.010    0.1251    0.7282
## A:B:C:D      1   0.002   0.002    0.0197    0.8902
## Residuals 16   1.299   0.081
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the ANOVA table above, we can see that the factor effects of A, B, C, D, AB, AC, and ABC are significant since their p-values are less than 0.05. Then, we can conclude that these factors affect cracking at significant level of 0.05.

(c)

```
titanium.data <- data.frame(Crack.len=c(R1, R2), A=A, B=B, C=C, D=D)
summary(lm(Crack.len~A*B*C+D, data=titanium.data))

##
## Call:
## lm(formula = Crack.len ~ A * B * C + D, data = titanium.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -0.3859 -0.1945  0.0245  0.1798  0.4069
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.98806    0.04518 265.369  <2e-16 ***
## A            1.50944    0.04518  33.413  <2e-16 ***
## B            1.98794    0.04518  44.005  <2e-16 ***
## C           -1.79812    0.04518 -39.803  <2e-16 ***
## D            0.97888    0.04518  21.668  <2e-16 ***
## A:B          0.96706    0.04518  21.407  <2e-16 ***
## A:C         -2.00388    0.04518 -44.358  <2e-16 ***
## B:C          0.04800    0.04518   1.063   0.299
## A:B:C        1.56875    0.04518  34.726  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2555 on 23 degrees of freedom
## Multiple R-squared:  0.9974,Adjusted R-squared:  0.9965
## F-statistic:  1092 on 8 and 23 DF,  p-value: < 2.2e-16
```

From part b, we found that the factor effects of A, B, C, D, AB, AC, and ABC are significant. Then, the regression model can be written as

$$\hat{y} = 11.988 + 1.509A + 1.988B - 1.798C + 0.979D + 0.967AB - 2.004AC + 0.048BC + 1.569ABC$$

where $\hat{y}$ is the estimated length of crack, A is the pouring temperature, B is the titanium content, C is the heat treatment method, D is the amount of grain refiner used.

(d)

```
par(mfrow = c(1,2))
res <- aov(Crack.len~A*B*C*D, data=titanium.data)
plot(res,1)
plot(res,2)
```
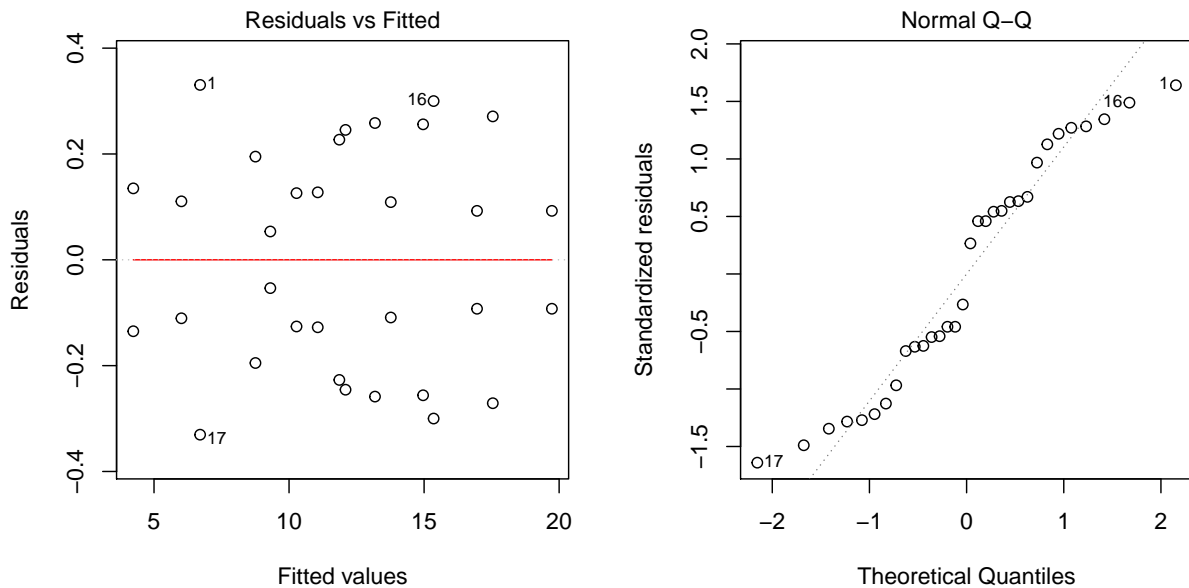
From the residuals vs fitted values and normality plots, we will conclude that there is nothing unusual with the residuals.

(f)

```
#A, B, C, D, AB, AC, and ABC significant
#AB interaction, AC interaction
par(mfrow = c(2, 2))
with(titanium.data, interaction.plot(x.factor = A,
        trace.factor = B, response = Crack.len, fun = mean, type = "b",
        legend = TRUE, pch = c(1,2,3),
        main = "Interaction Plot(Pouring Temperature-Titanium Content)",
        xlab = "Pouring Temperautre(A)", ylab = "Avg. length of crack(in mm*10^-2)"))
with(titanium.data, interaction.plot(x.factor = A,
        trace.factor = C, response = Crack.len, fun = mean, type = "b",
        legend = TRUE, pch = c(1,2,3),
        main = "Interaction Plot(Pouring Temperature-Heat Treatment Method)",
        xlab = "Pouring Temperautre(A)", ylab = "Avg. length of crack(in mm*10^-2)"))
plotmeans(Crack.len ~ D, data = titanium.data,
        xlab="Factor D", ylab="Avg. length of crack(in mm*10^-2)",
        p=.68, main = "Plot of Main effect D", barcol = "black")
```

Interaction Plot(Pouring Temperature–Titanium Content)

Interaction Plot(Pouring Temperature–Heat Treatment Method)

Plot of Main effect D
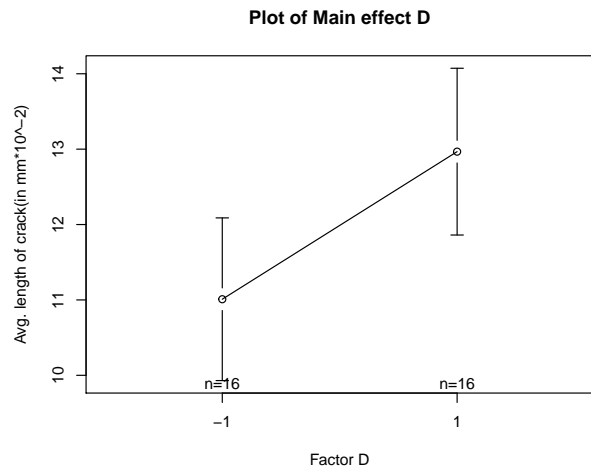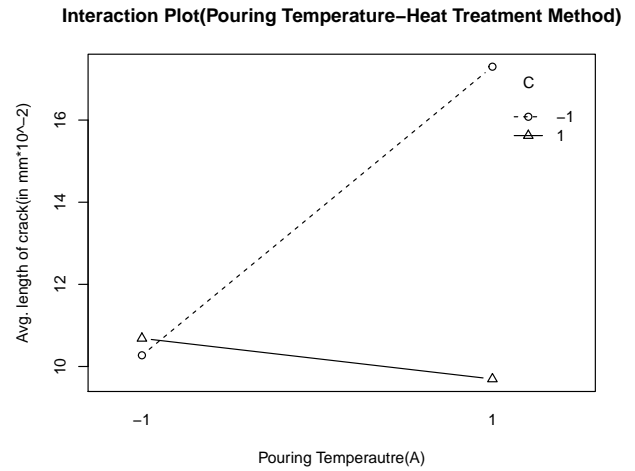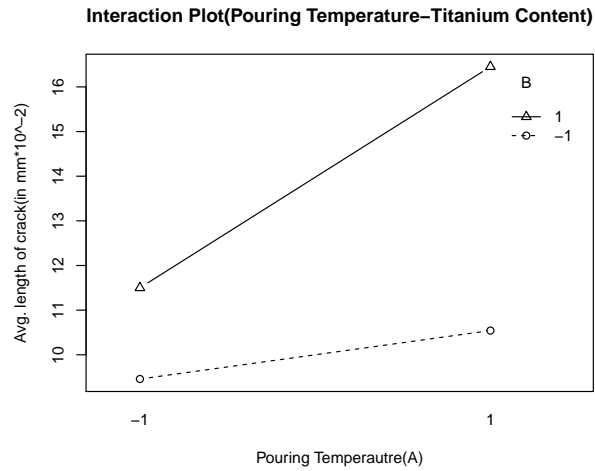
Since we are interested in the smallest length of crack, then from the interaction and main effect plots, we would recommend using lower levels of A, B, C, and D give the smallest length of crack.
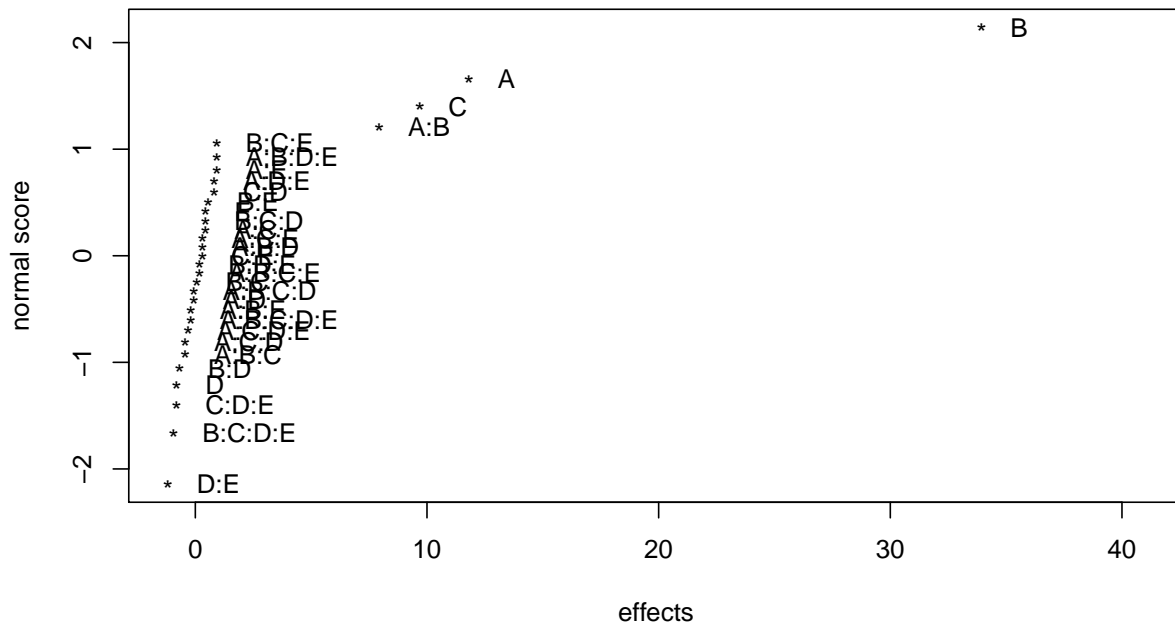
Question 6.30

(a)

```
# 2^5; 32 runs; no replicate
A <- rep(c(-1, 1), 16)
B <- rep(c(-1, -1, 1, 1), 8)
C <- rep(c(rep(-1,4), rep(1,4)), 4)
D <- c(rep(-1,8), rep(1,8), rep(-1,8), rep(1,8))
E <- c(rep(-1,16), rep(1,16))
plant.data <- data.frame(Yield=c(7, 9, 34, 55, 16, 20, 40, 60, 8, 10, 32,
                                 50, 18, 21, 44, 61, 8, 12, 35, 52, 15, 22,
                                 45, 65, 6, 10, 30, 53, 15, 20, 41, 63),
                         A=A, B=B, C=C, D=D, E=E)
# str(plant.data)
fit <- lm(Yield~A*B*C*D*E, data=plant.data)
par(mfrow=c(1,1))
DanielPlot(fit)
```


Based on the result of Daniel's plot, we can see that the factor effects of A, B, C, and AB appear to be large.

(b)

```
plant.data[,2:6] <- lapply(plant.data[,2:6], factor)
anova.fit <- aov(Yield~A*B+C, data=plant.data)
summary(anova.fit)

##              Df Sum Sq Mean Sq F value   Pr(>F)
## A            1   1116    1116   382.3  < 2e-16 ***
## B            1   9214    9214  3155.3  < 2e-16 ***
## C            1    751     751   257.1 2.53e-15 ***
## A:B          1    504     504   172.6 3.04e-13 ***
## Residuals   27     79       3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From part a, we found that the significant factors are A, B, C, and the interaction between A and B. Then, we will use these to conduct an ANOVA test, as shown above. Based on the ANOVA result, we can see that all factors are significant since their p-values as less than 0.05.

(c)

```
plant.data <- data.frame(Yield=c(7, 9, 34, 55, 16, 20, 40, 60, 8, 10, 32,
                                 50, 18, 21, 44, 61, 8, 12, 35, 52, 15, 22,
                                 45, 65, 6, 10, 30, 53, 15, 20, 41, 63),
                         A=A, B=B, C=C, D=D, E=E)
summary(lm(Yield~A*B+C, data=plant.data))

##
## Call:
## lm(formula = Yield ~ A * B + C, data = plant.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7812 -1.2812 -0.3438  1.2500  2.7812
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  30.5312     0.3021  101.07  < 2e-16 ***
```

```
## A                  5.9063      0.3021    19.55  < 2e-16 ***
## B                 16.9688      0.3021    56.17  < 2e-16 ***
## C                  4.8438      0.3021    16.03 2.53e-15 ***
## A:B                3.9687      0.3021    13.14 3.04e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.709 on 27 degrees of freedom
## Multiple R-squared:  0.9932,Adjusted R-squared:  0.9922
## F-statistic: 991.8 on 4 and 27 DF,  p-value: < 2.2e-16
```
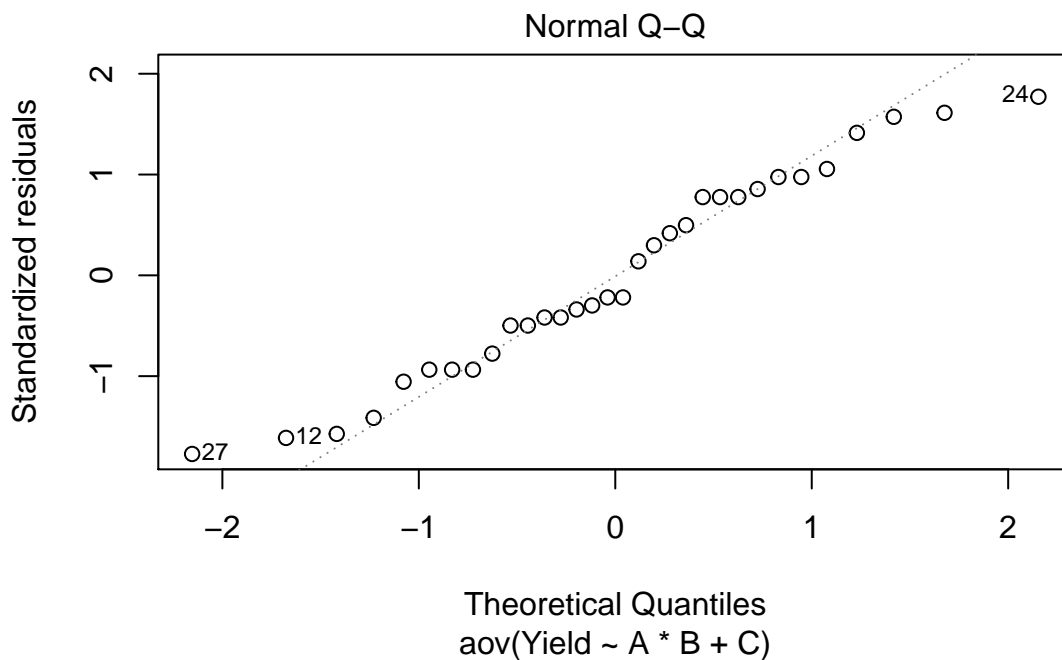
The regression model can be written as

$$\hat{y} = 30.531 + 5.906A + 16.969B + 4.844C + 3.969AB$$

where $\hat{y}$ is the estimated chemical yield, A is the apearture setting, B is the exposure time, C is the development time, and AB is the interaction between A and B.
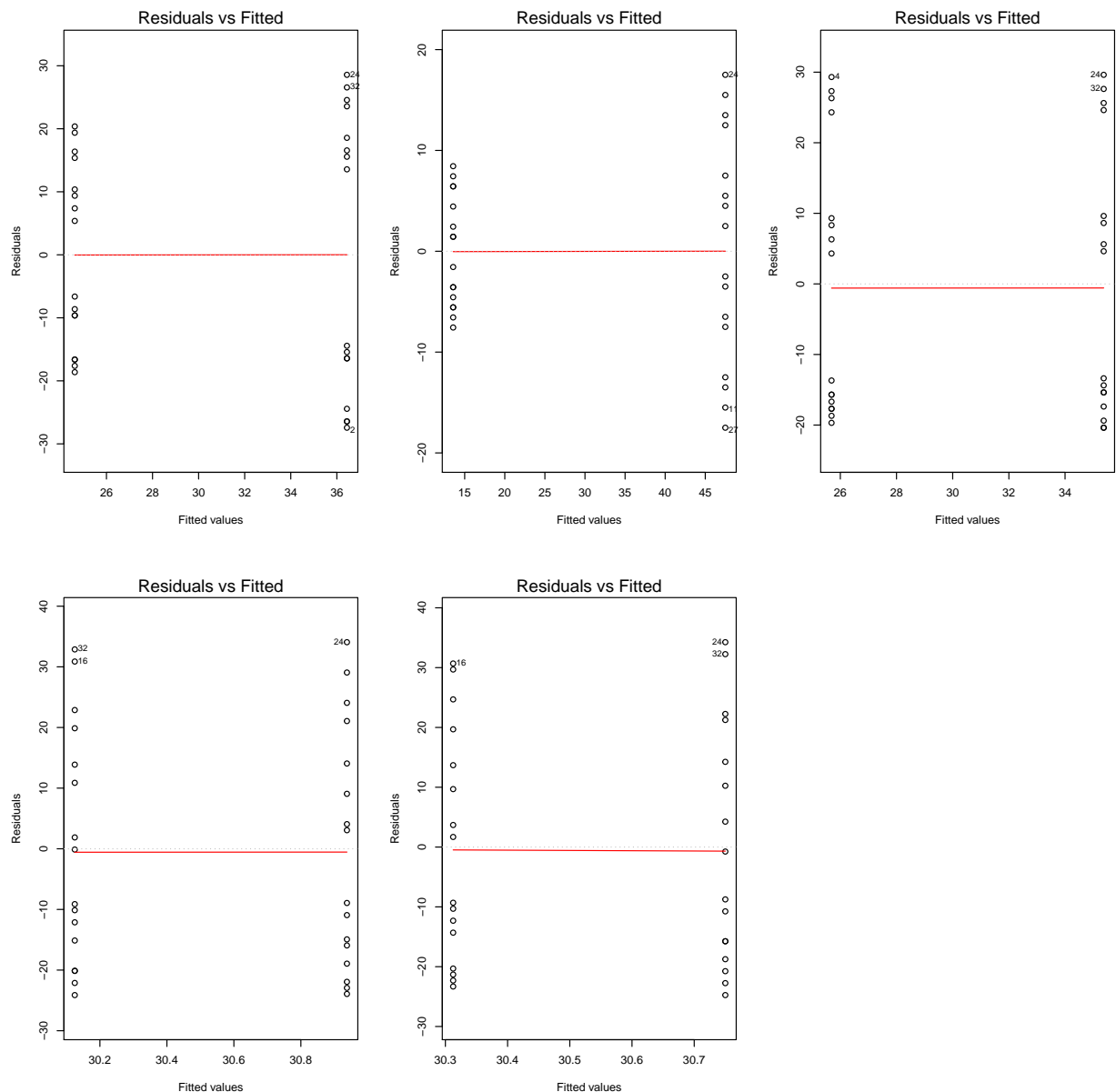
(d)

```
plot(anova.fit, 2)
```


The normality plot looks nothing unusual, then we can say that the assumption is satisfied.

(e)

```
par(mfrow = c(2,3))
plant.data[,2:6] <- lapply(plant.data[,2:6], factor)
plot(aov(Yield~A, data=plant.data), 1)
plot(aov(Yield~B, data=plant.data), 1)
plot(aov(Yield~C, data=plant.data), 1)
plot(aov(Yield~D, data=plant.data), 1)
plot(aov(Yield~E, data=plant.data), 1)
```


From the residuals vs fitted values plot, we can conclude that the constant variance assumption is not

satisfied since we only observed residual points only at the edges of the plots.

(f)

```
with(plant.data, interaction.plot(x.factor = A,
        trace.factor = B, response = Yield, fun = mean, type = "b",
        legend = TRUE, pch = c(1,2,3),
        main = "Interaction Plot(Aperture Setting-Exposure Time)",
        xlab = "Aperture Setting(A)", ylab = "Avg. Yield"))
```

Interaction Plot(Aperture Setting–Exposure Time)
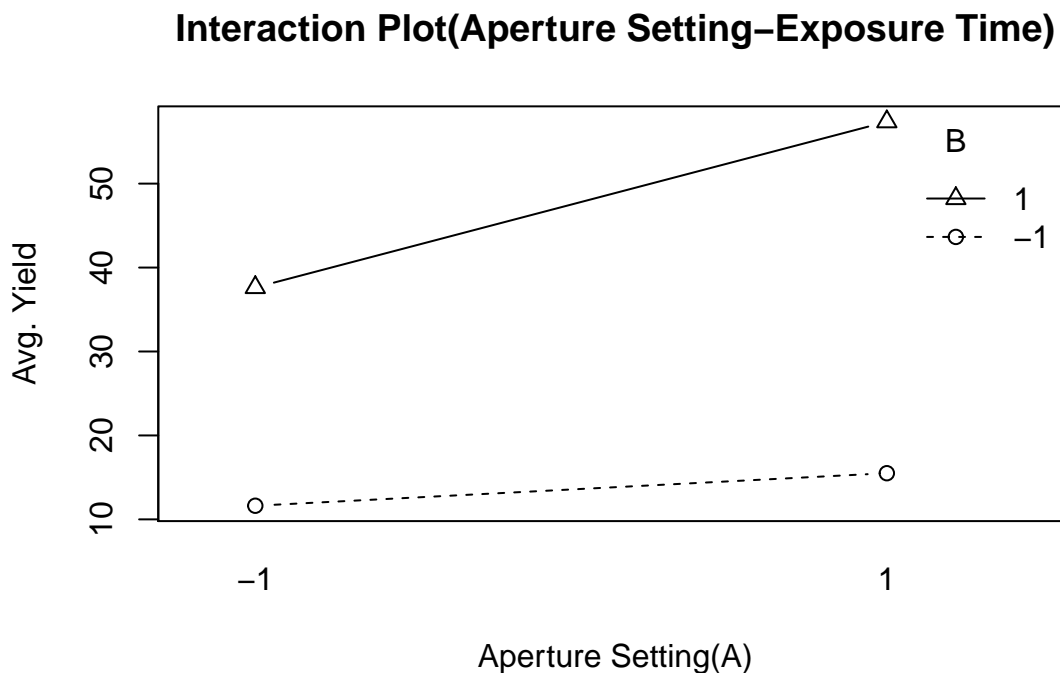
From the previous findings, we know that AB interaction is significant. And from the interaction plot, we can see that the factor B has a greater effect on chemical yield at both low and high levels than factor A.

(g)

```
#A, B, C, AB
par(mfrow=c(1,2))
with(plant.data, interaction.plot(x.factor = A,
        trace.factor = B, response = Yield, fun = mean, type = "b",
        legend = TRUE, pch = c(1,2,3),
        main = "Interaction Plot(Aperture Setting-Exposure Time)",
        xlab = "Aperture Setting(A)", ylab = "Avg. Yield"))
```

```
plotmeans(Yield ~ C, data = plant.data,
          xlab="Factor C", ylab="Avg. Yield",
          p=.68, main = "Plot of Main effect C", barcol = "black")
```

Interaction Plot(Aperture Setting–Exposure Time)   **Plot of Main effect C**
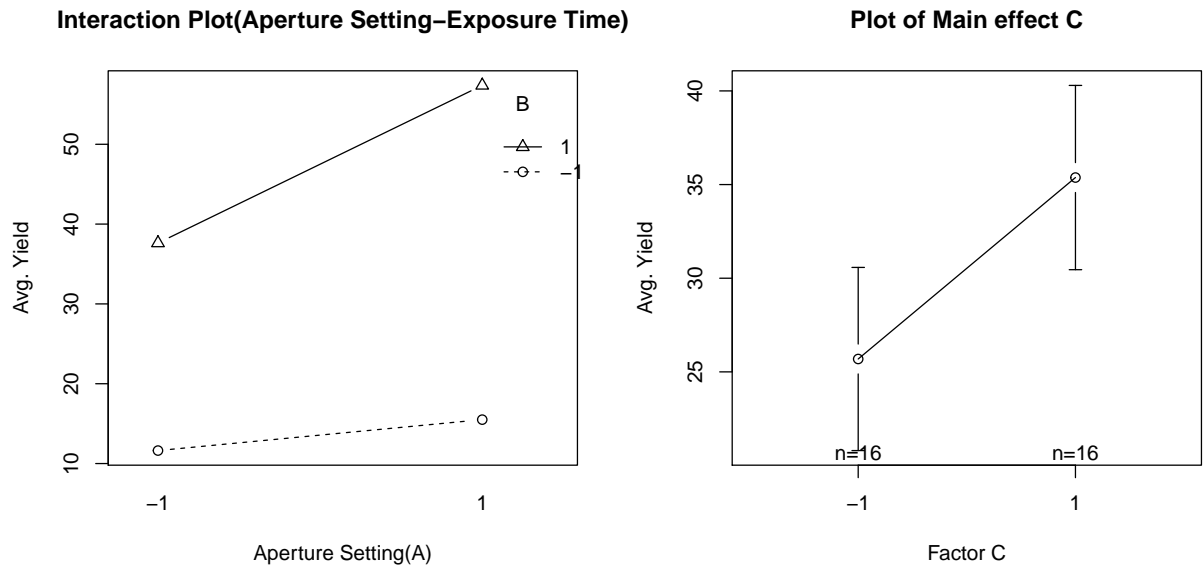
Based on the plots above, we would recommend using higher level of A, higher level of B and higher level of C give the highest yield. That says, we suggest using large aperture setting, 20% above nominal exposure time, and 45s development time.
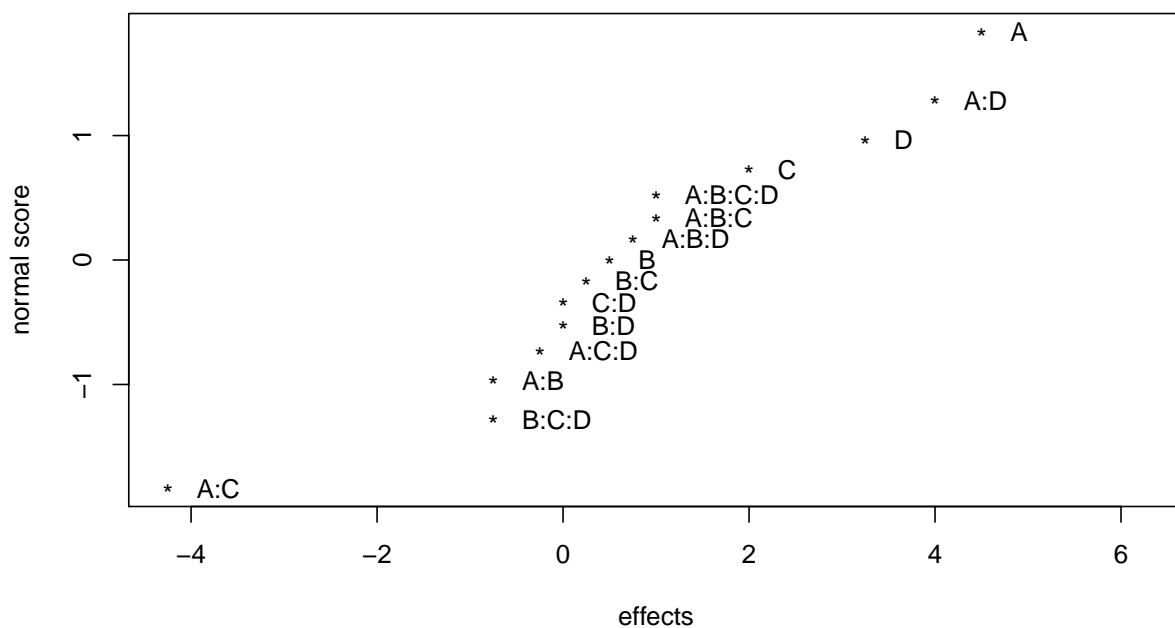
Question 6.32

(a)

```r
# 2^4; 16 runs; no replicate

A <- c(1, -1, -1, -1, -1, -1, 1, -1, 1, -1, 1, 1, 1, -1, 1, 1)
B <- c(1, 1, -1, -1, -1, -1, -1, 1, -1, 1, -1, 1, 1, 1, 1, -1)
C <- c(1, -1, 1, 1, -1, -1, 1, -1, -1, 1, -1, 1, -1, 1, -1, 1)
D <- c(-1, 1, -1, 1, -1, 1, -1, -1, -1, 1, 1, 1, -1, -1, 1, 1)


yield.data <- data.frame(Yield=c(15, 13, 17, 19, 12, 10, 15, 13, 18, 17,
                                 25, 23, 16, 20, 24, 21),
                         A=A, B=B, C=C, D=D)


fit <- lm(Yield~A*B*C*D, data=yield.data)
par(mfrow=c(1,1))
DanielPlot(fit)
```


From the result of Daniel's plot, we can conclude that the factors A, C, D, and interactions AC, AD appear to have larger effects.

(b)

```
yield.data[, 2:5] <- lapply(yield.data[, 2:5], factor)
anova.fit <- aov(Yield~A*C+A*D, data=yield.data)
summary(anova.fit)

##              Df Sum Sq Mean Sq F value   Pr(>F)
## A            1  81.00   81.00  49.846 3.46e-05 ***
## C            1  16.00   16.00   9.846 0.010549 *
## D            1  42.25   42.25  26.000 0.000465 ***
## A:C          1  72.25   72.25  44.462 5.58e-05 ***
## A:D          1  64.00   64.00  39.385 9.19e-05 ***
## Residuals   10  16.25    1.63
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Using the results from part a to conduct a ANOVA test. From the ANOVA table above, we can see that all the factors are significant since their p-values are less than 0.05.

(c)

```
yield.data[, 2:5] <- lapply(yield.data[, 2:5], as.numeric)

yield.data$A[yield.data$A==1] <- 2.5
yield.data$A[yield.data$A==2] <- 3

yield.data$B[yield.data$B==1] <- 0.14
yield.data$B[yield.data$B==2] <- 0.18

yield.data$C[yield.data$C==1] <- 60
yield.data$C[yield.data$C==2] <- 80

yield.data$D[yield.data$D==1] <- 225
yield.data$D[yield.data$D==2] <- 250

# str(yield.data)
summary(lm(Yield~A*C+A*D, data=yield.data))

##
```

```
## Call:
## lm(formula = Yield ~ A * C + A * D, data = yield.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6250 -0.9375  0.1250  0.8750  1.3750
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 209.1250    71.3621   2.930 0.015028 *
## A           -83.5000    25.8433  -3.231 0.009005 **
## C             2.4375     0.3520   6.925 4.07e-05 ***
## D            -1.6300     0.2816  -5.788 0.000176 ***
## A:C          -0.8500     0.1275  -6.668 5.58e-05 ***
## A:D           0.6400     0.1020   6.276 9.19e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.275 on 10 degrees of freedom
## Multiple R-squared:  0.9443,Adjusted R-squared:  0.9165
## F-statistic: 33.91 on 5 and 10 DF,  p-value: 5.856e-06
```
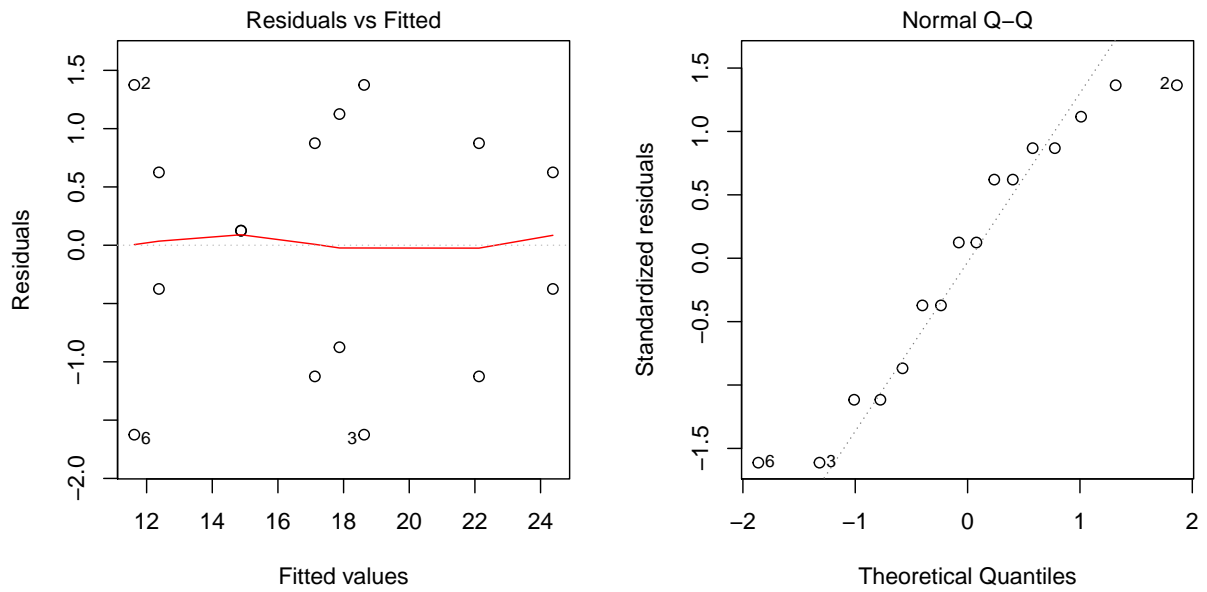
After recoding the factors with their corresponding numerical values, the regression model can be written as

$$\hat{y} = 209.125 - 83.500A + 2.438C - 1.630D - 0.850AC + 0.640AD$$

where $\hat{y}$ is the estimated yield, A is the time, B is the concentration, C is the pressure, and D is the temperature.

(d)

```
par(mfrow=c(1,2))
plot(anova.fit, 1)
plot(anova.fit, 2)
```

From the residuals vs fitted values and normality plots, we can conclude that there is nothing unusual with the residuals.

Question 7.3

```
#2^4; 16 runs; 2 Blocks
A <- rep(c(-1, 1), 8)
B <- rep(c(-1, -1, 1, 1), 4)
C <- rep(c(rep(-1,4), rep(1,4)),2)
D <- c(rep(-1,8), rep(1,8))


R1 <- c(7.037, 14.707, 11.635, 17.273, 10.403, 4.368, 9.360, 13.440, 8.561,
        16.867, 13.876, 19.824, 11.846, 6.125, 11.190, 15.653)
R2 <- c(6.376, 15.219, 12.089, 17.815, 10.151, 4.098, 9.253, 12.923, 8.951,
        17.052, 13.658, 19.639, 12.337, 5.904, 10.935, 15.053)


titanium.data <- data.frame(Crack.len=c(R1, R2), Blocks=rep(c(rep(1,16), rep(2,16))),
                            A=A, B=B, C=C, D=D)
```

```
lm.fit <- lm(Crack.len~Blocks+A*B*C*D, data=titanium.data)
DanielPlot(lm.fit)
```
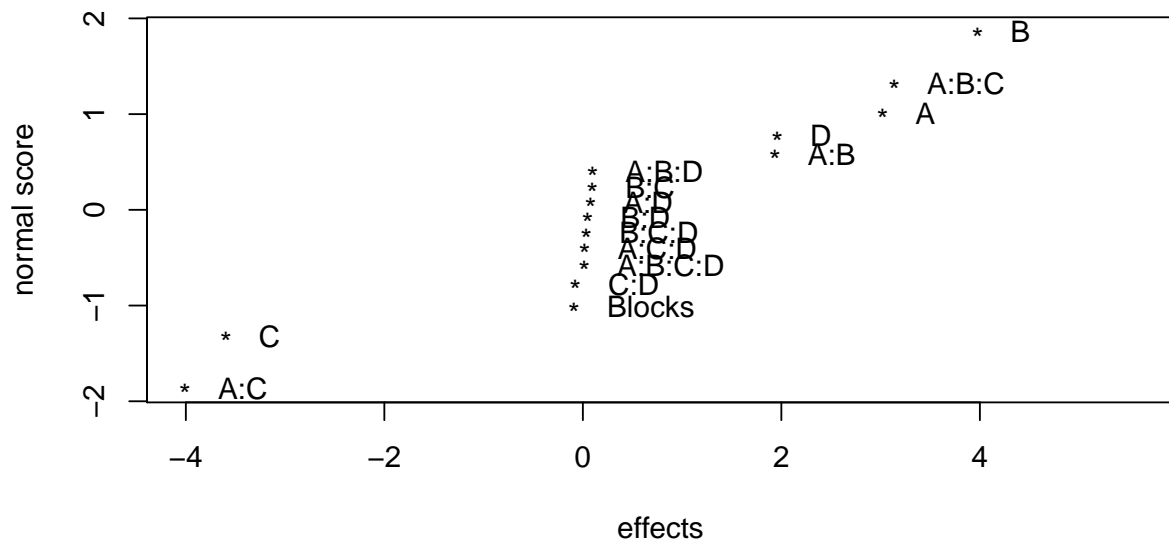
```
titanium.data[,2:6] <- lapply(titanium.data[,2:6], factor)
# str(titanium.data)
anova(lm(Crack.len~Blocks+A*B*C*D, data=titanium.data))

## Analysis of Variance Table
##
## Response: Crack.len
##            Df  Sum Sq Mean Sq   F value     Pr(>F)
## Blocks      1   0.016   0.016    0.1853     0.6730
## A           1  72.909  72.909  852.5942 1.240e-14 ***
## B           1 126.461 126.461 1478.8275 < 2.2e-16 ***
## C           1 103.464 103.464 1209.9066 9.310e-16 ***
## D           1  30.662  30.662  358.5639 6.957e-12 ***
## A:B         1  29.927  29.927  349.9622 8.289e-12 ***
## A:C         1 128.496 128.496 1502.6344 < 2.2e-16 ***
## B:C         1   0.074   0.074    0.8622     0.3678
## A:D         1   0.047   0.047    0.5475     0.4708
## B:D         1   0.018   0.018    0.2089     0.6542
## C:D         1   0.047   0.047    0.5529     0.4686
## A:B:C       1  78.751  78.751  920.9150 7.017e-15 ***
## A:B:D       1   0.077   0.077    0.8985     0.3582
## A:C:D       1   0.003   0.003    0.0342     0.8557
## B:C:D       1   0.010   0.010    0.1187     0.7352
## A:B:C:D     1   0.002   0.002    0.0187     0.8931
## Residuals 15   1.283   0.086
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the Daniel's plot and ANOVA test result, we can conclude that the significant factor effects are A, B, C, D, AB, AC, and ABC. And the blocking effect is not significant since its p-value greater than 0.05.

Question 7.7

7.7   ABCDE confounded with blocks.

The defining contrast to ABCDE is: $L = X_1 + X_2 + X_3 + X_4 + X_5$

$a: L = 1(1) + 1(0) + 1(0) \overset{+0+0}{=} 1 \equiv 1 \pmod{2}$

$b: L = 1(0) + 1(1) + 1(0) \overset{+0+0}{=} 1 \equiv 1 \pmod{2}$

$ab: L = 1(1) + 1(1) + 1(0) \overset{+0+0}{=} 2 \equiv 0 \pmod{2}$

$\vdots$

$abcde: L = 1(1) + 1(1) + 1(1) + 1(1) + 1(1) = 5 \equiv 1 \pmod{2}$

Therefore,

Run	Trt Combination (1)	Block		Run	Trt Comb.	Block
1	(1)	1		17	e	2
2	a	2		18	ae	1
3	b	2		19	be	1
4	ab	1		20	abe	2
5	c	2		21	ce	1
6	ac	1		22	ace	2
7	bc	1		23	bce	2
8	abc	2		24	abce	1
9	d	2		25	de	1
10	ad	1		26	ade	2
11	bd	1		27	bde	2
12	abd	2		28	abde	1
13	cd	1		29	cde	2
14	acd	2		30	acde	1
15	bcd	2		31	bcde	1
16	abcd	1		32	abcde	2

From 6.30, we found that the significant factors are A, B, C, and the interaction between A and B. And it is shown with the Daniel's plot below. Then, by conducting an ANOVA test with the confounded blocks, we will have the following result.

```
A <- rep(c(-1, 1), 16)
B <- rep(c(-1, -1, 1, 1), 8)
C <- rep(c(rep(-1,4), rep(1,4)), 4)
D <- c(rep(-1,8), rep(1,8), rep(-1,8), rep(1,8))
E <- c(rep(-1,16), rep(1,16))
#Confounded w/ blocks
Blocks <- c(1, 2, 2, 1, 2, 1, 1, 2, 2, 1, 1, 2, 1, 2, 2, 1,
            2, 1, 1, 2, 1, 2, 2, 1, 1 , 2, 2, 1, 2, 1, 1, 2)
plant.data <- data.frame(Yield=c(7, 9, 34, 55, 16, 20, 40, 60, 8, 10, 32,
```
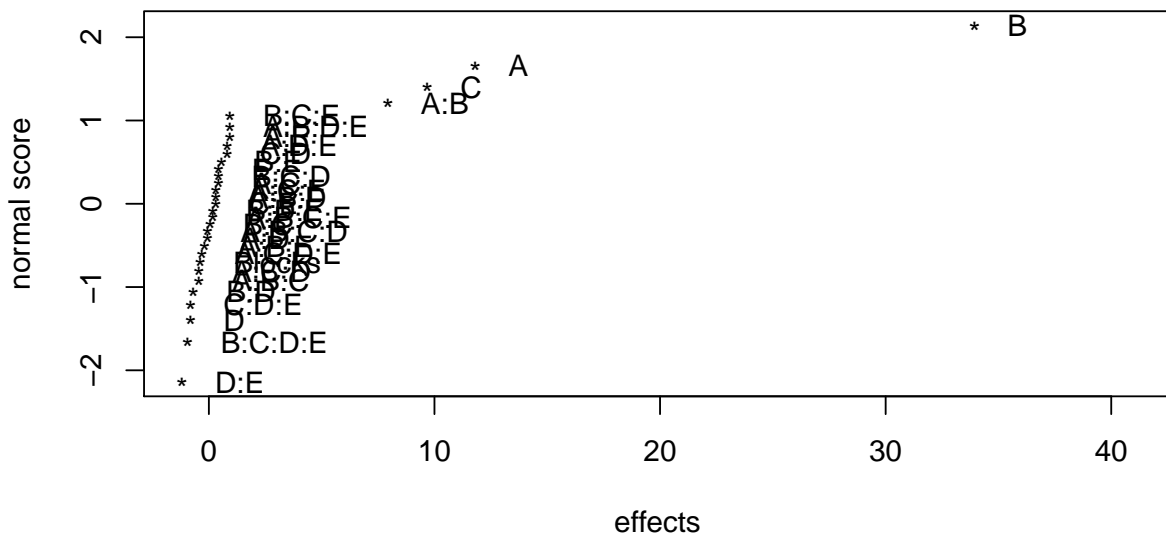
```
                              50, 18, 21, 44, 61, 8, 12, 35, 52, 15, 22,
                              45, 65, 6, 10, 30, 53, 15, 20, 41, 63),
                    Blocks=Blocks, A=A, B=B, C=C, D=D, E=E)


lm.fit <- lm(Yield~Blocks+A*B*C*D*E, data=plant.data)
DanielPlot(lm.fit)
```



```
plant.data[,2:7] <- lapply(plant.data[,2:7], factor)
# str(plant.data)


anova(lm(Yield~Blocks+A*B+C, data=plant.data))

## Analysis of Variance Table
##
## Response: Yield
##            Df Sum Sq Mean Sq   F value    Pr(>F)
## Blocks      1    0.3     0.3    0.0931    0.7627
## A           1 1116.3  1116.3  369.4296 < 2.2e-16 ***
## B           1 9214.0  9214.0 3049.3532 < 2.2e-16 ***
## C           1  750.8   750.8  248.4686 8.027e-15 ***
## A:B         1  504.0   504.0  166.8075 8.080e-13 ***
```

```
## Residuals 26    78.6      3.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the ANOVA result, we can conclude that the confounded blocking effect is not significant since its p-value is greater than 0.05.

Question 7.8

7-8 2^5 factorial design confounded in 4 blocks will have
$2^{5-2} = 8$ runs in each block.

We can choose blocks with ABC and CDE confounded, then
$(ABC)(CDE) = ABC^2DE = ABDE$ is also confounded.

Two defining contrasts are $L_1 = X_1 + X_2 + X_3$
$L_2 = X_3 + X_4 + X_5$.

| Block 1 $L_1=0, L_2=0$	Block 2 $L_1=1, L_2=0$	Block 3 $L_1=0, L_2=1$	Block 4 $L_1=1, L_2=1$
(1), ab,	a, b,	ac, bc,	c, abc,
acd, bcd,	cd, abcd,	d, abd,	ad, bd,
ace, bce,	ce, abce,	e, abe,	ae, be,
de, abde	ade, bde	acde, bcde	cde, abcde

```
A <- rep(c(-1, 1), 16)

B <- rep(c(-1, -1, 1, 1), 8)

C <- rep(c(rep(-1,4), rep(1,4)), 4)

D <- c(rep(-1,8), rep(1,8), rep(-1,8), rep(1,8))

E <- c(rep(-1,16), rep(1,16))

#Confounded w/ blocks

Blocks <- c(1, 2, 2, 1, 4, 3, 3, 4, 3, 4, 4, 3, 2, 1, 1, 2,
            3, 4, 4, 3, 2, 1, 1, 2, 1, 2, 2, 1, 4, 3, 3, 4)

# table(Blocks)

plant.data <- data.frame(Yield=c(7, 9, 34, 55, 16, 20, 40, 60, 8, 10, 32,
                                 50, 18, 21, 44, 61, 8, 12, 35, 52, 15, 22,
                                 45, 65, 6, 10, 30, 53, 15, 20, 41, 63),
                         Blocks=Blocks, A=A, B=B, C=C, D=D, E=E)


lm.fit <- lm(Yield~Blocks+A*B*C*D*E, data=plant.data)

DanielPlot(lm.fit)
```
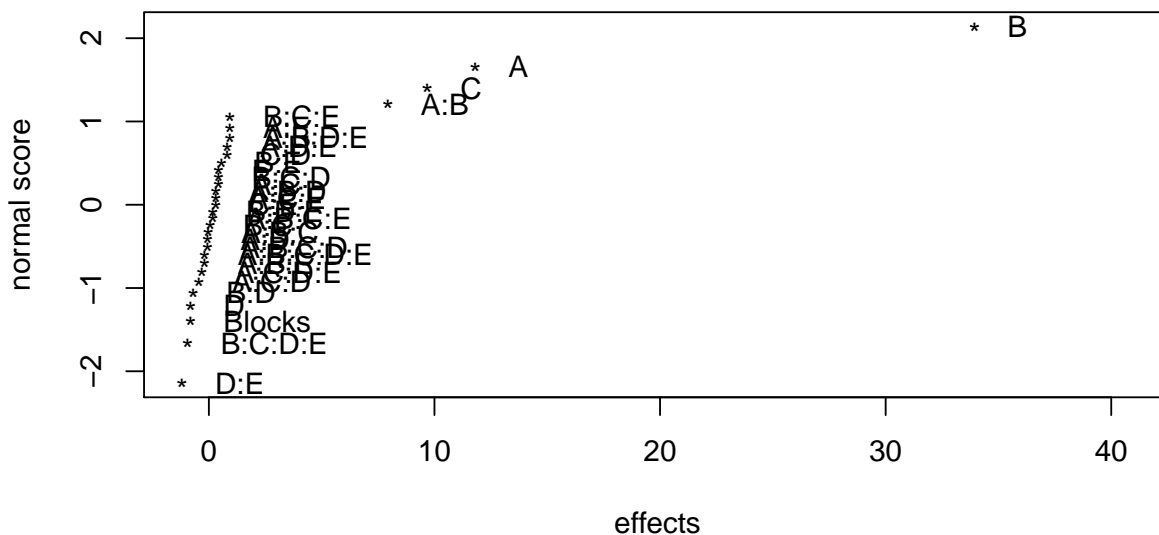
```r
plant.data[,2:7] <- lapply(plant.data[,2:7], factor)
# str(plant.data)


anova(lm(Yield~Blocks+A*B+C, data=plant.data))

## Analysis of Variance Table
##
## Response: Yield
##            Df Sum Sq Mean Sq   F value    Pr(>F)
## Blocks      3   13.8     4.6    1.7038    0.1929
## A           1 1116.3  1116.3  412.1654 < 2.2e-16 ***
## B           1 9214.0  9214.0 3402.1038 < 2.2e-16 ***
## C           1  750.8   750.8  277.2115 1.096e-14 ***
## A:B         1  504.0   504.0  186.1038 8.411e-13 ***
## Residuals 24   65.0     2.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

By repeating problem 7.7, we use four blocks this time. Then, based on the ANOVA result, we can conclude that the confounded blocking effect is not significant since its p-value is greater than 0.05.

Question 7.9

7.9   ACDE, BCD Confounded blocks.

Then, $(ACDE)(BCD) = ABC^2D^2E = ABE$ also confounded.

Two defining contrasts are:

$$L_1 = X_1 + X_3 + X_4 + X_5$$

$$L_2 = X_2 + X_3 + X_4$$

| Block 1 (0,0)	Block 2 (1,0)	Block 3 (0,1)	Block 4 (1,1)
(1), ae,	a, e,	b, abe,	c, ace,
cd, abc,	acd, bc,	bcd, ac,	d, ab,
acde, bce,	cde, abce,	abcde, ce,	ade, be,
abd, bde	bd, abde	ad, de	abcd, bcde

```
A <- rep(c(-1, 1), 16)

B <- rep(c(-1, -1, 1, 1), 8)

C <- rep(c(rep(-1,4), rep(1,4)), 4)

D <- c(rep(-1,8), rep(1,8), rep(-1,8), rep(1,8))

E <- c(rep(-1,16), rep(1,16))

#Confounded w/ blocks

Blocks <- c(1, 2, 3, 4, 4, 3, 2, 1, 4, 3, 2, 1, 1, 2, 3, 4,
            2, 1, 4, 3, 3, 4, 1, 2, 3, 4, 1, 2, 2, 1, 4, 3)

# table(Blocks)

plant.data <- data.frame(Yield=c(7, 9, 34, 55, 16, 20, 40, 60, 8, 10, 32,
                                 50, 18, 21, 44, 61, 8, 12, 35, 52, 15, 22,
                                 45, 65, 6, 10, 30, 53, 15, 20, 41, 63),
                         Blocks=Blocks, A=A, B=B, C=C, D=D, E=E)


lm.fit <- lm(Yield~Blocks+A*B*C*D*E, data=plant.data)

DanielPlot(lm.fit)
```
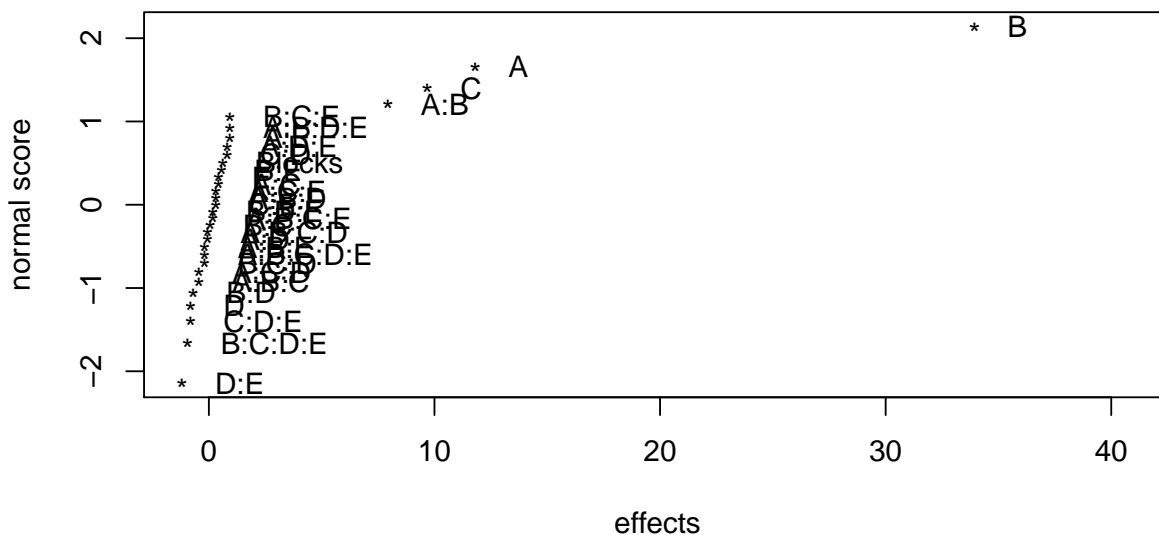
```r
plant.data[,2:7] <- lapply(plant.data[,2:7], factor)


anova(lm(Yield~Blocks+A*B+C, data=plant.data))

## Analysis of Variance Table
##
## Response: Yield
##            Df Sum Sq Mean Sq  F value     Pr(>F)
## Blocks      3    2.6     0.9   0.2721     0.8449
## A           1 1116.3  1116.3 351.3541  7.758e-16 ***
## B           1 9214.0  9214.0 2900.1541 < 2.2e-16 ***
## C           1  750.8   750.8 236.3115  6.357e-14 ***
## A:B         1  504.0   504.0 158.6459  4.556e-12 ***
## Residuals  24   76.2     3.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Similarly, from the Daniel's plot and ANOVA test result, we will conclude that the factor effects of A, B, C, and AB are significant, while the confounded blocking effect is not significant.