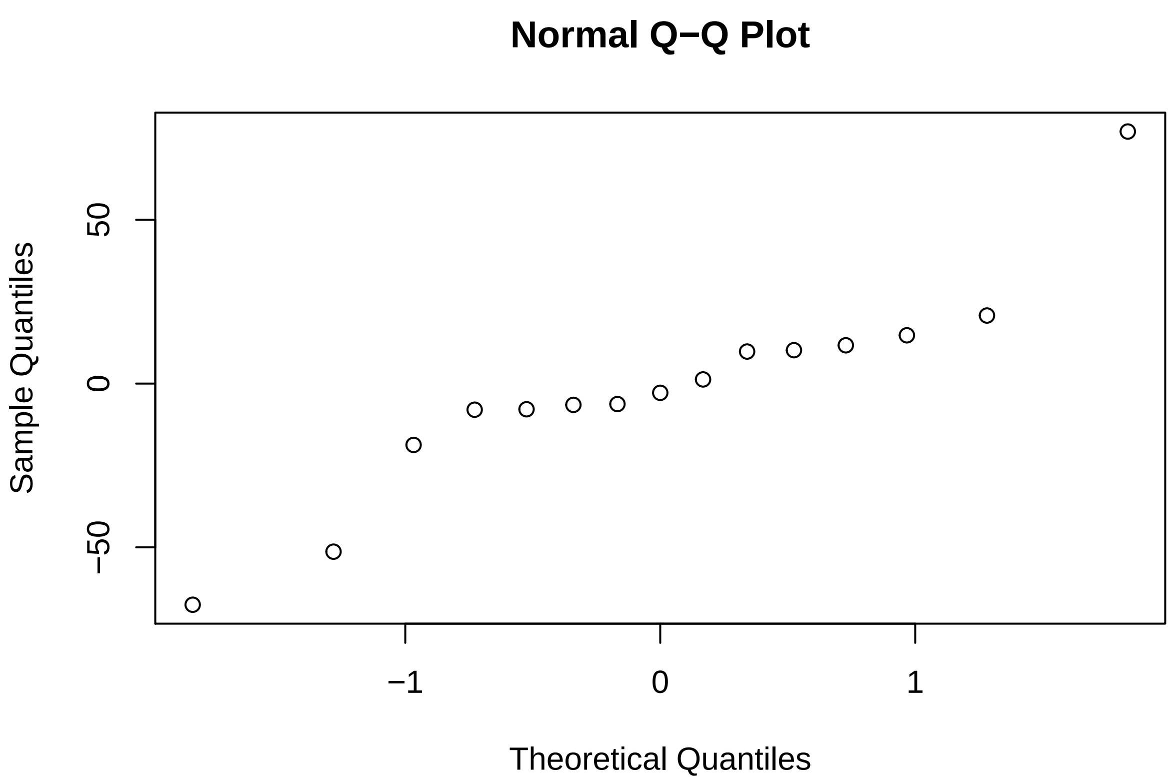
Homework 6

6.17

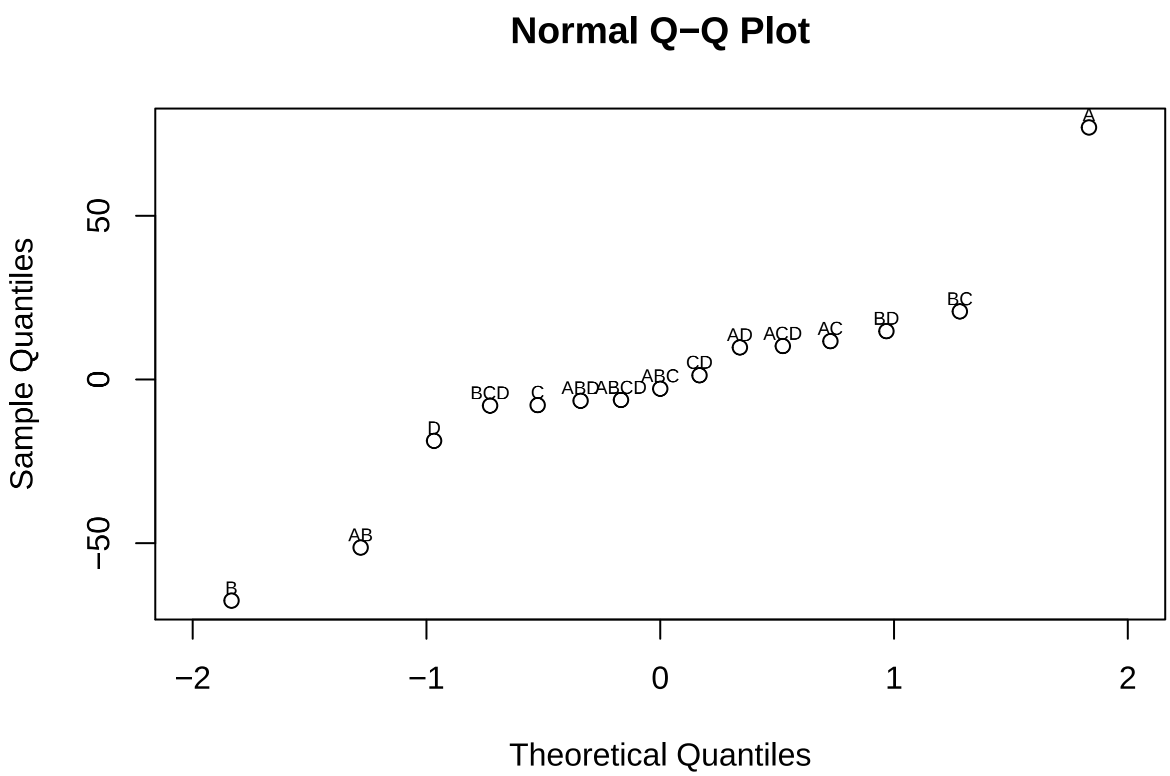
```
effects <- c("A","B","C","D","AB","AC","AD","BC",
             "BD","CD","ABC","ABD","ACD","BCD","ABCD")
values <- c(76.95,-67.52,-7.84,-18.73,-51.32,
            11.69,9.78,20.78,14.74,1.27,-2.82,
            -6.5,10.2,-7.98,-6.25)
df1 <- data.frame(effects, values)
```

a

```
data_sort <- df1[order(df1$value),]
x <- qqnorm(data_sort$values)[1]$x
```




```
qqnorm(data_sort$values, xlim=c(-2,2))
text(sort(x), data_sort$values+4, labels=data_sort$effects, cex=0.6)
```



b

The effects A, B, and AB are significant according to the Normal QQ plot. Therefore a model could include these effects only.

```
beta_A  <- df1$values[df1$effects == "A"]  / 2
beta_B  <- df1$values[df1$effects == "B"]  / 2
beta_AB <- df1$values[df1$effects == "AB"] / 2
```

6.24

```
library(readxl)
df <- read_excel("~/vmshare/stats 101b datasets/Chapter 6 (P.24).xlsx")
#read_excel("/Users/Earle/Downloads/Chapter 6 (P.24).xlsx")
df$Class <- as.factor(df$Class)
df$Type  <- as.factor(df$Type)
df$Price <- as.factor(df$Price)
A <- -as.level2(df$Class)
B <- as.level2(df$Type)
C <- as.level2(df$Price)

df2 <- cbind(A,B,C,Number_of_orders=df$`Number of Orders` )
df2 <- as.data.frame(df2)
```

a

The factors which are significant are C, AB, AC, BC with p-values: 0.0085163, 0.0056019, 0.0004176, and 0.0037282 respectively.

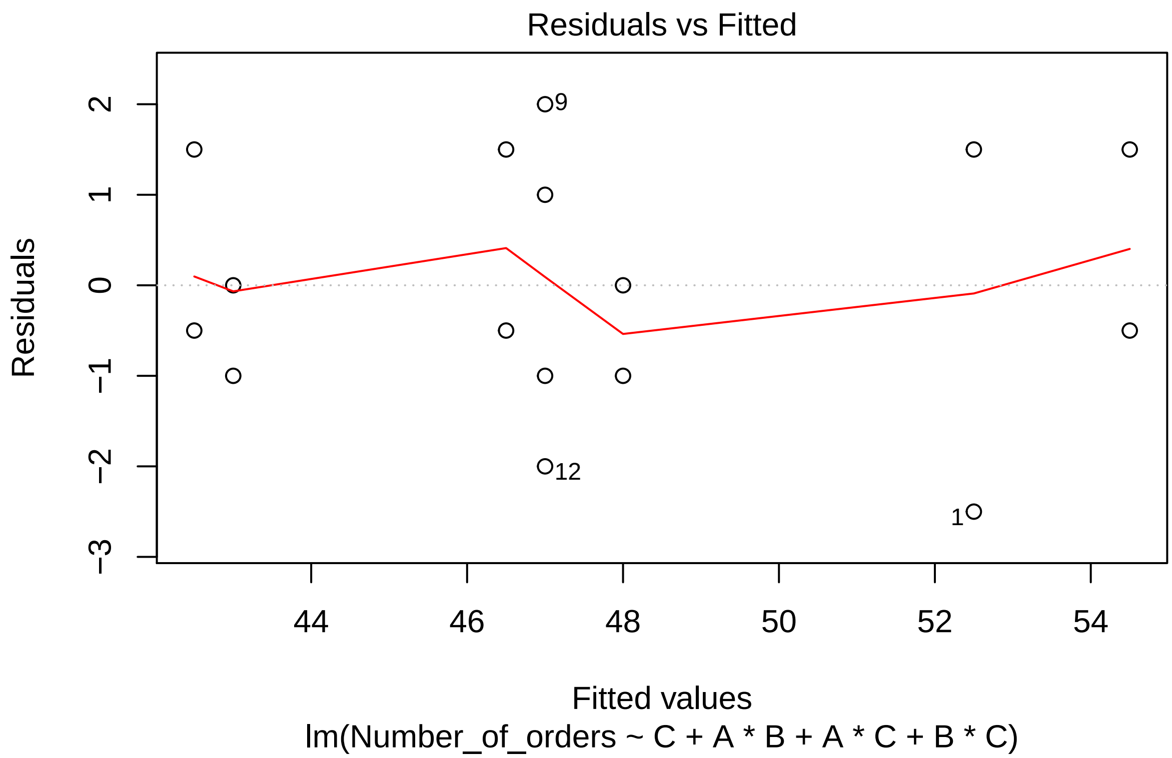
```
m <- lm(Number_of_orders ~ A*B*C,df2)
anova(m)
```

```
## Analysis of Variance Table
##
## Response: Number_of_orders
##           Df Sum Sq Mean Sq F value    Pr(>F)
## A          1  12.25   12.25  4.0833 0.0779708 .
## B          1   2.25    2.25  0.7500 0.4116944
## C          1  36.00   36.00 12.0000 0.0085163 **
## A:B        1  42.25   42.25 14.0833 0.0056019 **
## A:C        1 100.00  100.00 33.3333 0.0004176 ***
## B:C        1  49.00   49.00 16.3333 0.0037282 **
## A:B:C      1   4.00    4.00  1.3333 0.2815369
## Residuals  8  24.00    3.00
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

b

The residual plot does not show any indication of non-constant variance. The normal Q-Q plot shows that the residuals are not following the normal distribution.

```
#par(mfrow=c(2,2))
m <- lm(Number_of_orders ~ C + A*B + A*C + B*C,df2)
plot(m,1:2)
```

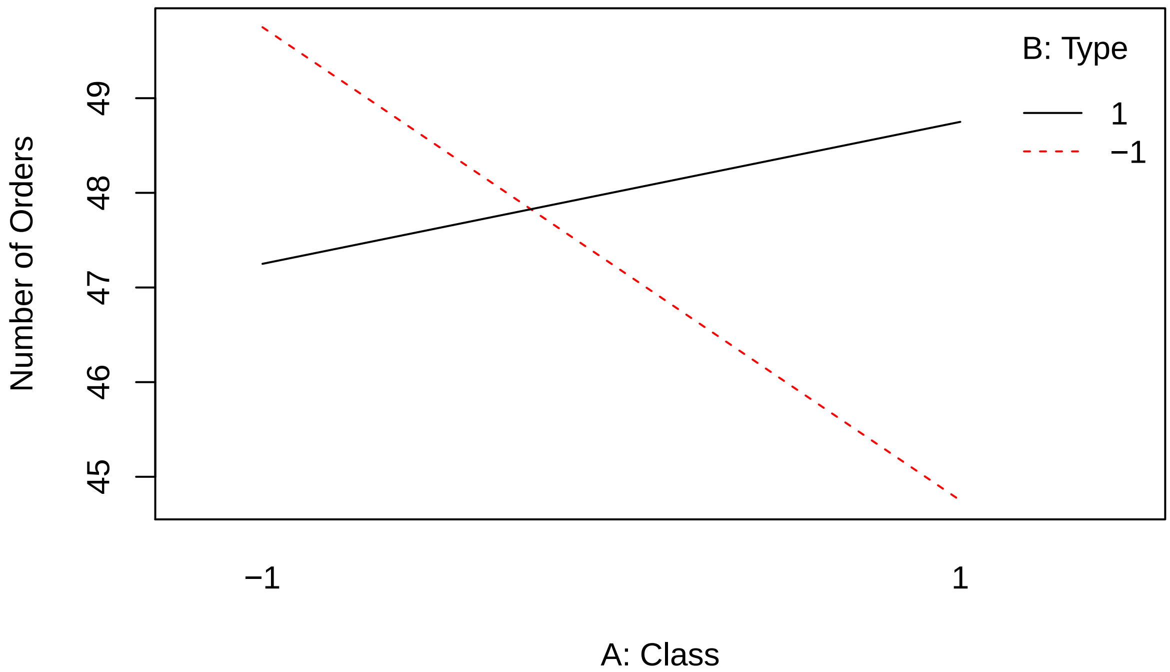


!

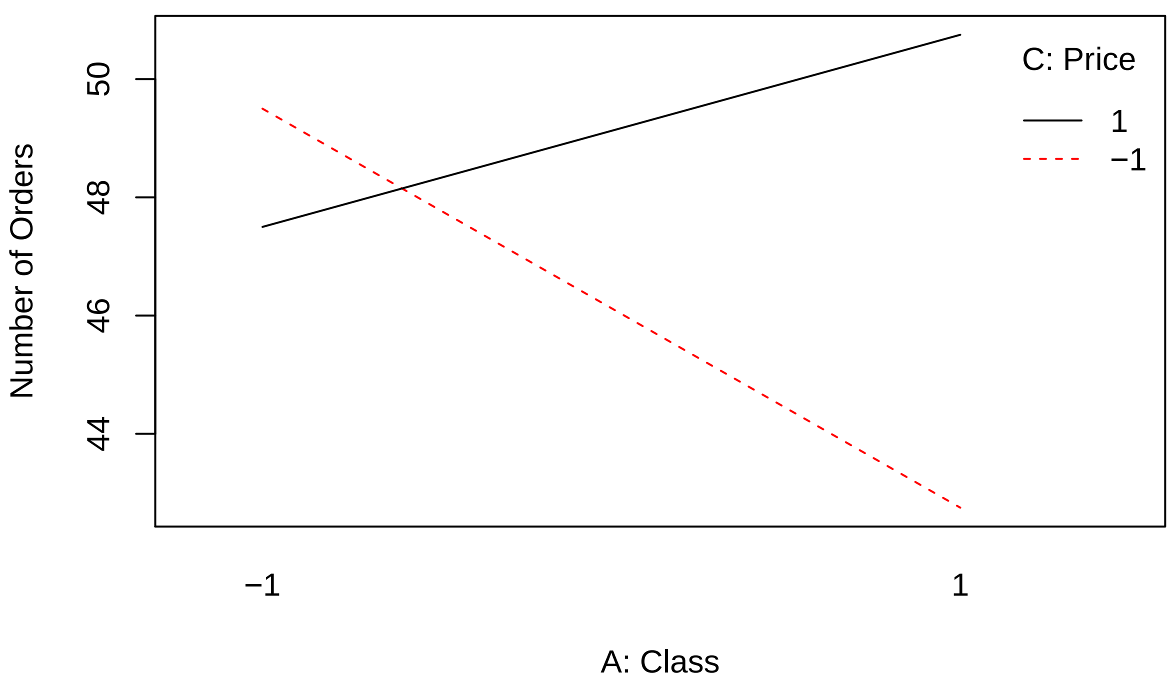
c

According to the interaction plots, If we need to use the price of \$19.95 the I recommend 3^{rd} class mail with black and white brochures. However if we need to use price \$24.95 then I recommend 1^{st} class mail with color brochures.

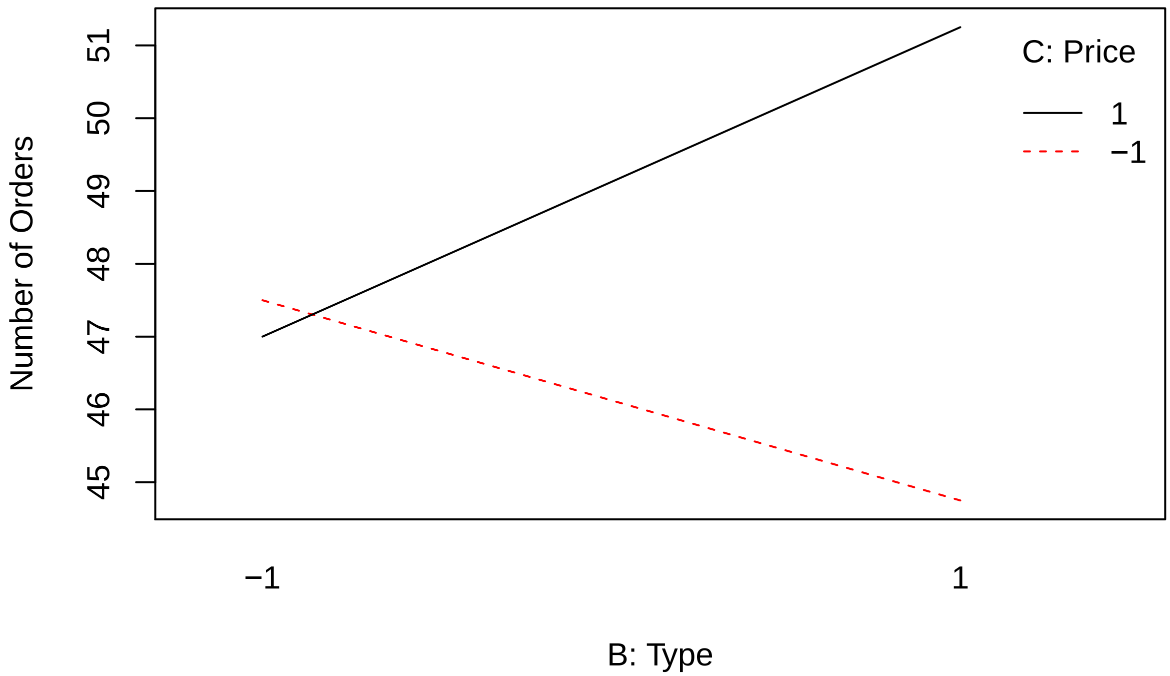
```
#par(mfrow=c(2,2))
interaction.plot(df2$A,df2$B,df2$Number_of_orders,
                 xlab = "A: Class", ylab="Number of Orders",
                 trace.label = "B: Type",col=c("red", "black"))
```




```
interaction.plot(df2$A,df2$C,df2$Number_of_orders,
                 xlab = "A: Class", ylab="Number of Orders",
                 trace.label = "C: Price",col=c("red", "black"))
```




```
interaction.plot(df2$B,df2$C,df2$Number_of_orders,
                 xlab = "B: Type", ylab="Number of Orders",
                 trace.label = "C: Price",col=c("red", "black"))
```



6.30

```
library(readxl)
#df <-  read_excel("/Users/Earle/Downloads/Test_book_prob.xlsx")
df <- read_excel("~/vmshare/stats 101b datasets/Ch6(P30).xlsx")
df$`Pan Material` <- as.factor(df$`Pan Material`)
df$`Stirring Method` <- as.factor(df$`Stirring Method`)
df$`Mix Brand` <- as.factor(df$`Mix Brand`)
A <- -as.level2(df$`Pan Material`)
B <- -as.level2(df$`Stirring Method`)
C <- -as.level2(df$`Mix Brand`)

df2 <- cbind(A,B,C,Scrumptiousness = df$Scrumptiousness)
df2 <- as.data.frame(df2)
```

a)

The ANOVA indicates that the most significant factor is the pan matrial. Creating model with this we see that a glass pan plays the significant role in scrumptiousness.

```
m <- lm(Scrumptiousness ~ A*B*C, df2)
anova(m)
```

```
## Analysis of Variance Table
##
## Response: Scrumptiousness
##            Df Sum Sq Mean Sq F value   Pr(>F)
```

```
## A          1  72.25  72.250 11.9527 0.001049 **
## B          1  18.06  18.062  2.9882 0.089385 .
## C          1   0.06   0.063  0.0103 0.919370
## A:B        1   0.06   0.062  0.0103 0.919370
## A:C        1   1.56   1.562  0.2585 0.613154
## B:C        1   1.00   1.000  0.1654 0.685751
## A:B:C      1   0.25   0.250  0.0414 0.839584
## Residuals 56 338.50   6.045
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
m <- lm(Scrumptiousness ~ A, df2)
summary(m)
```

```
##
## Call:
## lm(formula = Scrumptiousness ~ A, data = df2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.625 -1.500  0.375  1.406  6.500
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   11.562      0.301   38.41  < 2e-16 ***
## A              1.062      0.301    3.53  0.00079 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.408 on 62 degrees of freedom
## Multiple R-squared:  0.1673, Adjusted R-squared:  0.1539
## F-statistic: 12.46 on 1 and 62 DF,  p-value: 0.0007901
```

b)

No we do not have 8 replicates, each tester is tasting the same brownie. The appropriate method is to use the probability plots to determine which is significant.

c)

Average:

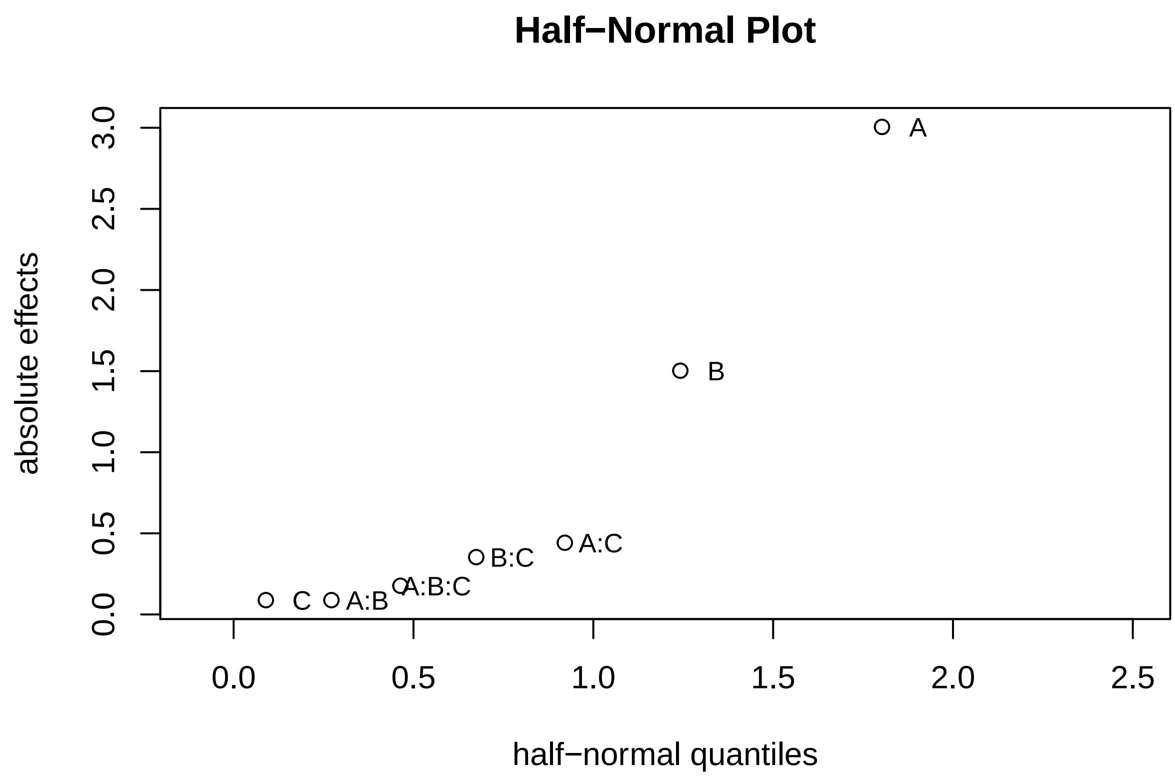
Factors A and B affect the mean of scrumptiousness.

```r
m <- lm(Scrumptiousness ~ A*B*C, df2)
effects <- m$coefficients[-1]*2
xy <- cbind(model.matrix(m)[,-1], scrumptiousness = df2$Scrumptiousness)

xy_avg <- matrix(nrow =8, ncol = 8)
colnames(xy_avg) <- colnames(xy)
for(i in 1:8){
  xy_avg[i, 1:7] <- print(xy[(i-1)*8+1,1:7])
```

```r
  xy_avg[i, 8] <- mean(xy[ ((i-1)*8+1):((i-1)*8+8) ,8] )
}
```

```
##      A     B     C   A:B   A:C   B:C A:B:C
##     -1    -1    -1     1     1     1    -1
##      A     B     C   A:B   A:C   B:C A:B:C
##      1    -1    -1    -1    -1     1     1
##      A     B     C   A:B   A:C   B:C A:B:C
##     -1     1    -1    -1     1    -1     1
##      A     B     C   A:B   A:C   B:C A:B:C
##      1     1    -1     1    -1    -1    -1
##      A     B     C   A:B   A:C   B:C A:B:C
##     -1    -1     1     1    -1    -1     1
##      A     B     C   A:B   A:C   B:C A:B:C
##      1    -1     1    -1     1    -1    -1
##      A     B     C   A:B   A:C   B:C A:B:C
##     -1     1     1    -1    -1     1    -1
##      A     B     C   A:B   A:C   B:C A:B:C
##      1     1     1     1     1     1     1
```

```r
xy_avg <- as.data.frame(xy_avg)

A <- xy_avg$A
B <- xy_avg$B
C <- xy_avg$C

m <- lm(xy_avg$scrumptiousness ~ A*B*C)

halfnormalplot(m$effects[-1], label=T)
```

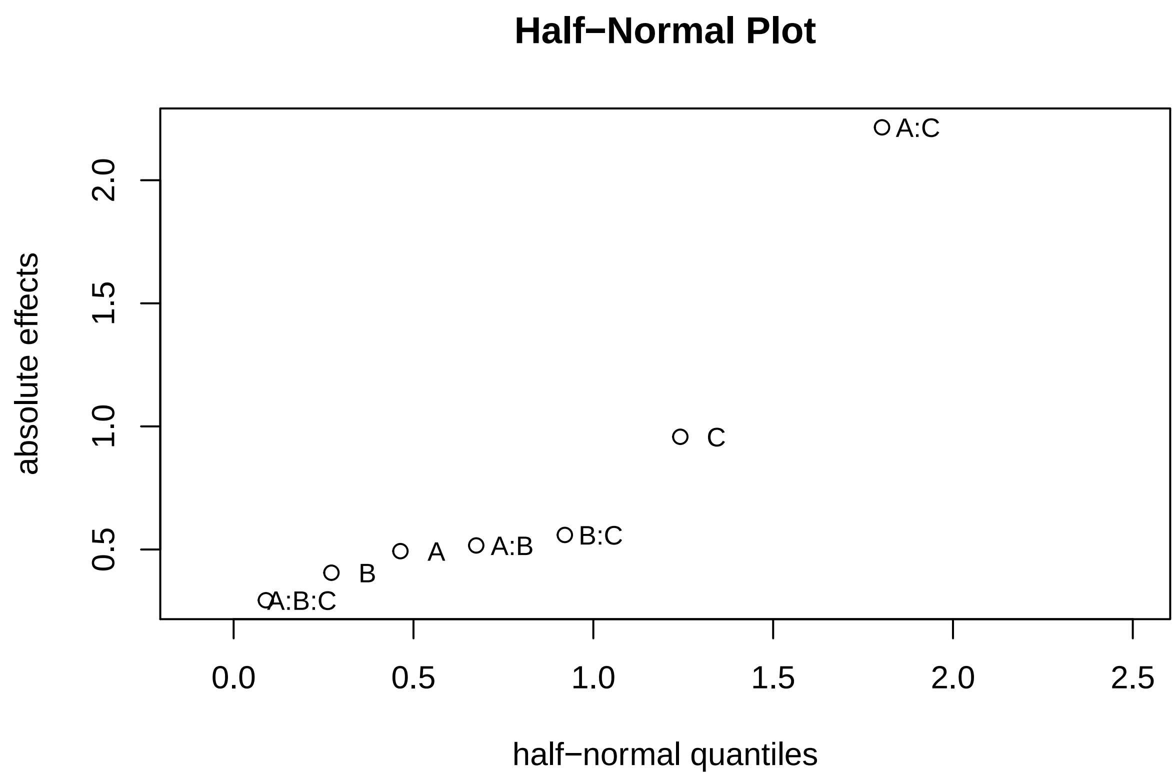


Standard Deviation: Factors AC and C affect the standard deviations of scrumptiousness.

```
xy_sd <- matrix(nrow =8, ncol = 8)
colnames(xy_sd) <- colnames(xy)
for(i in 1:8){
  xy_sd[i, 1:7] <- print(xy[(i-1)*8+1,1:7])
  xy_sd[i, 8] <- sd(xy[ ((i-1)*8+1):((i-1)*8+8) ,8] )
}
```

```
##     A     B     C   A:B   A:C   B:C A:B:C
##    -1    -1    -1     1     1     1    -1
##     A     B     C   A:B   A:C   B:C A:B:C
##     1    -1    -1    -1    -1     1     1
##     A     B     C   A:B   A:C   B:C A:B:C
##    -1     1    -1    -1     1    -1     1
##     A     B     C   A:B   A:C   B:C A:B:C
##     1     1    -1     1    -1    -1    -1
##     A     B     C   A:B   A:C   B:C A:B:C
##    -1    -1     1     1    -1    -1     1
##     A     B     C   A:B   A:C   B:C A:B:C
##     1    -1     1    -1     1    -1    -1
##     A     B     C   A:B   A:C   B:C A:B:C
##    -1     1     1    -1    -1     1    -1
##     A     B     C   A:B   A:C   B:C A:B:C
##     1     1     1     1     1     1     1
```

```
xy_sd <- as.data.frame(xy_sd)

A <- xy_sd$A
B <- xy_sd$B
C <- xy_sd$C

m <- lm(xy_sd$scrumptiousness ~ A*B*C)

halfnormalplot(m$effects[-1], label=T)
```



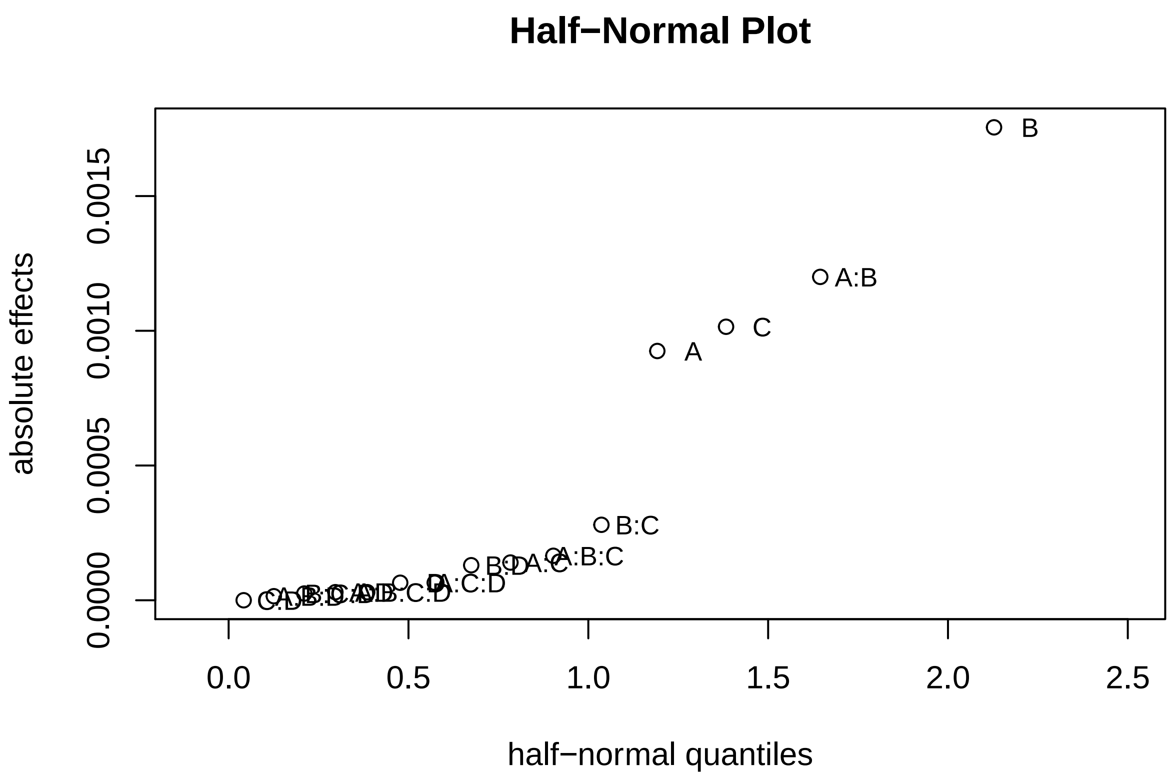
6.35

```
#df4 <- read_excel("/Users/Earle/Downloads/Chapter6_p35.xlsx")
df <- read_excel("~/vmshare/stats 101b datasets/CH6(P35).xlsx")
df$`Tool Angle` <- as.factor(df$`Tool Angle`)
df$Viscosity <- as.factor(df$Viscosity)
df$`Feet Rate` <- as.factor(df$`Feet Rate`)
df$`Cutting Fluid` <- as.factor(df$`Cutting Fluid`)

A <- as.level2(df$`Tool Angle`)
B <- as.level2(df$Viscosity)
C <- as.level2(df$`Feet Rate`)
D <- as.level2(df$`Cutting Fluid`)

df2 <- cbind(A,B,C,D, surface_roughness = df$`Surface Roughness`)
df2 <- as.data.frame(df2)
```

a

The half normal probability indicates that A, B, C, and AB significant.

```
m <- lm(surface_roughness ~ A*B*C*D,df2)
halfnormalplot(m$effects[2:16],label=T)
```



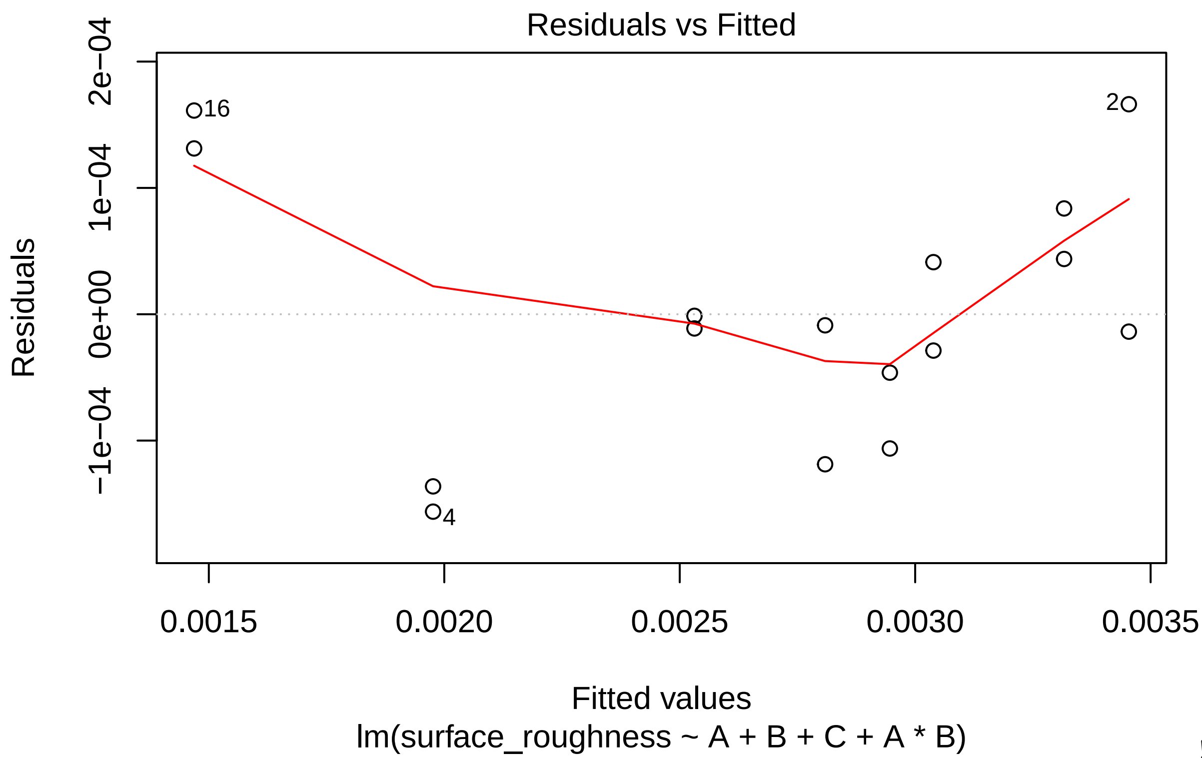
b

The residual plot show a trend line which is an indication of non-constant variance. The normal QQ plot shows that the resiuals are approximately following the normal distribution.

```
m <- lm(surface_roughness ~ A + B + C + A*B ,df2)
anova(m)
```

```
## Analysis of Variance Table
##
## Response: surface_roughness
##           Df     Sum Sq    Mean Sq F value    Pr(>F)
## A          1 8.5562e-07 8.5562e-07  61.425 7.936e-06 ***
## B          1 3.0800e-06 3.0800e-06 221.114 1.249e-08 ***
## C          1 1.0302e-06 1.0302e-06  73.960 3.263e-06 ***
## A:B        1 1.4400e-06 1.4400e-06 103.377 6.261e-07 ***
## Residuals 11 1.5322e-07 1.3930e-08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(m,1:2)
```

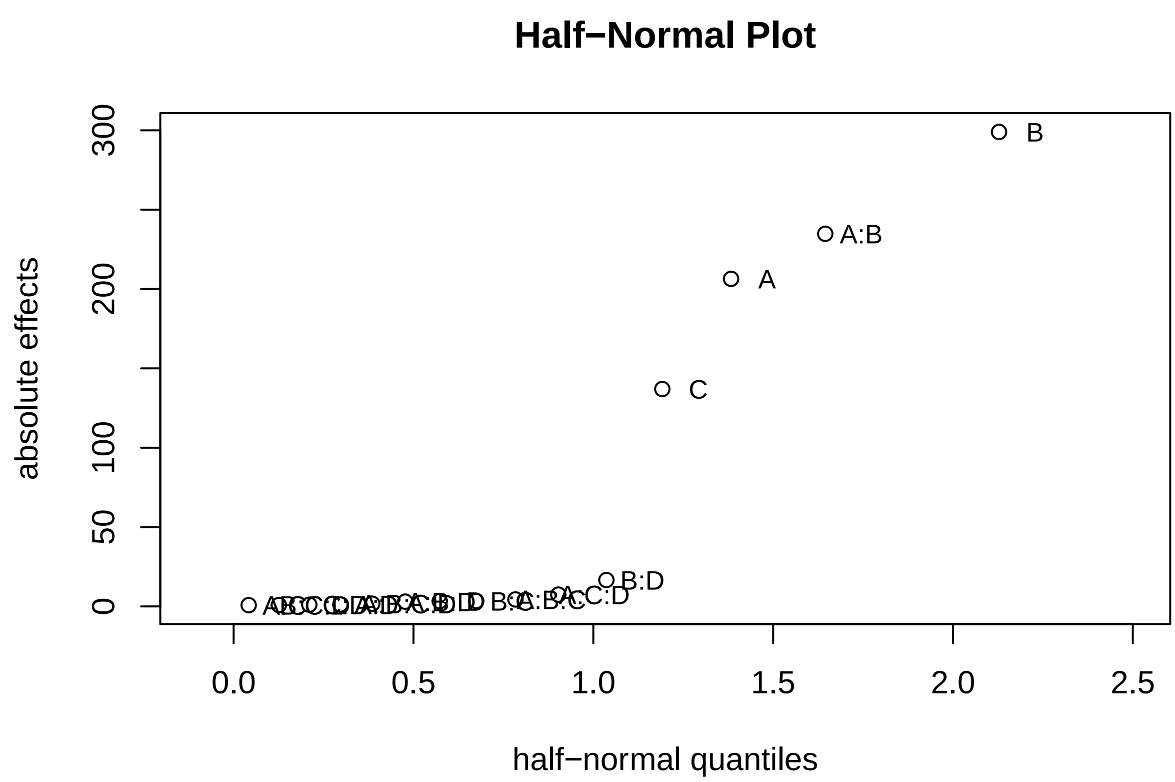


!

c

The transformation eliminated the trend line that was present in the earlier untransformed analysis. It also made the residuals follow the normal distribution more closely.

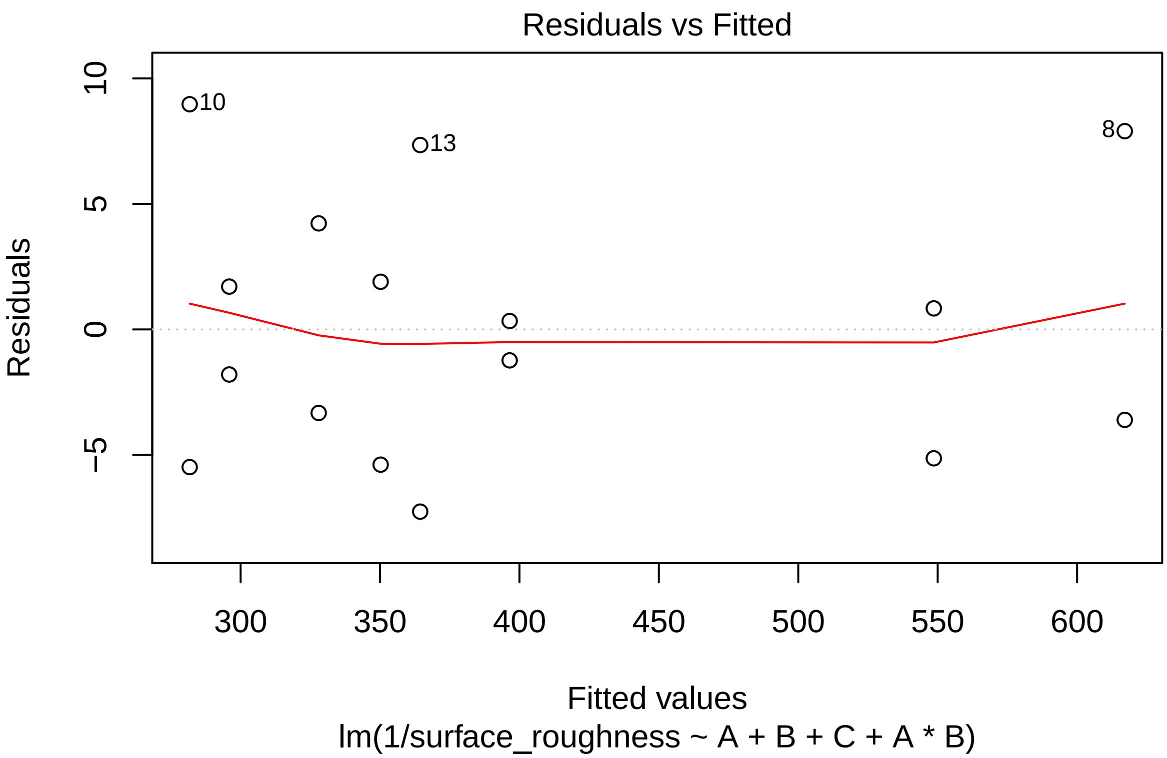
```
m <- lm(1/surface_roughness ~ A*B*C*D, df2)
halfnormalplot(m$effects[2:16],label=T)
```




```r
m <- lm(1/surface_roughness ~ A+B+C+A*B, df2)
anova(m)
```

```
## Analysis of Variance Table
##
## Response: 1/surface_roughness
##           Df Sum Sq Mean Sq F value    Pr(>F)
## A          1  42611   42611 1205.11 1.359e-12 ***
## B          1  89386   89386 2527.99 2.367e-14 ***
## C          1  18762   18762  530.63 1.168e-10 ***
## A:B        1  55130   55130 1559.16 3.332e-13 ***
## Residuals 11    389      35
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plot(m,1:2)
```



!

d

$$\frac{1}{Surface\ Roughness} = 397.81 + 51.61A + 74.74B + 34.24C + 58.70AB$$

6.45

$$2 * \hat{\beta} \pm t_{\frac{\alpha}{2}, df_E} \times se(\hat{\beta}) * 2 \leq 1.5 \quad (1)$$

$$2 * \hat{\beta} \pm t_{\frac{\alpha}{2}, df_E} \times \sqrt{V(2\hat{\beta})} * 2 \leq 1.5 \quad (2)$$

$$2 * \hat{\beta} \pm t_{\frac{\alpha}{2}, df_E} \times \sqrt{4V(\hat{\beta})} * 2 \leq 1.5 \quad (3)$$

$$2 * \hat{\beta} \pm t_{\frac{\alpha}{2}, df_E} \times 2 * \sqrt{\frac{\sigma^2}{n * 2^k}} * 2 \leq 1.5 \quad (4)$$

$$2 * \hat{\beta} \pm t_{\frac{\alpha}{2}, df_E} \times 4 * \sqrt{\frac{1}{2n}} \leq 1.5 \quad (5)$$

$$2 * \hat{\beta} \pm t_{0.975, 8n-8} \times \sqrt{\frac{8}{n}} \leq 1.5 \quad (6)$$

$$(7)$$

7.14

Since we have two replicates then we should have two block, one for each replicate.