Chapter 6 (8th Edition): 6.1, 6.3, 6.5, 6.6, 6.10, 6.11, 6.17, 6.22, 6.23, 6.25, 6.36, 6.37

Jeremy Ling & Emmanuel Mejia

April 16, 2018

```
# loading libraries
library(car)
library(gplots)
```

```
## Warning: package 'gplots' was built under R version 3.4.4

##
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':
##
##     lowess
```

6.1

An engineer is interested in the effects of cutting speed (A), tool geometry (B), and cutting angle (C) on the life (in hours) of a machine tool. Two levels of each factor are chosen, and three replicates of a 2^3 factorial design are run. The results are as follows:

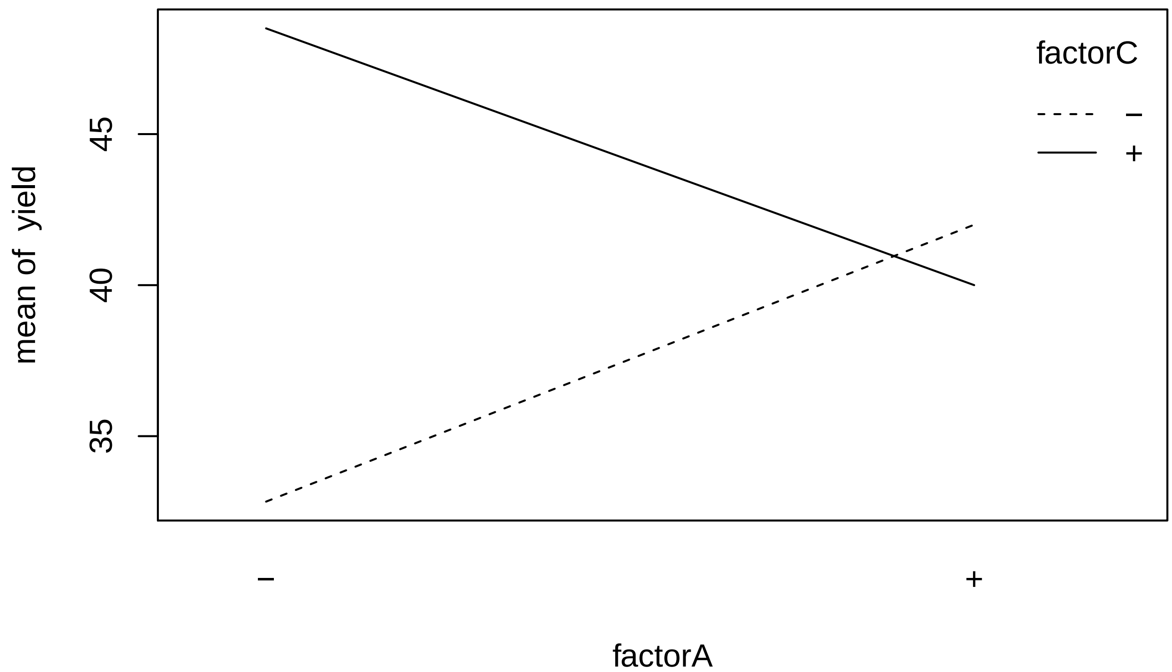
(a) Estimate the factor effects. Which effects appear to be large?

```
# creating data table
factorA = rep(c("-","+","-","+","-","+","-","+"), times = 3)
factorB = rep(c("-","-","+","+","-","-","+","+"), times = 3)
factorC = rep(c("-","-","-","-","+","+","+","+"), times = 3)
Rep = rep(c("I", "II", "III"), each = 8)
yield = c(22,32,35,55,44,40,60,39,31,43,34,47,45,37,50,41,25,29,50,46,38,36,54,47)

cutting.speed.long = data.frame(factorA, factorB, factorC, Rep, yield)

# defining coded
coded=function(x) #a function to code variable x
{
  ifelse(x=="+", 1, -1)
}

# linear regression
cutting.speed.lm=lm(yield ~ coded(factorA) * coded(factorB) * coded(factorC), cutting.speed.long)
summary(cutting.speed.lm)
```

```
##
## Call:
## lm(formula = yield ~ coded(factorA) * coded(factorB) * coded(factorC),
##     data = cutting.speed.long)
##
## Residuals:
##    Min     1Q Median     3Q    Max
```

```
## -5.667 -3.500 -1.167  3.167 10.333
##
## Coefficients:
##                                              Estimate Std. Error t value
## (Intercept)                                   40.8333     1.1211  36.421
## coded(factorA)                                 0.1667     1.1211   0.149
## coded(factorB)                                 5.6667     1.1211   5.054
## coded(factorC)                                 3.4167     1.1211   3.048
## coded(factorA):coded(factorB)                 -0.8333     1.1211  -0.743
## coded(factorA):coded(factorC)                 -4.4167     1.1211  -3.939
## coded(factorB):coded(factorC)                 -1.4167     1.1211  -1.264
## coded(factorA):coded(factorB):coded(factorC)  -1.0833     1.1211  -0.966
##                                              Pr(>|t|)
## (Intercept)                                   < 2e-16 ***
## coded(factorA)                               0.883680
## coded(factorB)                               0.000117 ***
## coded(factorC)                               0.007679 **
## coded(factorA):coded(factorB)                0.468078
## coded(factorA):coded(factorC)                0.001172 **
## coded(factorB):coded(factorC)                0.224475
## coded(factorA):coded(factorB):coded(factorC) 0.348282
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.492 on 16 degrees of freedom
## Multiple R-squared:  0.7696, Adjusted R-squared:  0.6689
## F-statistic: 7.637 on 7 and 16 DF,  p-value: 0.0003977
```

```r
# interaction plot
with(cutting.speed.long, interaction.plot(factorA, factorC, yield))
```



The effects of tool geometry and cutting angle are statistically significant. While cutting speed alone isn't statistically significant, its interaction with cutting angle is. Therefore cutting speed should remain in the model.

(b) Use the analysis of variance to confirm your conclusions for part (a).

```
# ANOVA test
cutting.speed.aov=aov(yield ~ factorA * factorB * factorC, cutting.speed.long)
summary(cutting.speed.aov)
```

```
##                         Df Sum Sq Mean Sq F value   Pr(>F)    
## factorA                  1    0.7     0.7   0.022 0.883680    
## factorB                  1  770.7   770.7  25.547 0.000117 ***
## factorC                  1  280.2   280.2   9.287 0.007679 ** 
## factorA:factorB          1   16.7    16.7   0.552 0.468078    
## factorA:factorC          1  468.2   468.2  15.519 0.001172 ** 
## factorB:factorC          1   48.2    48.2   1.597 0.224475    
## factorA:factorB:factorC  1   28.2    28.2   0.934 0.348282    
## Residuals               16  482.7    30.2                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
mse=summary(cutting.speed.aov)[[1]][8,3]
mse
```

```
## [1] 30.16667
```

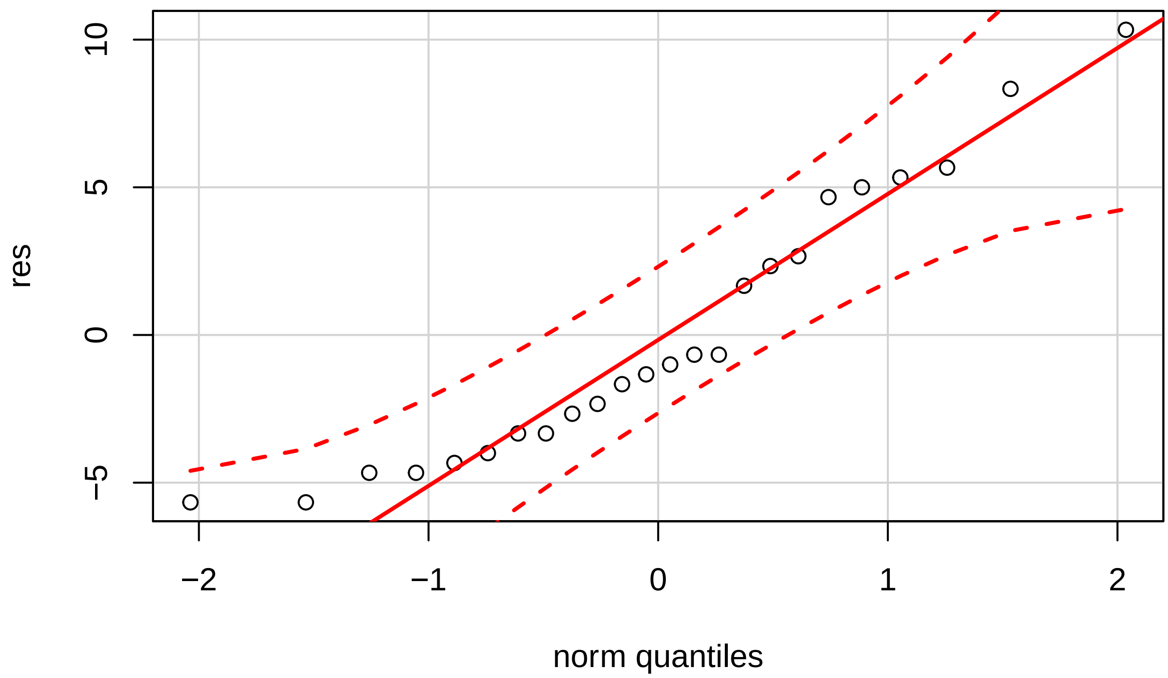
The variables that we find statistically significant also match our results from estimating factor effects in part a.

(c) Write down a regression model for predicting tool life (in hours) based on the results of this experiment.

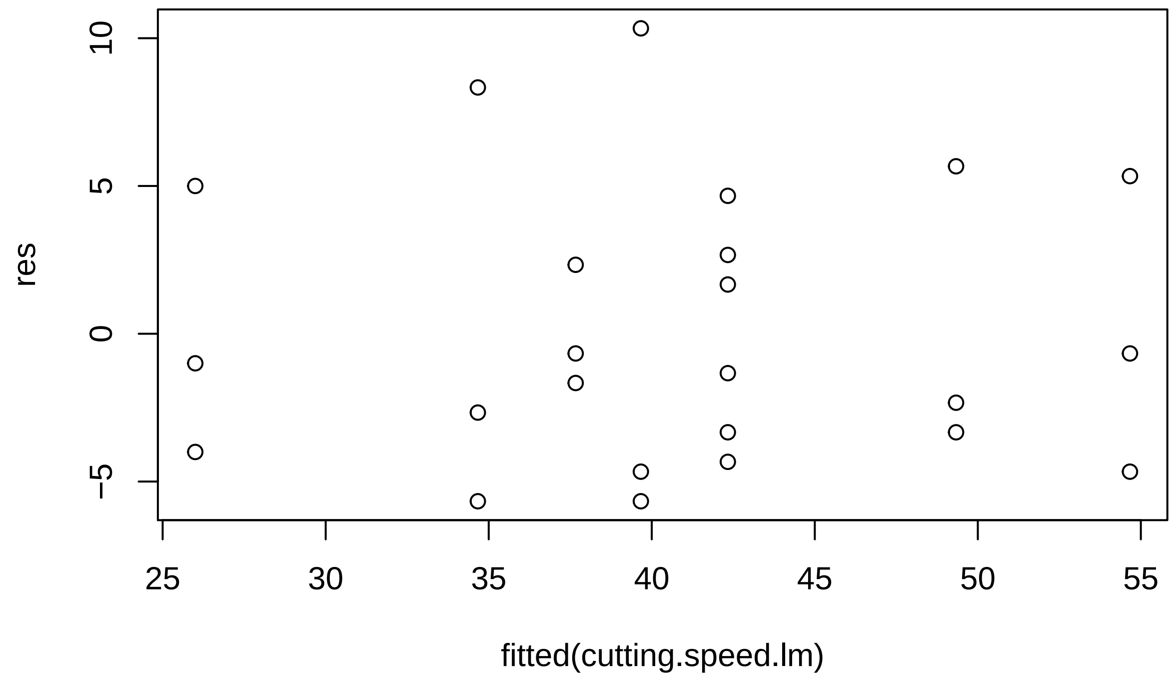
$$y = \beta_0 + \beta_1 X_1 + \beta_2 + X_2 + \beta_3 + X_3 + \beta_{1,3} X_1 X_3$$

(d) Analyze the residuals. Are there any obvious problems?

```
res=cutting.speed.long$yield-fitted(cutting.speed.lm)
qqPlot(res)
```




```
plot(fitted(cutting.speed.lm), res)
```



We take a look at our normality plot and can state that normality is good. We take a look at our residual plot and see no patterns, our model is good.

(e) On the basis of an analysis of main effect and interaction plots, what coded factor levels of A, B, and C would you recommend using?

Because the coefficient for factorB is positive, cutting angle should be high. In addition, the interaction plot reveals that lower cutting speed and higher life of a machine tool also produce a higher yield.

6.3

Find the standard error of the factor effects and approximate 95 percent confidence limits for the factor effects in Problem 6.1. Do the results of this analysis agree with the conclusions from the analysis of variance?

```r
#checking Standard error=sqrt(mse/N)
n=3;a=b=c=2;N=a*b*c*n
alpha=0.05
sqrt(mse/N)
```

```
## [1] 1.121135
```

```r
#consturct CI for regression coefficient (example, for coded(A))
se=sqrt(mse/N)
df=a*b*c*(n-1)
hat.beta1=cutting.speed.lm$coefficients[2]
CI.beta=hat.beta1+c(-1,1)*qt(alpha/2,df,lower.tail = F)*se
```

```r
CI.beta
```

```
## [1] -2.210034  2.543367
```

```r
2*CI.beta #CI for main effect A
```

```
## [1] -4.420068  5.086735
```

Standard Error is 1.12 and the confidence interval for the factor effects are (-4.42, 5.0867).

6.5

A router is used to cut locating notches on a printed circuit board. The vibration level at the surface of the board as it is cut is considered to be a major source of dimensional variation in the notches. Two factors are thought to influence vibration: bit size (A) and cutting speed (B). Two bit sizes (1/16 and 1/8 in.) and two speeds (40 and 90 rpm) are selected, and four boards are cut at each set of conditions shown below. The response variable is vibration measured as the resultant vector of three accelerometers (x, y, and z) on each test circuit board.

(a) Analyze the data from this experiment.

```r
# creating data table
A <- rep(c("-","+","-","+"), times = 4)
B <- rep(c("-","-","+","+"), times = 4)
Rep <- rep(c("I","II","III","IV"), each = 4)
Vibes <- c(18.2, 27.2, 15.9, 41.0, 18.9, 24.0, 14.5, 43.9, 12.9, 22.4, 15.1, 36.3, 14.4, 22.5, 14.2, 39
router.long <- data.frame(A, B, Rep, Vibes)

# defining coded
coded=function(x) #a function to code variable x
{
  ifelse(x=="+", 1, -1)
}

# linear regression
router.lm=lm(Vibes ~ coded(A) * coded(B), router.long)
summary(router.lm)
```

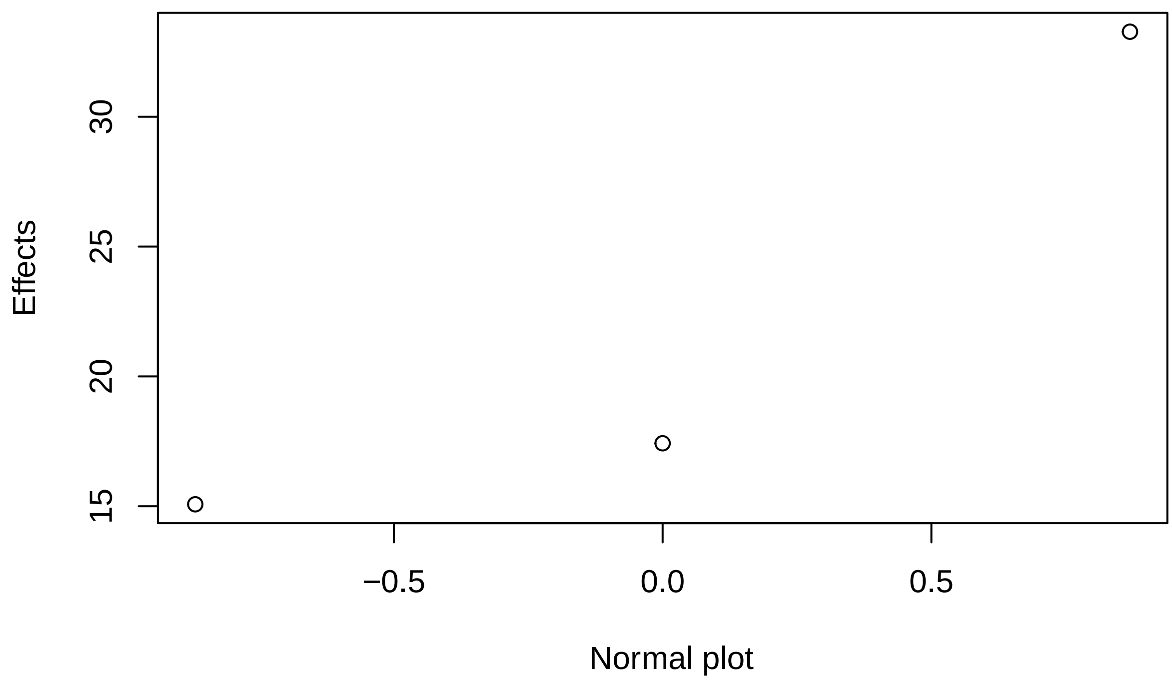
```
##
## Call:
## lm(formula = Vibes ~ coded(A) * coded(B), data = router.long)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.975 -1.550 -0.200  1.256  3.625
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        23.8312     0.6112  38.991 5.22e-14 ***
## coded(A)            8.3187     0.6112  13.611 1.17e-08 ***
## coded(B)            3.7687     0.6112   6.166 4.83e-05 ***
## coded(A):coded(B)   4.3562     0.6112   7.127 1.20e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.445 on 12 degrees of freedom
```

```
## Multiple R-squared:  0.9581, Adjusted R-squared:  0.9476
## F-statistic: 91.36 on 3 and 12 DF,  p-value: 1.569e-08
```

Our linear regression reveals that both treatments are statistically significant, with both variables positively correlated with vibration levels. In addition, there is evidence of interaction between the two.

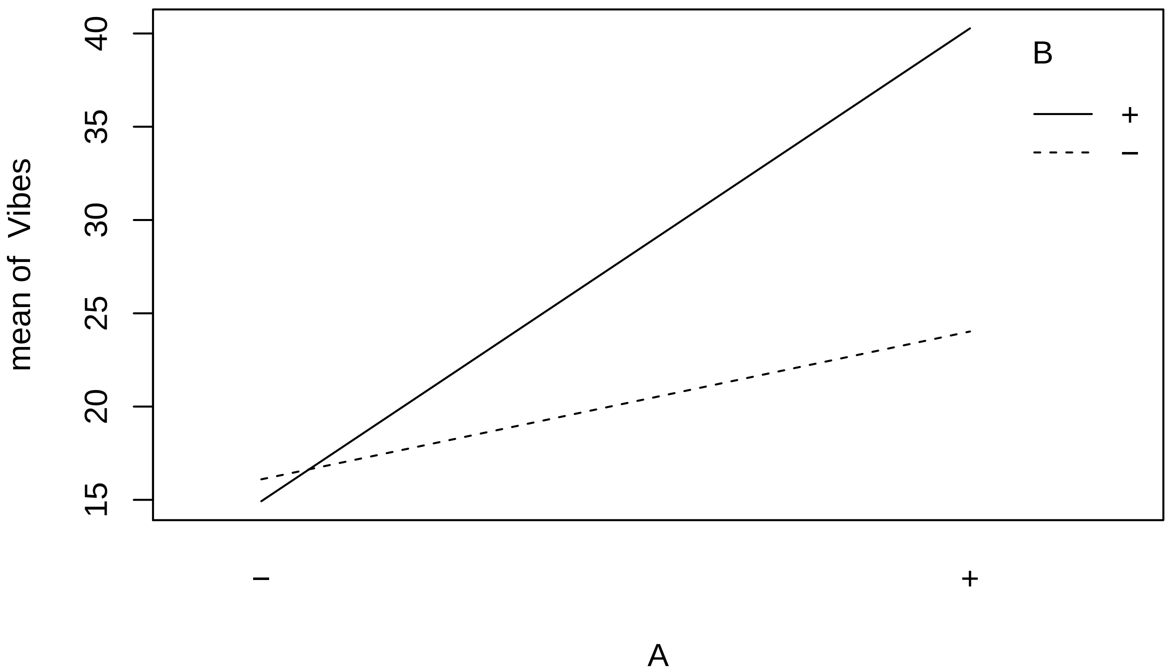
(b) Construct a normal probability plot of the residuals, and plot the residuals versus the predicted vibration level. Interpret these plots.

```r
router.aov = aov(Vibes ~ coded(A) * coded(B), router.long)
qqnorm(router.aov, full=T)
```



(c) Draw the AB interaction plot. Interpret this plot. What levels of bit size and speed would you recommend for routine operation?

```r
# interaction plot
with(router.long, interaction.plot(A, B, Vibes))
```



This plot reaffirms the notion that there is an interaction effect present between both variables. We'd recommend a $\frac{1}{16}$ in. bit size and 40rpm speed to minimize vibrations in this operation.

6.6

Reconsider the experiment described in Problem 6.1. Suppose that the experimenter only performed the eight trials from replicate I. In addition, he ran four center points and obtained the following response values: 36, 40, 43, 45.

(a) Estimate the factor effects. Which effects are large?

(b) Perform an analysis of variance, including a check for pure quadratic curvature. What are your conclusions?

(c) Write down an appropriate model for predicting tool life, based on the results of this experiment. Does this model differ in any substantial way from the model in Problem 6.1, part (c)?

(d) Analyze the residuals.

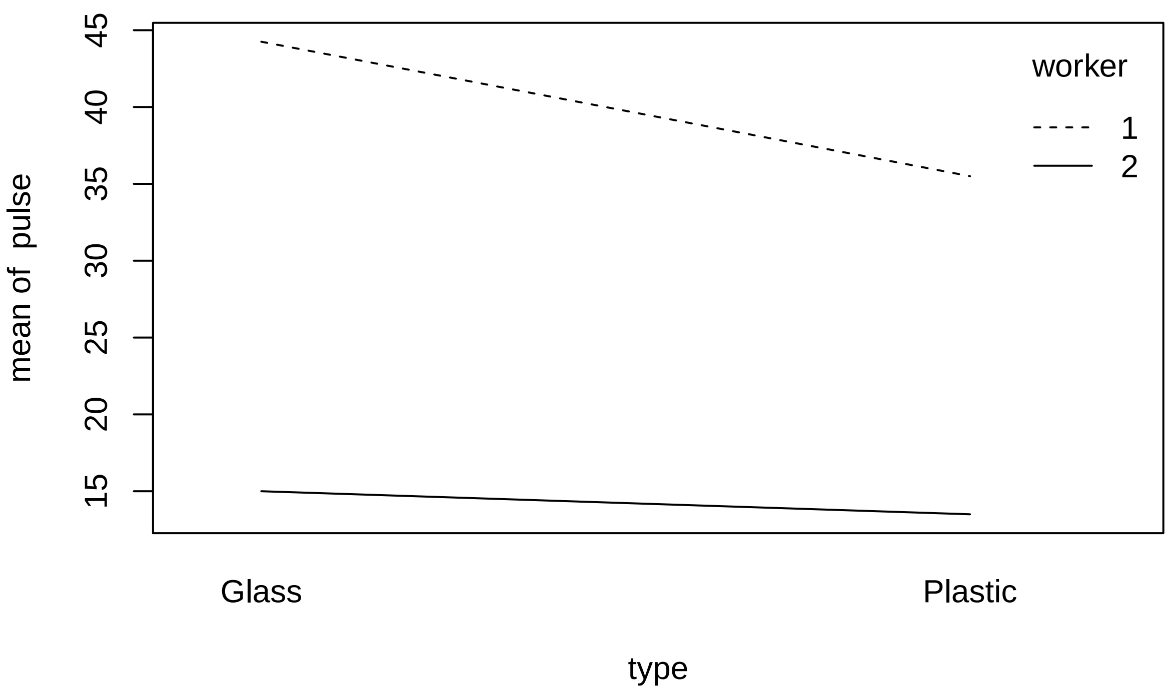
(e) What conclusions would you draw about the appropriate operating conditions for this process?

6.10

In Problem 6.9, the engineer was also interested in potential fatigue differences resulting from the two types of bottles. As a measure of the amount of effort required, he measured the

elevation of the heart rate (pulse) induced by the task. The results follow. Analyze the data and draw conclusions. Analyze the residuals and comment on the model's adequacy.

```r
# creating data table
type <- rep(c("Glass", "Plastic"), each = 4)
worker <- rep(c("1", "2"), each = 8)
pulse <- c(39, 45, 58, 35, 44, 35, 42, 21, 20, 13, 16, 11, 13, 10, 16, 15)
bottle.long <- data.frame(type, worker, pulse)

# interaction plot
with(bottle.long, interaction.plot(type, worker, pulse))
```




```r
# defining coded
coded=function(x) #a function to code variable x
{
  ifelse(x=="Glass" | x=="1", 1, -1)
}

# linear regression
bottle.lm=lm(pulse ~ coded(type) * coded(worker), bottle.long)
summary(bottle.lm)
```

```
##
## Call:
## lm(formula = pulse ~ coded(type) * coded(worker), data = bottle.long)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max 
## -14.500  -3.625   0.125   3.125  13.750 
## 
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                   27.062      1.902  14.227 7.11e-09 ***
## coded(type)                    2.563      1.902   1.347    0.203    
## coded(worker)                 12.812      1.902   6.736 2.09e-05 ***
## coded(type):coded(worker)      1.812      1.902   0.953    0.359    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 7.609 on 12 degrees of freedom
## Multiple R-squared:  0.8003, Adjusted R-squared:  0.7504 
## F-statistic: 16.03 on 3 and 12 DF,  p-value: 0.0001693
```

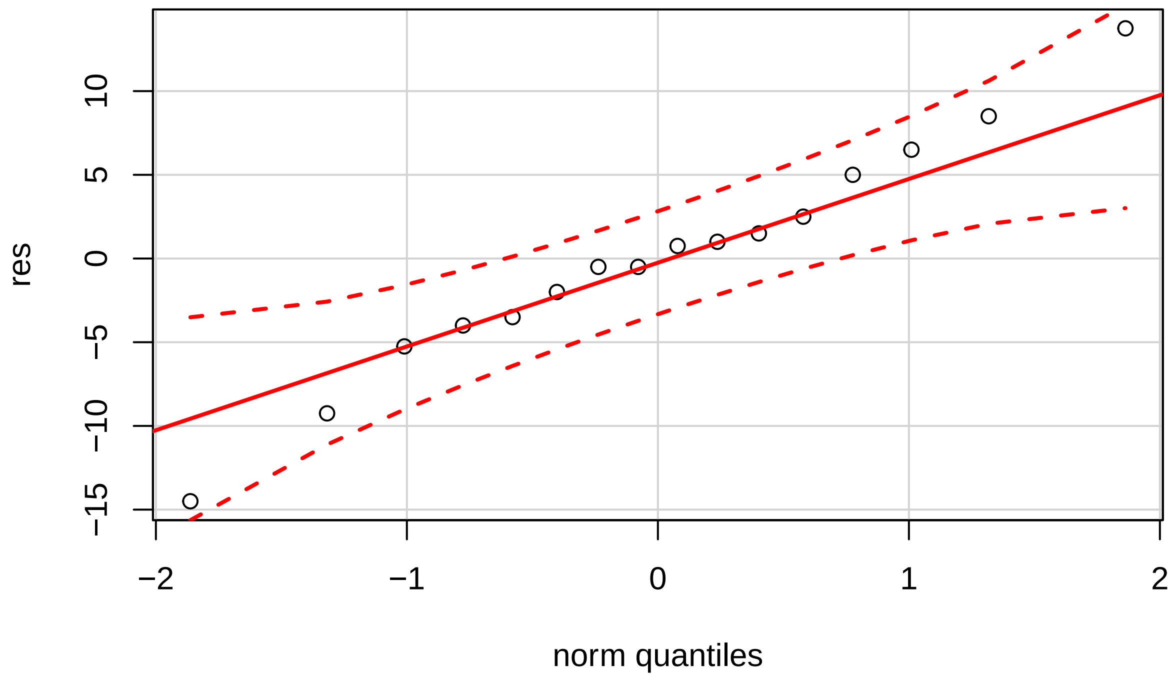
```r
# ANOVA test
bottle.aov=aov(pulse ~ type * worker, bottle.long)
summary(bottle.aov)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## type         1  105.1   105.1   1.815    0.203    
## worker       1 2626.6  2626.6  45.367 2.09e-05 ***
## type:worker  1   52.6    52.6   0.908    0.359    
## Residuals   12  694.7    57.9                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

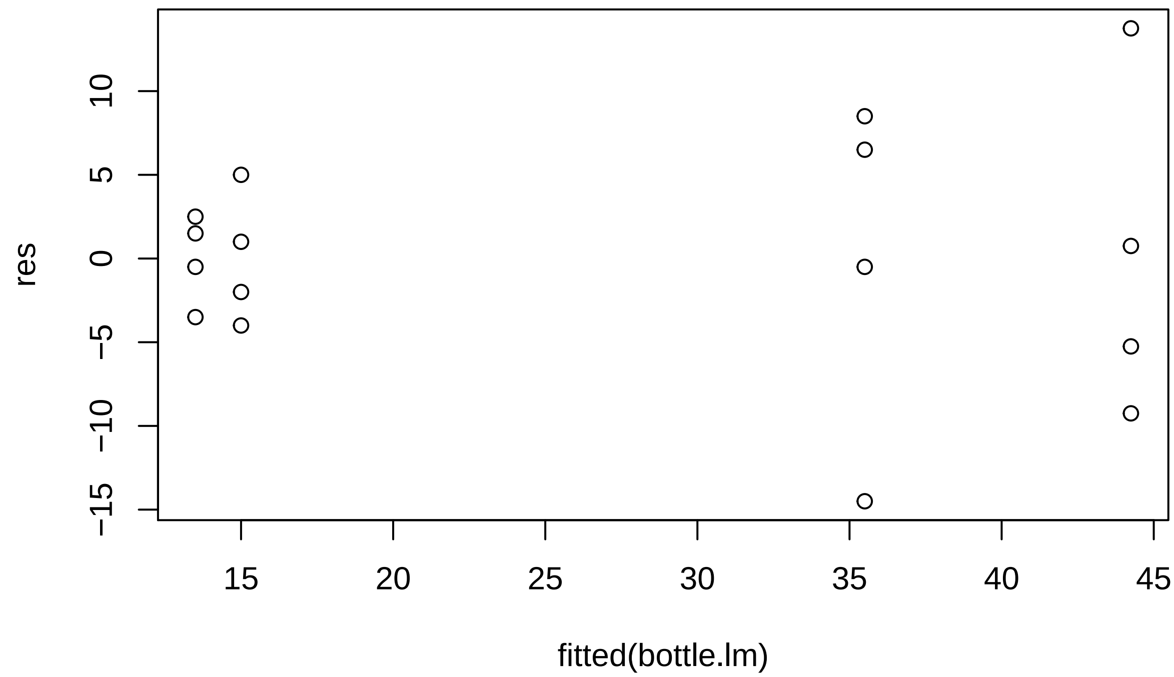
```r
mse=summary(bottle.aov)[[1]][4,3]
mse
```

```
## [1] 57.89583
```

```r
# checking model adequacy
res=bottle.long$pulse-fitted(bottle.lm)
qqPlot(res)
```




```
plot(fitted(bottle.lm), res)
```



When estimating factor effects, we don't suspect any interaction effects between bottle type and worker. Both our linear model and ANOVA test suggest that worker is statistically significant in predicting heart rate when performing the task, and reaffirm the conclusion we made in our interaction plot. When checking model adequacy, our qqplot reveals that our data is normally distributed, while the variances of residuals in our residual plot don't seem homogenous. As a result, the model may not be adequate.

6.11

Calculate approximate 95 percent confidence limits for the factor effects in Problem 6.10. Do the results of this analysis agree with the analysis of variance performed in Problem 6.10?

```
#checking Standard error=sqrt(mse/N)
n=4;a=b=2;N=a*b*n
alpha=0.05
sqrt(mse/N)
```

```
## [1] 1.902233
```

```
#consturct CI for regression coefficient (example, for coded(A))
se=sqrt(mse/N)
df=a*b*(n-1)

hat.beta1=bottle.lm$coefficients[2]
CI.beta=hat.beta1+c(-1,1)*qt(alpha/2,df,lower.tail = F)*se
CI.beta
```

```
## [1] -1.582109  6.707109
```

```
2*CI.beta #CI for main effect A
```

```
## [1] -3.164218 13.414218
```

```
hat.beta1=bottle.lm$coefficients[3]
CI.beta=hat.beta1+c(-1,1)*qt(alpha/2,df,lower.tail = F)*se
CI.beta
```

```
## [1]  8.667891 16.957109
```

```
2*CI.beta #CI for main effect B
```

```
## [1] 17.33578 33.91422
```

```
hat.beta1=bottle.lm$coefficients[4]
CI.beta=hat.beta1+c(-1,1)*qt(alpha/2,df,lower.tail = F)*se
CI.beta
```

```
## [1] -2.332109  5.957109
```

```
2*CI.beta #CI for main effect AB
```

```
## [1] -4.664218 11.914218
```

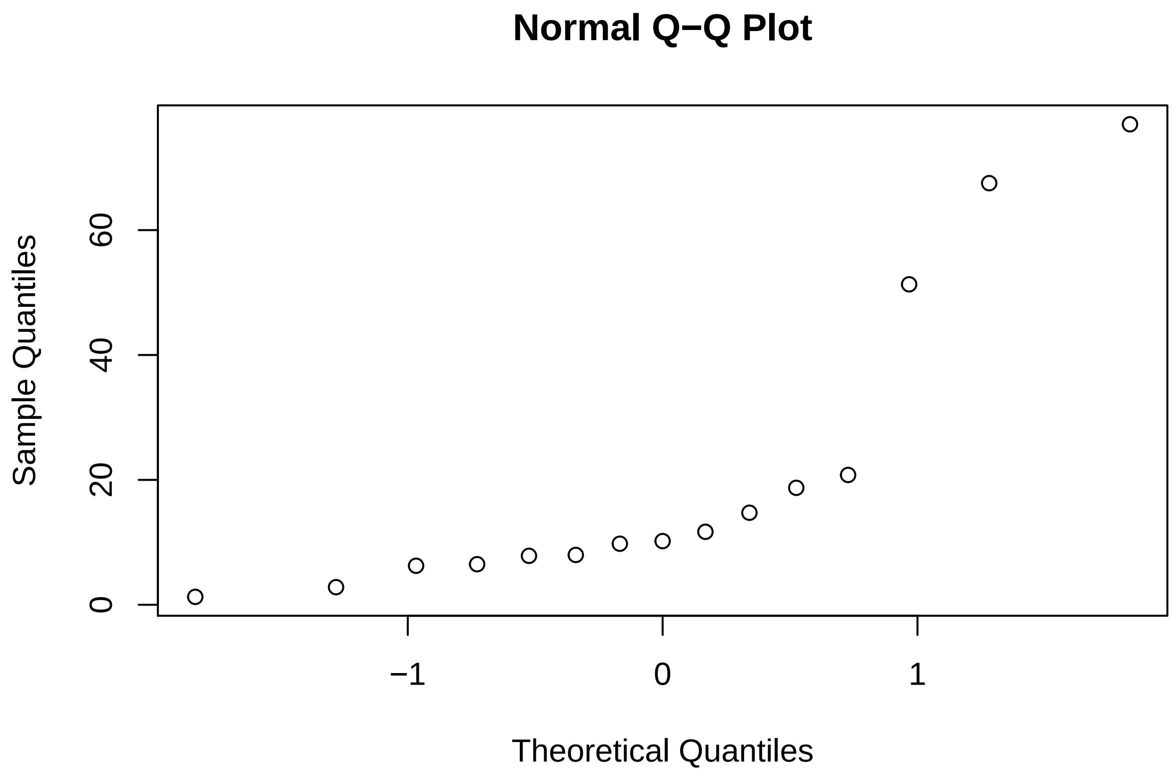
Confidence interval for main effect A is (-3.16, 13.4), for main effect B is (17.33, 33.91), and for interaction effect AB is (-4.66, 11.9). The only confidence interval the doesn't contain 0 is the one associated with 'worker'. The results agree with those receive from the ANOVA test we ran earlier.

6.17

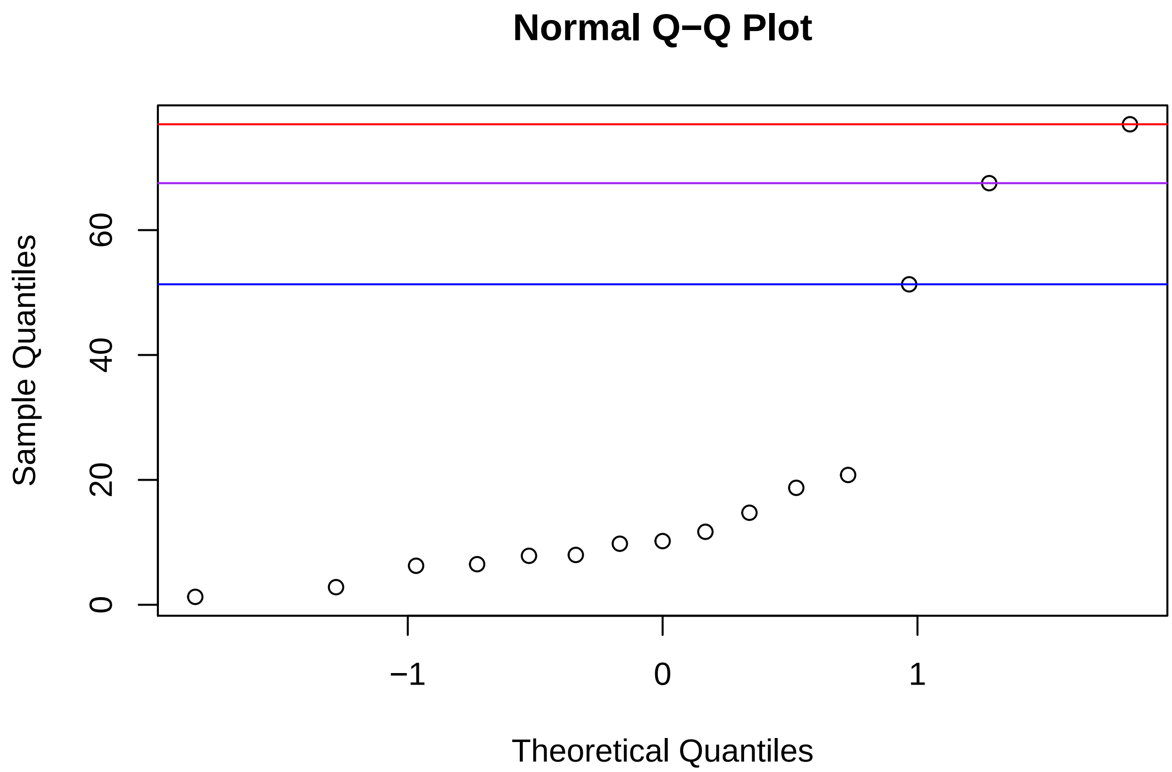
An experimenter has run a single replicate of a 2^4 design. The following effect estimates have been calculated:

(a) Construct a normal probability plot of these effects.

```
letter = c("A","B","C","D","AB","AC","AD","BC","BD","CD","ABC","ABD","ACD","BCD","ABCD")
number = c(76.95,-67.52,-7.84,-18.73,-51.32,11.69,9.78,20.79,14.74,1.27,-2.82,-6.50,10.20,-7.98,-6.25)
experiment = data.frame(letter, number)

experiment.aov = aov(number ~ coded(letter), experiment)


plot = qqnorm(abs(number))
```




```r
## click at the "outlier" points and then click "Finish" button


lister = sort(abs(number))

qqnorm(lister)
abline(h = 76.95, col = "red")
abline(h = 67.52, col = "purple")
abline(h = 51.32, col = "blue")
```




```
#factor A, B, and AB
```

(b) Identify a tentative model, based on the plot of the effects in part (a).

$$\hat{y} = \beta_0 + 75.95x_a + 67.52x_b + 51.32x_{ab}$$

6.22

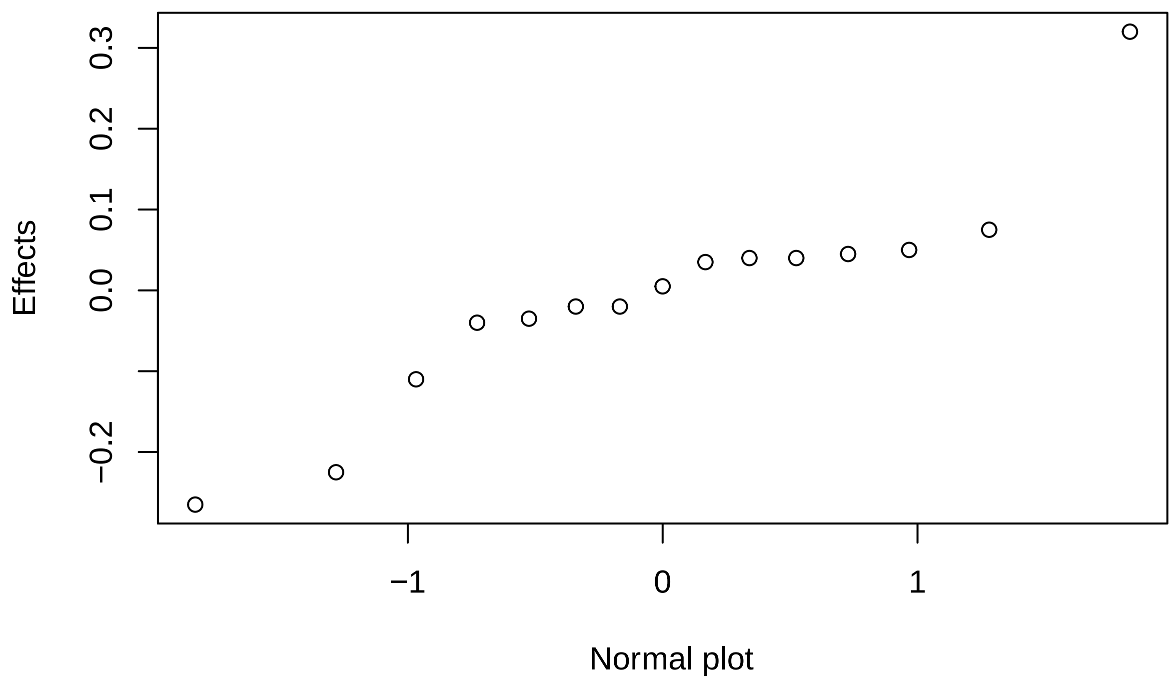
Semiconductor manufacturing processes have long and complex assembly flows, so matrix marks and automated 2d-matrix readers are used at several process steps throughout factories. Unreadable matrix marks negatively affect factory run rates because manual entry of part data is required before manufacturing can resume. A 2^4 factorial experiment was conducted to develop a 2d-matrix laser mark on a metal cover that protects a substrate-mounted die. The design factors are A = laser power (9 and 13 W), B = laser pulse frequency (4000 and 12,000 Hz), C = matrix cell size (0.07 and 0.12 in.), and D = writing speed (10 and 20 in./sec), and the response variable is the unused error correction (UEC). This is a measure of the unused portion of the redundant information embedded in the 2d-matrix. A UEC of 0 represents the lowest reading that still results in a decodable matrix, while a value of 1 is the highest reading. A DMX Verifier was used to measure UEC. The data from this experiment are shown in Table P6.5.

```
Standard.Order = c(8,10,12,9,7,15,2,6,16,13,5,14,1,3,4,11)
Run.Order = c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16)
A = Laser.Power = c(1,1,1,-1,-1,-1,1,1,1,-1,-1,1,-1,-1,1,-1)
B = Pulse.Freq = c(1,-1,1,-1,1,1,-1,-1,1,-1,-1,-1,-1,1,1,1)
C = Cell.Size = c(1,-1,-1,-1,1,1,-1,1,1,1,1,1,-1,-1,-1,-1)
```

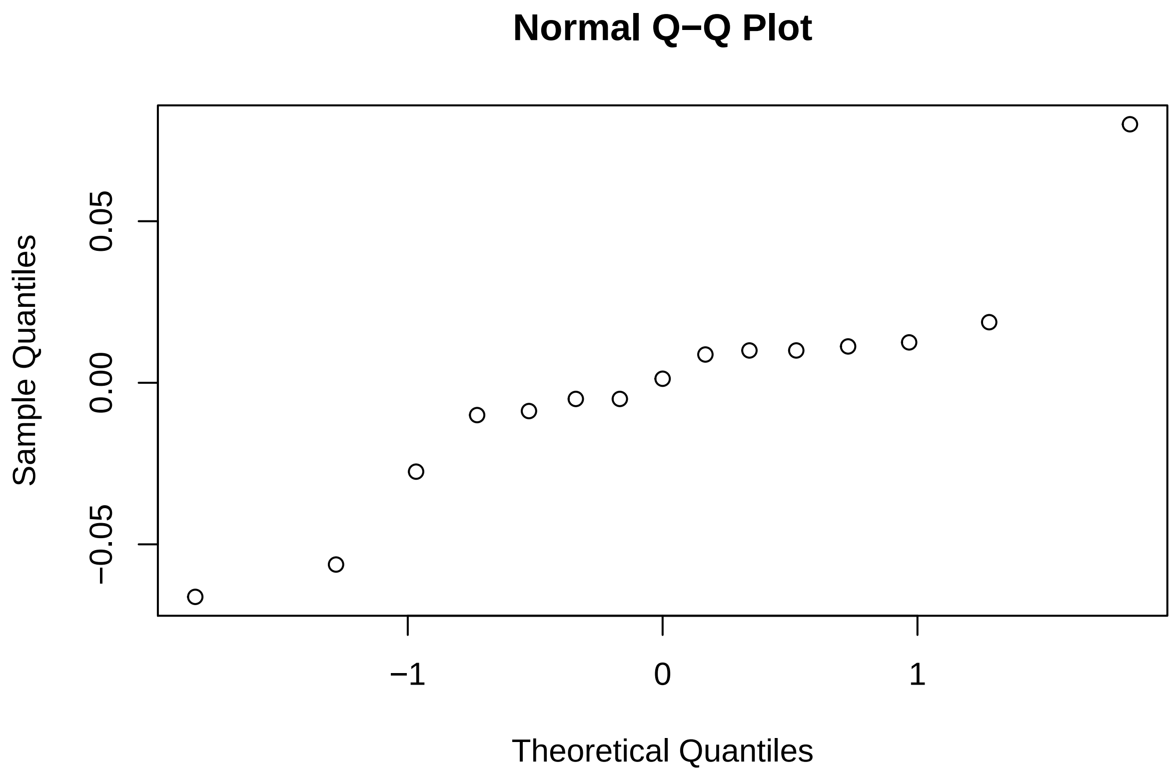
```
D = Writing.Speed = c(-1,1,1,1,-1,1,-1,-1,1,1,-1,1,-1,-1,-1,1)
UEC = c(0.8,0.81,0.79,0.6,0.65,0.55,0.98,0.67,0.69,0.56,0.63,0.65,0.75,0.72,0.98,0.63)
error = data.frame(Standard.Order,Run.Order,A,B,C,D,UEC)
```

(a) Analyze the data from this experiment. Which factors significantly affect UEC?

```
error.lm = lm(UEC ~A*B*C*D, error)
qqnorm(aov(UEC ~ A * B * C * D, error), label=T, full=T)
```




```
coef=error.lm$coefficients[-1]
variables=names(coef)
plot=qqnorm(coef)
variables[identify(plot)]
```




```
## character(0)
```

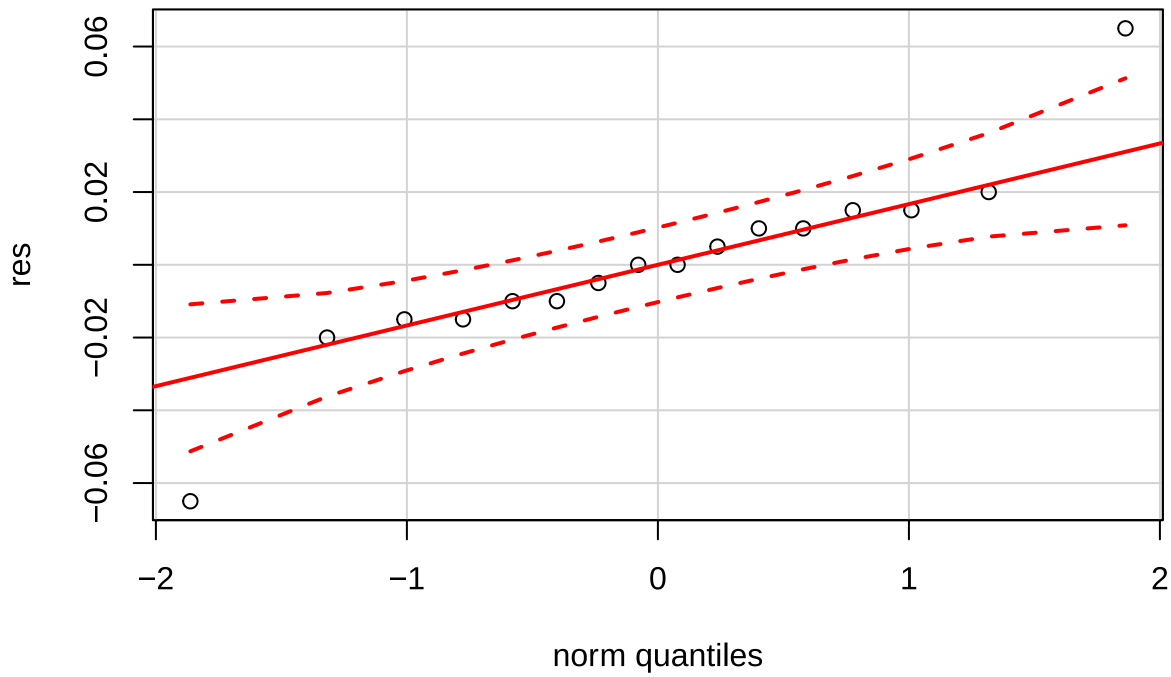
```
##new model
error.aov = aov(UEC ~ A+C+D, error)
summary(error.aov)
```

```
##             Df  Sum Sq Mean Sq F value   Pr(>F)    
## A            1 0.10240 0.10240   40.52 3.58e-05 ***
## C            1 0.07022 0.07022   27.79 0.000197 ***
## D            1 0.05063 0.05063   20.03 0.000758 ***
## Residuals   12 0.03033 0.00253                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

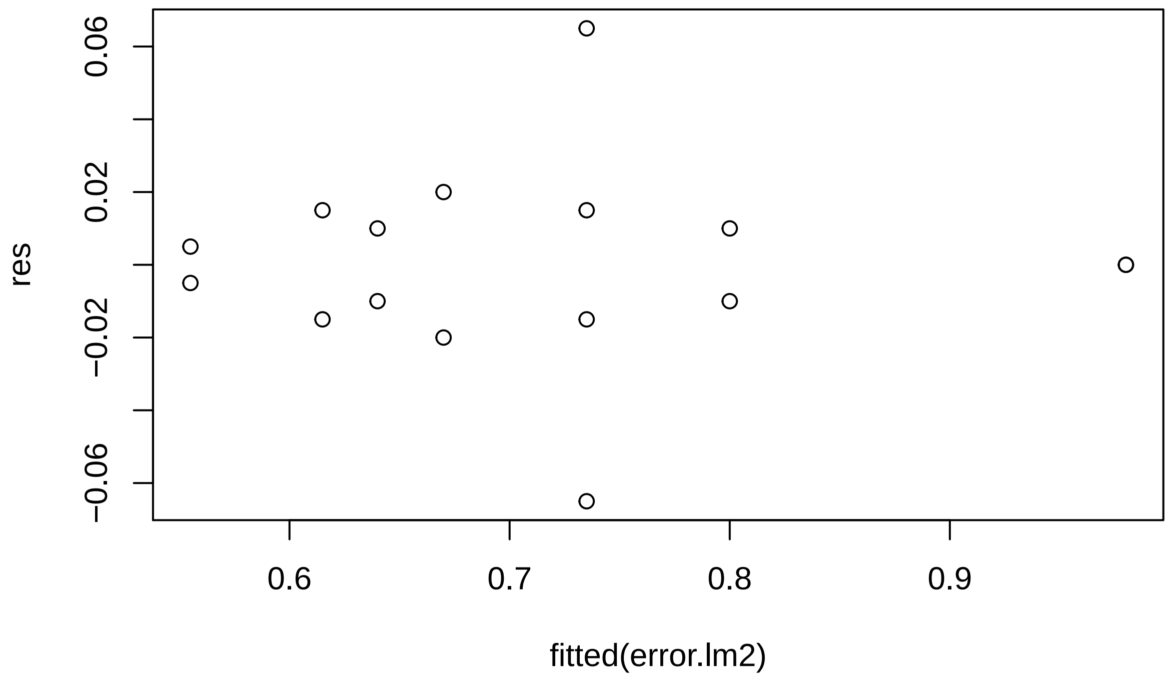
With the help of the normal probability plot we are able to identify that "A", "C", "D" are important. After choosing my factors, I plug them into a ANOVA and discover that all factors are significantly different.

(b) Analyze the residuals from this experiment. Are there any indications of model inadequacy?

```
error.lm2 = lm(UEC~ A*C*D, error)
res=error$UEC-fitted(error.lm2)
library(car)
qqPlot(res)
```




```
plot(fitted(error.lm2), res)
```



After studying the normality and residuals we can state that normality is good and residuals are patternless and random. We may conclude that model is good.

6.23

Reconsider the experiment described in Problem 6.20. Suppose that four center points are available and that the UEC response at these four runs is 0.98, 0.95, 0.93, and 0.96, respectively. Reanalyze the experiment incorporating a test for curvature into the analysis. What conclusions can you draw? What recommendations would you make to the experimenters?

6.25

Consider the single replicate of the 2^4 design in Example 6.2. Suppose that we had arbitrarily decided to analyze the data assuming that all three- and four-factor interactions were negligible. Conduct this analysis and compare your results with those obtained in the example. Do you think that it is a good idea to arbitrarily assume interactions to be negligible even if they are relatively high-order ones?

6.36

Resistivity on a silicon wafer is influenced by several factors. The results of a 2^4 factorial experiment performed during a critical processing step is shown in Table P6.10.

6.37

Continuation of Problem 6.36. Suppose that the experimenter had also run four center points along with the 16 runs in Problem 6.36. The resistivity measurements at the center points are 8.15, 7.63, 8.95, and 6.48. Analyze the experiment again incorporating the center points. What conclusions can you draw now?