Stat 5309 Semester Project

Tom Wilson

May 10th, 2019

1

An article in the AT&T Technical Journal (March/April 1986, Vol. 65, pp. 39-50) describes the application of two-level factorial designs to integrated circuit manufacturing. A basic processing step is to grow an epitaxial layer on polished silicon wafers. The wafers mounted on a susceptor are positioned inside a bell jar, and chemical vapors are introduced. The susceptor is rotated and heat is applied until the epitaxial layer is thick enough. An experiment was run using two factors: arsenic flow rate (A) and deposition time (B). Four replicates were run, and the epitaxial layer thickness was measured in um. The data are shown below:

Replicate Factor Levels

```
flow_rates <- c(0.55,0.59)
depo_times <- c(10,15)
epitaxial_data <- expand.grid(flow_rate=flow_rates,
                              depo_time=rep(depo_times,4))
epitaxial_data <- cbind(epitaxial_data,
                        thickness=c(14.037,13.880,14.821,14.888,
                                    16.165,13.860,14.757,14.921,
                                    13.972,14.032,14.843,14.415,
                                    13.907,13.914,14.878,14.932))
```

a

Estimate the factor effects.

```
epitaxial_model <- aov(formula = thickness~flow_rate*depo_time,
                       data=epitaxial_data)
epitaxial_model$coefficients %>% kable()
```

	x
(Intercept)	37.62656
flow_rate	-43.11875
depo_time	-1.48735
flow_rate:depo_time	2.81500

b

Conduct an analysis of variance. Which factors are important?

```
summary(epitaxial_model)
```

```
##                     Df Sum Sq Mean Sq F value Pr(>F)
## flow_rate            1  0.403  0.4026   1.262 0.2833
## depo_time            1  1.374  1.3736   4.305 0.0602 .
## flow_rate:depo_time  1  0.317  0.3170   0.994 0.3386
```

```
## Residuals           12  3.828  0.3190
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

deposition time is significant while flow rate and the interaction are not siginifcant.

c

Write down a regression equation that could be used to predict epitaxial layer thickness over the region of arsenic flow rate and deposition time used in this experiment.

$thickness = 37.627 - 43.119 flow rate - 1.148deposition time $

Build a RSM model (2nd order, 1st order with interaction). Choose one which works.

```
thickness_rsm <- rsm( thickness ~ FO(depo_time,flow_rate) + TWI(depo_time,flow_rate) , data=epitaxial_da
summary(thickness_rsm)
```

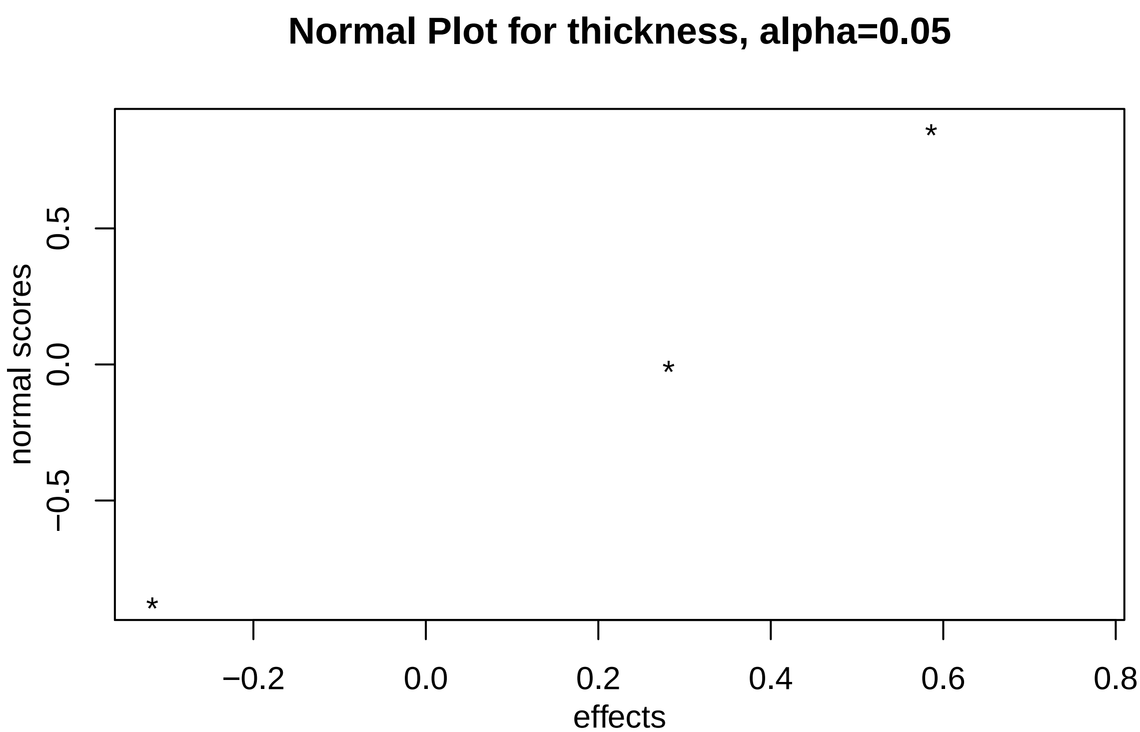
```
##
## Call:
## rsm(formula = thickness ~ FO(depo_time, flow_rate) + TWI(depo_time,
##     flow_rate), data = epitaxial_data)
##
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           37.6266    20.5334  1.8325   0.0918 .
## depo_time             -1.4874     1.6108 -0.9234   0.3740
## flow_rate            -43.1188    36.0014 -1.1977   0.2542
## depo_time:flow_rate    2.8150     2.8242  0.9967   0.3386
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Multiple R-squared:  0.3535, Adjusted R-squared:  0.1918
## F-statistic: 2.187 on 3 and 12 DF,  p-value: 0.1425
##
## Analysis of Variance Table
##
## Response: thickness
##                           Df Sum Sq Mean Sq F value Pr(>F)
## FO(depo_time, flow_rate)   2 1.7762       1  2.7836 0.1016
## TWI(depo_time, flow_rate)  1 0.3170       0  0.9935 0.3386
## Residuals                 12 3.8285       0
## Lack of fit                0 0.0000     Inf
## Pure error                12 3.8285       0
##
## Stationary point of response surface:
##  depo_time  flow_rate
## 15.3174956  0.5283659
##
## Eigenanalysis:
## eigen() decomposition
## $values
## [1]  1.4075 -1.4075
##
## $vectors
##                 [,1]       [,2]
```

```
## depo_time 0.7071068 -0.7071068
## flow_rate 0.7071068  0.7071068
```

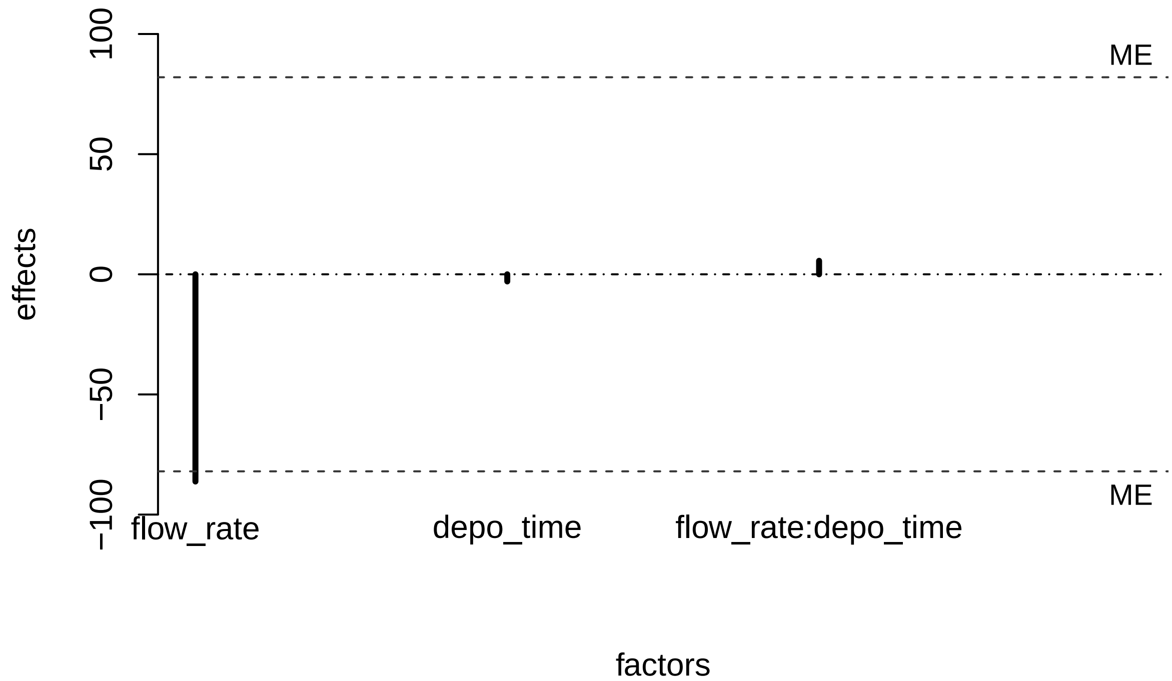
Perform Daniel plot and Lenth plot. What is the model 's R-square.

R-squared = 0.19

```r
DanielPlot(epitaxial_model)
```




```r
LenthPlot(epitaxial_model, alpha = 0.05, plt =TRUE, limits = TRUE)
```




```
##      alpha        PSE         ME        SME
##   0.050000   6.453525  81.999810 242.293942
```

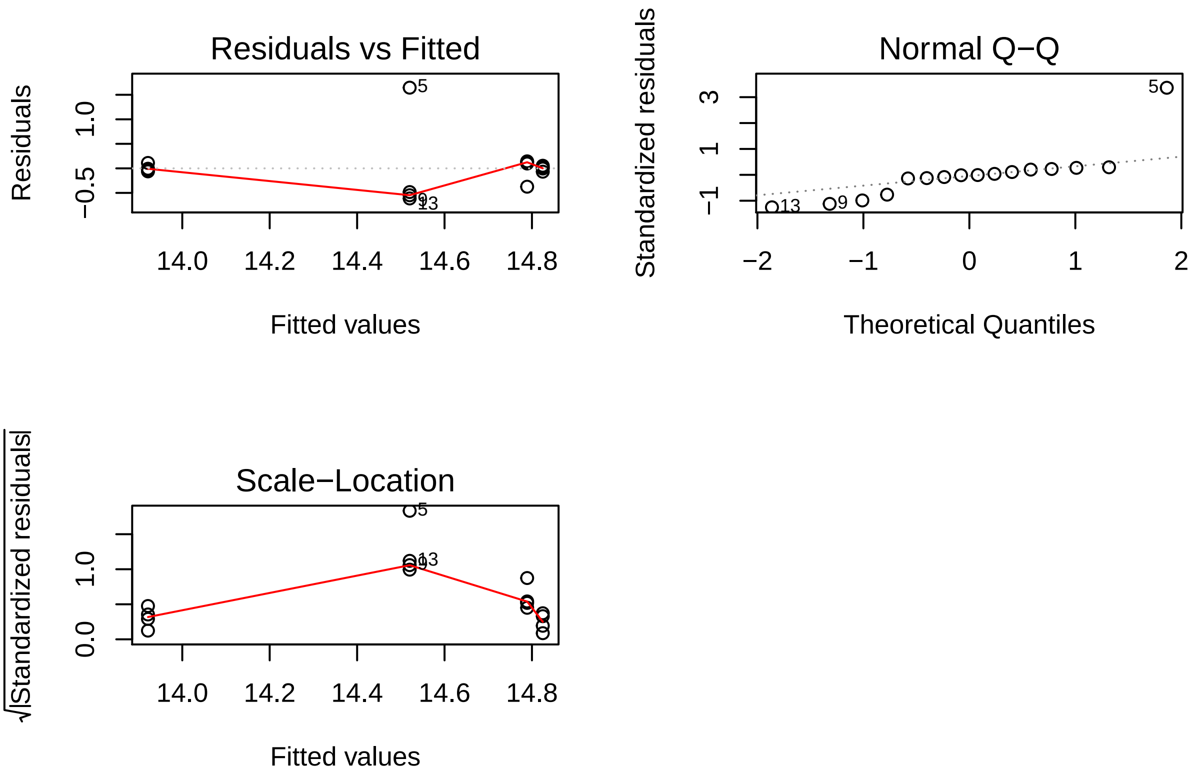
Flow rate has a large effect while depo time has a small effect.

d

Analyze the residuals. Are there any residuals that should cause concern?

```r
par(mfrow=c(2,2))
plot(epitaxial_model)
```

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

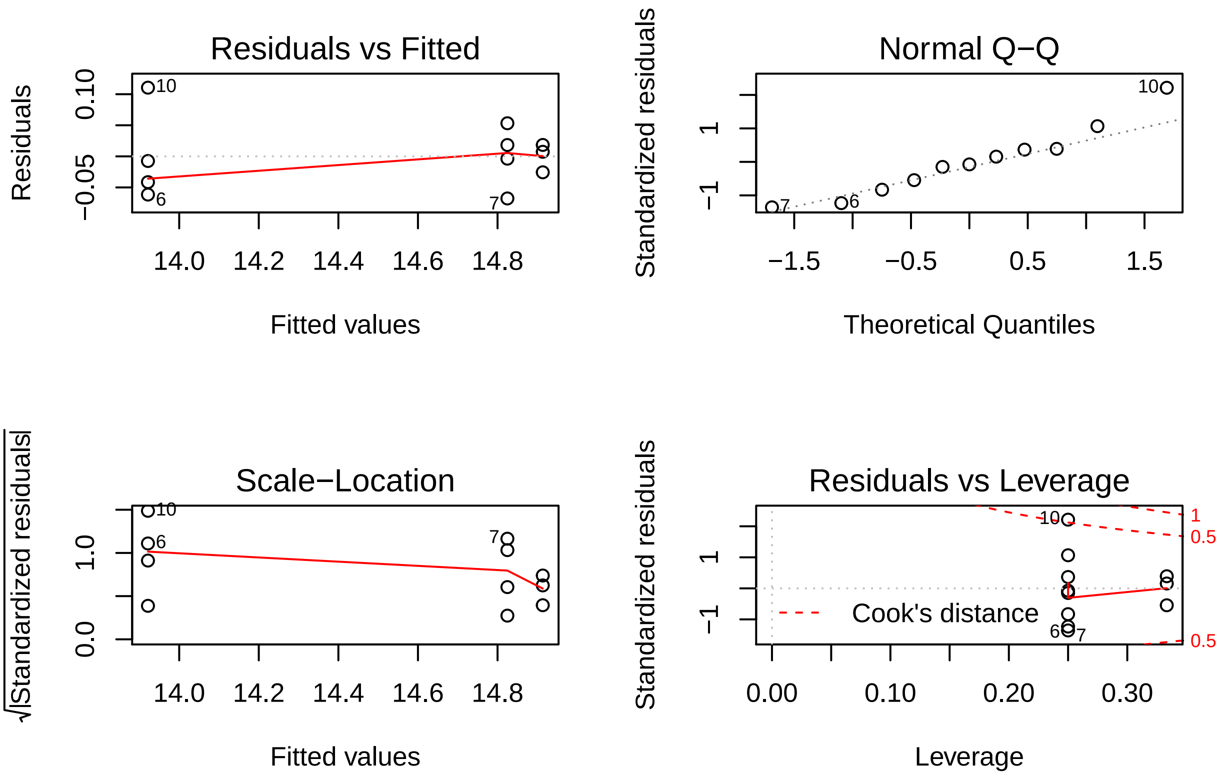


There is 1 very serious large outlier and a few less serious small outliers.

Find the cook distance. Take out the outlier(s). Rebuild the rsm model on new data.

```
is_not_outlier <- cooks.distance(epitaxial_model)<0.025
epitaxial_outlier_removed <- rsm(formula = thickness~FO(flow_rate,depo_time)+TWI(flow_rate,depo_time),
                                data=epitaxial_data[is_not_outlier,])
```

```
par(mfrow=c(2,2))
plot(epitaxial_outlier_removed)
```



e

Discuss how you might deal with the potential outlier found in part (d).

- Analyse the data with and without outliers removed.
- create an indication of outliers and analyze its effect on thickness.
- Cross validate

f

Perform a canonical analysis. Do a contour plot. Any optimal response?

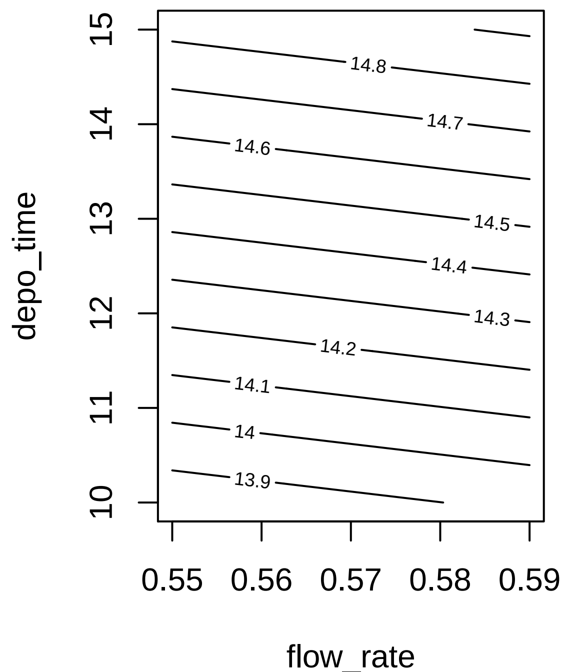
```r
canonical(thickness_rsm)
```

```
## $xs
##  depo_time  flow_rate
## 15.3174956  0.5283659
##
## $eigen
## eigen() decomposition
## $values
## [1]  1.4075 -1.4075
##
## $vectors
##                 [,1]       [,2]
```

```
## depo_time 0.7071068 -0.7071068
## flow_rate 0.7071068  0.7071068
```

```
par(mfrow=c(1,2))
contour(epitaxial_outlier_removed,~flow_rate+depo_time)
```

```
## Warning in predict.lm(lmobj, newdata = newdata): prediction from a rank-
## deficient fit may be misleading
```



There is a maximum thickness as depo time = 15.3174956 flow rate = 0.5283659 contours are planar, so the design space is not suffcient to find a global extremum.

2

A nickel-titanium alloy is used to make components for jet turbine aircraft engines. Cracking is a potentially serious problem in the final part, because it can lead to nonrecoverable failure. A test is run at the parts producer to determine the effect of four factors on cracks. The four factors are pouring temperature (A), titanium content (B), heat treatment method (C), and amount of grain refiner used (D). Two replicates of a 2ˆ4 design are run, and the length of crack in mm $\times 10^{-2}$? induced in a sample coupon subjected to a standard test is measured. The data are shown in the following table:

```
low_high <- c('-','+')
cracking_data <- expand.grid(temp=low_high,
                             content=low_high,
                             method=low_high,
                             refiner=rep(low_high,2))
cracking_data <- cbind(cracking_data,
```

```
                    length=c(7.037,14.707,11.635,17.273,
                             10.403,4.368,9.360,14.440,
                             8.561,16.867,13.876,19.824,
                             11.846,6.125,11.190,15.653,
                             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))
cracking_data %>% kable()
```

temp	content	method	refiner	length
-	-	-	-	7.037
+	-	-	-	14.707
-	+	-	-	11.635
+	+	-	-	17.273
-	-	+	-	10.403
+	-	+	-	4.368
-	+	+	-	9.360
+	+	+	-	14.440
-	-	-	+	8.561
+	-	-	+	16.867
-	+	-	+	13.876
+	+	-	+	19.824
-	-	+	+	11.846
+	-	+	+	6.125
-	+	+	+	11.190
+	+	+	+	15.653
-	-	-	-	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

a

Estimate the factor effects, Which factor effects appear to be large? Change the factors name (Temp,Content,Method,Refiner,Length).

```
cracking_model <- aov(formula=length~temp*content*method*refiner, data=cracking_data)
coef(cracking_model) %>% kable()
```

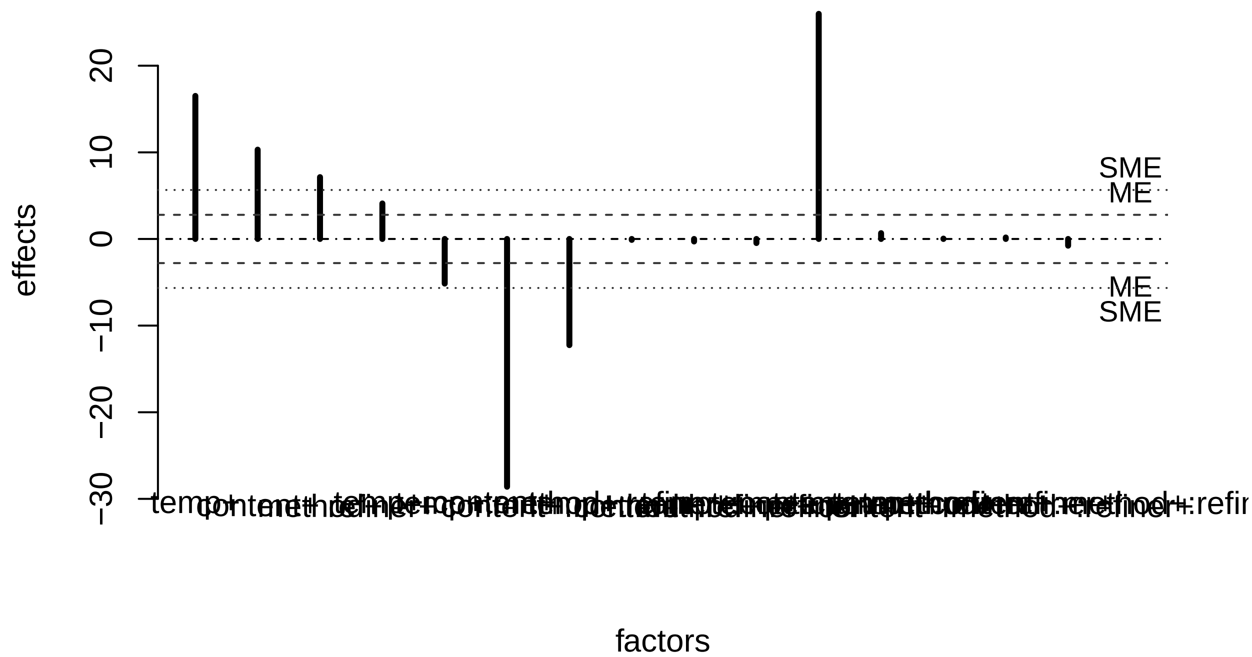
	x
(Intercept)	6.7065
temp+	8.2565
content+	5.1555
method+	3.5705
refiner+	2.0495
temp+:content+	-2.5745
temp+:method+	-14.3005
content+:method+	-6.1260
temp+:refiner+	-0.0530
content+:refiner+	-0.1445
method+:refiner+	-0.2350
temp+:content+:method+	12.9935
temp+:content+:refiner+	0.3355
temp+:method+:refiner+	0.0200
content+:method+:refiner+	0.0860
temp+:content+:method+:refiner+	-0.3870

```r
summary(cracking_model)
```

```
##                             Df Sum Sq Mean Sq F value   Pr(>F)    
## temp                         1  75.96   75.96 524.861 1.17e-13 ***
## content                      1 130.47  130.47 901.505 1.69e-15 ***
## method                       1  99.90   99.90 690.282 1.38e-14 ***
## refiner                      1  28.74   28.74 198.558 1.94e-10 ***
## temp:content                 1  31.89   31.89 220.368 8.93e-11 ***
## temp:method                  1 124.52  124.52 860.407 2.44e-15 ***
## content:method               1   0.20    0.20   1.389    0.256    
## temp:refiner                 1   0.00    0.00   0.011    0.918    
## content:refiner              1   0.00    0.00   0.013    0.911    
## method:refiner               1   0.16    0.16   1.074    0.315    
## temp:content:method          1  81.92   81.92 566.050 6.48e-14 ***
## temp:content:refiner         1   0.01    0.01   0.070    0.795    
## temp:method:refiner          1   0.02    0.02   0.104    0.751    
## content:method:refiner       1   0.01    0.01   0.040    0.844    
## temp:content:method:refiner  1   0.02    0.02   0.129    0.724    
## Residuals                   16   2.32    0.14                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There are significant interactions between pouring temperature, titanium content, and heat treatment method. temp, content, method, refiner, temp:content, temp:method are also significant.

```r
LenthPlot(cracking_model, alpha = 0.05, plt =TRUE)
```




```
##    alpha      PSE       ME      SME
## 0.050000 1.083750 2.785868 5.655713
```

These 6 coefficients are large: content+:method+ -6.126 temp+:method+ -14.3005 temp+:content+:method+ 12.9935 temp+ 8.2565 content+ 5.1555 method+ 3.5705

b

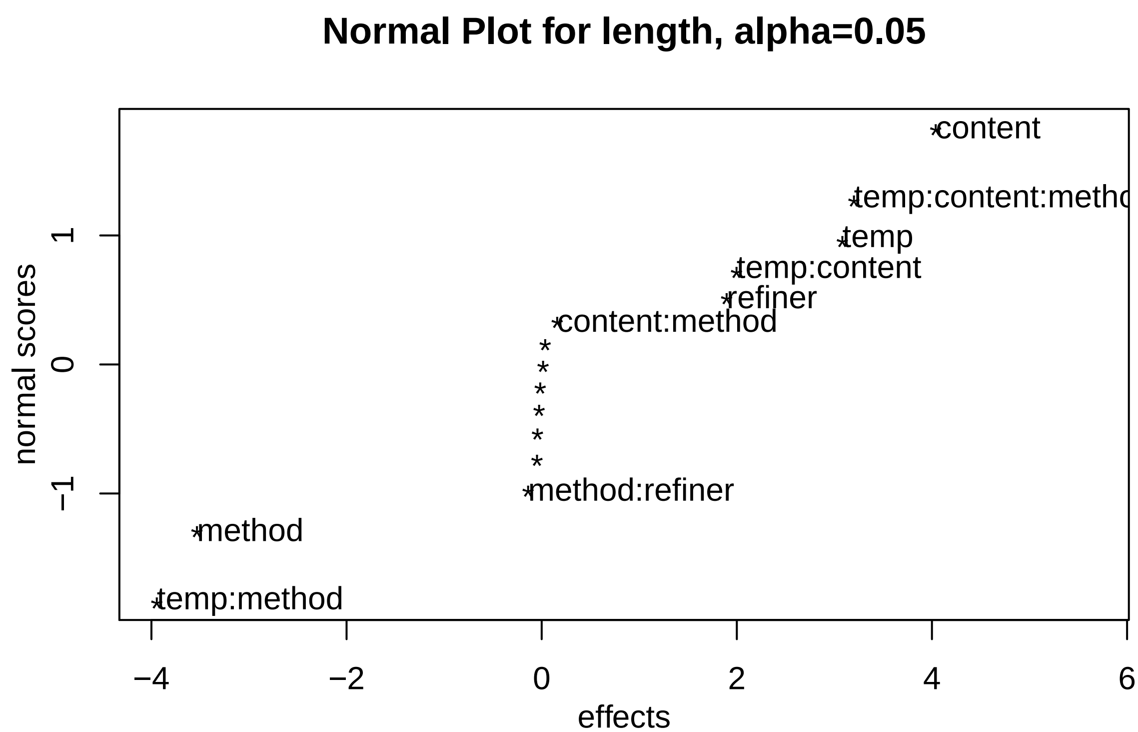
Conduct an analysis of variance. Do any of the factors affect cracking? Use $\alpha = 0.05$. Perform effects Daniel plot and Lenth plot.

```r
summary(cracking_model)
```

```
##                              Df Sum Sq Mean Sq F value   Pr(>F)    
## temp                          1  75.96   75.96 524.861 1.17e-13 ***
## content                       1 130.47  130.47 901.505 1.69e-15 ***
## method                        1  99.90   99.90 690.282 1.38e-14 ***
## refiner                       1  28.74   28.74 198.558 1.94e-10 ***
## temp:content                  1  31.89   31.89 220.368 8.93e-11 ***
## temp:method                   1 124.52  124.52 860.407 2.44e-15 ***
## content:method                1   0.20    0.20   1.389    0.256    
## temp:refiner                  1   0.00    0.00   0.011    0.918    
## content:refiner               1   0.00    0.00   0.013    0.911    
## method:refiner                1   0.16    0.16   1.074    0.315    
## temp:content:method           1  81.92   81.92 566.050 6.48e-14 ***
## temp:content:refiner          1   0.01    0.01   0.070    0.795    
## temp:method:refiner           1   0.02    0.02   0.104    0.751    
```

```
## content:method:refiner       1   0.01    0.01   0.040    0.844
## temp:content:method:refiner  1   0.02    0.02   0.129    0.724
## Residuals                   16   2.32    0.14
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
DanielPlot(cracking_model)
```



c

Write down a regression model that can be used to predict crack length as a function of the significant main effects and interactions you have identified in part (b). Build a RSM model (2nd order, 1st order with interaction). Choose one which works.

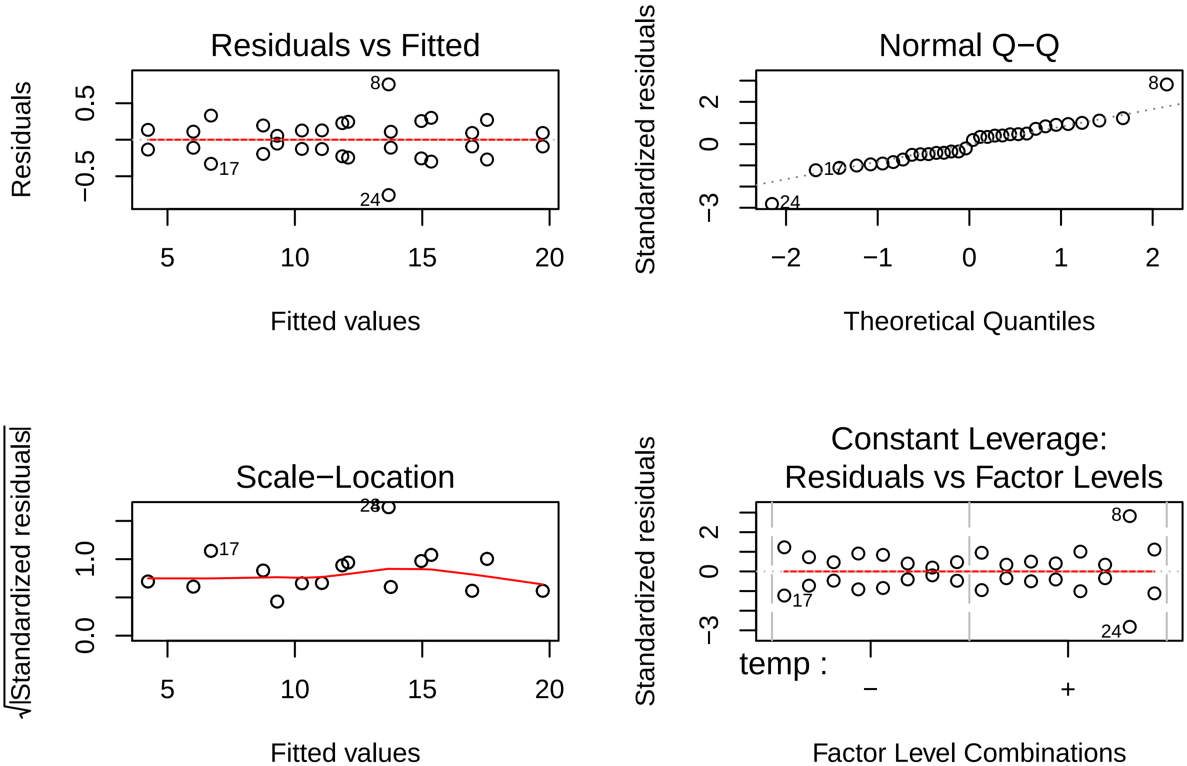
$$\begin{gathered} crack\ length = 8.2565 pouring\ temperature + 5.1555 titanium\ content + \\ 3.5705 heat\ treatment\ method + 2.0495 grain\ refiners - \\ 2.5745 pouring\ temperature \cdot titanium\ content - 14.3005 pouring\ temperature \cdot heat\ treatment\ method - \\ 6.1260 titanium\ content \cdot heat\ treatment\ method + 12.9935 pouring\ temperature \cdot titanium\ content \cdot heat\ treatment\ method \end{gathered}$$

cannot create rsm without numeric values for high and low settings of each factor.

d

Analyze the residuals from this experiment. Take out outliers, if any.

```r
par(mfrow=c(2,2))
plot(cracking_model)
```



The residuals are homoscedastic. No serious outliers.

e

Is there an indication that any of the factors affect the variability in cracking?

No

f

What recommendations would you make regarding process operations? Use interaction and/or main effect plots to assist in drawing conclusions. Perform a canonical analysis on the model. Is there an optimal response? Perform contour plot of Temp and Content.

Recommend to focus on temp, content, and method. Leverage the interaction between them.

3

Consider the three-variable central composite design shown below. Analyze the data and draw conclusions, assuming that we wish to maximize conversion (y,) with activity (y:) between 55 and 60.

```
conversion_data <- ccd(3, n0 = c(4,2), alpha = "rotatable", randomize = FALSE)
colnames(conversion_data) <- c('run','std.roder','time','temp','catalyst','Block')
conversion_data <- cbind(conversion_data,
                         conversion = c(74.00,51.00,88.00,70.00,
                                        71.00,90.00,66.00,97.00,
                                        81.00,75.00,76.00,83.00,
                                        76.00,79.00,85.00,97.00,
                                        35.00,81.00,80.00,91.00),
                         activity = c(53.20,62.90,53.40,62.60,
                                      57.30,67.90,59.80,67.80,
                                      59.20,60.40,59.10,60.60,
                                      59.10,65.90,60.00,60.70,
                                      57.40,63.20,60.80,38.90))
conversion_data %>% kable()
```

run	std.roder	time	temp	catalyst	Block	conversion	activity
1	1	-1.000000	-1.000000	-1.000000	1	74	53.2
2	2	1.000000	-1.000000	-1.000000	1	51	62.9
3	3	-1.000000	1.000000	-1.000000	1	88	53.4
4	4	1.000000	1.000000	-1.000000	1	70	62.6
5	5	-1.000000	-1.000000	1.000000	1	71	57.3
6	6	1.000000	-1.000000	1.000000	1	90	67.9
7	7	-1.000000	1.000000	1.000000	1	66	59.8
8	8	1.000000	1.000000	1.000000	1	97	67.8
9	9	0.000000	0.000000	0.000000	1	81	59.2
10	10	0.000000	0.000000	0.000000	1	75	60.4
11	11	0.000000	0.000000	0.000000	1	76	59.1
12	12	0.000000	0.000000	0.000000	1	83	60.6
1	1	-1.681793	0.000000	0.000000	2	76	59.1
2	2	1.681793	0.000000	0.000000	2	79	65.9
3	3	0.000000	-1.681793	0.000000	2	85	60.0
4	4	0.000000	1.681793	0.000000	2	97	60.7
5	5	0.000000	0.000000	-1.681793	2	35	57.4
6	6	0.000000	0.000000	1.681793	2	81	63.2
7	7	0.000000	0.000000	0.000000	2	80	60.8
8	8	0.000000	0.000000	0.000000	2	91	38.9

a

Estimate the factor effects. Which factors appear to be large?

b

Perform an analysis of variance. Do any factor affects . Use

c

Build a RSM models (choose a model which works). Daniel plot/Lenth plot.

d

Perform a residual analysis. Take out any outlyers.

e

Perform a canonical analysis. Any optimal response. Do a contour plot of Time-Temperature, Time-Catalyst, Temp-Catalyst.

4

The following data were collected by a chemical engineer. The response y is filtration time, x__1 ; is temperature, and x__3 ; is pressure. Fit a second-order model.

```
fitration_data <- data.frame(x1=c(-1,-1,1,1,-1.414,1.414,0,0,0,0,0,0,0),
                             x2=c(-1,1,-1,1,0,0,-1.414,1.414,0,0,0,0,0),
                             y =c(54,45,32,47,50,53,47,51,41,39,44,42,40))
fitration_data %>% kable()
```

x1	x2	y
-1.000	-1.000	54
-1.000	1.000	45
1.000	-1.000	32
1.000	1.000	47
-1.414	0.000	50
1.414	0.000	53
0.000	-1.414	47
0.000	1.414	51
0.000	0.000	41
0.000	0.000	39
0.000	0.000	44
0.000	0.000	42
0.000	0.000	40

a

What operating conditions would you recommend if the objective is to minimize the filtration time?

b

What operating conditions would you recommend if the objective is to operate the process at a mean filtration rate very close to 46?