Dec 04, 2022

STAT 6750 Applied Project

PRAGYA KHANAL

Bowling Green State University

Abstract

The main objective of this project was to maximize the yield in (%) of the chemical process with the use of 8 additives or factors. The yield could be anywhere between 0% to 100% but the preference is for higher values. We had a budget of 75 experimental runs, and we had to achieve the maximum within those runs. All the eight factors were continuous and limited to the interval [0,10]. I used fractional factorial, definitive screening and central composite design in this journey of obtaining the maximum response. Comparing these three methods, I chose Central Composite Design as my final design.

Section 1:

Considering all the 8 factors, from A to H, a full factorial model would be 2^8 with 256 runs. We had a budget of 75 runs only. So, we had to lower the size of our experiment and use the runs wisely in order to find the significant factors and maximize the response.

1.1

First of all, I decided to run a fractional factorial design of 2^{8-4} which would have 16 runs. I also added 4 center points to this fractional factorial design because I wanted to have enough degree of freedom to check for second order effect. Also, center points would increase the probability of detecting significant factors and estimate the variability. So, I had decided 20 runs for my first design.

As I selected my design, my job then was to set the levels of each factors making sure that they are within 0-10. I thought of setting each of the factor's levels differently rather than using the same numbers. I did this because I thought it would not be reasonable to set each factors level at the same points as each factors might have different input and characteristics. And it was mentioned to us that the amount of each additive may be set independent of other additives. So, I chose the set of levels for each of them randomly still having four of the factors set above 5. However, I made a typo while sending my design to the professor which costed me 2 extra runs to correct. I would say I used 22 runs for my first design.

After getting the data back, I analyzed the data first with rsm function which gave me two significant factors A and D in a first order model. I could not apply rsm function for second order model as some coefficients were aliased. But I checked the SSquad separately and found that there was no curvature. The highest predicted response value under the model generated from this data was below 40 and the adjusted R square was 0.6372. Predicted yield was pretty low than what I had expected it to be.

1.2

Few days back before we got to begin this project, we had learnt about definitive screening design (DSD) in the class. An example analysis was also illustrated on this kind of design. As DSD aims towards studying several factors in the hopes of finding a small number of important ones, I could relate it to the current project and I was eager to use this design in my experiment. It would also help me with the preparation of my final exams. For DSD, I had 2*8+1 = 17 runs. Only one center point was used. This time I increased the gap between the levels and set all the factors to 4 at lower level and 8 at higher level.

The response yield had not increased much. After I received back the response, I fit a first order model to check the significance of the factors which gave me factor A and E significant. Then I also checked the two factor interactions which were insignificant. The second order model I fitted also did not give me significant quadratic effects. But the p-values in this model were higher than the previous model. I could not see greater possibilities with those high p-values to get significant results if I continued following the same design. The highest predicted response within the levels set for this design would be below 50.

1.3

Combining the results of my first and second design, I got A, D and E significant. A and D were significant in the first design and A and E were significant in the second design. So, I was in this position of suspecting A, D and E to be significant factors for my model. At this point, I took suggestion from my professor on how I could move forward with these results with me. Following the suggestion, I decided to create a Central Composite Design. In this CCD, I would have amount of additives set to only the factors I considered significant meaning that B, C, F, G and H would be set to zero. This design had 6 axial points and 4 center points. I chose the natural units 4 and 6 as my lower and upper levels for A, D and E because I thought it would make a relatively balanced design. I could always increase and decrease the levels with my remaining runs.

The responses I received had decreased the yields in average. But, I was yet to fit models and check them. The first order model gave me A and D significant. No two factor interactions were significant. With a second order model, I could find only quadratic effect of D to be significant. E was not significant at all. I considered A and D only from now. I used contour plot to see the maximum and path of steepest ascent to get the predicted y values. Under this model and level settings, the highest yield would be around 45. A had a positive linear effect and D had a negative linear effect. But the quadratic effect for D was positive. I also performed a canonical analysis together with contour plots to determine my next set of additives.

1.4

The above CCD design result gave me an idea that factor E turned out to be insignificant. Now, I did not have enough runs to explore other factors in detail. So, I decided to go with A and D only for my final model, setting the rest of the factors to 0. I adjusted the level of A as 5 and 9 and D as 0 and 4 for my final design following the canonical analysis and contour plots with hopes of getting higher yield this time.

The responses I received were comparatively better this time although still not too high. I fit a first order model using rsm function which gave me A and D significant as expected. The interaction effect between A and B was also significant. However, the second order model did not find any quadratic effects significant. So, my final model had first order effect of A and D and an interaction factor between them. I used the contour plot and path of steepest ascent again which gave me highest yield of 56.076 within my settings of the additives.

Section 2:

2.1

The final model I used to determine the best settings to use is given below:

$$\hat{y} = 47.5 - 2.4463\ A - 7.3875\ B + 2.19\ AB$$

```
```{r}
CR1.rsmnew <- rsm(y ~ FO(x1,x2) + TWI(x1,x2), data=CR1)
summary(CR1.rsmnew)
```
```

```
Call:
rsm(formula = y ~ FO(x1, x2) + TWI(x1, x2), data = CR1)

             Estimate Std. Error  t value  Pr(>|t|)
(Intercept) 47.50000    0.40568 117.0879 3.164e-14 ***
x1          -2.44627    0.62885  -3.8901  0.004609 **
x2          -7.38752    0.62885 -11.7477 2.520e-06 ***
x1:x2        2.19000    0.70266   3.1167  0.014299 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

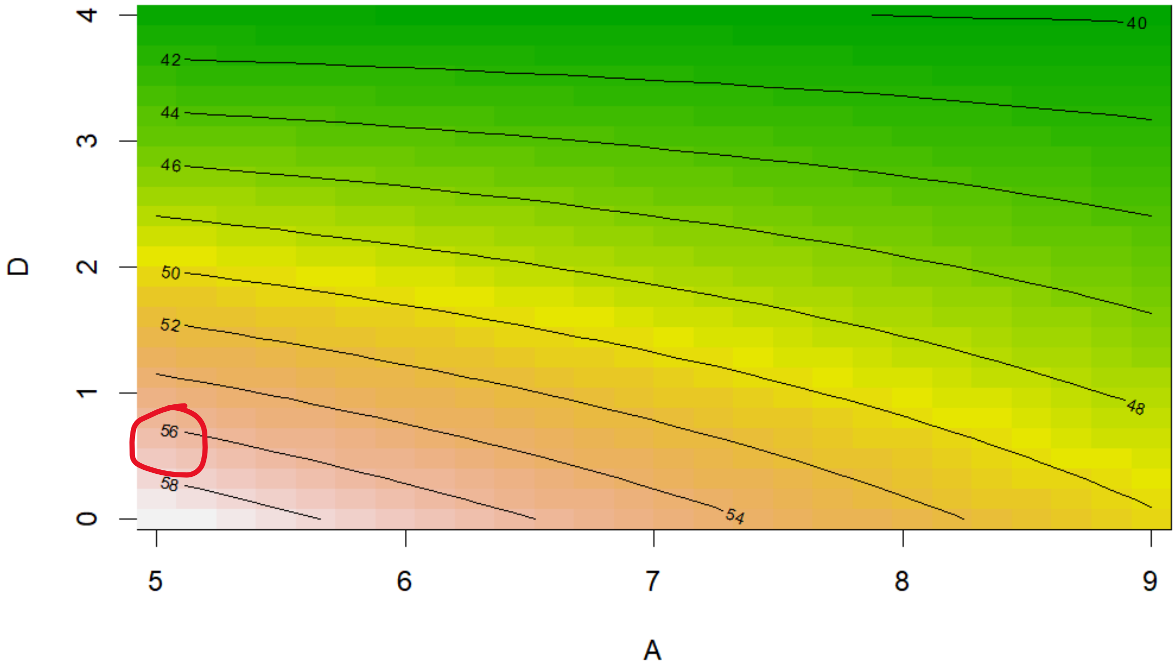
Multiple R-squared:  0.9532,    Adjusted R-squared:  0.9356
F-statistic: 54.29 on 3 and 8 DF,  p-value: 1.161e-05
```

In this model, A, D and interaction between A and D are significant.

2.2

To check the best setting, I used a contour plot in the above model which reflects the best setting as:

A=5, B=0, C=0, D=0.7, E=0, F=0, G=0, H=0



From the above contour plot, we can observe that the yield is 56 when A is set at 5 and D is set at 0.7 approximately.

2.3

I used ‘predict’ function to determine a 95% prediction yield in the above setting I specified.

````
```{r}
predict(CR1.rsmnew, df, interval="prediction")
```

       fit      lwr      upr
1 56.17166 52.23751 60.10581
````

Fitted value is 56.17166, lower value is 52.23751 and upper value is 60.10581.

Section 3

Submission 1

	A	B	C	D	E	F	G	H	pragya1
[1,]	8	2	6	3	4	7	5	9	46.55
[2,]	8	2	6	3	4	7	5	9	46.55
[3,]	9	3	5	2	5	8	4	8	46.90
[4,]	9	1	7	2	5	6	4	10	44.65
[5,]	7	3	5	4	3	8	6	8	41.60
[6,]	9	3	5	4	3	6	4	10	40.56
[7,]	7	3	5	3	5	6	6	10	49.89
[8,]	7	3	7	4	5	6	4	8	43.39
[9,]	9	3	7	3	3	6	6	8	40.88
[10,]	9	1	5	4	5	6	6	8	40.91
[11,]	8	2	6	3	4	7	5	9	46.52
[12,]	9	1	5	2	3	8	6	10	41.91
[13,]	7	1	7	4	3	6	6	10	42.90
[14,]	9	3	7	4	5	8	6	10	40.72
[15,]	7	3	7	2	3	8	4	10	50.06
[16,]	7	1	7	2	5	8	6	8	50.06
[17,]	7	1	5	4	5	8	4	10	45.70
[18,]	9	1	7	4	3	8	4	8	40.28
[19,]	7	1	5	2	3	6	4	8	49.94
[20,]	8	2	6	3	4	7	5	9	47.37

I made a typographical error in my first submission on 7th and 9th runs. I could not use this design until I corrected those runs. So, I had to make a new submission with those runs corrected.

Section 4

Submission 2

	A	B	C	D	E	F	G	H	pragya2
[1,]	8	2	6	3	4	7	5	9	46.55
[2,]	8	2	6	3	4	7	5	9	46.55
[3,]	9	3	5	2	5	8	4	8	46.90
[4,]	9	1	7	2	5	6	4	10	44.65
[5,]	7	3	5	4	3	8	6	8	41.60
[6,]	9	3	5	4	3	6	4	10	40.56
[7,]	7	3	5	2	5	6	6	10	49.50
[8,]	7	3	7	4	5	6	4	8	43.39
[9,]	9	3	7	2	3	6	6	8	42.61
[10,]	9	1	5	4	5	6	6	8	40.91
[11,]	8	2	6	3	4	7	5	9	46.52
[12,]	9	1	5	2	3	8	6	10	41.91
[13,]	7	1	7	4	3	6	6	10	42.90
[14,]	9	3	7	4	5	8	6	10	40.72
[15,]	7	3	7	2	3	8	4	10	50.06
[16,]	7	1	7	2	5	8	6	8	50.06
[17,]	7	1	5	4	5	8	4	10	45.70
[18,]	9	1	7	4	3	8	4	8	40.28
[19,]	7	1	5	2	3	6	4	8	49.94
[20,]	8	2	6	3	4	7	5	9	47.37

I corrected those runs and received the above set of responses.

Then, I checked for the first order effects which gave me the result below.

```
Call:
lm.default(formula = y ~ A + B + C + D + E + F + G + H + AB +
    AC + AD + AE + AF + AG + AH)

Residuals:
      1       2       3       4       5       6       7       8       9      10      11      12
 1.6830  1.6830 -0.4701 -0.4701 -0.4701 -0.4701 -0.4701 -0.4701 -0.4701 -0.4701  1.6530 -0.4701
     13      14      15      16      17      18      19      20
-0.4701 -0.4701 -0.4701 -0.4701 -0.4701 -0.4701 -0.4701  2.5030

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 44.86700    0.47695  94.070 7.66e-08 ***
A           -2.29563    0.53325  -4.305   0.0126 *
B           -0.14687    0.53325  -0.275   0.7966
C           -0.27937    0.53325  -0.524   0.6280
D           -2.38938    0.53325  -4.481   0.0110 *
E            0.88063    0.53325   1.651   0.1740
F            0.25688    0.53325   0.482   0.6552
G           -0.78813    0.53325  -1.478   0.2135
H            0.15188    0.53325   0.285   0.7899
AB           0.31063    0.53325   0.583   0.5915
AC          -0.18937    0.53325  -0.355   0.7404
AD           0.90562    0.53325   1.698   0.1647
AE           0.31312    0.53325   0.587   0.5886
AF           0.09437    0.53325   0.177   0.8681
AG          -0.20813    0.53325  -0.390   0.7162
AH          -0.29313    0.53325  -0.550   0.6118
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.133 on 4 degrees of freedom
Multiple R-squared:  0.9236,    Adjusted R-squared:  0.6372
F-statistic: 3.224 on 15 and 4 DF,  p-value: 0.1333
```

\

In the linear model, factor A and factor D were significant with no significant interaction effects.

Another lowest p value was for factor E but it was not significant.

```
 y_bar_f  y_bar_c  SS_quad      SS_E     MS_E
44.26600 46.49500 15.68981 63.83550 21.27850
            SS         Fvec     p_val
A  10.53978828 0.4953257176 0.5322694
B   0.04314453 0.0020276115 0.9669138
C   0.15610078 0.0073360801 0.9371400
D  11.41822578 0.5366085853 0.5168855
E   1.55100078 0.0728905130 0.8046805
F   0.13196953 0.0062020129 0.9421880
G   1.24228203 0.0583820303 0.8246459
H   0.04613203 0.0021680114 0.9657886
AB  0.19297578 0.0090690500 0.9301355
AC  0.07172578 0.0033708100 0.9573527
AD  1.64031328 0.0770878249 0.7993177
AE  0.19609453 0.0092156182 0.9295755
AF  0.01781328 0.0008371493 0.9787348
AG  0.08663203 0.0040713411 0.9531375
AH  0.17184453 0.0080759702 0.9340570
   15.68981368 0.7373552499 0.4536458
   63.83550000           NA        NA
```

When I checked SSquad, none of the quadratic effects were significant. The highest predicted response value under the model generated from this data was below 40.

Section 5

Submission 3

	A	B	C	D	E	F	G	H	pragya3
[1,]	6	4	8	8	4	8	8	8	48.74
[2,]	6	8	4	4	8	4	4	4	51.52
[3,]	4	6	4	8	8	8	8	4	46.73
[4,]	8	6	8	4	4	4	4	8	45.50
[5,]	4	4	6	8	8	4	4	8	44.55
[6,]	8	8	6	4	4	8	8	4	43.11
[7,]	8	4	8	6	8	8	4	4	47.56
[8,]	4	8	4	6	4	4	8	8	38.04
[9,]	4	4	8	4	6	4	8	4	44.38
[10,]	8	8	4	8	6	8	4	8	47.94
[11,]	8	4	4	4	8	6	8	8	49.67
[12,]	4	8	8	8	4	6	4	4	36.72
[13,]	4	8	8	4	8	8	6	8	46.34
[14,]	8	4	4	8	4	4	6	4	45.92
[15,]	8	8	8	8	8	4	8	6	54.51
[16,]	4	4	4	4	4	8	4	6	41.76
[17,]	6	6	6	6	6	6	6	6	43.95

With this submission of a DSD design, there was a slight increase in yield. The result from a linear model is given below.

```
Call:
lm(formula = y2 ~ A + B + C + D + E + F + G + H)

Residuals:
    Min      1Q  Median      3Q     Max
-2.9024 -1.8724  0.4126  0.4576  4.4276

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  26.5002     7.7205   3.432  0.00892 **
A             1.2746     0.4523   2.818  0.02257 *
B            -0.1571     0.4523  -0.347  0.73726
C             0.0775     0.4523   0.171  0.86822
D             0.1011     0.4523   0.223  0.82880
E             1.4675     0.4523   3.244  0.01180 *
F            -0.0800     0.4523  -0.177  0.86402
G             0.3439     0.4523   0.760  0.46888
H             0.1729     0.4523   0.382  0.71232
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.385 on 8 degrees of freedom
Multiple R-squared:  0.7082,	Adjusted R-squared:  0.4165
F-statistic: 2.427 on 8 and 8 DF,  p-value: 0.1156
```

From the above linear regression model, we could see that factors A and E are significant. Now, we fit a model using rsm function to test for curvature.

```
```{r}
#fit first order model (FO)
CR1.rsm <- rsm(y2 ~ FO(x1,x2), data = CR1)
summary(CR1.rsm)
```
```

```
Call:
rsm(formula = y2 ~ FO(x1, x2), data = CR1)

            Estimate Std. Error t value  Pr(>|t|)
(Intercept) 45.70235    0.65664 69.6006 < 2.2e-16 ***
x1           2.54929    0.72358  3.5232  0.003376 **
x2           2.93500    0.72358  4.0562  0.001179 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Multiple R-squared:  0.6734,	Adjusted R-squared:  0.6267
F-statistic: 14.43 on 2 and 14 DF,  p-value: 0.0003964

Analysis of Variance Table

Response: y2
            Df  Sum Sq Mean Sq F value    Pr(>F)
FO(x1, x2)   2 211.583 105.792 14.4328 0.0003964
Residuals   14 102.619   7.330
Lack of fit  6  56.267   9.378  1.6185 0.2580659
Pure error   8  46.352   5.794

Direction of steepest ascent (at radius 1):
       x1        x2
0.6557556 0.7549733

Corresponding increment in original units:
       A        E
1.311511 1.509947
```

Lack of fit was not significant from the above result.

Section 6

Submission 4

	A	B	C	D	E	F	G	H	pragya4
[1,]	4.00	0	0	4.00	4.00	0	0	0	44.07
[2,]	6.00	0	0	4.00	4.00	0	0	0	45.39
[3,]	4.00	0	0	6.00	4.00	0	0	0	37.44
[4,]	6.00	0	0	6.00	4.00	0	0	0	43.00
[5,]	4.00	0	0	4.00	6.00	0	0	0	41.30
[6,]	6.00	0	0	4.00	6.00	0	0	0	47.35
[7,]	4.00	0	0	6.00	6.00	0	0	0	38.53
[8,]	6.00	0	0	6.00	6.00	0	0	0	43.08
[9,]	6.73	0	0	5.00	5.00	0	0	0	45.62
[10,]	3.27	0	0	5.00	5.00	0	0	0	34.79
[11,]	5.00	0	0	6.73	5.00	0	0	0	42.55
[12,]	5.00	0	0	3.27	5.00	0	0	0	46.75
[13,]	5.00	0	0	5.00	6.73	0	0	0	44.82
[14,]	5.00	0	0	5.00	3.27	0	0	0	41.67
[15,]	5.00	0	0	5.00	5.00	0	0	0	42.30
[16,]	5.00	0	0	5.00	5.00	0	0	0	41.78
[17,]	5.00	0	0	5.00	5.00	0	0	0	39.98
[18,]	5.00	0	0	5.00	5.00	0	0	0	41.45

I was in this position of suspecting A, D and E to be significant factors for my model from previous results. So, rest of the factors were set at zero. I used rsm function to get the first order model below:

```{r}

#fit first order model (FO)
CR1.rsm <- rsm(y ~ FO(x1,x2,x3), data = CR1)
summary(CR1.rsm)

```

```
Call:
rsm(formula = y ~ FO(x1, x2, x3), data = CR1)

            Estimate Std. Error  t value  Pr(>|t|)
(Intercept) 42.32611    0.41507 101.9739 < 2.2e-16 ***
x1           2.58948    0.47088   5.4992 7.834e-05 ***
x2          -1.66783    0.47088  -3.5419  0.003253 **
x3           0.41539    0.47088   0.8821  0.392593
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Multiple R-squared:  0.7568,    Adjusted R-squared:  0.7047
F-statistic: 14.52 on 3 and 14 DF,  p-value: 0.0001403

Analysis of Variance Table

Response: y
               Df  Sum Sq Mean Sq F value    Pr(>F)
FO(x1, x2, x3)  3 135.097  45.032 14.5216 0.0001403
Residuals      14  43.415   3.101
Lack of fit    11  40.444   3.677  3.7122 0.1538329
Pure error      3   2.971   0.990

Direction of steepest ascent (at radius 1):
        x1         x2         x3
 0.8331674 -0.5366279  0.1336509

Corresponding increment in original units:
         A          D          E
 0.8331674 -0.5366279  0.1336509
```

From the above first order model, A and D were significant.

Then I checked for two factor interactions if any which turned out to be insignificant.

```r
  {r}

#adapt to add the interaction
#TWI = two-way interaction
CR1.rsmi <- update(CR1.rsm, . ~ . + TWI(x1,x2,x3))
summary(CR1.rsmi)
```

```
Call:
rsm(formula = y ~ FO(x1, x2, x3) + TWI(x1, x2, x3), data = CR1)

            Estimate Std. Error t value  Pr(>|t|)
(Intercept) 42.32611    0.45091 93.8690 < 2.2e-16 ***
x1           2.58948    0.51154  5.0621 0.0003651 ***
x2          -1.66783    0.51154 -3.2604 0.0075932 **
x3           0.41539    0.51154  0.8120 0.4339949
x1:x2        0.34250    0.67636  0.5064 0.6225798
x1:x3        0.46500    0.67636  0.6875 0.5060022
x2:x3        0.24750    0.67636  0.3659 0.7213574
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Multiple R-squared:  0.7745,    Adjusted R-squared:  0.6515
F-statistic: 6.296 on 6 and 11 DF,  p-value: 0.004426

Analysis of Variance Table

Response: y
                Df  Sum Sq Mean Sq F value    Pr(>F)
FO(x1, x2, x3)   3 135.097  45.032 12.3050 0.0007739
TWI(x1, x2, x3)  3   3.158   1.053  0.2877 0.8333890
Residuals       11  40.257   3.660
Lack of fit      8  37.285   4.661  4.7057 0.1150129
Pure error       3   2.971   0.990

Stationary point of response surface:
       x1        x2        x3
 4.000216 -9.193884  1.203072

Stationary point in original units:
        A         D         E
 9.000216 -4.193884  6.203072

Eigenanalysis:
eigen() decomposition
$values
[1]  0.3555461 -0.1154264 -0.2401197

$vectors
         [,1]       [,2]       [,3]
x1 -0.6292974  0.1841181  0.7550400
x2 -0.5078559 -0.8328251 -0.2201926
x3 -0.5882748  0.5220181 -0.6176001
```

Then I fit a second order model to check if there are any quadratic effects.

```{r}
CR2.rsm <- rsm(y ~ SO(x1, x2,x3), data = CR1)
summary(CR2.rsm)
```

```
Call:
rsm(formula = y ~ SO(x1, x2, x3), data = CR1)

            Estimate Std. Error t value  Pr(>|t|)
(Intercept) 41.37787    0.68427 60.4702 6.215e-12 ***
x1           2.58948    0.36594  7.0762 0.0001044 ***
x2          -1.66783    0.36594 -4.5576 0.0018557 **
x3           0.41539    0.36594  1.1351 0.2891922
x1:x2        0.34250    0.48385  0.7079 0.4991323
x1:x3        0.46500    0.48385  0.9610 0.3646766
x2:x3        0.24750    0.48385  0.5115 0.6228034
x1^2        -0.42684    0.36976 -1.1544 0.2816642
x2^2         1.05834    0.36976  2.8623 0.0210782 *
x3^2         0.58890    0.36976  1.5927 0.1499023
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.9161,    Adjusted R-squared:  0.8216
F-statistic: 9.701 on 9 and 8 DF,  p-value: 0.001958

Analysis of Variance Table

Response: y
                Df  Sum Sq Mean Sq F value    Pr(>F)
FO(x1, x2, x3)   3 135.097  45.032 24.0442 0.0002345
TWI(x1, x2, x3)  3   3.158   1.053  0.5621 0.6549498
PQ(x1, x2, x3)   3  25.273   8.424  4.4981 0.0395392
Residuals        8  14.983   1.873
Lack of fit      5  12.012   2.402  2.4256 0.2482698
Pure error       3   2.971   0.990

Stationary point of response surface:
       x1         x2         x3
 2.4698566  0.5572479 -1.4448901

Stationary point in original units:
       A        D        E
7.469857 5.557248 3.555110

Eigenanalysis:
eigen() decomposition
$values
[1]  1.1221020  0.5888413 -0.4905401

$vectors
         [,1]       [,2]        [,3]
x1 -0.1473933  0.1629692  0.97555945
x2 -0.9473838 -0.3066195 -0.09191496
x3 -0.2841462  0.9377769 -0.19958806
```

Quadratic effect of factor D only turned out to be significant in this model. So, now I decided to not include E in the further analysis.

```{r}
CR2.rsm <- rsm(y ~ SO(x1, x2), data = CR1)
summary(CR2.rsm)
```

```
Call:
rsm(formula = y ~ SO(x1, x2), data = CR1)

            Estimate Std. Error t value  Pr(>|t|)
(Intercept) 42.05081    0.56043 75.0328 < 2.2e-16 ***
x1           2.58948    0.38104  6.7959 1.914e-05 ***
x2          -1.66783    0.38104 -4.3771 0.0009009 ***
x1:x2        0.34250    0.50381  0.6798 0.5095196
x1^2        -0.56543    0.37419 -1.5111 0.1566486
x2^2         0.91975    0.37419  2.4579 0.0301519 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.8635,    Adjusted R-squared:  0.8066
F-statistic: 15.18 on 5 and 12 DF,  p-value: 7.871e-05

Analysis of Variance Table

Response: y
            Df  Sum Sq Mean Sq F value    Pr(>F)
FO(x1, x2)   2 132.684  66.342 32.6714 1.395e-05
TWI(x1, x2)  1   0.938   0.938  0.4622   0.50952
PQ(x1, x2)   2  20.523  10.261  5.0534   0.02558
Residuals   12  24.367   2.031
Lack of fit  3   5.430   1.810  0.8602   0.49608
Pure error   9  18.937   2.104

Stationary point of response surface:
       x1        x2
2.4275267 0.4546932

Stationary point in original units:
       A        D
7.427527 5.454693

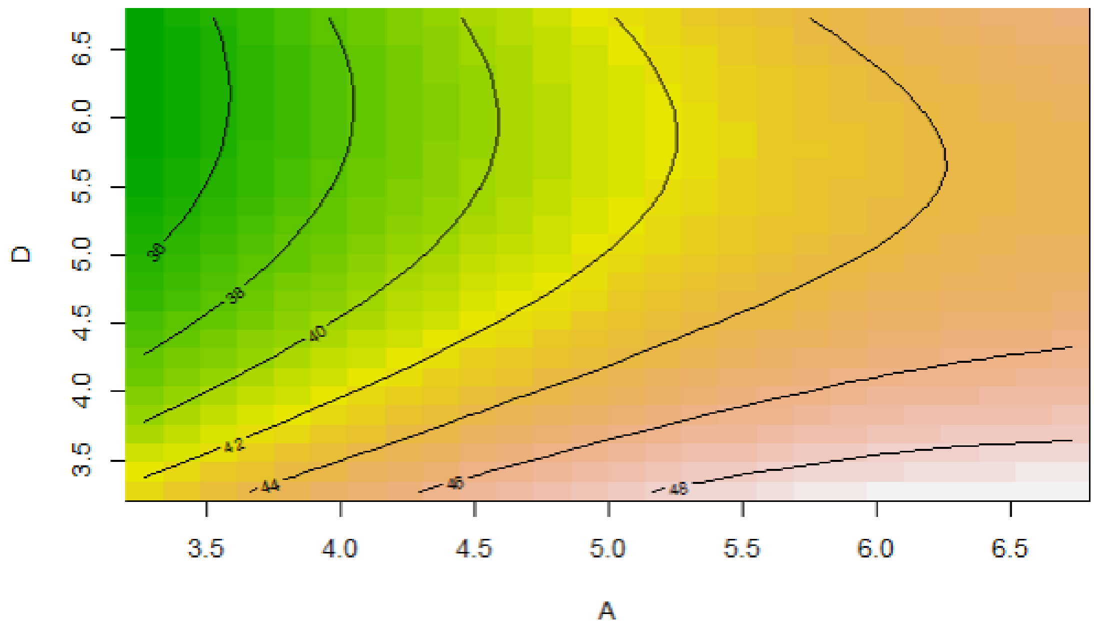
Eigenanalysis:
eigen() decomposition
$values
[1]  0.9392386 -0.5849236

$vectors
        [,1]       [,2]
x1 0.1130822 -0.9935856
x2 0.9935856  0.1130822
```

The reduced hierarchical model had the following:

$$\hat{y} = 42.05081 + 2.58948\,A - 1.66783\,D + 0.91975\,D^2$$

Using the above model, I got a contour plot.



This plot suggests the highest yield to be 48 when A=6.5 and D = 3.5. But that is outside the range of the values I have set for A and D. If I try to work a little above 6.5 for A and below 3.5 for D, I may get a higher yield.

```{r}
df=data.frame(x1=(7.5-5)/1,x2=(2-5)/1)
predict(CR2.rsm, df, interval="prediction")
```

```
       fit      lwr      upr
1 55.70303 42.86868 68.53738
```

```{r}
df=data.frame(x1=(8.5-5)/1,x2=(1-5)/1)
predict(CR2.rsm, df, interval="prediction")
```

```
       fit     lwr      upr
1 60.77973 37.2938 84.26567
```

I also ran a code for steepest ascent which gave me the output below.

path <- steepest(CR2.rsm)

path

	dist	x1	x2	|	A	D	|	yhat
	<dbl>	<dbl>	<dbl>	<fctr>	<dbl>	<dbl>	<fctr>	<dbl>
1	0.0	0.000	0.000	|	5.000	5.000	|	42.051
2	0.5	0.341	-0.365	|	5.341	4.635	|	43.557
3	1.0	0.488	-0.873	|	5.488	4.127	|	45.191
4	1.5	0.523	-1.406	|	5.523	3.594	|	47.162
5	2.0	0.515	-1.933	|	5.515	3.067	|	49.554
6	2.5	0.488	-2.452	|	5.488	2.548	|	52.389
7	3.0	0.452	-2.965	|	5.452	2.035	|	55.678
8	3.5	0.410	-3.476	|	5.410	1.524	|	59.440
9	4.0	0.365	-3.984	|	5.365	1.016	|	63.666
10	4.5	0.317	-4.488	|	5.317	0.512	|	68.339
11	5.0	0.268	-4.993	|	5.268	0.007	|	73.503

11 rows

From the above result, I could have picked 63 or 68 or even 73 as the maximum yield but the additive amount for D is out of range of levels that I have set. The coded units also show an imbalance.

```
```{r}
df=data.frame(x1=(5.3-5)/1,x2=(0.5-5)/1)
predict(CR2.rsm, df, interval="prediction")
```
```

```
       fit      lwr      upr
1 68.44455 51.85014 85.03896
```

The last thing I tried here was canonical analysis.

canonical.path(CR2.rsm)

	dist <dbl>	x1 <dbl>	x2 <dbl>	<fctr>	A <dbl>	D <dbl>	<fctr>	yhat <dbl>
1	-5.0	1.862	-4.513		6.862	0.487		68.294
2	-4.5	1.919	-4.016		6.919	0.984		63.830
3	-4.0	1.975	-3.520		6.975	1.480		59.845
4	-3.5	2.032	-3.023		7.032	1.977		56.321
5	-3.0	2.088	-2.526		7.088	2.474		53.268
6	-2.5	2.145	-2.029		7.145	2.971		50.684
7	-2.0	2.201	-1.532		7.201	3.468		48.570
8	-1.5	2.258	-1.036		7.258	3.964		46.929
9	-1.0	2.314	-0.539		7.314	4.461		45.754
10	-0.5	2.371	-0.042		7.371	4.958		45.049
11	0.0	2.428	0.455		7.428	5.455		44.815
12	0.5	2.484	0.951		7.484	5.951		45.049
13	1.0	2.541	1.448		7.541	6.448		45.753
14	1.5	2.597	1.945		7.597	6.945		46.928
15	2.0	2.654	2.442		7.654	7.442		48.572
16	2.5	2.710	2.939		7.710	7.939		50.686
17	3.0	2.767	3.435		7.767	8.435		53.265
18	3.5	2.823	3.932		7.823	8.932		56.318
19	4.0	2.880	4.429		7.880	9.429		59.842
20	4.5	2.936	4.926		7.936	9.926		63.835
21	5.0	2.993	5.423		7.993	10.423		68.299

Since the eigen values were negative, the stationery point was a saddle point. So, the response has decreased in the beginning up to 11th step and then increased till the end. The highest response I could consider here was 68.294 or 63.835 but the units for A and D are away from what I had set.

```
```{r}
df=data.frame(x1=(6.8-5)/1,x2=(0.5-5)/1)
predict(CR2.rsm, df, interval="prediction")
```
```

```
       fit      lwr      upr
1 68.23577 48.90547 87.56608
```

```
```{r}
df=data.frame(x1=(7.9-5)/1,x2=(9.9-5)/1)
predict(CR2.rsm, df, interval="prediction")
```
```

```
       fit      lwr      upr
1 63.58269 36.78219 90.3832
```

I think I cannot use the above fitted value as my maximum y because the values for A and D are away from what I had set. So, now did a new submission by adjusting the points following the above prediction results.

Section 7

Submission 5

```
        A B C    D E F G H pragya5
 [1,] 5.00 0 0 0.00 0 0 0 0   60.27
 [2,] 5.00 0 0 4.00 0 0 0 0   40.45
 [3,] 9.00 0 0 0.00 0 0 0 0   50.12
 [4,] 9.00 0 0 4.00 0 0 0 0   39.06
 [5,] 8.41 0 0 2.00 0 0 0 0   47.28
 [6,] 5.59 0 0 2.00 0 0 0 0   48.24
 [7,] 7.00 0 0 3.41 0 0 0 0   42.62
 [8,] 7.00 0 0 0.59 0 0 0 0   51.15
 [9,] 7.00 0 0 2.00 0 0 0 0   48.07
[10,] 7.00 0 0 2.00 0 0 0 0   45.46
[11,] 7.00 0 0 2.00 0 0 0 0   49.57
[12,] 7.00 0 0 2.00 0 0 0 0   47.71
```

After receiving the response, I fit a first order model using rsm.

```
#fit first order model (FO)
CR1.rsm <- rsm(y ~ FO(x1,x2), data = CR1)
summary(CR1.rsm)
```

```
Call:
rsm(formula = y ~ FO(x1, x2), data = CR1)

             Estimate Std. Error t value  Pr(>|t|)
(Intercept) 47.50000    0.56914 83.4590 2.578e-14 ***
x1          -2.44627    0.88224 -2.7728   0.02165 *
x2          -7.38752    0.88224 -8.3736 1.534e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Multiple R-squared:  0.8963,    Adjusted R-squared:  0.8733
F-statistic:  38.9 on 2 and 9 DF,  p-value: 3.721e-05
```

```{r}
CR1.rsmnew <- rsm(y ~ FO(x1,x2) + TWI(x1,x2), data=CR1)
summary(CR1.rsmnew)
```

```
Call:
rsm(formula = y ~ FO(x1, x2) + TWI(x1, x2), data = CR1)

            Estimate Std. Error  t value  Pr(>|t|)
(Intercept) 47.50000    0.40568 117.0879 3.164e-14 ***
x1          -2.44627    0.62885  -3.8901  0.004609 **
x2          -7.38752    0.62885 -11.7477 2.520e-06 ***
x1:x2        2.19000    0.70266   3.1167  0.014299 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.9532,    Adjusted R-squared:  0.9356
F-statistic: 54.29 on 3 and 8 DF,  p-value: 1.161e-05
```

Main effects of A and D and the interaction effect between them was significant.

Then, I checked for any second order effects.

```
Call:
rsm(formula = y ~ SO(x1, x2), data = CR1)

            Estimate Std. Error  t value  Pr(>|t|)
(Intercept) 47.55305    0.63871  74.4519 3.952e-10 ***
x1          -2.44627    0.70749  -3.4577   0.01351 *
x2          -7.38752    0.70749 -10.4419 4.524e-05 ***
x1:x2        2.19000    0.79052   2.7703   0.03241 *
x1^2         0.81650    1.67866   0.4864   0.64394
x2^2        -0.94397    1.67866  -0.5623   0.59426
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared:  0.9556,    Adjusted R-squared:  0.9185
F-statistic:  25.8 on 5 and 6 DF,  p-value: 0.0005478
```

The second order model was negligible, so I had to move forward with the model including main effects and interaction effect only.

$$\hat{y} = 47.5 - 2.4463\ A - 7.3875\ B + 2.19\ AB$$

Finally, using contour plot I found the points A=5 and D=0.7 (while staying under the range I had used in my design). The output in this setting was 56. I used predict function to get the predicted yield in 95% confident interval which resulted in fitted value as 56.17166, lower value as 52.23751 and upper value as 60.10581. (details and contour plot shown in section 2)

Section 8

This project was a very constructive, meaningful, and amazing experience for me. I could do the implication of the concepts and learnings of last two semesters. I learnt how we could use the designs in experiments in a real world. I always feared of getting worser results of the responses in each submission. The constraint of 75 runs was disturbing yet giving me a feeling of being involved in an actual cost restriction which would be a important concern in real world experiments. I think now I can explain the fractional factorial, definitive screening, and central composite designs with examples to interested groups. I have also gained a confidence of handling future projects independently. I also did some mistakes in this project, typographical errors which costed me few runs. This gave me a lesson that how resources value a lot in an experiment and how wisely and prudently we must utilize them to optimize the results. Also, I learnt to combine results of different designs to make decisions ahead. I am thankful to my professor for supporting me with understanding the concepts and moving forward.