CSCI 688
Homework 7a

Megan Rose Bryant
Department of Mathematics
William and Mary

November 19, 2014

8.1 *Suppose that in the chemical process development experiment described in Problem 6.7, it was only possible to run a one-half fraction of the 2^4 design. Construct the design and perform the statistical analysis, using the data from replicate 1.*
We will use the following 2^{4-1} design with the defining relation $I = ABCD$.

```
Fractional Factorial Design
Factors:  4   Base Design:         4, 8   Resolution:   IV
Runs:     8   Replicates:             1   Fraction:    1/2
Blocks:   1   Center pts (total):     0

Design Generators: D = ABC
Defining Relation:  I = ABCD

Alias Structure
I + ABCD
A + BCD
B + ACD
C + ABD
D + ABC
AB + CD
AC + BD
AD + BC

Design Table
Run  A  B  C  D
  1  -  -  -  -
  2  +  -  -  +
  3  -  +  -  +
  4  +  +  -  -
  5  -  -  +  +
  6  +  -  +  -
  7  -  +  +  -
  8  +  +  +  +
```

We see from our initial analysis shown below that factors A, AB, and AD seem to be the most significant. We chose these factors because A is aliased with a three factor interaction, which we assume to be more negligible. Furthermore, AB and AD are considered to be significant over their aliases since they include factor A, which is believed to be the most significant.

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Model	7	448.000	100.00%	448.000	64.000	*	*
Linear	4	324.000	72.32%	324.000	81.000	*	*
A	1	288.000	64.29%	288.000	288.000	*	*
B	1	2.000	0.45%	2.000	2.000	*	*
C	1	32.000	7.14%	32.000	32.000	*	*
D	1	2.000	0.45%	2.000	2.000	*	*
2-Way Interactions	3	124.000	27.68%	124.000	41.333	*	*
A*B	1	72.000	16.07%	72.000	72.000	*	*
A*C	1	2.000	0.45%	2.000	2.000	*	*
A*D	1	50.000	11.16%	50.000	50.000	*	*
Error	0	*	*	*	*		
Total	7	448.000	100.00%				

Therefore, we will rerun the analysis focusing on these factors and including factors B and D for heirarchy.

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Model	5	414.000	92.41%	414.000	82.800	4.87	0.179
Linear	3	292.000	65.18%	292.000	97.333	5.73	0.152
A	1	288.000	64.29%	288.000	288.000	16.94	0.054
B	1	2.000	0.45%	2.000	2.000	0.12	0.764
D	1	2.000	0.45%	2.000	2.000	0.12	0.764
2-Way Interactions	2	122.000	27.23%	122.000	61.000	3.59	0.218
A*B	1	72.000	16.07%	72.000	72.000	4.24	0.176
A*D	1	50.000	11.16%	50.000	50.000	2.94	0.228
Error	2	34.000	7.59%	34.000	17.000		
Total	7	448.000	100.00%				

Model Summary

S	R-sq	R-sq(adj)	PRESS	R-sq(pred)
4.12311	92.41%	73.44%	544	0.00%

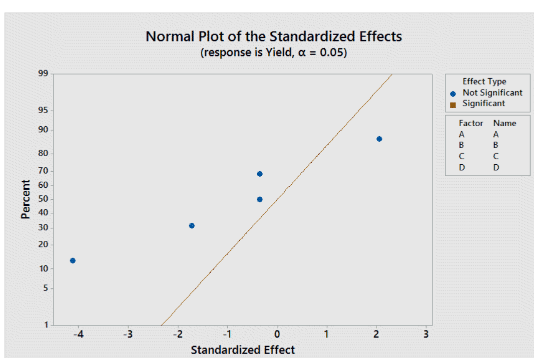
Coded Coefficients

Term	Effect	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant		85.00	1.46	(78.73, 91.27)	58.31	0.000	
A	-12.00	-6.00	1.46	(-12.27, 0.27)	-4.12	0.054	1.00
B	-1.00	-0.50	1.46	(-6.77, 5.77)	-0.34	0.764	1.00
D	-1.00	-0.50	1.46	(-6.77, 5.77)	-0.34	0.764	1.00
A*B	6.00	3.00	1.46	(-3.27, 9.27)	2.06	0.176	1.00
A*D	-5.00	-2.50	1.46	(-8.77, 3.77)	-1.71	0.228	1.00

Regression Equation in Uncoded Units

Yield = 85.00 - 6.00 A - 0.50 B - 0.50 D + 3.00 A*B - 2.50 A*D

We see from our new analysis of variance that our revised model has an R^2 of 92.41%, meaning that we have accounted for a fair amount of variance in the model, but there is still room for improvement. This is further evidenced by the fact that none of our factors are significant at the 5% significance level. We see this in the normal probability plot below.



Therefore, we shall run the model once more, this time considering only factor A.

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Model	1	288.0	64.29%	288.0	288.00	10.80	0.017
Linear	1	288.0	64.29%	288.0	288.00	10.80	0.017
A	1	288.0	64.29%	288.0	288.00	10.80	0.017
Error	6	160.0	35.71%	160.0	26.67		
Total	7	448.0	100.00%				

Model Summary

S	R-sq	R-sq(adj)	PRESS	R-sq(pred)
5.16398	64.29%	58.33%	284.444	36.51%

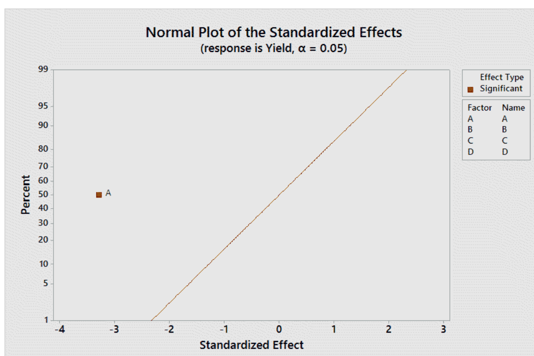
Coded Coefficients

Term	Effect	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant		85.00	1.83	(80.53, 89.47)	46.56	0.000	
A	-12.00	-6.00	1.83	(-10.47, -1.53)	-3.29	0.017	1.00

Regression Equation in Uncoded Units

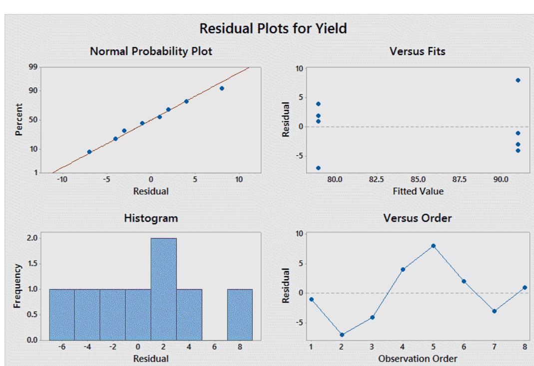
Yield = 85.00 - 6.00 A

This indicates that factor A is significant in determining the yield output. This is supported by the normal probability plot below.



We see that while A has become significant in this model at the 5% level, our R^2 has decreased dramatically. Therefore, we would recommend that the confirming experiment of the non-primal fraction be run so that we may further improve our understanding.

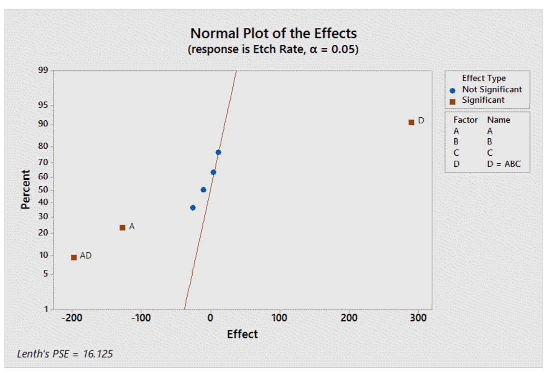
An analysis of the $4-in-1$ residual plots does not reveal any reasons to question our normality and equal variance assumptions.



8.3 *Consider the plasma etch experiment described in Example 6.1. Suppose that only a one-half fraction of the design could be run. Set up the design and analyze the data.*
We will use the followin 2^{3-1},2-replicate design. Note tha $D = ABC$.

```
Defining Relation:  I = ABC

Aliases

I + ABCD
A + BCD
B + ACD
C + ABD
D + ABC
AB + CD
AC + BD
AD + BC

Design Table
Run  A  B  C D
  1  - - - -
  2  + + - -
  3  + - + -
  4  - + + -
  5  + - - +
  6  - + - +
  7  - - + +
```

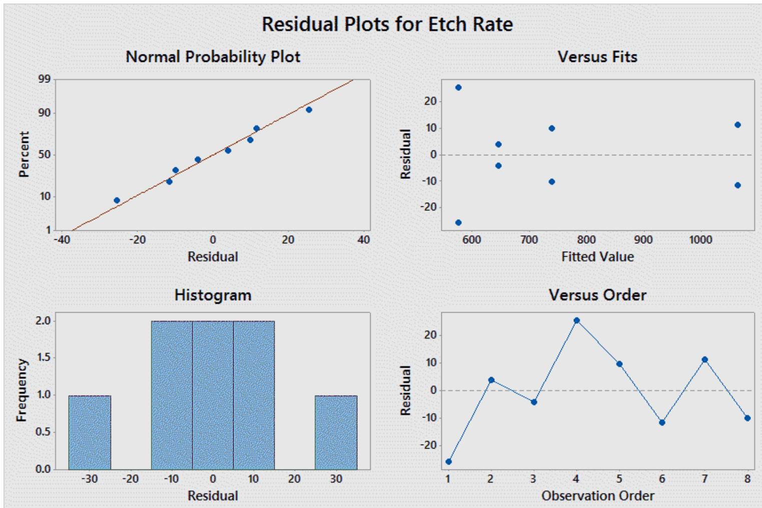
We see from the initial normal probability plot of effects included below that A, D, and AD are the significant factors.




```
Analysis of Variance
Source                   DF  Seq SS  Contribution  Adj SS  Adj MS  F-Value  P-Value
Model                     3  279051        99.36%  279051   93017   207.05    0.000
  Linear                  2  201039        71.58%  201039  100519   223.75    0.000
    A                     1   32258        11.49%   32258   32258    71.80    0.001
    D                     1  168781        60.10%  168781  168781   375.69    0.000
  2-Way Interactions      1   78012        27.78%   78012   78012   173.65    0.000
    A*D                   1   78012        27.78%   78012   78012   173.65    0.000
Error                     4    1797         0.64%    1797     449
Total                     7  280848       100.00%
Model Summary
```

```
     S    R-sq  R-sq(adj)  PRESS  R-sq(pred)
21.1955  99.36%     98.88%   7188      97.44%
```

We see in the above analysis of variance that factors A, D, and AD are significant. Now we must analyze the residual plots.



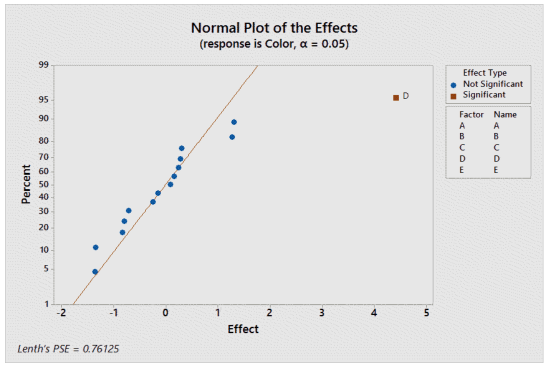
The versus fits give us some cause for concern as the left-most points hint at a possible inequality of variance. The normal probablity plot of effects is satisfactory and doesn't give us any reason to question our normality assumptions.

8.6 *R.D. ("Experimenting with a Large Number of Variables" in Experiments in Industry: Design, Analysis and Interpretation of Results, by R.D. Snee, L.B. Hare, and J.B. Trout, Editors ASQC, 1985) describes an experiment in which a* 2^{5-1} *design with* $I = ABCDE$ *was used to investigate the effects of five factors on the color of a chemical product The factors are* A = *solvent/reactant,* B = *catalyst/reactant,* C = *temperature,* D = *reactant purity, and* E = *reactant pH. The results obtained were as follows.*

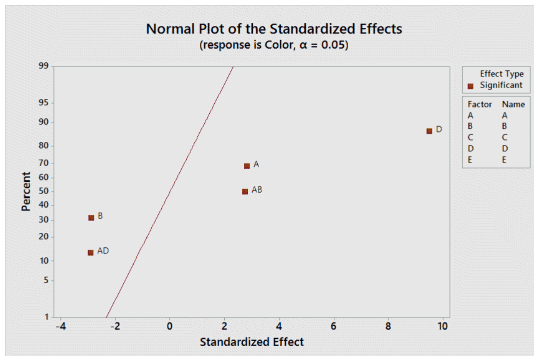
$$
\begin{array}{rr}
e = -0.63 & d = 6.79 \\
a = 2.51 & ade = 5.47 \\
b = -2.68 & bde = 3.45 \\
abe = 1.66 & abd = 5.68 \\
c = 2.06 & cde = 5.22 \\
ace = 1.22 & acd = 4.38 \\
bce = -2.09 & bcd = 4.30 \\
abc = 1.93 & abcde = 4.05
\end{array}
$$

a) *Prepare a normal probability plot of the effects. Which effects seem active?*

We see from our intial analysis that factors A, B, D, AB, and AD appear to be active (significant), however factor A is the only factor to register as such in this model.



If we re-analyze the model focusing on only these factors, the following normal probablity plot of effects is generated.



We see now that factors A, B, D, AB, and AD all register as significant in the reduced model. This is supported by the following analysis of variance performed using the reduced model.

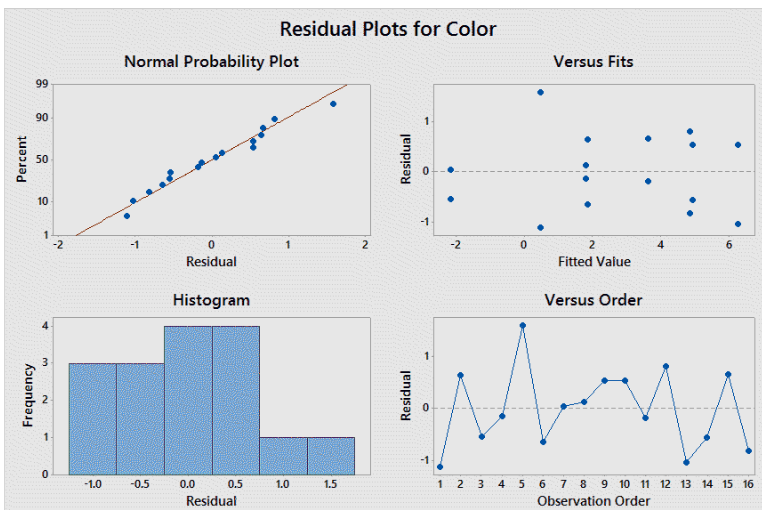
Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Model	5	106.039	92.46%	106.039	21.2078	24.53	0.000
Linear	3	92.192	80.39%	92.192	30.7308	35.54	0.000
A	1	6.864	5.99%	6.864	6.8644	7.94	0.018
B	1	7.182	6.26%	7.182	7.1824	8.31	0.016
D	1	78.146	68.14%	78.146	78.1456	90.37	0.000
2-Way Interactions	2	13.847	12.07%	13.847	6.9233	8.01	0.008
A*B	1	6.503	5.67%	6.503	6.5025	7.52	0.021
A*D	1	7.344	6.40%	7.344	7.3441	8.49	0.015
Error	10	8.647	7.54%	8.647	0.8647		
Total	15	114.686	100.00%				

b) *Calculate the residuals. Construct a normal probability plot of the residuals and plot the residuals versus the fitted values. Comment on the plots.* The following residuals and fits were calculated using Minitab.

Treatment Combination	Color	FITS	RESI
E	-0.63	0.4725	-1.1025
A	2.51	1.8625	0.6475
B	-2.68	-2.1425	-0.5375
ABE	1.66	1.7975	-0.1375
C	2.06	0.4725	1.5875
ACE	1.22	1.8625	-0.6425
BCE	-2.09	-2.1425	0.0525
ABC	1.93	1.7975	0.1325
D	6.79	6.2475	0.5425
ADE	5.47	4.9275	0.5425
BDE	3.45	3.6325	-0.1825
ABD	5.68	4.8625	0.8175
CDE	5.22	6.2475	-1.0275
ACD	4.38	4.9275	-0.5475
BCD	4.30	3.6325	0.6675
ABCD	4.05	4.8625	-0.8125

These residuals were used to construct the following graphs.



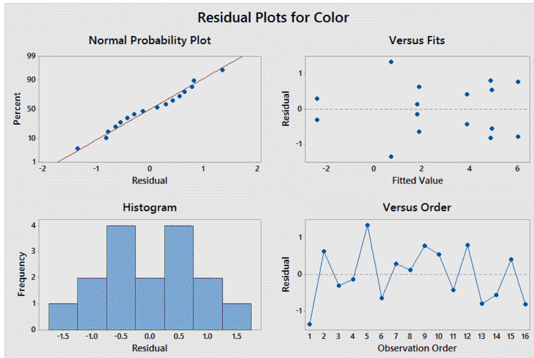
We see that the residual plots are satisfactory and do not give us any reason to question our normality and equal variance assumptions.

c) *If any factors are negligible, collapse the 2^{5-1} design into a full factorial in the active factors. Comment on the resulting design, and interpret the results.*

Since we found factors A,B,D, AB, and AD significant, we can collapse the 2^{5-1} into a full factorial 2^3. The results of this collapse are shown below. It is possible to relabel D as C, however we have chosen to keep the original labeling format for consistentcy.

```
Analysis of Variance

Source                DF   Seq SS  Contribution   Adj SS   Adj MS  F-Value  P-Value
Model                  7  106.510        92.87%  106.510  15.2156    14.89    0.001
  Linear               3   92.192        80.39%   92.192  30.7308    30.07    0.000
    A                  1    6.864         5.99%    6.864   6.8644     6.72    0.032
    B                  1    7.182         6.26%    7.182   7.1824     7.03    0.029
    D                  1   78.146        68.14%   78.146  78.1456    76.46    0.000
  2-Way Interactions   3   14.087        12.28%   14.087   4.6956     4.59    0.038
    A*B                1    6.503         5.67%    6.503   6.5025     6.36    0.036
    A*D                1    7.344         6.40%    7.344   7.3441     7.19    0.028
    B*D                1    0.240         0.21%    0.240   0.2401     0.23    0.641
  3-Way Interactions   1    0.230         0.20%    0.230   0.2304     0.23    0.648
    A*B*D              1    0.230         0.20%    0.230   0.2304     0.23    0.648
Error                  8    8.177         7.13%    8.177   1.0221
Total                 15  114.686       100.00%


Model Summary
      S    R-sq  R-sq(adj)    PRESS  R-sq(pred)
1.01099  92.87%     86.63%  32.7072      71.48%

Coded Coefficients
Term      Effect    Coef  SE Coef       95% CI        T-Value  P-Value   VIF
Constant           2.707    0.253  ( 2.125,  3.290)    10.71    0.000
A          1.310   0.655    0.253  ( 0.072,  1.238)     2.59    0.032  1.00
B         -1.340  -0.670    0.253  (-1.253, -0.087)    -2.65    0.029  1.00
D          4.420   2.210    0.253  ( 1.627,  2.793)     8.74    0.000  1.00
A*B        1.275   0.638    0.253  ( 0.055,  1.220)     2.52    0.036  1.00
A*D       -1.355  -0.677    0.253  (-1.260, -0.095)    -2.68    0.028  1.00
B*D        0.245   0.123    0.253  (-0.460,  0.705)     0.48    0.641  1.00
A*B*D     -0.240  -0.120    0.253  (-0.703,  0.463)    -0.47    0.648  1.00

Regression Equation in Uncoded Units
Color = 2.707 + 0.655 A - 0.670 B + 2.210 D + 0.638 A*B - 0.677 A*D + 0.123 B*D - 0.120 A*B*D
```

We see that the same factors are active in the collapsed model as were in the original model, though the SS_E has decreased slightly to accomodate the addition of $B, BD, andABD$.
Furthermore, we note that in the residual plots included below there is no significant cause for concern. However, the normal probability plot is potentially exhibiting a pattern which could indicate curvature that was unaccounted for. This should be looked into further to determine if the model cannot be made to be more accurate.



8.7 *An article written by J.J. Pignatiello Jr. and J.S. Ramberg in the Journal of Quality Technology (Vol. 17, 1985, pp. 198-206) describes the use of a replicated fractional factorial to investigate the effect of five factors on the free height of leaf springs used in an automotive application. The factors are A = furnace temperature, B = heating time, C = transfer time, D = hold down time, and E = quench oil temperature. The data are shown in the table below.*

A	B	C	D	E	Free Height		
-	-	-	-	-	7.78	7.78	7.81
+	-	-	+	-	8.15	8.18	7.88
-	+	-	+	-	7.50	7.56	7.50
+	+	-	-	-	7.59	7.56	7.75
-	-	+	+	-	7.54	8.00	7.88
+	-	+	-	-	7.69	8.09	8.06
-	+	+	-	-	7.56	7.52	7.44
+	+	+	+	-	7.56	7.81	7.69
-	-	-	-	+	7.50	7.25	7.12
+	-	-	+	+	7.88	7.88	7.44
-	+	-	+	+	7.50	7.56	7.50
+	+	-	-	+	7.63	7.75	7.56
-	-	+	+	+	7.32	7.44	7.44
+	-	+	-	+	7.56	7.69	7.62
-	+	+	-	+	7.18	7.18	7.25
+	+	+	+	+	7.81	7.50	7.59

a) *Write out the alias structure for this design. What is the resolution of this design?*

The defining relation of this design is $I = ABCD$ which means that is Resolution IV. The alias structure is as follows.

$$I+ABCD$$
$$A+BCD$$
$$B+ACD$$
$$C+ABD$$
$$D+ABC$$
$$E+ABCDE$$
$$AB+CD$$
$$AC+BD$$
$$AD+BC$$
$$AE+BCDE$$
$$BD+ACDE$$
$$CE+ABDE$$
$$DE+ABCE$$

b) *Analyze the data. What factors influence the mean free height?*

We see from the below intial analysis of variance with two-factor effects that the factors A, B, D, E, and BE are significant influences on the mean free height.

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Model	12	2.21646	76.67%	2.21646	0.184705	9.59	0.000
Linear	5	1.83846	63.60%	1.83846	0.367692	19.08	0.000
A	1	0.70325	24.33%	0.70325	0.703252	36.50	0.000
B	1	0.32177	11.13%	0.32177	0.321769	16.70	0.000
C	1	0.02950	1.02%	0.02950	0.029502	1.53	0.224
D	1	0.09992	3.46%	0.09992	0.099919	5.19	0.029
E	1	0.68402	23.66%	0.68402	0.684019	35.50	0.000
2-Way Interactions	7	0.37800	13.08%	0.37800	0.054000	2.80	0.020
A*B	1	0.01050	0.36%	0.01050	0.010502	0.55	0.465
A*C	1	0.00002	0.00%	0.00002	0.000019	0.00	0.975
A*D	1	0.00630	0.22%	0.00630	0.006302	0.33	0.571
A*E	1	0.04877	1.69%	0.04877	0.048769	2.53	0.121
B*E	1	0.28060	9.71%	0.28060	0.280602	14.56	0.001
C*E	1	0.01300	0.45%	0.01300	0.013002	0.67	0.417
D*E	1	0.01880	0.65%	0.01880	0.018802	0.98	0.330
Error	35	0.67432	23.33%	0.67432	0.019266		
Lack-of-Fit	3	0.04726	1.63%	0.04726	0.015752	0.80	0.501
Pure Error	32	0.62707	21.69%	0.62707	0.019596		
Total	47	2.89078	100.00%				

Therefore, we shall reanalyze the data to focus on significant factors and reduce noise.

Analysis of Variance

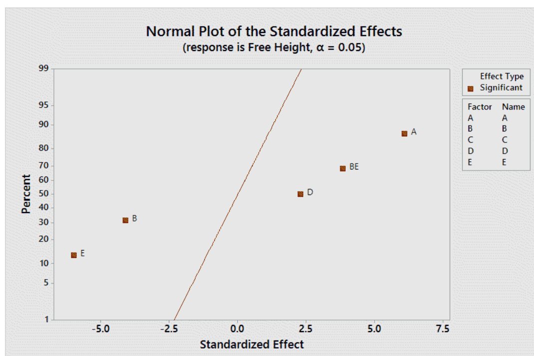
Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value

Model	5	2.08956	72.28%	2.08956	0.41791	21.91	0.000
Linear	4	1.80896	62.58%	1.80896	0.45224	23.71	0.000
A	1	0.70325	24.33%	0.70325	0.70325	36.86	0.000
B	1	0.32177	11.13%	0.32177	0.32177	16.87	0.000
D	1	0.09992	3.46%	0.09992	0.09992	5.24	0.027
E	1	0.68402	23.66%	0.68402	0.68402	35.86	0.000
2-Way Interactions	1	0.28060	9.71%	0.28060	0.28060	14.71	0.000
B*E	1	0.28060	9.71%	0.28060	0.28060	14.71	0.000
Error	42	0.80122	27.72%	0.80122	0.01908		
Lack-of-Fit	10	0.17415	6.02%	0.17415	0.01742	0.89	0.554
Pure Error	32	0.62707	21.69%	0.62707	0.01960		
Total	47	2.89078	100.00%				

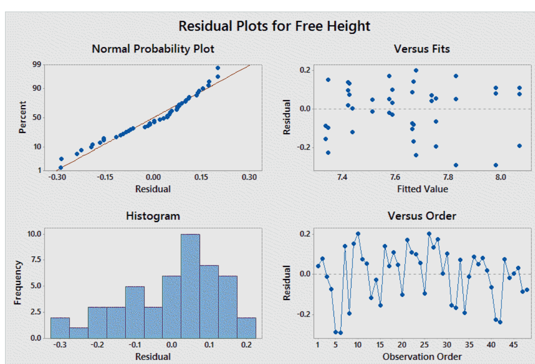
Model Summary

S	R-sq	R-sq(adj)	PRESS	R-sq(pred)
0.138118	72.28%	68.98%	1.04649	63.80%

It is clear that the factors A, B, D, E, and BE are all significant at the 5% level. Our normal probability plot of the data supports this conclusion.



Now, we must analyze the residuals for signs of abnormality.

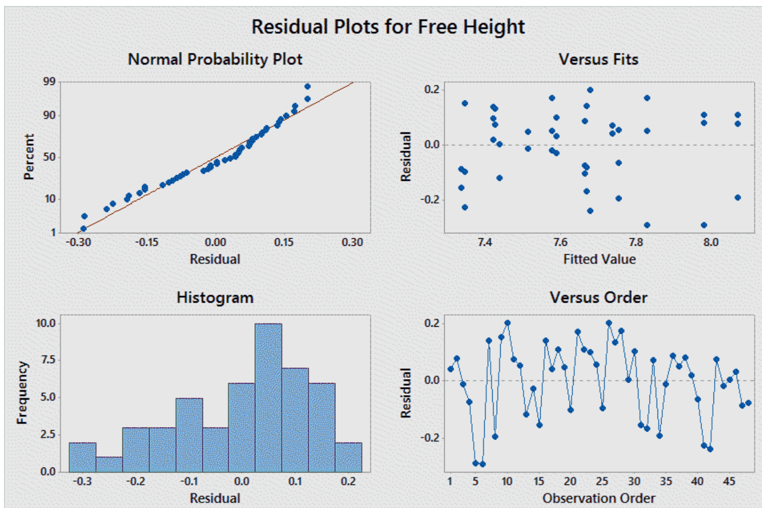


There is nothing particularly concerning regarding the residual plots and we have no reason to question our assumptions of normality and equal variances.

c) *Calculate the range and standard deviation of the free height for each run. Is there any indication that any of these factors affects variability in the free height.*

Term	Effect	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant		7.6256	0.0200	(7.5850, 7.6663)	380.62	0.000	
A	0.2421	0.1210	0.0200	(0.0804, 0.1617)	6.04	0.000	1.00
B	-0.1638	-0.0819	0.0200	(-0.1225, -0.0412)	-4.09	0.000	1.00
C	-0.0496	-0.0248	0.0200	(-0.0655, 0.0159)	-1.24	0.224	1.00
D	0.0913	0.0456	0.0200	(0.0050, 0.0863)	2.28	0.029	1.00
E	-0.2387	-0.1194	0.0200	(-0.1600, -0.0787)	-5.96	0.000	1.00
A*B	-0.0296	-0.0148	0.0200	(-0.0555, 0.0259)	-0.74	0.465	1.00
A*C	0.0013	0.0006	0.0200	(-0.0400, 0.0413)	0.03	0.975	1.00
A*D	-0.0229	-0.0115	0.0200	(-0.0521, 0.0292)	-0.57	0.571	1.00
A*E	0.0637	0.0319	0.0200	(-0.0088, 0.0725)	1.59	0.121	1.00
B*E	0.1529	0.0765	0.0200	(0.0358, 0.1171)	3.82	0.001	1.00
C*E	-0.0329	-0.0165	0.0200	(-0.0571, 0.0242)	-0.82	0.417	1.00
D*E	0.0396	0.0198	0.0200	(-0.0209, 0.0605)	0.99	0.330	1.00

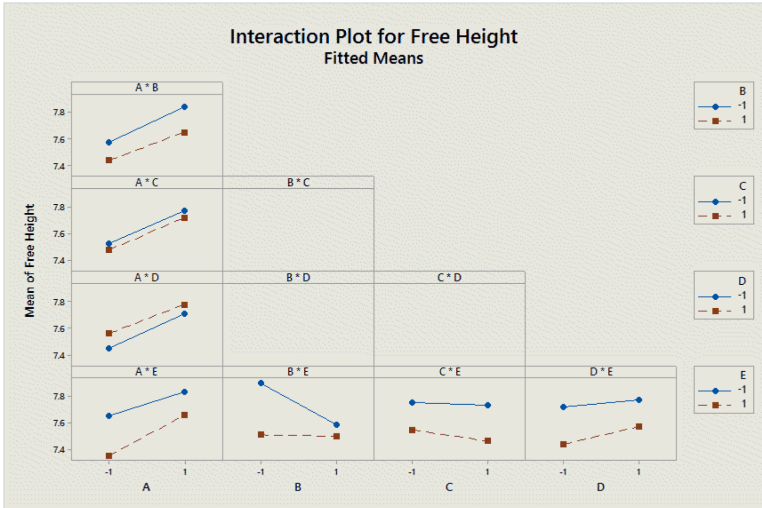
d) *Analyze the residuals from this experiment, and comment on your findings.* The following residual plots do not show anything of particular concern. They appear to be satisifactory.



e) *Is this the best possible design for five factors in 16 runs? Specifically, can you find a fractional design for five factors in 16 runs with a higher resolution than this one?*

No. This is not the best possible design. A design with 5 factors (i.e. 2^5 factorial) can have a maximum resolution of V. This can be acheived by setting the generator to $I = ABCDE$, a five-letter word.

8.9 *Consider the leaf spring experiment in Problem 8.7. Suppose that factor E (quench oil temperature) is very difficult to control during manufacturing. Where would you set factors A, B, C, and D to reduce variability in the free height as much as possible regardless of the quench oil temperature?*



Wee see from the above interaction of effects graph that we should run the process with A high, B and C low, and D at whichever level is more economical. This will reduce the variability in the free height as much as possible regardless of where E is set.