CSCI 688
Homework 5

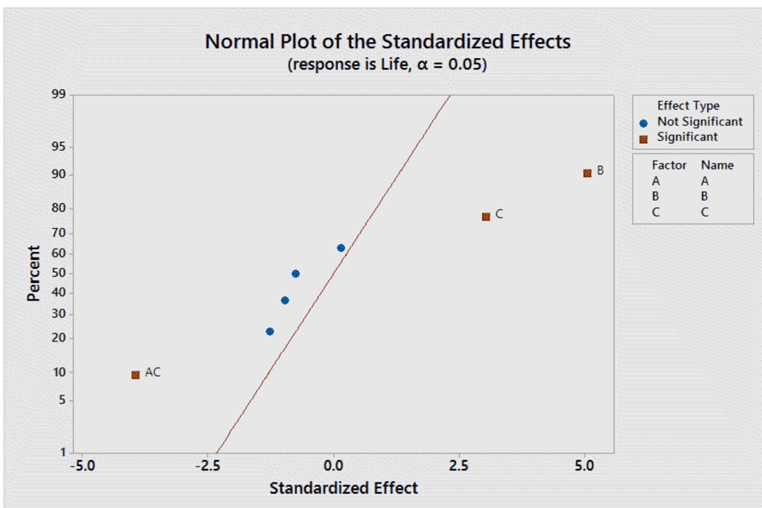
Megan Rose Bryant
Department of Mathematics
William and Mary

April 24, 2015

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	B	C	Treatment Combination	Replicate I	Replicate II	Replicate III
-	-	-	(1)	22	31	25
+	-	-	a	32	43	29
-	+	-	b	35	34	50
+	+	-	ab	55	47	46
-	-	+	c	44	45	38
+	-	+	ac	40	37	36
-	+	+	bc	60	50	54
+	+	+	abc	39	41	47

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



We can tell that the effects AC, B, and C are most likely to be significant since they are the farthest from the distribution line.

Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		40.83	1.12	36.42	0.000	
A	0.33	0.17	1.12	0.15	0.884	1.00
B	11.33	5.67	1.12	5.05	0.000	1.00
C	6.83	3.42	1.12	3.05	0.008	1.00
A*B	-1.67	-0.83	1.12	-0.74	0.468	1.00
A*C	-8.83	-4.42	1.12	-3.94	0.001	1.00
B*C	-2.83	-1.42	1.12	-1.26	0.224	1.00
A*B*C	-2.17	-1.08	1.12	-0.97	0.348	1.00

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

```
Factorial Regression: Life versus A, B, C

Analysis of Variance

Source               DF   Adj SS   Adj MS  F-Value  P-Value
Model                 7  1612.67  230.381     7.64    0.000
  Linear              3  1051.50  350.500    11.62    0.000
    A                 1     0.67    0.667     0.02    0.884
    B                 1   770.67  770.667    25.55    0.000
    C                 1   280.17  280.167     9.29    0.008
  2-Way Interactions  3   533.00  177.667     5.89    0.007
    A*B               1    16.67   16.667     0.55    0.468
    A*C               1   468.17  468.167    15.52    0.001
    B*C               1    48.17   48.167     1.60    0.224
  3-Way Interactions  1    28.17   28.167     0.93    0.348
    A*B*C             1    28.17   28.167     0.93    0.348
Error                16   482.67   30.167
Total                23  2095.33


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
5.49242  76.96%     66.89%      48.17%


Coded Coefficients

Term      Effect   Coef  SE Coef  T-Value  P-Value   VIF
Constant          40.83     1.12    36.42    0.000
A           0.33   0.17     1.12     0.15    0.884  1.00
B          11.33   5.67     1.12     5.05    0.000  1.00
C           6.83   3.42     1.12     3.05    0.008  1.00
A*B        -1.67  -0.83     1.12    -0.74    0.468  1.00
A*C        -8.83  -4.42     1.12    -3.94    0.001  1.00
B*C        -2.83  -1.42     1.12    -1.26    0.224  1.00
A*B*C      -2.17  -1.08     1.12    -0.97    0.348  1.00
```

We see from the analysis of variance that B, C, and AC are significant with p-values of $0.000, 0.008$, and 0.001, respectively. This supports our conclusions from part (a).
This reduces to the following model:

```
Factorial Regression: Life versus A, B, C

Analysis of Variance

Source               DF   Adj SS   Adj MS  F-Value  P-Value
Model                 4  1519.67  379.917    12.54    0.000
  Linear              3  1051.50  350.500    11.57    0.000
    A                 1     0.67    0.667     0.02    0.884
```

B	1	770.67	770.667	25.44	0.000
C	1	280.17	280.167	9.25	0.007
2-Way Interactions	1	468.17	468.167	15.45	0.001
A*C	1	468.17	468.167	15.45	0.001
Error	19	575.67	30.298		
Lack-of-Fit	3	93.00	31.000	1.03	0.407
Pure Error	16	482.67	30.167		
Total	23	2095.33			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
5.50438	72.53%	66.74%	56.16%

Coded Coefficients

Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		40.83	1.12	36.34	0.000	
A	0.33	0.17	1.12	0.15	0.884	1.00
B	11.33	5.67	1.12	5.04	0.000	1.00
C	6.83	3.42	1.12	3.04	0.007	1.00
A*C	-8.83	-4.42	1.12	-3.93	0.001	1.00

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

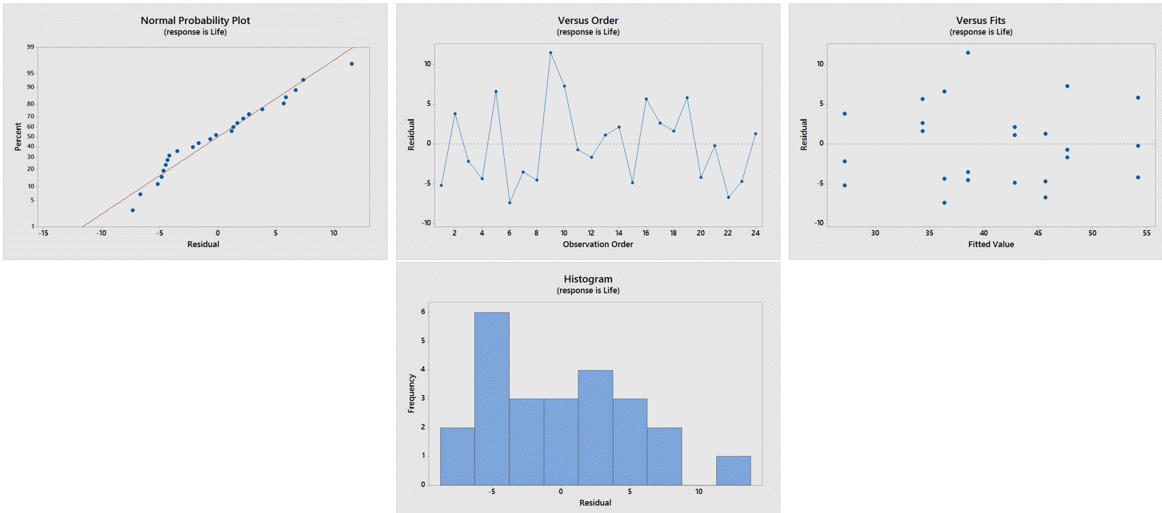
Coded Coefficients

Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		40.83	1.12	36.34	0.000	
A	0.33	0.17	1.12	0.15	0.884	1.00
B	11.33	5.67	1.12	5.04	0.000	1.00
C	6.83	3.42	1.12	3.04	0.007	1.00
A*C	-8.83	-4.42	1.12	-3.93	0.001	1.00

Regression Equation in Uncoded Units

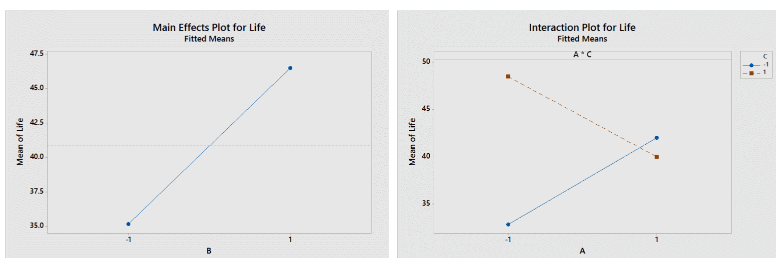
Life = 40.83 + 0.17 A + 5.67 B + 3.42 C - 4.42 A*C

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



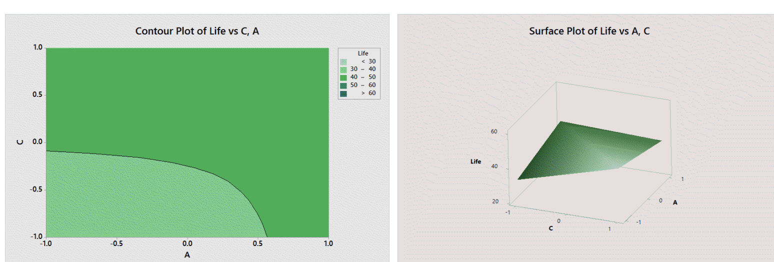
There is nothing in the residual plots that make us question our assumptions.

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.*



We can see from the main effect plot that factor B is having a positive effect, which means that we should have B at a high level. We see from the interaction affect that life is maximized when A is at the low level and C is at the high level. Therefore, to maximize life, we must set B at high, A at low, and C at high.

2. *Reconsider part (c) of Problem 6.1. Use the regression model to generate response surface and conour plots of the tool life response Interpret these plots. Do they provide insight regarding the desirable operating conditions fo this process.*



3. *Find the standard error of the factor effects and approximate 95% 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?*

$$SE_{Effect} = \sqrt{\frac{1}{n2^{k-2}}S^2} = \sqrt{\frac{1}{(3)(2^{3-2})} * 30.167} = 2.242$$

Term	Effect	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant		40.83	1.12	(38.46, 43.21)	36.42	0.000	
A	0.33	0.17	1.12	(-2.21, 2.54)	0.15	0.884	1.00
B	11.33	5.67	1.12	(3.29, 8.04)	5.05	0.000	1.00
C	6.83	3.42	1.12	(1.04, 5.79)	3.05	0.008	1.00
A*B	-1.67	-0.83	1.12	(-3.21, 1.54)	-0.74	0.468	1.00
A*C	-8.83	-4.42	1.12	(-6.79, -2.04)	-3.94	0.001	1.00
B*C	-2.83	-1.42	1.12	(-3.79, 0.96)	-1.26	0.224	1.00
A*B*C	-2.17	-1.08	1.12	(-3.46, 1.29)	-0.97	0.348	1.00

These minitab generated confidence intervals concur with our previous conclusions since 0 is not in the 95% confidence intervals of terms B, C, and AC, which shows that they are all significant factors.

7. *An experiment was performed to improve the yield of a chemical process. Four factors were selected, and two replicates of a completely randomized experiment were run. The results are shown in the following table.*

a.) *Estimate the factor of effects.*

Coded Coefficients

Term	Effect	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant		82.781	0.489	(81.744, 83.818)	169.24	0.000	
A	-9.063	-4.531	0.489	(-5.568, -3.494)	-9.26	0.000	1.00
B	-1.313	-0.656	0.489	(-1.693, 0.381)	-1.34	0.198	1.00
C	-2.687	-1.344	0.489	(-2.381, -0.307)	-2.75	0.014	1.00
D	3.938	1.969	0.489	(0.932, 3.006)	4.02	0.001	1.00
A*B	4.062	2.031	0.489	(0.994, 3.068)	4.15	0.001	1.00
A*C	0.688	0.344	0.489	(-0.693, 1.381)	0.70	0.492	1.00
A*D	-2.188	-1.094	0.489	(-2.131, -0.057)	-2.24	0.040	1.00
B*C	-0.562	-0.281	0.489	(-1.318, 0.756)	-0.57	0.573	1.00
B*D	-0.188	-0.094	0.489	(-1.131, 0.943)	-0.19	0.850	1.00
C*D	1.687	0.844	0.489	(-0.193, 1.881)	1.72	0.104	1.00
A*B*C	-5.187	-2.594	0.489	(-3.631, -1.557)	-5.30	0.000	1.00
A*B*D	4.687	2.344	0.489	(1.307, 3.381)	4.79	0.000	1.00
A*C*D	-0.937	-0.469	0.489	(-1.506, 0.568)	-0.96	0.352	1.00
B*C*D	-0.937	-0.469	0.489	(-1.506, 0.568)	-0.96	0.352	1.00
A*B*C*D	2.437	1.219	0.489	(0.182, 2.256)	2.49	0.024	1.00

b.) *Prepare an analysis of variance table and determine which factors are important in explaining yield.*

Full Factorial Design

```
Factors:   4   Base Design:        4, 16uns:     32   Replicates:              2
Blocks:    1   Center pts (total):      0
```

```
All terms are free from aliasing.

Factorial Regression: Yield versus A, B, C, D

Analysis of Variance
```

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Model	15	1504.97	92.47%	1504.97	100.331	13.10	0.000
Linear	4	852.63	52.39%	852.63	213.156	27.84	0.000
A	1	657.03	40.37%	657.03	657.031	85.82	0.000
B	1	13.78	0.85%	13.78	13.781	1.80	0.198
C	1	57.78	3.55%	57.78	57.781	7.55	0.014
D	1	124.03	7.62%	124.03	124.031	16.20	0.001
2-Way Interactions	6	199.69	12.27%	199.69	33.281	4.35	0.009
A*B	1	132.03	8.11%	132.03	132.031	17.24	0.001
A*C	1	3.78	0.23%	3.78	3.781	0.49	0.492
A*D	1	38.28	2.35%	38.28	38.281	5.00	0.040
B*C	1	2.53	0.16%	2.53	2.531	0.33	0.573
B*D	1	0.28	0.02%	0.28	0.281	0.04	0.850
C*D	1	22.78	1.40%	22.78	22.781	2.98	0.104
3-Way Interactions	4	405.12	24.89%	405.12	101.281	13.23	0.000
A*B*C	1	215.28	13.23%	215.28	215.281	28.12	0.000
A*B*D	1	175.78	10.80%	175.78	175.781	22.96	0.000
A*C*D	1	7.03	0.43%	7.03	7.031	0.92	0.352
B*C*D	1	7.03	0.43%	7.03	7.031	0.92	0.352
4-Way Interactions	1	47.53	2.92%	47.53	47.531	6.21	0.024
A*B*C*D	1	47.53	2.92%	47.53	47.531	6.21	0.024
Error	16	122.50	7.53%	122.50	7.656		
Total	31	1627.47	100.00%				

```
Model Summary
```

S	R-sq	R-sq(adj)	PRESS	R-sq(pred)
2.76699	92.47%	85.42%	490	69.89%

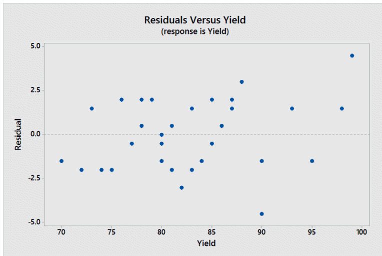
Based on the analysis of variance, we can conclude that factors A, C, D, AB,AD,ABC, ABD, and ABCD were all significant since they had p-values less than our alpha of 0.05.

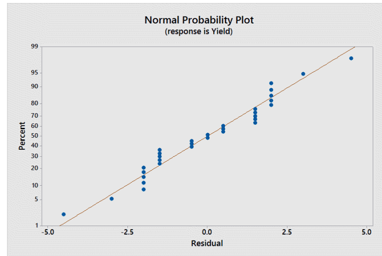
c.) *Write down a regression model for predicting yield, assuming that all four factors were varied over the range from* -1 *to* $+1$ *(in coded units).*

```
Regression Equation in Uncoded Units

Yield = 82.781 - 4.531 A - 0.656 B - 1.344 C + 1.969 D + 2.031 A*B + 0.344 A*C - 1.094 A*D
        - 0.281 B*C - 0.094 B*D + 0.844 C*D - 2.594 A*B*C + 2.344 A*B*D - 0.469 A*C*D
        - 0.469 B*C*D + 1.219 A*B*C*D
```

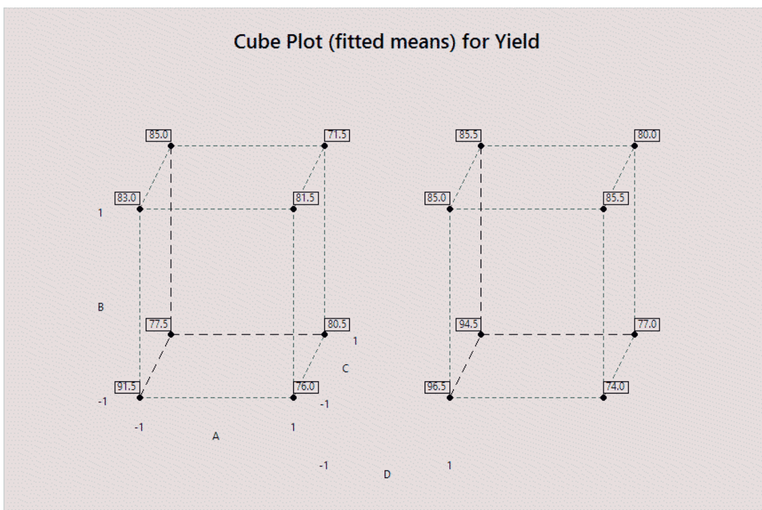
d.) *Plot the residuals versus the predicted yield and on a normal probability scale. Does the residual analysis appear satisfactory.*





We see one outlier in the Residuals Versus Yield plot which could skew the results very slightly. Since it is only one outlier, however, we can continue with our analysis.

e.) *Two three-factor interactions, ABC and ABD, apparently have large effects Draw a cube plot in the factors A, B, and C with the average yields shown at each corner. Repeat using the factors A, B, and D. Do these two plots aid in data interpretation? Where would you recommend that the process be run with respect to the four variables?*



So, we would run the process at A low, B low, C low, and D high for a yield of 94.5.

15. *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 affect of four factors on cracks. The four factors are pouring temperature (A), titanium content (B), heat treatment method (C), and amount of grain refiner (D). Two replicates of a 2^4 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 Table P6.2.*

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

Coded Coefficients

Term	Effect	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant		11.9881	0.0504	(11.8813, 12.0948)	238.04	0.000	
A	3.0189	1.5094	0.0504	(1.4027, 1.6162)	29.97	0.000	1.00
B	3.9759	1.9879	0.0504	(1.8812, 2.0947)	39.47	0.000	1.00
C	-3.5962	-1.7981	0.0504	(-1.9049, -1.6914)	-35.70	0.000	1.00
D	1.9577	0.9789	0.0504	(0.8721, 1.0856)	19.44	0.000	1.00
A*B	1.9341	0.9671	0.0504	(0.8603, 1.0738)	19.20	0.000	1.00
A*C	-4.0077	-2.0039	0.0504	(-2.1106, -1.8971)	-39.79	0.000	1.00
A*D	0.0765	0.0383	0.0504	(-0.0685, 0.1450)	0.76	0.459	1.00
B*C	0.0960	0.0480	0.0504	(-0.0588, 0.1548)	0.95	0.355	1.00
B*D	0.0473	0.0236	0.0504	(-0.0831, 0.1304)	0.47	0.645	1.00
C*D	-0.0769	-0.0384	0.0504	(-0.1452, 0.0683)	-0.76	0.456	1.00
A*B*C	3.1375	1.5687	0.0504	(1.4620, 1.6755)	31.15	0.000	1.00
A*B*D	0.0980	0.0490	0.0504	(-0.0578, 0.1558)	0.97	0.345	1.00
A*C*D	0.0191	0.0096	0.0504	(-0.0972, 0.1163)	0.19	0.852	1.00
B*C*D	0.0356	0.0178	0.0504	(-0.0889, 0.1246)	0.35	0.728	1.00
A*B*C*D	0.0141	0.0071	0.0504	(-0.0997, 0.1138)	0.14	0.890	1.00

A, B, C, D, AB, AC, ABC all are appear to be signficant.

b.) *Conduct an analysis of variance. Do any of the factors affect cracking? Use* $\alpha = 0.05$.

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Model	15	570.948	99.77%	570.948	38.063	468.99	0.000
Linear	4	333.496	58.28%	333.496	83.374	1027.28	0.000
A	1	72.909	12.74%	72.909	72.909	898.34	0.000
B	1	126.461	22.10%	126.461	126.461	1558.17	0.000
C	1	103.464	18.08%	103.464	103.464	1274.82	0.000
D	1	30.662	5.36%	30.662	30.662	377.80	0.000
2-Way Interactions	6	158.609	27.72%	158.609	26.435	325.71	0.000
A*B	1	29.927	5.23%	29.927	29.927	368.74	0.000
A*C	1	128.496	22.45%	128.496	128.496	1583.26	0.000
A*D	1	0.047	0.01%	0.047	0.047	0.58	0.459
B*C	1	0.074	0.01%	0.074	0.074	0.91	0.355
B*D	1	0.018	0.00%	0.018	0.018	0.22	0.645
C*D	1	0.047	0.01%	0.047	0.047	0.58	0.456
3-Way Interactions	4	78.841	13.78%	78.841	19.710	242.86	0.000
A*B*C	1	78.751	13.76%	78.751	78.751	970.33	0.000
A*B*D	1	0.077	0.01%	0.077	0.077	0.95	0.345
A*C*D	1	0.003	0.00%	0.003	0.003	0.04	0.852
B*C*D	1	0.010	0.00%	0.010	0.010	0.13	0.728
4-Way Interactions	1	0.002	0.00%	0.002	0.002	0.02	0.890
A*B*C*D	1	0.002	0.00%	0.002	0.002	0.02	0.890
Error	16	1.299	0.23%	1.299	0.081		
Total	31	572.246	100.00%				

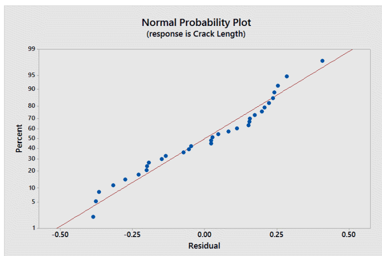
As we see above, A, B, C, D, AB, AC, and ABC all appear to be significant because they have p-values less than 0.05, our alpha.

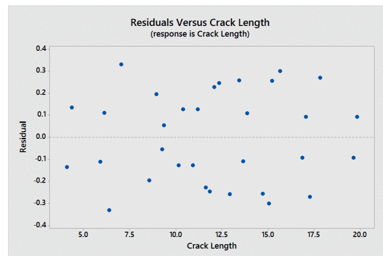
c.) *Write down a regression model that can be used to predict crack length as a function of the significant main effects and interactions you have identified in part (b).*
A reduced regression model is

```
Regression Equation in Uncoded Units

Crack Length = 11.9881 + 1.5094 A + 1.9879 B - 1.7981 C + 0.9789 D + 0.9671 A*B - 2.0039 A*C
               + 0.0480 B*C + 1.5687 A*B*C
```

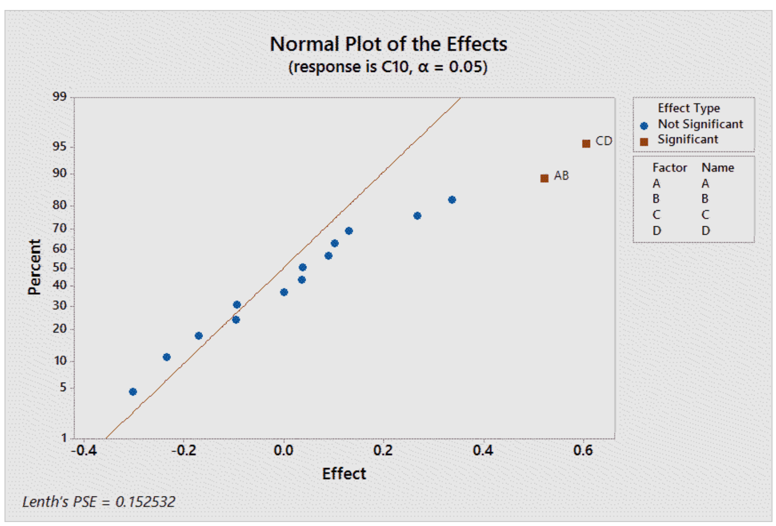
d.) *Analyze the residuals from this experiment.*





Both the normal probability plot and the versus plot don't give us any reason to question our assumptions.

e.) *Is there an indication that any of the factors affect the variability in cracking.*
First, we must analyze the variability of the factor replicates.



The normal probability plot of the effects indicates that the factors AB and CD are significant. Now, we must conduct an analysis of the variance of the variability.

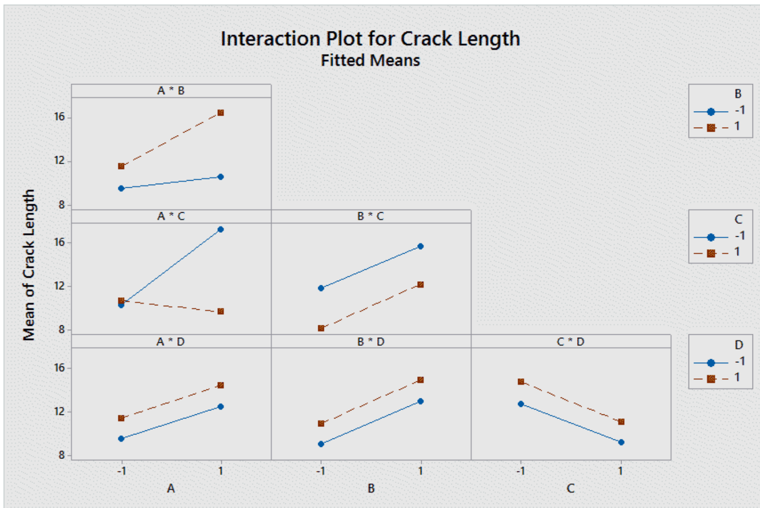
Analysis of Variance

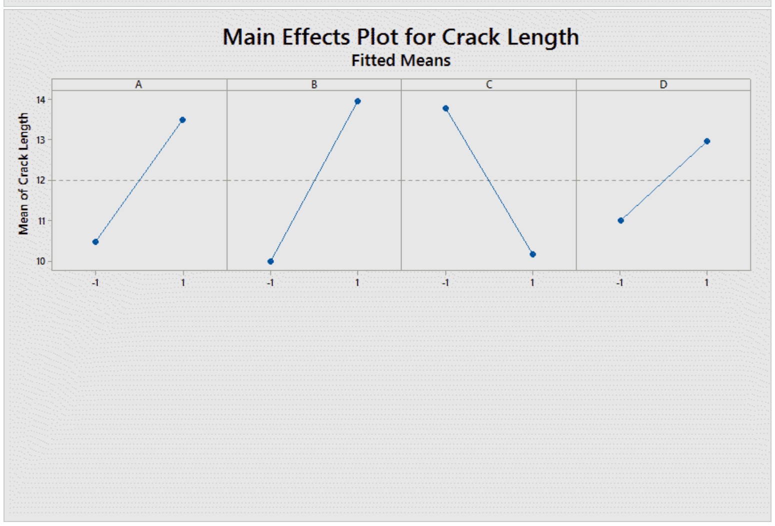
Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Model	2	0.14386	63.81%	0.14386	0.071931	11.46	0.001
2-Way Interactions	2	0.14386	63.81%	0.14386	0.071931	11.46	0.001
A*B	1	0.06266	27.79%	0.06266	0.062658	9.98	0.008
C*D	1	0.08120	36.02%	0.08120	0.081204	12.94	0.003

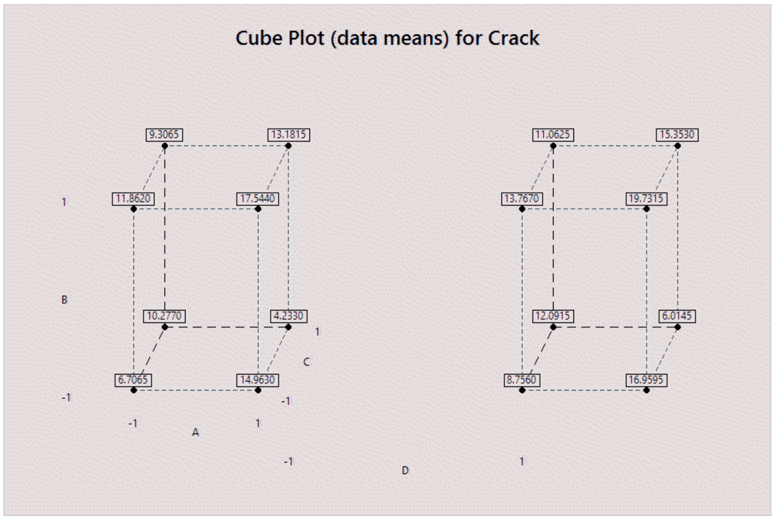
Error	13	0.08158	36.19%	0.08158	0.006275
Total	15	0.22544	100.00%		

The ANOVA verifies that AB and CD are both significant factors on cracking variability.

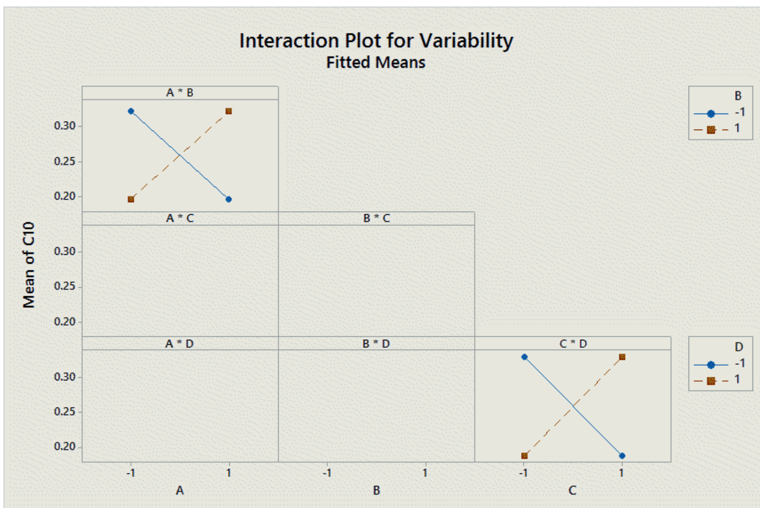
f.) *What recommendations would you make regarding process operations? Use interaction and/or main effect plots to assist in drawing conclusions.*
We need to examine the Main Effects, Interaction, and Cube plots of the crack length.







Now, we should look at the interaction plots for crack length variability.



We see from the plots on crack length that A, B, and C should be set at the high level and D can be either high or low. However, from the variability of crack length plots, we see that C should be set at high and D should be set at low. This will minimize crack length and variability.

16. *One of the variables in the experiment described in Problem 6.15, heat treatment method (C), is a categorical variable. Assume that the remanining factors are continuous.*

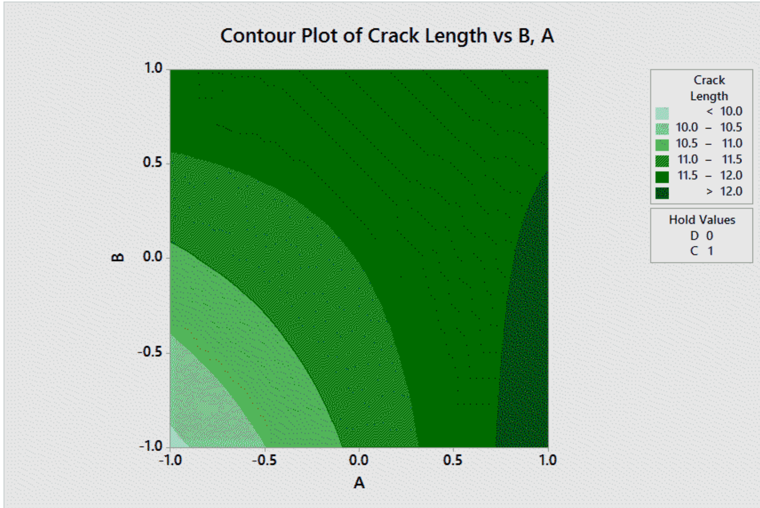
a.) *Write two regression models for predicting crack length, one for each level of the heat treatment method variable. What differences, if any, do you notice in these two equations?*

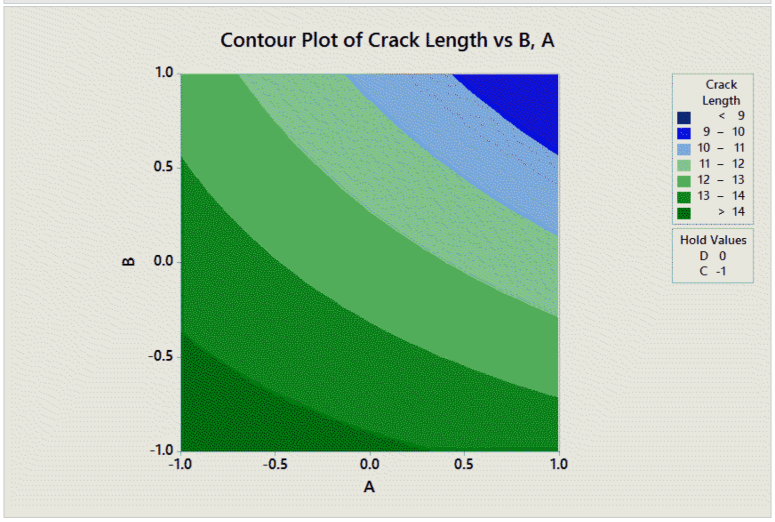
```
Regression Equation

C
-1   C9 = 12.47 - 0.270 A - 0.651 B + 0.916 D

1    C9 = 11.51 - 0.270 A - 0.651 B + 0.916 D
```

b.) *Generate appropriate response surface contour plots for the two regression models in part (a).*





c.) *What set of conditions would you recommend for the factors A, B, and D if you use heat treatment method $C = +$.*

If we were using C high, then the best choice would be to have A high, B low, and D low.

d.) *Repeat part (c) assuming that you wish to use heat treatment method $C = -$.*

If we were using C low, then the best choice would be to have A low, B low, and D low.

20. *Semiconductor manufacturing process 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 highes reading. A DMX Verifier was used to measure UEC. The data from this experiment are shown in Table P6.5.*

a.) *Analyze the data from this experiment. What factors significantly affect UEC?*
We see from the following normal probability plot of effects that the factors A, C, D, and AC are significant. This is supported by the following effects table from minitab.

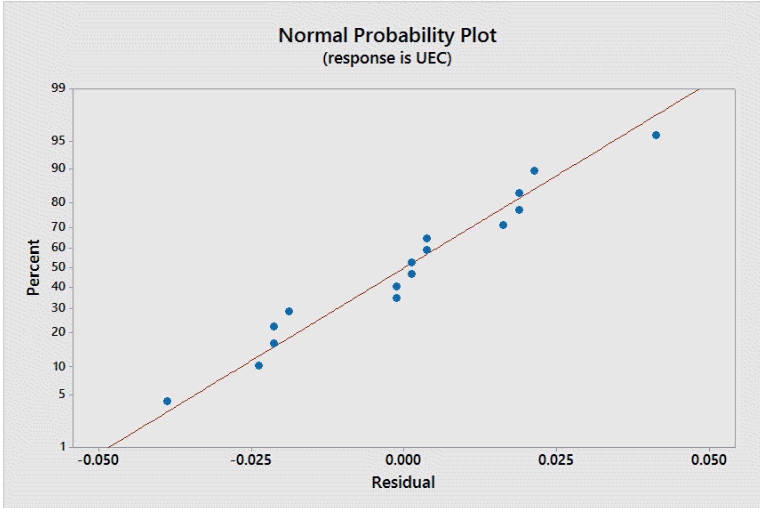
Coded Coefficients

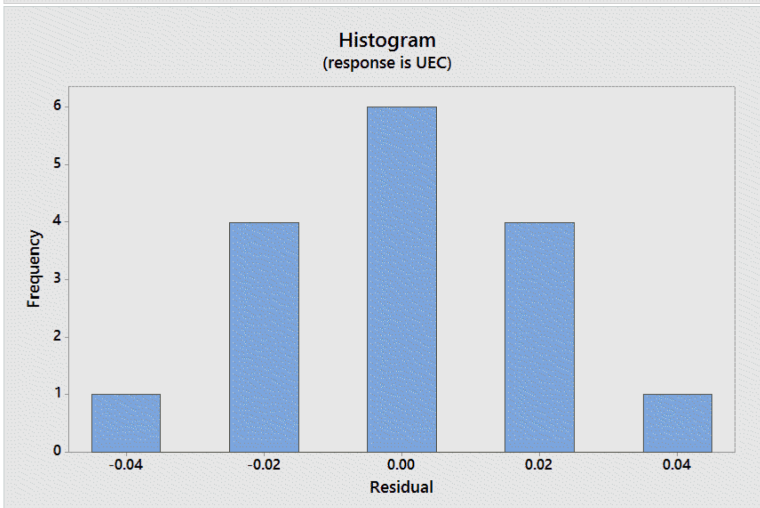
Term	Effect	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant		0.71625	0.00900	(0.69312, 0.73938)	79.61	0.000	
A	0.16000	0.08000	0.00900	(0.05687, 0.10313)	8.89	0.000	1.00
B	0.02000	0.01000	0.00900	(-0.01313, 0.03313)	1.11	0.317	1.00
C	-0.13250	-0.06625	0.00900	(-0.08938, -0.04312)	-7.36	0.001	1.00
D	-0.11250	-0.05625	0.00900	(-0.07938, -0.03312)	-6.25	0.002	1.00
A*B	0.01750	0.00875	0.00900	(-0.01438, 0.03188)	0.97	0.375	1.00
A*C	-0.05500	-0.02750	0.00900	(-0.05063, -0.00437)	-3.06	0.028	1.00
A*D	-0.01000	-0.00500	0.00900	(-0.02813, 0.01813)	-0.56	0.602	1.00
B*C	0.02500	0.01250	0.00900	(-0.01063, 0.03563)	1.39	0.223	1.00
B*D	-0.01000	-0.00500	0.00900	(-0.02813, 0.01813)	-0.56	0.602	1.00
C*D	0.03750	0.01875	0.00900	(-0.00438, 0.04188)	2.08	0.092	1.00

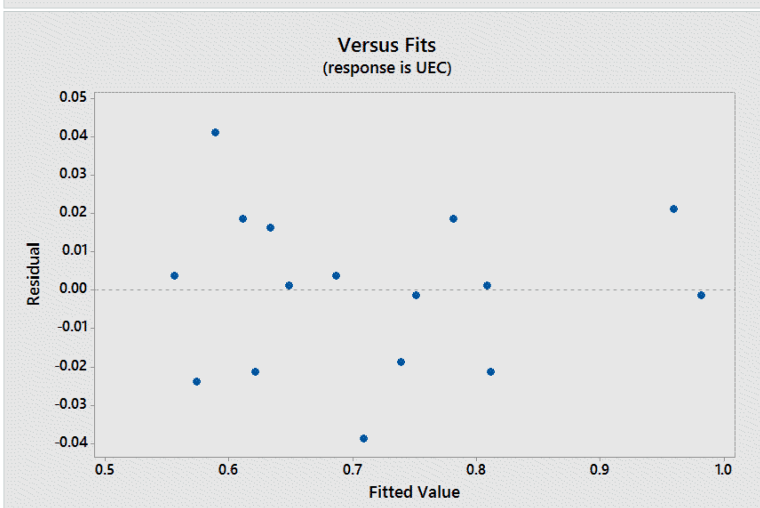
Analysis of Variance

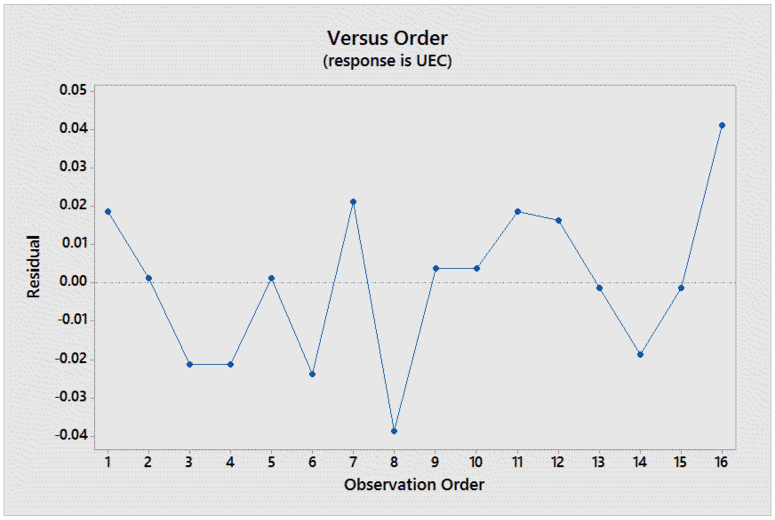
Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Model	10	0.247100	97.45%	0.247100	0.024710	19.08	0.002
Linear	4	0.224850	88.67%	0.224850	0.056212	43.41	0.000
A	1	0.102400	40.38%	0.102400	0.102400	79.07	0.000
B	1	0.001600	0.63%	0.001600	0.001600	1.24	0.317
C	1	0.070225	27.69%	0.070225	0.070225	54.23	0.001
D	1	0.050625	19.96%	0.050625	0.050625	39.09	0.002
2-Way Interactions	6	0.022250	8.77%	0.022250	0.003708	2.86	0.134
A*B	1	0.001225	0.48%	0.001225	0.001225	0.95	0.375
A*C	1	0.012100	4.77%	0.012100	0.012100	9.34	0.028
A*D	1	0.000400	0.16%	0.000400	0.000400	0.31	0.602
B*C	1	0.002500	0.99%	0.002500	0.002500	1.93	0.223
B*D	1	0.000400	0.16%	0.000400	0.000400	0.31	0.602
C*D	1	0.005625	2.22%	0.005625	0.005625	4.34	0.092
Error	5	0.006475	2.55%	0.006475	0.001295		
Total	15	0.253575	100.00%				

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







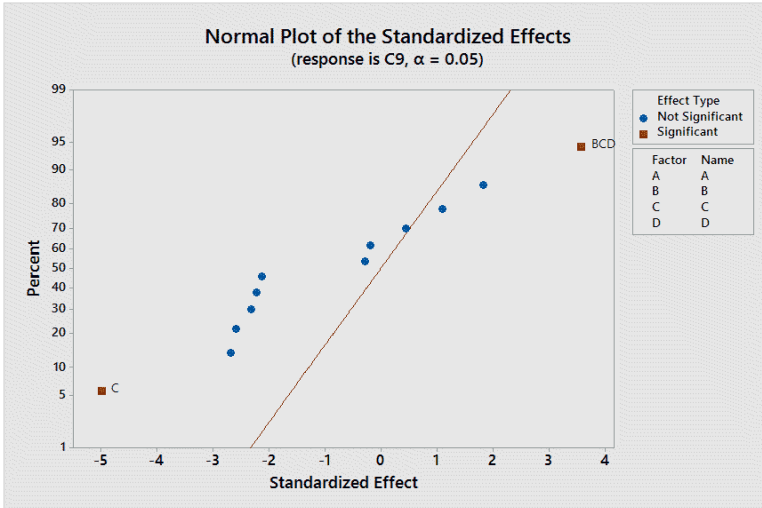


The residuals do not give us reason to question any of our assumptions.

21. *Reconsider the experiment described in Problem 6.20. Suppose that four center points are available and 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?*

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	16	1.88637	0.11790	39.99	0.006
Blocks	3	0.03538	0.01179	4.00	0.142
Linear	4	0.11175	0.02794	9.48	0.047
A	1	0.01323	0.01323	4.49	0.124
B	1	0.01563	0.01563	5.30	0.105
C	1	0.07290	0.07290	24.73	0.016
D	1	0.01000	0.01000	3.39	0.163
2-Way Interactions	5	0.05535	0.01107	3.75	0.153
A*B	1	0.00022	0.00022	0.08	0.800
A*C	1	0.01960	0.01960	6.65	0.082
A*D	1	0.00010	0.00010	0.03	0.866
B*D	1	0.01440	0.01440	4.88	0.114
C*D	1	0.02102	0.02102	7.13	0.076
3-Way Interactions	2	0.04163	0.02081	7.06	0.073
A*B*C	1	0.00360	0.00360	1.22	0.350
B*C*D	1	0.03803	0.03803	12.90	0.037
4-Way Interactions	1	0.00063	0.00063	0.21	0.677
A*B*C*D	1	0.00063	0.00063	0.21	0.677
Curvature	1	1.64164	1.64164	556.80	0.000
Error	3	0.00884	0.00295		
Total	19	1.89522			



We see fron the analysis of variance and normal probability plot of effects that the factors bcd and c are significant.

28. ***The scrumptious brownie experiment.*** *The author is an engineer by training and a firm believer in learning by doing. I have taught experimental design for many years to a wide variety of audiences and have always assigned the planning, conduct, and analysis of an actual experiment to the class participants. The participants seem to enjoy this practical experience and always learn a great deal from it. This problem uses the results of an experiment performed by Gretchen Krueger at Arizona State University. There are many different ways to bake brownies. The purpose of this experiment was to determine how the pan material, the brand of brownie mix, and the stirring method affect the scrumptiousness of brownies. The factor levels were*

Factor	**Low(-)**	**High(+)**
A = pan material	Glass	Alumninum
B = stirring method	Spoon	Mixer
C = brand of mix	Expensive	Cheap

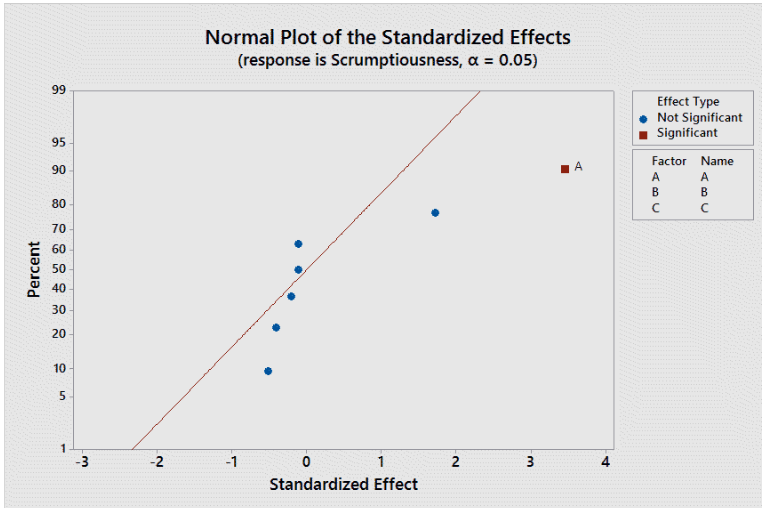
The response variable was scrumptiousness, a subjective measure derived from a questionnaire given to the subjects who sampled each batch of brownies. (The questionnaire dealt wit such issues as taste, appearance, consistency, aroma, and so forth.) An eight-person test panel sampled each batch and filled out the questionairre. The design matrix and the response data are as follows.

a.) *Analyze the data from this experiment as if there were eight replicates of a* 2^3 *design. Comment on the results.*

```
Analysis of Variance

Source                DF   Adj SS    Adj MS  F-Value  P-Value
Model                  7   93.250   13.3214     2.20    0.047
  Linear               3   90.375   30.1250     4.98    0.004
    A                  1   72.250   72.2500    11.95    0.001
    B                  1   18.062   18.0625     2.99    0.089
    C                  1    0.063    0.0625     0.01    0.919
  2-Way Interactions   3    2.625    0.8750     0.14    0.933
    A*B                1    0.062    0.0625     0.01    0.919
```

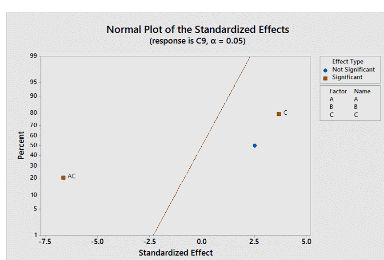
A*C	1	1.562	1.5625	0.26	0.613
B*C	1	1.000	1.0000	0.17	0.686
3-Way Interactions	1	0.250	0.2500	0.04	0.840
A*B*C	1	0.250	0.2500	0.04	0.840
Error	56	338.500	6.0446		
Total	63	431.750			



We see from both thee analysis of variance and the normal probability plot of the effects that the only factor that is significant is A. However, the factor B is close to being significant and might be significant in a reduced model.

b.) *Is the analysis in part (a) the correct approach? There are only eight batches; do we really have eight replicates of a* 2^3 *factorial design?*
Based on what we have been told about the model, we can conclude that this is not the correct approach. This is because the replicates are not action replicates. They are the same batch of brownies being tasted by different tasters. The ANOVA approach is innapropriate since it doesn't account for variation in batches.

c.) *Analyze the average and standard deviation of the scrumptiousness ratings. Comment on the results. Is this analysis more appropriate than the one in part (a)? Why or why not?*
Standard Deviation:

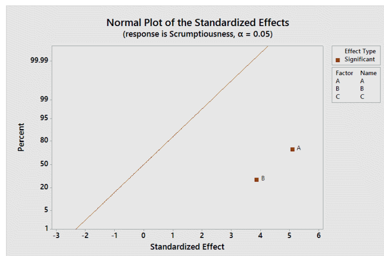


Analysis of Variance for Ln(C9)

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Model	3	20.683	94.11%	20.683	6.8943	21.30	0.006

Linear	2	6.442	29.31%	6.442	3.2211	9.95	0.028
A	1	2.086	9.49%	2.086	2.0857	6.44	0.064
C	1	4.357	19.82%	4.357	4.3566	13.46	0.021
2-Way Interactions	1	14.241	64.80%	14.241	14.2408	43.99	0.003
A*C	1	14.241	64.80%	14.241	14.2408	43.99	0.003
Error	4	1.295	5.89%	1.295	0.3237		
Total	7	21.978	100.00%				

Average:




```
Analysis of Variance

Source    DF  Adj SS  Adj MS  F-Value  P-Value
Model      2  10.352  5.1758    20.45    0.004
  Linear   2  10.352  5.1758    20.45    0.004
    A      1   6.570  6.5703    25.96    0.004
    B      1   3.781  3.7813    14.94    0.012
Error      5   1.266  0.2531
Total      7  11.617
```

We see that factors A and B affect the mean of the scrumptiousness and AC affects the variability of scrumptiousness. This is a better model than part a since it attempts to control for batch variability, which gives a better estimate of the error.

36. *Often the fitted regression model from a 2^k factorial design is used to make predictions at points of interest in the design space. Assume that the model contains all main effects and two-factor interactions.*

a.) *Find the variance of the predicted response $\hat{y}$ at a point $x_1, x_2, \ldots, x_k$ in the design space. Hint: Remember that the x's are coded variables and assume a 2^k design with an equal number of replicates n at each design point so that the variance of a regression coefficient $\hat{\beta}$ is $\sigma^2/(n2^k)$ and that the covariance between any pair of regression coefficients is zero.*
Let $x = [x_1, x_2, \ldots, x_k]$. A basic form of the model is

$$\hat{y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k$$

Now, we know that the variance of the predicted variables will follow the form

$$\begin{aligned} V[\hat{y}(x)] &= V(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k) \\ &= V(\hat{\beta}_0) + V(\hat{\beta}_1 x_1) + V(\hat{\beta}_2 x_2) + \ldots + V(\hat{\beta}_k x_k) \\ &= \frac{\sigma^2}{n2^k}(1 + \sum_{i=1}^{k} x_i^2) \end{aligned}$$

b.) *Use the result in part (a) to find an equation for* $100(1-\alpha)$ *percent confidence interval on the true mean response at the point* $x_1, x_2, \ldots, x_k$ *in design space.*
We know that the basic form for the $100(1-\alpha)$ percent confidence interval on the true mean is

$$\hat{y}(x) - t_{\alpha/2,df}\sqrt{V[\hat{y}(x)]} \leq y(x) \leq \hat{y}(x) + t_{\alpha/2,df}\sqrt{V[\hat{y}(x)]}$$

However, we can substitute our derived value of the variance of the predicted response.

$$\hat{y}(x) - t_{\alpha/2,df}\sqrt{\frac{\sigma^2}{n2^k}(1+\sum_{i=1}^{k} x_i^2)} \leq y(x) \leq \hat{y}(x) + t_{\alpha/2,df}\sqrt{\frac{\sigma^2}{n2^k}(1+\sum_{i=1}^{k} x_i^2)}$$