Meenakshi Nagarajan

U00618268

EGR 7050 Design and Analysis of Engineering experiments

Homework 8

1. *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:*

Treatment Combination	Replicate		Treatment Combination	Replicate	
	I	II		I	II
(1)	90	93	d	98	95
a	74	78	ad	72	76
b	81	85	bd	87	83
ab	83	80	abd	85	86
c	77	78	cd	99	90
ac	81	80	acd	79	75
bc	88	82	bcd	87	84
abc	73	70	abcd	80	80

Solution:

a. *Estimate the factor effects*

Response Yield

Summary of Fit

RSquare	0.92473
RSquare Adj	0.854164
Root Mean Square Error	2.766993
Mean of Response	82.78125
Observations (or Sum Wgts)	32

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	15	1504.9688	100.331	13.1045
Error	16	122.5000	7.656	Prob > F
C. Total	31	1627.4688		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	82.78125	0.48914	169.24	<.0001*
A[+]	-4.53125	0.48914	-9.26	<.0001*
B[+]	-0.65625	0.48914	-1.34	0.1984
C[+]	-1.34375	0.48914	-2.75	0.0143*
D[+]	1.96875	0.48914	4.02	0.0010*
A[+]*B[+]	2.03125	0.48914	4.15	0.0007*
A[+]*C[+]	0.34375	0.48914	0.70	0.4923
B[+]*C[+]	-0.28125	0.48914	-0.57	0.5733
A[+]*D[+]	-1.09375	0.48914	-2.24	0.0399*
B[+]*D[+]	-0.09375	0.48914	-0.19	0.8504
C[+]*D[+]	0.84375	0.48914	1.72	0.1038
A[+]*B[+]*C[+]	-2.59375	0.48914	-5.30	<.0001*
A[+]*B[+]*D[+]	2.34375	0.48914	4.79	0.0002*
A[+]*C[+]*D[+]	-0.46875	0.48914	-0.96	0.3522
B[+]*C[+]*D[+]	-0.46875	0.48914	-0.96	0.3522
A[+]*B[+]*C[+]*D[+]	1.21875	0.48914	2.49	0.0241*

a) This gives the estimates of factors and interactions. Factors A, C, D, AB, AD, ABC, ABD, ABCD are significant at 5% level.

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
A	1	1	657.03125	85.8163	<.0001*
B	1	1	13.78125	1.8000	0.1984
C	1	1	57.78125	7.5469	0.0143*
D	1	1	124.03125	16.2000	0.0010*
A*B	1	1	132.03125	17.2449	0.0007*
A*C	1	1	3.78125	0.4939	0.4923
B*C	1	1	2.53125	0.3306	0.5733
A*D	1	1	38.28125	5.0000	0.0399*
B*D	1	1	0.28125	0.0367	0.8504
C*D	1	1	22.78125	2.9755	0.1038
A*B*C	1	1	215.28125	28.1184	<.0001*
A*B*D	1	1	175.78125	22.9592	0.0002*
A*C*D	1	1	7.03125	0.9184	0.3522
B*C*D	1	1	7.03125	0.9184	0.3522
A*B*C*D	1	1	47.53125	6.2082	0.0241*

Fig. 1 Effect estimates

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

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	9	1461.5313	162.392	21.5300
Error	22	165.9375	7.543	Prob > F
C. Total	31	1627.4688		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	82.78125	0.485496	170.51	<.0001*
A[+]	-4.53125	0.485496	-9.33	<.0001*
B[+]	-0.65625	0.485496	-1.35	0.1902
C[+]	-1.34375	0.485496	-2.77	0.0112*
D[+]	1.96875	0.485496	4.06	0.0005*
A[+]*B[+]	2.03125	0.485496	4.18	0.0004*
A[+]*D[+]	-1.09375	0.485496	-2.25	0.0346*
A[+]*B[+]*C[+]	-2.59375	0.485496	-5.34	<.0001*
A[+]*B[+]*D[+]	2.34375	0.485496	4.83	<.0001*
A[+]*B[+]*C[+]*D[+]	1.21875	0.485496	2.51	0.0199*

The same factors are found to be significant at 5% level except factor B. But this is included in the model to preserve hierarchy.

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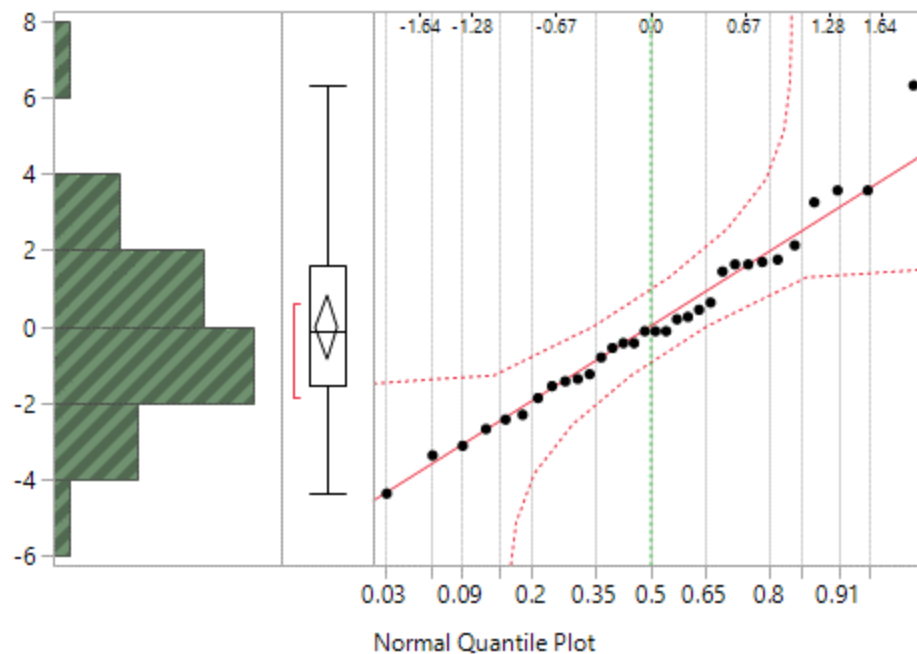

Fig. 2 Normal quantile plot

Not all points lie close to the line. One point is away from the straight line. Therefore, this does not support normality assumption.
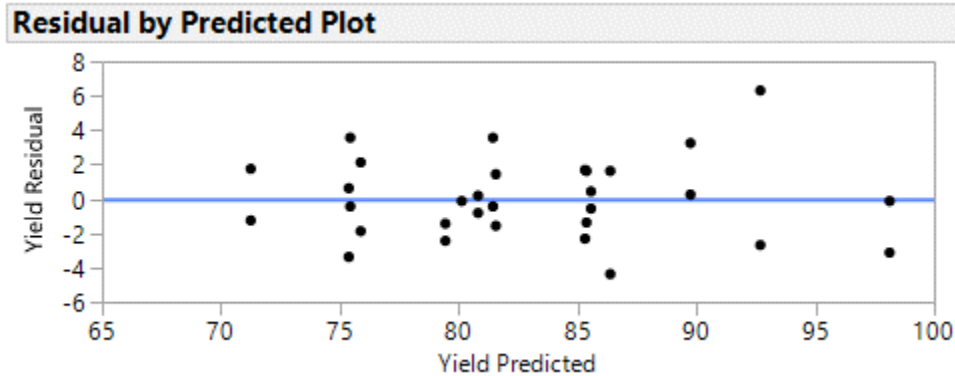

Fig. 3 Residual vs. Predicted

It shows that there is a relationship between size of residuals and predicted values. Thus, assumptions are violated.

2. *In a process development study on yield, four factors were studied, each at two levels: time (A), concentration (B), pressure (C), and temperature (D). A single replicate of a 24 design was run, and the resulting data are shown in Table P6.7.*

Run Number	Actual run order	A	B	C	D	Yield (lbs)		Factor levels	
								Low(-)	High (+)
1	5	-	-	-	-	12	A(h)	2.5	3.0
2	9	+	-	-	-	18	B(%)	14	18
3	8	-	+	-	-	13	C(psi)	60	80
4	13	+	+	-	-	16	D(÷C)	225	250
5	3	-	-	+	-	17			
6	7	+	-	+	-	15			
7	14	-	+	+	-	20			
8	1	+	+	+	-	15			

9	6	-	-	-	+	10			
10	11	+	-	-	+	25			
11	2	-	+	-	+	13			
12	15	+	+	-	+	24			
13	4	-	-	+	+	19			
14	16	+	-	+	+	21			
15	10	-	+	+	+	17			
16	12	+	+	+	+	23			

Solution:

a. *Construct a normal probability plot of the effect estimates. Which factors appear to have large effects?*

Response Yield

Summary of Fit

RSquare	1
RSquare Adj	.
Root Mean Square Error	.
Mean of Response	17.375
Observations (or Sum Wgts)	16

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
A	1	1	81.000000	.	.
B	1	1	1.000000	.	.
C	1	1	16.000000	.	.
D	1	1	42.250000	.	.
A*B	1	1	2.250000	.	.
A*C	1	1	72.250000	.	.
B*C	1	1	0.250000	.	.
A*D	1	1	64.000000	.	.
B*D	1	1	0.000000	.	.
C*D	1	1	0.000000	.	.
A*B*C	1	1	4.000000	.	.
A*B*D	1	1	2.250000	.	.
A*C*D	1	1	0.250000	.	.
B*C*D	1	1	2.250000	.	.
A*B*C*D	1	1	4.000000	.	.

Effect Screening

The parameter estimates have equal variances.
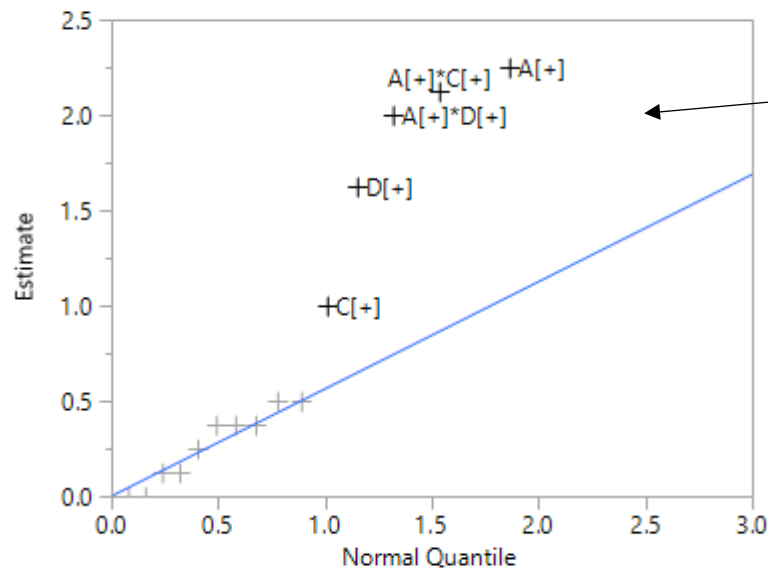
The parameter estimates are not correlated.

Lenth PSE

0.5625

Orthog t Test used Pseudo Standard Error

Normal Plot

Half Normal Plot

From half-normal plot of effect estimate, we could see that effects A, C, D and interactions AC and AD appear to have large effects.

Blue line has slope equal to Lenth's PSE.

Fig. 4 Response plot

c. *Conduct an analysis of variance using the normal probability plot in part (a) for guidance in forming an error term. What are your conclusions?*

Response Yield

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	2	0.250000	0.12500	0.0625
Pure Error	8	16.000000	2.00000	**Prob > F**
Total Error	10	16.250000		0.9399
				Max RSq
				0.9452

Summary of Fit

RSquare	0.944302
RSquare Adj	0.916452
Root Mean Square Error	1.274755
Mean of Response	17.375
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	275.50000	55.1000	33.9077
Error	10	16.25000	1.6250	**Prob > F**
C. Total	15	291.75000		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	17.375	0.318689	54.52	<.0001*
A[+]	2.25	0.318689	7.06	<.0001*
C[+]	1	0.318689	3.14	0.0105*
D[+]	1.625	0.318689	5.10	0.0005*
A[+]*C[+]	-2.125	0.318689	-6.67	<.0001*
A[+]*D[+]	2	0.318689	6.28	<.0001*

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
A	1	1	81.000000	49.8462	<.0001*
C	1	1	16.000000	9.8462	0.0105*
D	1	1	42.250000	26.0000	0.0005*
A*C	1	1	72.250000	44.4615	<.0001*
A*D	1	1	64.000000	39.3846	<.0001*

For $\alpha = 0.05$, these effects A,C,D,AC and AD are significant.

Cube Plot
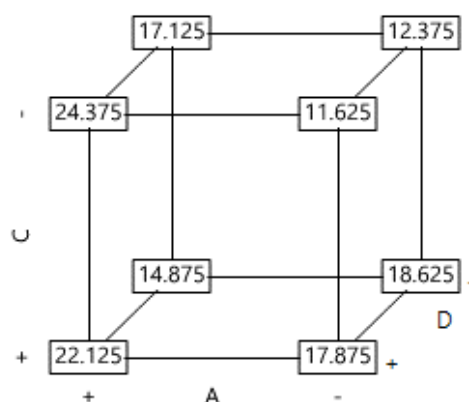
d.   Write down a regression model relating yield to the important process variables.

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	17.375	0.318689	54.52	<.0001*
A[+]	2.25	0.318689	7.06	<.0001*
C[+]	1	0.318689	3.14	0.0105*
D[+]	1.625	0.318689	5.10	0.0005*
A[+]*C[+]	-2.125	0.318689	-6.67	<.0001*
A[+]*D[+]	2	0.318689	6.28	<.0001*

These give the coefficients of the regression model.

Regression model:

$$Yield = 17.375 + 2.25 * A + 1.00 * C + 1.625 * D - 2.125 * A * C + 2.00 * A * D$$

$$Yield = 17.375 + 2.25 * Time + 1.00 * Pressure + 1.625 * Temperature - 2.125 * Time * Pressure + 2.00 * Time * Temperature$$

e.   Analyze the residuals from this experiment. Does your analysis indicate any potential problems?


Normal Quantile Plot
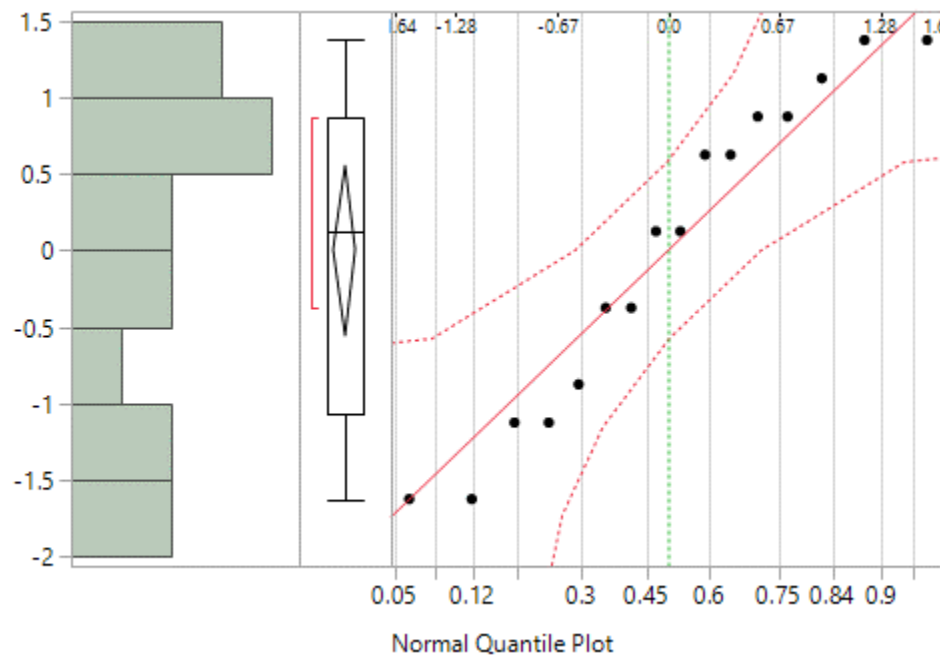
All points lie close to the line and within error bounds. Hence, normal probability plot supports the normality assumption.

Meenakshi Nagarajan
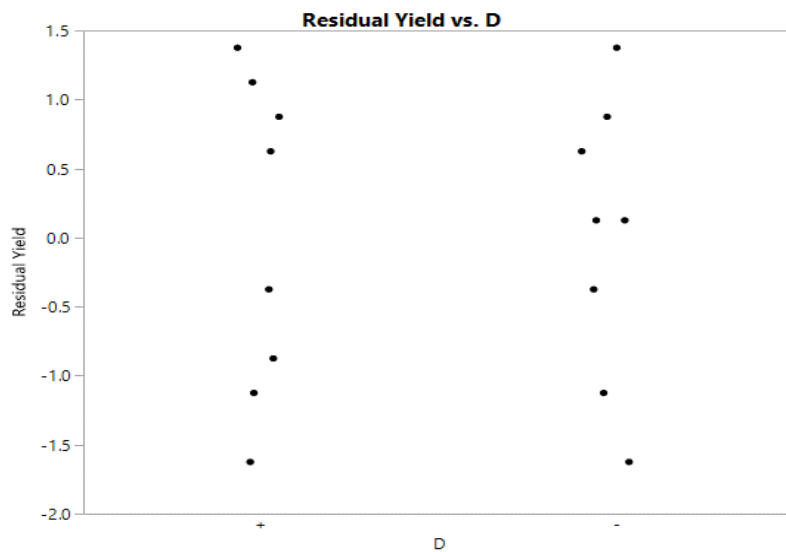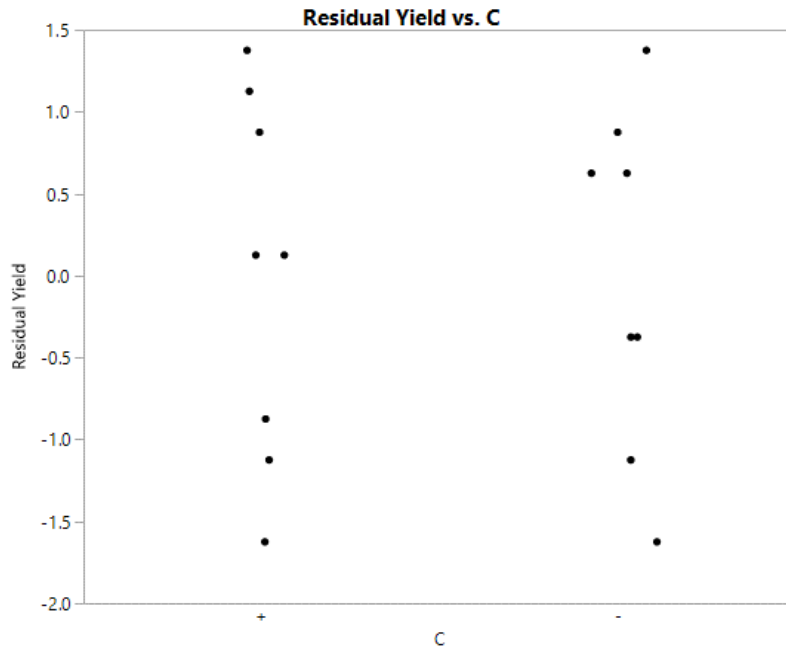
U00618268

Residual by Predicted Plot
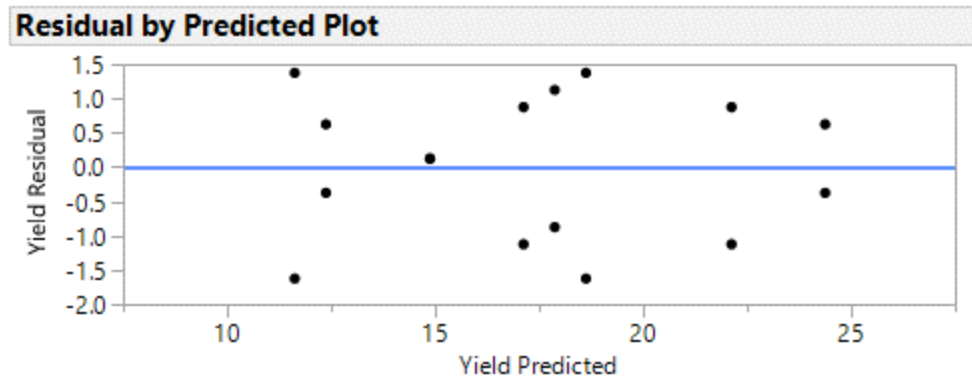
It shows similar range of residuals across all the fitted values. Therefore, we could not see any problems with the variance of the residuals.
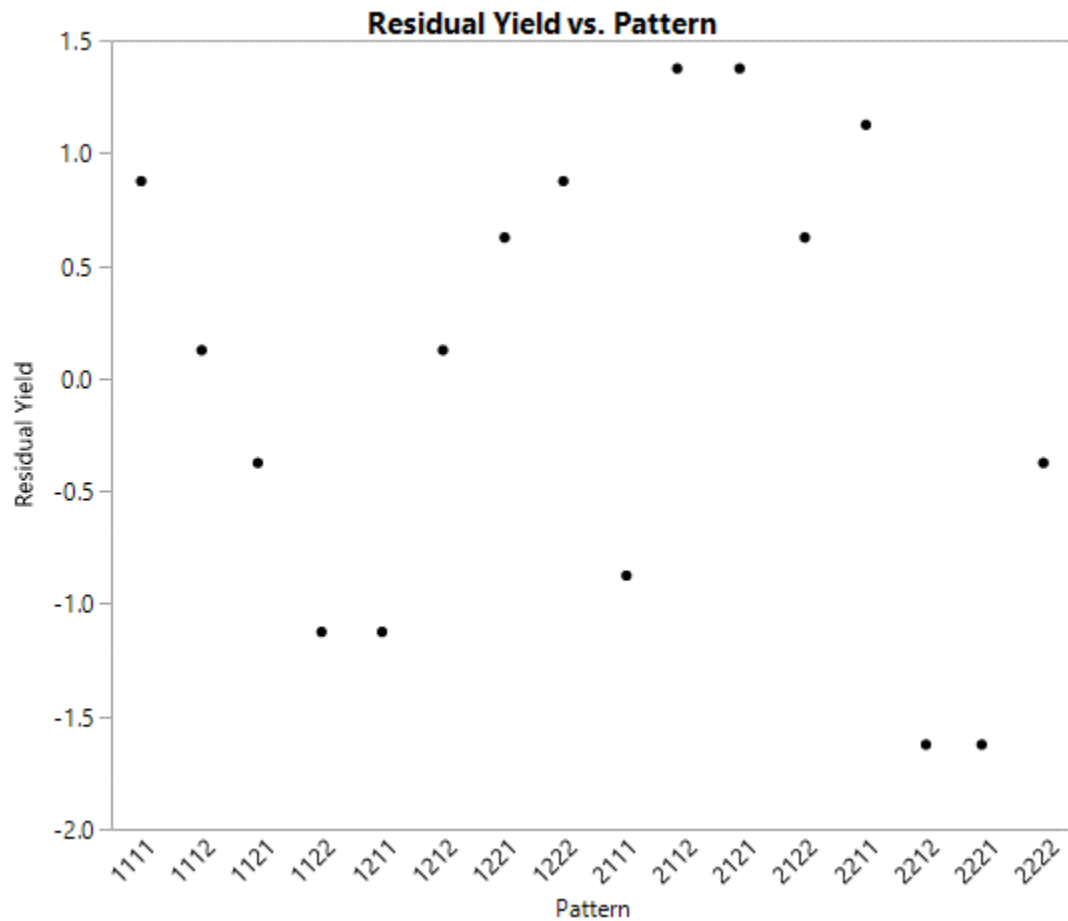
This shows no unusual pattern and there no problems with the variance of residuals.

Meenakshi Nagarajan

U00618268

Residual Yield vs. A

Residual Yield vs. B
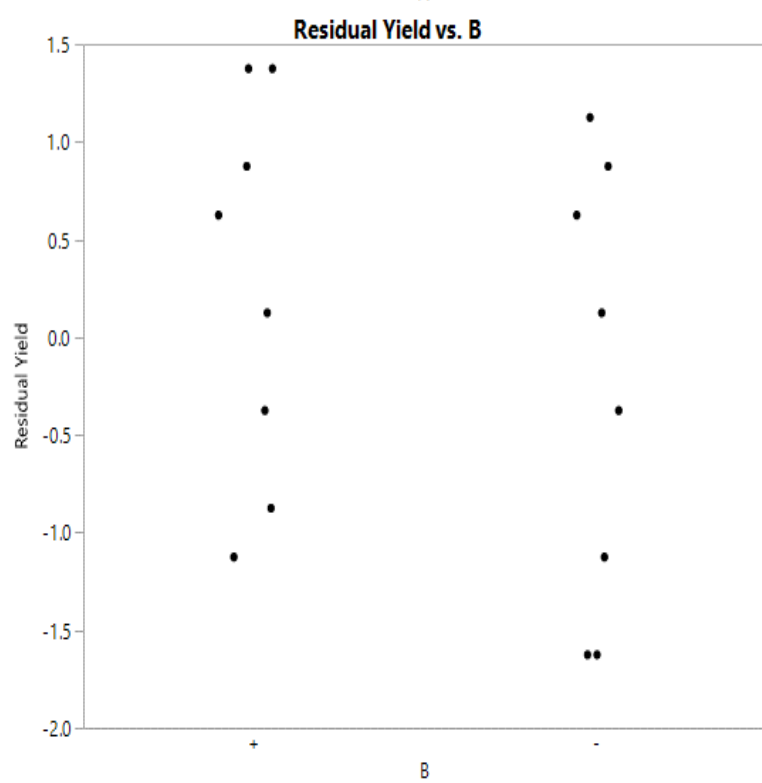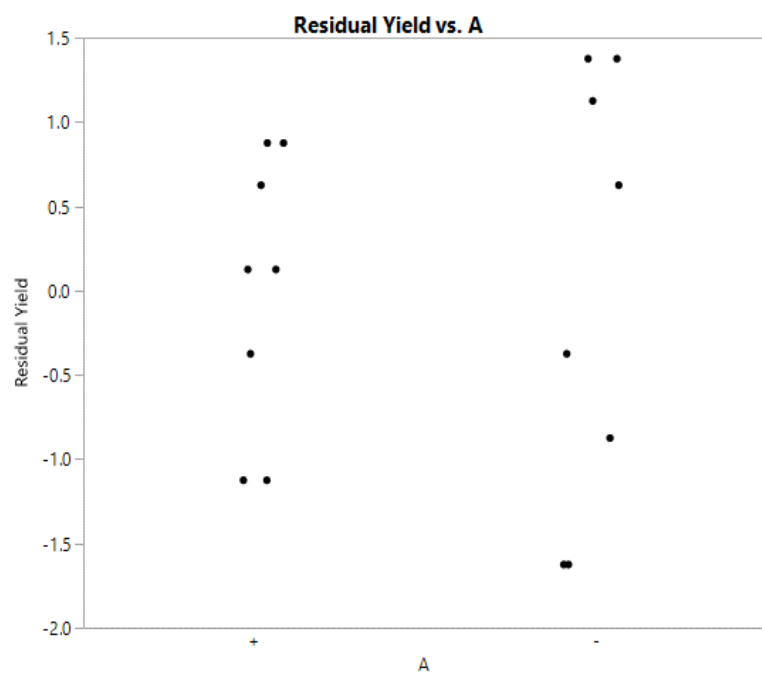
The range of residuals across two levels of each of the four factors are same.

3. *Consider the data from the first replicate of Problem 6.7. Construct a design with two blocks of eight observations each with ABCD confounded. Analyze the data.*

Treatment	Replicate		Treatment	Replicate	
Combination	I	II	Combination	I	II
(1)	90	93	d	98	95
a	74	78	ad	72	76
b	81	85	bd	87	83
ab	83	80	abd	85	86

c	77	78	cd	99	90
ac	81	80	acd	79	75
bc	88	82	bcd	87	84
abc	73	70	abcd	80	80

Solution:

Two blocks of eight observations each with ABCD confounded can be written as:

Block 1	Block 2
(1) = 90	a=74
ab=83	b=81
ac=81	c=77
bc=88	d=98
ad=72	abc=73
bd=87	abd=85
cd=99	acd=79
abcd=80	bcd=87
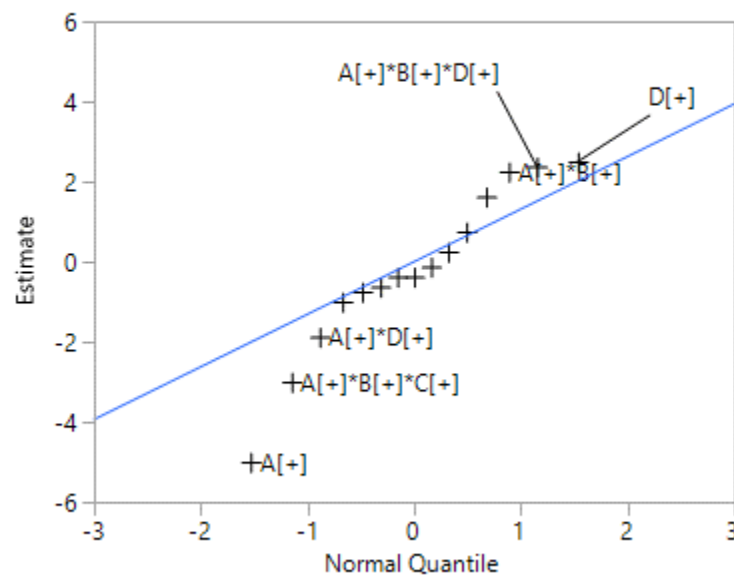
Blue line has slope equal to Lenth's PSE.

From the normal plot, significant factors can be identified as A, ABC, AD, AB, ABD, D. Factors B, AC, BC and BD are included to preserve hierarchy.

Response Yield

Summary of Fit

RSquare	0.973691
RSquare Adj	0.868455
Root Mean Square Error	2.901149
Mean of Response	83.375
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	12	934.50000	77.8750	9.2525
Error	3	25.25000	8.4167	Prob > F
C. Total	15	959.75000		0.0463*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	83.375	0.725287	114.95	<.0001*
A[+]	-5	0.725287	-6.89	0.0063*
B[+]	-0.375	0.725287	-0.52	0.6408
C[+]	-0.375	0.725287	-0.52	0.6408
D[+]	2.5	0.725287	3.45	0.0410*
A[+]*B[+]*C[+]	-3	0.725287	-4.14	0.0256*
A[+]*D[+]	-1.875	0.725287	-2.59	0.0814
A[+]*B[+]	2.25	0.725287	3.10	0.0532
A[+]*B[+]*D[+]	2.375	0.725287	3.27	0.0466*
A[+]*C[+]	0.25	0.725287	0.34	0.7531
B[+]*C[+]	-0.625	0.725287	-0.86	0.4522
B[+]*D[+]	-0.75	0.725287	-1.03	0.3772
Block[1]	-1.625	0.725287	-2.24	0.1109

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
A	1	1	400.00000	47.5248	0.0063*
B	1	1	2.25000	0.2673	0.6408
C	1	1	2.25000	0.2673	0.6408
D	1	1	100.00000	11.8812	0.0410*
A*B*C	1	1	144.00000	17.1089	0.0256*
A*D	1	1	56.25000	6.6832	0.0814
A*B	1	1	81.00000	9.6238	0.0532
A*B*D	1	1	90.25000	10.7228	0.0466*
A*C	1	1	1.00000	0.1188	0.7531
B*C	1	1	6.25000	0.7426	0.4522
B*D	1	1	9.00000	1.0693	0.3772
Block	1	1	42.25000	5.0198	0.1109

P values of A, D, ABC and ABD are less than $\alpha = 0.05$. These are significant at 5% level. This confirms our initial assessment.

Normal Plot

Normal Plot
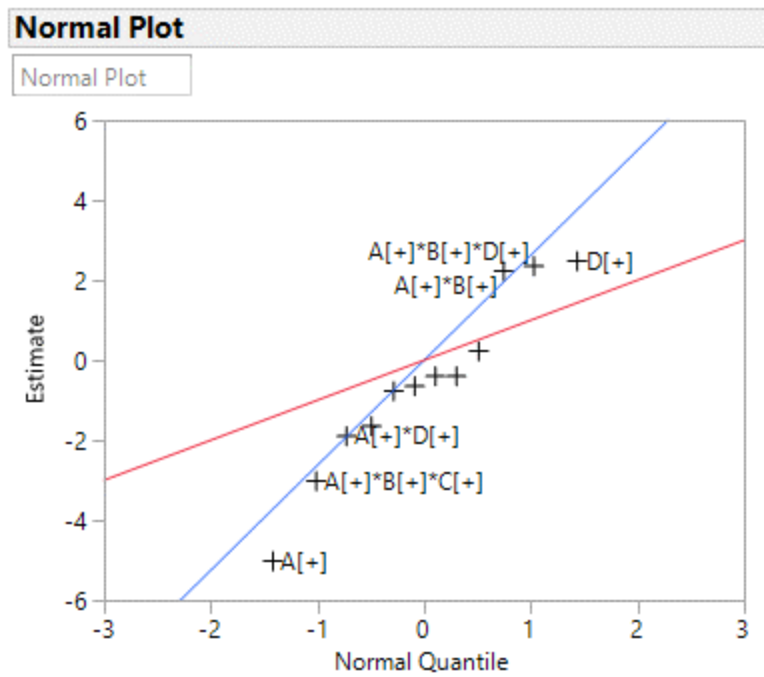
Blue line has slope equal to Lenth's PSE.
Red line has slope 1.