STAT 514 Design of Experiments — Homework 8

Problem 1

(a) First, the respective means for the 8 level combinations are listed in the following table

A	B	C	Mean
$-$	$-$	$-$	26.00
$+$	$-$	$-$	34.67
$-$	$+$	$-$	39.67
$+$	$+$	$-$	49.33
$-$	$-$	$+$	42.33
$+$	$-$	$+$	37.67
$-$	$+$	$+$	54.67
$+$	$+$	$+$	42.33

Now factorial effects A and AB can be calculated by

$$\begin{aligned}
A &= \frac{1}{4}\left[-\bar{y}(---)+\bar{y}(+--)-\bar{y}(-+-)+\bar{y}(++-)-\bar{y}(--+)+\bar{y}(+-+)-\bar{y}(-++)+\bar{y}(+++)\right] \\
&= \frac{1}{4}(-26.00+34.67-39.67+49.33-42.33+37.67-54.67+42.33) \\
&= 0.33, \\
AB &= \frac{1}{4}\left[\bar{y}(---)-\bar{y}(+--)-\bar{y}(-+-)+\bar{y}(++-)+\bar{y}(--+)-\bar{y}(+-+)-\bar{y}(-++)+\bar{y}(+++)\right] \\
&= \frac{1}{4}(26.00-34.67-39.67+49.33+42.33-37.67-54.67+42.33) \\
&= -1.67.
\end{aligned}$$

Other factorial effects can be calculated in a similar way. All the effects are summarized in the following table.

Effects	A	B	AB	C	AC	BC	ABC
Estimates	0.33	11.33	-1.67	6.83	-8.83	-2.83	-2.17

From the table, effects B, C and AC apprear to be large (significant).

(b) The ANOVA table from the SAS GLM procedure is shown below (model SS is replaced by the SS for individual effects), following which are the estimates of all the effects from this procedure.

Source	DF	Squares	Mean Square	F Value	Pr > F
A	1	0.6666667	0.6666667	0.02	0.8837
B	1	770.6666667	770.6666667	25.55	0.0001
C	1	280.1666667	280.1666667	9.29	0.0077
AB	1	16.6666667	16.6666667	0.55	0.4681
AC	1	468.1666667	468.1666667	15.52	0.0012
BC	1	48.1666667	48.1666667	1.60	0.2245
ABC	1	28.1666667	28.1666667	0.93	0.3483
Error	16	482.666667	30.166667		
Corrected Total	23	2095.333333			

R-Square	Coeff Var	Root MSE	y Mean
0.769647	13.45082	5.492419	40.83333

Parameter	Estimate	Standard Error	t Value	Pr > |t|
A	0.3333333	2.24227067	0.15	0.8837
B	11.3333333	2.24227067	5.05	0.0001
AB	-1.6666667	2.24227067	-0.74	0.4681
C	6.8333333	2.24227067	3.05	0.0077
AC	-8.8333333	2.24227067	-3.94	0.0012
BC	-2.8333333	2.24227067	-1.26	0.2245
ABC	-2.1666667	2.24227067	-0.97	0.3483

Notice that the significant effects from the ANOVA table are B, C and AC, which are consistent with my conclusions in part (a). Also, the estimates are the same.

(c) From (b), only effects B, C and AC are significant. So our model should include only these three effects plus the main effect A. If we introduce variables x_1, x_2 and x_3 as follows:

$$x_1 = \begin{cases} -1, & \text{if } A = -, \\ 1, & \text{if } A = +. \end{cases} \qquad x_2 = \begin{cases} -1, & \text{if } B = -, \\ 1, & \text{if } B = +. \end{cases} \qquad x_3 = \begin{cases} -1, & \text{if } C = -, \\ 1, & \text{if } C = +. \end{cases}$$

then the fitted regression model will have the following form with the coefficients equal to 1/2 of the corresponding effect estimates:

$$y = \bar{y}_{\dots} + \frac{A}{2}x_1 + \frac{B}{2}x_2 + \frac{C}{2}x_3 + \frac{AC}{2}x_1x_3.$$

Plugging in $\bar{y}_{\dots} = 40.83$ and the effect estimates from (a) gives the fitted regression model

$$y = 40.83 + 0.17x_1 + 5.67x_2 + 3.42x_3 - 4.42x_1x_3.$$

(d) The diagnostic plots in Figure 1 are: normal probability Q-Q plot, plot of residuals versus factor A (cutting speed), plot of residuals versus factor B (tool geometry), plot of residuals versus factor C (cutting angle), and plot of residuals versus predicted values. Also, the results of formal normality tests are listed below.

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.953002	Pr < W	0.3143
Kolmogorov-Smirnov	D	0.13291	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.059821	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.401988	Pr > A-Sq	>0.2500

The normal Q-Q plot and the normality tests shows that the normality assumption is valid. None of the four residual plots has shown unequal variances or potential outliers.

(e) The main effect plots for B and C are in Figure 2.

The interaction plot for A and C is in Figure 3.

We would like to choose the factor levels such that the life of a machine tool is maximized. According to the above three plots, the optimal levels would be $(A, B, C) = (-, +, +)$.

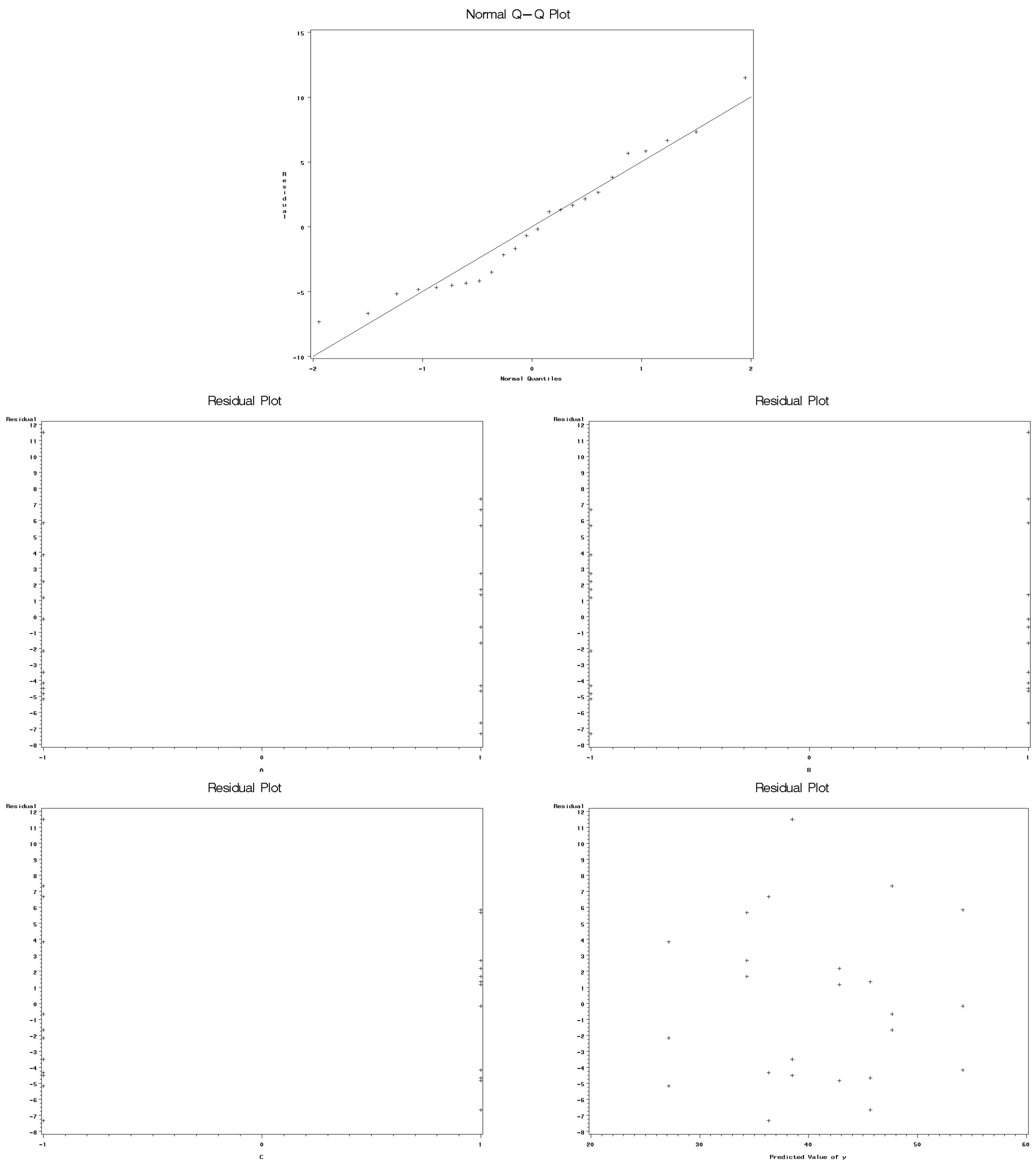

Figure 1: Diagnostic Plots

(f) In the regression model in part (c), if we set B to be at the low level ($x_2 = -1$) and the high level ($x_2 = 1$) respectively, we will end up with the following two models:

$$y = 35.17 + 0.17x_1 + 3.42x_3 - 4.42x_1x_3 \quad \text{when } x_2 = -1,$$
$$y = 46.5 + 0.17x_1 + 3.42x_3 - 4.42x_1x_3 \quad \text{when } x_2 = 1.$$

The contour plots generated from these two models are shown in Figure 4.

The trends in both plots suggest that to obtain the longest life for the machine tool, A (cutting speed) should be set at the low level and C (cutting angle) should be set at the high level. Also, the response values in the contour plot for high level of B are always larger than those in the contour plot for low level of B. So B (tool geometry) should be set at the high level. These results are consistent with what we found in part (e).

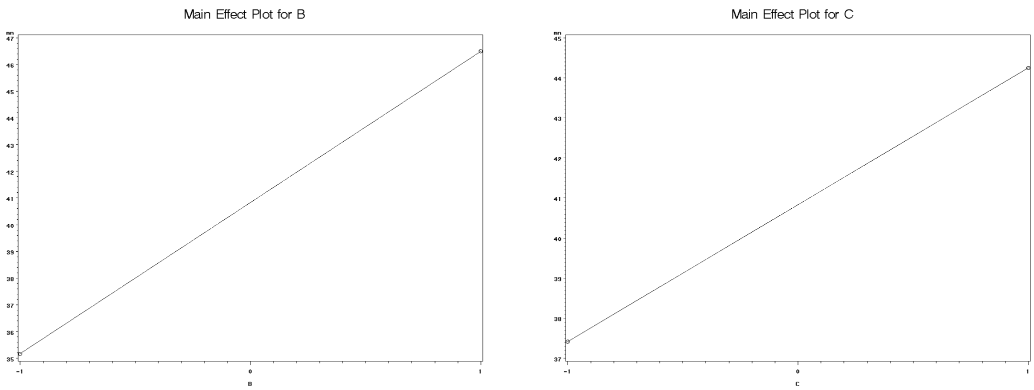

Figure 2: Main Effect Plots

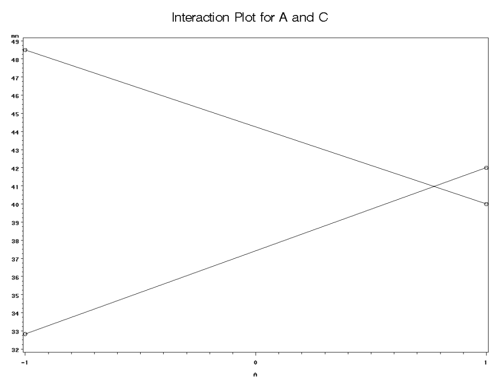

Figure 3: Interaction Plots

(g) The standard error of the factorial effects is caculated by

$$S.E. = \sqrt{\frac{MSE}{n2^{k-2}}} = \sqrt{\frac{30.17}{3 \cdot 2^{3-2}}} = 2.24. \quad \blacksquare$$

Problem 2

(a) The estimates of the factorial effects are summarized in the table below.

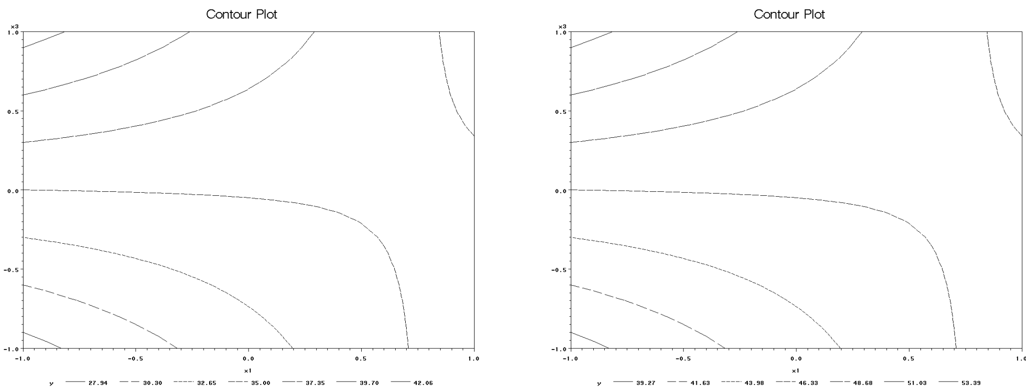

Figure 4: Contour Plots for Problem 1(f)

Effect	Estimate
A	-101.625
B	-1.625
C	7.375
D	306.125
AB	-7.875
AC	-24.875
AD	-153.625
BC	-43.875
BD	-0.625
CD	-2.125
ABC	-15.625
ABD	4.125
ACD	5.625
BCD	-25.375
$ABCD$	-40.125

And the normal QQ plot of these effect estimates are given in Figure 5, from which we can see that effects A, D and AD are significant. All the other effects approximately lie on a straight line and thus are insignificant.

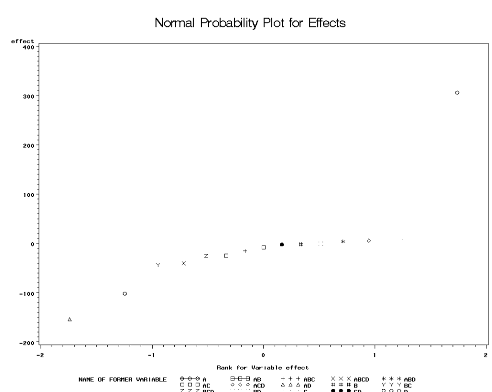

Figure 5: Normal Q-Q Plot for Effects

(b) The ANOVA result for the model including only effects A, D and AD is shown below (model SS replaced by the effect SS).

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
A	1	41310.5625	41310.5625	23.77	0.0004
D	1	374850.0625	374850.0625	215.66	<.0001
AD	1	94402.5625	94402.5625	54.31	<.0001
Error	12	20857.7500	1738.1458		
Corrected Total	15	531420.9375			

R-Square	Coeff Var	Root MSE	y Mean
0.960751	5.372129	41.69108	776.0625

All the three p-values are very small (< 0.0005), so these three effects are significant which confirms my findings in (a).

(c) If we introduce variables $x_1$and x_4 as follows:

$$x_1 = \begin{cases} -1, & \text{if } A = -, \\ 1, & \text{if } A = +. \end{cases} \qquad x_4 = \begin{cases} -1, & \text{if } D = -, \\ 1, & \text{if } D = +. \end{cases}$$

then the regression model relating the etch rate to the significant process variables A, D and AD is

$$\begin{aligned} y &= \bar{y} + \frac{A}{2}x_1 + \frac{D}{2}x_4 + \frac{AD}{2}x_1x_4 \\ &= 776.0625 - 50.8125x_1 + 153.0625x_4 - 76.8125x_1x_4, \end{aligned}$$

where the overall mean $\bar{y}$ and the effects estimates are obtained from output in (b) and result in (a) respectively.

(d) The diagnostic plots in Figure 6 are: normal probability Q-Q plot, plot of residuals versus factor A (anode-cathode gap), plot of residuals versus factor D (power applied to the cathode), and plot of residuals versus predicted values. Also, the results of formal normality tests are listed below.

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.969676	Pr < W	0.8333
Kolmogorov-Smirnov	D	0.113048	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.045931	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.268611	Pr > A-Sq	>0.2500

The normal Q-Q plot and the normality tests shows that the normality assumption is valid. None of the four residual plots has shown unequal variances or potential outliers. Hence the model fits well.

(e) Only effects A, D and AD are significant. So this 2^4 design can be projected to a 2^2 design, whose ANOVA table is already given in part (b).

(f) The interacton plot for A and D is in Figure 7.

The plot shows that A and D interact with each other: if D is set at low level, then the response mean increases as A changes from low level to high level; whereas, if D is at high level, then the response mean decreases as A changes from low level to high level.

(g) The contour plot of the etch rate using the model in (c) is as shown in Figure 8.

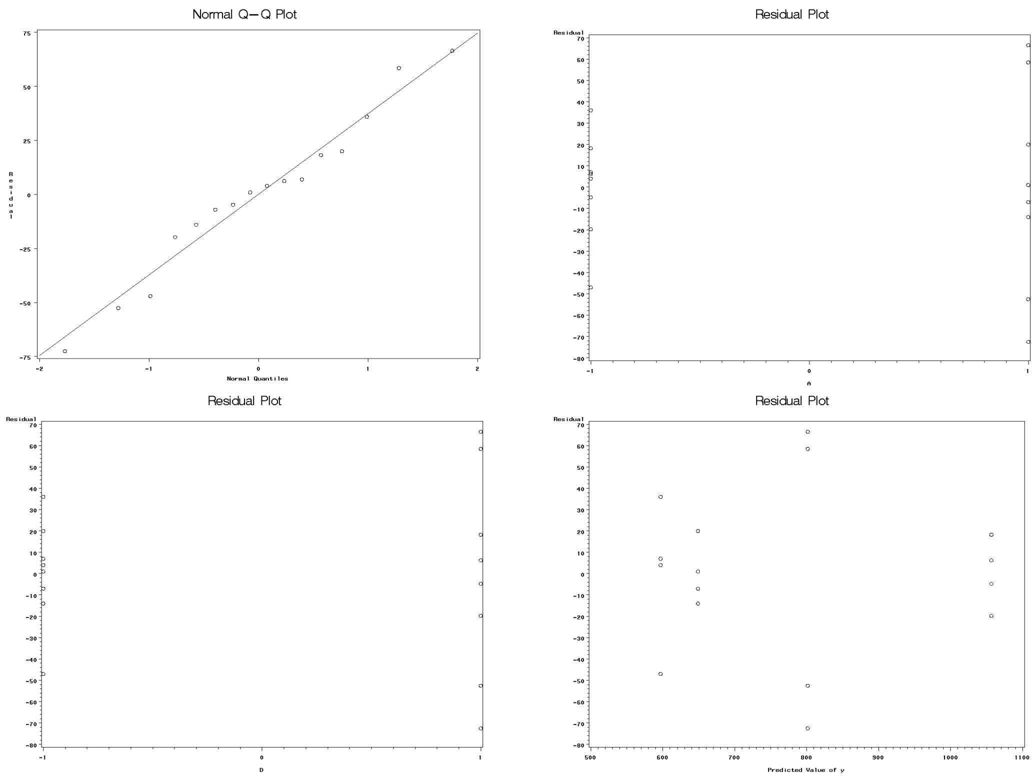

Figure 6: Diagnostic Plots

The trend in the contour plot suggests that to maximize the etch rate, we should set the levels as $(A, D) = (-, +)$.

(h) According to the trend in the contour plot in part (g), the contour corresponding to etch rate $= 800$ should be between the contours for etch rate $= 757.91$ and etch rate $= 826.88$, and thus will approximately pass the point $(1, 1)$ in the plot. This suggests that the levels should be set as $(A, D) = (+, +)$ to attain the target etch rate 800. $\blacksquare$

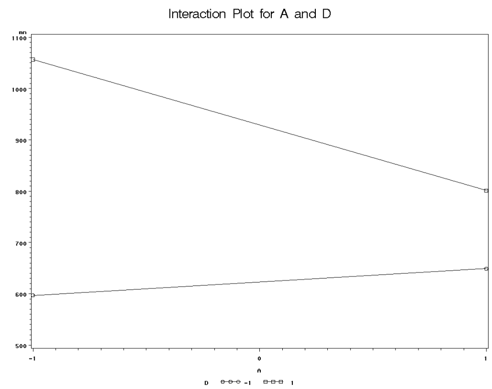

Figure 7: Interaction Plot for A and D

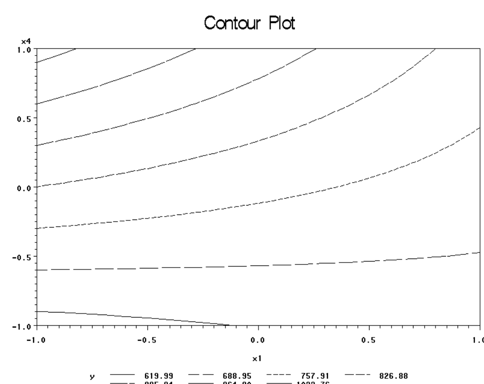

Figure 8: Contour Plot