



Problem Set #5

Issued: Wednesday, March 12, 2003

Due: Monday, March 31, 2003

Problem 1

A process engineer is about to begin a comprehensive study to determine the effects of five variables on the etch rate of polysilicon.

- (a) if a two level, five factor factorial design with no replicates were used, how many runs would be made?
- (b) If σ^2 is the experimental error variance of an individual observation, what is the variance of the main effect?
- (c) Write an expression for the 99% confidence interval for the main effect of variable 1, in terms of calculated averages and the experimental error standard deviation σ .
- (d) On the basis of some previous work it is believed that $\sigma=200 \text{ \AA}/\text{min}$. If the experimenter wants 99% confidence intervals for the main effect and interactions whose length are equal to $50 \text{ \AA}/\text{min}$, (i.e. the upper limit minus the lower limit of the confidence interval is equal to $50 \text{ \AA}/\text{min}$), how many replications of the 2^5 factorial design will be required? (Note: this may take some thought.)

Problem 2

Tests were carried out on a newly designed chemical-mechanical polisher. Four variables were studied as follows:

Variable	-	+
A - pad conditioning	before polish	during polish
B - polishing pad type	stacked	hard
C - table temperature	chilled	room temp.
D - slurry recirculation	none	engaged

The immediate object was to find the effect of these alternatives on the total thickness of oxide removed (in microns) during an extended polish cycle. The following results were obtained

A	B	C	D	oxide removed (μm)
-	+	+	+	0.82
-	-	+	-	0.17
-	-	-	+	0.62

A	B	C	D	oxide removed (μm)
+	-	-	-	0.30
+	-	+	+	0.68
+	+	+	-	0.50
-	+	-	-	0.38
+	+	-	+	0.93

(a) Stating any assumptions, analyze the data using a multivariate analysis of variance (MANOVA) table approach.

(b) A MANOVA table uses the F distribution. Describe equivalent tests that make use of the student-t distribution to check significance of effects. To what level of confidence is factor B significant? Factor C?

(c) Comment on the differences and similarities of this analysis compared to the basic factorial analysis method in which effects are calculated, together with the standard error of those effects.

Problem 3

The goal is to develop a response surface model for some response y as a function of two input or independent variables x_1 and x_2 . A central composite design with 5 center point replications has been performed, and the results summarized in Table 1 (where we have already normalized x_1 and x_2 to the interval -1 to +1).

Table 1: Central Composite Design and Results

Experiment	x_1	x_2	Comment	y
1	-1	-1	FF	3.7
2	-1	1	FF	0.1
3	1	-1	FF	5.5
4	1	1	FF	2.2
5	-1.41421	0	Axial	4.9
6	1.414214	0	Axial	7.5
7	0	-1.41421	Axial	2.9
8	0	1.414214	Axial	-3.2
9	0	0	Center-Ax	-0.2
10	0	0	Center-Ax	0.2
11	0	0	Center-Ax	0.2
12	0	0	Center-Ax	0.4
13	0	0	Center-Ax	-0.3

The purpose of this problem is for you to gain familiarity with a statistical or data analysis package of your choice. Reasonable alternatives include Matlab (especially with the statistics toolbox) or Splus, both of which are available on Athena. The JMP software (available on PCs and Macs) is very good for exploratory analysis. Spreadsheet programs (such as Excel) also provide somewhat more limited but usually sufficient data and statistical analysis options and graphics plotting capability.

(a) Construct a response surface model using a linear least squares fit for a model of the form:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Your analysis package should be capable of expressing the results of the regression model in an ANOVA format. What are your estimates for the three model parameters? Are the coefficients significant? What is the overall R^2 for your model?

(b) Is there any evidence of lack of fit in your model? Plot the residuals versus y , versus x_1 and versus x_2 . Does the form of the residuals suggest any additional model terms?

(c) Construct a response surface model using a linear least squares fit to the form:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2$$

Note: some packages do not explicitly support construction of higher order models; you may need to create additional columns holding the values $x_1 x_2$, x_1^2 , and x_2^2 to supply data for the regressions.

Now what are your estimates for these model coefficients? Which, if any, of the model coefficients are significant? What is the overall R^2 for your model now? Is there any evidence of lack of fit?

(d) In a pure factorial design (particularly one without replicates) if a problem arises during the execution of the experiment and we have missing data, the simple factor effect analysis is no longer easily accomplished, because we no longer have a balanced experiment. Response surface designs, however, are fairly robust to incomplete experiments or small amounts of missing data. Assume that experiment #1 failed to complete and this data point is missing. Refit the model of part (c), and compare the model coefficients to those of part (c).

Acknowledgments: Problems are inspired by C. Spanos, and by Box, Hunter, and Hunter.