# Design of Engineering Experiments Part 6 – The $2^k$ Factorial Design

- Text reference, Chapter 6
- **Special case** of the general factorial design; *k* factors, all at two levels
- The two levels are usually called **low** and **high** (they could be either quantitative or qualitative)
- Very widely used in industrial experimentation
- Form a basic "building block" for other very useful experimental designs
- Special (short-cut) methods for analysis

# The $2^k$ Factorial Design

- A complete replicate of such a design requires $2 \times 2 \times \ldots \times 2 = 2^k$ observations.
- Assume (1) the factors are fixed, (2) the designs are completely randomized, (3) the usual normality assumptions are satisfied.
- Widely used in **factor screening experiments** when many factors are to be investigated since it provides the smallest number of runs with which *k* factors can be studied in a complete factorial design.

# The Simplest Case: The $2^2$

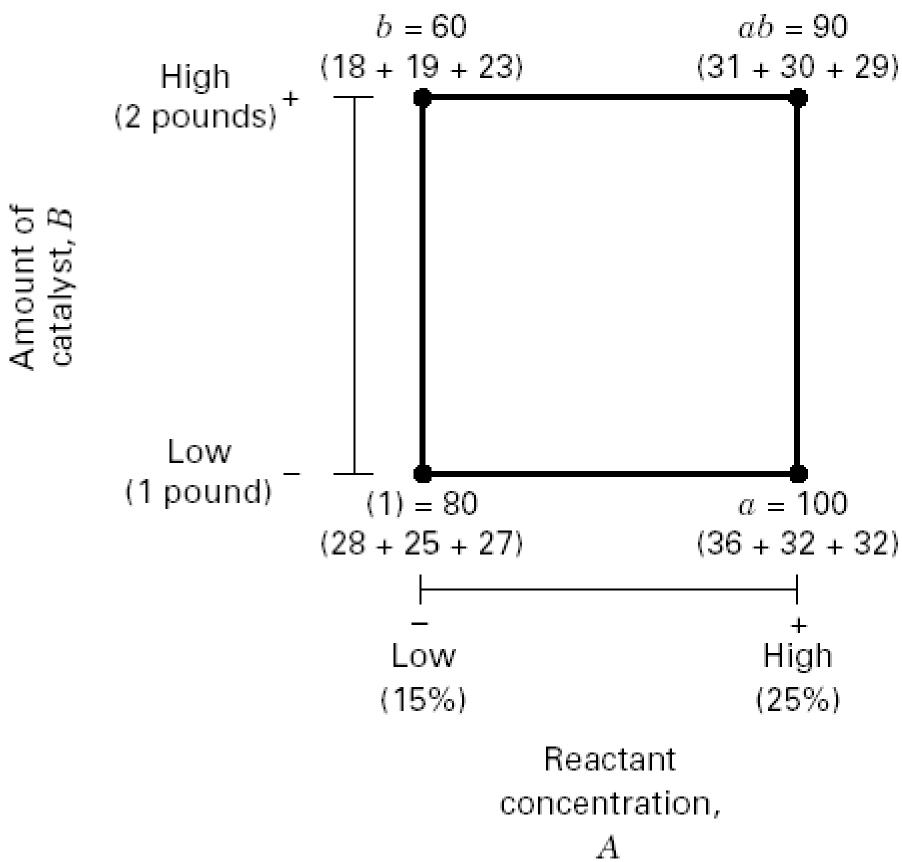



**Figure 6-1** Treatment combinations in the $2^2$ design.

“-” and “+” denote the low and high levels of a factor, respectively.

(1) - - (a) + -

(b) - + (ab) + +

Low and high are arbitrary terms

Geometrically, the four runs form the corners of a square

Factors can be quantitative or qualitative, although their treatment in the final model will be different

# Chemical Process Example

Study the effect of the reactant concentration (A) and the amount of the catalyst (B) on the conversion (yield) of a chemical process.

| *Run* | *A* | *B* | $y_{i1}$ | $y_{i2}$ | $y_{i3}$ | $\bar{y}_{i.}$ |
|---|---|---|---|---|---|---|
| 1 | − | − | 28.00 | 25.00 | 27.00 | 26.67 |
| 2 | + | − | 36.00 | 32.00 | 32.00 | 33.33 |
| 3 | − | + | 18.00 | 19.00 | 23.00 | 20.00 |
| 4 | + | + | 31.00 | 30.00 | 29.00 | 30.00 |

*A* = reactant concentration, *B* = catalyst amount, *y* = recovery

# Analysis Procedure for a Factorial Design

- Estimate factor **effects to examine the magnitude and direction to determine which variables are likely to be important.**
- **Formulate** model
  - With replication, use full model
  - With an unreplicated design, use normal probability plots
- Statistical **testing** (ANOVA)
- **Refine** the model
- Analyze **residuals** (graphical)
- **Interpret** results

# Estimation of Factor Effects

$$A = \bar{y}_{A^+} - \bar{y}_{A^-}$$

$$= \frac{(\text{ab - b}) + (\text{a - (1)})}{2}$$

$$= \frac{1}{2}[(ab + a - b - (1)]$$

$$B = \bar{y}_{B^+} - \bar{y}_{B^-}$$

$$= \frac{(\text{ab - a}) + (\text{b - (1)})}{2}$$

$$= \frac{1}{2}[(ab + b - a - (1)]$$

$$AB = \frac{(\text{ab - b}) - (\text{a - (1)})}{2}$$

$$= \frac{1}{2}[(ab + (1) - a - b]$$

See textbook, pg. 205-206 For **manual** calculations

The effect estimates are: *A* = 8.33, *B* = -5.00, *AB* = 1.67

Practical interpretation?

The effect of A (reactant concentration) is positive. The effect of B (catalyst) is negative.

The interaction effect appears to be small relative to the two main effects.

| | I | A | B | AB |
|---|---|---|---|---|
| (1) | + | - | - | + |
| a | + | + | - | - |
| b | + | - | + | - |
| ab | + | + | + | + |

## $2^2$ design

| *Obs* | *Yield* | $X_1$ | $X_2$ |
|---|---|---|---|
| 1 | 28 | −1 | −1 |
| 2 | 25 | −1 | −1 |
| 3 | 27 | −1 | −1 |
| 4 | 36 | 1 | −1 |
| 5 | 32 | 1 | −1 |
| 6 | 32 | 1 | −1 |
| 7 | 18 | −1 | 1 |
| 8 | 19 | −1 | 1 |
| 9 | 23 | −1 | 1 |
| 10 | 31 | 1 | 1 |
| 11 | 30 | 1 | 1 |
| 12 | 29 | 1 | 1 |

Table 1: The regression coding

ntgomery

$$x_1 = \frac{\text{Conc} - (\text{Conc}_{\text{low}} + \text{Conc}_{\text{high}})/2}{(\text{Conc}_{\text{high}} - \text{Conc}_{\text{low}})/2}$$

$$= \begin{cases} \dfrac{15-(15+25)/2}{(25-15)/2} = -1 \\ \dfrac{25-(15+25)/2}{(25-15)/2} = 1 \end{cases}$$

$$x_2 = \frac{\text{Catalyst} - (\text{Catalyst}_{\text{low}} + \text{Catalyst}_{\text{high}})/2}{(\text{Catalyst}_{\text{high}} - \text{Catalyst}_{\text{low}})/2}$$

$$= \begin{cases} \dfrac{1-(1+2)/2}{(2-1)/2} = -1 \\ \dfrac{2-(1+2)/2}{(2-1)/2} = 1 \end{cases}$$

## $2^2$ design

| A | B | AB |
|---|---|---|
| 8.33 | -5.00 | 1.67 |

| Coef | Estimate | SE | t | Pr(>\|t\|) |
|---|---|---|---|---|
| I | 27.50 | 0.57 | 48.14 | 0.00 |
| A | 4.17 | 0.57 | 7.29 | 0.00 |
| B | −2.50 | 0.57 | −4.38 | 0.00 |
| AB | 0.83 | 0.57 | 1.46 | 0.18 |

| | Df | SS | MS | F | Pr(>F) |
|---|---|---|---|---|---|
| A | 1.00 | 208.33 | 208.33 | 53.19 | 0.00 |
| B | 1.00 | 75.00 | 75.00 | 19.15 | 0.00 |
| AB | 1.00 | 8.33 | 8.33 | 2.13 | 0.18 |
| Error | 8.00 | 31.33 | 3.92 | | |

I (intercept) is the grand average of all 12 observations, and the regression coefficients for A,B and AB are one-half the corresponding factor effect estimates.

The least squares estimates of the model parameters $\beta$ are chosen to minimize the sum the squares of the model errors:

$$L = \sum_{i=1}^{4} \left( y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_{12} x_{i1} x_{i2} \right)^2$$

It is straightforward to show that the least squares normal equations are

$$4\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{4} x_{i1} + \hat{\beta}_2 \sum_{i=1}^{4} x_{i2} + \hat{\beta}_{12} \sum_{i=1}^{4} x_{i1} x_{i2} = (1) + a + b + ab$$

$$\hat{\beta}_0 \sum_{i=1}^{4} x_i + \hat{\beta}_1 \sum_{i=1}^{4} x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^{4} x_{i1} x_{i2} + \hat{\beta}_{12} \sum_{i=1}^{4} x_{i1}^2 x_{i2} = -(1) + a - b + ab$$

$$\hat{\beta}_0 \sum_{i=1}^{4} x_{i2} + \hat{\beta}_1 \sum_{i=1}^{4} x_{i1} x_{i2} + \hat{\beta}_2 \sum_{i=1}^{4} x_{i2}^2 + \hat{\beta}_{12} \sum_{i=1}^{4} x_{i1} x_{i2}^2 = -(1) - a + b + ab$$

$$\hat{\beta}_0 \sum_{i=1}^{4} x_{i1} x_{i2} + \hat{\beta}_1 \sum_{i=1}^{4} x_{i1}^2 x_{i2} + \hat{\beta}_2 \sum_{i=1}^{4} x_{i1} x_{i2}^2 + \hat{\beta}_{12} \sum_{i=1}^{4} x_{i1}^2 x_{i2}^2 = (1) - a - b + ab$$

Now since $\sum_{i=1}^{4} x_{i1} = \sum_{i=1}^{4} x_{i2} = \sum_{i=1}^{4} x_{i1} x_{i2} = \sum_{i=1}^{4} x_{i1}^2 x_{i2} = \sum_{i=1}^{4} x_{i1} x_{i2}^2 = 0$ because the design is orthogonal, the normal equations reduce to a very simple form:

$$4\hat{\beta}_0 = (1) + a + b + ab$$

$$4\hat{\beta}_1 = -(1) + a - b + ab$$

$$4\hat{\beta}_2 = -(1) - a + b + ab$$

$$4\hat{\beta}_{12} = (1) - a - b + ab$$

The solution is

$$\hat{\beta}_0 = \frac{[(1) + a + b + ab]}{4}$$

$$\hat{\beta}_1 = \frac{[-(1) + a - b + ab]}{4}$$

$$\hat{\beta}_2 = \frac{[-(1) - a + b + ab]}{4}$$

$$\hat{\beta}_{12} = \frac{[(1) - a - b + ab]}{4}$$

# A regression model for predicting yield based on ANOVA analysis

Coefficients:

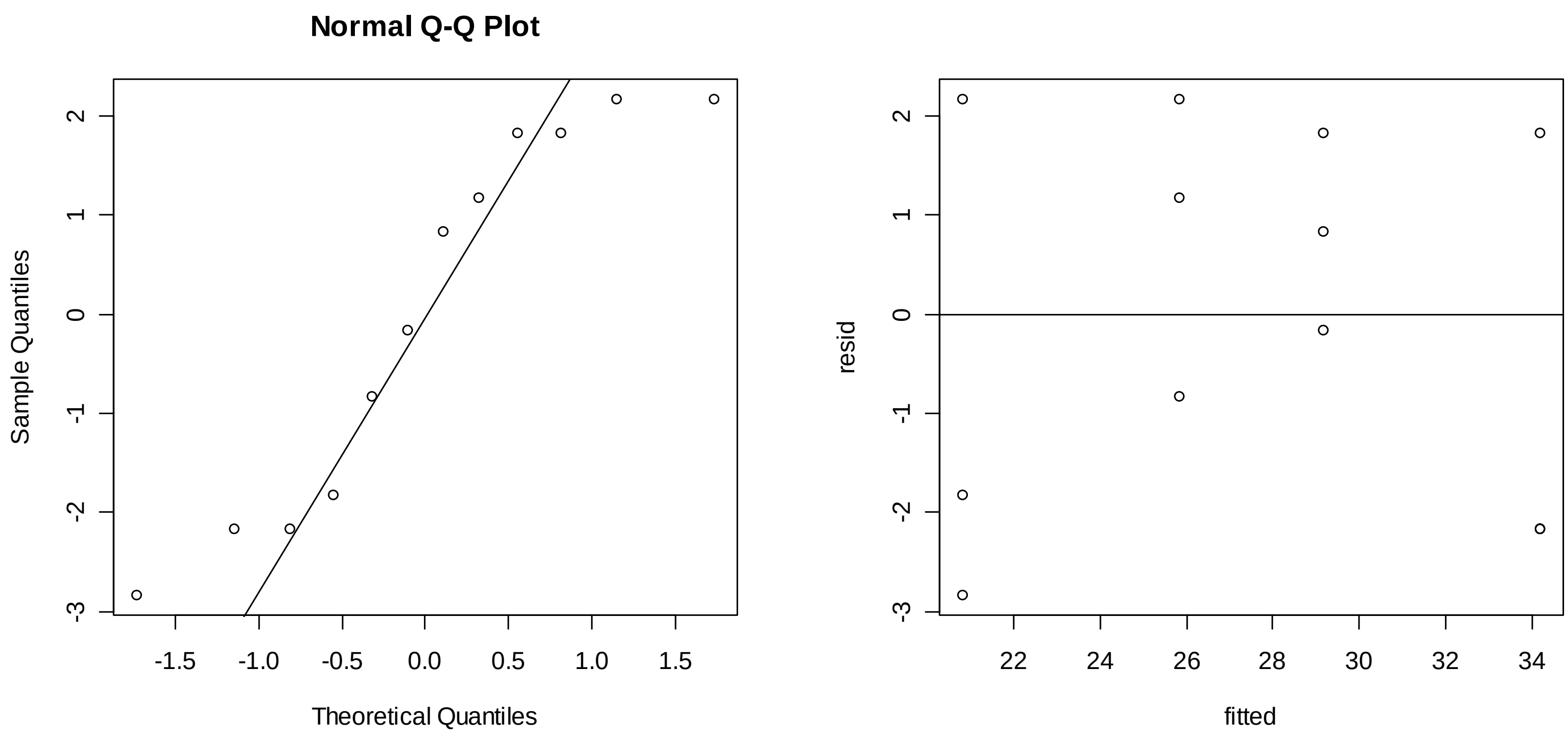
| | Estimate | Std. Error | t value | Pr(>\|t\|) | |
|---|---|---|---|---|---|
| (Intercept) | 27.500 | 0.606 | 45.377 | 6.13e-12 | *** |
| X1 | 4.167 | 0.606 | 6.875 | 7.27e-05 | *** |
| X2 | -2.500 | 0.606 | -4.125 | 0.00258 | ** |

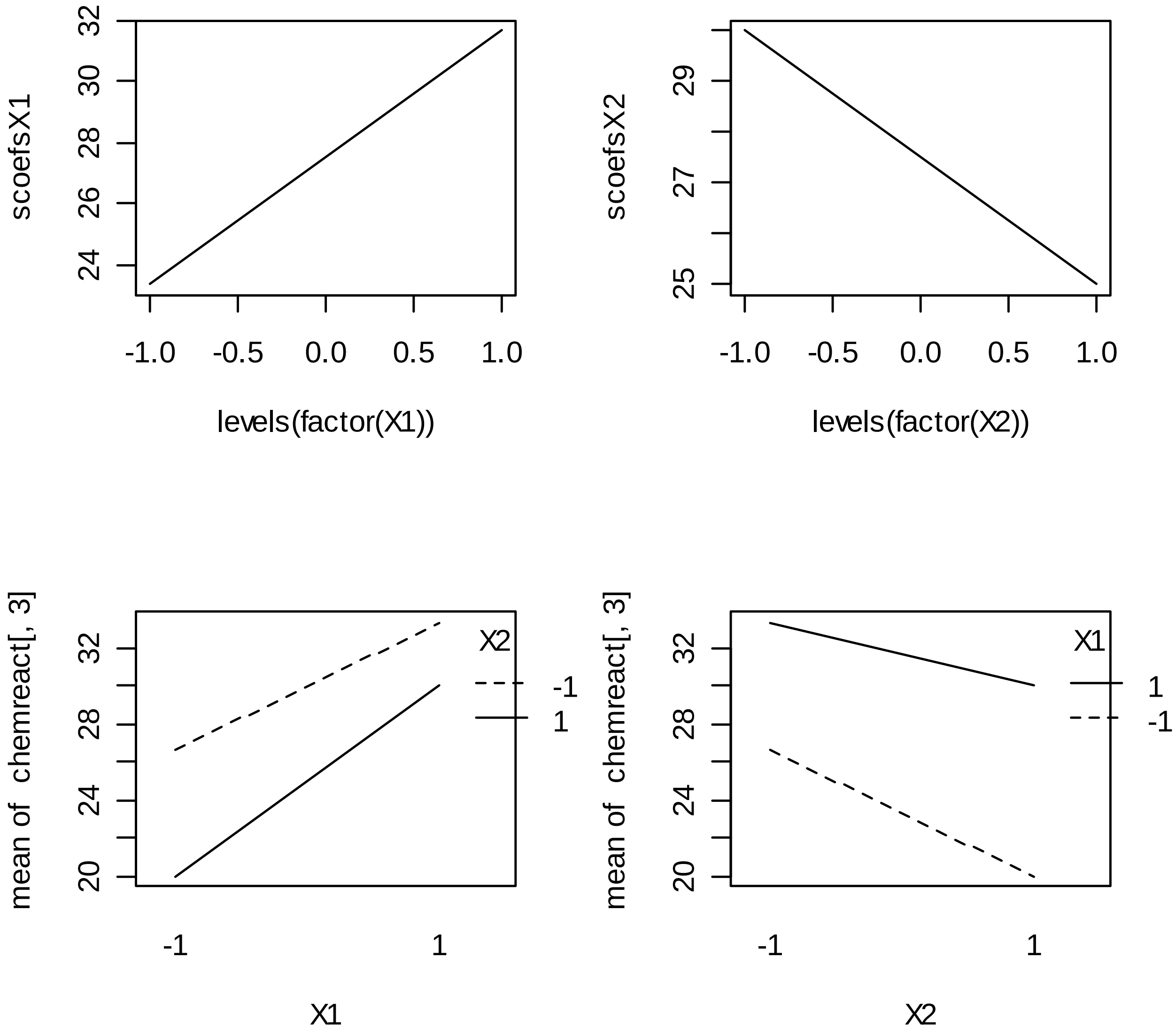
---

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$$\hat{y} = 27.5 + 4.167X_1 + (-2.5)X_2$$

# Residuals and Diagnostic Checking

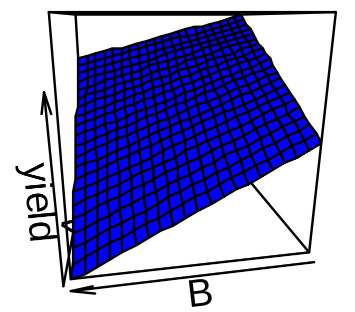


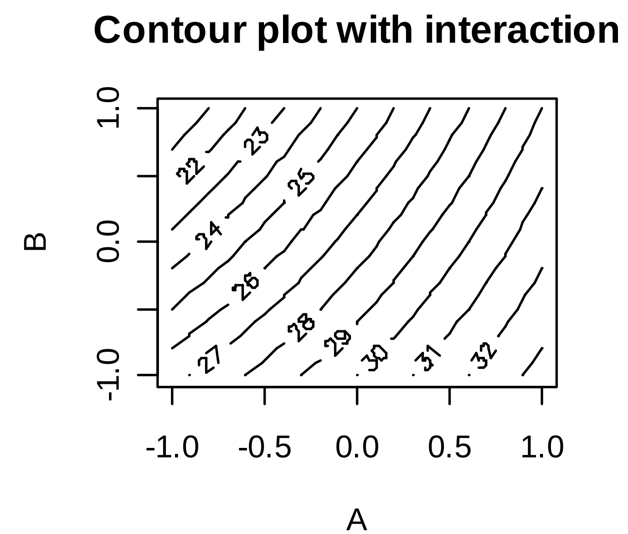


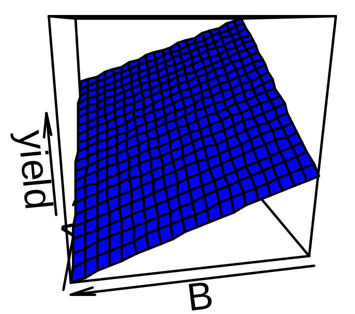

Since Concentration has positive effect and Catalyst has negative effect, set Concentration at the high level and Catalyst at low level to increase yield.

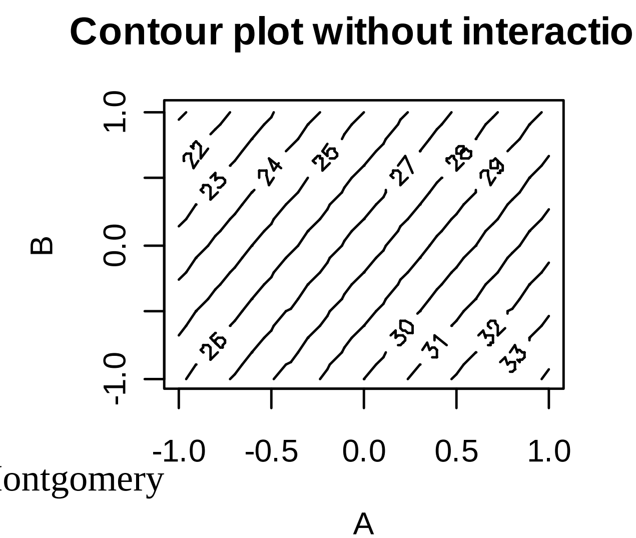
# The Response Surface

The contour plot also provide the insight that setting concentration at high level and catalyst at low level would increase yield

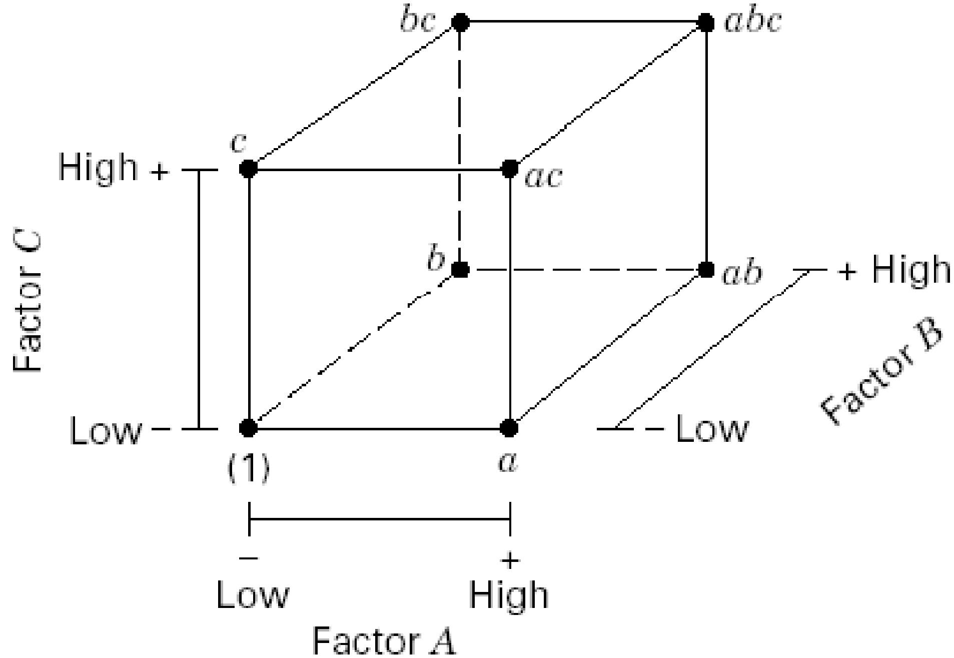


# The $2^3$ Factorial Design




(a) Geometric view

| Run | Factor A | B | C |
|---|---|---|---|
| 1 | − | − | − |
| 2 | + | − | − |
| 3 | − | + | − |
| 4 | + | + | − |
| 5 | − | − | + |
| 6 | + | − | + |
| 7 | − | + | + |
| 8 | + | + | + |

(b) The design matrix

**Figure 6-4** The $2^3$ factorial design.

# An Example of a $2^3$ Factorial Design

Table 6-4 The Plasma Etch Experiment, Example 6-1

| | Coded Factors | | | Etch Rate | | | Factor Levels | | |
|---|---|---|---|---|---|---|---|---|---|
| Run | $A$ | $B$ | $C$ | Replicate 1 | Replicate 2 | Total | | Low (−1) | High (+1) |
| 1 | −1 | −1 | −1 | 550 | 604 | (1) = 1154 | $A$ (Gap, cm) | 0.80 | 1.20 |
| 2 | 1 | −1 | −1 | 669 | 650 | $a$ = 1319 | $B$ ($C_2F_6$ flow, SCCM) | 125 | 200 |
| 3 | −1 | 1 | −1 | 633 | 601 | $b$ = 1234 | $C$ (Power, W) | 275 | 325 |
| 4 | 1 | 1 | −1 | 642 | 635 | $ab$ = 1277 | | | |
| 5 | −1 | −1 | 1 | 1037 | 1052 | $c$ = 2089 | | | |
| 6 | 1 | −1 | 1 | 749 | 868 | $ac$ = 1617 | | | |
| 7 | −1 | 1 | 1 | 1075 | 1063 | $bc$ = 2178 | | | |
| 8 | 1 | 1 | 1 | 729 | 860 | $abc$ = 1589 | | | |

$A$ = gap, $B$ = Flow, $C$ = Power, $y$ = Etch Rate

# Table of – and + Signs for the $2^3$ Factorial Design (pg. 214)

Table 6-3 Algebraic Signs for Calculating Effects in the $2^3$ Design

| Treatment Combination | Factorial Effect | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *I* | *A* | *B* | *AB* | *C* | *AC* | *BC* | *ABC* |
| (1) | + | − | − | + | − | + | + | − |
| *a* | + | + | − | − | − | − | + | + |
| *b* | + | − | + | − | − | + | − | + |
| *ab* | + | + | + | + | − | − | − | − |
| *c* | + | − | − | + | + | − | − | + |
| *ac* | + | + | − | − | + | + | − | − |
| *bc* | + | − | + | − | + | − | + | − |
| *abc* | + | + | + | + | + | + | + | + |

# Properties of the Table

- Except for column *I*, every column has an equal number of + and – signs
- The sum of the product of signs in any two columns is zero
- Multiplying any column by *I* leaves that column unchanged (identity element)
- The product of any two columns yields a column in the table:

$$A \times B = AB$$

$$AB \times BC = AB^2C = AC$$

- **Orthogonal** design
- Orthogonality is an important property shared by all factorial designs

# Estimation of Factor Effects

Coefficients:

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 76.062 | 11.865 | 65.406 | 3.32e-12 *** |
| A | -50.813 | 11.865 | -4.282 | 0.002679 ** |
| B | 3.687 | 11.865 | 0.311 | 0.763911 |
| C | 153.062 | 11.865 | 12.900 | 1.23e-06 *** |
| A:B | -12.437 | 11.865 | -1.048 | 0.325168 |
| A:C | -76.812 | 11.865 | -6.474 | 0.000193 *** |
| B:C | -1.063 | 11.865 | -0.090 | 0.930849 |
| A:B:C | 2.812 | 11.865 | 0.237 | 0.818586 |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 47.46 on 8 degrees of freedom

Multiple R-Squared: 0.9661, Adjusted R-squared: 0.9364

F-statistic: 32.56 on 7 and 8 DF, p-value: 2.896e-05

# Model Summary Statistics (pg. 222)

- **Standard error** of model coefficients (full model)

$$se(\hat{\beta}) = \sqrt{V(\hat{\beta})} = \sqrt{\frac{\sigma^2}{n2^k}} = \sqrt{\frac{MS_E}{n2^k}} = \sqrt{\frac{2252.56}{2(8)}} = 11.87$$

- **Confidence interval** on model coefficients

$$\hat{\beta} - t_{\alpha/2,df_E} se(\hat{\beta}) \le \beta \le \hat{\beta} + t_{\alpha/2,df_E} se(\hat{\beta})$$

# ANOVA Summary – Full Model

Table 6-6 Analysis of Variance for the Plasma Etching Experiment

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | $F_0$ | *P*-Value |
|---|---|---|---|---|---|
| Gap (*A*) | 41,310.5625 | 1 | 41,310.5625 | 18.34 | 0.0027 |
| Gas flow (*B*) | 217.5625 | 1 | 217.5625 | 0.10 | 0.7639 |
| Power (*C*) | 374,850.0625 | 1 | 374,850.0625 | 166.41 | 0.0001 |
| *AB* | 2475.0625 | 1 | 2475.0625 | 1.10 | 0.3252 |
| *AC* | 94,402.5625 | 1 | 94,402.5625 | 41.91 | 0.0002 |
| *BC* | 18.0625 | 1 | 18.0625 | 0.01 | 0.9308 |
| *ABC* | 126.5625 | 1 | 126.5625 | 0.06 | 0.8186 |
| Error | 18,020.5000 | 8 | 2252.5625 | | |
| Total | 531,420.9375 | 15 | | | |

# Refine Model – Remove Nonsignificant Factors

Analysis of Variance Table

Response: EtchRate

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) | |
|---|---|---|---|---|---|---|
| A | 1 | 41311 | 41311 | 23.767 | 0.0003816 | *** |
| C | 1 | 374850 | 374850 | 215.661 | 4.951e-09 | *** |
| A:C | 1 | 94403 | 94403 | 54.312 | 8.621e-06 | *** |
| Residuals | 12 | 20858 | 1738 | | | |

---

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Model Coefficients – Reduced Model

Coefficients:

| | Estimate | Std. Error | t value | Pr(>\|t\|) | |
|---|---|---|---|---|---|
| (Intercept) | 776.06 | 10.42 | 74.458 | < 2e-16 | *** |
| A | -50.81 | 10.42 | -4.875 | 0.000382 | *** |
| C | 153.06 | 10.42 | 14.685 | 4.95e-09 | *** |
| A:C | -76.81 | 10.42 | -7.370 | 8.62e-06 | *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 41.69 on 12 degrees of freedom
Multiple R-Squared: 0.9608, Adjusted R-squared: 0.9509
F-statistic: 97.91 on 3 and 12 DF, p-value: 1.054e-08

## Model Summary Statistics for Reduced Model (pg. 222)

- $R^2$ and adjusted $R^2$

$$R^2 = \frac{SS_{Model}}{SS_T} = \frac{5.106\times 10^5}{5.314\times 10^5} = 0.9608$$

$$R^2_{Adj} = 1 - \frac{SS_E / df_E}{SS_T / df_T} = 1 - \frac{20857.75/12}{5.314\times 10^5 /15} = 0.9509$$

- $R^2$ for prediction (based on PRESS)

$$R^2_{\text{Pred}} = 1 - \frac{PRESS}{SS_T} = 1 - \frac{37080.44}{5.314\times 10^5} = 0.9302$$

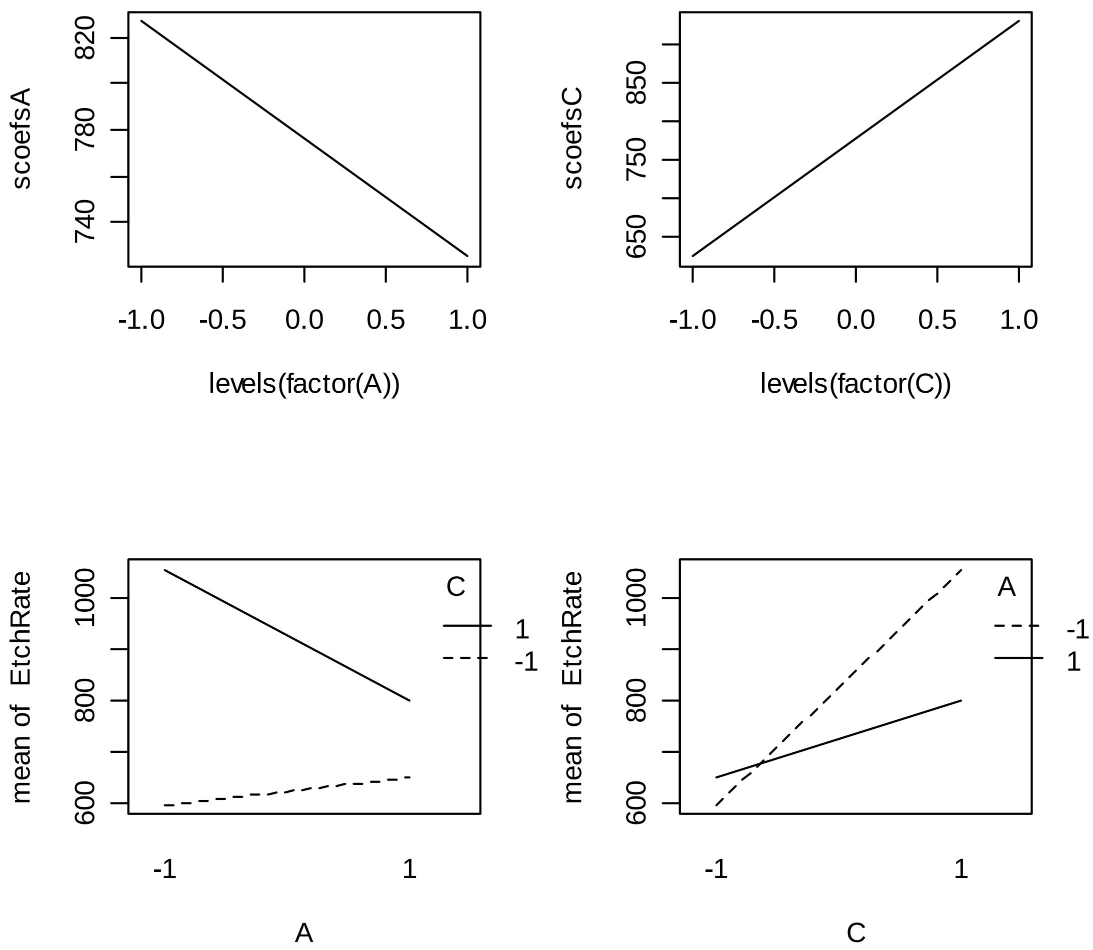
# The Regression Model

**Final Equation in Terms of Coded Factors:**

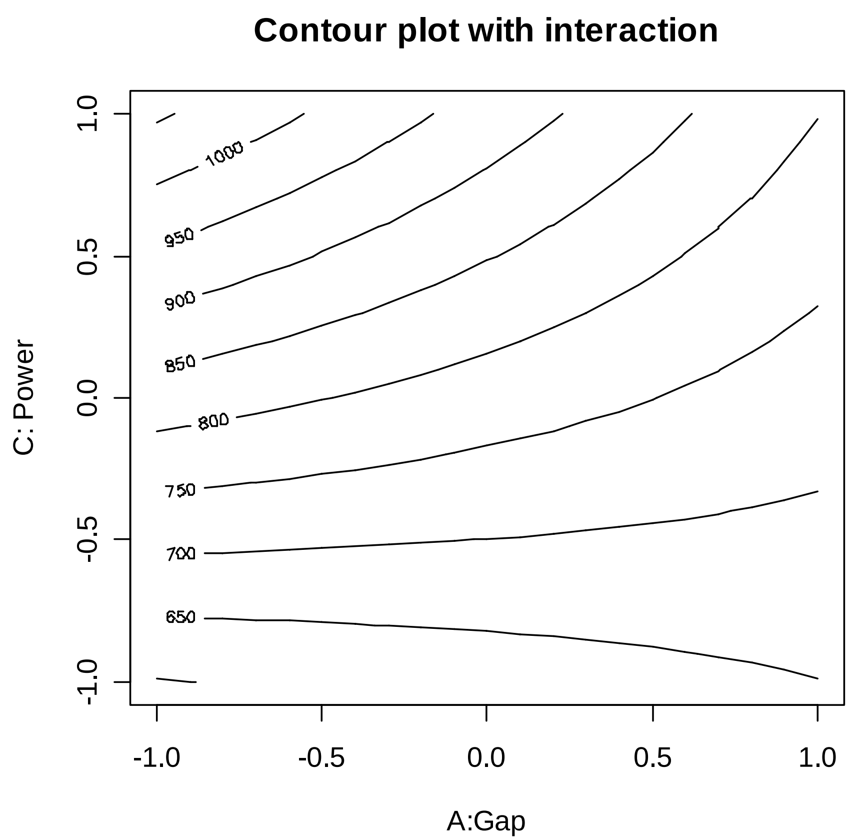
| | |
|---|---|
| Etch rate | = |
| +776.06 | |
| −50.81 | * *A* |
| +153.06 | * *C* |
| −76.81 | * *A* * *C* |

**Final Equation in Terms of Actual Factors:**

| | |
|---|---|
| Etch rate | = |
| −5415.37500 | |
| +4354.68750 | * Gap |
| +21.48500 | * Power |
| −15.36250 | * Gap * Power |




The interaction plot reveals that etch rate would be maximized with C(Power) at high level and A (Gap) at low level




- Because the model contains interaction, the contour lines are curved
- If it is desired to operate this process so that the etch rate is close to 900,. The contour plot shows that several combinations of gap and power will satisfy this objective
- However, it will be necessary to control both of these variables very precisely.

## $2^4$ design

| *Run* | *A* | *B* | *C* | *D* | $y_{i1}$ | $y_{i2}$ | $\bar{y}_{i\cdot}$ |
|---|---|---|---|---|---|---|---|
| 1 | − | − | − | − | 7.04 | 6.38 | 6.71 |
| 2 | + | − | − | − | 14.71 | 15.22 | 14.96 |
| 3 | − | + | − | − | 11.63 | 12.09 | 11.86 |
| 4 | + | + | − | − | 17.27 | 17.82 | 17.54 |
| 5 | − | − | + | − | 10.40 | 10.15 | 10.28 |
| 6 | + | − | + | − | 4.37 | 4.10 | 4.23 |
| 7 | − | + | + | − | 9.36 | 9.25 | 9.31 |
| 8 | + | + | + | − | 13.44 | 12.92 | 13.18 |
| 9 | − | − | − | + | 8.56 | 8.95 | 8.76 |
| 10 | + | − | − | + | 16.87 | 17.05 | 16.96 |
| 11 | − | + | − | + | 13.88 | 13.66 | 13.77 |
| 12 | + | + | − | + | 19.82 | 19.64 | 19.73 |
| 13 | − | − | + | + | 11.85 | 12.34 | 12.09 |
| 14 | + | − | + | + | 6.12 | 5.90 | 6.01 |
| 15 | − | + | + | + | 11.19 | 10.94 | 11.06 |
| 16 | + | + | + | + | 15.65 | 15.05 | 15.35 |

## $2^4$ design

Main effects

| A | B | C | D |
|---|---|---|---|
| 3.02 | 3.98 | −3.60 | 1.96 |

Two-way interactions

| AB | AC | AD | BC | BD | CD |
|---|---|---|---|---|---|
| 1.93 | −4.01 | 0.08 | 0.10 | 0.05 | −0.08 |

Three-way interactions

| ABC | ABD | ACD | BCD |
|---|---|---|---|
| 3.14 | 0.10 | 0.02 | 0.04 |

Four-way interaction

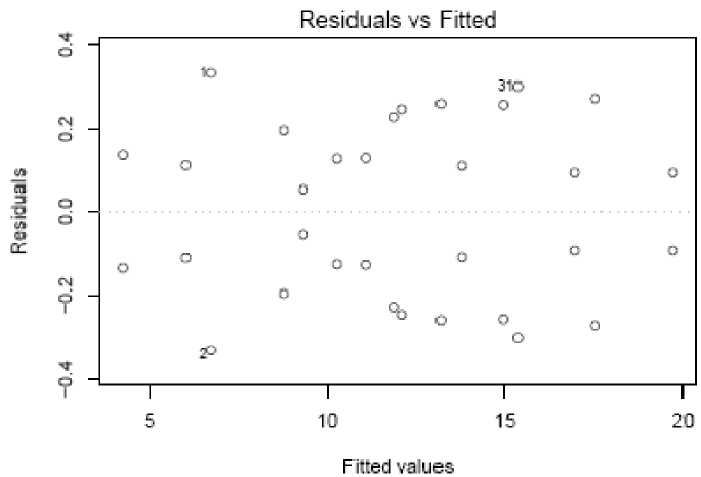
| ABCD |
|---|
| 0.01 |

## $2^4$ design

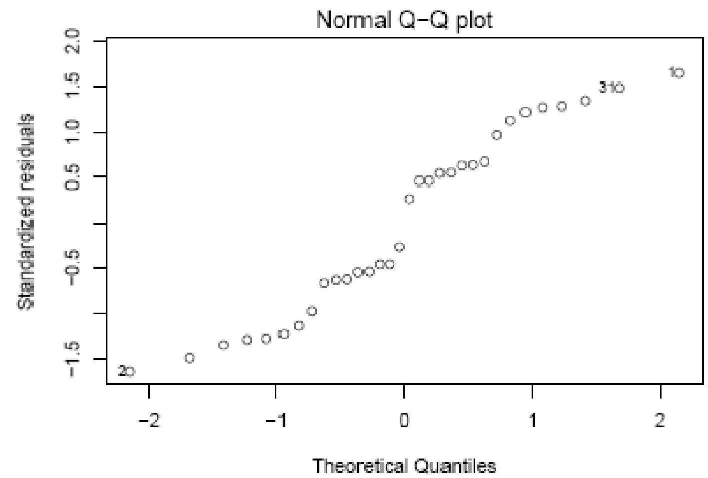
| Coef | Estimate | SE | t | Pr(>\|t\|) |
|---|---|---|---|---|
| I | 11.99 | 0.05 | 238.04 | 0.00 |
| A | 1.51 | 0.05 | 29.97 | 0.00 |
| B | 1.99 | 0.05 | 39.47 | 0.00 |
| C | −1.80 | 0.05 | −35.70 | 0.00 |
| D | 0.98 | 0.05 | 19.44 | 0.00 |
| AB | 0.97 | 0.05 | 19.20 | 0.00 |
| AC | −2.00 | 0.05 | −39.79 | 0.00 |
| AD | 0.04 | 0.05 | 0.76 | 0.46 |
| BC | 0.05 | 0.05 | 0.95 | 0.35 |
| BD | 0.02 | 0.05 | 0.47 | 0.65 |
| CD | −0.04 | 0.05 | −0.76 | 0.46 |
| ABC | 1.57 | 0.05 | 31.15 | 0.00 |
| ABD | 0.05 | 0.05 | 0.97 | 0.35 |
| ACD | 0.01 | 0.05 | 0.19 | 0.85 |
| BCD | 0.02 | 0.05 | 0.35 | 0.73 |
| ABCD | 0.01 | 0.05 | 0.14 | 0.89 |

## $2^4$ design

| | Df | SS | MS | F | Pr(>F) |
|---|---|---|---|---|---|
| A | 1 | 72.909 | 72.909 | 898.34 | 0.000 |
| B | 1 | 126.461 | 126.461 | 1558.17 | 0.000 |
| C | 1 | 103.464 | 103.464 | 1274.82 | 0.000 |
| D | 1 | 30.662 | 30.662 | 377.80 | 0.000 |
| AB | 1 | 29.927 | 29.927 | 368.74 | 0.000 |
| AC | 1 | 128.496 | 128.496 | 1583.26 | 0.000 |
| AD | 1 | 0.047 | 0.047 | 0.577 | 0.46 |
| BC | 1 | 0.074 | 0.074 | 0.908 | 0.36 |
| BD | 1 | 0.018 | 0.018 | 0.220 | 0.645 |
| CD | 1 | 0.047 | 0.047 | 0.583 | 0.46 |
| ABC | 1 | 78.751 | 78.751 | 970.33 | 0.000 |
| ABD | 1 | 0.077 | 0.077 | 0.95 | 0.345 |
| ACD | 1 | 0.003 | 0.003 | 0.04 | 0.852 |
| BCD | 1 | 0.010 | 0.010 | 0.13 | 0.728 |
| ABCD | 1 | 0.002 | 0.002 | 0.02 | 0.890 |
| Error | 16 | 1.299 | 0.081 | | |



Residuals vs Fitted
Residuals
Fitted values

Normal Q-Q plot
Standardized residuals
Theoretical Quantiles