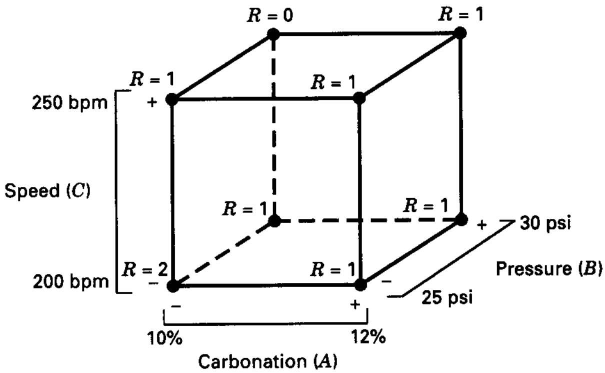

Figure 6-8 Ranges of fill height deviation for Example 6-1.

to run? One way to answer the question is to look at the **range** of fill height deviations for each of the eight runs in the 2^3 design. These ranges are plotted on the cube in Figure 6-8. Notice that the ranges are about the same for all eight runs in the design. Consequently, there is no strong evidence indicating that some of the process variables directly affect the variability in fill height deviation in the process.

6-4 THE GENERAL 2^k DESIGN

The methods of analysis that we have presented thus far may be generalized to the case of a **2^k factorial design**, that is, a design with k factors each at two levels. The statistical model for a 2^k design would include k main effects, $\binom{k}{2}$ two-factor interactions, $\binom{k}{3}$ three-factor interactions, . . . , and one k-factor interaction. That is, for a 2^k design the complete model would contain $2^k - 1$ effects. The notation introduced earlier for treatment combinations is also used here. For example, in a 2^5 design *abd* denotes the treatment combination with factors A, B, and D at the high level and factors C and E at the low level. The treatment combinations may be written in **standard order** by introducing the factors one at a time, with each new factor being successively combined with those that precede it. For example, the standard order for a 2^4 design is (1), *a*, *b*, *ab*, *c*, *ac*, *bc*, *abc*, *d*, *ad*, *bd*, *abd*, *cd*, *acd*, *bcd*, and *abcd*.

The general approach to the statistical analysis of the 2^k design is summarized in Table 6-8. As we have indicated previously, a computer software package is usually employed in this analysis process.

The sequence of steps in Table 6-8 should, by now, be familiar. The first step is to estimate factor effects and examine their signs and magnitudes. This gives the experi-

Table 6-8 Analysis Procedure for a 2^k Design

1. Estimate factor effects
2. Form initial model
3. Perform statistical testing
4. Refine model
5. Analyze residuals
6. Interpret results

menter preliminary information regarding which factors and interactions may be important, and in which directions these factors should be adjusted to improve the response. In forming the initial model for the experiment, we usually choose the **full model**, that is, all main effects and interactions, provided that at least one of the design points has been replicated (in the next section, we discuss a modification to this step). Then in step 3, we use the analysis of variance to formally test for significance of main effects and interaction. Table 6-9 shows the general form of an analysis of variance for a 2^k factorial design with n replicates. Step 4, refine the model, usually consists of removing any nonsignificant variables from the full model. Step 5 is the usual residual analysis to check for model adequacy and to check assumptions. Sometimes model refinement will occur after residual analysis, if we find that the model is inadequate or assumptions are badly violated. The final step usually consists of graphical analysis—either main effect or interaction plots, or response surface and contour plots.

Although the calculations described above are almost always done with a computer, occasionally it is necessary to manually calculate an effect estimate or sum of squares for an effect. To estimate an effect or to compute the sum of squares for an effect, we must first determine the contrast associated with that effect. This can always be done by using a table of plus and minus signs, such as Table 6-2 or Table 6-3. However, for large values of k this is awkward, and we can use an alternate method. In general, we determine the contrast for effect $AB \cdots K$ by expanding the right-hand side of

$$\text{Contrast}_{AB\cdots K} = (a \pm 1)(b \pm 1) \cdots (k \pm 1) \tag{6-21}$$

Table 6-9 Analysis of Variance for a 2^k Design

Source of Variation	Sum of Squares	Degrees of Freedom
k main effects		
A	SS_A	1
B	SS_B	1
$\vdots$	$\vdots$	$\vdots$
K	SS_K	1
$\binom{k}{2}$ two-factor interactions		
AB	SS_{AB}	1
AC	SS_{AC}	1
$\vdots$	$\vdots$	$\vdots$
JK	SS_{JK}	1
$\binom{k}{3}$ three-factor interactions		
ABC	SS_{ABC}	1
ABD	SS_{ABD}	1
$\vdots$	$\vdots$	$\vdots$
IJK	SS_{IJK}	1
$\vdots$	$\vdots$	$\vdots$
$\binom{k}{k} = 1$ k-factor interaction		
$ABC \cdots K$	$SS_{ABC\cdots K}$	1
Error	SS_E	$2^k(n-1)$
Total	SS_T	$n2^k - 1$

In expanding Equation 6-21, ordinary algebra is used with "1" being replaced by (1) in the final expression. The sign in each set of parentheses is negative if the factor is included in the effect and positive if the factor is not included.

To illustrate the use of Equation 6-21, consider a 2^3 factorial design. The contrast for *AB* would be

$$\begin{aligned}\text{Contrast}_{AB} &= (a - 1)(b - 1)(c + 1)\\ &= abc + ab + c + (1) - ac - bc - a - b\end{aligned}$$

As a further example, in a 2^5 design, the contrast for *ABCD* would be

$$\begin{aligned}\text{Contrast}_{ABCD} &= (a - 1)(b - 1)(c - 1)(d - 1)(e + 1)\\ &= abcde + cde + bde + ade + bce\\ &\quad + ace + abe + e + abcd + cd + bd\\ &\quad + ad + bc + ac + ab + (1) - a - b - c\\ &\quad - abc - d - abd - acd - bcd - ae\\ &\quad - be - ce - abce - de - abde - acde - bcde\end{aligned}$$

Once the contrasts for the effects have been computed, we may estimate the effects and compute the sums of squares according to

$$AB \cdots K = \frac{2}{n2^k}(\text{Contrast}_{AB \cdots K}) \tag{6-22}$$

and

$$SS_{AB \cdots K} = \frac{1}{n2^k}(\text{Contrast}_{AB \cdots K})^2 \tag{6-23}$$

respectively, where n denotes the number of replicates. There is also a tabular algorithm due to Dr. Frank Yates that can occasionally be useful for manual calculation of the effect estimates and the sums of squares. Refer to the supplemental text material for this chapter.

6-5 A SINGLE REPLICATE OF THE 2^k DESIGN

For even a moderate number of factors, the total number of treatment combinations in a 2^k factorial design is large. For example, a 2^5 design has 32 treatment combinations, a 2^6 design has 64 treatment combinations, and so on. Because resources are usually limited, the number of replicates that the experimenter can employ may be restricted. Frequently, available resources only allow a **single replicate** of the design to be run, unless the experimenter is willing to omit some of the original factors.

An obvious risk when conducting an experiment that has only one run at each test combination is that we may be fitting a model to noise. That is, if the response y is highly variable, misleading conclusions may result from the experiment. The situation is illustrated in Figure 6-9*a*. In this figure, the straight line represents the true factor effect. However, because of the random variability present in the response variable (represented by the shaded band), the experimenter actually obtains the two measured responses represented by the dark dots. Consequently, the estimated factor effect is close to zero, and the experimenter has reached an erroneous conclusion concerning this factor. Now if

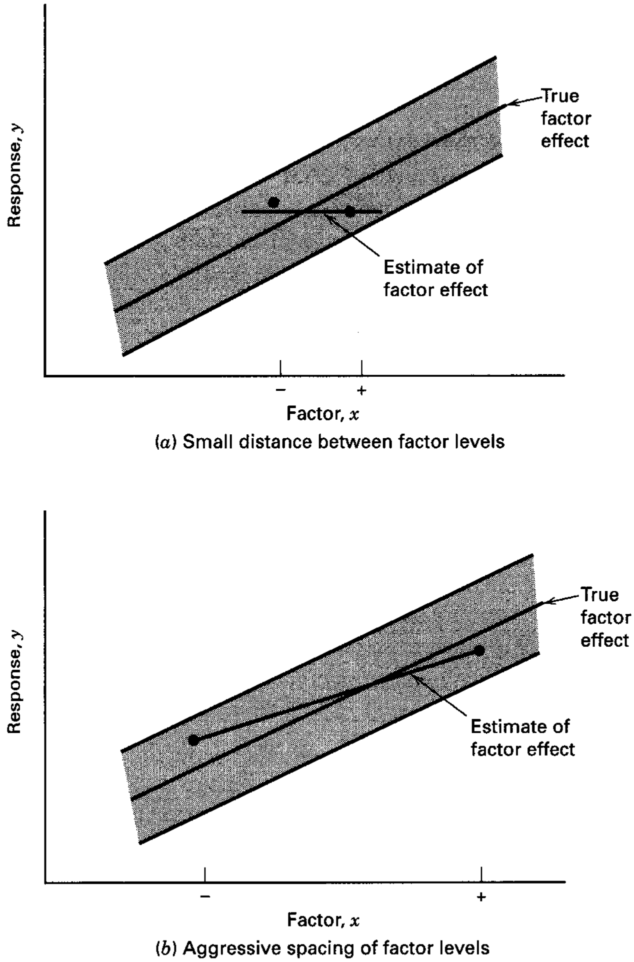

Figure 6-9 The impact of the choice of factor levels in an unreplicated design.

there is less variability in the response, the likelihood of an erroneous conclusion will be smaller. Another way to ensure that reliable effect estimates are obtained is to increase the distance between the low (−) and high (+) levels of the factor, as illustrated in Figure 6-9*b*. Notice that in this figure, the increased distance between the low and high factor levels results in a reasonable estimate of the true factor effect.

The single-replicate strategy is often used in screening experiments when there are relatively many factors under consideration. Because we can never be entirely certain in such cases that the experimental error is small, a good practice in these types of experiments is to spread out the factor levels aggressively. You might find it helpful to re-read the guidance on choosing factor levels in Chapter 1.

A single replicate of a 2^k design is sometimes called an **unreplicated factorial**. With only one replicate, there is no internal estimate of error (or "pure error"). One approach to the analysis of an unreplicated factorial is to assume that certain high-order interactions are negligible and combine their mean squares to estimate the error. This is an appeal to the **sparsity of effects principle**; that is, most systems are dominated by some of the main effects and low-order interactions, and most high-order interactions are negligible.

When analyzing data from unreplicated factorial designs, occasionally real high-order interactions occur. The use of an error mean square obtained by pooling high-order interactions is inappropriate in these cases. A method of analysis attributed to Daniel (1959) provides a simple way to overcome this problem. Daniel suggests examining a **normal probability plot** of the estimates of the effects. The effects that are negligible are normally distributed, with mean zero and variance σ^2 and will tend to fall along a straight line on this plot, whereas significant effects will have nonzero means and will not lie along the straight line. Thus the preliminary model will be specified to contain those effects that are apparently nonzero, based on the normal probability plot. The apparently negligible effects are combined as an estimate of error.

EXAMPLE 6-2 ....

A Single Replicate of the 2^4 Design

A chemical product is produced in a pressure vessel. A factorial experiment is carried out in the pilot plant to study the factors thought to influence the filtration rate of this product. The four factors are temperature (A), pressure (B), concentration of formaldehyde (C), and stirring rate (D). Each factor is present at two levels. The design matrix and the response data obtained from a single replicate of the 2^4 experiment are shown in Table 6-10 and Figure 6-10. The 16 runs are made in random order. The process engineer is interested in maximizing the filtration rate. Current process conditions give filtration rates of around 75 gal/h. The process also currently uses the concentration of formaldehyde, factor C, at the high level. The engineer would like to reduce the formaldehyde concentration as much as possible but has been unable to do so because it always results in lower filtration rates.

We will begin the analysis of this data by constructing a normal probability plot of the effect estimates. The table of plus and minus signs for the contrast constants for the

Table 6-10 Pilot Plant Filtration Rate Experiment

Run Number	Factor A	B	C	D	Run Label	Filtration Rate (gal/h)
1	−	−	−	−	(1)	45
2	+	−	−	−	a	71
3	−	+	−	−	b	48
4	+	+	−	−	ab	65
5	−	−	+	−	c	68
6	+	−	+	−	ac	60
7	−	+	+	−	bc	80
8	+	+	+	−	abc	65
9	−	−	−	+	d	43
10	+	−	−	+	ad	100
11	−	+	−	+	bd	45
12	+	+	−	+	abd	104
13	−	−	+	+	cd	75
14	+	−	+	+	acd	86
15	−	+	+	+	bcd	70
16	+	+	+	+	$abcd$	96

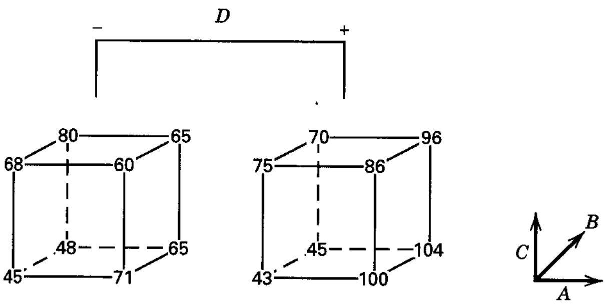

Figure 6-10 Data from the pilot plant filtration rate experiment for Example 6-2.

2^4 design are shown in Table 6-11 (page 248). From these contrasts, we may estimate the 15 factorial effects and the sums of squares shown in Table 6-12 (page 249).

The normal probability plot of these effects is shown in Figure 6-11 (page 249). All of the effects that lie along the line are negligible, whereas the large effects are far from the line. The important effects that emerge from this analysis are the main effects of A, C, and D and the AC and AD interactions.

The main effects of A, C, and D are plotted in Figure 6-12*a* (page 250). All three effects are positive, and if we considered only these main effects, we would run all three factors at the high level to maximize the filtration rate. However, it is always necessary to examine any interactions that are important. Remember that main effects do not have much meaning when they are involved in significant interactions.

The AC and AD interactions are plotted in Figure 6-12*b*. These interactions are the key to solving the problem. Note from the AC interaction that the temperature effect is very small when the concentration is at the high level and very large when the concentration is at the low level, with the best results obtained with low concentration and high temperature. The AD interaction indicates that stirring rate D has little effect at low temperature but a large positive effect at high temperature. Therefore, the best filtration rates would appear to be obtained when A and D are at the high level and C is at the low level. This would allow the reduction of the formaldehyde concentration to a lower level, another objective of the experimenter.

Design Projection

Another interpretation of the effects in Figure 6-11 is possible. Because B (pressure) is not significant and all interactions involving B are negligible, we may discard B from the experiment so that the design becomes a 2^3 factorial in A, C, and D with two replicates. This is easily seen from examining only columns A, C, and D in the design matrix shown in Table 6-10 and noting that those columns form two replicates of a 2^3 design. The analysis of variance for the data using this simplifying assumption is summarized in Table 6-13 (page 250). The conclusions that we would draw from this analysis are essentially unchanged from those of Example 6-2. Note that by projecting the single replicate of the 2^4 into a replicated 2^3, we now have both an estimate of the ACD interaction and an estimate of error based on what is sometimes called **hidden replication.**

Table 6-11 Contrast Constants for the 2^4 Design

	A	B	AB	C	AC	BC	ABC	D	AD	BD	ABD	CD	ACD	BCD	$ABCD$
(1)	−	−	+	−	+	+	−	−	+	+	−	+	−	−	+
a	+	−	−	−	−	+	+	−	−	+	+	+	+	−	−
b	−	+	−	−	+	−	+	−	+	−	+	+	−	+	−
ab	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+
c	−	−	+	+	−	−	+	−	+	+	−	−	+	+	−
ac	+	−	−	+	+	−	−	−	−	+	+	−	−	+	+
bc	−	+	−	+	−	+	−	−	+	−	+	−	+	−	+
abc	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−
d	−	−	+	−	+	+	−	+	−	−	+	−	+	+	−
ad	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+
bd	−	+	−	−	+	−	+	+	−	+	−	−	+	−	+
abd	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−
cd	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+
acd	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−
bcd	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−
abcd	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Table 6-12 Factor Effect Estimates and Sums of Squares for the 2^4 Factorial in Example 6-2

Model Term	Effect Estimate	Sum of Squares	Percent Contribution
A	21.625	1870.56	32.6397
B	3.125	39.0625	0.681608
C	9.875	390.062	6.80626
D	14.625	855.563	14.9288
AB	0.125	0.0625	0.00109057
AC	−18.125	1314.06	22.9293
AD	16.625	1105.56	19.2911
BC	2.375	22.5625	0.393696
BD	−0.375	0.5625	0.00981515
CD	−1.125	5.0625	0.0883363
ABC	1.875	14.0625	0.245379
ABD	4.125	68.0625	1.18763
ACD	−1.625	10.5625	0.184307
BCD	−2.625	27.5625	0.480942
ABCD	1.375	7.5625	0.131959

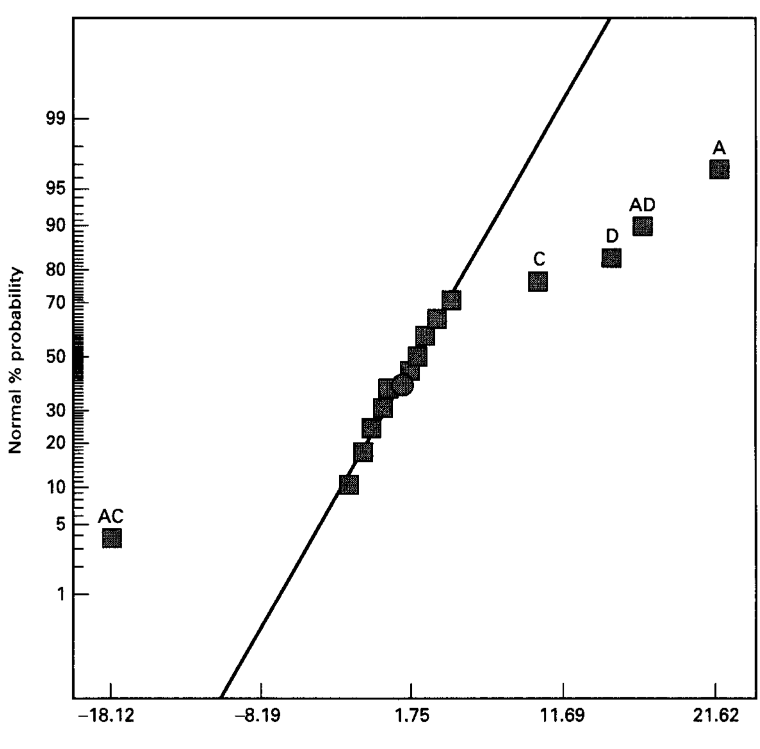

Figure 6-11 Normal probability plot of the effects for the 2^4 factorial in Example 6-2.

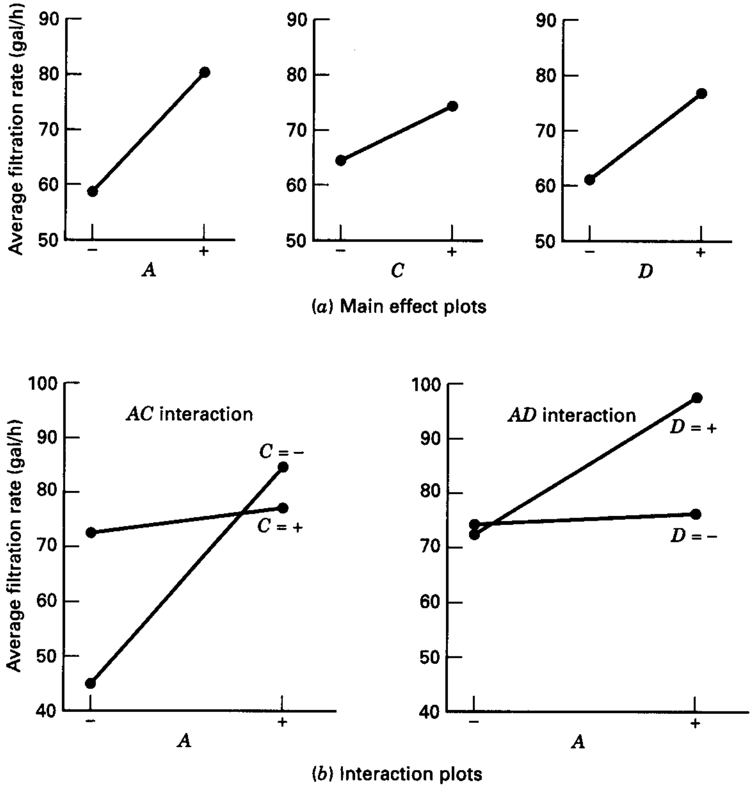

Figure 6-12 Main effect and interaction plots for Example 6-2.

Table 6-13 Analysis of Variance for the Pilot Plant Filtration Rate Experiment in A, C, and D

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
A	1870.56	1	1870.56	83.36	<0.0001
C	390.06	1	390.06	17.38	<0.0001
D	855.56	1	855.56	38.13	<0.0001
AC	1314.06	1	1314.06	58.56	<0.0001
AD	1105.56	1	1105.56	49.27	<0.0001
CD	5.06	1	5.06	<1	
ACD	10.56	1	10.56	<1	
Error	179.52	8	22.44		
Total	5730.94	15			

The concept of projecting an unreplicated factorial into a replicated factorial in fewer factors is very useful. In general, if we have a single replicate of 2^k design, and if h ($h < k$) factors are negligible and can be dropped, then the original data correspond to a full two-level factorial in the remaining $k - h$ factors with 2^h replicates.

Diagnostic Checking

The usual diagnostic checks should be applied to the residuals of a 2^k design. Our analysis indicates that the only significant effects are $A = 21.625$, $C = 9.875$, $D = 14.625$, $AC = -18.125$, and $AD = 16.625$. If this is true, the estimated filtration rates are given by

$$\hat{y} = 70.06 + \left(\frac{21.625}{2}\right) x_1 + \left(\frac{9.875}{2}\right) x_3 + \left(\frac{14.625}{2}\right) x_4 - \left(\frac{18.125}{2}\right) x_1 x_3 + \left(\frac{16.625}{2}\right) x_1 x_4$$

where 70.06 is the average response and the coded variables x_1, x_3, x_4 take on values between -1 and $+1$. The predicted filtration rate at run (1) is

$$\begin{aligned}\hat{y} &= 70.06 + \left(\frac{21.625}{2}\right)(-1) + \left(\frac{9.875}{2}\right)(-1) + \left(\frac{14.625}{2}\right)(-1) \\ &\quad - \left(\frac{18.125}{2}\right)(-1)(-1) + \left(\frac{16.625}{2}\right)(-1)(-1) \\ &= 46.22\end{aligned}$$

Because the observed value is 45, the residual is $e = y - \hat{y} = 45 - 46.22 = -1.22$. The values of y, $\hat{y}$, and $e = y - \hat{y}$ for all 16 observations follow.

	y	$\hat{y}$	$e = y - \hat{y}$
(1)	45	46.22	−1.22
a	71	69.39	1.61
b	48	46.22	1.78
ab	65	69.39	−4.39
c	68	74.23	−6.23
ac	60	61.14	−1.14
bc	80	74.23	5.77
abc	65	61.14	3.86
d	43	44.22	−1.22
ad	100	100.65	−0.65
bd	45	44.22	0.78
abd	104	100.65	3.35
cd	75	72.23	2.77
acd	86	92.40	−6.40
bcd	70	72.23	−2.23
$abcd$	96	92.40	3.60

A normal probability plot of the residuals is shown in Figure 6-13 on the next page. The points on this plot lie reasonably close to a straight line, lending support to our conclusion

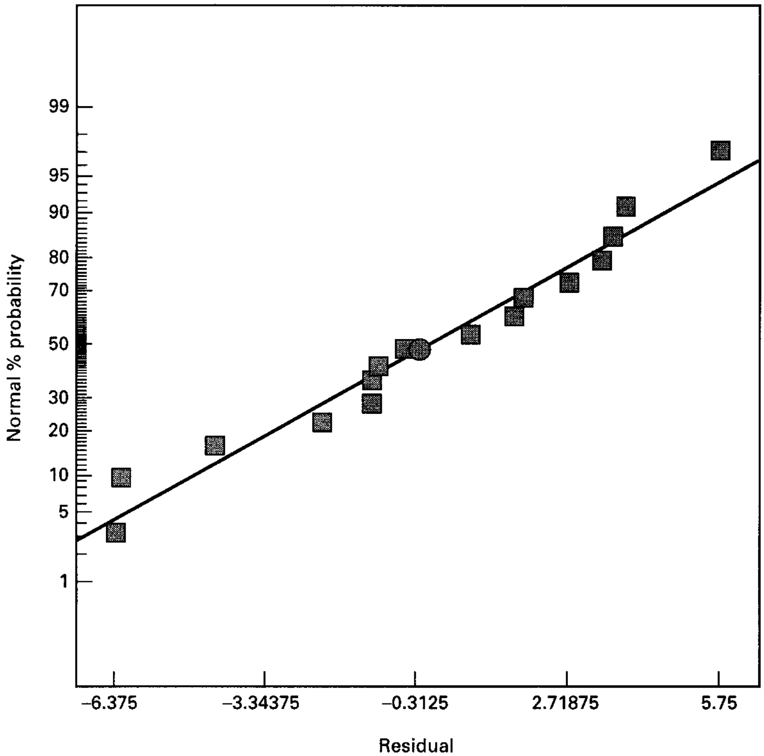

Figure 6-13 Normal probability plot of residuals for Example 6-2.

that A, C, D, AC, and AD are the only significant effects and that the underlying assumptions of the analysis are satisfied.

The Response Surface

We used the interaction plots in Figure 6-12 to provide a practical interpretation of the results of this experiment. Sometimes we find it helpful to use the response surface for this purpose. The response surface is generated by the regression model

$$\hat{y} = 70.06 + \left(\frac{21.625}{2}\right) x_1 + \left(\frac{9.875}{2}\right) x_3 + \left(\frac{14.625}{2}\right) x_4 - \left(\frac{18.125}{2}\right) x_1 x_3 + \left(\frac{16.625}{2}\right) x_1 x_4$$

Figure 6-14*a* shows the response surface contour plot when stirring rate is at the high level (i.e., $x_4 = 1$). The contours are generated from the above model with $x_4 = 1$, or

$$\hat{y} = 77.3725 + \left(\frac{38.25}{2}\right) x_1 + \left(\frac{9.875}{2}\right) x_3 - \left(\frac{18.125}{2}\right) x_1 x_3$$

Notice that the contours are curved lines because the model contains an interaction term.

Figure 6-14*b* is the response surface contour plot when temperature is at the high level (i.e., $x_1 = 1$). When we put $x_1 = 1$ in the regression model we obtain

$$\hat{y} = 80.8725 - \left(\frac{8.25}{2}\right) x_3 + \left(\frac{31.25}{2}\right) x_4$$

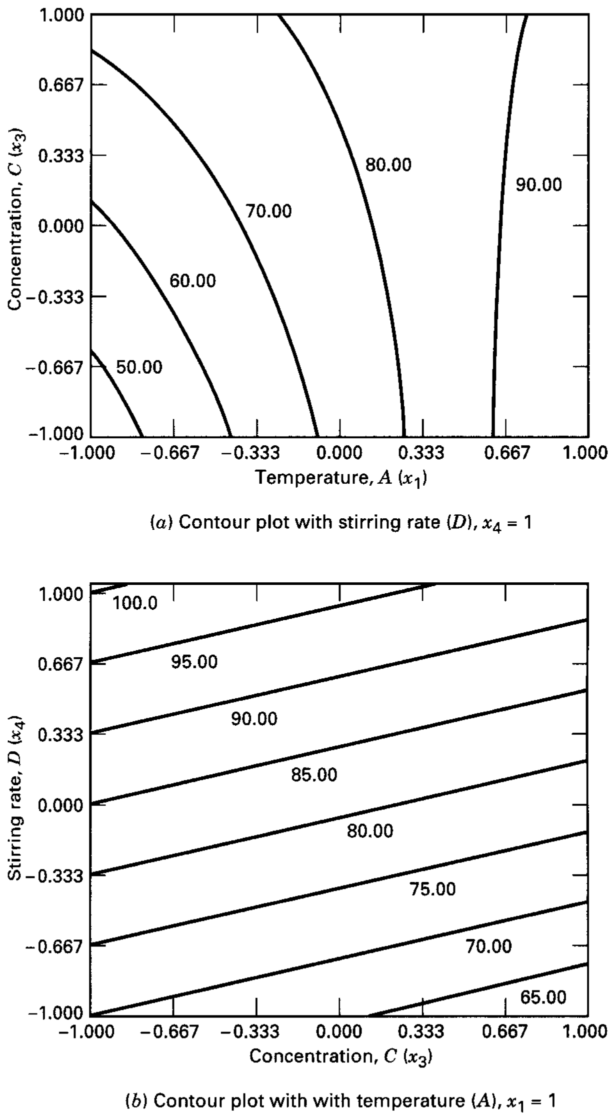

Figure 6-14 Contour plots of filtration rate, Example 6-2.

These contours are parallel straight lines because the model contains only the main effects of factors C (x_3) and D (x_4).

Both contour plots indicate that if we want to maximize the filtration rate, variables A (x_1) and D (x_4) should be at the high level and that the process is relatively robust to concentration C. We obtained similar conclusions from the interaction graphs.

The Half-Normal Plot of Effects

An alternative to the normal probability plot of the factor effects is the **half-normal plot.** This is a plot of the absolute value of the effect estimates against their cumulative normal probabilities. Figure 6-15 (page 254) presents the half-normal plot of the effects for Example 6-2. The straight line on the half-normal plot always passes through the origin and should also pass close to the fiftieth percentile data value. Many analysts feel that the half-normal plot is easier to interpret, particularly when there are only a few effect

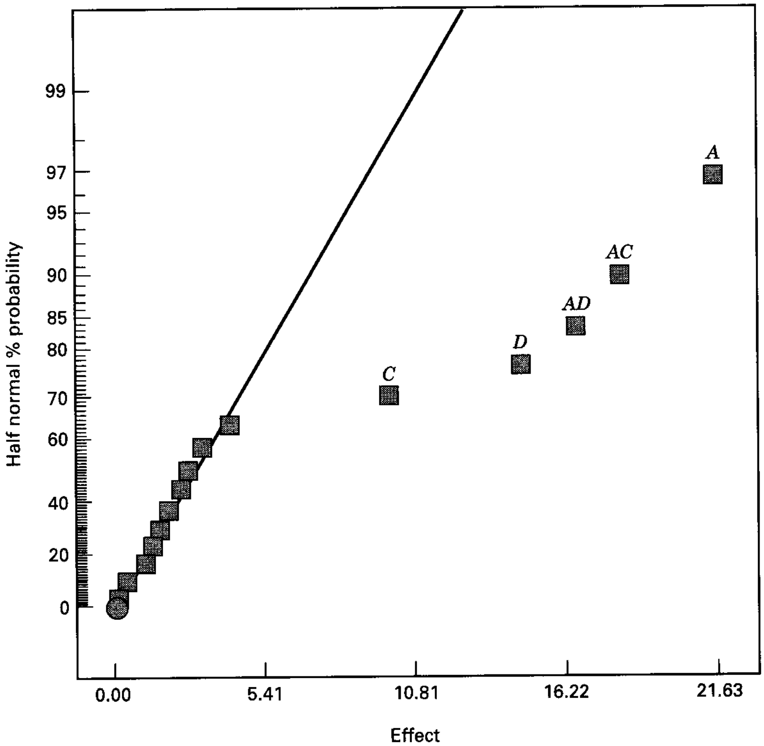

Figure 6-15 Half-normal plot of the factor effects from Example 6-2.

estimates such as when the experimenter has used an eight-run design. Some software packages will construct both plots.

Other Methods for Analyzing Unreplicated Factorials

The standard analysis procedure for an unreplicated two-level factorial design is the normal (or half-normal) plot of the estimated factor effects. However, unreplicated designs are so widely used in practice that many formal analysis procedures have been proposed to overcome the subjectivity of the normal probability plot. Hamada and Balakrishnan (1998) compared some of these methods. They found that the method proposed by Lenth (1989) has good power to detect significant effects. It is also easy to implement, and as a result, it is beginning to appear in some software packages for analyzing data from unreplicated factorials. We give a brief description of Lenth's method.

Suppose that we have m contrasts of interest, say $c_1, c_2, \ldots, c_m$. If the design is an unreplicated 2^k factorial design, these contrasts correspond to the $m = 2^k - 1$ factor effect estimates. The basis of Lenth's method is to estimate the variance of a contrast from the smallest (in absolute value) contrast estimates. Let

$$s_0 = 1.5 \times \text{median}(|c_j|)$$

and

$$PSE = 1.5 \times \text{median}(|c_j| : |c_j| < 2.5 s_0)$$

PSE is called the "pseudo standard error," and Lenth shows that it is a reasonable estimator of the contrast variance when there are not many active (significant) effects. The *PSE* is used to judge the significance of contrasts. An individual contrast can be compared to the **margin of error**

$$ME = t_{0.025,d} \times PSE$$

where the degrees of freedom are defined as $d = m/3$. For inference on a group of contrasts Lenth suggests using the **simultaneous margin of error**

$$SME = t_{\gamma,d} \times PSE$$

where the percentage point of the t distribution used is $\gamma = 1 - (1 + 0.95^{1/m})/2$.

To illustrate Lenth's method, consider the 2^4 experiment in Example 6-2. The calculations result in $s_0 = 1.5 \times |-2.625| = 3.9375$ and $2.5 \times 3.9375 = 9.84375$, so

$$\begin{aligned} PSE &= 1.5 \times |1.75| = 2.625 \\ ME &= 2.571 \times 2.625 = 6.75 \\ SME &= 5.219 \times 2.625 = 13.70 \end{aligned}$$

Now consider the effect estimates in Table 6-12. The *SME* criterion would indicate that the four largest effects (in magnitude) are significant, because their effect estimates exceed *SME*. The main effect of C is significant according to the *ME* criterion, but not with respect to *SME*. However, because the *AC* interaction is clearly important, we would probably include C in the list of significant effects. Notice that in this example, Lenth's method has produced the same answer that we obtained previously from examination of the normal probability plot of effects.

Several authors [see Hamada and Balakrishnan (1998), Loughin (1998), Loughin and Noble (1997), and Larntz and Whitcomb (1998)] have observed that Lenth's method fails to control type I error rates, and that simulation methods can be used to calibrate his procedure. Larntz and Whitcomb (1998) suggest replacing the original *ME* and *SME* multipliers with **adjusted multipliers** as follows:

Number of Contrasts	7	15	31
Original *ME*	3.764	2.571	2.218
Adjusted *ME*	2.295	2.140	2.082
Original *SME*	9.008	5.219	4.218
Adjusted *SME*	4.891	4.163	4.030

These are in close agreement with the results in Ye and Hamada (2000).

In general, the Lenth method is a clever and useful procedure. However, we recommend using it as a **supplement** to the usual normal probability plot of effects, not as a replacement for it.

Bisgaard (1998–1999) has provided a nice graphical technique, called a **conditional inference chart**, to assist in interpreting the normal probability plot. The purpose of the graph is to help the experimenter in judging significant effects. This would be relatively easy if the standard deviation σ were known, or if it could be estimated from the data. In unreplicated designs, there is no internal estimate of σ, so the conditional inference chart is designed to help the experimenter evaluate effect magnitude for a *range* of standard deviation values. Bisgaard bases the graph on the result that the standard error of an effect in a two-level design with N runs (for an unreplicated factorial, $N = 2^k$) is

$$\frac{2\sigma}{\sqrt{N}}$$

where σ is the standard deviation of an individual observation. Then ± 2 times the standard error of an effect is

$$\pm\frac{4\sigma}{\sqrt{N}}$$

Once the effects are estimated, plot a graph as shown in Figure 6-16, with the effect estimates plotted along the vertical, or y-axis. In this figure, we have used the effect estimates from Example 6-2.The horizontal, or x-axis, of Figure 6-16 is a standard deviation (σ) scale. The two lines are at

$$y = +\frac{4\sigma}{\sqrt{N}} \quad \text{and} \quad y = -\frac{4\sigma}{\sqrt{N}}$$

In our example, $N = 16$, so the lines are at $y = +\sigma$ and $y = -\sigma$. Thus, for any given value of the standard deviation σ we can read off the distance between these two lines as an approximate 95 percent confidence interval on the negligible effects.

In Figure 6-16, we observe that if the experimenter thinks that the standard deviation is between 4 and 8, then factors A, C, D, and the AC and AD interactions are significant. If he or she thinks that the standard deviation is as large as 10, factor C may not be

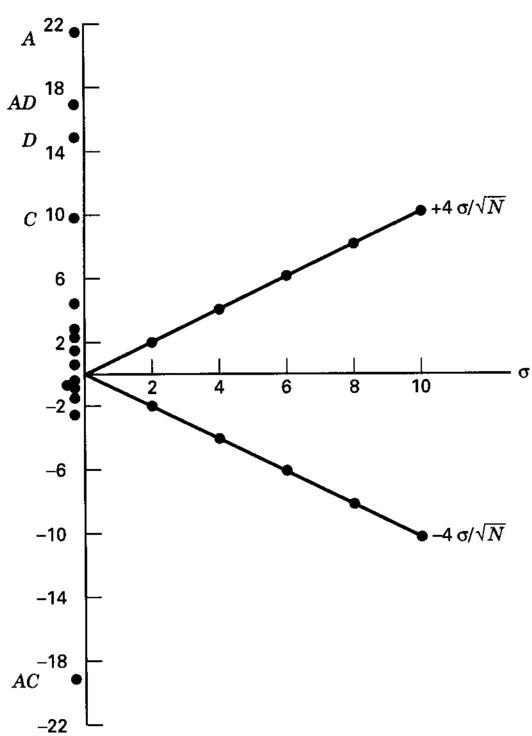

Figure 6-16 A conditional inference chart for Example 6-2.

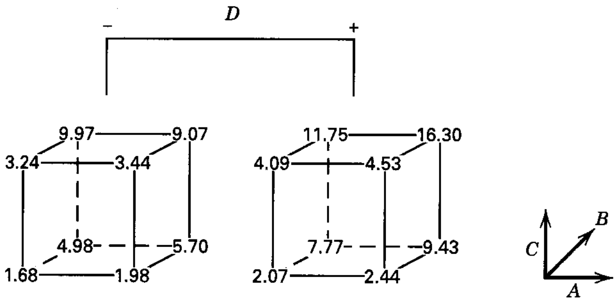

Figure 6-17 Data from the drilling experiment of Example 6-3.

significant. That is, for any given assumption about the magnitude of σ, the experimenter can construct a "yardstick" for judging the approximate significance of effects. The chart can also be used in reverse. For example, suppose that we were uncertain about whether factor C is significant. The experimenter could then ask whether it is reasonable to expect that σ could be as large as 10 or more. If it is unlikely that σ is as large as 10, then we can conclude that C is significant.

We now present three instructive examples of unreplicated 2^k factorial designs.

EXAMPLE 6-3 ....

Data Transformation in a Factorial Design

Daniel (1976) describes a 2^4 factorial design used to study the advance rate of a drill as a function of four factors: drill load (A), flow rate (B), rotational speed (C), and the type of drilling mud used (D). The data from the experiment are shown in Figure 6-17.

The normal probability plot of the effect estimates from this experiment is shown in Figure 6-18. Based on this plot, factors B, C, and D along with the BC and BD interactions

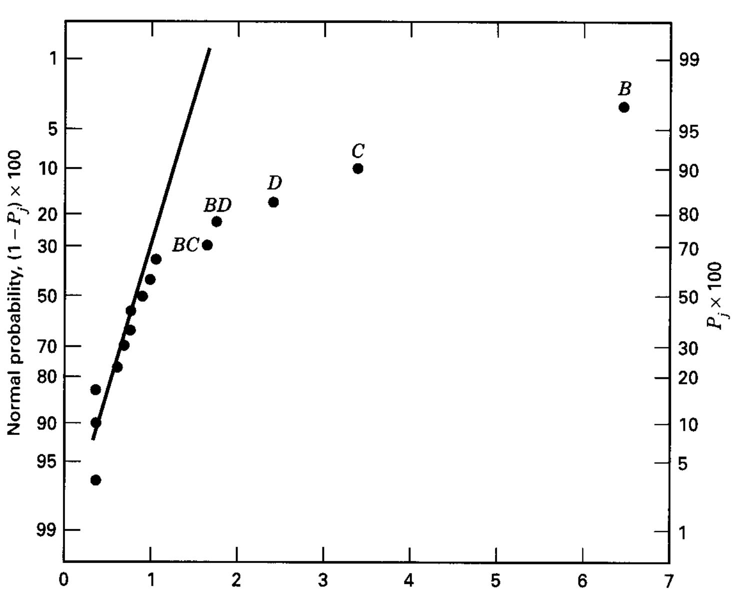

Figure 6-18 Normal probability plot of effects for Example 6-3.

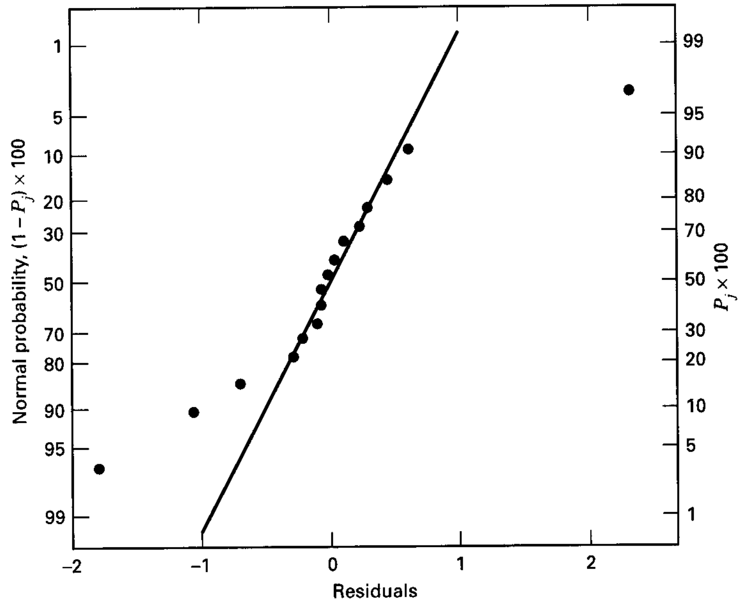

Figure 6-19 Normal probability plot of residuals for Example 6-3.

require interpretation. Figure 6-19 is the normal probability plot of the residuals and Figure 6-20 is the plot of the residuals versus the predicted advance rate from the model containing the identified factors. There are clearly problems with normality and equality of variance. A data transformation is often used to deal with such problems. Because the response variable is a rate, the log transformation seems a reasonable candidate.

Figure 6-21 presents a normal probability plot of the effect estimates following the

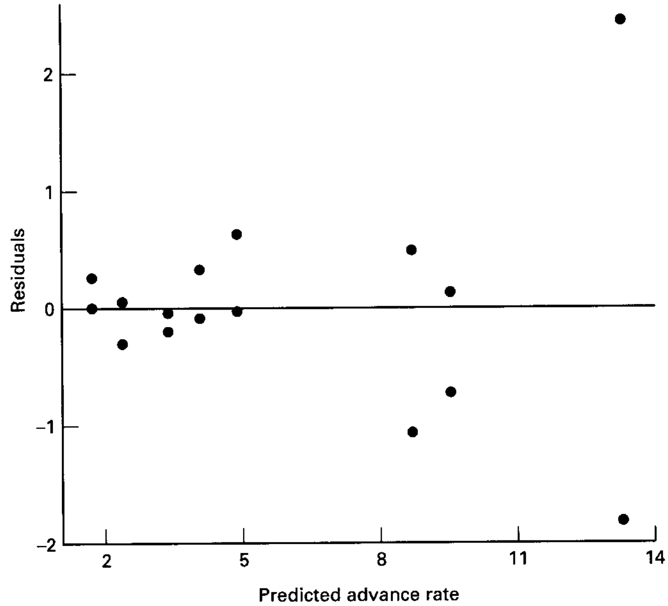

Figure 6-20 Plot of residuals versus predicted rate for Example 6-3.

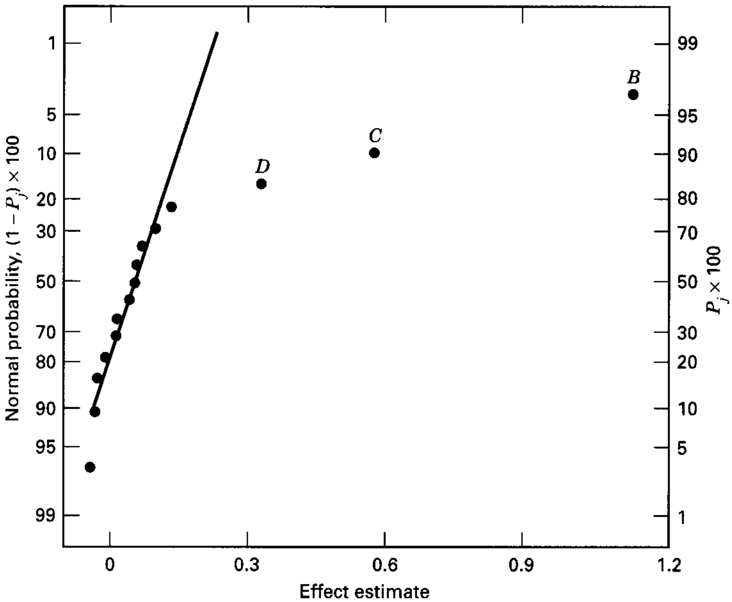

Figure 6-21 Normal probability plot of effects for Example 6-3 following log transformation.

transformation $y^* = \ln y$. Notice that a much simpler interpretation now seems possible, because only factors B, C, and D are active. That is, expressing the data in the correct metric has simplified its structure to the point that the two interactions are no longer required in the explanatory model.

Figures 6-22 and 6-23 present, respectively, a normal probability plot of the residuals

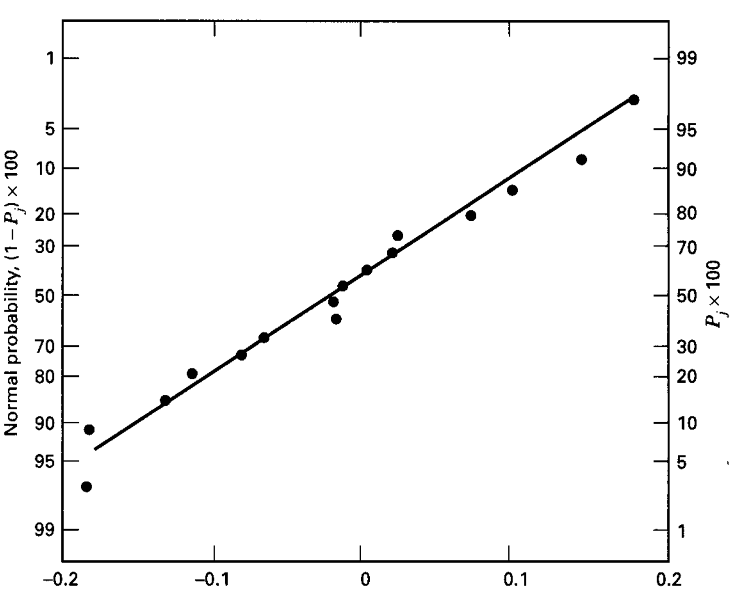

Figure 6-22 Normal probability plot of residuals for Example 6-3 following log transformation.

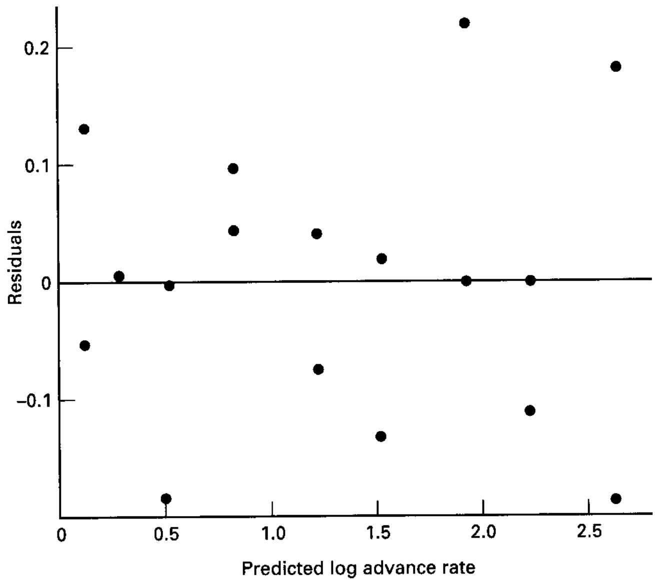

Figure 6-23 Plot of residuals versus predicted rate for Example 6-3 following log transformation.

and a plot of the residuals versus the predicted advance rate for the model in the log scale containing B, C, and D. These plots are now satisfactory. We conclude that the model for $y^* = \ln y$ requires only factors B, C, and D for adequate interpretation. The analysis of variance for this model is summarized in Table 6-14. The model sum of squares is

$$\begin{aligned} SS_{\text{Model}} &= SS_B + SS_C + SS_D \\ &= 5.345 + 1.339 + 0.431 \\ &= 7.115 \end{aligned}$$

and $R^2 = SS_{\text{Model}}/SS_T = 7.115/7.288 = 0.98$, so the model explains about 98 percent of the variability in the drill advance rate.

EXAMPLE 6-4

Location and Dispersion Effects in an Unreplicated Factorial

A 2^4 design was run in a manufacturing process producing interior side-wall and window panels for commercial aircraft. The panels are formed in a press, and under present

Table 6-14 Analysis of Variance for Example 6-3 Following the Log Transformation

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
B (Flow)	5.345	1	5.345	381.79	<0.0001
C (Speed)	1.339	1	1.339	95.64	<0.0001
D (Mud)	0.431	1	0.431	30.79	<0.0001
Error	0.173	12	0.014		
Total	7.288	15			

Factors	Low (−)	High (+)
A = Temperature (°F)	295	325
B = Clamp time (min)	7	9
C = Resin flow	10	20
D = Closing time (s)	15	30

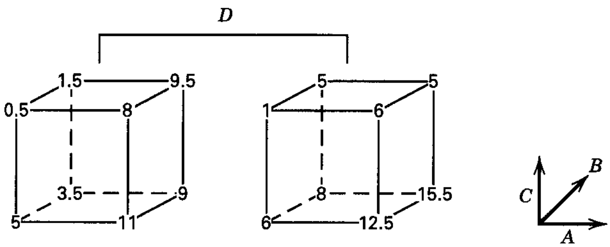

Figure 6-24 Data for the panel process experiment of Example 6-4.

conditions the average number of defects per panel in a press load is much too high. (The current process average is 5.5 defects per panel.) Four factors are investigated using a single replicate of a 2^4 design, with each replicate corresponding to a single press load. The factors are temperature (A), clamp time (B), resin flow (C), and press closing time (D). The data for this experiment are shown in Figure 6-24.

A normal probability plot of the factor effects is shown in Figure 6-25. Clearly the two largest effects are $A = 5.75$ and $C = -4.25$. No other factor effects appear to be large, and A and C explain about 77 percent of the total variability, so we conclude that lower temperature (A) and higher resin flow (C) would reduce the incidence of panel defects.

Careful residual analysis is an important aspect of any experiment. A normal probability plot of the residuals showed no anomalies, but when the experimenter plotted the residuals versus each of the factors A through D, the plot of residuals versus B (clamp

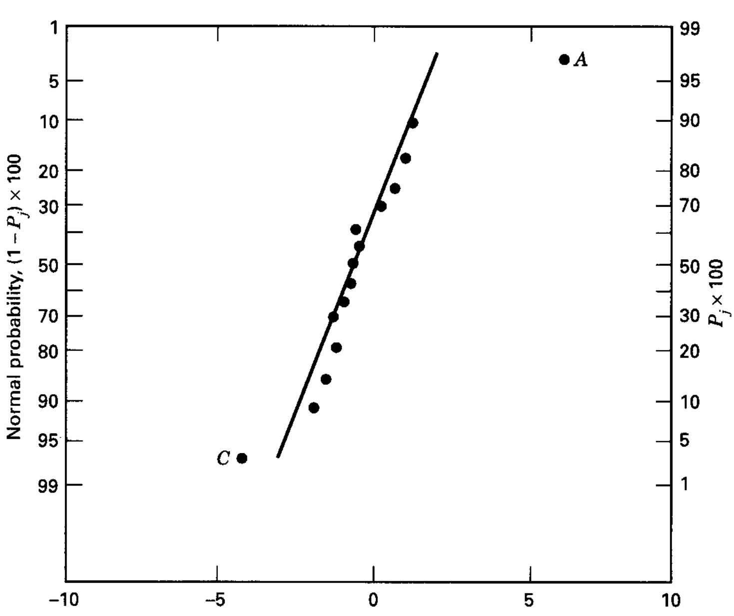

Figure 6-25 Normal probability plot of the factor effects for the panel process experiment of Example 6-4.

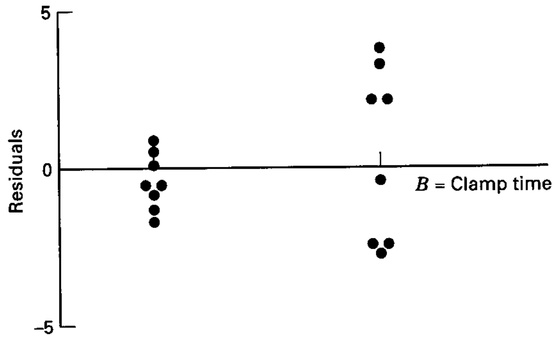

Figure 6-26 Plot of residuals versus clamp time for Example 6-4.

time) presented the pattern in Figure 6-26. This factor, which is unimportant insofar as the average number of defects per panel is concerned, is very important in its effect on process variability, with the lower clamp time resulting in less variability in the average number of defects per panel in a press load.

The dispersion effect of clamp time is also very evident from the **cube plot** in Figure 6-27, which plots the average number of defects per panel and the range of the number of defects at each point in the cube defined by factors A, B, and C. The average range when B is at the high level (the back face of the cube in Figure 6-27) is $\overline{R}_{B^+} = 4.75$, and when B is at the low level it is $\overline{R}_{B^-} = 1.25$.

As a result of this experiment, the engineer decided to run the process at low temperature and high resin flow to reduce the average number of defects, at low clamp time to reduce the variability in the number of defects per panel, and at low press closing time (which had no effect on either location or dispersion). The new set of operating conditions resulted in a new process average of less than one defect per panel.

The residuals from a 2^k design provide much information about the problem under study. Because residuals can be thought of as observed values of the noise or error, they often give insight into process variability. We can systematically examine the residuals from an unreplicated 2^k design to provide information about process variability.

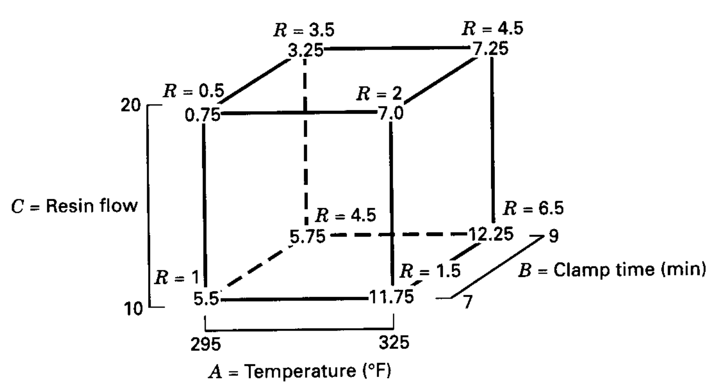

Figure 6-27 Cube plot of temperature, clamp time, and resin flow for Example 6-4.

Table 6-15 Calculation of Dispersion Effects for Example 6-4

Run	A	B	AB	C	AC	BC	ABC	D	AD	BD	ABD	CD	ACD	BCD	$ABCD$	Residual
1	−	−	+	−	+	+	−	−	+	+	−	+	−	−	+	−0.94
2	+	−	−	−	−	+	+	−	−	+	+	+	+	−	−	−0.69
3	−	+	−	−	+	−	+	−	+	−	+	+	−	+	−	−2.44
4	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+	−2.69
5	−	−	+	+	−	−	+	−	+	+	−	−	+	+	−	−1.19
6	+	−	−	+	+	−	−	−	−	+	+	−	−	+	+	0.56
7	−	+	−	+	−	+	−	−	+	−	+	−	+	−	+	−0.19
8	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	2.06
9	−	−	+	−	+	+	−	+	−	−	+	−	+	+	−	0.06
10	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+	0.81
11	−	+	−	−	+	−	+	+	−	+	−	−	+	−	+	2.06
12	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−	3.81
13	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	−0.69
14	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	−1.44
15	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	3.31
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−2.44
$S(i^+)$	2.25	2.72	2.21	1.91	1.81	1.80	1.80	2.24	2.05	2.28	1.97	1.93	1.52	2.09	1.61	
$S(i^-)$	1.85	0.83	1.86	2.20	2.24	2.26	2.24	1.55	1.93	1.61	2.11	1.58	2.16	1.89	2.33	
F_i^*	0.39	2.37	0.34	−0.28	−0.43	−0.46	−0.44	0.74	0.12	0.70	−0.14	0.40	−0.70	0.28	−0.74	

Consider the residual plot in Figure 6-26. The standard deviation of the eight residuals where B is at the low level is $S(B^-) = 0.83$, and the standard deviation of the eight residuals where B is at the high level is $S(B^+) = 2.72$. The statistic

$$F_B^* = \ln \frac{S^2(B^+)}{S^2(B^-)} \tag{6-24}$$

has an approximate normal distribution if the two variances $\sigma^2(B^+)$ and $\sigma^2(B^-)$ are equal. To illustrate the calculations, the value of F_B^* is

$$\begin{aligned} F_B^* &= \ln \frac{S^2(B^+)}{S^2(B^-)} \\ &= \ln \frac{(2.72)^2}{(0.83)^2} \\ &= 2.37 \end{aligned}$$

Table 6-15 (on the previous page) presents the complete set of contrasts for the 2^4 design along with the residuals for each run from the panel process experiment in Example 6-4. Each column in this table contains an equal number of plus and minus signs, and we can calculate the standard deviation of the residuals for each group of signs in each column, say $S(i^+)$ and $S(i^-)$, $i = 1, 2, \ldots, 15$. Then

$$F_i^* = \ln \frac{S^2(i^+)}{S^2(i^-)} \qquad i = 1, 2, \ldots, 15 \tag{6-25}$$

is a statistic that can be used to assess the magnitude of the **dispersion effects** in the experiment. If the variance of the residuals for the runs where factor i is positive equals the variance of the residuals for the runs where factor i is negative, then F_i^* has an approximate normal distribution. The values of F_i^* are shown below each column in Table 6-15.

Figure 6-28 is a normal probability plot of the dispersion effects F_i^*. Clearly, B is an important factor with respect to process dispersion. For more discussion of this pro-

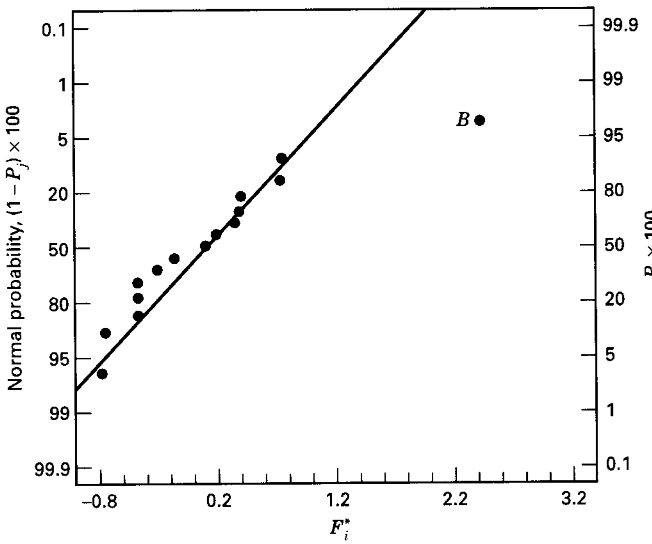

Figure 6-28 Normal probability plot of the dispersion effects F_i^* for Example 6-4.

cedure, see Box and Meyer (1986) and Myers and Montgomery (1995). Also, in order for the model residuals to properly convey information about dispersion effects, the **location model** must be correctly specified. Refer to the supplemental text material for this chapter for more details and an example.

EXAMPLE 6-5

Duplicate Measurements on the Response

A team of engineers at a semiconductor manufacturer ran a 2^4 factorial design in a vertical oxidation furnace. Four wafers are "stacked" in the furnace, and the response variable of interest is the oxide thickness on the wafers. The four design factors are temperature (A), time (B), pressure (C), and gas flow (D). The experiment is conducted by loading four wafers into the furnace, setting the process variables to the test conditions required by the experimental design, processing the wafers, and then measuring the oxide thickness on all four wafers. Table 6-16 presents the design and the resulting thickness measurements. In this table, the four columns labeled "Thickness" contain the oxide thickness measurements on each individual wafer, and the last two columns contain the sample average and sample variance of the thickness measurements on the four wafers in each run.

The proper analysis of this experiment is to consider the individual wafer thickness measurements as **duplicate measurements**, and not as replicates. If they were really replicates, each wafer would have been processed individually on a single run of the furnace. However, because all four wafers were processed together, they received the treatment factors (that is, the levels of the design variables) *simultaneously*, so there is much less variability in the individual wafer thickness measurements than would have been observed if each wafer was a replicate. Therefore, the **average** of the thickness measurements is the correct response variable to initially consider.

Table 6-17 (page 266) presents the effect estimates for this experiment, using the average oxide thickness $\bar{y}$ as the response variable. Note that factors A and B and the AB interaction have large effects that together account for nearly 90 percent of the variability

Table 6-16 The Oxide Thickness Experiment

Standard Order	Run Order	A	B	C	D	Thickness				$\bar{y}$	s^2
1	10	−1	−1	−1	−1	378	376	379	379	378	2
2	7	1	−1	−1	−1	415	416	416	417	416	0.67
3	3	−1	1	−1	−1	380	379	382	383	381	3.33
4	9	1	1	−1	−1	450	446	449	447	448	3.33
5	6	−1	−1	1	−1	375	371	373	369	372	6.67
6	2	1	−1	1	−1	391	390	388	391	390	2
7	5	−1	1	1	−1	384	385	386	385	385	0.67
8	4	1	1	1	−1	426	433	430	431	430	8.67
9	12	−1	−1	−1	1	381	381	375	383	380	12.00
10	16	1	−1	−1	1	416	420	412	412	415	14.67
11	8	−1	1	−1	1	371	372	371	370	371	0.67
12	1	1	1	−1	1	445	448	443	448	446	6
13	14	−1	−1	1	1	377	377	379	379	378	1.33
14	15	1	−1	1	1	391	391	386	400	392	34
15	11	−1	1	1	1	375	376	376	377	376	0.67
16	13	1	1	1	1	430	430	428	428	429	1.33

Table 6-17 Effect Estimates for Example 6-5, Response Variable Is Average Oxide Thickness

Model Term	Effect Estimate	Sum of Squares	Percent Contribution
A	43.125	7439.06	67.9339
B	18.125	1314.06	12.0001
C	−10.375	430.562	3.93192
D	−1.625	10.5625	0.0964573
AB	16.875	1139.06	10.402
AC	−10.625	451.563	4.12369
AD	1.125	5.0625	0.046231
BC	3.875	60.0625	0.548494
BD	−3.875	60.0625	0.548494
CD	1.125	5.0625	0.046231
ABC	−0.375	0.5625	0.00513678
ABD	2.875	33.0625	0.301929
ACD	−0.125	0.0625	0.000570753
BCD	−0.625	1.5625	0.0142688
ABCD	0.125	0.0625	0.000570753

in average oxide thickness. Figure 6-29 is a normal probability plot of the effects. From examination of this display, we would conclude that factors *A*, *B*, and *C* and the *AB* and *AC* interactions are important. The analysis of variance display for this model is shown in Table 6-18.

The model for predicting average oxide thickness is

$$\hat{y} = 399.19 + 21.56x_1 + 9.06x_2 - 5.19x_3 + 8.44x_1x_2 - 5.31x_1x_3$$

The residual analysis of this model is satisfactory.

The experimenters are interested in obtaining an average oxide thickness of 400 Å, and product specifications require that the thickness must lie between 390 and 410 Å. Figure 6-30 (page 268) presents two contour plots of average thickness, one with factor C (or x_3), pressure, at the low level (that is, $x_3 = -1$) and the other with C (or x_3) at the high level (that is, $x_3 = +1$). From examining these contour plots, it is obvious that there are many combinations of time and temperature (factors *A* and *B*) that will produce acceptable results. However, if pressure is held constant at the low level, the operating "window" is shifted toward the left, or lower, end of the time axis, indicating that lower cycle times will be required to achieve the desired oxide thickness.

It is interesting to observe the results that would be obtained if we **incorrectly** consider the individual wafer oxide thickness measurements as replicates. Table 6-19 (page 269) presents a full model analysis of variance based on treating the experiment as a replicated 2^4 factorial. Notice that there are many significant factors in this analysis, suggesting a much more complex model than we found when using the average oxide thickness as the response. The reason for this is that the estimate of the error variance in Table 6-19 is too small ($\hat{\sigma}^2 = 6.12$). The residual mean square in Table 6-19 reflects the variability between wafers *within* a run and variability *between* runs. The estimate of error obtained from Table 6-18 is much larger, $\hat{\sigma}^2 = 17.61$, and it is primarily a measure of the between-run variability. This is the best estimate of error to use in judging the significance of process variables that are changed from run to run.

A logical question to ask is: What harm results from identifying too many factors as

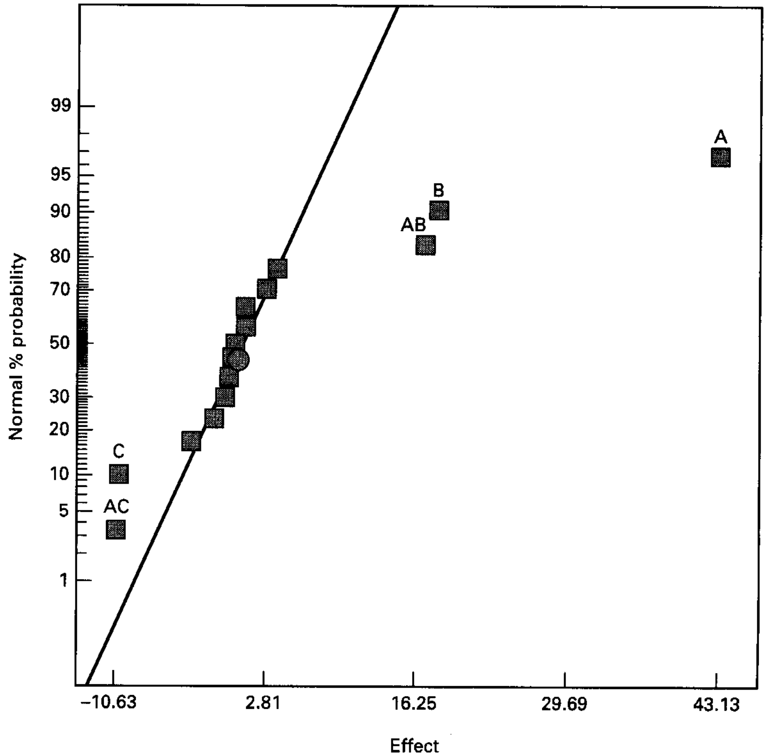

Figure 6-29 Normal probability plot of the effects for the average oxide thickness response, Example 6-5.

Table 6-18 Analysis of Variance (From Design-Expert) for the Average Oxide Thickness Response, Example 6-5

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Model	10774.31	5	2154.86	122.35	<0.000
A	*7439.06*	*1*	*7439.06*	*422.37*	*<0.000*
B	*1314.06*	*1*	*1314.06*	*74.61*	*<0.000*
C	*430.56*	*1*	*430.56*	*24.45*	*0.0006*
AB	*1139.06*	*1*	*1139.06*	*64.67*	*<0.000*
AC	*451.56*	*1*	*451.56*	*25.64*	*0.0005*
Residual	176.12	10	17.61		
Cor Total	10950.44	15			

Std. Dev.	4.20	*R*-Squared	0.9839
Mean	399.19	Adj. *R*-Squared	0.9759
C.V.	1.05	Pred. *R*-Squared	0.9588
PRESS	450.88	Adeq. Precision	27.967

Factor	Coefficient Estimate	DF	Standard Error	95% CI Low	95% CI High
Intercept	399.19	1	1.05	396.85	401.53
A-Time	21.56	1	1.05	19.22	23.90
B-Temp	9.06	1	1.05	6.72	11.40
C-Pressure	−5.19	1	1.05	−7.53	−2.85
AB	8.44	1	1.05	6.10	10.78
AC	−5.31	1	1.05	−7.65	−2.97

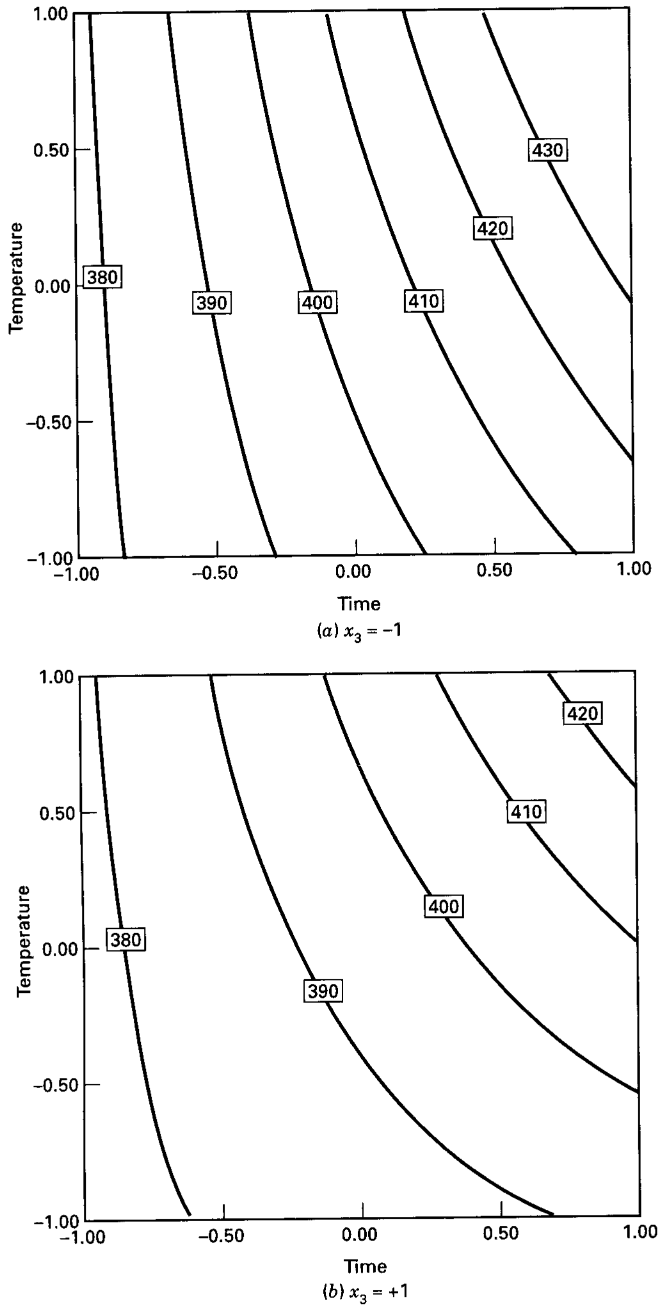

Figure 6-30 Contour plots of average oxide thickness with pressure (x_3) held constant.

important? as the incorrect analysis in Table 6-19 would certainly do. The answer is that trying to manipulate or optimize the unimportant factors would be a waste of resources, and it could result in adding unnecessary variability to *other* responses of interest.

When there are duplicate measurements on the response, there is almost always useful information about some aspect of process variability contained in these observa-

Table 6-19 Analysis of Variance (From Design-Expert) of the Individual Wafer Oxide Thickness Response

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Model	43801.75	15	2920.12	476.75	<0.0001
A	*29756.25*	*1*	*29756.25*	*4858.16*	*<0.0001*
B	*5256.25*	*1*	*5256.25*	*858.16*	*<0.0001*
C	*1722.25*	*1*	*1722.25*	*281.18*	*<0.0001*
D	*42.25*	*1*	*42.25*	*6.90*	*0.0115*
AB	*4556.25*	*1*	*4556.25*	*743.88*	*<0.0001*
AC	*1806.25*	*1*	*1806.25*	*294.90*	*<0.0001*
AD	*20.25*	*1*	*20.25*	*3.31*	*0.0753*
BC	*240.25*	*1*	*240.25*	*39.22*	*<0.0001*
BD	*240.25*	*1*	*240.25*	*39.22*	*<0.0001*
CD	*20.25*	*1*	*20.25*	*3.31*	*0.0753*
ABD	*132.25*	*1*	*132.25*	*21.59*	*<0.0001*
ABC	*2.25*	*1*	*2.25*	*0.37*	*0.5473*
ACD	*0.25*	*1*	*0.25*	*0.041*	*0.8407*
BCD	*6.25*	*1*	*6.25*	*1.02*	*0.3175*
ABCD	*0.25*	*1*	*0.25*	*0.041*	*0.8407*
Residual	294.00	48	6.12		
Lack of Fit	*0.000*	*0*			
Pure Error	*294.00*	*48*	*6.13*		
Cor. Total	44095.75	63			

tions. For example, if the duplicate measurements are multiple tests by a gauge on the same experimental unit, then the duplicate measurements give some insight about *gauge capability*. If the duplicate measurements are made at different locations on an experimental unit, they may give some information about the *uniformity* of the response variable across that unit. In our example, because we have one observation on each of four experimental units that have undergone processing together, we have some information about the *within-run* variability in the process. This information is contained in the variance of the oxide thickness measurements from the four wafers in each run. It would be of interest to determine if any of the process variables influence the within-run variability.

Figure 6-31 (page 270) is a normal probability plot of the effect estimates obtained using $\ln(s^2)$ as the response. Recall from Chapter 3 that we indicated that the log transformation is generally appropriate for modeling variability. There are not any strong individual effects, but factor A and the BD interaction are the largest. If we also include the main effects of B and D to obtain a hierarchical model, then the model for $\ln(s^2)$ is

$$\widehat{\ln(s^2)} = 1.08 + 0.41x_1 - 0.40x_2 + 0.20x_4 - 0.56x_2x_4$$

The model accounts for just slightly less than half of the variability in the $\ln(s^2)$ response, which is certainly not spectacular as empirical models go, but it is often difficult to obtain exceptionally good models of variances.

Figure 6-32 (page 270) is a contour plot of the predicted variance (not the log of the predicted variance) with pressure x_3 at the low level (recall that this minimizes cycle time) and gas flow x_4 at the high level. This choice of gas flow gives the lowest values of predicted variance in the region of the contour plot.

The experimenters here were interested in selecting values of the design variables that gave a mean oxide thickness within the process specifications and as close to 400 Å as possible, while simultaneously making the within-run variability small, say $s^2 \leq 2$.

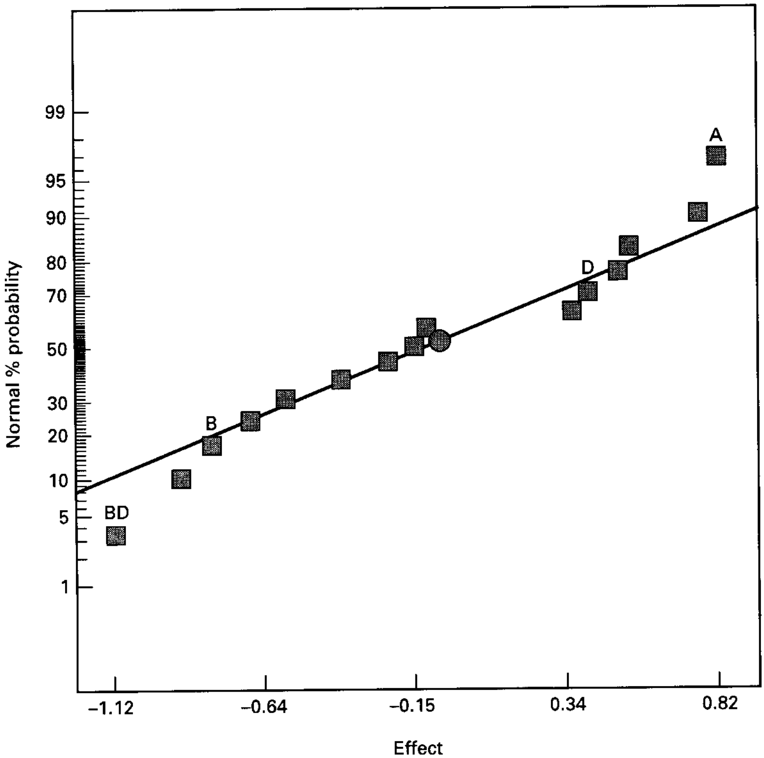

Figure 6-31 Normal probability plot of the effects using ln (s^2) as the response, Example 6-5.

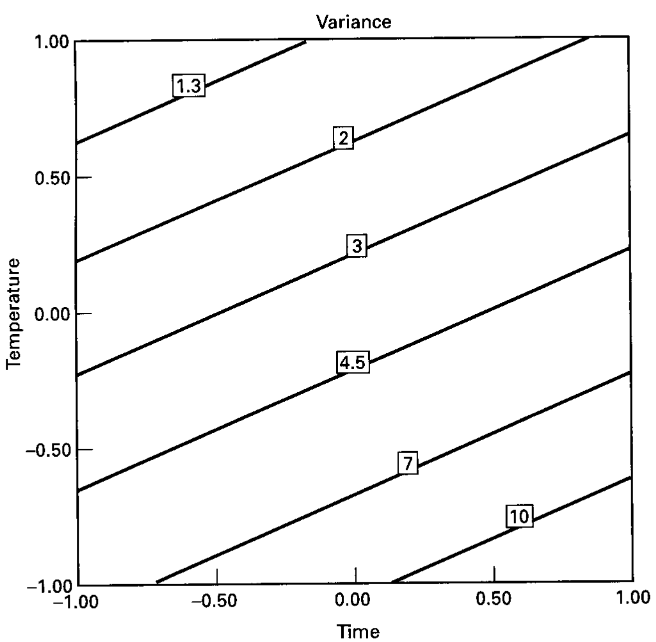

Figure 6-32 Contour plot of s^2 (within-run variability) with pressure at the low level and gas flow at the high level.

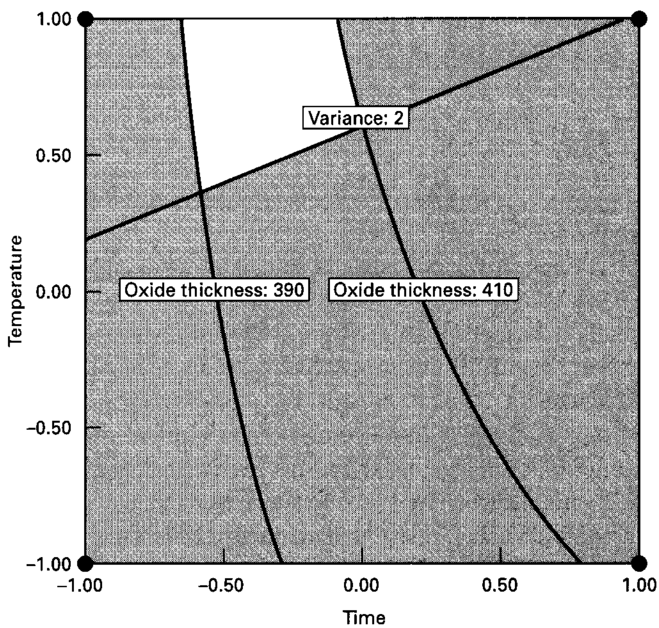

Figure 6-33 Overlay of the average oxide thickness and s^2 responses with pressure at the low level and gas flow at the high level.

One possible way to find a suitable set of conditions is to overlay the contour plots in Figures 6-30 and 6-32. The overlay plot is shown in Figure 6-33, with the specifications on mean oxide thickness and the constraint $s^2 \leqslant 2$ shown as contours. In this plot, pressure is held constant at the low level and gas flow is held constant at the high level. The open region near the upper left center of the graph identifies a feasible region for the variables time and temperature.

This is a simple example of using contour plots to study two responses simultaneously. We will discuss this problem in more detail in Chapter 11.

6-6 THE ADDITION OF CENTER POINTS TO THE 2^k DESIGN

A potential concern in the use of two-level factorial designs is the assumption of **linearity** in the factor effects. Of course, perfect linearity is unnecessary, and the 2^k system will work quite well even when the linearity assumption holds only very approximately. In fact, we have noted that if **interaction terms** are added to a main effects or first-order model, resulting in

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum\sum_{i<j} \beta_{ij} x_i x_j + \epsilon \tag{6-26}$$

then we have a model capable of representing some curvature in the response function. This curvature, of course, results from the twisting of the plane induced by the interaction terms $\beta_{ij} x_i x_j$.