CHAPTER 5

Fractional Factorial Designs with Two Levels

The numerical value of 2^k increases rapidly as k increases, so the number of design points for a 2^k design could be quite large when there are more than just a few factors of interest. As Steinberg and Hunter (1984) indicated, there might be as many as 50 or 100 potentially important factors in some applications.

A 2^k full factorial allows an experimenter to estimate all of the $2^k - 1$ effects, but when k is greater than 3 or 4, most of these effects will be high-order interaction effects that would not be significant, as the higher the order of an interaction effect, the less likely it is to be significant. Therefore, using a full factorial would be wasteful when the value of k is not small. A *fractional factorial* has, as the name suggests, a fraction of the number of design points of the corresponding full factorial, with the objective being to not waste design points that allow the estimation of effects that will almost certainly not be real.

We shall focus attention on two-level fractional factorial designs as this type of fractional factorial has been used extensively in practice. Fractional factorials were first presented by Finney (1945) and popularized in the landmark papers of Box and Hunter (1961a,b). We shall adopt the notation used by the latter.

A two-level fractional factorial can be written in the general form 2^{k-p}, where as indicated previously, k denotes the number of factors, and the fraction of the full 2^k factorial that is to be run is $1/2^p$. The *Resolution* of a fractional factorial design, first discussed in detail by Box, Hunter, and Hunter (1978), who assigned resolution numbers to categories of effects that could be estimated, indicates the effects that can be estimated with the design. Specifically resolution III indicates that only main effects can be estimated in the absence of interaction effects; IV indicates that only main effects are estimable even if two-factor interactions exist (provided that higher-order interactions do not exist); and V means that both main effects and two-factor interactions are estimable, provided that higher-order interactions do not exist. We

Modern Experimental Design By Thomas P. Ryan

will use notation such as 2^{6-2}_{IV} in representing fractional factorial designs, which here is a $1/4$ fraction of a 2^6 design that is of resolution IV. In the foregoing we will sometimes indicate the resolution of a design in this manner.

The term *defining relation* is explained and illustrated in Section 5.1. Briefly, it is composed of one or more interactions, whose inclusion in the defining relation determines effects that are confounded (i.e., confused). With this in mind, we have the following.

IMPORTANT POINT

The Resolution of a fractional factorial design is generally defined as the length of the shortest word in the defining relation, with the number of words in the defining relation determined by the degree of fractionization. This definition will suffice in almost all cases, although exceptions can be found when blocking is involved (see Giesbrecht and Gumpertz, 2004, pp. 381–382).

The last part of the Important Point will be illustrated by the examples in Sections 5.1 and 5.2. (*Note*: Fractional factorial designs are "orthogonal arrays," meaning that the columns for the effects to be estimated are orthogonal. There are also orthogonal arrays that are not fractional factorials. It is common for the "strength" of an orthogonal array to be given. This is simply one less than the resolution of the design.)

It is very important to realize that the properties of a fractional factorial design (or any design in general) for all the factors and effects that are estimable with a given design is not what is important, as generally less than half of the effects that are being examined will be significant. Therefore, the focus must be on the properties of the design for the effects that are significant. Thus, the *projective properties* of designs are important, as a design is projected onto the number of dimensions corresponding to the number of significant factors. For example, the properties of a design for investigating six factors are not particularly important when only three factors turn out to be significant. The projective properties of 2^{k-p} designs are discussed in Section 5.11.

The term "regular design" is used extensively in the design literature. A full factorial design is a regular design because all the effects can be estimated independently of each of the other effects. The 2^{k-p} designs are also regular designs because although not all effects are estimable, the effects that are not estimable are completely confounded with other effects. A design for which at least some effects are neither independently estimable nor completely confounded is called a *nonregular design*. The classification of designs in this way is of interest because we would naturally like to know whether there is at least partial information on all effects that are of interest.

5.1 2^{k-1} DESIGNS

Thus, a 2^{3-1} design would be a $1/2$ fraction of a 2^3 design. Sixteen point designs (so that $k - p = 4$) are the ones that have been used most often in industry, although

designs with a much higher number of points are sometimes needed, especially when there is a large number of factors. An example with 47 factors is discussed in Section 5.12.

For simplicity, however, we shall first illustrate a 2^{3-1} design, which, although of very limited usefulness, does have value for illustrative purposes. We should first recognize that with only four design points (here we are assuming that the design is not replicated), we will have only three df, so we can estimate only three effects. Which three effects do we choose to estimate? Although in rare instances a two-factor interaction will be of more interest to an experimenter than a main effect, we would generally choose to estimate main effects over interactions, if we had to select one over the other. Thus, the logical choice would be to estimate the three main effects: A, B, and C. Before we can undertake that task, however, we must determine what four design points to use. We cannot just randomly select four treatment combinations from the eight that are available. For example, we obviously could not estimate the main effect of A if we happened to select four treatment combinations in which A was at the high level and none in which A was at the low level, and similarly for B and C. Thus, we would clearly want to have two treatment combinations in which A is at the high level and two in which A is at the low level, and the same for B and C.

With a little trial and error we could obtain four treatment combinations that satisfy this property without too much difficulty, but it would obviously be preferable to use some systematic approach. Whenever a $1/2$ fraction is used, we have to select one effect to "confound" with the difference between the two fractions, that is, that particular effect would be estimated by the difference of the averages of the treatment combinations in each fraction (which of course is the way that we would logically estimate the difference between the two fractions). If we have to "give up" the estimate of one effect in this way (which is obviously what we are doing since we will run only one of the two fractions), it would be logical to select the highest-order interaction to relinquish. For a 2^3 design that is obviously the ABC interaction.

One simple way to construct the two $1/2$ fractions (from which one could be randomly selected, although this is normally not done in practice) would be to assign those treatment combinations with an even number of letters in common with ABC to one fraction, and those with an odd number of letters in common with ABC to the other fraction. This has the desired effect of creating the two fractions with ABC confounded between the two fractions, which can be explained as follows. As in the case of a full factorial design, the sign of an interaction for a particular treatment combination is a product of the signs of the factors represented by the treatment combination. The "even" fraction must have all minus signs because the number of letters in common with ABC must be either 0 or 2. In each case there will be an odd number of minus signs (e.g., the treatment combination ab has two letters in common with ABC, and $(+1)(+1)(-1) = -1$, so the "even" fraction will be the "minus fraction").

The two fractions are given in Figure 5.1.

There is a *defining relation* associated with each fraction, with $I = -ABC$ for the minus fraction and $I = ABC$ for the plus fraction. The first defining relation, which corresponds to the second fraction, states that the estimate of the ABC interaction is

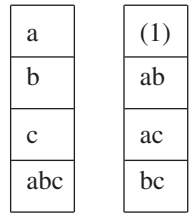

Figure 5.1 Two 1/2 fractions of 2^3.

confounded with minus the estimate of the overall mean. That is, we obtain a minus for each treatment combination (e.g., *ab* is $(+)(+)(-) = -$), so the observations for this fraction would estimate the negative of the *ABC* effect. Of course we would estimate the mean using all plus signs, and "*I*" refers to the mean effect. Thus, the estimate of the mean effect will be the negative of the estimate of the *ABC* effect.

One way to gain insight into the 2^{3-1} design is to write out the manner in which each effect is estimated using the full 2^3 design and then look at the treatment combinations in Figure 5.1 that comprise each fraction. Although Yates' algorithm, which can be used for obtaining effect estimates, no longer has much practical value since effect estimates are routinely obtained using software, we can use it effectively here. The treatment combinations are written in a specified order, and for a 2^3 design the order is given in Table 5.1, using the data from Daniel (1976) that was used previously in Section 4.5 in illustrating conditional effects and the potential hazards of using high-order interactions in estimating σ^2.

Here we will use it to illustrate half fractions, and since we have data from the full factorial, we will be able to compare the results from the two fractions. For

TABLE 5.1 Yates' Algorithm Calculations Using the Data in Daniel (1976)

Treatment Combination	Response	(1)	(2)	(3)	Column (3) Representation
(1)	297	597	834	1374	(Sum of all responses)
a	300	237	540	62	$a + ab + ac + abc$ $-b - c - bc - (1)$
b	106	355	28	−530	$b + ab + bc + abc$ $-a - c - ac - (1)$
ab	131	185	34	54	$c + ab + abc + (1)$ $-a - b - ac - bc$
c	177	3	−360	−294	$c + ac + bc + abc$ $-a - b - ab - (1)$
ac	178	25	−170	6	$b + ac + abc + (1)$ $-a - c - ab - bc$
bc	76	1	22	190	$a + bc + abc + (1)$ $-b - c - ac - ab$
abc	109	33	32	10	$a + b + c + abc - ac$ $-ab - bc - (1)$

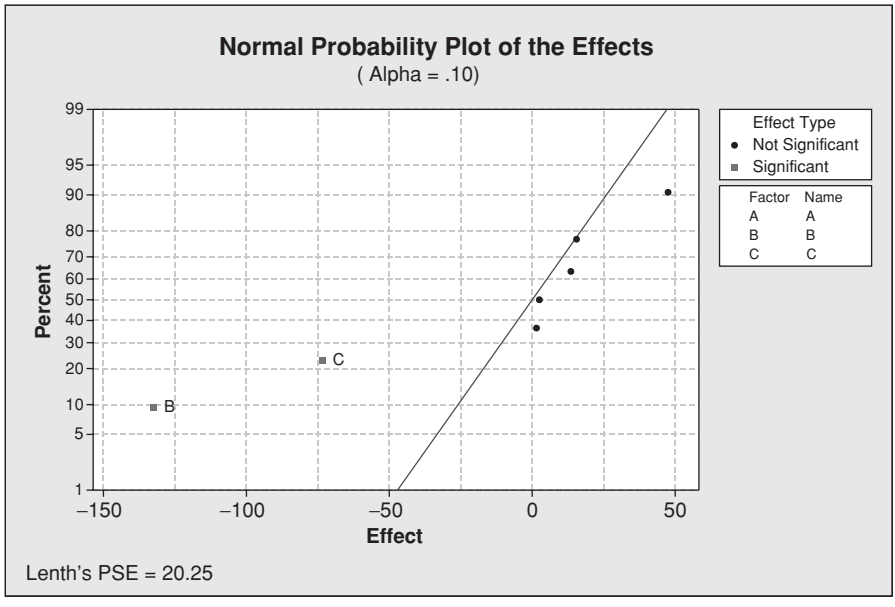

Figure 5.2 Normal probability plot of Daniel (1976) data.

convenience, the normal probability plot analysis of that data, using Lenth's (1989) method for determining significant effects, is repeated (Fig. 5.2), since it will be referred to in subsequent sections. (Haaland and O'Connell (1995) concluded that Lenth's (1989) method is a reasonable approach.)

The plot shows only the B and C effects to be significant. (It is worth noting that even though the BC interaction effect doesn't show as being significant, it is more than half the C effect and more than $1/3$ the B effect, thus rendering those estimates almost meaningless.)

Using adjacent treatment combinations in Table 5.1, there are four pairs, and additions and subtractions are performed in pairs, with the first number in each pair subtracted from the second number in the pair. This should be clear from inspection of the column labeled (1). The number of times that this set of operations is performed is equal to the number of factors, with the last column in Table 5.1 indicating how each number in column (3) is obtained.

Those numbers are used in obtaining the estimates of the various effects. For example, the AB interaction effect is estimated as $54/4 = 13.5$, with the number 54 being used because that is in the row for the *ab* treatment combination and 4 is used as the divisor because there are four plus signs and four minus signs.

It can be observed that the 10, which is the last number in column (3) and corresponds to the *abc* treatment combination, is obtained by adding the response values of the treatment combinations in the first fraction in Figure 5.1 and subtracting the response values of the treatment combinations in the second fraction. This shows that

the *ABC* interaction is confounded with the difference between the two fractions, as was stated previously.

Before proceeding any further, we should think about what we are giving up in addition to an estimate of the *ABC* interaction when we run one of the fractions in Figure 5.1. We assume that we are relinquishing information on the three two-factor interactions by using one of the fractions in Figure 5.1 (since we have only three df), but we also assume that we will be able to estimate the three main effects.

Let's verify this by looking at some of the other parenthetical expressions in Table 5.1. Assume that we have randomly selected the first fraction in Figure 5.1 to run. The sum and difference of the various treatment combinations beside the number -294 indicate that we would estimate the main effect of C by $(abc + c - a - b)/2 = (109 + 177 - 300 - 106)/2 = -60$, which does not differ greatly from -73.5 that was obtained using all eight treatment combinations. Further, we can see from Table 5.1 that, using the first fraction, we would also estimate AB by $(abc + c - a - b)/2$. What this means is that the estimate of the C effect is *confounded* (i.e., confused) with the AB effect.

(*Note*: This is complete confounding. In subsequent chapters, and especially in Section 13.4.1, we will speak of *partial aliasing*. As the name suggests, this is a state of affairs between orthogonality and complete confounding. With the former, the dot product of two columns that represent effects to be estimated is zero; with the latter the dot product is n because the signs are the same. With partial aliasing the dot product is something between 0 and n. If the sum is $n/3$, then the correlation (using the term somewhat loosely) between the two columns is $1/3$; if the sum is $-n/3$, then the correlation is $-1/3$. It should be noted, however, that the term partial *confounding* is used in conjunction with designs that are run in blocks to mean that an effect is confounded with some blocks but not with others. That is different from what is discussed in this section, so that terminology is not used here.)

Similarly, it could be shown that A is confounded with BC and B is confounded with AC, as can be verified from Table 5.1. Fortunately, there are easier ways to determine which effects are confounded. One way is to write out how each effect is estimated using plus and minus signs, and then identify those effects that have the same configuration of signs. The other method is much easier and is simply a matter of multiplying each effect by the effect that was confounded with the difference of the two fractions (*ABC* in this example), and removing any letter whose exponent is a 2. This applies to *any* two-level fractional factorial. This is modular arithmetic and it is modulo 2 because the factors have two levels. To see why a letter whose exponent is 2 is removed, notice that if we square a column of numbers that are all $+1$ and -1, the result is a column of numbers that are all $+1$, and any such column multiplied by any effect will simply give the column corresponding to that effect.

Thus, $A(ABC) = A^2BC = BC$, $B(ABC) = AB^2C = AC$, and $C(ABC) = ABC^2 = AB$. The effects that are confounded with each other are said to be *aliases* of each other, and the set of such aliases is said to be the *alias structure*.

Another way to view the alias structure is to use Yates' algorithm after filling in zeros for the treatment combinations that are in the fraction that is not used. Factors that are aliased will then have the same totals in column (3), as the reader is asked to demonstrate for these data in Exercise 5.3.

This is not a recommended approach for determining the alias structure, however, as effect estimates (and, hence, numbers in column (3)) can be the same without the effects being confounded. It is also far more time-consuming than the multiplication approach just illustrated. It is simply another way of viewing the alias structure. Of course software will readily provide the alias structure for a fractional factorial, and obviously that is the preferred approach for obtaining the alias structure.

We saw from the analysis of the full factorial that the AB interaction was not significant, and we can also see from Table 5.1 that the estimate of C plus the estimate of AB equals -60 that we just obtained using the four treatment combinations in the first fraction. Thus, we are actually estimating $C + AB$ rather than just C, and the extent to which our estimate of C is contaminated depends upon the size of the AB interaction. Here there is no serious problem because the AB interaction is not large.

What if we had randomly selected the other fraction? A little arithmetic would reveal that our estimate of C would be -87 and that we would really be estimating $C-AB$.

Although moderately large, the BC interaction was declared not significant by the normal probability plot analysis of Figure 5.2, and we now know that BC is aliased with A. Therefore, we can see to what extent the estimate of the A effect is contaminated by the presence of a moderate BC interaction. Again assuming that we had used the first fraction, our estimate of the A effect would be obtained from $(abc + a - b - c)/2 = (109 + 300 - 106 - 177)/2 = 63$, which differs dramatically from our estimate of 15.5 obtained from the full factorial. As the reader might suspect, we are actually estimating $A + BC$ (and $A - BC$ with the other fraction), so that, in this example, we would erroneously conclude that there is a strong A effect when in fact there was not. (We remember, however, that our detective work did reveal that A was somewhat influential when B was at its high level.)

The upshot of all of this is that when we run a fractional factorial we do take a risk, and the severity of the risk depends upon the order of the interactions that are lost. First-order interactions (i.e., involving two factors) and the second-order interaction were lost in the 2^{3-1} example; so that design should be considered only if there is a strong prior belief that none of the interactions will be significant. Even then, the advantage of the fractional factorial would be minimal, as four design points would be run instead of eight—not much of a saving.

The picture changes considerably, however, when there are more than three factors. What about a 2^{4-1}? The alias structure would certainly be more palatable in that the fractions could be constructed in such a way that the main effects would be aliased with second-order interactions, but, unfortunately, the first-order interactions are aliased in pairs. The $ABCD$ interaction would be confounded with the difference between the two fractions, and we would then have $A = BCD$, $B = ACD$, ..., $AB = CD$, $AC = BD$, and so on. Although it is a small design, 8-point designs have been used in practice—an application in which a 2^{4-1} design was used was discussed by Hill and Wiles (1975).

From a practical standpoint, however, we should ask whether or not such small designs should be used. When Design-Expert is used to construct a 2^{4-1} design, the user receives a *warning message* which states that effects must be at least 2 standard deviations in size in order to be detected. (Actually the discussion of a 2^3 design in

TABLE 5.2 The Two 2^{3-1} Fractions

First Fraction				Second Fraction			
Treatment Combination	A	B	C	Treatment Combination	A	B	C
a	+	−	−	(1)	−	−	−
b	−	+	−	ab	+	+	−
c	−	−	+	ac	+	−	+
abc	+	+	+	bc	−	+	+

Section 4.6 applies here since the number of design points is the same. Recall that it was stated there that 2.29σ is the smallest effect that can be detected, assuming that σ is known.) Lynch (1993) gave tables for minimum detectable effects for 2^{k-p} designs, with most of the entries for a small number of runs based on the assumption that there were at least three degrees of freedom for estimating σ. Those tables show that for $\alpha = .05$ and a power of .90 for detecting significant effects, the minimum detectable effect (in standard deviation units) for a 2^{4-1} design is 3.5. A 2^{5-1} design fares somewhat better as the minimum detectable effect is 2.5, but this will also be too large for many practitioners. (Of course these results cannot be compared directly with results stated under the assumption that σ is known. The result of 2.29σ given earlier in this paragraph is certainly more in line with the warning message given by Design-Expert.)

One way to view the 2^{3-1} design—and this applies in a general way to all 2^{k-p} designs—is that each fraction contains a full factorial, namely a 2^2 design in any pair of the factors. This can be seen by writing out the treatment combinations as shown in Table 5.2.

Notice that all four combinations of plus and minus signs occur with each pair of factors in each fraction. This result extends generally in that there is a full factorial in $k - p$ factors in each fraction whenever a 2^{k-p} design is constructed. The importance of this is that if, by chance, only $k - p$ factors seem to be important, there is a full factorial in those factors that can be used for analysis. If the number of apparently significant factors, w, is less than $k - p$, then there is a replicated 2^w design within the 2^{k-p} design. This is illustrated in Section 5.11.

5.1.1 Which Fraction?

Software for DOE generally provides the user with the capability to specify a particular fraction. For example, in MINITAB the FRACTION subcommand for the FFDESIGN command can be used to specify the fraction that is to be used. The following commands would be used to select the second fraction in Figure 5.1.

```
MTB> FFDESIGN 3 4 ;
SUBC> FRAC 2.
```

Without the subcommand, MINITAB will simply list the first fraction, as the fraction that contains all plus signs for the confounded interaction(s) is the default fraction in MINITAB, and is termed the principal fraction, in general. JMP automatically

produces the same fraction but does give the user certain options, as discussed in Section 5.2.1. Design-Expert allows any desired fraction to be constructed when the Make Generators Editable option is invoked. The user can then specify the generator(s). For the 2^{3-1} design being discussed in this section, the user would enter $C = -AB$ to produce the other fraction (i.e., not the principal fraction).

The obvious question to address at this point is "Does it make any difference which fraction we use?" We can address this question relative to (a) having no prior information, and (b) having certain prior information. If we have no prior information, we would have no reason to select one fraction so as to not "miss anything," whereas prior information might lead us to select a particular fraction. For example, Daniel (1976, p. 41) stated that in his experience interactions are usually "one-cell interactions." That is if an additive model is fit to the data, only one cell will usually have a large residual that stands out. If desired, a significance test could be used to determine a "large residual." This is discussed by Daniel (1976, p. 40) and elsewhere.

If we anticipated this and had no other a priori information, we would want to select the fraction that included that corner point. For example, let's assume an experiment was run and the *AC* interaction was found to be significant, and that this was due primarily to a large value for the *abc* treatment combination. If this outcome had been anticipated, then the fraction that contains treatment combination *abc* could have been (nonrandomly) selected. If, however, this was not anticipated and the other fraction was used, then there is a strong possibility that the conclusion regarding the *AC* interaction would be incorrect. Therefore, prior information should be utilized in design construction if it is available and reliable.

If, however, no prior information is available and in fact no unusual values would have been observed if the full factorial had been run, then one or more conclusions from using one fraction that differ from the set of conclusions obtained using the other fraction would expectedly be due only to random variation.

This obviously raises the question of how much random variation might be expected between the results obtained using the two fractions. This question is addressed in Section 5.1.4.

5.1.2 Effect Estimates and Regression Coefficients

The manner in which effect estimates and the regression coefficients for the selected model are obtained for a full factorial was explained in Section 4.1 and in Appendix B of Chapter 4. It works essentially the same way for a fractional factorial, except that we cannot estimate all interaction effects up to one order less than the number of factors. Specifically, we know that we cannot estimate the *ABC* interaction in a 2^{3-1} design since the interaction is confounded with the difference between the two fractions.

5.1.3 Alias Structure

Let's return to the experiment described by Daniel (1976) and look at the numbers in a somewhat different way, still thinking about what would have happened if a 2^{3-1} design had been used.

In particular, let's look for extreme observations and see what their effect will be. With only eight observations, as given in Table 5.1, it is easy to see that there are two numbers that are much larger than any of the other numbers. We would be especially concerned if these numbers were in the same fraction, but that does not happen as treatment combination a is in the principal fraction and (1) is in the other fraction.

Thus if a 2^{3-1} design had been used instead of the 2^3 design we would have had one large number with which to be concerned. What is common to those two treatment combinations is that the other two factors are each at their low levels. The comparison of, say, b versus (1) and c versus (1) suggests that the main effects B and C may be significant (as they are), as each effect estimate is the average of four differences and one of the differences is very large relative to the magnitude of the numbers. Since hardly any designed experiments are run flawlessly, these two large observations should have been checked out. There is no evidence from Daniel (1976), however, that the experiment resulted in any bad data.

Therefore, we will assume that all of the data points are valid. The estimates of the effects from each fraction, some of which were given in Section 5.1, are as follows.

Fraction			
I = ABC		I = −ABC	
Effect	Estimate	Effect	Estimate
A + BC	63	A + (−BC)	−32
B + AC	−131	B + (−AC)	−134
C + AB	−60	C + (−AB)	−87

Notice that the A effect differs greatly between the two fractions, which results from the fact that the two large values are in different fractions, as mentioned previously, with one at the high level of A and the other at the low level. This means that when the A effect estimate is computed, the large value at the high level of A has a plus sign attached to it in the computation, whereas the other large value has a minus sign attached to it, in the other fraction. Although a 2^{3-1} design is not a particularly useful design, as stated previously, this does illustrate how the estimates of effects from the two fractions can differ greatly on data from an actual experiment.

Furthermore, since the experiment was actually run as a 2^3, we know the separate main effect and interaction effect estimates. These estimates can be obtained from the results given above, as we might expect. For example, since $B + AC = -131$ and $B + (-\text{AC}) = -134$, if we add the two equations together we obtain $2B = -165$, so the estimate of the B effect from the two fractions combined is $-265/2 = -132.5$. Similarly, if we subtract $C + (-AB) = -87$ from $C + AB = -60$, we obtain $2AB = 27$, so the estimate of the AB interaction effect is $27/2 = 13.5$.

As is explained in Section 5.4, if we ran one of these fractions and then later decided to run the other fraction because we became concerned about possible interaction effects, we would be using a *foldover design*. That is, we would be "folding over" the first fraction (i.e., changing all of the signs in the first fraction) to obtain the second fraction, with the foldover design being the full 2^3 design.

In this case, folding over one of the fractions is the appropriate thing to do because there is a large *BC* interaction effect (the estimate being 47.5), which causes the estimates from the two fractions to differ greatly since the estimate of the *A* effect in one fraction is really the estimate of $A + BC$, whereas it is $A - BC$ in the other fraction. Obviously the estimates would not differ by very much if the *BC* effect were small.

As a practical matter, if a fractional factorial were run and the sign of an effect was unanticipated, with the magnitude of the effect not being small, more runs, such as in a foldover design, should be made. (Note that this problem caused by a large interaction effect is different from the problem of a large interaction effect necessitating an analysis using conditional effects.)

5.1.4 What if I Had Used the Other Fraction?

Cochran and Cox (1958) gave an example of a field experiment performed in 1936 at the famous Rothamsted Experimental Station that had four factors, each at two levels. The experimental area was divided into two blocks of land, with each block containing eight plots because the experimenters suspected that the blocks of land might be dissimilar. So the experiment was run with two blocks and the *ABCD* effect was confounded with the difference between the blocks.

Since it can be shown that there is no evidence of a block effect (see Section 4.15 for the analysis of factorial designs run in blocks), we will proceed to analyze the data as two half fractions of a 2^4 design. We would look at only the main effect estimates since the two-factor interactions are confounded in pairs, with the main effects confounded with three-factor interactions.

I = ABCD		I = −ABCD	
Effect	Estimate	Effect	Estimate
A	−6.25	A	4.75
B	−5.75	B	−10.25
C	2.25	C	−1.75
D	−4.25	D	−0.25

Although a normal probability plot analysis is somewhat shaky with only eight data points, using such an analysis with the data in the second fraction results in the *B* effect identified as significant, whereas no effect is identified as being significant for the first fraction. Thus, the conclusion depends on which fraction is used.

One problem we can immediately recognize is that we can't realistically perform a conditional effects analysis for a resolution IV design because we can't determine how to split the data to look at the conditional effects. This is because we perform the splits by looking at the magnitude of the two-factor interactions, but the two-factor interactions are confounded in pairs in resolution IV designs.

It is instructive to compare the results obtained using each fraction with the results obtained using all of the data and analyzing it as a 2^4 design. When this is done, the normal probability plot analysis identifies the main effect of *B* as the only significant effect.

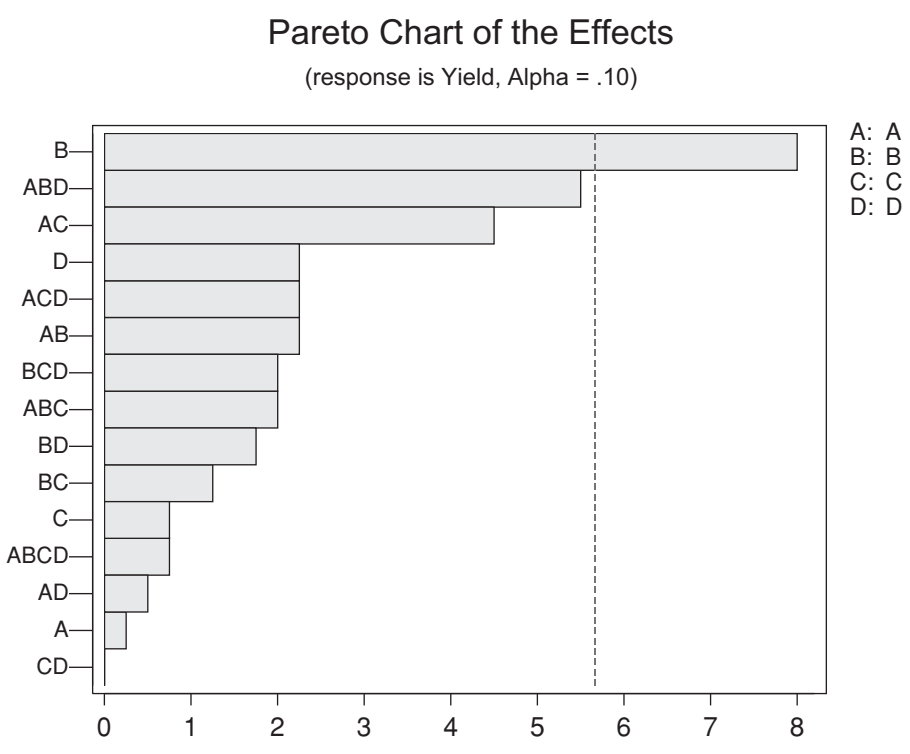

Figure 5.3 Pareto chart of effect estimates.

What we see with this analysis that of course isn't available with either fraction is that the *BCD* interaction is the second largest effect, as is shown in the Pareto chart of the effects that is given in Figure 5.3. (Of course this interaction is confounded with the *A* effect in each fraction, and we know that three-factor interactions are generally not real.)

If the 2^{4-1} design had been used, it would have probably seemed safe to have the *A* effect confounded with the *BCD* interaction, but we see from the full analysis that the *BCD* interaction is just barely below the line that separates effects that are judged to be significant from those that are not so judged. Similarly, if we do a normal probability plot analysis and use $\alpha = .11$ in MINITAB instead of the default value of .10, the *BCD* interaction shows as being significant. Therefore, it is not an effect that can be safely ignored.

It is of interest to determine the effect, if any, that the large *BCD* interaction has on the *B* effect estimate. Although normally we don't think of going directly from a three-factor interaction to conditional main effects, we can do so in two steps. The large *BCD* interaction means that all two-factor interactions involving those letters are unreliable. The largest of these interactions is the *BC* interaction, whose conditional effects, splitting on the *D* factor, are $2.25 \pm (-5.50)$. We would then obtain the conditional *B* effects from the four observations that produce the larger of these

two numbers (7.75), which occurs at the low level of D. Specifically, the conditional effects of B are -14 at the low level of C and 1.5 at the high level of C.

Although each of these last two numbers is computed from averages of only two observations, and would both be more meaningful and have a smaller variance if the design were replicated, the numbers do suggest that there may be a considerable B effect at the low level of C and the low level of D.

This analysis simply shows the region of the factor space where the B effect is the greatest; this information is not available with traditional analyses. Of course the B effect was significant in the full factorial analysis and in the second of the two half fractions, so this simply adds to that information.

It would be of greater interest to discover a conditional main effect of a similar magnitude, also based on four numbers, for which the main effect was not significant in the full factorial or with either of the half fractions. From inspection of Figure 5.3 we can see that this is not going to happen, although we note in particular that the small unconditional A and C effects in the full analysis are very misleading because of the large AC interaction. We can observe that the conditional main effects of A and C, with the data split, in turn, on each, will not differ much from plus and minus the AC interaction since both main effect estimates are close to zero.

Those conditional main effects would each be computed using averages based on four numbers. We could, if desired, obtain the conditional main effects based on averages computed using two numbers each. We will not, however, obtain a conditional main effect of C that is anywhere near the largest conditional main effect of B since the unconditional effect of C is close to zero.

Nevertheless, to illustrate the computations, our starting point would again be the BCD interaction and we would split on D and again focus on the low level of that factor for the same reason that we did so in looking at the conditional main effects of B. This leads to -4.0 and 7.5 as the conditional main effects of C. It should be noted that these numbers bear no relationship to the unconditional main effects since only half of the data are being used. This differs from the case when the conditional main effects are obtained from only a single split because in that case all of the data are being used.

Accordingly, we could, if desired, designate conditional main effects in accordance with the number of splits that were performed in producing the estimates, so these conditional effects for B and C might be called *two-split conditional main effects*.

Although the largest, in absolute value, two-split conditional main effect of C of 7.5 is only about half of the largest two-split conditional main effect of B, it is nevertheless large enough to tell us that we should not overlook factor C.

We will pursue the topics of the "fraction effect" and multi-split conditional main effects further in Section 5.2.

5.2 2^{k-2} DESIGNS

In this section we will use an experiment performed using a 2^5 design and analyze the data as four 2^{5-2} designs and also as two 2^{5-1} designs to further illustrate and reinforce

the concepts described in Section 5.1. This is somewhat akin to seeing how certain methodology works when we know the right answers, which is generally the case only when simulated data are used. Of course we don't exactly have the "answers" when a 2^5 design is used, but this has enough experimental runs, especially when hidden replication is considered, to give an experimenter reasonable confidence in the results.

Example 5.1

Kempthorne (1973, p. 267) gave the results of an experiment performed at the Rothamsted Experimental Station to determine the effect of certain fertilizers on the yield of mangolds. A 2^5 design was used and the five factors were sulphate of ammonia, superphosphate, muriate of potash, agricultural salt, and dung. Each of these was a presence–absence factor in that one of the two levels of each factor was "none." The experiment was apparently run in four randomized blocks of size 8, and we will analogously analyze each of the blocks as if each block were a $1/4$ fraction, in addition to analyzing the data as having coming from a 2^5 design without blocking, which would be defensible if there were no block effect, as there wasn't in this example.

Kempthorne (1973) did not do a full analysis of these data, but did construct an error term with 13 df using the 13 three-, four-, and five-factor interactions that are unconfounded with blocks. Of course, this analysis was performed over 50 years ago as it was in the original 1952 edition of the book. Today it would be better not to assume that all three-factor interactions, in particular, do not exist. In fact, we will see for these data that some high-order interactions show as being significant in a normal probability plot analysis.

We will first analyze the data as an unblocked 2^5 design, ignoring the way that the design was run. (Assume that this information was initially lost.) The normal probability plot for this analysis is given in Figure 5.4.

We should be suspicious of these results simply because a four-factor interaction is identified as being significant, in addition to two three-factor interactions. Now assume that the blocking was discovered. The proper analysis is then to isolate the block effect and the three effects that are confounded with the difference between blocks: the *ABD*, *BCE*, and *ACDE* interactions. Of course this will affect the assessment of significant effects in two ways: the *BCE* interaction will not be a significant effect, as it was previously, since it will be confounded with blocks, and the pseudo-error term implicit in the assessment of significant effects will be affected by the fact that the very small *ABD* interaction won't be part of the pseudo-error computation since it is confounded with blocks.

The "corrected" results are given in Figure 5.5.

We note that the *ACE* interaction is still significant, and since three-factor interactions generally are not significant, this should make us somewhat suspicious of the data. Kempthorne (1973) gave the *bd* treatment combination as having a yield of 864 in his Table 14.11, and that was the value used in these two analyses. However, 964 was the value for that treatment combination given in Table 14.10. Furthermore, the

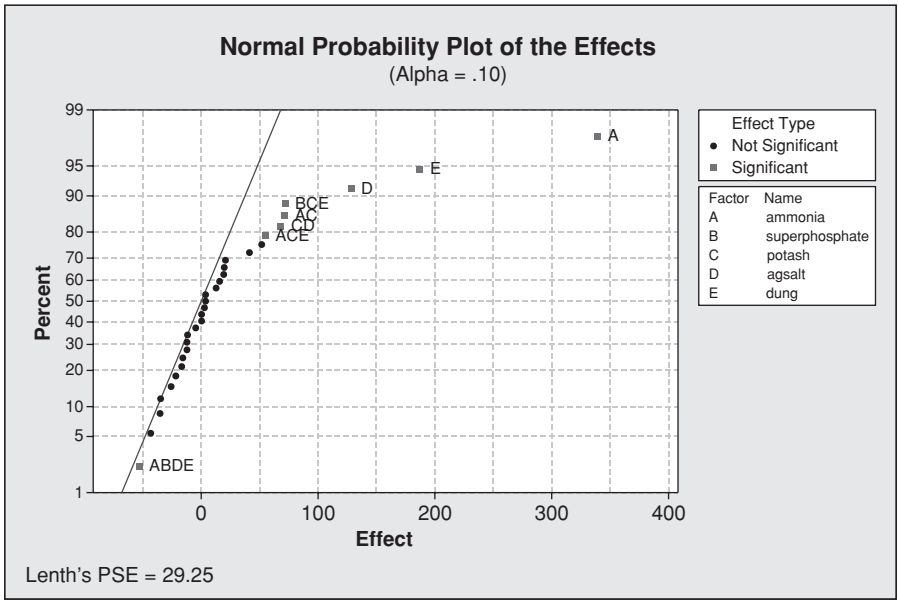

Figure 5.4 Analysis of Kempthorne data—ignoring blocking.

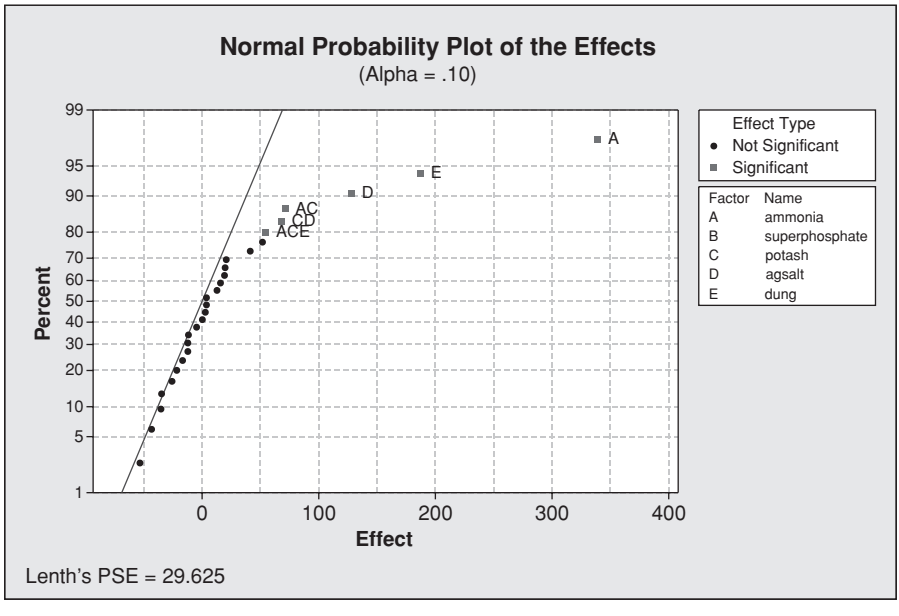

Figure 5.5 Analysis of Kempthorne data—with blocking.

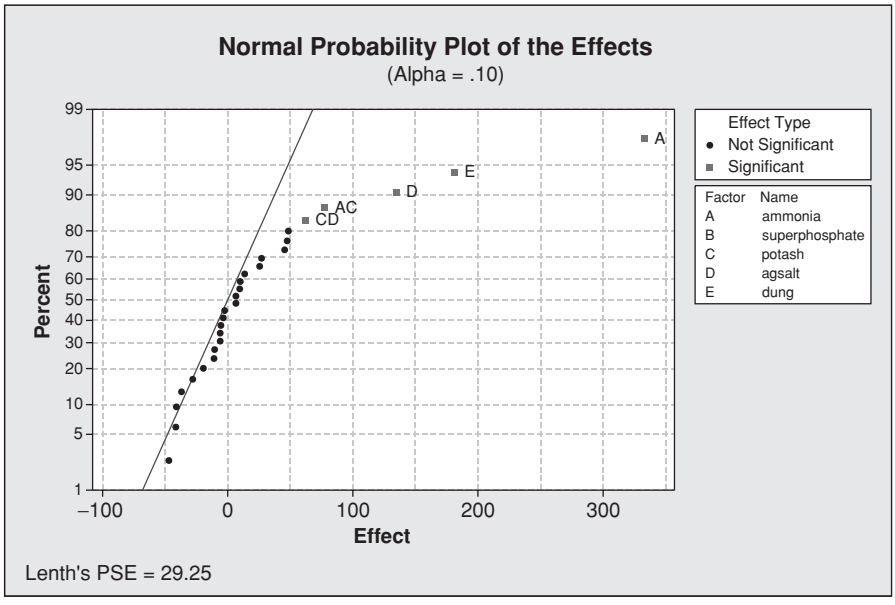

Figure 5.6 Kempthorne data—error corrected.

numerical results given in his Table 14.12 are those that result when 964 is used, so this appears to be the correct value. The normal probability plot with the corrected value is shown in Figure 5.6.

These results look more reasonable than the previous results. When the individual fractions are analyzed, the results are as follows. For the first fraction, only the *A* effect shows as being significant in the normal probability plot analysis, even though several other points are well off the line. No effect is identified as significant in the second fraction, despite the fact that there are some large effect estimates, and the same thing happens with the third fraction. Only the *A* effect is identified as significant in the fourth fraction.

Thus, there is agreement only between the first and fourth fractions and second and third fractions. All of these fractions fail to identify four effects in the full factorial analysis that were significant, and of course the second and third fractions miss all five effects. Why does this occur? Notice that there are many effect estimates in Figure 5.6 that are practically on the line. As has been emphasized previously (e.g., Section 4.10), a normal probability plot analysis works only if there is a sufficient number of small effect estimates. The normal plots for each of the four fractions are quite different from Figure 5.6, with no point being very close to the line. The alias structure is not the culprit as none of the significant effects in Figure 5.6 is aliased (with each other) in any of the fractions.

This example is important because it illustrates that (1) the results obtained from different fractions can easily disagree, and (2) the results from the fractions can differ greatly from the result that would be obtained from the full factorial as the

TABLE 5.3 Treatment Combinations of Four 1/4 Fractions of a 2^5 ($I = ABC = BDE = ACDE$)

(1)	(2)	(3)	(4)
(1)	*ab*	*a*	*b*
ac	*acd*	*c*	*cd*
de	*ace*	*abcd*	*ce*
acde	*d*	*abce*	*bde*
abd	*e*	*bd*	*abc*
abe	*bc*	*be*	*ae*
bcd	*bcde*	*ade*	*ad*
bce	*abde*	*cde*	*abcde*

necessary requirement of effect sparsity for a normal probability plot analysis to work is obviously more likely to be met when a small number of effect estimates are plotted than when there is a much larger number of effects.

5.2.1 Basic Concepts

In general, when the fraction is of the form $1/2^p$, there will be 2^{p-1} effects that must be confounded with the difference between the fractions, and their product(s) will also be confounded. Thus, for the 2^{5-2} design in the preceding example, we must select two effects to confound, and their product will also be confounded. So in that example, *ACDE* was the product of *ABC* and *BCE*.

Since, as stated in the Important Point at the beginning of the chapter, the resolution of a two-level fractional factorial is defined as the number of letters in the shortest word in the defining relation, it might seem as though we should select *ABCDE* and one of the four-factor interactions. This would be disastrous, however, as their product would contain only a single letter, and we would consequently lose a main effect. Similarly, a little reflection should indicate that any pair of four-factor interactions will have three factors in common, and will thus produce a two-factor interaction when multiplied together. We can, however, choose a pair of three-factor interactions in such a way that the product is a four-factor interaction. For example, *ABC* and *BDE* would produce *ACDE* when multiplied together. (There is no advantage in selecting any particular pair, as only main effects are estimable, anyway.)

The alias structure is then obtained by multiplying each of the effects by the three defining contrasts. For example, $A = BC = ABDE = CDE$. The four 1/4 fractions could be constructed by determining the treatment combinations that have (1) an even number of letters in common with both *ABC* and *BDE*, (2) an even number in common with *ABC* and an odd number in common with *BDE*, (3) an odd number in common with *ABC* and an even number in common with *BDE* (i.e., the reverse of (2)), and (4) an odd number in common with both *ABC* and *BDE*.

The four fractions are given in Table 5.3.

The five main effects would be estimated in the usual way—the average of the four values in which the factor is at the high level minus the average of the four values in which the factor is at the low level. For example, if the first fraction were used,

the main effect of A would be estimated by $(ac + abe + abd + acde)/4 - [bcd + bce + de + (1)]/4$.

The estimation of the interaction effects with this design also follows the same rule that is used in estimating interaction effects with full factorials and other fractional factorials. That is, the average of the response values for which the product of the signs of the factors that comprise the interaction is positive is computed, and from that average is subtracted the average of the response values for which the product of the signs is negative.

It should be noted that with this fractional factorial there will be two df for the residual, so the plotting of effect estimates on normal probability paper would not be absolutely essential, although generally desirable since the two df correspond to two pairs of two-factor interactions that are not being estimated. Rarely will there be sufficient prior information to strongly suggest that none of the four interactions is real.

It was stated previously that obtaining the design configuration by enumerating the treatment combinations is somewhat laborious when there are more than just a few factors. For 1/4 fractions, however, that is the most straightforward approach. Another approach is described by Box et al. (1978, p. 397), who show how to obtain a 2^{5-2} design from a 2^{7-4} design by dropping columns from the latter. Of course we prefer to use software to enumerate all of the fractions.

It is of course highly desirable for experimenters to understand the mechanics of design construction and effect estimation so that they can have the knowledge about when they should construct custom designs rather than routinely depending on the default options in statistical software to construct designs. Once that knowledge has been acquired, experimenters can confidently rely on the extensive design construction capabilities of readily available statistical software.

For example, in MINITAB the principal fraction for the design with defining relation $I = ABD = ACE = BCDE$ is constructed and stored with the commands

```
MTB> FFDESIGN 5 8;
SUBC> XMATRIX C1-C5.
```

Note that the principal fraction is the one that has all plus signs in the defining relation. Another one of the four fractions would have the defining relation $I = ABD = -ACE = -BCDE$, with of course the sign for the third interaction being the product of the signs for the first two interactions.

As stated in Section 5.1, fractional factorial designs in the 2^{k-p} series can be viewed as full factorials in $k - p$ factors, so if for some reason a different alias structure were desired, such as to avoid confounding effects that are believed a priori as being possibly significant, it is easy to construct such a "custom design" with practically any software. For example, in MINITAB, a 2^{5-2} design with the alias structure in the example in Section 5.2 could be constructed with the commands

```
MTB > FFDESIGN 3 8;
SUBC > ADD D = AB E = AC;
SUBC > XMATRIX C1 - C5.
```

It is also possible to specify the alias structure in JMP by specifying how, in this case, the two additional variables are defined by using the Change Generating Rules option when the "Screening Designs" capability is used.

5.3 DESIGNS WITH $k - p = 16$

Snee (1985) considered the 2^{5-1}, 2^{6-2}, 2^{7-3}, and 2^{8-4} designs for the study of five, six, seven, and eight factors, respectively, to be the most useful fractional factorial designs. Accordingly, it is useful to examine these designs for their power to detect real effects, using the tables in Lynch (1993). Table 3 (for $\alpha = .05$) shows that if $\beta = .90$, as in Chapter 4, the minimum detectable difference between factor level means is 2.5σ, assuming that at least three degrees of freedom are used to estimate σ, with these degrees of freedom obviously coming from high-order interactions if the design is not replicated. Of course, replicating a fractional factorial design would seem to defeat the purpose of using such a design, but Table 3 in Lynch (1993) reveals that the minimum detectable difference in factor level means is $k\sigma$ with $k = 1.2, 1.3, 1.4$, and 1.8 for five, six, seven, and eight factors, respectively. Thus, if 32 design points could be afforded, these 16-point fractional factorial designs should be replicated whenever users consider 2.5σ to be unacceptable, as will often be the case.

The obvious alternative to doing this would be to use a 2^5 design for five factors, a 2^{6-1} design for six factors, a 2^{7-2} design for seven factors, and a 2^{8-3} design for eight factors—all of which are 32-point designs.

Which approach is preferable? If there are hard-to-change factors or debarred observations (see Section 4.19 and Section 13.8, respectively), then it would obviously be better to use 16 factor–level combinations than 32. The resolution of these designs must be considered, however. The respective comparisons are 2^{5-1} versus 2^5, 2^{6-2} versus 2^{6-1}, 2^{7-3} versus 2^{7-2}, and 2^{8-4} versus 2^{8-3}. The results in terms of resolution are V versus not defined, IV versus VI, IV versus IV, and IV versus IV. Thus, except in the case of six factors, for which the loss in resolution would be deemed unacceptable if a design must be at least resolution V, using two replicates of a 16-point fractional factorial design is not a bad idea.

5.3.1 Normal Probability Plot Methods when $k - p = 16$

Haaland and O'Connell (1995) gave specific recommendations for normal probability plot methods used with fractional factorial designs with 15 estimable effects. Recalling the general form of Eq. (4.3) that was given in Section 4.10,

$$\widehat{\sigma}_{\text{PSE}}(q, b) = a_{\text{PSE}}(q, b) \cdot \text{median}\,\{|\widehat{\theta}_i| : |\widehat{\theta}_i| \le b \cdot s_0(q)]\} \tag{5.1}$$

and Eq. (4.4), which was

$$s_0(q) = a_0(q) \cdot \text{quantile}\,\{q; |\widehat{\theta}_i|\, i = 1, \ldots, k\} \tag{5.2}$$

for simplicity they recommended selecting q such that quantile $\{q; |\widehat{\theta}_i| i = 1, \ldots, k\} = |\widehat{\theta}|_{(m)}$, with the latter denoting the mth order statistic of the absolute values of the effect estimates. If more than 6 of the 15 estimable effects are expected to be real effects (which seems unlikely to occur very often), they recommend using $m = 7$ and $b = 1.25$ in Eq. (5.1). If very few real effects are expected, the test given by Dong (1993) is recommended, with Lenth's (1989) method recommended whenever a moderate number of effects are expected. Of course, there may not be any a priori beliefs about the number of real effects. If so, then Lenth's (1989) method was recommended.

A much more extensive study was performed by Hamada and Balakrishnan (1998), who preferred the methods of Daniel (1959), Schneider, Kasperski, and Weissfeld (1993), Zahn (1975), Berk and Picard (1991), and the half-normal plot version of Loh (1992) for up to 6 real effects out of 15 estimable effects. Certainly the latter should be the usual scenario. (Their study and the entire paper are available online at http://www.stat.sinica.edu.tw/statistica/oldpdf/A8n11.pdf.)

More recently, Miller (2005) proposed a totally different type of (nongraphical) procedure, one that could be used for arbitrary n, which utilizes a selection score in arriving at the best model through implicitly comparing all possible candidate models. The author showed with simulation that the method is often superior to Lenth's method. Nevertheless, new methods such as this one are not likely to be used to any extent unless they are implemented in statistical software.

5.3.2 Other Graphical Methods

Certain novel (unpublished) graphical methods for experimental design data have been developed by Jim Filliben of the National Institute of Standards and Technology (NIST). For example, in an internal NIST publication (J. J. Filliben, K. Inn, and R. Radford, "Troubleshooting plutonium contamination in NIST radionuclide shellfish measurements via experiment design," 2004) an ordered dataplot is given in which the ordered Y-values are plotted against the treatment combinations. When there are extreme observations—and three such observations were shown in their graph—the reader can easily see if those observations appear to be associated with certain factor levels, or combinations of factor levels.

5.4 UTILITY OF SMALL FRACTIONAL FACTORIALS VIS-À-VIS NORMAL PROBABILITY PLOTS

It should be apparent from previous sections that small, unreplicated, fractional factorials, such as a 2^{4-1}_{IV} design, can present problems when a normal probability plot analysis must be employed. This is because there must be a sufficient number of small effect estimates for the normal plot approach to work, as has been stated previously, but that requirement will often not be met when there simply aren't very many effects that are estimable.

It is entirely possible that some two-factor interactions exist when a 2^{4-1}_{IV} design is used. If so, since these interactions are confounded in pairs, there might not be more than one or two small interaction pairs, which would cause the normal plot to likely fail if there were more than one significant main effect.

One way to avoid this problem is to simply not use these small designs and to be guided by results such as those given in Hamada and Balakrishnan (1998). Some of their recommendations were mentioned briefly in Section 5.3.1. They also discussed the difficulty in identifying the real effects and not selecting the null effects when designs have only eight points because of the substantial variability in the plots when all effects are null.

If experimentation is so expensive in a particular application that a small, unreplicated fractional factorial design must be used, thus necessitating the use of a normal probability plot approach, then selective replication might be employed, combined with checks on process control. Specifically, it is important that processes be maintained in a state of statistical control when experimentation is performed, as was stressed in Section 1.7. Therefore, runs at the standard operating conditions might be performed at the beginning, the end, and possibly in the middle of the experimental runs. Alternatively, the first run that is performed might be replicated in the middle of the experiment and at the end of the experiment, which would require one fewer run than if the standard conditions were replicated.

If an experimenter is daring, these replicated runs might be used to estimate the error term, which would be used in hypothesis tests, obviating a normal plot approach. There is a big risk in this approach, however, in that (1) it is not very practical to estimate a variance based on only a few replicated points, and (2) it would be necessary to assume that the variance is the same at each design point—an assumption that could not be checked.

The work of Carroll and Cline (1988) and others on variance estimation in regression is somewhat applicable here as they showed that at least 8–10 replications of a design point are necessary in order to obtain a good estimate of the variance at that point.

Therefore, it would be unwise to try to estimate the error variance from a few replicates of either the standard conditions or one of the design points. These replicates should be used simply to provide a rough check on process control, nothing more. Unfortunately, however, probably only a tiny fraction of experiments is run with process control checks. This is undoubtedly due in part to the fact that experimenters in general are not experts in process control and are probably not well aware of the extent to which experiments can be undermined when relevant processes go out of statistical control.

Of course the replications must be true replications and not just multiple readings, with the distinction between the two having been made in Section 1.4.2. This means that factor levels must generally be reset, even though they are to be reset to the same value. If true replicates are not obtained, then an analysis using the average of the repeated values as a single observation would be appropriate. If it is uncertain whether or not true replicates have been obtained, it would be wise to analyze the data using both approaches. This was done by Ryan (2003) in analyzing some lead recovery

data for the NIST. If the results differ considerably, then it would be best to draw conclusions from the analysis obtained using the averages.

5.5 DESIGN EFFICIENCY

As discussed by Mee (2004), Daniel (1956) defined the degrees-of-freedom efficiency of designs for estimating all main effects and two-factor interactions as

$$\text{df-efficiency} = \frac{\text{Number of effects estimated}}{\text{Total degrees of freedom}}$$

For k factors there are obviously k main effects and $k(k-1)/2$ interaction effects, so the total number of effects is $0.5k(k+1)$. Since the total degrees of freedom is always $n-1$, df-efficiency can thus be written as

$$\text{df-efficiency} = \frac{0.5\,k(k+1)}{n-1}$$

Designs that are df-efficient are designs for which the fraction is close to 1.0, which would be 100 percent efficiency. A 2^{5-1} design is 100 percent efficient since numerator and denominator are both 15, this being the only commonly used fractional factorial design in the 2^{k-p} series that is 100 percent efficient. As shown by Figure 1 in Mee (2004), most 2^{k-p} designs have efficiencies of about 50 percent and most designs when k is large have efficiencies well below 50 percent. As discussed by Mee (2004), nonorthogonal designs have been proposed as an efficient alternative.

Mee (2004) gives design recommendations for $5 \leq k \leq 15$ with the 2^{5-1} design being the only 2^{k-p} design recommended.

5.6 RETRIEVING A LOST DEFINING RELATION

As pointed out by Bisgaard (1993), it is not uncommon for information on the defining relation to be lost, thus necessitating that it be determined in some manner. Various researchers have addressed this problem, with Margolin (1967) being among the first to do so. His method is also described in John (1971, p. 160).

Bisgaard (1993) gave a simple, matrix-based method for retrieving a defining relation that is more efficient than previous methods. The method utilizes the alias matrix, which was originally given by Box and Wilson (1951). Assume that we have fit a model of the form $\mathbf{Y} = \mathbf{X}_1\mathbf{b}_1 + \mathbf{e}$, but the true model is $\mathbf{Y} = \mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2\mathbf{b}_2 + \mathbf{e}$, with $\mathbf{X}_1$ denoting the matrix whose columns contain the values of the terms used in the model and $\mathbf{X}_2$ denoting the corresponding columns for terms that are in the true model but not in the fitted model. These matrices will have $+1$ and -1 for their entries.

Using standard regression and linear model expressions, $E(\widehat{\mathbf{b}}_1) = (\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_1)^{-1}\mathbf{X}_1^{\mathrm{T}}E(\mathbf{Y}) = (\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_1)^{-1}\mathbf{X}_1^{\mathrm{T}}(\mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2\mathbf{b}_2) = \mathbf{b}_1 + (\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_1)^{-1}\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_2\mathbf{b}_2 = \mathbf{b}_1 + \mathbf{A}\mathbf{b}_2$, with $\mathbf{A} = (\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_1)^{-1}\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_2$ denoting the alias matrix.

The defining relation can be easily determined by adapting this result, although not in an obvious way. Specifically, let $\mathbf{X}_1$ represent the k factors in the design matrix and let $\mathbf{X}_2$ represent all possible interactions formed from the factors in $\mathbf{X}_1$ that are of the order that would be used in constructing a specific design. (The last point will be explained later.)

To illustrate, consider the following 2^{5-2}_{III} design.

A	B	C	D	E
−1	−1	−1	1	1
1	−1	−1	−1	−1
−1	1	−1	−1	1
1	1	−1	1	−1
−1	−1	1	1	−1
1	−1	1	−1	1
−1	1	1	−1	−1
1	1	1	1	1

As usual the first three columns form the basis columns, as they constitute a full factorial in $5-2 = 3$ factors. We could easily use trial and error to determine how D and E were formed, but such an approach would be laborious for at least a moderate number of factors. So let $\mathbf{X}_1$ be composed of the first three columns and let $\mathbf{X}_2$ be composed of all two-factor interactions since we know that such designs are constructed using two-factor interactions to represent the fourth and fifth factors.

Straightforward matrix multiplication, using software, yields

$$\mathbf{A} = \begin{matrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

Multiplying this matrix times the vector of 10 interactions (*AB* through *DE*)

$$\begin{pmatrix} AB \\ AC \\ AD \\ AE \\ BC \\ BD \\ BE \\ CD \\ CE \\ DE \end{pmatrix}$$

shows that $D = AB$ and $E = AC$, in addition to the aliasing of the first three factors, which is irrelevant as we need only these two to determine that the defining relation is $I = ABD = ACE = BCDE$. Of course in this simple example we can observe that $D = AB$ and $E = AC$ from inspection of the five columns that form the design matrix.

5.7 MINIMUM ABERRATION DESIGNS AND MINIMUM CONFOUNDED EFFECTS DESIGNS

Unless there is only a small number of factors to investigate, a fraction smaller than a $1/2$ fraction will be necessary. For example, a 2^{7-1} design requires 64 points, which may be prohibitive if experimentation is expensive, so a 2^{7-2} design or a 2^{7-3} design might have to be used. The defining relation for such designs is more involved than the defining relation for a $1/2$ fraction, as the number of words (effects) in the defining relation is given by the order of the fraction plus the number of generalized interactions of those effects. For example, the highest possible resolution for a 2^{7-2} design is IV, and one possible defining relation for a 2^{7-2}_{IV} design is $I = ABCDE = DEFG = ABCFG$, with the latter being the product (modulo 2) of the first two effects.

The resolution of a fractional factorial design is given by the length of the shortest word, so many 2^{7-2}_{IV} designs could be constructed, one of which would have the defining relation $I = ADEG = BCDEFG = ABCF$. Which design is preferable? The first design is preferable because there would be fewer main effects confounded with three-factor interactions, and more importantly, there would be fewer two-factor interactions confounded among themselves. This is because the word of the shorter length occurs fewer number of times with the first design. For a given resolution and a given number of factors, the design that has the fewest number of words of the shortest length is the *minimum aberration* design, with the term due to Fries and Hunter (1980).

Minimum aberration designs have been automatically generated by Design-Expert (*Source*: http://www.statease.com/e6ug/DE06-Features-Design.pdf (screen 2)) and the two-level fractional factorial designs generated by JMP are also minimum aberration (*Source*: JMP help file for "screening designs"). Although "minimum aberration" is not covered in the MINITAB help file material, the two-level fractional factorial designs are probably minimum aberration.

It is worth noting that minimum aberration designs do not necessarily maximize the number of unconfounded effects. In the example given above the number of two-factor interactions that are not confounded with other two-factor interactions is minimized by the minimum aberration design, so this result may seem counterintuitive. The result is well known, however, and is discussed by, for example, Wu and Hamada (2000, p. 176), who showed that the minimum aberration 2^{9-4}_{IV} design has 8 two-factor interactions that are not confounded with other two-factor interactions, whereas another 2^{9-4}_{IV} design has 15 such two-factor interactions.

Consequently, the *number of clear effects* was proposed as an additional criterion by Chen, Sun, and Wu (1993), whereas others have suggested that the "maximum number

of unconfounded effects" be the criterion that is used (see http://www.statsoftinc.com/pdf/MinimumAberrationPaper.pdf). Of course, the maximum number of clear effects and the maximum number of unconfounded effects would be the same thing. Designs that are efficient in terms of the minimum aberration criterion and other objectives are discussed by Butler (2005). Jacroux (2003) gave a two-stage procedure that first identifies 2^{k-p} designs with the maximum number of estimable effects, and within this class the minimum aberration design is found.

In comparing the best design according to each criterion, assuming that the designs have the same resolution, the design with the minimum number of confounded effects should be chosen.

For a 2^{k-p} design there aren't many different combinations of k and p for which the application of the two criteria will result in the selection of different designs, however. To illustrate, if $p = 1$, the same design results because there is only one word. For $p = 2$, if $k = 5$, the design cannot exceed resolution III because there is obviously no way to construct the columns for factors D and E from the 2^3 design for factors A, B, and C that is the interaction of three factors because there is only one triplet, ABC. Furthermore, if we let, say, $D = ABC$ and $E = AB$, then the generalized interaction is C, so we would not be able to estimate the main effect of that factor and the design would not even be of resolution III. So there is no option but to construct D and E using two-factor interactions, so that each word will be of length 3.

For a 2^{6-2}_{IV} design, the generators for E and F cannot each be a word of length 4 because there is only one such word. Furthermore, one of the two generators cannot be that word because the other generator would have three letters in common with that word, so the generalized interaction would be a three-factor interaction and the design would thus be only resolution III. Therefore, E and F must each be formed using three-factor interactions, and when this is done, the generalized interaction will also be a word of length 4. To see this, note that the generators for E and F will have to differ by one letter for the generalized interaction to have four letters (e.g., $E = ABC$ and $F = ABD$) and they will differ by exactly one letter regardless of which pair is chosen because three of the four letters are being selected. It follows that the defining relation will have two words of length 4 that differ by two letters, so the generalized interaction consists of the two different letters from the first word and the two different letters from the second word. It also follows that no two-factor interaction is estimable because six of the four-factor interactions can be formed from the letters of the first word, five from the second one (since one is accounted for by the first word), and four from the third word (since one is accounted for by each of the first two words). The sum of these three numbers is 15, which is the number of two-factor interactions.

There is no unique 2^{6-2}_{IV} minimum aberration design, and since only the six main effects are estimable regardless of the 2^{6-2}_{IV} design that is used, there is thus no unique design that gives the minimum number of unconfounded effects.

We won't pursue this type of analysis any further and will conclude the discussion by stating that the minimum aberration design and the design with the maximum number of unconfounded effects will generally be the same, as for only a few designs will they differ. (Wu and Hamada (2000) identified six combinations of k and p for

which the two criteria lead to different designs: 2_{IV}^{9-4}, 2_{IV}^{13-7}, 2_{IV}^{14-8}, 2_{IV}^{15-9}, 2_{IV}^{16-10}, and 2_{IV}^{17-11}. The difference for the last design is the greatest as the design that maximizes the number of unconfounded effects allows 31 of the 136 two-factor interactions to be estimable, none of which can be estimated with the minimum aberration design.)

Another proposed criterion, which has been recommended when many two-factor interactions are expected, is to minimize the maximum length of two-factor interaction alias chains (Mee and Block, 2006).

5.8 BLOCKING FACTORIAL DESIGNS

The example in Section 5.2 is an example of a blocked factorial, although that example was used to illustrate variable results between fractions of a single set of data, and also to illustrate how and why the results from the fractions will generally not agree with the result from the full factorial.

In this section we revisit blocking factorial designs, which was also covered in Section 4.15, this time considering the topic relative to fractional factorials and serving as a lead-in to the blocking of fractional factorial designs. Factorial designs are generally blocked because of practical considerations, such as not being able to make all the runs on one day and, in general, not being able to make all the runs under the same conditions. As with a randomized block design (Section 3.1), the blocking should be effective, meaning that we want the conditions to be essentially the same within each block if they are different blocks.

Blocking an unreplicated factorial means that we will have $b - 1$ fewer interactions that are estimable, with b representing the number of blocks, since interactions will be confounded with differences between blocks. As in the selection of the defining relation of a fractional factorial, we don't want to relinquish the estimation of interaction effects that may be significant.

In that respect, we should be guided by the same principles that are used in constructing fractions. For example if we run a 2^4 design in two blocks, we would confound the *ABCD* interaction with the difference between blocks, just as we select that interaction to confound between the difference between the two fractions so as to maximize the resolution of the fractional factorial. The main difference of course is that when we use a fractional factorial we use only one of the fractions, whereas when we block a factorial we use all the fractions. A blocked factorial is "kinder" on a normal probability plot analysis, however, because considerably more effects are estimated with a blocked factorial than with a fractional factorial, so with a blocked factorial a better pseudo-error estimate is likely to result.

Consider what happens when a 2^{k-2} design is used. Two extra columns are created and three effects are confounded with the estimate of the mean. With a 2^{k-2} design we can regard the four fractions as blocks with three effects confounded with the differences between the blocks. Similarly, for a 2^{k-3} design three additional columns would be created and seven interaction effects would be lost because there are four

generalized interactions between the effects corresponding to the three columns: $\binom{3}{2} + \binom{3}{3}$.There are of course eight available fractions with a 2^{k-3} design, and we can regard these as representing eight blocks, with seven interaction effects confounded with the differences between the blocks. And so on.

The bottom line relative to this analogy is that we can construct the blocks analogous to the way that we construct fractional factorials—the difference is simply that we use all the "fractions." Whether or not we hope to see a block effect is likely to depend upon the trouble we go through in constructing the blocks. If the blocks are days, with the blocks formed because the runs must be apportioned to different days, the blocking is natural and does not require any special effort. On the other hand, if the blocking results from careful consideration of differences in raw materials, the blocking may require considerable effort and we would then prefer to see a block effect, just as is the case when a randomized complete block (RCB) design is used. As with an RCB design, there is an assumption of no blocks $\times$ treatment interactions. That is, if raw material problems cause the yield for the second block to be less than the yield for the other blocks, it should uniformly affect all treatment combinations in the second block. If that is not the case, then the usual statistical analysis that is performed will be flawed.

Wu and Hamada (2000) gave a table of blocking schemes for 2^k factorials in 2^q blocks in their Appendix 3A.

5.8.1 Blocking Fractional Factorial Designs

The need to block a fractional factorial also frequently arises. For example, with a 2^{5-1} design there are 16 observations that would have to be run under the same conditions and that may not be possible.

Bisgaard (1994) described an application in which a team of engineers that was working on developing a new motor wanted to estimate the effect on fuel efficiency of a new carburetor design, new manifold design, and so on—eight design changes (factors) were involved, with the levels being "new" versus "present." Sixteen runs were to be made, so this was to be a 2^{8-4} design. There were 10 parallel test stands available at the time of the experiment, and parallel testing would of course reduce the time required to conduct the experiment.

As stated by Bisgaard (1994), an additional factor is required to block a fractional factorial, and in this case the additional (ninth) factor was obvious: the test sites, with two such sites used. Since the 16 runs are made in two blocks of 8 runs each, the question arises as to what effects are lost of the effects that can be estimated with the 2^{8-4} design? Bisgaard (1994) gave tables that can be used for constructing such designs. For the 2^{8-4}_{IV} design run in two blocks of size 8, Table 1 lists $\mathsf{B} = ABCD$ as the generator for the blocks, with B being the blocking factor. Since the design is only of resolution IV, the $ABCD$ interaction is not estimable anyway; this simply indicates the assignment of the treatment combinations to the two blocks. The obvious question is, "Why use this particular generator for the blocks?" As with generators for fractional factorials without blocking, there generally won't be a single optimal

choice of a generator. We should note that the design resolution drops from IV to III when the blocking is performed, even though the blocking factor is confounded with a four-factor interaction. There is a simple explanation for this. The generators for factors E, F, G, and H are all three-factor interactions, but there are only $\binom{4}{3} = 4$ of them, so all of them are used. If a two-factor interaction were used, the resolution would, by definition, be III, so nothing would be gained in terms of resolution as the same general level of aliasing would exist as when $\mathsf{B} = ABCD$.

Bisgaard (1994) also illustrated blocking a 2^{4-1}_{IV} design. Of course the fraction would be constructed using $D = ABC$, so the defining relation is $I = ABCD$. If this design is run in two blocks, what should be the generator for the blocking factor? Clearly it cannot be another three-factor interaction, because the generalized interaction with ABCD would be a two-factor interaction (e.g., if $\mathsf{B} = BCD$, then $I = ABCD = BCD\mathsf{B} = A\mathsf{B}$ so that A be confounded with the blocking factor B). Thus, a two-factor interaction (any one will do) would have to be used, so that the blocking factor would thus be confounded with this interaction and the resolution would be reduced from IV to III. That is unavoidable.

Example 5.2 Case Study

Chokshi (2000) gave a case study in which a 2^{4-1}_{IV} design was run in two blocks. This was a rocket engine fabrication experiment. The components of the rocket thrust chamber are brazed together to form the rocket thrust chamber. There are several thousand braze points and the objective is to come as close as possible to 100 percent of the maximum feasible number of braze points. The percent coverage was considered unsatisfactory, which was the motivation for the experiment.

Six factors were identified from engineering considerations and the use of Pareto chart methods, but the team decided to initially study only four of these factors. Each chamber could handle only four test combinations, so two chambers were used, with chamber thus being the blocking factor. The defining relation for the 2^{4-1}_{IV} design before the blocking was $I = ABCD$ and the blocking factor was confounded with the BC interaction, so that the full defining relation with the blocking was $I = ABCD = \mathsf{B}BC = \mathsf{B}AB$. Since the defining relation shows that the blocking factor is confounded with 2 two-factor interactions, the resolution drops from IV to III.

Curiously, there was one percentage that was far above 100 (115.46) among the eight data points and there was no explanation for this. Each percentage was computed as the average braze length per tube in inches divided by the available braze length per tube in inches, with each average computed from either 180 or 360 measurements. The only way the percentage could exceed 100 would be for at least one braze length to exceed the available braze length. The braze coverage measurements were obtained by X-ray. One set of 180 measurements was listed in the paper and there appears to be at least one other error as one of the listed coverage figures is .225, whereas almost all the other numbers exceed .80 and the next smallest value is .65.

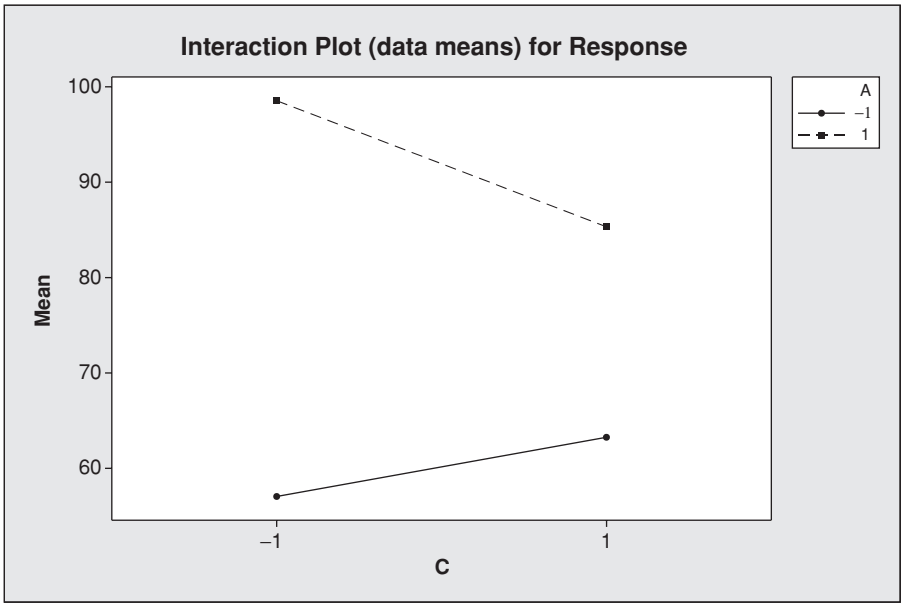

Figure 5.7 *AC* interaction plot for braze coverage data.

Nevertheless, we will use the numbers as given since we have no way of correcting any errors. The numbers for each treatment combination are as follows.

Treatment Combination	Response Value(%)
(1)	46.42
bc	34.94
ad	115.46
abcd	78.01
cd	91.60
bd	67.70
ac	92.64
ab	81.64

There was interest in examining the *AB* and *AC* interactions, in addition to the main effects. These interactions were declared not significant, however, as was the *C* main effect, and these effects were used to create an error term with three degrees of freedom for the ANOVA that was used. It is not clear how this was determined since this is an unreplicated fractional factorial. The *C* factor (which was the number of wires, 0 and 2) does have a noticeable effect at the high level of factor *A*, as can be seen from Figure 5.7.

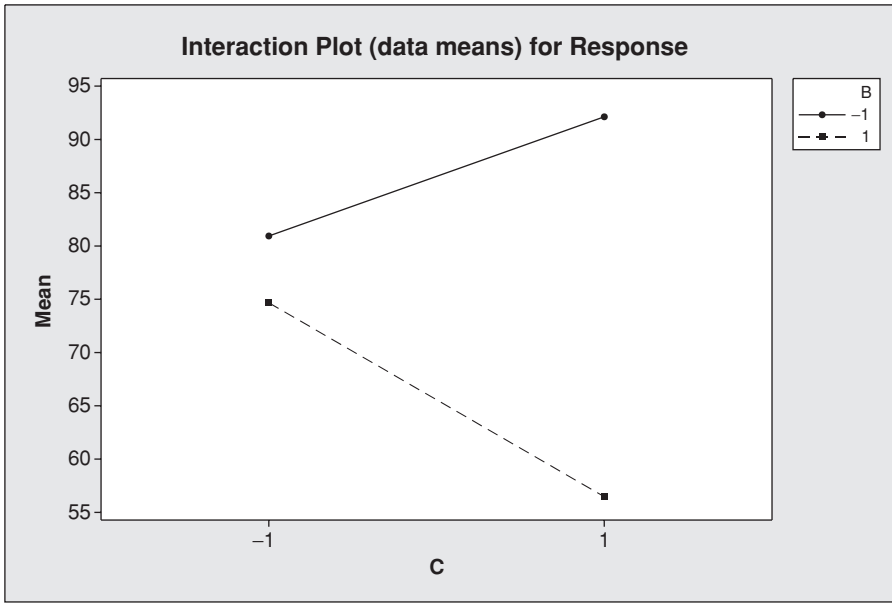

Figure 5.8 *BC* interaction plot for braze coverage data.

Since the slope of one line in Figure 5.7 is positive and that of the other is negative and the absolute values of the slopes do not differ greatly, the interaction effect estimate must be close to zero, which obscures the obvious effect of factor C at the high level of A. The vertical distances between the lines at each level of C being considerably different shows that a single main effect number for factor A will also be misleading.

The lines in Figure 5.7 were stated as being "nearly parallel" in the paper, but they clearly aren't, and the degree of departure from parallelism necessitates a conditional effects analysis.

The BC interaction is even more extreme, as can be seen from Figure 5.8.

This interaction was not discussed in the paper, probably because it was confounded with the block effect by the manner in which the design was constructed. The question that obviously must be addressed is, "Is this the block effect or the BC interaction effect, or some of both?" If there is no reason to suspect that there should be a test chamber effect—and on the surface that would seem to be a reasonable assumption—then this would seemingly be the BC interaction effect. In any event, this is something that should have been resolved with additional runs because the conclusion was that there was no C effect, whereas there is very clearly a C effect at each level of B. Since the slopes of the line segments in Figure 5.8 have opposite signs and the absolute values are almost equal, the conditional effects are almost offsetting, resulting in a very small C main effect, which was -3.51. The conditional main effects of C are -18.20 and 11.18. The former was not much less than the B effect estimate, -20.96,

that was declared significant, although the numbers are not directly comparable since the estimators have different variances.

Nevertheless a conditional effects analysis should have been performed before declaring that the C effect was not significant. Although the results of the experiment led to changes that resulted in annual savings of more than $40,000 and management became convinced of the power of experimental design, additional work should have been performed. In particular, the magnitude of the (block + BC) interaction effect cannot be ignored if there is no reason to suspect a sizable block effect, so additional runs should be performed to unconfound these two effects. (*Note*: It is not possible to perform a conditional BC effects analysis because B and C are confounded within each block, so an interaction plot would contain only two points instead of four. Furthermore, even if this were not the problem, there would not be enough data to perform the analysis since each block has only four observations. If the experimenters concluded that there should not have been a block effect, one obvious way to perform additional runs would be to fix the level of the blocking factor and use a replicated 2^2 design for factors B and C, using enough replicates to be able to detect a significant interaction.)

If further experimentation revealed that there really is a (sizable) BC interaction, a conditional effects analysis would have to be performed for factors B and C. There is also evidence of a failure to perform cleaning of the data that were used to compute the coverage percentages used as the response values.

Example 5.3 Case Study

Another case study that involved blocking a fractional factorial was given by Ledolter (2002–2003), which has some very interesting features. As in Example 5.2, there was natural blocking as the experiment had to be carried out over several days.

The experimental setting involved a viscose fiber that was being produced and interest centered on fiber strength and fiber elongation. Ten factors were investigated and a 2_{IV}^{10-5} design was used. Of course, only main effects are estimable with this design, as two-factor interactions are confounded amongst themselves. The experiment was performed over a period of 10 days, so these are the natural blocks. Ignoring the latter could inflate the error term and cause significant effects to not be identified as such.

In addition to the 32 factorial design points, there were two centerpoints, which were run on days 5 and 10. In addition to providing some evidence of experimental error, the placement of these centerpoints could provide a check on possible out-of-control process conditions that could contaminate the results. For each of the two centerpoints, the middle value was used for nine of the factors, but three levels (not two) of the other factor (stretch) were varied. Only one of the three levels was used in the other design points, however, that being 20 percent, with the other levels at the two centerpoints being 15 and 25 percent. Although two levels of the Stretch factor were used at each design point, the levels differed as 10 percent was used in combination with either 20 or 30 percent at some design points.

As pointed out by Ledolter (2002–2003), the three observations cannot be considered to be independent because the measurements were made on the same material. Therefore, this is not true experimental error.

A decision must of course be made regarding the composition of the blocks, which determines the effects that are not estimable. The days with the centerpoint runs became blocks 5 and 10 and these blocks contained only those runs. This means that the other 32 runs were split into eight blocks, with four runs per block. The composition of the viscose was determined by three factors, and the eight combinations of levels of those factors determined the eight blocks. This means that the effects of those factors and their interactions could not be estimated if individual block effects are to be determined because all those effects have the same level within each block. One benefit from this, however, is that the two-factor interactions of the remaining seven factors all become estimable, along with the main effects, of course. This essentially has the effect of "splitting up" the two-factor interactions into a group of estimable interactions and a group of nonestimable interactions, whereas with the 2_{IV}^{10-5} design no such grouping is formed, directly or indirectly, with certain two-factor interactions estimable only if the two-factor interactions with which they are confounded do not exist.

One shortcoming of this approach is that the main effects that are likely to be real generally aren't known before the experiment is conducted; to determine this is the objective of the experiment.

The data from the experiment are given in Ledolter (2002–2003, pp. 315–316). Readers may wish to perform their own analyses of the data and compare their results with those given by Ledolter (2002–2003).

Loeppky and Sitter (2002) give additional information about blocked fractional factorial designs and discuss the analysis of data from such designs.

5.8.1.1 Blocks of Size 2

A block size of 2 is often appropriate, such as when human subjects are involved (two eyes, two hands, two feet, etc.). When this block size is used for either a full or a fractional factorial, the usual approach is to have the set of second observations for the blocks be a foldover of the set of first observations. (One type of foldover design is a mirror image of the original design, which is then used in combination with that design. Foldover designs are discussed in Section 5.9.) This permits the estimation of all main effects free of block effects, but only main effects are estimable and all the factor levels must be changed to produce the mirror image. This will clearly be a serious disadvantage if there are any hard-to-change factors.

Draper and Guttman (1997) proposed a design strategy with more runs ($(k - p) \times 2^{k-p}$ runs for a 2^{k-p} design) that allows all effects to be estimated free of block effects. For example, in the simple case of a 2^2 design, four blocks, with two observations per block, are needed to estimate the A and B main effects and the AB interaction. Note that this is twice as many observations that are needed without blocking, but this is unavoidable if blocking is necessary and all effects are to be estimated. The following statement is especially important if either A or B or both are hard-to-change factors: "One choice, for example, would be to use pairings that require a change of only one

factor level per pair." That is, only one factor-level change would be required within each block.

These designs might be cost prohibitive if $k - p$ is not small, requiring, for example, 64 runs for four factors. Consequently, these designs will be most useful when $k - p$ is small, such as in a highly fractionated factorial when k is not small. Draper and Guttman (1997) do not discuss such designs, however. They did illustrate how a 2^{4-1} design would be blocked, noting that there are 28 possible blocks of size 2 to select from, since $\binom{8}{2} = 28$ with $2^{4-1} = 8$. They noted that 12 of these are needed to estimate all effects, so 24 runs are necessary. They did not give the 12 pairs (blocks) and clearly such enumeration would be laborious, in general, since as Draper and Guttman (1997) pointed out, the 12 cannot be arbitrarily chosen. Consequently, it is almost imperative that software be used to construct the blocks.

As might be expected, however, software is not generally available for this purpose. For example, MINITAB cannot be used to block a 2^{k-p} design in such a way that the number of runs will exceed 2^{k-p}, and this is also true of JMP and Design-Expert.

5.9 FOLDOVER DESIGNS

It is often necessary to make additional experimental runs in order to resolve ambiguities. For example, assume that a 2^{6-2}_{IV} design was used and the AB interaction effect was judged significant. Since it is aliased with the CE interaction plus two four-factor interactions with the particular fraction that was used (defining relation of $I = ABCE = BCDF = ADEF$), this raises the question of whether the AB interaction or the CE interaction is a real effect, or are they both real effects.

Assume that a priori there is no reason to believe that any particular two-factor interaction is apt to be significant or not significant. Therefore, there is a need to make additional runs to dealias these two interactions, and of course the objective should be to make the minimum number of runs necessary to accomplish this.

We will return to this example near the end of this section after first discussing foldover designs in general.

A standard *foldover design* (also called a mirror image foldover design; see, e.g., http://www.itl.nist.gov/div898/handbook/pri/section3/pri3381.htm) results when a 2^{k-1} design is "folded over"; that is, all of the signs in the fraction are reversed and a new fraction of the same size is combined with the original fraction. Obviously this must produce the full 2^k design because by reversing all of the signs we create 2^{k-1} additional treatment combinations, and since there is a total of 2^k treatment combinations in the full factorial, the combination of the treatment combinations in the two half fractions must constitute the full 2^k design.

Although this should be apparent, another way to view this, which will be helpful when we consider foldovers of 2^{k-2} designs, is to recall that a 2^{k-1} design is a full factorial in $k - 1$ factors with the signs of the kth factor obtained as the product of the signs of the $k - 1$ factors in one fraction, and the negative of that product in the other fraction. For example, if we constructed a 2^{5-1}_{V} design we have $E = ABCD$ in

one fraction and $E = -ABCD$ in the other fraction. If we combine the two fractions we have the full 2^5 design.

Note also the following. If we write out a 2^3 design, we notice that the second half of the treatment combinations (when written in Yates order) are the sign reversals of the treatment combinations in the first half, with the pairing of the treatment combinations being (4, 5), (3, 6), (2, 7), and (1, 8). Of course these halves are not the usual half fractions of the 2^3 design, however, which of course are formed by constructing the 2^2 design and letting $C = -AB$ for one fraction and $C = AB$ for the other fraction. If, however, we take one fraction and reverse all the signs, that will produce the other fraction and of course this must be true regadless of how the first fraction is constructed. This same type of relationship will exist when $k > 3$.

This equivalence does not extend to smaller fractions, however, because more than one additional column is constructed. For example, consider a 2^{5-2}_{III} design. The two additional columns beyond the 2^3 design would logically be constructed as $D = AB$ and $E = AC$. Changing all of the signs in this fraction so as to create a second fraction would result in $D = -AB$ and $E = -AC$. Thus, the defining relation for the first fraction is $I = ABD = ACE = BCDE$ and for the second fraction $I = -ABD = -ACE = BCDE$. Thus, $BCDE$ is the only interaction that is common to both defining relations, so not surprisingly the defining relation for the design that results from combining the two fractions is $I = BCDE$. (In essence, we are adding the components of the defining relations with the components that have opposite signs since they add to zero; see John (1971, p. 161) for a related discussion/illustration.) Of course a 2^{5-1}_{V} design should be constructed so that the defining relation is $I = ABCDE$, so the half fraction that results from folding over the $1/4$ fraction is not a useful one as most of the two-factor interactions are confounded with other two-factor interactions, whereas two-factor interactions would be confounded with three-factor interactions if a $1/2$ fraction were constructed in the usual way.

Thus, we have seen that combining $1/4$ fractions does not produce a useful $1/2$ fraction, a result that should not be surprising since none of the words in the defining relations for the two $1/4$ fractions were of length 5, whereas the defining relation for the $1/2$ fraction consists of a single word of length 5.

Similar problems would be encountered in folding over $1/4$ fractions of larger designs, in that we would not obtain one of the $1/2$ fractions with the maximum resolution. For example, a 2^{6-2}_{IV} design could be constructed, if done so by hand or with software, by constructing the 2^4 full factorial design and then defining the two additional factors as $E = ABC$ and $F = BCD$. The defining relation is then $I = ABCE = BCDF = ADEF$. One of the other fractions has $E = ABC$ and $F = -BCD$ for a defining relation of $I = ABCE = -BCDF = -ADEF$. The defining relation will contain, in addition to other terms, the interaction that has the same sign in each fraction. Thus, $ABCE$ is part of the defining relation of the combination of the two fractions. The defining relation of the optimal half fraction is $I = ABCDEF$, so we see very quickly that this will not produce the optimal half fraction, nor would this happen if we merged any of the other five pairs of $1/4$ fractions, as the reader can easily verify in Exercise 5.1.

This is not to suggest that foldover designs aren't useful, as the purpose is to de-alias effects when at least one effect in a pair is significant. The point to be

made is simply that the design that results from a foldover will generally not have the highest resolution possible for that number of design points, and could be considerably less than the highest resolution possible, as in the preceding example. A related disadvantage of a (full) foldover, as pointed out by Wu and Hamada (2000, p. 175), is that the number of effects to de-alias is not in proportion to the size of the original experiment, so a full foldover will be quite inefficient, relative to the optimal design, when the original experiment has a moderate-to-large number of runs. Essentially the same point was made by Mee and Peralta (2000), who stated that a full foldover design following a 2^{k-p} design is degree-of-freedom inefficient.

These shortcomings in foldover designs mandate the consideration of semifoldover designs, which are covered in the next section.

5.9.1 Semifolding

Let's return to the problem posed at the beginning of Section 5.9, with a pair of two-factor interactions to be dealiased. We would like to be able to perform the dealiasing using only half as many observations as would be used in a foldover design, and this can be accomplished with the *semifolding* technique, the term being due to Barnett, Czitrom, John, and León (1997), although the general idea was apparently first proposed by Daniel (1962).

First, let's consider our options for dealiasing the *AB* interaction from the *CE* interaction. Technically, if we wanted to estimate one more effect, we need just one additional run. This approach is illustrated by Montgomery (1997, p. 553). Although this does have the advantage of economy, there is a rather heavy price that must be paid. One disadvantage is the loss of orthogonality, which is also lost when semifolding is used since the fraction that results from the use of semifolding is an irregular fractional factorial (i.e., it is not in the form 2^{k-p}). So it is a question of degree, but we don't want to have a serious departure from orthogonality. In Montgomery's example, the two interactions that he wishes to separate are very highly correlated, as the correlation between the columns in the design is .80. The only difference in the two columns occurs at the additional run and it would be very risky to try to separate them using a single run. Furthermore, with an odd number of observations, each effect estimate would be computed from unbalanced data as each effect estimate will be computed from five observations at one level and four observations at the other level.

With the semifolding technique, eight additional runs would be made rather than the 16 runs that would have been made if a foldover design had been used, assuming that a 16-run fractional factorial had originally been run. There are many possible combinations of those runs that might be selected, as discussed in detail by Mee and Peralta (2000), who provided guidance. They emphasize that it is advantageous to view semifolding as a two-step process with the first step being to select the (full) foldover fraction from which half the runs will be used. Second, a level of a factor (or interaction) is selected (e.g., the high level of factor *E*), which defines the half of the runs in the full foldover fraction that will constitute the foldover. Mee and Peralta (2000) refer to this as "subsetting" on the effect.

How are those runs selected? There are 15 estimable effects for the 2^{6-2}_{IV} design and one of the 15 would be selected to "semifold on," with the treatment combinations

selected that have the same sign relative to the selected estimable effect. Since there are two possibilities for each of the 15 estimable effects, there are thus (15)(2) = 30 possibilities for each fraction, and thus (30)(3) = 90 possibilities altogether since there are three other fractions.

Assume that when the original design was run, the principal fraction was used so that the defining relation was $I = ABCE = BCDF = ADEF$, and a semifold is to be performed using the fraction with the defining relation $I = -ABCE = BCDF = -ADEF$. Operationally, however, it isn't necessary to select the fraction that is to be used, and in fact the user interface of Design-Expert does not permit such a selection, as the user instead is asked to select a factor on which to semifold. If we select factor A to (semi)fold on, then we reverse the sign of A in each of the 16 runs. That is, we indirectly let $A = -BCE$, which algebraically is obviously equivalent to $E = -ABC$, so it is equivalent to using the generator $E = -ABC$, which leads to the desired defining relation. Thus, selecting factor A to fold on is equivalent to specifying the defining relation $I = -ABCE = BCDF = -ADEF$.

Which of the 15 effects should the fold be performed on? The position taken by Mee and Peralta (2000) is that in general it is best not to fold on a two-factor interaction, as doing so will result in fewer estimable two-factor interactions than can be achieved by folding on a main effect. A simplification of their approach, based on the assumptions of effect heredity and three-factor and higher-order interactions not being real, was given by Wang and Lee (2006).

One advantage of folding on a factor for the current example is that all two-factor interactions containing that factor are dealiased. This result holds for 2^{k-p}_{IV} designs in general; Mee and Peralta (2000) sketch a proof of this. Included in their proof is the result that subsetting on a main effect results in the same estimable functions of two-factor interactions as does any full foldover fraction that is obtained by reversing the sign of a single factor. (Of course this is not how a full (mirror image) foldover is performed, however.)

Whereas the recommendation of Mee and Peralta (2000) to fold on a main effect is generally good advice, we don't have a great need to estimate interactions when sets of confounded interactions are not significant in the original analysis. That is, we don't need to separate interactions whose sum is not significant. One problem with folding on a main effect is that we would create a major imbalance in the precision of the estimation of that particular main effect. For example, assume that the fold was performed on factor A. This means that with the eight additional runs, factor A will be at either the high level or at the low level for all of those runs, so that when these runs are combined with the original 16 runs, factor A will be at the high level 16 times or at the low level 16 times, and at the other level only 8 times. This means that the average effect at one level will be estimated with a standard error of $\sigma/4$ and with a standard error of $\sigma/2.8$ at the other level.

Of course this is undesirable but is not a serious problem, whereas some desirable properties, including orthogonality, must be relinquished regardless of how the semifolding is performed. Of course such semifolding could be justified if the factor that is folded on is not significant when the original fraction, in this case the 2^{6-2}_{IV}, is analyzed, but the general recommendation is to fold on a factor whose interactions

one wants to estimate, and usually a significant two-factor interaction will be accompanied by at least one of the main effects of those factors being significant. (Of course it was observed in Chapter 4 in the discussions of conditional effects that a large interaction effect can cause main effects to be small, so there are exceptions.)

We don't have an imbalance problem with a main effect if we fold on a two-factor interaction. If we decide to fold on the *DF* interaction, we would use the eight runs for which *DF* was either $+1$ or -1. If we select the former, the design is as given below.

Row	A	B	C	D	E	F
1	−1	−1	−1	−1	−1	−1
2	1	−1	−1	−1	1	−1
3	−1	1	−1	−1	1	1
4	1	1	−1	−1	−1	1
5	−1	−1	1	−1	1	1
6	1	−1	1	−1	−1	1
7	−1	1	1	−1	−1	−1
8	1	1	1	−1	1	−1
9	−1	−1	−1	1	−1	1
10	1	−1	−1	1	1	1
11	−1	1	−1	1	1	−1
12	1	1	−1	1	−1	−1
13	−1	−1	1	1	1	−1
14	1	−1	1	1	−1	−1
15	−1	1	1	1	−1	1
16	1	1	1	1	1	1
17	−1	1	−1	−1	−1	−1
18	1	1	−1	−1	1	−1
19	−1	−1	1	−1	−1	−1
20	1	−1	1	−1	1	−1
21	−1	1	−1	1	−1	1
22	1	1	−1	1	1	1
23	−1	−1	1	1	−1	1
24	1	−1	1	1	1	1

The design is not orthogonal as three pairs of main effect estimates are correlated (two with a correlation of .333 and one with a correlation of $-.333$), but this is part of the price that is paid when less than 16 additional design points are added to the original 16 points. The interaction effect estimates for *AB* and *CE* are also correlated with a correlation of .333. All the effects are estimated with 12 runs at the high level and 12 at the low level. That is somewhat offset, however, by the fact that folding on factor *A* results in only three of the effect estimates being correlated, instead of four when we fold on the *DF* interaction.

All things considered, there is little to choose for this design between folding on *DF* and folding on *A* if our objective is simply to de-alias the *AB* and *CE* interactions. If, however, certain pairs of confounded two-factor interactions are close to being judged significant, it would be desirable to disentangle them and take a closer look.

TABLE 5.4 Data from Barnett et al. (1997)

A	*B*	*C*	*D*	*E*	*F*	Log (etch uniformity)
−1	−1	−1	−1	−1	−1	2.4
−1	1	−1	−1	1	1	2.31
−1	−1	1	−1	1	1	2.16
−1	1	1	−1	−1	−1	2.22
−1	−1	−1	1	−1	1	1.16
−1	1	−1	1	1	−1	1.59
−1	−1	1	1	1	−1	1.76
−1	1	1	1	−1	1	1.06
1	−1	−1	−1	1	−1	1.13
1	1	−1	−1	−1	1	1.28
1	−1	1	−1	−1	1	1.28
1	1	1	−1	1	−1	2.04
1	−1	−1	1	1	1	−0.22
1	1	−1	1	−1	−1	3.71
1	−1	1	1	−1	−1	4.26
1	1	1	1	1	1	0.41
0	0	0	0	0	0	1.36
0	0	0	0	0	0	1.65

This would favor folding on a main effect since that will permit the estimation of more two-factor interactions, as discussed previously. Of course we would also want to do so if there were multiple confounded interactions that were judged to be significant, as in the following example.

Example 5.4

In the example given by Barnett et al. (1997), the authors analyzed data from an experiment in which a 2^{6-2}_{IV} design was used plus two centerpoints for a total of 18 runs, with the design and the data (on three response variables) given in their Table 17.2, and are given here, in Table 5.4, with the generators for factors E and F being $E = ABC$ and $F = BCD$.

This was a semiconductor etch uniformity experiment and the objective was to certify a new vapor phase etching process for 200-mm wafers. With data available on three response variables, the experimenters elected to use log (etch uniformity) as the response variable in their analyses. The factors were $A =$ revolutions per minute, $B =$ pre-etch total flow, $C =$ pre-etch vapor flow, $D =$ etch total flow, $E =$ etch vapor flow, and $F =$ amount of oxide etched.

There are 13 main effects and two-factor interactions that are estimable with the 2^{6-2}_{IV} design, leaving two degrees of freedom for estimating the error variance. The analysis of Barnett et al. (1997) showed six significant effects (as judged by $p < .05$), four of which were interaction effects. When the two centerpoint runs are used, however, there are only five significant effects, as the AC interaction, which is confounded

with the *BE* interaction, has a p-value of .06. The p-value for the *AC* interaction doubled when the centerpoints were added because the estimate of σ^2 was approximately four times larger with the centerpoints than without them. This difference resulted from the fact that there was a degree of freedom for curvature when the centerpoints were used and although the curvature was not significant (the p-value was .085), the curvature sum of squares was large enough to inflate the estimate of σ^2. (We may note that the estimate of σ^2 using just the pure error, with one degree of freedom, is close to what it is when all four available degrees of freedom are used for the estimation.)

Even though it isn't necessary to use Lenth's method since there are degrees of freedom available for estimating the error term, it is still of interest to see what results are obtained with this approach. Six effects are declared significant, in agreement with the results given by Barnett et al. (1997). Thus, there is agreement between two of the methods, but data are being discarded when each method is used.

Regardless of the results that one goes by, there are major problems, as for example, the significant *AE* interaction is confounded with the *BC* and *DF* interactions, although the engineer stated before the experiment that any *DF* interaction would be negligible. There were three other interactions involving factor *A* such that the sum of the interaction and the two-factor interaction that it was confounded with were significant. Thus, there were 9 two-factor interactions that were potentially significant, so semifolding should be performed in a manner so as to isolate as many of these as possible. By semifolding on factor *A*, Barnett et al. (1997) were able to isolate seven of the nine.

There is a price for doing so, however. The *BC* and *DF* interactions are still confounded, but in this example that is not a problem as long as the engineer is correct in believing that any *DF* interaction would be negligible. If the engineer's belief were incorrect, however, this would create subsequent analysis problems. A real problem is that there are correlations between some of the effects that were identified in the first stage of the analysis that are .333 in absolute value. This precludes a straightforward analysis using t-statistics and p-values. Barnett et al. (1997) attempted to circumvent this problem by using stepwise regression. The results of the stepwise regression given in their Table 17.5 have many t-statistics that are much less than 1, however, which is not the way stepwise regression is generally used.

If stepwise regression is to be performed, it would logically be done using the main effects that were identified in the first stage of the analysis as candidate factors in addition to the set of seven interactions that can now be estimated separately from the group of nine that were aliased in the first stage. Thus, the candidate predictors would be E, F, AC, BE, AD, EF, AE, BC, AF, and DE, with DF excluded since it is confounded with BC. (Note that four of these interactions involve factor *A* even though the main effect of *A* was not significant. We would, however, expect to see some effect of factor *A* when we look at conditional effects. We address this issue is the second stage of the analysis.)

Variable selection is actually not a good strategy when variables are highly correlated because the results will be very sensitive to small data perturbations, although there aren't any high correlations with these data. Since stepwise regression can easily give misleading results under certain conditions, it is advisable to use more than

one variable selection approach, and perhaps use all possible regressions. If we use the latter here we do not have a clear-cut choice for a model. Certainly a good model, though, would be one with the following five terms: F, AF, E, BC, and EF. This produces an R^2 value of 87.2 percent. This differs sharply from the model selected by Barnett et al. (1997), which oddly includes five main effects that have extremely large p-values, including one that is .8683!

One possible criticism of our model with the five terms is that it is nonhierarchical since the main effects of A, B, and C are not included. The fact that interactions involving these factors entered the model when stepwise regression was used suggests a possible strong need for a conditional effects analysis of these data, which is performed in Section 5.9.1.1.

Barnett et al. (1997) included a block term in their model that measures the difference between the first 18 runs and the 8 runs that comprised the semifold. The inclusion of a block term is of dubious value as that corresponds to using a blocking variable in the model when a randomized block design is used, with the blocking variable being an extraneous variable. Furthermore, although there was some evidence of a difference between variables with these data, a blocking variable cannot be "set" the way that a physical variable such as temperature would be set, so it is of no value in future applications of a regression model; it simply helps explain the variability in the response variable for the data that were used in developing the model.

Thus, the function of the block variable in the analysis is to remove from the error term an effect that is significant, and the block variable is selected when stepwise regression is used, with the result that R^2 increases to 90.8 percent when the block variable is included.

There are many applications of the semifoldover technique that are easily accessible. For example, a case study showing the utility of the technique given by Mark Anderson and Pat Whitcomb of Stat-Ease, Inc., can be found at http://www.statease.com/pubs/semifold.pdf, and other case studies are described by Mark Anderson at http://www.statease.com/pubs/breaddoe.pdf and at http://pubs.acs.org/subscribe/journals/tcaw/11/i02/html/02comp.html, with the latter also having appeared in print (Anderson, 2002).

5.9.1.1 Conditional Effects

Many practitioners would object to using a model with E, F, AF, BC, and EF because this is not a hierarchical model, as noted previously, but recall the discussion in Section 4.18 regarding hierarchical and nonhierarchical models. Barnett et al. (1997, pp. 241–242) address this issue, pointing out that the corresponding main effects "have little meaning" when interactions are significant. While this is certainly true, we need to look at conditional effects and in this case we need to look at the conditional main effects of A, B, and C. Since the block effect was significant, we might calculate two conditional main effects: one computed using only the first 16 runs (centerpoints have no effect on effect estimates) and one computed using all 24 runs. The second conditional main effect should not differ greatly from the former; if it does, that would be cause for concern and further investigation.

The A effect would of course be obtained by splitting the data on the F factor since AF is the only interaction involving factor A that is significant. For the first 16 runs, the conditional effects of A are -0.985 at high F and 0.7925 at low F. These are comparatively large effect estimates as they exceed in absolute value all the regular effect estimates that were judged to be significant, although we should keep in mind that these conditional effect estimates have a variance that is twice the variance of the regular effect estimates since they are based on half as many data points. Nevertheless, the magnitude of these conditional effects cannot be ignored, and we should conclude that there really is an A effect. For the 26 points the conditional effects of A are -0.6325 at high F and 0.8575 at low F. Although there isn't much difference between the two sets of conditional effects at low F, there is a considerable difference at high F.

Similarly the B and C conditional effects would be obtained by splitting the data on C to obtain the B conditional effects and on B to obtain the C conditional effects. These are also comparatively large, as we might expect. For example, for the 16 data points the B conditional effects are 1.105 at low C and -0.9325 at high C.

In determining if certain significant effects are real, an *ad hoc* approach that would probably be sufficient in many situations would be to simply compare the conditional effects with the whole effect analysis obtained using a normal probability plot, recognizing that the variances of the effect estimates differ by a factor of 2. Note that this was essentially done in analyzing the data from Daniel (1976). That is, we might say that a conditional effect is "large" if it is at least equal to the smallest whole effect that was significant.

From a modeling perspective, a (regular) regression model could not be constructed using a mixture of conditional effects and regular effects because different amounts of data are used in constructing each. Furthermore, different data (with the same number of observations) are used when the data are split on different factors, so a model could not be constructed using a standard approach even with a constant number of observations. (Recall from Example 4.4, however, that we can approach this in a somewhat indirect way and obtain the appropriate model coefficients.)

Despite problems in trying to model conditional effects in a conventional manner, it is important to compute conditional effects because they will exist when there are large interactions. This is by definition because a first-order (i.e., two-factor) interaction effect estimate is simply the difference of the conditional effect estimates for each factor divided by $2r$, with r denoting the number of replicates. This is, in essence, the definition of a two-factor interaction (see Appendix A to Chapter 4). Thus, the difference in the two conditional effect estimates is the same for each factor.

Of course a large difference does not necessarily mean that each conditional effect is large in absolute value, however, as one of the conditional effects could theoretically be zero.

In general, the computation and consideration of conditional effects is just as important in fractional factorial designs as it is in factorial designs. As another example, Myers and Montgomery (1995, p. 142) analyzed data from a 2^{4-1} design and concluded that factor B is not significant and can be dropped from consideration because the main effect estimate is only 1.50. The conditional effects were 20.5 and -17.5, however, which were not small as all the main effect estimates were less than 20.

5.9.1.2 Semifolding a 2^{k-1} Design

If we semifold a 2^{k-1} design, we will create a 3/4 fraction of the 2^k, with 3/4 fractions of 2^k designs due to John (1961, 1962) and also described in John (1971, p. 163). John used the notation, slightly altered here to accommodate the notation of this chapter, $3(2^{k-2})$ to designate a 3/4 fraction of a 2^k design, the simplest practical design of which would be the $3(2^{4-2})$ design. These designs are discussed in more detail in Section 5.10.

There are many ways to construct a 2^{4-2} design, depending upon the choice of effects in the defining relation, but once those effects are chosen, there are four fractions, defined by the four combinations of plus and minus signs preceding the effects in the defining relation.

For example, one of the four fractions could be defined as $I = AB = ACD = BCD$, with the last effect of course being the generalized interaction of the first two effects. Then, fixing *AB* and *ACD* as the effects in the defining relation (and thus *BCD* also), the other three fractions would be obtained by the other three combinations of signs for the first two effects, with the combinations being $+/-$, $-/+$, and $-/-$. One way to view the treatment combinations that comprise each fraction is to recognize that these sign combinations correspond to even/odd, even/even, odd/odd, and odd/even, as stated previously. Thus, the first fraction would consist of the treatment combinations that have an even number of letters in common with *AB* and an even number of letters in common with *ACD*. There are 16 treatment combinations altogether and it should be apparent that from those 16 the following treatment combinations comprise the even/even fraction: *ab*, *c*, *d*, and *abcd*.

Assume, for the sake of the example that we will use, that the 3/4 replicate is constructed by omitting the even/odd fraction. This gives the following treatment combinations, which are listed in the same order that they were given in the case study of Prat and Tort (1989) that will be discussed shortly.

Treatment Combination	A	B	C	D
ac	−	+	−	+
bcd	+	−	−	−
ad	−	+	+	−
b	+	−	+	+
abcd	−	−	−	−
c	+	+	−	+
ab	−	−	+	+
d	+	+	+	−
acd	−	+	−	−
bc	+	−	−	+
a	−	+	+	+
bd	+	−	+	−

We immediately observe that this is not an orthogonal design since the correlation between the first two columns is $-1/3$. This is not a small correlation, but of course factor *A* might not be significant and the five other pairs of factors are uncorrelated. If we look at two-factor interactions, however, there are nonzero correlations involving

all the other factors. For example, the correlation between the BC and D columns is also $-1/3$.

Of course a regression approach can still be used to obtain the coefficients for a fitted model, but a normal probability plot for identifying significant effects is, strictly speaking, inapplicable since the effect estimates are correlated.

Another way to view this design is to recognize that the design is equivalent to first constructing a 2^{4-1} design with $I = BCD$ and then semifolding on the AB interaction. This can be easily seen by recognizing that the first eight treatment combinations all have an odd number of letters in common with BCD and the last four treatment combinations have the same sign (minus) for the AB interaction. Also observe that these last four treatment combinations all have an even number of letters in common with BCD and thus are part of the other half fraction defined by $I = -BCD$.

Of course we always construct a 2^{k-1} design by confounding the highest-order interaction between the two fractions, so the obvious question is whether the design would be better by constructing the $1/2$ fraction in this manner and then semifolding on the AB interaction. The alias structure is actually worse when this approach is used, however, as all four of the factors are involved in nonzero correlations (A and B are correlated and C and D are correlated).

Example 5.5

The case study of Prat and Tort (1989) alluded to previously is an excellent example of some of the problems that are encountered in the statistical design of experiments (DOE). A plant experiment was run with the objective of determining how to improve the process of manufacturing a particular brand of pet food. As is generally the case with designed experiments, there was more than one response variable of interest, and observations were made on four response variables: powder in the product, powder in the process, yield, and consumption.

The experiment was preceded by lengthy discussions extending over a few weeks, which included having to convince the plant manager of the desirability of conducting the experiment. A pet food is of course a mixture, and the possibility of using a mixture design (see Section 13.12) was put off to a later date because the number of runs that such a design would require was considered unacceptably high for the first experiment.

The combination of economical considerations and engineering knowledge of the process resulted in the decision to use four factors, each at two levels. The factors were the formula, conditioning temperature, flow, and compression zone in die.

There were some constraints that served as complicating factors. One constraint was that the experiment had to be run in the real plant (i.e., not a pilot plant), so naturally another requirement was that salable product should be produced. Another requirement was that all of the experimental runs had to be performed in one day, which meant that no more than 13 runs could be made.

If we consider 2^{4-p} designs, the last requirement is a severe restriction as a 2^{4-1}_{III} design does not permit the estimation of two-factor interactions, whereas the objectives of the experiment stipulated that all two-factor interactions should be considered, and there was to be no confounding of main effects and two-factor interactions.

A 12-run Plackett–Burman design (see Section 13.4.1) would not be a possibility since this is a resolution III design and main effects are heavily confounded with two-factor interactions. Consequently, one is led to consider a $3(2^{4-2})$ design and the one that was used in the Prat and Tort (1989) study was given earlier in this section.

A major problem is that the fourth factor, compression zone in die, was hard to change. Prat and Tort presented the engineers with two possibilities, one of which was to use a split-plot design with the fourth factor defining the main plots. The other option was to use complete randomization. As a compromise, the hard-to-change (HTC) factor was changed four times during the day, whereas seven changes would have been necessary if the experiment had been run in accordance with the design layout given earlier in the section.

There is no indication of whether the particular run sequence was randomly selected from among all run sequences that require four changes of the HTC factor. There is a large number of ways in which the four changes could occur. The most obvious general arrangement of the runs would be to alternate the level in such a way that four groups of 3 have the same level. For example, the ordering that was actually used was $---+++---+++$. If we fix the treatment combinations that are to be in each group of three, there are $(3!)^4 = 1{,}296$ possible orderings of treatment combinations that have this pattern, and the same number that start with $+++$ and end with $---$, again fixing the treatment combinations in each group. Of course there are $\binom{6}{3} = 20$ ways of choosing the three for each group that have a plus sign, and the same number for a minus sign. Thus, for these two logical patterns there are $(20)(20)(2)(1296) = 1{,}136{,}800$ possible orderings of four groups of three. Randomly selecting one of these would be reasonable, but it seems unlikely that this occurred since there was no discussion in Prat and Tort (1989) of the sequence being determined randomly.

Prat and Tort (1989) proved that the design is resolution V by determining the alias structure, but we should keep in mind that this is a nonorthogonal design and correlations that are $1/3$ in absolute value are not small. Consequently, the results of a normal probability plot analysis of estimated effects should not necessarily be strictly followed since the plot is for independent effects.

Prat and Tort concluded that only the second and fourth factors "seem to have a significant effect on the mean" (for the first response variable), with no interactions standing out in their normal probability plot of effect estimates. We should note that their plot was the type of plot that was typically performed during that era, with no decision rule that selects significant effects, as in Lenth (1989), for example. When we employ the latter, however, no effects are identified as being significant. This is obviously due to the fact that the effect sparsity requirement is not met as six of the ten estimable effects have t-statistics ranging in absolute value from 1.94 to 4.53. The effect estimate of the BC interaction is exceeded in absolute value only by the effect estimate of D, so an examination of the conditional effects of B and C should be performed, and especially for factor C since it was judged to be nonsignificant.

One problem with a nonorthogonal design is that the effect estimates will not be equal to twice the regression coefficients *if* we computed each effect estimate using all

the observations. That isn't done, however, which undoubtedly seems counterintuitive. John (1971, p. 161) indicated that A is estimable from the second and third fractions (i.e., the even/even and odd/odd fractions), and *not* from using all 12 observations. This can be seen by adding the defining relations together for the three quarters of the 2^4 that are used. Doing so produces $I = -AB = -ACD = -BCD$. This means that A is confounded with B, CD, and $ABCD$. Confounding with $ABCD$ is of no concern, but confounding with B and CD is of concern. This "defining relation" for the design allows us to see which effects are partially aliased, with the correlations between the partially aliased effects through first-order interactions all being $-1/3$ because of the minus signs in the "defining relation" and because we are summing the defining relations over three fractions to obtain the minus sign.

We don't need to literally use only parts of the data for estimating each effect, however, as John (1971, p. 161) showed that the least squares estimates for the effects using all the data—and hence the effects estimates being twice those coefficients—are the same effect estimates that would be obtained by selecting the appropriate data subsets, computing the effect estimates one at a time and averaging the estimates. (See John, 1971, pp. 161–162 for an explanation of why this occurs.) Thus, the effect estimates are the "correct" ones from uncorrelated effect estimates.

Regarding which effects are judged significant, there is an important point to be made that is undoubtedly better understood today than it was when Prat and Tort (1989) analyzed these data. Specifically, the line on a normal probability plot of effect estimates must go through the point (0, 0) if these correspond to (normal score, effect estimate), or through (50 percent, 0) if the vertical axis is the normal cumulative probability scale; the line cannot be drawn in such a way as to connect or nearly connect the majority of points that are plotted. Theoretically, it is possible, although unlikely, that the majority of the effect estimates could be large and be such that they line up as almost a straight line on the plot. Then if a line were drawn through these points, we would indirectly be declaring the most significant effects not significant.

Notice that no (standardized) effect is declared significant (i.e., no point is labeled) even though one point in particular is very far from the line. (The plotted points are just t-statistics instead of effect estimates so as to make a point.) The reason for no effect being declared significant can be seen from the fact that only two points are close to the line, signifying that effect sparsity does not exist. Stated differently, only four of the t-statistics are less than 1.94 in absolute value. Thus, the approach breaks down with these data as there are not enough small effects to combine to create a pseudo-error term. Of course if we tilted the line in Figure 5.9, the line would almost pass through all but two of the points, but then the line would not come close to going through the point (50 percent, 0).

The obvious question to ask then is "Which effects are significant?" If we tried to answer this question by constructing a Pareto chart of effect estimates (as in Section 5.1.4, for example), we would have the same problem relative to the dotted line that is used for determining effect significance, as there are not enough small effect estimates to provide a reliable estimate of σ. Stated differently, the Pareto chart shown in Figure 5.10 shows no evidence of "the vital few and the trivial many." Rather the chart comes close to exhibiting almost perfect stair steps.

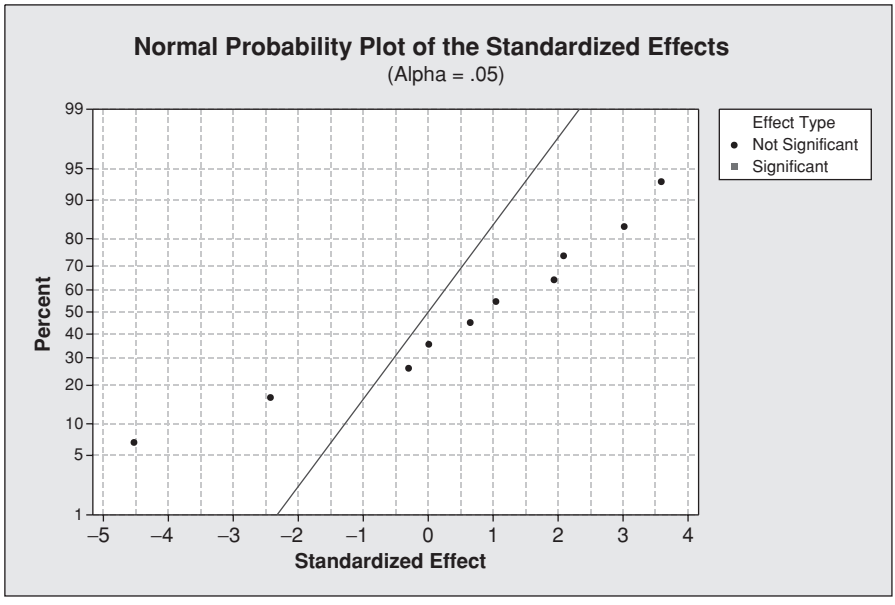

Figure 5.9 Normal plot of (standardized) effect estimates.

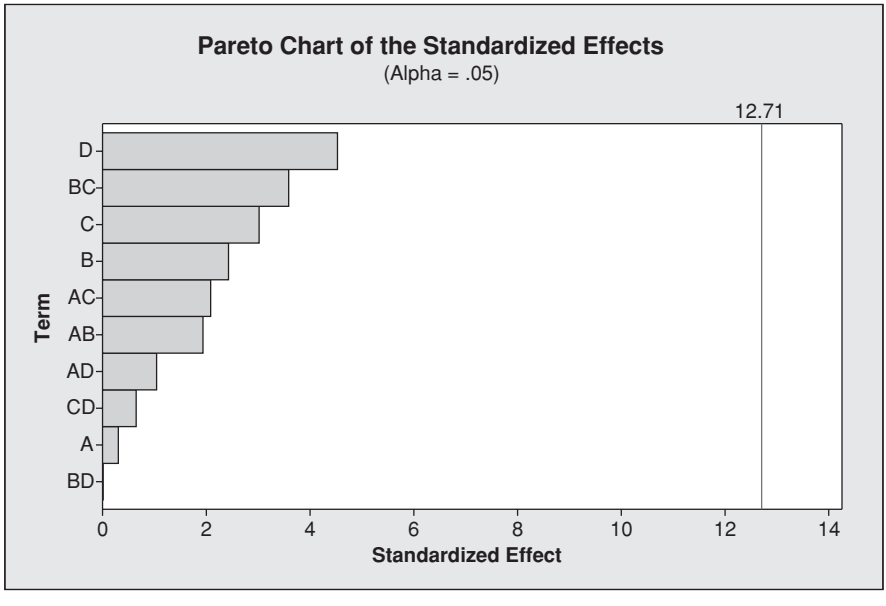

Figure 5.10 Pareto chart of the standardized effect estimates.

Furthermore, the analysis is complicated by the fact that the *BC* interaction is large and the *AC* interaction is large relative to the *C* effect, and that interaction dwarfs the *A* effect. This means that an analysis of conditional main effects should be performed, and we can see from Figure 5.10 that one of the conditional main effects of *A* will be moderately large. The *BC* interaction being larger than the *B* and *C* main effects means that there will be a conditional main effect for each of the factors *B* and *C* that will be large, and the largest conditional main effect of *C* will be slightly larger than the largest conditional main effect of factor *D* since there is no large interaction involving factor *D*.

We miss this if we look only at main effects, and indeed Prat and Tort (1989) stated "The conclusion is then that only the compression zone of the die has an important effect. . . ."

This is a difficult dataset to analyze because of the absence of effect sparsity, the presence of large interactions, and the apparent absence of randomization in the determination of the run sequence. These problems render a traditional analysis implausible and the large interactions necessitate the use of conditional main effect estimates.

Should a different design have been used? A split-plot design, with the hard-to-change factor being the main plot factor, was considered, but was not used, apparently because of concern that a lurking variable could be confounded with the hard-to-change factor. Although that could certainly happen, there is always this danger when a split-plot design is used to accommodate a hard-to-change factor, but is probably not a sufficient reason for ruling out a split-plot design unless there is strong reason to suspect a particular lurking variable.

The problems involved in this application of experimental design are by no means atypical, and illustrate that the statistical DOE and the analysis of the resultant data are by no means routine.

5.9.1.3 General Strategy?

It is inadvisable to try to formulate a general strategy for semifolding because so many factors must be considered. Clearly the advantage of folding on a factor instead of an interaction may be slight at best if suspected interactions do not involve a single factor but rather are spread over the factors. The level of the factor that is used in selecting the semifold should generally be the "best" level in the sense that this will aid in the maximization or minimization, whichever is desired, of the response and allow the variance of the response at that level to be minimized.

5.9.1.4 Semifolding with Software

Although semifolding is reasonably simple and straightforward, manual semifolding would be tedious and impractical for a large number of factors, and we would also like to avoid it for even a moderate number of factors.

Design-Expert 7.0 has considerable semifolding capability, which can be employed with a Plackett–Burman design as well as with fractional factorial designs. Indeed, semifolding a Plackett–Burman design will result if, for example, the user specifies a minimum run resolution IV design and that design turns out to be a Plackett–Burman

design. If a resolution V design has been constructed, semifolding is not an option and only a full foldover can be used.

Semifolding can also be used with a resolution III design, which of course confounds main effects with two-factor interactions. Such designs are generally used as screening designs; so one possible use of a semifoldover would be to increase the resolution of the design so that main effects can be de-aliased from two-factor interactions. This may not be easily done, however. For example, Mee and Peralta (2000) state that only one of the 31 foldover fractions of the 2_{III}^{9-5} design used by Hsieh and Goodwin (1986) would increase the resolution.

5.10 JOHN'S 3/4 DESIGNS

Designs that are a 3/4 fraction of a 2^k design were mentioned briefly in Section 5.9.1.2. These designs, which were proposed by John (1961, 1962) and also presented in John (1971), deserve consideration because they are efficient designs in that the available degrees of freedom is not much greater than the number of main effects and two-factor interactions to be investigated. John's work was motivated by a practical problem. As described in John (2003), he had a client for whom he recommended a 2^3 design, but when the experimental material arrived, there was enough material only for six runs. In another application, an experiment had to be planned for four factors, each at two levels, but there was enough raw material only for 12 runs. So the 3/4 fraction was born. Because these are nonorthogonal designs, the degree of correlation between the effect estimates must be considered.

Design-Expert constructs a 3/4 fraction when the Irregular Fraction option is selected and the number of factors is 4, 5, or 6. For 7 factors the fraction is 3/8; for 8 factors the fraction is 3/16; and for 9, 10, and 11 factors the fraction is 1/8, 1/16, and 3/64, respectively.

The 3/4 fraction of a 2^4 design that the software produces is given below.

A	B	C	D
1.00	1.00	−1.00	−1.00
1.00	−1.00	−1.00	1.00
1.00	1.00	1.00	−1.00
−1.00	−1.00	1.00	−1.00
1.00	1.00	−1.00	1.00
−1.00	−1.00	−1.00	−1.00
−1.00	1.00	−1.00	1.00
−1.00	1.00	1.00	1.00
−1.00	−1.00	−1.00	1.00
−1.00	1.00	1.00	−1.00
1.00	−1.00	1.00	1.00
1.00	−1.00	1.00	−1.00

Design-Expert gives the alias structure for this design, which is given below.

```
              4 Factors: A, B, C, D

              Design Matrix Evaluation for Factorial 2FI Model

     Factorial Effects Aliases
[Est. Terms] Aliased Terms
     [Intercept]= Intercept - 0.5 * ABC - 0.5 * ABD
     [A] = A - ACD
     [B] = B - BCD
     [C] = C
     [D] = D
     [AB] = AB
     [AC] = AC - BCD
     [AD] = AD - BCD
     [BC] = BC - ACD
     [BD] = BD - ACD
     [CD] = CD - 0.5 * ABC - 0.5 * ABD
```

Of course with 12 runs we can estimate 11 effects. Design-Expert lists 10 effects on the left, with the *ABCD* interaction not listed in the alias set. This is because interactions beyond second order are not displayed with the default, so the feature that displays them would have to be selected.

This is a reasonable design because two of the main effects are unaliased and the two that are aliased are each aliased with a three-factor interaction. Similarly, one two-factor interaction is unaliased and the others that are aliased are each aliased with a three-factor interaction. It may look odd that the intercept is part of the alias structure. This occurs because the columns for *ABC* and *ABD* do not sum to zero, as the reader can observe by constructing them or by simply performing the appropriate multiplication and summation with the design matrix above. In general, the effects that are aliased with the intercept are those whose column does not sum to zero.

Mee (2004) recommends the use of John's $(3/4)2^{8-2}$ design for $k = 8$ (48 points) and the $(3/4)2^{11-4}$ design for $k = 11$ (96 points). (Note that this notation differs from the notation used by John, 1971.) Both of these designs can be generated using Design-Expert.

The alias structure for the $(3/4)2^{8-2}$ design given by Design-Expert is listed below.

```
              8 Factors: A, B, C, D, E, F, G, H

              Design Matrix Evaluation for Factorial 2FI Model

     Factorial Effects Aliases
[Est. Terms]  Aliased Terms
     [Intercept] = Intercept - 0.333 * ACF - 0.333 * BEG -
0.333 * CEH - 0.333 * DFG
     [A] = A - 0.5 * BDH - 0.5 * CDG - 0.5 * EFH
     [B] = B - 0.5 * ADH - 0.5 * CGH - 0.5 * DEF
     [C] = C - ADG
```

```
[D] = D - 0.5 * ACG - 0.5 * BEF
[E] = E - BDF
[F] = F - AEH
[G] = G - BCH
[H] = H - 0.5 * AEF - 0.5 * BCG
[AB] = AB - 0.5 * AEG - 0.5 * BCF + CDE + FGH
[AC] = AC - AEH + BDE
[AD] = AD + BCE - CDF - EGH
[AE] = AE - 0.5 * ACH + BCD - 0.5 * CEF - 0.5 * DGH
[AF] = AF - ADG + BGH
[AG] = AG - 0.5 * ADF + BFH - 0.5 * CFG - 0.5 * DEH
[AH] = AH + BFG - CFH - DEG
[BC] = BC + ADE - 0.5 * BEH - 0.5 * CEG - 0.5 * DFH
[BD] = BD + ACE - CFH - DEG
[BE] = BE + ACD - BCH
[BF] = BF + AGH - 0.5 * BDG - 0.5 * CDH - 0.5 * EFG
[BG] = BG + AFH - BDF
[BH] = BH + AFG - CDF - EGH
[CD] = CD + ABE - 0.5 * ADF - 0.5 * CFG - 0.5 * DEH
[CE] = CE + ABD - 0.5 * AEF - 0.5 * BCG
[CF] = CF - 0.5 * BDH - 0.5 * CDG - 0.5 * EFH
[CG] = CG - CDF - EGH
[CH] = CH - BDF
[DE] = DE + ABC - 0.5 * BDG - 0.5 * CDH - 0.5 * EFG
[DF] = DF - BCH
[DG] = DG - AEH
[DH] = DH - 0.5 * AEG - 0.5 * BCF
[EF] = EF - CFH - DEG
[EG] = EG - 0.5 * ADH - 0.5 * CGH - 0.5 * DEF
[EH] = EH - ADG
[FG] = FG + ABH - 0.5 * ACG - 0.5 * BEF
[FH] = FH + ABG - 0.5 * ACH - 0.5 * CEF - 0.5 * DGH
[GH] = GH + ABF - 0.5 * BEH - 0.5 * CEG - 0.5 * DFH
```

We observe that all the main effects are aliased with two-factor interactions, and the two-factor interactions are aliased with three-factor interactions. Despite the complicated alias structure, the design is resolution V and obviously has good properties. (We may note that the 2^{8-2} design is also resolution V, so there is no loss in resolution in using 3/4 of this number of design points.)

The $(3/4)2^{11-4}$ design with 96 points, which is described in detail in the next section, has a similar alias structure in the sense that main effects and two-factor interactions are aliased with three-factor interactions.

If appropriate software is unavailable, there are various ways in which these designs can be constructed, one way being the semifolding technique as was discussed in Section 5.9.1.2. The 96-point design for $k = 11$ recommended by Mee (2004) is obviously not a 3/4 fraction of a 2^{11} design, as such a design would have over 1000 points. The obvious question then is of what design is this a 3/4 fraction? John (1971)

describes how the design is constructed, stating that it is obtained by thinking of the four quarters of a 2^7 design and dropping the quarter defined by $I = AB = CDEFG = ABCDEFG$, with those treatment combinations eliminated by John (1969) for which $AB = CDEFG = -1$. (Note of course that this is not the way that a 2^{7-2} design would be constructed as the main effects of A and B would be confounded with this design.) Then four additional factors are added to the 3/4 of a 2^7 design by defining them as $H = ABDEF$, $J = ABCEG$, $K = AEFG$, and $L = BCDE$.

How good is this design, remembering that the design is nonorthogonal? One obvious problem is that A and B will not be orthogonal because of the way that the design was constructed. The dot product of A and B is -32 because 32 points have been excluded, for which the dot product is $+32$. Since there are 96 design points and the sum of all columns is zero, it follows that the correlation between the A and B columns is $-1/3$. The correlation between all other pairs of main effect estimates is zero. Thus, although John (1971) describes this as a resolution V design, it is a nonorthogonal resolution V design. The complete alias sets for this design are given in John (1969).

The obvious question to ask at this point is could the partial aliasing of main effects be avoided by, for example, omitting the quarter defined by $I = ABC = DEFG = ABCDEFG$ and defining the four additional factors the same as before? Although this might seem to be an improvement, Mee (2004) concluded that when the partial aliasing of all effects is considered, the design given by John (1969) is the best possible design.

The notation used by Mee (2004) is preferable to the notation used by John since $(3/4)2^{11-4}$ clearly indicates that this is 3/4 of a 2^{11-4} design, which is 3/64 of a 2^{11} design. Furthermore, it makes no difference whether we first construct a 2^{11-4} and then delete a specified 1/4 of the points, or following the implied steps given by John (1971, p. 167), delete 1/4 of the 2^7 and then construct the four additional factors.

Readers interested in learning more about the various types of foldover designs are referred to the sources cited previously, in addition to Li and Mee (2002), Li and Lin (2003), and Li, Lin, and Ye (2003). The first two papers are concerned with optimal foldover plans for 2^{k-p} designs with optimality determined by the minimum aberration criterion (see Section 5.7) for the combined designs. The third paper is concerned with optimal foldover plans for nonregular designs, including 12-run and 20-run designs. The authors showed that the full foldover is the optimal design for all 12-run and 20-run orthogonal designs. Other papers on foldover designs and their properties that may be of interest include Ye and Li (2003), Montgomery and Runger (1996), and Miller and Sitter (2001).

5.11 PROJECTIVE PROPERTIES OF 2^{k-p} DESIGNS

Assume that a 2^{6-2}_{IV} design has been used and it was determined that three of the six factors seem to be important. The design contains a replicated full factorial in those factors and thus has the advantage of a replicated full factorial over a fractional factorial. Of course we know that the design is also a full factorial in the first four

factors (A–D), as those factors were used to form the full factorial from which the two additional columns were generated. But it is not a full factorial in any other set of four factors, whereas it is a full factorial in *any* set of three factors.

The general result, which is due to Box and Hunter (1961a,b), is that a 2^{k-p} design of resolution R contains a full factorial in $R - 1$ factors. This results applies almost trivially to a 2^{3-1} resolution III design since it is constructed from a 2^2 design, which is virtually of no practical value, but the real applicability of the result is for designs of resolution IV and V.

This is an important result because all the factors that are examined in a design are almost certainly not going to be significant. Of course there is no guarantee that the number of significant factors will turn out to be one less than the resolution of the design, but there is a good chance of this happening when the number of factors is small to moderate. At the other extreme, assume that a 2_{V}^{5-1} design has been used and only three factors seem to be important. The design in those factors is then a replicated 2^3 design. This can be seen from the following 2_{V}^{5-1} design. It is easily seen that this is a 2^3 design in factors B, C, and D with two replicates, but note that this is also true for each of the other nine 2^3 designs that could be constructed from these points.

A	B	C	D	E
−1	−1	−1	−1	1
1	−1	−1	−1	−1
−1	1	−1	−1	−1
1	1	−1	−1	1
−1	−1	1	−1	−1
1	−1	1	−1	1
−1	1	1	−1	1
1	1	1	−1	−1
−1	−1	−1	1	−1
1	−1	−1	1	1
−1	1	−1	1	1
1	1	−1	1	−1
−1	−1	1	1	1
1	−1	1	1	−1
−1	1	1	1	−1
1	1	1	1	1

5.12 SMALL FRACTIONS AND IRREGULAR DESIGNS

As stated in Section 5.3, Snee (1985) considered the 2^{5-1}, 2^{6-2}, 2^{7-3}, and 2^{8-4} designs for the study of five, six, seven, and eight factors, respectively, to be the most useful fractional factorial designs, and provided an application of 2^{5-1} and 2^{7-3} designs. There is clearly a need for fractions smaller than $1/2$ and $1/4$ fractions when the number of factors under consideration is at least moderate. For example, a 2^{8-1} design has 127 degrees of freedom available for estimating eight main effects, 28 two-factor interactions, and if desired, 56 three-factor interactions. The latter are

seldom significant, but even if we estimated all of the three-factor interactions (3fi's), we would still have 35 degrees of freedom that are unused and thus could be used to estimate the standard deviation of the error term.

This is wasteful of degrees of freedom, so something smaller than a $1/2$ fraction would be a better choice. In general, when a $(1/p)$th fraction of a 2^k design is used, the defining relation will contain p effects and all generalized interactions of those effects. We can use the fact that the sum of binomial coefficients in the binomial expansion is 2^p to arrive at the number of generalized interactions. That number is obtained as $\binom{p}{2} + \binom{p}{3} + \cdots \binom{p}{p}$, which equals $2^p - p - 1$. Therefore, the total number of effects in the defining relation is $(2^p - p - 1) + p = 2^p - 1$. Obviously this number grows rapidly with p, so the alias structure is quite involved when p is at least moderate, but that complexity is a small price to pay for the savings in time and cost that are realized by using a small fraction.

We should be ever mindful of the minimum detectable effects with 16-point designs, as that minimum might be unacceptable. For example, for a 2^{5-1} design, if $\alpha = .05$ and the power is .90, the minimum detectable effect is 2.5σ, assuming that at least three degrees of freedom are used to estimate σ. This result is from Table 3 of Lynch (1993), from which it is apparent that 32-run 2^{k-p} designs have much better detection properties. A quick summary of power for various numbers of factors and numbers of runs is given in Figure 2 of a Stat-Teaser newsletter from Stat-Ease, Inc., that is available at http://www.statease.com/news/news0409.pdf.

Oftentimes we will prefer to use irregular fractions but we still must consider the size of the design relative to the sizes of effects that we would like to detect. For example, assume that there are eight factors to investigate and we want to be able to estimate all two-factor interactions (2fi's). A 2_{V}^{8-2} design could be constructed, and although this would have good detection properties, 64 runs would be used to estimate eight main effects and 28 2fi's, thus leaving 27 unused degrees of freedom. When we think about powers of 2, we have 2, 4, 8, 16, 32, 64, 128, ..., with the gaps between the numbers becoming greater as we move up the line. Here we need at least 36 degrees of freedom, but we don't need 63 degrees of freedom.

We lose orthogonality when we use irregular fractions such as $3/4$ fractions, but if that loss is small, it can be more than offset by the gain in cost by using far fewer experimental runs.

As an admittedly extreme example, but one that did occur in practice, Mee (2004) refers to a ballistic missile simulation experiment requiring the estimation of all main effects and two-factor interactions for 47 factors. Obviously a design with a very large number of runs is required, which is generally not a major problem with computer/simulation experiments, as there are 1,128 effects to be estimated. If we tried to bracket this number with powers of 2, and thus provide a sufficient number of degrees of freedom, we would need 2,048 runs. Thus, several hundred additional runs would be made beyond what is necessary if a 2^{k-p} design were used.

There are various efficiency measures for judging the efficiency of a design, which are based upon a particular model assumption. As discussed in Section 5.5, following Daniel (1956), Mee (2004) discussed df-efficiency for assessing the efficiency of a design, under the assumption that the true model contains only main effects and 2fi's

in the factors that are used in the design. Of course the true model is never known, and in particular, all of these effects won't be significant, but this is a useful measure because it gives an upper bound of the efficiency relative to the effects that are judged significant.

Using the expression for df-efficiency given in Section 5.5, a two-level resolution V design for estimating the 1,363 effects with 2,048 runs that was alluded to earlier would have a df-efficiency of

$$\begin{aligned}\text{df-efficiency} &= \frac{\text{Number of effects to be estimated}}{\text{Total number of degrees of freedom}} \\ &= \frac{0.5(k)(k+1)}{n-1} \\ &= 1128/2047 \\ &= .551\end{aligned}$$

for the 2^{47-36} design. If the df-efficiency were .50, this would mean that the power of 2 in 2^{k-p} was 1 more than necessary, so we can think of this as a lower bound on df-efficiency, with designs that have a df-efficiency value close to this being highly df inefficient.

The researchers initially used a 2^{47-38}_{IV} design and then followed that with 17 groups of runs so as to separate the aliased 2fi's. Due to the complexity of such a sequence of runs for separating the aliased 2fi's, the investigators later ran, under a different scenario, a 2^{47-35}_{V} design, even though this has 4,096 design points. Since each run took 30 minutes, an extra 2,000 runs, say, would require an additional 1,000 hours! Therefore, a design with fewer design points would obviously be highly desirable.

As shown in Hedayat, Sloane, and Stufken (1999), it is possible to construct a resolution V design with $n = 2048$ and 47 factors, although the design would have to be constructed using nonlinear error-correcting codes. This approach for constructing the 47-factor design is also explained in the appendix of Mee (2004). Since the method is rather involved, it will not be explained here; the interested reader is referred to one of these sources for details. Mee (2004) recommended the use of either this design or a nonorthogonal subset of the design.

Benefits in the form of df-efficiency can also be realized for other numbers of factors by using an irregular design. For example, designs that are smaller than the regular resolution V designs can be constructed for various values of k for $n = 128$, 256, 2048, and 4096 design points, as described in Hedayat et al. (1999).

5.13 AN EXAMPLE OF SEQUENTIAL EXPERIMENTATION

As an example of the sequential use of designs, we will examine the three sequential experiments described by Eibl, Kess, and Pukelsheim (1992). The response variable was the coating thickness resulting from a painting operation, and the objective was to identify the factors that influenced the value of the response variable and to determine

the settings of those factors so as to achieve a target value of 0.8 mm. (The thickness had been varying between 2.0 mm and 2.5 mm.)

Interestingly, the first experiment was a *replicated* fractional factorial, with the design consisting of four replications of a 2_{III}^{6-3} design. The reason given for replicating the design was that prior data had indicated that the standard deviation of the error term might be large. Recall from Eq. (1.1) in Section 1.4.3 that the necessary sample size to detect a difference in factor levels of a specified amount Δ is a function of σ, so a large value of σ can dictate the use of a large number of observations for each factor level.

Randomization was not used in the first experiment because the machine operators felt that it would be too time-consuming. They were subsequently convinced that this was not the case, however, so randomization was used in the second and third experiments. This did not include, however, resetting the factor levels after each run, as was done in the first experiment. Paint viscosity was a hard-to-change factor and was not varied randomly in the first experiment. Instead, the low level was used for the first 16 runs and the high level used for the last 16 runs. Those sets of runs were made one week apart but there was no analysis of a possible time effect or apparent consideration of any lurking variables that might be confounded with a possible time effect.

All four observations (replicates) of one treatment combination were much higher than any of the other observations, suggesting the possibility of a lurking variable or at least an interaction effect resulting from the specific combination of factor levels.

Only main effects are estimable with the design that was used in the first experiment, and four of the six main effects were identified as being significant. The target value for the response variable was 0.8 mm, a number that was exceeded by 30 of the 32 observations. Consequently, the experimental region was changed for the second experiment.

The design used for the second experiment was a foldover of the first experiment, except that only four factors were used instead of six, and the levels of three of the four factors were adjusted to force the response values in the direction of the target value. Accordingly, the factor levels in the second experiment were stated relative to the levels of the first experiment (e.g., 0 and -2 rather than the customary -1 and $+1$). Two replications were used since 16 runs could be made in one day, with the first eight runs and the last eight runs each being randomized.

The first factor, belt speed, was adjusted in the direction opposite of that suggested by the coefficient of the factor from the first experiment Based on their knowledge of the belt apparatus, the authors felt that they should reduce the belt speed in the second experiment, rather than increase it. Their conjecture was not supported by the data from the second experiment, however.

The coefficient of the first factor was, oddly, the same in the fitted model from the second experiment as that in the fitted model from the first experiment. The other coefficients differed noticeably from those in the first fitted model, although the data from the first experiment were added to the data from the second experiment in obtaining the fitted model.

The choice of levels for the second experiment turned out to be not much better than the choice for the first experiment as 14 of the 16 observations were above the target value.

For the third experiment, paint viscosity was fixed at its low level, a seemingly reasonable decision since the two observations below the target value occurred at that level, with most observations well above the target value. Therefore, a 2^3 design was used with two replications, so the number of observations was again 16. The factor levels were changed again with an eye toward driving the response values toward the target value. The experimenters overcorrected somewhat, however, because 15 of the observations were below the target value, although 5 were between 0.74 and 0.79.

The coefficients in the final fitted model, using all 64 observations, differed very little from the coefficients in the second fitted model. From the final fitted model, which had an R^2 value of .93, the experimenters determined factor-level combinations that produced a predicted response value equal to the target value and indicated two such combinations that had small standard deviations of the predicted response value.

5.13.1 Critique of Example

The use of three experiments performed in sequence with information obtained from one experiment used in designing the following experiment is highly commendable. There are, however, some important points that should be made. It is questionable how much influence the data from the first two experiments should have on the coefficients of the fitted model since the factor levels were such that many of the response values were more than twice the target value. This is especially problematic since the first experiment had 32 observations, some of which were almost three times the target value, whereas the third experiment had only 16 observations.

The nonorthogonality of the design that results when the three sets of design values are pieced together should also be of concern since the correlations given below are not small enough for the nonorthogonality issue to be dismissed.

```
Correlations:  A, B, C, D

          A          B          C
B      -0.187
C      -0.187      0.342
D      -0.086      0.281      0.281
```

Strictly speaking, the nonzero correlations mean that the coefficients in the fitted model are not directly interpretable. Consequently it is perhaps not surprising that one of the two sets of factor-level combinations recommended by the authors is outside the experimental region because the suggested level of the first factor is outside the range of levels used in the 64 runs. The other recommended combination was not one of the 24 combinations used in the experiment, and although each factor level was within its range over the 24 combinations, that does not mean that the set of combinations is within the experimental region for a nonorthogonal design. (In general, there is no exact way to determine if design points are inside or outside the experimental region for such designs when there are more than two factors.)

The experimenters found the *BC* interaction to be significant in the third experiment, but analyzing the data using ANOVA, since the design is replicated, does not

show the *BC* interaction to be significant. The analysis does show, however, that the *A* effect is not significant. Furthermore, since factor *D* was not varied in the third experiment, its coefficient and significance or nonsignificance cannot be compared with the results obtained using all the data.

Thus, there are some major concerns about the conclusions that were drawn from the experiment. In particular, questions must be raised about the wisdom of combining data from experiments in which the response values were not close to the target value. Furthermore, there are problems in trying to use the numerical values of the coefficients in determining optimum operating conditions when there is more than slight nonorthogonality that results from combining the factor levels for the three experiments. Finally, extrapolating beyond the experimental region can be risky and it can be difficult to determine when this has occurred with a nonorthogonal design.

Despite these problems, the authors reported that the information obtained from the experiments resulted in the target value of 0.8 mm being "achieved" with a standard deviation of 0.1 mm. There is no discussion of whether or not 0.8 is the average of the post-experimentation thickness values, although the authors may have assumed that their wording implies this. If 0.8 was the approximate average, then the experimentation was successful in leading to the distribution of thickness values being centered at the target value, which was clearly a vast improvement over the range of 2.0–2.5 mm for the thickness that existed before the experimentation.

We might prefer the post-experimentation standard deviation to be somewhat smaller, and perhaps the standard deviation could be reduced by seeking to identify potentially influential factors that were not included in the experimentation.

The results from this experiment were quite useful in helping to achieve the desired target value when a new paint was introduced. (The standard deviation of the thickness values for the new paint was not given.)

5.14 INADVERTENT NONORTHOGONALITY—CASE STUDY

Nonorthogonal designs, such as John's 3/4 fractions, can be quite useful. Users must be careful not to inadvertently create nonorthogonality, however. This can easily happen, as was illustrated by Ryan (2003). In a lead recovery experiment performed by the NIST in 2001, the objective was to identify the factors that influence the extent of lead recovery from paint. There were 112 paint specimens available for the experiment and 7 specimens could be handled by the sonicator on each run. Since $112/7 = 16$ and five factors were investigated, it follows that a 2^{5-1} would be one possible choice for the design.

The design that was run, however, was one for which the 2^5 design was split into two sets of 16 design points, with one set replicated three times and the other set replicated four times. Despite this imbalance between replicates, there is no problem with orthogonality as long as the 2^5 design is split into two 2^{5-1} designs. That was not what was done, however, so there was a small departure from orthogonality, but it was not enough to cause any major problem with the analysis of the data. This

does illustrate, however, that orthogonality can be "lost" through design if care is not exercised, and the loss could be consequential.

5.15 FRACTIONAL FACTORIAL DESIGNS FOR NATURAL SUBSETS OF FACTORS

In recent years there has been interest in developing designs in recognition of the fact that many experiments are used for factors that would be logically grouped into subsets. Yates and Mee (2000) proposed two methods of constructing such designs, which would have application in assembly experiments, in particular. Bisgaard (1997) also proposed an approach for constructing such designs for tolerancing assembly products.

For example, several factors might be identified that probably affect one part of a machine assembly and different factors might be identified that likely affect a second part of the assembly. It should be apparent that constructing a design for subsets of factors will restrict the number of feasible treatment combinations. For example, if there are 10 factors in two subsets of 5 factors each, no treatment combination can contain more than five letters. The method proposed by Bisgaard (1997) stipulates that a design be selected for each subset of factors, and then the product array of the designs is formed, analogous to Taguchi's idea of a product array in robust parameter design (see Section 8.4). Since a product array generally results in an unnecessarily large design (see, e.g., Ryan, 2000, p. 447), it is desirable to use only some portion of the treatment combinations in the product array. Bisgaard (1997) referred to this process as *post-fractionation* and the process of obtaining the fraction for each subset as *pre-fractionation*. Yates and Mee (2000) provided tables of designs for 16 runs and 32 runs that gave the pre- and post-fractionation for the number of factors ranging from 6 to 15 and up to five subsets. Each design is either a resolution III design or a resolution IV design. Although the designs actually aren't listed, they are designated by the plan number in the catalog of designs given by Chen et al. (1993).

For example, their Table 1 shows that if eight factors are grouped into two subsets of four factors, a 2^{4-1} design would be used for each design, and then a $1/4$ fraction of the product array would be constructed, for a total of 16 runs. The design is resolution III. If a resolution IV design were desired, a $1/2$ fraction of the product array would be constructed, giving 32 runs. (Of course this is the same number of design points in a 2^{8-3}_{IV} design when there is no restriction on treatment combinations.)

To illustrate the construction of these designs, assume that there are three factors of interest for each of two stages of an assembly operation and a 2^6 design would result in two many runs. A 2^{3-1} design could be used in each stage. This would require 16 runs, which would generally not be an excessive number. Therefore, assume that there is no post-fractionation. With A–C denoting the factors in the first stage and D–F the factors in the second stage, the first fraction would logically be constructed using $I = ABC$ and the second fraction would be constructed using $I = DEF$. The product array, which would be a 2^{6-2}, would then have a defining relation obtained

from the product of the defining relations for each fraction. That is $(I + ABC)(I + DEF) = I + ABC + DEF + ABCDEF$. Notice that this is a resolution III design whereas it is possible to construct a 2^{6-2}_{IV} design. Thus, there is a loss of resolution but that is unavoidable because the treatment combinations that would be used when a 2^{6-2}_{IV} design is employed are not available when only three factors are varied in each experiment. This is not to suggest that inefficiency results when fractions are used in stages, as the post-fractionization will generally result in a saving.

To illustrate the latter, assume that there are four factors in each of two stages with a 2^{4-1} design used in each stage, with the resultant 64-run product array considered to have an unsatisfactorily high number of runs so that post–fractionation will be necessary. With the letters A–D denoting the first subset of factors and E–H denoting the second subset, the 2^{8-2} product array has 64 runs, so a 1/4 fraction of this is needed to reduce the design to 16 runs. Before the post-fractionation, the defining relation would be $(I + ABCD)(I + EFGH) = I + ABCD + EFGH + ABCDEFGH$. This is obviously a resolution IV design, which is the highest resolution possible. We need a 1/4 fraction of this design, so the result will be a 2^{8-4} design, with resolution IV being possible for a single-stage 2^{8-4} design. We will see if that resolution is possible for the two-stage design. We cannot use $I = ABCD = EFGH = ABCDEFGH$ as the defining relation for the 1/4 fraction since the treatment combinations have already been split on $ABCD$ and $EFGH$, but after some trial and error we see that we can construct a defining relation $I + ACEF + BCEH + ABFH$ such that a three-factor interaction is not involved in the product of this with $I + ABCD + EFGH + ABCDEFGH$. Thus, the full defining relation written in "alphabetical order" is $(I + ABCD + EFGH + ABCDEFGH)(I + ACEF + BCEH + ABFH) = I + ABCD + ABEG + ABFH + ACEF + ACGH + ADEH + ADFG + BCEH + BCFG + BDEF + BDGH + CDEG + CDFH + EFGH + ABCDEFGH$. (We can see that the product should have 15 interactions because it has to be one less than 4×4 since $I \times I = I$, with 4 being the number of components in each defining relation.)

The treatment combinations for the design are given in Table 5.5.

Looking at the treatment combinations and remembering that factors A–D will be studied in the first stage and then factors E–H in the second stage, we see that there are eight distinct treatment combinations for each set of four factors, which of course is the way that it should be, since a 2^{4-1} design was used for each set in the pre-fractionation stage. We can also note that for each treatment combination in either set, the pair of corresponding treatment combinations in the other set (i.e., in the same row of the overall design) form a foldover pair (e.g., note the first two rows of the design).

Before the post-fractionation there are of course eight treatment combinations in the second group for each treatment combination in the first group since it is a product array. Then a 1/4 fraction is taken so it stands to reason that there should be two treatment combinations in the second group for each treatment combination in the first group after the fractionation.

These designs would be extremely laborious to construct by hand and they apparently cannot be constructed with widely available software, so practitioners should rely on the tables of Yates and Mee (2000) and the designs of Chen et al. (1993) as

TABLE 5.5 $(2^{4-1} \times 2^{4-1})/2^2$ Design

Row	A	B	C	D	E	F	G	H
1	−1	−1	−1	−1	−1	−1	−1	−1
2	1	1	1	1	−1	−1	−1	−1
3	−1	1	1	−1	1	−1	−1	1
4	1	−1	−1	1	1	−1	−1	1
5	1	1	−1	−1	−1	1	−1	1
6	−1	−1	1	1	−1	1	−1	1
7	1	−1	1	−1	1	1	−1	−1
8	−1	1	−1	1	1	1	−1	−1
9	1	−1	1	−1	−1	−1	1	1
10	−1	1	−1	1	−1	−1	1	1
11	1	1	−1	−1	1	−1	1	−1
12	−1	−1	1	1	1	−1	1	−1
13	−1	1	1	−1	−1	1	1	−1
14	1	−1	−1	1	−1	1	1	−1
15	−1	−1	−1	−1	1	1	1	1
16	1	1	1	1	1	1	1	1

much as possible, as well as the tables of Block and Mee (2005), who enumerated all 128-run resolution IV designs.

5.16 RELATIONSHIP BETWEEN FRACTIONAL FACTORIALS AND LATIN SQUARES

Latin square designs were covered in Section 3.3. There are relationships between Latin square designs and fractional factorial designs that are well known. For example, Kempthorne (1973, pp. 279–283) showed that a replicated 2^3 design can be arranged in two or three replicated 4×4 Latin squares. These cannot be constructed in such a way that an entire replicate is blocked, however, so such designs might be best viewed as being appropriate for field experiments with the need to block on soil fertility in each direction.

Similarly, Wu and Hamada (2000, p. 280) discuss the relationship between 5^{k-p} designs and 5×5 Latin squares and their variations. (Fractional factorial designs with more than two levels are discussed in Chapter 6.)

A less-known relationship between fractional factorials and Latin squares was given by Copeland and Nelson (2000), who showed that a standard $t \times t$ Latin square with $t = 2^k$ corresponds to a 2^{3k-k} fractional factorial. For example, a standard 4×4 Latin square corresponds to a 2^{6-2} fractional factorial. Such a relationship is by no means obvious since a Latin square is for one factor of interest whereas the 2^{6-2} design is obviously for six factors.

Copeland and Nelson (2000) stated that a Latin square configuration is easier for a nonstatistically oriented experimenter to understand, and obviously that is the case

since fractional factorials are not intuitive. Apart from simplicity, however, there is an obvious question of what can be gained by knowing this relationship.

This raises the question of what would happen if we tried to analyze a Latin square as a fractional factorial, rather than going in the other direction and simply using a Latin square to represent a fractional factorial. That won't work in general, as Copeland and Nelson (2000) point out that there is only one such standard Latin square of each order that has this correspondence.

This creates somewhat of a problem relative to randomization as ideally a Latin square would be selected at random.

5.17 ALTERNATIVES TO FRACTIONAL FACTORIALS

People in many different fields have used statistically designed experiments and many people outside the statistics profession have written about design and attempted to contribute to the state of the art, sometimes with poor results.

Over the years there have been various types of experimental designs presented in books and journal articles that have poor properties compared to fractional factorials and other well-accepted designs, but unfortunately these alternative designs have gained acceptance in industry to a considerable extent, simply due to the industrial stature of the developers of these designs.

Included in this group of designs are those credited to Dorian Shainin. Undoubtedly many people in industry have used Shainin's variable search technique, which was recommended by Shainin when there are more than four factors. These are tiny designs (relative to the number of factors), and if they were fractional factorials they would be called highly fractionated designs. The methodology is not based on sound principles, however, and in their study of the designs, Ledolter and Swersey (1997) concluded, "We find that there is little reason to abandon the traditional and well-studied fractional factorial designs in favor of the Variables Search design strategy." More generally, Nelson (2004) stated, "The inclusion of Shainin requires a rather broad definition of DOE," with "DOE" representing design of experiments. See also the discussion of Shainin's methods in Woodall (2000).

5.17.1 Designs Attributed to Genichi Taguchi

The statistical designs of Genichi Taguchi have been much debated, as have the statistical methods that he proposed. Ryan (1988) was apparently the first to point out that some of Taguchi's designs correspond to suboptimal fractional factorials (suboptimal in the sense that the resolution of the design is not optimized). Only a few of the designs that have been attributed to Taguchi were actually invented by Taguchi, and some misinformed writers have even stated that Taguchi invented design of experiments!

The designs advocated by Taguchi are commonly known as orthogonal arrays, some of which correspond to fractional factorials. As stated at the beginning of this chapter, all fractional factorials are orthogonal arrays but not all orthogonal arrays

are fractional factorials. Taguchi's experimental design contributions are discussed in detail in Chapter 8 in this book and also in Chapter 14 of Ryan (2000).

One highly beneficial result of Taguchi's foray into experimental design has been that he has sparked a considerable amount of research to develop better methods. The most striking of these developments is that his work has motivated researchers to work on designs for identifying factors that influence variability rather than just the mean, and on robust designs that allow the inclusion of noise factors. (These are not model-robust designs, but rather the objective is to use designs that enable the identification of factor levels such that performance of what is being measured will not vary greatly under variations in the levels of noise factors that are uncontrollable during regular production.) These designs and design issues are discussed in detail in Chapter 8.

5.18 MISSING AND BAD DATA

Missing data in full factorial designs were discussed in Section 4.12 and the effects of bad data were illustrated in Section 4.10 and discussed further in Sections 4.12.1 and 4.12.2. Missing data and bad data will have the same deleterious effects in fractional factorial designs: loss of orthogonality and can cause erroneous decisions to be reached, as was illustrated for a full factorial design in Section 4.10.

Srivastava, Gupta, and Dey (1991) examined the robustness of resolution III fractional factorial designs to one missing observation, in addition to doing this for other designs. (Their definition of *robustness* is that a design is robust if all parameters are still estimable for a given model in the presence of a missing observation.) They found that the consequence of a missing observation varied, depending on which observation was missing (see also Srivastava, Gupta, and Dey, 1990).

5.19 PLACKETT–BURMAN DESIGNS

It should be noted that a Plackett–Burman design, which is a two-level design, is a useful alternative to a 2^{k-p} design, especially when there is a desire to use something between 16 and 32 design points, which of course cannot be done with 2^{k-p} designs. Within this interval, Plackett–Burman designs can be constructed for 20 and 24 design points, as well as for larger and smaller numbers of points. The designs are discussed in detail in Section 13.4.1. (They are not discussed in this chapter in part due to the length of the chapter.)

5.20 SOFTWARE

Virtually all statistical software packages have the capability for fractional factorial designs, including SAS, MINITAB, JMP, D. o. E. Fusion Pro, and Design-Expert. The latter has a nice interface that shows, in particular, a grid of the 2^{k-p} designs that are available with bright color shading to indicate the resolution of each design and

the grid constructed using the number of factors and the number of design points. MINITAB has a similarly informative interface that shows the same information and additionally shows the full factorial designs that are available, with the available Plackett–Burman designs (covered in Section 13.4.1) shown on the same display. JMP shows available designs once the user indicates the number of continuous and categorical factors that are to be used in the design.

Books on experimental design also give tables of 2^{k-p} designs, including the design generators. Robinson (2003) pointed out that there is a lot of redundant information in those tables, however, and provided the same information in a more concise manner in his Table 2.

Not all statistical software packages have the capability for all the design methods presented in the chapter. For example, Design-Expert 7.0 has the capability for semifolding, but MINITAB Release 14 does not, nor does most statistical software, in general. (It should be noted, however, that the former does not permit the foldover to be selected by doing the foldover on an interaction, although the instances in which there will be a need for doing so seem to be rather limited, following the discussion in Mee and Peralta, 2000.)

Design-Expert also has the capability for John's 3/4 designs discussed in Section 5.10 and given in John (1961), including the two designs recommended by Mee (2004). The irregular designs that can be produced by Design-Expert are those given by Addelman (1961). These designs cannot be produced using either MINITAB or JMP. Specifically, JMP gives many options for a screening design with 11 factors at two levels, but none has 96 runs, which is the number of runs for a $(3/4)2^{11-4}$ design. JMP also cannot produce a design for 11 two-level factors with 96 runs when its Custom design option is used.

Design-Expert 7.0 can produce minimum run resolution IV designs for 5-50 factors, and resolution V designs for estimating main effects and two-factor interactions for 6-30 factors. Some care must be exercised in using these designs, however, because they are not orthogonal designs and thus are not fractional factorials.

For example, the minimum run resolution IV design for five factors has pairwise correlations of $-.33$ and .33 for 4 of the 10 pairwise correlations. These are not small correlations. (For six factors, in 14 runs, all of the correlations are $-.143$. For seven factors in 16 runs all of the correlations are 0 because this is a 2_{IV}^{7-3} design with defining relation $I = ABDF = ABEG = ACDG$. For eight factors in 18 runs, the correlations are all $-.111$ and .111, etc.)

The pairwise correlations for the minimum run resolution V designs are generally lower because the minimum number of runs is higher. For example, the design for six factors has 22 runs and the correlations between the columns for the factors are .091 and $-.091$. These of course are small correlations, however. The correlation structure for the design for seven factors in 30 runs is disturbing, however, because five of the correlations are either .2 or $-.2$ and one correlation is $-.33$. For eight factors in 38 runs, three of the correlations are .158 in absolute value; the others are .053 in absolute value. For nine factors in 46 runs, the correlations are lower, and so on.

Although numerical output is both useful and standard, and Design-Expert has a unique feature of permitting the selection of backward elimination, a variable selection

technique, to arrive at the terms to use in the model as an alternative to a normal probability plot, graphical methods can often be more informative. For example, a Pareto chart of effect estimates, as in Figure 5.3, is more informative than having the same information displayed numerically without the effect estimates being ordered by magnitude. That graph was produced using MINITAB; a Pareto chart of effect estimates can also be produced with Design-Expert.

One unique feature of Design-Expert is that it does a Box–Cox transformation analysis on the response variable, providing a confidence interval on the power transformation parameter, λ, and on the basis of that result, making a recommendation as to whether or not a transformation is needed. This is part of the expert systems software flavor that Design-Expert possesses, and it may be the only commercial software that warns users of the poor effect detection capability of eight-run fractional factorials, for example. (The warning is also given for the 2^3 design, as was stated in Chapter 4.) Similarly, it may be the only commercial software (at least among well-known software) that gives advice to the user (provided that the Annotated ANOVA option is selected), as was illustrated in Section 4.18. Given below is advice when a 2^{5-2} design was constructed:

> Aliases are calculated based on your response selection, taking into account missing datapoints, if necessary. Watch for aliases among terms you need to estimate.

Then, when hypothetical data were analyzed for that design, the following messages resulted:

> Proceed to Diagnostic Plots (the next icon in progression). Be sure to look at the:
>
> **1)** Normal probability plot of the studentized residuals to check for normality of residuals.
> **2)** Studentized residuals versus predicted values to check for constant error.
> **3)** Externally Studentized Residuals to look for outliers, i.e., influential values.
> **4)** Box-Cox plot for power transformations.
>
> If all the model statistics and diagnostic plots are OK, finish up with the Model Graphs icon.

D. o. E. Fusion Pro also gives advice and has more of an expert systems flavor than Design-Expert in the sense that users are guided toward an appropriate path leading to design selection based on the answers they provide to questions that they are asked when the Design Navigator Wizard option is followed.

The extensive comparison study of leading software with experimental design capability by Reece (2003) was mentioned in Section 2.3 and also in Section 4.18. It is mentioned again here because the software in the study is rated on its capability regarding full factorial and fractional factorial designs, as well as foldover designs and design augmentation, in general. Regarding the latter, Design-Expert and JMP received the highest possible score on design augmentation among the 11 software packages in the comparison study, in addition to D. o. E. Fusion Pro. The aforementioned features of

Design-Expert in addition to various types of graphs, including diagnostic plots with some unusual features that can be produced dynamically, seem to justify its rating.

The software packages that received the highest rating in the Reece (2003) study in the category of "Factorial amd Fractional Factorial Designs" were, in alphabetical order, D. o. E. Fusion Pro, Design-Expert, JMP, MINITAB, Statgraphics Plus, and Statistica.

5.21 SUMMARY

Just as factorial designs are useful in a wide variety of application fields, so are fractional factorial designs. Although we associate their use with manufacturing, industry, and the physical sciences in general, they have also been used extensively in the social sciences, with examples given by Stolle, Robbennolt, Patry, and Penrod (2002).

The primary advantage of fractional factorial designs is they permit effects that may be significant to be estimated with a reasonable number of runs. However, since fractional factorial designs that are used most often in practice have a small number of runs, such as 16, it is important to consider the minimum detectable difference in factor level means for a selected design before the design is used. It is also important to use designs that are degrees-of-freedom efficient so that degrees of freedom are not wasted.

Resolution III designs are frequently used in the first stage of experimentation for the purpose of identifying factors that seem important. Such an approach will work well, provided that interactions are small. A resolution IV or resolution V design might be used in the second stage. If the former is used, a foldover design might be needed to de-alias certain two-factor interactions. There are various types of foldover designs, including a mirror image foldover. Semifoldover designs, as discussed in Section 5.9.1, are useful for de-aliasing two-factor interactions, although it appears as though virtually no statistical software or DOE software has full semifoldover capability.

Although Lenth's (1989) normal probability plot method will generally be a good choice, it won't always be the best choice, especially if a ratio of real effects to total estimable effects exceeds 0.4. Therefore, it would be useful if the user had at least a rough idea of the number of real effects to anticipate.

Missing and/or bad data can occur when any design is used. Although imputing missing data simplifies computation, the use of artificial data would be unsettling to many experimenters.

Even when there are sufficient resources and time to conduct a full factorial with at least a moderate number of runs, a fractional factorial would be the preferred choice if there was reason to believe that problems might ensue, such as mechanical failures, that would prevent all of the runs in the full factorial from being performed. If there were no problems, the additional runs that would comprise the full factorial could be made (such as folding over a half fraction to produce the full factorial). If problems ensue during the experiment, however, such that, for example, only half of the planned runs could be made, orthogonality will almost certainly be lost and the

"design" for the runs that are actually made may be very poor. Recall the discussion in Section 1.4.1 and the quote from Peter John regarding an experience early in his career when he was working as an industrial statistician. In that example, a 2^4 design was planned but only eight runs could be made because of machinery failure. Thus, half of the data were "missing." If the runs had been ordered such that the first eight runs constituted a half fraction, a meaningful analysis could be performed. If there were no problems after the first eight runs, the other eight runs might then be made, or perhaps, as John suggested, four of the runs could be made to form a 3/4 fraction, as discussed in Section 5.10.

Conditional effects analyses are just as important when fractional factorials are used as when full factorials are used, as large interactions can also render main effect estimates meaningless for fractional factorial designs. The situation can be more complicated, as large interactions may be confounded with main effects that are estimated, with the interactions not estimated and thus not detected.

Finally, although the use of two-level fractional factorial designs does predominate in industry, if there are very many factors, there will generally be at least one or two factors for which there will be interest in three or more levels. Then mixed factorials will have to be considered, and these designs are covered in Section 6.4, with mixed fractional factorial designs covered in Section 6.5. (Note that many designs that are referred to as "Taguchi designs" are mixed factorials, as discussed in Chapter 8.)

REFERENCES

Addelman, S. (1961). Irregular fractions of the 2^n factorial experiments. *Technometrics*, **3**, 479–496.

Anderson, M. (2002). The knead for speed. *Today's Chemist at Work*, **11**(2), 21, 23–24.

Barnett, J., V. Czitrom, P. W. M. John, and R. V. León (1997). Using fewer wafers to resolve confounding in screening experiments. In *Statistical Case Studies for Industrial Process Improvement*, Chap. 17 (V. Czitrom and P. D. Spagon, eds.). Philadelphia: Society for Industrial and Applied Mathematics.

Berk, K. N. and R. R. Picard (1991). Significance tests for saturated orthogonal arrays. *Journal of Quality Technology*, **23**, 79–89.

Bisgaard, S. (1993). A method for identifying defining contrasts for 2^{k-p} experiments. *Journal of Quality Technology*, **25**(1), 28–35.

Bisgaard, S. (1994). Blocking generators for small 2^{k-p} designs. *Journal of Quality Technology*, **26**(4), 288–296.

Bisgaard, S. (1997). Designing experiments for tolerancing assembled products. *Technometrics*, **39**, 142–152.

Block, R. and R. W. Mee (2005). Resolution IV designs with 128 runs. *Journal of Quality Technology*, **37**(4), 282–293.

Box, G. E. P. and J. S. Hunter (1961a). The 2^{k-p} fractional factorial designs, Part I. *Technometrics*, **3**, 311–351.

Box, G. E. P. and J. S. Hunter (1961b). The 2^{k-p} fractional factorial designs, Part II. *Technometrics*, **3**, 449–458.

Box, G. E. P., W. G. Hunter, and J. S. Hunter (1978). *Statistics for Experimenters*. New York: Wiley.

Box, G. E. P. and K. B. Wilson (1951). On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society, Series B*, **13**, 1–45.

Butler, N. A. (2005). Classification of efficient two-level fractional factorial designs of Resolution IV or more. *Journal of Statistical Planning and Inference*, **131**, 145–159.

Carroll, R. J. and D. B. H. Cline (1988). An asymptotic theory for weighted least squares with weights estimated by replication. *Biometrika*, **75**, 35–43.

Chen, J., D. X. Sun, and C. F. J. Wu (1993). A catalog of two-level and three-level fractional factorial designs with small runs. *International Statistical Review*, **61**, 131–145.

Chokshi, D. (2000). Design of experiments in rocket engine fabrication. In *Proceedings of the 54th Annual Quality Congress*, American Society for Quality, Milwaukee, WI. (http://www.asq.org/members/news/aqc/54_2000/14037.pdf)

Cochran, W. G. and G. M. Cox (1958). *Experimental Designs*. New York: Wiley.

Copeland, K. A. F. and P. R. Nelson (2000). Latin squares and two-level fractional factorial designs. *Journal of Quality Technology*, **32**(4), 432–439.

Daniel, C. (1956). Fractional replication in industrial research. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 8, pp. 87–98.

Daniel, C. (1959). Use of half normal plots in interpreting factorial two-level experiments. *Technometrics*, **1**, 311–341.

Daniel, C. (1962). Sequences of fractional replicates in the 2^{p-q} series. *Journal of the American Statistical Association*, **67**, 403–429.

Daniel, C. (1976). *Applications of Statistics to Industrial Experimentation*. New York: Wiley.

Dong, F. (1993). On the identification of active contrasts in unreplicated fractional factorials. *Statistica Sinica*, **3**, 209–217.

Draper. N. R. and I. Guttman (1997). Two-level factorial and fractional factorial designs in blocks of size two. *Journal of Quality Technology*, **29**(1), 71–75.

Eibl, S., V. Kess, and F. Pukelsheim (1992). Achieving a target value for a manufacturing process: A case study. *Journal of Quality Technology*, **24**, 22–26.

Finney, D. J. (1945). Fractional replication of factorial arrangements. *Annals of Eugenics*, **12**, 291–301.

Fries, A. and W. G. Hunter (1980). Minimum aberration 2^{k-p} designs. *Technometrics*, **22**, 601–608.

Giesbrecht, F. G. and M. L. Gumpertz (2004). *Planning, Construction and Statistical Analysis of Comparative Experiments*. Hoboken, NJ: Wiley.

Haaland, P. D. and M. A. O'Connell (1995). Inference for contrast-saturated fractional factorials. *Technometrics*, **37**, 82–93.

Hamada, M. and N. Balakrishnan (1998). Analyzing unreplicated factorial experiments: A review with some new proposals. *Statistica Sinica*, **8**(1), 31–35.

Hedayat, A. S., N. J. A. Sloane, and J. Stufken (1999). *Orthogonal Arrays*. New York: Springer.

Hill, W. J. and R. A. Wiles (1975). Plant experimentation (PLEX). *Journal of Quality Technology*, **7**, 115–122.

Hsieh, P. and W. Goodwin (1986). Sheet molded compound process improvement. In *Fourth Symposium on Taguchi Methods*. Dearborn, MI: American Supplier Institute, pp.13–21.

Jacroux, M. (2003). A modified minimum aberration criterion for selecting regular 2^{m-k} fractional factorial designs. *Journal of Statistical Planning and Inference*, **126**(1), 325–336.

John, P. W. M. (1961). Three-quarter replicates of 2^4 and 2^5 designs. *Biometrics*, **17**, 319–321.

John, P. W. M. (1962). Three-quarter replicates of 2^n designs. *Biometrics*, **18**, 172–184.

John, P. W. M. (1969). Some non-orthogonal fractions of 2^n designs. *Journal of the Royal Statistical Society*, Series B, **31**, 270–275.

John, P. W. M. (1971). *Statistical Design and Analysis of Experiments*. New York: Macmillan.

John, P. W. M. (2003). Plenary presentation at *the 2003 Quality and Productivity Research Conference*, IBM T. J. Watson Research Center, Yorktown Heights, NY, May 21–23, 2003. (The talk is available at http://www.research.ibm.com/stat/qprc/papers/Peter_John.pdf.)

Kempthorne, O. (1973). *Design and Analysis of Experiments*. New York: Robert E. Krieger Publishing Co. (copyright held by John Wiley & Sons, Inc.)

Ledolter, J. (2002–2003). A case study in design of experiments: Improving the manufacture of viscose fiber. *Quality Engineering*, **15**(2), 311–322.

Ledolter, J. and A. Swersey (1997). Dorian Shainin's variables search procedure: A critical assessment. *Journal of Quality Technology*, **29**(3), 237–247.

Lenth, R. V. (1989). Quick and easy analysis of unreplicated factorials. *Technometrics*, **31**, 469–473.

Li, W. and D. K. J. Lin (2003). Optimal foldover plans for two-level fractional factorial designs. *Technometrics*, **45**(2), 141–149.

Li, W., D. K. J. Lin, and K. Q. Ye (2003). Optimal foldover plans for two-level nonregular orthogonal designs. *Technometrics*, **45**(4), 347–351.

Li, W. and R. W. Mee (2002). Better foldover fractions for Resolution III 2^{k-p} designs. *Technometrics*, **44**(3), 278–283.

Loeppky, J. L. and R. R. Sitter (2002). Analyzing unreplicated blocked or split-plot fractional factorial designs. *Journal of Quality Technology*, **34**(3), 229–243.

Loh, W. Y. (1992). Identification of active contrasts in unreplicated factorial experiments. *Computational Statistics and Data Analysis*, **14**, 135–148.

Lynch, R. O. (1993). Minimum detectable effects for 2^{k-p} experimental plans. *Journal of Quality Technology*, **25**(1), 12–17.

Margolin, B. H. (1967). Systematic methods of analyzing 2^m3^n factorial experiments with applications. *Technometrics*, **9**, 245–260.

Mee, R. W. (2004). Efficient two-level designs for estimating all main effects and two-factor interactions. *Journal of Quality Technology*, **36**, 400–412.

Mee, R. W. and R. M. Block (2006). Constructing regular 2^{k-p} designs by bounding length of alias chains. Technical Report No. 2006-1, Department of Statistics, Operations, and Management Science, University of Tennessee. (http://stat.bus.utk.edu/techrpts)

Mee, R. W. and M. Peralta (2000). Semifolding 2^{k-p} designs. *Technometrics*, **42**, 122–134.

Miller, A. (2005). The analysis of unreplicated factorial experiments using all possible comparisons. *Technometrics*, **47**(1), 51–63.

Miller, A. and R. R. Sitter (2001). Using the folded-over 12-run Plackett–Burman design to consider interactions. *Technometrics*, **43**(1), 44–55.

Montgomery, D. C. (1997). *Design of Experiments*, 4th ed. New York: Wiley.

Montgomery, D. C. and G. C. Runger (1996). Foldovers of 2^{k-p} Resolution IV experimental designs. *Journal of Quality Technology*, **28**, 446–450.

Myers, R. H. and D. C. Montgomery (1995). *Response Surface Methodology: Process and Product Optimization using Designed Experiments* (2nd ed., 2002). New York: Wiley.

Nelson, L. S. (2004). *Review of Quality Control*, 7th ed. by D. H. Besterfield. *Journal of Quality Technology*, **36**(2), 241–242.

Prat, A. and X. Tort (1989). Case study: Experimental design in a pet food manufacturing company. Report No. 37. Center for Quality Improvement, University of Wisconsin-Madison.

Reece, J. E. (2003). Software to support manufacturing systems. In *Handbook of Statistics*, Vol. 22, Chap. 9 (R. Khattree and C. R. Rao, eds.). Amsterdam: Elsevier Science B. V.

Robinson, L. W. (2003). Concise experimental designs. *Quality Engineering*, **15**(3), 403–406.

Ryan, T. P. (1988). Taguchi's approach to experimental design: Some concerns. *Quality Progress*, **21**, 34–36 (May).

Ryan, T. P. (2000). *Statistical Methods for Quality Improvement*, 2nd ed. New York: Wiley.

Ryan T. P. (2003). *Analysis of Lead Recovery Data*. Gaithersburg, MD: Statistical Engineering Division, National Institute of Standards and Technology. (Case study available at http://www.itl.nist.gov/div898/casestud/casest3f.pdf)

Schneider, H., W. J. Kasperski, and L. Weissfeld (1993). Finding significant effects for unreplicated fractional factorials using the *n* smallest contrasts. *Journal of Quality Technology*, **25**, 18–27.

Snee, R. D. (1985). Experimenting with a large number of variables. In *Experiments in Industry: Design, Analysis, and Interpretation of Results*, pp. 25–35 (R. D. Snee, L. B. Hare, and J. R. Trout, eds.). Milwaukee, WI: Quality Press.

Srivastava, R., V. K. Gupta, and A. Dey (1990). Robustness of some designs against missing observations. *Communications in Statistics, Theory and Methods*, **19**(1), 121–126.

Srivastava, R., V. K. Gupta, and A. Dey (1991). Robustness of some designs against missing data. *Journal of Applied Statistics*, **18**(3), 313–318.

Steinberg, D. M. and W. G. Hunter (1984). Experimental design: Review and Comment. *Technometrics*, **26**, 71–97 (discussion: pp. 98–130).

Stolle, D. P., J. K. Robbennolt, M. Patry, and S. D. Penrod (2002). Fractional factorial designs for legal psychology. *Behavioral Sciences and the Law*, **20**(1–2), 5–17.

Wang, P. C. and C. Lee (2006). Strategies for semi-folding fractional factorial designs. *Quality and Reliability Engineering International*. **22**(3), 265–273.

Woodall, W. H. (2000). Controversies and contradictions in statistical process control. *Journal of Quality Technology*, **42**, 341–350; discussion: 351–378.

Wu, C. F. J. and M. Hamada (2000). *Experiments: Planning, Analysis, and Parameter Design Optimization*. New York: Wiley.

Yates, P. and R. W. Mee (2000). Fractional factorial designs that restrict the number of treatment combinations for factor subsets. *Quality and Reliability Engineering International*, **16**, 343–354.

Ye, K. Q. and W. Li (2003). Some properties of blocked and unblocked foldovers of 2^{k-p} designs. *Statistica Sinica*, **13**(2), 403–408.

Zahn, D. A. (1975). Modifications of and revised critical values for the half-normal plot. *Technometrics*, **17**, 189–200.

EXERCISES

5.1 Show that no combination of pairs of 2^{6-2} designs produces a 2^{6-1} design with the maximum resolution.

5.2 Explain why the variances of effect estimates in 2^{k-p} designs are equal regardless of the number of observations that are missing.

5.3 Show that the alias structure of the 2^{3-1} design in Section 5.1 can be obtained using Yates' algorithm in the manner described in that section by using zeros for the treatment combinations that are not in the fraction that is used.

5.4 There are 16 treatment combinations used when a 2^{5-1} design is run with $I = ABCDE$. Which one of the following treatment combinations is used in error with the other four (correct) treatment combinations: *ab*, *bcde*, *acd*, *bc*, and *abce*?

5.5 Explain why a 2^{5-1} design should not be run with *ABDE* confounded with the difference between the two fractions. What would you do instead?

5.6 An example of sequential experimentation was given in Section 5.9. A follow-up experiment is often necessary to dealias effects, as was discussed in Section 5.7 and subsequent sections. Snee (1985, references) described an experiment conducted to determine the factors affecting viscosity measurements in an analytical laboratory, as the variation in those measurements was thought to be too great. The chemists felt that five factors should be used in the experiment, but were less enthusiastic about two other factors that were under consideration. Snee stated, "They agreed to the seven variable test when it was pointed out to them that these two variables (spindle and protective lid) would not involve any additional runs beyond the sixteen required to test the first five variables." That is, a 2^{7-3} design was used whereas the chemists might have preferred a 2^{5-1} design. Do you believe that the former is a suitable substitute for the latter? What would you say if there was a strong a priori belief that the two additional factors would not likely be significant?

When the experiment was run, spindle had the largest effect (!). There was also a large interaction component (which represented $X_1X_4 + X_3X_6 + X_5X_7$). Since X_1 and X_7 had very small effects, it was believed that X_1X_4 and X_5X_7 might not be significant. Do you agree with that assessment, especially when viewed in the context of conditional effects? Accordingly, a 2^2 design was run using only X_3 and X_6 in an effort to assess the magnitude of the X_3X_6 interaction. A moderate interaction was observed and the predicted response values of the four additional runs were close to the predicted values from the model based on the first 16 runs. The interaction graph was interpreted in terms

of the best combination of levels of X_3 and X_6 and this is where the analysis ended. Would you have proceeded further? Explain.

5.7 Assume that a 2^{4-1} design has been run using the treatment combinations (1), *ab*, *ac*, *bc*, *ad*, *bd*, *cd*, and *abcd*.

(a) What was the defining contrast?

(b) Was it necessary to use the methodology given in Section 5.4 in answering the question? Why, or why not?

(c) What would you recommend if the treatment combination *bc* is an impossible combination of factor levels?

5.8 Assume that four factors, each at two levels, are studied with the design points given by the following treatment combinations: (1), *ab*, *bc*, *abd*, *acd*, *bcd*, *d*, and *ac*.

(a) What is the defining contrast?

(b) Could the design be improved using the same number of design points?

(c) In particular, which three main effects are confounded with two-factor interactions?

5.9 Assume that a 2^{5-2} design with three replications has been used.

(a) What is the smallest possible value of the F-statistic for testing the significance of the A effect and when will that occur? What is the largest possible value?

(b) Construct an example with six observations at each level that will produce this minimum value.

(c) Fractional factorial designs are generally not replicated. Assume that you have pointed this out to the experimenter, who responds that he was not willing to trust the result of a normal probability plot analysis because he expected more than a few effects to be significant. Is this a reasonable reason for using a replicated fractional factorial? Explain.

5.10 A design with six factors was run and the alias structure involving factor E is $E = ABCDF$. How many design points were there (assuming no replication)?

5.11 Critique the following statement: "I don't understand why textbooks give alias structures such as $AB = DE$. The effects are almost certainly not equal, and no hypothesis test of their equality is being performed. Therefore, I don't see any point in listing these effects as being equal and believe that this is misleading. Instead, this should be written as $AB + DE$ since that is what is being estimated, and similarly for other aliased effects."

5.12 Given below is the alias structure for a fractional factorial design with seven factors, each at two levels.

Alias Structure

$I + ABCE + ABFG + ACDG + ADEF + BCDF + BDEG + CEFG$

$A + BCE + BFG + CDG + DEF + ABCDF + ABDEG + ACEFG$
$B + ACE + AFG + CDF + DEG + ABCDG + ABDEF + BCEFG$
$C + ABE + ADG + BDF + EFG + ABCFG + ACDEF + BCDEG$
$D + ACG + AEF + BCF + BEG + ABCDE + ABDFG + CDEFG$
$E + ABC + ADF + BDG + CFG + ABEFG + ACDEG + BCDEF$
$F + ABG + ADE + BCD + CEG + ABCEF + ACDFG + BDEFG$
$G + ABF + ACD + BDE + CEF + ABCEG + ADEFG + BCDFG$
$AB + CE + FG + ACDF + ADEG + BCDG + BDEF + ABCEFG$
$AC + BE + DG + ABDF + AEFG + BCFG + CDEF + ABCDEG$
$AD + CG + EF + ABCF + ABEG + BCDE + BDFG + ACDEFG$
$AE + BC + DF + ABDG + ACFG + BEFG + CDEG + ABCDEF$
$AF + BG + DE + ABCD + ACEG + BCEF + CDFG + ABDEFG$
$AG + BF + CD + ABDE + ACEF + BCEG + DEFG + ABCDFG$
$BD + CF + EG + ABCG + ABEF + ACDE + ADFG + BCDEFG$
$ABD + ACF + AEG + BCG + BEF + CDE + DFG + ABCDEFG$

(a) What is the order of the fraction that was used?
(b) How would you explain to a manager who knows very little about experimental design what this alias structure means in practical terms?

5.13 Consider a 2^{4-1} design constructed so as to maximize the resolution. If the AB interaction is very close to zero, what will be the relationship between the conditional effects of factor D when the data are split on factor C?

5.14 Explain the condition(s) under which the selection of a one-half fraction of a full factorial design should not be chosen randomly.

5.15 Explain why a minimum aberration design isn't necessarily the best design to use.

5.16 What is the resolution of any fractional factorial design that has k factors and $k + 1$ design points?

5.17 Consider a 2^{8-4} design.
(a) How many words are in the defining relation?
(b) What is the maximum resolution of the design?
(c) Explain how a novice might erroneously construct the design such that the resolution is one less than the maximum?
(d) Explain to that person why it is desirable to construct a design so that it has maximum resolution. Are there any conditions under which you would recommend the use of a design that did not have maximum resolution?

5.18 Assume that an engineer constructed the following design, with "−" and "+" used in place of −1 and +1 for the sake of brevity.

−	−	−	+	−
+	−	−	−	−
−	+	−	−	+
+	+	−	+	+
−	−	+	−	+
+	−	+	+	+
−	+	+	+	−
+	+	+	−	−

What is the name of the design, in the form 2^{k-p}? Would you have constructed the design differently? If so, give the design.

5.19 What are the dangers of using highly fractionated designs?

5.20 Explain how you would construct a 2^{6-2} design to be run in two blocks.

5.21 Would you analyze with a normal probability plot the data from an experiment that used a 2^{5-2} design? Why, or why not? If not, how would you analyze the data?

5.22 For the data given in Example 5.2, construct the *AC* interaction plot using factor *A* on the horizontal axis. Does this plot suggest a weaker or stronger interaction than the interaction plot given in Figure 5.7? Explain.

5.23 Assume that an experimenter decides to use all four main effects in a model for the data in Example 5.2. The R^2 value is 85.9 percent for that model. Would you recommend the use of R^2 for that scenario, considering the number of data points and the number of terms in the model? Explain. Would you recommend that all four of those terms be used in the model? Why, or why not?

5.24 In his paper "Establishing optimum process levels of suspending agents for a suspension product" (*Quality Engineering*, **10**(2), 347–350), 1997–1998) A. Gupta used a design that was stated as "the L_{16} orthogonal-array design." (This is Taguchi-type terminology; see Chapter 8.) The design for five factors, each at two levels, is given below in the original units.

A	B	C	D	E
8	50	0.2	0.4	Usual
8	50	0.4	0.4	Modified
8	60	0.4	0.4	Usual
8	60	0.2	0.4	Modified
16	50	0.4	0.4	Usual

16	50	0.2	0.4	Modified
16	60	0.2	0.4	Usual
16	60	0.4	0.4	Modified
8	50	0.2	0.6	Usual
8	50	0.4	0.6	Modified
8	60	0.4	0.6	Usual
8	60	0.2	0.6	Modified
16	50	0.4	0.6	Usual
16	50	0.2	0.6	Modified
16	60	0.2	0.6	Usual
16	60	0.4	0.6	Modified

Since there are 16 runs, this must be "some" 2^{5-1} design. What is the resolution of the design and why is this a suboptimal fractional factorial? Convert this to a 2^{5-1} design with the maximum possible resolution by changing the numbers in one of the columns.

5.25 Given below are the runs for a two-level design with five factors. What was the defining relation?

A	B	C	D	E
−	−	−	−	−
−	−	−	+	+
−	−	+	−	+
−	−	+	+	−
−	+	−	−	+
−	+	−	+	−
−	+	+	−	−
−	+	+	+	+
+	−	−	−	−
+	−	−	+	+
+	−	+	−	+
+	−	+	+	−
+	+	−	−	+
+	+	−	+	−
+	+	+	−	−
+	+	+	+	+

5.26 Explain the difference between complete confounding and partial aliasing. Assume the use of a fractional factorial design that is not run in blocks.

5.27 What is the maximum possible resolution of a 2^{8-3} design? Give one possible defining relation that would produce that resolution. What is the *minimum* possible resolution? Give one possible defining relation that would give this resolution and explain why the design would never be used.

5.28 Consider the design given in Exercise 5.18. Let the columns represent the factors A, B, C, D, and E (as usual), and assume that when the design was run, the values of the response variable were, in order of the rows in the design, 16.2, 18.4, 17.3, 19.4, 15.9, 17.1, 18.0, and 16.8. Analyze the data and present your conclusion. Do you believe that at least one follow-up experiment is necessary? Explain.

5.29 Referring to Yates and Mee (2000), construct a design with a reasonable number of treatment combinations for which there are seven two-level factors, with four factors in the first subset and three factors in the second subset.

5.30 The fact that an interaction that is the product of the two selected interactions which are selected to be part of the defining relation when a 1/4 fraction is used is also part of the defining relation is not necessarily intuitive. To see that this does in fact occur, proceed as follows for the following example. Construct a 2^{8-1} design, using of course $I = ABCDEFGH$. Then split this design in half by using only those treatment combinations for which $ABCD = 1$. (This can be easily done using MINITAB, for example.). Look at the value of $EFGH$ for each of those 64 treatment combinations. What must that value be? Is this what happens? Explain.

5.31 In his article "In the soup: A case study to identify contributors to filling variability" (*Journal of Quality Technology*, **20**(1), 1988), L. B. Hare describes an experiment performed with the objective of minimizing the filling variation among pockets of dry soup mix. The use of control charts had shown that the variation was unacceptable. Five factors were examined in a 2^{5-1} design: number of ports, temperature, mixing time, batch weight, and delay. (The latter is the delay in days between mixing and packaging. For temperature, "C" means that cool water was used and "A" means that no water was running through the mixing jacket.) The design and the values of the response variable are given below.

Row	A	B	C	D	E	Y
1	1	C	60	2000	7	0.78
2	3	C	80	2000	7	1.10
3	3	A	60	1500	1	1.70
4	3	C	80	1500	1	1.28
5	1	A	60	1500	7	0.97
6	1	C	80	1500	7	1.47
7	1	A	60	2000	1	1.85
8	3	A	80	2000	1	2.10
9	1	A	80	2000	7	0.76
10	3	A	60	2000	7	0.62
11	1	C	80	2000	1	1.09
12	1	C	60	1500	1	1.13

13	3	C	60	1500	7	1.25
14	3	A	80	1500	7	0.98
15	3	C	60	2000	1	1.36
16	1	A	80	1500	1	1.18

(a) Analyze the data by using appropriate computer software. Is a conditional effects analysis necessary? (Notice which effects are the three largest effects.)

(b) Consider your analysis versus the hand analysis given in the paper. Are most of your effect estimates a constant multiple of the effect estimates given in the paper? What does this suggest about one error that was made in the paper (several errors were made in the analysis)?

5.32 Explain or show why a $(2^{3-1} \times 2^{4-1})/2^1$ assembly design could not be resolution IV.

5.33 There is an upper limit on the number of experimental runs in a design constructed using various statistical software (e.g., with MINITAB the limit is 128 runs). Assume that you are using MINITAB, for which operations on columns can be easily performed, and you need to construct a 2^{9-1} design, as you need the design for an application in which experimental runs are very inexpensive. If you construct a 2^{9-2} design (i.e., 128 runs) and then fold it over using appropriate column operations in MINITAB, you would obtain a 2^{9-1} design. Would you want to use this design, however? Explain.

5.34 Assume that your boss is distrustful of fractional factorials, preferring instead to see "the whole thing, not a fraction." You ran a 2^{5-1} design last week but now he insists that the full factorial be used. You could run the other $1/2$ fraction and add it to the results from the first fraction, or you could start over and run the 2^5 design. Assume that it will be two weeks before the other $1/2$ fraction can be run and that certain conditions could change by then. How will you proceed?

5.35 The projective properties of a 2^{k-p} design were of relevance in a study described by T. Jørgensen and T. Naes in their paper "A design and analysis strategy for situations with uncontrolled raw material variation" (*Journal of Chemometrics*, **18**, 45–52, 2004). A cheese experiment was conducted that involved 6 two-level factors in a 2^{6-1} design with eight raw material blocks. Only the first four factors were analyzed in the paper, however, because the response variable is measured before the final process steps are performed, which involve the other two factors. *If* blocking had not been used, what effects would be estimable and would be unconfounded when the first four factors are analyzed? How would you describe the design when used for only these factors? Now, given your answers to these questions, how should the blocks be viewed and indeed what is one way that the blocks could have been constructed so as

to minimize the confounding of effects with blocks that might be significant? Read the paper, if desired, to learn more of this somewhat unusual use of a fractional factorial.

5.36 (Harder problem) Using a table of the signs for a 2^4 design, construct a 2^{4-1} design with 12 blocks of size 2 such that the same effects are estimable that are estimable with the 2^{4-1} design when run without blocking.

5.37 J. V. Stephenson and D. Drabenstadt (*Surface Mount Technology*, November 1999) described sequential experimentation relative to a solder paste printing process. Eleven factors were used in the first experiment and it was stated that the design was resolution IV. The 32 treatment combinations (so the design is a 2^{11-6}) are given at the following Web site: http://smt.pennnet.com/Articles/Article_Display.cfm?Section=Archives&Subsection=Display&ARTICLE_ID=111949&KEYWORD=printing, and the defining relation was not given. Use the method of Bisgaard (1993) or any other method you prefer and determine the defining relation.

5.38 Continuing Exercise 5.37, a second experiment was performed with five factors that were identified as important in the first experiment. A 2^{5-1} design was used so that the design was resolution V. The data for the second experiment (the data were not given for the first experiment) were given in the article at the URL stated in Exercise 5.37. The authors stated that one main effect and 2 two-factor interactions were significant, with print pressure being the significant main effect and the interactions were between print pressure and print speed and print pressure and snap off. Do you agree? If so, are you concerned about the fact that there are significant interactions involving the factor that is significant since a regression equation was developed for predictive purposes in the optimization stage? If so, what would you do differently, if anything. If you are not concerned, explain why.

5.39 B. N. Gawande and A. Y. Patkar (*Biotechnology and Bioengineering*, **64**(2), 168–173) described two experiments in which the effect of dextrin, peptone, yeast extract, ammonium dihydrogen orthophosphate, and magnesium sulfate on enzyme production was investigated in two fractional factorial experiments. In the first experiment, six observations were made at the center of the design. The observations, which were not given in the article, produced a standard error of the response of 0.144. The centerpoint replicates of course permit the construction of significance tests and it was found that all of the main effects and one interaction effect were significant when the data from the 2^{5-1} design were analyzed. (The numbers suggest that there should have been two significant interaction effects, contrary to the impression created by the article.) Significance was declared when an effect estimate exceeded twice its standard error. The design and the data are given below.

Row	Detrin	Pepton	Yeast Extract	$NH_4H_2PO_4$	$MgSO_4 \cdot 7H_2O$	Activity
1	-1	-1	-1	-1	-1	3.97
2	-1	-1	-1	+1	+1	5.99
3	-1	-1	+1	-1	+1	4.13
4	-1	-1	+1	+1	-1	5.59
5	-1	+1	-1	-1	+1	5.18
6	-1	+1	-1	+1	-1	6.47
7	-1	+1	+1	-1	-1	5.12
8	-1	+1	+1	+1	+1	6.53
9	+1	-1	-1	-1	-1	5.39
10	+1	-1	-1	+1	+1	5.25
11	+1	-1	+1	-1	+1	5.39
12	+1	-1	+1	+1	-1	6.06
13	+1	+1	-1	-1	+1	4.98
14	+1	+1	-1	+1	-1	6.74
15	+1	+1	+1	-1	-1	5.66
16	+1	+1	+1	+1	+1	8.42

(a) Analyze the data. In particular, note that the response values when the first, second, and fourth factors are at the low level are considerably smaller than the other values. Is this possibly suggestive of anything in particular?

(b) Compare the number of significant effects that you obtained with the number obtained by the authors of the article. What do you think accounts for the discrepancy? Which set of results would you be inclined to use?

(c) Of the 15 estimable effects, about how many effects, recalling the discussion in Section 4.11, would you expect to be significant without even looking at the data?

A second fractional factorial was used (actually the mirror image foldover of the first design), with the design points centered at the best design point from the runs that were made along the path of steepest ascent (see Section 10.4). The method of steepest ascent was also used after the second design. We return to this study in Exercise 10.12 in Chapter 10 and critique the response surface analysis.

5.40 In referring to the 2^{9-5}_{III} design given by Hsieh and Goodwin (1986), mentioned in Section 5.9.1.4, with generators $E = -ABCD$, $F = -CD$, $G = ACD$, $H = -BC$, and $J = ABC$, Mee and Peralta (2000, references) stated in part " . . . a better design exists. . . ." Explain why they would make such a statement. If you agree, give an example of a superior design.

5.41 (Requires knowledge of regression) It was stated in Section 5.20 that Design-Expert allows the user the option of selecting the model terms to use by using backward elimination, a variable selection technique that is used in regression analysis. Is this of any value when the design is orthogonal, however (as are most of the designs in the chapter), or will the model be the same as when the results of t-tests are used to arrive at the model?

5.42 Consider Example 5.3. If there are no restrictions on effects to be used in constructing a 2^{10-5} design to be run in eight blocks of size 4, is it possible to construct the design so that no main effect is confounded with a block estimate? Use appropriate software to construct the design and describe, in general, the alias structure.