CHAPTER 6

Designs with More Than Two Levels

Although two-level designs are the most popular and simplest designs to use, experimenters will oftentimes be interested in using more than two levels. In this chapter we consider full and fractional factorials with more than two levels, including full and incomplete mixed factorials. We do not consider Box–Behnken designs, which have three levels. These designs are intended to be used as response surface designs, so they are covered in Section 10.8.

6.1 3^k DESIGNS

Designs with three levels (a middle level in addition to low and high) permit the investigation of quadratic effects, although 3^k or 3^{k-p} designs might require too many runs, depending on the values of k and p.

Designs in the 3^k series can be generated rather easily using most statistical software, but not so with 3^{k-p} designs. This is discussed later. Given below is a 3^2 design that was created using the general full factorial design capability in MINITAB, with the runs randomized.

Row	A	B
1	1	1
2	3	2
3	3	1
4	3	3
5	1	2
6	2	3
7	1	3
8	2	1
9	2	2

(*Note*: There are various ways of denoting the three levels. One alternative way is to use –1, 0, and 1 to denote the three levels and another way is to use the numbers 0, 1, and 2, which is what is used later in the chapter, beginning with Section 6.1.1. The reason for this choice should become apparent in this section.) The following is MINITAB output for hypothetical response data from using this 3^2 design with the data simulated as $Y = A + e/20$, with e denoting a random error term that is $N(0, 1)$.

```
General Linear Model: Response versus

Factor Type Levels Values

Analysis of Variance for Response, using Adjusted SS

for Tests

Source   DF   Seq SS   Adj SS   Adj MS     F      P
 A        1   5.8535   0.1219   0.1219   26.55  0.014
 A*A      1   0.0015   0.0015   0.0015   0.34   0.603
 B        1   0.0002   0.0047   0.0047   1.02   0.388
 B*B      1   0.0081   0.0081   0.0081   1.76   0.277
 A*B      1   0.0017   0.0017   0.0017   0.38   0.582
 Error    3   0.0138   0.0138   0.0046
 Total    8   5.8788

 S = 0.0677626   R-Sq = 99.77%   R-Sq(adj) = 99.38%

    Term           Coef     SE Coef     T      P
    Constant    -0.1591    0.2757    -0.58   0.604
    A            1.0572    0.2052     5.15   0.014
    A*A         -0.02779   0.04792   -0.58   0.603
    B            0.2067    0.2052     1.01   0.388
    B*B         -0.06351   0.04792   -1.33   0.277
    A*B          0.02083   0.03388    0.61   0.582
```

Here A^*A denotes the quadratic component of factor A and similarly for B, and A^*B is the interaction term. Note that in this chapter we will denote interactions in this way rather than the more customary AB notation because one way to decompose the A^*B interaction is to decompose it into components that are designated as AB and AB^2, each with two degrees of freedom when A and B each have two levels. These components don't have any physical meaning but under certain conditions they can be related to another method of decomposing the interaction that does have physical meaning. This is discussed in Section 6.1.1.

The "adjusted" sums of squares and adjusted mean squares are just the usual sums of squares and mean squares. That is, the sums of squares are the values for each of the effects when the other effects are in the model. The sequential sums of squares are the sums of squares for the effects when the terms are entered in the order in which the effects are listed in the table. These two sets of columns are different because the model components are not orthogonal, as they are when two-level designs are used, which obviates the use of both sequential and adjusted sums of squares. In particular, notice that the two sum of squares for the A effect differ greatly, due to the correlation between A and A^*A, in particular. (*Note*: These numbers are obtained without any "orthogonal polynomial" approach. That is, the columns for the quadratic terms are not transformed, such as by centering, in any way. In general, some type of centering should be used, however (see Ryan, 1997, p. 251).)

Another way to decompose the A^*B interaction is in terms of the trend components. Since each main effect has two degrees of freedom, which allows the linear and quadratic effects to be estimated, with one degree of freedom for each, it logically follows that the interaction of two 3-level factors would have four components with one degree of freedom each that results from "crossing" the linear and quadratic components, just as one obtains the four treatment combinations in a 2^2 design. That is, the four components are linear $\times$ linear, linear $\times$ quadratic, quadratic $\times$ linear, and quadratic $\times$ quadratic. These are referred to as the trend components of interaction.

Of course the (linear) main effect of A is significant, as expected, and no other terms are significant, which is also as expected since we know the true model because these are simulated data.

In the output above, each factor has two df because of the three levels, with, as stated, one df used to estimate the linear effect and the other used to estimate the quadratic effect. Then why does the A^*B interaction have only one df instead of four since the df for the interaction term must be the product of the df for each factor that comprises the interaction? The different components of the A^*B interaction are explained in detail in the next section, but for the moment we will note that the interaction cannot be broken down using MINITAB unless the interaction components are entered physically when the model is specified, and only the trend components can be entered (i.e., AB and AB^2 cannot be entered). The same is true of the linear and quadratic effects of a factor. In general, any term that is to appear in the ANOVA table must be specified in the model, which would obviously be cumbersome if there were more than a few factors. Design-Expert will display the quadratic components in the ANOVA table and they can also be specified in the model. Confidence intervals are given for all the parameters corresponding to the nonlinear components if there is at least one df for the error term and the fitted model with the parameter estimates is also given. MINITAB will also give the confidence intervals and parameter estimates. MINITAB does not provide the capability for a normal probability plot for effect estimates with its general linear model (GLM) capability, which would be used for the types of designs covered in this chapter, nor does Design-Expert.

This is unfortunate because a normal probability plot can be constructed for *any* design, provided that the plotted effect estimates are uncorrelated and have the same variance. This is discussed further in Section 6.1.2.

Even when quadratic effects are displayed, the advantage of being able to investigate quadratic effects is offset somewhat by the complexity of interactions and the difficulty in interpreting them. When all factors have only two levels, each interaction has one degree of freedom, regardless of the order of the interaction. The number of degrees of freedom for an interaction term when all factors have three levels is 2^c, with c denoting the number of factors involved in the interaction. The interpretation of even a simple interaction such as the A^*B interaction in a 3^2 design is somewhat involved, partly because of the various possible ways of decomposing the four degrees of freedom for the interaction.

6.1.1 Decomposing the A^*B Interaction

As stated, one way to decompose the A^*B interaction is to decompose it into AB and AB^2. These have also been referred to as the *I–J* components. To understand these components, it is helpful to understand modular arithmetic, which, recall, was used in Section 5.1 for two-level fractional factorials.

To facilitate this, we will use 0, 1, and 2 to represent the three levels. (In principle, any set of three consecutive integers could be used.) With this designation, the column for AB is $A + B$, modulo 3, and AB^2 is $A + 2B$, modulo 3. (Modulo 3 means that $1 + 2 = 0$ and $2 + 2 = 1$, which is analogous to our base 10 (i.e., "modulo 10") number system if we consider only the units digit, since the latter is 0 for $1 + 9$, 1 for $2 + 9$, etc.) Thus, we have the following:

A	B	AB	AB^2
0	0	0	0
1	0	1	1
2	0	2	2
0	1	1	2
1	1	2	0
2	1	0	1
0	2	2	1
1	2	0	2
2	2	1	0

Orthogonality of these columns could be verified by subtracting 1 from each number (so as to make the column sums zero, as is the case with two-level designs) and showing that all column dot products are zero, as the reader is asked to show in Exercise 6.9. With column sums of zero, the columns for A^2 and B^2 would each be $1 -2\ 1\ 1 -2\ 1\ 1 -2\ 1$. The reader can verify that each of these two columns is orthogonal to what the four columns above become after 1 is subtracted from each number. As is probably self-evident, the columns for A^2 and B^2 cannot be obtained by transformation of the columns for A and B (using modular arithmetic or otherwise) since those columns measure the *linear* effect of A and B, respectively. That is, the numbers 1, 2, 3 form a linear progression, whereas a graph of $Y = 1, 2, 1$ against a sequence index would show quadrature. Similarly, if the value of the response variable at the middle level of a factor is quite different from the values at the low level and the high level, respectively, the linear combination $1(Y_{\text{low}}) - 2(Y_{\text{middle}}) + 1(Y_{\text{high}})$ will be quite different from zero.

Of course we should note that referring to the linear effect of a factor makes sense only if the factor is quantitative. It wouldn't make any sense to speak of the linear effect of a qualitative factor because the level designations are arbitrary. Thus, the terms low, middle, and high in referring to the levels would not make any sense unless there was an ordinal scale involved (such as army ranks).

As stated, the components AB and AB^2 have no physical meaning but there is a connection between these components and the trend components. Specifically, Ryan (1981) showed that the sum of squares for the AB and AB^2 components are equal when exactly one of the trend components of the interaction is nonzero. There is a related discussion in Sun and Wu (1994).

The interaction plots have a definite type of configuration for each of these four cases, and prior information that one of the four cases may be very likely to occur, such as the interaction being strictly linear × linear, may be available. When a 3^2 design is run in blocks of size 3, the usual approach is to confound either AB or AB^2 with the difference between blocks. If AB^2 were large and AB quite small, a serious mistake would be to confound the larger component, as that could lead to the false conclusion that there is no interaction. It would be comforting to know that nothing is lost when one of these two components is selected to be confounded. Further research is needed in this area for other 3-level designs and for mixed factorials.

6.1.2 Inference with Unreplicated 3^k Designs

Unreplicated 3^k designs, like unreplicated 2^k and 2^{k-p} designs, do not have any degrees of freedom for an error term unless some interaction components are not estimated and instead are used to create an error term. Unlike 2^k and 2^{k-p} designs, however, software developers have not provided the capability for normal plots for use with factorial designs with three (or more) levels, as indicated previously. There is no reason why such plots cannot be constructed, however. Indeed, a normal probability plot could be constructed for *any* set of continuous data or any set of statistics computed from such data, provided the assumptions of independence and equal variances are met. Similarly, the method of Lenth (1989) could be applied to any unreplicated factorial, but this has generally not been done by software developers. Indeed, both MINITAB and Design-Expert utilize Lenth's method only for two-level designs.

Wu and Hamada (2000, pp. 222, 226) do give two half-normal plots for effect estimates from a 3^{4-1} design with three replications and conclude for one of them that the linear effect of the first factor may be significant, the conclusion being confirmable by using Lenth's method, which the authors leave as an exercise. Of course we much prefer to see normal probability plots of effect estimates identifying significant effects for us, rather than having to use hand computations to assess significance. (Designs in the 3^{k-p} series are discussed in Section 6.3.) Although it is, of course, easier to restrict consideration to two-level designs, the absence of normal probability plot methods for general factorial designs seems to be a common shortcoming of statistical software and software specifically for design of experiments (DOEs). Interestingly, this shortcoming was not mentioned in the software survey of Reece (2003).

Example 6.1

An application in which a 3^2 design with three replicates of the centerpoint (which unlike a 2^k design happens to be one of the points in the design) was given by Vázquez and Martin (1998). The objective was to "optimize the growth of the yeast *Phaffia* *rhodozyma* in continuous fermentation using peat hydrolysates as substrate." The two factors in the experiment were dilution rate at levels of 0.13, 0.23, and 0.33, and pH at levels of 5, 7, and 9. There were four response variables: biomass concentration at steady-state conditions (g/L), substrate concentration at steady-state conditions (g/L), biomass yield (g/g), and biomass volumetric productivity [g/(L z h)].

We will use the first of these response variables for illustration, with the data given below. (Vázquez and Martin (1998) stated that the response values given below are the averages of three numbers; the 33 original observations were not given.)

A	B	Y
-1	-1	3.8
-1	0	4.0
-1	+1	1.4
0	-1	5.1
0	0	4.8
0	0	4.9
0	0	4.9
0	+1	1.9
+1	-1	4.5
+1	0	3.8
+1	+1	0.8

(Note that *A* and *B* are orthogonal, as must be the case, just as columns representing the levels of two-level factors are orthogonal.)

We saw in Chapter 4 that the definition of a main effect for two-level factorials was quite intuitive as we simply calculate the average response value at the high level of a factor and subtract from that average the average response value at the low level of the factor.

The way to measure the effect of a factor may not be obvious when the factor has three levels, however, and indeed various alternative definitions have been proposed and can be proposed. We need an estimate of the *linear effect* and the *quadratic effect* of each factor. The middle level can be viewed as the "additional" level that is used to detect quadrature, so it would be ignored when the linear effect is estimated. Thus, *one way* to define the linear effect of a 3-level factor is as the average response at the high level of the factor minus the average response at the low level. An alternative definition, which will give one-half of the value obtained using the intuitive definition, is to bring the middle level into play and define the linear effect as the average of two linear effects—middle minus low and high minus middle. This gives (average at high level–average at low level)/2, since the middle level drops out in averaging the two linear effects. We could make a good argument for each definition but it is the second definition that is generally used.

A similar debate could occur regarding the definition of the quadratic effect of a factor since all three levels must be used directly. In order for the effect of a 3-level factor to be 100 percent linear, the difference between the average response at the high level and the average response at the middle level must be the same as the difference between the average response at the middle level and the average response at the low level. If not, the line segment that connects the three averages will not be a straight line and there will thus be evidence of quadrature. Therefore, the quadratic effect should seemingly be measured by the difference of these two differences, but instead it is defined as the (less intuitive) difference of the differences divided by 2.

To summarize, let $\overline{A_0}$, $\overline{A_1}$, and $\overline{A_2}$ denote the average response value at the low, middle, and high levels, respectively, for factor A, and let A_L and A_Q denote the linear and quadratic effects, respectively, of factor A. Then

$$A_L = \frac{\overline{A_2} - \overline{A_0}}{2}$$

and

$$A_Q = \frac{\overline{A_2} - 2\overline{A_1} + \overline{A_0}}{2}$$

and similarly for factor B.

For the current example, $A_L = -0.017$, $A_Q = -2.54$, $B_L = -1.55$, and $B_Q = -1.40$. We could similarly define four interaction effect estimates, $A_L B_L$, $A_L B_Q$, $A_Q B_L$, and $A_Q B_Q$, which would be computed using coefficients that are products of the component coefficients.

The sum of squares for the terms in the model can, for a balanced design, be obtained simply using products of the columns for A and B times Y and divided by the sum of the squares of the numbers in the column that is multiplied times Y. It is more complicated for unbalanced data, however, as in this example, and a complete discussion would be beyond the intended scope of this text.

Vázquez and Martin (1998) analyzed these data by fitting a regression model with linear and quadratic terms in A and B and the interaction term. All terms except the linear term in A are (highly) significant when this regression is run.

Critique

It is not clear why an unbalanced design was used; it appears, though, that this may have been done to facilitate a lack-of-fit test from the repeated runs. This was unnecessary, however, as the 33 original data points could, assuming they were available, be used and that would permit a lack-of-fit test. Furthermore, the design would be balanced as it would be a 3^2 design with three replicates, thus avoiding the headaches that are caused by unbalanced data.

Example 6.2

Clark (1999) described an experiment in which a 3^3 design with three replicates was used in a resistance welding operation experiment. The three factors studied were current, weld time, and pressure. The data were not given so we cannot analyze it, but we can critique the design and the methods of analysis. The three-factor interaction was not investigated, despite the fact that there were plenty of degrees of freedom for doing so. Eighteen degrees of freedom are needed to estimate the main effects and two-factor interactions, so with an unreplicated 3^3 design there would be eight degrees of freedom for estimating the error term, which are the eight degrees of freedom for the A^*B^*C interaction.

Only three degrees of freedom were listed for "interaction" in the ANOVA table, however, whereas the number should have been 12 since there are four degrees of freedom for each of the three interactions. This analysis was performed in MINITAB and it seems likely that only the linear $\times$ linear component of the interaction was broken out in the ANOVA table. Ignoring most of the degrees of freedom for the interaction is very risky. Two of the three factors were significant at the .01 level in terms of the linear effect and one quadratic component was also significant, as well as one of the two-factor interactions. The latter should necessitate a conditional effects approach (see Section 6.2).

The R^2 value was only .806 with all nine fitted terms in the model, suggesting that the 81 observations might have been used more efficiently to look at additional factors, especially since one of the three linear effects and two of the three quadratic effects were significant. There were two huge standardized residuals (-5.12 and 5.13) on consecutive observations, but this was not noted in the paper. It is interesting that these observations are in opposite directions relative to their respective fitted values. A possible explanation is that the tooling process was badly out of control at those two observations. If so, then all the effect estimates are invalid and some of them might be off by a considerable margin.

A Pareto chart effect analysis was performed and the model was reduced to six terms, which was then analyzed to determine optimum operating conditions. Those conditions soon ceased to be optimum because variation in tip height and taper created problems with the process. This underscores the importance of maintaining processes in a state of statistical quality control, as was emphasized in Section 1.7 for this particular experiment.

After a state of statistical quality control has been reached and maintained, future experimentation might best be performed with fewer observations, which might not only greatly reduce the cost of experimentation, but the use of 81 observations could result in tests of significance that will detect significant effects that do not have practical significance. A better choice for a design might be either a response surface design (Chapter 10), a 3^{k-p} design if more factors are to be examined and all have three levels, or perhaps a mixed fractional factorial (Section 6.5) if differing numbers of levels are to be used for the factors. Indeed, Bisgaard (1997) in a nontechnical article expressed a preference for response surface designs over three-level designs for technological applications.

6.2 CONDITIONAL EFFECTS

We need to examine conditional effects for 3-level designs just as we do for 2-level designs. Assume that an experiment with a 3^2 design has been run and the results are

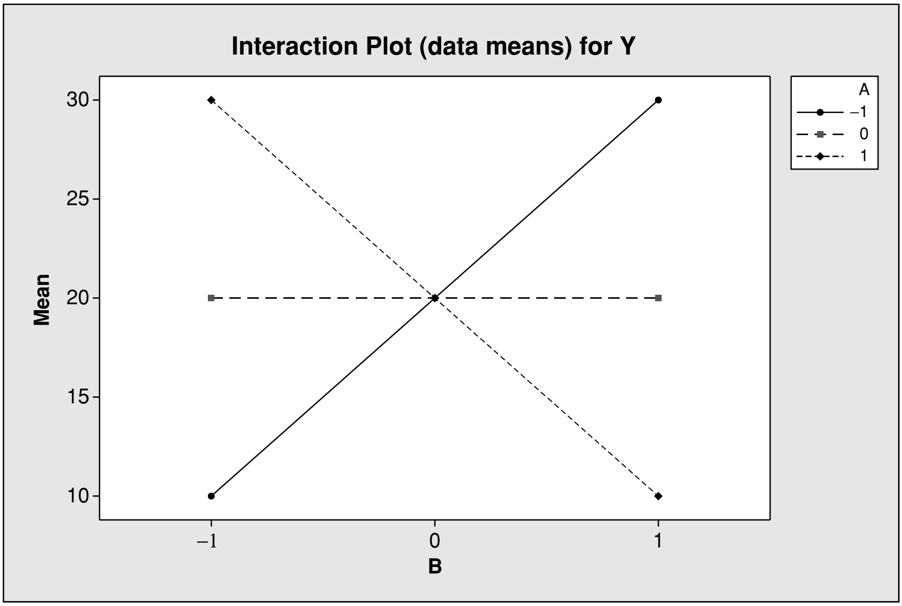

Figure 6.1 Interaction plot for 3^2 design.

as follows:

A	B	Y
-1	-1	10
-1	0	20
-1	1	30
0	-1	20
0	0	20
0	1	20
1	-1	30
1	0	20
1	1	10

The data are plotted in Figure 6.1.

If we computed A_L, A_Q, B_L, and B_Q, we would find that each one is zero. (Figure 6.1 is the counterpart to Fig. 4.2 that was for a 2^2 design.) Thus, if we ran these data through software that produced an ANOVA table, we would find that the sum of squares for each effect is zero.

Given below is the MINITAB output for these data (using the GLM command and using only A and B as the model terms). Even though the linear and quadratic sum of squares are not shown separately, in this case it is unnecessary because if the sum is zero, each of the two component parts must be zero.

```
General Linear Model: Y versus A, B

Factor   Type    Levels  Values
A        fixed     3     -1, 0, 1
B        fixed     3     -1, 0, 1

Analysis of Variance for Y, using Adjusted SS for Tests

Source   DF   Seq SS   Adj SS   Adj MS     F       P
A        2       0.0      0.0      0.0  0.00   1.000
B        2       0.0      0.0      0.0  0.00   1.000
Error    4     400.0    400.0    100.0
Total    8     400.0

S = 10   R-Sq = 0.00%   R-Sq(adj) = 0.00%
```

The A^*B interaction is not displayed in the table because it was not specified in the model. Since this is an unreplicated factorial, it is confounded with the error term.

It is clear from Figure 6.1 that A and B each have a strong linear effect when viewed in the context of conditional effects; the conditional main effects simply add to zero as one is -20 and the other is 20. (These numbers result from conditioning, in turn, on the low and the high levels of each factor.)

Of course this is the most extreme case, and the zero sums of squares, which correspond to the main effect estimates being zero, are a sizable red flag. The point, however, is that the use of conditional effects when factors have more than two levels should match the use of conditional effects for two-level designs.

In general, the need to consider conditional effects applies to *every* type of factorial design, and also to designs with blocking since there is no guarantee that the assumption of no block $\times$ treatments interaction in a randomized complete block design will be met, and if the assumption is not met, then the rank ordering of the treatments within a block might differ considerably over the blocks. Then the conditional effects, which might be defined as the deviations of treatment i within each block from the average response for the observations in that block, could differ considerably over the blocks.

6.3 3^{k-p} DESIGNS

Although a 3^k design generally requires too many runs, a 3^{k-p} design reduces that number by a factor of 1/3—a sizable reduction. Nevertheless, historically these designs have not been viewed favorably, especially relative to other designs, covered in Chapter 10, for estimating second-order effects because the 3^{k-p} designs have been assumed to be relatively inefficient. In particular, Box and Wilson (1951) made this argument.

Recall from Section 5.9.1.4, however, that 2_{III}^{k-p} designs are used as screening designs (i.e., the focal point is identifying significant main effects). There is no reason why 3^{k-p} designs cannot be used for the same purpose in a two-stage procedure, with possible interaction effects investigated in the second stage, just as is done when a 2_{III}^{k-p} design is used in the first stage. This is essentially the point that was made by Cheng and Wu (2001), who pointed out, for example, that up to 13 factors could be studied with a 3^{k-p} design with 27 runs. This is apparent because each factor would have 2 df and $13(2) = 26$, the available number of degrees of freedom. Thus, the design is saturated relative to the estimation of single-factor effects. It should be noted that Cheng and Wu (2001) proposed a method that permits factor screening and model fitting in the same stage, thus reducing the two-stage procedure to a single stage. This is discussed further in Section 10.1.

As discussed in Section 6.7, statistical software generally does not provide the capability for generating many of the designs discussed in this chapter, especially a 3^{13-10}.

It is also worth noting that many 3^{k-p} designs will not have enough df for fitting the full second-order model, anyway. For example, there are eight df for estimating effects with a 3^{3-1} design, but since each two-factor interaction has four df, there are only enough df to estimate the linear and quadratic effects of each factor, with two df left. Thus, a full 3^3 factorial would have to be used to fit the full second-order model: 6 df for the main effects (linear and quadratic) and 12 df for the 3 two-factor interactions, and there would be 8 df for the three-factor interaction if that were fit.

A 3^{4-1} design will *not* work, however, as 8 df are needed to estimate the main effects and 24 df are needed for all of the two-factor interactions. A 3^{5-1} design will work, as 80 df far exceeds the 10 that are needed for main effects and the 40 that are needed for the two-factor interactions. A 3^{5-2} design will obviously not have enough degrees of freedom, however. Similarly, a 3^{6-1} design will work, as will a 3^{6-2}, but not a 3^{6-3}. It should be clear that as we let the number of factors increase, we will not be able to fit a full second-order model with a reasonable number of design points.

Of course designs with three-level factors are used when there are three levels of interest. Three levels don't necessarily have to be used at the screening stage, however, so we have to think about what the justification would be for doing so. The general belief is that when quadrature is present, it is generally not present in the absence of linearity, and indeed models with polynomial terms are built on this assumption.

Assume that we have a temperature factor and we are primarily interested in 300° and 500°F. If we elect to additionally use 400°F at the screening stage, we are implicitly assuming that if the factor has an effect, the quadrature will dominate the linearity, as would occur, for example, if the plot of the response values against temperature has a V-shape configuration.

On the other hand, if 300°F and 400°F are the primary levels of interest but 500°F is chosen for use as a "test for linearity," the three levels would be quite revealing if the response value for the middle temperature is slightly higher than that for the lowest temperature, but the response is much lower at 500°F (as in the plot of factor D in Figure 5.1 in Wu and Hamada, 2000). In this case the linear effect will dominate the quadratic effect, but we wouldn't have been able to detect the linear effect without the use of the third level.

So the justification for using a 3^{k-p} design as a screening design will hinge largely upon whether or not a possible quadratic effect might be expected to dominate a possible linear effect, and whether or not the third level is within the range of the two levels if a 2^{k-p} design had been used instead.

Certain concepts that were discussed in Chapter 5 for 2^{k-p} designs also apply to the 3^{k-p} series. Specifically, the resolution of a 3^{k-p} design is determined in the same way as for a 2^{k-p} design, and we may also speak of minimum aberration 3^{k-p} designs, which are discussed in Section 6.7.

Of course this implies that the 3^{k-p} designs have the same type of projective properties that the 2^{k-p} designs possess; for example, a 3^{4-1} design would project into a full 3^3 design in any combinations of three factors, regardless of whether $D = ABC$ or $D = ABC^2$, as the reader is asked to show in Exercise 6.10.

Projective properties for 3^{k-p} designs are somewhat complicated in general, however. The interested reader is referred to Theorem 1 of Cheng and Wu (2001). Part of that theorem states that if we start with a 3^{k-p} design and project it onto q factors, then the projected design is either a fractional factorial design or a 3^q design with $3^{k-p}/3^q$ replicates.

6.3.1 Understanding 3^{k-p} Designs

To try to simplify a complex subject as much as possible, let's consider a 3^{2-1} design. This would not be a practical design because with only two degrees of freedom only one of the main effects could be estimated, but we need to start with a very simple design because things can become complicated in a hurry. There are two degrees of freedom associated with the three possible fractions of the 3^2 design, so we need to confound either AB or AB^2 with the estimate of the mean. Since we must relinquish information on one of the two components, it would be nice to know when these will be approximately equal in magnitude. As stated in Section 6.1.1, they will be equal when only one of the four trend components of the interaction is nonzero. Since the interaction plots for these four scenarios have distinct forms, as the reader is asked to show in Exercise 6.14, the experimenters might provide input as to whether or not any of these scenarios would be likely.

Clearly, we don't want to confound a large component and then conclude from the estimable small component that the interaction is not significant, nor do we want to do the reverse and conclude that an interaction is significant when it may be borderline at best.

Assume for the sake of illustration that we elect to confound AB^2. Recall that for the 2^{k-p} series the arithmetic was performed modulo 2 because the designs had two levels. For the 3^{k-p} series the arithmetic is modulo 3. For that arithmetic, the AB^2 interaction is represented by $x_1 + 2x_2 = 0$, 1, and 2 for the three fractions, with x_1

representing the first factor and x_2 representing the second factor. Unlike the 2^{k-p} series, we cannot use the presence or absence of a factor to represent the level of a factor since there are three levels. Therefore, *for the sake of constructing the fractions*, the levels are denoted as 0, 1, and 2.

Thus, with AB^2 confounded with the estimate of the mean, the three fractions, which would be the three blocks if a 3^2 design were run in blocks of size 3, are

00	02	01
11	10	12
22	21	20

with the fractions corresponding to, in order, $x_1 + 2x_2 = 0$, 1, and 2. If *AB* had been confounded with the estimate of the mean, the fractions would have been

00	01	02
12	10	11
21	22	20

Since an exponent can be a 2, the reader may wonder why A^2B and A^2B^2 are not being considered. The algebraic representation of the first is $2x_1 + x_2$ and $2x_1 + 2x_2$ for the second. The reader can observe that the first expression will produce the same three fractions as was produced by the expression $x_1 + 2x_2$, and $2x_1 + 2x_2$ produces the same set of fractions as was produced by $x_1 + x_2$. Thus, there is a reason for not using expressions for which the exponent of the first factor is not a 1. Furthermore, the squaring of such expressions ceases to be a "black box operation" when we see what underlies such actions.

It is important to note that, as was implied earlier, we have information on one component of the A^*B interaction—the one that is not selected for the defining contrast. That is, we have lost half of the information about the interaction. When we have more than two factors and a $1/3$ fraction is used, we retain information on interaction components that are not confounded, and for the $3^{k-\mathrm{p}}$ series we will have some information on components of interactions that are selected to be confounded with the mean.

There is an analogous interaction breakdown when there are three factors. Since the A^*B^*C interaction has (2)(2)(2) = 8 df, there are other components of the interaction, with the full set being ABC, ABC^2, AB^2C, and AB^2C^2, each of which has 2 df. Notice that the exponent of A is one, by convention, and the combinations of exponents of B and C are all possible combinations of each exponent being either 1 or 2.

6.3.2 Constructing 3^{k-p} Designs

The construction and understanding of these designs is complicated by the fact that the factors involved in interactions can have an exponent of 2. For example, as stated in Section 6.1, the A^*B interaction has four degrees of freedom, two each for AB and AB^2.

These designs are constructed analogous to the way 2^{k-p} designs are constructed. That is, a full factorial is constructed in $k-p$ factors, with one or more generators used to construct the additional column(s). For example, a 3^{4-1} design would be constructed by first constructing the 3^3 design and then defining the fourth factor, D, to be equal to one of the components of the A^*B^*C interaction, such as $D = ABC^2$. More specifically, the level of D in each experimental run would be determined as $D = x_1 + x_2 + 2x_3$ (mod 3), with x_1, x_2, and x_3 denoting the levels of A, B, and C, respectively in the 3^3 design, and "mod 3" indicating that modulo 3 arithmetic is performed, as stated previously.

Of course there are three possible fractions for a given choice of the effect to use in constructing the factor D, just as there are two possible fractions in the 2^{k-1} series. The three fractions are defined by $x_1 + x_2 + 2x_3 = 0$, 1, and 2, mod 3. Unlike the 2^{k-1} series, however, the use of, say, $D = ABC$ to generate the additional factor does not mean that the defining relation is $I = ABCD$.

This type of lack of correspondence will be explained shortly, but first we will apply the modular arithmetic approach to a 2^{4-1} design. Letting $D = ABC$ means $x_4 = x_1 + x_2 + x_3$, so $x_1 + x_2 + x_3 - x_4 = 0$. We need to remove the $-x_4$ because we need a positive coefficient for that term, which can be accomplished by adding $2x_4$ to each side of the equation. This is equivalent to adding zero to each side because $2x_4 = 0$ (mod 2), regardless of whether $x_4 = 0$ or 1. This produces $x_1 + x_2 + x_3 + x_4 = 0 = I$, so $I = ABCD$.

Modular arithmetic was not used in explaining the 2^{k-p} designs in Chapter 5 because it wasn't needed there. This can be done here, however, to serve as a bridge that may make certain aspects of the 3^{k-p} series easier to understand. Consider the simple case of a 2^{3-1} design constructed by first constructing the 2^2 design and then letting $C = AB$. Using 0 and 1 to denote the two levels since modular arithmetic is being used, the design is thus represented as

A	B	C
0	0	0
1	0	1
0	1	1
1	1	0

Note that $x_3 = x_1 + x_2$ and that $x_1 + x_2 + x_3 = 0$ (mod 2). (Note that if we wanted to translate this design back to the $(+1, -1)$ level designations, we would have to let 0 = 1 and $1 = -1$. This substitution would produce the first of the two fractions in Figure 5.1 in Section 5.1, with the treatment combinations in the reverse order.

Returning to the 3^{3-1} and 3^{4-1} design explanations, for the former $C = AB$ so $x_3 = x_1 + x_2$ and $x_1 + x_2 - x_3 = 0$. Notice that in this instance adding $2x_3$ to each side of the equation won't work because, for example, $2(1) = 2$ (mod 2), not zero. Therefore, we must add $3x_3$ to each side of the equation. This works because $3(0) = 3(1) = 3(2) = 0$ (mod 2). Therefore, $x_1 + x_2 + 2x_3 = 0$, so $I = ABC^2$, even though $C = AB$. The design is thus as follows:

A	B	C
0	0	0
0	1	1
0	2	2
1	0	1
1	1	2
1	2	0
2	0	2
2	1	0
2	2	1

Notice that $ABC^2 = 0(= I)$. We have thus illustrated that we cannot obtain the defining relation directly from the expression for the additional factor for three-level designs. The 3^{4-1} design can be similarly constructed, as the reader is asked to show in Exercise 6.22.

6.3.3 Alias Structure

The alias structure for a 3^{k-1} design is determined differently from the way it is determined for the 2^{k-1} series, and similarly for 3^{k-p} designs in general. Specifically, for a 3^{k-1} design the alias structure is determined by multiplying each effect by the effect that is confounded with the mean and the square of the effect that is confounded with the mean.

Therefore, for the 3^{2-1} design in the current example, if we use AB^2 as the effect that is confounded with the mean and pick one of those three fractions (preferably at random), we know that A, B, and AB must share the two degrees of freedom for estimating effects, and thus be confounded. Thus, $A(AB^2) = A^2B^2 = AB$ and $A(AB^2)^2 = B^4 = B$, so that $A = AB = B$. Similarly, $B(AB^2) = A$ and $B(AB^2)^2 = A^2B^2 = AB$, so that $B = A = AB$, as of course determined with the first set of multiplications.

Again, this is not a practical design; the intent was only to illustrate the construction of a 3^{k-p} design with the smallest number of factors possible for simplicity.

As an aside, it is worth noting the expression given by Wu and Hamada (2000, p. 216) for the maximum number of factors that can be studied with a 3^{k-p} design. The expression is $(3^{k-p} - 1)/2$, which does give 1 for the previous example as there are two degrees of freedom for estimating effects, and A and B each have two degrees of freedom. The way this result should be viewed, so that it doesn't seem nonsensical since k appears to be fixed in $k - p$, is to view $k - p$ as the fixed quantity, not k.

6.3.4 Constructing a 3^{3-1} Design

As a second example, we consider the construction of a 3^{3-1} design for which we know that we *could* estimate four factor main effects with that number of design points, but of course we have only three factors. With three factors, we would logically select one of the four components of the A^*B^*C interaction (ABC, AB^2C, ABC^2, and

AB^2C^2) to confound with the estimate of the mean. Note that with four choices for the interaction and three choices for the fraction since this is a 1/3 fraction, there are 12 fractions from which one is selected.

Assume that ABC is selected. Then obviously each factor effect will be confounded with a two-factor interaction component and part of the three-factor interaction component since, for example, $A(ABC) = AB^2C^2$ and $A(ABC)^2 = B^2C^2 = BC$.

Of course, there are eight degrees of freedom and A, B, and C account for six of them. The other two degrees of freedom are represented by the alias string $AB^2 = AC^2 = BC^2$, as the reader is asked to show in Exercise 6.3.

Giesbrecht and Gumpertz (2004, p. 384) comment on the 3^{3-1}design, stating "We must warn the reader that while this fractional plan using 9 of 27 treatment combinations has proved to be quite popular in some circles, it is not really a good experimental plan. The problem is that the main effects are all aliased with two-factor interactions." Although the warning does seem to be appropriate, the statement is somewhat misleading. The problem here is somewhat analogous to running a 3^2 design in blocks of size 3. If we confound AB with the block differences, we still have information on AB^2. Specifically, the eight df are broken down as blocks (2), A(2), B(2), and AB^2(2).

For the 3^{3-1} design with ABC confounded, the BC component of the B^*C interaction might be much smaller than the BC^2 component, in which case the fact that A is confounded with BC might not be a major problem, and similarly for the other two factors. Although there are at present no theoretical results to provide the basis for a strategic approach, if the sum of squares associated with an alias string of two-factor interaction components is large, it might be desirable to look at the trend components of the interactions and look at the relative magnitude of those components. There is a need for research in this area, but at present it would be best to be cautious and heed the warning of Giesbrecht and Gumpertz (2004) while recognizing that the aliasing of main effects with two-factor interaction components won't always be a problem.

The construction of the design can proceed analogous to the way that a 2^{k-p} design is constructed; that is, a 3^2 design could first be constructed with the additional factor created as $C = AB$ or $C = AB^2$. Assume that the first option is used. If this were to be a 2^{3-1} design, then $I = ABC$. This does not apply to the 3^{k-p} series, however, as was illustrated.

6.3.5 Need for Mixed Number of Levels

Of course one problem with the 3^{k-p} designs is the large jumps in run size for the various plans, from 9 to 27 to 81 and beyond. Eighty-one runs will be viewed by most experimenters as too many runs, yet the tables of Wu and Hamada (2000, pp. 250–251) show that there is only one 27-run 3^{k-p} design, the 3^{4-1}, that is resolution IV. (See Chen, Sun, and Wu (1993) for the complete enumeration of 27-run 3^{k-p} designs, upon which Table 5a.2 in Wu and Hamada (2000) is based. This work has been extended by Xu (2005), who gives a large number of 3^{k-p} designs for up to 729 runs.)

There shouldn't be a need, however, to use three levels for every factor when there is at least a moderate number of factors, as some factors may be qualitative with only two levels of interest. There may also be a strong belief that a given quantitative factor has a linear effect over the range of levels that are of interest and are considered feasible. Therefore, a fraction of a design with a mixture of the number of levels for the factors will often be the appropriate choice. Such designs are discussed in Section 6.5.

6.3.6 Replication of 3^{k-p} Designs?

While we normally don't think about replicating a 2^{k-p} design, although an example of one was given in Section 5.13, replicating a 3^{k-p} design is actually not a bad idea. One reason for doing so is the lack of a normal probability plot capability; of course that problem can be avoided simply by replicating the design. Although this is probably the most compelling reason for the replication, we should note that we have partial information on interactions whenever a 3^{k-p} design is used since not all components are confounded, as stated previously. If we are fortunate enough to have the "right" (i.e., largest) components unconfounded, then by using replication we obtain more precise estimates of those components. The downside is that the number of runs is of course $c(3^{k-p})$, with c denoting the number of replicates. This number will grow rapidly with c if 3^{k-p} is not small.

6.4 MIXED FACTORIALS

Frequently there is interest in using designs for which the number of levels for the factors is not the same, and as noted in Section 6.3.3, the number of runs with a 3^{k-p} design will be large when $k - p$ is not very small. A design with differing numbers of levels of the factors is called a *mixed factorial*.

A mixed (full) factorial (also called an asymmetrical factorial) in two factors is of the general form $a^{k_1}b^{k_2}$, with k_1 and k_2 greater than or equal to 1, and $2 \leq a < b$. Mixed factorial designs have been discussed, in particular, by Addelman (1962) and in Chapter 18 of Kempthorne (1973), although the treatment in these sources is somewhat mathematical. A more recent source, albeit with somewhat limited treatment, is Giesbrecht and Gumpertz (2004). Hinkelmann and Kempthorne (2005) devote a chapter to confounding in mixed factorial designs, which they term asymmetrical factorial designs.

Example 6.3

The simplest example of a mixed factorial is a 2×3 design. Sahin and Örnderci (2002) investigated the effect of Vitamin C (two levels) and chromium picolinate (three levels) and concluded that the high level of Vitamin C (250 mg) combined with either the high or middle level of chromium "can positively influence the performance of laying eggs reared under low ambient temperature." There were 180 laying hens, all 32 weeks old, which were used in the experiment, and they were divided into six groups of 30 hens each. This provides a considerable amount of power for detecting effects—which would be too much power in a typical application. In many if not most applications, it would be impractical, impossible, or simply too costly to obtain this many observations for each treatment combination. There were actually three response variables, so this example is revisited in Exercise 12.10 of Chapter 12. Experiments can be run inexpensively in many fields of application, however, and in such cases a full mixed factorial may be affordable.

In this application, 48 design points could be afforded, and in fact all of the runs could be made in one day. One problem with mixed factorials is that in many, if not most, applications, if k_2 in the general expression for a mixed factorial given earlier is at least 2, the number of design points may be prohibitive. We can use a fractional mixed factorial design, but the design won't necessarily be orthogonal.

Consider the 3×2^4 design and assume that only 24 runs can be afforded. The necessary fractionation must obviously be in regard to the two-level design in order to produce half the original points, and also because one cannot fractionate on a single 3-level factor, as multiple 3-level factors would have to be involved. Clearly, a $3 \times 2^{4-1}$ design would produce the desired number of points and would be an orthogonal design. If, however, something other than 6, 12, or 24 points were desired, the design would not be orthogonal. This becomes clear if we recognize that a design with any other number of points would not be of the form $3 \times 2^{4-p}$. For example, a design with 14 points wouldn't work because 14 is not a multiple of 3. Similarly, 15 would not work because $2^{4-p} \neq 5$, for any integral value of p, and so on.

We encounter the same type of analysis problem with mixed factorials as we do with three-level designs in that we can't construct a normal probability plot for a mixed factorial with commonly used statistical software.

6.4.1 Constructing Mixed Factorials

A standard way of constructing mixed factorials is by using the method of collapsing levels, due to Addelman (1962), such as collapsing a 3-level factor into a 2-level factor. Giesbrecht and Gumpertz (2004) discuss how a $2^2 3^2$ design would be constructed by using this approach, which starts with a 3^4 design and ends with a $2^2 3^2$ design with 81 runs, so that some of the runs are repeated. Of course the "extra" runs would have to be extracted if the experimenter wanted to use only 36 runs.

For a relatively small mixed factorial such as this, we could simply construct a 3^2 design and then construct a 2^2 design at each point in the 3^2 design. Of course the easiest approach would be to use software. The following $2^2 3^2$ design was created using the "full factorial" DOE option in JMP, with the runs not randomized. It is easy to see the 2^2 design at each point in the 3^2 design.

```
11--   1   1   -1   -1
11-+   1   1   -1    1
11+-   1   1    1   -1
11++   1   1    1    1
12--   1   2   -1   -1
12-+   1   2   -1    1
12+-   1   2    1   -1
12++   1   2    1    1
```

```
13--  1  3  -1  -1
13-+  1  3  -1   1
13+-  1  3   1  -1
13++  1  3   1   1
21--  2  1  -1  -1
21-+  2  1  -1   1
21+-  2  1   1  -1
21++  2  1   1   1
22--  2  2  -1  -1
22-+  2  2  -1   1
22+-  2  2   1  -1
22++  2  2   1   1
23--  2  3  -1  -1
23-+  2  3  -1   1
23+-  2  3   1  -1
23++  2  3   1   1
31--  3  1  -1  -1
31-+  3  1  -1   1
31+-  3  1   1  -1
31++  3  1   1   1
32--  3  2  -1  -1
32-+  3  2  -1   1
32+-  3  2   1  -1
32++  3  2   1   1
33--  3  3  -1  -1
33-+  3  3  -1   1
33+-  3  3   1  -1
33++  3  3   1   1
```

6.4.2 Additional Examples

A $3^2 2^3$ design was used in one of the two experiments described by Inman, Ledolter, Lenth, and Niemi (1992). The experiment involved a Baird spectrometer. Control charts of spectrometer readings had exhibited instability, so an experiment was conducted to try to identify the causes of the instability. Spectrometers must be frequently calibrated, especially when there is evidence of instability, but if the causes of the instability could be identified and removed, many of the time-consuming recalibrations could be eliminated.

It is worth noting in this experiment that the response variable is the same variable that is out of control. This is different from the situation where factors that are extraneous to the experiment are out of control and might be affecting the measurements of the response variable. When the response variable and the out-of-control variable are the same, there is the possibility of the results of the experiment being misleading if the factors that affect the response variable have not been identified. For example, what if the experiment is performed and the R^2 value is only .62? This means that 38 percent of the variability in the response measurements has not been explained. The question then must be addressed as to whether this is common cause variability or is there (unacceptable) variability due to assignable causes that have not been accounted for by the factors in the experiment. In general, if a model is wrong, the expected value of the effect estimates will generally not be equal to the effect that we are trying to estimate.

It is possible, although unlikely, that variability due to extraneous factors that are out of control could even go completely undetected. For example, process shifts could conceivably exactly coincide with changes in factor levels, so that changes in the values of the response variable could be erroneously attributed to changes in the levels of the factors in the experiment. Of course this is very unlikely to occur exactly, but something approximating this scenario could occur and ruin an experiment.

There were five factors in the experiment under discussion: room temperature of the laboratory (three levels), sharpness of the counterelectrode tip (two levels), the boron nitride disk (two levels—new and used), cleanliness of the entrance window (clean and in use for a week), and placement of the sample with respect to the boron nitride disk.

The 72 runs in the experiment were blocked on temperature, with temperature changes made 8 hours before the experimental runs were made at each level of temperature. The 24 runs made within each block were completely randomized. It was necessary to block on temperature because it was not possible to change temperature between individual runs; that is, temperature was a "hard-to-change" (actually, impossible to change) factor, which occurs much more frequently than is probably realized.

Wang (1999) gave an application of a 2×3^7 design in 18 runs to Poisson data. This was a blackening experiment conducted in an electric company using a three-layer oven. Thirty masks from each layer in the oven were collected for examining the number of defects in each mask. The response variable was the number of defects summed over the 30 masks in each layer and Wang (1999) gave the data for the upper layer.

It would be reasonable to assume a Poisson distribution for the defects, although all assumptions should be tested. Using the average number of defects per mask would be one possible approach, assuming that the average is approximately normally distributed. This would not work very well, however, if the probability of a defect was extremely small. Another approach would be to transform the data, such as by using some form of a square root transformation.

Example 6.4 — Case Study

We consider the second experiment described by Sheesley (1985) as this has some interesting features. The experiment involved lead wires for incandescent lamps. The objective was to determine if lead wires produced by a new process would perform in a superior way, relative to the current process, during the lamp-making process in terms of feeding into the automatic equipment.

The experiment involved five factors as Sheesley (1985) considered replications to be the fifth factor. We don't normally consider replications to be a factor, but in this case that does seem appropriate because the replications were made over days,

TABLE 6.1 Lead Wire Experiment Data

Lead Type	Plant	Machine	Shift	Replications			
C	A	S	1	24.6	38.5	34.1	16.4
C	A	S	2	35.8	20.1	44.1	47.1
C	A	S	3	35.3	18.1	15.3	44.2
C	A	H	1	24.4	29.2	39.8	28.2
C	A	H	2	17.0	33.5	22.3	29.2
C	A	H	3	35.2	25.4	24.2	19.0
C	B	S	1	17.6	19.2	21.4	22.7
C	B	S	2	18.3	18.0	19.9	23.2
C	B	S	3	10.8	39.4	23.7	19.6
C	B	H	1	40.5	37.8	25.1	49.4
C	B	H	2	18.9	24.7	19.6	24.9
C	B	H	3	6.5	16.6	67.8	30.3
T	A	S	1	17.0	20.8	26.0	23.7
T	A	S	2	11.8	22.0	22.4	23.5
T	A	S	3	21.0	52.4	20.6	12.4
T	A	H	1	35.5	13.1	21.7	30.1
T	A	H	2	12.1	30.5	21.3	22.1
T	A	H	3	17.4	23.4	28.8	18.0
T	B	S	1	12.4	7.8	25.6	11.2
T	B	S	2	28.1	16.7	23.7	21.5
T	B	S	3	11.5	16.8	26.9	18.9
T	B	H	1	6.7	18.6	12.7	13.1
T	B	H	2	8.7	19.0	34.4	27.7
T	B	H	3	12.7	9.1	31.2	17.7

C and T denote the control and test lead types; A and B are the two plant types; S and T are the standard and high-speed machines, and 1, 2, 3 denote the three shifts, with the replications listed in order within the treatment combinations for the other four factors.

so we might view replications as being a proxy for days. The other four factors were shift (three different shifts), lead type (current process, new process), plant (A, B), and machine (standard, high speed). Thus, this is a $2^3 \times 3 \times 4$ design, with all factors fixed except replications. The response variable was the average number of leads missed per running hour. The data, 96 observations, are given in Table 6.1.

Of course the first step that we would take in analyzing such data would be to look for bad data since this occurs with almost all experiments. A dotplot of the data reveals one extreme data point, 67.8, which is probably a bad data point since this value is considerably greater than the sum of the other three replications in the 12th row in Table 6.1. Sheesley (1985) did not do any preliminary data analysis, so there is no discussion of this data point. Consequently, we will keep an eye on the influence of the data point during the analyses and, if necessary, suggest corrective action, that being the best that we can do at this point in time. (The 52.4 in the 15th row of Table 6.1 also looks somewhat suspicious since it is only slightly less than the sum of the other three numbers in that row.)

The output using the GLM command in MINITAB is given below, with the fitted model containing the largest three-factor interaction but no other three-factor interaction. (In the expected mean squares section, Q [number] denotes the fixed effect of the corresponding numbered term.)

General Linear Model: Y versus T, P, M, S, R

Factor	Type	Levels	Values
T	fixed	2	C, T
P	fixed	2	A, B
M	fixed	2	H, S
S	fixed	3	1, 2, 3
R	random	4	1, 2, 3, 4

Analysis of Variance for Y, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
T	1	1201.3	1201.3	1201.3	47.70	0.006
P	1	404.3	404.3	404.3	4.03	0.138
M	1	32.7	32.7	32.7	1.06	0.378
S	2	1.9	1.9	1.0	0.02	0.977
R	3	639.8	639.8	213.3	**	
T*P	1	7.5	7.5	7.5	0.07	0.796
T*M	1	48.5	48.5	48.5	0.44	0.511
T*S	2	172.4	172.4	86.2	0.78	0.464
T*R	3	75.6	75.6	25.2	0.23	0.877
P*M	1	208.9	208.9	208.9	1.88	0.175
P*S	2	14.2	14.2	7.1	0.06	0.938
P*R	3	300.9	300.9	100.3	0.90	0.445
M*S	2	227.0	227.0	113.5	1.02	0.366
M*R	3	92.2	92.2	30.7	0.28	0.842
S*R	6	251.6	251.6	41.9	0.38	0.890
T*P*M	1	291.9	291.9	291.9	2.63	0.110
Error	62	6880.8	6880.8	111.0		
Total	95	10851.3				

** Denominator of F-test is zero.

S = 10.5347   R-Sq = 36.59%   R-Sq(adj) = 2.84%

Unusual Observations for Y

Obs	Y	Fit	SE Fit	Residual	St Resid
33	39.4000	21.8333	6.2694	17.5667	2.07 R
39	52.4000	27.6708	6.2694	24.7292	2.92 R
60	67.8000	38.0750	6.2694	29.7250	3.51 R

R denotes an observation with a large standardized residual.

Expected Mean Squares, using Adjusted SS

	Source	Expected Mean Square for Each Term
1	T	(17) + 12.0000 (9) + Q[1, 6 , 7 , 8 , 16]
2	P	(17) + 12.0000 (12) + Q[2, 6 , 10 , 11 , 16]

```
 3   M       (17) + 12.0000 (14) + Q[3, 7 , 10 , 13 , 16]
 4   S       (17) +  8.0000 (15) + Q[4, 8 , 11 , 13]
 5   R       (17) +  8.0000 (15)+ 12.0000 (14) + 12.0000 (12)
                 + 12.0000 (9)+ 24.0000 (5)
 6   T*P     (17) + Q[6, 16]
 7   T*M     (17) + Q[7, 16]
 8   T*S     (17) + Q[8]
 9   T*R     (17) + 12.0000 (9)
10  P*M     (17) + Q[10, 16]
11  P*S     (17) + Q[11]
12  P*R     (17) + 12.0000 (12)
13  M*S     (17) + Q[13]
14  M*R     (17) + 12.0000 (14)
15  S*R     (17) + 8.0000 (15)
16  T*P*M   (17) + Q[16]
17  Error   (17)

Error Terms for Tests, using Adjusted SS


 Source Error DF Error MS Synthesis of Error MS
  1   T      3.00    25.2  (9)
  2   P      3.00   100.3  (12)
  3   M      3.00    30.7  (14)
  4   S      6.00    41.9  (15)
  5   R      3.05   *      (9) + (12) + (14)
                           + (15) - 3.0000 (17)
  6   T*P    62.00  111.0  (17)
  7   T*M    62.00  111.0  (17)
  8   T*S    62.00  111.0  (17)
  9   T*R    62.00  111.0  (17)
 10   P*M    62.00  111.0  (17)
 11   P*S    62.00  111.0  (17)
 12   P*R    62.00  111.0  (17)
 13   M*S    62.00  111.0  (17)
 14   M*R    62.00  111.0  (17)
 15   S*R    62.00  111.0  (17)
 16   T*P*M  62.00  111.0  (17)


 Variance Components, using Adjusted SS

          Estimated
 Source     Value
 R           14.502
 T*R         -7.150
 P*R         -0.891
 M*R         -6.687
 S*R         -8.631
 Error      110.981
```

There are some important observations that should be made from these results:

(1) The Lead Type (T) factor is the only one that is close to being significant, and it is highly significant.

(2) The denominator of the F-test for testing the significance of the Replications factor is given as zero. (This will be explained shortly.)

(3) The R^2 for the model is low since Lead Type is the only significant factor.

(4) Four estimated variance components are negative. A variance component cannot, by definition, be negative. Special estimation methods must be employed to prevent negative estimates, which will otherwise frequently occur when effects are far from being significant.

(5) The two suspicious observations that were mentioned previously have very large standardized residuals, 3.51 and 2.92. Of course a data point is an outlier relative to the model that is used and here the model has many non-significant terms. The standardized residuals are even larger, however (4.03 and 3.20, respectively), when the model contains only Lead Type. Action would normally be taken regarding such extreme points. In this case, however, there are 96 data points so one or two extreme points are not likely to have much effect, especially since none of the effects are close to the dividing line between significance and non-significance, although two of the three-factor interactions are larger than we might prefer.

The message from the computer output that the denominator degrees of freedom is zero for testing the Replications factor undoubtedly looks strange. This can be explained as follows. There are not always exact tests for effects; oftentimes a pseudo-error term must be created by using Satterthwaite's (1946) procedure, which provides an approximate test rather than one based on distribution theory, and sometimes strange things can happen when this is done.

Satterthwaite's procedure uses linear combinations of mean squares. To understand the mechanics, consider the discussion of expected mean squares (EMSs) for a 2^2design in Appendix C of Chapter 4. It was stated that $E(\mathrm{MS}_{AB}) = \sigma^2 + 2\,\sigma^2_{AB}$ and $E(\mathrm{MS_E}) = \sigma^2$. To test the hypothesis that $\sigma^2_{AB} = 0$, we thus construct an F-test as $F = \mathrm{MS}_{AB}/\mathrm{MS_E}$, so that if $\sigma^2_{AB} = 0$, the numerator and denominator are estimating the same thing. In general, the expected value of the numerator minus the expected value of the denominator should be equal to a multiple of what is being tested. Now assume that we have a design for which it is not possible to accomplish this using one mean square in the numerator and one in the denominator, but rather a linear combination is necessary for both the numerator and the denominator. These linear combinations should be constructed using the guiding principle that the expected value of the numerator minus the expected value of the denominator is equal to a multiple of what is to be tested, as stated.

It will generally give a fractional degrees of freedom, which in this case is 0.49, and is treated as zero by MINITAB. PROC GLM in SAS also uses Satterthwaite's procedure.

Since there is obviously no shift effect and the Replications factor is not significant, we would be tempted to analyze the other three factors as having come from a 2^3 design. Although such ad hocery is not recommended, in general, in this case the change in the conclusion is that the plant effect is borderline significant with such an analysis.

We will illustrate Satterthwaite's procedure by using an example that is simpler than the one just used. Specifically, one of the sample datasets that comes with MINITAB is EXH_AOV, which contains sets of columns of data of unequal length that are used in the ANOVA online help. Four of those columns constitute a useful example for illustrating Satterthwaite's procedure, however. There are two factors at three levels (one fixed and one random) and one fixed factor at two levels. So it is a 2×3^2 design with Thickness being the response variable and Time, Operator, and Setting the factors. The MINITAB output is given below.

```
ANOVA: Thickness versus Time, Operator, Setting

Factor     Type     Levels   Values
Time       fixed    2        1, 2
Operator   random   3        1, 2, 3
Setting    fixed    3        35, 44, 52

 Analysis of Variance for Thickness

 Source                 DF        SS       MS       F    P
 Time                    1       9.0      9.0    0.29  0.644
 Operator                2    1120.9    560.4    4.91  0.090 x
 Setting                 2   15676.4   7838.2   73.18  0.001
 Time*Operator           2      62.0     31.0    1.29  0.369
 Time*Setting            2     114.5     57.3    2.39  0.208
 Operator*Setting        4     428.4    107.1    4.46  0.088
 Time*Operator*Setting   4      96.0     24.0    7.08  0.001
 Error                  18      61.0      3.4
 Total                  35   17568.2

 x Not an exact F-test.

 S = 1.84089 R-Sq = 99.65% R-Sq(adj) = 99.32%

                               Variance   Error   Expected Mean Square for Each
    Source                     component   term   Term (using unrestricted model)
1   Time                                    4     (8) + 2 (7) + 6 (4) + Q [1,5]
2   Operator                      37.194    *     (8) + 2 (7) + 4 (6) + 6 (4)+12(2)
3   Setting                                 6     (8) + 2 (7) + 4 (6) + Q[3,5]
4   Time*Operator                  1.167    7     (8) + 2 (7) + 6 (4)
5   Time*Setting                            7     (8) + 2 (7) + Q[5]
6   Operator*Setting              20.778    7     (8) + 2 (7) + 4 (6)
7   Time*Operator*Setting         10.306    8     (8) + 2 (7)
8   Error                          3.389          (8)

*Synthesized Test.

 Error Terms for Synthesized Tests
                                             Synthesis of
 Source        Error DF    Error MS    Error MS
 2 Operator    3.73        114.1       (4) + (6) - (7)
```

Consider the EMSs. Notice that for each effect except the second one (i.e., Operator), it is possible to find an effect whose EMS is a subset of the EMS of the effect that is to be tested, with the only difference being a term in the effect that is being tested. For example, we can see that Time should be tested against Time*Operator; the latter should be tested against Time*Operator*Setting, and so on. We need to construct a denominator that contains every term in EMS(Operator) except $\sigma^2_{\text{operator}}$. We see that we can accomplish this by using Time*Operator + Operator*Setting − Time*Operator*Setting. Thus, this combination of mean squares forms the denominator for testing the Operator effect, realizing that this is just an approximate test.

Example 6.5

One of the sample datasets, PANCAKE.MTW, which comes with the MINITAB software is data from a 2 × 4 experiment with three replications conducted to study the effect of two factors on the quality of pancakes. The two factors are supplement (present or absent) and four levels of whey. Obviously, the first factor is fixed and we will also assume that the second factor is fixed. Three experts were asked to rate the quality of the pancake and the average of the three ratings was used as the value of the response variable. This process was performed three times for each treatment combination, providing a total of 24 observations.

The analysis of the data is as follows

```
ANOVA: Quality versus Supplement, Whey

Factor       Type   Levels  Values
Supplement   fixed     2    1, 2
Whey         fixed     4    0, 10, 20, 30

Analysis of Variance for Quality

Source       DF       SS      MS        F      P
Supplement    1   0.5104  0.5104     2.31  0.145
Whey          3   6.6912  2.2304    10.08  0.000
Error        19   4.2046  0.2213
Total        23  11.4062

S = 0.470419   R-Sq = 63.14%   R-Sq(adj) = 55.38%
```

The four levels of the whey factor permit the fitting of second- and higher-order effects, but the plot of the response variable against whey in Figure 6.2 shows no evidence of a nonlinear effect.

Although the scatterplot does not show any evidence of a nonlinear relationship between the response variable and whey, the clear separation that results in two distinct groups of three values at the first two values of whey is something for which an explanation should be sought. This separation tends to reduce the R^2 value, which is low partly because of the separation and partly because the supplement factor is not significant.

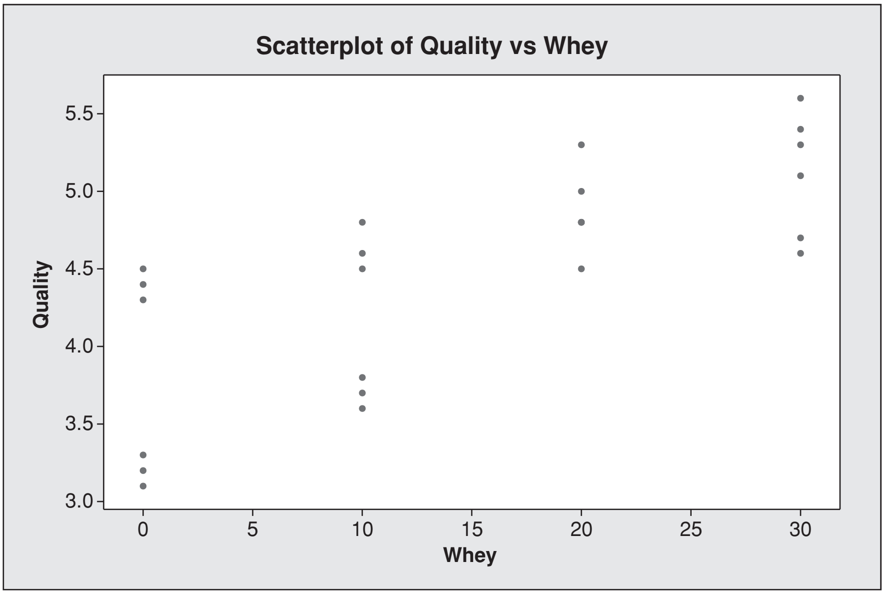

Figure 6.2 Scatterplot of the pancake experiment data.

6.5 MIXED FRACTIONAL FACTORIALS

Because of the number of runs required for a mixed full factorial, a mixed fractional factorial is often used. For example, a $2^{3-1}3^{3-1}$ design was used in the case study of Hale-Bennett and Lin (1997). (We won't analyze the data from that study here because the individual data values were not given.)

Such a design can of course be viewed analogous to the way that a $2^2 3^2$ is viewed; that is, the $2^{3-1}3^{3-1}$ design can be viewed as a 2^{3-1} design at each point of a 3^{3-1} design (or the reverse). Notice that 216 design points would be required if fractionation were not used, whereas only 36 points are needed for the mixed fractional factorial.

Although tables and software are of course available for constructing such designs, the $2^{3-1}3^{3-1}$ design could be easily constructed by hand. First, a 3^{3-1} design would be constructed by confounding one component of the A^*B^*C interaction. Of course we know that main effects are confounded with two-factor interactions in the 2^{3-1} design since $I = ABC$.

When we concatenate the two designs we can think of the combined design as a replicated 3^{3-1} design and as a replicated 2^{3-1} design. Replication has no effect on estimability, so the estimability for the separate designs is the same as that for the combined design, and the alias structure cannot be more palatable than it is for the separate designs.

The bottom line is that 36 design points provide for the estimation of the six main effects plus certain two-factor interactions, and the main effects are estimable only if two-factor interactions for the two-level factors and the appropriate two-factor interaction components for the three-level factors do not exist. In general, fractions of three-level designs and designs that include fractions of three-level designs are not particularly good designs. (A related discussion of fractions of the three-level factorials is given by Giesbrecht and Gumpertz (2004, pp. 384–385)).

One problem with designs like the $2^{3-1}3^{3-1}$ design is that the interaction components eat up multiple degrees of freedom (as contrasted with two-level designs for which each interaction has one degree of freedom). When these degrees of freedom are added, the total and hence the run size become at least moderate in size.

In addition to the case study of Hale-Bennett and Lin (1997), a detailed analysis of the design and resultant data are given by Wu and Hamada (2000, pp. 271–278), and the reader is referred to these sources for further details.

Designs with differing numbers of levels are generally presented in the literature as orthogonal arrays (OAs), and many tables of OAs are available. This of course encourages the blind use of these designs, which is certainly undesirable, although the alias structure for incomplete factorial designs with more than two levels is generally complex. Consequently, most practitioners would probably be better off avoiding it, but not completely avoiding an understanding of OAs and their strengths and weaknesses.

Some tables of $4^m 2^{k-p}$ designs are given at http://iems.northwestern.edu/~bea/articles/Appendix.PDF.

6.6 ORTHOGONAL ARRAYS WITH MIXED LEVELS

It was mentioned at the beginning of Chapter 5 that fractional factorial designs are OAs, but not all OAs are fractional factorials. Specifically, not all OAs are regular designs, meaning that not all effects to be estimated can be estimated independently.

Xu, Cheng, and Wu (2004) give an application of an OA with mixed levels that is not a regular design, using experimental data originally given by King and Allen (1987). The objective of the experiment described in the latter had radio frequency choke as the response variable and the objectives were to identify the factors that affected the choke readings and to identify the best settings of those factors. (Recall the definitions of a "regular design" and a nonregular design given at the start of Chapter 5. For each pair of factorial effects in a regular design, the effects are either independent or completely aliased. Designs for which this condition is not met are nonregular designs.)

There was one 2-level factor and seven 3-level factors in 18 runs, with each run replicated twice. The design is given below. (*Note*: This design can be generated with various statistical software packages. MINITAB was used to create this design, creating it as an L18 Taguchi design. Such designs are discussed in Chapter 8.)

A	B	C	D	E	F	G	H
0	0	0	0	0	0	0	0
0	0	1	1	1	1	1	1
0	0	2	2	2	2	2	2
0	1	0	0	1	1	2	2
0	1	1	1	2	2	0	0
0	1	2	2	0	0	1	1
0	2	0	1	0	2	1	2
0	2	1	2	1	0	2	0
0	2	2	0	2	1	0	1
1	0	0	2	2	1	1	0
1	0	1	0	0	2	2	1
1	0	2	1	1	0	0	2
1	1	0	1	2	0	2	1
1	1	1	2	0	1	0	2
1	1	2	0	1	2	1	0
1	2	0	2	1	2	0	1
1	2	1	0	2	0	1	2
1	2	2	1	0	1	2	0

How do we know this could not be a $2 \times 3^{6-4}$ design and thus be a regular design? In order for this to be true, a 3^{6-4} design in the 6 three-level factors would have to be crossed with the two levels of the other factor, which means that there would have to be replicate rows for those six factors. We can see that there are no such replicate rows, so it could not be this design. Since there is no other way to form 18 runs with one 2-level factor and six 3-level factors using a factorial/fractional factorial representation, the design cannot be a mixed fractional factorial or a combination of a factor with a fractional factorial.

We can also verify that the design is not a regular design by computing correlations between effect estimates. The 17 degrees of freedom for estimating effects could be used to estimate the linear effect of the two-level factor and the linear and quadratic effects of the three-level factors, with four df left, which could be used to estimate one two-factor interaction involving three-level factors. Or, interactions could be estimated among the factors in lieu of estimating quadratic effects. The correlation between C and the GH interaction is .50 and there are many other correlations between 0 and 1 since this is a nonregular design.

Since interactions are not orthogonal to main effects with this design but interaction terms might be more likely to be real than pure quadratic effects for these data, the question arises as to how to proceed. As when a supersaturated design (Section 13.4.2) is used, a logical strategy is to use some variable selection approach such as stepwise regression to identify what seem to be the most important effects and then estimate those effects.

Xu et al. (2004) used an alternative two-stage approach, however, first performing only a main effects analysis and identifying four factors, B, E, G, and H, as being significant at the .01 level. At this point, stepwise regression might have been applied to all possible terms in those four factors, with the available degrees of freedom used to estimate the effects that were selected.

Xu et al. (2004) did not use that approach, however, but rather fit the full second-order model in the four factors, although there would not be enough degrees of freedom to do that if an ANOVA approach were taken. There *are* enough degrees of freedom, however, if only the linear $\times$ linear component of each interaction is estimated. Significant effects were then identified through the use of t-tests and the effects that were identified were B, E, G, H, BE, EG, EH, GH, and E^2.

This approach is quite similar to the two-stage approach that is used in response surface methodology (Chapter 10) for model identification, except that in the latter two designs (i.e., two experiments) are used instead of one. Cheng and Wu (2001) argued that their two-stage analysis is a useful alternative to this response surface methodology approach. This is discussed further in Section 10.1.

There are potential problems with either the stepwise regression or the t-test approach, however, as stepwise regression can be undermined by correlations between variables, as discussed in Section 13.4.2, and correlations can also cause t-test results to be misleading.

Since first-order effects generally dominate second-order effects in the same factors, the Cheng and Wu strategy may be expected to work well, in general. There could be problems, however, when main effects are moderately correlated with two-factor interactions. With this example, the GH interaction was found by Xu et al. (2004) to be significant at the .01 level. The correlation between that effect estimate and the C effect estimate is 0.5, however, which raises the question of how much of the GH interaction is due to the C effect. We cannot answer questions of this type without running another experiment unless we employ methods such as those of Chevan and Sutherland (1991) to obtain a measure of the "independent effects" of effects that are correlated.

The data for this experiment are given in Exercise 6.25 and the reader is asked to analyze it there. An algorithm for constructing orthogonal and nearly orthogonal arrays with mixed levels was given by Xu (2002).

6.7 MINIMUM ABERRATION DESIGNS AND MINIMUM CONFOUNDED EFFECTS DESIGNS

Minimum aberration designs and minimum confounded effects designs were discussed in Section 5.7 in the context of 2^{k-p} designs. We can also discuss such designs when there are more than two levels and when the number of levels is mixed. The discussion must necessarily be different, however, for the following reason. With 2^{k-p} designs, each interaction has one degree of freedom, so the interaction is either estimable or not estimable, whereas, as has been explained, this is not the case with a 3^{k-p} design. Just as with the 2^{k-p} series, a minimum aberration 3^{k-p} design isn't necessarily the best design to use. In particular, Sun and Wu (1994) explain why designs that are not minimum aberration may be preferred for $k = 6$, 7, and 8.

6.8 FOUR OR MORE LEVELS

Not infrequently, a multiple-factor design is used in which at least one of the factors has four levels. This will generally occur when the four-level factor(s) are qualitative. Of course the required number of runs of a full factorial may be prohibitive if there are very many factors and some of them have four levels. Consequently, a fractional mixed factorial design or an orthogonal main effect design may have to be used. Lorenzen (1993) showed how a variety of orthogonal main effect plans could be easily constructed, including those with four or more levels. By using the catalog of 33 designs combined with some simple rules, a total of 7172 distinct orthogonal main effect plans could be created that have up to six levels and 50 runs.

As discussed by Wu and Hamada (2000, p. 258), the simplest way to construct a design with a mixture of two-level and four-level factors that is not a full factorial is to start with a saturated 2^{k-p} design and use the *method of replacement*. For each four-level factor this entails replacing three of the two-level factors by using the four combinations to designate the levels of the four-level factor. (The third factor is a product of the first two in their example; hence there are four combinations, not eight. Viewed differently, a 2^{k-p} design contains a replicated 2^{k-p-1} design.) If the construction began with the 2^{7-4} design, the resultant design is a 4-level factor combined with a 2^{4-1}. This method has been referred to as using *pseudofactors* by Giesbrecht and Gumpertz (2004).

Full factorial designs in four or more factors can be easily created with almost all statistical software packages; fractional factorials are another matter. The construction of fractional factorials and various types of custom designs by JMP are discussed in the next section.

Assume that we wish to construct a 4^{3-1} design, which of course would have 16 runs. JMP produces the following design when the user selects 16 runs:

A	B	C
0	0	0
0	1	2
0	2	1
0	3	3
1	0	3
1	1	0
1	2	2
1	3	1
2	0	1
2	1	0
2	2	2
2	3	3
3	0	2
3	1	1
3	2	3
3	3	0

As we would expect, there is a full factorial in A and B. We note that the levels for C cannot be obtained from the levels of A and B, however, unlike the case with two-level and three-level fractional factorials. That is, we cannot use $C = AB^s$ for $s = 1$, 2, or 3 and obtain the levels of C given in the design above. Does this mean that a 4^{3-1} design cannot be constructed? It can't be done with JMP. The default option for three four-level factors is 16 design points. When that option is used, JMP generates a D-optimal design (see Section 13.7 for a discussion of various types of optimal designs). A similar problem is encountered when Design-Expert is used, as no 4^{3-1} design can be constructed, but a D-optimal design with 41 runs is constructed when the D-optimal design option is selected.

It is possible to construct a 4^{3-1} design, however, using the same general approach as is used in constructing 2^{k-p} and 3^{k-p} designs. If we let $C = AB$, we obtain the following design:

A	B	C
0	0	0
0	1	1
0	2	2
0	3	3
1	0	1
1	1	2
1	2	3
1	3	0
2	0	2
2	1	3
2	2	0
2	3	1
3	0	3
3	1	0
3	2	1
3	3	2

It can be observed that $A + B + 3C = 0 \pmod 4$ for every combination so this is one of the four fractions (the others summing to 1, 2, and 3). Multiplying both sides of $C = AB$ by C^3, we obtain $I = ABC^3$ as *part* of the defining relation. From this point the development does not parallel the development for 2^{k-p} and 3^{k-p} designs because 4 is not a prime number. For notational simplification, let $A + B + 3C = 0 \pmod 4$ be denoted by (113). Then the rest of the defining relation is obtained as 2(113), mod 4, and 3(113), mod 4. This produces the set (113), (222), and (331), so the entire defining relation is $I = ABC^3 = A^2B^2C^2 = A^3B^3C$. (Note that the exponent of A is allowed to be something other than 1 resulting from the fact that 4 is not a prime number.) The reader can verify that each of the last two components of the defining relation also give a sum of zero, mod 4, when applied to each of the 16 treatment combinations.

The (complicated) alias structure is then easily obtained from this defining relation and part of it is as follows:

$$
\begin{aligned}
A &= A^2BC^3 = A^3B^2C^2 = B^3C \\
A^2 &= A^3BC^3 = B^2C^2 = AB^3C \\
A^3 &= BC^3 = AB^2C^2 = A^2B^3C \\
B &= AB^2C^3 = A^2B^3C^2 = A^3C \\
B^2 &= AB^3C^3 = A^2C^2 = A^3BC \\
B^3 &= A^2C^3 = A^2BC^2 = A^3B^2C \\
C &= AB = A^2B^2C^3 = A^3B^3C^2 \\
C^2 &= ABC = A^2B^2 = A^3B^3C^3 \\
C^3 &= ABC^2 = A^2B^2C = A^3B^3 \\
AC &= A^2B = A^3B^2C^3 = B^3C^2 \\
BC &= AB^2 = A^2B^3C^3 = A^3C^2
\end{aligned}
$$

With 16 runs we, of course, have 15 degrees of freedom for estimating effects. Each main effect has three degrees of freedom and there are three factors. This leaves six df for estimating two-factor interactions, but the components of these interactions are so badly confounded that estimating them would generally be unwise. Although the main effects can be estimated, this is contingent upon the interaction components with which they are confounded not being significant.

It is also possible, of course, to construct designs for factors with five or more levels. Fractional factorials with five levels are easier to deal with than are fractional factorials with six levels because 5 is a prime number whereas 6 is not. The construction of an r^{k-p} design with r a prime number of at least 5 is discussed by Giesbrecht and Gumpertz (2004, p. 282).

6.9 SOFTWARE

Virtually all software packages with experimental design capability can be used to construct designs with more than two levels, including mixed factorials. For example, the "General Full Factorial Design" capability in MINITAB can be used to create full factorial designs with any number of levels up to 100, provided that the number of factors does not exceed 15 and the number of runs does not exceed 100,000. Similarly, the "General Factorial Design" capability in Design-Expert can be used to create a design with as many as 999 levels but the number of factors cannot exceed 12.

Neither MINITAB nor Design-Expert can be used to generate a fraction of a design with more than two levels, however, which means that neither a 3^{k-p} design nor a fraction of a mixed factorial can be generated. Design-Expert will, however, create a design that is D-optimal. For example, if five 3-level factors and two 2-level factors are specified, Design-Expert creates a design with 79 points. Of course there is no way to obtain that number of design points by fractionation since 79 is a prime number. The alias structure for the design is, as one might guess, extremely complex. Correlations between effect estimates can be obtained if that option is selected.

JMP has limited capability for generating 3^{k-p} designs, with only designs that are Taguchi OAs being available. (The latter are discussed in Chapter 8.) For example, a 3^{3-1} design can be selected, which is listed as an L_9 Taguchi design. If we convert the levels given in the JMP output to the level designation used in Section 6.1.1, we have the design as follows:

A	B	C
0	0	0
0	1	1
0	2	2
1	0	1
1	1	2
1	2	0
2	0	2
2	1	0
2	2	1

Using modulo arithmetic as in Section 6.1.1, we can see that $C = AB$ since $A + B = C$, mod 3, with AB being one of the two components of the A^*B interaction. The eight degrees of freedom that are available for estimating effects can thus be used to estimate A (2 df), B (2 df), C (2 df), and AB^2 (2 df).

The defining contrast is $I = ABC^2$, which can be obtained by multiplying each side of the equation $C = AB$ by C^2, and it can be observed that $A + B + 2C = 0$ (mod 3), for each of the nine treatment combinations. The design is thus resolution III, with main effects confounded with two-factor interaction components. The full alias structure (not given by JMP) is thus as follows:

$$A = ABC^2 = BC^2$$

$$B = AB^2C^2 = AC^2$$

$$C = AB = ABC$$

$$AC = AB^2 = BC$$

There would have been 26 df for estimating effects if the 3^3 design had been used; we can see how the 26 df are broken down when the 3^{3-1} design is used. Two df are of course confounded with the mean, with the remaining 24 df confounded in four sets of three, as shown above, with each component in each set of three having 2 df.

Thus, one component of the A^*B^*C interaction is confounded with the estimate of the mean, and the other three components are confounded with main effect estimates, which are each confounded with a component of a two-factor interaction.

A 3^{4-1} design and a 3^{4-2} design can also be constructed by JMP. These are listed as the L_{27} and the L_9, respectively, but only the L_{27} is available when there are more than four factors. Somewhat similarly, a 3^{4-2} design can also be constructed as an L_9, using Design-Expert.

There are many 3^{k-p} designs in existence, as listed in Wu and Hamada (2000, pp. 250–255); see also Chen et al. (1993). Design-Expert can generate two other $3^{k-\text{p}}$ designs as OA*s*, which we will write in OA notation as the $L_{27}(3^{13})$ and $L_{27}(3^{22})$.

The 3^{4-2} design is as follows, again converting the JMP and Design-Expert notation into the notation used in the preceding example:

A	B	C	D
0	0	0	0
0	1	1	1
0	2	2	2
1	0	1	2
1	1	2	0
1	2	0	1
2	0	2	1
2	1	0	2
2	2	1	0

As in the construction of 2^{k-p} designs, we can see that there is a full 3^2 factorial in the first two factors. We can also observe that $C = AB$ and $D = AC$. Therefore, $I = ABC^2 = ACD^2 = AB^2D = BCD$, with the third component obtained as (ABC^2) (ACD^2) and the last component obtained as (ABC^2) $(ACD^2)^2$. From this defining relation, the full alias structure can be obtained, as the reader is asked to obtain in Exercise 6.17.

JMP also has very limited capabilities for fractions of mixed factorials that are cataloged as screening designs. Specifically, only four fractions of mixed factorials are available as screening designs in JMP, and these are listed as OAs: (1) the L_{18} array due to Peter John that can be used for at most one 2-level factor and up to seven 3-level factors, (2) the L_{18} array due to Chakravarty that can be used for up to three 2-level factors and six 3-level factors, (3) the L_{18} array due to Hunter that can be used for eight 2-level factors and four 3-level factors, and (4) the L_{36} array that can be used for eleven 2-level factors and twelve 3-level factors. These designs are all orthogonal, although all of them are not balanced. For example, an L_{18} array for five 3-level factors and two 2-level factors cannot be balanced because the four combinations of the two-level factors cannot all occur an equal number of times because $18/4$ is not an integer.

(It is possible to create balanced designs in JMP, however, such as by using the "Custom Design" option and selecting "Default.") It is also possible to use the custom design capability in JMP to create D-optimal designs (covered in Section 13.7) for certain numbers of design points.

Design-Expert has greater capability than JMP for constructing fractions of mixed factorials. For example, for a mixture of two-level and three-level factors, Design-

Expert will also construct a $L_{36}(2^{11} \times 3^{12})$, $L_{36}(2^3 \times 3^{13})$, and $L_{54}(2 \times 3^{25})$. There is also a $L_{32}(2 \times 4^9)$ design that can be used for one factor at two levels and up to nine factors at four levels, and an $L_{50}(2 \times 5^{11})$ that can be used for one 2-level factor and up to eleven 5-level factors.

D. o. E. Fusion Pro has some capability for three-level and mixed level designs that are not full factorials, although in the case of the latter the properties of the design are not given. More specifically, only 3^k and mixed full factorial designs are available unless one selects the “Design Navigator Wizard” option and lets the design be generated automatically—without even specifying the number of design points. When the user lists five categorical factors, three at three levels and two at two levels, the following design is generated when no points are to be replicated and there is to be no blocking.

```
 Experiment Design - Experiment 1

          Var. A   Var. B   Var. C   Var. D   Var. E
Run No.
1           A        A        C        B        A
2           B        A        A        C        B
3           B        A        B        B        A
4           B        B        C        A        A
5           A        B        A        B        B
6           C        B        B        B        A
7           A        B        B        C        A
8           A        A        A        A        A
9           C        A        A        A        A
10          C        B        C        C        B
11          A        A        B        A        B
```

The User’s Guide gives the following explanation as to how the design points are selected: “D.o.E. FUSION generates model-robust designs based on a combination of D-, A-, G-, and V-Optimality. Its algorithms and approach result in a design that is not strictly optimal according to any one letter goal, but meets the requirements of (1) good coverage of the design space, including the interior, (2) low predictive variances of the design points, and (3) low model term coefficient estimation errors.” Thus, although the G-efficiency of the generated designs is given, and indeed is given when the user selects “Design Menu Wizard” and then selects “mixed level,” this is *not* the (sole) criterion under which a design is constructed. (Of course the G-efficiency is given as 100 when a mixed full factorial is generated.)

Although it is nice to have designs generated automatically, the user needs to know more than the G-efficiency. Specifically, the alias structure should be known, at least generally since alias structures for three-level and mixed-level designs are not particularly tidy, so viewing an entire alias set when there are more than a few factors could be more confusing than enlightening.

Perhaps there is some commercial software that will generate 4^{k-p} designs, but if so, I am not aware of it.

6.10 CATALOG OF DESIGNS

Since not all statistical software packages have the capability for 3^{k-p} designs and fractions of mixed factorials, some users may have to resort to catalogs of these designs. The downside of doing so, however, is that the catalogs are not constructed in such a way that all of the factor-level combinations are listed, so that cutting and pasting won't work. Consequently, the designs would have to be entered manually, thus leading to the possibility of errors.

Nevertheless, the catalogs can be quite useful. One such catalog, by Xu (2005), lists three-level fractional factorial designs (available at http://www.stat.ucla.edu/~hqxu/pub/ffd/ffd3a.pdf). Designs with 27, 81, 243, and 729 runs are given. This is a more extensive list of designs than were given in Wu and Hamada (2000) and Chen et al. (1993). In particular, the designs listed in Wu and Hamada (2000) do not go beyond 81 runs.

6.11 SUMMARY

It will frequently be necessary to use a design with more than two levels, especially when at least one of the factors is qualitative rather than quantitative. One impediment to the use of designs given in this chapter and elsewhere is that software for constructing many of the designs is not readily available. In particular, there is hardly any statistical software that will generate a wide variety of 3^{n-k} designs. Consequently, catalogs of designs such as were mentioned in Section 6.10 take on greater importance than would be the case if software packages were readily available.

The interpretation of computer output for designs with more than two levels is complicated by the fact that interaction components can be decomposed into either trend components or *I–J* components, and the latter have no physical meaning. Alias structures are also rather complicated for fractional factorial designs with three or more levels.

A conditional effects analysis should be performed when there are large interactions, just as when a 2^{k-p} design is used.

REFERENCES

Addelman, S. (1962). Orthogonal main effect plans for asymmetrical fractional factorial experiments. *Technometrics*, **4**(1), 47–58.

Bisgaard, S. (1997). Why three-level designs are not so useful for technological applications. *Quality Engineering*, **9**(3), 545–550.

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

Buckner, J., B. L. Chin, and J. Henri (1997). Prometrix RS35e gauge study in five two-level factors and one three-level factor. In *Statistical Case Studies for Industrial Process Improvement*, Chap. 2 (V. Czitrom and P. D. Spagon, eds.). Philadelphia:

Society for Industrial and Applied Mathematics, and Alexandria, VA: American Statistical Association.

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.

Cheng, S.-W. and C. F. J. Wu (2001). Factor screening and response surface exploration. *Statistica Sinica*, **11**, 553–580; discussion: 581–604. (available online at http://www3.stat.sinica.edu.tw/statistica/oldpdf/A11n31.pdf)

Chevan, A. and M. Sutherland (1991). Hierarchical partitioning. *The American Statistician*, **45**, 90–96.

Clark, J. B. (1999). Response surface modeling for resistance welding. In *Annual Quality Congress Proceedings*, American Society for Quality, Milwaukee, WI.

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

Gupte, M. and P. Kulkarni (2003). A study of antifungal antibiotic production by *Thermomonospora* sp MTCC 3340 using full factorial design. *Journal of Chemical Technology and Biotechnology*, **78**(6), 605–610.

Hale-Bennett, C. and D. K. J. Lin (1997). From SPC to DOE: A case study at Meco, Inc. *Quality Engineering*, **9**(3), 489–502.

Heaney, M. D., W. A. Lidy, E. G. Rightor, and C. G. Barnes (2000). Analysis of a difficult factorial designed experiment from polyurethane product research. *Quality Engineering*, **12**(3), 425–438.

Hinkelmann, K. and O. Kempthorne (2005). *Design and Analysis of Experiments, Volume 2: Advanced Experimental Design*. Hoboken, NJ: Wiley.

Inman, J., J, Ledolter, R. V. Lenth, and L. Niemi (1992). Two case studies involving an optical emission spectrometer. *Journal of Quality Technology*, **24**(1), 27–36.

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

King, C. and L. Allen (1987). Optimization of winding operation for radio frequency chokes. In *Fifth Symposium on Taguchi Methods*, Dearborn, MI: American Supplier Institute, pp. 67–80.

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

Lorenzen, T. J. (1993). Making orthogonal main effect designs useful. General Motors Research Report 8025. General Motors Research and Development Center, Warren, MI.

Peyton, B. M. and W. G. Characklis (1993). A statistical analysis of the effect of substrate utilization and shear stress on the kinetics of biofilm detachment. *Biotechnology and Bioengineering*, **41**(7), 728–735.

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

Ryan, T. P. (1981). Relationships between the trend components and I-J components of interaction in a 3^2 design. Unpublished manuscript.

Ryan, T. P. (1997). *Modern Regression Methods*. New York: Wiley.

Sahin, K. and M. Örnderci (2002). Optimal dietary concentrations of Vitamin C and chromium for alleviating the effect of low ambient temperature on serum insulin, corticosterone, and some blood metabolites in laying hens. *The Journal of Trace Elements in Experimental Medicine*, **15**(3), 153–161.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, **2**, 110–114.

Sheesley, J. H. (1985). Use of factorial designs in the development of lighting products. In *Experiments in Industry: Design, Analysis, and Interpretation of Results*, pp. 47–57. (R. D. Snee, L. B. Hare, and J. R. Trout, eds.). Milwaukee, WI: Quality Press.

Sun, D. X. and C. F. J. Wu (1994). Interaction graphs for three-level fractional factorial designs. *Journal of Quality Technology*, **26**, 297–307.

Taguchi, G. (1987). *System of Experimental Design*. White Plains, NY: UNIPUB/Kraus International Publications.

Vázquez, M. and A. M. Martin (1998). Optimization of Phaffia rhodozyma continuous culture through response surface methodology. *Biotechnology and Bioengineering*, **57**(3), 314–320.

Wang, P. C. (1999). Comparisons of the analysis of Poisson data. *Quality and Reliability Engineering International*, **15**, 379–383.

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

Xu, H. (2002). An algorithm for constructing orthogonal and nearly-orthogonal arrays with mixed levels and a small number of runs. *Technometrics*, **44**, 356–368.

Xu, H. (2005). A catalogue of three-level fractional factorial designs. *Metrika*, **62**, 259–281. (preprint available online at http://www.stat.ucla.edu/∼hqxu/pub/ffd/ffd3a.pdf and see the online supplement (catalogs) at http://www.stat.ucla.edu/∼hqxu/pub/ffd)

Xu, H., S. W. Cheng, and C. F. J. Wu (2004). Optimal projective three-level designs for factor screening and interaction detection. *Technometrics*, **46**, 280– 292.

EXERCISES

6.1 After reading a few articles on experimental design, but not really being knowledgeable in the subject, a scientist proclaims that he is going to use a mixed factorial: a 1×3 design. Explain why this is really not a mixed factorial. What type of design is it? Would it be of any practical value if the design were unreplicated? Explain. Would your answer be the same if the design were replicated? Explain.

6.2 Buckner, Chin, and Henri (1997, references) described an experiment that investigated the effects of one 3-level factor and five 2-level factors in 16 runs. The design was a modification of a 2^{6-2} design to accommodate the three-level factor and would be properly called an "Addelman design." It is interesting to note that the number of design points was determined from a calculation such as is performed using Eq. (1.1), although the authors used a different value for α and a different value for the power of the test.

(a) What can be said about the levels of the three-level factor?

(b) Two response variables were examined in an ANOVA analysis of the data and none of the six factors was even close to being significant for either response variable. Regarding this, the authors stated, “The Addleman [sic] design has a more complex confounding scheme than ordinary fractional factorials, but since no factors appeared important, confounding was not considered in this application.” Do you agree with that position?

(c) For the response variable, Uniformity, the sum of squares for one of the factors, Prior Film, was given as 0.000000000. Since the values for this response variable were all two-digit numbers, what is the only way that this number could be correct? Would you be inclined to suspect that it may be incorrect? Explain.

6.3 Derive the alias string for the 3^{3-1} design that was discussed in Section 6.3.4, with part of the alias string given in that section.

6.4 It was shown in Section 4.2.1 that it will generally be a good idea to construct interaction plots for two-level factors in such a way that each of two factors involved in a two-factor interaction is, in turn, the horizontal axis label, since the message of the extent of the interaction can be quite different for the two plots. This is also a good idea for three-level factors. For the data shown in Figure 6.1, will the plot have a different configuration if factor A is plotted on the horizontal axis? If so, construct the plot. If not, explain why not.

6.5 The sample dataset PENDULUM.MTW that comes with the MINITAB software contains data from a 2×4 experiment for which an estimate of gravity is the response variable. A weight (heavy or light) is suspended at the end of a string. The time (T) required for a single cycle of the pendulum is related to the length (L) of the string by the equation $T = (2\pi/g)(\sqrt{L})$, so g can be solved for a given combination of T and L. Four different lengths were used: 60, 70, 80, and 90 cm. Three replicates were used so that there were 24 observations altogether. Analyze the data and draw conclusions.

6.6 Consider the 3^{9-6} design and accompanying response values in Table 9 in Cheng and Wu (2001, references), for the electric wire experiment given by Taguchi (1987, p. 423, references) experiment that they described. (This is available online at http://www3.stat.sinica.edu.tw/statistica/oldpdf/A11n31.pdf.) Perform your analysis of the data, including computing conditional effects if necessary, and compare your results with the results given by the authors. Do the results agree? If not, justify your results and approach. Does the two-stage approach with a single design seem appropriate for this example? Explain.

6.7 Peyton and Characklis (1993, references) examined the effects of substrate utilization and shear stress on the rate of biofilm detachment, using three levels for each factor. Read the article, analyze the data, and draw appropriate conclusions.

6.8 Gupte and Kalkarni (2003, references) used a 3^3 design to investigate the effects of three factors, with three levels used for each factor. The three factors used in the experiment were selected as a result of previous one-factor-at-a-time experimentation. Does this two-stage approach seem reasonable? Read the article and determine whether or not you agree with the authors' conclusions and state whether you would have used a different design strategy.

6.9 Show that all the dot products alluded to in the discussion in Section 6.1.1 are zero.

6.10 Verify the statement made in Section 6.3 that a 3^{4-1} design projects into a full 3^3 for any combination of three factors, with either $D = ABC$ or $D = ABC^2$.

6.11 It was stated in Section 6.3 that it is conventional to list the components of an interaction, such as the A^*B^*C interaction, by listing the components with the first factor in the interaction always having an exponent of 1. Assume that someone decides to break with convention and creates a 3^{4-1} design by letting $D = A^2B^2C$ instead of $D = ABC^2$. Will the designs be different? If so, are there any conditions under which one of the designs might be preferred over the other one? Explain.

6.12 Consider a 3^2 design that is run in blocks of size 3, with the *AB* component of the A^*B interaction confounded with the difference between blocks. Even though it is unconventional, assume that the blocks are constructed using A^2B^2 rather than *AB*. Will this result in the same set of three blocks being constructed as when *AB* is confounded with the difference between blocks?

6.13 Consider Exercise 6.11. Does it follow from the result in that exercise that the three 3^{4-1} fractions that result from using $D = A^2B^2C$ to construct the fractions are the same set of three fractions that result when $D = ABC^2$ is used?

6.14 Using the data in Example 6.1, compute the four interaction graphs for a 3^2 design that correspond to $A_{\text{linear}}B_{\text{linear}}$, $A_{\text{linear}}B_{\text{quadratic}}$, $A_{\text{quadratic}}B_{\text{linear}}$, and $A_{\text{quadratic}}B_{\text{quadratic}}$, and comment on the relationship between them.

6.15 Can a 3^{7-p} design be constructed with a reasonable number of runs for fitting a full second-order model? If so, construct the design. If not, explain why it can't be done.

6.16 Give the design columns for the $4 \times 2^{4-1}$ design that was alluded to in Section 6.8, using the method for constructing the design that was discussed therein.

6.17 Obtain the alias structure for the 3^{4-2} design whose defining relation was given in Section 6.9.

6.18 Use JMP or other software to construct a 3^{4-1} design and determine the defining relation and full alias structure.

6.19 Assume that you have three quantitative variables and an experimenter tells you that you should use a 4^3 design. Would this be a reasonable choice? Why, or why not? Assume that you start to use JMP but you discover that the software will permit you to construct the design only if the factors are categorical (i.e., qualitative). Why do you think this is the case?

6.20 Assume that an unreplicated 4^2 design is used. Give the degrees of freedom breakdown.

6.21 An experimenter has four factors to be investigated and wants to use three levels for three of the factors and two levels for the other factor. Economic considerations restrict the design to at most 48 experimental runs, however. The experimenter believes it is highly probable that some interactions will exist but isn't sure which interactions could be real. What design would you suggest?

6.22 Construct a 3^{5-1}design using the same general approach that was used for constructing the 3^{3-1} design in Section 6.3.2.

6.23 What is the resolution of a 3^{10-5} design? Is this a practical design for fitting a second-order model? Explain.

6.24 An application of a 3^2 design is given at http://www.smta.org/files/SMTAI02-Wang_Paul.pdf. The three levels of one factor, pressure, were 10, 14, and 18 (Kgf), and the levels of the other factor, speed, were 15, 25, and 50 (mm/s). Thus, the factor levels were not equally spaced. Would this have created a problem when the data were analyzed? Explain.

6.25 The data for the experiment described in Section 6.6 are as follows, with the first pair of observations for the first treatment combination, the next pair for the second treatment combination, and so on: 106.20, 107.70, 104.20, 102.35, 85.90, 85.90, 101.15, 104.96, 109.92, 110.47, 108.91, 108.91, 109.76, 112.66, 97.20, 94.51, 112.77, 113.03, 93.15, 92.83, 97.25, 100.60, 109.51, 113.28, 85.63, 86.91, 113.17, 113.45, 104.85, 98.87, 113.14, 113.78, 103.19, 106.46, 95.70, and 97.93.

(a) Do any of the data look suspicious? (*Hint*: Look at the data in the appropriate pairs.)

(b) If you identified one or more data points that you believe are probably in error, will the apparent error(s) likely affect the analysis? Explain. What would be a possible explanation for any suspected errors, recalling what has been discussed about factor level resettings in previous chapters?

(c) Analyze the data using whatever methods you prefer and compare the effects you select with the effects selected by Xu et al. (2004, references), as given in Section 6.6.

6.26 Heaney, Lidy, Rightor, and Barnes (2000, references) gave an example of a $3 \times 4 \times 5$ experiment that was questionably designed with the subsequent deleterious effects of that compounded by problems that occurred when the experiment was conducted. Indeed the design used was questioned by the authors in their "Conclusions" section. Read the article as well as the first referenced article if you find that necessary and can gain access to it.

(a) Would you have designed the experiment differently? If so, how?

(b) Would you have used a different data analysis approach in an effort to recover from the problems that occurred when the experiment was performed? If so, what would you have done differently? If not, why not?