CHAPTER 13

Miscellaneous Design Topics

The field of experimental design is extremely broad and there is not sufficient space in any design book to cover every important topic in detail. This chapter covers various topics that might be considered special topics because they aren't covered in any depth, if even mentioned, in most design books, but that doesn't lessen their importance. Some of these topics are covered herein in some detail; other topics are only briefly discussed. Many topics are covered and there is a large number of references at the end of the chapter to point readers in appropriate directions for further reading.

13.1 ONE-FACTOR-AT-A-TIME DESIGNS

For decades practitioners have been strongly advised to use factorial and other multifactor designs rather than varying one factor at a time while holding the other factors at fixed levels, as they have done for decades and still continue to do. An example of this is a case study given by Chokshi (2000), who stated "While this is a departure from traditional approaches such as "vary-one-factor-at-a-time" and "trial-and-error" approaches. . . ." Czitrom (1999) stated that many scientists and engineers will continue to use one-factor-at-a-time (OFAT) experiments ". . . until they understand the advantages of designed experiments over OFAT experiments, and until they learn to recognize OFAT experiments so they can avoid them." (This paper is available on the Internet at http://www.amstat.org/publications/tas/czitrom.pdf.) Many additional sources could be quoted to show that varying one factor at a time continues to be used in a large fraction of all designed experiments. Voelkel (2005) considered the efficiency of fractional factorial designs relative to OFAT experiments, under the somewhat restrictive condition that there are no interactions, and found that it is desirable to scale back the levels of fractional designs from the levels that would have been used if an OFAT design had been employed. McDaniel and Ankenman (2000) stated that "some one-factor-at-a-time experimentation may hold promise in the small factor change problem because they can detect significant main effects and

Modern Experimental Design By Thomas P. Ryan

may allow for fewer factor changes." (The small factor change problem is that where an experimenter seeks a certain amount of improvement in the response variable, such as moving it closer to a target value, while changing the factor levels as little as possible.) A point to be made is that opinions differ regarding the efficacy and potential of OFAT experimentation. Another important point is that there is a difference between OFAT experimentation and OFAT *designs*. It is obvious that most writers are referring to the former, whereas the latter should not automatically be dismissed.

We thus need to make a distinction between OFAT experimentation used by practitioners and the OFAT designs that have appeared in the literature, especially in recent years. To illustrate, consider the example given in Table 3 of Czitrom (1999). There were two factors: time (in seconds) and temperature (in degrees Celsius) with 8, 9, and 10 being the values used for time and 980, 1000, and 1020 the values used for temperature. Six runs were given in the following order: (10, 980), (10, 1000), (10, 1020), (8, 1000), (9, 1000), and (10, 1000). (Here we might assume that the combination of time and temperature in use was (10, 1000).)

Notice that only three more runs would be required to investigate all combinations of time and temperature, and there are only five distinct runs in the OFAT design since the (10, 1000) combination is repeated. This is the type of OFAT design/experimentation that practitioners are likely to use, especially practitioners who are unaware of the fact that OFAT designs with better structure are given in the literature. Better OFAT designs are discussed in the remainder of this section.

One of the main arguments against an OFAT approach is that interactions cannot be detected. This has been discussed and illustrated in many books, including Montgomery and Runger (2003, p. 509). Arguments against OFAT experimentation have been based on the assumption that factor changes for all factors can be easily made. This is very often not the case, however, and within the past 10 years this has been emphasized in the literature, as in Ganju and Lucas (1997) and Wang and Jan (1995), and is discussed in Section 13.2.

Consider the simplest case of two factors of interest but one is almost impossible to vary. This then reduces to a one-factor experiment, for which a t-test could be performed, assuming multiple observations per level of the factor that can be varied. Similarly, a three-factor experiment reduces to a two-factor experiment, and so on.

When we recognize that hard-to-change factors occur very frequently in practice, we must either (a) consider the use of factorial designs with the minimum number of factor changes for hard-to-change factors, or (b) consider designs other than factorial designs.

We should also recognize that interactions *can* be estimated with OFAT designs. The OFAT designs date from at least Daniel (1973) and have recently been "revived" by Qu and Wu (2005). Daniel (1973) considered five types of OFAT designs but not every type will be considered here (see also, Daniel, 1994).

In a standard OFAT design with two-level factors, there will be $(n^2 + n + 6)/2$ runs for n factors. Specifically, there will be one run with all of the factors at the high level, subsequent runs in which each factor in turn is set at the low level while the other factors remain at the high level, one run with all factors set at the high level, and $(n - 1)(n - 2)/2$ runs with ith and jth factors at the high level and the other

factors at the low level $1 \leq i < j \leq n - 1$. (Note that the latter is *not* the same as having all pairs of factors at the high level with the other factors at the low level since $j \leq n - 1$.)

All of the points in these designs will not be used to estimate a given effect, however, as the name of the designs would suggest. Indeed, the standard OFAT design has 13 points for four factors, and we cannot have factorial designs for two-level factors that have an odd number of points since the number of design points at the high and low levels for each factor would be unequal. Thus, we have to view these designs differently from the way we view factorial designs.

Instead, the user of such designs would make comparisons as they occur, and there is some discussion of this in Daniel (1973).

Consider the discussion of conditional effects in Section 4.2.1 and consider the fact that the lack of effect sparsity for conditional effects would cause a problem with normal probability plots of conditional effects *if* such a plot could be constructed, since Lenth's method and all other methods break down when there is not effect sparsity. What happens when there are many conditional effects and we use a OFAT design?

With either OFAT approach, conditional effects could be obtained only by conditioning on all of the other factors being at their high level (or their low level with the alternative approach). Although this might sometimes be of interest, this would be stronger conditioning than conditioning on a level of a factor and might be called treatment combination conditioning rather than factor-level conditioning. In essence, the conditional effects that are obtained would be the same as the effect estimates. Obviously, there would be essentially no power to detect effects when only two numbers are compared, so replication would be almost essential unless the resultant numbers were to differ greatly (and both be valid values).

In a strict OFAT design, one-factor level is changed at each run (unlike a standard OFAT design), with the treatment combinations progressing in $n + 1$ runs from all factors being at the high level to all factors being at the low level. Thus, this type of OFAT design has fewer run-to-run factor-level changes than does a standard OFAT design. As with the standard OFAT design, conditional effects would have to be obtained from two runs, the difference being that with the strict OFAT design, the other factors would not always be at their high (low) level, but rather the treatment combinations in the pairs of runs that would be used in computing the conditional effects would differ over the pairs.

Obviously, only main effects can be estimated with these designs of $n + 1$ runs for n two-level factors, but it is also possible to construct OFAT designs of resolution IV and resolution V. Specifically, a resolution IV design would be produced by folding over an OFAT design. (As described in Section 5.9, a foldover design results from combining a design with its mirror image.)

The construction of strict OFAT designs of resolution V is discussed by Qu and Wu (2005), with such designs having originally been given by Daniel (1973). In particular, Daniel (1973) gave a strict OFAT design for five factors and 16 runs, with the treatment combinations used in the following order: (1), a, ab, abc, $abcd$, $abcde$, $bcde$, cde, de, e, ae, ade, $acde$, acd, ac, and ace. Notice that literally one factor level is changed between successive runs. Also notice that the structure of this design

differs considerably from the structure of the design that was given at the beginning of the section. The design given by Daniel (1973) would be an ideal design if all of the factors were hard to change, although such a state of affairs would certainly be extremely rare.

How might such a design be used? The main effect of factor A could be estimated after the second run, although that estimate would be confounded with all interactions involving A. The next opportunity to estimate that main effect comes after the seventh run, as the sixth and seventh runs could be paired with the first two, and the estimate from those four runs would not be confounded with any interactions. Similarly, the other four main effects could be estimated, in addition to 3 of the two-factor interactions, with the other 7 two-factor interactions confounded in groups of 2 and 3.

What has been gained, if anything, relative to a 2_{V}^{5-1} design? The obvious gain is a considerable reduction in the number of factor-level changes, although the amount of the reduction, of course, depends on the number of factor changes that would be dictated by the randomization of runs when the fractional factorial design is used. In other ways, this design is strongly inferior to the fractional factorial design as the effect estimates from the latter have a smaller variance since all 16 points are used to estimate each effect and all of the two-factor interactions are estimable.

One of the touted advantages of OFAT designs is the possibility of early termination of the experiment, thus saving resources. Assume that for the strict OFAT that was just considered, the difference between the response values at (1) and at a differ by an order of magnitude. If no errors were made, this would suggest a possibly large A effect, although this effect is confounded with many interaction effects. If the experimenter proceeds and estimates the main effect after the seventh run and the estimate is still very large, this would seem to constitute clear evidence of a large A effect, with no further experimentation involving factor A apparently necessary, although further experimentation with A would be necessary to investigate interaction effects involving factor A.

Daniel (1973) indicated the effect and linear combinations of effects that were estimable after each run, and it should be realized that one cannot use additional runs in estimating a given effect. For example, if we look at the entire set of runs, factor B is at the high level for 5 of the runs and at the low level for the other 11 runs. Thus, we will have this type of imbalance if we proceed beyond run 10. Daniel (1973) used a different approach, however, and indicated that factor B was estimable after run 8, by which point 3 runs were at the high level and 5 runs were at the low level. Daniel (1973) is not using all eight data points in the estimation of the B effect, however.

Qu and Wu (2005) gave an example of a six-factor strict OFAT design of resolution V with 22 runs and, as in the example of Daniel (1973), indicated the effects that are estimated after each run. The runs in order were (1), $a, ab, abc, abcd, abcde, abcdef$, $bcdef, cdef, def, ef, f, af, aef, adef, acdef, acde, acd, ac, acf, acef$, and ace. The first four main effects, A–D, are estimated after runs 8–11, respectively, E and F are both estimated after run 12, and either one or two interactions are estimated after each of the succeeding runs. Qu and Wu (2005) indirectly gave a measure of the nonorthogonality of OFAT designs by giving the D-efficiency for strict and

standard OFAT designs with the number of factors between 4 and 10 and compared these designs against some other small, economical designs (see Section 13.7.1 for a discussion of *D*-efficiency). *D*-efficiency values indirectly measure nonorthogonality since only orthogonal designs have a *D*-efficiency value of 1.0.

The advantages of OFAT designs under certain conditions must be weighed against the fact that there is no randomization of runs since the runs must be made in a prescribed order and there is not a fixed number of runs as in factorial designs. Consequently, changing environmental conditions could undermine the results, which is why it is even more important that statistical process control checks (such as the check runs discussed in Section 4.14) be used with OFAT designs than with factorial designs, although such checks should certainly be used with both types of designs, and preferably with virtually all types of designs.

All things considered, OFAT designs should be considered when experimental runs are extremely expensive and when there are hard-to-change factors. The designs should be used with caution, however, because any small design has shortcomings and OFAT designs are nonorthogonal designs.

13.2 COTTER DESIGNS

Cotter designs (Cotter, 1979) are similar to OFAT plans and are intended for use when there are many factors, few resources, and it is believed that interactions may exist. The designs have $2k + 2$ runs for k factors. The second "2" is composed of one run with all of the factors at the high level and one run with all of the factors at the low level. The $2k$ runs result from each of the factors in turn set at the high level while the other factors are set at the low level, and vice versa.

This design should be considered only for $k \geq 4$ because for $k = 3$ it produces a 2^3 design (as might be expected because $2k + 2 = 2^k$ when $k = 3$), and for $k = 2$ it would produce a 2^2 design with two points replicated. The design for $k = 4$ is given below:

A	B	C	D
-1	-1	-1	-1
1	-1	-1	-1
-1	1	-1	-1
-1	-1	1	-1
-1	-1	-1	1
-1	1	1	1
1	-1	1	1
1	1	-1	1
1	1	1	-1
1	1	1	1

Designs that have an "ad hoc" look to them and a small number of design points relative to the number of factors will generally be nonorthogonal, and these designs

are nonorthogonal. In general, although the designs are in JMP, for example, and the design given above was created using JMP, they are of very questionable value and the default in JMP has them suppressed from the list of screening designs.

They should not be used where there are many factors, although this is when they were intended to be used. This can be explained as follows. The correlation between any pair of factors, and thus between the main effects estimates is

$$\frac{(2k + 2) - 2(4)}{2k + 2} = \frac{2k - 6}{2k + 2}$$

for $k \geq 4$. This result can be explained as follows: For any pair of columns, the number of rows for which the product of the corresponding column elements is not 1 (i.e., is -1) is 4. The number of dot products that are 1 is thus reduced by that number, and the dot product is reduced further by the fact that the sum of the dot products for the pairs with unlike signs is 4. (Hence, 2(4).) This explains the numerator of the fraction, which also explains the correlation since the denominator is fixed.

Notice that the correlation approaches 1 as $k \to \infty$, with the correlations being $1/3$ when $k = 5$. Even this correlation is too large and the criticisms of the deleterious effects of moderate correlations that have been made about some supersaturated designs (see Section 13.4.2) can also be applied to Cotter designs.

Another problem with these designs is that although they are meant to be used when interactions are suspected, only a small percentage of interactions can be estimated when k is large because $\binom{k}{2}$ grows at a much faster rate than $2k + 2$. For example, $\binom{k}{2} = 45$ when $k = 10$, but $2k + 2 = 22$ and the 10 main effect estimates must be obtained from the 21 degrees of freedom.

13.3 ROTATION DESIGNS

A relatively new class of designs was proposed by Bursztyn and Steinberg (2001, 2002). These designs are obtained by rotating (hence, the name) a two-level factorial design. Specifically, let D denote an $n \times k$ design matrix for a two-level fractional factorial design and let R be any $k \times k$ orthogonal matrix. Then the design $D^* = (1/c)DR$ is a rotation design, with c a scaling constant that scales the design to the unit cube. Some choices of R may be more useful than others, and Steinberg and Burszytn (2001) state that a useful choice for R is $R = HSH'$ when $k = 2^t$, with H denoting the standard Hadamard matrix for estimating the effects with the 2^t design and S is a block diagonal matrix. (A Hadamard matrix is a square, orthogonal matrix with entries of $+1$ and -1.)

A primary feature of these rotation designs is that they are intended to be used for factor screening and fitting a model higher than a first-order model in the same stage, thus obviating the usual two-stage procedure.

13.4 SCREENING DESIGNS

The term "screening design" is used often in the literature. As the name implies, the objective with a screening design is to "screen out" seemingly unimportant factors so that the apparently important factors can be identified and used in subsequent experiments as part of a sequential experimentation strategy. An example of such a strategy using a screening design is given in Chokshi (2000).

The emphasis is thus on main effects when these designs, which generally have two levels, are used, and the successful use of these designs requires that main effects dominate interaction effects.

Screening designs have a small number of design points relative to the number of factors. A design with $(k + 1)$ runs for k factors is often called an *orthogonal main effects plan* and it can also be called a *saturated design*, the latter being a design that just has enough degrees of freedom to estimate the effects that one wishes to estimate. There are also designs with fewer than $(k + 1)$ runs and these are called *supersaturated designs*, which are covered in Section 13.4.2.

One analysis problem that must be addressed when designs with a small number of runs relative to the number of factors are used is the determination of significant effects. It is entirely possible that a majority of the main effects will be significant in a particular application, which will create a problem if Lenth's method or any one of the alternative methods is used. With a supersaturated design, a technique such as forward selection must be used to identify the main effects that are to be estimated since not all of them can be estimated. Although such an approach is not mandatory when a saturated design is used, it would not be a bad idea to use both a variable selection approach and a normal probability plot approach and compare the results, especially when there is a small number of runs.

Assume that there are seven candidate factors/main effects and eight runs are to be used, with each factor at two levels. This is a 2^{7-4} design, a $(1/16)$th fraction of a 2^7 design. Since only main effects are estimable, these are thus aliased with the interaction effects—all of them. Specifically, each main effect is aliased with 3 two-factor interactions, plus higher-order interactions. Now let's add a factor so that we have eight factors. Can we construct a saturated design with nine experimental runs? Obviously not if we still want to use two levels because 9 is not a multiple of 2. Notice that 9 is also not a power of 2, but we can construct designs with a number of runs not a power of 2, provided that the number of runs is a multiple of 4. Such designs are presented in the next section.

13.4.1 Plackett–Burman Designs

We can construct two-level (orthogonal) designs with 12, 20, 24, 28, and larger numbers of runs, up to 100. These are called *Plackett–Burman* (PB) *designs* and were given by Plackett and Burman (1946). The projective properties of these designs were given by Lin and Draper (1992), Tyssedal (1993), and Box and Tyssedal (1996a).

These designs were actually given for studying up to k factors in $N = (k + 1)$ runs, but they are equivalent to 2^{k-p} designs when N is a power of 2, with $2^{k-p} = k + 1$. Examples include the 2^{7-4} and 2^{15-11} designs. Statistical software is readily available for constructing these designs, as Design-Expert will construct the design for (up to) 11, 19, 23, 27, and 31 factors, with the number of runs of course being one more than each of these numbers. MINITAB will construct PB designs for 12, 20, 24, 28, 32, 36, 40, 44, and 48 runs, and of course up to one less for the number of factors for each design with the indicated number of runs. Thus, PB designs can be constructed with MINITAB for 2–47 factors, using the appropriate run size. Of course the run size must be an even number, as the number of low and high levels of each factor would not otherwise be equal.

The value of these designs is that they provide (possibly) economical alternatives to 2^{k-p} designs when runs are quite expensive. That is, with the 2^{k-p} series we have 8, 16, 32, 48, ... runs, whereas the PB designs give us three possibilities between 16 and 32 runs, and one design between 8 and 16 runs.

These designs are somewhat controversial. In particular, Anderson and Whitcomb (2000; http://www.statease.com/pubs/aqc2004.pdf) recommend that they be avoided and Montgomery, Borror, and Stanley (1997) criticize the design as having a complex alias structure (which is true). Snee (1985b) offers an opposing view, however, and Hamada and Wu (1992) and Wu and Hamada (2000, p. 356) provide an analysis strategy that permits the estimation of some interactions. We will later show how all of the two-factor interactions involving factors that are declared to be significant can be estimated for a particular scenario.

The designs have an easily discernible pattern. To illustrate, for $n = 8$ the first row of the design has the following signs: $+ + + - + - -$. The next six rows are obtained by successively moving the sign on the far right to the far left, with the last row being all minus signs. Thus, the design for n = 8, generated using MINITAB, is as follows:

Run	A	B	C	D	E	F	G
1	1	-1	-1	1	-1	1	1
2	1	1	-1	-1	1	-1	1
3	1	1	1	-1	-1	1	-1
4	-1	1	1	1	-1	-1	1
5	1	-1	1	1	1	-1	-1
6	-1	1	-1	1	1	1	-1
7	-1	-1	1	-1	1	1	1
8	-1	-1	-1	-1	-1	-1	-1

As stated earlier in this section, this design must be a 2^{k-p} design since the number of runs is a power of 2. This is not obvious, however, because of the way the runs are listed by MINITAB. If we look at the first three columns, however, we can see that this does constitute a 2^3 design in those factors, which becomes apparent if we put the runs in standard order, as follows:

Run	A	B	C	D	E	F	G
8	-1	-1	-1	-1	-1	-1	-1
1	1	-1	-1	1	-1	1	1
6	-1	1	-1	1	1	1	-1
2	1	1	-1	-1	1	-1	1
7	-1	-1	1	-1	1	1	1
5	1	-1	1	1	1	-1	-1
4	-1	1	1	1	-1	-1	1
3	1	1	1	-1	-1	1	-1

Now it is clear that the first three columns comprise a 2^3 design. (Actually, any column triplet constitutes a 2^3 design.) It may also be apparent that the generators for the other four factors are $D = -AB$, $E = -BC$, $F = -AE$, and $G = -AC$. Thus, the defining relation is $I = -ABD = -BCE = -AEF = -ACG$ plus all generalized interactions of these four effects. Of course this isn't the way we would normally construct a 2^{7-4} design, but we can see that the PB design with eight runs is in fact one of these designs.

Designs with larger values of n are constructed in the same manner, as can be seen for the following PB design for $n = 12$, which was also generated using MINITAB but which cannot be a 2^{k-p} design because 12 is not a power of 2.

Run	A	B	C	D	E	F	G	H	J	K	L
1	+	-	+	-	-	-	+	+	+	-	+
2	+	+	-	+	-	-	-	+	+	+	-
3	-	+	+	-	+	-	-	-	+	+	+
4	+	-	+	+	-	+	-	-	-	+	+
5	+	+	-	+	+	-	+	-	-	-	+
6	+	+	+	-	+	+	-	+	-	-	-
7	-	+	+	+	-	+	+	-	+	-	-
8	-	-	+	+	+	-	+	+	-	+	-
9	-	-	-	+	+	+	-	+	+	-	+
10	+	-	-	-	+	+	+	-	+	+	-
11	-	+	-	-	-	+	+	+	-	+	+
12	-	-	-	-	-	-	-	-	-	-	-

Assume that the design for $n = 8$ has been used for the seven indicated factors and only the main effects for factors A, B, and C were found to be significant using Lenth's method (see Section 5.1). This tells us all we need to know about the 3 two-factor interactions involving those factors because, as stated, $D = -AB$, $E = -BC$, and $G = -AC$. That is, the interactions are not significant because the main effects with which they are confounded are not significant.

When all seven factors are used, the fact that D, E, and G are not significant is important information because the interpretation of the main effects of A, B, and C is relatively straightforward in the absence of two-factor interaction effects involving

those factors because the conditional effects will not differ greatly when interaction effects are small (see Chapter 4, Appendix A). Of course if the main effects for D, E, and G were judged significant, there would then be a dilemma because of the confounding. The *absence* of significant effects for these factors is helpful in this case because declaring the main effect of factor E to be not significant is the same as declaring to be not significant the two-factor interaction (involving factors B and C) that is confounded with factor E.

We have thus seen how a PB design can be used advantageously relative to interactions, even though screening designs such as this type of design are generally viewed as being for main effect determination almost exclusively.

In general, users of experimental designs should always remember that it will rarely be necessary to estimate all of the effects that can be estimated with a design of at least moderate size, so columns in a design matrix can often be used for something other than the original purpose.

The alias structure for a PB design with eight runs is relatively straightforward, although somewhat involved, since this is actually a fractional factorial design. The alias structure is much more complicated when the run size is 12 (and in general when the run size is not a power of 2), as in the 12-run design each main effect is partially aliased with several two-factor interactions. (The term *partially aliased* is used here to designate that an effect is neither orthogonal to nor confounded with another effect, but rather can be expressed as a linear combination of effects, possibly a long string of effects. For example, the full alias structure for the PB design with 12 runs and 11 factors is extremely long, with the *first part* of the (partial) alias structure for factor E given by Design-Expert as $[E] = E - 0.333 * AB - 0.333 * AC + 0.333 * AD - 0.333 * AF - 0.333 * AG - 0.333 * AH - 0.333 * AJ + 0.333 * AK + 0.333 * AL - 0.333 * BC + 0.333 * BD \ldots$. (The entire partial alias structure up through three-factor interactions is not given here as it would occupy about half a page.)

Partial aliasing is actually worse than complete confounding relative to what was illustrated for the 8-run design because complete confounding would allow us to make the kinds of substitutions that were made in the 8-run design illustration. For example, the AB interaction could not be estimated using one of the columns in the matrix for the 12-run design because the column that results from multiplying together the A and B columns in the design is not given by any of the columns in the matrix.

This does not mean that the AB interaction could not be estimated, however, because if A and B are orthogonal (as of course they are since the PB design is orthogonal), then AB is orthogonal to A and B. This is easy to prove: Let a_i denote the ith observation in the column of the design matrix for factor A and similarly let b_i denote the ith observation in the column of the design matrix for factor B. Since the columns of the PB design matrix are orthogonal, $\sum_{i=1}^{n} a_i b_i = 0$. Since $a_i b_i$ is $(AB)_i$, it follows that AB is orthogonal to A if $\sum_{i=1}^{n} a_i b_i a_i = 0$. Of course we may write this as $\sum_{i=1}^{n} a_i^2 b_i = (1) \sum_{i=1}^{n} b_i$, since each $a_i^2 = 1$. It follows that $(1) \sum_{i=1}^{n} b_i = 0$, since $\sum_{i=1}^{n} b_i = 0$. Thus, AB is orthogonal to A and the companion proof that AB is orthogonal to B should be obvious. Of course by definition, AB cannot be orthogonal to every column, however, and in fact $AB = -F$.

We can estimate as many effects as we have degrees of freedom for estimating effects; an effect that we estimate doesn't necessarily have to be an effect that we intended to accommodate with our choice of design. More specifically, it is worth noting (although it may be quite obvious) that, for example, AB is not part of the design matrix, but it can be part of the model matrix, and in general any term can be part of the model matrix as long as the factors that comprise it are included in the design matrix, provided the degrees of freedom are sufficient to accommodate it.

There is a problem, however, if, for example, the use of a 12-run PB design for 11 factors results in factors A, B, C, and D identified as significant, with the AB and CD interactions believed to be important and added to the model. The problem is that doing so creates correlations between the following effect estimates: CD and A, CD and B, AB and C, and AB and D. Thus, orthogonality is lost when we go beyond having one 2-factor interaction and the main effects of the two factors that comprise the interaction estimated.

The absolute value of these correlations is only $1/3$, however. Hamada and Wu (1992) gave an analysis strategy for use with PB designs that they recommended for use when correlations are small to moderate, with the strategy involving the use of forward selection to identify significant effects. They used some examples of datasets to illustrate their approach.

PB designs for $n = 20$, 24, and 28 are given by Wu and Hamada (2000, pp. 331–332) Since a PB design is a special class of an orthogonal array design, it can be found among the large collection of orthogonal array designs at http://www.research.att.com/~njas/oadir. Specifically, the design is given at http://www.research.att.com/~njas/oadir/oa.12.11.2.2.txt. (*Note*: In order to see that the PB designs given at the latter URL are equivalent to the design produced by MINITAB, it is necessary to use the correspondence $0 = +$ and $1 = -$, with the levels at the Web site given by 0 and 1 and the MINITAB levels of course given by $+$ and $-$. Of course to construct these designs we only need to know the first row of the design, as explained previously.

Nelson (1982) recommended the use of a foldover PB design to deal with the complex alias structure of the PB design and to produce a resolution IV design. Ahuja, Ferreira, and Moreira (2004) used this idea in an experiment that they performed and termed the design the reverse PB design, a name that was not used by Nelson (1982). This is discussed further in the next section.

It is also possible to use half fractions of PB designs, as discussed by Lin (1993a). See Samset and Tyssedal (1998) for an extensive catalog of PB designs of various sizes and their foldovers and half fractions. This includes a nice summarization of the correlation structure for each design. This technical report can be accessed through the Internet. See the URL given in the References section.

13.4.1.1 Projection Properties of Plackett–Burman Designs

PB designs have very good projection properties, as is extolled by Box, Hunter, and Hunter (2005, pp. 283–284). Lin and Draper (1992) and Box and Bisgaard (1993) showed that some of the saturated PB designs whose number of runs is not a power of 2 have better projection properties than do the saturated fractional factorial designs.

Properties of PB designs are discussed extensively by Samset and Tyssedal (1998). The smaller designs have projectivity of 3, meaning that the designs contain at least a full 2^3 design. The larger PB designs, such as for 68, 72, 80, and 84 runs, have projectivity 4, meaning that they contain at least a full 2^4 design. As pointed out by Samset and Tyssedal (1998), a half fraction of a PB design of projectivity 4 has projectivity 3. Of course if we take a half fraction of a PB design that is almost saturated, then we obtain a design that is supersaturated (see Section 13.4.2), which requires the use of special methods for the data analysis, such as stepwise regression.

13.4.1.2 Applications

PB designs are used extensively in industry, but they are also used extensively in biotechnology, as is apparent from the literature, such as Balusu, Paduru, Seenayya, and Reddy (2004).

Another example is the use of a PB design for investigating shipworm bacterium that was given by Ahuja et al. (2004), including a folded over PB design, which the authors called a "reverse Plackett–Burman design," following the recommendation of Nelson (1982) of "reversing" the signs in the PB design in order to de-alias effects. (Ahuja et al. (2004) seem to have invented the name and it seems more appropriate to use conventional terminology, especially since the term "foldover" is well established.") Other peculiar terminology included "... 80 runs (20 runs in the original design and their 20 replicates, and 20 runs in the fold-over design with 20 replicates)." If we read this literally, the design would have 800 runs, not 80. The design is actually a replicated PB design with its foldover also replicated.

The experiment involved 18 factors, so a design with 20 runs was chosen to permit the estimation of all of the main effects. The replication does not, of course, permit the estimation of any more effects than are possible without the replication, as the replication simply creates a pure error term in the ANOVA table with 40 degrees of freedom. The foldover, however, does convert the design to a resolution IV design, so that main effects are clear of two-factor interactions.

There were some errors made in the analysis of the data and the discussion of the design. In particular, the authors don't seem to understand that the foldover of a PB design creates a resolution IV design, even though they recognize that the main effects and two-factor interactions are not confounded, even partially, with the foldover design that they used. In particular, they stated that an alternative design would have been a resolution IV design with 96 runs (a design with 48 original runs that are replicated). The type of design to which the authors are referring is not clear.

Another error, which resulted from hand computation instead of using software, is that the authors gave the expression for the mean square of a factor as equal to $(\text{effect estimate})^2/20$, whereas the correct expression is $20*(\text{effect estimate})^2$. This error resulted in some very small mean squares being computed with the consequence that no F-statistic exceeded 2.78 and some of the statistics were practically zero. The authors concluded that only three of the factors were significant, but that was also an error because, with the error term having 61 degrees of freedom and each factor

of course having one degree of freedom since there are two levels for each factor, $F_{1,61,.05} = 4.0$. Thus, technically, the authors should have concluded that nothing was significant.

Finding that only 3 of the 18 factors were significant led the authors to state "Thus the analysis of Plackett–Burman design alone may lead to erroneous conclusions." The problem was actually the authors' analysis. As long as the design is carried out properly and the experimenters have done a good job of brainstorming and selecting a set of factors to study, the analysis of main effects using a resolution IV design should not be misleading. A correct analysis cannot be performed by the reader because the 80 data values were not given, but such an analysis would presumably show more than three significant factors. The lesson to be learned here is that it is best to rely on software, especially when something other than a very simple design is being used.

There are many other examples in the literature of the use of a PB design in biotechnology, including an experiment described by Wen and Chen (2001). They used a PB design to identify significant factors and then used a central composite design for the factors that emerged as being significant in the analysis of the PB design data (see Exercise 13.7).

Case Study

Another example of the use of a PB design, this time in the training of engineers, was given by Anthony (2002). I used catapult experiments to train industrial personnel in experimental design, as have other short course instructors and consultants. The general idea is to use a small catapult (which catapults a ping-pong ball) to identify and illustrate effects that influence the distance the ping-pong ball is catapulted. In this particular training exercise, the team of engineers identified seven factors that were deemed worthy of study. These factors were ball type/color, rubber band type, stop position, peg height, hook position, cup position, and release angle. The design that was selected was a PB design with 12 runs. The design layout for these seven factors, denoted by *A*–*G* respectively, is given below, in addition to the values of the response variable, *Y*.

A	B	C	D	E	F	G	Y
Y	BR	4	1	1	5	170	119
Y	BL	1	4	1	5	Full	161
P	BL	4	1	4	5	Full	253
Y	BR	4	4	1	6	Full	249
Y	BR	1	4	4	5	170	114
Y	BL	4	1	4	6	Full	227
P	BL	4	4	1	6	170	214
P	BR	4	4	4	5	170	327
P	BR	1	4	4	6	Full	304
Y	BR	1	1	4	6	170	60
P	BL	1	1	1	6	170	18
P	BR	1	1	1	5	Full	78

This is how the design was presented in Anthony (2002), but there is a not-so-obvious error in the design layout, which becomes apparent when we try to analyze the data. The design should be orthogonal but it isn't, which is the message that one receives when trying to analyze the data using, say, MINITAB. For factor B, one of the *BRs* must be a *BL* so that there are six of each. (Note that this is an example of what was discussed in Section 4.10 regarding bad data.)

But which one is incorrect? If this were the only error in the layout (which it is), we could determine the error using a process of elimination since for each pair of factors, each pair of levels must occur three times. Starting with factor A, we conclude that the error must be in row 1, 4, 5, or 10; looking at factor C tells us that it must be 5, 9, 10, or 12; from examining factor D, we can tell that it must be 4, 5, 8, or 9. At this point, we have determined that if there is only one error, it has to be that the *BR* in row 5 must be a BL because that is the only number that is common to all three sets of numbers. Of course there could be more than one error, however, so we need to continue with the other factors. Doing so results in the sets (5, 8, 9, 10), (1, 5, 8, 12), and (1, 5, 8, 10) for factors E, F, and G, respectively. Since "5" also appears in each of these sets (and is the only number to appear in all six sets), we know that there is only one error, which we have identified. We could also identify the error by generating the design with software and matching that up with the design in the article. This also leads to the identification of the single error.

Without knowledge to the contrary, we will assume that this was just a printing error that occurred when the article was printed and that the experiment was properly conducted. In his analysis of these data, Anthony (2002) presented a Pareto chart of the standardized effect estimates, with $\alpha = .10$ as the cutoff for judging significant effects. Five of the seven main effects were judged significant, with the model containing only main effects. (The same result is obtained when a normal probability plot is used since both methods in MINITAB are based on Lenth's pseudostandard error.)

Although this result would seem to suggest that the team of engineers was very successful in identifying which factors to study, in practice we would not expect to see such a high proportion of effects be significant.

Therefore, we need to see why this happened. There are 11 degrees of freedom available for estimating effects, but only seven of them were used; the others being used to estimate the error term. If we try to use the other four degrees of freedom to estimate four interaction terms, we have a problem because the interactions are not orthogonal to (all of) the main effects. Furthermore, the correlations can be moderate in size, which can create problems with interpretation. (E.g., adding the AE interaction to the model creates correlations of .33 in absolute value between that interaction and most of the main effects). The AE interaction plot is given in Figure 13.1.

Despite the difficulty of interpreting this interaction because of its partial aliasing with several main effects, one thing is inarguable: the three greatest distances all occurred with the pink ball and the high setting on the hook position—something that is unlikely to occur due only to chance. It is also worth noting that there is (apparently)

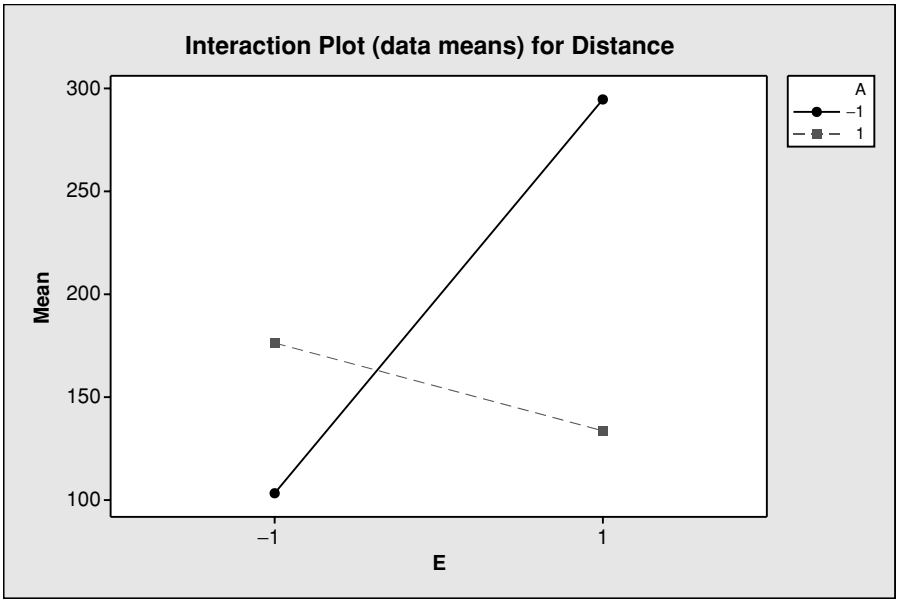

Figure 13.1 Interaction plot showing extreme interaction.

a very large hook position (E) effect with the pink ball. Since the decision was made to run a second experiment using only the pink ball and varying the other four factors in a factorial design, we would expect to see a sizable hook position effect in that experiment.

If we follow Hamada and Wu (1992) and Wu and Hamada (2000), we can employ stepwise regression with a combination of main effects and interactions in an effort to gain insight into possible interaction effects. By using a combination of their three-step approach and stepwise regression applied to all main effects and two-factor interactions, we arrive at a model with the following effects: AE, EG, F, AB, CE, and AC, BF, and AF, with the terms entering the model in this order. This model, which has nine terms, has an R^2 of .9998, which is considerably higher than the R^2 of .9657 for the model with the seven main effects (plus the constant term). Thus, the model with almost all interaction terms explains essentially all of the variation in the distances, but we have to realize that estimating eight effects with 11 degrees of freedom is spreading the data very thin. Of course a model with almost all interaction terms would be highly suspect as an algorithmic selection of such a model would beg for a conditional effects analysis.

Nevertheless, these results suggest that a subsequent experiment (and there was one) should permit the estimation of these seven interactions so as to determine if these interactions are real. This was not done, however, because only the other four factors from the initial experiment (C, D, E, and G) were used in a 2^4 with two replicates follow-up experiment, with factor F set at the high level and, as mentioned, the pink ping-pong ball used.

An oddity occurs with the data from the PB design because if simple linear regressions are run using F, AE, and EG in turn as the predictor variables, the R^2 values are .0003, .374, and .069, respectively, which adds to .4433. When all three terms are used in a regression model, $R^2 = .929$, which is obviously far in excess of .4443. This is an example of "the whole is greater than the sum of the parts," as discussed by Ryan (1997, p. 177) and is similar to the example given therein. This peculiarity can occur only when effect estimates are correlated, as they are here.

So does this mean that there really is an F effect, even though there appeared to be no chance of that when only main effects were considered? The "partial alias" structure for a 12-run design with seven factors is given by Wu and Hamada (2000, p. 351). (This is also automatically produced by the Design-Expert software.) Assume, for the sake of illustration, that the only significant effects were F, AE, and EG, which is not far from the actual situation since they account for 93% of the variation in the response. Then, from Wu and Hamada (2000), $E(\widehat{F}) = F - \frac{1}{3}AE - \frac{1}{3}EG$. Since both signs are negative, these interactions could create the false impression that F is not significant if the true F effect is positive and of about the same order as the sum of $-\frac{1}{3}AE$ and $-\frac{1}{3}EG$. Of course we don't know the true effect nor can we estimate it unambiguously here, but this is a plausible explanation.

In the follow-up experiment (i.e., the 2^4 with two replications), the effects that were found to be significant were, in terms of the original letter designations: D, G, E, C, DE and CG, in the order of effect size. It is worth noting that Anthony (2002) used a Pareto effect analysis of these data, just as was done with the first dataset, although that isn't necessary here since the design is replicated. In fact, an analysis using ANOVA also shows a significant CDE interaction, although the p-value at .044 is borderline.

Since the objective with such an experiment is to determine the factor-level combinations that maximize distance, the use of the 2^4 design is questionable, as is the decision to fix the levels of A and F for that experiment, as interactions involving factor A of the magnitude of the one shown in Figure 13.1 could be missed with that strategy. Since the A effect was strongly significant in the first experiment and the non-significance of the F effect was apparently due to its partial aliasing with certain large two-factor interactions, it might have been better to include at least one of these two factors in a $1/2$ fraction so as to, in particular, determine the F effect untangled from interactions and to see if any two-factor interactions involving factor A were significant in the follow-up experiment. A 2_{V}^{6-1} design would have allowed the estimation of all two-factor interactions, and if unreplicated would have required the same number of runs as the replicated 2^4 design.

13.4.2 Supersaturated Designs

These are designs for which the number of design points is less than the number of terms in a model. This of course means that not all effects can be estimated, and it also means that the design is not orthogonal.

Such a design might seem to have little value, but these designs, with the idea first proposed by Satterthwaite (1959) for use in random balance designs and studied

systematically by Booth and Cox (1962), can be used advantageously when a small number of significant effects is expected. Since this early work, Allen and Bernshteyn (2003) studied the then-existing supersaturated designs and that motivated them to develop an improved class of designs that maximized the probability of identifying active factors. Liu, Ruan, and Dean (2005) subsequently constructed a class of supersaturated designs by augmenting the k-circulant designs of Liu and Dean (2004). The new class of designs contains the designs of Eskridge, Gilmour, Mead, Butler, and Travnicek (2004) as a special case.

In recent years there has been research interest in the construction of three-level and mixed-level designs that are supersaturated. The former are discussed in Yamada and Lin (1999), Yamada, Ikebe, Hashiguchi, and Niki (1999), and in Section 15.5 of Hinkelmann and Kempthorne (2005). Mixed-level supersaturated designs are discussed in Fang, Lin, and Ma (1999), Fang, Lin, and Liu (2003), and Xu and Wu (2005).

Some care must be exercised in the selection of a design, however, as a supersaturated design that departs considerably from an orthogonal design could produce misleading results. This can especially occur if the departure from orthogonality is more than slight, as found by Abraham, Chipman, and Vijayan (1999). In particular, stepwise regression is one method that has been used for identifying the effects that should be estimated, but stepwise regression can easily fail to make the appropriate determination when the correlations between the columns in the design are not quite small. Accordingly, Abraham et al. (1999) recommended that all possible regressions be used instead of stepwise regression (as had been recommended by some researchers) in trying to identify significant effects. (See Ryan (1997) for information on these regression methods.)

One analysis method that is presently advocated (see Lin, 2003) is to use a nonconvex penalized least squares approach, as described in Li and Lin (2002, 2003), with the second of the last two papers available at http://www.sinica.edu.tw/∼jds/JDS-134.pdf. It should be noted that this method, which requires special software, is really an estimation method rather than a variable selection method. Nevertheless, it performed well in the numerical study performed by Li and Lin (2003). More recently, Lu and Wu (2004) proposed a method of successive dimensionality reduction that they claim is easier to understand and implement than other recently proposed methods.

13.4.2.1 Applications

Two applications of supersaturated designs were described by Bandurek (1999). We will discuss the first of these and critique the analysis, which was rather peculiar.

Booth Dispensers Ltd. manufactures soft drink coolers and cooler carbonators for use in vending machines. One major customer for one of the carbonators needed to be convinced that the product would work in their vending machines. A team of people from both the supplier and the customer identified 20 factors that could affect performance. An orthogonal array with eight runs was selected for the experiment. These are too few runs for a statistical analysis, such as by using stepwise regression as the correlations would undermine such an approach. The experimenters seemed to realize this, so they ordered the responses from best to worst, with the level of each

factor recorded on the sheet of paper along with the response value. The experimenters wanted to see if any factors would have level 1 at the top of the pile and level 2 at the bottom, which would suggest to them that such factors were significant. They "shuffled the deck" before looking at the results, then tried to combine engineering judgment with the results of the experiment. Although the results one year later apparently suggested that the experiment was a success, the analysis was quite crude and is not recommended. It is somewhat reminiscent of Taguchi's "pick-the-winner strategy" and Shainin's variables search approach, both of which have low power for detecting significant effects. (See Ledolter and Swersey (1997) for information on the latter.)

13.4.3 Lesser Known Screening Designs

Lin (2003) presents some screening designs that are not well known, one of which is *p-efficient* designs. These are saturated designs introduced by Lin (1993a,b) that are reasonably efficient for estimating the parameters of projective submodels, from whence the design derives its name. The designs are near-orthogonal. Box and Tyssedal (1996b) considered the projective properties of two-level 16-run orthogonal arrays that are due to Hall (1961). See also the simplex designs of Mee (2002).

In general, as has been stressed in this book beginning with the start of Chapter 5, properties of the design that corresponds to the initial model are not what's important since a model with fewer terms than the number of effects that can be investigated with the design used is almost certainly going to be used due to effect sparsity. The properties of the design for the model that is actually used is obviously what is important.

13.5 DESIGN OF EXPERIMENTS FOR ANALYTIC STUDIES

Investigations involving experimental designs are generally enumerative studies; that is, the objective is to assess the current state of affairs regarding the relationship between a response variable and one or more factors. With an analytic study the objective is to predict the future, with the terms *analytic studies* and *enumerative studies* due to Deming (1953, 1975). See also Hahn and Meeker (1993).

In an enumerative study, blocks are generally assumed to be selected at random, and block $\times$ treatment interactions are assumed not to exist. In analytic studies, however, blocks should be selected to cover extreme conditions and to identify block $\times$ treatment interactions. That is, if, say, process yields are to be predicted, we would want to know how important certain process variables are when the operator on duty is the company's best, versus the importance of those variables when the worst operator is on duty. Thus, it would be much better to deliberately select the particular operators that will serve as blocks rather than simply selecting them at random. For more discussion of blocking in analytic studies, see León and Mee (2000).

13.6 EQUILEVERAGE DESIGNS

An unstated objective in experimental design is that the design points should exert equal, or at least nearly equal, influence on the effect estimates, and equivalently on the regression coefficients in the model representation.

Of course whether or not an observation is influential will depend in part on the Y-coordinate, but we can improve the chances of not having any influential points by using design points that are equally influential; that is, by constructing the design so that the design points have equal leverages.

A leverage value reflects the distance that a point is from the "middle" of the group of points. With this semidefinition in mind (a formal definition will soon be given), it is easy to see why common designs such as two-level full and fractional factorial designs are equileverage designs since the design points are at the vertices of a hyperrectangular region, such as in Figure 4.5, and are thus equidistant from the center, which would have coordinates of zero for all the factors.

It is important to keep in mind that leverage values are determined by the model that is fit, however, not just by the design. Therefore, we cannot speak solely of leverages for particular designs. As we have seen in previous chapters, designs allow various models to be fit and it is the model in conjunction with the design that determines the leverage values.

To illustrate this point, assume that a 2^3 design is employed and only main effects are used in the model. Let the vector $\boldsymbol{x}_i$ denote a column whose first value is 1 and remaining values are the ith value for each of the terms in the model, and let $\boldsymbol{X}$ denote the model matrix, whose first column is a column of 1s and other columns contain the design values for the terms in the model. The n leverages for the n observations are then given by $\boldsymbol{x}_i^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{x}_i$. For example, if only main effects are fit using the data from the 2^3 design, the leverages are all 0.5, whereas if the two-factor interactions are added to the model, the leverages become 0.875. In an equileverage design the leverages are p/n, with p denoting the number of parameters in the model. When only the main effects are fit, that number is 4, so the leverages are $4/8 = 0.5$. When the three 2-factor interactions are added, the leverages are then $7/8 = 0.875$. For a saturated design, the leverages of course are 1.0, although leverages for a saturated design are not given by certain software packages (e.g., MINITAB).

If we construct a design whose leverages are equal or at least almost equal, we reduce the likelihood of having any influential data points, as a point could then only be influential through its Y-value. Despite the fact that this is obviously desirable, researchers have almost completely ignored equileverage designs as there are apparently only two research papers in which the term has been used, and only one paper devoted to the construction of such designs.

Equileverage designs are not without weaknesses, however, as they cannot be used to detect nonlinearity since the points are on the faces of a sphere and no points lie in the interior of the sphere. Consequently, an equileverage design would have to be modified slightly to provide nonlinear detection capability.

It is important that equileverage designs be considered and this can be motivated by the fact that various diagnostics break down when leverages are 1.00, as occurs

with some very common designs such as some designs given by Box and Wilson (1951), Box and Behnken (1960), and Roquemore (1976), as discussed by Jensen (2000).

13.6.1 One Factor, Two Levels

The simplest and most intuitive way to introduce and understand equileverage designs is to consider a design for a single factor with two levels. Thus, starting from the model for one factor (see Section 2.1.1):

$$Y_{ij} = \mu + A_i + \epsilon_{ij}$$

We may equivalently write the model as a simple linear regression model:

$$Y_{ij} = \beta_0 + \beta_1 X_i + \epsilon_{ij}$$

In simple linear regression, it is well known that the leverages may be written as

$$h_{ii} = \frac{1}{n} + \frac{(X_i - \overline{X})^2}{\sum_{j=1}^{n}(X_j - \overline{X})^2}$$

Obviously, the leverages will be equal only if all of the X_i are equidistant from $\overline{X}$, and that will happen only if n is even and half of the design points are at one value and the other half at the other value.

It can also be seen that $\text{Var}(\widehat{\beta}_1)$ is minimized with $n/2$ design points at each of the two X_i values, as becomes apparent when we recall that $\text{Var}(\widehat{\beta}_1) = \sigma^2 / \sum(\text{X} - \overline{X})^2$. It could also be shown that $\text{Var}(\widehat{\beta}_0)$ is independent of the choice of the X_i. (As an aside, if a no-intercept model had been fit, the leverages would still be equal but would of course be different from the leverages when the intercept is in the model.)

Unfortunately, however, this design would not permit the detection of nonlinearity, and we generally want to use designs that allow us to detect model misspecification since all models are wrong. (Model-robust designs are discussed briefly in Section 13.20.) If, however, we start from an equileverage design, and then make suitable modifications to facilitate the detection of possible nonlinearity, we will be able to avoid high leverage points.

13.6.2 Are Commonly Used Designs Equileverage?

Since high leverages are outliers in the design space, we might not suspect that there are designs with high leverage values. It is true that many of the commonly used designs are equileverage designs. For example, 2^k and 2^{k-p} designs are equileverage designs for any model that can be fit with one of these designs provided that no centerpoints are used, and the same can be said of Box–Behnken designs without centerpoints. Similarly, central composite designs, discussed in Section 10.5, are

equileverage designs if no centerpoints are used, but of course centerpoints are used with central composite designs.

In general, the equileverage property is lost for factorial designs with more than two levels (but not always), as well as for other types of designs with more than a few levels. For example, the leverages of a 5^2 design for fitting only the linear terms are .04, .06, .08, .12, .14, and .20. Thus, the leverages differ considerably, although .20 would not be considered a high leverage value using any reasonable rule of thumb.

Given below is a table of designs and the corresponding leverages, with the number after the leverage value indicating how many times the indicated leverage value occurs, with the design used to fit only linear terms with no interactions.

Design	Leverages
2×3	(.5833)4; (.3333)2
$2 \times 2 \times 3$	(.25)4; (.375)8
$2 \times 2 \times 2 \times 3$	(.1667)8; (.2292)16

Notice as we move down the table that the design moves closer to being equileverage as the factor that spoils the equileverage property (i.e., the last one) becomes a progressively smaller part of the design.

Each of the last two designs is an equileverage design for fitting the main effects and all two-factor interactions as the leverages are all 0.8333 for the first design and 0.625 for the second design. (There are not enough degrees of freedom to fit the interaction term in the first model as doing so would use all of the five degrees of freedom.)

13.7 OPTIMAL DESIGNS

Although not as popular as two-level full and fractional factorial designs, optimal designs have been used in many applications and are especially useful where there are restricted regions of operability (see Section 13.8). Various statistical software packages have the capability of producing designs that are optimal for a given optimality criterion. (These are discussed in the next section.)

Optimal designs are not without their critics, however, and an optimal design won't necessarily be the best design to use in a given situation. For example, consider a single covariate and a response variable, and a simple linear regression model is posited as being a good model that represents their relationship. The design that minimizes the variance of the slope estimate has half of the design points at one value of the covariate and the other half at the other value, as indicated previously. A regression equation (fitted line) would not be very practical as, in particular, it would not allow for nonlinearity between the two points to be tested. A similar point is made by Fedorov and Leonov (2005) regarding logistic regression models. Matthews and James (2005) make the important point that in general applications of optimal designs, the number of distinct points will usually be limited, so there may not be enough distinct data points to permit model verification.

In addition to designs that are optimal in terms of a statistical criterion, designs can also be constructed that are optimal in terms of cost. These are considered in Section 13.12.

13.7.1 Alphabetic Optimality

Some of the properties that a good design should possess were listed in Section 1.2. There has been much discussion in the literature about construction of designs that meet a certain optimality criterion, and the collection of such criteria has been informally referred to as alphabetic optimality. For example, letting $\mathbf{X}$ denote the design matrix, for a collection of $m > n$ candidate design points, the set of n points that maximizes $|\mathbf{X}'\mathbf{X}|$ is the D-optimal design, *for that set of candidate points*.

Since the design matrix is part of the criterion, this means that a design that meets the criterion is D-optimal only if the model that is assumed is correct, but of course the model is never correct.

So there are two problems: a design can be D-optimal only for the candidate points that the experimenter has in mind, but the guarantee of D-optimality for those points hinges on the false assumption that the model is correct. (See Myers and Montgomery (1995, p. 384) for a somewhat similar but more extensive discussion than that given here.) It is important to note, however, that the criticism of optimality strictly applying only for the postulated model has motivated research on model-robust optimal designs, so not all optimal designs have this model-dependency weakness. In particular, DuMouchel and Jones (1994) addressed the model-dependency problem for D-optimal designs.

Another problem with D-optimal designs, as pointed out by Montgomery, Loredo, Jearkpaporn, and Testik (2002), is that if we start with a standard design such as a 2^3, which is in fact the D-optimal design for 3 two-level factors, as can be verified with experimental design software such as Design-Expert, and then request that one point be added at a time to satisfy the D-optimal criterion, the points that are added won't necessarily be the points that an experimenter would want to add. Specifically, if a few points were to be added to a 2^3 design, conventional thinking is that the points should be added to the center so that lack of fit can be tested and the experimental error variance can be estimated. This does not happen, however, as when one point is added, the coordinates are $(-1, -1, -1)$ which of course is a corner point. If another point were added, it would have the same coordinates, as would a third added point and a fourth added point. Thus, the corner point continues to be replicated. This does produce a pure error estimate with four degrees of freedom for $n = 12$, but we might question having that estimate obtained from a point that is well removed from some of the other points rather than being at the center of the design space. Certainly, the latter would generally be preferable.

For these reasons, a D-optimal design is certainly subject to criticism. One alternative is to use a Bayesian D-optimal design, which can be used to check for adequacy of the fitted model. Let k denote the number of primary terms that are to be used in the model, and let q denote the number of potential terms. A Bayesian D-optimal design maximizes the determinant of $(\boldsymbol{X}^{\mathbf{T}}\boldsymbol{X} + \boldsymbol{K})$, with K a $(p + q) \times (p + q)$ diagonal

matrix whose first p diagonal elements are equal to zero and last q diagonal elements equal to 1. The latter is how the Bayesian D-optimal design uses the potential terms to force in runs that permit possible model inadequacy of the model containing only the primary terms to be checked.

This design can be generated using JMP and is illustrated in the help file for "Bayesian D-optimal Designs" for the simple case of a 2^2 design with the potential terms being the two pure quadratic terms. When five points are requested instead of four and the estimation of the pure quadratic terms specified as "if possible," JMP adds a centerpoint. If six points are requested, an additional point with coordinates (1, 0) is added.

There are other optimality criteria (in particular, G-optimality, A-optimality, E-optimality, I-optimality, and L-optimality), which are discussed in detail in books on optimal design such as Silvey (1980) and Pukelsheim (1993)), but alphabetic optimality has received criticism over the years (and understandably so), such as from Box (1982) and Myers (1999). The latter, in discussing response surface methodology, stated "... in many cases alphabetic optimality is thus not appropriate." Box et al. (2005, p. 460) stated, "The attempt to compress the information about a design into a single number seems mistaken."

As explained by Atkinson and Bailey (2001), the basic optimum design ideas were set forth by Kiefer (1959), whose paper read before the Royal Statistical Society apparently had a rough reception, especially in the written comments.

This is not to suggest that optimum designs should be abandoned, as optimal mixture designs, in particular, have proven useful (Piepel, 2004). In general, optimal designs are valuable when there are constraints on factor-level combinations, as always occurs with mixture designs since the component factors must sum to 100 percent, in addition to the type of constraints discussed in the next section that can occur with nonmixture designs. Under such circumstances, the commonly used designs are no longer applicable, at least not without modification, so automated generation of designs can be quite useful.

Thus, optimal designs are no longer deserving of all the criticism that these designs have received. We should keep in mind, however, that sequential use of designs is usually preferred, whenever possible, and when we emphasize the sequential nature of experimentation, we logically reduce attention on the need for any one design to be optimum in any sense.

Optimal designs must be constructed by computer and one of the best ways of doing so is by using GOSSET, an extremely comprehensive experimental design program developed by Neil Sloane and Ron Hardin at AT&T. Originally proprietary software, it is now freeware. It does not run on all major platforms, however, as it runs only on the Unix, Linux, and Mac operating systems. It can be used to generate designs that are I-, A-, D-, or E-optimal. Other notable software programs that will generate optimal designs include Design-Expert, which will generate D-optimal designs using the CONVERT algorithm (Piepel, 1988) to find the vertices of the design. For example, the design given in Table 13.1 for five factors at two levels each was generated using Design-Expert, assuming a model with all main effects and two-factor interactions. Notice that 21 design points were generated from $2^5 = 32$ candidate points, the number

TABLE 13.1 D-optimal Design for five Factors Generated Using Design-Expert

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

of points being determined by the algorithm and not by the user. (This is not true for all software packages that generate optimal designs, however.)

Of course the design is not orthogonal, as it obviously could not be for an odd number of design points and an even number of levels, but the departure from an orthogonal design is not great. (This can be demonstrated by using the five columns for the design and computing the "correlation matrix." Specifically, all of the pairwise correlations are .028 and −.028.)

A method for generating large D-efficient designs was given by Kuhfeld and Tobias (2005). (D-efficiency is explained at the end of this paragraph.) The authors presented "... experimental design tools that originated in industrial statistics but have been significantly extended to handle problems in marketing research." The approach they gave is claimed to generate 115,208 orthogonal arrays. The Appendix of their paper contains a large number of the orthogonal array parent designs, ranging from 4 runs to 513. (Note that these designs were termed D-efficient rather than D-optimal. It is commonplace to describe a "near-optimal" design, using a measure of the efficiency of the design relative to a, perhaps hypothetical, orthogonal design of the same size. Of course for a very large number of factors and a huge number of design points, we should not expect to be able to construct an optimal design with any algorithm. The authors refer to one instance in which their macro ran for 2 hours and produced a

design with a D-efficiency of 80.521%. A formal definition of D-efficiency is given by Waterhouse (2005, p. 9). Simply stated, it is the determinant of a given design matrix divided by the determinant of the D-optimal design, for a given size, with that fraction raised to the $1/p$ power, with p denoting the number of parameters to be estimated.

Heredia-Langner, Carlyle, Montgomery, Borror, and Runger (2003) reviewed the proposed methods for generating optimal designs and presented a genetic algorithm, used in computer science, which has certain advances over other methods. In particular, the candidate points need not be enumerated. Berger and Wong (2005) is a relatively small edited work that contains contributed chapters on applications of optimal designs in various fields. In Chapter 10, "The Optimal Design of Blocked Experiments in Industry," the authors, P. Goos, L. Tack, and M. Vandebroek, contend that it is important to incorporate cost considerations in optimal design theory, while admitting "... the literature on that topic is very scarce." They do provide an example in which they illustrate how an optimal design can be constructed that uses cost information when cost information on hard-to-change factors is available, at least in ratio form relative to the costs of factors that are not hard to change.

It should be noted that there are many potential applications of optimal designs when not all of the relevant factors can be controlled. This issue is addressed by Lopez-Fidalgo and Garcet-Rodríguez (2004) and in the papers that they cited. Applications of optimal designs in various industries are discussed briefly in Section 13.7.2.

13.7.2 Applications of Optimal Designs

Optimal designs seem to be used somewhat infrequently when experimental designs are employed, at least in some fields of application. For example, Dette, Melas, and Strigul (2005) state that "... at present, there are only a few real applications of optimal designs in microbiological practice." An exception is the use of D-optimal designs for identifying parameters of the nitrification model for activated sludge batch experiments (Ossenbruggen, Spanjers, and Klapwik, 1996). Other examples of applications of optimal designs for microbiological models, with references, are given in Table 6.2 of Dette et al. (2005). Matthews and James (2005) made several cautionary comments, similar to those in Section 13.7.1, in discussing the application of optimal designs in the measurement of cerebral blood flow using the Kety–Schmidt technique. In particular, they made the point that the model used in their application was relatively new, so they would prefer to have more design points than the limited number of points that generally occurs with optimal designs. When there is not much, if any, evidence to support a particular model, a design such as a uniform design (see Section 13.9.1) might be a better choice.

There are various other fields in which optimal designs have been used, as indicated by Berger and Wong (2005). These include toxicology, pharmacokinetic studies, bioavailability studies for compartmental models, rhythmometry, cancer research, and various other fields. Although this may sound impressive, we should keep in mind that there is a huge number of statistically designed experiments and undoubtedly experiments in which optimal designs are used are a very small fraction of the total number of experiments.

13.8 DESIGNS FOR RESTRICTED REGIONS OF OPERABILITY

Implicit in the use of factorial and fractional factorial designs is the tacit assumption that the design points correspond to feasible operating conditions. Frequently, this will not be the case, however, as certain combinations of factor levels may not be feasible, and might even cause hazardous conditions, as in chemical reactions. Kennard and Stone (1969) were apparently the first to discuss irregular experimental regions in the literature; Snee (1985a) gave 10 examples of experiments with irregular regions and illustrated those regions graphically; Czitrom (2003, pp. 20–24) described three experiments for which there was a restricted factor space; and Bingham (1997) described an application to fastener quality in which certain factor levels could not be used together. Similarly, Cheng and Li (1993) described an experiment conducted in Taiwan in 1990 in which no measurements could be made when all of the factors were at their lowest levels because a bar-code verifier couldn't detect a concentration. Montgomery et al. (2002) discussed experimental designs for restricted regions in terms of design efficiences. Giesbrecht and Gumpertz (2004, p. 413) state "In practice, it is more common to find severe constraints that limit the investigation to irregular regions." Those constraints will usually be in the form of linear inequality constraints, but Reece (2003, p. 386) reports encountering complex constraint relationships between factors that require second-order expressions to describe.

Example 13.1

Reece and Shenasa (1997) gave an example of an experiment for which there was a restricted region of operability that unfortunately was not discovered until the design for the experiment had been constructed. Specifically, a 2_{IV}^{7-3} design with five centerpoint replicates was constructed. (It is worth noting, especially relative to the message in Section 1.7, that prior analysis of data from the relevant process showed that it was stable and suitable for experimentation.)

Before the experiment was run, the pumping capacity of the furnace system that was to be used in the experiment was checked. It was found that the system could not provide 150 MTorr units of pressure when the total flow of gases was 400 SCCM. Consequently, the four treatment combinations that had this combination of levels of these two factors had to be either removed or modified. The team of experimenters decided to modify those four treatment combinations in such a way that the total flow of gasses was set at 200 whenever the pressure was 150. Six addition runs were added, using the RS/1 software, so as to create a D-optimal design. Only five of these runs were used, however, due to time constraints.

Example 13.2

Another example of a restricted factor space (or *debarred observations*, using terminology, although perhaps not the best, which has been used in the literature) was

given by Buncick, Ralston, Denton, and Bisgaard (2000). There were five factors, and a 2_{V}^{5-1} design with $I = ABCDE$ was originally selected for use. Four of the 16 treatment combinations had to be eliminated, however, for the following reason. Two of the factors were heat (none, 160) and substrate holder rotation (off, on). When there was no rotation, it was not possible to accurately position the substrate over the sputter source after the substrate was heated. Of course this creates an unbalanced design so the decision was made to use only the runs at the high level of the sputter deposition power setting. The number of runs used would thus be eight (half of 16) minus the number of debarred observations that were at the high level of that factor, which was 2. Thus, the number of runs was cut to 6. The runs were analyzed as two 2_{III}^{3-1} designs with two overlapping runs. Thus, the debarred observations presented a complication that required a method of dealing with it.

Since restricted factor spaces frequently occur, methods of constructing designs under such conditions should be carefully considered. Kennard and Stone (1969) used their CADEX algorithm to select design points to uniformly cover the experimental region. This does not mean, however, that we can automatically obtain a design with good properties by using such algorithms, as a good design may not be feasible. This is illustrated later in the section.

Zahran, Anderson-Cook, Myers, and Smith (2003) considered the modification of a 2^2 factorial design to accommodate restricted design spaces. Specifically, they considered the case where the design point $(-1, 1)$ is not feasible and must be replaced. Replacing that point but retaining the other three points will result in a nonorthogonal design, so the experimenter will have to choose between the nonorthogonality that results from replacing only the impossible point, or changing three of the four points but retaining orthogonality. The latter is achieved by moving just outside the region of nonoperability, then changing two of the other three points so as to create a rectangle.

For example, assume that the impossible point is (1, 1), as in Zahran et al. (2003). In order to create an orthogonal design with minimal disturbance of the coordinates of two of the other three points, we would need to know more than just the fact that (1, 1) is an impossible point. For example, is (.9, .9) also an impossible point?

Assume that the region of inoperability is given by the square that is enclosed by the points (.95, .95), (.95, 1), (1, .95), and (1, 1). A very simple way to construct an orthogonal design would be to use, for example, the points (.94, .94), (−.94, .94), (.94,−.94), and (−.94, −.94), which would then of course be rescaled so that the $(-1, +1)$ notation could be used. This would be accomplished by using a scale factor of 0.94. Of course with the rescaling, the points would correspond to the factor values in the original units to which the (−.94, +.94) notation corresponds in the coded units.

To illustrate, assume that the original units were temperature (°F), with values of 300 and 400, and pressure (in psi) with values of 200 and 300. Then .94 for temperature would correspond to $.94(50) + 350 = 397$ and −.94 would correspond to $-.94(50) + 350 = 303$. The pressure values would be similarly obtained and can be shown to be 203 and 297, as should be intuitively apparent.

A possible objection to this design from an experimenter would be that all of the design points are odd numbers, and another possible objection is that each coordinate of the other three points is changed, albeit slightly.

Zahran et al. (2003) did not consider this approach to altering the original design points. Instead, they stated that an orthogonal design could be constructed using the design points as $(-1, -1)$, $(-1, a)$, $(a, -1)$, and (a, a). But such a design, *without recentering*, is not orthogonal unless $a = 1$, which would be the original design. The recentering would have to convert the center of the design from the original (0, 0) to $[(a - 1)/2, (a - 1)/2]$. Then the scaling factor would have to be $(a + 1)/2$, with the latter being half the distance between the coordinates for each factor. Then by using the coordinate values for each factor (i.e., a and -1), we can easily see that we obtain the usual $+1$ and -1 for each coordinate value with the appropriate scaling.

Specifically, $[(a - (a - 1)/2]/[(a + 1)/2] = 1$, and $[(-1 - (a - 1)/2]/[(a + 1)/2] = -1$. Thus, the design points given by Zahran et al. (2003) really do constitute an orthogonal design, but only if we view these points as not the coded values but rather view them as coordinate points which when properly coded will produce the usual $+1$ and -1 values. One way to view the design points given by Zahran et al. (2003) is to view them as uncoded values (since the center is not (0, 0)), so that the same coding convention would be applied as is applied when the starting point is original units of factors. Another way of looking at it is the following: if the four points are the corner points of a rectangle, then an orthogonal design can be constructed with appropriate centering and scaling. Think of it this way, if we have two variables (factors) whose pairs of values form a rectangle, then there is complete absence of a linear relationship between the variables. Then the correlation between the two variables would be zero, which would mean that the variables could be made orthogonal with appropriate center and scaling.

One thing to keep in mind is that if centerpoints are to be used, they will not be the original centerpoints because those points will not be in the center of the rescaled design. For example, using the temperature and pressure values, the original center would have been 350 for temperature and 250 for pressure. If, as in the example above, we use (.94, .94) as the coded values, then those coordinates would correspond to (397, 297). Similarly, the other points would have original-unit coordinates of (300, 200), (300, 297), and (397, 200). The center of $[(a - 1)/2, (a - 1)/2]$ is thus $(-.003, -.003)$. This must of course correspond to (348.5, 248.5) in the original units, and we can see that the latter is the center for this set of four points.

We can avoid these entanglements if we simply work with the raw units. Zahran and Anderson-Cook (2003), who present a more general equation for determining the coordinates of the point that replaces the ineligible point, gave an example with two factors that we will use for illustration. The design region was stated as 250–300 for the first factor and 50–75 for the second factor, but values of the first factor above 270 and values of the second factor greater than 70 were believed to cause problems. This simply means that a 2^2 design would be constructed using 250 and 270 for the first factor and 50 and 70 for the second factor.

The JMP-generated 2^2 design for this scenario is given below:

Run	X1	X2
1	250	50
2	270	50
3	250	70
4	270	70

This design can be generated by starting with the desired factor levels and then entering the appropriate constraints, or more directly, simply starting with (250, 270) and (50, 70), respectively, as the levels of the two factors.

Zahran and Anderson-Cook (2003) gave the restricted region in terms of coordinates corresponding to coordinates of the endpoints of the unrestricted region being (0, 0), (0, 1), (1, 0), and (1, 1). The corresponding coordinates of the restricted region were shown to be (0, 0), (0, 0.8), (0.7, 0), and (0.7, 0.8). (Similarly, the coordinates could be written using the more customary $(-1, 1)$ range, which would be $(-1, -1)$, $(-1, 0.6)$, $(0.4, -1)$, and $(0.4, 0.6)$. Of course these numbers also look a bit unorthodox, and coding would have to be performed to produce the more customary set of four coordinates, with each coordinate being either 1 or -1.

Regardless of how we write the coordinates, we have a 2^2 design—and a rectangular region—so the design is orthogonal. If the coordinates are written as in the last set, however, the values would have to be converted to the raw units before the experiment could be run. Since the penultimate set of coordinates would be obtained as $X1^* = (X1 - 250)/(200/7)$ and $X2^* = (X2 - 50)/(25)$, the values of $X1$ would thus be obtained as $(200/7)X1^* + 250$, and the values of $X2$ would be obtained as $25X2^* + 50$. Again, we can avoid the unnecessary algebra just by working with the raw units.

Zahran et al. (2003) also considered ways to construct a design with four or five points to accommodate the restricted region that satisfy each of a few other optimality criteria. Specifically, they focused attention on GH-optimality and Q-optimality, in addition to D-optimality. The Gendex DOE toolkit can be used to construct designs for restricted regions.

This issue of a restricted factor space was also addressed by Cheng and Li (1993), who considered the construction of orthogonal fractional factorial designs under such conditions. Consider the case of a single combination of factor levels that is infeasible. If this combination involved all of the factors, the solution to the problem would be trivial: simply do not use the fraction that contains the infeasible combination.

The problem can be complex, however, if not all of the factors are involved in the combination. For example, assume that a 2_{V}^{5-1} design is to be used for factors A–E and that the high levels of factors A and C cannot be used together, this being the only restriction. Of course, "high" is arbitrary, so the first question to ask is "can the high levels be reduced slightly and the pair of high levels still be used together?" If a drastic reduction would be necessary, then it would be best to determine if this combination can be avoided by using one of the two fractions. Of course the defining relation would be $I = ABCDE$. There are eight combinations with A and C at their

high levels, so the question then is whether or not they would all be in the same fraction.

There are at least two ways to look at this. Since the two fractions are defined by the treatment combinations that have an odd number of letters in common with the defining relation and an even number of letters in common with it, respectively, we can quickly see that not all of the eight treatment combinations will be in one fraction (e.g., ac is even but abc is odd). Thus, the defining relation $I = ABCDE$ won't work, nor would any other reasonable choice for the defining relation. Using Proposition 1 of Cheng and Li (1993), we arrive at the conclusion that the defining relation would have to be $I = -\text{AC}$, which of course would be totally unacceptable as this would confound main effects. Thus, we cannot construct a 2^{5-1}_{V} design.

Of course a 2^{5-2}_{IV} design is a less-desirable design, but some experimentation (or trial and error) would show that for each possible defining relation that produces a set of four fractions, the requirement of not having A and C at their high levels cannot be met. Stated differently, none of the four fractions, which of course each contains eight treatment combinations, contains all eight of the ineligible combinations. Thus, there is not a practical fractional factorial that would accommodate this restriction, as we would have to drop back to an entirely unsatisfactory defining relation such as $I = A = C = AC$ in order to be able to construct a fraction that contains the eight treatment combinations. (Here the eight treatment combinations with A and C at their high levels would be in one fraction, the odd–odd fraction, so any of the other three fractions would work, although being totally impractical since we would not be able to estimate the main effects of A and C.)

Is the problem of constructing fractional factorials to accommodate unacceptable combinations harder when there are more factors involved in the ineligible combinations? Not necessarily. After much work, Cheng and Li (1993) showed that there is one 2^{5-1} design that will accommodate two ineligible combinations, one involving three of the five factors and the other involving four of the factors.

Clearly, an algorithmic approach is needed for determining designs that will accommodate ineligible combinations, and Cheng and Li (1993) described how the Franklin and Bailey (1977) algorithm would have to be modified to accomplish that. Determining feasible designs in the presence of ineligible combinations is a difficult task, especially in the relative absence of software that will accomplish this.

JMP, however, permits the specification of a restricted design space and this can be used to specify ineligible combinations. As a simple example, consider a 2^2 design, which also happens to be D-optimal on the interval $(-1, 1)$ on each factor. If the "Custom Design" feature is invoked and "D-optimal" is selected, and if the constraint $X1 + X2 \geq -1$ is entered, then the design point $(-1, -1)$ is replaced by the point $(0, -1)$. Now consider a 2^4 design that must be modified as the scientists say that both $X1$ and $X2$ cannot be at the high level because of physical considerations. Again the Custom Design feature must be selected and when the constraint $X1 + X2 \leq 1$ is entered, the design that is returned is the 2^4 design except that three values of $X2$ are changed from 1 to 0 when $X1 = 1$. If a D-optimal design is specified with this constraint and JMP is used, the design that is produced is the one that is given below:

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

Notice that the sum of the levels of the first two factors is indeed less than two and that there is great imbalance in the levels of the first two factors, especially the first factor, for which the high level occurs only four times in the 16 runs. This is the type of price that is paid when a D-optimal design is used with constraints. Notice also that the correlation between the columns for the first two factors is −.447, with the correlation induced by the constraint. (This is the only nonzero correlation, however.)

As indicated previously, just having a suitable algorithm won't necessarily be sufficient, as a suitable orthogonal design may not exist. Thus, nonorthogonal designs will undoubtedly have to be considered in many applications. Supersaturated designs might also have to be considered because if there are too many debarred combinations, the number of feasible combinations might be less than the number of parameters in a tentative model. Of course the magnitude of the correlations between the parameter estimates will have to be assessed and carefully considered if such a design is to be used, and effect estimates will be sacrificed since effect estimates don't have any meaning for nonorthogonal designs.

One potential danger when infeasible treatment combinations exist and a suitable design does not exist is that if we back off too far from a region of infeasibility because no practical design exists with the ineligible combinations, we may fail to detect an interaction or two that exists. As a simple example, assume that treatment combination ab is infeasible with a 2^2 design and that an interaction plot would show two lines that would be a considerable distance apart in the vicinity of the ab combination (but not at the other end) if we backed off only very slightly from the raw unit values to which ab corresponds and then constructed a 2^2 design for a slightly smaller factor space. Now assume that we back off considerably from that point so that the factor space is slightly more than half of the original space. Now the lines may not have nearly the vertical separation in the vicinity of the "new" ab combination as was the case previously.

For example, consider Figure 13.2. The moderate interaction is due to the ab combination. If we think about shrinking the factor space so that the factor-level

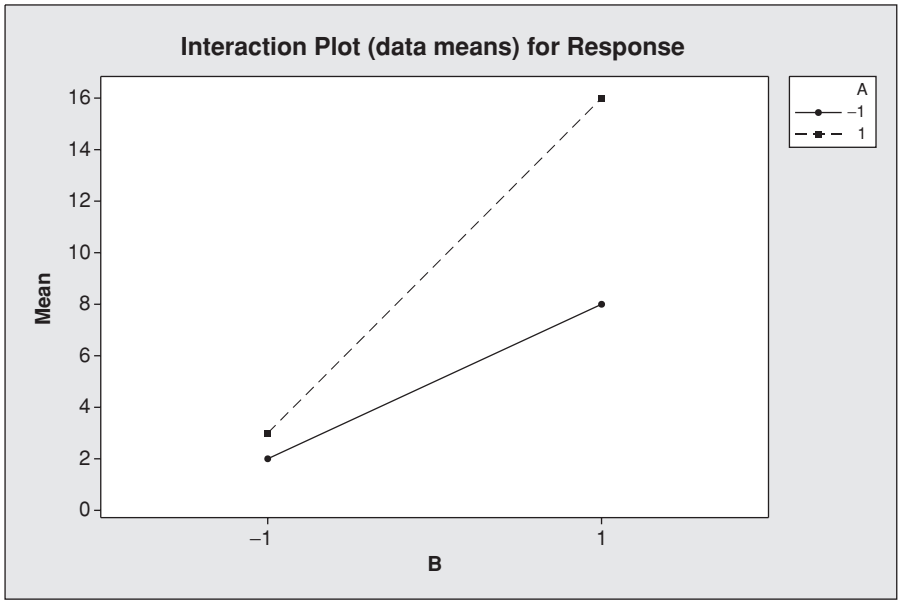

Figure 13.2 2^2 design showing moderate interaction.

combinations (1, 1) and (1, −1) are replaced by (0, 1) and (0, −1), respectively, before rescaling, then there would be only a weak interaction signal.

Obviously, there is a need for research on this topic so that the best design for a given combination of debarred combinations can be identified. In particular, for fractional factorial designs it would be of interest to identify the best fractional factorial design that can be constructed for a given set of debarred combinations, and compare the properties of that design against the properties of the design that results from excluding the debarred combinations from the 2^k runs.

13.9 SPACE-FILLING DESIGNS

Space-filling designs are useful when there is little or no information about the effects of factors on the response but the true model is suspected to be highly nonlinear. These designs are useful for modeling systems that are deterministic or near-deterministic. For such modeling situations, variability is of little or no concern because variability is believed to be practically nonexistent.

A *uniform design* is one type of space-filling design. When there is no prior information, the points might as well be regularly spaced over the design region, which is essentially what is accomplished with these designs. These designs, when used in conjunction with nonparametric regression and possibly semiparametric regression, are potentially beneficial in many applications and industries, including the

pharmaceutical, biotechnological, chemical, and process industries. They are also useful in computer experiments when the postulated model is quite complex.

13.9.1 Uniform Designs

This is one type of space-filling design that was originally proposed by K.-T. Fang and Y. Wang in 1980 and presented in Fang (1980) and Wang and Fang (1981). As stated by Fang, Lin, Winker, and Zhang (2000), the design can be considered to be a "nonparametric experimental design." The design was originally motivated by three system engineering projects in 1978 (Fang and Lin, 2003). According to Professor Fang, uniform designs have been used in various fields such as chemistry and chemical engineering, pharmaceutics, quality engineering, system engineering, survey design, computer sciences, and natural sciences. Fang et al. (2000) gave a case study of "launching a dynamic system for a one-off ship-based detector."

The concept of the uniform design is quite simple. To illustrate, assume that we have a single factor and n design points are to be used, with the points uniformly spaced. Assume that the design space on the single factor is (0, 1). The points are to be $(1/n)$ units apart, so that the distance between the first point and the nth point is thus $(n-1)/n$. Since the set of n points should obviously be centered in the interval (0, 1), it follows that with a denoting the first point, $a + (n-1)/n + b = 1$, with b denoting the distance between the nth point and 1. Similarly, $a - b = 0$. Solving these two equations gives $a = 1/2n$. Thus, the design points are $\frac{1}{2n}, \frac{3}{2n}, \frac{5}{2n}, \ldots, \frac{2n-1}{2n}$, which can be written as $\frac{2i-1}{2n}$, $i = 1, 2, \ldots, n$.

If we wanted to use the usual $(-1, 1)$ as the design space for the factor, then the distance between the points should be $2/n$ (double the previous distance since the design space has been doubled), with the first equation now being $a + 2(n-1)/n + b = 1$ and the second equation to be solved is $a - b = -1$. Solving these two equations thus gives $a = -\frac{(n-1)}{n}$. The design points are thus $-\frac{(n-1)}{n}, -\frac{n-3}{n}, \ldots, \frac{n-1}{n}$, which can be written in the form $-\frac{(n-(2i-1)}{n}$, $i = 1, 2, \ldots n$.

Note that the selection of the number of levels of each factor to use determines the number of runs since each level is used only once.

Many uniform designs are also orthogonal designs; many others are near-orthogonal. Although the uniform design is a simple and easy-to-understand concept in one dimension, it is neither in more than one dimension. In fact, as pointed out by Fang and Lin (2003), the definition of a uniform design that they give in their Eq. (5.1) is not uniquely satisfied when there are at least two factors.

Their definition of a uniform design is as follows. Let s denote the number of factors and define the experimental region as the unit cube $C^s = [0, 1]^s$. Let $\mathcal{P} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\} \subset C^s$ denote the set of n experimental points, with $\boldsymbol{x}_i$ denoting a vector of values that gives the ith value of each factor. Let M be a measure of uniformity of $\mathcal{P}$ with a small value of M desirable since such values represent better uniformity than larger values. Let $Z(n, s)$ denote the collection of sets of n points on C^s. A set $\mathcal{P}^* \in Z(n, s)$ is termed a uniform design if it has the minimum M-value over $Z(n,s)$.

As Fang and Lin (2003) explain, many different measures of uniformity have been defined (see pp. 142–143 of their chapter).

The construction of uniform designs for at least two factors requires either software, a catalog, or an applet. One of the latter is given at http://www.math.hkbu.edu.hk/UniformDesign. The output from the latter is in the form that orthogonal arrays are usually given, with 1 being the lowest level for each factor, and the highest level is n. This is the same form that is used by Fang and Lin (2003), and the designs are referred to as U-type (uniform) designs.

If the user prefers to see the design given such that each level is contained in the interval $(-1, 1)$, then it would be better to use appropriate software. For example, JMP allows the user to specify the interval when space-filling designs are constructed, and the following JMP output gives a uniform design for four points and two factors, with the design space for each factor specified as $(-1, 1)$. (JMP has only somewhat recently added uniform design capability.)

```
Space Filling Design

Factors

X1   Continuous

X2   Continuous

2 Factors
Space Filling Uniform Design

Factor Settings

Run        X1          X2
1       -0.73363   -0.24587
2        0.73363    0.24587
3       -0.24587    0.73363
4        0.24587   -0.73363
```

If the interval (1, 4) is specified, JMP constructs the design given below:

```
Factor Settings

Run        X1          X2
1       2.86881     1.39956
2       1.39956     2.13119
3       3.60044     2.86881
4       2.13119     3.60044
```

This is different from the U-type uniform design mentioned previously, thus illustrating the nonuniqueness of uniform designs for at least two factors.

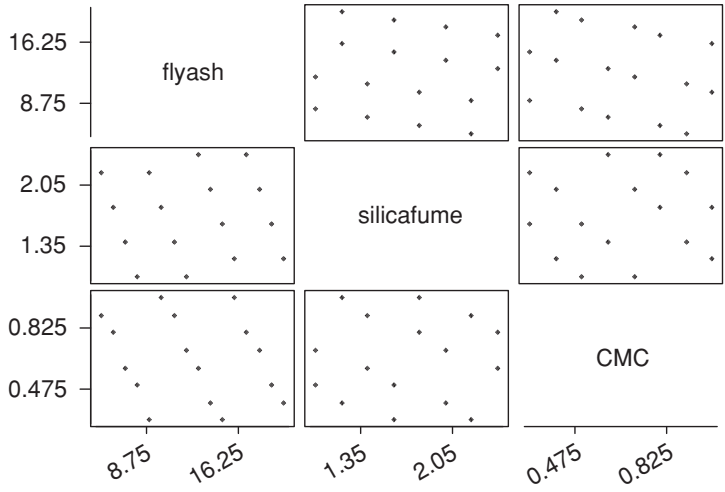

Figure 13.3 Uniform design configuration for design in Tang et al. (2004).

Case Study

Tang, Li, Chan, and Lin (2004) described a successful application of a uniform design in product formation in the cement manufacturing industry. The objective in the experiment they described was to determine an optimal composition of materials so that the cement grouting material that was formed had the desired properties. There were two response variables, one a measure of compressive strength and the other a coefficient of bleeding, and there were four factors: the percentages of fly ash, silica fume, carboxyl methyl cellulose, and cement. Since these are percentages, this is really a mixture experiment with only three of the percentages free to vary. Therefore, the design consisted of the first three factors.

A uniform design was used because 16 levels were used for the first factor, 8 levels for the second factor, and 8 factors for the third factor. The large number of levels for each factor rules out traditional mixture designs, such as those discussed in Section 13.11. Instead, a uniform design $U_{16}(16 \times 8^2)$ with 16 runs was used, with each factor constrained as follows:

$$5 \le x_1 \le 20 \qquad 1 \le x_2 \le 2.4 \qquad 0.3 \le x_3 \le 1.0$$

The uniformity of the design in two dimensions can be seen in the matrix scatterplot as given in Figure 13.3.

The run numbers given in Table 1 of Tang et al. (2004) show that the runs were not randomized as the percentage for the first factor, fly ash, ran consecutively from 5 to 20. Factors having differing numbers of levels makes this a somewhat unorthodox uniform design that would have to be constructed using special software or tables. JMP, for example, which has the capability for uniform design construction, cannot construct uniform designs with differing numbers of factor levels. This can be done in

Design-Expert, but technically the latter does not construct uniform designs. Rather, it creates "distance-based" designs that attempt to spread the design points evenly over the region but it does so by maximizing the distance between points. The designs that are generated have coordinates that are quite different from the coordinates in the designs produced by JMP. The Design-Expert documentation advises against the use of these designs, however. The differing numbers of levels of course means that some levels of the factors with eight levels will be repeated.

We will not analyze the data from the experiment in this section since this was a mixture experiment, but will note that the apparent lack of randomization regarding the runs does, strictly speaking, invalidate the model hypothesis tests that were performed for each of the two response variables, although that is not of any great concern since the p-values were of the order of 10^{-10} and smaller and these tests were not a major component of the article.

To what extent are uniform designs used in practice? This may be debatable but the following quote given in Freeman (2004, p. 38), which is available online at http://cipd.mit.edu/education/thesis_files/ion_freeman_thesis_2004.pdf with this quote resulting from one of the phone interviews of practitioners, is worth pondering: "Orthogonal array can require too many experiments, and not give the best coverage. I don't need orthogonality; I need the best information. Uniform design techniques are far more popular in the practice." The last statement is probably not true but may be true 20 years from now.

13.9.1.1 From Raw Form to Coded Form

There is an easy and obvious correspondence between the levels of a factor in raw units and in coded units when a U-type uniform design is used. For example, let's assume that an experimenter wishes to use temperature levels of 340, 370, 400, and 430°F. The coded values would be simply 1, 2, 3, and 4. If 420 were to be used instead of 430, a uniform design would not be possible because of the unequal spacing that would result from the switch.

We have a different situation if we use the values in the JMP output as those values are not equally spaced on each factor since this is not a U-type uniform design. Therefore, the temperature levels of 340, 370, 400, and 430 cannot be transformed to those coded values since the temperature values are equally spaced. There is such a small difference, however, that these temperature values could be used anyway.

Uniform space-filling designs essentially mimic the continuous uniform distribution for a given factor, so we can think of points for two factors as essentially mimicking a bivariate continuous uniform distribution.

13.9.2 Sphere-Packing Designs

Uniform designs can be contrasted with "sphere-packing" designs, which maximize the minimum Euclidean distance between the points. See, for example, the discussion given at http://www.warwick.ac.uk/statsdept/staff/JEHS/jehpack.htm. For example, a sphere-packing design for two factors and four points is simply the 2^2 design; for three factors and eight points, it is the 2^3 design; and so on. Of course, this is because

the design points for these factorial designs are at the vertices of the spherical region and are thus the maximum distance apart. The answer is not so obvious if the number of design points is not of the form 2^k, however.

For example, the following is the sphere-packing design for six runs and two factors given by JMP:

```
X1 Continuous
X2 Continuous

2 Factors
Space Filling Sphere Packing

Factor Settings

Run        X1          X2
1      -0.33333     1.00000
2       1.00000     1.00000
3       1.00000    -1.00000
4      -0.33333    -1.00000
5      -1.00000     0.00000
6       0.33333    -0.00000
```

13.9.3 Latin Hypercube Design

The Latin hypercube design (LHD) is the other commonly used space-filling design, and it is essentially a compromise between the uniform design and the sphere-packing design. A design of this type is constructed by dividing the desired range for each factor into n intervals and then randomly selecting a point within each interval. A randomly selected point for the first factor is then matched with a randomly selected point for the second factor and subsequent factors, and then the process is repeated $n - 1$ times until the coordinates for all n points have been obtained, with the number of levels of each factor equal to the number of runs. The design is due to McKay, Beckman, and Conover (1979).

The design for four points and two factors is given below:

```
X1 Continuous
X2 Continuous

2 Factors
Space Filling Latin Hypercube

Factor Settings

Run        X1          X2
1      -0.33333    -1.00000
2      -1.00000     0.33333
3       0.33333     1.00000
4       1.00000    -0.33333
```

Notice that the levels of each factor are equally spaced, which is one of the characteristics of these designs. Notice also that the number of levels of each factor is equal to the number of runs, just as is the case with the uniform design. Although it differs in this respect from the sphere-packing design, it shares with that design the maximizing of the Euclidean distance between design points, but subject to the constraint of having even spacing between the factor levels.

The design with four points is orthogonal because the points form a parallelogram, which will generally not be the case if we use more design points for the two factors, especially if there is an odd number of design points, and similarly when more factors are used.

Case Study

Bailey and Jones (2004) gave a case study that illustrated the use of an LHD, and which could also be viewed as an example that illustrates (many) debarred combinations. The case study, which the authors stated that they intentionally simplified to illustrate space-filling designs and which was perhaps also simplified for a general audience, involved the use of a polymerase chain reaction (PCR). The main purpose of this reaction is to make many copies of native DNA, with the response variable being the amount of the product DNA fragment that is formed. There were eight factors to be studied: primers A, B, and C, annealing temperature, $MgCl_2$, dNTP, polymerase, and cycles. As stated by the authors, the reaction involves repeatedly heating and cooling the reaction mixture, inducing the DNA to melt, and separating the double strands.

Although the scientists could be guided by experience in choosing reasonable ranges for the factors, few combinations of factor levels work for a particular PCR. Since they don't know which combinations will work, they need to sample a large number of levels of each factor.

Accordingly, the design that was used was an LHD with 48 chosen as the number of runs to be made. Of course the objective is to uniformly spread the points over the eight-dimensional space. Although the eight-dimensional coverage is obviously not completely captured by a scatterplot matrix, the matrix that was given in Bailey and Jones (2004) and which is given in Figure 13.4 does provide two-dimensional glimpses into the uniformity.

It is worth noting that some of the factors, including annealing temperature, are hard to change, and Bailey and Jones (2004) pointed out that the actual runs may have differed from those in the design. Because of the way the design points are constructed, replicated points and/or replicated coordinates are rare because the coordinates are generated at random. Consequently, it is not possible to minimize the number of factor-level changes and still adhere to the design. If factors are very difficult or impractical to change, an experimenter might be expected to simply deviate from the design, with the LHD not being an appropriate design in the first place when there are hard-to-change factors.

Latin hypercube designs and related designs are compared in terms of prediction error by Morris, Mitchell, and Ylvisaker (1993). Although LHDs can be generated using some statistical software, such as JMP and MathWorks using the model-based calibration toolbox (see http://www.mathworks.com/products/mbc/

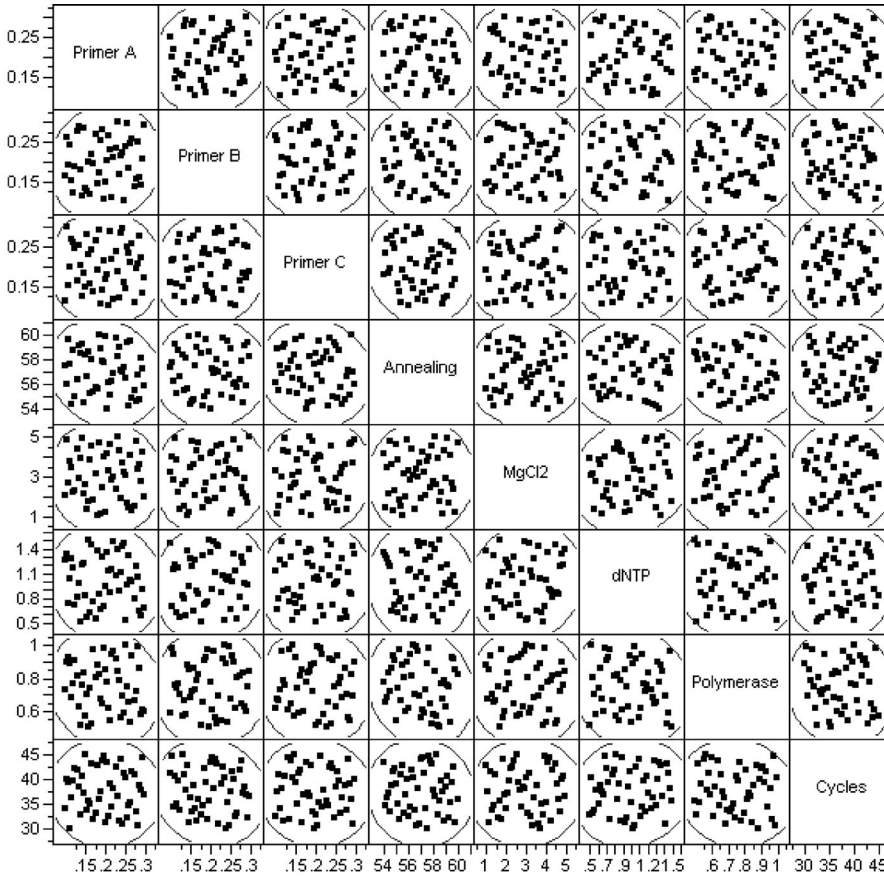

Figure 13.4 Scattering of points in Latin hypercube design.

description1.html and page 67 of http://www.mathworks.com/access/helpdesk/help/pdf_doc/mbc/mbcmodel.pdf), tables of LHDs for up to 22 factors can be found in Cioppa (2002).

13.10 TREND-FREE DESIGNS

Recall the discussion in Section 1.7 about the importance of processes being in a state of statistical control, as the lack of this could undermine a designed experiment in which complete randomization was used.

Consider the following example. Assume that a 2^2 design has been used, without replication, and the randomized run order was (1), b, a, ab. Thus, the combinations at the high level of factor A were run last. A trend, for whatever reason, which causes the response values to increase without a change in the experimental conditions could produce a false signal that factor A was significant.

Assume that control cannot be achieved, and quite frankly, achieving control such that there is an unchanging mean and variance would be extremely difficult, in general. One way to try to overcome the problem would be to use a trend-free design. With such designs, treatments are assigned to experimental units in a systematic manner rather than by using randomization.

There has been comparatively little work on such designs, which essentially began with the work of Cox (1951). A significant early work was Daniel and Wilcoxon (1966), who gave sequences of treatment combinations for full and fractional factorial two-level designs for use in the presence of time trends. Their work was not extended for many years and Daniel (1982, personal communication) wondered if it would ever be extended. Finally, John (1990) introduced a foldover method developed from their plans and discussed the trend-free properties of systematic run orders based on their method. Trend-free block designs were introduced by Bradley and Yeh (1980) and in 1983, they showed that such designs do not always exist. Bradley, Yeh, and Notz (1985) introduced nearly trend-free block designs. Other work in the area of trend-free designs includes Hill (1960).

Trend-free designs can be constructed with the Gendex DOE toolkit.

13.11 COST-MINIMIZING DESIGNS

The explicit consideration of developing designs to minimize costs has received very little attention in the literature, just as cost minimization relative to statistical methodology in general has received only scant attention. Significant early work on cost-minimizing experimental designs was performed by Draper and Stoneman (1968) and Dickinson (1974). Subsequent work followed, most of which is summarized by Kim (1997), who also considered trend-free designs. This dissertation is available at http://scholar.lib.vt.edu/theses/available/etd-7497-15232/unrestricted/etd.PDF. Bisgaard (2000) provides the framework for cost comparisons between split-plot designs and fully randomized designs and gives some illustrative examples. Tack and Vandebroek (2002) considered cost-efficient block designs; Park, Montgomery, Fowler, and Borror (2006) considered cost-constrained G-efficient response surface designs, and Allen, Yu, and Bernshteyn (2000) described a response surface design application when there were cost constraints.

As mentioned in Section 13.8.1, Goos, Tack, and Vandebroek (2005) state that it is important to incorporate cost considerations in optimal design theory but admit that there is not much literature on the subject.

Hard-to-change factors were discussed in Section 4.19. Certainly the prevalence of such factors should motivate the development of cost-efficient designs for hard-to-change factors.

13.12 MIXTURE DESIGNS

Mixture designs are used, as the name practically implies, to aid in the determination of an optimal combination of components for which the total amount is fixed, with

100 percent being the total if the components are expressed as percentages, as is the usual practice.

Since there is a constraint on the total mixture and there will also generally be constraints on the individual components, the experimental space is not rectangular or hyperrectangular. Thus, the types of designs that were given in previous chapters, such as Chapters 4 and 5, are not applicable.

Cornell (2002) is an excellent source of information on mixture designs and the reader is referred to that source for comprehensive information on these designs. Software that is available for the design and analysis of mixture experiments includes MIXSOFT (see, e.g., http://members.aol.com/mixsoft).

13.12.1 Optimal Mixture Designs or Not?

As stated in Section 13.7, optimal designs have been very controversial but have been used advantageously in mixture experiments (Piepel, 2004, personal communication). Mixture designs are used extensively in pharmaceutical applications and there are many documented applications of the use of optimal designs, including Bodea and Leucuta (1997). Some critical comments about optimal designs as mixture designs have been voiced by Myers and Montgomery (1995, p. 595).

13.12.2 ANOM

Just as ANOM can be used for various common designs, as illustrated in Chapters 2 and 4 and discussed in Chapter 3, it can also be used for assessing the significance of components in an axial mixture design, as illustrated by Nelson (1993) and in Chapter 7 of Nelson, Wludyka, and Copeland (2005).

13.13 DESIGN OF MEASUREMENT CAPABILITY STUDIES

An important use of experimental designs is in assessing measurement capability. A typical study involves parts and operators, with parts being random and operators usually random, but fixed if the operators used in the experiment are the same ones who work on the process. Burdick, Borror, and Montgomery (2003) discuss the design of a measurement system study in their review paper, with such a study referred to more specifically as a gage repeatability and reproducibility study.

13.14 DESIGN OF COMPUTER EXPERIMENTS

Engineers frequently encounter complex systems with computer experiments required to optimize such systems. Classical designs, such as factorial designs, are useful when there is a large sample size. Space-filling designs are often an attractive alternative, however, because models used in computer experiments do tend to be rather complex. Chen, Tsui, Barton, and Allen (2003) devote considerable space to the discussion of such designs in computer experiments, and their review paper would be a good

starting point for the study of experimental designs in computer experiments, as would the book by Santner, Williams, and Notz (2003). Simpson, Lin, and Chen (2001) compared space-filling designs and discussed sampling strategies for computer experiments. The latter simply stated that designs for computer experiments should "provide information about all portions of the experimental region," and "allow one to fit a variety of models." More recently, Jin, Chen, and Sudjianto (2005) presented an algorithm that was purported to save considerably on computing time for constructing optimal designs for computer experiments.

A useful user's guide (literally) for designing computer experiments, written at a basic level, is Kleijnen, Sanchez, Lucas, and Cioppa (2005), which is available online and which has a five-page online supplement that contains design matrices for the designs used in the authors' study (see the references for both). See also the comparison study in Bursztyn and Steinberg (2006). For additional reading on this topic, see Fang, Li, and Sudjianto (2006), which includes material on space-filling designs and other topics. See also the case study of Liu and Fang (2006).

13.15 DESIGN OF EXPERIMENTS FOR CATEGORICAL RESPONSE VARIABLES

Although measurement data provide more information than categorical data, the latter (such as go, no-go data) are often used due to the expense of making accurate measurements, faulty measurement systems, and so on. One problem that can occur when data are dichotomized is that the resultant data may not be informative. That is, the data may be highly asymmetric, with the data being almost all of one type. It would obviously be difficult to model data that are, say, 99 percent "go" and 1 percent "no-go," as a design with a sizable number of runs would be required to virtually ensure that some no-go data were obtained. Otherwise, the experiment would be a waste of time. Joseph and Wu (2004) gave an approach, incorporating experimental design, that is intended to overcome this problem and the interested reader is referred to their paper for details, as it won't be discussed here since it is outside the scope of this book. Another article that may be of interest is Bisgaard and Weiss (2002–2003), who took an experiment performed by Taguchi with the response variable being function/malfunction and performed an alternative analysis that provided greater insight.

13.16 WEIGHING DESIGNS AND CALIBRATION DESIGNS

These designs are probably most often associated with the National Institute of Standards and Technology (NIST) and with K. S. Banerjee. In particular, Banerjee (1975) is a well-known book on weighing designs, as is Jones and Schoonover (2002), written by two veteran NIST workers. These books are recommended, as are the seminal articles of Banerjee (1950, 1951), in addition to the many articles of Bronislaw Ceranka,

in particular. There is considerable literature on weighing designs and these designs and calibration designs are covered in the next two sections.

13.16.1 Calibration Designs

The general idea with these designs is to determine the weight of one or more objects. The designs are referred to as *calibration designs* by NIST because their objective is to assign the correct weight to a test item by comparing it to a reference standard, whose weight is assumed to be known. There is an extensive catalog of these designs in the NIST *e-Handbook of Statistical Methods*. There are designs for different purposes and for different applications; see Section 2.3.4 of the *e-Handbook* at http://www.itl.nist.gov/div898/handbook/mpc/section3/mpc34.htm.

Just as statistical process control methods are used in conjunction with the designs discussed in previous chapters, so too are they used with calibration designs, as is discussed in Section 2.3.5 of the *e-Handbook*.

The general form of the solution that provides the estimates of the test weights for a calibration design is a restrained least squares method described in Zelen (1962), with the restraint (or "constraint") being the known weight of a reference standard if only one is used, or the sum of the weights of the reference standards if more than one standard is used. See also Lawson and Hanson (1995). The details are also given in Cameron, Croarkin, and Raybold (1977) and are as follows.

Assume that what is known is a linear combination of weights of reference standards of the form

$$r_1\beta_1 + r_2\beta_2 + \cdots + r_k\beta_k = m$$

with the β_i being the weights that are to be estimated (individually) and the r_i are constants. When least squares without constraints is used, the sum of the squared deviations between the observed and fitted values is minimized. When constrained least squares is used, the method of Lagrangian multipliers is used so that the quantity to be minimized is, say, W, with

$$W = \text{sum of squared deviations} + \lambda(r_1\beta_1 + r_2\beta_2 + \cdots + r_k\beta_k - m)$$

The constrained least squares normal equations are then those that result from minimizing W.

Example 13.3

We will use the simplest example of a calibration design, a (1, 1, 1), to illustrate the design and the mechanics involved. The "(1, 1, 1)" refers to one reference standard that is known, plus a check standard and a test standard with unknown values. The design matrix and observed values, Y_1, Y_2, and Y_3, are as follows:

Reference standard	Check Standard	Test item	Observations
1	-1	0	Y_1
1	0	-1	Y_2
0	1	-1	Y_3

Here the runs are the rows, just as is the case with other statistical designs. Notice that the layout is essentially a balanced incomplete block (BIB) design as each pair of treatments occurs once in each "block" when we view the rows as blocks. A restraint on the solution is, of course, not imposed when a BIB design is used, however.

The solution, which consists of estimates of the unknown values of the check standard and test item, cannot be obtained using standard software, however, because the reference standard is assumed known and is thus not something to be estimated. This is somewhat analogous to estimating main effects and interactions for an experiment with a factorial design when one of the main effects is assumed known.

The solution might seem to be simply the solution to a system of linear equations subject to a restraint that represents the known reference standard, i.e., just by using straight algebra. That isn't the case, however, unless the system of equations is the set of constrained least squares equations. Indeed, as indicated earlier in this section, that is what is done. Furthermore, it should be apparent that the equations that would result from using the design and the data without least squares will generally be inconsistent.

The mechanics for this example won't be given here because they are somewhat involved and also because they are available on the Web (see http://www.itl.nist.gov/div898/handbook/mpc/section3/mpc3321.htm).

Using what is provided therein, it can be shown that the estimate for the test item is

$$\text{Test item estimate} = \frac{1}{3}(-Y_1 - 2Y_2 - Y_3) + \text{Reference standard}$$

Readers are also referred to http://www.itl.nist.gov/div898/handbook/mpc/section3/mpc332.htm for general information on solutions to calibration designs.

13.16.2 Weighing Designs

We may also use these type of designs without reference standards. When this is done, it is more appropriate to refer to the designs as weighing designs, as Banerjee (1975) did in the title of his book. That is, there are k objects to be weighed and a design must be constructed to permit their weights to be estimated.

In general, the criteria for good weighing designs depend partly on the intended use for the resulting values, as explained by Cameron et al. (1977). For example, if the weights are to be used independently of each other, it would be desirable to use a design that gives the smallest variance of each of the effect estimators, whereas if the weights are to be used in combination, it would then be appropriate to use a design that minimizes the variance of a certain linear combination or gives small variances for linear combinations that are likely to be used.

Cameron et al. (1977) gave the following categories of weighing designs: (1) designs for nominally equal groups, (2) designs for the 2, 2, . . . , 1, 1, 1, . . . series, (3) the 5, 3, 2, 1 and 5, 2, 2, 1 series, (4) binary and miscellaneous series, and (5) designs for direct reading with constant load. (The numbers, such as 5, 3, 2, 1, refer to the relationships between the weights in a reference set. For example, the weights might be 5 grams, 3 grams, 2 grams, and 1 gram, or 500, 300, 200, and 100.)

Regarding the first group, designs for nominally equal groups, the (1, 1, 1) calibration design was given in Section 13.16.1. The notation is used differently to designate a weighing design. For the latter, the notation refers to the reference set consisting of three equal weights. The first restraint, as given by Cameron et al. (1977), is that the sum of the first two weights is known and the second restraint is that the third weight is known.

Since there are three weights and all pairwise comparisons are made, there must be three observations since $\binom{3}{2} = 3$. Similarly there are $\binom{4}{2} = 6$ observations for four equal weights, $\binom{5}{2} = 10$ observations for five equal weights, and so on.

Designs for trend elimination were given by Cameron et al. (1977) in juxtaposition to the designs with equal weights. These designs are due to Cameron and Hailes (1974). These designs are used when responses are time dependent due to temperature and atmospheric changes.

The second group of designs, the 2, 2, . . . , 1, 1, 1, . . . series does not possess the simplicity and symmetry of the designs for equal weights. This should be apparent from the 2, 1, 1, 1 design that is given below:

Observations	2	1	1	1
Y_1	+	−		−
Y_2	+		−	−
Y_3	+	−	−	
Y_4		+		−
Y_5		+	−	
Y_6			+	−

The most commonly used designs have been those in the third group, the 5, 3, 2, 1 and 5, 2, 2, 1 designs. The 5, 3, 2, 1, 1 design, which is due to Hayford (1893), is the following:

Observations	5	3	2	1	1
Y_1	+	−	−	+	−
Y_2	+	−	−	−	+
Y_3	+	−	−		
Y_4	+	−		−	−
Y_5		+	−	−	
Y_6		+	−		−
Y_7			+	−	−
Y_8				+	−

Various examples of weighing designs can be found in Banerjee (1951), in particular.

There has been a considerable amount of research on weighing designs since Banerjee (1975). In particular, Craigen (1996) provided a table of weighing design matrices that has since had an addition or two. (A missing entry in that table was provided by Arasu and Torban, 1999.)

13.17 DESIGNS FOR ASSESSING THE CAPABILITY OF A SYSTEM

Hubele, Beaumariage, Baweja, and Hong (1994) discussed the use of experimental designs to assess the capability of a system. The authors described a designed experiment that was used to help determine the capability of a vision system at AlliedSignal. A design with blocking was used, but unlike most block designs, the blocks were not random effects, but rather corresponded to systematic effects. The result was a better understanding of system performance.

13.18 DESIGNS FOR NONLINEAR MODELS

All the designs presented to this point in the book presume the use of a linear model (i.e., a model that is linear in the parameters). The true, unknown model is almost certainly nonlinear, but a linear model is often a suitable proxy for the nonlinear model. In linear model applications of experimental design, the objectives are to determine what effects are important and to estimate those effects. In nonlinear applications, however, the objectives are to construct designs that facilitate model discrimination and parameter estimation. The work on designs for nonlinear models starts from Box and Lucas (1959) and there is a vast amount of literature on the subject, especially in regard to the construction of optimum designs. More recent references on designs for nonlinear models include Dette, Melas and Wong (2005), who constructed efficient designs for the Michaelis–Menten enzyme kinetic model, Hackl, Muller, and Atkinson (2001), Fedorov and Hackl (1997), and Atkinson and Haines (1996).

13.19 MODEL-ROBUST DESIGNS

The term "model-robust designs" refers to designs that are robust to model misspecification, as in Li and Nachtsheim (2000). One of the disadvantages of the designs that they proposed, however, is that the designs are frequently not orthogonal designs, although the departure from orthogonality is generally very small. They found that their designs were "surprisingly robust" to model misspecification, whereas fractional factorial designs were generally found to not be robust. Some useful information relative to this work can be found at http://webpages.csom.umn.edu/oms/wli/Optimal/html/robust.htm and a catalog of the designs can be found at http://home.hccnet.nl/kees.duineveld/fd.en.ffd2/factorial.html. Other recent work in this topic area

includes Goos, Kobilinsky, O'Brien, and Vandebroek (2005), Heredia-Langner, Montgomery, Carlyle, and Borror (2004), and Cooney and Verseput (2000).

13.20 DESIGNS AND ANALYSES FOR NON-NORMAL RESPONSES

There are many response variables that are not normally distributed, such as percentages. A non-normal response variable can often be transformed to be approximately normal. For example, count data are not normally distributed, although data with large counts should be approximately normally distributed. A standard transformation for count data is

$$y = \frac{\sqrt{c} + \sqrt{c+1}}{2}$$

with c denoting a count. This approach was used by Bisgaard and Fuller (1994–1995) in analyzing data given by Hsieh and Goodwin (1986), for which the response variable was the number of defects. Similarly, an arcsine transformation could be used for a binomial response variable.

If transformation to near-normality is not possible, then the question of how to analyze data from a designed experiment with a non-normal response variable must be addressed. Lewis, Montgomery, and Myers (2001a) gave several examples of designed experiments with non-normal responses and showed that using a generalized linear model (GLM) approach is an excellent alternative to transforming the data. Furthermore, Lewis et al. (2001b) stated that a GLM approach is often superior to a transformation approach for a confidence interval on the mean response and for a prediction interval. They also showed, however, that the results are highly sensitive to the proper selection of the link function. So a GLM approach is not devoid of potential problems. For information on generalized linear models, readers are referred to McCullagh and Nelder (1989) and Dobson (2001).

13.21 DESIGN OF MICROARRAY EXPERIMENTS

The design of microarray experiments is a new area that is growing rapidly. A microarray has been defined in many different ways but the bottom line is that it is used to study genes. One particular definition, given at http://www.medterms.com/script/main/art.asp?articlekey=30712, is "A tool used to sift through and analyze the information contained within a genome. A microarray consists of different nucleic acid probes that are chemically attached to a substrate, which can be a microchip, a glass slide or a microsphere-sized bead." Microarray experiments are thus biological experiments, with important implications also for the agricultural and pharmaceutical sciences. Experimental designs are necessary for assessing the many sources of uncertainty in microarray experiments.

Classical experimental designs would not be helpful in microarray experiments because biologists are not interested in main effects and interaction effects. Therefore, new experimental approaches had to be developed to answer questions such as the following that are of interest to biologists: (1) "Which mRNA samples should be competitively hybridized on the same slide?," and (2) "How many times should each slide by replicated?"

Since this is a new field, research papers on the subject date from the early years of this century. Pioneering, primary work on designs for microarray experiments was performed by Kerr and Churchill (2001) and Glonek and Solomon (2004), who used concepts from optimal design theory to suggest efficient designs for some of the common cDNA microarray experiments. More specifically, Glonek and Solomon (2004) proposed optimal designs for factorial and time course experiments, which arise frequently in microarray experiments. The approach enabled efficient designs to be constructed subject to the information available on effects of most interest to biologists, as well as the number of arrays available for the experiment and other practical constraints. They showed that their designs are superior to the popular reference designs, which they claim to be inefficient.

Most biologists performing simple gene expression comparisons, such as between treated and control cells, carry out replicate experiments on different slides. Speed and Yang (2002) referred to these as *technical replicates*, using the term to refer to the situation where the target mRNA in each hybridization was from the same RNA extraction, which is labeled independently for each hybridization. Since estimates of differential gene expression based on technical replicates tend to be correlated, the correlation structure must be taken into account and Speed and Yang (2002) did so.

The popularity of microarray experiments has spawned various books that contain material on methods for designing such experiments. These include Simon, Korn, McShane, Radmacher, Wright, and Zhao (2004) and Allison, Page, Beasley, and Edwards (2006). Significant articles include Wit, Nobile, and Khanin (2005), Scholtens, Miron, Merchant, Miller, Miron, Iglehart, and Gentleman (2004), and Emptage, Hudson-Curtis, and Sen (2003).

13.22 MULTI-VARI PLOT

A *multi-vari plot* is a technique that would undoubtedly be used more extensively with experimental design data if it were better known. As the name essentially implies, the idea is to identify sources of variation, but the method is not well known outside the group of people who are knowledgeable in statistical quality improvement techniques. Thus, the thrust is different from the well-known techniques that seek to identify factors that affect location.

The plot is similar to an ANOM plot in that the extreme values within cells of a design are connected, thus showing the variation within each cell of, say, a factorial design, and of course that variation is assumed to be equal across the cells. An example of a multi-vari plot for a one-factor design is given by Ryan (2000, p. 330), and examples are also given in Delott and Gupta (1990).

Czitrom, de Amezua Ayala, and Sanchez (1998) used a multi-vari plot in a somewhat different manner, however, as all of the 120 observations for a replicated 2×3 factorial design (in a split-plot arrangement) are shown as circles. As with plotting techniques, the best way to plot the data is the way that provides the most insightful information.

13.23 EVOLUTIONARY OPERATION

Although sequential experimentation often produces satisfactory results, the dynamics may be continually changing so that what is optimal today may not be optimal in three months, and/or the response surface may be sufficiently complicated that determining the best combination of factor levels may be extremely difficult. Consequently, frequent, almost continuous, experimentation may be needed. Evolutionary Operation (EVOP) can be used in such situations.

Designed experiments are performed off-line, so as not to interfere with production processes. EVOP is for online experimentation, however. Since the experimentation is done online, only small changes in factor levels are made from current levels. Historically, only a small number of two-level factors have generally been used in EVOP programs, with apparently only full factorials employed, perhaps with blocking. There are various forms of EVOP that have been proposed, with the most frequently used approach undoubtedly being Box-EVOP, so named for its inventor, George Box (1957). This entails the use of a 2^k design with a decision made after a specified number of cycles to change or not change the factor levels, with the decision determined by whether or not the 2σ confidence intervals for the main effect estimates include zero.

Significant interactions create problems for EVOP, just as they do when 2^k designs are used, in general. When interactions cannot be "transformed away," Ryan (2000, p. 481) suggests looking at the best combination of factor levels and making a decision based on the average for that combination. Although this isn't the same as using conditional effects, it is in the same spirit of breaking the data down to a finer level than is used when ANOVA is performed.

Although there is not very much in the literature on EVOP, Hoerl (1998) listed EVOP as a component of a typical Six Sigma black-belt training program. Despite this, it seems safe to assume that EVOP is still underutilized in industry, just as Hahn and Dershowitz (1974) concluded that it was underutilized over 30 years ago. Another recommended source is Box and Draper (1969).

13.24 SOFTWARE

Whereas most of the experimental design methods presented in previous chapters are readily available in the leading statistical software, this is not the case for some of the methods presented in this chapter. For example, in the comparison study of Reece (2003), which was mentioned in previous chapters, the author rated the software in

regard to "Designs with Factor Constraints," with each software package receiving a rating of 1 to 5. Four of the packages did not receive a rating because they did not have that capability at the time of the survey; six received a rating of 4, with only one, RS/Discover, receiving a rating of 5. JMP and Design-Expert were among the well-known packages that received a rating of 4.

As we might expect, there was a high correlation between these ratings and the ratings for "Algorithmic Designs" since, for example, an algorithmic design is typically constructed when there are factor constraints, as was discussed in Section 13.8. Of course opinions vary regarding important experimental design capabilities for software. In discussing MINITAB, Reece (2003) stated, "Not supported are the algorithmic designs needed for some forms of sequential experimentation or for dealing with constrained process spaces. The absence of these designs is a significant contributing factor to rating the software 'Good' as opposed to a higher evaluation." (These comments were in regard to Version 13.31 of MINITAB.) Release 14 and Release 15 of MINITAB *do* have capability for D-optimal designs, however, which is for use with response surface designs and mixture designs and can be generated in either menu mode or command mode. For the latter, the command is OPTDESIGN and the input is a set of candidate points, with the user specifying the number of points for the D-optimal design. The output from that command is the set of row numbers from the set of candidate points that comprise the D-optimal design. Although computing D-optimal and other optimal designs can be computer intensive, the OPTDESIGN command is, for example, executed almost instantly when used by this writer to obtain a design with 16 points from 100 candidate points, with the coded levels of each of the three factors being -1, 0, and 1. Various information is provided in addition to the design, including leverage information. It is useful to see if there are any large leverages and to see whether or not the leverages are about the same.

D-optimal designs can also be constructed with Design-Expert 7.0, for up to 30 factors. Of course specifying a D-optimal design for a large number of factors can cause the program to run for more than a short time, however. Design-Expert 7.0 also allows multiple linear constraints to be specified that can be used to construct a restricted region of operability, to which a D-optimal design could be applied. This was illustrated at http://www.statease.com/x7pup/answers/Prob%207-2.pdf.

13.25 SUMMARY

There were some designs presented in this chapter that are not well known, and consequently they are seldom used. Other designs, however, have been used extensively. Of course the extent to which any particular type of design is used is determined primarily by software capability. Plackett–Burman designs are used extensively and this is undoubtedly due to the fact that the designs can be constructed using MINITAB, Design-Expert, and JMP, plus other software. At the other extreme, uniform designs, Doehlert designs, and Cotter designs have for the most part not found their way into statistical software, although it was stated in Section 13.9.1 that JMP fairly recently added uniform design capability.

REFERENCES

Abraham, B., H. Chipman, and K. Vijayan (1999). Some risks in the construction and analysis of supersaturated designs. *Technometrics*, **41**(2), 135–141.

Ahuja, S. K., G. M. Ferreira, and A. R. Moreira (2004). Application of Plackett–Burman design and response surface methodology to achieve exponential growth for aggregated shipworm bacterium. *Biotechnology and Bioengineering*, **85** (6), 666–675.

Allen, T. T. and M. Bernshteyn (2003). Supersaturated designs that maximize the probability of identifying active factors. *Technometrics*, **45**, 90–97.

Allen, T., L. Yu, and M. Bernshteyn (2000). Low-cost response surface methods applied to the design of plastic fasteners. *Quality Engineering*, **12**(4), 583–591.

Allison, D. B., G. R. Page, T. M. Beasley, and J. W. Edwards, eds. (2006). *DNA Microarrays and Related Genomics Techniques: Design, Analysis, and Interpretation of Experiments*. New York: Chapman and Hall/CRC.

Anderson, M. and P. Whitcomb (2000). *DOE Simplified: Practical Tools for Effective Experimentation*. Portland, OR: Productivity, Inc.

Anthony, J. (2002). Training for design of experiments using a catapult. *Quality and Reliability Engineering International*, **18**, 29–35.

Arasu, K. T. and D. Torban (1999). New weighing matrices of weight 25. *Journal of Combinatorial Designs*, **7**, 11–15.

Atkinson, A. C. and R. A. Bailey (2001). One hundred years of the design of experiments on and off the pages of *Biometrika. Biometrika*, **88**, 53–97.

Atkinson, A. C. and L. M. Haines (1996). Designs for nonlinear and generalized linear models. In *Handbook of Statistics*, Vol. 13: *Design and Analysis of Experiments*, Chap. 14, pp. 437–475 (S. Ghosh and C. R. Rao, eds.). Amsterdam: Elsevier Science B.V.

Bailey, M. and B. Jones (2004). Space filling designs for pre-screen experiments. *JMPer Cable*, Issue 13, Winter, 8–10. SAS Institute Inc., Cary, NC.

Balusu, R., R. M. R. Paduru, G. Seenayya, and G. Reddy (2004). Production of ethanol from *Clostridium thermocellum* SS19 in submerged fermentation: Screening of nutrients using Plackett–Burman design. *Applied Biochemistry and Biotechnology*, **117**(3), 133–142.

Bandurek, G. (1999). Practical application of supersaturated designs. *Quality and Reliability Engineering International*, **15**, 123–133.

Banerjee, K. S. (1950). How balanced incomplete block designs may be made to furnish orthogonal estimates in weighing designs. *Biometrika*, **37**, 50–58.

Banerjee, K. S. (1951). Some observations on the practical aspects of weighing designs. *Biometrika*, **38**, 248–251.

Banerjee, K. S. (1975). *Weighing Designs for Chemistry, Medicine, Economics, Operations research, Statistics*. New York: Dekker.

Berger, M. P. F. and W. K. Wong, eds. (2005). *Applied Optimal Designs*. Hoboken, NJ: Wiley.

Bingham, T. C. (1997). An approach to developing multi-level fractional factorial designs. *Journal of Quality Technology*, **29**(4), 370–380.

Bisgaard, S. (2000). The design and analysis of $2^{k-p} \times 2^{q-r}$ split-plot experiments. *Journal of Quality Technology*, **32**, 39–56.

Bisgaard, S. and H. T. Fuller (1994–1995). Analysis of factorial experiments with defects or defectives as the response. *Quality Engineering*, **7**(2), 429–443.

Bisgaard, S. and P. Weiss (2002–2003). Experimental design applied to identify defective parts in assembled products. *Quality Engineering*, **15**(2), 347–350.

Bodea, A. and S. E. Leucuta (1997). Optimization of hydrophilic matrix tablets using a D-optimal design. *International Journal of Pharmacy*, **153**, 247–255.

Booth, K. H. V. and D. R. Cox (1962). Some systematic supersaturated designs. *Technometrics*, **4**, 489–495.

Box, G. E. P. (1957). Evolutionary operation: A method for increasing industrial productivity. *Applied Statistics*, **6**, 81–101.

Box, G. E. P. (1982). Seminar talk. Department of Statistics, University of Wisconsin-Madison.

Box, G. E. P. and D. Behnken (1960). Some new three level designs for the study of quantitative variables. *Technometrics*, **2**(4), 455–475.

Box, G. E. P. and S. Bisgaard (1993). What can you find out from 12 experimental runs? *Quality Engineering*, **5**, 663–668.

Box, G. E. P. and N. R. Draper (1969). *Evolutionary Operation*. New York: Wiley.

Box, G. E. P. and H. L. Lucas (1959). Design of experiments in nonlinear situations. *Biometrika*, **46**, 77–90.

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

Box, G. E. P. and J. Tyssedal (1996a). Projective properties of certain orthogonal arrays. *Biometrika*, **83**(4), 950–955. (This article is also available as Report No. 116, Center for Quality and Productivity Improvement, University of Wisconsin-Madison and can be downloaded at http://www.engr.wisc.edu/centers/cqpi/reports/pdfs/r116.pdf.)

Box, G. E. P. and J. Tyssedal (1996b). Sixteen run designs under high projectivity for factor screening. *Communications in Statistics: Simulation and Computation*, **30**(2), 217–228. (This article is also available as "Projective properties of the sixteen run two-level orthogonal arrays", Report No. 135, Center for Quality and Productivity Improvement, University of Wisconsin-Madison and can be downloaded at http://www.engr.wisc.edu/centers/cqpi/reports/pdfs/r135.pdf.)

Box, G. E. P., J. S. Hunter, and W. G. Hunter (2005). *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd ed. Hoboken, NJ: Wiley.

Bradley, R. A. and C. M. Yeh (1980). Trend-free block designs: Theory. *The Annals of Statistics*, **8**, 883–893.

Bradley, R. A. and C. M. Yeh (1983). Trend-free block designs: Existence and construction results. *Communications in Statistics, Theory and Methods*, **12**, 1–24.

Bradley, R. A., W. I. Notz, and C. M. Yeh (1985). Nearly trend-free block designs. *Journal of the American Statistical Association*, **80**, 985–992.

Buncick, M. C., A. R. K. Ralston, D. D. Denton, and S. Bisgaard (2000). Characterization of a plasma sputter deposition process by fractional factorial design. *Quality Engineering*, **12**, 371–385.

Bundick, R. K., C. M. Borror, and D. C. Montgomery (2003). A review of methods for measurement systems capability. *Journal of Quality Technology*, **35**, 342–354.

Bursztyn, D. and D. M. Steinberg (2001). Rotation designs for experiments in high-bias situations. *Journal of Statistical Planning and Inference*, **97**(2), 399–414.

Bursztyn, D. and D. M. Steinberg (2002). Rotation designs: Orthogonal first-order designs with higher order projectivity. *Applied Stochastic Models in Business and Industry*, **18**(3), 197–206.

Bursztyn, D. and D. M. Steinberg (2006). Comparison of designs for computer experiments. *Journal of Statistical Planning and Inference*, **136**, 1103–1119.

Cameron, J. M. and G. E. Hailes (1974). Designs for the calibration of small groups of standards in the presence of drift. NBS Technical Note 844, National Bureau of Standards, United States Department of Commerce.

Cameron, J. M., M. C. Croarkin, and R. C. Raybold (1977). Designs for the calibration of standards of mass. NBS Technical Note 952, National Bureau of Standards, United States Department of Commerce.

Chen, V. C. P., K.-L. Tsui, R. R. Barton, and J. K. Allen (2003). A review of design and modeling in computer experiments. In *Handbook of Statistics*, Vol. 22, Chap. 7 (R. Khattree and C. R. Rao, eds.). Amsterdam: Elsevier Science B. V.

Cheng, C.-S. and C.-C. Li (1993). Constructing orthogonal fractional factorial designs when some factor-level combinations are debarred. *Technometrics*, **35**(3), 277–283.

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

Cioppa, T. M. (2002). *Efficient Nearly Orthogonal and Space-Filling Experimental Designs for High-Dimensional Complex Models*. (Ph.D. Dissertation). Monterey, CA: Operations Research Department, Naval Postgraduate School. (available online at http://library.nps.navy.mil/uhtbin/Hyperion-image/02sep_Cioppa_PhD.pdf)

Cooney, G. A., Jr. and R. P. Verseput (2000). The design of model-robust experiments. *Quality and Reliability Engineering International*, **16**, 373–389.

Cornell, J. A. (2002). *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data*, 3rd ed. New York: Wiley.

Cotter, S. (1979). A screening design for factorial experiments with interactions. *Biometrika*, **66**(2), 317–320.

Cox, D. R. (1951). Some systematic experimental designs. *Biometrika*, **38**, 312–323.

Craigen, R. (1996). Weighing matrices and conference matrices. In *CRC Handbook of Combinatorial Designs*, pp. 496–504. (C. J. Colbourn and J. H. Dinitz, eds.). Boca Raton, FL: CRC Press.

Czitrom, V. (1999). One-factor-at-a-time versus designed experiments. The *American Statistician*, **53**(2), 126–131.

Czitrom, V., J. C. de Amezua Ayala, and A. R. Sanchez (1998). Matching implant doses during design of experiments. *Quality and Reliability Engineering International*, **14**, 211–217.

Czitrom, V. (2003). Guidelines for selecting factors and factor levels for an industrial designed experiment. In *Statistics in Industry*, *Handbook of Statistics*, Vol. 22 (R. Khattree and C. R. Rao, eds.). Amsterdam: Elsevier Science B. V.

Daniel, C. (1973). One-at-a-time plans. *Journal of the American Statistical Association*, **68**, 353–360 (June).

Daniel, C. (1994). Factorial one factor at a time experiments, dedicated to the memory of W. Edwards Deming. *The American Statistician*, **48**, 132–135.

Daniel, C. and F. Wilcoxon (1966). Factorial 2^{n-p} plans robust against linear and quadratic trends. *Technometrics*, **8**, 259–278.

Delott, C. and P. Gupta (1990). Characterization of copperplating process for ceramic substrates. *Quality Engineering*, **2**(3), 269–284.

Deming, W. E. (1953). On the distinction between enumerative and analytic surveys. *Journal of the American Statistical Association*, **48**, 244–245.

Deming, W. E. (1975). On probability as a basis for action. *The American Statistician*, **29**, 146–152.

Dette, H., V. B. Melas, and N. Strigul (2005). Design of experiments for microbiological models. In *Applied Optimal Designs*, Chap. 6 (M. P. F. Berger and W. K. Wong, eds.). Hoboken, NJ: Wiley.

Dette, H., V. B. Melas, and W. K. Wong (2005). Optimal design for goodness-of-fit of the Michaelis-Menten enzyme kinetic function. *Journal of the American Statistical Association*, **100**(472), 1370–1381.

Dickinson, A. W. (1974). Some run orders requiring a minimum number of factor level changes for the 2^4 and 2^5 factorial designs. *Technometrics*, **16**, 31–37.

Dobson, A. J. (2001). *An Introduction to Generalized Linear Models*, 2nd ed. London: Chapman Hall/CRC.

Draper, N. R. and D. M. Stoneman (1968). Factor changes and linear trends in eight-run two-level factorial designs. *Technometrics*, **10**, 301–311.

DuMouchel, W. and B. Jones (1994). A simple Bayesian modification of D-optimal designs to reduce the dependence on an assumed model. *Technometrics*, **36**(1), 37–47.

Emptage, M. R., B. Hudson-Curtis, and K. Sen (2003). Treatment of microarray experiments as split-plot designs. *Journal of Biopharmaceutical Statistics*, **13**, 159–178.

Eskridge, K. M., S. G. Gilmour, R. Mead, N. A. Butler, and D. A. Travnicek (2004). Large supersaturated designs. *Journal of Statistical Computation and Simulation*, **74**, 525–542.

Fang, K.-T. (1980). The uniform design: Application of number-theoretic methods in experimental design. *Acta Mathematicae Applacatae Sinica*, **3**, 363–372.

Fang, K.-T. and D. K. J. Lin (2003). Uniform experimental designs and their applications in industry. In *Handbook of Statistics*, Vol. 22, Chap. 4 (R. Khattree and C. R. Rao, eds.). Amsterdam: Elsevier Science B. V.

Fang, K.-T., R. Li, and A. Sudjianto (2006). *Design and Modeling for Computer Experiments*. New York: Chapman and Hall.

Fang, K.-T., D. K. J. Lin, and C. X. Ma (1999). On the construction of multi-level supersaturated designs. *Journal of Statistical Planning and Inference*, **86**, 239–252.

Fang, K.-T., D. K. J. Lin, and M. Q. Liu (2003). Optimal mixed-level supersaturated design. *Metrika*, **58**, 279–291.

Fang, K.-T., D. K. J. Lin, P. Winker, and Y. Zhang (2000). Uniform design: Theory and application. *Technometrics*, **42**(3), 237–248.

Fedorov, V. V. and P. Hackl (1997). *Model-Oriented Design of Experiments*. New York: Springer-Verlag.

Fedorov, V. V. and S. L. Leonov (2005). Response driven designs in drug development. In *Applied Optimal Designs*, Chap. 5 (M. P. F. Berger and W. K. Wong, eds.). Hoboken, NJ: Wiley.

Franklin, M. F. and R. A. Bailey (1977). Selection of defining contrasts and confounded effects in two-level experiments. *Applied Statistics*, **25**, 64–67.

Freeman, I. C. (2004). *Objective Comparison of Design of Experiments Strategies in Design and Observations in Practice*. (Master of Science in Engineering and Management Thesis). Cambridge, MA: Engineering Systems Division, Massachusetts Institute of Technology.

Ganju, J. and J. M. Lucas (1997). Bias in test statistics when restrictions in randomization are caused by factors. *Communications in Statistics – Theory and Methods*, **26**(1), 47–63.

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

Glonek, G. F. V. and P. J. Solomon (2004). Factorial and time course designs for cDNA microarray experiments. *Biostatistics*, **5**, 89–111.

Goos, P., L. Tack, and M. Vandebroek (2005). The optimal design of blocked experiments in industry. In *Applied Optimal Designs*, Chap. 10 (M. P. F. Berger and W. K. Wong, eds.). Hoboken, NJ: Wiley.

Goos, P., A. Kobilinsky, T. O'Brien, and M. Vandebroek (2005). Model-robust and model-sensitive designs. *Computational Statistics and Data Analysis*, **49**(1), 201–216.

Hackl, P., W. G. Muller, and A. C. Atkinson, eds. (2001): *MODA 6: Advances in Model-Oriented Design and Analysis: Proceedings of the 6th International Workshop on Model-Oriented Design and Analysis*. Santa Clara, CA: Telos, The Electronic Library of Science. (Telos is an imprint of Springer-Verlag, New York, NY.)

Hahn, G. J. and A. F. Dershowitz (1974). Evolutionary operation today—some survey results and observations. *Applied Statistics*, **23**(2), 214–218.

Hahn, G. J. and W. Q. Meeker (1993). Assumptions for statistical inference. *The American Statistician*, **47**, 1–11.

Hall, M. J. (1961). Hadamard matrices of order 16. *Jet Propulsion Laboratory*, Research Summary 1, 21–26.

Hamada, M. and C. F. J. Wu (1992). Analysis of designed experiments with complex aliasing. *Journal of Quality Technology*, **24**(3), 130–137.

Hayford, J. A. (1893). On the least squares adjustment of weighings. U. S. Coast & Geodetic Survey Report for 1892, Appendix 10, USCGS.

Heredia-Langner, A., W. M. Carlyle, D. C. Montgomery, C. M. Borror, and G. C. Runger (2003). Genetic algorithms for the construction of D-optimal designs. *Journal of Quality Technology*, **35**(1), 28–46.

Heredia-Langner, A., D. C. Montgomery, W. M. Carlyle, and C. M. Borror (2004). Model-robust optimal designs: A genetic algorithm approach. *Journal of Quality Technology*, **36**(3), 263–279.

Hill, H. M. (1960). Experimental designs to adjust for time trend. *Technometrics*, **2**, 67–82.

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

Hoerl, R. W. (1998). Six Sigma and the future of the quality profession. *Quality Progress*, **31**(6), 35–42.

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

Hubele, N. F., T. Beaumariage, G. Baweja, and S.-C. Hong (1994). Using experimental design to assess the capability of a system. *Journal of Quality Technology*, **26**(1), 1–11.

Jensen, D. (2000). The use of studentized diagnostics in regression. *Metrika*, **52**, 213–223.

Jin, R., W. Chen, and A. Sudjianto (2005). An efficient algorithm for constructing optimal design of computer experiments. *Journal of Statistical Planning and Inference*, **34**(1), 268–287.

John, P. W. M. (1990). Time trend and factorial experiments. *Technometrics*, **32**, 275–282.

Jones, F. E. E. and R. M. Schoonover (2002). *Handbook of Mass Measurement*. Boca Raton, FL: CRC Press.

Joseph, V. R. and C. F. J. Wu (2004). Failure amplification method: An information maximization approach to categorical response optimization. *Technometrics*, **46**(1), 1–12; discussion: 12–31.

Kennard, R. W. and L. Stone (1969). Computer aided design of experiments. *Technometrics*, **11**, 137–148.

Kerr, M. K. and G. A. Churchill (2001). Experimental design for gene expression microarrays. *Biostatistics*, **2**, 183–201.

Kiefer, J. (1959). Optimum experimental designs (with discussion). *Journal of the Royal Statistical Society*, *Series B*, **21**, 272–319.

Kim, K. (1997). *Construction and Analysis of Trend-Free Factorial Designs under a General Cost Structure*. (Ph.D. dissertation). Blacksburg, VA: Department of Statistics, Virginia Polytechnic Institute and State University.

Kleijnen, J. P. C., S. M. Sanchez, T. W. Lucas, and T. M. Cioppa (2005). A user's guide to the brave new world of designing simulation experiments. *INFORMS Journal on Computing*, **17**(3), 263–289. (available online at http://www.nps.navy.mil/orfacpag/resumePages/sanchsLinkedPages/UserGuideSimExpts.pdf and the 5-page online supplement is available at http://joc.pubs.informs.org/Supplements/Kleijnen.pdf.)

Kuhfeld, W. F. and R. D. Tobias (2005). Large factorial designs for product engineering and marketing research applications. *Technometrics*, **47**, 132–141.

Lawson, C. L. and R. J. Hanson (1995). *Solving Least Squares Problems* (*Classics in Applied Mathematics Series*, *Vol. 15*). Philadelphia: Society for Industrial and Applied Mathematics.

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

León, R. V. and R. W. Mee (2000). Blocking multiple sources of error in small analytic studies. *Quality Engineering*, **12**(4), 497–502.

Lewis, S. L., D. C. Montgomery, and R. H. Myers (2001a). Examples of designed experiments with nonnormal responses. *Journal of Quality Technology*, **33**(3), 265–278.

Lewis, S. L., D. C. Montgomery, and R. H. Myers (2001b). Confidence interval coverage for designed experiments analyzed using generalized linear models. *Journal of Quality Technology*, **33**(3), 279–292.

Li, R. Z. and D. K. J. Lin (2002). Data analysis of supersaturated design. *Statistics and Probability Letters*, **59**, 135–144.

Li, R. and D. K. J. Lin (2003). Analysis methods for supersaturated design: Some comparisons. *Journal of Data Analysis*, **3**, 103–121.

Li, W. and C. J. Nachtsheim (2000). Model-robust factorial designs. *Technometrics*, **42**, 345–352.

Lin, D. K. J. (1993a). A new class of supersaturated designs. *Technometrics*, **35**, 28–31.

Lin, D. K. J. (1993b). Another look at first-order saturated designs: The *p*-efficient designs. *Technometrics*, **35**, 284–292.

Lin, D. K. J. (2003). Industrial experimentation for screening. In *Handbook of Statistics*, Vol. 22, Chap. 2 (R. Khattree and C. R. Rao, eds.). Amsterdam: Elsevier Science B. V.

Lin, D. K. J. and N. R. Draper (1992). Projective properties of Plackett and Burman designs. *Technometrics*, **34**, 423–428.

Liu, M. Q. and K. T. Fang (2006). A case study in the application of supersaturated designs to computer experiments. *Acta Scientiarum Mathematicarum* **26**(B4), to appear. (A preprint is available at http://www.math.nankai.edu.cn/keyan/pre/preprint06/06-06.pdf.)

Liu, Y. and A. M. Dean (2004). k-circulant supersaturated designs. *Technometrics*, **46**, 32–43.

Liu, Y., S. Ruan, and A. M. Dean (2006). Construction and analysis of Es^2 efficient supersaturated designs. *Journal of Statistical Planning and Inference* (to appear). (A preprint of this paper is available at http://www.stat.ohio-state.edu/~amd/papers/liu-ruan-dean-rev10.pdf.)

Lopez-Fidalgo, J. and S. A. Garcet-Rodríguez (2004). Optimal experimental designs when some independent variables are not subject to control. *Journal of the American Statistical Association*, **99**, 1190–1199.

Lu, X. and X. Wu (2004). A strategy of searching active factors in supersaturated screening experiments. *Journal of Quality Technology*, **36**, 392–399.

Matthews, J. N. S. and P. W. James (2005). Restricted optimal design in the measurement of cerebral blood flow using the Kety–Schmidt technique. In *Applied Optimal Designs*, Chap. 8 (M. P. F. Berger and W. K. Wong, eds.). Hoboken, NJ: Wiley.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models*, 2nd ed. New York: Chapman and Hall/CRC.

McKay, M. D., R. J. Beckman, and W. J. Conover (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, **21**(2), 239–245.

McDaniel, W. R. and B. E. Ankenman (2000). Comparing experimental design strategies for quality improvement with minimal changes to factor levels. *Quality and Reliability Engineering International*, **16**, 355–362.

Mee, R. (2002). Three-level simplex designs and their role in sequential experimentation. *Journal of Quality Technology*, **34**, 152–164.

Montgomery, D. C. and G. C. Runger (2003). *Applied Statistics and Probability for Engineers*, 3rd ed. New York: Wiley.

Montgomery, D. C., C. M. Borror, and J. D. Stanley (1997). Some cautions in the use of Plackett–Burman designs. *Quality Engineering*, **10**(2), 371–381.

Montgomery, D. C., E. N. Loredo, D. Jearkpaporn, and M. C. Testik (2002). Experimental designs for constrained regions. *Quality Engineering*, **14**(4), 587–601.

Morris, M. D., T. J. Mitchell, and D. Ylvisaker (1993). Bayesian design and analysis of computer experiments: Use of derivatives in surface prediction. *Technometrics*, **35**, 243–255.

Myers, R. H. (1999). Response surface methodology—current status and future directions. *Journal of Quality Technology*, **31**(1), 30–44.

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

Nelson, L. S. (1982). Technical aids: Extreme screening designs. *Journal of Quality Technology*, **14**(2), 99–100.

Nelson, P. R. (1993). Additional uses for the analysis of means and extended tables of critical values. *Technometrics*, **35**(1), 61–71.

Nelson, P. R., P. S. Wludyka, and K. A. F. Copeland (2005). *The Analysis of Means*: *A Graphical Method for Comparing Means, Rates, and Proportions*. Philadelphia: Society for Industrial and Applied Mathematics; Alexandria, VA: American Statistical Association.

Ossenbruggen, P. J., H. Spanjers, and A. Klapwik (1996). Assessment of a two-step nitrification model for activated sludge. *Water Research*, **30**(4), 939–953.

Park, Y., D. C. Montgomery, J. W. Fowler, and C. M. Borror (2006). Cost-constrained G-efficient response surface designs for cuboidal regions. *Quality and Reliability Engineering International*, **22**(2), 121–139.

Piepel, G. F. (1988). Programs for generating extreme vertices and centroids of linearly constrained experimental regions. *Journal of Quality Technology*, **20**(2), 125–139.

Plackett, R. L. and J. P. Burman (1946). The design of optimum multifactorial experiments. *Biometrika*, **33**, 305–325.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. New York: Wiley.

Qu, X. and C. F. J. Wu (2005). One-factor-at-a-time designs of Resolution V. *Journal of Statistical Planning and Inference*, **131**, 407–416.

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

Reece, J. E. and M. Shenasa (1997). Characterization of a vertical furnace chemical vapor deposition (CVD) silicon nitride process. In *Statistical Case Studies for Industrial Process Improvement*, pp. 429–468. (V. Czitrom and P. D. Spagon, eds.). Philadelphia: Society for Industrial and Applied Mathematics; Alexandria, VA: American Statistical Association.

Roquemore, K. G. (1976). Hybrid designs for quadratic response surfaces. *Technometrics*, **18**, 419–423.

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

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

Samset, O. and J. Tyssedal (1998). Repeat and mirror-image patterns and correlation structures for Plackett–Burman designs, their foldovers and half fractions. Technical Report #13, Department of Mathematical Sciences, The Norwegian University of Science and Technology, Norway. (The report is accessible from the list of technical reports at http://www.math.ntnu.no/∼sam/reports.html.)

Santner, T., B. Williams, and W. Notz (2003). *The Design and Analysis of Computer Experiments*. New York: Springer-Verlag.

Satterthwaite, F. (1959). Random balance experimentation (with discussion). *Technometrics*, **1**, 111–137.

Scholtens, D., A. Miron, F. M. Merchant, A. Miller, P. L. Miron, J. D. Iglehart, and R. Gentleman (2004). Analyzing factorial designed microarray experiments. *Journal of Multivariate Analysis*, **90**, 19–43.

Simon, R. M., E. L. Korn, L. M. McShane, M. D. Radmacher, G. W. Wright, and Y. Zhao (2004). *Design and Analysis of DNA Microarray Investigations*. New York: Springer.

Simpson, T.W., D. K. J. Lin, and W. Chen (2001). Sampling strategies for computer experiments: Design and analysis. *International Journal of Reliability and Applications*, **2**(3), 209–240.

Silvey, S. D. (1980). *Optimal Design*. London: Chapman and Hall.

Snee, R. D. (1985a). Computer-aided design of experiments—some practical experiences. *Journal of Quality Technology*, **17**, 222–236.

Snee, R. D. (1985b). Experimenting with a large number of variables. In *Experiments in Industry*: *Design, Analysis, and Interpretation of Results*. Milwaukee, WI: American Society for Quality Control.

Speed, T. P. and T. H. Yang (2002). Direct versus indirect designs for cDNA microarray experiments. *Sankhya, Series A*, **64**, 707–721. (This article is available at http://www.biostat.ucsf.edu/jean/Papers/Sankhya02.pdf.)

Steinberg, D. M. and D. Bursztyn (2001). Discussion of "Factor Screening and Response Surface Exploration" by Cheng and Wu. *Statistica Sinica*, **11**, 596–599.

Tack, L. and M. Vandebroek (2002). Trend resistant and cost-efficient block designs with fixed or random block effects. *Journal of Quality Technology*, **34**, 422–436.

Tang, M., J. Li, L.-Y. Chan, and D. K. J. Lin (2004). Application of uniform design in the formation of cement mixtures. *Quality Engineering*, **16**(3), 461–474.

Tyssedal, J. (1993). Projections in the 12-run Plackett–Burman design. Report No. 106, Center for Quality and Productivity Improvement, University of Wisconsin-Madison. This report is downloadable at http://www.engr.wisc.edu/centers/cqpi/reports/pdfs/r106.pdf.

Voelkel, J. G. (2005). Efficiencies of fractional factorial designs. *Technometrics*, **47**(4), 488–494.

Wang, P. C. and H. W. Jan (1995). Designing 2-level factorial experiments using orthogonal arrays when the run order is important. *Statistician*, **44**(3), 379–388.

Wang, Y. and K.-T. Fang (1981). A note on uniform distribution and experimental design. *KeXue TongBao*, **26**, 485–489.

Waterhouse, T. H. (2005). *Optimal Experimental Design for Nonlinear and Generalised Linear Models*. (Ph. D. Thesis). Brisbane, Australia: School of Physical Sciences, University of Queensland. (This is available at http://www.maths.uq.edu.au/∼thw/research/thesis.pdf.)

Wen, Z.-Y. and F. Chen (2001). Application of statistically based experimental designs for the optimization of eicosapentaenoic acid production by the diatom *Nitzschia laevis. Biotechnology and Bioengineering*, **75**(2), 159–169.

Wit, E., A. Nobile, and R. Khanin (2005). Near-optimal designs for dual channel microarray studies. *Applied Statistics*, **54**, 817–830.

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

Xu, H. and C. F. J. Wu (2005). Construction of optimal multi-level supersaturated designs. *Annals of Statistics*, **33**(6), 2811–2836. (also available at http://www.stat.ucla.edu/∼hqxu/pub/XuWu2005.pdf)

Yamada, S. and D. K. J. Lin (1999). Three-level supersaturated designs. *Statistics and Probability Letters*, **45**, 31–39.

Yamada, S., Y. T. Ikebe, H. Hashiguchi, and N. Nicki (1999). Construction of three-level supersaturated design. *Journal of Statistical Planning and Inference*, **81**, 183–194.

Zahran, A. and C. M. Anderson-Cook (2003). A general equation and optimal design for a 2-factor restricted region. *Statistics and Probability Letters*, **64**, 9–16.

Zahran, A , C. M. Anderson-Cook, R. H. Myers, and E. P. Smith (2003). Modifying 2^2 factorial designs to accommodate a restricted design space. *Journal of Quality Technology*, **35**(4), 387–392.

Zelen, M. (1962). Linear estimation and related topics. In *Survey of Numerical Analysis*, pp. 558–577 (J. Todd, ed.) New York: McGraw-Hill.

EXERCISES

13.1 Critique the following statement: "I don't see the need to set a factor to a neutral position and then putting it back where it was before I changed the level to a neutral position. This seems much to-do about nothing."

13.2 Explain how you would design an experiment if you had three factors to examine in the experiment and one of the three was a hard-to-change factor.

13.3 Close-to-orthogonal designs, such as supersaturated designs, are often (appropriately) used. When this is done, it is necessary to consider the correlations between the factors. Explain why these are not true correlations.

13.4 What is one potential danger in using a supersaturated design?

13.5 Construct a standard OFAT design for four factors and compute the correlations between the columns of the design. Do you consider the latter to be unacceptably large? Why, or why not?

13.6 Respond to the following statement: "I don't see any point in using an equileverage design because whether or not points are influential will depend partly on the response values, and they of course cannot be controlled."

13.7 Wen and Chen (2001, references) initially used a Plackett-Burman (PB) design to investigate the factors that affected eicosapentaeonic acid (EPA) production by the diatom *Nitzschia laervis* in heterotrophic conditions. A 20-run PB design was used to investigate the effects of 13 factors. Since the number of available degrees of freedom exceeded the number of factors in the study, the experimenters created dummy columns D_1–D_6 "to evaluate the standard errors of the experiment." Are the dummy columns necessary? Why or why not? They stated that "only confidence levels above 90 percent were accepted in the experiment." This is not common statistical jargon. What did they actually mean? Once they identified the factors that they declared significant, they used a central composite design to investigate these factors further, using a 2^{5-1} design with 6 centerpoints and 10 axial points. Does this seem like a reasonable strategy? Comment.

13.8 Using the methodology given at the stated URL, derive the expression for the test item estimate that was given in Section 13.16.

13.9 What will be the leverage values if all the estimable effects in a 2^{5-1} design are estimated and are used in a model?

13.10 Determine using either software or by direct computation the leverages for a 2×4 design when only linear terms are fit, then compute the leverages

when the linear, quadratic, and cubic terms in the second factor are fit. Comment.

13.11 Assume that you have used MINITAB or other software to construct a PB design for 15 factors in 16 runs. Since 16 is a power of 2, the design must be a full factorial in a subset of these factors, and the PB design must thus be a fractional factorial. What must be the size of this subset. Show this after first appropriately rearranging the runs, if necessary. Is the fractional factorial that the PB design is equivalent to inferior in any way (e.g., in terms of minimum aberration or maximum number of clear effects) to any other fractional factorial design of this type that could be constructed? Explain.

13.12 State the form of the fractional factorial design to which a PB design with 32 runs is equivalent.

13.13 Assume that a PB design with 24 runs has been constructed to investigate the effect of each of 23 factors. Would it be appropriate to perform a conditional effects analysis, along the lines of what was illustrated in Chapter 5? Why, or why not? If this would not be possible, although a PB design is used because of run size economy, would replicating the design permit a conditional effects analysis? In general, what must be true before such an analysis can be performed?

13.14 Consider the problem posed in Example 13.2. We can label the factors that were involved in the debarred observations as D and E, referring to substrate holder and heat, respectively. If the other half fraction had been used (i.e., with $I = -ABCDE$), would the problem still have existed, and if so, would the number of debarred observations still have been 4? Explain.

13.15 What is the motivation for using a "reverse" PB design.

13.16 Since a PB design is a small-run design, what would be the motivation for using a half-fraction of such a design? Consider a half-fraction of a 12-run PB design and the correlation structure given in Table 82 at www.math.ntnu.no/preprint/statistics/1998/S13-1998.ps . Would you recommend such a design? Why, or why not?