Ronald Christensen
Professor of Statistics
Department of Mathematics and Statistics
University of New Mexico

Topics in Experimental Design

Springer

Preface

This (non)book assumes that the reader is familiar with the basic ideas of experimental design and with linear models. I think most of the applied material should be accessible to people with MS courses in regression and ANOVA but I had no hesitation in using linear model theory, if I needed it, to discuss more technical aspects.

Over the last several years I've been working on revisions to my books *Analysis of Variance, Design, and Regression* (*ANREG*); *Plane Answers to Complex Questions: The Theory of Linear Models* (*PA*); and *Advanced Linear Modeling* (*ALM*). In each of these books there was material that no longer seemed sufficiently relevant to the themes of the book for me to retain. Most of that material related to Experimental Design. Additionally, due to Kempthorne's (1952) book, many years ago I figured out p^f designs and wrote a chapter explaining them for the very first edition of *ALM*, but it was never included in any of my books. I like all of this material and think it is worthwhile, so I have accumulated it here. (I'm not actually all that wild about the recovery of interblock information in BIBs.)

A few years ago my friend Chris Nachtsheim came to Albuquerque and gave a wonderful talk on Definitive Screening Designs. Chapter 5 was inspired by that talk along with my longstanding desire to figure out what was going on with Placett-Burman designs.

I'm very slowly working on R code for this material. See http://www.stat.unm.edu/~fletcher/R-TD.pdf. Also, a retypeset version of the first edition of ANREG (*ANREG-I*) is available at http://www.stat.unm.edu/~fletcher/anreg.pdf. R computer code for the second edition of ANREG is available at http://www.stat.unm.edu/~fletcher/Rcode.pdf.

As I have mentioned elsewhere, the large numbers of references to myself are as much due to sloth as ego.

Ronald Christensen
Albuquerque, New Mexico
November 12, 2016
March 30, 2019

Contents

Chapter 1
Fundamentals

Three fundamental ideas in experimental design are *replication*, *blocking*, and *randomization*. Replication is required so that we have a measure of the variability of the observations. Blocking is used to reduce experimental variability. Blocking has inspired a number of standard designs including randomized complete blocks, Latin squares, balanced incomplete blocks, Youden squares, balanced lattice designs, and partially balanced incomplete blocks. Randomly assigning treatments to experimental units gives one a philosophical basis for inferring that the treatments *cause* the results observed in the experiment, cf. Peirce and Jastrow (1885) and Fisher (1935).

An extremely useful concept in experimental design is the use of treatments with factorial structure. For example, if you are interested in the effect of two factors, say the effect of alcohol and the effect of sleeping pills, rather than performing separate experiments on each, you can incorporate both factors into the treatments of a single experiment. Briefly, suppose we are interested in two levels of alcohol, say, a_0 – no alcohol and a_1 – a standard dose of alcohol, and we are also interested in two levels of sleeping pills, say, s_0 – no sleeping pills and s_1 – a standard dose of sleeping pills. A factorial treatment structure involves forming 4 treatments, a_0s_0 – no alcohol, no sleeping pills, a_0s_1 – no alcohol, sleeping pills, a_1s_0 – alcohol, no sleeping pills, a_1s_1 – alcohol and sleeping pills. A *factorial treatment structure* obtains when the treatments in an experiment consist of all possible combinations of the levels of some factors.

The point of using factorial treatment structures is that they allow one to look very efficiently at the main effects of the factors but also that they allow examination of interactions between the factors. If there is no interaction, i.e., if the effect of alcohol does not depend on whether or not the subject has taken sleeping pills, you can learn as much from running one experiment using factorial treatment structure on, say, 20 people as you can from two separate experiments, one for alcohol and one for sleeping pills, each involving 20 people. This is a 50% savings in the number of observations needed. Second, if interaction exists, i.e., if the effect of alcohol depends on the amount of sleeping pills a person has taken, you can study that interaction in an experiment with factorial treatment structure but you cannot study

interaction in an experiment that was solely devoted to looking at the effects of alcohol.

An experiment such as this involving two factors each at two levels is referred to as a $2 \times 2 = 2^2$ factorial treatment structure. Note that 4 is the number of treatments we end up with. If we had 3 levels of sleeping pills, it would be a 2×3 factorial, thus giving 6 treatments. If we had three levels of both alcohol and sleeping pills it would be a $3 \times 3 = 3^2$. If we had three factors, say, alcohol, sleeping pills, and benzedrine each at 2 levels, we would have a $2 \times 2 \times 2 = 2^3$ structure. If each factor were at 3 levels, say, none of the drug, a standard dose, and twice the standard dose, and we made up our treatments by taking every combination of the levels of each factor, we would get $3 \times 3 \times 3 = 3^3 = 27$ treatments. More generally, if we have f factors each at p levels we have a p^f factorial treatment structure.

Christensen (1996) and (2015) contain more information on the fundamentals of experimental design including more on factorial treatment structures and more on replication, blocking, and randomization. There are lots of excellent books on Experimental Design many of which are cited in the references above.

EXAMPLE 1.0.1. Hare (1988) reported results on an experiment with five factors each at two levels for $2^5 = 32$ factor combinations (treatments). The issue was excessive variability in the taste of a dry soup mix. The source of variability was identified as a particular component of the mix called the 'intermix' containing flavorful ingredients such as salt and vegetable oil. Intermix is made in a large mixer. Factor A is the number of ports for adding vegetable oil to the mixer. This was set at either 1 (a_0) or 3 (a_1). Factor B is the temperature of the mixer. The mixer can be cooled by circulating water through the mixer jacket (b_0) or the mixer can be used at room temperature (b_1). Factor C is the mixing time, 60 seconds (c_0) or 80 seconds (c_1). Factor D is the size of the intermix batch, either 1500 pounds (d_0) or 2000 pounds (d_1). Factor E is the delay between making the intermix and using it in the final soup mix. The delay is either 1 day (e_0) or 7 days (e_1). The experimenters chose to collect data on only 16 of the 32 treatments. We will examine the data in Chapters 2 and 5.

In general, I contend that if interaction exists, it is the only thing worth looking at. In general, if interaction exists, main effects have no useful meaning. In particular, you could have an interaction between two factors, say alcohol and sleeping pills, where responses are low when you use both alcohol and sleeping pills or use neither alcohol nor sleeping pills but responses are high when you use either one but not the other. In such a case, neither the main effect for alcohol nor the main effect for sleeping pills may look important, because their average effects may be unimportant, *despite* the importance of knowing the exact combination of alcohol and sleeping pill used. When looking at more than two factors, interactions get even more complicated. For example, the nature of an alcohol by sleeping pill interaction could change depending on one's use of benzedrine. Christensen et al. (2010, Subsection 7.4.7) and Christensen (2015, Section 9.9) contain extensive discussions of interaction.

To examine all the possible interactions, one needs to look at all of the treatments defined by the factorial structure. To draw valid conclusions about the interactions, one needs replications from which to estimate variability. When the number of treatments is large, that can be expensive. If you want to save money, you have to give something up. Obviously, you should give up the things that you think are unlikely to be important.

Screening designs are used to look at many factors using relatively few observations; fewer observations than the number of treatments in a factorial structure. Screening designs are based on the hope that most potential interactions will be unimportant.

Historically, screening designs were first developed for f factors each at 2 levels, i.e. 2^f designs. These allowed for consideration of some interactions and for blocking. In fact, it would seem that the ideas were first developed as methods for dividing the treatments into blocks. For example, with $2^4 = 16$ drug treatments being applied to rats, one might want to divide the treatments into groups of four so that they can be applied to a block of four rats from a single litter. Once you have the treatments divided into blocks, an easy leap recognizes that a single block can be used as a screening design. We consider such designs in Chapter 2. Chapter 3 extends the results of Chapter 2 for f factors each at p levels, i.e. p^f designs, where p is a prime number. Chapter 4 considers designs involving factors with different numbers of levels. Chapter 5 discusses screening designs in general with emphasis on Plackett-Burman designs and Definitive Screening Designs. Chapter 6 examines response surface designs and their analysis. Chapter 7 discusses the recovery of interblock information in balanced incomplete block designs.

1.1 Notation

The notation follows that of my other books. An $r \times c$ matrix of 1s is denoted J_r^c with $J_r \equiv J_r^1$ and because n typically denotes the number of observation, $J \equiv J_n$.

Chapter 2
2^f Factorial Systems

A version of this material appeared in the first edition of ANREG. It references *ANREG-I* frequently: www.stat.unm.edu/~fletcher/anreg.pdf.

The use of treatment structures involving f factors each at 2 levels, i.e., 2^f factorials, has been popular in agriculture from early in the 20th century and became very popular in industrial applications during the last quarter of the 20th century. In particular, these can make for useful screening experiments in which one seeks to find out which of many potential factors have important effects on some process. Even when f is of moderate size, the number of treatments involved can get large in a hurry. For example, $2^5 = 32$ and $2^{10} = 1024$. With 1024 treatments, we probably do not want to look at all of them, it is just too much work. Fractional factorial designs have been developed that allow us to look at a subset of the 1024 treatments while still being able to isolate the most important effects. With only, say, 32 treatments, we may be willing to look at all of those treatments, perhaps even replicate the treatments. But if we want to use blocking to reduce variability, we may be incapable of finding blocks that allow us to apply 32 treatments within each block, hence we may be incapable of constructing reasonable randomized complete block designs. Confounding is a method for creating blocks of sizes smaller than 32 that still allow us to examine the important effects. Most books on experimental design including Christensen (1996, Chapter 17) discuss 2^f factorials in detail.

Confounding is a method of designing a factorial experiment that allows incomplete blocks, i.e., blocks of smaller size than the full number of factorial treatments. In *fractional replication* an experiment has fewer observations than the full factorial number of treatments. A basic idea in experimental design is ensuring adequate replication to provide a good estimate of error. Fractional replication not only fails to get replication – it fails to provide even an observation on every factor combination. Not surprisingly, fractional replications present new challenges in analyzing data.

In this chapter, we will informally use concepts of modular arithmetic, e.g., $7 \bmod 5 = 2$ where 2 is the remainder when 7 is divided by 5. Modular arithmetic is crucial to more advanced discussions of confounding and fractional replication,

but its use in this chapter will be minimal. To help minimize modular arithmetic, we will refer to 0 as an even number.

In a 2^f experiment, the treatments have $2^f - 1$ degrees of freedom and they are broken down into $2^f - 1$ effects each with one degree of freedom.

EXAMPLE 2.0.1. *A 2^4 factorial structure*
Consider a 2^4 experiment with factors A, B, C, D at levels a_0, a_1, b_0, b_1, c_0, c_1, and d_0, d_1, respectively. There are $2^4 = 16$ treatment combinations, so there are 15 degrees of freedom for treatments. The sources in the ANOVA table can be broken down as follows

Source	*df*	Source	*df*	Source	*df*
A	1	*AB*	1	*ABC*	1
B	1	*AC*	1	*ABD*	1
C	1	*AD*	1	*ACD*	1
D	1	*BC*	1	*BCD*	1
		BD	1	*ABCD*	1
		CD	1		

Since each effect has a single degree of freedom, it can be identified with a contrast among the 16 treatments. □

2.0.1 Effect contrasts in 2^f factorials

The simplest way to understand confounding and fractional replication in 2^f systems is in terms of contrasts corresponding to the different effects. As was just seen, each effect in a 2^f has one degree of freedom and thus each effect corresponds to a single contrast. We now review the correspondence between contrasts and factorial effects.

EXAMPLE 2.0.2. *A 2^2 experiment*
Consider a 2^2 experiment with factors A and B at levels a_0, a_1 and b_0, b_1, respectively. The coefficients of the contrasts that correspond to the main effects and interaction are given below

Treatment	*A*	*B*	*AB*
a_0b_0	1	1	1
a_0b_1	1	−1	−1
a_1b_0	−1	1	−1
a_1b_1	−1	−1	1

Christensen (1996, Example 11.1.1) examined a 2^2 experiment and showed that these contrasts give the same sums of squares as the analysis of variance table methods for obtaining the sums of squares for A, B, and AB.

The contrast coefficients are determined by the subscripts in the treatment combinations. The A contrast coefficient is 1 for any treatment that has an a subscript of 0 and -1 for any treatment that has an a subscript of 1. In other words, the A contrast is 1 for a_0b_0 and a_0b_1 and -1 for a_1b_0 and a_1b_1. Similarly the B contrast is 1 for any treatment that has a b subscript of 0 and -1 for any treatment that has an b subscript of 1, i.e., a_0b_0 and a_1b_0 have coefficients of 1 and a_0b_1 and a_1b_1 have coefficients of -1. The AB contrast involves both factors, so the subscripts are added. Treatments with an even total, 0 or 2, get contrast coefficients of 1, while treatments with an odd total for the subscripts get -1. Thus a_0b_0 and a_1b_1 get 1s and a_0b_1 and a_1b_0 get -1s. Actually, the key is modular arithmetic. For 2^f factorials, the contrast coefficients are determined by an appropriate sum of the subscripts modulo 2. Thus *any sum that is an even number is 0 mod 2 and any odd sum is 1 mod 2*. □

EXAMPLE 2.0.3. *A 2^3 experiment*
Consider a 2^3 experiment with factors A, B, and C. The contrast coefficients for main effects and interactions are given below.

Treatment	A	B	C	AB	AC	BC	ABC
$a_0b_0c_0$	1	1	1	1	1	1	1
$a_0b_0c_1$	1	1	−1	1	−1	−1	−1
$a_0b_1c_0$	1	−1	1	−1	1	−1	−1
$a_0b_1c_1$	1	−1	−1	−1	−1	1	1
$a_1b_0c_0$	−1	1	1	−1	−1	1	−1
$a_1b_0c_1$	−1	1	−1	−1	1	−1	1
$a_1b_1c_0$	−1	−1	1	1	−1	−1	1
$a_1b_1c_1$	−1	−1	−1	1	1	1	−1

Once again the contrast coefficients are determined by the subscripts of the treatment combinations. The A contrast has 1s for a_0s and -1s for a_1s; similarly for B and C. The AB contrast is determined by the sum of the a and b subscripts. The sum of the a and b subscripts is even, either 0 or 2, for the treatments $a_0b_0c_0$, $a_0b_0c_1$, $a_1b_1c_0$, $a_1b_1c_1$, so all have AB contrast coefficients of 1. The sum of the a and b subscripts is 1 for the treatments $a_0b_1c_0$, $a_0b_1c_1$, $a_1b_0c_0$, $a_1b_0c_1$, so all have coefficients of -1. The AC contrast is determined by the sum of the a and c subscripts and the BC contrast is determined by the sum of the b and c subscripts. The ABC contrast is determined by the sum of the a, b, and c subscripts. The sum of the a, b, and c subscripts is even, either 0 or 2, for the treatments $a_0b_0c_0$, $a_0b_1c_1$, $a_1b_0c_1$, $a_1b_1c_0$, so all have ABC coefficients of 1. The sum of the a, b, and c subscripts is odd, 1 or 3, for the treatments $a_0b_0c_1$, $a_0b_1c_0$, $a_1b_0c_0$, $a_1b_1c_1$, so all have coefficients of -1. □

EXAMPLE 2.0.4. *A 2^4 experiment*
Consider a 2^4 experiment with factors A, B, C, and D. The contrast coefficients are given in Tables 2.1 and 2.2. Again the contrast coefficients are determined by the subscripts of the treatments. The A, B, C, and D contrasts are determined by the subscripts of a, b, c, and d, respectively. The AB, AC, AD, BC, BD, and CD contrasts are determined by the sums of the appropriate pair of subscripts. The ABC, ABD, ACD, and BCD contrasts are determined by the sum of the three appropriate

subscripts. The coefficients of the *ABCD* contrast are determined by the sum of all four subscripts. As before, the contrast coefficient is 1 if the appropriate value or sum equals 0 mod 2 (is even) and is -1 if it equals 1 mod 2 (is odd). □

Table 2.1 Main effect and second-order interaction contrast coefficients for a 2^4 factorial.

Treatment	*A*	*B*	*C*	*D*	*AB*	*AC*	*AD*	*BC*	*BD*	*CD*
$a_0b_0c_0d_0$	1	1	1	1	1	1	1	1	1	1
$a_0b_0c_0d_1$	1	1	1	−1	1	1	−1	1	−1	−1
$a_0b_0c_1d_0$	1	1	−1	1	1	−1	1	−1	1	−1
$a_0b_0c_1d_1$	1	1	−1	−1	1	−1	−1	−1	−1	1
$a_0b_1c_0d_0$	1	−1	1	1	−1	1	1	−1	−1	1
$a_0b_1c_0d_1$	1	−1	1	−1	−1	1	−1	−1	1	−1
$a_0b_1c_1d_0$	1	−1	−1	1	−1	−1	1	1	−1	−1
$a_0b_1c_1d_1$	1	−1	−1	−1	−1	−1	−1	1	1	1
$a_1b_0c_0d_0$	−1	1	1	1	−1	−1	−1	1	1	1
$a_1b_0c_0d_1$	−1	1	1	−1	−1	−1	1	1	−1	−1
$a_1b_0c_1d_0$	−1	1	−1	1	−1	1	−1	−1	1	−1
$a_1b_0c_1d_1$	−1	1	−1	−1	−1	1	1	−1	−1	1
$a_1b_1c_0d_0$	−1	−1	1	1	1	−1	−1	−1	−1	1
$a_1b_1c_0d_1$	−1	−1	1	−1	1	−1	1	−1	1	−1
$a_1b_1c_1d_0$	−1	−1	−1	1	1	1	−1	1	−1	−1
$a_1b_1c_1d_1$	−1	−1	−1	−1	1	1	1	1	1	1

Table 2.2 Higher order interaction contrast coefficients for a 2^4 factorial.

Treatment	*ABC*	*ABD*	*ACD*	*BCD*	*ABCD*
$a_0b_0c_0d_0$	1	1	1	1	1
$a_0b_0c_0d_1$	1	−1	−1	−1	−1
$a_0b_0c_1d_0$	−1	1	−1	−1	−1
$a_0b_0c_1d_1$	−1	−1	1	1	1
$a_0b_1c_0d_0$	−1	−1	1	−1	−1
$a_0b_1c_0d_1$	−1	1	−1	1	1
$a_0b_1c_1d_0$	1	−1	−1	1	1
$a_0b_1c_1d_1$	1	1	1	−1	−1
$a_1b_0c_0d_0$	−1	−1	−1	1	−1
$a_1b_0c_0d_1$	−1	1	1	−1	1
$a_1b_0c_1d_0$	1	−1	1	−1	1
$a_1b_0c_1d_1$	1	1	−1	1	−1
$a_1b_1c_0d_0$	1	1	−1	−1	1
$a_1b_1c_0d_1$	1	−1	1	1	−1
$a_1b_1c_1d_0$	−1	1	1	1	−1
$a_1b_1c_1d_1$	−1	−1	−1	−1	1

Most books on experimental design contain a discussion of confounding and fractional replication for 2^f treatment structures. Daniel (1976), Box, Hunter, and

Hunter (1978), and Box and Draper (1987) are excellent books that focus on industrial applications.

2.1 Confounding

Confounding involves creating blocks that are smaller than the total number of treatments. Thus, confounding is a method for arriving at an incomplete block design. However, we will see that the analysis of confounding designs remains simple. For example, the analysis is considerably simpler than the balanced incomplete block analysis in Christensen (1996, Section 16.2).

EXAMPLE 2.1.1. *Confounding in a* 2^3 *experiment*
Suppose we have three drugs that we wish to investigate simultaneously. Each drug is a factor; the levels are either no dose of the drug or a standard dose. There are $2^3 = 8$ treatment combinations. The drugs will be applied to a certain type of animal. To reduce variation, we may want to use different litters of animals as blocks. However, it may be difficult to find litters containing 8 animals. On the other hand, litters of size 4 may be readily available. In such a case, we want to use four treatments on one litter and the other four treatments on a different litter. There are 70 ways to do this. We need a systematic method of choosing the treatments for each litter that allows us to perform as complete an analysis as possible.

To examine the application of the treatments from a 2^3 factorial in blocks of size 4, recall that the 2^3 has 8 treatments, so $1/2$ the treatments will go in each block. The table of contrast coefficients for a 2^3 factorial is repeated below.

Treatment	A	B	C	AB	AC	BC	ABC
$a_0b_0c_0$	1	1	1	1	1	1	1
$a_0b_0c_1$	1	1	−1	1	−1	−1	−1
$a_0b_1c_0$	1	−1	1	−1	1	−1	−1
$a_0b_1c_1$	1	−1	−1	−1	−1	1	1
$a_1b_0c_0$	−1	1	1	−1	−1	1	−1
$a_1b_0c_1$	−1	1	−1	−1	1	−1	1
$a_1b_1c_0$	−1	−1	1	1	−1	−1	1
$a_1b_1c_1$	−1	−1	−1	1	1	1	−1

We need to divide the treatments into two groups of size 4 but every contrast does this. The two groups of four are those treatments that have contrast coefficients of 1 and those that have -1. Thus we can use any contrast to define the blocks. Unfortunately, the contrast we choose will be lost to us because it will be *confounded* with blocks. In other words, we will not be able to tell what effects are due to blocks (litters) and what effects are due to the defining contrast. We choose to define blocks using the *ABC* contrast because it is the highest order interaction. Typically, it is the least painful to lose. The *ABC* contrast defines two groups of treatments

ABC coefficients	
$ABC(1)$	$ABC(-1)$
$a_0b_0c_0$	$a_0b_0c_1$
$a_0b_1c_1$	$a_0b_1c_0$
$a_1b_0c_1$	$a_1b_0c_0$
$a_1b_1c_0$	$a_1b_1c_1$

Each group of treatments is used in a separate block. The four treatments labeled $ABC(1)$ will be randomly assigned to the animals in one randomly chosen litter and the four treatments labeled $ABC(-1)$ will be randomly assigned to the animals in another litter. Recall that all information about ABC has been lost because it is confounded with blocks.

As indicated earlier, we could choose any of the contrasts to define the blocks. Typically, we use high order interactions because they are the effects that are most difficult to interpret and thus the most comfortable to live without. For illustrative purposes, we also give the blocks defined by the BC contrast.

BC coefficients	
$BC(1)$	$BC(-1)$
$a_0b_0c_0$	$a_0b_0c_1$
$a_0b_1c_1$	$a_0b_1c_0$
$a_1b_0c_0$	$a_1b_0c_1$
$a_1b_1c_1$	$a_1b_1c_0$

If the subjects of the drug study are humans, it will be difficult to obtain 'litters' of four, but it may be practical to use identical twins. We now have 8 treatments that need to be divided into blocks of 2 units. Each block will consist of $1/4$ of the treatments. Since each contrast divides the treatments into two groups of four, if we use two contrasts we can divide each group of four into 2 groups of two. We take as our first contrast ABC and as our second contrast AB. The four treatments with ABC coefficients of 1 are $a_0b_0c_0$, $a_0b_1c_1$, $a_1b_0c_1$, and $a_1b_1c_0$. These can be divided into 2 groups of two, depending on whether their AB coefficient is 1 or -1. The two groups are $a_0b_0c_0$, $a_1b_1c_0$ and $a_0b_1c_1$, $a_1b_0c_1$. Similarly, the $ABC(-1)$ group, $a_0b_0c_1$, $a_0b_1c_0$, $a_1b_0c_0$, and $a_1b_1c_1$ can be divided into $a_0b_0c_1$, $a_1b_1c_1$ and $a_0b_1c_0$, $a_1b_0c_0$ based on the AB coefficients. In tabular form we get

$ABC(1)$		$ABC(-1)$	
$AB(1)$	$AB(-1)$	$AB(1)$	$AB(-1)$
$a_0b_0c_0$	$a_1b_0c_1$	$a_0b_0c_1$	$a_1b_0c_0$
$a_1b_1c_0$	$a_0b_1c_1$	$a_1b_1c_1$	$a_0b_1c_0$

To get blocks of size 2, we confounded two contrasts, ABC and AB. Thus we have lost all information on both of these contrasts. It turns out that we have also lost all information on another contrast, C. Exactly the same four blocks would be obtained if we confounded ABC and C.

$ABC(1)$		$ABC(-1)$	
$C(1)$	$C(-1)$	$C(-1)$	$C(1)$
$a_0b_0c_0$	$a_1b_0c_1$	$a_0b_0c_1$	$a_1b_0c_0$
$a_1b_1c_0$	$a_0b_1c_1$	$a_1b_1c_1$	$a_0b_1c_0$

Similarly, if we had confounded AB and C, we would obtain the same four blocks. Note that with four blocks, there are three degrees of freedom for blocks. Each contrast has one degree of freedom, so there must be three contrasts confounded with blocks.

Given the two defining contrasts ABC and AB, there is a simple way to identify the other contrast that is confounded with blocks. The confounding is determined by a form of modular multiplication where any even power is treated as 0; thus $A^2 = A^0 = 1$ and $B^2 = 1$. Multiplying the defining contrasts gives

$$ABC \times AB = A^2B^2C = C,$$

so C is also confounded with blocks.

Typically, we want to retain information on all main effects. The choice of ABC and AB for defining contrasts is poor because it leads to complete loss of information on the main effect C. We would do better to choose AB and BC. In that case, the other confounded contrast is

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

which is another two-factor interaction. Using this confounding scheme, we get information on all main effects. The blocking scheme is given below.

$AB(1)$		$AB(-1)$	
$BC(1)$	$BC(-1)$	$BC(1)$	$BC(-1)$
$a_0b_0c_0$	$a_0b_0c_1$	$a_1b_0c_0$	$a_1b_0c_1$
$a_1b_1c_1$	$a_1b_1c_0$	$a_0b_1c_1$	$a_0b_1c_0$

□

EXAMPLE 2.1.2. *Confounding in a* 2^4 *experiment*
The $ABCD$ contrast was given in Table 2.2. Dividing the treatments into two groups based on their $ABCD$ coefficients defines two blocks of size 8,

$ABCD$ coefficients	
$ABCD(1)$	$ABCD(-1)$
$a_0b_0c_0d_0$	$a_0b_0c_0d_1$
$a_0b_0c_1d_1$	$a_0b_0c_1d_0$
$a_0b_1c_0d_1$	$a_0b_1c_0d_0$
$a_0b_1c_1d_0$	$a_0b_1c_1d_1$
$a_1b_0c_0d_1$	$a_1b_0c_0d_0$
$a_1b_0c_1d_0$	$a_1b_0c_1d_1$
$a_1b_1c_0d_0$	$a_1b_1c_0d_1$
$a_1b_1c_1d_1$	$a_1b_1c_1d_0$

To define four blocks of size 4 requires choosing two defining contrasts. To obtain four blocks, $ABCD$ is not a good choice for a defining contrast because if we choose the second contrast as a three-factor effect, we also confound a main effect, e.g.,

$$ABCD \times ABC = A^2B^2C^2D = D.$$

Similarly, if we choose the second contrast as a two-factor effect, we lose another two-factor effect, e.g.,

$$ABCD \times AB = A^2B^2CD = CD$$

However, if we choose two three-factor effects as defining contrasts, we lose only one two-factor effect, e.g.,

$$ABC \times BCD = AB^2C^2D = AD.$$

The four blocks for the confounding scheme based on ABC and BCD are given below.

$ABC(1)$		$ABC(-1)$	
$BCD(1)$	$BCD(-1)$	$BCD(1)$	$BCD(-1)$
$a_0b_0c_0d_0$	$a_0b_0c_0d_1$	$a_0b_0c_1d_1$	$a_0b_0c_1d_0$
$a_0b_1c_1d_0$	$a_0b_1c_1d_1$	$a_0b_1c_0d_1$	$a_0b_1c_0d_0$
$a_1b_0c_1d_1$	$a_1b_0c_1d_0$	$a_1b_0c_0d_0$	$a_1b_0c_0d_1$
$a_1b_1c_0d_1$	$a_1b_1c_0d_0$	$a_1b_1c_1d_0$	$a_1b_1c_1d_1$

The treatment groups can be checked against the contrasts given in Table 2.2.

If we wanted blocks of size 2 we would need three defining contrasts, say ABC, BCD, and ACD. Blocks of size 2 imply the existence of 8 blocks, so 7 degrees of freedom must be confounded with blocks. To obtain the other confounded contrasts, multiply each pair of defining contrasts and multiply all three defining contrasts together. Multiplying the pairs gives $ABC \times BCD = AD$, $ABC \times ACD = BD$, and $BCD \times ACD = AB$. Multiplying all three together gives $ABC \times BCD \times ACD = AD \times ACD = C$. □

Consider the problem of creating 16 blocks for a 2^f experiment. Since $16 = 2^4$, we need 4 defining contrasts. With 16 blocks there are 15 degrees of freedom for blocks, hence 15 contrasts confounded with blocks. Four of these 15 are the defining contrasts. Multiplying distinct pairs of defining contrasts gives 6 implicitly confounded contrasts. There are 4 distinct triples that can be made from the defining contrasts; multiplying the triples gives 4 more confounded contrasts. Multiplying all four of the defining contrasts gives the fifteenth and last confounded contrast.

We now consider the analysis of data obtained from a confounded 2^f design.

EXAMPLE 2.1.3. *Analysis of a 2^3 in blocks of four with replication*
Yates (1935) presented data on a 2^3 agricultural experiment involving yields of peas when various fertilizers were applied. The three factors were a nitrogen fertilizer (N), a phosphorous fertilizer (P), and a potash fertilizer (K). Each factor consisted of two levels, none of the fertilizer and a standard dose. It was determined that homogenous blocks of land were best obtained by creating six squares each containing four plots. Thus we have $2^3 = 8$ treatments, blocks of size 4, and six available blocks. By confounding one treatment contrast, we can obtain blocks of size 4. With six blocks, we can have 3 replications of the treatments. The confounded contrast was chosen to be the NPK interaction, so the treatments in one block are $n_0p_0k_0$, $n_1p_1k_0$, $n_1p_0k_1$, and $n_0p_1k_1$ and the treatments in the other block are $n_1p_0k_0$, $n_0p_1k_0$, $n_0p_0k_1$, and $n_1p_1k_1$. The data are given in Table 2.3. The table displays the original geographical layout of the experiment with lines identifying blocks and replications.

Each pair of rows in the table are a replication with the left and right halves identifying blocks. In each replication, the set of four treatments to be applied to a block is randomly decided and then within each block the four treatments are randomly assigned to the four plots.

Table 2.3 Yates's confounded pea data.

$n_0p_0k_0$(56.0) $n_1p_1k_0$(59.0)	$n_0p_0k_1$(55.0) $n_1p_1k_1$(55.8)
$n_0p_1k_1$(53.2) $n_1p_0k_1$(57.2)	$n_1p_0k_0$(69.5) $n_0p_1k_0$(62.8)
$n_0p_1k_1$(48.8) $n_0p_0k_0$(51.5)	$n_0p_1k_0$(56.0) $n_1p_1k_1$(58.5)
$n_1p_0k_1$(49.8) $n_1p_1k_0$(52.0)	$n_0p_0k_1$(55.5) $n_1p_0k_0$(59.8)
$n_0p_1k_0$(44.2) $n_1p_1k_1$(48.8)	$n_1p_0k_1$(57.0) $n_1p_1k_0$(62.8)
$n_0p_0k_1$(45.5) $n_1p_0k_0$(62.0)	$n_0p_0k_0$(46.8) $n_0p_1k_1$(49.5)
Three replications with *NPK* confounded in each.	

The analysis of these data is straightforward; it follows the usual pattern. The mean square and sum of squares for blocks is obtained from the six block means. Each block mean is the average of 4 observations. The sum of squares for a main effect, say, N, can be obtained from the two nitrogen means, each based on 12 observations, or equivalently, it can be obtained from the contrast

Treatment	N
$n_0p_0k_0$	1
$n_0p_0k_1$	1
$n_0p_1k_0$	1
$n_0p_1k_1$	1
$n_1p_0k_0$	−1
$n_1p_0k_1$	−1
$n_1p_1k_0$	−1
$n_1p_1k_1$	−1

applied to the 8 treatment means which are obtained by averaging over the 3 replications. The contrast for *NPK* was confounded with blocks, so it should not appear in the analysis; the one degree of freedom for *NPK* is part of the five degrees of freedom for blocks.

The complete analysis of variance is given in Table 2.4. It is the result of fitting the model

$$y_{hijk} = \mu + \beta_h + \nu_i + \rho_j + \kappa_k + (\nu\rho)_{ij} + (\nu\kappa)_{ik} + (\rho\kappa)_{jk} + \varepsilon_{hijk} \tag{1}$$

where β_h, $h = 1,\ldots,6$ indicates a block effect and ν, ρ, and κ indicate effects relating to N, P, and K respectively. Every effect in the analysis of variance has one degree of freedom, so there is no need to investigate contrasts beyond what is given

Table 2.4 Analysis of variance.

Source	*df*	*SS*	*MS*	*F*	*P*
Blocks	5	343.30	68.66	4.45	0.016
N	1	189.28	189.28	12.26	0.004
P	1	8.40	8.40	0.54	0.475
K	1	95.20	95.20	6.17	0.029
NP	1	21.28	21.28	1.38	0.263
NK	1	33.13	33.13	2.15	0.169
PK	1	0.48	0.48	0.03	0.863
Error	12	185.29	15.44		
Total	23	876.36			

in the ANOVA table. The only effects that appear significant are those for N and K. The evidence for an effect due to the nitrogen-based fertilizer is quite clear. The means for the nitrogen treatments are

N	n_0	n_1
	52.067	57.683

so the addition of nitrogen increases yields. The evidence for an affect due to potash is somewhat less clear. The means are

K	k_0	k_1
	56.867	52.883

Surprisingly (to a city boy like me), application of potash actually decreases pea yields.

Fitting the analysis of variance model (2.1.1) provides residuals that can be evaluated in the usual way. Figures 2.1 and 2.2 contain residual plots. Except for some slight curvature at the very ends of the normal plot, the residuals look good. Remember that these are plots of the residuals, not the standardized residuals, so residual values greater than 3 do not necessarily contain a suggestion of outlying points. □

Computer commands

R commands appear in the document mentioned in the preface.

Minitab 16's 'glm' command gives a simple way of obtaining the analysis of these data. The glm command does not recognize the orthogonality in the design, so it reports two types of sums of squares for each term. However, the orthogonality ensures that the values are identical for the two types. This particular analysis

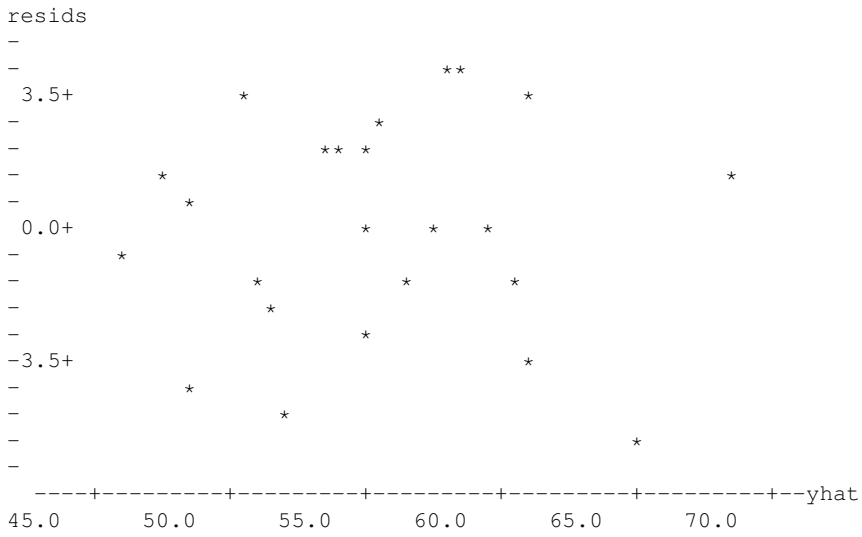

Fig. 2.1 Plot of residuals versus predicted values, pea data.

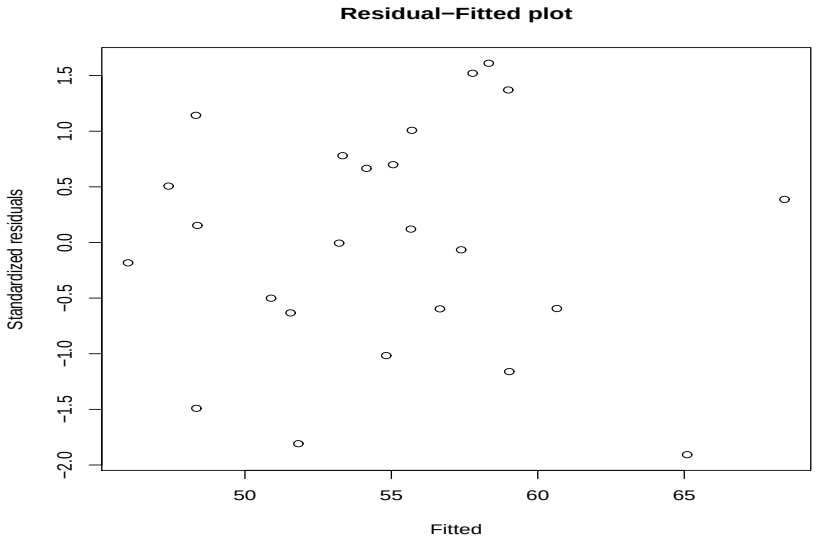

Fig. 2.2 Plot of residuals versus predicted values, pea data.

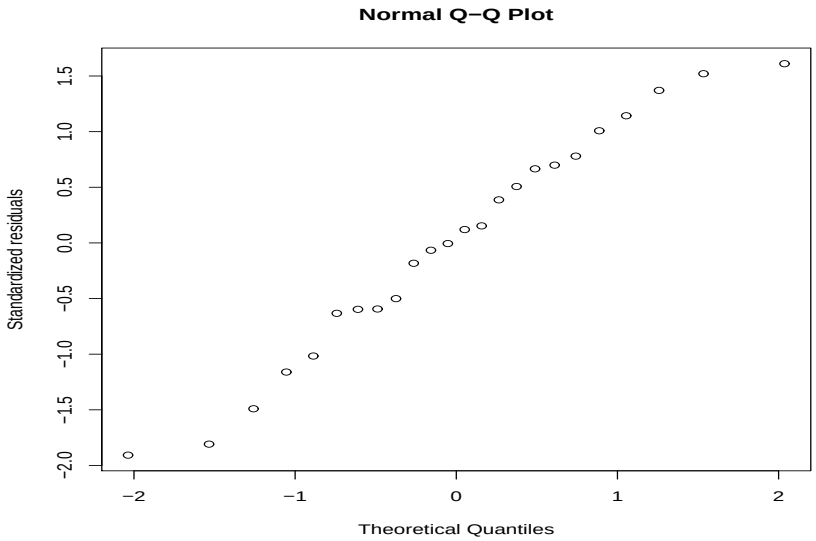

Fig. 2.3 Normal plot of residuals, pea data, $W' = 0.981$.

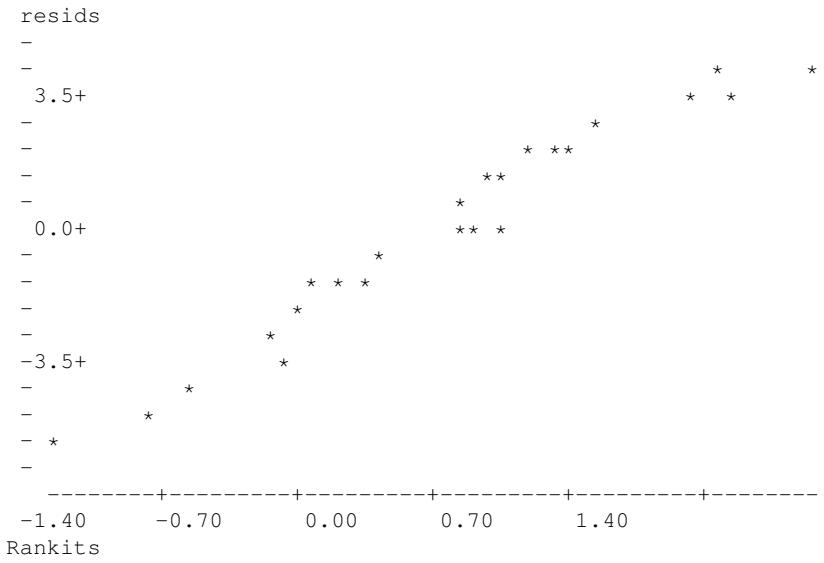

Fig. 2.2 Normal plot of residuals, pea data, $W' = 0.980$.

could also be run using the 'ancova' command, but the other analyses in this chapter require glm.

```
MTB > names c1 'y' c2 'Blocks' c3 'N' c4 'P' c5 'K'
MTB > glm c1=c2 c3|c4|c5 - c3*c4*c5;
SUBC> resid c10;
SUBC> fits c11;
SUBC> means c3 c5.
```

I became unhappy with how Minitab 18 dealt with unbalanced ANOVAs. I have not yet examined Minitab 19.

2.1.1 Split plot analysis

EXAMPLE 2.1.4. *Split plot analysis of a* 2^3 *in blocks of four with replication*
Example 2.1.3 gives the standard analysis of confounded data with replications. However, more information can be extracted. Yates' design is very similar to a split plot design. We have three replications and we can think of each block as a whole plot. The whole plots are randomly assigned one of two treatments but the treatments have a peculiar structure. One 'treatment' applied to a whole plot consists of having the $NPK(1)$ treatments assigned to the whole plot and the other 'treatment' is having the $NPK(-1)$ treatments assigned to a whole plot. The one degree of freedom for whole plot treatments is due to the difference between these sets of treatments. This difference is just the *NPK* interaction. The analysis in Example 2.1.3 is the subplot analysis and remains unchanged when we consider applying a split plot analysis to the data. Recall that in a subplot analysis each whole plot is treated like a block; that is precisely what we did in Example 2.1.3. We need only perform the whole plot analysis to complete the split plot analysis.

The 6 blocks in Example 2.1.3 are now considered as whole plots. The 5 degrees of freedom for blocks are the 5 degrees of freedom total for the whole plot analysis. These 5 degrees of freedom can be decomposed into 2 degrees of freedom for comparing the three replications, 1 degree of freedom for *NPK* interaction, i.e., whole plot treatments, and 2 degrees of freedom for replication by whole plot treatment interaction. *The replication by whole plot treatment interaction is just the error term for the whole plot analysis*. Note that in this model, rather than thinking about having 6 fixed block effects, we have 3 fixed replication effects, a fixed *NPK* effect, and a random error term that distinguishes the blocks within the replications.

The necessary means for the whole plot analysis are given below.

$N=4$	NPK level		
Rep.	$NPK(1)$	$NPK(-1)$	Rep. means
1	56.350	60.775	58.5625
2	50.525	57.450	53.9875
3	54.025	50.125	52.0750
NPK means	$53.63\bar{3}$	$56.11\bar{6}$	54.8750

The three rep. means are averages over the 8 observations in each replication. The mean square for replications is obtained in the usual way as 8 times the sample variance of the rep. means. The mean square for NPK interaction is obtained from the $NPK(1)$ mean and the $NPK(-1)$ mean. Each of these means is averaged over 12 observations. The mean square for NPK is 12 times the sample variance of the NPK means. The interaction (whole plot error) sum of squares is found by subtracting the NPK and replication sums of squares from the blocks sum of squares found in Example 2.1.3. The whole plot analysis of variance is given in Table 2.5. There is no evidence for an NPK interaction.

Table 2.5 Whole plot analysis of variance.

Source	df	SS	MS	F
Reps	2	177.80	88.90	1.38
NPK	1	37.00	37.00	0.58
Error	2	128.50	64.25	
Total	5	343.30	68.66	

In this experiment, there are not enough whole plots to provide a very powerful test for the NPK interaction. There are only 2 degrees of freedom in the denominator of the F test. If we had 5 replications of a 2^3 in blocks of size 2 rather than blocks of size 4, i.e., if we confounded three contrasts with blocks, the whole plot analysis would have 4 degrees of freedom for reps., three 1 degree of freedom effects for the confounded contrasts, and 12 degrees of freedom for whole plot error. The 12 error degrees of freedom come from pooling the rep. by effect interactions. □

2.1.2 Partial confounding

In Example 2.1.3, we considered a 2^3 in blocks of 4 with three replications. The same contrast NPK was confounded in all replications, so, within the subplot analysis, all information was lost on the NPK contrast. Something must be lost when a treatment effect is confounded with blocks, but with multiple replications it is not necessary to give up all the information on NPK. Consider a 2^3 experiment with factors A, B, and C that is conducted in blocks of size 4 with two replications. It would be natural to confound ABC with blocks. However, instead of confounding

ABC with blocks in both replications, we could pick another contrast, say, *BC*, to confound in the second replication. The design is given below.

Replication 1		Replication 2	
$ABC(1)$	$ABC(-1)$	$BC(1)$	$BC(-1)$
$a_0b_0c_0$	$a_0b_0c_1$	$a_0b_0c_0$	$a_0b_0c_1$
$a_0b_1c_1$	$a_0b_1c_0$	$a_0b_1c_1$	$a_0b_1c_0$
$a_1b_0c_1$	$a_1b_0c_0$	$a_1b_0c_0$	$a_1b_0c_1$
$a_1b_1c_0$	$a_1b_1c_1$	$a_1b_1c_1$	$a_1b_1c_0$

In the first replication we give up all the information on *ABC* but retain information on *BC*. In the second replication we give up all the information on *BC* but retain information on *ABC*. Thus we have partial information available on both *ABC* and *BC*.

The process of confounding different contrasts in different replications is known as *partial confounding* because contrasts are only partially confounded with blocks. In some replications they are confounded but in others they are not. Thus partial information is available on contrasts that are partially confounded.

When more than one confounded contrast is needed to define blocks of an appropriate size, some contrasts can be totally confounded, while others are only partially confounded. We now consider an example using these ideas.

EXAMPLE 2.1.5. We again analyze pea yield data with the same 2^3 fertilizer treatments considered in Examples 2.1.3 and 2.1.4. This analysis involves blocks of size 2 with two replications. Both replications have *NPK* confounded, but the first rep. has *PK* (and thus *N*) confounded, while the second has *NK* (and thus *P*) confounded. The data are given in Table 2.6 with lines used to identify blocks. Again, pairs of rows denote replications.

Table 2.6 Pea data with partial confounding.

$n_1p_0k_0$(59.8)	$n_0p_0k_1$(55.5)	$n_1p_0k_1$(57.2)	$n_0p_1k_1$(53.2)
$n_1p_1k_1$(58.5)	$n_0p_1k_0$(56.0)	$n_1p_1k_0$(59.0)	$n_0p_0k_0$(56.0)
$n_0p_1k_1$(49.5)	$n_0p_0k_0$(46.8)	$n_0p_1k_0$(62.8)	$n_1p_0k_0$(69.5)
$n_1p_1k_0$(62.8)	$n_1p_0k_1$(57.0)	$n_1p_1k_1$(55.8)	$n_0p_0k_1$(55.0)
The first replication has *NPK*, *PK*, and *N* confounded.			
The second replication has *NPK*, *NK*, and *P* confounded.			

The mean square and sum of squares for blocks is obtained from the eight block means. Each of these means is the average of 2 observations. The sums of squares for error is obtained by subtraction. The sums of squares for treatment effects are more complicated. We use contrasts to compute them. First, *NPK* is confounded in both replications, so no information is available on it. The effects of *K* and *NP* are not confounded in either replication, so they can be estimated from both. The contrasts, means, and sums of squares are given below.

Treatment	K	NP	Means
$n_0p_0k_0$	1	1	51.40
$n_0p_0k_1$	−1	1	55.25
$n_0p_1k_0$	1	−1	59.40
$n_0p_1k_1$	−1	−1	51.35
$n_1p_0k_0$	1	−1	64.65
$n_1p_0k_1$	−1	−1	57.10
$n_1p_1k_0$	1	1	60.90
$n_1p_1k_1$	−1	1	57.15
SS	60.0625	15.210	

The sum of squares for K is

$$SS(K) = \frac{[-15.50]^2}{8/2} = 60.0625$$

where −15.50 is the estimated contrast, 8 is the sum of the squared contrast coefficients, and the means are averages of 2 observations, one from each replication.

The effects of N and PK are confounded in the first replication but P and NK are not. Thus P and NK can be evaluated from the first replication.

Treatment	P	NK	Rep. 1
$n_0p_0k_0$	1	1	56.0
$n_0p_0k_1$	1	−1	55.5
$n_0p_1k_0$	−1	1	56.0
$n_0p_1k_1$	−1	−1	53.2
$n_1p_0k_0$	1	−1	59.8
$n_1p_0k_1$	1	1	57.2
$n_1p_1k_0$	−1	−1	59.0
$n_1p_1k_1$	−1	1	58.5
SS	0.405	0.005	

The sum of squares for P is

$$SS(P) = \frac{[-1.8]^2}{8/1} = 0.405$$

where −1.8 is the estimated contrast and the treatment 'means' used in the contrasts are just the observations from the first replication.

Note that the sums of squares for the partially confounded contrasts involve a multiplier of 1, because they are averages of one observation, rather than the multiplier of 2 that is used for the contrasts that have full information. This is the price paid for partial confounding. An estimated contrast with full information would have a sum of squares that is twice as large as the same estimated contrast that is

confounded in half of the replications. This sort of thing happens generally when using partial confounding but the size of the multiplicative effect depends on the exact experimental design. In this example the factor is 2 because the sample sizes of the means are twice as large for full information contrasts as for partial information contrasts.

Similarly, the effects of N and PK can be obtained from the second replication but not the first.

Treatment	N	PK	Rep. 2
$n_0p_0k_0$	−1	−1	46.8
$n_0p_0k_1$	−1	1	55.0
$n_0p_1k_0$	−1	1	62.8
$n_0p_1k_1$	−1	−1	49.5
$n_1p_0k_0$	1	−1	69.5
$n_1p_0k_1$	1	1	57.0
$n_1p_1k_0$	1	1	62.8
$n_1p_1k_1$	1	−1	55.8
SS	120.125	32.000	

The analysis of variance table is given in Table 2.7. Too much blocking has been built into this experiment. The F statistic for blocks is only 0.85, so the differences between blocks are not as substantial as the error. The whole point of blocking is that the differences between blocks should be greater than error so that by isolating the block effects we reduce the experimental error. The excessive blocking has also reduced the degrees of freedom for error to 2, thus ensuring a poor estimate of the variance. A poor error estimate reduces power; it is difficult to establish that effects are significant. For example, the F statistic for N is 4.92; this would be significant at the 0.05 level if there were 11 degrees of freedom for error, but with only 2 degrees of freedom the P value is just 0.157. □

Table 2.7 Analysis of variance.

Source	df	SS	MS	F	P
Blocks	7	145.12	20.73	0.85	—
N	1	120.12	120.12	4.92	0.157
P	1	0.40	0.40	0.02	0.909
K	1	60.06	60.06	2.46	0.257
NP	1	15.21	15.21	0.62	0.513
NK	1	0.01	0.01	0.00	0.990
PK	1	32.00	32.00	1.31	0.371
Error	2	48.79	24.40		
Total	15	421.72			

In general, *effects are evaluated from all replications in which they are not confounded*. If we had 3 replications with NPK confounded in each and PK and N

confounded only in the first, the sums of squares for N and PK would be based on treatment means averaged over both rep. 2 and rep. 3.

Computer commands

When blocks are listed first in the model, the sequential *sum of squares is appropriate for blocks*. R has no problem fitting this.

Minitab 16's glm command can again provide the analysis, however, the column with the block indices must appear immediately after the equals sign in the statement of the glm model. Minitab provides both sequential and adjusted sums of squares. For all effects other than blocks, these are the same.

```
MTB > names c1 'Blocks' c2 'N' c3 'P' c4 'K' c5 'y'
MTB > glm c5=c1 c2|c3|c4 - c2*c3*c4
```

2.2 Fractional replication

Consider a chemical process in which there are seven factors that affect the process. These could be percentages of four chemical components, temperature of the reaction, choice of catalyst, and choice of different machines for conducting the process. If all of these factors are at just two levels, there are $2^7 = 128$ treatment combinations. If obtaining data on a treatment combination is expensive, running 128 treatment combinations may be prohibitive. If there are just three levels for each factor, there are $3^7 = 2187$ treatment combinations. In fractional replication, one uses only a fraction of the treatment combinations. Of course, if we give up the full factorial, we must lose information in the analysis of the experiment. In fractional replication the treatments are chosen in a systematic fashion so that we lose only the higher-order interactions. (I, for one, do not look forward to trying to interpret 6- and 7-factor interactions anyway.) To be more precise, we do not actually give up the high order interactions, we give up our ability to distinguish them from the main effects and lower order interactions. Effects are *aliased* with other effects. The contrasts we examine involve both a main effect or lower order interaction and some higher-order interactions. Of course we assume the high order interactions are of no interest and treat the contrasts as if they involve only the main effects and low order interactions. That is the whole point of fractional replication. While fractional replication is of most interest when there are large numbers of factors, we will illustrate the techniques with smaller numbers of factors.

EXAMPLE 2.2.1. *A* $1/2$ *rep. of a* 2^3 *experiment*
We now examine the construction of a $1/2$ replicate of a 2^3 factorial. The 2^3 has 8 treatments, so a $1/2$ replicate involves 4 treatments. Recall the table of contrast coefficients for a 2^3 factorial.

Treatment	A	B	C	AB	AC	BC	ABC
$a_0b_0c_0$	1	1	1	1	1	1	1
$a_0b_0c_1$	1	1	−1	1	−1	−1	−1
$a_0b_1c_0$	1	−1	1	−1	1	−1	−1
$a_0b_1c_1$	1	−1	−1	−1	−1	1	1
$a_1b_0c_0$	−1	1	1	−1	−1	1	−1
$a_1b_0c_1$	−1	1	−1	−1	1	−1	1
$a_1b_1c_0$	−1	−1	1	1	−1	−1	1
$a_1b_1c_1$	−1	−1	−1	1	1	1	−1

Every contrast divides the treatments into two groups of four. We can use any contrast to define the $1/2$ rep. We choose to use *ABC* because it is the highest order interaction. The *ABC* contrast defines two groups of treatments

ABC coefficients	
$ABC(1)$	$ABC(-1)$
$a_0b_0c_0$	$a_0b_0c_1$
$a_0b_1c_1$	$a_0b_1c_0$
$a_1b_0c_1$	$a_1b_0c_0$
$a_1b_1c_0$	$a_1b_1c_1$

These are just the two blocks obtained when confounding a 2^3 into two blocks of 4. Each of these groups of four treatments comprises a $1/2$ replicate. It is irrelevant which of the two groups is actually used. These $1/2$ replicates are referred to as resolution III designs because the defining contrast involves three factors.

The $1/2$ rep. involves only four treatments, so there can be at most three orthogonal treatment contrasts. All seven effects cannot be estimated. The aliasing of effects is determined by the modular multiplication illustrated earlier. To determine the aliases, multiply each effect by the defining contrast *ABC*. For example, to find the alias of A, multiply

$$A \times ABC = A^2BC = BC$$

where any even power is treated as 0, so $A^2 = A^0 = 1$. Thus A and BC are aliased; we cannot tell them apart; they are two names for the same contrast. Similarly,

$$BC \times ABC = AB^2C^2 = A.$$

The aliasing structure for the entire $1/2$ rep. based on *ABC* is given below.

Effect	$\times ABC$	Alias
A	=	BC
B	=	AC
C	=	AB
AB	=	C
AC	=	B
BC	=	A
ABC		—

In this experiment, we completely lose any information about the defining contrast, *ABC*. In addition, we cannot tell the main effects from the two-factor interactions. If we had no interest in two-factor interactions, this design would be fine. Generally, if there are only three factors each at two levels, there is little reason not to perform

the entire experiment. As mentioned earlier, fractional replication is primarily of interest when there are many factors so that even a fractional replication involves many observations.

Another way to examine aliasing is by looking at the table of contrast coefficients when we use only $1/2$ of the treatments. We consider the $1/2$ rep. in which the treatments have *ABC* coefficients of 1. The contrasts, when restricted to the treatments actually used, have the following coefficients:

Treatment	A	B	C	AB	AC	BC	ABC
$a_0b_0c_0$	1	1	1	1	1	1	1
$a_0b_1c_1$	1	−1	−1	−1	−1	1	1
$a_1b_0c_1$	−1	1	−1	−1	1	−1	1
$a_1b_1c_0$	−1	−1	1	1	−1	−1	1

All of the columns other than *ABC* still define contrasts in the four treatment combinations; each column has two 1s and two −1s. However, the contrast defined by A is identical to the contrast defined by its alias, *BC*. In fact, this is true for each contrast and its alias.

Consider now the choice of a different contrast to define the $1/2$ rep. Instead of *ABC*, we might choose *AB*. Again, we lose all information about the defining contrast *AB* and we have aliases involving the other effects. The *AB* contrast defines two groups of treatments

AB coefficients	
$AB(1)$	$AB(-1)$
$a_0b_0c_0$	$a_0b_1c_0$
$a_0b_0c_1$	$a_0b_1c_1$
$a_1b_1c_0$	$a_1b_0c_0$
$a_1b_1c_1$	$a_1b_0c_1$

Each of these groups of four treatments comprises a $1/2$ replicate. Again, it is irrelevant which of the two groups is actually used. Both groups determine resolution II designs because the defining contrast involves two factors.

The aliasing is determined by modular multiplication with the defining contrast *AB*. To find the alias of A multiply

$$A \times AB = A^2B = B.$$

The alias of *BC* is

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

The aliasing structure for the entire $1/2$ rep. based on *AB* is given below.

Effect	$\times AB$	Alias
A	=	B
B	=	A
C	=	ABC
AB		—
AC	=	BC
BC	=	AC
ABC	=	C

With this $1/2$ rep., we do not even get to estimate all of the main effects because A is aliased with B. □

EXAMPLE 2.2.2. *A* $1/4$ *replicate of a* 2^4 *experiment*
In Example 2.1.2 we considered confounding ABC and BCD in a 2^4 experiment. The four blocks are given below.

$ABC(1)$		$ABC(-1)$	
$BCD(1)$	$BCD(-1)$	$BCD(1)$	$BCD(-1)$
$a_0b_0c_0d_0$	$a_0b_0c_0d_1$	$a_0b_0c_1d_1$	$a_0b_0c_1d_0$
$a_0b_1c_1d_0$	$a_0b_1c_1d_1$	$a_0b_1c_0d_1$	$a_0b_1c_0d_0$
$a_1b_0c_1d_1$	$a_1b_0c_1d_0$	$a_1b_0c_0d_0$	$a_1b_0c_0d_1$
$a_1b_1c_0d_1$	$a_1b_1c_0d_0$	$a_1b_1c_1d_0$	$a_1b_1c_1d_1$

With ABC and BCD defining the blocks,

$$ABC \times BCD = AB^2C^2D = AD$$

is also confounded with blocks. Any one of these blocks can be used as a $1/4$ replicate of the 2^4 experiment. The smallest defining contrast is the two-factor effect AD, so this $1/4$ replicate is a resolution II design.

The aliasing structure of the $1/4$ rep. must account for all three of the defining contrasts. An effect, say A, is aliased with

$$A \times ABC = A^2BC = BC,$$

$$A \times BCD = ABCD,$$

and

$$A \times AD = A^2D = D.$$

Thus we cannot tell main effects A from main effects D, from BC interaction, or from $ABCD$ interaction. After all, what do you expect when you take four observations to learn about 16 treatments? Similar computations show that

$$B = AC = CD = ABD$$

and

$$C = AB = BD = ACD.$$

This is the complete aliasing structure for the $1/4$ rep. There are 4 observations, so there are 3 degrees of freedom for treatment effects. We can label these effects as A, B, and C with the understanding that we cannot tell aliases apart, so we have no idea if an effect referred to as A is really due, entirely or in part, to D, BC, or $ABCD$. □

Fractional replication is primarily of value when you have large numbers of treatments, require information only on low order effects, can assume that high order effects are negligible, and can find a design that aliases low order effects with high order effects.

EXAMPLE 2.2.3. *Fractional replication of a* 2^8 *experiment*
A 2^8 experiment involves eight factors, A through H, and 256 treatments. It may be impractical to take that many observations. Consider first a $1/8 = 2^{-3}$ replication. This involves only $2^{8-3} = 32$ treatment combinations, a much more manageable number than 256. A $1/8 = 2^{-3}$ rep. requires 3 defining contrasts, say $ABCD$, $EFGH$, and $CDEF$. Multiplying pairs of the defining contrasts and multiplying all three of the contrasts give the other contrasts that implicitly define the $1/8$ rep. The other implicit defining contrasts are $ABED$, $CDGH$, $ABCDEFGH$, and $ABGH$. Note that the smallest defining contrast has four terms, so this is a resolution IV design.

The aliases of an effect are obtained from multiplying the effect by all 7 of the defining contrasts; e.g., for A the aliases are

$$\begin{aligned} A = A(ABCD) = A(EFGH) = A(CDEF) = A(ABED) \\ = A(CDGH) = A(ABCDEFGH) = A(ABGH) \end{aligned}$$

or simplifying

$$A = BCD = AEFGH = ACDEF = BED = ACDGH = BCDEFGH = BGH.$$

With a resolution IV design, it is easily seen that main effects are only aliased with three-factor and higher-order effects. A two-factor effect, say AB, has aliases

$$\begin{aligned} AB = AB(ABCD) = AB(EFGH) = AB(CDEF) = AB(ABED) \\ = AB(CDGH) = AB(ABCDEFGH) = AB(ABGH) \end{aligned}$$

or simplifying

$$AB = CD = ABEFGH = ABCDEF = ED = ABCDGH = CDEFGH = GH.$$

Unfortunately, at least some two-factor effects are aliased with other two-factor effects in a resolution IV design.

If we had constructed a $1/4$ replicate, we could have chosen the defining contrasts in such a way that two-factor effects were only aliased with three-factor and higher-order effects. For example, the defining contrasts $ABCDE$ and $DEFGH$ determine such a design. The additional defining contrast is $ABCDE(DEFGH) = ABCFGH$. The smallest defining effect involves 5 factors, so this has resolution V. In computing aliases, a two-factor term is multiplied by a five-factor or greater term. The result is at least a three-factor term. Thus two-factor effects are aliased with 3 or higher-order effects. Similarly, main effects are aliased with 4 or higher-order effects. A $1/4$ replicate of a 2^8 experiment can provide information on all main effects and all two-factor interactions under the assumption of no 3 or higher-order interaction effects. □

As mentioned, the $1/4$ replicate given above is known as a resolution V design because the smallest defining contrast involved a five-factor interaction. As should now be clear, in general, *the resolution of a* 2^f *fractional replication is the order*

of the smallest defining contrast. To keep main effects from being aliased with one another, one needs a resolution III or higher design. To keep both main effects and two-factor effects from being aliased with one another, one needs a resolution V or higher design.

2.2.1 Fractional replication with confounding

The two concepts of fractional replication and confounding can be combined in designing an experiment. To illustrate fractional replication with confounding we consider a subset of the 2^3 data in Table 2.6. The subset is given in Table 2.8. This is the first half of the first replication in Table 2.6. The fractional replication is based on NPK. All of the observations have NPK contrast coefficients of -1. The confounding is based on PK. The first block has PK contrast coefficients of 1 and the second block has PK contrast coefficients of -1.

Table 2.8 Pea data.

$n_1p_0k_0$(59.8)	$n_0p_0k_1$(55.5)
$n_1p_1k_1$(58.5)	$n_0p_1k_0$(56.0)
The fractional replication is based on NPK.	
Confounding is based on PK.	

The aliasing structure for the $1/2$ rep. based on NPK is given below.

Effect	$\times NPK$	Alias
N	=	PK
P	=	NK
K	=	NP
NP	=	K
NK	=	P
PK	=	N
NPK		—

Blocks are confounded with PK and PK is aliased with N, so N is also confounded with blocks. With only 4 observations, we can compute sums of squares for only 3 effects. Ignoring the two-factor interactions, those effects are blocks, P, and K.

Perhaps the simplest way to perform the analysis of such designs is to begin by ignoring the blocking. If the blocking is ignored, the analysis is just that of a fractional factorial and can be conducted as discussed in the next section. After computing all the sums of squares ignoring blocks, go back and isolate the effects that are confounded with blocks. In this example, the fractional factorial ignoring blocks gives sums of squares for $N = PK$, $P = NK$, and $NP = K$. Then observe that the sum of squares for N is really the sum of squares for blocks.

2.3 Analysis of unreplicated experiments

One new problem we have in a fractional replication is that there is no natural estimate of error because there is no replication. We don't even have observations on every factor combination, much less multiple observations on treatments. We present two ways to proceed, one is to assume that higher-order interactions do not exist, the other is based on a graphical display of the effects that is similar in spirit to a normal plot.

EXAMPLE 2.3.1. We consider the $1/2$ rep. of a 2^5 that was reported by Hare (1988) and introduced in Chapter 1. The issue is excessive variability in the taste of a dry soup mix. The source of variability was identified as a particular component of the mix called the 'intermix' containing flavorful ingredients such as salt and vegetable oil.

From each batch of intermix, the original data are groups of 5 samples taken every 15 minutes throughout a day of processing. Thus each batch yields data for a balanced one-way analysis of variance with $N = 5$. The data actually analyzed are derived from the ANOVAs on different batches. There are two sources of variability in the original observations, the variability within a group of 5 samples and variability that occurs between 15 minute intervals. From the analysis of variance data, the within group variability is estimated with the MSE and summarized as the estimated 'capability' standard deviation

$$s_c = \sqrt{MSE}.$$

The 'process' standard deviation was defined as the standard deviation of an individual observation. The standard deviation of an observation incorporates both the between group and the within group sources of variability. The estimated process standard deviation is taken as

$$s_p = \sqrt{MSE + \frac{MSGrps - MSE}{5}},$$

where the 5 is the number of samples taken at each time, cf. Christensen (1996, Subsection 12.4.2). These two statistics, s_c and s_p, are available from every batch of soup mix prepared and provide the data for analyzing batches. The $1/2$ rep. of a 2^5 involves different ways of making batches of soup mix. The factors in the design were discussed in Example 1.0.1 and are repeated below. For now, we analyze only the data on s_p.

The two blocks obtained by confounding $ABCDE$ in a 2^5 are reported in Table 2.9. Table 2.9 also presents an alternative form of identifying the treatments. In the alternative form, only the letters with a subscript of 1 are reported. Hare's experiment used the block consisting of treatments with $ABCDE$ contrast coefficients of 1.

There are five factors involved in the experiment. Intermix is made in a large mixer. Factor A is the number of ports for adding vegetable oil to the mixer. This

Table 2.9 $1/2$ reps. from a 2^5 based on $ABCDE$.

$ABCDE(1)$	Treatment	$ABCDE(-1)$	Treatment
$a_0b_0c_0d_0e_0$	(1)	$a_0b_0c_0d_0e_1$	e
$a_0b_0c_0d_1e_1$	de	$a_0b_0c_0d_1e_0$	d
$a_0b_0c_1d_0e_1$	ce	$a_0b_0c_1d_0e_0$	c
$a_0b_0c_1d_1e_0$	cd	$a_0b_0c_1d_1e_1$	cde
$a_0b_1c_0d_0e_1$	be	$a_0b_1c_0d_0e_0$	b
$a_0b_1c_0d_1e_0$	bd	$a_0b_1c_0d_1e_1$	bde
$a_0b_1c_1d_0e_0$	bc	$a_0b_1c_1d_0e_1$	bce
$a_0b_1c_1d_1e_1$	bcde	$a_0b_1c_1d_1e_0$	bcd
$a_1b_0c_0d_0e_1$	ae	$a_1b_0c_0d_0e_0$	a
$a_1b_0c_0d_1e_0$	ad	$a_1b_0c_0d_1e_1$	ade
$a_1b_0c_1d_0e_0$	ac	$a_1b_0c_1d_0e_1$	ace
$a_1b_0c_1d_1e_1$	acde	$a_1b_0c_1d_1e_0$	acd
$a_1b_1c_0d_0e_0$	ab	$a_1b_1c_0d_0e_1$	abe
$a_1b_1c_0d_1e_1$	abde	$a_1b_1c_0d_1e_0$	abd
$a_1b_1c_1d_0e_1$	abce	$a_1b_1c_1d_0e_0$	abc
$a_1b_1c_1d_1e_0$	abcd	$a_1b_1c_1d_1e_1$	abcde

was set at either 1 (a_0) or 3 (a_1). Factor B is the temperature of the mixer. The mixer can be cooled by circulating water through the mixer jacket (b_0) or the mixer can be used at room temperature (b_1). Factor C is the mixing time, 60 seconds (c_0) or 80 seconds (c_1). Factor D is the size of the intermix batch, either 1500 pounds (d_0) or 2000 pounds (d_1). Factor E is the delay between making the intermix and using it in the final soup mix. The delay is either 1 day (e_0) or 7 days (e_1). Table 2.10 contains the data along with the aliases for a $1/2$ rep. of a 2^5 based on $ABCDE$. The order in which the treatments were run was randomized and they are listed in that order. Batch number 7 contains the standard operating conditions. Note that this is a resolution V design: all main effects are confounded with four-factor interactions and all two-factor interactions are confounded with three-factor interactions. If we are prepared to assume that there are no three- or four-factor interactions, we have estimates of all the main effects and two-factor interactions.

One way to perform the analysis is to compute the sums of squares for each contrast, however the simplest way to obtain an analysis is to let a computer program do most of the work. The R language does this easily, see the computing commands document. Occasionally, when using software written only for balanced analysis of variance, one needs to trick the program into doing the fitting. If we could drop one of the factors, we would have observed a full factorial (without replication) on the remaining factors. For example, *if we dropped factor E, and thus dropped the e terms from all the treatment combinations in Table 2.10, we would have observations on all 16 of the treatment combinations in the* 2^4 *defined by A, B, C, and D*. It is easy to find computer programs that will analyze a full factorial. Table 2.11 gives the results of an analysis in which we have ignored the presence of factor E. Table 2.11 contains two columns labeled 'Source'. The one on the left gives the sources from the full factorial on A, B, C, and D; the one on the right replaces the higher-order

Table 2.10 Hare's $1/2$ rep. from a 2^5 based on $ABCDE$.

Batch	Treatment		s_c	s_p	Aliases
1	$a_0b_0c_0d_1e_1$	de	0.43	0.78	$A = BCDE$
2	$a_1b_0c_1d_1e_1$	acde	0.52	1.10	$B = ACDE$
3	$a_1b_1c_0d_0e_0$	ab	0.58	1.70	$C = ABDE$
4	$a_1b_0c_1d_0e_0$	ac	0.55	1.28	$D = ABCE$
5	$a_0b_1c_0d_0e_1$	be	0.58	0.97	$E = ABCD$
6	$a_0b_0c_1d_0e_1$	ce	0.60	1.47	$AB = CDE$
7	$a_0b_1c_0d_1e_0$	bd	1.04	1.85	$AC = BDE$
8	$a_1b_1c_1d_1e_0$	abcd	0.53	2.10	$AD = BCE$
9	$a_0b_1c_1d_1e_1$	bcde	0.38	0.76	$AE = BCD$
10	$a_1b_1c_0d_1e_1$	abde	0.41	0.62	$BC = ADE$
11	$a_0b_0c_1d_1e_0$	cd	0.66	1.09	$BD = ACE$
12	$a_0b_0c_0d_0e_0$	(1)	0.55	1.13	$BE = ACD$
13	$a_1b_0c_0d_0e_1$	ae	0.65	1.25	$CD = ABE$
14	$a_1b_1c_1d_0e_1$	abce	0.72	0.98	$CE = ABD$
15	$a_1b_0c_0d_1e_0$	ad	0.48	1.36	$DE = ABC$
16	$a_0b_1c_1d_0e_0$	bc	0.68	1.18	

interactions from the full factorial with their lower order aliases. Table 2.11 also contains a ranking of the sizes of the sums of squares from smallest to largest.

Table 2.11 ANOVA for s_p.

Source	Source	df	SS	Rank
A	A	1	0.0841	10
B	B	1	0.0306	7
C	C	1	0.0056	4
D	D	1	0.0056	3
AB	AB	1	0.0009	1
AC	AC	1	0.0361	8
AD	AD	1	0.0036	2
BC	BC	1	0.0182	5
BD	BD	1	0.1056	12
CD	CD	1	0.0210	6
ABC	DE	1	0.3969	13
ABD	CE	1	0.0729	9
ACD	BE	1	0.6561	14
BCD	AE	1	0.0930	11
$ABCD$	E	1	0.8836	15
Total	Total	15	2.4140	

One method of analysis is to assume that no higher-order interactions exist and form an error term by pooling the estimable terms that involve only higher-order interactions. A particular term involves only higher-order interactions if the term and all of its aliases are high order interactions. What we mean by high order interac-

tions is intentionally left ill-defined to maintain flexibility. In this design, unless you consider second-order interactions as higher-order, there are no terms involving only higher-order interactions. Most often, higher-order interactions are taken to be interactions that only involve three or more factors, but in designs like this, one *might* be willing to consider two-factor interactions as higher-order to obtain an error term for testing main effects. (I personally would not be willing to do it with these data.) Often terms that involve only three and higher-order interactions are pooled into an error, but in designs with more factors and many high order interactions, one might wish to estimate three-factor interactions and use only terms involving four or more factors in a pooled error.

If we assume away all two-factor and higher-order interactions for the present data, the ANOVA table becomes that displayed in Table 2.12. With this error term, only factor E appears to be important. As we will see later, most of the important effects in these data seem to be interactions, so the error term based on no interactions is probably inappropriate.

Table 2.12 Analysis of variance on s_p for Hare's data.

Source	*df*	*SS*	*MS*	*F*
A	1	0.0841	0.0841	0.60
B	1	0.0306	0.0306	0.22
C	1	0.0056	0.0056	0.04
D	1	0.0056	0.0056	0.04
E	1	0.8836	0.8836	6.29
Error	10	1.4044	0.1404	
Total	15	2.4140		

Rather than assuming away higher-order interactions, Daniel (1959) proposed an alternative method of analysis based on an idea similar to normal plotting. Recall that in a normal plot, the data from a single sample are ordered from smallest to largest and plotted against the *expected order statistics* from a standard normal distribution. In other words, the smallest observation in a sample of size, say, 13 is plotted against the expected value for the smallest observation in a sample of size 13 from a $N(0,1)$ distribution. The second smallest observation is plotted against the expected value for the second smallest observation in a sample of size 13 from a $N(0,1)$ distribution, and so on. This plot should approximate a straight line if the data are truly normal, the slope of the plot estimates the standard deviation of the population, and the intercept estimates the population mean. One approach to a graphical analysis of 2^f experiments is to perform a normal plot on the estimated contrasts. Daniel (1959) used a plot of the absolute values of the estimated contrasts. Here we discuss a graphical method of analysis for unreplicated and fractional factorials that applies a similar idea to the sums of squares in Table 2.11.

Assume a standard ANOVA model with independent $N(0,\sigma^2)$ errors. The analysis looks for departures from the assumption that none of the factors have an effect

on the observations. Under the assumption of no effects, every mean square gives an estimate of σ^2 and every sum of squares has the distribution

$$\frac{SS}{\sigma^2} \sim \chi^2(1),$$

where the degrees of freedom in the χ^2 are 1 because each effect has 1 degree of freedom. Moreover, the sums of squares are independent, so in the absence of treatment effects, the sums of squares form a random sample from a $\sigma^2\chi^2(1)$ distribution. If we order the sums of squares from smallest to largest, the ordered sums of squares should estimate the expected order statistics from a $\sigma^2\chi^2(1)$ distribution. Plotting the ordered sums of squares against the expected order statistics, we should get an approximate straight line through the origin with a slope of 1. In practice, we cannot obtain expected values from a $\sigma^2\chi^2(1)$ distribution because we do not know σ^2. Instead, we plot the ordered sums of squares against the expected order statistics of a $\chi^2(1)$ distribution. This plot should be an approximate straight line through the origin with a slope of σ^2.

Table 2.13 contains the statistics necessary for the χ^2 plot of the 15 effects from Hare's data. Figure 2.3 contains the plot. The $\chi^2(1)$ scores in Table 2.13 are approximate expected order statistics. They are computed by applying the inverse of the $\chi^2(1)$ cumulative distribution function to the values $i/(n+1)$, where i goes from 1 to 15 and $n = 15$. This is easily done in R and Minitab. Table 2.13 also contains partial sums of the ordered sums of squares; these values will be used in the next section.

Table 2.13 $\chi^2(1)$ scores, ordered sums of squares, and partial sums of the sums of squares for Hare's (1988) data.

$\chi^2(1)$ scores	Ordered *SS*	Partial sums
0.00615	0.0009	0.0009
0.02475	0.0036	0.0045
0.05626	0.0056	0.0101
0.10153	0.0056	0.0157
0.16181	0.0182	0.0339
0.23890	0.0210	0.0549
0.33539	0.0306	0.0855
0.45494	0.0361	0.1216
0.60283	0.0729	0.1945
0.78703	0.0841	0.2786
1.02008	0.0930	0.3716
1.32330	0.1056	0.4772
1.73715	0.3969	0.8741
2.35353	0.6561	1.5302
3.46977	0.8836	2.4138

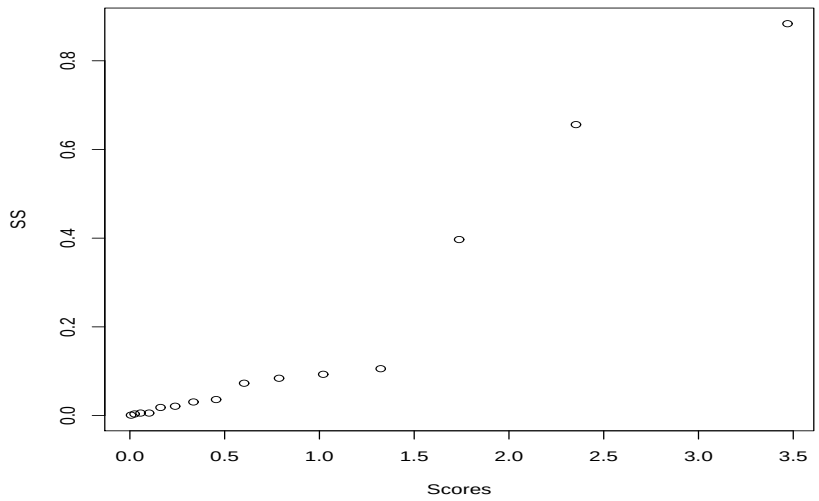

Fig. 2.3 $\chi^2(1)$ plot of sums of squares.

The key to the graphical analysis is that nonnegligible treatment effects cause the mean square to estimate something larger than σ^2. The sums of squares for nonnegligible effects should show up in the plot as inappropriately large values. The lower 12 observations in Figure 2.3 seem to fit roughly on a line, but the three largest observations seem to be inconsistent with the others. These three observations correspond to the most important effects in the data. From the rankings in Table 2.11, we see that the important effects are E, BE, and DE.

We need to evaluate the meaning of the important effects. There is no question of breaking things down into contrasts because all of the effects already have only one degree of freedom. We need only interpret the meanings of the specific effects. The largest effect is due to E, the delay in using the intermix. However, this effect is complicated by interactions involving the delay.

The means for the four combinations of B and E are given below.

$N=4$	B	
E	b_0	b_1
e_0	1.215	1.7075
e_1	1.150	0.8325

The BE interaction is due to the fact that running the mixer at room temperature, b_1, increases variability if the intermix is used after one day, e_0, but decreases variability if the intermix is used a week later, e_1. However, the variability under delay is

smaller for both B levels than the variability for immediate use with either B level. This suggests delaying use of the intermix.

The means for the four combinations of D and E are given below.

$N = 4$	D	
E	d_0	d_1
e_0	1.3225	1.6000
e_1	1.1675	0.8150

A large batch weight, d_1, causes increased variability when the intermix is used immediately but decreased variability with use delayed to 7 days. Again, it is uniformly better to delay.

Figure 2.4 contains a plot of the remaining sums of squares after deleting the three largest effects. The plot indicates that the four largest values are somewhat larger than the remaining effects. The fourth through seventh largest effects are BD, AE, A, and CE. These may be important but the results are less clear. Figure 2.5 is an alternative to Figure 2.4. Figure 2.4 simply dropped the three largest cases in Figure 2.3 to give a better view of the remainder of the plot. In Figure 2.5 the three largest sums of squares from Figure 2.3 are dropped but the expected order statistics are recomputed for a sample of size $15 - 3 = 12$.

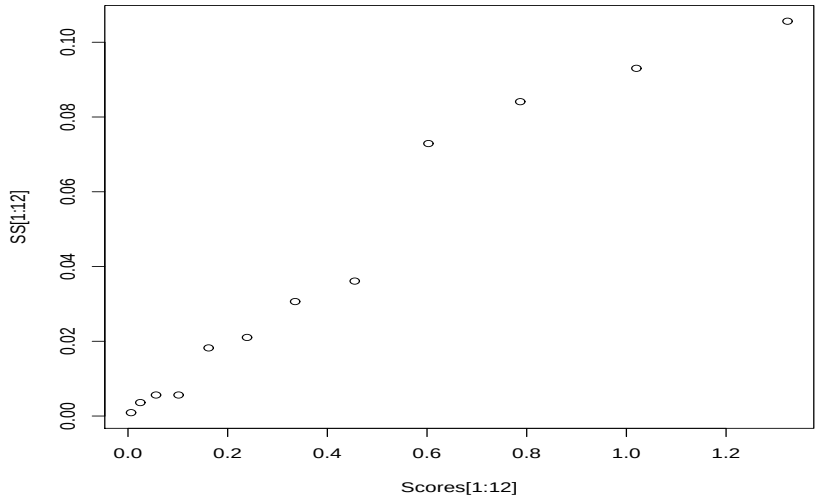

Fig. 2.4 $\chi^2(1)$ plot of sums of squares, largest 3 cases deleted.

The suggestions of treatment effects in these plots are not sufficiently clear to justify their use in the analysis. Recalling that the current process is batch 7 with

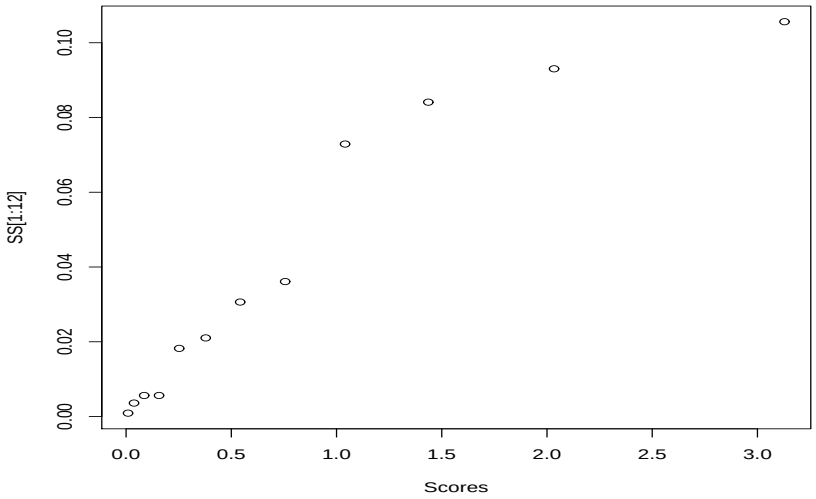

Fig. 2.5 $\chi^2(1)$ plot of sums of squares, largest 3 cases deleted, expected order statistics recomputed.

one vegetable oil port, room temperature mixing, 60 seconds mixing time, 2000 pound batches, and a 1 day delay, we would recommend changing to a 7 day delay. □

Computing commands

R commands are given in the separate document.

When I originally wrote this, I could not get Minitab to give me the output directly. Below are Minitab commands for obtaining the analysis. The data file had eight columns, the first six were indicators for batch and factors A, B, C, D and E, respectively. Columns 7 and 8 contained the data on s_c and s_p.

```
MTB > names c8 'y' c2 'a' c3 'b' c4 'c' c5 'd' c6 'e'
MTB > anova c8=c2|c3|c4|c5 - c2*c3*c4*c5
MTB > note     AFTER SEEING THE ANOVA, ENTER THE SUMS
MTB > note     OF SQUARES INTO c10.
MTB > set c10
DATA> 841 306 56 56 9 361 36 182
DATA> 1056 210 3969 729 6561 930 8836
DATA> end
```

```
MTB > let c10=c10/10000
MTB > note     CONSTRUCT CHI-SQUARED SCORES AND PLOT.
MTB > rank c10 c11
MTB > let c11=c11/16
MTB > invcd  f c11 c12;
SUBC> chisquare 1.
MTB > plot c10 c12
```

Note that $c6$ was not used in the anova command. Factor E was dropped to deal with the fractional nature of the factorial. Minitab's ANOVA command requires an error term to exist in the model. The command given above specifies a full factorial model ($c2|c3|c4|c5$) but subtracts out the *ABCD* interaction ($c2*c3*c4*c5$) and sets it equal to the error. Thus, Minitab's error term is actually the *ABCD* interaction. The command 'set $c10$' is used to create a data column that contains the sums of squares for the various effects. The commands involving $c11$ and $c12$ are used to get the approximate expected order statistics from a $\chi^2(1)$ and to plot the ordered sums of squares against the expected order statistics. After identifying the important effects, the ANOVA command can be repeated with various factors deleted to obtain the necessary means tables.

2.3.1 Balanced ANOVA computing techniques

When using software written only for balanced analysis of variance, the technique of computing the sums of squares in a fractional factorial by dropping factors and performing a full factorial analysis on the remaining factors is quite general, but when choosing factors to be dropped *every defining contrast must involve at least one dropped factor*. For example, if we used *ABCD* as a defining contrast in a $1/2$ rep. of a 2^5, we must drop A, B, C, or D to compute a full factorial. Dropping any of these will give all 16 treatment combinations in a 2^4 based on E and the other three factors. On the other hand, dropping E does not give all 16 treatments combinations that are present in a 2^4 based on factors A, B, C, and D, nor does dropping E give the appropriate sums of squares. In particular, a factorial analysis with factors A, B, C, and D normally has terms for both A and *BCD*, but these are aliased in the $1/2$ rep. based on *ABCD*. Thus the full factorial cannot be computed. (If you are confused, do Exercise 2.6.7.)

In a $1/4$ rep, two defining contrasts are used with another contrast also lost. In the analysis of a $1/4$ rep, two factors are dropped and a full factorial is computed on the remaining factors. Again, at least one of the dropped factors must be in each of the three defining contrasts. For example, in a 2^5 with defining contrasts *BCD*, *CDE* and implicitly *BE*, we could not drop the two factors C and D because the defining contrast *BE* contains neither of these. In particular, dropping C and D leads to a factorial on A, B, and E, but the main effects for B and E are aliased. Similarly, the factor A cannot be one of those dropped to obtain a full factorial. Since A is not

contained in any of the defining contrasts, the other dropped factor would have to be in all three. This is impossible because if a factor is in two defining contrasts, when they are multiplied to obtain the third defining contrast that factor will not be present.

Generally, in a $1/2^s$ replication of a 2^f factorial there are $2^{f-s}-1$ distinct groups of effects that are aliased. We need to find the sum of squares for each group. To do this we drop s appropriately chosen factors and compute a full factorial analysis on the remaining factors. The effects in this analysis of a 2^{f-s} factorial represent all of the alias groups in the $1/2^s$ replication. We merely have to identify the lowest order, and thus most interesting, effects in each group of aliases. Having at least one dropped factor in every defining contrast ensures that the effects arising in the 2^{f-s} factorial are all aliased only with effects that involve a dropped factor and thus are not aliased with any other effect in the 2^{f-s} factorial. Therefore all the effects in the 2^{f-s} factorial are in distinct alias groups and we have sums of squares for every alias group.

2.4 More on graphical analysis

Normal and χ^2 graphs of effects give valuable information on the relative sizes of effects but it is difficult to judge which effects are truly important and which could be the result of random variation. Many such plots give the false impression that there are a number of important effects. The problem is that we tend to see what we look for. In a normal plot of, say, regression residuals, we look for a straight line and are concerned if the residuals obviously contradict the assumption of normality. When analyzing a saturated linear model, we expect to see important effects, so instead of looking for an overall line, we focus on the extreme order statistics. Doing so can easily lead us astray.

In this section we consider two methods for evaluating significant effects in a χ^2 plot. The first method was originally suggested by Holms and Berrettoni (1969). It uses Cochran's (1941) test for homogeneity (equality) of variances. The second is similar in spirit to the simulation envelopes suggested by Atkinson (1981) for evaluating normal plots of regression residuals. The methods are introduced in relation to Hare's data but they apply quite generally. Both methods provide envelopes for the χ^2 plots. A χ^2 plot that goes outside the envelope suggests the existence of significant effects. Box and Meyer (1986), Lenth (1989), and Berk and Picard (1991) proposed alternative methods for the analysis of contrasts in unreplicated factorials and Lenth (2015) argues against methods based on plotting effects.

2.4.1 Multiple maximum tests

Cochran's test for homogeneity of variances applies when there are, say, $n \geq 2$ independent $\chi^2(r)$ estimates of a variance. In a balanced one-way ANOVA with a treatments, N observations per group, and independent $N(0, \sigma^2)$ errors, Cochran's test applies to the individual group variances s_i^2. Cochran's n equals a and his r is $N-1$. In this case, Cochran's test statistic is the maximum of the variance estimates divided by the sum of the variance estimates. The test is rejected for large values of the statistic. In analyzing Hare's unreplicated factorial, if there are no significant effects, the 15 sums of squares in Hare's data are independent $\chi^2(1)$ estimates of σ^2. Thus Cochran's procedure can be applied with $n = 15$ and $r = 1$ to test the hypothesis that all the sums of squares are estimating the same variance.

Cochran's test is best suited for detecting a single variance that is larger than the others. (Under the alternative, large terms *other than* the maximum get included in the total, making it more difficult to detect unusual behavior in the maximum.) In analyzing unreplicated linear models we often expect more than one significant effect. In the spirit of the multiple range tests used for comparing pairs of means in analysis of variance (e.g. Christensen, 1996, Chapter 6), we use Cochran's test repeatedly to evaluate the ordered sums of squares. Thus we define C_j as the jth smallest of the sums of squares divided by the sum of the j smallest sums of squares. From Table 2.13, the values of C_j are obtained by taking the ordered sums of squares and dividing by the partial sums of the ordered sums of squares. Each value of C_j is then compared to an appropriate percentile of Cochran's distribution based on having j estimates of the variance. Note that such a procedure does not provide any control of the experimentwise error rate for the multiple comparisons. Weak control can be achieved by first performing an overall test for equality of variances and then evaluating individual C_js only if this overall test is significant. One such choice could be Cochran's test for the entire collection of sums of squares, but that seems like a poor selection. As mentioned, Cochran's test is best at detecting a single unusual variance; having more than one large variance (as we often expect to have) reduces the power of Cochran's test. To control the experimentwise error rate, it is probably better to use alternative tests for equality of variances such as Bartlett's (1937) or Hartley's (1938) tests, see Snedecor and Cochran (1980). Currently the most popular tests seem to be the Breusch-Pagan/Cook-Weisberg test and Levene's test.

Table 2.14 gives the values of the C_j statistics and various percentiles of Cochran's distribution. Note that the 13th largest effect exceeds the 0.15 percentage point and is almost significant at the 0.10 level. While a significance level of 0.15 is not commonly thought to be very impressive, in an unreplicated experiment one would not expect to have a great deal of power, so the use of larger α levels may be appropriate. Having concluded that the 13th effect is significant, it is logical to conclude that all larger effects are also significant. Once we have a single significant effect, testing the larger effects makes little sense. For larger effects, the sum of squares in the denominator of Cochran's statistic is biased by the inclusion of the sum of squares for the 13th effect. Moreover, if the 13th effect is so large as to be identified

as significant, effects that are even larger should also be significant. Note that for Hare's data the test of the 14th effect is also significant at the 0.15 level. This simply compounds the evidence for the significance of the 14th effect.

Table 2.14 Percentiles for Cochran's statistic with $r = 1$ and Cochran's statistics for Hare's data.

j	0.01	0.05	0.10	0.15	C_j
2	0.9999	0.9985	0.9938	0.9862	0.80000
3	0.9933	0.9670	0.9344	0.9025	0.55446
4	0.9677	0.9065	0.8533	0.8096	0.35669
5	0.9279	0.8413	0.7783	0.7311	0.53687
6	0.8826	0.7808	0.7141	0.6668	0.38251
7	0.8377	0.7270	0.6598	0.6139	0.35789
8	0.7945	0.6798	0.6138	0.5696	0.29688
9	0.7549	0.6385	0.5742	0.5320	0.37481
10	0.7176	0.6020	0.5399	0.4997	0.30187
11	0.6837	0.5697	0.5100	0.4716	0.25027
12	0.6528	0.5411	0.4834	0.4469	0.22129
13	0.6248	0.5152	0.4598	0.4249	0.45407
14	0.5987	0.4921	0.4386	0.4053	0.42877
15	0.5749	0.4711	0.4196	0.3876	0.36606

The commonly available tables for Cochran's distribution are inadequate for the analysis just given. The α level percentage points in Table 2.14 were obtained by evaluating the inverse of the cumulative distribution function of a $Beta(r/2, r(j-1))$ distribution at the point $1-\alpha/j$. These are easily obtained in R and Minitab. Cochran (1941) notes that these values are generally a good approximation to the true percentage points and that they are exact whenever the true percentage point is greater than 0.5. Moreover, the true significance level corresponding to a nominal significance level of α in Table 2.14 is at most α and at least $\alpha - \alpha^2/2$, so the true significance level associated with the 0.15 values listed in Table 2.14 is between 0.13875 and 0.15 for $j = 10, \ldots, 15$ and is exactly 0.15 for $j = 2, \ldots, 9$.

Most tests for equality of variances, including Cochran's test, are notoriously sensitive to nonnormality — so much so that they are rarely used in practice. However the analysis of variance F test is not noted for extreme sensitivity to nonnormality, even though it is a test for the equality of two variances. This is probably because the numerator mean square is computed from sample means and sample means tend to be reasonably normal. The current application of Cochran's test should benefit in the same way. The sums of squares in this example are essentially computed from the difference between two sample means each based on 8 observations. Thus the sensitivity to nonnormality should be mitigated. Of course for nonnormal data the sums of squares are unlikely to be independent, but the estimated effects are still uncorrelated.

The multiple maximum procedure is easily incorporated into χ^2 plots. Figure 2.6 contains the $\chi^2(1)$ plot for Hare's data along with an upper envelope. The upper envelope is the product of the Cochran 15% points and the partial sums from Ta-

ble 2.13. The 13th and 14th largest sums of squares exceed the upper envelope, indicating that the corresponding maximum tests are rejected.

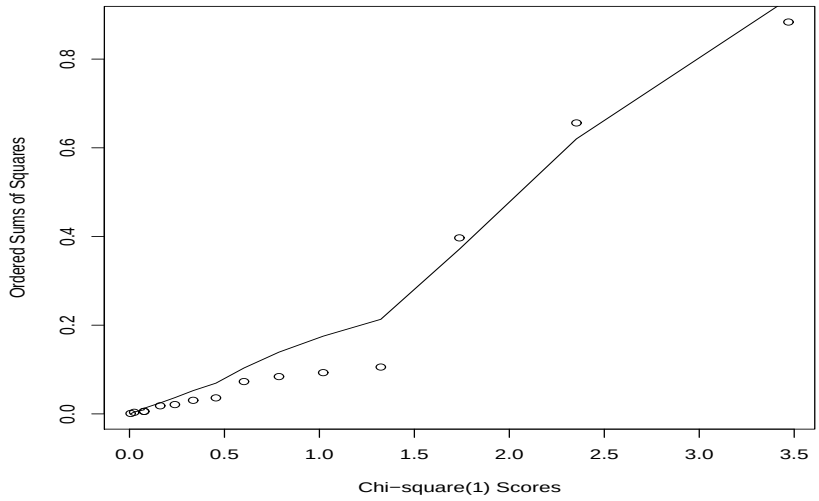

Fig. 2.6 $\chi^2(1)$ plot for Hare's data with 15% Cochran upper envelope.

2.4.2 Simulation envelopes

Figure 2.7 contains the $\chi^2(1)$ plot for Hare's data with a simulation envelope. Actually, the plot uses the standardized sums of squares, i.e., the sums of squares divided by the total sum of squares. Obviously, dividing each sum of squares by the same number has no effect on the visual interpretation of the plot. The simulation envelope is based on 99 analyses of randomly generated standard normal data. The upper envelope is the maximum for each order statistic from the 99 replications and the lower envelope is the minimum of the replications. Performing 99 analyses of a 2^4 experiment is time consuming; it is computationally more efficient just to take 99 random samples from a $\chi^2(1)$ distribution, standardize them, and order them. Unlike Atkinson's (1981) residual envelopes for regression, having divided all the sums of squares by the total, the same envelopes can be used for any subsequent analysis of 15 sums of squares each having one degree of freedom.

There are two prominent features in Figure 2.7. The 12th effect is barely below the lower envelope and the 14th effect is barely above the upper envelope. All of the sums of squares have been divided by the total, so the low value for the 12th effect

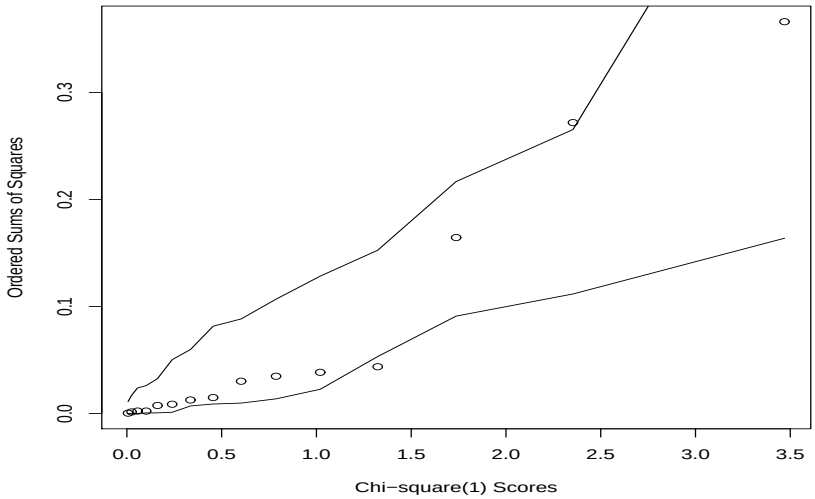

Fig. 2.7 $\chi^2(1)$ plot for Hare's data with simulated envelope.

indicates that the sum of squares for the 12th effect has been divided by a number that is too large. In other words, the sum of squares total is too large to be consistent with the 12 smallest sums of squares. This indicates that there must be significant effects among the 3 largest terms. The simulation envelope does not indicate which of the 3 larger terms are real effects; violation of the envelope only suggests that the envelope is inappropriate, i.e., that there are real effects. Visual interpretation of the graph must be used to identify the important effects.

2.4.3 Outliers

It is frequently suggested that outliers can be spotted from patterns in the plots. Consider a single outlier. Every effect contrast is the sum of 8 observations minus the sum of the other 8, thus if the outlier is not on the treatment $a_0b_0c_0d_0e_0$, the outlier is added into either 7 or 8 effect contrasts and is subtracted in the others. In a normal plot, such an outlier should cause a jump in the level of the line involving about half of the effects. In a χ^2 plot, a single outlier should cause all of the effects to look large, i.e., the intercept of the plot should not be zero. If two outliers exist in the data, they should cancel each other in about half of the effect contrasts and compound each other in the other effects. Thus, half of the effects should appear to be important in the plot. In my experience the natural variation in the plots is large enough that it is difficult to identify even very large outliers based on these facts.

2.5 Lenth's method

2.6 Augmenting designs for factors at two levels

When running a sequence of experiments, one may find that a particular 2^f experiment or fractional replication is inadequate to answer the relevant questions. In such cases it is often possible to add more points to the design to gain the necessary information. After running an initial fractional replication, a particular factor, say A, may be identified as being of primary importance. It is then of interest to estimate the main effect A and all two-factor interactions involving A without their being aliased with other two-factor interactions.

Box and Draper (1987, p. 156) suggest a method for adding points to the original design to achieve this end. Christensen and Huzurbazar (1996) also treat this method and provide a proof. Consider the resolution III, $1/16$th replication of a 2^7 experiment defined by $ABD(1)$, $ACE(1)$, $BCF(1)$, $ABCG(1)$. Box and Draper suggest augmenting the design with the $1/16$th replication $ABD(-1)$, $ACE(-1)$, $BCF(1)$, $ABCG(-1)$. The idea is that 1s have been changed to -1s in all defining contrasts that include A.

Together, the two $1/16$th replications define a $1/8$th replication. The nature of this $1/8$th replication can be explored by adding an imaginary factor H to the experiment. As H is imaginary, obviously it can have no effect on the responses. Any effects that involve H are simply error. For example, the DH interaction examines whether the effect for D changes from the low to high level of H. As there are no low and high levels of H, any observed change must be due to random error. The treatments included in the augmented $1/8$th rep. of the 2^7 given previously are identical to those obtained from the $1/16$th rep. of the 2^8 defined by $ABDH(1), ACEH(1), BCF(1), ABCGH(1)$. Here H has been added to any defining effect that includes A. In fact, if we want to consider the augmented $1/8$th rep. as occurring in two blocks, the imaginary factor H is the effect confounded with blocks.

Now consider the aliasing structure of this 2^8 design. In constructing any 2^f design in 16 blocks, there are $16-1$ effects confounded with the blocks, i.e., with the possible fractional replications. All of these effects are involved in determining the aliases of the $1/16$th rep. To get the complete set of 15 defining effects we must multiply the nominal defining effects together and multiply products of the nominal defining effects with other nominal defining effects. Using an asterisk $(*)$ to identify the nominal defining effects and performing the multiplications systematically, the complete set of defining effects is

$$ABDH^*, ACEH^*, BCDE, BCF^*, ACDFH, ABEFH, DEF, ABCGH^*,$$

$$CDG, BEG, AFGH, ADEGH, BDFG, CEFG, ABCDEFGH.$$

This is still a resolution III design but now every defining effect that includes A has at least three other factors, one of which is the imaginary H. Thus multiplying A times the defining effects we see that A is not aliased with any two-factor or main

effects. Moreover, a two-factor effect involving A, say AB, is only aliased with other two-factor effects that involve H. In the case of AB, the only two-factor effect that it is aliased with is DH. But H is imaginary, so two-factor effects involving A are not aliased with any real two-factor effects.

Box and Draper (1987) and Christensen and Huzurbazar (1996) provide a similar solution to the problem of augmenting a design to allow estimation of all main effects unaliased with two-factor effects. Again consider the resolution III, $1/16$th replication of a 2^7 experiment defined by $ABD(1), ACE(1), BCF(1), ABCG(1)$. For this problem they suggest adding the $1/16$th fraction defined by $ABD(-1), ACE(-1), BCF(-1), ABCG(1)$. Here 1s have been changed to -1s in the defining effects that involve an odd number of factors. This augmented $1/8$th rep. design is equivalent to adding an imaginary factor H and using the $1/16$th rep. of the 2^8 experiment defined by $ABDH(1), ACEH(1), BCFH(1), ABCG(1)$. In this approach, any defining effect with an odd number of terms has H added to it. As before, any effects that involve H are error. Adding H in this manner has changed the resolution III design into a resolution IV design. Thus all main effects are aliased with three-factor or higher terms. As before, if we view the augmented $1/8$th rep. as occurring in two blocks, H is the effect confounded with blocks.

2.7 Exercises

EXERCISE 2.7.1. Analyze Hare's s_c data that was given in Table 2.10.

EXERCISE 2.7.2. To consider the effect of a possible outlier, reanalyze Hare's s_p data, changing the largest value, the 2.10 in Table 2.10, into the second largest value, 1.85.

EXERCISE 2.7.3. Reanalyze Hare's s_c data after identifying and deleting the possible outlier. Does having an outlier in that particular batch suggest anything?

EXERCISE 2.7.4. Consider a 2^6 factorial. Give a good design for performing this in blocks of sixteen. Try to avoid confounding main effects and two-factor interactions with blocks.

EXERCISE 2.7.5. Consider a 2^6 factorial. Give a good design for performing a $1/4$ replication. Try to avoid aliasing main effects with each other and with two-factor interactions. Also try to avoid aliasing two-factor interactions with other two-factor interactions.

EXERCISE 2.7.6. Consider a 2^6 factorial. Give a good design for performing a $1/2$ replication in blocks of 16. Do not confound main effects or two-factor interactions with blocks. Try to avoid aliasing main effects with each other and with

two-factor interactions. Also try to avoid aliasing two-factor interactions with other two-factor interactions.

EXERCISE 2.7.7. Consider a $1/2$ rep. of a 2^5 with factors A, B, C, D, E and $ABCD$ defining the $1/2$ rep. Write down the treatment combinations in the $1/2$ rep. Now drop factor D from the analysis. In particular, write down all the treatment combinations in the $1/2$ rep. but delete all the d terms from the treatments. Does this list contain all the treatments in a 2^4 on factors A, B, C, E? Now return to the original $1/2$ rep., drop factor E, and list the treatment combinations. Does this list contain all the treatments in a 2^4 on factors A, B, C, D?

Chapter 3
p^f Factorial Treatment Structures

Except for the Taguchi material, I originally wrote this in the '90s but it appears for the first time.

In this chapter we examine extensions of the ideas used with 2^f factorials to situations where we have f factors each at p levels where p is a prime number. The general approach will obviously apply to 2^f factorials, but beyond that, the most important special case will be 3^f factorials. We also look briefly at 5^f factorials and some cross factorials such as a $2 \times 2 \times 3 \times 3$ involving two different prime numbers (2 and 3) and a 3×4 which involves a nonprime number of factor levels (4).

A key tool in this discussion is the use of modular arithmetic, e.g., $5 \bmod 3 = 2$ where 2 is the remainder when 5 is divided by 3. Similarly, $13 \bmod 5 = 3$, because 3 is the remainder when 5 is divided into 13. Modular arithmetic is applied to the subscripts defining the treatments. For a 3^2 experiment involving factor A with levels a_0, a_1, a_2 and factor B with levels b_0, b_1, b_2, the treatments are denoted a_0b_0, a_0b_1, a_0b_2, $a_1b_0, \ldots, a_2b_2$. In a 3^f, each effect is based on using modular arithmetic to divide the treatments into 3 groups. The analysis involves performing a one-way ANOVA on the three groups, thus every effect has 2 degrees of freedom. These two degree of freedom effects are then related to the usual main effects and interactions that are used in a two factor experiment. Similarly, in a 5^f we use modular arithmetic to create 5 groups.

Section 1 introduces the use of modular arithmetic in 3^f factorials to define 2 degree of freedom effects and relates them to the usual effects. Section 2 illustrates how the modular arithmetic determines a linear model that has the nice properties that we exploit in the analysis. In other words, it illustrates how it is that we can perform a valid analysis by simply looking at the one-way ANOVAs that are performed on each group defined by the modular arithmetic. Section 3 introduces the concepts of confounding and fractional replication for 3^f factorials. Section 4 introduces the work of Genichi Taguchi. Section 5 gives an example of the analysis of a 3^3 experiment that involves confounding. Section 6 illustrates how the concepts extend to 5^f factorials. Chapter 4 looks at mixtures of prime powers, e.g., a $2 \times 2 \times 3 \times 3$ factorial and powers of primes, e.g., an example that involves a factor with $4 = 2^2$ levels.

3.1 3^f Factorials

In a 2^f factorial, life is simple because every effect has only one degree of freedom. The situation is more complicated in 3^fs, 5^fs, and p^fs. We begin by considering the breakdown of the treatment sum of squares in a 3^f.

EXAMPLE 3.1.1. *A 3^4 experiment*
Consider a 3^4 experiment with factors A, B, C, D at levels a_0, a_1, a_2, b_0, b_1, b_2, c_0, c_1, c_2, and d_0, d_1, d_2, respectively. There are $3^4 = 81$ treatment combinations so there are 80 degrees of freedom for treatments. The usual breakdown of the treatments line in the ANOVA is given on the left of Table 3.1.

Table 3.1 Analysis of Variance Tables for 3^4

Traditional		Extended					
Source	*df*	Source	*df*	Source	*df*	Source	*df*
A	2	A	2	ABC	2	$ABCD$	2
B	2	B	2	ABC^2	2	$ABCD^2$	2
C	2	C	2	AB^2C	2	ABC^2D	2
D	2	D	2	AB^2C^2	2	ABC^2D^2	2
$A*B$	4	AB	2	ABD	2	AB^2CD	2
$A*C$	4	AB^2	2	ABD^2	2	AB^2CD^2	2
$A*D$	4	AC	2	AB^2D	2	AB^2C^2D	2
$B*C$	4	AC^2	2	AB^2D^2	2	$AB^2C^2D^2$	2
$B*D$	4	AD	2	ACD	2		
$C*D$	4	AD^2	2	ACD^2	2		
$A*B*C$	8	BC	2	AC^2D	2		
$A*B*D$	8	BC^2	2	AC^2D^2	2		
$A*C*D$	8	BD	2	BCD	2		
$B*C*D$	8	BD^2	2	BCD^2	2		
$A*B*C*D$	16	CD	2	BC^2D	2		
		CD^2	2	BC^2D^2	2		

With 2^f factorials, it is convenient to break everything into effects with $2-1=1$ degree of freedom. Each degree of freedom corresponds to a contrast or, equivalently, a one-way ANOVA between 2 groups. In 3^f factorials, it is convenient to break everything into $3-1=2$ degree of freedom effects. Such a breakdown is given on the right of Table 3.1. Every set of 2 degrees of freedom corresponds to performing a one-way analysis of variance among three groups. Later, we will use modular arithmetic to identify the three groups. In general, a p^f factorial is broken up into sets of $p-1$ degrees of freedom. Each set of $p-1$ degrees of freedom corresponds to a one-way ANOVA between p groups. Again, modular arithmetic is used to identify the p groups. The technique will be demonstrated on a 3^3.

EXAMPLE 3.1.2. 3^3 *Factorial*
Let the factors be A, B, and C with levels a_0, a_1, a_2, b_0, b_1, b_2, and c_0, c_1, c_2, respec-

tively. The usual breakdown of the treatment degrees of freedom is given on the left of Table 3.2. The sums of squares are found in the usual way. The extended breakdown is given on the right of Table 3.2. It is this breakdown that we will describe in detail. Each sum of squares is computed from a one-way analysis of variance on three groups. We focus on identifying the appropriate groups.

Table 3.2 Analysis of Variance Table for 3^3

Source	*df*	Source	*df*
A	2	A	2
B	2	B	2
C	2	C	2
$A*B$	4	AB	2
		AB^2	2
$A*C$	4	AC	2
		AC^2	2
$B*C$	4	BC	2
		BC^2	2
$A*B*C$	8	ABC	2
		ABC^2	2
		AB^2C	2
		AB^2C^2	2

As usual, the sum of squares for A is found by doing a one-way analysis of variance on three sample means, the mean of all observations that got treatment a_0, the mean of all observations that got treatment a_1, and the mean of all observations that got treatment a_2. There are three treatments so A has 2 degrees of freedom. The main effects for B and C work similarly.

The standard way of getting the sum of squares for the $A*B$ interaction involves doing a one-way analysis of variance on nine sample means, the means of all observations that got treatments a_0b_0, a_0b_1, a_0b_2, a_1b_0, a_1b_1, a_1b_2, a_2b_0, a_2b_1, and a_2b_2. This has 8 degrees of freedom. From this one-way ANOVA, the sum of squares for A and the sum of squares for B are subtracted, leaving the $A*B$ interaction sum of squares with $8-2-2=4$ degrees of freedom.

In the new analysis, the $A*B$ interaction term is broken into two interaction terms, one is called AB while the other is called AB^2. The AB term defines a group of three sample means. These three means have a one-way ANOVA performed on them yielding 2 degrees of freedom and a sum of squares. Similarly, the AB^2 term defines another group of three sample means for which a one-way is performed yielding 2 degrees of freedom and a sum of squares. By pooling together the degrees of freedom and sums of squares from the new AB and AB^2 terms, we can reconstruct the old $A*B$ term with 4 degrees of freedom.

The analysis of the new AB and AB^2 terms is straightforward, just two one-way ANOVAs. The trick is in specifying the observations that go into forming the three sample means. The specification involves the use of modular arithmetic applied to

the subscripts that identify the treatments. Specifically, in a 3^f factorial, the arithmetic is performed mod 3. In other words, arithmetic is performed in the usual way to get a number but the final answer is reported as the remainder when this number is divided by 3. For the AB interaction, the a and b subscripts are added together. After adjusting the numbers mod 3, we have three groups based on the sum of the subscripts. In other words, if we let x_1, x_2, and x_3 denote the subscripts of a treatment, the process involves finding $(1)x_1+(1)x_2+(0)x_3$ mod 3; the result is always either 0, 1 or 2. The three groups are determined by this value. The 0 group consists of all treatments that yield a 0. The 1 group has all treatments that yield a 1. The 2 group has all treatments that yield a 2. The process is illustrated in Table 3.3. The groups are also reported in Table 3.4. Without replication, each treatment yields one y value. The 9 y values for group 0 are averaged to find a sample mean for group 0. Similarly, sample means are computed for groups 1 and 2. The mean square for AB, say $MS(AB)$, is the sample variance of the three means times 9, the number of observations in each mean. The mean squared multiplied by $(3-1)$ gives $SS(AB)$, the sum of squares for AB.

Table 3.3 AB Interaction Groups

	subscripts	$AB=A^1B^1C^0$	Group
Treatment	(x_1,x_2,x_3)	$(1)x_1+(1)x_2+(0)x_3$	(mod 3)
$a_0b_0c_0$	$(0,0,0)$	0	0
$a_0b_0c_1$	$(0,0,1)$	0	0
$a_0b_0c_2$	$(0,0,2)$	0	0
$a_0b_1c_0$	$(0,1,0)$	1	1
$a_0b_1c_1$	$(0,1,1)$	1	1
$a_0b_1c_2$	$(0,1,2)$	1	1
$a_0b_2c_0$	$(0,2,0)$	2	2
$a_0b_2c_1$	$(0,2,1)$	2	2
$a_0b_2c_2$	$(0,2,2)$	2	2
$a_1b_0c_0$	$(1,0,0)$	1	1
$a_1b_0c_1$	$(1,0,1)$	1	1
$a_1b_0c_2$	$(1,0,2)$	1	1
$a_1b_1c_0$	$(1,1,0)$	2	2
$a_1b_1c_1$	$(1,1,1)$	2	2
$a_1b_1c_2$	$(1,1,2)$	2	2
$a_1b_2c_0$	$(1,2,0)$	3	0
$a_1b_2c_1$	$(1,2,1)$	3	0
$a_1b_2c_2$	$(1,2,2)$	3	0
$a_2b_0c_0$	$(2,0,0)$	2	2
$a_2b_0c_1$	$(2,0,1)$	2	2
$a_2b_0c_2$	$(2,0,2)$	2	2
$a_2b_1c_0$	$(2,1,0)$	3	0
$a_2b_1c_1$	$(2,1,1)$	3	0
$a_2b_1c_2$	$(2,1,2)$	3	0
$a_2b_2c_0$	$(2,2,0)$	4	1
$a_2b_2c_1$	$(2,2,1)$	4	1
$a_2b_2c_2$	$(2,2,2)$	4	1

Table 3.4 *AB* Groups

$(1)x_1+(1)x_2+(0)x_3$ mod 3 Groups		
0	1	2
$a_0b_0c_0$	$a_0b_1c_0$	$a_0b_2c_0$
$a_0b_0c_1$	$a_0b_1c_1$	$a_0b_2c_1$
$a_0b_0c_2$	$a_0b_1c_2$	$a_0b_2c_2$
$a_1b_2c_0$	$a_1b_0c_0$	$a_1b_1c_0$
$a_1b_2c_1$	$a_1b_0c_1$	$a_1b_1c_1$
$a_1b_2c_2$	$a_1b_0c_2$	$a_1b_1c_2$
$a_2b_1c_0$	$a_2b_2c_0$	$a_2b_0c_0$
$a_2b_1c_1$	$a_2b_2c_1$	$a_2b_0c_1$
$a_2b_1c_2$	$a_2b_2c_2$	$a_2b_0c_2$

The AB^2 groups are found by adding the a subscript to twice the b subscript. Computation of the groups for AB^2 is illustrated in Table 3.5. The groups are also given in Table 3.6. The computations for $MS(AB^2)$ and $SS(AB^2)$ are similar to the computations for $MS(AB)$ and $SS(AB)$. Recall that $SS(AB)$ and $SS(AB^2)$, each with 2 degrees of freedom, can be added to get $SS(A*B)$.

Confounding and fractional replication are based on the 2 degree of freedom effects. The analysis of data from experiments designed using these concepts requires the ability to compute the sums of squares for certain 2 degree of freedom effects.

Each effect determines the coefficients of the formula to be evaluated. The formula is then applied to the subscripts identifying each treatment to identify treatment groups. In general, the exponents of the effect determine the formula. The effect AB has $AB = A^1B^1C^0$ so the formula is $(1)x_1+(1)x_2+(0)x_3$ mod 3. The effect AB^2 has $AB^2 = A^1B^2C^0$ so the formula is $(1)x_1+(2)x_2+(0)x_3$ mod 3. The effect B has $B = A^0B^1C^0$ and thus the formula is $(0)x_1+(1)x_2+(0)x_3$ mod 3. The effect AB^2C has $AB^2C = A^1B^2C^1$ and formula $(1)x_1+(2)x_2+(1)x_3$ mod 3. All of the groups for the effects A, AB, AB^2, ABC, ABC^2, AB^2C, and AB^2C^2 are given in Table 3.7. Notice that the effects always have a first superscript of 1, e.g., AB^2C but never A^2B^2C. This has to do with redundancy in the formulas and will be discussed further in the subsection on Fractional Replication.

Table 3.5 AB^2 Interaction Groups

Treatment	subscripts (x_1,x_2,x_3)	$AB^2 = A^1B^2C^0$ $(1)x_1+(2)x_2+(0)x_3$	Group (mod 3)
$a_0b_0c_0$	$(0,0,0)$	0	0
$a_0b_0c_1$	$(0,0,1)$	0	0
$a_0b_0c_2$	$(0,0,2)$	0	0
$a_0b_1c_0$	$(0,1,0)$	2	2
$a_0b_1c_1$	$(0,1,1)$	2	2
$a_0b_1c_2$	$(0,1,2)$	2	2
$a_0b_2c_0$	$(0,2,0)$	4	1
$a_0b_2c_1$	$(0,2,1)$	4	1
$a_0b_2c_2$	$(0,2,2)$	4	1
$a_1b_0c_0$	$(1,0,0)$	1	1
$a_1b_0c_1$	$(1,0,1)$	1	1
$a_1b_0c_2$	$(1,0,2)$	1	1
$a_1b_1c_0$	$(1,1,0)$	3	0
$a_1b_1c_1$	$(1,1,1)$	3	0
$a_1b_1c_2$	$(1,1,2)$	3	0
$a_1b_2c_0$	$(1,2,0)$	5	2
$a_1b_2c_1$	$(1,2,1)$	5	2
$a_1b_2c_2$	$(1,2,2)$	5	2
$a_2b_0c_0$	$(2,0,0)$	2	2
$a_2b_0c_1$	$(2,0,1)$	2	2
$a_2b_0c_2$	$(2,0,2)$	2	2
$a_2b_1c_0$	$(2,1,0)$	4	1
$a_2b_1c_1$	$(2,1,1)$	4	1
$a_2b_1c_2$	$(2,1,2)$	4	1
$a_2b_2c_0$	$(2,2,0)$	6	0
$a_2b_2c_1$	$(2,2,1)$	6	0
$a_2b_2c_2$	$(2,2,2)$	6	0

Table 3.6 AB^2 Groups

$(1)x_1+(2)x_2+(0)x_3$ *mod* 3 Groups		
0	1	2
$a_0b_0c_0$	$a_0b_2c_0$	$a_0b_1c_0$
$a_0b_0c_1$	$a_0b_2c_1$	$a_0b_1c_1$
$a_0b_0c_2$	$a_0b_2c_2$	$a_0b_1c_2$
$a_1b_1c_0$	$a_1b_0c_0$	$a_1b_2c_0$
$a_1b_1c_1$	$a_1b_0c_1$	$a_1b_2c_1$
$a_1b_1c_2$	$a_1b_0c_2$	$a_1b_2c_2$
$a_2b_2c_0$	$a_2b_1c_0$	$a_2b_0c_0$
$a_2b_2c_1$	$a_2b_1c_1$	$a_2b_0c_1$
$a_2b_2c_2$	$a_2b_1c_2$	$a_2b_0c_2$

Table 3.7 Some Effect Groups for a 3^3 Factorial

Treatment	A	AB	AB^2	ABC	ABC^2	AB^2C	AB^2C^2
$a_0b_0c_0$	0	0	0	0	0	0	0
$a_0b_0c_1$	0	0	0	1	2	1	2
$a_0b_0c_2$	0	0	0	2	1	2	1
$a_0b_1c_0$	0	1	2	1	1	2	2
$a_0b_1c_1$	0	1	2	2	0	0	1
$a_0b_1c_2$	0	1	2	0	2	1	0
$a_0b_2c_0$	0	2	1	2	2	1	1
$a_0b_2c_1$	0	2	1	0	1	2	0
$a_0b_2c_2$	0	2	1	1	0	0	2
$a_1b_0c_0$	1	1	1	1	1	1	1
$a_1b_0c_1$	1	1	1	2	0	2	0
$a_1b_0c_2$	1	1	1	0	2	0	2
$a_1b_1c_0$	1	2	0	2	2	0	0
$a_1b_1c_1$	1	2	0	0	1	1	2
$a_1b_1c_2$	1	2	0	1	0	2	1
$a_1b_2c_0$	1	0	2	0	0	2	2
$a_1b_2c_1$	1	0	2	1	2	0	1
$a_1b_2c_2$	1	0	2	2	1	1	0
$a_2b_0c_0$	2	2	2	2	2	2	2
$a_2b_0c_1$	2	2	2	0	1	0	1
$a_2b_0c_2$	2	2	2	1	0	1	0
$a_2b_1c_0$	2	0	1	0	0	1	1
$a_2b_1c_1$	2	0	1	1	2	2	0
$a_2b_1c_2$	2	0	1	2	1	0	2
$a_2b_2c_0$	2	1	0	1	1	0	0
$a_2b_2c_1$	2	1	0	2	0	1	2
$a_2b_2c_2$	2	1	0	0	2	2	1

3.2 Column Space Considerations

Consider a 3^2 factorial. The ANOVA table can be broken up in two ways. The usual method with 4 degrees of freedom for interaction and the new method in which the interaction is divided into two 2 degree of freedom parts.

Source	df	Source	df
A	2	A	2
B	2	B	2
$A*B$	4	AB	2
		AB^2	2

Table 3.8 gives the treatment groups that are determined by the AB and AB^2 interaction groups. The groups from Table 3.8 can be rearranged into the form

		B		
	0	1	2
A	0	$AB(0)AB^2(0)$	$AB(1)AB^2(2)$	$AB(2)AB^2(1)$
	1	$AB(1)AB^2(1)$	$AB(2)AB^2(0)$	$AB(0)AB^2(2)$
	2	$AB(2)AB^2(2)$	$AB(0)AB^2(1)$	$AB(1)AB^2(0)$.

Note that this arrangement is really a Graeco-Latin square. Each of $AB(0)$, $AB(1)$, and $AB(2)$ appears in every row and in every column, the same is true for $AB^2(0)$, $AB^2(1)$, and $AB^2(2)$, moreover each of $AB(0)$, $AB(1)$, and $AB(2)$ appears exactly once with every AB^2 group.

Table 3.8 AB and AB^2 Interaction Groups

	subscripts	$AB = A^1B^1$	Group	$AB^2 = A^1B^2$	Group
Treatment	(x_1, x_2)	$(1)x_1 + (1)x_2$	(mod 3)	$(1)x_1 + (2)x_2$	(mod 3)
a_0b_0	$(0,0)$	0	0	0	0
a_0b_1	$(0,1)$	1	1	2	2
a_0b_2	$(0,2)$	2	2	4	1
a_1b_0	$(1,0)$	1	1	1	1
a_1b_1	$(1,1)$	2	2	3	0
a_1b_2	$(1,2)$	3	0	5	2
a_2b_0	$(2,0)$	2	2	2	2
a_2b_1	$(2,1)$	3	0	4	1
a_2b_2	$(2,2)$	4	1	6	0

The usual linear model for a 3×3 factorial with interaction but no replication is

$$\begin{bmatrix} y_{00} \\ y_{01} \\ y_{02} \\ y_{10} \\ y_{11} \\ y_{12} \\ y_{20} \\ y_{21} \\ y_{22} \end{bmatrix} = X \begin{bmatrix} \mu \\ \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \beta_0 \\ \beta_1 \\ \beta_2 \\ (\alpha * \beta)_{00} \\ (\alpha * \beta)_{01} \\ (\alpha * \beta)_{02} \\ (\alpha * \beta)_{10} \\ (\alpha * \beta)_{11} \\ (\alpha * \beta)_{12} \\ (\alpha * \beta)_{20} \\ (\alpha * \beta)_{21} \\ (\alpha * \beta)_{22} \end{bmatrix} + e$$

where

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Building a model based on a grand mean and effects for each of the A, B, AB, and AB^2 groupings gives

$$\begin{bmatrix} y_{00} \\ y_{01} \\ y_{02} \\ y_{10} \\ y_{11} \\ y_{12} \\ y_{20} \\ y_{21} \\ y_{22} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \beta_0 \\ \beta_1 \\ \beta_2 \\ (\alpha\beta)_0 \\ (\alpha\beta)_1 \\ (\alpha\beta)_2 \\ (\alpha\beta^2)_0 \\ (\alpha\beta^2)_1 \\ (\alpha\beta^2)_2 \end{bmatrix} + e$$

Note that the column space of the model matrix is the same in each model, i.e., $\mathbf{R}^9$. Since the columns corresponding to μ, the αs, and the βs are the same in each model, the interaction space — which is the space orthogonal to that spanned by μ, the αs, and βs — must be the same for each model. However, after adjusting for the grand mean, i.e. using Gram-Schmidt to orthogonalize with respect to the column of 1s, the columns corresponding to the $(\alpha\beta)$ terms are orthogonal to the columns for the α terms, the β terms, and also the $(\alpha\beta^2)$ terms. Similarly, the columns for the $(\alpha\beta^2)$ terms are orthogonal to all of the other sets of columns. To see this, observe that after adjusting everything for a column of 1s, the X matrix becomes

$$\begin{bmatrix}
1 & 2 & -1 & -1 & 2 & -1 & -1 & 2 & -1 & -1 & 2 & -1 & -1 \\
1 & 2 & -1 & -1 & -1 & 2 & -1 & -1 & 2 & -1 & -1 & -1 & 2 \\
1 & 2 & -1 & -1 & -1 & -1 & 2 & -1 & -1 & 2 & -1 & 2 & -1 \\
1 & -1 & 2 & -1 & 2 & -1 & -1 & -1 & 2 & -1 & -1 & 2 & -1 \\
1 & -1 & 2 & -1 & -1 & 2 & -1 & -1 & -1 & 2 & 2 & -1 & -1 \\
1 & -1 & 2 & -1 & -1 & -1 & 2 & 2 & -1 & -1 & -1 & -1 & 2 \\
1 & -1 & -1 & 2 & 2 & -1 & -1 & -1 & -1 & 2 & -1 & -1 & 2 \\
1 & -1 & -1 & 2 & -1 & 2 & -1 & 2 & -1 & -1 & -1 & 2 & -1 \\
1 & -1 & -1 & 2 & -1 & -1 & 2 & -1 & 2 & -1 & 2 & -1 & -1
\end{bmatrix}.$$

The last three columns corresponding to the $(\alpha\beta^2)$ terms are orthogonal to everything else, and columns 8, 9, 10 corresponding to the $(\alpha\beta)$ terms are orthogonal to everything else. Each of these sets of 3 columns must be in the interaction space because they are orthogonal to the columns for μ, the αs, and βs. Since they are also orthogonal to each other, they are breaking the interaction into 2 orthogonal pieces. Finally, the interaction columns were originally just group indicators, so the sums of squares for the one-way ANOVA on these groups will give the sum of squares appropriate for these sets of orthogonal columns.

3.3 Confounding and Fractional Replication

Having identified the groups associated with each 2 degree of freedom effect, we can now consider the issues of confounding and fractional replication. Confounding is a process for determining incomplete blocks, i.e., blocks that do not contain all of the treatments, but confounding retains a relatively simple analysis for the data. Fractional replication is a process for constructing experiments that do not contain all of the factorial treatments but retains an analysis of the data that allows one to identify important effects.

3.3.1 Confounding

We begin with a discussion of confounding. We can pick an effect, say, ABC to define blocks of size 9 in a 3^3 experiment. The effect ABC divides the treatments into three groups, the three groups define three blocks of 9 treatments. The groups were identified in Table 3.7 and the blocks are given in Table 3.9. Alternatively, we could use AB^2C^2 to define three different blocks of 9. These are given in Table 3.10.

Table 3.9 ABC Defining Three Blocks of Nine

Block 0	Block 1	Block 2
$a_0b_0c_0$	$a_0b_0c_1$	$a_0b_0c_2$
$a_0b_1c_2$	$a_0b_1c_0$	$a_0b_1c_1$
$a_0b_2c_1$	$a_0b_2c_2$	$a_0b_2c_0$
$a_1b_0c_2$	$a_1b_0c_0$	$a_1b_0c_1$
$a_1b_1c_1$	$a_1b_1c_2$	$a_1b_1c_0$
$a_1b_2c_0$	$a_1b_2c_1$	$a_1b_2c_2$
$a_2b_0c_1$	$a_2b_0c_2$	$a_2b_0c_0$
$a_2b_1c_0$	$a_2b_1c_1$	$a_2b_1c_2$
$a_2b_2c_2$	$a_2b_2c_0$	$a_2b_2c_1$

With ABC defining blocks, two degrees of freedom are confounded with the three blocks but the other 6 degrees of freedom involving $A*B*C$ interactions are available in ABC^2, AB^2C, and AB^2C^2. It is at this point, and not before, that the analysis requires computation of sums of squares for individual 2 degree of freedom interaction effects. (Although a good linear model program that fits blocks before interactions should give the correct 6 degree of freedom interaction term.)

Table 3.10 AB^2C^2 Defining Three Blocks of Nine

Block $AB^2C^2(0)$	Block $AB^2C^2(1)$	Block $AB^2C^2(2)$
$a_0b_0c_0$	$a_0b_0c_2$	$a_0b_0c_1$
$a_0b_1c_2$	$a_0b_1c_1$	$a_0b_1c_0$
$a_0b_2c_1$	$a_0b_2c_0$	$a_0b_2c_2$
$a_1b_0c_1$	$a_1b_0c_0$	$a_1b_0c_2$
$a_1b_1c_0$	$a_1b_1c_2$	$a_1b_1c_1$
$a_1b_2c_2$	$a_1b_2c_1$	$a_1b_2c_0$
$a_2b_0c_2$	$a_2b_0c_1$	$a_2b_0c_0$
$a_2b_1c_1$	$a_2b_1c_0$	$a_2b_1c_2$
$a_2b_2c_0$	$a_2b_2c_2$	$a_2b_2c_1$

We can also use two effects to create 9 blocks of size three. If we use ABC and AB^2C^2 to define blocks, one group consists of the three treatments in the $ABC(0)$

group that are also in the $AB^2C^2(0)$ group. Another group of three has both $ABC(0)$ and $AB^2C^2(1)$. The 9 blocks of three run through $ABC(i)$ and $AB^2C^2(j)$ for $i, j = 0, 1, 2$. Table 3.11 gives the 9 blocks defined by ABC and AB^2C^2.

Table 3.11 ABC, AB^2C^2 Defining Nine Blocks of Three

	$ABC(0)$	$ABC(1)$	$ABC(2)$
$AB^2C^2(0)$	$a_0b_0c_0$	$a_2b_0c_2$	$a_1b_0c_1$
	$a_0b_1c_2$	$a_2b_1c_1$	$a_1b_1c_0$
	$a_0b_2c_1$	$a_2b_2c_0$	$a_1b_2c_2$
$AB^2C^2(1)$	$a_2b_0c_1$	$a_1b_0c_0$	$a_0b_0c_2$
	$a_2b_1c_0$	$a_1b_1c_2$	$a_0b_1c_1$
	$a_2b_2c_2$	$a_1b_2c_1$	$a_0b_2c_0$
$AB^2C^2(2)$	$a_1b_0c_2$	$a_0b_0c_1$	$a_2b_0c_0$
	$a_1b_1c_1$	$a_0b_1c_0$	$a_2b_1c_2$
	$a_1b_2c_0$	$a_0b_2c_2$	$a_2b_2c_1$

As in 2^f factorials, when more than one effect is used to define groups (blocks), other effects also determine the same groups (are confounded with blocks). With 9 blocks, there are 8 degrees of freedom for blocks. The defining effects, ABC and AB^2C^2, are confounded with blocks but these account for only 4 degrees of freedom. There are 4 additional degrees of freedom for blocks and since each effect has 2 degrees of freedom, two additional effects are implicitly confounded with blocks. Again as in 2^f factorials, the implicitly confounded effects can be identified using modular multiplication, however, the multiplication works differently. Exponents are now evaluated modulo 3. Multiplying the two defining effects gives

$$ABC \times AB^2C^2 = A^2B^3C^3 = A^2B^0C^0 = A^2.$$

We never allow the first superscript in an effect to be 2. To eliminate such terms in a 3^f experiment, raise the term to the $3-1$ power,

$$A^2 = (A^2)^2 = A^4 = A.$$

(More on this issue of the first superscript later.) Thus when confounding both ABC and AB^2C^2, the main effect A is also confounded. We can see this directly from Table 3.11, where the level of A is constant in every block, i.e., the a subscript is the same in every block. One more effect is implicitly confounded with blocks; it can be identified by multiplying one defining effect times the square of the other effect, i.e.,

$$ABC \times (AB^2C^2)^2 = A^3B^5C^5 = B^2C^2 = (B^2C^2)^2 = B^4C^4 = BC.$$

It is irrelevant which effect is squared, the result is the same either way.

Performing a 3^f experiment in 27 blocks of size 3^{f-3} requires the use of three defining effects. These are effects such as $ABCD$, ABC^2D, and AB^2CD^2, but we will refer to them generically as α, β, and γ. With 27 blocks there are 26 degrees of

freedom for blocks. Each effect has 2 degrees of freedom, so there are 13 effects confounded with blocks. Three of the effects are the defining effects α, β, and γ. The other ten effects can be obtained through multiplying pairs of factors,

$$\alpha(\beta),\ \alpha(\beta)^2,\ \alpha(\gamma),\ \alpha(\gamma)^2,\ \beta(\gamma),\ \beta(\gamma)^2$$

and multiplying all three factors,

$$\alpha\beta\gamma,\ \alpha\beta\gamma^2,\ \alpha\beta^2\gamma,\ \alpha\beta^2\gamma^2.$$

Just to illustrate the method of finding the effects that are confounded, suppose AB, CD and EF are the defining effects. The effects that are implicitly confounded are $ABCD$, $AB(CD)^2$, $ABEF$, $AB(EF)^2$, $CDEF$, $CD(EF)^2$, $ABCDEF$, $ABCD(EF)^2$, $AB(CD)^2EF$, $AB(CDEF)^2$.

3.3.2 Fractional Replication

Any one of the three blocks in Table 3.9 defines a $1/3$ rep. of the 3^3 factorial based on ABC. Similarly, any one of the three blocks in Table 3.10 defines a $1/3$ rep. of the 3^3 factorial based on AB^2C^2. Similarly, any one of the nine blocks in Table 3.11 defines a $1/9$ rep. of the 3^3 factorial based on ABC and AB^2C^2.

In fractional replication, many effects will be lost. In a $1/2$ rep. of a 2^f factorial, about $1/2$ of the effects are lost to aliasing. In a $1/3$ rep. of a 3^f factorial, about $2/3$s of the effects are lost to aliasing. It is vital to identify the aliasing structure in fractional replications. To do this, we need to delve a bit deeper into modular arithmetic.

Recall that the first term in all of our effects always has a power of 1, e.g., AB^2C^2 but never $A^2B^2C^2$, also BC^2 but never B^2C. The first term always has a power of 1 because having any larger power is redundant. The groups defined by an effect, say, AB^2C^2 are determined by the value of

$$z = [(1)x_1 + (2)x_2 + (2)x_3] \mod 3$$

where (x_1, x_2, x_3) are the subscripts of the treatments (a, b, c). If we double the coefficients in the equation we get

$$\begin{aligned} z &= [(2)x_1 + (4)x_2 + (4)x_3] \mod 3 \\ &= [(2)x_1 + (1)x_2 + (1)x_3] \mod 3. \end{aligned}$$

Here we have also adjusted the coefficients using modular arithmetic. For example, it is easy to see that with an integer x, $4x \bmod 3 = x \bmod 3$. The groups defined by this second equation correspond to

$$A^2B^4C^4 = A^2B^1C^1 = A^2BC,$$

where we are again using modular multiplication. The point is that the two sets of equations for AB^2C^2 and A^2BC give identical groups. Every treatment that is in group 0 for AB^2C^2 is also in group 0 for A^2BC. Every treatment that is in group 1 for AB^2C^2 is in group 2 for A^2BC and every treatment that is in group 2 for AB^2C^2 is in group 1 for A^2BC. Thus,

$$AB^2C^2 \equiv A^2BC.$$

Generally, if we take any power of an effect where the power is a positive integer less than $p = 3$, we still have the same effect because the new modular equation gives the same groups. To have a unique way of writing each effect, we never allow the first superscript in an effect to be greater than 1. When given an effect with a lead term having an exponent of 2, squaring the term reduces the exponent without changing the effect, e.g.,

$$A^2BC = \left(A^2BC\right)^2 = A^4B^2C^2 = AB^2C^2.$$

Modular multiplication and the equivalences between effects are crucial to identifying aliases. In a $1/3$ rep. based on ABC, the aliases of an effect are determined from multiplying the effect by ABC and $(ABC)^2$. For example, the main effect A is aliased with two other effects

$$A \times ABC = A^2BC = \left(A^2BC\right)^2 = A^4B^2C^2 = AB^2C^2$$

and

$$A \times (ABC)^2 = A^3B^2C^2 = B^2C^2 = \left(B^2C^2\right)^2 = B^4C^4 = BC.$$

The interaction effect BC^2 is aliased with two effects

$$BC^2 \times ABC = AB^2C^3 = AB^2$$

and

$$BC^2 \times (ABC)^2 = A^2B^3C^4 = A^2C = \left(A^2C\right)^2 = A^4C^2 = AC^2.$$

Tables 3.12 and 3.13 give all the multiplications for determining the aliasing structure for the $1/3$ rep. based on ABC. The aliasing structure of this example can be determined from *either* Table 3.12 *or* 3.13 by simply listing all of the equalities. The aliasing structure reduces to

$$A = AB^2C^2 = BC,$$

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

$$C = ABC^2 = AB,$$

and

$$AB^2 = AC^2 = BC^2$$

with the defining effect ABC lost (confounded with the grand mean).

Table 3.12 First Aliasing Table for a $1/3$ Rep. of a 3^3 Based on ABC

Source	$\times ABC$	
A	$= A^2BC$	$= AB^2C^2$
B	$= AB^2C$	$= AB^2C$
C	$= ABC^2$	$= ABC^2$
AB	$= A^2B^2C$	$= ABC^2$
AB^2	$= A^2B^3C$	$= AC^2$
AC	$= A^2BC^2$	$= AB^2C$
AC^2	$= A^2BC^3$	$= AB^2$
BC	$= AB^2C^2$	$= AB^2C^2$
BC^2	$= AB^2C^3$	$= AB^2$
ABC	$= A^2B^2C^2$	$= ABC^*$
ABC^2	$= A^2B^2C^3$	$= AB$
AB^2C	$= A^2B^3C^2$	$= AC$
AB^2C^2	$= A^2B^3C^3$	$= A$

Table 3.13 Second Aliasing Table for a $1/3$ Rep. of a 3^3 Based on ABC

Source	$\times A^2B^2C^2$	
A	$= A^3B^2C^2$	$= BC$
B	$= A^2B^3C^2$	$= AC$
C	$= A^2B^2C^3$	$= AB$
AB	$= A^3B^3C^2$	$= C$
AB^2	$= A^3B^4C^2$	$= BC^2$
AC	$= A^3B^2C^3$	$= B$
AC^2	$= A^3B^2C^4$	$= BC^2$
BC	$= A^2B^3C^3$	$= A$
BC^2	$= A^2B^3C^4$	$= AC^2$
ABC	$= A^3B^3C^3$	$=$ —
ABC^2	$= A^3B^3C^4$	$= C$
AB^2C	$= A^3B^4C^3$	$= B$
AB^2C^2	$= A^3B^4C^4$	$= BC$

A $1/9$ rep. of a 3^3 can be obtained by using just one of the 9 blocks given in Table 3.11. Typically one would not perform a $1/9$th rep. of something as small as a 3^3, but we will discuss the issues involved to illustrate the principles. They extend easily to general 3^f systems. Table 3.11 uses the defining effects ABC and AB^2C^2. As with 2^f systems, there are additional effects lost when more than one defining effect is used. The effects

$$ABC \times AB^2C^2 = A^2B^3C^3 = A^2 = A$$

and

$$ABC \times \left(AB^2C^2\right)^2 = A^3B^5C^5 = B^2C^2 = BC.$$

are also (implicit) defining effects. Note that squaring both terms in the multiplications just leads to redundancies, e.g.,

$$(ABC)^2 \times AB^2C^2 = BC.$$

To find the aliases for an effect, say B, we need to multiply the effect by all four of the defining effects and by the squares of all the defining effects. Thus B has the following aliases,

$$\begin{aligned} B \times ABC &= AB^2C \\ B \times (ABC)^2 &= AC \\ B \times AB^2C^2 &= AC^2 \\ B \times \left(AB^2C^2\right)^2 &= ABC^2 \\ B \times A &= AB \\ B \times A^2 &= AB^2 \\ B \times BC &= BC^2 \\ B \times B^2C^2 &= C \end{aligned}$$

or

$$B = C = AB = AB^2 = AC = AC^2 = BC^2 = ABC^2 = AB^2C.$$

In this example, all of the effects that are not used in defining the $1/9$th rep. are aliased with each other. In a $1/9$th rep. of a 3^3, only three treatments are used so there are only 2 degrees of freedom for treatments. Every effect has two degrees of freedom so there is only one effect available; it can be called anything except one of the effects that define the fractional rep.

EXAMPLE 3.3.1. *A 3^5 Factorial*
With only 27 treatments in a 3^3, it would not be of much value to consider a $1/9$th rep. However a 3^5 has 243 treatments, so a $1/9$th rep. reduces the problem to a more manageable 27 treatments. With factors A, B, C, D, E, we might use ABC^2D and BC^2DE^2 as defining effects. The groups are defined using the ABC^2D equation

$$z_1 = [x_1 + x_2 + 2x_3 + x_4 + 0x_5] \text{ mod } 3$$

and the BC^2DE^2 equation

$$z_2 = [0x_1 + x_2 + 2x_3 + x_4 + 2x_5] \text{ mod } 3$$

where $(x_1, x_2, x_3, x_4, x_5)$ denotes the subscripts of a treatment combination and the coefficients in the equations are determined by the superscripts in the defining effects. The treatments that have, say, $z_1 = 1$ are referred to as the $ABC^2D(1)$ group with similar definitions for $z_1 = 0, 2$ and groups $ABC^2D(0)$ and $ABC^2D(2)$. BC^2DE^2 groups are denoted in the same way based on the values of z_2. Thus $a_0b_0c_0d_0e_0$ is in the group defined by $ABC^2D(0)$ and $BC^2DE^2(0)$, while $a_0b_1c_2d_1e_1$ is in the group

defined by $ABC^2D(0)$ and $BC^2DE^2(2)$, and $a_1b_1c_1d_1e_1$ is in the group defined by $ABC^2D(2)$ and $BC^2DE^2(0)$.

The two defining effects ABC^2D and BC^2DE^2 determine 9 blocks of 27 treatments. Any one of the blocks can be used as a $1/9$ rep. In a $1/9$ rep., other effects are implicit defining effects, namely

$$BC^2DE^2 \times ABC^2D = AB^2C^2D^2E$$

and

$$BC^2DE^2 \times \left(ABC^2D\right)^2 = A^2B^3C^6D^3E^2 = A^2E^2 = AE.$$

An effect, say B, has the following aliases,

$$\begin{aligned} B \times ABC^2D &= AB^2C^2D \\ B \times \left(ABC^2D\right)^2 &= AC^2D \\ B \times BC^2DE^2 &= BCD^2E \\ B \times \left(BC^2DE^2\right)^2 &= CD^2E \\ B \times AB^2C^2D^2E &= AC^2D^2E \\ B \times \left(AB^2C^2D^2E\right)^2 &= ABC^2D^2E \\ B \times AE &= ABE \\ B \times A^2E^2 &= AB^2E \end{aligned}$$

or

$$B = ABE = AB^2E = AC^2D = CD^2E = AB^2C^2D = AC^2D^2E = BCD^2E = ABC^2D^2E.$$

3.4 Taguchi's Orthogonal Arrays

Byrne and Taguchi (1989) and Lucas (1994) considered an experiment on the force y, measured in pounds, needed to pull tubing from a connector. Large values of y are good. The *controllable factors* in the experiment are as follows: A — Interference (Low, Medium, High), B — Wall Thickness (Thin, Medium, Thick), C — Ins. Depth (Shallow, Medium, Deep), D — Percent Adhesive (Low, Medium, High). The stated factor levels are all ordered but not numerical. The data are given in Table 3.14.

The design is a standard one proposed by Genichi Taguchi that he called L_9. The 9 is because it involves 9 treatment combinations. The design is a fractional factorial involving four factors each at three levels. Specifically, it is a 1/9th rep of a $3 \times 3 \times 3 \times 3$ design. With $3^4 = 81$ treatment combinations and only 9 observations, it must be a 1/9th rep. A p^{-r} fractional replication of a p^f factorial structure is sometimes referred to as a p^{f-r} design, so this is a 3^{4-2}. We will see that the design allows one to estimate all of the main effects. It also *only allows* estimation of the

Table 3.14 Taguchi $3 \times 3 \times 3 \times 3$ Fractional Design

Run	a	b	c	d	Observations							
1	0	0	0	0	15.6	9.5	16.9	19.9	19.6	19.6	20.0	19.1
4	0	1	1	1	15.0	16.2	19.4	19.6	19.7	19.8	24.2	21.9
7	0	2	2	2	16.3	16.7	19.1	15.6	22.6	18.2	23.3	20.4
3	1	0	1	2	18.3	17.4	18.9	18.6	21.0	18.9	23.2	24.7
2	1	1	2	0	19.7	18.6	19.4	25.1	25.6	21.4	27.5	25.3
5	1	2	0	1	16.2	16.3	20.0	19.8	14.7	19.6	22.5	24.7
8	2	0	2	1	16.4	19.1	18.4	23.6	16.8	18.6	24.3	21.6
6	2	1	0	2	14.2	15.6	15.1	16.8	17.8	19.6	23.2	24.4
9	2	2	1	0	16.1	19.9	19.3	17.3	23.1	22.7	22.6	28.6

main effects and assumes the absence of any interactions. This is characteristic of the designs that Taguchi typically suggested for controllable factors.

We begin by identifying the defining effects. In general a p^{f-r} design requires r (unaliased) explicit defining effects to divide the original p^f factor combinations into p^r groups each containing p^{f-r} factor combinations. The treatments in this design satisfy the two subscript constraints,

$$x_2 + x_3 + x_4 \bmod 3 = 0$$

and

$$x_1 + x_2 + 2x_3 \bmod 3 = 0.$$

In other words, the defining effects can be taken as BCD and ABC^2. Implicit defining effects are then $BCD \times ABC^2 = AB^2C^3D = AB^2D$ and $BCD \times (ABC^2)^2 = A^2B^3C^5D = A^2C^2D = (A^2C^2D)^2 = A^4C^4D^2 = ACD^2$. To double check this, the corresponding modular subscript equations are

$$x_1 + 2x_2 + x_4 \bmod 3 = 0$$

and

$$x_1 + x_3 + 2x_4 \bmod 3 = 0.$$

These are both satisfied by the treatments employed. Having identified all of the defining effects, we see that this is a resolution III design, so no main effects are aliased with each other. The main effects account for 8 degrees of freedom, so with 9 observations, no other treatment effects can be examined.

To check the aliasing more specifically,

$$A(ABC^2) = A^2BC^2 = AB^2C,$$

$$A(ABC^2)^2 = B^2C = BC^2,$$

$$A(BCD) = ABCD,$$

$$A(BCD)^2 = AB^2C^2D^2,$$
$$A(AB^2D) = A^2B^2D = ABD^2,$$
$$A(AB^2D)^2 = B^4D^2 = BD^2,$$
$$A(ACD^2) = A^2CD^2 = AC^2D,$$
$$A(ACD^2)^2 = C^2D = CD^2,$$

so A is not confounded with other main effects. Similar patterns hold for for B, C, and D.

3.4.1 Listed Designs

Taguchi apparently liked this particular $1/9$th rep, but I don't see how it could be better than a huge number of other choices that are resolution III designs. In particular, it cannot be better than the other eight $1/9$th replications that are determined by the same defining effects.

Taguchi provided L_q designs for $q = 4, 8, 12, 16, 18, 25, 27, 32, 36, 50, 54, 64, 81$. In other words, for $q = 2^2, 2^3, 2^2\times 3, 2^4, 2\times 3^2, 5^2, 3^3, 2^5, 2^2\times 3^2, 2\times 5^2, 2\times 3^3, 2^6, 3^4$. Our current theory handles all of these except for those involving products of powers of different prime numbers, which are examined in Chapter 4. (5^f structures are illustrated in Section 6 but the theory is the same as for 3^f.)

Taguchi discusses how, for example, his L_{18} and L_{27} structures provide resolution III designs for $2\times 3^{7-5}$ and 3^{13-10} treatment structures. These require 5 and 10 explicit defining effects, respectively. When using a large number of explicit defining effects, there are even more implicit defining effects. With so many defining effects, it can be difficult to achieve a resolution III design. Construction of L_{18} is considered in the next chapter. We now address finding competitors to L_{27}.

For a 3^{13-10}, the factors are A through M. In general, to get a 3^{13-10}, we need 10 defining effects. With $10 = \binom{5}{2}$, one way to proceed is by partitioning the factors into 5 sets, say, ABC, DEF, GHI, JK, LM. In an effort to get a resolution III design, we want these all to involve 3 factors, so change the sets to ABC, DEF, GHI, AJK, DLM. We create defining effects by combining pairs of these, i.e., $ABCDEF$, $ABCGHI$, A^2BCJK, $ABCDLM$, $DEFGHI$, $ADEFJK$, D^2EFLM, $AGHIJK$, $DGHILM$, $ADJKLM$. Because the defining effects all contain at least 5 distinct factors and every pair of defining effects has at least 3 distinct entries, the implicit defining effects should all include at least three factors, which means that no main effects are aliased with other main effects.

Think of this process as involving $A\tilde{B} \equiv ABC$, $D\tilde{E} \equiv DEF$, $\tilde{G} \equiv GHI$, $A\tilde{J} \equiv AJK$, $D\tilde{L} \equiv DLM$ and defining effects $A\tilde{B}D\tilde{E}$, $A\tilde{B}\tilde{G}$, $A^2\tilde{B}\tilde{J}$, $A\tilde{B}D\tilde{L}$, $D\tilde{E}\tilde{G}$, $AD\tilde{E}\tilde{J}$, $D^2\tilde{E}\tilde{L}$, $A\tilde{G}\tilde{J}$, $AD\tilde{G}\tilde{J}$, $D\tilde{G}\tilde{L}$, $AD\tilde{J}D\tilde{L}$. I think you can see that products will always give at least a three factor interaction.

3.4.2 Inner and Outer Arrays

The four factors associated with Table 3.4.14 are what Taguchi called *control factors* because they are (should be) relatively easy for the experimenter to control. It turns out that the 8 observations presented for every treatment combination are not true replications. They are the result of a 2^3 design on three *noise factors* that can be controlled only with exceptional effort. In reality this is a 1/9th rep of a $2^3 \times 3^4$ experiment. The traditional analysis of such designs is examined in the next chapter however Taguchi's proposed data analysis ignores the 2^3 structure and treats the 8 observations on the noise factors as a random sample to be summarized before being analyzed. Taguchi called the design for the control factors the *inner array* and the design for the noise factors the *outer array*. In our example, the inner array is the L_9 and the outer array is L_8.

The summarization of the outer array data has been one source of criticism for Taguchi's methods. Since the noise factors are beyond practical control, the data generated for them are treated as replications rather than as observations on factors. (In essence, Taguchi treats them as a stratified random sample with one observation per strata and equal weights on each strata — ignoring within stratum variability). My opinion is that this is a sensible approach if the levels of the noise factors are well chosen to represent the full spectrum of noise levels encountered in practice. The sample variance computed over the noise factors should supply a valuable measure of how bad the variability could be during standard operations and, similarly, the mean is a not inappropriate summary measure of central tendency.

3.4.2.1 Split Plot Designs

While it is possible to run an inner-outer array experiment in any sort of design, *the very nature of inner and outer arrays suggests that they would most often be run as split plot experiments!*

If the outer array noise factors are the most difficult to control, it seems likely that one would (randomly pick a set of noise factors from the outer array and) do what is necessary to control them and, while controlled at that level, run all of the inner array treatment combinations (in random order). That makes the outer array into whole plot treatments and the inner array into subplot treatments. The analysis of such a data collection scheme will be illustrated in the next chapter because in the Byrne-Taguchi example, the inner array is a 3^{4-2} but the outer array is a 2^3, so the design involves a mixture of prime powers as discussed in the next chapter.

The other possibility is to take the inner array as the whole plot treatments and the outer array as subplot treatments. That involves fixing an inner-array factor combination and manipulating the noise factors so that all of the outer array is observed before moving on to the next inner array combination. *Treating the inner array as whole plot treatments leads to the traditional Taguchi analysis that is illustrated in this section.* As discussed by Christensen (1996, 2015), in split plot models the whole plot treatments (here the inner array) can be analyzed by averaging over the

subplots (outer array). This implicitly ignores any interactions between the inner array factors and the outer array factors, which may be reasonable to do when the outer array contains only noise factors.

Another advantage to making the inner array the whole plot treatments is that variability associated with changing the inner array factor combinations is not particularly relevant to the eventual application of the data. The ultimate idea is to settle on a particular inner array combination and to use that in the future. The relevant variability for future use is the variability associated with the uncontrollability of noise factors, i.e., the variance within each whole plot.

We will see in the next subsection that if the observations are y_{io}, where i indicates an inner array factor combination and o indicates an outer array factor combination, the Taguchi analysis is performed by averaging (over o) one or more of the dependent variable values y_{io}, y_{io}^2, $1/y_{io}^2$, $(y_{io}-\bar{y}_{i\cdot})^2$, or more often transforming these averages. Note that the mean of $(y_{io}-\bar{y}_{i\cdot})^2$ is the method of moments estimator for the variance within whole plots.

Our description of the Byrne-Taguchi data does not specify how the data were collected, so we will illustrate both split plot analyses on these data. (The inclusion of the variable "Run" actually hints that the inner array may constitute the whole plot treatments making the analysis given in this section the correct one.) The analysis for treating the outer array as whole plot treatments is reserved for the next chapter because of the specific form of this outer array.

3.4.3 Signal-to-Noise Ratios

Taguchi's primary emphasis was that processes should be designed to minimize variability while staying on target. He operationalized this idea through the analysis of *signal-to-noise ratios*.

Although the eight observations for each (inner array) factor combination in Table 3.14 were obtained in a systematic fashion, in the Taguchi analysis those outer array observations are treated as a random sample, even though they clearly are not. In the analysis, the multiple observations on each inner array factor combination are summarized in some way prior to analysis. The obvious summarization is the sample mean. Additional obvious summary measures are the sample variance, sample standard deviation, or the logarithms of those values.

One of Taguchi's most controversial ideas was to summarize the outer array data using "signal-to-noise ratios." The idea is to maximize the appropriate signal-to-noise ratio. Again let the observations be y_{io} with i — inner array and o — outer array, $o = 1,\ldots,N$. For minimizing a response y his signal-to-noise ratio is defined as

$$SN_{\min,i} \equiv -\log\left(\sum_{o=1}^{N} y_{io}^2/N\right) = -\log\left(\bar{y}_{i\cdot}^2 + \frac{N-1}{N}s_i^2\right).$$

Because of the minus sign, to make this large you need both $\bar{y}_{i\cdot}$ and s_i^2 to be small but of course there are many other functions of $\bar{y}_{i\cdot}$ and s_i^2 that could accomplish the same thing. For maximizing a response his signal-to-noise ratio is

$$SN_{\max,i} \equiv -\log\left(\sum_{o=1}^{N} 1/y_{io}^2 N\right),$$

apparently because maximizing y is the same as minimizing $1/y$ (for positive y). For minimizing variability around a target his signal-to-noise ratio is

$$SN_{\text{tar},i} \equiv \log\left(\frac{\bar{y}_{i\cdot}^2}{s_i^2}\right) = \log(\bar{y}_{i\cdot}^2) - \log(s_i^2).$$

The apparent rationale is that if $\bar{y}_{i\cdot}$ always remains close to the target, you just want to minimize s_i^2. An apparently common approach is to divide the control factors into two groups. First identify control factors that affect the signal-to-noise ratio and use them to maximize it. Then use the control factors that do not have much affect on the SN ratio to try to put the process on target.

3.4.4 Taguchi Analysis

Table 3.15 contains three summary statistics to be used in a Taguchi analysis of the Byrne-Taguchi data. We fit a main-effects model to each of $\bar{y}_{i\cdot}$, $\log(s_i)$, and $SN_{\max,i}$. For each dependent variable we constructed two χ^2 plots. The first is a $\chi^2(2)$ plot for the main-effect sums of squares. The second plot is a $\chi^2(1)$ plot based on treating the factor subscripts (associated with ordered levels) as regression variables and fitting quadratic polynomials in the main effects. (Quadratic effects really just measure nonlinearity.) This gives sums of squares for a linear (e.g. a) and quadratic (e.g. aa) contrast in each main effect. Figures 3.1, 3.2, and 3.3 contain the χ^2 plots for the different dependent variables.

I don't see any clear evidence for the existence of main effects in either the mean or the log-standard-deviation plot. But I can imagine someone else arguing that all or nearly all of the effects are important. For the signal-to-noise ratio I again see no *clear* evidence of effects but some evidence for one or possibly two contrasts having effects. The two largest sums of squares are for the linear effect in C and the quadratic effect in A.

To make sense of any important effects we would look at means plots. These are given in Figures 3.4, 3.5, and 3.6. We will not discuss the means plots for the means or log standard deviations of the outer array data because they displayed no obvious main effects. For the signal-to-noise ratio the two largest $\chi^2(1)$ values were for the curvature in A and the linear effect in C. Since interest is in maximizing the signal-to-noise ratio, the recommendation would be to pick the middle level of A and, despite the importance of the linear effect in C (which really only looks at the

Table 3.15 Bryne-Taguchi 3^{4-2} Design

Run					Outer Array Summaries		
i	a	b	c	d	$\bar{y}_{i\cdot}$	s_i^2	$SN_{\max,i}$
1	0	0	0	0	17.5250	13.050714	5.532040
4	0	1	1	1	19.4750	8.447857	5.876575
7	0	2	2	2	19.0250	8.313571	5.833544
3	1	0	1	2	20.1250	6.747857	5.964675
2	1	1	2	0	22.8250	11.747857	6.195688
5	1	2	0	1	19.2250	11.422143	5.831468
8	2	0	2	1	19.8500	8.908571	5.920132
6	2	1	0	2	18.3375	14.248393	5.717851
9	2	2	1	0	21.2000	15.585714	6.021715

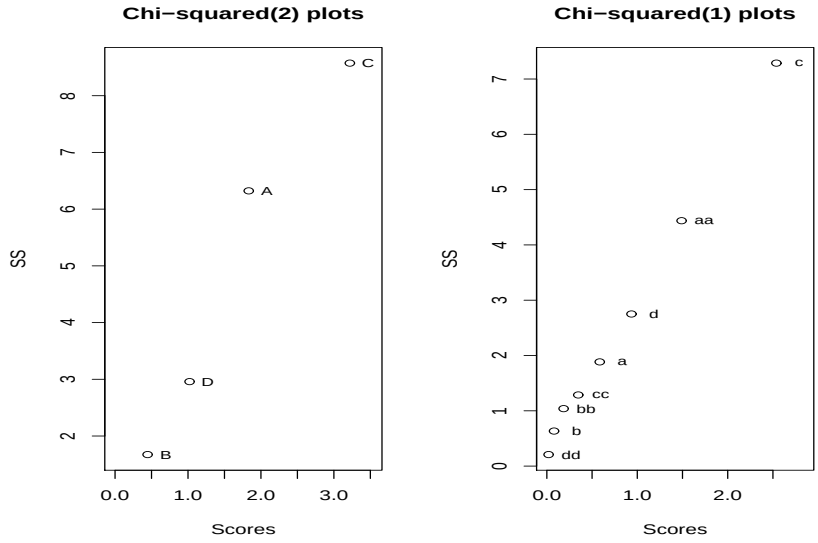

Fig. 3.1 χ^2 plots for main effects: Means of outer array.

difference between the low and high levels), it looks like either the high or middle level of C should work reasonably well.

In practice you have to play off the experimental results against the production costs of various techniques. For example, if two levels have roughly the same effects, obviously you would choose the more inexpensive level. If two levels have radically different costs, it is harder to decide whether improved performance is worth the cost.

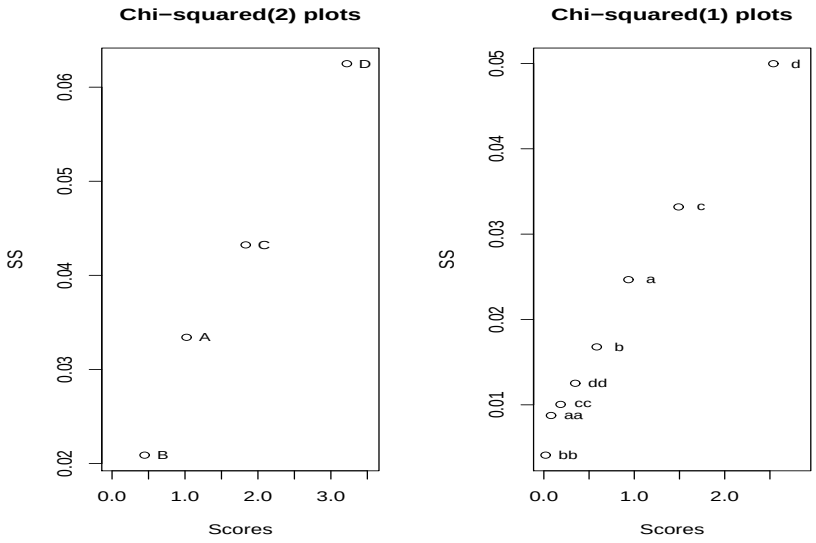

Fig. 3.2 χ^2 plots for main effects: Log standard deviations of outer array.

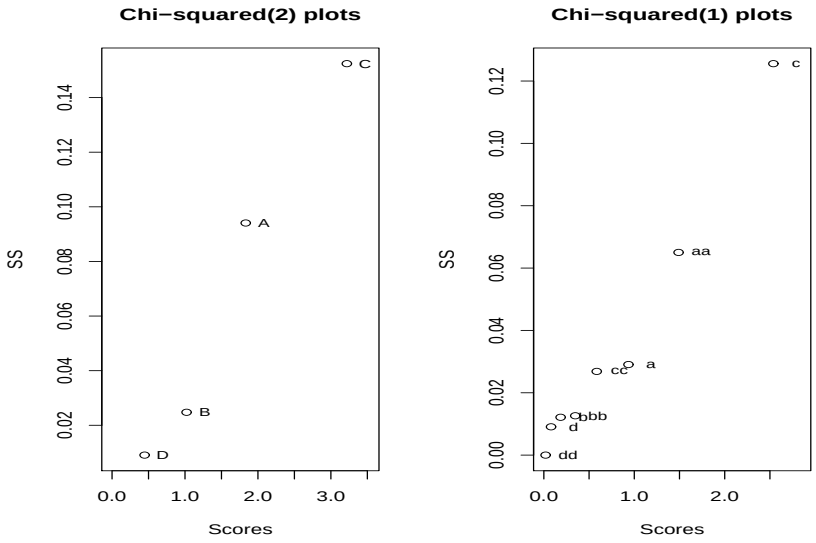

Fig. 3.3 χ^2 plots for main effects: $SN_{\max}$ of outer array.

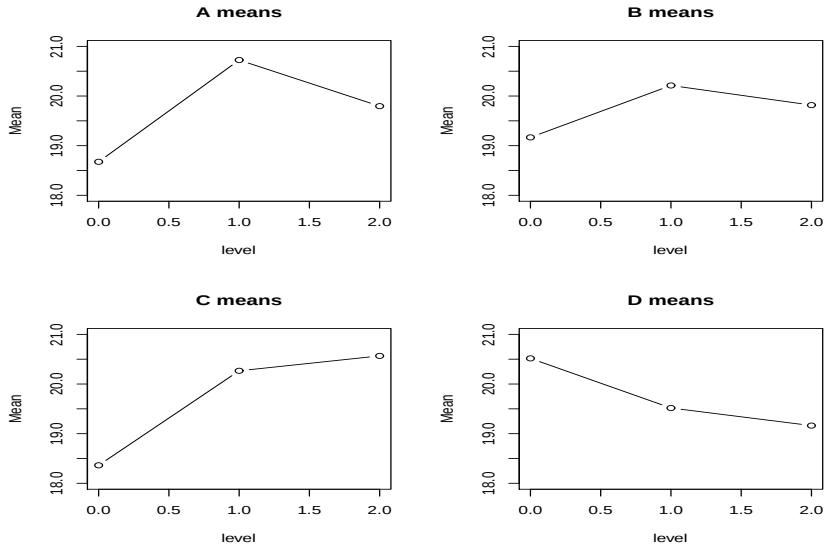

Fig. 3.4 Means plots for main effects: Means of Taguchi outer array.

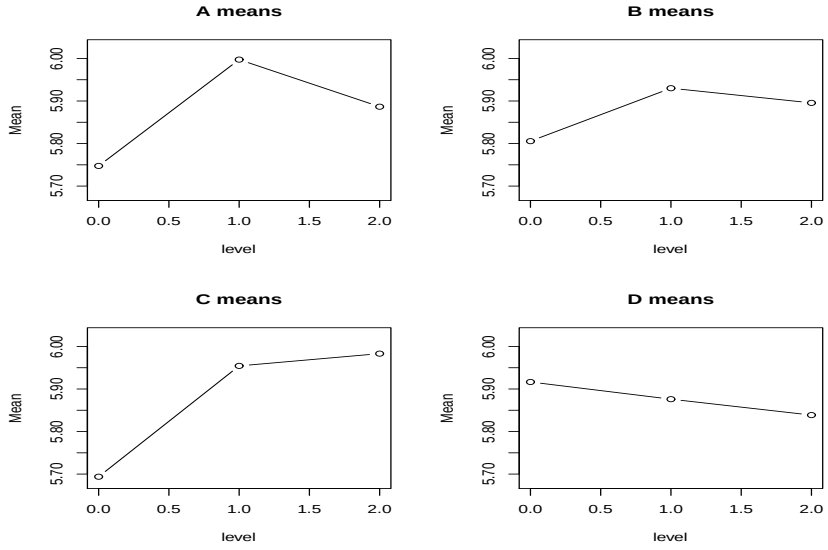

Fig. 3.5 Means plots for main effects: Log standard deviations of Taguchi outer array.

3.5 Analysis of a Partially Confounded 3^3

The data in Table 3.16 are adapted from Kempthorne (1952) to illustrate appropriate methods of analysis. The experiment concerns the effects of three factors on the

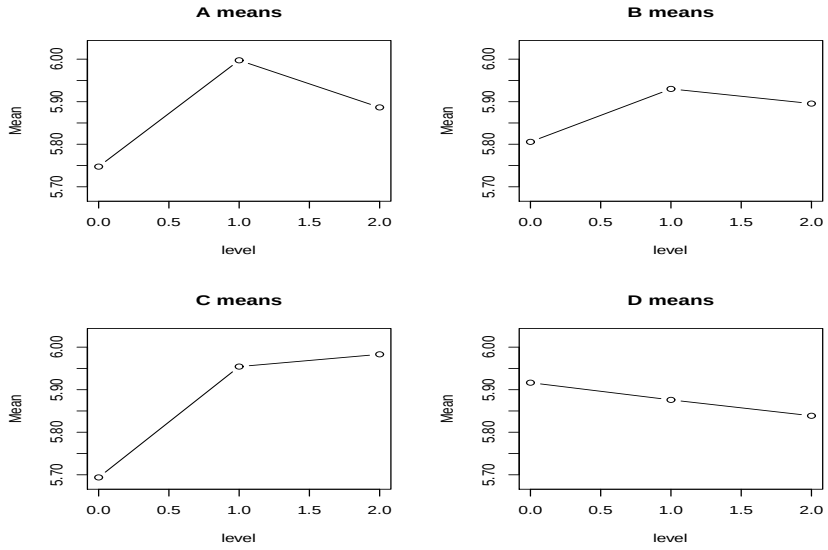

Fig. 3.6 Means plots for main effects: $SN_{\max}$ of Taguchi outer array.

sugar content in sugar beets. The dependent variable is $(\%\text{sugar} - 16)100$. The three factors are D – the sowing date, S – the spacings of the rows, and N – the amount of fertilizer (sulphate of ammonia) applied. In Table 3.16 the treatments are identified by their subscripts alone. The experiment was performed in 2 replications. Each replication has 9 blocks of size 3. In the first replication DS^2N and SN are confounded with blocks. In the second replication DS^2N^2 and DN are confounded. Thus we introduce *partial confounding* for 3^fs in this analysis.

The defining effects for the blocks in Rep. 1 are DS^2N and SN. It follows that

$$DS^2N \times SN = DN^2$$

and

$$DS^2N \times (SN)^2 = DS$$

are also confounded with blocks in Rep. 1. Similarly, the defining effects for the blocks in Rep. 2 are DS^2N^2 and DN, so DS and SN^2 are also confounded with blocks in Rep. 2. Note that DS is confounded with blocks in both replications, so there is no information available on DS.

One version of the analysis of variance table is given as Table 3.17. In this version the interactions are *not* broken into 2 degree of freedom effects. The model involves Blocks and a full factorial analysis on the three factors. The blocks are just listed from 1 to 18. To get the correct analysis from good software, the blocks *must* be fitted before the treatment factors.

Table 3.16 Sugar Beet Data

Rep. 1						Rep. 2					
dsn	*y*	*dsn*	*y*	*dsn*	*y*	*dsn*	*y*	*dsn*	*y*	*dsn*	*y*
000	79	110	62	102	79	000	47	202	42	012	10
121	44	022	−2	011	105	211	43	021	36	220	108
212	59	201	53	220	56	122	30	110	44	101	33
100	85	001	100	111	105	112	39	100	88	121	79
221	36	210	105	020	50	020	21	011	53	002	18
012	70	122	79	202	50	201	39	222	39	210	88
200	59	222	21	211	96	102	44	212	42	200	27
112	13	101	47	120	65	010	33	120	105	111	4
021	70	010	30	002	85	221	68	001	27	022	56

In Rep. 1 the defining effects for blocks are DS^2N and SN with DN^2 and DS also confounded. In Rep. 2 the defining effects for blocks are DS^2N^2 and DN with SN^2 and DS also confounded.

In Minitab 16, the output from the "glm" command provided two sets of degrees of freedom: model degrees of freedom, i.e., what one would normally expect to have for degrees of freedom, and reduced degrees of freedom. *The reduced degrees of freedom are appropriate.* When the model and reduced degrees of freedom differ, the reduced number had a + after it to emphasize the difference. For example, under normal conditions the $D*S$ interaction would have 4 degrees of freedom, however $D*S$ has been decomposed into DS, which is completely confounded with blocks, and DS^2. The 2 degrees of freedom in the table for $D*S$ are just the two degrees of freedom for DS^2. In Minitab 18, my experience with similar problems makes me suspect that the output will look pretty, but be useless. I haven't actually checked it yet. Minitab 19 will be released soon. Hope springs eternal.

Table 3.17 Analysis of Variance for Confounded Sugar Beet Experiment

Analysis of Variance				
Source	*df*	*SS*	*MS*	*F*
Blks	17	17416.00	1024.47	0.81
D	2	838.78	419.39	0.33
S	2	60.78	30.39	0.02
N	2	4177.33	2088.67	1.65
$D*S$	2	238.78	119.39	0.09
$D*N$	4	1692.22	423.06	0.33
$S*N$	4	898.89	224.72	0.18
$D*S*N$	8	2724.22	340.53	0.27
Error	12	15178.33	1264.86	
Total	53	43225.33		

The most remarkable thing about Table 3.17 is the fact that the F statistics are so uniformly small. If I were getting paid to analyze these data, I would have a serious

talk with the people who conducted the experiment to see whether some explanation of this could be found. In the present context, our primary concern is not the data but the *process* of analyzing the data.

3.5.1 The Expanded ANOVA Table

We now consider the expanded ANOVA table in which the interactions are broken into 2 degree of freedom effects. The expanded table is presented as Table 3.18. In Table 3.18, the first line of Table 3.17 has been broken into Reps and Blocks within Reps. The mean square for Blocks in Table 3.17 is computed from the 18 block means. Each block mean is the average of 3 observations. The mean square for Reps in Table 3.18 is computed from the two replication means. Each Rep. mean is averaged over 27 observations. To compute the sum of squares for Blocks within Reps, subtract the sum of squares for Reps from the sum of squares for Blocks in Table 3.17. The lines for D, S, and N are identical in the two ANOVA tables; they are computed in the usual way. The three means for any main effect are each averages over 18 observations.

Table 3.18 Analysis of Variance for Confounded Sugar Beet Experiment

Analysis of Variance				
Source	*df*	*SS*	*MS*	*F*
Reps	1	3552.67	3552.67	2.81
Blks(Reps)	16	13863.33	866.46	0.69
D	2	838.78	419.39	0.34
S	2	60.78	30.39	0.02
N	2	4177.33	2088.66	1.65
DS^2	2	238.78	119.39	0.09
DN/Rep 1	2	578.67	289.33	0.23
DN^2/Rep 2	2	1113.56	556.78	0.44
SN/Rep 2	2	170.89	85.44	0.07
SN^2/Rep 1	2	728.00	364.00	0.29
DSN	2	552.33	276.17	0.22
DSN^2	2	446.33	223.17	0.18
DS^2N/Rep 2	2	1381.56	690.78	0.55
DS^2N^2/Rep 1	2	344.00	172.00	0.14
Error	12	15178.33	1264.86	
Total	53	43225.33		

The analysis begins to get more complicated when we consider the interactions. The effects DN^2, SN, and DS^2N are confounded in Rep. 1 but not in Rep. 2, so we can obtain mean squares and sums of squares for these effects from Rep. 2. Each of these effects defines three groups of treatments. The sugar content observations

from Rep. 2 are listed below for each group along with their means. The means are averages over 9 observations.

Groups and Means from Rep. 2

DN^2			SN			DS^2N		
0	1	2	0	1	2	0	1	2
47	43	30	47	30	43	47	30	43
42	44	36	36	44	42	44	42	36
33	10	108	10	33	108	108	10	33
21	39	39	39	39	21	39	21	39
39	88	53	88	39	53	53	88	39
79	18	88	79	88	18	79	88	18
33	68	44	68	33	44	44	68	33
42	105	27	42	27	105	42	27	105
4	56	27	27	56	4	56	4	27
$37.\bar{7}$	$52.\bar{3}$	$50.\bar{2}$	$48.\bar{4}$	$43.\bar{2}$	$48.\bar{6}$	$56.\bar{8}$	42.0	$41.\bar{4}$

The mean square for DN^2 is 9 times the sample variance of $37.\bar{7}$, $52.\bar{3}$, and $50.\bar{2}$. The mean squares for SN and DS^2N are found similarly.

The effects DS^2N^2, DN, and SN^2 are confounded in Rep. 2 but not in Rep. 1 so we can obtain mean squares and sums of squares for these effects from Rep. 1. Again, each of the effects defines three groups of treatments. The sugar content observations from Rep. 1 are listed below for each group along with their means. The means are again averages over 9 observations and the mean square for each effect is 9 times the sample variance of the three group means.

Groups and Means from Rep. 1

DN			SN^2			DS^2N^2		
0	1	2	0	1	2	0	1	2
79	59	44	79	59	44	79	44	59
53	62	−2	−2	53	62	62	53	−2
79	105	56	105	56	79	56	105	79
36	85	70	85	70	36	70	85	36
79	100	105	79	100	105	79	105	100
50	50	105	105	50	50	50	50	105
13	70	59	59	13	70	70	13	59
30	21	47	21	47	30	47	21	30
96	65	85	96	65	85	96	85	65
$57.\bar{2}$	$68.\bar{5}$	$63.\bar{2}$	$69.\bar{6}$	57.0	$62.\bar{3}$	$67.\bar{6}$	$62.\bar{3}$	59.0

There are three interaction effects remaining: DS^2, DSN, and DSN^2. None of these effects are confounded in either replication so their mean squares are computed from the complete data. The observations in the three groups for DS^2 are given below with their means.

DS^2 Groups and Means from Complete Data

$DS^2(0)$			$DS^2(1)$			$DS^2(2)$		
79	12	39	59	70	21	44	59	39
62	21	39	−2	47	88	53	30	53
56	85	18	79	96	88	105	65	79
36	47	68	85	43	44	70	30	33
100	44	27	105	36	42	79	42	105
105	108	4	50	33	56	50	10	27
	$52.8\bar{3}$			$57.7\bar{7}$			$54.0\bar{5}$	

The mean square for DS^2 is 18 times the sample variance of the three means. The observations in the three groups for DSN are given below with their means.

DSN Groups and Means from Complete Data

$DSN(0)$			$DSN(1)$			$DSN(2)$		
79	70	39	44	13	39	59	59	21
53	21	39	−2	30	88	62	47	53
79	65	88	56	96	79	105	85	18
70	47	44	85	43	33	36	30	68
105	36	105	100	42	27	79	44	42
105	10	4	50	108	56	50	33	27
	$58.8\bar{3}$			$54.8\bar{3}$			51.00	

Finally, the observations in the three groups for DSN^2 are given below with their means.

DSN^2 Groups and Means from Complete Data

$DSN^2(0)$			$DSN^2(1)$			$DSN^2(2)$		
79	13	39	59	70	39	44	59	21
−2	47	53	53	30	88	62	21	39
105	65	88	56	85	18	79	96	79
36	47	68	85	30	33	70	43	44
105	42	105	79	36	42	100	44	27
50	33	56	105	108	4	50	10	27
	$57.1\bar{6}$			$56.6\bar{6}$			$50.8\bar{3}$	

In general, *an effect is estimated from all replications in which it is not confounded*.

This particular experiment has $dfE = 12$, but if it had not involved replications we would need an alternative form of analysis such as χ^2 plots.

3.5.2 Interaction Contrasts

Typically, when an interaction effect is of even marginal significance, we investigate it further by looking at interaction contrasts. Christensen (2011, Sec. 7.2.1) and Christensen (1996, Chapters 11, 12) discuss interaction contrasts in detail. To define a contrast in, say, the $D*N$ interaction we typically choose a contrast in D and a contrast in N and combine them to form an interaction contrast. For example, if we use the main effect contrasts $2n_0 - n_1 - n_2$ and $d_0 - d_1$, the interaction contrast coefficients are the 9 numbers in the body of of the table

		N		
		n_0	n_1	n_2
D	contrasts	2	−1	−1
d_0	1	2	−1	−1
d_1	−1	−2	1	1
d_2	0	0	0	0

The 9 interaction contrast coefficients are obtained by multiplying the D contrast coefficients by the N contrast coefficients. The 9 contrast coefficients in the body of this table correspond to the 9 $D*N$ combinations $d_0n_0, d_0n_1, \ldots, d_2n_2$, so the estimated contrast is the linear combination of the D-N mean values (averaged over Reps. and S) determined by the contrast coefficients.

In general, an interaction contrast is defined by contrast coefficients

	N		
D	n_0	n_1	n_2
d_0	q_{00}	q_{01}	q_{02}
d_1	q_{10}	q_{11}	q_{12}
d_2	q_{20}	q_{21}	q_{22}

where the sum of the q_{ij}s over each row and over each column equals 0. We now examine how this approach to interaction contrasts relates to the analysis just given for a confounded 3^f.

Consider again the $D*N$ interaction. It has been decomposed into two parts, DN and DN^2. At one level, it is very easy to find interaction contrasts in DN and DN^2. For example, DN defines three groups $DN(0)$, $DN(1)$, and $DN(2)$ with sample means from Rep. 1 of $57.\bar{2}$, $68.\bar{5}$, and $63.\bar{2}$, respectively. To obtain the DN sum of squares, we simply performed a one-way ANOVA on the three group means. Continuing our analogy with one-way ANOVA, DN has two degrees of freedom so we can find two orthogonal contrasts in DN. In particular, we can define orthogonal contrasts in the DN groups, say

$$DN_1 \equiv (3)[DN(0)] + (-3)[DN(1)] + (0)[DN(2)], \tag{1}$$

which compares 3 times the $DN(0)$ group mean with 3 times the $DN(1)$ group mean, and

$$DN_2 \equiv (3)[DN(0)] + (3)[DN(1)] + (-6)[DN(2)],$$

which is *equivalent* to comparing the average of groups $DN(0)$ and $DN(1)$ with the $DN(2)$ group mean. The reason for including the factors of 3 in the contrasts will become clear later. Estimates and sums of squares for these contrasts are computed in the usual way based on the sample means for the three groups. The computations will be illustrated later.

We can also define orthogonal contrasts in DN^2, say

$$DN_1^2 \equiv (3)[DN^2(0)] + (-3)[DN^2(1)] + (0)[DN^2(2)]$$

and

$$DN_2^2 \equiv (3)[DN^2(0)] + (3)[DN^2(1)] + (-6)[DN^2(2)].$$

The appropriate means for estimating the DN^2 contrasts are obtained from Rep. 2.

At issue is the correspondence between these contrasts in DN and DN^2 and contrasts in the four degree of freedom interaction $D*N$ that we usually consider. The key to the correspondence is in identifying the treatment groups for DN and DN^2 relative to the 9 combinations of a level of D with a level of N. These correspondences are given below.

DN Groups				DN^2 Groups			
	N				N		
D	n_0	n_1	n_2	D	n_0	n_1	n_2
d_0	0	1	2	d_0	0	2	1
d_1	1	2	0	d_1	1	0	2
d_2	2	0	1	d_2	2	1	0

For example, any treatment with d_1n_2 corresponds to DN group $[1+2] \bmod 3 = 0$ and DN^2 group $[1+(2)2] \bmod 3 = 2$.

The contrast DN_1 in equation (1) compares 3 times group $DN(0)$ to 3 times $DN(1)$, we can rewrite the contrast as

DN_1 Contrast			
	N		
D	n_0	n_1	n_2
d_0	1	−1	0
d_1	−1	0	1
d_2	0	1	−1

Here the three treatments in group $DN(0)$ have been assigned 1s, the three treatments in group $DN(1)$ have been assigned -1s, and the three treatments in group $DN(2)$ have been assigned 0s. This contrast compares the sum of the 3 treatments in the $DN(0)$ group with the sum of the 3 treatments in the $DN(1)$ group. We will call this *the $D*N$ version of the DN_1 contrast*. Note that the contrast given is indeed an interaction contrast in $D*N$, because the contrast coefficients in each row and column sum to 0.

The $D*N$ version of DN_2 can be constructed similarly by assigning 1s to the treatments in $DN(0)$ and $DN(1)$ and -2s to the treatments in $DN(2)$.

DN_2 Contrast			
	N		
D	n_0	n_1	n_2
d_0	1	1	−2
d_1	1	−2	1
d_2	−2	1	1

The DN^2 contrasts, DN_1^2 and DN_2^2 can also be written in their $D*N$ versions.

DN_1^2 Contrast				DN_2^2 Contrast			
	N				N		
D	n_0	n_1	n_2	D	n_0	n_1	n_2
d_0	1	0	−1	d_0	1	−2	1
d_1	−1	1	0	d_1	1	1	−2
d_2	0	−1	1	d_2	−2	1	1

It is a simple matter to check that these four contrasts in the $D*N$ interaction are orthogonal. Incidentally, this confirms that the DN and DN^2 effects are orthogonal. Orthogonality of all the 2 degree of freedom effects was assumed throughout our analysis.

Just as the DN contrasts can be written in two ways, we can also obtain the estimates in two ways. Using the group means obtained from Rep. 1, the estimate of $DN_1 \equiv (3)[DN(0)] + (-3)[DN(1)] + (0)[DN(2)]$ is

$$\widehat{DN_1} \equiv (3)57.\bar{2} + (-3)68.\bar{5} + (0)63.\bar{2} = -34.$$

The estimate of $DN_2 \equiv (3)[DN(0)] + (3)[DN(1)] + (-6)[DN(2)]$ is obtained similarly. Estimates of DN_1^2 and DN_2^2 use the group means obtained from Rep. 2. The variance of $\widehat{DN_1}$ is

$$\text{Var}(\widehat{DN_1}) = \sigma^2[3^2 + (-3)^2 + 0^2]/9 = 2\sigma^2$$

where the 9 in the denominator of the second term is the number of observations that the DN group means are averaged over. The standard error for the estimate of DN_1 is

$$\text{SE}(\widehat{DN_1}) = \sqrt{MSE[3^2 + (-3)^2 + 0^2]/9}$$

The sum of squares for DN_1 is

$$\frac{-34^2}{[3^2 + (-3)^2 + 0^2]/9} = 578.$$

Similar computations apply to DN_2, DN_1^2, and DN_2^2.

Alternatively, we could apply the $D*N$ version of DN_1 to the means table

D-N Means from Rep. 1

$N = 3$	N		
D	n_0	n_1	n_2
d_0	53.0	$91.\bar{6}$	51.0
d_1	$70.\bar{6}$	$65.\bar{3}$	57.0
d_2	$73.\bar{3}$	$61.\bar{6}$	$43.\bar{3}$

The estimated DN_1 contrast in $D*N$ form is

$$\widehat{DN_1} = 53.0 - 91.\bar{6} - 70.\bar{6} + 57.0 + 61.\bar{6} - 43.\bar{3} = -34.$$

This is exactly the estimate obtained from the other method. The variance of $\widehat{DN_1}$ in $D*N$ form is

$$\text{Var}(\widehat{DN_1}) = \sigma^2[1^2+(-1)^2+(-1)^2+1^2+1^2+(-1)^2]/3 = 2\sigma^2$$

where the 3 in the denominator of the middle term is the number of observations in each mean. The sum of squares is

$$\frac{-34^2}{[1^2+(-1)^2+(-1)^2+1^2+1^2+(-1)^2]/3} = 578$$

Of course the variance and the sums of squares from the two methods are identical.

We can construct other $D*N$ interaction contrasts from the $D*N$ forms of the DN and DN^2 contrasts. If we add the DN_1 and DN_1^2 contrast coefficients we get

	N		
D	n_0	n_1	n_2
d_0	2	−1	−1
d_1	−2	1	1
d_2	0	0	0

This is not only a contrast in $D*N$ but it even corresponds to our usual way of constructing interaction contrasts from main effect contrasts. It combines $2n_0 - n_1 - n_2$ and $d_0 - d_1$.

The new contrast is $DN_1 + DN_1^2$, so the estimate of the new contrast is just $\widehat{DN_1} + \widehat{DN_1^2}$. Recall that $\widehat{DN_1}$ is estimated from Rep. 1 and $\widehat{DN_1^2}$ is estimated from Rep. 2. Because DN_1 and DN_1^2 are orthogonal contrasts, $\widehat{DN_1}$ and $\widehat{DN_1^2}$ are independent, the variance of the new estimate is

$$\text{Var}(\widehat{DN_1} + \widehat{DN_1^2}) = \text{Var}(\widehat{DN_1}) + \text{Var}(\widehat{DN_1^2}) = 2\sigma^2 + 2\sigma^2,$$

and the standard error is $\sqrt{4MSE}$.

3.6 5^f Factorials

We now briefly consider f factors each at 5 levels, i.e., 5^f's. In this case, as *in all p^f structures with p a prime number, the methods for 3^f factorials extend easily*.

A 5^f can be broken down into groups of $5-1=4$ degrees of freedom. Consider two factors A and B at 5 levels, say, a_0, a_1, a_2, a_3, a_4, and b_0, b_1, b_2, b_3, b_4. There are $5^2 = 25$ treatment combinations $a_i b_j$ for $i, j = 0, 1, 2, 3, 4$. The breakdown into 4 degree of freedom effects is illustrated in Table 3.19.

In a simple extension of our discussion for 3^f factorials, the groups of treatments determined by, say, AB^4 correspond to the five values of

$$z = x_1 + 4x_2 \bmod 5$$

where x_1 and x_2 are the subscripts for the treatment combinations. Table 3.20 gives the groups for all six of the 4 degree of freedom effects. Table 3.21 rewrites the

Table 3.19 Analysis of Variance for a 5^2

Source	df	Source	df
A	4	A	4
B	4	B	4
$A*B$	16	AB	4
		AB^2	4
		AB^3	4
		AB^4	4

treatment groups determined by AB^4. Table 3.21 provides a scheme for assigning the 25 treatments to blocks of size five with AB^4 confounded.

Table 3.20 Effect Groups for a 5^2 Factorial

treatment	A	B	AB	AB^2	AB^3	AB^4
a_0b_0	0	0	0	0	0	0
a_0b_1	0	1	1	2	3	4
a_0b_2	0	2	2	4	1	3
a_0b_3	0	3	3	1	4	2
a_0b_4	0	4	4	3	2	1
a_1b_0	1	0	1	1	1	1
a_1b_1	1	1	2	3	4	0
a_1b_2	1	2	3	0	2	4
a_1b_3	1	3	4	2	0	3
a_1b_4	1	4	0	4	3	2
a_2b_0	2	0	2	2	2	2
a_2b_1	2	1	3	4	0	1
a_2b_2	2	2	4	1	3	0
a_2b_3	2	3	0	3	1	4
a_2b_4	2	4	1	0	4	3
a_3b_0	3	0	3	3	3	3
a_3b_1	3	1	4	0	1	2
a_3b_2	3	2	0	2	4	1
a_3b_3	3	3	1	4	2	0
a_3b_4	3	4	2	1	0	4
a_4b_0	4	0	4	4	4	4
a_4b_1	4	1	0	1	2	3
a_4b_2	4	2	1	3	0	2
a_4b_3	4	3	2	0	3	1
a_4b_4	4	4	3	2	1	0

Any one of the blocks in Table 3.21 provides a $1/5$ replicate of the 5^2 experiment based on AB^4. An effect such as A is aliased with

$$A \times AB^4 = A^2B^4 = (A^2B^4)^3 = A^6B^{12} = AB^2$$
$$A \times (AB^4)^2 = A^3B^8 = (A^3B^8)^2 = A^6B^{16} = AB$$

Table 3.21 AB^4 Groups for a 5^2 Factorial

$AB^4(0)$	$AB^4(1)$	$AB^4(2)$	$AB^4(3)$	$AB^4(4)$
a_0b_0	a_0b_4	a_0b_3	a_0b_2	a_0b_1
a_1b_1	a_1b_0	a_1b_4	a_1b_3	a_1b_2
a_2b_2	a_2b_1	a_2b_0	a_2b_4	a_2b_3
a_3b_3	a_3b_2	a_3b_1	a_3b_0	a_3b_4
a_4b_4	a_4b_3	a_4b_2	a_4b_1	a_4b_0

$$A \times (AB^4)^3 = A^4B^{12} = (A^4B^{12})^4 = A^{16}B^{48} = AB^3$$
$$A \times (AB^4)^4 = A^5B^{16} = B.$$

The alias group is

$$A = B = AB = AB^2 = AB^3.$$

All effects other than AB^4 are aliased together. AB^4 is lost because it defines the $1/5$ rep. Of course a $1/5$ rep. of a 5^2 contains only 5 treatments, so it has only 4 degrees of freedom for treatment effects. Each effect has 4 degrees of freedom, so there could not be more than one treatment effect available in the $1/5$ rep.

A 5^3 in blocks of 5 requires two defining effects, say AB^4 and BC^2, to be confounded with blocks. There are 125 treatments in a 5^3, so with blocks of size 5, there are 25 blocks and 24 degrees of freedom for blocks. Since each effect has 4 degrees of freedom, there must be 6 effects confounded with blocks. These are AB^4, BC^2, and

$$AB^4 \times BC^2 = AB^5C^2 = AC^2,$$
$$AB^4 \times (BC^2)^2 = AB^6C^4 = ABC^4,$$
$$AB^4 \times (BC^2)^3 = AB^7C^6 = AB^2C,$$
$$AB^4 \times (BC^2)^4 = AB^8C^8 = AB^3C^3.$$

The key idea is that one effect is multiplied by all powers of the other effect, up to the $5-1$ power.

If we had done the multiplications reversing the orders of the defining effects, the computations would be more complicated but we would get the same set of confounded effects. For example,

$$BC^2 \times (AB^4)^2 = A^2B^9C^2 = A^2B^4C^2 = (A^2B^4C^2)^3 = A^6B^{12}C^6 = AB^2C.$$

Note that to change $A^2B^4C^2$ into something with a leading exponent of 1, we cubed it, i.e., $A^2B^4C^2 = (A^2B^4C^2)^3$. Any effect raised to a positive integer power less than $p = 5$ remains the same effect, because the modular equation defining groups continues to give the same groups. We cubed $A^2B^4C^2$ because we recognized that $(A^2)^3 = A^6 = A$.

Finally, consider a $1/25$ rep. of a 5^4. For simplicity take AB^4 and BC^2 as defining effects. We have already multiplied AB^4 by the appropriate powers of BC^2 so the

complete set of effects defining the fractional replication is AB^4, BC^2, AC^2, ABC^4, AB^2C, and AB^3C^3. An effect, say A, is aliased with

$$A \times AB^4,\ A \times (AB^4)^2,\ A \times (AB^4)^3,\ A \times (AB^4)^4,$$

$$A \times BC^2,\ A \times (BC^2)^2,\ A \times (BC^2)^3,\ A \times (BC^2)^4,$$

$$A \times AC^2,\ A \times (AC^2)^2,\ A \times (AC^2)^3,\ A \times (AC^2)^4,$$

$$A \times ABC^4,\ A \times (ABC^4)^2,\ A \times (ABC^4)^3,\ A \times (ABC^4)^4,$$

$$A \times AB^2C,\ A \times (AB^2C)^2,\ A \times (AB^2C)^3,\ A \times (AB^2C)^4,$$

and

$$A \times AB^3C^3,\ A \times (AB^3C^3)^2,\ A \times (AB^3C^3)^3,\ A \times (AB^3C^3)^4.$$

When using an unreplicated or fractionally replicated 5^f, an analysis can be based on $\chi^2(4)$ plots of the effects.

Chapter 4
Mixed Factor Levels

We briefly mention some extensions of the methods presented in Chapter 3. These involve treatment structures that are not simply powers of prime numbers. We begin by constructing Taguchi's L_{18} which is a $2 \times 3^{7-5}$ design. In the second section we look at a $2^2 \times 3^2$ in six blocks. Next we reexamine Byrne and Taguchi's $2^3 \times 3^{4-2}$. The final section examines a 3×4 design.

For additional information on mixed factor levels, see Kempthorne (1952) or possibly Hinkleman and Kempthorne (2005).

4.1 Fractionated Partial Confounding

Consider a group of seven factors each at three levels, say B, C, D, E, F, G, H. (We have omitted factor A for now so that we can incorporate it later.)

In this discussion we refer to the *effect group*, say, $CDEFGH(z)$, as all factor combinations that satisfy the subscript equation

$$z = 0x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 \bmod 3.$$

Similarly, $EF^2GH^2(z)$ denotes all factor combinations that satisfy

$$z = 0x_2 + 0x_3 + 0x_4 + x_5 + 2x_6 + x_7 + 2x_8 \bmod 3.$$

Consider a 3^{7-5} design, i.e. a 3^{-5} fractional replication of the 3^7 factorial structure. In particular, consider the 3^{7-5} defined by the intersection of the five defining effect groups $CDEFGH(0)$, $BDE^2(0)$, $DEH(0)$, $BFG^2(0)$, and $EF^2GH^2(0)$. Using the modular subscript equations associated with these five effect groups, it is not hard to see that the nine treatments involved are

$$T_1 = \begin{bmatrix} b & c & d & e & f & g & h \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 2 & 2 & 2 & 2 & 2 & 2 \\ 1 & 0 & 0 & 1 & 1 & 2 & 2 \\ 1 & 1 & 1 & 2 & 2 & 0 & 0 \\ 1 & 2 & 2 & 0 & 0 & 1 & 1 \\ 2 & 0 & 1 & 0 & 2 & 1 & 2 \\ 2 & 1 & 2 & 1 & 0 & 2 & 0 \\ 2 & 2 & 0 & 2 & 1 & 0 & 1 \end{bmatrix}.$$

Now consider an alternative 3^{7-5} design that uses three of the same defining effects and two of the same effect groups $CDEFGH(0)$ and $BDE^2(0)$ but this time for the DEH effect we use $DEH(1)$. The other two defining effect groups are $BF^2G(0)$ and $EFG^2H^2(2)$. Again, it is not hard to check that the following treatments

$$T_2 = \begin{bmatrix} b & c & d & e & f & g & h \\ 0 & 0 & 2 & 2 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 2 & 2 & 1 \\ 0 & 2 & 1 & 1 & 0 & 0 & 2 \\ 1 & 0 & 1 & 2 & 0 & 2 & 1 \\ 1 & 1 & 2 & 0 & 1 & 0 & 2 \\ 1 & 2 & 0 & 1 & 2 & 1 & 0 \\ 2 & 0 & 2 & 1 & 2 & 0 & 1 \\ 2 & 1 & 0 & 2 & 0 & 1 & 2 \\ 2 & 2 & 1 & 0 & 1 & 2 & 0 \end{bmatrix}$$

satisfy the associated subscript equations.

The idea is to run both of these 3^{7-5} fractional replications. Obviously, all information is lost on $CDEFGH$, BDE^2, and DEH because they are confounded with the grand mean in both fractional replications. From the first 3^{7-5} one could potentially learn about BF^2G and EFG^2H^2 whereas these are lost in the second fractional replication. Likewise, from the second fractional replication one could potentially learn about BFG^2 and EF^2GH^2. But in a design this small (this highly fractionated), with the extensive aliasing involved, what one typically cares about is that both fractions be resolution III designs so that all of the main effects can be estimated. Actually checking that these are resolution III designs, by finding all of the implicit defining effects, is quite tedious.

If both fractional replications are resolution III designs, why use *both* of them? So that we can incorporate a two-level factor A into the design. We will look at one level of A in the first fractional rep. and the other level of A in the second fractional rep. Using the subscripts 0, 2 to denote the levels of A, the treatment subscripts for this new $2 \times 3^{7-5}$ design are

$$T = \begin{bmatrix} a & b & c & d & e & f & g & h \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 2 & 2 & 2 & 2 & 2 & 2 \\ 0 & 1 & 0 & 0 & 1 & 1 & 2 & 2 \\ 0 & 1 & 1 & 1 & 2 & 2 & 0 & 0 \\ 0 & 1 & 2 & 2 & 0 & 0 & 1 & 1 \\ 0 & 2 & 0 & 1 & 0 & 2 & 1 & 2 \\ 0 & 2 & 1 & 2 & 1 & 0 & 2 & 0 \\ 0 & 2 & 2 & 0 & 2 & 1 & 0 & 1 \\ 2 & 0 & 0 & 2 & 2 & 1 & 1 & 0 \\ 2 & 0 & 1 & 0 & 0 & 2 & 2 & 1 \\ 2 & 0 & 2 & 1 & 1 & 0 & 0 & 2 \\ 2 & 1 & 0 & 1 & 2 & 0 & 2 & 1 \\ 2 & 1 & 1 & 2 & 0 & 1 & 0 & 2 \\ 2 & 1 & 2 & 0 & 1 & 2 & 1 & 0 \\ 2 & 2 & 0 & 2 & 1 & 2 & 0 & 1 \\ 2 & 2 & 1 & 0 & 2 & 0 & 1 & 2 \\ 2 & 2 & 2 & 1 & 0 & 1 & 2 & 0 \end{bmatrix}.$$

This turns out to be Taguchi's famous L_{18} design. More frequently Taguchi's design is presented using the subscripts -1, 0, 1 rather than 0, 1, 2, thus

$$T = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 0 & -1 & -1 & 0 & 0 & 1 & 1 \\ -1 & 0 & 0 & 0 & 1 & 1 & -1 & -1 \\ -1 & 0 & 1 & 1 & -1 & -1 & 0 & 0 \\ -1 & 1 & -1 & 0 & -1 & 1 & 0 & 1 \\ -1 & 1 & 0 & 1 & 0 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & -1 & 1 & 1 & 0 & 0 & -1 \\ 1 & -1 & 0 & -1 & -1 & 1 & 1 & 0 \\ 1 & -1 & 1 & 0 & 0 & -1 & -1 & 1 \\ 1 & 0 & -1 & 0 & 1 & -1 & 1 & 0 \\ 1 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \\ 1 & 0 & 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & 1 & -1 & 1 & 0 & 1 & -1 & 0 \\ 1 & 1 & 0 & -1 & 1 & -1 & 0 & 1 \\ 1 & 1 & 1 & 0 & -1 & 0 & 1 & -1 \end{bmatrix}.$$

Again, I do not see any advantage to using L_{18} rather than using any other two resolution III 3^{7-5} designs, at least no advantage other than that L_{18} is well known and easy to find.

4.2 More Mixtures of Prime Powers

Suppose we have a $2 \times 2 \times 3 \times 3$ factorial treatment structure that needs to be put in blocks of six units. There are 36 treatments, so there would be six blocks of six units each. Any one of these blocks could be used as a $1/6$ replication. Think of the $2 \times 2 \times 3 \times 3$ factorial as combining two sets of treatments: a 2×2 factorial and a 3×3 factorial. One simple way to get a $1/6$ replication is to take a $1/2$ rep. of the 2×2 and a $1/3$ rep. of the 3×3. If the factors are A, B, C, and D, we can use the AB interaction to define the two blocks a_0b_0, a_1b_1 and a_0b_1, a_1b_0 and the CD^2 interaction to define the three blocks c_0d_0, c_1d_1, c_2d_2 and c_0d_2, c_1d_0, c_2d_1 and c_0d_1, c_1d_2, c_2d_0. Combining each AB block with each CD^2 block gives a collection of six blocks. Note that the blocking structure also gives six different $1/6$th fractional replications for the $2 \times 2 \times 3 \times 3$ factorial. The blocks are given in Table 4.1.

Table 4.1 $2 \times 2 \times 3 \times 3$ with AB and CD^2 Confounded

Confounded		CD^2		
		$CD^2(0)$	$CD^2(1)$	$CD^2(2)$
AB	$AB(0)$	Blk. 1	Blk. 2	Blk. 3
		$a_0b_0c_0d_0$	$a_0b_0c_0d_2$	$a_0b_0c_0d_1$
		$a_0b_0c_1d_1$	$a_0b_0c_1d_0$	$a_0b_0c_1d_2$
		$a_0b_0c_2d_2$	$a_0b_0c_2d_1$	$a_0b_0c_2d_0$
		$a_1b_1c_0d_0$	$a_1b_1c_0d_2$	$a_1b_1c_0d_1$
		$a_1b_1c_1d_1$	$a_1b_1c_1d_0$	$a_1b_1c_1d_2$
		$a_1b_1c_2d_2$	$a_1b_1c_2d_1$	$a_1b_1c_2d_0$
	$AB(1)$	Blk. 4	Blk. 5	Blk. 6
		$a_0b_1c_0d_0$	$a_0b_1c_0d_2$	$a_0b_1c_0d_1$
		$a_0b_1c_1d_1$	$a_0b_1c_1d_0$	$a_0b_1c_1d_2$
		$a_0b_1c_2d_2$	$a_0b_1c_2d_1$	$a_0b_1c_2d_0$
		$a_1b_0c_0d_0$	$a_1b_0c_0d_2$	$a_1b_0c_0d_1$
		$a_1b_0c_1d_1$	$a_1b_0c_1d_0$	$a_1b_0c_1d_2$
		$a_1b_0c_2d_2$	$a_1b_0c_2d_1$	$a_1b_0c_2d_0$

In addition to AB and CD^2, the design in Table 4.1 has $AB \times CD^2 = ABCD^2$ confounded with blocks, i.e., $ABCD^2$ is implicitly a defining effect for each $1/6$ rep. The effects defining blocks, AB, CD^2 and $ABCD^2$, have 1, 2, and 2 degrees of freedom respectively. These account for the 5 degrees of freedom for blocks.

Aliasing for any of the $1/6$ rep.s goes as

$$A \times AB = B,$$

$$A \times CD^2 = ACD^2,$$

$$A \times ABCD^2 = BCD^2.$$

Somewhat similarly,

$$C = (AB)C = C(CD^2) = C(CD^2)^2 = C(ABCD^2) = C(AB)(CD^2)^2$$

which simplifies to

$$C = ABC = CD = D = ABCD = ABD.$$

Note that, for example, $CD = (CD)^2$ so in simplifying we use the fact that $ABC^2D^2 = ABCD$. Finally,

$$AC = AC(AB) = AC(CD^2) = AC(CD^2)^2 = AC(ABCD^2) = AC(AB)(CD^2)^2,$$

which simplifies to

$$AC = BC = ACD = AD = BCD = BD.$$

Now considering something larger, say, a $2^4 \times 3^3$ in 36 blocks with ABC, ABD, EFG, and EF^2G confounded. Implicitly, we get the defining effects

$$ABC \times ABD = CD$$

$$EFG \times EF^2G = E^2G^2 = EG$$

$$EFG \times (EF^2G)^2 = E^3F^5G^3 = F^2 = F$$

$$ABC \times EFG = ABCEFG, \quad ABC \times EG = ABCEG$$

$$ABC \times EF^2G = ABCEF^2G, \quad ABC \times F = ABCF$$

$$ABD \times EFG = ABDEFG, \quad ABD \times EG = ABDEG$$

$$ABD \times EF^2G = ABDEF^2G, \quad ABD \times F = ABDF$$

$$CD \times EFG = CDEFG, \quad CD \times EG = CDEG$$

$$CD \times EF^2G = CDEF^2G, \quad CD \times F = CDF$$

These account for all 35 degrees of freedom between blocks. Each of ABC, ABD, and CD have one degree of freedom. The other 16 effects have two degrees of freedom.

4.3 Taguchi's L_9 with an Outer Array

In Section 3.4 we examined Byrne and Taguchi's data which has a 3^{4-2} inner array and a 2^3 outer array so overall the design is a $2^3 \times 3^{4-2}$, a mixture of prime powers. The earlier discussion was not explicit about the outer array. The complete set of factors and levels are given below.
A: Interference (Low, Medium, High)
B: Wall Thickness (Thin, Medium, Thick)

C: Ins. Depth (Shallow, Medium, Deep)
D: Percent Adhesive (Low, Medium, High)
E: Condition Time (24hr, 120hr)
F: Condition Temp. (72°F, 150°F)
G: Condition R. H. (25%, 75%)
The data are given in Table 4.2.

Table 4.2 Taguchi $2^3 \times 3^{4-2}$ design

	Control Factors				Noise Factors							
Run	a	b	c	d	$e_0f_0g_0$	$e_0f_0g_1$	$e_0f_1g_0$	$e_0f_1g_1$	$e_1f_0g_0$	$e_1f_0g_1$	$e_1f_1g_0$	$e_1f_1g_1$
1	0	0	0	0	15.6	9.5	16.9	19.9	19.6	19.6	20.0	19.1
4	0	1	1	1	15.0	16.2	19.4	19.6	19.7	19.8	24.2	21.9
7	0	2	2	2	16.3	16.7	19.1	15.6	22.6	18.2	23.3	20.4
3	1	0	1	2	18.3	17.4	18.9	18.6	21.0	18.9	23.2	24.7
2	1	1	2	0	19.7	18.6	19.4	25.1	25.6	21.4	27.5	25.3
5	1	2	0	1	16.2	16.3	20.0	19.8	14.7	19.6	22.5	24.7
8	2	0	2	1	16.4	19.1	18.4	23.6	16.8	18.6	24.3	21.6
6	2	1	0	2	14.2	15.6	15.1	16.8	17.8	19.6	23.2	24.4
9	2	2	1	0	16.1	19.9	19.3	17.3	23.1	22.7	22.6	28.6

Our earlier discussion emphasized that inner-outer array experiments are naturally conducted as split plot experiments and we illustrated the traditional Taguchi analysis that treats the inner array as the whole plot treatments in split plot design. Here we focus on treating the outer array as whole plot treatments. As mentioned earlier, this seems a more natural procedure because with the outer array noise factors being more difficult to control, it seems natural to (randomly pick a set of noise factors from the outer array and) control them and, while controlled at some level, to run all of the inner array treatment combinations (in random order).

As discussed in Christensen (1996, 2015) in a split plot model the subplot treatments (here the inner array) can be analyzed as a randomized complete block experiment in which each whole plot is treated as a block for the subplot treatments. In the current situation, that means that every level of the outer array would merely define a block for the purpose of examining the inner array. Since Taguchi's interest is exclusively in the inner array, for Taguci's purposes the analysis should simply be a randomized complete block analysis with each level of the outer array defining a block.

Table 4.3 gives the appropriate ANOVA table for this interpretation of the Byrne-Taguchi data. Similar to the signal-to-noise ratio analysis in Section 3.4, factors A and C look to be important. Evaluating the main effects involves looking at the means plots in Figure 3.4.

Also as discussed in Christensen (1996, 2015), one of the subtleties of split plot designs is that the whole plot treatment by subplot treatment interaction can be teased out of the randomized-complete-block-on-subplot-treatments error term.

Table 4.3 Analysis of Variance for Byrne-Taguchi Data

Analysis of Variance					
Source	df	SS	MS	F	P
Blks	7	447.02	63.860	14.7602	0.000
A	2	50.58	25.288	5.8450	0.005
B	2	13.38	6.692	1.5467	0.222
C	2	68.59	34.297	7.9271	0.001
D	2	23.67	11.837	2.7359	0.074
Error	56	242.29	4.327		
Total	71	845.54			

When there are no replications on the whole plot treatments, the entire subplot error will be eaten up by the whole plot treatment by subplot treatment interaction. But the point of a randomized block experiment is that the treatment effects have to be prominent enough that they are distinguishable above any block by treatment interaction, so in this case, we want the inner array effects to be prominent enough that they are distinguishable above the outer array by inner array interaction. Ideally the blocks would be a representative (random) sample of possible blocks, just as the outer array is chosen to be a representative (but not random) sample of possible noise factors.

We have presented two distinct analyses of these data. As always with designed experiments, *the appropriate analysis is determined by how the experiment was physically performed!*

If we had a $2^3 \times 3^{4-2}$ design without the inner-outer array structure, we would have to fit the model with all the E, F, G main effects and interactions, A, B, C, D main effects, and the interactions between the E, F, G terms and the A, B, C, D main effects. This involves $39 = (8 \times 5) - 1$ terms involving either 1 or 2 degrees of freedom and no error term.

4.4 Factor Levels that are Powers of Primes

To this point we have dealt with factor levels that are prime numbers. If the number of levels for one of the factors is a power of a prime number, we can maintain a simple analysis. Consider a 3×4 factorial treatment structure. Let factor A have levels a_0, a_1, a_2 and let factor B have levels b_0, b_1, b_2, b_3. The number of levels for B is 4, which is a power of a prime number, i.e., $4 = 2^2$. We can artificially replace factor B with two factors C and D each with two levels. We identify the treatment combinations as follows:

B	b_0	b_1	b_2	b_3
CD	c_0d_0	c_0d_1	c_1d_0	c_1d_1

The analysis can now be conducted as either a 3×4 in A and B or as a $3 \times 2 \times 2$ in A, C, and D. The analysis of variance tables are given in Table 4.4 Note that the three degrees of freedom for B correspond to the C, D, and CD terms in the $3 \times 2 \times 2$ analysis and that the six degrees of freedom for AB correspond to the AC, AD, and ACD terms in the $3 \times 2 \times 2$ analysis.

Table 4.4 Analysis of Variance Table for 3×4

Source	df	Source	df
A	2	A	2
B	3	C	1
$A*B$	6	D	1
		AC	2
		AD	2
		CD	1
		ACD	2

The C, D, and CD contrasts are given in Table 4.5. One of these can be used to confound the design into blocks of size 6; two confounding contrasts will give blocks of size 3. One of these contrasts along with the A effect can be used to define blocks of size 2. Similarly, $1/2$, $1/4$, and $1/6$ fractional replications can be defined. Note that the C main effect contrast, $c_0 - c_1$ is equivalent to $(b_0 + b_1) - (b_2 + b_3)$. Similarly, $d_0 - d_1$ is equivalent to $(b_0 + b_2) - (b_1 + b_3)$ while the interaction contrast $c_0d_0 - c_0d_1 - c_1d_0 + c_0d_0$ is equivalent to $(b_0 + b_3) - (b_1 + b_2)$.

Table 4.5 Contrasts for 3×4

A,C,D treatment	C	D	CD	A,B treatment
$a_0c_0d_0$	1	1	1	a_0b_0
$a_0c_0d_1$	1	−1	−1	a_0b_1
$a_0c_1d_0$	−1	1	−1	a_0b_2
$a_0c_1d_1$	−1	−1	1	a_0b_3
$a_1c_0d_0$	1	1	1	a_1b_0
$a_1c_0d_1$	1	−1	−1	a_1b_1
$a_1c_1d_0$	−1	1	−1	a_1b_2
$a_1c_1d_1$	−1	−1	1	a_1b_3
$a_2c_0d_0$	1	1	1	a_2b_0
$a_2c_0d_1$	1	−1	−1	a_2b_1
$a_2c_1d_0$	−1	1	−1	a_2b_2
$a_2c_1d_1$	−1	−1	1	a_2b_3

Table 4.6 Contrasts for 3×4

A,C,D treatment	AC	AD	ACD	A,B treatment
$a_0c_0d_0$	1	1	1	a_0b_0
$a_0c_0d_1$	1	−1	−1	a_0b_1
$a_0c_1d_0$	−1	1	−1	a_0b_2
$a_0c_1d_1$	−1	−1	1	a_0b_3
$a_1c_0d_0$	1	1	1	a_1b_0
$a_1c_0d_1$	1	−1	−1	a_1b_1
$a_1c_1d_0$	−1	1	−1	a_1b_2
$a_1c_1d_1$	−1	−1	1	a_1b_3
$a_2c_0d_0$	1	1	1	a_2b_0
$a_2c_0d_1$	1	−1	−1	a_2b_1
$a_2c_1d_0$	−1	1	−1	a_2b_2
$a_2c_1d_1$	−1	−1	1	a_2b_3

Chapter 5
Screening Designs

This chapter is new.

In this chapter we examine general designs for efficiently screening through many factors. For factorial treatment structures, *screening designs* can be set up to examine all of the factors' main effects efficiently but without the cost and trouble of examining all of the factorial treatments. Often, especially in industrial experiments, there are so many factors that using a complete factorial structure becomes prohibitive because there are just too many treatments to consider. We will see that blocking can be accomplished by treating blocks as additional factors in the experiment.

Replication tends to get short shrift in screening designs. It largely consists of pretending that interaction does not exist and using estimates of the nonexistent interaction to estimate variability. Alternatively, we have looked at some graphical methods for analyzing data without replications.

5.1 Designs at Two Levels

Suppose we have 8 factors each at two levels. The number of factor combinations (treatments) is $2^8 = 256$. That is a lot of treatments, especially if you plan to perform replications in order to estimate error. If you only want to estimate the 8 factorial main effects, in theory you can do that with as few as 9 observations. Nine observations means 9 degrees of freedom, which can be allocated as one for each factor's main effect and one for fitting the grand mean (intercept).

With only 5 factors each at two levels, in theory we could get estimates of all the treatment main effects from as little as 6 observations. Hare's experiment from Chapter 2 used 16 observations to estimate the $\binom{5}{1} = 5$ main effects and $\binom{5}{2} = 10$ two factor interactions. In practice, the smallest number of observations for examining 5 factors each at two levels that has nice properties is 8. We will return to this issue later.

There are many schemes in common use for identifying the treatments in factorial structures. For Hare's experiment with 5 factors each at two levels we denoted the treatments using lower case letters and subscripts. The lower case letters really just provide an ordering for the factor subscripts so we know that, say, the third subscript corresponds to the third factor, C. Given the ordering, the subscripts contain all the information about treatments. In other words, a 16×5 matrix of 0s and 1s identifies the treatments. Another convenient scheme of subscripting is to replace the 0s with -1s, and that is sometimes reduced to reporting just plus and minus signs. Yet another way of identifying treatments is to write down only the treatment letters that have a subscript of 1 (and not write down the subscripts). All of these schemes are illustrated in Table 5.1 and the last scheme also appeared in Table 2.10.

Table 5.1 Alternative treatment identifications for Hare's intermix variability data.

1	$a_0b_0c_0d_1e_1$	00011	-1	-1	-1	1	1	$---++$	de
2	$a_1b_0c_1d_1e_1$	10111	1	-1	1	1	1	$+-+++$	acde
3	$a_1b_1c_0d_0e_0$	11000	1	1	-1	-1	-1	$++---$	ab
4	$a_1b_0c_1d_0e_0$	10100	1	-1	1	-1	-1	$+-+--$	ac
5	$a_0b_1c_0d_0e_1$	01001	-1	1	-1	-1	1	$-+--+$	be
6	$a_0b_0c_1d_0e_1$	00101	-1	-1	1	-1	1	$--+-+$	ce
7	$a_0b_1c_0d_1e_0$	01010	-1	1	-1	1	-1	$-+-+-$	bd
8	$a_1b_1c_1d_1e_0$	11110	1	1	1	1	-1	$++++-$	abcd
9	$a_0b_1c_1d_1e_1$	01111	-1	1	1	1	1	$-++++$	bcde
10	$a_1b_1c_0d_1e_1$	11011	1	1	-1	1	1	$++-++$	abde
11	$a_0b_0c_1d_1e_0$	00110	-1	-1	1	1	-1	$--++-$	cd
12	$a_0b_0c_0d_0e_0$	00000	-1	-1	-1	-1	-1	$-----$	(1)
13	$a_1b_0c_0d_0e_1$	10001	1	-1	-1	-1	1	$+---+$	ae
14	$a_1b_1c_1d_0e_1$	11101	1	1	1	-1	1	$+++-+$	abce
15	$a_1b_0c_0d_1e_0$	10010	1	-1	-1	1	-1	$+--+-$	ad
16	$a_0b_1c_1d_0e_0$	01100	-1	1	1	-1	-1	$-++--$	bc

A screening design focuses on main effects. We can get the main effect information out of the data by fitting the model $Y = X\beta + e$ where X is a matrix of numerical subscript values together with an initial column of 1s.

EXAMPLE 5.1.1. For Hare's experiment from Example 2.3.1, Y is one of the last two columns of Table 2.10 and the main effects only model X has a first column of 1s and then the rest of X consists of the treatment subscripts from Table 2.10 (Table 5.1), i.e.,

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

An equivalent but alternative method of writing the model matrix replaces the subscript 0 with the subscript -1,

$$\tilde{X} = \begin{bmatrix} +1 & -1 & -1 & -1 & +1 & +1 \\ +1 & +1 & -1 & +1 & +1 & +1 \\ +1 & +1 & +1 & -1 & -1 & -1 \\ +1 & +1 & -1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 & -1 & +1 \\ +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & -1 & +1 & -1 & +1 & -1 \\ +1 & +1 & +1 & +1 & +1 & -1 \\ +1 & -1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & -1 & +1 & +1 \\ +1 & -1 & -1 & +1 & +1 & -1 \\ +1 & -1 & -1 & -1 & -1 & -1 \\ +1 & +1 & -1 & -1 & -1 & +1 \\ +1 & +1 & +1 & +1 & -1 & +1 \\ +1 & +1 & -1 & -1 & +1 & -1 \\ +1 & -1 & +1 & +1 & -1 & -1 \end{bmatrix}, \tag{1}$$

in a model $Y = \tilde{X}\gamma + e$. The matrix $\tilde{X}$ has the useful mathematical property that $\tilde{X}'\tilde{X} = 16I_6$, which makes the linear model easy to analyze. In particular, the estimate of γ is $\hat{\gamma} = (1/16)\tilde{X}'Y$.

For analyzing the s_p data, fitting either the X or $\tilde{X}$ model gives.

Analysis of Variance for Hare's s_p

Source	*df*	*SS*	*MS*	*F*	*P*
Regression	5	1.0096	0.2019	1.44	0.292
Residual Error	10	1.4044	0.1404		
Total	15	2.4140			

Fitting the $\tilde{X}$ model gives

Table of Coefficients for Hare's s_p

Predictor	$\hat{\gamma}_k$	SE($\hat{\gamma}_k$)	t	P
Constant	1.22625	0.09369	13.09	0.000
A	0.07250	0.09369	0.77	0.457
B	0.04375	0.09369	0.47	0.651
C	0.01875	0.09369	0.20	0.845
D	−0.01875	0.09369	−0.20	0.845
E	−0.23500	0.09369	−2.51	0.031

Fitting the X model gives estimates and standard errors, other than the intercept, that are twice as large but gives the same t statistics and P values. Specifically, for $k = 1, \ldots, 5$, $\hat{\beta}_k = 2\hat{\gamma}_k$ and $\text{SE}(\hat{\beta}_k) = 2\text{SE}(\hat{\gamma}_k)$. For, say, factor E, the estimated change in going from the low treatment level e_0 to the high treatment level is $2(-0.235) = -0.470$ with a standard error of $2(0.09369) = 0.18738$.

Again regardless of the model, we can divide the *SSReg* into one degree of freedom for each main effect.

Source	df	SS
A	1	0.0841
B	1	0.0306
C	1	0.0056
D	1	0.0056
E	1	0.8836

These sums of squares, divided by the *MSE*, are equal to the square of the t statistics from the Table of Coefficients. Because of the special structure (orthogonality) of $\tilde{X}$, unlike standard regression problems, neither the sums of squares nor the t statistics change if you drop any other main effects out of the model.

From this analysis, factor E, the time waited before using the intermix, has a much larger sum of squares and a much larger t statistic than any of the other factors, so it would seem to be the most important factor. As we saw in Chapter 2, the design Hare used allows examination of all the two-factor interactions and we found that two-factor interactions are important in these data but our current focus is on screening designs for looking at just main effects. □

We mentioned earlier that, in theory, estimating all the main effects in a 2^5 factorial treatment structure requires only 6 observations and that a good practical design can be obtained using only 8. The last 5 columns of the following model matrix determines one good design for evaluating just the five main effects.

$$\tilde{X}_* = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & -1 & -1 & -1 \\ +1 & +1 & -1 & +1 & +1 & -1 \\ +1 & +1 & -1 & -1 & -1 & +1 \\ +1 & -1 & +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 & +1 & +1 \\ +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & -1 & -1 & -1 & +1 & -1 \end{bmatrix}. \tag{2}$$

The treatments corresponding to $\tilde{X}_*$ are

$$\begin{bmatrix} a_1b_1c_1d_1e_1 \\ a_1b_1c_{-1}d_{-1}e_{-1} \\ a_1b_{-1}c_1d_1e_{-1} \\ a_1b_{-1}c_{-1}d_{-1}e_1 \\ a_{-1}b_1c_1d_{-1}e_{-1} \\ a_{-1}b_1c_{-1}d_1e_1 \\ a_{-1}b_{-1}c_1d_{-1}e_1 \\ a_{-1}b_{-1}c_{-1}d_1e_{-1} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} a_1b_1c_1d_1e_1 \\ a_1b_1c_0d_0e_0 \\ a_1b_0c_1d_1e_0 \\ a_1b_0c_0d_0e_1 \\ a_0b_1c_1d_0e_0 \\ a_0b_1c_0d_1e_1 \\ a_0b_0c_1d_0e_1 \\ a_0b_0c_0d_1e_0 \end{bmatrix}.$$

In the next section we examine where such designs originate.

5.2 Theory for Designs at Two Levels

For experiments having all factors at two levels, Plackett and Burman (1946) proposed using normalized Hadamard matrices to define screening designs, i.e., groups of factorial treatments that provide nice estimates of the main effects for all factors. A *Hadamard matrix* H is an $n \times n$ square matrix that consists of the numbers ± 1 for which $\frac{1}{\sqrt{n}}H$ is an *orthonormal* (more often called *orthogonal*) *matrix*. In other words,

$$H'H = HH' = nI.$$

Recall that permuting either the rows or columns of an orthonormal matrix gives another orthonormal matrix.

The number of rows n in a Hadamard matrix needs to be 1, 2, or a multiple of 4 and even then it is not clear that Hadamard matrices always exist. One Hadamard matrix of order 8 is

$$H = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & -1 & -1 & -1 & -1 & +1 \\ +1 & +1 & -1 & +1 & +1 & -1 & -1 & -1 \\ +1 & +1 & -1 & -1 & -1 & +1 & +1 & -1 \\ +1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 \\ +1 & -1 & +1 & -1 & +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 & -1 & +1 & -1 & +1 \\ +1 & -1 & -1 & -1 & +1 & -1 & +1 & +1 \end{bmatrix}. \tag{1}$$

One Hadamard matrix of order 12 is

$$H = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 \\ +1 & -1 & -1 & +1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 \\ +1 & +1 & -1 & -1 & +1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 & +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 & +1 & +1 & -1 \\ +1 & -1 & -1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 & +1 & +1 \\ +1 & +1 & -1 & -1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 & +1 \\ +1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 & -1 & +1 & -1 \\ +1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 & -1 & +1 \\ +1 & +1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 & -1 \end{bmatrix}.$$

A *normalized* Hadamard matrix has the form

$$H = [J, T],$$

where J is a column of 1s. The submatrix T, or any subset of its columns, can be used to define treatments (treatment subscripts) for analysis of variance problems involving many factors each at two levels in which our interest lies only in main effects.

For example, if we have f factors each at two levels, an $n \equiv f + 1$ dimensional normalized Hadamard matrix H, if it exists, determines a set of treatments whose observation allows us to estimate all f of the main effects. Randomly associate each of the f factors with one of the f columns in T. T provides the subscripts associated with each treatment to be observed. Indeed, the normalized Hadamard matrix becomes $\tilde{X}$ in the linear model for main effects, $Y = \tilde{X}\gamma + e$. This is a smallest design that allows us to estimate the grand mean and all f of the factor main effects. But remember, Hadamard matrices do not exist for all values $f + 1$. Except for the trivial case of $f = 1$, normalized Hadamard matrixes only exist when $f + 1$ is a multiple of 4.

More often we have f factors and choose $n > f + 1$ as the size of the normalized Hadamard matrix H. Again, excluding the initial column of 1s, randomly associate each factor with one of the remaining $n - 1$ columns of H. From the Hadamard matrix, extract the matrix $\tilde{X} = [J, T]$ that consists of the column of 1s followed by (in any convenient order) the columns associated with the factors. Because $H'H =$

nI_n, we have $\tilde{X}'\tilde{X} = nI_{f+1}$ and the perpendicular projection operator onto $C(\tilde{X})$ is $\tilde{M} = \frac{1}{n}\tilde{X}\tilde{X}'$. T provides the subscripts associated with the treatments to be observed. Assuming no interactions, the model $Y = \tilde{X}\gamma + e$ involves n observations, provides $n - r(\tilde{X}) = n - (f+1) = n - f - 1$ degrees of freedom for estimating the error, as well as provides estimates of the f main effects and the intercept.

If we take $n >> f+1$, we should be able to do much more than merely examine main effects, e.g., examine at least some interactions. But in general, it is difficult to know what more we can do, i.e., what interactions we can look at. In Chapter 2 we examined in detail the special case of 2^{-s} replications of 2^f factorial structures These involve $n = 2^{f-s}$ dimensional Hadamard matrices. For screening main effects, we would want to choose s to get 2^{f-s} as close as possible to $f+1$. But as demonstrated in Chapter 2, this special case with $n = 2^{f-s}$ is extremely useful if we want to keep careful track of which interactions can be estimated and which cannot.

EXAMPLE 5.2.1. In Hare's example, the matrix $\tilde{X}$ in equation (5.1.1) defined the model matrix for the main-effects linear model and its last five columns defined the treatments used. $\tilde{X}$ consists of the first 6 columns of the following normalized Hadamard matrix:

$$\tilde{H} = \begin{bmatrix}
+1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 & -1 & -1 & +1 & -1 & -1 & -1 & -1 & +1 \\
+1 & +1 & -1 & +1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & +1 & +1 \\
+1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 & -1 & -1 & -1 & -1 & -1 & +1 & +1 & +1 \\
+1 & +1 & -1 & +1 & -1 & -1 & -1 & +1 & -1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\
+1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 & -1 & -1 & +1 & +1 & -1 & -1 \\
+1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 & +1 & -1 & -1 & +1 & -1 & -1 & +1 & -1 \\
+1 & -1 & +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 & -1 & +1 & -1 \\
+1 & +1 & +1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 & +1 & +1 & -1 & +1 & -1 & -1 \\
+1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\
+1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 & +1 & +1 & -1 & +1 & +1 & -1 & -1 & +1 \\
+1 & -1 & -1 & +1 & +1 & -1 & +1 & -1 & -1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 \\
+1 & -1 & -1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\
+1 & +1 & -1 & -1 & -1 & +1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 & +1 & -1 & -1 \\
+1 & +1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 & +1 & +1 & -1 & +1 & -1 & +1 & -1 \\
+1 & +1 & -1 & -1 & +1 & -1 & -1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 \\
+1 & -1 & +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 & +1
\end{bmatrix}.$$

□

If examining main effects was the only goal, Hare could have gotten by with examining only the 8 factor combinations defined by the last five columns of $\tilde{X}_*$ in equation (5.1.2). The matrix $\tilde{X}_*$ consists of the first six columns of the Hadamard matrix in equation (5.2.1).

If Hare had 7 factors each at two levels, a smallest (orthogonal) design for obtaining all main effects takes $\tilde{X}$ equal to the Hadamard matrix in (5.2.1) [or some other normalized Hadamard matrix of the same size]. Alternatively, Hare could have stuck with 16 treatments and used, say, columns 2 through 8 of $\tilde{H}$ to define the factor combinations. That would be a perfectly good design for looking only at main effects. But Hare was also interested in two-factor interactions and the 7th and 8th

columns of $\tilde{H}$ happen to be associated with the AB and AC interactions. (More on this later.) Using the 7th and 8th columns to help define treatments for 7 factors would mean loosing the ability to estimate the AB and AC interactions: estimating these interactions is something that Hare could do with only 5 factors but something that typically is not a priority in a screening design.

Permuting the rows of a Hadamard matrix gives another Hadamard matrix which is, for our purposes, equivalent to the first. The rows define the treatments we want, and permuting the rows does not change the collection of treatments. Also, permuting the rows does not change that $\tilde{X}'\tilde{X} = nI_{f+1}$. We just have to make sure that we apply the same permutation to the rows of Y. We could also permute the columns of either H or $\tilde{X}$, as long as we remember what factor is associated with each column.

5.2.1 Blocking

Returning to Hare's experiment with 5 factors and 16 observations (factor combinations), suppose Hare had wanted to run the experiment in four blocks of size 4. The first 6 columns of $\tilde{H}$ define the intercept and treatments, any other two columns of $\tilde{H}$ could be used to define the four blocks of size 4. Let's use the last two columns to define blocks. The last two columns define pairs of numbers $(1,1)$, $(1,-1)$, $(-1,1)$, $(-1,-1)$ that will define the blocks.

Deleting the columns that we are not using and rearranging the rows of $\tilde{H}$ so that the pairs of numbers from the last two columns are grouped, and introducing some extra space to focus on the five columns that define the treatments, gives

$$\begin{bmatrix}
+1 & +1 & -1 & +1 & +1 & +1 & +1 & +1 \\
+1 & +1 & +1 & -1 & -1 & -1 & +1 & +1 \\
+1 & -1 & -1 & -1 & -1 & -1 & +1 & +1 \\
+1 & -1 & +1 & +1 & +1 & +1 & +1 & +1 \\
+1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 \\
+1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 \\
+1 & +1 & +1 & +1 & -1 & +1 & +1 & -1 \\
+1 & +1 & -1 & -1 & +1 & -1 & +1 & -1 \\
+1 & -1 & +1 & +1 & -1 & -1 & -1 & +1 \\
+1 & -1 & -1 & -1 & +1 & +1 & -1 & +1 \\
+1 & +1 & -1 & +1 & -1 & -1 & -1 & +1 \\
+1 & +1 & +1 & -1 & +1 & +1 & -1 & +1 \\
+1 & -1 & +1 & -1 & -1 & +1 & -1 & -1 \\
+1 & +1 & -1 & -1 & -1 & +1 & -1 & -1 \\
+1 & +1 & +1 & +1 & +1 & -1 & -1 & -1 \\
+1 & -1 & -1 & +1 & +1 & -1 & -1 & -1
\end{bmatrix}. \tag{2}$$

We can read off the blocking structure from this matrix. Block one consists of the the treatments (using ± 1 subscripts) corresponding to rows of $\tilde{H}$ in which the last two columns are (1,1). The subscripts come from the first four rows of the previous

matrix,

$$\begin{bmatrix} +1 & -1 & +1 & +1 & +1 \\ +1 & +1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 \\ -1 & +1 & +1 & +1 & +1 \end{bmatrix},$$

so changing the -1 subscripts back to 0s, the treatments are

$$\begin{bmatrix} a_1b_0c_1d_1e_1 \\ a_1b_1c_0d_0e_0 \\ a_0b_0c_0d_0e_0 \\ a_0b_1c_1d_1e_1 \end{bmatrix}.$$

In the second block the treatment subscripts correspond to rows of $\tilde{H}$ where the last two columns are $(1,-1)$:

$$\begin{bmatrix} -1 & -1 & +1 & -1 & +1 \\ -1 & +1 & -1 & +1 & -1 \\ +1 & +1 & +1 & -1 & +1 \\ +1 & -1 & -1 & +1 & -1 \end{bmatrix}.$$

The third block has $(-1,1)$ in the last two columns,

$$\begin{bmatrix} -1 & +1 & +1 & -1 & -1 \\ -1 & -1 & -1 & +1 & +1 \\ +1 & -1 & +1 & -1 & -1 \\ +1 & +1 & -1 & +1 & +1 \end{bmatrix}.$$

And the last block has $(-1,-1)$,

$$\begin{bmatrix} -1 & +1 & -1 & -1 & +1 \\ +1 & -1 & -1 & -1 & +1 \\ +1 & +1 & +1 & +1 & -1 \\ -1 & -1 & +1 & +1 & -1 \end{bmatrix}.$$

Thus the other three blocks are

$$\begin{bmatrix} a_0b_0c_1d_0e_1 \\ a_0b_1c_0d_1e_0 \\ a_1b_1c_1d_0e_1 \\ a_1b_0c_0d_1e_0 \end{bmatrix}, \begin{bmatrix} a_0b_1c_1d_0e_0 \\ a_0b_0c_0d_1e_1 \\ a_1b_0c_1d_0e_0 \\ a_1b_1c_0d_1e_1 \end{bmatrix}, \begin{bmatrix} a_0b_1c_0d_0e_1 \\ a_1b_0c_0d_0e_1 \\ a_1b_1c_1d_1e_0 \\ a_0b_0c_1d_1e_0 \end{bmatrix}.$$

Exercise 5.1 Show that these blocks have CE and DE confounded with them. What else is confounded with them? (There is another defining effect and a number of aliases.)

The model matrix for the main effects with blocking model can be taken as

$$\tilde{X} = \begin{bmatrix} +1 & 1 & 0 & 0 & +1 & -1 & +1 & +1 & +1 \\ +1 & 1 & 0 & 0 & +1 & +1 & -1 & -1 & -1 \\ +1 & 1 & 0 & 0 & -1 & -1 & -1 & -1 & -1 \\ +1 & 1 & 0 & 0 & -1 & +1 & +1 & +1 & +1 \\ +1 & 0 & 1 & 0 & -1 & -1 & +1 & -1 & +1 \\ +1 & 0 & 1 & 0 & -1 & +1 & -1 & +1 & -1 \\ +1 & 0 & 1 & 0 & +1 & +1 & +1 & -1 & +1 \\ +1 & 0 & 1 & 0 & +1 & -1 & -1 & +1 & -1 \\ +1 & 0 & 0 & 1 & -1 & +1 & +1 & -1 & -1 \\ +1 & 0 & 0 & 1 & -1 & -1 & -1 & +1 & +1 \\ +1 & 0 & 0 & 1 & +1 & -1 & +1 & -1 & -1 \\ +1 & 0 & 0 & 1 & +1 & +1 & -1 & +1 & +1 \\ +1 & 0 & 0 & 0 & -1 & +1 & -1 & -1 & +1 \\ +1 & 0 & 0 & 0 & +1 & -1 & -1 & -1 & +1 \\ +1 & 0 & 0 & 0 & +1 & +1 & +1 & +1 & -1 \\ +1 & 0 & 0 & 0 & -1 & -1 & +1 & +1 & -1 \end{bmatrix}$$

Here the second through fourth columns account for blocks and the last five columns account for factors A, B, C, D, E. The order of listing the treatments has changed from Table 5.1, so the row order of listing the variability measures s_c and s_p would also need to change. The physical act of blocking would almost certainly change the data from that observed in Table 5.1, but if the data from the blocked experiment were the same as those reported, the estimates and sums of squares for the main effects would also remain the same. Blocking should change the estimate of error. The whole point of blocking is to isolate substantial effects due to blocks and remove them from the error that would have occurred without blocking. In Hare's experiment, it takes a day to run one treatment, so blocks might be run in different weeks or even different months.

If we wanted two blocks of size 8, we would have used only one column (not previously used for treatments or the intercept) of the Hadamard matrix $\tilde{H}$ to define blocks. If we wanted 8 blocks of size 2, we would have used three (not previously used) columns to define blocks.

Now let's examine blocking with $f = 5$ factors and $n = 8$ factor combinations. In this example, the last two columns of the Hadamard matrix in equation (5.2.1) are used to define 4 blocks of size 2. Rearranging the rows of (5.2.1) gives

$$\begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & -1 & -1 & -1 & +1 & -1 & +1 & +1 \\ +1 & +1 & -1 & -1 & -1 & +1 & +1 & -1 \\ +1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 \\ +1 & -1 & -1 & +1 & -1 & +1 & -1 & +1 \\ +1 & +1 & +1 & -1 & -1 & -1 & -1 & +1 \\ +1 & +1 & -1 & +1 & +1 & -1 & -1 & -1 \\ +1 & -1 & +1 & -1 & +1 & +1 & -1 & -1 \end{bmatrix}$$

from which we gets the blocks

$$\begin{bmatrix} +1 & +1 & +1 & +1 & +1 \\ -1 & -1 & -1 & +1 & -1 \end{bmatrix},$$

$$\begin{bmatrix} +1 & -1 & -1 & -1 & +1 \\ -1 & +1 & +1 & -1 & -1 \end{bmatrix},$$

$$\begin{bmatrix} -1 & -1 & +1 & -1 & +1 \\ +1 & +1 & -1 & -1 & -1 \end{bmatrix},$$

$$\begin{bmatrix} +1 & -1 & +1 & +1 & -1 \\ -1 & +1 & -1 & +1 & +1 \end{bmatrix}.$$

Unfortunately, using these blocks will lose us the ability to look at the main effect for factor D because D is at the same level in every block. There are 8 observations, so 8 degrees of freedom. There are 4 degrees of freedom for the blocks and the intercept, which leaves only 4 degrees of freedom for estimating effects, but we have 5 main effects to estimate, so we must lose something.

Using a Hadamard matrix to determine blocks can be done without worrying about which (interaction) effects are being confounded with blocks (if you think that is a good thing). A major virtue of the approach in Chapter 2 is that it allows us to keep track of such things. The major virtue of arbitrary Hadamard matrices is that you are not restricted to taking n as a power of 2.

5.2.2 Hare's Full model

The Hadamard matrix $\tilde{H}$ was (implicitly) used by Hare to determine a group of 16 treatments to examine from a 2^5 factorial structure; treatments that provide a clean analysis of main effects. From our discussion in this chapter, $\tilde{H}$ could have been used to define a design and a main-effects model for up to 15 factors.

The normalized Hadamard matrix $\tilde{H}$ was actually constructed using the methods of Chapter 2 for a 2^5. This allows us to identify each of the 10 columns of the matrix not used in the main-effects model with a particular two-factor interaction. Fitting the linear model $Y = \tilde{H}\delta + e$ fits the data perfectly, leaving 0 degrees of freedom for error, and leading to the analysis of Chapter 2.

The trick, in using this model, is identifying what effect each column represents. The first 6 columns correspond to the intercept and the factor main effects, the other 10 columns are two-factor interactions and are obtained by multiplying two main-effects columns elementwise. In other words, column 7 is the AB interaction column because it is obtained from multiplying column 2 (factor A) times column 3 (factor B) elementwise. In the first row, columns 2 and 3 take the values -1 and -1, so column 7 is $-1 \times -1 = 1$. In the second row, columns 2 and 3 take the values 1 and -1, so column 7 is $1 \times -1 = -1$. In the last row, columns 2 and 3 are -1 and 1, so column 7 is $-1 \times 1 = -1$. There are 10 distinct pairs of main effects, hence 10 two-factor intereactions. Many ANOVA and regression programs have this method, or an equivalent process, automated. Incidentally, in Chapter 2 we saw that the AB

interaction effect is indistinguishable from the *CDE* interaction effect. Note that column 7 (*AB*) is -1 times the elementwise product of columns 4, 5, and 6; the columns associated with *C*, *D*, and *E*.

Many computer programs (*not* R's `lm`) disallow fitting models with 0 dfE, so we deleted the last column of $\tilde{H}$ before fitting the model. The last column corresponds to the *DE* interaction.

Analysis of Variance for s_p

Source	*df*	*SS*	*MS*	*F*	*P*
Regression	14	2.0171	0.1441	0.36	0.881
Residual Error	1	0.3969	0.3969		
Total	15	2.4140			

The sums of squares can be broken down into 15 individual terms associated with main effects and two-factor interactions. These numbers are just the elements of the last 15 terms of the vector $\tilde{H}'Y$, squared, and divided by $n = 16$. (We have ignored the contribution from the intercept.) Again, the Error term is labeled as *DE*.

Source	df	SS	Source	df	SS	Source	df	SS
A	1	0.0841	*AB*	1	0.0009	*BD*	1	0.1056
B	1	0.0306	*AC*	1	0.0361	*BE*	1	0.6561
C	1	0.0056	*AD*	1	0.0036	*CD*	1	0.0210
D	1	0.0056	*AE*	1	0.0930	*CE*	1	0.0729
E	1	0.8836	*BC*	1	0.0182	*DE*	1	0.3969

In Chapter 2 we saw that all of these terms are indistinguishable from higher-order interaction terms. In the main-effects only analysis of Section 1, we noted that *E* looked to be the most important effect. While that remains true in this more expansive model, the sum of squares for *BE* is of a similar size to that of *E* and the sum of squares for *DE* is not inconsiderable.

5.2.3 Construction of Hadamard Matrices

Hadmard matrices have very nice design properties but you have to be able to find them. In practice, you just look them up. But there are a variety of ways to construct Hadamards. For an $n \times n$ Hadamard to exist, n has to be 1, 2, or a multiple of 4.

For $n = 1$, $H = \pm 1$. For $n = 2$, some Hadamards are

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Given any two Hadamard matrices, it is easy to see that their *kronecker product* is Hadamard:

$$\begin{aligned}[H_1 \otimes H_2][H_1 \otimes H_2]' &= [H_1 \otimes H_2][H_1' \otimes H_2'] \\ &= [H_1 H_1' \otimes H_1 H_2'] = [n_1 I_{n_1} \otimes n_2 I_{n_2}] = n_1 n_2 I_{n_1 n_2}.\end{aligned}$$

Paley's method of constructing Hadamards uses Jacobsthal matrices. A $q \times q$ *Jacobsthal matrix* Q has 0s on the diagonal, ± 1s elsewhere, and has the properties: (a) $QQ' = qI_q - J_q^q$ and (b) $QJ_q = J_q'Q = 0$. From (a) $\frac{1}{q}QQ'$ is the *ppo (perpendicular projection operator)* onto $C(J_q)^\perp$, so the property $J_q'Q = 0$ must be true. Clearly, if Q is Jacobsthal, $-Q$ is also.

If $q \bmod 4 = 3$, Q will be skew symmetric, i.e. $-Q = Q'$. In that case, an $n = q+1$ dimensional Hadamard matrix can be obtained from

$$H = \begin{bmatrix} 1 & J_q' \\ J_q & Q - I_q \end{bmatrix} \qquad \text{or} \qquad H = \begin{bmatrix} 1 & J_q' \\ -J_q & -Q + I_q \end{bmatrix}$$

or

$$H = I_n + \begin{bmatrix} 0 & J_q' \\ -J_q & Q \end{bmatrix} = \begin{bmatrix} 1 & J_q' \\ -J_q & Q + I_q \end{bmatrix}.$$

If $q \bmod 4 = 1$, Q will be symmetric and Hadamards of dimension $n = 2(q+1)$ can be constructed by replacing elements of

$$\begin{bmatrix} 0 & J_q' \\ J_q & Q \end{bmatrix}$$

by certain 2 dimensional Hadamards.

Finally, you can construct Hadamards from *conference matrices*. (In the next section we will find a different use for conference matrices.) A $q \times q$ conference matrix has 0 on the diagonal, ± 1 elsewhere, and $C'C = (q-1)I$. Examples of conference matrices are

$$C_1 = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & -1 \\ 1 & -1 & 0 & 1 \\ 1 & 1 & -1 & 0 \end{bmatrix}, \quad C_2 = \begin{bmatrix} 0 & -1 & -1 & -1 \\ 1 & 0 & -1 & 1 \\ 1 & 1 & 0 & -1 \\ 1 & -1 & 1 & 0 \end{bmatrix}, \tag{3}$$

and

$$C_3 = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & -1 & -1 & 1 \\ 1 & 1 & 0 & 1 & -1 & -1 \\ 1 & -1 & 1 & 0 & 1 & -1 \\ 1 & -1 & -1 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 & 1 & 0 \end{bmatrix}. \tag{4}$$

C_2 is skew symmetric and C_3 is symmetric.

If you think about what it takes for the columns of a conference matrix to be orthogonal, it is pretty easy to see that the dimension q of a conference matrix has to be an even number. Skew symmetric conference matrices have q a multiple of 4, with the other even numbers giving symmetric conference matrices if they exist. For example, conference matrices are not known to exist for $q = 22, 34$.

If C is skew symmetric, a Hadamard matrix is $H = I + C$. If C symmetric,

$$H = \begin{bmatrix} C+I & C-I \\ C-I & -C-I \end{bmatrix}$$

is Hadamard.

We are not going to consider how one goes about constructing either Jacobsthal and conference matrices.

5.3 Designs for Three Levels

Screening designs are used as a first step in identifying important factors. They rely on the hope that any interactions that exist will not mask the important main effects.

Suppose we have f factors each at 3 levels. As before, we use capital letters to denote factors and small letters to denote levels. With three levels, we have two useful subscripting options for identifying levels. We can identify the levels of factor A as a_0, a_1, a_2 or as a_{-1}, a_0, a_1. The first subscripting option was used in Chapters 3 and 4 (because it facilitates modular arithmetic). The second option is used here.

Screening designs focus on main effects but now a main effect has 2 degrees of freedom. In this section, we will assume that the three levels are quantitative levels, which means that we can associate the 2 main effect degrees of freedom with one linear term and one quadratic term. Moreover, we assume that the levels are equally spaced, so that the actual levels might just as well be the subscripts $-1, 0, 1$. Looking at the linear effect involves comparing the treatments with the 1 subscript to treatments with the -1 subscript. Often the linear effect is considered more important than the quadratic effect. In a comparable two-factor screening design, the linear effects *are* the main effects. We will consider designs that focus on the linear effect but also retain the ability to estimate the quadratic effect.

In Chapter 3 we discussed fractional replications of 3^f factorial structures. It is often possible to create designs that allow independent estimation of all main effects that involve observing $3^2 = 9$, or $3^3 = 27$, or $3^4 = 81$ factor combinations. The designs in the next subsection offer more flexibility in terms of numbers of observations but lose the independence of quadratic effects.

5.3.1 Definitive Screening Designs

As discussed in the previous section, a $q \times q$ conference matrix C has 0 on the diagonal, ± 1 elsewhere, and $C'C = (q-1)I$. As such, every row of a conference matrix can be identified with a factor combination.

Jones and Nachtsheim (2011) introduced *definitive screening designs (DSDs)* that allow one to examine both the linear and quadratic main effects efficiently. The treatments are defined by the matrix

$$T = \begin{bmatrix} C \\ -C \\ 0 \end{bmatrix},$$

where 0 indicates a $1 \times q$ row vector of 0s. Again, the rows of T consist of the subscripts for the factor combinations to be observed. The number of observations is obviously $n = 2q+1$. For example, using C_2 from equation (5.2.3), the DSD is defined by

$$T = \begin{bmatrix} 0 & -1 & -1 & -1 \\ 1 & 0 & -1 & 1 \\ 1 & 1 & 0 & -1 \\ 1 & -1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ -1 & 0 & 1 & -1 \\ -1 & -1 & 0 & 1 \\ -1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

This determines the treatments

$$\begin{bmatrix} a_0b_{-1}c_{-1}d_{-1} \\ a_1b_0c_{-1}d_1 \\ a_1b_1c_0d_{-1} \\ a_1b_{-1}c_1d_0 \\ a_0b_1c_1d_1 \\ a_{-1}b_0c_1d_{-1} \\ a_{-1}b_{-1}c_0d_1 \\ a_{-1}b_1c_{-1}d_0 \\ a_0b_0c_0d_0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} a_1b_0c_0d_0 \\ a_2b_1c_0d_2 \\ a_2b_2c_1d_0 \\ a_2b_0c_2d_1 \\ a_1b_2c_2d_2 \\ a_0b_1c_2d_0 \\ a_0b_0c_1d_2 \\ a_0b_2c_0d_1 \\ a_1b_1c_1d_1 \end{bmatrix}.$$

Relating these back to Chapter 3, note that the 0,1,2 treatments happen to be solutions to the subscript equations

$$(x_2+x_3+x_4) \bmod = 0; \quad \text{and} \quad (x_1+0x_2+2x_3+x_4) \bmod = 1,$$

so this happens to be a 3^{4-2} design with BCD and AC^2D as defining effects, hence also ABD^2 and AB^2C. Most DSDs are not examples of the ideas in Chapter 3. Most of them do not have n as a power of 3.

Each of the confounding effects lacks one of the 4 factors, so that determines the structure of this design if we dropped a factor. For example, if we use this design only on factors A, B, and C, it would be a $1/3$ rep of a 3^3 that confounds AB^2C. Dropping C rather than D or dropping B rather than D works similarly. If we use this design only on factors B, C, and D, it would be a $1/3$ rep of a 3^3 that confounds BCD.

A linear-main-effects-only model is $Y = [J,T]\beta + e$. A model with all main effects is $Y = [J,T,T^{(2)}]\gamma + e$ wherein, if $T \equiv [t_{ij}]_{n\times q}$, then $T^{(2)} \equiv [t_{ij}^2]$. The definitive screening design for q factors has $n = 2q+1$ observations with n effects to esti-

mate, i.e., an intercept, q linear main effects, and q quadratic main effects. So the main-effects model is a saturated model with 0 degrees of freedom for Error. As discussed earlier, conference matrices have to have an even numbered dimension, so the available sizes for DSDs are $n = 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 49, \ldots$, i.e., one more than a multiple of 4. You might expect $n = 45$ in this list but recall that there is no conference matrix (known) for $q = 22$.

One way to avoid having a saturated main-effects model is to use a DSD based on larger order conference matrices. For $f = 5$ factors, the following matrix uses the DSD design based on C_3, the 6 dimensional conference matrix given in (5.2.4), but deleting the fourth column.

$$T = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & -1 & 1 \\ 1 & 1 & 0 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 0 & 1 \\ 1 & 1 & -1 & 1 & 0 \\ 0 & -1 & -1 & -1 & -1 \\ -1 & 0 & -1 & 1 & -1 \\ -1 & -1 & 0 & 1 & 1 \\ -1 & 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & 0 & -1 \\ -1 & -1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

This design will provide 2 degrees of freedom for error, or for looking at interactions.

EXAMPLE 5.3.1. *It would be nice to have a real data example here.*

5.3.2 Some Linear Model Theory

Definitive screening designs seem to be the nicest of a class of designs for three factors based on weighing matrices. A q dimensional *weighing matrix* W takes values $-1, 0, 1$ and has $W'W = wI = WW'$ for some w. Hadamard and Conference matrices are weighing with respective ws of q and $q-1$. The number of 0s in each column and row of W is $q - w$. The class of designs we will consider is

$$T = \begin{bmatrix} W \\ -W \\ 0 \end{bmatrix},$$

so they are a generalization of definitive screening designs. Hadamard matrices have estimability problems when used in three-factor designs. In any case, the number of

observations in these designs is

$$n = 2q + 1.$$

To use these designs with $f < q$ just choose a selection submatrix of I_q, say $S_{q\times f}$, that throws out $q - f$ columns of I_q and replace T with TS.

Write T in terms of its columns, rows and elements as

$$T = [T_1, \ldots, T_q] = \begin{bmatrix} t_1' \\ \vdots \\ t_n' \end{bmatrix} = [t_{ij}].$$

Also define $T^{(2)}$ to be the matrix consisting of the squares of the elements in T, i.e.,

$$T^{(2)} \equiv [t_{ij}^2].$$

The quadratic-main-effects linear model associated with this design is

$$Y = [J, T, T^{(2)}] \begin{bmatrix} \delta_0 \\ \delta_1 \\ \delta_2 \end{bmatrix} + e,$$

where δ_0 is a scalar with δ_1 and δ_2 being q vectors. After centering $T^{(2)}$, we get the equivalent linear model

$$Y = \left[J, T, \left(I - \frac{1}{n}JJ'\right)T^{(2)}\right] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + e. \tag{1}$$

Note that for $f < q$, the quadratic-main-effects model is just

$$Y = [J, TS, T^{(2)}S] \begin{bmatrix} \delta_0 \\ \delta_1 \\ \delta_2 \end{bmatrix} + e,$$

where δ_1 and δ_2 are now f vectors.

We will see that the basic nature of these designs makes the linear effects orthogonal to the quadratic effects and the intercept. The use of a weighing matrix makes the linear effects orthogonal with each other. Typically the quadratic effects are mildly correlated with each other. We will find the correlation structure of the quadratic effects for a definitive screening design and discuss the difficulties of doing that for general designs based on weighing matrices.

Note that, because all entries in T are $-1, 0, 1$, any column of $T^{(2)}$, say $T_k^{(2)}$, has the property that $J'T_k^{(2)} = T_k'T_k$ and this equals n minus the number of 0s in T_k, i.e. $2w$. In particular, for these designs

$$T^{(2)} \equiv \begin{bmatrix} W^{(2)} \\ W^{(2)} \\ 0 \end{bmatrix}$$

and in particular for the definitive screening design, since the only 0s in a conference matrix are on the diagonal,

$$T^{(2)} = \begin{bmatrix} J_q J_q' - I_q \\ J_q J_q' - I_q \\ 0 \end{bmatrix} = \begin{bmatrix} J_q J_q' \\ J_q J_q' \\ J_q' \end{bmatrix} - \begin{bmatrix} I_q \\ I_q \\ J_q' \end{bmatrix} = J_n J_q' - \begin{bmatrix} I_q \\ I_q \\ J_q' \end{bmatrix}. \tag{2}$$

For a general treatment matrix T consisting of -1s, 0s, and 1s, to estimate all of the parameters in model (1) we need the columns of the model matrix to be linearly independent. Linear independence requires $n \geq 2f+1$ but can break down even if that condition is satisfied. In general, to estimate all of the effects, we need T to satisfy the following properties.

(i) There must be at least one -1, 0, and 1 in each column of T.
(ii) For every ordered pair (j,k), $j,k = 1,\ldots,f$, $j \neq k$, there exists an i such that $t_{ij} = 0$ and $t_{ik} \neq 0$.

If (i) fails, either the missing level in T_k is 0, in which case $T_k^{(2)} = J$, or the missing level is ± 1, in which case $T_k^{(2)} \propto T_k$. If (ii) fails for both (j,k) and (k,j), $T_j^{(2)} = T_k^{(2)}$, so we cannot estimate both quadratic main effects.

The nice analysis properties of these designs follows from the fact that the submatrices of the model matrix in (1) are all orthogonal. In particular,

$$\left[J, T, (I - \frac{1}{n}JJ')T^{(2)}\right]' \left[J, T, (I - \frac{1}{n}JJ')T^{(2)}\right] = \begin{bmatrix} n & 0 & 0 \\ 0 & 2wI & 0 \\ 0 & 0 & T^{(2)}\left(I - \frac{1}{n}JJ'\right)T^{(2)} \end{bmatrix}. \tag{3}$$

To have a truly orthogonal design, we need the matrix in (2) to be diagonal, not just block diagonal. Even without W being a weighing matrix, the design structure would make the matrix block diagonal, but with W being weighing, the matrix is almost diagonal. In particular, the block diagonal structure facilitates finding the matrix inverse, which is more important to the analysis than the original matrix. We now verify equation (3).

Clearly, $J'J = n$. From the definition of T in terms of W, $J_n'T = J_q'W - J_q'W + 0 = 0_{1\times q}$, and $J'(I - \frac{1}{n}JJ')T^{(2)} = 0$. Also

$$T'T = \begin{bmatrix} W \\ -W \\ 0 \end{bmatrix}' \begin{bmatrix} W \\ -W \\ 0 \end{bmatrix} = 2W'W = 2wI_q,$$

where only the last equality depends on W being weighing and, since $J_n'T = 0$,

$$T'(I - \frac{1}{n}JJ')T^{(2)} = T'T^{(2)} = \begin{bmatrix} W \\ -W \\ 0 \end{bmatrix}' \begin{bmatrix} W^{(2)} \\ W^{(2)} \\ 0 \end{bmatrix} = W'W^{(2)} - W'W^{(2)} = 0. \quad (4)$$

So we have established that the matrix is diagonal except for the last block. The more difficult task is to find

$$T^{(2)\prime}(I - \frac{1}{n}JJ')T^{(2)} = T^{(2)\prime}T^{(2)} - \left[\frac{1}{n}JJ'T^{(2)}\right]' \left[\frac{1}{n}JJ'T^{(2)}\right].$$

First

$$\left[\frac{1}{n}JJ'T^{(2)}\right] = \frac{2w}{n}J_nJ_q',$$

because w is the number of ± 1s in a column of of W. It follows that

$$\left[\frac{1}{n}J_nJ_n'T^{(2)}\right]' \left[\frac{1}{n}J_nJ_n'T^{(2)}\right] = \frac{(2w)^2}{n}J_qJ_q'.$$

Finding $T^{(2)\prime}T^{(2)}$ seems like it needs to be done for every different weighing matrix. It depends on how the 0s in one column match up with the 0s in other columns.

For a definitive screening design,

$$T^{(2)\prime}T^{(2)} = \left(J_nJ_q' - \begin{bmatrix} I_q \\ I_q \\ J_q' \end{bmatrix}\right)' \left(J_nJ_q' - \begin{bmatrix} I_q \\ I_q \\ J_q' \end{bmatrix}\right) = 2I_q + (n+1-6)J_qJ_q'$$

and, as above with $w = q-1$,

$$\left[\frac{1}{n}J_nJ_q'T^{(2)}\right]' \left[\frac{1}{n}J_nJ_q'T^{(2)}\right] = \frac{[(2(q-1)]^2}{2q+1}J_qJ_q',$$

so

$$T^{(2)\prime}\left(I - \frac{1}{n}JJ'\right)T^{(2)} = 2I_q + \frac{2(q-4)}{2q+1}J_q^q = 2\left[I_q + \frac{q(q-4)}{2q+1}(1/q)J_q^q\right].$$

Since $(1/q)J_q^q$ is a ppo, $2\left[I + \frac{q(q-4)}{2q+1}(1/q)J_q^q\right]$ is easily inverted using Christensen (2011, Proposition 12.11.1), i.e., for an idempotent matrix P,

$$[aI + bP]^{-1} = \frac{1}{a}\left[I - \frac{b}{a+b}P\right].$$

Thus

$$\left[T^{(2)\prime}\left(I - \frac{1}{n}JJ'\right)T^{(2)}\right]^{-1} = \frac{1}{2}\left[I_q - \frac{q-4}{(q-1)^2}J_q^q\right].$$

Unless $q = 4$ so that $n = 9$, the quadratic effects will not be orthogonal. This seems to be the only (useful) case in which a DSD is also a 3^f fractional replication as discussed in Chapter 3. Jones and Nachtsheim (2011) indicate that the common correlation among quadratic terms is

$$\frac{1}{3} - \frac{1}{q-1} = \frac{q-4}{3(q-1)}$$

but I have not been able to reproduce that result.

For $f < q$, the covariance matrix for the quadratic coefficients is

$$\begin{aligned}\sigma^2 \left[S' T^{(2)\prime} \left(I - \frac{1}{n} JJ' \right) T^{(2)} S \right]^{-1} &= \frac{\sigma^2}{2} \left[I_f + \frac{f(q-4)}{2q+1} (1/f) J_f^f \right]^{-1} \\ &= \frac{\sigma^2}{2} \left[I_f - \frac{q-4}{(2q+1)+f(q-4)} J_f^f \right].\end{aligned}$$

In our example from the previous subsection with $q = 6$ and $f = 5$, the correlation between two squared terms is

$$\frac{-2/(13+10)}{1-2/(13+10)} = \frac{-2}{21}.$$

In equation (4) we showed that the linear main effects are orthogonal to their square terms. In fact, we can show that the linear main effects are orthogonal, not only to the square terms, but also to linear by linear interaction terms. Let $(T_k T_j)$ denote the vector that results from multiplying the vectors T_k and T_j elementwise. In particular, $(T_k T_k) = T_k^{(2)}$. To estimate the linear by linear interaction term between factors k and j, one simply includes the vector $(T_k T_j)$ as a column of the model matrix. Orthogonality follows from the fact that,

$$\begin{aligned}T_s'(T_k T_j) = \sum_{i=1}^{n} t_{is} t_{ik} t_{ij} = \sum_{i=1}^{2q} t_{is} t_{ik} t_{ij} = \sum_{i=1}^{q} t_{is} t_{ik} t_{ij} + \sum_{i=q+1}^{2q} t_{is} t_{ik} t_{ij} \\ = \sum_{i=1}^{q} t_{is} t_{ik} t_{ij} + \sum_{i=1}^{q} (-t_{is})(-t_{ik})(-t_{ij}) = \sum_{i=1}^{q} t_{is} t_{ik} t_{ij} - \sum_{i=1}^{q} t_{is} t_{ik} t_{ij} = 0.\end{aligned}$$

A similar argument shows that $T_s'(T_k^{(2)} T_j^{(2)}) = 0$, so the linear main effect terms are also orthogonal to the quadratic by quadratic interaction terms. Without additional conditions on W, the linear main effects do not appear to be orthogonal to linear by quadratic or quadratic by linear interactions.

Choices for W other than conference matrices are less promising. If W is a Hadamard matrix, neither of the estimability conditions (i) nor (ii) hold. In fact, with a Hadamard matrix we would get only 1 degree of freedom for estimating all q of the quadratic effects. In the next subsection, we also construct a weighting matrix W, with two 0s in each row and column, which aliases pairs of quadratic effects.

5.3.3 Weighing out of conference

Hadamard matrices are weighing but give

$$T^{(2)} = \begin{bmatrix} J_{2q}^f \\ 0 \end{bmatrix}, \quad \left(I - \frac{1}{n}JJ'\right) T^{(2)} = \begin{bmatrix} \frac{1}{n}J_{2q}^f \\ \frac{-(n-1)}{n}J_1^f \end{bmatrix},$$

both of which are rank 1 matrices, hence our earlier comment about only 1 degree of freedom for estimating quadratic effects, and

$$T^{(2)\prime}\left(I - \frac{1}{n}JJ'\right) T^{(2)} = \frac{n-1}{n}J_f^f.$$

Next we look at a weighing matrix for which $w = q-2$ and $T^{(2)\prime}\left(I - \frac{1}{n}JJ'\right) T^{(2)}$ is again singular and incapable of estimating all of the quadratic main effects. Consider

$$W_1 = \begin{bmatrix} 0 & 0 & -1 & -1 & -1 & -1 & -1 & -1 \\ 0 & 0 & -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 0 & 0 & -1 & -1 & 1 & 1 \\ 1 & -1 & 0 & 0 & -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 0 & 0 & -1 & -1 \\ 1 & -1 & 1 & -1 & 0 & 0 & -1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & 0 & 0 \\ 1 & -1 & -1 & 1 & 1 & -1 & 0 & 0 \end{bmatrix}.$$

This skew symmetric W_1 was constructed as $[C_2 \otimes H_0]$ where C_2 was the skew symmetric conference matrix given near the end of Section 2 and

$$H_0 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

is a symmetric Hadamard matrix. The problem is, if we square the elements of W_1 to get $W_1^{(2)} \equiv [w_{1ij}^2]$, unlike a conference matrix, the consecutive pairs of vectors become identical. Thus we could not be able to tell apart the quadratic terms for factors A and B, or for factors C and D, etc. In particular, using W_1 as W in our design T, causes a violation of condition (ii) for estimating quadratic effects, with $T^{(2)}$ not having full rank and $T^{(2)\prime}\left(I - \frac{1}{n}JJ'\right) T^{(2)}$ being singular. We could solve this problem by using every other column of W in W, but at that point we might just as well save ourselves some observations and just use C_2 to define the design.

The multiplier w associated with a weighing matrix needs to be q minus the common number of 0s in each row. (The jth diagonal element of WW' is the number of nonzero elements in the jth row.) Mathematically, Hadamard matrices are weighing matrices with $w = q$ and conference matrices are weighing matrices with $w = q-1$. Conference matrices C are required to have 0's down the diagonal. Permuting the rows of C gives a weighting matrix with one 0 in each row and column. For design purposes, a conference matrix and a weighing matrix with one 0 per

row are equivalent. What is relevant to the design is the number of 0s in each row (or column) of W. As with Hadamard and conference matrices, there are serious mathematical questions about when weighing matrices exist, cf. Koukouvinos and Seberry (1997). In particular, it would be interesting to know if any weighing matrices exist with $w < (q-1)$ that satisfy property (ii) and allow estimation of all quadratic effects. While perhaps not terribly practical, it would be of theoretical interest to examine the relationship between a possible DSD with $n = 81 = 2(20)+1$ and the $n = 81 = 3^{20-16}$ fractional replication associated with Chapter 3. It would also be nice to establish similar results for DSDs with $n = 3^3 = 27$ observations, but unfortunately DSDs do not exist for $n = 27$.

5.4 Notes

I found some of the course notes on Bill Cherowitzo's webpage (math.ucdenver.edu/~wcherowi) useful, especially
http://math.ucdenver.edu/~wcherowi/courses/m6406/m6406f.html and
http://math.ucdenver.edu/~wcherowi/courses/m6023/m6023f.html. A useful book is Stinson (2003).

Chapter 6
Response Surface Maximization

A version of this material appeared in the first two editions of ALM.

One purpose of response surface methodologies is to maximize or minimize a response function. The response is a function of some input variables that are controllable, call these $\xi = (\xi_1, \ldots, \xi_f)'$. Denote the response function

$$\mu(\xi) \equiv \mu(\xi_1, \ldots, \xi_f).$$

Often, $\mu(\xi)$ is thought of as the output of an industrial process that has $\xi_1, \ldots, \xi_f$ as inputs to the process. The response function is unknown, so we need to estimate it. In fact, even the form of the response function is unknown, so we will approximate $\mu(\xi)$ with linear or quadratic functions of ξ.

To estimate the response function, we need to collect data that relate the ξ variables to the response. Information about the unknown function $\mu(\xi)$ is obtained by selecting values for ξ and making observations

$$y = \mu(\xi) + \varepsilon, \tag{1}$$

where ε is an unobservable error assumed to have $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = \sigma^2$. Often, ε is assumed to have a normal distribution. It may be necessary to transform y to make these assumptions approximately true. An experimental design consists of specifying a set of ξ values at which to take observations. The observations are generally assumed to be independent.

Often, experimental designs are used that specify either two levels or three equally spaced levels for each variable ξ_j. These are 2^f or 3^f designs or fractional replications of the full designs. Each ξ_j variable corresponds to a factor, and we can choose the levels for each factor. In general, we can call the levels $\xi_{0j} < \xi_{1j}$ in the two-level case and $\xi_{0j} < \bar{\xi}_{\cdot j} < \xi_{1j}$ in the equally spaced three-level case. Note that with three equally spaced levels, the middle level $\bar{\xi}_{\cdot j}$ must be the average of the low level and the high level. It is common practice in response surface methods to transform all the predictor variables in these designs so that 0 becomes the middle value of each transformed variable and the extreme values become ± 1. Specifically,

define

$$x_{kj} = 2\frac{\xi_{kj} - \bar{\xi}_{\cdot j}}{\xi_{1j} - \xi_{0j}} \qquad k = 0, 1.$$

In general we transform any arbitrary ξ_j value into $x_j = 2(\xi_j - \bar{\xi}_{\cdot j})/(\xi_{1j} - \xi_{0j})$ and define $x = (x_1, \ldots, x_f)'$. We can now consider the response function in terms of the transformed variables and write

$$\mu(x) \equiv \mu(x_1, \ldots, x_f).$$

The primary advantage of this redefinition of ξ into x is that the information collected from any experiment will always be collected about the center point $(0, \ldots, 0)$ as measured in x. This may seem like a small gain for all the trouble involved in the transformation, but ultimately it is probably worthwhile.

If $\mu(x)$ were a known function, the standard approach to finding a maximum or minimum involves finding a critical point by setting the partial derivatives equal to zero and investigating properties of the matrix of second partial derivatives to determine whether the critical point is a maximum, minimum, or saddlepoint. Without knowing $\mu(x)$, this does not work. Instead, to find a maximum, one typically performs a series of experiments each of which leads one to look at values of the x variables that increase the response. In each experiment, we approximate $\mu(x)$ over the range of the observed data with a polynomial and use the fitted approximating polynomial to estimate the direction in which the yield increases fastest. We then take observations in the direction of most rapid increase until no more increase is obtained. Another full experiment is conducted about the ξ value that currently gives the highest yield. These ξ values in the new experiment are transformed into *new* x values with 0 as the center point. The new experiment indicates a new direction of maximum increase to follow or, if we are already near the maximum response, an estimate of the x values, and thus the ξ values, that give maximum yield.

The process is rather like climbing a mountain on a foggy day. You cannot see the peak, you can only see what is nearby, and from that information you try to get to the top as quickly as possible.

In this chapter, we discuss only the problem of finding the maximum response. Methods for finding a minimum are similar, or one could minimize $\mu(x)$ by maximizing $-\mu(x)$. Section 1 discusses approximations to the true response function. Section 2 examines the use of linear approximating functions and the method of steepest ascent. Section 3 discusses the fitting of quadratic polynomials. Section 4 presents an introduction to the interpretation of quadratic response functions. Throughout, we make extensive use of multivariable calculus. (Appendix A in *ALM-III* contains many of the necessary results in the notation used here.) There are many fine books on response surface methodologies. Box and Draper (1987) gives an excellent and comprehensive discussion.

6.1 Approximating Response Functions

Depending on information previously obtained experimentally, one of two polynomial approximations to the response function $\mu(x)$ is used. When we are far from the conditions (x_j values) that give maximum yield, we can often use a first-order polynomial to approximate the response surface. The first-order Taylor approximation about the center vector $x = 0$ of the data is

$$\begin{aligned}\mu(x) &\doteq \mu(0) + \sum_{j=1}^{f} \left[\frac{\partial \mu(x)}{\partial x_j}\Big|_{x_j=0} \right] x_j \\ &= \mu(0) + [\mathbf{d}\mu(0)]x \\ &= \mu(0) + x'[\mathbf{d}\mu(0)]',\end{aligned}$$

where $\mathbf{d}\mu(0)$ is the $1 \times f$ row vector of partial derivatives $\partial\mu(x)/\partial x_j$ evaluated at the vector $x = 0$ (i.e., $x_1 = 0, \ldots, x_f = 0$).

We do not know $\mu(x)$, so we do not know the partial derivatives; they are just some unknown values. Identify

$$\beta_0 \equiv \mu(0),\ \beta_1 \equiv \frac{\partial \mu(x)}{\partial x_1}\Big|_{x=0},\ \ldots,\ \beta_f \equiv \frac{\partial \mu(x)}{\partial x_f}\Big|_{x=0}$$

and write

$$\begin{aligned}\mu(x) &\doteq \beta_0 + \sum_{j=1}^{f} \beta_j x_j \\ &= \beta_0 + x'\beta_*,\end{aligned}$$

where

$$\beta_* = (\beta_1, \ldots, \beta_f)'.$$

Applying equation (6.0.1) to each observation $i = 1, \ldots, n$ from some design gives

$$\begin{aligned}y_i &= \mu(x_{i1}, \ldots, x_{iq}) + \varepsilon_i \\ &\doteq \beta_0 + \sum_{j=1}^{f} \beta_j x_{ij} + \varepsilon_i.\end{aligned}$$

The approximation is a multiple regression model that we know how to fit.

As we get closer to the maximum, the response surface must curve, so a first-order polynomial becomes inadequate to approximate the surface. Second-order polynomials are needed. Recall again that we are approximating $\mu(x)$ over the range of values in a designed experiment and that x is defined so that the center of the design is 0. The second-order Taylor approximation about 0 is

$$\mu(x) \doteq \mu(0) + x'\mathbf{d}\mu(0)' + x'[\mathbf{d}^2\mu(0)]x/2,$$

where $\mathbf{d}\mu(0)$ was defined previously and $\mathbf{d}^2\mu(0)$ is the $f \times f$ matrix of second partial derivatives evaluated at the vector $x = 0$. The element of $\mathbf{d}^2\mu(0)$ in the ith row and jth column is $\partial^2\mu(x)/\partial x_i \partial x_j$ evaluated at $x = 0$.

Again, we do not know $\mu(x)$, so we do not know the derivatives and we write

$$\mu(x) \doteq \beta_0 + x'\beta_* + x'Bx,$$

where again

$$\beta_* = (\beta_1, \ldots, \beta_f)' = \mathbf{d}\mu(0)'$$

and now we define

$$B \equiv \begin{bmatrix} \beta_{11} & \beta_{12}/2 & \beta_{13}/2 & \cdots & \beta_{1q}/2 \\ \beta_{12}/2 & \beta_{22} & \beta_{23}/2 & \cdots & \beta_{2q}/2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_{1q}/2 & \beta_{2q}/2 & \beta_{3q}/2 & \cdots & \beta_{qq} \end{bmatrix} \equiv \frac{1}{2}\mathbf{d}^2\mu(0).$$

With this definition of B, the approximation becomes

$$\mu(x) \doteq \beta_0 + \sum_{j=1}^{f} \beta_j x_j + \sum_{j=1}^{f} \sum_{k \geq j} \beta_{jk} x_j x_k = \beta_0 + x'\beta_* + x'Bx.$$

This is a much more succinct quadratic approximation than the nonparametric regression polynomial approximation examined in, say, *ALM-III*, Chapter 1 which would use

$$\mu(x) \doteq \sum_{j_1=0}^{2} \cdots \sum_{j_f=0}^{2} \beta_{j_1 \cdots j_f} x_1^{j_1} \cdots x_f^{j_f}.$$

Response surface methods are designed to find you the shortest route to the top of the mountain. Nonparameteric methods are designed to map the entire mountain. One of those things is easier to do than the other.

Applying equation (6.0.1) to each observation from some design gives

$$\begin{aligned} y_i &= \mu(x_{i1}, \ldots, x_{iq}) + \varepsilon_i \\ &\doteq \beta_0 + \sum_{j=1}^{f} \beta_j x_{ij} + \sum_{j=1}^{f} \sum_{k \geq j} \beta_{jk} x_{ij} x_{ik} + \varepsilon_i. \end{aligned}$$

The approximation is a multiple (polynomial) regression model that we know how to fit.

Before using approximate response surface models, they need to be checked. In practice, this means checking on the adequacy of the corresponding multiple regression models. The assumptions of the regression model should be examined with residual plots. In addition, it is important to check whether the order of the polynomial approximation is appropriate. In checking whether the approximate model helps to explain the data, it is not sufficient to perform the standard regression F test ($MSReg/MSE$) and check for significance. Box and Draper (1987, Section 8.2), in

discussing unpublished work by Box and Wetz, suggest that one should not attempt to interpret a fitted response surface unless the F statistic is 10 times greater than the F percentage point that defines an appropriate test. The approximate model also needs to be tested for lack of fit. Perhaps a higher-order polynomial is needed to give an adequate response surface. Frequently, we do not have enough data to test the approximate model against a complete higher-order model, but often parts of a higher-order model can be fitted and we test what we can. It should be remembered that just because a higher-order effect is statistically significant, it does not follow that the higher-order effect is of practical importance. It may be possible to ignore statistically significant higher-order effects because they have little practical effect on the estimated response surface.

Often, transformations of the response or predictor variables can substantially improve the fit of approximating polynomials. The need for a transformation of y can be examined as in any regression problem; see Christensen (1996, Section 7.10 or 2015, Section 7.3). Box and Draper (1987) provide an extensive discussion of transformations.

6.2 First-Order Models and Steepest Ascent

With f input variables, the design used for estimating a first-order model is often a 2^f design or a fractional replication of a 2^f design. The two levels are arrived at by specifying a middle location m_j and a spread s_j for each factor (input variable) ξ_j; then, the two levels of ξ_j are defined by

$$\pm 1 = \frac{\xi_j - m_j}{s_j}.$$

The entire analysis for a given design is typically conducted on the transformed variables

$$x_j = \frac{\xi_j - m_j}{s_j}$$

that only take on the values 1 and -1.

In this section, we assume that the response function is

$$\begin{aligned}\mu(x) &= \beta_0 + \sum_{j=1}^{f} \beta_j x_j \\ &= \beta_0 + x'\beta_*,\end{aligned}$$

where

$$\beta_* = (\beta_1, \ldots, \beta_f)'.$$

In fact, this is only an approximation. Moreover, we can only estimate β_0 and β_* by fitting

$$y_i = \beta_0 + \sum_{j=1}^{f} \beta_j x_{ij} + \varepsilon_i$$

to data obtained experimentally.

The purpose of a first-order model is to indicate the *direction* in which the response can be increased most rapidly. A direction is defined by the vector $x = (x_1, \ldots, x_f)'$. We need to find values for x that increase the response most rapidly. In other words, we need to find x so that $x'\beta_* = \sum_{j=1}^{f} \beta_j x_j$ is as large as possible.

We really only want the direction of most rapid increase (*steepest ascent*), so that we can take observations to explore that direction. When looking for the best direction, we need to have a standard length for the x vectors we consider. If the vectors are not of standard length, the issue of best direction gets confused with the length of the vector. Obviously, a vector x with extremely large x_j components has a tendency to produce more extreme values for $x'\beta$ than a vector with moderate x_j values. We specify a standard length of 1, that is, any direction vector x that we consider is required to have

$$x'x = \sum_{j=1}^{f} x_j^2 = 1.$$

(Actually, $x'x$ is the squared length of x, but the length is one, so the square does not matter.)

We can now use a well-known result called the Cauchy–Schwartz inequality to find the direction of steepest ascent. We are trying to maximize $x'\beta_* = \sum_{j=1}^{f} \beta_j x_j$ subject to the constraint that $x'x = 1$. The Cauchy–Schwartz inequality states that

$$(x'\beta_*)^2 = \left(\sum_{j=1}^{f} \beta_j x_j\right)^2 \le \left(\sum_{j=1}^{f} \beta_j^2\right)\left(\sum_{j=1}^{f} x_j^2\right) = (\beta_*'\beta_*)(x'x).$$

Because $x'x = 1$, we have

$$(x'\beta_*)^2 \le (\beta_*'\beta_*).$$

We can attain the maximum value of the inequality by picking

$$x = \frac{1}{\sqrt{\beta_*'\beta_*}}\beta_*.$$

This gives

$$(x'\beta_*)^2 = \left(\frac{\beta_*'\beta_*}{\sqrt{\beta_*'\beta_*}}\right)^2 = \frac{(\beta_*'\beta_*)^2}{\beta_*'\beta_*} = \beta_*'\beta_*.$$

Thus, the upper bound of the inequality is actually achieved. The direction of steepest ascent is $\beta_*/\sqrt{\beta_*'\beta_*}$ or, ignoring the restriction to vectors of length 1, the direction of steepest ascent is β_*.

In practice, the first-order model is estimated from experimental data to give $\hat{\beta}_0$ and $\hat{\beta}_*$. The estimated direction of steepest ascent is $\hat{\beta}_*$. The procedure is then to

take observations along the direction of steepest ascent, for example, at

$$x = \frac{1}{\sqrt{\hat{\beta}_*'\hat{\beta}_*}}\hat{\beta}_*, \frac{2}{\sqrt{\hat{\beta}_*'\hat{\beta}_*}}\hat{\beta}_*, \frac{3}{\sqrt{\hat{\beta}_*'\hat{\beta}_*}}\hat{\beta}_*, \ldots,$$

continuing as long as the corresponding observed response $y = \mu(x) + \varepsilon$ continues to increase. Once the response starts to drop, run another experiment centered near the x value that generated the largest observed response. Note that the sequence of x values given previously must be transformed back into ξ values before the experimenters will know where to take the new observations. It is also true that the direction of steepest ascent depends on how one chooses to define the transformation between ξ and x.

EXAMPLE 6.2.1. In this example we will collect initial data, fit a first-order model, then follow the direction of steepest ascent to a new center point. At the new center point we will collect data in stages until we reach an appropriate conclusion about what to do (either again follow the direction of steepest ascent or decide to fit a a second-order model).

The factors in the investigation are: A — nitrogen, B — phosphorous, C — potassium, D — manganese. In the initial experiment, it was decided to set each factor at two levels: 0 units and 2 units. The initial design was a $1/2$ rep. of a 2^4 design using $ABCD$ to define the $1/2$ rep. To obtain an estimate of error and to check the fit of the linear polynomial model, the design was augmented with four points at the x center $(0,0,0,0)$, that is, four points each receiving 1 unit of every factor. It was decided to run the experiment in two blocks of six treatments. The $1/2$ rep. was confounded into blocks of four using the $AB = CD$ interaction and two center points were added to each block. For example, with a_0 denoting $x_1 = -1$ and b_1 denoting $x_2 = 1$, the basic $1/2$ rep. of the 2^4 confounded in blocks of four follows.

$$\begin{array}{c} ABCD(1) \\ \hline AB(1) \mid AB(-1) \\ \hline a_0b_0c_0d_0 \mid a_0b_1c_0d_1 \\ a_0b_0c_1d_1 \mid a_0b_1c_1d_0 \\ a_1b_1c_0d_0 \mid a_1b_0c_0d_1 \\ a_1b_1c_1d_1 \mid a_1b_0c_1d_0 \end{array}$$

Note that with $ABCD$ as the defining effect, the main effects are aliased with three-factor effects. We assume that three-factor effects and the four-factor effect are negligible. The two-factor effect $AB = CD$ is confounded with blocks. The other two-factor effects are aliased as $AC = BD$ and $AD = BC$. The addition of the center points has no effect on this aliasing. The center points make this an unbalanced ANOVA design with each factor at three levels: 0, 1, and 2 units. To a large extent, we can analyze the data as a $1/2$ rep. of a 2^4, ignoring the center points. The most unusual aspect of the design is that the blocks are confounded not only with AB and CD but also with aspects of the center points.

The center points provide both the estimate of error and a measure of lack of fit. They need to be handled correctly. Within each block, the six treatments should

be performed in random order and on randomly chosen material. There may be a temptation to run the two center points one after the other because they use the same settings of the process. This is unacceptable. It is important that the two center points be subject to all of the variability involved in any other two runs. If a run involves, say, shutting off machines and readjusting them, the center points must be subjected to all of the same procedures. If they are not, the estimate of error provided by the center points is invalid for comparing the treatment effects.

In this example, the levels a_0 and a_1 correspond to nitrogen levels of $\xi_{01} = 0$ and $\xi_{11} = 2$. These are transformed into $x_{01} = -1$ and $x_{11} = 1$ for use in the first-order polynomial model. Similarly, factors B, C, and D correspond to variables x_2, x_3, and x_4, respectively. The first-order model incorporates block effects, so the model is

$$y = \beta_0 + \gamma_0 Blk + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon. \tag{1}$$

The parameter vector $\beta_* = (\beta_1, \beta_2, \beta_3, \beta_4)'$ determines the direction of steepest ascent. The blocking variable *Blk* consists of 0's and 1's with the 1's indicating the observations from the first block. The *SSReg* for model (1) can be divided into two parts, one for blocks and one for the linear terms associated with β_*. Table 6.1 gives the data and the predictor variables for the regression along with other predictor variables to be discussed later.

Table 6.1 Data and model matrix for initial $1/2$ replication.

y	x_1	x_2	x_3	x_4	x_1x_3	x_1x_4	*Blks*	*Ctr*
8.8117	−1	−1	−1	−1	1	1	1	0
11.7345	−1	−1	1	1	−1	−1	1	0
10.8053	−1	1	−1	1	1	−1	0	0
14.1937	−1	1	1	−1	−1	1	0	0
9.3778	1	−1	−1	1	−1	1	0	0
11.7957	1	−1	1	−1	1	−1	0	0
10.2977	1	1	−1	−1	−1	−1	1	0
13.9054	1	1	1	1	1	1	1	0
12.6100	0	0	0	0	0	0	1	1
12.0802	0	0	0	0	0	0	1	1
11.9820	0	0	0	0	0	0	0	1
11.7558	0	0	0	0	0	0	0	1

To test model (1) for lack of fit, ideally we would test it against the second-order model

$$\begin{aligned} y = {} & \beta_0 + \gamma_0 Blk + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \\ & + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{44} x_4^2 \\ & + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 \\ & + \beta_{23} x_2 x_3 + \beta_{24} x_2 x_4 + \beta_{34} x_3 x_4 + \varepsilon. \end{aligned}$$

We do not have sufficient data in Table 6.1 to carry out such a test. We have essentially two levels on each factor, so we cannot test all of the quadratic effects, the β_{jj}'s. The cross terms β_{ij} correspond to interactions, so some of them can be examined. The predictors x_1x_3 and x_1x_4 correspond to the *AC* and *AD* interactions. These interactions are aliased with *BD* and *BC* respectively, which in turn correspond to x_2x_4 and x_2x_3. Thus, β_{13} is aliased with β_{24}, and β_{14} is aliased with β_{23}; these regression coefficients have no separate identities. Moreover, the predictors x_1x_2 and x_3x_4 cannot be isolated from the block effects. A regression model that we can actually fit is

$$y = \beta_0 + \gamma_0 Blk + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 + \varepsilon.$$

It is merely a whim that we use x_1x_3 and x_1x_4 rather than x_2x_4 and x_2x_3. In fact, $x_1x_3 \equiv x_1 \times x_3$, so it is easily seen that $x_1x_3 = x_2x_4$ and that $x_1x_4 = x_2x_3$.

While we have essentially two levels on each factor, the existence of the center points gives a third level for each factor but without maintaining factorial treatment structure. The center points can be used to test lack of fit by examining whether they are consistent with the rest of the model. We define a center variable *Ctr* similar to the block variable *Blk*. *Ctr* consists of 0's and 1's with ones identifying the points at the center. The first-order model defines a plane in four dimensions, and the average of the center points should lie near that plane if the first-order model fits well. A significant effect due to the center points suggests curvature and thus that the plane is an inadequate model for the response surface. Including the effect for the center points in the model gives

$$y = \beta_0 + \gamma_0 Blk + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 + \gamma_1 Ctr + \varepsilon. \quad (2)$$

The *SSReg* for this model has 8 degrees of freedom and it can be divided into components with 1 degree of freedom for blocks, 4 degrees of freedom for the linear terms, 2 degrees of freedom for the interactions, and 1 degree of freedom for the center effect. The blocks and linear terms are identical to those obtained from model (1).

Including a center point effect is equivalent to including one $\beta_{jj}x_j^2$ term. Note that for any j, x_j^2 is 0 if the point is a center point and 1 otherwise. Thus, the variable *Ctr* satisfies $Ctr = 1 - x_j^2$ and in model (2) $\gamma_1 = -\beta_{jj}$. This argument does not depend on j, so the β_{jj} terms must be aliased with each other.

The fitted regression equation is

$$\hat{y} = 11.4 - 0.078 Blk - 0.021 x_1 + 0.935 x_2 + 1.54 x_3 + 0.091 x_4 - 0.036 x_1 x_3 + 0.207 x_1 x_4 + 0.742 Ctr.$$

A more complete look at the parameter estimates is available from Table 6.2. The effect of blocks seems minor, as do the effects of factors A and D (x_1 and x_4). Factors B and C (x_2 and x_3) have substantial effects. The interactions have little effect. The

center points contain some suggestion of lack of fit for the first-order model but the evidence is not clear cut.

Table 6.2 Model (2) estimates for initial $1/2$ replication.

Table of Coefficients				
Predictor	$\hat{\beta}_j$	SE($\hat{\beta}_j$)	t	P
Constant	11.4045	0.2087	54.64	0.000
Blocks	−0.0785	0.2640	−0.30	0.786
x_1	−0.0211	0.1617	−0.13	0.905
x_2	0.9353	0.1617	5.78	0.010
x_3	1.5421	0.1617	9.54	0.002
x_4	0.0905	0.1617	0.56	0.615
x_1x_3	−0.0357	0.1617	−0.22	0.839
x_1x_4	0.2069	0.1617	1.28	0.291
Center	0.7418	0.2800	2.65	0.077

Table 6.3 gives a sequential analysis of variance table for the regression model. Again, we see that blocks have little effect. The linear terms have a large effect; as noted earlier, this is mostly due to factors B and C. The interaction is negligible, while the evidence of lack of fit from the center points is questionable. The coefficient of determination for the model is $R^2 = 97.8\%$.

Table 6.3 Model (2) analysis of variance for initial $1/2$ replication.

Analysis of Variance					
Source	df	SS	MS	F	P
Blocks	1	0.0185	0.0185	0.09	0.786
Linear	4	26.0919	6.5230	31.20	0.009
Interaction	2	0.3528	0.1764	0.84	0.513
Center	1	1.4672	1.4672	7.02	0.077
Error	3	0.6274	0.2091		
Total	11	28.5577			

Table 6.4 gives a detailed listing of the sums of squares for the regression model and the corresponding sums of squares from an analysis of variance performed on the same data. In the ANOVA, the error was broken into two parts. Each block had two observations on the center of the design, so each block gives one degree of freedom for pure error. The other degree of freedom for error comes from the block by center points interaction. As usual in a blocking experiment, interactions involving blocks are used as error. This table gives much the same information as Table 6.2 about the importance of different factors.

As discussed at the end of Section 1, we need to verify that the estimated first-order model satisfies the assumptions, that it does not display sufficient lack of fit

Table 6.4 Sums of squares for initial $1/2$ replication.

Regression			ANOVA		
Source	*df*	*SS*	Source	*df*	*SS*
Blocks	1	0.0185	Blocks ($AB = CD$)	1	0.0185
Center	1	1.4672	Center	1	1.4672
x_1	1	0.0036	A	1	0.0036
x_2	1	6.9982	B	1	6.9982
x_3	1	19.0245	C	1	19.0245
x_4	1	0.0656	D	1	0.0656
x_1x_3	1	0.0102	$AC = BD$	1	0.0102
x_1x_4	1	0.3426	$AD = BC$	1	0.3426
Error	3	0.6274	Blks∗Center	1	0.4615
			Pure Error	2	0.1659
Total	11	28.5577	Total	11	28.5577

to invalidate conclusions drawn from it, and that it is sufficiently informative to interpret. We have already seen that the interactions do not suggest lack of fit and that the center points do not display a convincing lack of fit. Figure 6.1 contains the residual versus predicted plot; it appears to be alright. The normal plot in Figure 6.2 is not wonderful but one needs to make allowance for the high dependency among the residuals; there are 12 residuals but only 3 degrees of freedom for error. As for the issue of whether the estimated first-order model is sufficiently informative, the rule of thumb is that the F statistic for the linear effects should be 10 times the value of the F percentile in the test. Appropriate F percentiles are

$$F(.90,4,3) = 5.34 \quad \text{and} \quad F(.95,4,3) = 9.12.$$

The observed F value of 31.20 from Table 6.3 does not meet the criterion. We noted earlier that factors A and D seem to have little effect. Consider what happens when we test only factors B and C. From Table 6.4, the F statistic becomes

$$F = \frac{(6.9982 + 19.0245)/2}{0.2091} = \frac{13.01135}{0.2091} = 62.23.$$

Appropriate F percentiles for this test are

$$F(.90,2,3) = 5.46 \quad \text{and} \quad F(.95,2,3) = 9.28.$$

We are now in the ballpark, $62.23 > 10(5.46)$. The two substantial factors suggest that we can proceed with some hope of getting reasonable results. Based on this analysis, we should perhaps drop factors A and D from further consideration. Instead, we retain them because of the possibility that factors A and D have an effect in some other area of the design space (i.e., for some other values of the factor levels). The next step in the procedure is to move away from the current area of experimentation toward an area that provides greater yields. A and D may become important.

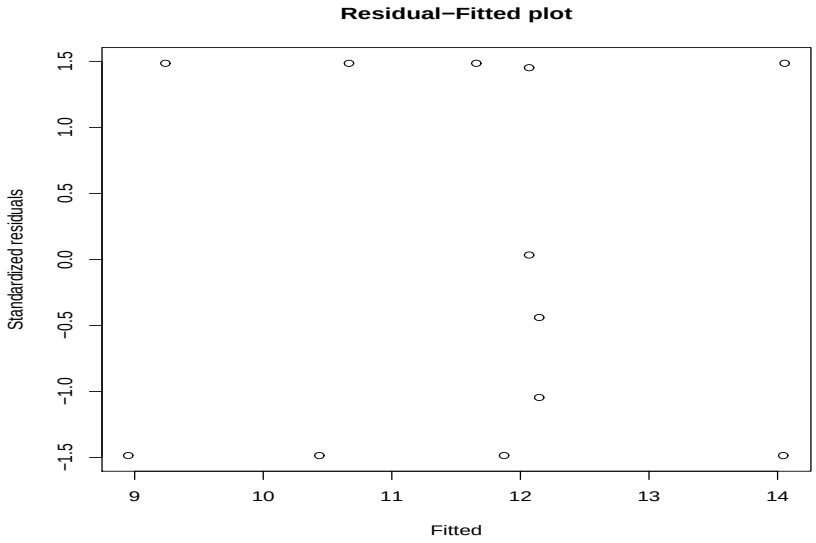

Fig. 6.1 Standardized residuals versus predicted values, initial $1/2$ replication.

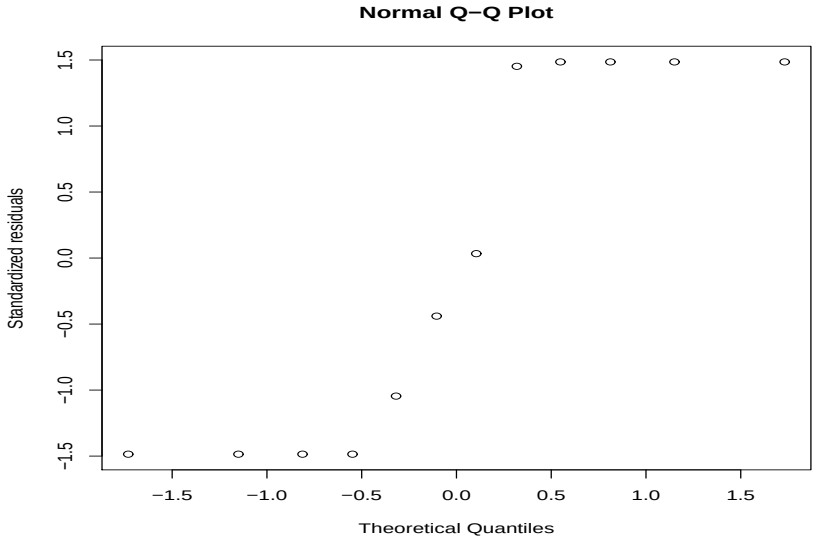

Fig. 6.2 Normal plot of standardized residuals, $W' = 0.806$, initial $1/2$ replication.

Dropping the lack-of-fit terms from the model (i.e., the interactions and center effect), the regression equation becomes

$$\hat{y} = 11.4 - 0.078Blk - 0.021x_1 + 0.935x_2 + 1.54x_3 + 0.091x_4.$$

The estimated regression coefficients do not change when the lack-of-fit terms are dropped. The nature of the design and the model ensure this.

The direction of steepest ascent does not depend on the intercept or blocks but only on

$$\hat{\beta}_* = (-0.0211, 0.9353, 1.5421, 0.0905)'.$$

The normalizing constant is $\sqrt{\hat{\beta}_*'\hat{\beta}_*} = 1.805961$, so the normalized direction of steepest ascent is

$$(-0.012, 0.52, 0.85, 0.05).$$

Note that changes in this direction yield little change in the levels of factors A and D. Table 6.5 gives yields and predictor variables for a series of observations in the direction of steepest ascent. Table 6.5 also gives the corresponding ξ values that are necessary to obtain the observations. The sequence of observations was stopped when a drop in yield was obtained.

Table 6.5 Observations in the direction of steepest ascent.

y	x_1	x_2	x_3	x_4	ξ_1	ξ_2	ξ_3	ξ_4
13.9371	−0.012	0.52	0.85	0.05	0.988	1.52	1.85	1.05
14.7492	−0.023	1.04	1.71	0.10	0.977	2.04	2.71	1.10
15.0789	−0.035	1.55	2.56	0.15	0.965	2.55	3.56	1.15
16.2788	−0.047	2.07	3.42	0.20	0.953	3.07	4.42	1.20
16.6521	−0.058	2.59	4.27	0.25	0.942	3.59	5.27	1.25
17.3583	−0.070	3.11	5.12	0.30	0.930	4.11	6.12	1.30
16.9928	−0.082	3.63	5.98	0.35	0.918	4.63	6.98	1.35

The maximum yield obtained is at $x = (-0.070, 3.11, 5.12, 0.30)'$, or equivalently $\xi = (0.930, 4.11, 6.12, 1.30)'$. It was decided to center the next design at $\xi = (1, 4, 6, 1.5)'$ and to use spreads of $(.5, 1, 1, .5)$. Thus, $\xi_{01} = 1 - .5 = .5$ and $\xi_{11} = 1 + .5 = 1.5$. Similarly, $\xi_{02} = 4 - 1 = 3$ and $\xi_{12} = 4 + 1 = 5$. We again use the $1/2$ rep. of the 2^4 design using *ABCD* to define the $1/2$ rep. As the blocking had little effect, we no longer incorporate it. Thus, we can now estimate the $AB = CD$ interaction by using x_1x_2 as a predictor variable. As before, the design includes four center points. With no blocks in the design, the four center points provide three degrees of freedom for pure error.

The data and model matrix for this second $1/2$ rep. are given in Table 6.6 along with data from a third $1/2$ rep. with different scale factors that will be discussed later. Note that many of the observations from the new design are in the high 16s.

Yields have improved considerably from those reported in Table 6.1 and are consistent with the yields in Table 6.5.

Table 6.6 Data and design for recentered and recentered-rescaled $1/2$ replications.

y	x_1	x_2	x_3	x_4	x_1x_3	x_1x_4	x_1x_2	Ctr	Rescaled $1/2$ Rep.
16.1984	−1	−1	−1	−1	1	1	1	0	12.5737
16.9490	−1	−1	1	1	−1	−1	1	0	14.6108
16.5140	−1	1	−1	1	1	−1	−1	0	13.2115
16.8413	−1	1	1	−1	−1	1	−1	0	16.6965
16.3669	1	−1	−1	1	−1	1	−1	0	11.7133
16.6894	1	−1	1	−1	1	−1	−1	0	14.7368
16.8586	1	1	−1	−1	−1	−1	1	0	13.1926
16.3623	1	1	1	1	1	1	1	0	16.9687
17.0425	0	0	0	0	0	0	0	1	16.7286
16.6733	0	0	0	0	0	0	0	1	16.8469
16.5159	0	0	0	0	0	0	0	1	16.7771
16.6081	0	0	0	0	0	0	0	1	16.1963

Tables 6.7 and 6.8 present ANOVA tables for four designs all centered at $\xi = (1,4,6,1.5)'$. The first three use the spreads $(.5,1,1,.5)$. The actual data for these three are given in Section 3. We consider the designs in turn. From Table 6.7, the recentered $1/2$ rep. just discussed is inadequate for drawing inferences because there is almost no effect due to the linear terms.

In an attempt to obtain data worth interpreting, the $1/2$ rep. was augmented with additional observations to obtain one complete replication of a 2^4 design with four center points. The full replication allows estimation of all six two-factor interactions, so all of the predictor variables x_ix_j can be included in the model. The five degrees of freedom for higher-order interactions are pooled with the three degrees of freedom for pure error to obtain the MSE for the model. The ANOVA table for the full replication is given in Table 6.7. Again, none of the effects are significant. The linear effects are not significant at the $\alpha = .10$ level and thus are a far cry from exceeding ten times the significance level. Nonetheless, from Table 6.8 we begin to see variables x_2 and x_3 appearing as important, just as they did earlier.

We again augmented the design to obtain more informative data. This time, we duplicated the full 2^4 plus four center points design. From Table 6.7, the linear effects are now nearly significant at the 0.01 level but they are still far from having the overwhelming significance required for further interpretation. More to the point, we now have a clear indication of lack of fit. The F value of 7.86 for testing the center points against the rest of the model exceeds the percentile $F(.99,1,28) = 7.64$. This suggests that we need at least a quadratic approximating function. The design needs to be augmented to allow fitting of a second-order polynomial. This is discussed in the next section. It is interesting to note from Table 6.8 that the most significant effects are those for x_2, x_3, x_2x_3, and the center points. In fact, all of the effects other than these have sums of squares that are smaller than the MSE.

Table 6.7 Analysis of variance for augmented designs.

Recentered Half Replicate Analysis of Variance					
Source	*df*	*SS*	*MS*	*F*	*P*
Linear	4	0.14542	0.03636	0.68	0.651
Interaction	3	0.38896	0.12965	2.43	0.242
Center	1	0.03373	0.03373	0.63	0.485
Error	3	0.15995	0.05332		
Total	11	0.72806			

Recentered Full Replicate Analysis of Variance					
Source	*df*	*SS*	*MS*	*F*	*P*
Linear	4	0.79931	0.19983	2.69	0.109
Interaction	6	0.48957	0.08160	1.10	0.437
Center	1	0.15155	0.15155	2.04	0.191
Error	8	0.59432	0.07429		
Total	19	2.03474			

Recentered Two Replicates Analysis of Variance					
Source	*df*	*SS*	*MS*	*F*	*P*
Linear	4	1.42594	0.35649	3.92	0.012
Interaction	6	0.45671	0.11418	1.26	0.307
Center	1	0.71427	0.71427	7.86	0.009
Error	28	2.54492	0.09089		
Total	39	7.14182			

Recentered-Rescaled $1/2$ Replicate Analysis of Variance					
Source	*df*	*SS*	*MS*	*F*	*P*
Linear	4	24.2434	6.0609	68.25	0.003
Interaction	3	0.9312	0.3104	3.50	0.165
Center	1	15.6717	15.6717	176.48	0.001
Error	3	0.2663	0.0888		
Total	11	41.1127			

In retrospect, the replicated 2^4 experiment was constructed in three blocks: an initial $1/2$ rep. with center points, the completion of the $1/2$ rep., and the replication of the full factorial. Typically, such blocks are subject to different experimental conditions, so an effect for blocks should be included in the regression model. This is easily done by including predictor variables for blocks similar to the variable that identifies the center points. The predictor variable for the second block consists of zeros and ones with ones identifying observations in the second block. The predictor variable for the third block is similar. The first block is identified by default as those observations not belonging to any of the other blocks. In our example, the process generating the data was identical across blocks, so no block effects were necessary. *In practice, having a process that does not change with blocks is extremely unusual*. It is also more common when adding a second half replicate to include the same number of center points as were in the first half rep.

Table 6.8 Effect sums of squares.

Source	*df*	1/2 Rep. *SS*	Full Rep. *SS*	2 Reps. *SS*	Rescaled 1/2 Rep. *SS*
x_1	1	0.00635	0.00028	0.08203	0.0289
x_2	1	0.01735	0.34631	2.54006	5.1756
x_3	1	0.10216	0.45271	0.77505	18.9785
x_4	1	0.01956	0.00001	0.02880	0.0604
x_1x_2	1	0.00023	0.02977	0.04842	0.1219
x_1x_3	1	0.19582	0.03704	0.02933	0.2040
x_1x_4	1	0.19291	0.06967	0.01950	0.6053
x_2x_3	1		0.12758	0.35619	
x_2x_4	1		0.18781	0.00064	
x_3x_4	1		0.03770	0.00263	
Center	1	0.03373	0.15155	0.71427	15.6717
MSE		0.05332	0.07429	0.09089	0.0888

There is substantial work involved in completing the 2^4 design and then replicating it. It is probably a good idea to see whether we inadvertently set the design in a region of suboptimal yield. To check on this, we centered another design at $\xi = (1,4,6,1.5)'$ but expanded the spreads to $(1,3,5,1)$. Again, we used no blocking and a $1/2$ rep. with four center points. Summaries of the analysis are given in Tables 6.7 and 6.8 as the recentered-rescaled $1/2$ rep. More importantly, Table 6.6 contains the yields in the recentered-rescaled $1/2$ rep. The yields are uniformly smaller than those in the original $1/2$ rep except at $x = (1,1,1,1)$ and at two of the center points. This suggests that increasing the levels of the factors may still increase yield slightly, but a more complete analysis requires a more extensive design and a more extensive polynomial model. □

6.3 Fitting Quadratic Models

In Example 6.2.1, we considered two replicates of a 2^4 design with four center points per replicate. The data are given on the right of Table 6.9 with the first $1/2$ rep. being the observations above the horizontal line in the "1st Rep" column. A first-order polynomial model was found to have significant lack of fit. We now consider augmenting the design to allow fitting of a second-order polynomial

$$y = \beta_0 + \sum_{j=1}^{4} \beta_j x_j + \sum_{j=1}^{4} \sum_{k \geq j} \beta_{jk} x_j x_k + \varepsilon.$$

To do this we add a *star design* along with four more center points. Star points are just the opposite of factorial treatment combinations. In a star design, you start at the center point $(0,0,0,0)'$ and change only one of the factors. Each factor is changed

by the same amount in both a positive and negative direction. We changed the factors by 2 units, thus we took new observations with x values such as $(2,0,0,0)'$, $(-2,0,0,0)'$, $(0,2,0,0)'$, $(0,-2,0,0)'$, and so on. For f factors the location of the star points is determined by $\pm\sqrt{f}$. We included two replications of the star design; the data are given on the left in Table 6.9. The complete design is a *central composite* containing the center points, the 2^4 design, and the star design.

Table 6.9 Data from central composite design.

Star Design					1st Rep.	2nd Rep.				
y	x_1	x_2	x_3	x_4	y	y	x_1	x_2	x_3	x_4
16.7165	2	0	0	0	16.1984	15.9140	−1	−1	−1	−1
16.9255	2	0	0	0	16.9490	16.6119	−1	−1	1	1
16.8088	0	2	0	0	16.5140	16.9104	−1	1	−1	1
17.0992	0	2	0	0	16.8413	16.6026	−1	1	1	−1
16.4714	0	0	2	0	16.3669	15.9193	1	−1	−1	1
16.6335	0	0	2	0	16.6894	16.4512	1	−1	1	−1
17.3265	0	0	0	2	16.8586	16.7357	1	1	−1	−1
16.5691	0	0	0	2	16.3623	17.4326	1	1	1	1
16.5002	−2	0	0	0	17.0425	17.1692	0	0	0	0
16.8499	−2	0	0	0	16.6733	17.0099	0	0	0	0
15.8735	0	−2	0	0	16.5159	16.9727	0	0	0	0
15.9320	0	−2	0	0	16.6081	17.0657	0	0	0	0
16.6993	0	0	−2	0	16.0726	15.5941	−1	−1	−1	1
16.5193	0	0	−2	0	16.3502	16.3469	−1	−1	1	−1
16.8886	0	0	0	−2	16.3358	16.9532	−1	1	−1	−1
17.0260	0	0	0	−2	16.7109	17.0542	−1	1	1	1
17.1728	0	0	0	0	15.7130	16.2635	1	−1	−1	−1
16.3076	0	0	0	0	16.4223	16.3990	1	−1	1	1
16.5874	0	0	0	0	16.5337	17.3961	1	1	−1	1
16.5722	0	0	0	0	16.9590	17.0767	1	1	1	−1

We begin by fitting the quadratic model to the entire data. The fitted regression equation is

$$\begin{aligned}\hat{y} = {} & 16.8 + 0.0459x_1 + 0.275x_2 + 0.0990x_3 + 0.0192x_4 \\ & - 0.0366x_1^2 - 0.116x_2^2 - 0.0784x_3^2 + 0.0146x_4^2 \\ & + 0.0389x_1x_2 - 0.0303x_1x_3 - 0.0247x_1x_4 \\ & - 0.106x_2x_3 + 0.0045x_2x_4 + 0.0091x_3x_4.\end{aligned}$$

More detail on the parameters is given in Table 6.10. The coefficient of determination for the model is $R^2 = 58.8\%$.

Before interpreting the fitted model, we need to check assumptions, check for lack of fit, and check whether the fit is adequate for interpretation. Illustrating these methods is the point of the current section. Interpretation of this model is considered in Section 4.

Table 6.10 Second-order model on all factors.

Table of Coefficients				
Predictor	$\hat{\beta}$	SE($\hat{\beta}$)	t	P
Constant	16.8081	0.0864	194.50	0.000
x_1	0.04592	0.04321	1.06	0.294
x_2	0.27543	0.04321	6.37	0.000
x_3	0.09902	0.04321	2.29	0.027
x_4	0.01921	0.04321	0.44	0.659
x_1^2	−0.03658	0.04042	−0.90	0.370
x_2^2	−0.11649	0.04042	−2.88	0.006
x_3^2	−0.07837	0.04042	−1.94	0.059
x_4^2	0.01455	0.04042	0.36	0.720
x_1x_2	0.03890	0.05292	0.74	0.466
x_1x_3	−0.03027	0.05292	−0.57	0.570
x_1x_4	−0.02468	0.05292	−0.47	0.643
x_2x_3	−0.10550	0.05292	−1.99	0.052
x_2x_4	0.00447	0.05292	0.08	0.933
x_3x_4	0.00906	0.05292	0.17	0.865

To check whether the fit is adequate for interpretation, consider the analysis of variance in Table 6.11. The F statistic is 4.58 while the F percentile for an $\alpha = 0.10$ test is about 1.3; the statistic is nowhere near ten times greater than the percentile, even for this large α level. From Table 6.10, we see that only terms involving x_2 and x_3 have any substantial effect.

Table 6.11 Analysis of variance for second-order model on all factors.

Analysis of Variance					
Source	*df*	*SS*	*MS*	*F*	*P*
Regression	14	5.74740	0.41053	4.58	0.000
Error	45	4.03255	0.08961		
Total	59	9.77995			

Table 6.12 gives the analysis of variance for the quadratic model based only on factors B and C. The F statistic is now 13.5 which is nearly ten times greater than the 90th percentile of the $F(5,54)$ distribution, about 1.38. At least in the directions x_2 and x_3, it is probably worthwhile to interpret a fitted quadratic polynomial.

The fitted quadratic equation in x_2 and x_3 alone is

$$\hat{y} = 16.8 + 0.275x_2 + 0.0990x_3 - 0.114x_2^2 - 0.0756x_3^2 - 0.106x_2x_3.$$

More detail is given in Table 6.13. The fitted model gives $R^2 = 55.5\%$.

We also need to check the assumptions of the quadratic model. Figures 6.3 and 6.4 give standardized residual plots for the quadratic model on x_2 and x_3. Both look

Table 6.12 Analysis of variance for second-order model on factors B and C.

Analysis of Variance					
Source	*df*	*SS*	*MS*	*F*	*P*
Regression	5	5.4327	1.0865	13.50	0.000
Error	54	4.3473	0.0805		
Total	59	9.7799			

Table 6.13 Second-order model on factors B and C.

Table of Coefficients				
Predictor	$\hat{\beta}$	SE($\hat{\beta}$)	*t*	*P*
Constant	16.7861	0.0579	289.83	0.000
x_2	0.27543	0.04095	6.73	0.000
x_3	0.09902	0.04095	2.42	0.019
x_2^2	−0.11374	0.03762	−3.02	0.004
x_3^2	−0.07562	0.03762	−2.01	0.049
x_2x_3	−0.10550	0.05016	−2.10	0.040

quite good. Figures 6.5 and 6.6 give standardized residual plots for the quadratic model in all four factors. Again, both look quite good.

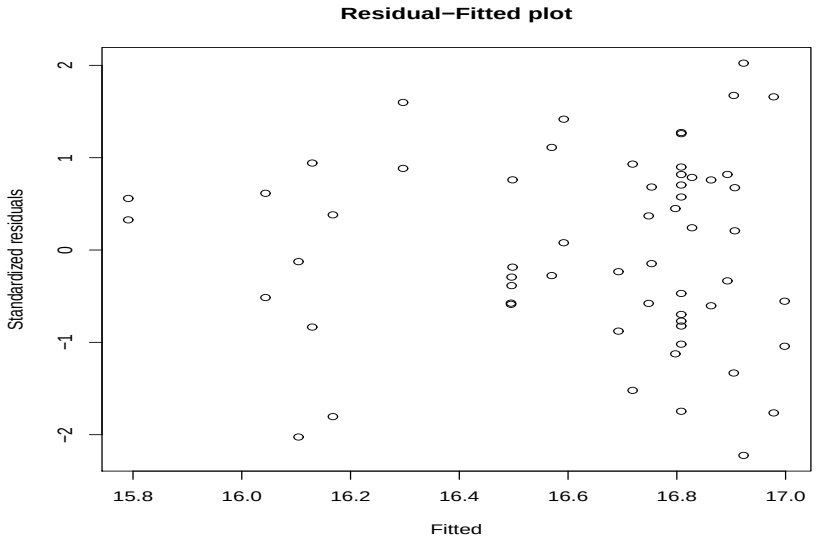

Fig. 6.3 Standardized residuals versus predicted values, quadratic model for factors B and C.

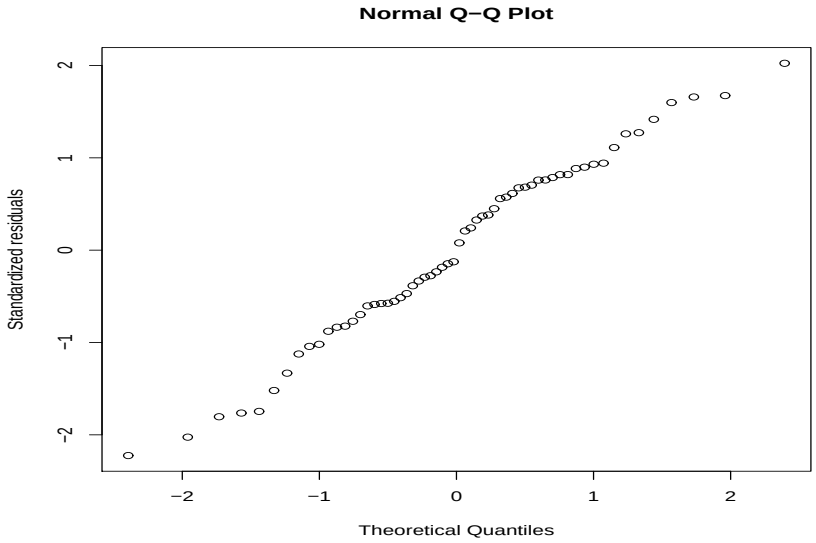

Fig. 6.4 Normal plot of standardized residuals, $W' = 0.990$, quadratic model for factors B and C.

Fig. 6.5 Standardized residuals versus predicted values, quadratic model for all factors.

Fig. 6.6 Normal plot of standardized residuals, $W' = 0.984$, quadratic model for all factors.

The last step before interpreting the model is to check for lack of fit. To do this, we added cubic terms and a center point effect to the model fitted to x_2 and x_3 only. The regression equation is

$$\hat{y} = 16.8 + 0.288x_2 + 0.212x_3 - 0.105x_2^2 - 0.0674x_3^2 - 0.106x_2x_3 \\ - 0.0063x_2^3 - 0.0566x_3^3 + 0.044Ctr,$$

with additional information on the parameters given in Table 6.14. The analysis of variance is given in Table 6.15. Table 6.16 gives the sums of squares for each term in the quadratic model and in the quadratic model with lack of fit.

From Table 6.15, we see little overall evidence of lack of fit. However, from Table 6.14 the effect of x_3^3 is of marginal significance. An analysis of the residuals shows that this cubic effect is due almost entirely to the four high leverage points with $(x_2, x_3) = (0, 2)$ or $(0, -2)$. The design was set up to be well behaved in all four of the x variables; when restricting the analysis to x_2 and x_3, the star points

Table 6.14 Second-order model with lack-of-fit terms for factors B and C.

Table of Coefficients				
Predictor	$\hat{\beta}$	SE($\hat{\beta}$)	t	P
Constant	16.7641	0.0811	206.74	0.000
x_2	0.28804	0.07022	4.10	0.000
x_3	0.21224	0.07022	3.02	0.004
x_2^2	−0.10548	0.04300	−2.45	0.018
x_3^2	−0.06736	0.04300	−1.57	0.123
x_2x_3	−0.10550	0.04966	−2.12	0.038
x_2^3	−0.00631	0.02867	−0.22	0.827
x_3^3	−0.05661	0.02867	−1.97	0.054
Ctr	0.0440	0.1147	0.38	0.703

Table 6.15 Analysis of variance for second-order model with lack-of-fit terms for B and C.

Analysis of Variance					
Source	*df*	*SS*	*MS*	*F*	*P*
Quadratic	5	5.43269	1.08654	13.77	0.000
Lack of Fit	3	0.32313	0.10771	1.37	0.262
Error	51	4.02413	0.07890		
Total	59	9.77995			

become high leverage points. Plotting the residuals against x_3 shows something like a linear trend because of these four residuals. Of course, linear trends cannot exist for residuals plotted against variables in the model; the trend is actually cubic. Given the nature of the quadratic surface as discussed later, any lack of fit should not have great influence on our conclusions.

Table 6.16 Sums of squares for the quadratic model with lack-of-fit terms, x_2 and x_3 only.

Quadratic Model			Lack of Fit		
Source	*df*	Seq. *SS*	Source	*df*	Seq. *SS*
x_2	1	3.64146	*Ctr*	1	0.01164
x_3	1	0.47059	x_2^3	1	0.00382
x_2^2	1	0.63915	x_3^3	1	0.30768
x_3^2	1	0.32530			
x_2x_3	1	0.35619			

The quadratic model in all four factors was also examined for lack of fit. Cubic terms and three- and four-factor interactions were added to the quadratic model. The sums of squares are given in Table 6.17. The analysis of variance is in Table 6.18. Cumulatively, there is no evidence for lack of fit. Only x_3^3 and $x_1x_3x_4$ display any hint of lack of fit. This second term is not significant when tested alone, and it involves two factors, x_1 and x_4, that have not displayed any lower-order effects.

Table 6.17 Second-order model with lack-of-fit terms for all factors.

Quadratic Model			Lack of Fit		
Source	*df*	Seq. *SS*	Source	*df*	Seq. *SS*
x_1	1	0.10120	$x_1x_2x_3$	1	0.01090
x_2	1	3.64146	$x_1x_2x_4$	1	0.00013
x_3	1	0.47059	$x_1x_3x_4$	1	0.22789
x_4	1	0.01770	$x_2x_3x_4$	1	0.03585
x_1^2	1	0.01576	$x_1x_2x_3x_4$	1	0.04324
x_2^2	1	0.66992	x_1^3	1	0.00213
x_3^2	1	0.36244	x_2^3	1	0.00382
x_4^2	1	0.01162	x_3^3	1	0.30768
x_1x_2	1	0.04842	x_4^3	1	0.01119
x_1x_3	1	0.02933			
x_1x_4	1	0.01950			
x_2x_3	1	0.35619			
x_2x_4	1	0.00064			
x_3x_4	1	0.00263			
Cross terms are interactions from the two replications of the 2^4.					

Table 6.18 Analysis of variance for second-order model with lack-of-fit terms for all factors.

Analysis of Variance					
Source	*df*	*SS*	*MS*	*F*	*P*
Quadratic	14	5.74740	0.41053	4.36	0.000
Lack of Fit	9	0.64283	0.07143	0.76	0.654
Error	36	3.38972	0.09416		
Total	59	9.77995			

All in all, it seems worthwhile to try to draw conclusions from the quadratic model on x_2 and x_3. Conclusions drawn from the quadratic model on all factors seem much more questionable. In the next section, we use both of these fitted models to illustrate methods for interpreting quadratic functions. It should be remembered that any conclusions drawn from the model on all four factors are questionable because of the relatively small regression F statistic.

6.4 Interpreting Quadratic Response Functions

In this section, we discuss methods for finding the maximum of a quadratic function and for examining the nature of a quadratic surface when the maximum or critical point is located far from the observed data. The discussion is considerably more sophisticated than earlier material as it relates to vector geometry. For $x' = (x_1, \ldots, x_f)$, consider a response function that is quadratic in the x_j's,

$$\mu(x) = \beta_0 + \sum_{j=1}^{f} \beta_j x_j + \sum_{j=1}^{f} \sum_{k \geq j} \beta_{jk} x_j x_k.$$

As before, we write

$$\beta_* = (\beta_1, \ldots, \beta_f)',$$

$$B = \begin{bmatrix} \beta_{11} & \beta_{12}/2 & \beta_{13}/2 & \cdots & \beta_{1q}/2 \\ \beta_{12}/2 & \beta_{22} & \beta_{23}/2 & \cdots & \beta_{2q}/2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_{1q}/2 & \beta_{2q}/2 & \beta_{3q}/2 & \cdots & \beta_{qq} \end{bmatrix},$$

and

$$\mu(x) = \beta_0 + x'\beta_* + x'Bx. \tag{1}$$

Again, we realize that this is only an approximation to the true response function and that β_0, β_*, and B must be estimated from experimental data.

To find a critical point, differentiate the response function (1) to give

$$[\mathbf{d}_x \mu(x) = \beta_* + 2Bx.$$

Setting the derivative equal to the 0 vector gives

$$\beta_* + 2Bx = 0, \tag{2}$$

so a critical point is located at the solution

$$x_0 \equiv -B^{-1}\beta/2. \tag{3}$$

From equation (2) note that

$$\beta_* = -2Bx_0,$$

so the value of the response function at the critical point is

$$\begin{aligned} \mu_0 &\equiv \mu(x_0) \\ &= \beta_0 + x_0'\beta_* + x_0'Bx_0 \\ &= \beta_0 - 2x_0'Bx_0 + x_0'Bx_0 \\ &= \beta_0 - x_0'Bx_0. \end{aligned}$$

This analysis assumes that an inverse matrix exists for B. In practice, estimates of B almost always have an inverse but the true matrix B may not. We will return to this issue in the next subsection.

Even the simple form of the response function in equation (1) is too complicated for a detailed analysis. For example, situations exist where changing any variable by itself leads to a decrease in yield but changing variables together increases yield. We need to rewrite the response function in its simplest form. First, recenter the quadratic function as

$$\begin{aligned}\mu(x) &= \beta_0 + x'\beta_* + x'Bx \\ &= \beta_0 - 2x'Bx_0 + x'Bx \\ &= \beta_0 - x_0'Bx_0 + x_0'Bx_0 - 2x'Bx_0 + x'Bx \\ &= \mu_0 + (x-x_0)'B(x-x_0).\end{aligned}$$

We now use the singular value decomposition to go a step further and write the symmetric matrix B as

$$B = PD(\lambda)P'$$

with P orthonormal (orthogonal) and $D(\lambda)$ diagonal. The elements of the vector λ are eigenvalues of B, and the columns of P are corresponding eigenvectors. Substituting for B gives

$$\begin{aligned}\mu(x) &= \mu_0 + (x-x_0)'PD(\lambda)P'(x-x_0) \\ &= \mu_0 + [P'(x-x_0)]D(\lambda)[P'(x-x_0)].\end{aligned}$$

Transforming x into

$$z = P'(x-x_0)$$

gives

$$\begin{aligned}\mu(x) &= \mu_0 + z'D(\lambda)z \\ &= \mu_0 + \sum_{j=1}^{f} \lambda_j z_j^2.\end{aligned} \tag{4}$$

Equation (4) is known as the *B canonical form* of the quadratic response function.

If the λ_j's that comprise the vector λ are all positive, the function increases as x differs from x_0, thus x_0 is the location of the minimum. If the λ_j's are all negative, the function decreases as x differs from x_0, so x_0 is the location of the maximum. If the λ_j's are both positive and negative, x_0 is the location of a saddlepoint. If the λ_j's were all negatives and zeros, a maximum could be attained for many different x values or the function could increase indefinitely. Having many values that maximize the function can be useful because in such situations the value of x can be chosen to minimize costs of production while attaining a maximum yield.

If any of the λ_j's are zero, the matrix B^{-1} does not exist and thus x_0 does not exist. The analysis just given breaks down. Such situations are discussed in the following subsection. In practice, estimated λ_j's are almost never zero but they are often close to zero. Investigating situations where some λ_j's equal zero sheds light on situations where their estimates are almost zero.

The location x_0 of a maximum is a key feature in modeling responses. If the maximum is attained close to the center of data collection (i.e., if x_0 is close to 0, we can have some faith in the estimated location). If x_0 is far from the center of data collection, it provides only a direction for further exploration. □

EXAMPLE 6.4.1. In the previous section, we completed checks on assumptions, on whether the fit is adequate for interpretation, and on lack of fit. In this example we examine the fitted quadratic model for x_2 and x_3. Recall that this model is both simpler and better estimated than the model with all four factors. The more complete model will be examined later.

Using the matrix notation of this chapter, Table 6.13 gives

$$\hat{\beta}_* = \begin{bmatrix} 0.27543 \\ 0.09902 \end{bmatrix} \quad \hat{B} = \begin{bmatrix} -0.11374 & -0.05275 \\ -0.05275 & -0.07562 \end{bmatrix}.$$

Applying equation (3) gives the location of the critical point as

$$\hat{x}_0 = (1.34097, -0.28069)',$$

with the corresponding value of the estimated response function

$$\hat{\mu}_0 = 16.9569.$$

The critical point is reasonably close to the center of the data $(x_2, x_3) = (0,0)$. Some of the design points are at the center, while the factorial points $(\pm 1, \pm 1)$ are a distance of

$$\sqrt{(\pm 1)^2 + (\pm 1)^2} = \sqrt{2} \doteq 1.414$$

from the center, and the star points are 2 units from the center. The distance of the critical point from the center is

$$\sqrt{\hat{x}_0' \hat{x}_0} = 1.37$$

units, so it is closer to the center than any of the actual observations other than those taken at the center.

Recall that to transform $\hat{x}_0$ from the x scale to the original ξ scale involves solving $x_j = (\xi_j - m_j)/s_j$ to get $\xi_j = m_j + s_j x_j$. In this problem, $m_2 = 4$, $s_2 = 1$, $m_3 = 6$, $s_3 = 1$, so $\hat{x}_0$ transforms into $\hat{\xi}_0 = (5.34097, 5.71931)'$.

The eigenvalues of $\hat{B}$ are given by

$$\hat{\lambda} = (-0.150768, -0.038592)'.$$

Both $\hat{\lambda}_j$'s are negative, so the critical point is a maximum and the estimated response surface gives a maximum achievable mean response of $\hat{\mu}_0 = 16.9569$. Of course, we have actually observed higher values for y, but we ascribe that to random variation.

The canonical form for the quadratic model is

$$\hat{y} = \hat{\mu}(z) = 16.9569 - 0.150768 z_1^2 - 0.038592 z_2^2,$$

where

$$z = P'(x - \hat{x}_0)$$

and

$$P = \begin{bmatrix} 0.818482 & -0.574533 \\ 0.574533 & 0.818482 \end{bmatrix}.$$

The maximum of $\hat{\mu}(z)$ is obtained at $z = (0,0)'$. Note that $\hat{\lambda}_1$ is about four times greater than $\hat{\lambda}_2$, so the model is $\sqrt{4} = 2$ times more sensitive to changes in the z_1 direction than to changes in the z_2 direction. Note that in Table 6.13 the standard errors for the quadratic terms $\hat{\beta}_{22}$ and $\hat{\beta}_{33}$ are both 0.03762. This value can also be used as a rough standard error for the $\hat{\lambda}_j$'s. By this criterion, $\hat{\lambda}_1$ is clearly different from zero, while $\hat{\lambda}_2$ shows little evidence of being different from zero.

The center of data collection as measured in the z variables is

$$z_c \equiv \begin{bmatrix} z_{1c} \\ z_{2c} \end{bmatrix} \equiv P'(0 - \hat{x}_0) = \\ - \begin{bmatrix} 0.818482 & -0.574533 \\ 0.574533 & 0.818482 \end{bmatrix} \begin{bmatrix} 1.34097 \\ -0.28069 \end{bmatrix} = \begin{bmatrix} -1.25882 \\ -0.54069 \end{bmatrix}. \quad (5)$$

If movement in the z_2 direction has relatively little effect, the most dramatic improvement in response with the *least* change in *current* operating conditions is obtained by moving from the current center $z_c = (z_{1c}, z_{2c})'$ to $(0, z_{2c})'$ — in other words, by moving from $z_c = (-1.25882, -0.54069)'$ to $(0, -0.54069)'$. We now find this point in terms of x. The matrix P is orthonormal, so $PP' = I$, $Pz = PP'(x - \hat{x}_0)$, and

$$x = Pz + \hat{x}_0.$$

We want to find

$$P \begin{pmatrix} 0 \\ z_{2c} \end{pmatrix} + \hat{x}_0.$$

From equation (5), writing P in terms of its columns, say $P = [P_1, P_2]$, gives $z_{2c} = -P_2'\hat{x}_0$, so the point we are looking for is

$$\begin{aligned} P \begin{bmatrix} 0 \\ -P_2'\hat{x}_0 \end{bmatrix} + \hat{x}_0 &= [P_1, P_2] \begin{bmatrix} 0 \\ -P_2'\hat{x}_0 \end{bmatrix} + [P_1, P_2] \begin{bmatrix} P_1' \\ P_2' \end{bmatrix} \hat{x}_0 \\ &= -P_2 P_2' \hat{x}_0 + P_1 P_1' \hat{x}_0 + P_2 P_2' \hat{x}_0 \\ &= P_1 P_1' \hat{x}_0. \end{aligned}$$

In this example,

$$P_1 P_1' \hat{x}_0 = \begin{bmatrix} 0.766339 \\ 0.537931 \end{bmatrix}.$$

(This is the perpendicular projection of $\hat{x}_0$ into $C(P_1)$.) Note that $P_1 P_1' \hat{x}_0$ is the scalar $P_1' \hat{x}_0$ times the vector P_1, so the most dramatic improvement in response with the least change in current operating conditions is obtained by moving a specified amount away from $x = (0,0)$ in the direction P_1. From this new point, we can investigate the effect of moving in the P_2 direction. In this example, additional improvement is possible by moving in the P_2 direction toward $\hat{x}_0$. When λ_2 is essentially

zero, no substantial improvement is possible by changes in the P_2 direction, so any operating conditions that give $z_{1c} = 0$ (i.e., any x values that give $0 = P_1'(x - \hat{x}_0)$ give optimal response). □

EXAMPLE 6.4.2. We now consider the quadratic model for all four factors. Even though the function is not well-estimated from our data, we will use the fitted polynomial to illustrate the ideas of interpreting second-order polynomials. In the matrix form used in this section, Table 6.10 gives the estimates of the coefficients of the quadratic response function as

$$\hat{\beta}_* = \begin{bmatrix} 0.04592 \\ 0.27543 \\ 0.09902 \\ 0.01921 \end{bmatrix}, \hat{B} = \begin{bmatrix} -0.03658 & 0.03890 & -0.03027 & -0.02468 \\ 0.03890 & -0.11649 & -0.10550 & 0.00447 \\ -0.03027 & -0.10550 & -0.07837 & 0.00906 \\ -0.02468 & 0.00447 & 0.00906 & 0.02910 \end{bmatrix}.$$

The critical point is located at

$$\hat{x}_0 = (-0.77552, -0.79934, 1.84107, -1.43821)',$$

with a response at the critical point of

$$\hat{\mu}_0 = 16.7576.$$

The distance of the critical point from the center of the design is $\sqrt{\hat{x}_0'\hat{x}_0} = 2.588$. The design points, other than centers, are all two units from the center, so the critical point is substantially farther from the center than any of the design points. This suggests that we should not put great faith in the precise location of the critical point. We can, however, use the fitted model, including the critical point, to inform us about the behavior of the fitted model in the region of data collection.

Transforming $\hat{x}_0$ from the x scale to the original ξ involves solving $\xi_j = m_j + s_j x_j$. In this problem, $(m_1, m_2, m_3, m_4) = (1, 4, 6, 1.5)$ and $(s_1, s_2, s_3, s_4) = (.5, 1, 1, 0.5)$, so $\hat{x}_0$ becomes $\hat{\xi}_0 = (0.61224, 5.34097, -0.28069, 0.780895)'$.

The eigenvalues of $\hat{B}$ are

$$\hat{\lambda} = (-0.205831, -0.069644, 0.053113, 0.020022)'.$$

The $\hat{\lambda}_j$'s are both positive and negative, so the critical point is a saddlepoint. The canonical form for the quadratic model is

$$\hat{y} = \hat{\mu}(z) = 16.7576 - 0.205831 z_1^2 - 0.069644 z_2^2 + 0.053113 z_3^2 + 0.020022 z_4^2,$$

where

$$z = P'(x - \hat{x}_0)$$

and

$$P = \begin{bmatrix} 0.072618 & 0.842179 & -0.497374 & 0.195141 \\ -0.774280 & -0.261909 & -0.390255 & 0.423787 \\ -0.626938 & 0.434557 & 0.470081 & -0.443996 \\ 0.046538 & 0.182478 & 0.615909 & 0.764978 \end{bmatrix}.$$

The critical point is obtained at $z = (0,0,0,0)'$. Note that $|\hat{\lambda}_1|$ is about three times greater than $|\hat{\lambda}_2|$, four times greater than $|\hat{\lambda}_3|$, and ten times greater than $|\hat{\lambda}_4|$. In particular, the model is $\sqrt{10} \doteq 3$ times more sensitive to changes in the z_1 direction than to changes in the z_4 direction. From Table 6.10, the standard errors for the quadratic terms $\hat{\beta}_{jj}$ are all 0.04042. This value can be used as a rough standard error for the $\hat{\lambda}_j$'s. By this standard, $\hat{\lambda}_1$ is clearly different from zero, while the other $\hat{\lambda}_j$'s show little evidence of being different from zero. For this reason we focus on movements only in the z_1 direction.

The center of data collection in the z variables is

$$\begin{aligned} z_c &\equiv \begin{bmatrix} z_{1c} \\ z_{2c} \\ z_{3c} \\ z_{4c} \end{bmatrix} \equiv P'(0 - \hat{x}_0) \\ &= -\begin{bmatrix} 0.072618 & 0.842179 & -0.497374 & 0.195141 \\ -0.774280 & -0.261909 & -0.390255 & 0.423787 \\ -0.626938 & 0.434557 & 0.470081 & -0.443996 \\ 0.046538 & 0.182478 & 0.615909 & 0.764978 \end{bmatrix} \begin{bmatrix} -0.77552 \\ -0.79934 \\ 1.84107 \\ -1.43821 \end{bmatrix} \\ &= \begin{bmatrix} 1.92586 \\ 0.51816 \\ -1.64286 \\ 0.14822 \end{bmatrix}. \end{aligned}$$

If movements in the z_2, z_3, and z_4 directions have relatively little effect, the most dramatic change in response with the least change in current operating conditions is obtained by moving from the current center $z_c = (z_{1c}, z_{2c}, z_{3c}, z_{4c})'$ to $(0, z_{2c}, z_{3c}, z_{4c})'$. Writing $P = [P_1, P_2, P_3, P_4]$, an argument similar to that in the previous example gives this new point in terms of x as

$$P_1 P_1' \hat{x}_0 = \begin{bmatrix} -0.047824 \\ 0.509919 \\ 0.412884 \\ -0.030649 \end{bmatrix}.$$

Again, the most dramatic change in response with the least change in current operating conditions is obtained by moving a specified amount away from $x = (0,0,0,0)$ in the direction P_1. From this point, we need to investigate the effect of moving in the P_2, P_3, and P_4 directions. The response may increase with changes in some or all of these directions, and the rate of increase typically varies with the direction. To analyze this phenomenon further requires an alternative canonical form discussed in the next subsection.

In both this example and Example 6.4.1, the direction to move for the most dramatic change in response with the least change in current operating conditions (i.e., the direction P_1), involves changing x_2 about 8 units for every 6 unit change in x_3 with little or no change in x_1 or x_4. Comparing the overall maximum of Example 6.4.1 to the current example, the point that maximized the response in x_2 and x_3 alone gives the estimated response $\hat{\mu}(0, 1.34097, -.28069, 0) = 17.0134$, which is greater than the saddlepoint response $\hat{\mu}_0$. □

EXAMPLE 6.4.3. We again consider the quadratic model for all factors, but this time we consider only $\hat{\lambda}_4$ as being not substantially different from zero. As before, the center of data collection in the z variables is

$$z_c \equiv \begin{bmatrix} z_{1c} \\ z_{2c} \\ z_{3c} \\ z_{4c} \end{bmatrix} \equiv P'(0 - \hat{x}_0),$$

with P and $\hat{x}_0$ as given in the previous example. This time, we consider only movement in the z_4 direction as having an insubstantial effect on response. The most dramatic change in response with the least change in current operating conditions is obtained by moving from the current center $z_c = (z_{1c}, z_{2c}, z_{3c}, z_{4c})'$ to $(0, 0, 0, z_{4c})'$. We now find this point in terms of x. Again writing $P = [P_1, P_2, P_3, P_4]$, an argument similar to that in Example 6.4.1 gives the point as

$$x = \sum_{j=1}^{3} P_j P_j' \hat{x}_0 = \begin{bmatrix} -0.305676 \\ 0.221018 \\ 0.772054 \\ 0.403639 \end{bmatrix},$$

where the vectors P_1, P_2, and P_3 are used because they correspond to zeroing out z_{1c}, z_{2c}, and z_{3c}. A matrix equivalence establishes that the point is also $(I - P_4 P_4')\hat{x}_0$, where the vector P_4 is used because it was not zeroed out. The most dramatic change in response with the least change in current operating conditions is obtained by moving a specified amount away from $(0, 0, 0, 0)$ to $P_1(P_1'\hat{x}_0) + P_2(P_2'\hat{x}_0) + P_3(P_3'\hat{x}_0)$. This involves movement in all three of the directions P_1, P_2, and P_3. From this point, we need to investigate the effect of moving in the P_4 direction. □

6.4.1 Noninvertible B

When B has no inverse, x_0 does not exist, so the canonical form (4) cannot be used. We can write an alternative canonical form that does not involve recentering but instead uses the transformation

$$\tilde{z} = P'x.$$

The center of the data is still at $(0,\ldots,0)'$ under this transformation. Observing that $PP' = I$ and $B = PD(\lambda)P'$ allows us to write $\mu(x) = \beta_0 + x'\beta_* + x'Bx$ as

$$\mu(\tilde{z}) = \beta_0 + \tilde{z}'P'\beta_* + \tilde{z}'D(\lambda)\tilde{z}.$$

Now, defining

$$\theta \equiv P'\beta_*$$

gives the *A canonical form*

$$\begin{aligned}\mu(\tilde{z}) &= \beta_0 + \tilde{z}'\theta + \tilde{z}'D(\lambda)\tilde{z} \\ &= \beta_0 + \sum_{j=1}^{f}\theta_j\tilde{z}_j + \sum_{j=1}^{f}\lambda_j\tilde{z}_j^2.\end{aligned}$$

Suppose now that all the λ_j's are negative except for $\lambda_{j'} = 0$. If $\theta_{j'}$ is also zero, then $\tilde{z}_{j'}$ has no effect on the response surface. There is a maximum value for

$$\beta_0 + \sum_{j\neq j'}^{f}\theta_j\tilde{z}_j + \sum_{j\neq j'}^{f}\lambda_j\tilde{z}_j^2$$

that depends only on the other $\tilde{z}_j$'s, so $\tilde{z}_{j'}$ can be chosen to minimize costs of production. Obviously, if more than one of the (λ_j, θ_j) pairs are both zero, a similar analysis holds. In these cases, $\mu(\tilde{z})$ is said to have a *stationary ridge*.

Now, suppose that all the λ_j's are negative except for $\lambda_{j'} = 0$, but $\theta_{j'} \neq 0$. For any fixed $\tilde{z}_{j'}$, we can maximize the function. We need only consider behavior in the $\tilde{z}_{j'}$ direction. If $\theta_{j'}$ is positive, the response will increase indefinitely as $\tilde{z}_{j'}$ increases, and if $\theta_{j'}$ is negative, the response will increase indefinitely as $\tilde{z}_{j'}$ decreases. This situation is known as a *rising ridge*.

EXAMPLE 6.4.4. Consider again the quadratic response function in x_2 and x_3. In practice, the smallest $\hat{\lambda}_j$ values are almost never zero, but they are often close enough to zero that the estimated response function behaves as if some $\hat{\lambda}_j$'s were zero. In this example, $\hat{\lambda}_2$ is reasonably small. We examine what can be learned from treating it as though it were zero.

In Example 6.4.1, we gave the matrices P and $\hat{\beta}_*$. Multiplying $P'\hat{\beta}_*$ gives

$$\hat{\theta} = \begin{bmatrix} 0.282325 \\ -0.077198 \end{bmatrix},$$

and the A canonical form becomes

$$\hat{y} = \hat{\mu}(\tilde{z}) = 16.7861 + 0.282325\tilde{z}_1 - 0.077198\tilde{z}_2 - 0.150768\tilde{z}_1^2 - 0.038592\tilde{z}_2^2.$$

It is easily seen that when the z transformation exists, $\tilde{z} = z - z_c$. In other words, the $\tilde{z}$ transformation takes the z transformation and recenters it at the orig-

inal x origin. In the z transform, the center of the data for this example was at $z_c = (-1.25882, -0.54069)'$.

We saw in Example 6.4.1 that by treating $\hat{\lambda}_2$ as zero, the point nearest the center of the data with maximum response was $(x_2, x_3) = (0.766339, 0.537931)$, or equivalently $z = (0, -0.54069)'$. In the $\tilde{z}$ transform, this is

$$\tilde{z} = z - z_c = (1.25882, 0)'.$$

From this point, we indicated that one needs to examine the behavior of the function in the x direction P_2. A change in this direction involves a change in $\tilde{z}_2$ but not in $\tilde{z}_1$. The estimated response when treating $\hat{\lambda}_2$ as zero and allowing only changes in the z_2 direction is

$$\hat{y} = \hat{\mu}(\tilde{z}) = 16.7861 + 0.282325(1.25882) - 0.077198\tilde{z}_2 - 0.150768(1.25882)^2.$$

Clearly, increasing the value of $\tilde{z}_2$ causes the response to decrease. To improve response from $(x_2, x_3) = (0.766339, 0.537931)$, we need to change $(x_2, x_3)'$ so that $\tilde{z}_2$ decreases. Recalling that $P_2 = (-0.574533, 0.818482)'$, a change of $\tilde{z}_2$ units moves from $(x_2, x_3) = (0.766339, 0.537931)$ to $(x_2, x_3) = (0.766339 - \tilde{z}_2 0.574533, 0.537931 + \tilde{z}_2 0.818482)$. □

EXAMPLE 6.4.5. Consider the quadratic response function in all four variables. In Example 6.4.2, we gave the matrices P and $\hat{\beta}_*$. Multiplying $P'\hat{\beta}_*$ gives

$$\hat{\theta} = \begin{bmatrix} -0.271111 \\ 0.013070 \\ -0.071948 \\ 0.096415 \end{bmatrix},$$

and the A canonical form becomes

$$\begin{aligned} \hat{y} = \hat{\mu}(\tilde{z}) = 16.8081 &- 0.271111\tilde{z}_1 + 0.013070\tilde{z}_2 - 0.071948\tilde{z}_3 + 0.096415\tilde{z}_4 \\ &- 0.205831\tilde{z}_1^2 - 0.069644\tilde{z}_2^2 + 0.053113\tilde{z}_3^2 + 0.020022\tilde{z}_4^2. \end{aligned}$$

We saw in Example 6.4.2 that by treating $\hat{\lambda}_2$, $\hat{\lambda}_3$, and $\hat{\lambda}_4$ as zero, the point nearest the center of the data (i.e., nearest $z_c = (1.92586, 0.51816, -1.64286, 0.14822)'$) that has maximum response is $z = (0, 0.51816, -1.64286, 0.14822)'$, or equivalently $(x_1, x_2, x_3, x_4)' = (-0.047824, 0.509919, 0.412884, -0.030649)'$. In the $\tilde{z}$ transform, this is

$$\tilde{z} = z - z_c = (-1.92586, 0, 0, 0)'.$$

From this point, one needs to examine the behavior of the function in the $\tilde{z}_2$, $\tilde{z}_3$, and $\tilde{z}_4$ directions. The estimated response when treating $\hat{\lambda}_2$, $\hat{\lambda}_3$, and $\hat{\lambda}_4$ as zero and allowing only changes in the $\tilde{z}_2$, $\tilde{z}_3$, and $\tilde{z}_4$ directions is

$$\begin{aligned} \hat{y} = \hat{\mu}(\tilde{z}) = 16.8081 &- 0.271111(-1.92586) \\ &+ 0.013070\tilde{z}_2 - 0.071948\tilde{z}_3 + 0.096415\tilde{z}_4 - 0.205831(-1.92586)^2. \end{aligned}$$

Clearly, increasing the value of $\tilde{z}_2$ causes the response to increase, as does increasing the value of $\tilde{z}_4$. Increasing the value of $\tilde{z}_3$ causes the response to decrease. The largest increases in response per unit change in $\tilde{z}_j$ come from increasing $\tilde{z}_4$, while decreasing $\tilde{z}_3$ is a reasonably close second. Very little occurs when changing $\tilde{z}_2$ because the coefficient 0.013070 is so small. □

Chapter 7
Recovery of Interblock Information in BIB Designs

A version of this material appeared in the first four editions of PA.

Consider the analysis of a balanced incomplete block (BIB) design in which blocks are random effects. This analysis is known as the *recovery of interblock information*. The BIB analysis in which blocks are fixed effects is known as the intrablock analysis and is discussed in PA Section 9.4.

We begin by fixing notation. The model for a BIB is

$$y_{ij} = \mu + \beta_i + \tau_j + e_{ij},$$

$i = 1,\ldots,b$, with $j \in D_i$ or, equivalently, $j = 1,\ldots,t$, with $i \in A_i$. Here β and τ indicate block and treatment effects, respectively, D_i is the set of treatment indices for block i, and A_j is the set of indices for blocks that contain treatment j. The model is written using matrices as

$$Y = J\mu + X\beta + Z\tau + e,$$

where μ, β, and τ are the grand mean, block effects vector, and treatment effects vector, respectively. The matrix X is the matrix of indicators for the blocks and can be written as

$$X = [x_{ij,m}], \qquad x_{ij,m} = \delta_{im}.$$

Here the columns of X are $m = 1,\ldots,b$, the pair ij identifies a row of the matrix, and δ_{ab} for any two symbols a and b is 1 if $a = b$ and 0 if $a \neq b$. Z is a matrix of indicators for the treatments and is defined as $Z = [z_{ij,r}]$ with $z_{ij,r} = \delta_{jr}$, $r = 1,\ldots,t$, and the pair ij denoting a row of the matrix.

Recall two fundamental relations necessary for a BIB,

$$rt = bk$$

and

$$(t-1)\lambda = r(k-1),$$

where r is the number of replications for each treatment, k is the number of units in each block, and λ is the number of times any two treatments occur in the same block.

In the mixed model, β is a random vector with $\mathrm{E}(\beta) = 0$, $\mathrm{Cov}(\beta) = \sigma_B^2 I_b$, and $\mathrm{Cov}(\beta, e) = 0$. In a slight change of notation write $\mathrm{Cov}(e) = \sigma_e^2 I_n$, where $n = rt = bk$. Combining the random effects, write $\eta = X\beta + e$ and the model as

$$Y = Z\tau + \eta, \quad \mathrm{E}(\eta) = 0, \quad \mathrm{Cov}(\eta) = \sigma_e^2 I_n + \sigma_B^2 XX'. \tag{1}$$

Note that we have dropped the grand mean, thus removing the overparameterization associated with the treatment effects. In other words, we are using the model $y_{ij} = \tau_j + \eta_{ij}$, where $\eta_{ij} \equiv \beta_j + e_{ij}$ is the random error term.

As in PA Chapter 11, write $\sigma^2 = \sigma_e^2 + \sigma_B^2$ and $\rho = \sigma_B^2/(\sigma_e^2 + \sigma_B^2)$. It follows that $\sigma_e^2 = \sigma^2(1-\rho)$ and $\sigma_B^2 = \sigma^2\rho$. A term that frequently appears in the analysis is the interblock (between cluster) error term,

$$\sigma^2[(1-\rho) + k\rho] = \sigma_e^2 + k\sigma_B^2.$$

With the notation given earlier, write

$$\mathrm{Cov}(\eta) = \sigma^2 V,$$

where, again as in Chapter 11,

$$\begin{aligned} V &= \left[(1-\rho)I + \rho XX'\right] \\ &= \left[(1-\rho)I + k\rho M\right] \end{aligned}$$

and M is the perpendicular projection operator onto $C(X)$.

7.1 Estimation

In this subsection we derive the BLUE of τ. From Section 2.7,

$$\begin{aligned} Z\hat{\tau} &= AY \\ &= Z\left(Z'V^{-1}Z\right)^{-1} Z'V^{-1}Y. \end{aligned}$$

Note that finding $\hat{\tau}$ is essentially equivalent to finding the oblique projection operator A. Given $\hat{\tau}$ we can easily find $Z\hat{\tau}$; thus we know AY. With AY known for any vector Y, the matrix A is completely characterized. Finding $\hat{\tau} = \left(Z'V^{-1}Z\right)^{-1} Z'V^{-1}Y$ requires computation of both V^{-1} and $\left(Z'V^{-1}Z\right)^{-1}$. These computations are facilitated by the following result.

Proposition 7.1.1. Let P be a projection operator (idempotent), and let a and b

be real numbers. Then

$$[aI+bP]^{-1} = \frac{1}{a}\left[I - \frac{b}{a+b}P\right].$$

PROOF.

$$\frac{1}{a}\left[I - \frac{b}{a+b}P\right][aI+bP] = \frac{1}{a}\left[aI+bP - \frac{ab}{a+b}P - \frac{b^2}{a+b}P\right] = I. \qquad \square$$

Using this result, we obtain two forms for V^{-1}:

$$V^{-1} = \frac{1}{1-\rho}\left[I - \frac{k\rho}{[(1-\rho)+k\rho]}M\right] = \frac{1}{1-\rho}\left[(I-M) + \frac{1-\rho}{[(1-\rho)+k\rho]}M\right]. \quad (2)$$

Both forms will be used frequently.

We now compute $Z'V^{-1}Z$ and $\left(Z'V^{-1}Z\right)^{-1}$. First note that the fundamental relation $(t-1)\lambda = r(k-1)$ implies

$$\lambda t = rk - (r-\lambda).$$

Using this equality and the characterizations

$$Z'Z = rI_t, \qquad Z'MZ = \frac{1}{k}\left[(r-\lambda)I + \lambda J_t^t\right]$$

(proven in PA Section 9.4), the first form for V^{-1} in equation (2), and writing $P_t = \frac{1}{t}J_t^t$, we get

$$\begin{aligned}
(Z'V^{-1}Z) &= \frac{1}{1-\rho}\left[Z'Z - \frac{k\rho}{[(1-\rho)+k\rho]}Z'MZ\right] \\
&= \frac{1}{1-\rho}\left[rI - \frac{\rho}{[(1-\rho)+k\rho]}\{(r-\lambda)I + \lambda tP_t\}\right] \\
&= \frac{1}{1-\rho}\left[\frac{r(1-\rho)}{[(1-\rho)+k\rho]}I + \frac{\lambda t\rho}{[(1-\rho)+k\rho]}(I-P_t)\right] \\
&= \frac{1}{(1-\rho)[(1-\rho)+k\rho]}[r(1-\rho)I + \lambda t\rho(I-P_t)].
\end{aligned}$$

From Proposition 7.11.1,

$$\begin{aligned}
(Z'V^{-1}Z)^{-1} &= \frac{[(1-\rho)+k\rho]}{r}\left[I - \frac{\lambda t\rho}{r(1-\rho)+\lambda t\rho}(I-P_t)\right] \quad (3) \\
&= \frac{[(1-\rho)+k\rho]}{r}\left[\frac{r(1-\rho)}{r(1-\rho)+\lambda t\rho}I + \frac{\lambda t\rho}{r(1-\rho)+\lambda t\rho}P_t\right].
\end{aligned}$$

Rather than computing $\hat{\tau} = (Z'V^{-1}Z)^{-1}Z'V^{-1}Y$ directly, it is convenient to decompose $Z'V^{-1}Y$ into intrablock and interblock components. The intrablock component is related to the fixed block effect analysis. The interblock component is what is left. The fixed block effect analysis is based on

$$Y'(I-M)Z = (Q_1, \ldots, Q_t)$$

where, cf. PA Section 9.4,

$$Q_m = \sum_{i \in A_m} (y_{im} - \bar{y}_{i\cdot}).$$

Define

$$Q = (Q_1, \ldots, Q_t)',$$

so $Q \equiv Z'(I-M)Y$. Similarly, define

$$W = (W_1, \ldots, W_t)',$$

where $W \equiv Z'MY$. M is the perpendicular projection operator for the one-way ANOVA in blocks (ignoring treatments) so

$$MY = [t_{ij}], \qquad t_{ij} = \bar{y}_{i\cdot},$$

with

$$\bar{y}_{i\cdot} \equiv \frac{1}{k} \sum_{j \in D_i} y_{ij}.$$

Z is a matrix of treatment indicators, so $Z'MY$ yields

$$W_j = \sum_{i \in A_j} \bar{y}_{i\cdot}.$$

In particular,

$$Q_j = \sum_{i \in A_j} (y_{ij} - \bar{y}_{i\cdot}) = \sum_{i \in A_j} y_{ij} - W_j.$$

In computing $\hat{\tau}$, we will also have occasion to use

$$\overline{W}_{\cdot} \equiv \frac{1}{t} \sum_{j=1}^{t} W_j$$

and the fact that

$$\overline{Q}_{\cdot} \equiv \frac{1}{t} \sum_{j=1}^{t} Q_j = \frac{1}{t} J_t' Z'(I-M)Y = \frac{1}{t} J_n'(I-M)Y = 0.$$

Using the second form of V^{-1} given in (2),

$$Z'V^{-1}Y = \frac{1}{1-\rho}Z'(I-M)Y + \frac{1}{(1-\rho)+k\rho}Z'MY \tag{4}$$
$$= \frac{1}{1-\rho}Q + \frac{1}{(1-\rho)+k\rho}W.$$

Finally, using (3) and (4),

$$\begin{aligned}\hat{\tau} &= (Z'V^{-1}Z)^{-1}Z'V^{-1}Y \\ &= \frac{[(1-\rho)+k\rho]}{r}\left[\frac{r(1-\rho)}{r(1-\rho)+\lambda t\rho}I + \frac{\lambda t\rho}{r(1-\rho)+\lambda t\rho}P_t\right]Z'V^{-1}Y \\ &= \frac{[(1-\rho)+k\rho](1-\rho)}{r(1-\rho)+\lambda t\rho}Z'V^{-1}Y + \frac{[(1-\rho)+k\rho]\lambda t\rho}{r[r(1-\rho)+\lambda t\rho]}P_tZ'V^{-1}Y \\ &= \frac{[(1-\rho)+k\rho]}{r(1-\rho)+\lambda t\rho}Q + \frac{(1-\rho)}{r(1-\rho)+\lambda t\rho}W + \frac{\lambda t\rho\overline{W}_{\cdot}}{r[r(1-\rho)+\lambda t\rho]}J_t.\end{aligned}$$

The last equality comes from $P_tQ = \overline{Q}_{\cdot}J_t = 0$ and $P_tW = \overline{W}_{\cdot}J_t$. In particular, an individual component of $\hat{\tau}$ is

$$\hat{\tau}_j = \frac{[(1-\rho)+k\rho]}{r(1-\rho)+\lambda t\rho}Q_j + \frac{(1-\rho)}{r(1-\rho)+\lambda t\rho}W_j + \frac{\lambda t\rho\overline{W}_{\cdot}}{r[r(1-\rho)+\lambda t\rho]}.$$

For purposes of comparing treatments, the term involving $\overline{W}_{\cdot}$, which is constant, can be dropped. Finally, the projection operator is characterized by

$$AY = Z\hat{\tau} = [t_{ij}], \qquad t_{ij} = \hat{\tau}_j.$$

There are three additional aspects of the analysis to consider. First, we need to consider testing the hypothesis $\tau_1 = \cdots = \tau_t$. Second, we need to examine contrasts. Third, we need to deal with the fact that our estimate of $\hat{\tau}$ is useless. The estimate depends on $\rho = \sigma_B^2/(\sigma_e^2+\sigma_B^2)$. This is an unknown parameter. Writing $\hat{\tau} = (\sigma^2/\sigma^2)\hat{\tau}$, and using $\sigma^2(1-\rho) = \sigma_e^2$ and $\sigma^2\rho = \sigma_B^2$, gives

$$\hat{\tau} = \frac{\sigma_e^2+k\sigma_B^2}{r\sigma_e^2+\lambda t\sigma_B^2}Q + \frac{\sigma_e^2}{r\sigma_e^2+\lambda t\sigma_B^2}W + \frac{\lambda t\sigma_B^2\overline{W}_{\cdot}}{r\sigma_e^2+\lambda t\sigma_B^2}J_t. \tag{5}$$

Model (1) is a mixed model, so the methods of this chapter can be used to estimate σ_e^2 and σ_B^2. The variance estimates can be substituted into (5) to give a usable estimate of τ. Traditionally, Henderson's Method 3 has been used to obtain variance estimates. The use of Henderson's Method 3 will be discussed in detail later. Tests of models and examination of contrasts will be discussed as if σ_e^2 and σ_B^2 (hence σ^2 and ρ) were known. A discussion in which only ρ is assumed known is also given. Throughout we assume that $\eta \equiv X\beta + e \sim N(0, \sigma^2V)$.

7.2 Model Testing

We desire to test model (1) against the reduced model

$$Y = J\mu + \eta. \tag{6}$$

In particular, we will show that an α level test rejects H_0 if

$$\frac{r\sigma_e^2 + \lambda t\sigma_B^2}{\sigma_e^2\left[\sigma_e^2 + k\sigma_B^2\right]} \sum_{i=1}^{t} (\hat{\tau}_j - \tilde{\tau}_\cdot) > \chi^2(1-\alpha, t-1), \tag{7}$$

where $\tilde{\tau}_\cdot = \sum_{j=1}^{t} \hat{\tau}_j/t$ is the mean of the $\hat{\tau}_j$s. The remainder of this subsection is devoted to showing that this is the appropriate test.

We begin by finding the BLUE of $J\mu$ from model (6). The BLUE is $J\hat{\mu} = A_0 Y = J\left(J'V^{-1}J\right)^{-1} J'V^{-1}Y$. However, if we show that $C(VJ) \subset C(J)$, we can apply PA Proposition 2.7.5 or Theorem 10.4.5 to see that the simple least squares estimate $J\hat{\mu}$ with $\hat{\mu} = \sum_{ij} y_{ij}/rt$ is the BLUE. Because $J \in C(X)$, $VJ = [(1-\rho)I + k\rho M]J = (1-\rho)J + k\rho J = [(1-\rho) + k\rho]J \in C(J)$.

From PA Corollary 3.8.3,

$$Y'(A - A_0)'V^{-1}(A - A_0)Y \big/ \sigma^2 \sim \chi^2(t-1, 0)$$

if and only if model (6) is true. We wish to show that the test statistic is identical to that used in (7). Our argument involves five algebraic identities. First,

$$\sum_{j=1}^{t} (\hat{\tau}_j - \tilde{\tau}_\cdot)^2 = \hat{\tau}'(I - P_t)\hat{\tau}.$$

Second, we show that $\hat{\mu} = \tilde{\tau}_\cdot$. Using the second form for V^{-1} in (2) gives $V^{-1}J = [(1-\rho) + k\rho]^{-1}J$; also recall that $J \in C(Z)$ so $AJ = J$. These equalities lead to the result.

$$\begin{aligned}
\tilde{\tau}_\cdot &= \frac{1}{t} J_t' \hat{\tau} = \frac{1}{rt} J'Z\hat{\tau} = \frac{1}{rt} J'AY \\
&= \frac{(1-\rho) + k\rho}{rt} J'V^{-1}AY \\
&= \frac{(1-\rho) + k\rho}{rt} J'A'V^{-1}Y \\
&= \frac{(1-\rho) + k\rho}{rt} J'V^{-1}Y \\
&= \frac{1}{rt} J'Y = \hat{\mu}.
\end{aligned}$$

Third,

$$\begin{aligned}\hat{\tau}'(I-P_t)\hat{\tau} &= \hat{\tau}'\hat{\tau} - t\left(\frac{1}{t}\hat{\tau}'J_t'\right)^2 \\ &= \hat{\tau}'\hat{\tau} - t\hat{\mu}^2 \\ &= [\hat{\tau} - J_t\hat{\mu}]'[\hat{\tau} - J_t\hat{\mu}].\end{aligned}$$

Fourth, recall from Section 9.4 that

$$Z'(I-M)Z = \frac{\lambda t}{k}(I-P_t),$$

and, finally, from Section 9.4, and because $r-\lambda = rk-\lambda t$,

$$\begin{aligned}Z'MZ &= \frac{1}{k}[(r-\lambda)I + \lambda t P_t] \\ &= \frac{1}{k}[rkI - \lambda t(I-P_t)].\end{aligned}$$

Using the second form of V^{-1} in (2),

$$\begin{aligned}&Y'(A-A_0)'V^{-1}(A-A_0)Y \\ &= [Z\hat{\tau} - J\hat{\mu}]'V^{-1}[Z\hat{\tau} - J\hat{\mu}] \\ &= \frac{1}{1-\rho}[Z\hat{\tau} - J\hat{\mu}]'(I-M)[Z\hat{\tau} - J\hat{\mu}] + \frac{1}{(1-\rho)+k\rho}[Z\hat{\tau} - J\hat{\mu}]'M[Z\hat{\tau} - J\hat{\mu}] \\ &= \frac{1}{1-\rho}\hat{\tau}'Z'(I-M)Z\hat{\tau} + \frac{1}{(1-\rho)+k\rho}[Z\hat{\tau} - ZJ_t\hat{\mu}]'M[Z\hat{\tau} - ZJ_t\hat{\mu}] \\ &= \frac{1}{1-\rho}\frac{\lambda t}{k}\hat{\tau}'(I-P_t)\hat{\tau} + \frac{1}{(1-\rho)+k\rho}[\hat{\tau} - J_t\hat{\mu}]'Z'MZ[\hat{\tau} - J_t\hat{\mu}] \\ &= \frac{1}{1-\rho}\frac{\lambda t}{k}\hat{\tau}'(I-P_t)\hat{\tau} + \frac{1}{(1-\rho)+k\rho}[\hat{\tau} - J_t\hat{\mu}]'\left\{\frac{1}{k}[rkI - \lambda t(I-P_t)]\right\}[\hat{\tau} - J_t\hat{\mu}] \\ &= \frac{\lambda t}{k}\hat{\tau}'(I-P_t)\hat{\tau}\left\{\frac{1}{1-\rho} - \frac{1}{(1-\rho)+k\rho}\right\} + \frac{r}{(1-\rho)+k\rho}[\hat{\tau} - J_t\hat{\mu}]'[\hat{\tau} - J_t\hat{\mu}] \\ &= \frac{k\rho}{[(1-\rho)+k\rho](1-\rho)}\frac{\lambda t}{k}\hat{\tau}'(I-P_t)\hat{\tau} + \frac{r}{(1-\rho)+k\rho}\hat{\tau}'(I-P_t)\hat{\tau} \\ &= \frac{r(1-\rho)+\lambda t\rho}{(1-\rho)[(1-\rho)+k\rho]}\hat{\tau}'(I-P_t)\hat{\tau} \\ &= \frac{r\sigma^2(1-\rho)+\lambda t\sigma^2\rho}{\sigma^2(1-\rho)[(1-\rho)+k\rho]}\hat{\tau}'(I-P_t)\hat{\tau} \\ &= \frac{r\sigma_e^2+\lambda t\sigma_B^2}{\sigma_e^2[(1-\rho)+k\rho]}\hat{\tau}'(I-P_t)\hat{\tau}.\end{aligned}$$

Dividing by σ^2 gives

$$\frac{Y'(A-A_0)'V^{-1}(A-A_0)Y}{\sigma^2} = \frac{r\sigma_e^2+\lambda t\sigma_B^2}{\sigma_e^2\left[\sigma_e^2+k\sigma_B^2\right]}\,\hat{\tau}'(I-P_t)\hat{\tau},$$

which is the test statistic in (7). In practice, estimates of σ_e^2 and σ_B^2 are used to compute both the multiplier and $\hat{\tau}$. The substitution is then ignored and the χ^2 test is conducted as if σ_e^2 and σ_B^2 were known.

7.3 Contrasts

Model (1) is a one-way ANOVA model with an unusual covariance structure. However, estimable functions do not depend on the covariance matrix, so contrasts are estimable. This is true regardless of whether μ is included as a parameter in the model. A contrast is a linear parametric function $\xi'\tau$ with $\xi'J_t = 0$. The estimate is $\xi'\hat{\tau} = \sum_{j=1}^{t}\xi_j'\hat{\tau}_j$, where $\hat{\tau}$ has already been characterized.

We need to compute the variance of the estimate. Recall that with $\xi'J_t = 0$ we have $\xi'P_t = 0$. Using the second form in (3),

$$\begin{aligned}
\mathrm{Var}\left(\xi'\hat{\tau}\right) &= \xi'\mathrm{Cov}(\hat{\tau})\xi \\
&= \sigma^2\xi'\left(Z'V^{-1}Z\right)^{-1}\xi \\
&= \sigma^2\frac{[(1-\rho)+k\rho](1-\rho)}{r(1-\rho)+\lambda t\rho}\xi'\xi + \sigma^2\frac{[(1-\rho)+k\rho]\lambda t\rho}{r[r(1-\rho)+\lambda t\rho]}\xi'P_t\xi \\
&= \frac{\left[\sigma^2(1-\rho)+k\sigma^2\rho\right]\sigma^2(1-\rho)}{r\sigma^2(1-\rho)+\lambda t\sigma^2\rho}\xi'\xi \\
&= \frac{\left[\sigma_e^2+k\sigma_B^2\right]\sigma_e^2}{r\sigma_e^2+\lambda t\sigma_B^2}\xi'\xi.
\end{aligned}$$

Note that the variance can also be written as

$$\mathrm{Var}(\xi'\hat{\tau}) = \sigma_e^2\frac{[(1-\rho)+k\rho]}{r(1-\rho)+\lambda t\rho}\xi'\xi;$$

this second form will be used in the next subsection. Under normality,

$$\xi'\hat{\tau} \sim N\left(\xi'\tau, \frac{\left[\sigma_e^2+k\sigma_B^2\right]\sigma_e^2}{r\sigma_e^2+\lambda t\sigma_B^2}\xi'\xi\right). \tag{8}$$

In practice, estimates of σ_e^2 and σ_B^2 are substituted to find $\hat{\tau}$ and the estimated variance. Tests and confidence intervals are conducted using the distribution (8), ignoring the fact that estimates have been substituted for σ_e^2 and σ_B^2.

7.4 Alternative Inferential Procedures

Traditionally, statistical inferences have been conducted using the distributions in (7) and (8). These are based on the incorrect assumption that both σ_e^2 and σ_B^2 are known. Some improvement is made by assuming that only $\rho = \sigma_B^2/(\sigma_e^2 + \sigma_B^2)$ is known while σ^2 is unknown. In particular, it follows from Section 11.1 that the model with both fixed block effects δ and random block effects β, i.e.,

$$Y = J\mu + X\delta + Z\tau + \eta,$$

provides an estimate of $\sigma^2(1-\rho) = \sigma_e^2$. This estimate is $\hat{\sigma}_e^2$, the mean squared error for the fixed block effect model of Section 9.4.

The key results are that, under model (1),

$$\frac{\hat{\sigma}_e^2}{\sigma_e^2} \sim \frac{\chi^2(rt-b-t+1)}{rt-b-t+1}$$

and $\hat{\sigma}_e^2$ is independent of $\hat{\tau}$. We show the independence and leave the distributional result to the reader:

Let P be the perpendicular projection operator onto $C(X,Z)$ so

$$\hat{\sigma}_e^2 = \frac{Y'(I-P)Y}{rt-b-t+1}.$$

Independence follows from Theorem 1.2.3 upon observing that

$$\begin{aligned}\operatorname{Cov}\big((I-P)Y,(A-A_0)Y\big) &= \sigma^2(I-P)\left[(1-\rho)I+k\rho M\right](A-A_0)\\ &= \sigma^2(1-\rho)(I-P)(A-A_0)+\sigma^2 k\rho(I-P)M(A-A_0)\\ &= 0\,.\end{aligned}$$

The last equality holds because $C(A-A_0) \subset C(X,Z) = C(P)$ so that $(I-P)(A-A_0) = 0$ and $(I-P)M = 0$.

A test of model (6) versus model (1) can be based on

$$\frac{r(1-\rho)+\lambda t\rho}{\hat{\sigma}_e^2\left[(1-\rho)+k\rho\right]}\,\frac{\hat{\tau}'(I-P_t)\hat{\tau}}{t-1} \sim F(t-1, rt-t-b-1). \tag{9}$$

This is true because the lefthand side equals

$$\frac{Y'(A-A_0)'V^{-1}(A-A_0)Y/\sigma^2(t-1)}{\hat{\sigma}_e^2/\sigma^2(1-\rho)},$$

which has the appropriate F distribution under H_0. To see the equality of the two statistics, examine the third to the last equality given earlier in the simplification of $Y'(A-A_0)'V^{-1}(A-A_0)Y$. Similarly, tests and confidence intervals can be based on

$$\frac{\xi'\hat{\tau}-\xi'\tau}{\sqrt{\hat{\sigma}_e^2\left[(1-\rho)+k\rho\right]/\left[r(1-\rho)+\lambda t\rho\right]}} \sim t(rt-t-b+1). \tag{10}$$

This uses the second form for $\mathrm{Var}(\xi'\hat{\tau})$ given earlier. To actually use (9) and (10), we need to estimate $\rho = \sigma_B^2/\left(\sigma_e^2+\sigma_B^2\right)$. If we estimate σ_B^2 and take $\hat{\rho} = \hat{\sigma}_B^2/\left(\hat{\sigma}_e^2+\hat{\sigma}_B^2\right)$, the inferential procedures will be identical to those based on (7) and (8), except that they will be based on the more realistic F and t distributions rather than the χ^2 and normal. Thus we have replaced the traditional analysis, which does not account for the estimation of either of the two unknown parameters σ_e^2 and σ_B^2, with an analysis that does not account for the estimation of only one parameter, ρ.

7.5 Estimation of Variance Components

The traditional analysis of a BIB with recovery of interblock information uses the variance component estimates of Henderson's Method 3, cf. Section 9. The estimate of σ_e^2 is just that described in the previous subsection. To estimate σ_B^2, let P_τ be the perpendicular projection operator onto $C(Z)$ and recall that P is the perpendicular projection operator onto $C(X,Z)$. Using Henderson's Method 3,

$$\hat{\sigma}_B^2 = \frac{\left[Y'(P-P_\tau)Y - \hat{\sigma}_e^2\,\mathrm{tr}(P-P_\tau)\right]}{\mathrm{tr}\left[X'(P-P_\tau)X\right]}.$$

All of these terms are easily computed. $Y'(P-P_\tau)Y = Y'PY - YP_\tau Y$. $Y'PY$ is available from the fixed block effect analysis. In particular,

$$Y'PY = SS(\text{Grand Mean}) + SS(\text{Blocks}) + SS(\text{Treatments After Blocks})$$

and

$$Y'P_\tau Y = SS(\text{Grand Mean}) + SS(\text{Treatments}),$$

where $SS(\text{Treatments})$ is just the sum of squares from a standard one-way ANOVA that ignores the blocks and the covariance structure. The term $\mathrm{tr}(P-P_\tau)$ is simply $b-1$. It is shown later that $\mathrm{tr}[X'(P-P_\tau)X] = t(r-1)$; thus

$$\hat{\sigma}_B^2 = \frac{SS(\text{Blocks after Treatments}) - \hat{\sigma}_e^2\,(b-1)}{t(r-1)}.$$

To see that $\mathrm{tr}\left[X'(P-P_\tau)X\right] = t(r-1)$, note that $\mathrm{tr}\left[X'(P-P_\tau)X\right] = \mathrm{tr}(X'PX) - \mathrm{tr}(X'P_\tau X) = \mathrm{tr}(X'X) - \mathrm{tr}(X'P_\tau X)$. However, $X'X = kI_b$, so $\mathrm{tr}(X'X) = bk = rt$. The trace of $X'P_\tau X$ is more complicated. From the one-way ANOVA, for any vector Y,

$$P_\tau Y = [t_{ij}], \qquad \text{where } t_{ij} = \frac{1}{r}\sum_{i\in A_j} y_{ij}.$$

The matrix $X = [X_1,\ldots,X_b]$ has

$$X_m = [v_{ij}], \qquad \text{where } v_{ij} = \delta_{im},$$

for $m = 1, \ldots, b$; so applying P_τ to X_m gives

$$P_\tau X_m = [t_{ij}], \qquad \text{where } t_{ij} = \frac{1}{r} \sum_{i \in A_j} \delta_{im} = \frac{1}{r} \delta_m(A_j).$$

Recall that A_j is the set of indices for blocks that include treatment j so that $\delta_m(A_j)$ is 1 if block m contains treatment j, and 0 otherwise. This occurs if and only if treatment j is in block m, so $\delta_m(A_j) = \delta_j(D_m)$. Again, D_m is the set of indices for the treatments contained in block m. It follows that

$$\begin{aligned} X_m' P_\tau X_m &= [P_\tau X_m]' [P_\tau X_m] \\ &= \sum_{j=1}^{t} \sum_{i \in A_j} \frac{1}{r^2} \delta_m(A_j) \\ &= \frac{1}{r^2} \sum_{j=1}^{t} \delta_m(A_j) \sum_{i \in A_j} 1 \\ &= \frac{1}{r} \sum_{j=1}^{t} \delta_m(A_j) \\ &= \frac{1}{r} \sum_{j=1}^{t} \delta_j(D_m) \\ &= \frac{k}{r}, \end{aligned}$$

and therefore

$$\operatorname{tr}\left(X' P_\tau X\right) = \sum_{m=1}^{b} X_m' P_\tau X_m = \frac{bk}{r} = \frac{rt}{r} = t.$$

Combining results gives

$$\operatorname{tr}\left[X' (P - P_\tau) X\right] = rt - t = t(r-1).$$

Exercise 7.8 Find the REML estimates of σ_e^2 and σ_B^2.

Exercise 7.9 Do an interblock analysis of the BIB data of PA Example 9.4.1. An experiment was conducted to examine the effects of fertilizers on potato yields. Six treatments (A, B, C, D, E, and F) were used but blocks were chosen that contained only five experimental units. The experiment was performed using a balanced incomplete block design with six blocks. The potato yields (in pounds) along with the mean yield for each block are reported in Table 7.1.

Table 7.1 Potato Yields in Pounds for Six Fertilizer Treatments.

Block	Data					Block Means
1	*E* 583	*B* 512	*F* 661	*A* 399	*C* 525	536.0
2	*B* 439	*C* 460	*D* 424	*E* 497	*F* 592	482.4
3	*A* 334	*E* 466	*C* 492	*B* 431	*D* 355	415.6
4	*F* 570	*D* 433	*E* 514	*C* 448	*A* 344	461.8
5	*D* 402	*A* 417	*B* 420	*F* 626	*E* 615	496.0
6	*C* 450	*F* 490	*A* 268	*D* 375	*B* 347	386.0

The six treatments consist of all of the possible combinations of two factors. One factor was that a nitrogen-based fertilizer was either applied (n_1) or not applied (n_0). The other factor was that a phosphate-based fertilizer was either not applied (p_0), applied in a single dose (p_1), or applied in a double dose (p_2). In terms of the factorial structure, the six treatments are $A = n_0p_0$, $B = n_0p_1$, $C = n_0p_2$, $D = n_1p_0$, $E = n_1p_1$, and $F = n_1p_2$.

7.5.1 Recovery of interblock information

This was in my "notes" for PA-V.

Consider a variance component model with fixed effects τ and random effects γ,

$$Y = Z\tau + X_1\gamma + e. \tag{11}$$

As usual in a variance component model, take $\mathrm{E}(\gamma) = 0$, $\mathrm{E}(e) = 0$, $\mathrm{Cov}(\gamma) = \sigma_B^2 I$, and $\mathrm{Cov}(e) = \sigma_e^2 I_n$. We also assume that X_1 corresponds to indicator variables for balanced groups of size k. Write $\sigma^2 = \sigma_B^2 + \sigma_e^2$ and $\rho = \sigma_B^2/(\sigma_B^2 + \sigma_e^2)$. Thus, with M_1 the ppo onto $C(X_1)$,

$$\mathrm{Cov}(Y) = \sigma^2\{(1-\rho)I + k\rho M_1\}$$

Alternatively, we can write the GLS model

$$Y = Z\tau + \xi, \quad \mathrm{E}(\xi) = 0, \quad \mathrm{Cov}(\xi) = \sigma^2\{(1-\rho)I + k\rho M_1\}.$$

Typically we have to estimate ρ to get empirical GLS estimates. Generalized Split Plot (GSP) models are an exception. GSP models have

$$Z = [X_*, X_2], \quad C(X_*) \subset C(X_1), \quad C(Z) = C[X_*, (I - M_1)X_2].$$

Instead of empirical GLS, might fit the following two linear models. The interblock model (whole plot model)

$$M_1Y = M_1Z\tau + M_1\xi, \quad \mathrm{Cov}(M_1Y) = \sigma^2\{(1-\rho)I + k\rho\}M_1.$$

and the intrablock model (subplot model)

$$(I-M_1)Y = (I-M_1)Z\tau + (I-M_1)\xi, \quad \mathrm{Cov}[(I-M_1)Y] = \sigma^2(1-\rho)(I-M_1).$$

The intrablock model should give the same results as treating γ as fixed in model (1). Both of these models have least squares estimates as BLUEs and I think both allow statistical inferences for normal data. Moreover, $\mathrm{Cov}[M_1Y,(I-M_1)Y] = 0$.

We now combine the information from the two models. Let

$$M_* \equiv M_1Z(Z'M_1Z)^-Z'M_1, \qquad M_2 \equiv (I-M_1)Z[Z'(I-M_1)Z]^-Z'(I-M_1)$$

and

$$Z\tilde{\tau} = M_1Z\hat{\tau}_1 + (I-M_1)Z\hat{\tau}_2 = M_*Y + M_2Y.$$

I think one can do statistical inference on estimable functions of τ using the same approximate methods as for two samples with unequal variance. In a GSP, whole plot and subplot inferences use only one of M_* and M_2.

How does this compare to the empirical GLS estimate? When are they the same? Obviously for GSP models. It would be nice if for some weight between 0 and 1 we had

$$Z\hat{\tau} = \alpha M_1Z\hat{\tau}_1 + (1-\alpha)(I-M_1)Z\hat{\tau}_2$$

but I haven't had any luck with that, even for BIB designs. I guess α would need to be, for estimating $\eta'Z\tau$, something like

$$\frac{\frac{1}{\sigma^2\{(1-\rho)I+k\rho\}\eta'M_*\eta}}{\frac{1}{\sigma^2\{(1-\rho)I+k\rho\}\eta'M_*\eta} + \frac{1}{\sigma^2(1-\rho)I\eta'M_2\eta}}.$$

Good application is a covariate measured at subplot level in a split plot

References

Atkinson, A. C. (1981). Robustness, transformations and two graphical displays for outlying and influential observations in regression. *Biometrika*, **68**, 13–20.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society, A*, **160**, 268–282.

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

Box, G. E. P. and Draper, N. R. (1987). *Empirical Model-Building and Response Surfaces*. John Wiley and Sons, New York.

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

Box, G. E. P. and Meyer, R. D. (1986). An analysis for unreplicated fractional factorials. *Technometrics*, **28**, 11–18.

Byrne, D. M. and Taguchi, S. (1989). The Taguchi approach to parameter design. *Quality Progress*, **December**, 19-26.

Christensen, Ronald (1996). *Analysis of Variance, Design, and Regression: Applied Statistical Methods*. Chapman and Hall, London. http://www.stat.unm.edu/~fletcher/anreg.pdf

Christensen, R. (2011). *Plane Answers to Complex Questions: The Theory of Linear Models*, Fourth Edition. Springer-Verlag, New York.

Christensen, Ronald (2015). *Analysis of Variance, Design, and Regression: Linear Modeling for Unbalanced Data*, Second Edition. Chapman and Hall/CRC Pres, Boca Raton, FL.

Christensen, R., and Huzurbazar, A. V. (1996). A note on augmenting resolution III designs. *The American Statistician*, **50**, 175-177.

Cochran, W. G. (1941). The distribution of the largest of a set of estimated variances as a fraction of their total. *Annals of Eugenics*, **11**, 47–52.

Cochran, W. G. and Cox, G. M. (1957). *Experimental Designs*, Second Edition. John Wiley and Sons, New York.

Cox, D. R. (1958). *Planning of Experiments*. John Wiley and Sons, New York.

Cox, D. R. and Reid, N. (2000). The Theory of the Design of Experiments. Chapman and Hall/CRC, Boca Raton, FL.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press, Princeton.

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

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

Fisher, R. A. (1925). *Statistical Methods for Research Workers*, 14th Edition, 1970. Hafner Press, New York.

Fisher, R. A. (1935). *The Design of Experiments*, Ninth Edition, 1971. Hafner Press, New York.

Fisher, R. A. (1956). *Statistical Methods and Scientific Inference*, Third Edition, 1973. Hafner Press, New York.

Hare, L. B. (1988). In the soup: A case study to identify contributors to filling variability. *Journal of Quality Technology*, **20**, 36–43.

Hartley, H. O. (1938). Studentization and large sample theory. *Journal of the Royal Statistical Society, Supplement*, **5**, 80–88.

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

Hinkelmann, K. and Kempthorne, O. (2008). *Design and Analysis of Experiments: Volume 1, Introduction to Experimental Design*, Second Edition. John Wiley and Sons, Hoboken, NJ.

Holms, A. G. and Berrettoni, J. N. (1969). Chain-pooling ANOVA for two-level factorial replication-free experiments. *Technometrics*, **11**, 725–746.

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

Jones, B. and Nachtsheim, C. J. (2011). A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects. *Journal of Quality Technology*, **43**, 1-14.

Kempthorne, O. (1952). *Design and Analysis of Experiments*. Krieger, Huntington, NY.

Koukouvinos, Christos and Seberry, Jennifer (1997). Weighing matrices and their applications. *JSPI*, **62** 91-101.

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

Lenth, R.V. (2009). Response-surface methods in R, using rsm. *Journal of Statistical Software*, **32**, 1–17.

Lenth, R. V. (2015). The case against normal plots of effects (with discussion). *Journal of Quality Technology*, **47**, 91–97.

Lucas, James M. (1994). How to achieve a robust process using response surface methodology. *Journal of Quality Technology*, **26**, 248-260.

Peirce, Charles S. and Jastrow, Joseph (1885). On Small Differences of Sensation. *Memoirs of the National Academy of Sciences*, **3**, 75–83. This is often referenced as 1984, I suspect because the paper was read Oct. 17, 1984.

Plackett, R. L. and Burman, J. P. (1946). The Design of Optimum Multifactorial Experiments. *Biometrika*, **33**, 305-325

Snedecor, G. W. and Cochran, W. G. (1980). *Statistical Methods*, Seventh Edition. Iowa State University Press, Ames, IA.

Stinson, D. R. (2003). *Combinatorial Designs: Constructions and Analysis*. Springer, New York.

Yates, F. (1935). Complex experiments. *Journal of the Royal Statistical Society, Supplement*, **1**, 181–223.

Index