has expectation

$$E(CC) = \beta_{11} + \beta_{22} + \cdots + \beta_{kk}.$$

Thus if the CC contrast is compared to its standard error, a test for H_0: $\Sigma\beta_{ii} = 0$ versus H_1: $\Sigma\beta_{ii} \neq 0$ may be made.) Suppose you fitted a first-order model to each response. Would the center point average response be counted as one data point or as six data points. (The answer is six; why?) A reviewer of the original article suggested that the responses might have been transformed via

$$\omega_i = \sin^{-1}\left(\frac{y_i}{100}\right)^{1/2}$$

which is a transformation often appropriate when the data are percentages, as here, or proportions (when the division by 100 would not be needed). Is such a transformation useful in the sense that it renders a simpler interpretation of the factors' effects? Transform all the response values and repeat the factorial analysis to find out. (The y_2 data will be used again in Exercise 11.16 and will be combined there with additional data, for a second-order surface fitting exercise.)

4.17. The following estimates of effects resulted from a 2^5 experiment:

$E \leftarrow -224$	$BCD \leftarrow -18$	$BE \leftarrow 29$
$C \leftarrow -153$	$ABDE \leftarrow -14$	$DE \leftarrow 30$
$ABCDE \leftarrow -77$	$D \leftarrow -9$	$ABCE \leftarrow 31$
$ACE \leftarrow -58$	$B \leftarrow -6$	$BCE \leftarrow 39$
$AD \leftarrow -54$	$CD \leftarrow -4$	$ACDE \leftarrow 47$
$BC \leftarrow -53$	$ABC \leftarrow 0$	$AC \leftarrow 53$
$ABD \leftarrow -34$	$AE \leftarrow 2$	$ABCD \leftarrow 58$
$ACD \leftarrow -33$	$BD \leftarrow 7$	$AB \leftarrow 64$
$BDE \leftarrow -28$	$CDE \leftarrow 12$	$CE \leftarrow 83$
$ABE \leftarrow -22$	$BCDE \leftarrow 16$	$A \leftarrow 190$
	$ADE \leftarrow 21$	

(a) Plot these contrasts on normal probability paper.
(b) Which effects do you think are distinguishable from noise?
(c) Replot the remaining contrasts.
(d) Estimate the standard deviation of the contrasts from plot (c). See p. 131.

[Partial answers: (b) E, C, A. (d) Roughly 40.]

CHAPTER 5

Blocking and Fractionating 2^k Factorial Designs

5.1. BLOCKING THE 2^6 DESIGN

In the 2^6 design used above for illustration, the runs were made in a random order; the design was *fully randomized*. However, it might have been much better to run the arrangement in a series of *randomized blocks*. Suppose, for example, we expected trouble because of appreciable inhomogeneity of the raw material. Suppose further that a blender was available, large enough to produce homogeneous blends of raw material sufficient for making eight batches, that is, for conducting 8 of the 64 runs. Then the 64-run experiment might have been arranged using 8 blends (blocks) of homogeneous raw material within each of which 8 runs were made. The difficulty would be that differences between blends could bias the estimates of effects. R. A. Fisher found a way of circumventing this biasing problem by making the (often reasonable) assumption that the effects of blocks would be approximately additive, that is, that the differences between blends would cause the response results simply to be raised or lowered by a fixed (but unknown) amount.

Two Blocks

We introduce the method in a simpler context. Suppose first that we would like to make the 64 runs in two blocks (I and II) each with 32 runs. (Such an arrangement would be appropriate if we had a blender large enough to provide homogeneous raw material for 32 runs.) Consider the partial table of signs, Table 5.1, which shows the 2^6 design laid out in standard order. Suppose now that we allocate runs for which the product column **123456** has a minus sign to block I, and runs for which **123456** has a plus sign to block II. Then, following Fisher, we say that the block effect (denoted by B)

TABLE 5.1. Blocking the 2^6 design into two blocks using B = 123456.

Run number	Variable number 1	2	3	4	5	6	123456	Block
1	−	−	−	−	−	−	+	II
2	+	−	−	−	−	−	−	I
3	−	+	−	−	−	−	−	I
4	+	+	−	−	−	−	+	II
5	−	−	+	−	−	−	−	I
6	+	−	+	−	−	−	+	II
7	−	+	+	−	−	−	+	II
8	+	+	+	−	−	−	−	I
9	−	−	−	+	−	−	−	I
10	+	−	−	+	−	−	+	II
11	−	+	−	+	−	−	+	II
12	+	+	−	+	−	−	−	I
13	−	−	+	+	−	−	+	II
14	+	−	+	+	−	−	−	I
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
63	−	+	+	+	+	+	−	I
64	+	+	+	+	+	+	+	II

is completely *confounded* with the 123456 interaction, that is,

$$\mathbf{B} = \mathbf{123456}. \tag{5.1.1}$$

Thus we lose the ability to estimate (at least with the same accuracy) the 123456 interaction. However, in most practical contexts, this very high-order interaction will probably be negligible in size and of little interest.

Now the 2^k factorial designs have an important orthogonal* property whereby a sequence of signs corresponding to a particular effect is orthogonal to every other such sequence. The important implication of this in the present context is that altering the apparent 123456 effect by superimposing on it the difference between blocks (blends) does not change the estimate of any of the other effects.

* Two columns of ± signs, or two columns of numbers, are said to be *orthogonal* if the sum of the cross-products of corresponding signs, or of corresponding numbers, is zero. The geometrical implication is that the two columns represent two vectors in space which are orthogonal, that is, at right angles, to each other.

Exercise. Add 3 to each of the runs for which elements of the **123456** column are negative and subtract 7 from each of the runs for which they are positive, and confirm by recalculation that the estimates of all effects except for 123456 are unchanged.

Four Blocks; Eight Blocks

Let us now get a little more ambitious and try to arrange the design in four blocks each containing 16 runs. This can be done by confounding two high-order interactions with block contrasts. Suppose, as a first shot, we associate the contrasts 123456 and 23456 with blocks. (This might seem reasonable since we shall wish to confound interactions of the highest possible order with blocks.) Thus the *blocking generators* for our design will be

$$\mathbf{B}_1 = \mathbf{123456}, \qquad \mathbf{B}_2 = \mathbf{23456}, \tag{5.1.2}$$

say. The runs would then be allocated to the four blocks according to the following scheme:

$$\begin{array}{cc|cc} & & \multicolumn{2}{c}{\mathbf{B}_2 = \mathbf{23456}} \\ & & - & + \\ \hline \mathbf{B}_1 = \mathbf{123456} & - & I & II \\ & + & III & IV \end{array} \tag{5.1.3}$$

That is, runs would be allocated to the four blocks I, II, III, and IV as the signs associated with the columns **123456** and **23456** took the values (− −) (− +) (+ −) and (+ +). Unfortunately, a serious difficulty occurs with this arrangement. There are, of course, 3 df (independent contrasts) among the four blocks of 16 runs each. If we associate $\mathbf{B}_1$ and $\mathbf{B}_2$ with two of these contrasts, the third must be the interaction $\mathbf{B}_1 \times \mathbf{B}_2 = \mathbf{B}_1\mathbf{B}_2$. However, then the interaction $\mathbf{B}_1\mathbf{B}_2$ will be confounded with the "interaction" between **123456** and **23456**. To see what this interaction is, consider the following table of signs in standard order. (We write them as rows to conserve space.)

123456	+ − − +	− + + −	− + + −	···	− +
23456	− − + +	+ + − −	+ + − −	···	+ +
product of above	− + − +	− + − +	− + − +	···	− +

We see that the interaction between **123456** and **23456** is the main effect **1**. This arrangement would thus be a very poor one, because the main effect of

the first variable would be confounded with block differences. To make a better choice, we need to understand how to calculate interactions between complex effects. To see how to do this, we note that if we take any set of signs for any effect (say 1) and multiply those signs by the signs of the *same* effect, we obtain a row of +'s which we denote by the *identity* **I**.

$$\begin{array}{rcccccc} \mathbf{1} & -+-+ & -+-+ & -+-+ & \cdots & -+ \\ \mathbf{1} & -+-+ & -+-+ & -+-+ & \cdots & -+ \\ \mathbf{1}\times\mathbf{1}=\mathbf{1}^2=\mathbf{I} & ++++ & ++++ & ++++ & \cdots & ++ \end{array}$$

Thus, using the multiplication sign to imply the multiplication of the signs in corresponding positions in two rows (or two columns if the design is arranged vertically, as would be usual), we can write

$$\mathbf{1}\times\mathbf{1}=\mathbf{1}^2=\mathbf{I}, \qquad \mathbf{2}\times\mathbf{2}=\mathbf{2}^2=\mathbf{I}, \qquad \mathbf{12}\times\mathbf{12}=\mathbf{1}^2\mathbf{2}^2=\mathbf{I}^2=\mathbf{I}, \tag{5.1.4}$$

and so on. Also, multiplication of any contrast by the identity **I** leaves the contrast unchanged. Thus

$$\mathbf{1}\times\mathbf{I}=\mathbf{1}, \qquad \mathbf{23}\times\mathbf{I}=\mathbf{23}, \qquad \mathbf{12345}\times\mathbf{I}=\mathbf{12345}, \tag{5.1.5}$$

and so on. Applying this rule in the case of the design above, we have that

$$\mathbf{B}_1\mathbf{B}_2=\mathbf{123456}\times\mathbf{23456}=\mathbf{1}^2\mathbf{3}^2\mathbf{4}^2\mathbf{5}^2\mathbf{6}^2=\mathbf{1}\times\mathbf{I}^5=\mathbf{1}, \tag{5.1.6}$$

indicating that B_1B_2 and 1 are confounded.

An arrangement in four blocks of 16 runs so that all the interactions confounded with blocks are of the highest possible order may be obtained by using as generators two four-factor interactions in which only two symbols overlap. For example, if we choose

$$\mathbf{B}_1=\mathbf{1234}$$

$$\mathbf{B}_2=\quad\mathbf{3456},$$

then

$$\mathbf{B}_1\mathbf{B}_2=\mathbf{12\ \ 56}. \tag{5.1.7}$$

We can now write out, with the design arranged in standard order, as usual, rows of signs for $\mathbf{B}_1$ and $\mathbf{B}_2$ and allocate the runs to the four blocks corresponding to the sign combinations $(\mathbf{B}_1, \mathbf{B}_2) = (-,-), (-,+), (+,-),$

$(+,+)$, as follows:

$$\begin{array}{lcccccccccccccccc} \mathbf{B}_1=\mathbf{1234} & + & - & - & + & - & + & + & - & - & + & + & - & \cdots & - & + \\ \mathbf{B}_2=\mathbf{3456} & + & + & + & + & - & - & - & - & - & - & - & - & \cdots & + & + \\ \text{block number} & \text{IV} & \text{II} & \text{II} & \text{IV} & \text{I} & \text{III} & \text{III} & \text{I} & \text{I} & \text{III} & \text{III} & \text{I} & \cdots & \text{II} & \text{IV} \end{array}$$

On the assumption that the blocks contribute only additive effects, the main effects, two-factor interactions, and three-factor interactions will all remain unconfounded with any block effect.

If, as would often be acceptable, we are prepared to confound some three-factor interactions, we can split the 64 runs into eight blocks each containing eight runs. For example, choosing the generators

$$\mathbf{B}_1=\mathbf{1234}, \qquad \mathbf{B}_2=\mathbf{3456}, \qquad \mathbf{B}_3=\mathbf{136}, \tag{5.1.8}$$

we obtain the following confounding pattern for the seven contrasts among the eight blocks:

$$\begin{aligned} \mathbf{B}_1 &= \mathbf{1234} \\ \mathbf{B}_2 &= \quad\mathbf{3456} \\ \mathbf{B}_1\mathbf{B}_2 &= \mathbf{12\ \ 56} \\ \mathbf{B}_3 &= \mathbf{1\ 3\ \ 6} \\ \mathbf{B}_1\mathbf{B}_3 &= \mathbf{2\ \ 4\ 6} \\ \mathbf{B}_2\mathbf{B}_3 &= \mathbf{1\ \ 4\ 5} \\ \mathbf{B}_1\mathbf{B}_2\mathbf{B}_3 &= \mathbf{2\,3\ 5}. \end{aligned} \tag{5.1.9}$$

On the assumption that blocks have only additive effects, no main effect or two-factor interaction is confounded with any block effect.

Exercise. Allocate the 64 runs of a 2^6 design to eight blocks using the generators $\mathbf{B}_1 = \mathbf{1346}$, $\mathbf{B}_2 = \mathbf{1235}$, and $\mathbf{B}_3 = \mathbf{156}$ and write down the confounding pattern for the seven contrasts among the eight blocks. Is any main effect or two-factor interaction confounded with blocks? (No.)

A short table of useful confounding arrangements adapted from Box, Hunter, and Hunter (1978) is given in Table 5.2. Other tables will be found in Davies (1954), the National Bureau of Standards Applied Mathematics Series, No. 48 (1957), and McLean and Anderson (1984).

TABLE 5.2. Useful blocking arrangements for 2^k factorial designs

k = number of variables	Block size	Block generator	Interactions confounded with blocks
3	4	$B_1 = 123$	123
	2	$B_1 = 12, B_2 = 13$	12, 13, 23
4	8	$B_1 = 1234$	1234
	4	$B_1 = 124, B_2 = 134$	124, 134, 23
	2	$B_1 = 12, B_2 = 23, B_3 = 34$	12, 23, 34, 13, 1234, 24, 14
5	16	$B_1 = 12345$	12345
	8	$B_1 = 123, B_2 = 345$	123, 345, 1245
	4	$B_1 = 125, B_2 = 235, B_3 = 345$	125, 235, 345, 13, 1234, 24, 145
	2	$B_1 = 12, B_2 = 13, B_3 = 34, B_4 = 45$	12, 13, 34, 45, 23, 1234, 1245, 14, 1345 35, 24, 2345, 1235, 15, 25 that is, all 2fi and 4fi[a]
6	32	$B_1 = 123456$	123456
	16	$B_1 = 1236, B_2 = 3456$	1236, 3456, 1245
	8	$B_1 = 135, B_2 = 1256, B_3 = 1234$	135, 1256, 1234, 236, 245, 3456, 146
	4	$B_1 = 126, B_2 = 136, B_3 = 346, B_4 = 456$	126, 136, 346, 456, 23, 1234, 1245 14, 1345, 35, 246, 23456, 12356, 156, 25
	2	$B_1 = 12, B_2 = 23, B_3 = 34, B_4 = 45, B_5 = 56$	All 2fi, 4fi, and 6fi
7	64	$B_1 = 1234567$	1234567
	32	$B_1 = 12367, B_2 = 34567$	12367, 34567, 1245
	16	$B_1 = 123, B_2 = 456, B_3 = 167$	123, 456, 167, 123456, 2367, 1457, 23457
	8	$B_1 = 1234, B_2 = 567, B_3 = 345, B_4 = 147$	1234, 567, 345, 147, 1234567, 125, 237, 3467, 1456, 1357, 1267, 2356, 2457, 136, 246
	4	$B_1 = 127, B_2 = 237, B_3 = 347, B_4 = 457, B_5 = 567$	127, 237, 347, 457, 567, 13, 1234, 1245, 1256, 24, 2345, 2356, 3456, 35, 1234567, 46, 147, 13457, 13567, 12467, 12357, 257, 24567, 23467, 367, 15, 1456, 1346, 1236, 26, 167
	2	$B_1 = 12, B_2 = 23, B_3 = 34, B_4 = 45, B_5 = 56, B_6 = 67$	All 2fi, 4fi, and 6fi

[a] "fi" is an abbreviation for "factor interaction"; thus, for example, 2fi means two-factor interaction.

Source: Table 5.2 is reproduced and adapted with permission from the table on pp. 346–347 of *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*, by G. E. P. Box, W. G. Hunter, and J. S. Hunter, published by John Wiley & Sons, Inc., New York, 1978.

5.2. FRACTIONATING THE 2^6 DESIGN

To study the dyestuffs problem in Chapter 4, we used a 2^6 factorial design that required 64 runs. As it turned out, only main effects of three of the variables could be distinguished from the noise. Such a discovery naturally leads to the speculation as to whether so many runs were really needed. In this particular example, the standard deviation σ of the experimental error was rather large compared with the size of the effects of interest, so the full 64 runs *were* probably needed. However, if σ had been smaller, a design containing a smaller number of runs than the full $2^6 = 64$ might have proved adequate. A class of two-level designs requiring fewer runs than the full 2^k factorials is that of the 2^{k-p} *fractional factorials*.

Half-Fractions of the 2^6 Factorial

To illustrate the nature and the utility of such designs, consider again the problem of blocking the 2^6 factorial into two blocks, each of 32 runs, with the six-factor interaction **123456** (= **B**) used as the defining (or blocking) contrast. (See Section 5.1.) Let us temporarily augment our previous notation a little by using the symbol $\mathbf{I}_{32}$ to refer to a sequence of 32 plus signs. For each of the 32 runs in the first block, the signs of the six factors will multiply to give -1 and for each of the 32 runs in the second block, the signs will multiply to give $+1$. Thus,

$$\mathbf{123456} = -\mathbf{I}_{32}, \quad \text{for block I,}$$

$$\mathbf{123456} = \mathbf{I}_{32} \quad \text{for block II.} \tag{5.2.1}$$

Suppose now that, instead of performing all the 64 runs, we carried out only the 32 runs in, say, block II. These are the runs for which elements of **123456** are + in Table 5.1. Let us set out a table of signs for the calculation of main effects, two-factor interactions, three-, four-, five-, and six-factor interactions, just as for a full factorial design but using only block II runs, as shown in Table 5.3.

We could certainly use the table of signs to compute all the 63 main effects and interactions. Nevertheless, something would seem to be amiss because we can scarcely expect to compute 63 independent effects from only 32 observations. The explanation lies in the duplication that occurs in the table. We see, as expected, that the signs in the **123456** column are all identical to those of $\mathbf{I}_{32}$ (because we have deliberately arranged this). Notice also, however, that the signs for the interaction column **23456** are identical to those for the column **1**, or symbolically **1** = **23456**. Additionally **2** = **13456**, and so on. Let us denote by l_1, l_2, and so on, the various linear contrasts obtained by adding together all the response values with plus signs in the **1**, **2**, and so on, column and subtracting all those with minus signs, and dividing by the appropriate column divisor. Also* let us use an arrow pointing to the right ($\rightarrow$) to mean "is an estimate of." Then, for example, l_1 provides an estimate not just of the main effect of 1, nor of the interaction 23456, but of the sum of these, that is,

$$l_1 \rightarrow 1 + 23456. \tag{5.2.2}$$

*As in Chapter 4, the arrow pointing to the left ($\leftarrow$) will mean "is estimated by."

TABLE 5.3. Columns of signs for calculation of effects in a half-fraction of a 2^6 design

Run number from Table 5.1	I_{32}	Main effects						Two-factor interactions					Five-factor interactions				Six-factor interaction
		1	2	3	4	5	6	12	13	14	...	...	12456	13456	23456		123456
1	+	−	−	−	−	−	−	+	+	+	...	...	−	−	−		+
4	+	+	+	−	−	−	−	+	−	−	...		−	+	+		+
6	+	+	−	+	−	−	−	−	+	−	...		+	−	+		+
7	+	−	+	+	−	−	−	−	−	+	...		+	+	−		+
10	+	+	−	−	+	−	−	−	−	+	...		−	−	+		+
11	+	−	+	−	+	−	−	−	+	−	...		−	+	−		+
13	+	−	−	+	+	−	−	+	−	−	...		+	−	−		+
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...		...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...		...
64	+	+	+	+	+	+	+	+	+	+	...	...	+	+	+		+
Divisor	32	16	16	16	16	16	16	16	16	16	...	...	16	16	16		16

To establish this, we need only write out the formulas for the estimates of 1 and of 23456 from the full 2^6 and add the two together. It will be found that addition and cancellation leaves only the observations in block II with the signs of column **1** in Table 5.3 divided by the divisor 16. In similar fashion,

$$l_2 \to 2 + 13456 \tag{5.2.3}$$

and so on. The device that enables us quickly to identify all these *aliases* is supplied by the generator for this half-fraction of the 2^6 design which we can write as

$$\mathbf{I} = \mathbf{123456}, \tag{5.2.4}$$

(where now, for simplicity, we write **I** for $\mathbf{I}_{32}$). Equation (5.2.4) simply tells us to pick out runs from the full 64 whose product of signs **123456** is plus, same as the **I** column. Multiplying both sides of Eq. (5.2.4) by **1**, we have **1** = **23456** which implies that $l_1 \to 1 + 23456$.

All other existing alias relationships are obtained similarly; for example, on multiplying Eq. (5.2.4) by **123** we obtain

$$\mathbf{123} = \mathbf{456}, \qquad \text{implying that } l_{123} \to 123 + 456. \tag{5.2.5}$$

(Note that each l can be denoted by two equivalent names; for example, $l_{123} = l_{456}$.)

For quantitative variables, the various interactions 12, 345, ..., reflect the effects of mixed derivatives of the response function η, such as

$$\frac{\partial^2 \eta}{\partial x_1 \partial x_2}, \frac{\partial^3 \eta}{\partial x_3 \partial x_4 \partial x_5}, \ldots .$$

It would frequently be reasonable to assume that derivatives higher than the second could be ignored, that is, in a Taylor's series expansion of the response function, terms of third and of higher orders could be assumed to be negligible within the local region of experimentation. In these circumstances, we could, correspondingly, ignore interactions of three or more factors in the factorial analysis of the data. *Under such an assumption*, no confounding would in fact occur in the half-fraction **I** = **123456** of the 2^6 design above. When σ is not too large, so that fewer runs than a full factorial will provide adequate accuracy, fractional factorials are frequently very useful, particularly in experimental situations with more than three or four variables.

For our illustration above, we chose the half-fraction defined by $\mathbf{I} = \mathbf{123456}$, but we could equally well have used the other half-fraction defined by $-\mathbf{I} = \mathbf{123456}$ or, equivalently, $\mathbf{I} = -\mathbf{123456}$. In this case, the columns $\mathbf{1}$ and $-\mathbf{23456}$ will be identical and, denoting the linear contrasts defined by columns $\mathbf{1}, \mathbf{2}, \ldots,$ and so on by $l_1', l_2', \ldots,$ and so on, we shall find that

$$l_1' \to 1 - 23456, \qquad l_2' \to 2 - 13456, \ldots, \tag{5.2.6}$$

and so on. Notice that, if we perform *both* halves of the 2^6 design, we can recover all the individual main effects and interactions by averaging or halving and differencing. For example,

$$\tfrac{1}{2}(l_1 + l_1') \to \tfrac{1}{2}(1 + 23456 + 1 - 23456) = 1 \tag{5.2.7}$$

and this would be identical to the estimate supplied by the full factorial. Also, similarly,

$$\tfrac{1}{2}(l_1 - l_1') \to 23456, \tag{5.2.8}$$

again identical to that supplied by the full factorial.

A one-half replicate of a full factorial in k factors, that is, $\frac{1}{2}$ of a 2^k, or a 2^{-1} fraction of a 2^k, is often referred to as a 2^{k-1} fractional factorial. Thus the fractions we have described above are both 2^{6-1} fractional factorial designs.

Quarter Fractions of the 2^6 Factorial

Suppose now we ran only 16 of the 64 runs of the full 2^6 design corresponding to one block of the design when it is divided into four blocks (as in Section 5.1). This *quarter replicate* would be referred to as a 2^{6-2} fractional factorial design. There are four blocks from which to choose. Suppose we use the one for which $(\mathbf{B}_1, \mathbf{B}_2) = (+, +)$. Then for all the runs in this block,

$$\mathbf{B}_1 = \mathbf{I} = \mathbf{1234}, \qquad \mathbf{B}_2 = \mathbf{I} = \mathbf{3456}, \qquad \mathbf{B}_1\mathbf{B}_2 = \mathbf{I} = \mathbf{1256}. \tag{5.2.9}$$

We can multiply through this defining relation by $\mathbf{1}, \mathbf{2}, \ldots, \mathbf{12}, \mathbf{13}, \ldots$ to give the alias pattern. For example,

$$\begin{aligned} \mathbf{1} &= \mathbf{234} = \mathbf{13456} = \mathbf{256}, \\ \mathbf{2} &= \mathbf{134} = \mathbf{23456} = \mathbf{156}, \\ &\ldots \\ \mathbf{12} &= \mathbf{34} = \mathbf{123456} = \mathbf{56}, \\ &\ldots \end{aligned} \tag{5.2.10}$$

and so on. Ignoring interactions involving four or more factors we see that

the design allows estimates of 16 combinations of main effects and interactions such as, for example,

$$\begin{aligned} l_1 &\to 1 + (234 + 256), \\ l_2 &\to 2 + (134 + 156), \\ l_3 &\to 3 + (124 + 456), \\ l_4 &\to 4 + (123 + 356), \\ l_5 &\to 5 + (346 + 126), \\ l_6 &\to 6 + (345 + 125), \\ l_{12} &\to 12 + 34 + 56, \\ l_{13} &\to 13 + 24, \\ l_{14} &\to 14 + 23, \\ &\ldots \end{aligned} \tag{5.2.11}$$

(Note that the l subscripts can be those of any of the aliased effects. For example, $l_{12} = l_{34} = l_{56}$, and so on. Typically, we choose subscripts which occur earlier in the ordering $1, 2, \ldots, 12, 13, \ldots, 123, \ldots,$ eliminating aliased duplicates, but this is not essential.)

The estimation equations show that all main effects are confounded with third-order (or ignored higher-order) interactions, not with two-factor interactions or other main effects. Two-factor interactions are confounded with other two-factor interactions. The defining relation must indicate all the independent generators and their products in twos, threes, ..., all of which produce the identity $\mathbf{I}$. For example, suppose we had used the 2^{6-2} design defined by $(\mathbf{B}_1, \mathbf{B}_2) = (-, +)$, we would have had

$$\mathbf{B}_1 = -\mathbf{I} = \mathbf{1234}, \qquad \mathbf{B}_2 = \mathbf{I} = \mathbf{3456}, \qquad \mathbf{B}_1\mathbf{B}_2 = -\mathbf{I} = \mathbf{1256}, \tag{5.2.12}$$

which leads to the defining relation

$$\mathbf{I} = -\mathbf{1234} = \mathbf{3456} = -\mathbf{1256}, \tag{5.2.13}$$

and means that the estimates obtained would be as follows:

$$\begin{aligned} l_1' &\to 1 - 234 - 256, \\ l_2' &\to 2 - 134 - 156, \\ &\ldots \end{aligned} \tag{5.2.14}$$

and so on, ignoring interactions of four or more factors.

The variance of any of the estimates in (5.2.11) or (5.2.14) is

$$V(l) = 4\sigma^2/n \tag{5.2.15}$$

where n is the total number of individual observations (with variance σ^2) contributing to l. This formula applies to estimates throughout this chapter.

5.3. RESOLUTION OF A 2^{k-p} FACTORIAL DESIGN

The resolution R of a 2^{k-p} fractional factorial design is the length of the shortest word in the defining relation. For example, the 2^{6-1} design with defining relation **I = 123456** is of resolution VI. (Note that the resolution is expressed in Roman numerals.) The quarter fraction 2^{6-2} design with defining relation **I = 1234 = 3456 = 1256** is of resolution IV, and so on.

In this book we shall frequently use polynomial approximations of first or second degree. First-order designs to obtain data to estimate the coefficients in a first degree equation should therefore be of resolution at least III. This will ensure that no main effect is aliased with any other main effect, although main effects may (if R = III) be aliased with two-factor interactions. On the strict assumption that the response can be represented by a first degree equation, these two-factor interactions will all be zero. A resolution IV design provides added security because then no main effect would be aliased with either main effects or two-factor interactions.

We shall see later that second-order designs used to fit second degree polynomials can be formed by a suitable augmentation of two-level factorial or fractional factorial designs.

5.4. CONSTRUCTION OF 2^{k-p} DESIGNS OF RESOLUTION III AND IV

A design of resolution III can be constructed by *saturating* all the interactions of a full factorial design with additional factors. We illustrate this with an initial 2^3 design which can be saturated to produce a one-sixteenth replicate of a 2^7 design, namely a 2^{7-4} design. Consider again the columns of signs of Table 4.2 for the 2^3 factorial, now reproduced as Table 5.4.

If we use the **12** column to accommodate factor **4**, the **13** column similarly for factor **5**, the **23** column for **6**, and the **123** column for **7**, we have made the equivalances

$$\mathbf{4 = 12}, \quad \mathbf{5 = 13}, \quad \mathbf{6 = 23}, \quad \text{and} \quad \mathbf{7 = 123}. \tag{5.4.1}$$

TABLE 5.4. A saturated, resolution III, 2^{7-4} design.

I	1	2	3	4 = 12	5 = 13	6 = 23	7 = 123
+	−	−	−	+	+	+	−
+	+	−	−	−	−	+	+
+	−	+	−	−	+	−	+
+	+	+	−	+	−	−	−
+	−	−	+	+	−	−	+
+	+	−	+	−	+	−	−
+	−	+	+	−	−	+	−
+	+	+	+	+	+	+	+

Thus the generators of the design are

$$\mathbf{I = 124 = 135 = 236 = 1237}. \tag{5.4.2}$$

By multiplying these generators together in all possible ways, we obtain the defining relation for this 2^{7-4} design, and this leads to the alias relationships shown below. In this tabulation, we have ignored all interactions involving more than two factors for the sake of simplicity.

$$\begin{aligned}
l_1 &\to 1 + 24 + 35 + 67,\\
l_2 &\to 2 + 14 + 36 + 57,\\
l_3 &\to 3 + 15 + 26 + 47,\\
l_4 &\to 4 + 12 + 56 + 37,\\
l_5 &\to 5 + 13 + 46 + 27,\\
l_6 &\to 6 + 23 + 45 + 17,\\
l_7 &\to 7 + 34 + 25 + 16.
\end{aligned} \tag{5.4.3}$$

Eight-Run Designs for Fewer Than Seven Variables

When the number of variables examined is smaller than seven, useful designs can also be obtained by omitting one or more of the columns of the 2^{7-4} design or, more accurately, not using those columns for a variable. (The column is still used for estimating combinations of interactions but no longer specifies the levels of a variable.) The choice of which columns to

drop can make use of the judgment of the investigator that only certain specific interactions are likely to occur. For example, suppose we have five factors and a particular one of these, which we designate as 1, is thought likely to interact with two others, 2 and 3, say; apart from this no other interactions are likely. Thus we would seek a design which can estimate the main effects of five factors, including 1, 2, and 3, and also the interactions 12 and 13, assuming all other interactions are zero. Looking at Table 5.4, we see that the required design is obtained by associating our five factors with the columns **1**, **2**, **3**, **6**, and **7**. If our assumptions are correct, then the linear combinations will provide the following estimates:

$$\begin{aligned} &l_1 \to 1, \quad l_4 \to 12, \quad l_6 \to 6, \\ &l_2 \to 2, \quad l_5 \to 13, \quad l_7 \to 7. \\ &l_3 \to 3, \end{aligned} \tag{5.4.4}$$

For four variables, we can use variables 1, 2, 3, and 7 to obtain a design of resolution IV.

Adding Further Fractions

Sometimes, it will happen that the results from a highly fractionated design are ambiguous. In such a case we may wish to run a further fraction or fractions which will resolve the uncertainties created by the first. There are many possibilities; the following illustrations will indicate the general idea. We assume that the first fraction is the 2^{7-4} given by $\mathbf{I} = \mathbf{124} = \mathbf{135} = \mathbf{236} = \mathbf{1237}$ with estimates as in Eq. (5.4.3).

Possibility 1

Suppose that, from a view of the results of the first fraction, it seems likely that a particular factor is dominating, and we decide we would like to add a further fraction and estimate the main effect of the dominating factor and all its interactions with other factors, clear of two-factor aliases. We can achieve this by choosing the second fraction whose generators are like those of the first fraction but with the sign of the variable of interest reversed everywhere. For example, if variable **1** is the designated variable we use

$$\mathbf{I} = -\mathbf{124} = -\mathbf{135} = \mathbf{236} = -\mathbf{1237} \tag{5.4.5}$$

and obtain the individual runs of the design by writing an initial 2^3 factorial

in variables **1**, **2**, and **3** and then generating columns $\mathbf{4} = -\mathbf{12}$, $\mathbf{5} = -\mathbf{13}$, $\mathbf{6} = \mathbf{23}$, and $\mathbf{7} = -\mathbf{123}$. From the second fraction alone we can then estimate

$$\begin{aligned} l_1' &\to 1 - 24 - 35 - 67, \\ l_2' &\to 2 - 14 + 36 + 57, \\ l_3' &\to 3 - 15 + 26 + 47, \\ l_4' &\to 4 - 12 + 56 + 37, \\ l_5' &\to 5 - 13 + 46 + 27, \\ l_6' &\to 6 + 23 + 45 - 17, \\ l_7' &\to 7 + 34 + 25 - 16, \end{aligned} \tag{5.4.6}$$

which has the same basic alias structure as Eq. (5.4.3) but with certain signs reversed, as shown. It will now be seen that by combining the results from both fractions we can compute the estimates:

$$\begin{aligned} &\tfrac{1}{2}(l_1 - l_1') \to 24 + 35 + 67, && \tfrac{1}{2}(l_1 + l_1') \to 1, \\ &\tfrac{1}{2}(l_2 - l_2') \to 14, && \tfrac{1}{2}(l_2 + l_2') \to 2 + 36 + 57, \\ &\tfrac{1}{2}(l_3 - l_3') \to 15, && \tfrac{1}{2}(l_3 + l_3') \to 3 + 26 + 47, \\ &\tfrac{1}{2}(l_4 - l_4') \to 12, && \tfrac{1}{2}(l_4 + l_4') \to 4 + 56 + 37, \\ &\tfrac{1}{2}(l_5 - l_5') \to 13, && \tfrac{1}{2}(l_5 + l_5') \to 5 + 46 + 27, \\ &\tfrac{1}{2}(l_6 - l_6') \to 17, && \tfrac{1}{2}(l_6 + l_6') \to 6 + 23 + 45, \\ &\tfrac{1}{2}(l_7 - l_7') \to 16, && \tfrac{1}{2}(l_7 + l_7') \to 7 + 34 + 25. \end{aligned} \tag{5.4.7}$$

We thus see that, provided our assumption that interactions between more than two factors are negligible is true, we can isolate estimates of the main effect of factor 1 and all its interactions with other factors free from aliases.

Possibility 2

Suppose instead that we wished to run a second fraction that would provide all main effects clear of *all* two-factor interactions. This can be done by choosing a second fraction whose generators have the signs of all the variables reversed compared with the first one. For example, from Eq. (5.4.2) this procedure gives the fraction

$$\mathbf{I} = -\mathbf{124} = -\mathbf{135} = -\mathbf{236} = \mathbf{1237} \tag{5.4.8}$$

whose columns are obtained by writing down a 2^3 design in the initial columns **1**, **2**, and **3** and then setting $\mathbf{4} = -\mathbf{12}$, $\mathbf{5} = -\mathbf{13}$, $\mathbf{6} = -\mathbf{23}$, and $\mathbf{7} = \mathbf{123}$. Such a design produces the estimates

$$\begin{aligned}
l_1' &\to 1 - 24 - 35 - 67,\\
l_2' &\to 2 - 14 - 36 - 57,\\
l_3' &\to 3 - 15 - 26 - 47,\\
l_4' &\to 4 - 12 - 56 - 37,\\
l_5' &\to 5 - 13 - 46 - 27,\\
l_6' &\to 6 - 23 - 45 - 17,\\
l_7' &\to 7 - 34 - 25 - 16.
\end{aligned} \tag{5.4.9}$$

Combining these estimates with those of Eq. (5.4.3) allows us to separate the main effects from the groups of two-factor interactions as follows:

$$\begin{aligned}
&\tfrac{1}{2}(l_1 + l_1') \to 1, && \tfrac{1}{2}(l_1 - l_1') \to 24 + 35 + 67\\
&\tfrac{1}{2}(l_2 + l_2') \to 2, && \tfrac{1}{2}(l_2 - l_2') \to 14 + 36 + 57,\\
&\cdots && \cdots
\end{aligned} \tag{5.4.10}$$

and so on.

Designs of Resolution IV via "Foldover"

The foregoing combination design is of resolution IV and so the method of switching all signs provides a method of building up a resolution IV design from one of resolution III. The switching can be carried out directly on the sign pattern as follows. Suppose **A** represents the block of signs shown in Table 5.4 (ignoring the **I** column for the moment) for the 2^{7-4} design $\mathbf{I} = \mathbf{124} = \mathbf{135} = \mathbf{236} = \mathbf{1237}$. Then the signs

$$\begin{matrix} \mathbf{A} \\ -\mathbf{A} \end{matrix} \tag{5.4.11}$$

provide 16 lines of signs which give a 2^{7-3} design of resolution IV. The $-\mathbf{A}$ block will be exactly the second set of runs as in "Possibility 2" above but will occur in a different order. Such a design is usually called a *foldover* design because we have "folded over" the signs, that is, switched them all. We can actually do better than this, and obtain a 2^{8-4} design of resolution IV by associating an eighth factor with the initial $\mathbf{I}_8$ column of ones. The folded design can then be written

$$\begin{matrix} \text{runs 1–8} \\ \text{runs 9–16} \end{matrix} \begin{bmatrix} \mathbf{A} & \mathbf{I}_8 \\ -\mathbf{A} & -\mathbf{I}_8 \end{bmatrix}. \tag{5.4.12}$$

The design matrix of the 2^{8-4} resolution IV design obtained in this manner is shown in Table 5.5. Equation (5.4.13) shows the estimates that can be obtained from the columns whose subscripts are shown in the l's.

$$\begin{aligned}
l_1 &\to 1.\\
l_2 &\to 2,\\
l_3 &\to 3,\\
l_4 &\to 4,\\
l_5 &\to 5,\\
l_6 &\to 6,\\
l_7 &\to 7,\\
l_8 &\to 8,\\
l_{12} &\to 12 + 37 + 48 + 56 && (B_1),\\
l_{13} &\to 13 + 27 + 46 + 58 && (B_2),\\
l_{14} &\to 14 + 28 + 36 + 57 && (B_3),\\
l_{15} &\to 15 + 26 + 38 + 47 && (B_1B_2B_3),\\
l_{16} &\to 16 + 25 + 34 + 78 && (B_2B_3),\\
l_{17} &\to 17 + 23 + 68 + 45 && (B_1B_2),\\
l_{18} &\to 18 + 24 + 35 + 67 && (B_1B_3).
\end{aligned} \tag{5.4.13}$$

Exercise. Use the data shown in Table 5.5 and obtain the numerical estimates for the combinations of effects in Eq. (5.4.13). They are, in order, −0.7, −0.1, 5.5, −0.3, −3.8, −0.1, 0.6, 1.2, −0.6, 0.9, −0.4, 4.6, −0.3, −0.2, and −0.6. What do you conclude about the factors as a result of examining these figures? For additional details see Box, Hunter, and Hunter (1978), p. 399.

TABLE 5.5. A 2^{8-4} design of resolution IV, obtained here by the foldover technique

	Mold temperature	Moisture content	Holding pressure	Cavity thickness	Booster pressure	Cycle time	Gate size	Screw speed	Shrinkage
Run	1	2	3	4	5	6	7	8	y
1	−	−	−	+	+	+	−	+	14.0
2	+	−	−	−	−	+	+	+	16.8
3	−	+	−	−	+	−	+	+	15.0
4	+	+	−	+	−	−	−	+	15.4
5	−	−	+	+	−	−	+	+	27.6
6	+	−	+	−	+	−	−	+	24.0
7	−	+	+	−	−	+	−	+	27.4
8	+	+	+	+	+	+	+	+	22.6
9	+	+	+	−	−	−	+	−	22.3
10	−	+	+	+	+	−	−	−	17.1
11	+	−	+	+	−	+	−	−	21.5
12	−	−	+	−	+	+	+	−	17.5
13	+	+	−	−	+	+	−	−	15.9
14	−	+	−	+	−	+	+	−	21.9
15	+	−	−	+	+	−	+	−	16.7
16	−	−	−	−	−	−	−	−	20.3

Source: Table 5.5 is reproduced and adapted with permission from the table on p. 402 of *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*, by G. E. P. Box, W. G. Hunter, and J. S. Hunter, published by John Wiley & Sons, Inc., New York, 1978.

The foldover method of Eq. (5.4.12) is completely general and will provide a resolution IV design whenever **A** represents a block of signs designating a resolution III design.

Exercise. Write down the 2^{3-1} resolution III design given by **I = 123**. Associate a fourth variable with $\mathbf{I}_4$. Fold over the entire design. Confirm that the combination design is the resolution IV design **I = 1234**.

In the parentheses in Eq. (5.4.13) we show how to generate an interesting blocking arrangement which allows the 2^{8-4} design to be run in eight blocks of two runs each, without confounding main effects with blocks. The way in which this is done will be clear from Table 5.6 which shows three blocking variables associated with selected interactions. The eight possible sign combinations of the blocking variables provide the eight blocks. It is of interest to note that the two runs within any block are mirror images of each other. All the signs of any one run are reversed in its companion run.

TABLE 5.6. A 2_{IV}^{8-4} design in eight blocks of size two

	2_{IV}^{8-4} design								Block variable					Design rearranged in eight blocks										
Run	1	2	3	4	5	6	7	8	B_1 12	B_2 13	B_3 14	Block	1	2	3	4	5	6	7	8	B_1	B_2	B_3	Run
1	−	−	−	+	+	+	−	+	+	+	−	1	+	−	−	−	−	+	+	+	−	−	−	2
2	+	−	−	−	−	+	+	+	−	−	−		−	+	+	+	+	−	−	−				10
3	−	+	−	−	+	−	+	+	−	+	+	2	−	−	+	+	−	−	+	+	+	−	−	5
4	+	+	−	+	−	−	−	+	+	−	+		+	+	−	−	+	+	−	−				13
5	−	−	+	+	−	+	+	+	+	−	−	3	+	−	+	−	+	−	−	+	−	+	−	6
6	+	−	+	−	+	−	−	+	−	+	−		−	+	−	+	−	+	+	−				14
7	−	+	+	−	−	+	−	+	−	−	+	4	−	+	−	+	+	+	−	+	+	+	−	1
8	+	+	+	+	+	+	+	+	+	+	+		+	+	+	−	−	−	+	−				9
9	+	+	+	−	−	−	+	+	+	+	−	5	−	+	+	−	−	−	−	+	−	−	+	7
10	−	+	+	+	+	−	−	+	−	−	−		+	−	−	+	+	+	+	−				15
11	+	−	+	+	−	+	−	+	−	+	+	6	+	−	−	+	−	−	−	+	+	−	+	4
12	−	−	+	−	+	+	+	+	+	−	+		−	+	+	−	+	+	+	−				12
13	+	+	−	−	+	+	−	+	+	+	−	7	−	+	−	−	+	−	+	+	−	+	+	3
14	−	+	−	+	−	+	+	+	−	+	−		+	−	+	+	−	+	−	−				11
15	+	−	−	+	+	−	+	−	−	−	+	8	+	+	+	+	+	+	+	+	+	+	+	8
16	−	−	−	−	−	−	−	−	+	+	+		−	−	−	−	−	−	−	−				16

Source: Table 5.6 is reproduced and adapted with permission from the table on p. 406 of *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*, by G. E. P. Box, W. G. Hunter, and J. S. Hunter, published by John Wiley & Sons, Inc., New York, 1978.

Plackett and Burman Designs

Saturated resolution III two-level designs for up to $k = n - 1$ factors in $n = 2^q$ runs can be generated from any "core" 2^q factorial design by associating new variables with some or all of the interaction columns of the core design. Thus designs of this type for $n = 4, 8, 16, 32, 64, \ldots$ runs are easily generated. For intermediate values of n that are a multiple of 4 but *not* a power of 2, namely, $n = 12, 20, 24, 28, 36, \ldots$, saturated designs given by Plackett and Burman (1946) can be used. (In fact, the Plackett and Burman paper also covers $n = 2^q$ cases.) For $n = 12, 16, 20, 24$, for example, Plackett and Burman supply signs for the first row of the design matrix as follows:

$n = 12$	+ + − + + + − − − + −
$n = 16$	+ + + + − + − + + − − + − − −
$n = 20$	+ + − − + + + + − + − + − − − − + + −
$n = 24$	+ + + + + − + − + + − − + + − − + − + − − − −

To obtain the full design, cyclically permute $(n - 2)$ times the row given, and then add an nth row consisting of all minus signs.

Other constructions are different. For $n = 28$, for example, Plackett and Burman provide three 9×9 blocks of signs, A, B, and C, say, which, set side by side, form nine rows of width $k = 27$. Cyclic permutation to BCA and then CAB gives a total of 27 rows and a final row of minuses completes the $N = 28$ experimental runs. In all, Plackett and Burman (1946, 323–324) provide designs for n equal to a multiple of four for $n \leq 100$, except 92. For $n = 92$, see Baumert et al. (1962). Also see Draper (1985a).

All such designs are resolution III with complicated alias structures. Foldovers of the kinds illustrated in Eqs. (5.4.11) or (5.4.12) produce resolution IV designs. The mirror image pairs produced by foldover may be deployed in blocks of size two, if desired.

Exercise. Write out the Plackett and Burman designs for $n = 12$ and 16. Take the first six columns of the latter and label them $\mathbf{1}, \mathbf{2}, \ldots, \mathbf{6}$. What are the generators of this design. (Answer: Rearrange the first four columns in standard order. It is then clear that $\mathbf{5} = -\mathbf{14}$ and $\mathbf{6} = \mathbf{124}$, so the generators are $\mathbf{I} = -\mathbf{145} = \mathbf{1246}$.)

5.5. DESIGNS OF RESOLUTION V AND OF HIGHER RESOLUTION

When five or more factors are being considered, very useful designs of resolution V or higher are provided by half-fractions, the 2^{k-1} designs. For example, the 2^{5-1} design with generator $\mathbf{I} = \mathbf{12345}$ (i.e., we set $\mathbf{5} = \mathbf{1234}$

after writing down an initial 2^4 design in **1**, **2**, **3**, and **4**) is a 16-run design in which main effects are confounded only with four-factor interactions and two-factor interactions are confounded only with three-factor interactions. Thus, if the assumption that interactions between more than two factors are negligible is correct, the design allows unbiased estimation of all main effects and two-factor interactions. Designs of resolution five or more are of particular value as main building blocks of *composite designs* (presented later in this book) which allow estimation of all the terms in a second degree polynomial approximation. For eight factors or more, one-quarter replicates can produce designs of resolution V. Some useful resolution V designs and

TABLE 5.7. Construction and blocking of some designs of resolution V and higher so that no main effect or two-factor interaction is confounded with any other main effect or two-factor interaction

(1) Number of variables	(2) Number of runs	(3) Degree of fractionation	(4) Type of design	(5) Method of introducing "new" factors	(6) Blocking (with no main effect or 2fi confounded)	(7) Method of introducing blocks
5	16	$\frac{1}{2}$	2_{V}^{5-1}	$\pm\mathbf{5} = \mathbf{1234}$	Not available	
6	32	$\frac{1}{2}$	2_{VI}^{6-1}	$\pm\mathbf{6} = \mathbf{12345}$	Two blocks of 16 runs	$\mathbf{B}_1 = \mathbf{123}$
7	64	$\frac{1}{2}$	2_{VII}^{7-1}	$\pm\mathbf{7} = \mathbf{123456}$	Eight blocks of 8 runs	$\mathbf{B}_1 = \mathbf{1357}$ $\mathbf{B}_2 = \mathbf{1256}$ $\mathbf{B}_3 = \mathbf{1234}$
8	64	$\frac{1}{4}$	2_{V}^{8-2}	$\pm\mathbf{7} = \mathbf{1234}$ $\pm\mathbf{8} = \mathbf{1256}$	Four blocks of 16 runs	$\mathbf{B}_1 = \mathbf{135}$ $\mathbf{B}_2 = \mathbf{348}$
9	128	$\frac{1}{4}$	2_{VI}^{9-2}	$\pm\mathbf{8} = \mathbf{13467}$ $\pm\mathbf{9} = \mathbf{23567}$	Eight blocks of 16 runs	$\mathbf{B}_1 = \mathbf{138}$ $\mathbf{B}_2 = \mathbf{129}$ $\mathbf{B}_3 = \mathbf{789}$
10	128	$\frac{1}{8}$	2_{V}^{10-3}	$\pm\mathbf{8} = \mathbf{1237}$ $\pm\mathbf{9} = \mathbf{2345}$ $\pm\overline{\mathbf{10}} = \mathbf{1346}$	Eight blocks of 16 runs	$\mathbf{B}_1 = \mathbf{149}$ $\mathbf{B}_2 = \mathbf{12\overline{10}}$ $\mathbf{B}_3 = \mathbf{89\overline{10}}$
11	128	$\frac{1}{16}$	2_{V}^{11-4}	$\pm\mathbf{8} = \mathbf{1237}$ $\pm\mathbf{9} = \mathbf{2345}$ $\pm\overline{\mathbf{10}} = \mathbf{1346}$ $\pm\overline{\mathbf{11}} = \mathbf{1234567}$	Eight blocks of 16 runs	$\mathbf{B}_1 = \mathbf{149}$ $\mathbf{B}_2 = \mathbf{12\overline{10}}$ $\mathbf{B}_3 = \mathbf{89\overline{10}}$

Source: Table 5.7 is reproduced and adapted with permission from the table on p. 408 of *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*, by G. E. P. Box, W. G. Hunter, and J. S. Hunter, published by John Wiley & Sons, Inc., New York, 1978.

TABLE 5.8. Two-level fractional factorial designs for k variables and N runs (numbers in parentheses represent replication)

Number of variables k

Number of runs N	3	4	5	6	7
4	2_{III}^{3-1} $\pm 3 = 12$				
8	2^3	2_{IV}^{4-1} $\pm 4 = 123$	2_{III}^{5-2} $\pm 4 = 12$ $\pm 5 = 13$	2_{III}^{6-3} $\pm 4 = 12$ $\pm 5 = 13$ $\pm 6 = 23$	2_{III}^{7-4} $\pm 4 = 12$ $\pm 5 = 13$ $\pm 6 = 23$ $\pm 7 = 123$
16	2^3 2 times	2^4	2_{V}^{5-1} $\pm 5 = 1234$	2_{IV}^{6-2} $\pm 5 = 123$ $\pm 6 = 234$	2_{IV}^{7-3} $\pm 5 = 123$ $\pm 6 = 234$ $\pm 7 = 134$
32	2^3 4 times	2^4 2 times	2^5	2_{VI}^{6-1} $\pm 6 = 12345$	2_{IV}^{7-2} $\pm 6 = 1234$ $\pm 7 = 1245$
64	2^3 8 times	2^4 4 times	2^5 2 times	2^6	2_{VII}^{7-1} $\pm 7 = 123456$
128	2^3 16 times	2^4 8 times	2^5 4 times	2^6 2 times	2^7
	(16)	(8)	(4)	(2)	(1)

Number of variables k

Number of runs N	8	9	10	11
4				
8				
16	2_{IV}^{8-4} $\pm 5 = 234$ $\pm 6 = 134$ $\pm 7 = 123$ $\pm 8 = 124$	2_{III}^{9-5} $\pm 5 = 123$ $\pm 6 = 234$ $\pm 7 = 134$ $\pm 8 = 124$ $\pm 9 = 1234$	2_{III}^{10-6} $\pm 5 = 123$ $\pm 6 = 234$ $\pm 7 = 134$ $\pm 8 = 124$ $\pm 9 = 1234$ $\pm \overline{10} = 12$	2_{III}^{11-7} $\pm 5 = 123$ $\pm 6 = 234$ $\pm 7 = 134$ $\pm 8 = 124$ $\pm 9 = 1234$ $\pm \overline{10} = 12$ $\pm \overline{11} = 13$
32	2_{IV}^{8-3} $\pm 6 = 123$ $\pm 7 = 124$ $\pm 8 = 2345$	2_{IV}^{9-4} $\pm 6 = 2345$ $\pm 7 = 1345$ $\pm 8 = 1245$ $\pm 9 = 1235$	2_{IV}^{10-5} $\pm 6 = 1234$ $\pm 7 = 1235$ $\pm 8 = 1245$ $\pm 9 = 1345$ $\pm \overline{10} = 2345$	2_{IV}^{11-6} $\pm 6 = 123$ $\pm 7 = 234$ $\pm 8 = 345$ $\pm 9 = 134$ $\pm \overline{10} = 145$ $\pm \overline{11} = 245$ $(\frac{1}{128})$
64	2_{V}^{8-2} $\pm 7 = 1234$ $\pm 8 = 1256$	2_{IV}^{9-3} $\pm 7 = 1234$ $\pm 8 = 1356$ $\pm 9 = 3456$	2_{IV}^{10-4} $\pm 7 = 2346$ $\pm 8 = 1346$ $\pm 9 = 1245$ $\pm \overline{10} = 1235$	2_{IV}^{11-5} $\pm 7 = 345$ $\pm 8 = 1234$ $\pm 9 = 126$ $\pm \overline{10} = 2456$ $\pm \overline{11} = 1456$ $(\frac{1}{64})$
128	2_{VIII}^{8-1} $\pm 8 = 1234567$	2_{VI}^{9-2} $\pm 8 = 13467$ $\pm 9 = 23567$	2_{V}^{10-3} $\pm 8 = 1237$ $\pm 9 = 2345$ $\pm \overline{10} = 1346$	2_{V}^{11-4} $\pm 8 = 1237$ $\pm 9 = 2345$ $\pm \overline{10} = 1346$ $\pm \overline{11} = 1234567$ $(\frac{1}{32})$
	(1)	$(\frac{1}{2})$	$(\frac{1}{4})$	$(\frac{1}{8})$ $(\frac{1}{16})$

Source: Table 5.8 is reproduced and adapted with permission from the table on p. 410 of *Statistics for Experimenters: An Introduction to Design, Data analysis and Model Building*, by G. E. P. Box, W. G. Hunter, and J. S. Hunter, published by John Wiley & Sons, Inc., New York, 1978. For additional details see that book and "Minimum aberration 2^{k-p} designs," by A. Fries and W. G. Hunter *Technometrics*, **22**, 1980, 601–608, The numbers in parentheses show the replication factors in a "North-West" direction.

associated blocking arrangements are shown in Table 5.7. Note that the design resolution has been attached as a subscript, for example, 2_{V}^{8-2}.

A Table of 2^{k-p} Designs

Extensive tables of 2^{k-p} designs of various resolutions are available; see, for example, National Bureau of Standards (1957). A short but useful listing is given in Table 5.8.

Calculation of Effects Using Yates' Method

In the foregoing discussion, we have shown how two-level fractional factorial designs requiring $2^{k-p} = 2^q$ runs may be generated by first writing down a complete 2^q factorial as a "core" design and then using interaction columns of the 2^q design to accommodate additional factors. To analyze the resulting fractional factorial, Yates' method may be used to compute the $2^q - 1$ contrasts *from the complete core factorial* in the usual way. The string

of effects and interactions estimated by each of the $2^q - 1$ contrasts is then obtained from the alias pattern for the design. (See also Berger, 1972.)

5.6. APPLICATION OF FRACTIONAL FACTORIAL DESIGNS TO RESPONSE SURFACE METHODOLOGY

Suppose the expected response $E(y) = \eta$ is a function of k predictor variables $x_1, x_2, \ldots, x_k$, coded so that the center of the region of interest is at the origin $(0, 0, \ldots, 0)$. Consider a Taylor's series expansion

$$E(y) = \eta = \eta_0 + \sum_{i=1}^{k} \left[\frac{\partial \eta}{\partial x_i}\right]_0 x_i + \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} \left[\frac{\partial^2 \eta}{\partial x_i \, \partial x_j}\right]_0 x_i x_j + \cdots \tag{5.6.1}$$

where the subscript zero indicates evaluation at the origin $(0, 0, \ldots, 0)$. If we ignore terms of order higher than the first, the expansion yields the first degree (or first-order, or planar) approximation

$$\eta = \beta_0 + \sum_{i=1}^{k} \beta_i x_i. \tag{5.6.2}$$

If, in addition, we retain terms of second degree, we obtain the second degree (or second-order, or quadratic) approximation

$$\eta = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \sum_{j \geq i}^{k} \beta_{ij} x_i x_j. \tag{5.6.3}$$

We shall use these approximations a great deal in subsequent chapters and will refer to designs suitable for collecting data to estimate the β parameters in Eqs. (5.6.2) and (5.6.3) as *first-order designs* and *second-order designs*, respectively.

Clearly, if two-level fractional factorial designs are to be used as first-order designs, they must have resolution of at least III so that, if model (5.6.2) is valid, the main effect contrasts of the design will provide estimates of all the β_i unconfounded with one another.

When a two-level design is used as a building block in second-order composite designs, it is usually* of resolution at least V. Then, if model (5.6.3) is valid, the main effect and two-factor interaction contrasts of the two-level design provide estimates of all main effect coefficients β_i and of all

*However, also see Section 15.5.

interaction coefficients β_{ij} $(i \neq j)$, unconfounded with one another. (See also Sections 9.2, 13.8, 14.3, and 15.3.)

Because we are never sure about the adequacy of an approximation of selected degree (or order), we shall need to check lack of fit as well as to estimate the model parameters. Resolution III designs that are not saturated can allow estimation of suspect interactions. Also (as we shall see in Chapter 9) the addition of center points makes possible a general curvature check. Resolution IV designs provide safer first-order designs because first-order coefficients will not then be confused with any two-factor interactions. Although the latter are ignored in a first-order fitting, they may yet exist; a resolution IV design will enable combinations of them to be estimated as a check of lack of fit, if desired.

5.7. PLOTTING EFFECTS FROM FRACTIONAL FACTORIALS ON PROBABILITY PAPER

To distinguish real effects from noise, estimates of effects from fractional factorial designs may be plotted on normal probability paper (see Appendix 4B) in just the same way as for full factorial designs.

Numerical Example, 2_{IV}^{8-4} Design

Sixteen observations were obtained from a 2_{IV}^{8-4} design given by Box, Hunter, and Hunter (1978, Table 12.11). The mean of these was $\bar{y} = 19.75$ and there were 15 estimates in addition, namely of:

1. Eight main effects: $-0.7, -0.1, 5.5, -0.3, -3.8, -0.1, 0.6, 1.2$.
2. Seven sets of four two-factor interactions: $-0.6, 0.9, -0.2, -0.4, -0.6, -0.3, 4.6$

These effects are plotted on normal probability paper in Figure 5.1. The plot suggests that two main effects (of variables 3 and 5) and one combination of two-factor interactions (15 + 26 + 38 + 47) are distinguishable from the noise.

EXERCISES

5.1. Are the following statements true or false, for a 2_R^{k-p} design?
(a) $k \geq R$
(b) $k \geq R + p$
Give, in a few words, reasons for your answers.

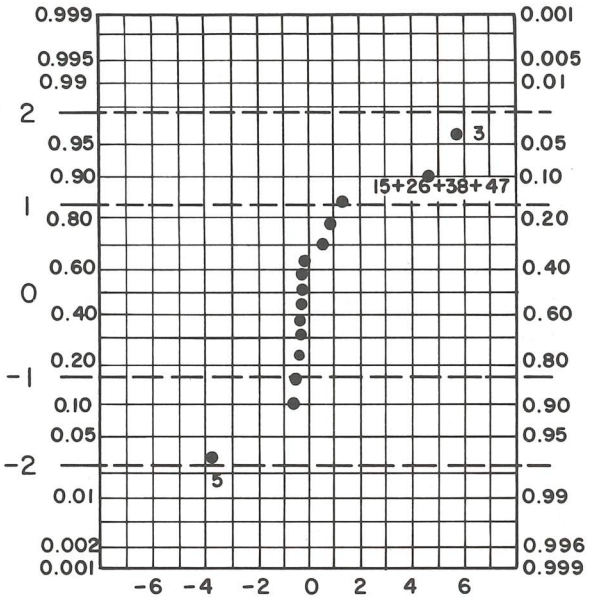

FIGURE 5.1. Plot of 15 estimated effects from a 2^{8-4} factorial on normal probability paper.

5.2. A 2_{III}^{5-2} design with generators $\mathbf{I} = \mathbf{1234} = \mathbf{135}$ is run.
(a) If it is *known* that factors 1, 2, and 3 all act independently of one another, and that factors 3, 4, and 5 all act independently of one another, what (ignoring interactions of three or more factors) can we estimate?
(b) Can you suggest a better 2_{III}^{5-2} design for these factors, or is the one given the best available?

5.3. For the construction of a one-quarter replicate of a 2^5 factorial, that is, a 2^{5-2} factorial, two generating relations are proposed:
(a) $\mathbf{I} = \mathbf{ABCDE} = \mathbf{BCDE}$
(b) $\mathbf{I} = \mathbf{ABC} = \mathbf{BCDE}$
Which design would you prefer, and why?

5.4. (a) Using the principle of foldover, show how the 2_{III}^{7-4} design generated by

$$\mathbf{I} = \mathbf{1234} = \mathbf{125} = \mathbf{146} = \mathbf{247}$$

can provide a 2_{IV}^{8-4} design.
(b) What are the generators of the new design?
(c) If this second design is again folded, but the **I** column is *not* folded with it this time, what type is the final design obtained?

5.5. Consider the 2_{IV}^{8-4} design generated by $\mathbf{I} = \mathbf{1235} = \mathbf{1246} = \mathbf{1347} = \mathbf{2348}$.

(a) By using a column that normally provides an estimate of the sum of four two-factor interactions, split the design into two blocks.
(b) Retaining two-factor interactions of all kinds show what estimates are available.
(c) If the block variable does not interact, show what estimates are available.

5.6. (a) Obtain a design for examining six variables in four blocks each containing four runs.
(b) What are the generators for the design (including blocking)?
(c) Find the defining relation (including blocking).
(d) Show what estimates can be found, assuming interactions between three or more factors are zero and that block variables do not interact with ordinary variables.
(*Hint*: Consider the 2_{IV}^{6-2} design generated by $\mathbf{I} = \mathbf{1235} = \mathbf{1246}$. Associate $\mathbf{B}_1 = \mathbf{12}$, $\mathbf{B}_2 = \mathbf{13}$, so that $\mathbf{B}_1\mathbf{B}_2 = \mathbf{23}$. The design is then generated by $\mathbf{I} = \mathbf{1235} = \mathbf{1246} = \mathbf{12B}_1 = \mathbf{13B}_2$, including blocking. In developing the estimates remember that although $\mathbf{B}_1\mathbf{B}_2\mathbf{5}$, for example, looks like a *three*-factor interaction, it must be regarded as a *two*-factor interaction between the blocking variable $\mathbf{B}_1\mathbf{B}_2$ and the ordinary variable **5**.)
(e) Is the choice suggested in the hint a good one? Explain the reasons for your answer. If you decide that better choices exist, provide one.

5.7. (a) Obtain a design for examining six variables in eight blocks each containing two runs so that the estimates of main effects are not influenced by block effects.
(b) What are the generators for the design (including blocking)?
(c) Find the defining relation.
(d) Show what estimates can be found assuming interactions between three or more factors are zero and that block variables do not interact with ordinary variables.
(*Hint*: Consider the 2_{IV}^{6-2} design generated by $\mathbf{I} = \mathbf{1235} = \mathbf{1246}$. Associate $\mathbf{B}_1 = \mathbf{12}$, $\mathbf{B}_2 = \mathbf{13}$ (so that $\mathbf{B}_1\mathbf{B}_2 = \mathbf{23}$), and $\mathbf{B}_3 = \mathbf{14}$. Then $\mathbf{B}_1\mathbf{B}_3 = \mathbf{24}$, $\mathbf{B}_2\mathbf{B}_3 = \mathbf{34}$, and $\mathbf{B}_1\mathbf{B}_2\mathbf{B}_3 = \mathbf{1234}$. The blocks are now defined. Remember that, for example, $\mathbf{B}_1\mathbf{B}_2\mathbf{B}_3\mathbf{5}$ has the status of a two-factor interaction.)

5.8. An experimenter begins an investigation with a 2^{5-2} design using generators $\mathbf{4} = \mathbf{123}$ and $\mathbf{5} = \mathbf{23}$. Later he performs the eight foldover runs obtained by reversing all the signs in the five-variable design table.
(a) What is the resolution of the combined 16-run design? What generators would produce this design?
(b) What better design might the experimenter have used had he known in advance that 16 runs would be made?

5.9. Construct a 16-run two-level fractional factorial design in six factors such that all main effects and all two-factor interactions associated with factor **1** are estimable.

5.10. If the design $\mathbf{I} = \mathbf{124} = \mathbf{135} = \mathbf{236}$ is followed by the design in which variables **1** and **2** have their signs switched, what is the alias structure of the 16-run design so formed?

5.11. An experimenter needs a 16-run two-level fractional factorial design that will estimate the main effects $1, 2, \ldots, 6$ of six variables as well as the interactions 12, 123, 13, 14, 15, 16, 23, 24 between them, assuming all interactions of three or more factors can be ignored except for 123. Does such a design exist? If so, what is its defining relation?

5.12. Verify that the 2^{8-4} resolution IV design can be used in all the following alternative ways. (Assume that all interactions between three or more factors can be neglected.)

(a) As a "main effects" design to obtain the main effects of eight factors clear of all two factor interactions.

(b) As in (a) but blocked into two, or four, or eight blocks.

(c) If there are three major variables, all of whose main effects, two-factor interactions, and three-factor interaction are needed, and eight minor variables all of whose interactions can be neglected, the design can be used to estimate all these. (*Hint*: Look at Eq. (5.4.13) and associate each $\mathbf{B}_i$ with a major variable $\mathbf{M}_i$, $i = 1, 2, 3$. Remember that all the two-factor interactions involving numbers, e.g., 12, 13, etc., are zero. Take the numbered variables as the minor variables.)

(d) As in (c) but with eight minor variables, two major variables, and in two blocks of eight runs each. (*Hint*: Replace $\mathbf{B}_1$, $\mathbf{B}_2$ and $\mathbf{B}_3$ in Eq. (5.4.13) by $\mathbf{M}_1$, $\mathbf{M}_2$, and $\mathbf{B}$.)

(e) As a design for eight minor variables and one major variable in four blocks of four runs each.

(f) As a screening design that will provide a complete factorial, twice over, in *any* set of three variables, if the other five variables have no effect. (*Hint*: If five variables are not effective we can "cross them out" in the defining relationship that leads to the alias structure. Since no words of length three or smaller appear in this—because the design is of resolution IV—all words are "deleted," no matter which five variables are chosen, implying that a full factorial design remains in the other three variables. Because the original design had 16 runs, the 2^3 is necessarily replicated.)

(g) As a screening design that will provide a complete factorial in some sets of four variables, or a 2^{4-1}, twice over, in all other sets of four variables, provided the remaining four variables *not* in the considered set have no effect. (*Hint*: If the chosen four variables form a word in the defining relationship, a replicated 2^{4-1} occurs. Otherwise we get a 2^4.)

5.13. Select a set of suitable generators for a 2_{V}^{10-3} design and write out the defining relation which provides the alias structure. Select a suitable blocking generator for division into two blocks. (*Hint*: See Table 5.7.)

5.14. A 2_{V}^{5-1} design was used to examine five factors A, B, C, D, and E. It was assumed that all interactions involving three or more variables could be neglected. From the resulting data, the following estimates (rounded to the nearest unit) were obtained:

$A \leftarrow$	0	$AB \leftarrow$	17	$BD \leftarrow$	-4
$B \leftarrow$	-8	$AC \leftarrow$	3	$BE \leftarrow$	8
$C \leftarrow$	14	$AD \leftarrow$	5	$CD \leftarrow$	14
$D \leftarrow$	67	$AE \leftarrow$	-2	$CE \leftarrow$	-5
$E \leftarrow$	4	$BC \leftarrow$	8	$DE \leftarrow$	0

Plot these results on normal probability paper and identify effects that appear to be important. Remove these, replot the rest, and estimate σ, where $V(y_i) = \sigma^2$. See Appendix 4B for additional details.

5.15. Identify the design given in Table E5.15 and perform an appropriate analysis. Assume $V(y_i) = \sigma^2 = 2$.

What factors appear to influence the response? What features need to be further elucidated? If you were allowed to perform just four more runs, what would they be? Why?

TABLE E5.15

x_1	x_2	x_3	x_4	y
1	-1	-1	-1	105
-1	1	-1	-1	107
-1	-1	1	-1	102
-1	-1	-1	1	104
1	1	1	-1	114
1	1	-1	1	111
1	-1	1	1	105
-1	1	1	1	107

5.16. What design is given in Table E5.16? Perform an appropriate factorial analysis of the data.

Would you like to have additional data to clarify what you have seen in this experiment? If four more runs were authorized what would you suggest? Why? If eight more runs were authorized what would you suggest? Why?

5.17. Consider a 2_{V}^{5-1} fractional factorial design defined by $\mathbf{I} = -\mathbf{12345}$. Suppose we wish to choose two blocking variables $\mathbf{B}_1$ and $\mathbf{B}_2$ so that the design can be divided into four blocks. Can this be done in such a way that all main effects

TABLE E5.16

x_1	x_2	x_3	x_4	x_5	y
−1	−1	1	−1	−1	18, 15
1	−1	−1	−1	1	16, 18
−1	1	−1	−1	1	18, 17
1	1	1	−1	−1	15, 16
−1	−1	1	1	1	17, 19
1	−1	−1	1	−1	31, 30
−1	1	−1	1	−1	18, 18
1	1	1	1	1	30, 27

and all two-factor interactions are *not* confounded with blocks? If your answer is yes, write out the design and show what estimates can be made.

5.18. (a) You have convinced an experimenter that she should use a 2^{7-1} design. Each run takes about 20 min, and 16 runs is the most she wishes to attempt during one day's work. Thus she wishes the design divided into four blocks of 16 runs each. Assume that the design will be generated using $\mathbf{I} = \mathbf{1234567}$ and blocked using blocking generators $\mathbf{B}_1 = \mathbf{1357}$, $\mathbf{B}_2 = \mathbf{1256}$. Write out the 16 runs in the $(\mathbf{B}_1, \mathbf{B}_2) = (+, +)$ block.

(b) After these 16 runs have been completed, the experimental unit suffers a breakdown. It will not be repaired until after the monthly meeting at which the experimenter has to make a report. Assuming that all interactions involving three or more of the seven factors are negligible, and that blocks do not interact with the factors, write out the 16 quantities that can be estimated using only the 16 runs currently available.

5.19. (Source: Toby J. Mitchell.) Ozzie Cadenza, holder of a Ph.D. in Statistics from the University of Wisconsin, and now owner and manager of Ozzie's Bar and Grill, recently decided to study the factors that influence the amount of business done at his bar.

At first, he did not know which factors were important and which were not, but he drew up the following list of six, which he decided to investigate by means of a fractional factorial experiment:

1. The amount of lighting in the bar.
2. The presence of free potato chips and chip dip at the bar.
3. The volume of the juke box.
4. The presence of Ozzie's favorite customer, a young lady by the name of Rapunzel Freeny. Miss Freeny was a real life-of-the-party type, continually chatting with the customers, passing around the potato chips, and so on, all of which made Ozzie feel that she had a real effect on the amount of his bar business.
5. The presence of a band of roving gypsies, who had formed a musical group called the Roving Gypsy Band and who had been hired by Ozzie to play a limited engagement there.
6. The effect of the particular bartender who happened to be on duty. There were originally three bartenders, Tom, Dick, and Harry, but Harry was fired so that each factor in the experiment would have only two levels.

Plus and minus levels were assigned to these six factors as follows:

	−	+
1.	Lights are dim	Lights are bright
2.	No chips are at the bar	Chips are at the bar
3.	Juke box is playing softly	Juke box is blaring loudly
4.	Miss Freeny stays at home	Miss Freeny is there
5.	Gypsies are not there	Gypsies are there
6.	Tom is the bartender	Dick is the bartender

Ozzie decided to perform one "run" every Friday night during the cocktail hour (4:30 to 6:30 P.M.). He thought he should try a fractional factorial with as few runs as possible, since he was never quite sure just when the Roving Gypsy Band would pack up and leave. He finally decided to use a member of the family of 2_{III}^{6-3} designs with principal generating relation $\mathbf{I} = \mathbf{124} = \mathbf{135} = \mathbf{236}$. (He had wanted to find a resolution III design which would be such that the juke box would never be blaring away while the gypsies were playing but found this requirement to be impossible.

(a) Why?

He did insist, however, that in *no* run of the experiment could variables **1**, **3**, and **5** attain their plus levels simultaneously. This restriction was made necessary by the annoying tendency of all the lights to fuse whenever the gypsies plugged in their electric zither at the same time the lights and the juke box were on full blast. Note that this restriction made it impossible for the principal member of the chosen family to be used.

(b) What members of the given family *does* this restriction allow?)

Ozzie settled on the generating relation. $\mathbf{I} = \mathbf{124} = -\mathbf{135} = \mathbf{236}$. The design matrix and the "response" (income in dollars) corresponding to each run are given in Table E5.19.

(c) Assuming third- and higher-order interactions are negligible, write down the estimates obtained from this experiment and tell what they estimate.

TABLE E5.19

1	2	3	4 = 12	5 = −13	6 = 23	y
−	−	−	+	−	+	265
+	−	−	−	+	+	155
−	+	−	−	−	−	135
+	+	−	+	+	−	205
−	−	+	+	+	−	195
+	−	+	−	−	−	205
−	+	+	−	+	+	125
+	+	+	+	−	+	315

(d) In a few sentences, tell what fraction you would perform next and why.

Being partial to Miss Freeny and encouraged by the results of the first fraction, Ozzie chose a second fraction which would give unaliased estimates of her and each of her interactions. The results of this second fraction, given with variables **1**, **2**, and **3** in standard order, were 135, 165, 285, 175, 205, 195, 295, 145.

(e) Write down the estimates obtained by combining the results of both fractions.

(f) In the light of these results, was the choice of the second fraction a wise one?

(g) Offer a brief conjecture which might explain the presence and direction of the interactions involving Miss Freeney.

5.20. (Source: Designed experiments, by K. R. Williams, *Rubber Age*, **100**, August 1968, 65–71.) Table E5.20 shows a 28-run Plackett and Burman type design for 24 variables together with the responses obtained from each run. Except for the last column, the CP row (cross-product) is the result of taking CPs of each column of signs with the y values and the EF row ("effect") is the CP row divided by 14 and rounded. In the last column, $\Sigma y = 3065$ and $\bar{y} = 109$. Confirm one or more of the effects calculations. If, a priori, it were believed that at least one-third of the factors were ineffective, but not known which factors these were, which factors would be regarded as effective ones?

5.21. A 2^3 factorial design has been partially replicated as shown in Table E5.21. Analyze the data using regression methods, with a "full factorial analysis" type model.

5.22. (Source: Sucrose-modified phenolic resins as plywood adhesives, by C. D. Chang and O. K. Kononenko, *Adhesives Age*, **5**, (7), July 1962, 36–40. Material reproduced and adapted with the permission of *Adhesives Age*. The

TABLE E5.20

Run Number	Factor Number 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Factor number 17	18	19	20	21	22	23	24	Response y
1	+	+	+	−	−	−	+	+	+	+	+	−	+	−	−	+	+	+	−	+	−	+	−	+	133
2	−	+	−	−	−	−	+	+	+	+	−	+	−	−	+	−	+	+	−	+	−	+	−	−	49
3	+	−	−	−	−	−	−	+	+	−	−	−	+	−	+	+	−	+	−	−	+	+	+	−	62
4	+	−	−	+	+	+	−	−	−	+	−	+	+	+	+	+	+	−	−	−	−	+	−	−	45
5	+	+	−	+	+	+	−	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+	−	+	88
6	+	+	−	−	+	+	−	−	−	+	+	−	+	−	−	+	−	+	+	+	+	−	−	+	52
7	−	−	+	+	−	+	+	+	−	+	−	+	+	−	+	+	−	+	+	+	+	−	−	−	300
8	−	−	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	−	+	−	+	+	+	56
9	−	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	47
10	−	−	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	88
11	−	−	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	116
12	+	+	+	+	−	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	83
13	+	+	+	+	−	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	193
14	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	230
15	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	51
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	82
17	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	32
18	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	58
19	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	201
20	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	56
21	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	97
22	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	53
23	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	276
24	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	145
25	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	130
26	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	55
27	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	160
28	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	127
CP	−365	−281	−95	511	175	−165	−189	415	213	199	−51	−221	−321	−433	−1209	−321	−599	−191	−49	−683	−117	−451	−167	−173	3065
EF	−26	−20	−7	37	13	−12	−14	30	15	14	−4	−16	−23	−31	−86	−23	−43	−14	−4	−49	−8	−32	−12	−12	109

[a] The level of run 8, factor 20 is here shown as +. In the source reference, the lower level appears; that seems to be a typographical error.

TABLE E5.21

x_1	x_2	x_3	y
−	−	−	46
+	−	−	61, 57
−	+	−	37, 45
+	+	−	56
−	−	+	37, 41
+	−	+	68
−	+	+	33
+	+	+	66, 68

data were obtained under a grant from the Sugar Research Foundation, Inc., now the World Sugar Research Organization, Ltd.) The object of a research study was to improve both the dry and wet strip-shear strength in pounds per square inch (psi) of Douglas fir plywood glued with a resin modified with sugar. Here, we shall examine only the dry strength data, however.

The first part of the investigation involved six reaction variables $x_1, x_2, \ldots, x_6$, each examined at two levels coded to -1 and 1. The variables are shown in Table E5.22a, together with the coding used for each variable.

A 16-run 2_{IV}^{6-2} design generated by $\mathbf{I} = \mathbf{1235} = \mathbf{2346}$ (so that $\mathbf{I} = \mathbf{1456}$, also) was performed, and it provided the responses in Table E5.22b. (The y values are average values of six to eight samples per run; the fact that this provides observations whose variances may vary slightly is ignored in what follows.) Five additional center-point runs at $(0, 0, \ldots, 0)$ provided a pure error estimate $s_e^2 = 95.30$ for $V(y) = \sigma^2$.

Perform a factorial analysis on these data, assuming all interactions involving three or more factors can be neglected, and so confirm the results in Table E5.22c. Note, in the left column of that table, that "other" denotes groups of interactions of order three or more. These should represent "error"

TABLE E5.22a. Levels of the six predictor variables examined

Variable (units), designation	Level −1	Level 1	Coding
Sucrose (g), S	43	71	$x_1 = (S - 57)/14$
Paraform (g), P	30	42	$x_2 = (P - 36)/6$
NaOH (g), N	6	10	$x_3 = (N - 8)/2$
Water (g), W	16	20	$x_4 = (W - 18)/2$
Temperature maximum (°C), T	80	90	$x_5 = (T - 85)/5$
Time at T°C, t	25	35	$x_6 = (t - 30)/5$

TABLE E5.22b. The design and the response values

x_1	x_2	x_3	x_4	x_5	x_6	y
−1	−1	−1	−1	−1	−1	162[a]
1	−1	−1	−1	1	−1	146
−1	1	−1	−1	1	1	182
1	1	−1	−1	−1	1	133
−1	−1	1	−1	1	1	228
1	−1	1	−1	−1	1	143
−1	1	1	−1	−1	−1	223
1	1	1	−1	1	−1	172
−1	−1	−1	1	−1	1	168
1	−1	−1	1	1	1	128
−1	1	−1	1	1	−1	175
1	1	−1	1	−1	−1	186
−1	−1	1	1	1	−1	197
1	−1	1	1	−1	−1	175
−1	1	1	1	−1	1	196
1	1	1	1	1	1	173

[a]Shown as 172 in the source reference; however, according to Dr. O. K. Kononenko, the 172 may be a typographical error.

if interactions between three or more factors are negligible, and so should not be statistically significant under this tentative assumption. The designation (124 +) stands for (124 + 345 + 136 + 2456) while (134 +) represents (134 + 245 + 126 + 3456). The superscript a's indicate estimated effects that exceed 2.78(s.e.) in modulus. (The standard error is based on five average readings, and thus on 4 df. Ninety-five percent of the t_4 distribution lies within ± 2.78 so that estimates exceeding 2.78 (s.e.) in modulus are those whose true values are tentatively judged to be nonzero.)

At this stage, it was decided to explore outward on a path of steepest ascent, ignoring the two significant interactions and this was successful, leading to the following tentative conclusions:

1. Variables x_5 and x_6 have shown little effect and may be dropped in future runs.
2. A good balance of dry and wet strength values was achieved around $(S, P, N, W) = (47, 36.6, 8.8, 23.8)$, and this point will be the center for a subsequent experiment.
3. A 27-run three-level rotatable design with blocking, equivalent to a rotated "cube plus star" design (see Section 15.4 and/or Box and Behnken, 1960a) with added center points, will be performed next, and a

TABLE E5.22c. The estimated factorial effects from the 2^{6-2} experiment

Effect name	Estimate
1	−34.375[a]
2	11.625
3	28.375[a]
4	1.125
5	1.875
6	−10.625
12 + 35	6.375
13 + 25	−10.875
14 + 56	15.875[a]
15 + 23 + 46	−6.375
16 + 45	−14.875[a]
24 + 36	3.875
26 + 34	−7.375
other (124 +)	6.125
other (134 +)	6.875
s.e.	4.88
2.78 (s.e.)	13.57

[a] Significant at $\alpha = 0.05$.

second-order surface will be fitted to the y values thus obtained. (For the continuation of this work, see Exercise 7.23. For steepest ascent, see Chapter 6.)

5.23. (Source: A first-order five variable cutting-tool temperature equation and chip equivalent, by S. M. Wu and R. N. Meyer, *J. Eng. Indus., Trans. ASME, Series B*, **87**, 1965, 395–400. Adapted with the permission of the copyright holder, The American Society of Mechanical Engineers, Copyright © 1965.) Fit a first-order model $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_5 x_5 + \varepsilon$ by least squares to the data in Table E5.23. The design is a 2_{V}^{5-1} design generated by $x_1 x_2 x_3 x_4 x_5 = 1$ where the x's are coded predictor variables. The actual response is T (temperature in °F) but the model should be fitted in terms of $y = \ln T$. Treat the two values at each set of conditions as repeat runs for pure error purposes. Test the model for lack of fit and, if none is revealed, test if all x's are needed. Decide what model is suitable for these data, check the residuals, especially against the testing order, and interpret what you have done. (Some complications stemming from blocking have been ignored in this exercise modification; see the source reference for additional details.)

5.24. Is it possible to obtain a 2_R^{k-p} design which provides clear estimates of the main effects of nine variables in 16 runs, assuming all interactions involving three or more variables are zero? If yes, provide the design. If no, provide the best 16-run design in the circumstances and specify how many two-factor

TABLE E5.23. Coded cutting conditions and temperature results

Testing order	x_1	x_2	x_3	x_4	x_5	Temperature results T_1	T_2
4	−1	−1	−1	−1	1	462	498
16	1	−1	−1	−1	−1	743	736
9	−1	1	−1	−1	−1	714	681
8	1	1	−1	−1	1	1070	1059
11	−1	−1	1	−1	−1	474	486
5	1	−1	1	−1	1	832	810
1	−1	1	1	−1	1	764	756
15	1	1	1	−1	−1	1087	1063
14	−1	−1	−1	1	−1	522	520
6	1	−1	−1	1	1	854	828
2	−1	1	−1	1	1	773	756
10	1	1	−1	1	−1	1068	1063
7	−1	−1	1	1	1	572	574
12	1	−1	1	1	−1	831	856
13	−1	1	1	1	−1	819	813
3	1	1	1	1	1	1104	1092

interactions would need to be zero to achieve the requirement of clear estimates of main effects.

5.25. Variables **4** and **7** are dropped from the 2_{III}^{7-4} design generated by $\mathbf{I} = \mathbf{124} = \mathbf{235} = \mathbf{136} = \mathbf{1237}$. What is the resulting design?

5.26. The data in Table E5.26 were obtained from the records of a medium-sized dairy company seeking to improve sales. Explain what design is being used,

TABLE E5.26

1	2	3	4	5	Response
Daily deliveries?	Roundsman wears uniform?	Older man?	City Round?	Delivery before noon?	Sales rating
NO	NO	YES	YES	NO	4
NO	NO	YES	NO	YES	3
YES	YES	YES	YES	YES	0
NO	YES	NO	YES	NO	3
YES	NO	NO	YES	YES	0
YES	YES	YES	NO	NO	4
YES	NO	NO	NO	NO	4
NO	YES	NO	NO	YES	3

analyze the data, and suggest what sensible next step(s) should be considered by the company.

5.27. (a) A 2_{III}^{7-4} design with generators $\mathbf{I} = -\mathbf{126} = \mathbf{135} = -\mathbf{237} = -\mathbf{1234}$ is used in a factory investigation of the time taken to pack 100 standard items. The variables investigated are given in Table E5.27. With variables **1**, **2**, and **3** in standard order, the observed packing times were, in minutes, 46.1, 55.4, 44.1, 58.7, 56.3, 18.9, 46.4, 16.4. Perform a standard analysis and state tentative conclusions.

(b) It was decided to perform a second fractional factorial, obtained from the first by reversing the signs of all variables. For these eight experiments, the packing times were (with **1**, **2** and **3** in order $+++, -++, +-+, --+, ++-, -+-, +--, ---$,) 41.8, 40.1, 61.5, 37.0, 22.9, 34.1, 17.7, 42.7. Perform a standard analysis.

(c) Combine the two fractions given above to obtain estimates of main effects and two factor interactions. Comment on the results. If you were the factory owner, what would you do?

TABLE E5.27

Variable	(−1, 1)
1: Foreman	(absent, present)
2: Sex of packer	(man, woman)
3: Time of day	(morning, afternoon)
4: Temperature	(normal, high)
5: Music	(none, piped in)
6: Age of packer	(under 25, 25 or over)
7: Factory location	(Los Angeles, New York)

TABLE E5.28

Variable Number						
1	**2**	**3**	**4**	**5**	**6**	Response, y
+	+	−	−	−	−	29
−	+	+	−	+	−	27
+	−	−	+	+	−	42
−	+	−	+	−	+	0
−	−	−	−	+	+	30
+	+	+	+	+	+	0
−	−	+	+	−	−	39

5.28. An industrial spy is disturbed while photographing experimental records. On later developing his film, he discovers he has only seven runs of a 2_R^{k-p} design, given in Table E5.28. How can he estimate the effects from the incomplete data? Do this. The information from these tests is vital to the spy's employer who is prepared to run another 2^{k-p} block if the results from the photographed data indicate this to be a worthwhile endeavor. Should additional tests be run? If so, what would you recommend?