8

Inference for Full and Fractional Factorial Studies

Chapter 7 began this book's exposition of inference methods for multisample studies. The methods there neither require nor make use of any special structure relating the samples. They are both widely applicable and practically informative tools. But Chapter 4 illustrated on an informal or descriptive level the engineering importance of discovering, interpreting, and ultimately exploiting structure relating a response to one or more other variables. This chapter begins to provide inference methods to support these activities.

This chapter builds on the descriptive statistics material of Section 4.3 and the tools of Chapter 7 to provide methods for full and fractional factorial studies. It begins with a discussion of some inference methods for complete two-way factorials. Then complete p-way factorial inference is considered with special attention to the 2^p case. Then two successive sections describe what is possible in the way of factorial inference from well-chosen fractions of a 2^p factorial. First, half fractions are considered, and then $1/2^q$ fractions for $q > 1$.

8.1 Basic Inference in Two-Way Factorials with Some Replication

This section considers inference from complete two-way factorial data in cases where there is some replication—i.e., at least one of the sample sizes is larger than 1. It begins by pointing out that the material in Sections 7.1 through 7.4 can often be useful in sharpening the preliminary graphical analyses suggested in Section 4.3. Then there is a discussion of inference based on the fitted two-way factorial effects defined in Chapter 4. These are used to develop both individual and simultaneous confidence interval methods.

8.1.1 One-Way Methods in Two-Way Factorials

Example 1 revives a case used extensively in Section 4.3.

Example 1
(Example 7, Chapter 4, revisited—page 163)

Joint Strengths for Three Different Joint Types in Three Different Woods

Consider again the wood joint strength study of Kotlers, MacFarland, and Tomlinson. Table 8.1 reorganizes the data given earlier in Table 4.11 into a 3×3 table showing the nine different samples of one or two joint strengths for all combinations of three woods and three joint types. The data in Table 8.1 have complete two-way factorial structure, and seven of the nine combinations represented in the table provide some replication.

Table 8.1
Joint Strengths for 3^2 Combinations of Joint Type and Wood

		Wood		
		1 (Pine)	2 (Oak)	3 (Walnut)
	1 (Butt)	829, 596	1169	1263, 1029
Joint	2 (Beveled)	1348, 1207	1518, 1927	2571, 2443
	3 (Lap)	1000, 859	1295, 1561	1489

The data in Table 8.1 constitute $r = 9$ samples of sizes 1 or 2. Provided the graphical and numerical checks of Section 7.1 reveal no obvious problems with the one-way model for joint strengths, all of the methods of Sections 7.2 through 7.4 can be brought to bear.

Error bars on interaction plots

One way in which this is particularly helpful is in indicating the precision of estimated means on interaction plots. Section 4.3 discussed how near-parallelism on such plots leads to simple interpretations of two-way factorials. By marking either individual or simultaneous confidence limits as **error bars** around the sample means on an interaction plot, it is possible to get a rough idea of the detectability or statistical significance of any apparent lack of parallelism.

Example 1
(continued)

The place to begin a formal analysis of the wood joint strength data is with consideration of the appropriateness of the one-way (normal distributions with a common variance) model for joint strength. Table 8.2 gives some summary statistics for the data of Table 8.1.

Residuals for the joint strength data are obtained by subtracting the sample means in Table 8.2 from the corresponding observations in Table 8.1. In this data set, the sample sizes are so small that the residuals will obviously be highly dependent. Those from samples of size 2 will be plus-and-minus a single number

Example 1 *(continued)*

Table 8.2
Sample Means and Standard Deviations for Nine Joint/Wood Combinations

		Wood		
		1 (Pine)	2 (Oak)	3 (Walnut)
	1 (Butt)	$\bar{y}_{11} = 712.5$ $s_{11} = 164.8$	$\bar{y}_{12} = 1{,}169$	$\bar{y}_{13} = 1{,}146$ $s_{13} = 165.5$
Joint	2 (Beveled)	$\bar{y}_{21} = 1{,}277.5$ $s_{21} = 99.7$	$\bar{y}_{22} = 1{,}722.5$ $s_{22} = 289.2$	$\bar{y}_{23} = 2{,}507$ $s_{23} = 90.5$
	3 (Lap)	$\bar{y}_{31} = 929.5$ $s_{31} = 99.7$	$\bar{y}_{32} = 1{,}428$ $s_{32} = 188.1$	$\bar{y}_{33} = 1489$

corresponding to that sample. Those from samples of size 1 will be zero. So there is reason to expect residual plots to show some effects of this dependence. Figure 8.1 is a normal plot of the 16 residuals, and its complete symmetry (with respect to the positive and negative residuals) is caused by this dependence.

Of course, the sample standard deviations in Table 8.2 vary somewhat, but the ratio between the largest and smallest (a factor of about 3) is in no way surprising based on these sample sizes of 2. (Even if only 2 rather than 7 sample variances were involved, since $9(= 3^2)$ is between the .75 and .9 quantiles of the $F_{1,1}$ distribution, the observed level of significance for testing the equality of the two underlying variances would exceed $.2 = 2(1 - .9)$.) And Figure 8.2, which is a plot of residuals versus sample means, suggests no trend in σ as a function of mean response, μ.

In sum, the very small sample sizes represented in Table 8.1 make definitive investigation of the appropriateness of the one-way normal model assumptions

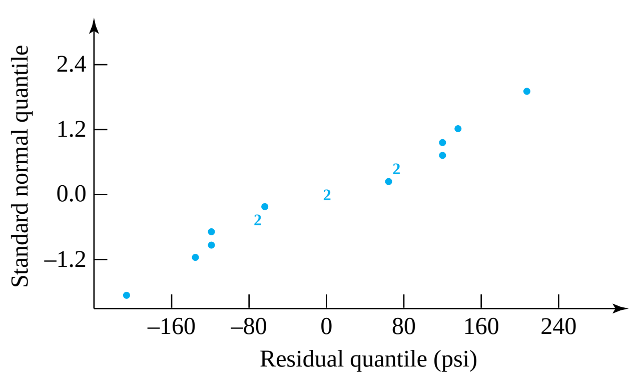

Figure 8.1 Normal plot of 16 residuals for the wood joint strength study

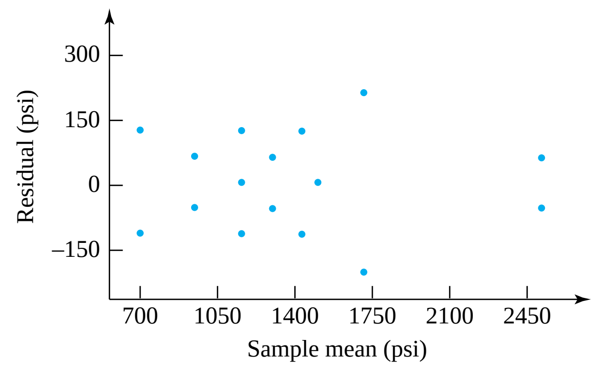

Figure 8.2 Plot of residuals versus sample means for the joint strength study

impossible. But the limited checks that are possible provide no indication of serious problems with operating under those restrictions.

Notice that for these data,

$$s_P^2 = \frac{(2-1)s_{11}^2 + (2-1)s_{13}^2 + (2-1)s_{21}^2 + \cdots + (2-1)s_{32}^2}{(2-1)+(2-1)+(2-1)+\cdots+(2-1)}$$

$$= \frac{1}{7}\left((164.8)^2 + (165.5)^2 + \cdots + (188.1)^2\right)$$

$$= 28{,}805\ (\text{psi})^2$$

So

$$s_P = \sqrt{28{,}805} = 169.7 \text{ psi}$$

where s_P has 7 associated degrees of freedom.

Then, for example, from formula (7.14) of Section 7.2, individual two-sided 99% confidence intervals for the combination mean strengths would have endpoints

$$\bar{y}_{ij} \pm 3.499(169.7)\frac{1}{\sqrt{n_{ij}}}$$

For the samples of size 1, this is

$$\bar{y}_{ij} \pm 593.9 \tag{8.1}$$

while for the samples of size 2, appropriate endpoints are

$$\bar{y}_{ij} \pm 419.9 \tag{8.2}$$

Example 1 *(continued)*

Figure 8.3 is an interaction plot (like Figure 4.22) enhanced with error bars made using limits (8.1) and (8.2). Notice, by the way, that the Bonferroni inequality puts the simultaneous confidence associated with all nine of the indicated intervals at a minimum of 91% ($.91 = 1 - 9(1 - .99)$).

The important message carried by Figure 8.3, not already present in Figure 4.22, is the relatively large imprecision associated with the sample means as estimates of long-run mean strengths. And that imprecision has implications regarding the statistical detectability of factorial effects. For example, by moving near the extremes on some error bars in Figure 8.3, one might find nine means within the indicated intervals such that their connecting line segments would exhibit parallelism. That is, the plot already suggests that the empirical interactions between Wood Type and Joint Type seen in these data may not be large enough to distinguish from background noise. Or if they are detectable, they may be only barely so.

The issues of whether the empirical differences between woods and between joint types are distinguishable from experimental variation are perhaps somewhat easier to call. There is consistency in the patterns "Walnut is stronger than oak is stronger than pine" and "Beveled is stronger than lap is stronger than butt." This, combined with differences at least approaching the size of indicated imprecisions,

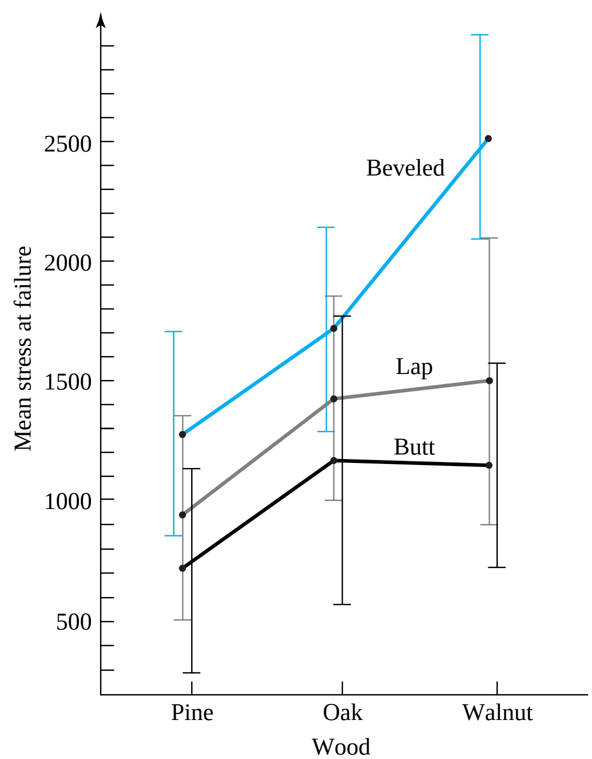

Figure 8.3 Interaction plot of mean joint strength with error bars based on individual 99% confidence intervals

suggests that firm statements about the main effects of Wood Type and Joint Type are likely possible.

The kind of analysis made thus far on the joint strength data is extremely important and illuminating. Our discussion will proceed to more complicated statistical methods for such problems. But these often amount primarily to a further refinement and quantification of the two-way factorial story already told graphically by a plot like Figure 8.3.

8.1.2 Two-Way Factorial Notation and Definitions of Effects

In order to discuss inference in two-way factorial studies, it is useful to modify the generic multisample notation used in Chapter 7. Consider combinations of factor A having I levels and factor B having J levels and use the triple subscript notation:

Two-way (triple subscript) notation

y_{ijk} = the kth observation in the sample from the ith level of A and jth plevel of B

Then for $I \cdot J$ different samples corresponding to the possible combinations of a level of A with a level of B, let

n_{ij} = the number of observations in the sample from the ith level of A and jth level of B

Use the notations $\bar{y}_{ij}$, $\bar{y}_{i.}$, and $\bar{y}_{.j}$ introduced in Section 4.3, and in the obvious way (actually already used in Example 1), let

s_{ij} = the sample standard deviation of the n_{ij} observations in the sample from the ith level of A and the jth level of B

This amounts to adding another subscript to the notation introduced in Chapter 7 in order to acknowledge the two-way structure. In Chapter 7, it was most natural to think of r samples as numbered $i = 1$ to r and laid out in a single row. Here it is appropriate to think of $r = I \cdot J$ samples laid out in the cells of a two-way table like Table 8.1 and named by their row number i and column number j.

In addition to using this notation for empirical quantities, it is also useful to modify the notation used in Chapter 7 for model parameters. That is, let

μ_{ij} = the underlying mean response corresponding to the ith level of A and jth level of B

The model assumptions that the $I \cdot J$ samples are roughly describable as independent samples from normal distributions with a common variance σ^2 can be written as

Two-way model statement

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk} \tag{8.3}$$

where the quantities $\epsilon_{111}, \ldots, \epsilon_{11n_{11}}, \epsilon_{121}, \ldots, \epsilon_{12n_{12}}, \ldots, \epsilon_{IJ1}, \ldots, \epsilon_{IJn_{IJ}}$ are independent normal $(0, \sigma^2)$ random variables. Equation (8.3) is sometimes called the **two-way (normal) model equation**. It is nothing but a rewrite of the basic one-way model equation of Chapter 7 in a notation that recognizes the special organization of $r = I \cdot J$ samples into rows and columns, as in Table 8.1.

The descriptive analysis of two-way factorials in Section 4.3 relied on computing row averages $\bar{y}_{i.}$ and column averages $\bar{y}_{.j}$ from the sample means $\bar{y}_{ij}$. These were then used to define fitted factorial effects. Analogous operations performed on the underlying or theoretical means μ_{ij} lead to appropriate definitions for theoretical factorial effects. That is, let

$$\begin{aligned}
\mu_{i.} &= \frac{1}{J}\sum_{j=1}^{J}\mu_{ij} \\
&= \text{the average underlying mean when factor A is at level } i \\
\mu_{.j} &= \frac{1}{I}\sum_{i=1}^{I}\mu_{ij} \\
&= \text{the average underlying mean when factor B is at level } j \\
\mu_{..} &= \frac{1}{IJ}\sum_{i,j}\mu_{ij} \\
&= \text{the grand average underlying mean}
\end{aligned}$$

Figure 8.4 shows these as row, column, and grand averages of the μ_{ij}. (This is the theoretical counterpart of Figure 4.21.)

Then, following the pattern established in Definitions 5 and 6 in Chapter 4 for sample quantities, there are the following two definitions for theoretical quantities.

Definition 1

In a two-way complete factorial study with factors A and B, the **main effect of factor A at its *i*th level** is

$$\alpha_i = \mu_{i.} - \mu_{..}$$

Similarly, the **main effect of factor B at its *j*th level** is

$$\beta_j = \mu_{.j} - \mu_{..}$$

These main effects are measures of how (theoretical) mean responses change from row to row or from column to column in Figure 8.4. The fitted main effects of Section 4.3 can be thought of as empirical approximations to them. It is a

		Factor B				
		Level 1	Level 2		Level J	
	Level 1	μ_{11}	μ_{11}	$\cdots$	μ_{1J}	$\mu_{1.}$
	Level 2	μ_{21}	μ_{22}	$\cdots$	μ_{2J}	$\mu_{2.}$
Factor A		$\vdots$	$\vdots$		$\vdots$	$\vdots$
	Level I	μ_{I1}	μ_{I2}	$\cdots$	μ_{IJ}	$\mu_{I.}$
		$\mu_{.1}$	$\mu_{.2}$	$\cdots$	$\mu_{.J}$	$\mu_{..}$

Figure 8.4 Underlying cell mean responses and their row, column, and grand averages

consequence of the form of Definition 1 that (like their empirical counterparts) main effects of a given factor sum to 0 over levels of that factor. That is, simple algebra shows that

$$\sum_{i=1}^{I} \alpha_i = 0 \quad \text{and} \quad \sum_{j=1}^{J} \beta_j = 0$$

Next is a definition of theoretical interactions.

Definition 2

In a two-way complete factorial study with factors A and B, the **interaction of factor A at its *i*th level and factor B at its *j*th level** is

$$\alpha\beta_{ij} = \mu_{ij} - (\mu_{..} + \alpha_i + \beta_j)$$

The interactions in a two-way set of underlying means μ_{ij} measure lack of parallelism on an interaction plot of the parameters μ_{ij}. They measure how much pattern there is in the theoretical means μ_{ij} that is not explainable in terms of the factors A and B acting individually. The fitted interactions of Section 4.3 are empirical approximations of these theoretical quantities. Small fitted interactions ab_{ij} indicate small underlying interactions $\alpha\beta_{ij}$ and thus make it justifiable to think of the two factors A and B as operating separately on the response variable.

Definition 2 has several simple algebraic consequences that are occasionally useful to know. One is that (like fitted interactions) interactions $\alpha\beta_{ij}$ sum to 0 over levels of either factor. That is, as defined,

$$\sum_{i=1}^{I} \alpha\beta_{ij} = \sum_{j=1}^{J} \alpha\beta_{ij} = 0$$

Another simple consequence is that upon adding $(\mu_{..} + \alpha_i + \beta_j)$ to both sides of the equation defining $\alpha\beta_{ij}$, one obtains a decomposition of each μ_{ij} into a grand mean plus an A main effect plus a B main effect plus an AB interaction:

$$\mu_{ij} = \mu_{..} + \alpha_i + \beta_j + \alpha\beta_{ij} \tag{8.4}$$

The identity (8.4) is sometimes combined with the two-way model equation (8.3) to obtain the equivalent model equation

A second statement of the two-way model

$$y_{ijk} = \mu_{..} + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk} \tag{8.5}$$

Here the factorial effects appear explicitly as going into the makeup of the observations. Although there are circumstances where representation (8.5) is essential, in most cases it is best to think of the two-way model assumptions in form (8.3) and just remember that the α_i, β_j, and $\alpha\beta_{ij}$ are simple functions of the $I \cdot J$ means μ_{ij}.

8.1.3 Individual Confidence Intervals for Factorial Effects

The primary new wrinkles in two-way factorial inference are

1. the drawing of inferences concerning the interactions and main effects, with
2. the possibility of finding A, B, or A and B "main effects only" models adequate to describe responses, and subsequently using such simplified descriptions in making predictions about system behavior.

Factorial effects are L's, fitted effects are corresponding $\hat{L}$'s

The basis of inference for the α_i, β_j, and $\alpha\beta_{ij}$ is that they are linear combinations of the means μ_{ij}. (That is, for properly chosen "c's," the factorial effects are "L's" from Section 7.2.) And the fitted effects defined in Chapter 4's Definitions 5 and 6 are the corresponding linear combinations of the sample means $\bar{y}_{ij}$. (That is, the fitted factorial effects are the corresponding "$\hat{L}$'s.")

Example 1 *(continued)*

To illustrate that the effects defined in Definitions 1 and 2 are linear combinations of the underlying means μ_{ij}, consider α_1 and $\alpha\beta_{23}$ in the wood joint strength study. First,

$$\alpha_1 = \mu_{1.} - \mu_{..}$$

$$= \frac{1}{3}(\mu_{11} + \mu_{12} + \mu_{13}) - \frac{1}{9}(\mu_{11} + \mu_{12} + \cdots + \mu_{32} + \mu_{33})$$

$$= \frac{2}{9}\mu_{11} + \frac{2}{9}\mu_{12} + \frac{2}{9}\mu_{13} - \frac{1}{9}\mu_{21} - \frac{1}{9}\mu_{22} - \frac{1}{9}\mu_{23} - \frac{1}{9}\mu_{31} - \frac{1}{9}\mu_{32} - \frac{1}{9}\mu_{33}$$

and a_1 is the corresponding linear combination of the $\bar{y}_{ij}$. Similarly,

$$\begin{aligned}
\alpha\beta_{23} &= \mu_{23} - (\mu_{..} + \alpha_2 + \beta_3) \\
&= \mu_{23} - \big(\mu_{..} + (\mu_{2.} - \mu_{..}) + (\mu_{.3} - \mu_{..})\big) \\
&= \mu_{23} - \mu_{2.} - \mu_{.3} + \mu_{..} \\
&= \mu_{23} - \frac{1}{3}(\mu_{21} + \mu_{22} + \mu_{23}) - \frac{1}{3}(\mu_{13} + \mu_{23} + \mu_{33}) \\
&\quad + \frac{1}{9}(\mu_{11} + \mu_{12} + \cdots + \mu_{33}) \\
&= \frac{4}{9}\mu_{23} - \frac{2}{9}\mu_{21} - \frac{2}{9}\mu_{22} - \frac{2}{9}\mu_{13} - \frac{2}{9}\mu_{33} + \frac{1}{9}\mu_{11} + \frac{1}{9}\mu_{12} \\
&\quad + \frac{1}{9}\mu_{31} + \frac{1}{9}\mu_{32}
\end{aligned}$$

and ab_{23} is the corresponding linear combination of the $\bar{y}_{ij}$.

Once one realizes that the factorial effects are simple linear combinations of the μ_{ij}, it is a small step to recognize that formula (7.20) of Section 7.2 can be applied to make confidence intervals for them. For example, the question of whether the lack of parallelism evident in Figure 8.3 is large enough to be statistically detectable can be approached by looking at confidence intervals for the $\alpha\beta_{ij}$. And quantitative comparisons between joint types can be based on confidence intervals for differences between the A main effects, $\alpha_i - \alpha_{i'} = \mu_{i.} - \mu_{i'.}$. And quantitative comparisons between woods can be based on differences between the B main effects, $\beta_j - \beta_{j'} = \mu_{.j} - \mu_{.j'}$.

The only obstacle to applying formula (7.20) of Section 7.2 to do inference for factorial effects is determining how the "$\sum c_i^2/n_i$" term appearing in the formula should look for quantities of interest. In the preceding example, a number of rather odd-looking coefficients c_{ij} appeared when writing out expressions for α_1 and $\alpha\beta_{23}$ in terms of the basic means μ_{ij}. However, it is possible to discover and write down general formulas for the sum $\sum c_{ij}^2/n_{ij}$ for some important functions of the factorial effects. Table 8.3 gives the relatively simple formulas for the balanced data case where all $n_{ij} = m$. The less pleasant general versions of the formulas are given in Table 8.4.

Table 8.3
Balanced Data Formulas to Use with Limits (8.6)

L	$\hat{L}$	$\sum_{i,j} \frac{c_{ij}^2}{n_{ij}}$
$\alpha\beta_{ij}$	ab_{ij}	$\frac{(I-1)(J-1)}{mIJ}$
α_i	a_i	$\frac{I-1}{mIJ}$
$\alpha_i - \alpha_{i'}$	$a_i - a_{i'}$	$\frac{2}{mJ}$
β_j	b_j	$\frac{J-1}{mIJ}$
$\beta_j - \beta_{j'}$	$b_j - b_{j'}$	$\frac{2}{mI}$

Armed with Tables 8.3 and 8.4, the form of individual confidence intervals for any of the quantities $L = \alpha\beta_{ij}, \alpha_i, \beta_j, \alpha_i - \alpha_{i'}$, or $\beta_j - \beta_{j'}$ is obvious. In the formula for confidence interval endpoints

Confidence limits for a linear combination of two-way factorial means

$$\boxed{\hat{L} \pm t s_{\text{P}} \sqrt{\sum_{i,j} \frac{c_{ij}^2}{n_{ij}}}} \qquad \textbf{(8.6)}$$

1. s_{P} is computed by pooling the $I \cdot J$ sample variances in the usual way (arriving at an estimate with $n - r = n - IJ$ associated degrees of freedom),
2. the fitted effects from Section 4.3 are used to find $\hat{L}$,
3. an appropriate formula from Table 8.3 or 8.4 is chosen to give the quantity under the radical, and
4. t from Table B.4 is chosen according to a desired confidence and degrees of freedom $\nu = n - IJ$.

Table 8.4
General Formulas to use with Limits (8.6)

L	$\hat{L}$	$\sum_{i,j} \frac{c_{ij}^2}{n_{ij}}$
$\alpha\beta_{ij}$	ab_{ij}	$\left(\frac{1}{IJ}\right)^2 \left(\frac{(I-1)^2(J-1)^2}{n_{ij}} + (I-1)^2 \sum_{j' \neq j} \frac{1}{n_{ij'}} + (J-1)^2 \sum_{i' \neq i} \frac{1}{n_{i'j}} + \sum_{i' \neq i, j' \neq j} \frac{1}{n_{i'j'}}\right)$
α_i	a_i	$\left(\frac{1}{IJ}\right)^2 \left((I-1)^2 \sum_j \frac{1}{n_{ij}} + \sum_{i' \neq i, j} \frac{1}{n_{i'j}}\right)$
$\alpha_i - \alpha_{i'}$	$a_i - a_{i'}$	$\frac{1}{J^2} \left(\sum_j \frac{1}{n_{ij}} + \sum_j \frac{1}{n_{i'j}}\right)$
β_j	b_j	$\left(\frac{1}{IJ}\right)^2 \left((J-1)^2 \sum_i \frac{1}{n_{ij}} + \sum_{i, j' \neq j} \frac{1}{n_{ij'}}\right)$
$\beta_j - \beta_{j'}$	$b_j - b_{j'}$	$\frac{1}{I^2} \left(\sum_i \frac{1}{n_{ij}} + \sum_i \frac{1}{n_{ij'}}\right)$

Example 2

A Synthetic 3 × 3 Balanced Data Example

To illustrate how easy it is to do inference for factorial effects when complete two-way factorial data are balanced, consider a 3×3 factorial with $m = 2$ observations per cell. (This is the way that the wood joint strength study of Example 1 was planned. It was only circumstances beyond the control of the students that conspired to produce the unbalanced data of Table 8.1 through the loss of two specimens.) In this hypothetical situation, s_P has degrees of freedom $\nu = n - IJ = mIJ - IJ = 2 \cdot 3 \cdot 3 - 3 \cdot 3 = 9$. Definitions 5 and 6 in Chapter 4 show how to compute fitted main effects a_i and b_j and fitted interactions ab_{ij}.

To, for example, make a confidence interval for an interaction $\alpha\beta_{ij}$, consult the first row of Table 8.3 and compute

$$\sum_{i,j} \frac{c_{ij}^2}{n_{ij}} = \frac{(I-1)(J-1)}{mIJ} = \frac{2 \cdot 2}{2 \cdot 3 \cdot 3} = \frac{2}{9} \quad \text{and} \quad \sqrt{\sum_{i,j} \frac{c_{ij}^2}{n_{ij}}} = .4714$$

Example 2
(*continued*)

Then choosing t (as a quantile of the t_9 distribution) to produce the desired confidence level, equation (8.6) shows appropriate confidence limits to be

$$ab_{ij} \pm ts_{\text{P}}(.4714)$$

As a second example of this methodology, consider the estimation of the difference in two factor B main effects, $L = \beta_j - \beta_{j'} = \mu_{.j} - \mu_{.j'}$. Consulting the last row of Table 8.3,

$$\sum_{i,j} \frac{c_{ij}^2}{n_{ij}} = \frac{2}{mI} = \frac{2}{2 \cdot 3} = \frac{1}{3} \quad \text{and} \quad \sqrt{\sum_{i,j} \frac{c_{ij}^2}{n_{ij}}} = .5774$$

Then again choosing t to produce the desired confidence level, equation (8.6) shows appropriate confidence limits to be

$$b_j - b_{j'} \pm ts_{\text{P}}(.5774)$$

that is,

$$\bar{y}_{.j} - \bar{y}_{.j'} \pm ts_{\text{P}}(.5774)$$

Example 1
(*continued*)

Consider making formal inferences for the factorial effects in the (unbalanced) wood joint strength. Suppose that inferences are to be phrased in terms of two-sided 99% individual confidence intervals and begin by considering the interactions $\alpha\beta_{ij}$.

Despite the students' best efforts to the contrary, the sample sizes in Table 8.1 are not all the same. So one is forced to use formulas in Table 8.4 instead of the simpler ones in Table 8.3. Table 8.5 collects the sums of reciprocal sample sizes appearing in the first row of Table 8.4 for each of the nine combinations of $i = 1, 2, 3$ and $j = 1, 2, 3$.

For example, for the combination $i = 1$ and $j = 1$,

$$\frac{1}{n_{11}} = \frac{1}{2} = .5$$

$$\frac{1}{n_{12}} + \frac{1}{n_{13}} = \frac{1}{1} + \frac{1}{2} = 1.5$$

$$\frac{1}{n_{21}} + \frac{1}{n_{31}} = \frac{1}{2} + \frac{1}{2} = 1.0$$

$$\frac{1}{n_{22}} + \frac{1}{n_{23}} + \frac{1}{n_{32}} + \frac{1}{n_{33}} = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{1} = 2.5$$

Table 8.5
Sums of Reciprocal Sample Sizes Needed in Making Confidence Intervals for Joint/Wood Interactions

i	j	$\frac{1}{n_{ij}}$	$\sum_{j'\neq j}\frac{1}{n_{ij'}}$	$\sum_{i'\neq i}\frac{1}{n_{i'j}}$	$\sum_{i'\neq i,\, j'\neq j}\frac{1}{n_{i'j'}}$
1	1	.5	1.5	1.0	2.5
1	2	1.0	1.0	1.0	2.5
1	3	.5	1.5	1.5	2.0
2	1	.5	1.0	1.0	3.0
2	2	.5	1.0	1.5	2.5
2	3	.5	1.0	1.5	2.5
3	1	.5	1.5	1.0	2.5
3	2	.5	1.5	1.5	2.0
3	3	1.0	1.0	1.0	2.5

The entries in Table 8.5 lead to values for $\sum c_{ij}^2/n_{ij}$ via the formula on the first row of Table 8.4. Then, since (from before) $s_{\mathrm{P}} = 169.7$ psi with 7 associated degrees of freedom, and since the .995 quantile of the t_7 distribution is 3.499, it is possible to calculate the plus-or-minus part of formula (8.6) in order to get two-sided 99% confidence intervals for the $\alpha\beta_{ij}$. In addition, remember that all nine fitted interactions were calculated in Section 4.3 and collected in Table 4.14 (page 170). Table 8.6 gives the $\sqrt{\sum c_{ij}^2/n_{ij}}$ values, the fitted interactions ab_{ij}, and the plus-or-minus part of two-sided 99% individual confidence intervals for the interactions $\alpha\beta_{ij}$.

To illustrate the calculations summarized in the third column of Table 8.6, consider the combination with $i = 1$ (butt joints) and $j = 1$ (pine wood). Since $I = 3$ and $J = 3$, the first row of Table 8.4 shows that for $L = \alpha\beta_{11}$

$$\sum \frac{c_{ij}^2}{n_{ij}} = \left(\frac{1}{3\cdot 3}\right)^2 \left(\frac{2^2\cdot 2^2}{2} + 2^2(1.5) + 2^2(1.0) + 2.5\right) = .2531$$

from which

$$\sqrt{\sum \frac{c_{ij}^2}{n_{ij}}} = \sqrt{.2531} = .5031$$

Consider the practical implications of the calculations summarized in Table 8.6. All but one of the intervals centered at an ab_{ij} with a half width given in the last column of the table would cover 0. Only for $i = 2$ (beveled joints) and $j = 3$ (walnut wood) is the magnitude of the fitted interaction big enough to put its

Example 1
(*continued*)

Table 8.6
99% Individual Two-Sided Confidence Intervals for Joint Type/Wood Type Interactions

i	j	$\sqrt{\sum \frac{c_{ij}^2}{n_{ij}}}$	ab_{ij} (psi)	$ts_{\mathrm{P}}\sqrt{\sum \frac{c_{ij}^2}{n_{ij}}}$ (psi)
1	1	.5031	105.83	298.7
1	2	.5720	95.67	339.6
1	3	.5212	−201.5	309.5
2	1	.4843	−155.67	287.6
2	2	.5031	−177.33	298.7
2	3	.5031	333.0	298.7
3	1	.5031	49.83	298.7
3	2	.5212	81.67	309.5
3	3	.5720	−131.5	339.6

associated confidence interval entirely to one side of 0. That is, most of the lack of parallelism seen in Figure 8.3 is potentially attributable to experimental variation. But that associated with beveled joints and walnut wood can be differentiated from background noise. This suggests that if mean joint strength differences on the order of 333 ± 299 psi are of engineering importance, it is not adequate to think of the factors Joint Type and Wood Type as operating separately on joint strength across all three levels of each factor. On the other hand, if attention was restricted to either butt and lap joints or to pine and oak woods, a "no detectable interactions" description of joint strength would perhaps be tenable.

To illustrate the use of formula (8.6) in making inferences about main effects on joint strength, consider comparing joint strengths for pine and oak woods. The rather extended analysis of interactions here and the character of Figure 8.3 suggest that the strength profiles of pine and oak across the three joint types are comparable. If this is so, estimation of $\beta_1 - \beta_2 = \mu_{.1} - \mu_{.2}$ amounts to more than the estimation of the difference in average (across joint types) mean strengths of pine and oak joints (pine minus oak). $\beta_1 - \beta_2$ is also the difference in mean strengths of pine and oak joints *for any of the three joint types individually*. It is thus a quantity of real interest.

Once again, since the data in Table 8.1 are not balanced, it is necessary to use the more complicated formula in Table 8.4 rather than the formula in Table 8.3 in making a confidence interval for $\beta_1 - \beta_2$. For $L = \beta_1 - \beta_2$, the last row of Table 8.4 gives

$$\sum_{i,j} \frac{c_{ij}^2}{n_{ij}} = \frac{1}{3^2}\left[\left(\frac{1}{2}+\frac{1}{2}+\frac{1}{2}\right)+\left(\frac{1}{1}+\frac{1}{2}+\frac{1}{2}\right)\right] = .3889$$

So, since from the fitted effects in Section 4.3

$$b_1 = -402.5 \text{ psi} \quad \text{and} \quad b_2 = 64.17 \text{ psi}$$

formula (8.6) shows that endpoints of a two-sided 99% confidence interval for $L = \beta_1 - \beta_2$ are

$$(-402.5 - 64.17) \pm 3.499(169.7)\sqrt{.3889}$$

that is,

$$-466.67 \pm 370.29$$

that is,

$$-836.96 \text{ psi} \quad \text{and} \quad -96.38 \text{ psi}$$

This analysis establishes that the oak joints are on average from 96 psi to 837 psi stronger than comparable pine joints. This may seem a rather weak conclusion, given the apparent strong increase in sample mean strengths as one moves from pine to oak in Figure 8.3. But it is as strong a statement as is justified in the light of the large confidence requirement (99%) and the substantial imprecision in the students' data (related to the small sample sizes and a large pooled standard deviation, $s_{\text{P}} = 169.7$ psi). If ± 370 psi precision for comparing pine and oak joint strength is not adequate for engineering purposes and large confidence is still desired, these calculations point to the need for more data in order to sharpen that comparison.

The computational unpleasantness of the previous discussion results from the fact that the data of Kotlers, MacFarland, and Tomlinson are unbalanced. Example 2 illustrated that with balanced data, "by hand" calculation is simple. Most statistical packages have routines that will eliminate the need for a user to grind through the most tedious of the computations just illustrated. Printout 1 is a MINITAB General Linear Model output for the wood strength study (which is part of Printout 6 of Chapter 4). The "Coef" values in that printout are (again) the fitted effects of Definitions 5 and 6 in Chapter 4. The "StDev" values are the quantities

$$s_{\text{P}}\sqrt{\sum_{i,j} \frac{c_{ij}^2}{n_{ij}}}$$

from formula (8.6) needed to make confidence limits for main effects and interactions. (The MINITAB printout lists this information for only $(I-1)$ factor A main effects, $(J-1)$ factor B main effects, and $(I-1)(J-1)$ A$\times$B interactions. Renaming levels of the factors to change their alphabetical order will produce a different printout giving this information for the remaining main effects and interactions.)

Printout 1 Estimated Standard Deviations of Joint Strength Fitted Effects (*Example 1*)

```
General Linear Model

Factor     Type Levels Values
joint     fixed      3 beveled butt    lap
wood      fixed      3     oak pine walnut

Analysis of Variance for strength, using Adjusted SS for Tests

Source        DF     Seq SS     Adj SS     Adj MS       F      P
joint          2    2153879    1881650     940825   32.67  0.000
wood           2    1641095    1481377     740689   25.72  0.001
joint*wood     4     468408     468408     117102    4.07  0.052
Error          7     201614     201614      28802
Total         15    4464996

Term                Coef     StDev        T      P
Constant         1375.67     44.22    31.11  0.000
  joint
beveled           460.00     59.63     7.71  0.000
butt             -366.50     63.95    -5.73  0.001
  wood
oak                64.17     63.95     1.00  0.349
pine             -402.50     59.63    -6.75  0.000
  joint*  wood
beveled oak      -177.33     85.38    -2.08  0.076
beveled pine     -155.67     82.20    -1.89  0.100
butt    oak        95.67     97.07     0.99  0.357
butt    pine      105.83     85.38     1.24  0.255
```

8.1.4 Tukey's Method for Comparing Main Effects (*Optional*)

Formula (8.6) is meant to guarantee *individual* confidence levels for intervals made using it. When interactions in a two-way factorial study are negligible, questions of practical engineering importance can usually be phrased in terms of comparing the various A or B main effects. It is then useful to have a method designed specifically to produce a *simultaneous* confidence level for the comparison of all pairs of A or B main effects. Tukey's method (discussed in Section 7.3) can be modified to produce simultaneous confidence intervals for all differences in α_i's or in β_j's. That is, two-sided simultaneous confidence intervals for all possible differences in A main effects $\alpha_i - \alpha_{i'} = \mu_{i.} - \mu_{i'.}$ can be made using endpoints

Tukey simultaneous confidence limits for all differences in A main effects

$$\bar{y}_{i.} - \bar{y}_{i'.} \pm \frac{q^*}{\sqrt{2}} s_{\mathrm{P}} \frac{1}{J} \sqrt{\sum_j \frac{1}{n_{ij}} + \sum_j \frac{1}{n_{i'j}}} \qquad \textbf{(8.7)}$$

where q^* is taken from Tables B.9 using $\nu = n - IJ$ degrees of freedom, number of means to be compared I, and the .95 or .99 quantile figure (depending whether 95% or 99% simultaneous confidence is desired). Expression (8.7) amounts to the specialization of formula (8.6) to $L = \alpha_i - \alpha_{i'}$ with t replaced by $q^*/\sqrt{2}$. When all $n_{ij} = m$, formula (8.7) simplifies to

Balanced data Tukey simultaneous confidence limits for all differences in A main effects

$$\boxed{\bar{y}_{i.} - \bar{y}_{i'.} \pm \frac{q^* s_{\text{P}}}{\sqrt{Jm}}} \tag{8.8}$$

Corresponding to formulas (8.7) and (8.8) are formulas for simultaneous two-sided confidence limits for all possible differences in B main effects $\beta_j - \beta_{j'} = \mu_{.j} - \mu_{.j'}$—namely,

Tukey simultaneous confidence limits for all differences in B main effects

$$\boxed{\bar{y}_{.j} - \bar{y}_{.j'} \pm \frac{q^*}{\sqrt{2}} s_{\text{P}} \frac{1}{I} \sqrt{\sum_i \frac{1}{n_{ij}} + \sum_i \frac{1}{n_{ij'}}}} \tag{8.9}$$

and

Balanced data Tukey simultaneous confidence limits for all differences in B main effects

$$\boxed{\bar{y}_{.j} - \bar{y}_{.j'} \pm \frac{q^* s_{\text{P}}}{\sqrt{Im}}} \tag{8.10}$$

where q^* is taken from Tables B.9 using $\nu = n - IJ$ degrees of freedom and number of means to be compared J.

Example 3

A 3 × 2 Factorial Study of Ultimate Tensile Strength for Drilled Aluminum Strips

Clubb and Goedken studied the effects on tensile strength of holes drilled in 6 in.-by-2 in. 2024–T3 aluminum strips .0525 in. thick. A hole of diameter .149 in., .185 in., or .221 in. was centered either .5 in. or 1.0 in. from the edge (and 3.0 in. from each end) of 18 strips. Ultimate axial stress was then measured for each on an MTS machine. $m = 3$ tests were made for each of the 3×2 combinations of hole size and placement. Mean tensile strengths (in pounds) obtained in the study are given in Table 8.7. Some plotting with the original data (not given here) shows that (except for some hint that hole size 3 strengths were less variable than the others) the one-way normal model assumptions provide a plausible description of tensile strength. We will proceed to use the assumptions (8.3) in what follows.

Example 3 *(continued)*

Table 8.7
Sample Means for 3 × 2 Size/Placement Combinations

		B *Placement*	
		1 (.5 in. from Edge)	2 (1.0 in. from Edge)
A *Size*	1 (.149 in.)	$\bar{y}_{11} = 5635.3$ lb	$\bar{y}_{12} = 5730.3$ lb
	2 (.185 in.)	$\bar{y}_{21} = 5501.0$ lb	$\bar{y}_{22} = 5638.0$ lb
	3 (.221 in.)	$\bar{y}_{31} = 5456.3$ lb	$\bar{y}_{32} = 5602.7$ lb

Pooling the $3 \cdot 2 = 6$ sample variances in the usual way produced

$$s_{\text{P}} = 106.7 \text{ lb}$$

with $\nu = mIJ - IJ = 3 \cdot 3 \cdot 2 - 3 \cdot 2 = 12$ associated degrees of freedom. Then consider summarizing the experimental results graphically. Notice that the P-R method for making simultaneous two-sided 95% confidence intervals for $r = 6$ means based on $\nu = 12$ degrees of freedom is (from formula (7.28) of Section 7.3) to use endpoints

$$\bar{y}_{ij} \pm 3.095 \frac{106.7}{\sqrt{3}}$$

for estimating each μ_{ij}. ($k_2^* = 3.095$ was obtained from Table B.8A.) This is approximately

$$\bar{y}_{ij} \pm 191$$

Figure 8.5 is an interaction plot of the $3 \times 2 = 6$ sample mean tensile strengths enhanced with ± 191 lb error bars.

The lack of parallelism in Figure 8.5 is fairly small, both compared to the absolute size of the strengths being measured and also relative to the kind of uncertainty about the individual mean strengths indicated by the error bars. Letting factor A be size and factor B be placement, it is straightforward to use the methods of Section 4.3 to calculate

$$\begin{aligned}
a_1 &= 88.9 & b_1 &= -63.1 \\
a_2 &= -24.4 & b_2 &= 63.1 \\
a_3 &= -64.4 & & \\
ab_{11} &= 15.6 & ab_{12} &= -15.6 \\
ab_{21} &= -5.4 & ab_{22} &= 5.4 \\
ab_{31} &= -10.1 & ab_{32} &= 10.1
\end{aligned}$$

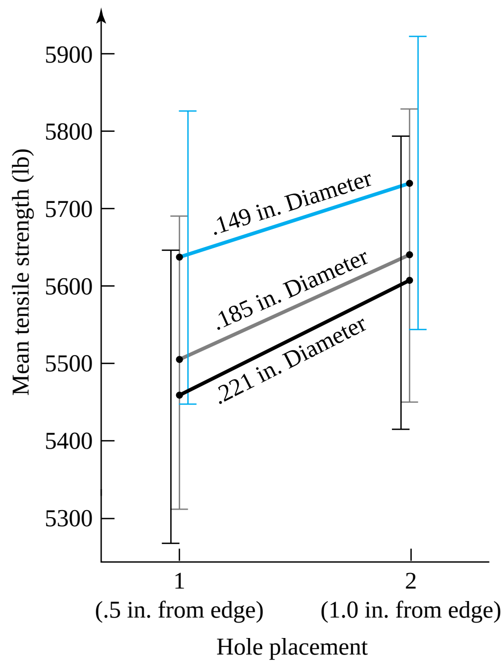

Figure 8.5 Interaction plot of aluminum strip sample means, enhanced with error bars based on 95% simultaneous confidence intervals

Then, since the data are balanced, one may use the formulas of Table 8.3 together with formula (8.6). So individual confidence intervals for the interactions $\alpha\beta_{ij}$ are of the form

$$ab_{ij} \pm t(106.7)\sqrt{\frac{(3-1)(2-1)}{3 \cdot 3 \cdot 2}}$$

that is,

$$ab_{ij} \pm t(35.6)$$

Clearly, for any sensible confidence level (producing t of at least 1), such intervals all cover 0. This confirms the lack of statistical detectability of the interactions already represented in Figure 8.5.

It thus seems sensible to proceed to consideration of the main effects in this tensile strength study. To illustrate the application of Tukey's method to factorial main effects, consider first simultaneous 95% two-sided confidence intervals for the three differences $\alpha_1 - \alpha_2$, $\alpha_1 - \alpha_3$, and $\alpha_2 - \alpha_3$. Applying formula (8.8)

Example 3
(continued)

with $\nu = 12$ degrees of freedom and $I = 3$ means to be compared, Table B.9A indicates that intervals with endpoints

$$\bar{y}_{i.} - \bar{y}_{i'.} \pm \frac{(3.77)(106.7)}{\sqrt{2 \cdot 3}}$$

that is,

$$\bar{y}_{i.} - \bar{y}_{i'.} \pm 164 \text{ lb}$$

are in order. No difference between the a_i's exceeds 164 lb. That is, if simultaneous 95% confidence is desired in the comparison of the hole size main effects, one must judge the students' data to be interesting—perhaps even suggestive of a decrease in strength with increased diameter—but nevertheless statistically inconclusive. To really pin down the impact of hole size on tensile strength, larger samples are needed.

To see that the Clubb and Goedken data do tell at least some story in a reasonably conclusive manner, finally consider the use of the last row of Table 8.3 with formula (8.6) to make a two-sided 95% confidence interval for $\beta_2 - \beta_1$, the difference in mean strengths for strips with centered holes as compared to ones with holes .5 in. from the strip edge. The desired interval has endpoints

$$b_2 - b_1 \pm t s_{\text{P}} \sqrt{\frac{2}{mI}}$$

that is,

$$63.1 - (-63.1) \pm 2.179(106.7)\sqrt{\frac{2}{3(3)}}$$

that is,

$$126.2 \pm 109.6$$

that is,

$$16.6 \text{ lb} \quad \text{and} \quad 235.8 \text{ lb}$$

Thus, although the students' data don't provide much precision, they *are* adequate to establish clearly the existence of some decrease in tensile strength as a hole is moved from the center of the strip towards its edge.

Formulas (8.7) through (8.10) are, mathematically speaking, perfectly valid providing only that the basic "equal variances, underlying normal distributions" model is a reasonable description of an engineering application. (Under the basic model (8.3), formulas (8.7) and (8.9) provide an actual simultaneous confidence at least as big as the nominal one, and when all $n_{ij} = m$, formulas (8.8) and (8.10) provide actual simultaneous confidence equal to the nominal one.) But in practical *terms*, the inferences they provide (and indeed the ones provided by formula (8.6) for individual differences in main effects) are not of much interest unless the interactions $\alpha\beta_{ij}$ have been judged to be negligible.

Nonnegligible interactions constitute a warning that the patterns of change in mean response, as one moves between levels of one factor, (say, B) are different for various levels of the other factor (say, A). That is, the pattern in the μ_{ij} is not a simple one generally describable in terms of the two factors acting separately. Rather than trying to understand the pattern in terms of main effects, something else must be done.

What if interactions are not negligible?

As discussed in Section 4.4, sometimes a transformation can produce a response variable describable in terms of main effects only. At other times, restriction of attention to part of a factorial produces a study (of reduced scope) where it makes sense to think in terms of main effects. (In Example 1, consideration of only butt and lap joints gives an arena where "negligible interactions" may be a sensible description of joint strength.) Or it may be most natural to mentally separate an $I \times J$ factorial into $I(J)$ different $J(I)$ level studies on the effects of factor B(A) at different levels of A(B). (The 3×3 wood joint strength study in Example 1 might be thought of as three different studies, one for each joint type, of the effects of wood type on strength.) Or if none of these approaches to analyzing two-way factorial data with important interactions is attractive, it is always possible to ignore the two-way structure completely and treat the $I \cdot J$ samples as arising from simply $r = I \cdot J$ unstructured different conditions.

Section 1 Exercises

1. The accompanying table shows part of the data of Dimond and Dix, referred to in Examples 6 (Chapter 1) and 9 (Chapter 3). The values are the shear strengths (in lb) for $m = 3$ tests on joints of various combinations of Wood Type and Glue Type.

Wood	Glue	Joint Shear Strengths
pine	white	130, 127, 138
pine	carpenter's	195, 194, 189
pine	cascamite	195, 202, 207
fir	white	95, 119, 62
fir	carpenter's	137, 157, 145
fir	cascamite	152, 163, 155

(a) Make an interaction plot of the six combination means and enhance it with error bars derived using the P-R method of making 95% simultaneous two-sided confidence intervals. (Plot mean strength versus glue type.)

(b) Compute the fitted main effects and interactions from the six combination sample means. Use these to make individual 95% confidence intervals for all of the main effects and interactions in this 2×3 factorial study. What do these indicate about the detectability of the various effects?

(c) Use Tukey's method for simultaneous comparison of main effects and give simultaneous 95% confidence intervals for all differences in Glue Type main effects.

2. B. Choi conducted a replicated full factorial study of the stopping properties of various types of bicycle tires on various riding surfaces. Three different Types of Tires were used on the bike, and three different Pavement Conditions were used. For each Tire Type/Pavement Condition combination, $m = 6$ skid mark lengths were measured. The accompanying table shows some summary statistics for the study. (The units are cm.)

	Dry Concrete	Wet Concrete	Dirt
Smooth Tires	$\bar{y}_{11} = 359.8$ $s_{11} = 19.2$	$\bar{y}_{12} = 366.5$ $s_{12} = 26.4$	$\bar{y}_{13} = 393.0$ $s_{13} = 25.4$
Reverse Tread	$\bar{y}_{21} = 343.0$ $s_{21} = 15.5$	$\bar{y}_{22} = 356.7$ $s_{22} = 37.4$	$\bar{y}_{23} = 375.7$ $s_{23} = 39.9$
Treaded Tires	$\bar{y}_{31} = 384.8$ $s_{31} = 15.4$	$\bar{y}_{32} = 400.8$ $s_{32} = 60.8$	$\bar{y}_{33} = 402.5$ $s_{33} = 32.8$

(a) Compute s_{P} for Choi's data set. What is this supposed to be measuring?

(b) Make an interaction plot of the sample means similar to Figure 8.3. Use error bars for the means calculated from individual 95% two-sided confidence limits for the means. (Make use of your value of s_{P} from (a).)

(c) Based on your plot from (b), which factorial effects appear to be distinguishable from background noise? (Tire Type main effects? Pavement Condition main effects? Tire $\times$ Pavement interactions?)

(d) Compute all of the fitted factorial effects for Choi's data. (Find the a_i's, b_j's, and ab_{ij}'s defined in Section 4.3.)

(e) If one wishes to make individual 95% two-sided confidence intervals for the interactions $\alpha\beta_{ij}$, intervals of the form $ab_{ij} \pm \Delta$ are appropriate. Find Δ. Based on this value, are there statistically detectable interactions here? How does this conclusion compare with your more qualitative answer to part (c)?

(f) If one wishes to compare Tire Type main effects, confidence intervals for the differences $\alpha_i - \alpha_{i'}$ are in order. Find individual 95% two-sided confidence intervals for $\alpha_1 - \alpha_2$, $\alpha_1 - \alpha_3$, and $\alpha_2 - \alpha_3$. Based on these, are there statistically detectable differences in Tire Type main effects here? How does this conclusion compare with your answer to part (c)?

(g) Redo part (f), this time using (Tukey) simultaneous 95% two-sided confidence intervals.

8.2 *p*-Factor Studies with Two Levels for Each Factor

The previous section looked at inference for two-way factorial studies. This section presents methods of inference for complete p-way factorials, paying primary attention to those cases where each of p factors is represented at only two levels.

The discussion begins by again pointing out the relevance of the one-way methods of Chapter 7 to structured (in this case, p-way factorial) situations. Next, the p-way factorial normal model, definitions of effects in that model, and basic confidence interval methods for the effects are considered, paying particular attention to the 2^p case. Then attention is completely restricted to 2^p studies, and a further method for identifying detectable (2^p factorial) effects is presented. For balanced 2^p studies, there follows a review of the fitting of reduced models via the reverse Yates algorithm and the role of residuals in checking their efficacy. Finally, confidence interval methods based on simplified models in balanced 2^p studies are discussed.

8.2.1 One-Way Methods in *p*-Way Factorials

The place to begin the analysis of p-way factorial data is to recognize that fundamentally one is just working with several samples. Subject to the relevance of the model assumptions of Chapter 7, the inference methods of that chapter are available for use in analyzing the data.

Example 4

A 2^3 Factorial Study of Power Requirements in Metal Cutting

In *Fundamental Concepts in the Design of Experiments*, C. R. Hicks describes a study conducted by Purdue University engineering graduate student L. D. Miller on power requirements for cutting malleable iron using ceramic tooling. Miller studied the effects of the three factors

Factor A	Tool Type	(type 1 or type 2)
Factor B	Tool Bevel Angle	(15° or 30°)
Factor C	Type of Cut	(continuous or interrupted)

on the power required to make a cut on a lathe at a particular depth of cut, feed rate, and spindle speed. The response variable was the vertical deflection (in mm) of the indicator needle on a dynamometer (a measurement proportional to the horsepower required to make the particular cut). Miller's data are given in Table 8.8.

The most elementary view possible of the power requirement data in Table 8.8 is as $r = 8$ samples of size $m = 4$. Simple summary statistics for these $2^3 = 8$ samples are given in Table 8.9.

To the extent that the one-way normal model is an adequate description of this study, the methods of Chapter 7 are available for use in analyzing the data of Table 8.8. The reader is encouraged to verify that plotting of residuals (obtained by subtracting the $\bar{y}$ values in Table 8.9 from the corresponding raw data values of

Table 8.8
Dynamometer Readings for 2^3 Treatment Combinations in a Metal Cutting Study

Tool Type	Bevel Angle	Type of Cut	y, Dynamometer Reading (mm)
1	15°	continuous	29.0, 26.5, 30.5, 27.0
2	15°	continuous	28.0, 28.5, 28.0, 25.0
1	30°	continuous	28.5, 28.5, 30.0, 32.5
2	30°	continuous	29.5, 32.0, 29.0, 28.0
1	15°	interrupted	28.0, 25.0, 26.5, 26.5
2	15°	interrupted	24.5, 25.0, 28.0, 26.0
1	30°	interrupted	27.0, 29.0, 27.5, 27.5
2	30°	interrupted	27.5, 28.0, 27.0, 26.0

Example 4 *(continued)*

Table 8.9
Summary Statistics for 2^3 Samples of Dynamometer Readings in a Metal Cutting Study

Tool Type	Bevel Angle	Type of Cut	$\bar{y}$	s
1	15°	continuous	28.250	1.848
2	15°	continuous	27.375	1.601
1	30°	continuous	29.875	1.887
2	30°	continuous	29.625	1.702
1	15°	interrupted	26.500	1.225
2	15°	interrupted	25.875	1.548
1	30°	interrupted	27.750	0.866
2	30°	interrupted	27.125	0.854

Table 8.8) reveals only one slightly unpleasant feature of the power requirement data relative to the potential use of standard methods of inference. When plotted against levels of the Type of Cut variable, the residuals for interrupted cuts are shown to be on the whole somewhat smaller than those for continuous cuts. (This phenomenon is also obvious in retrospect from the sample standard deviations in Table 8.9. These are smaller for the second four samples than for the first four.) But the disparity in the sizes of the residuals is not huge. So although there may be some basis for suspecting improvement in power requirement consistency for interrupted cuts as opposed to continuous ones, the tractability of the one-way model and the kind of robustness arguments put forth at the end of Section 6.3 once again suggest that the standard model and methods be used. This is sensible, provided the resulting inferences are then treated as approximate and real-world "close calls" are not based on them.

The pooled sample variance here is

$$s_{\mathrm{P}}^2 = \frac{(4-1)(1.848)^2 + (4-1)(1.601)^2 + \cdots + (4-1)(.854)^2}{(4-1)+(4-1)+\cdots+(4-1)} = 2.226$$

so

$$s_{\mathrm{P}} = 1.492 \text{ mm}$$

with $\nu = n - r = 32 - 8 = 24$ associated degrees of freedom. Then, for example, the P-R method of simultaneous inference from Section 7.3 produces two-sided simultaneous 95% confidence intervals for mean dynamometer readings with endpoints

$$\bar{y}_{ijk} \pm 2.969 \frac{1.492}{\sqrt{4}}$$

that is,

$$\bar{y}_{ijk} \pm 2.21 \text{ mm}$$

(There is enough precision provided by the data to think of the sample means in Table 8.9 as roughly "all good to within 2.21 mm.") And the other methods of Sections 7.1 through 7.4 based on s_{P} might be used as well.

8.2.2 *p*-Way Factorial Notation, Definitions of Effects, and Related Confidence Interval Methods

Section 8.1 illustrated that standard notation in two-way factorials requires triple subscripts for naming observations. In a general p-way factorial, "$(p+1)$-subscript" notation is required. As p grows, such notation quickly gets out of hand. As in Section 4.3 (on a descriptive level) the exposition here will explicitly develop only the general factorial notation for $p = 3$, leaving the reader to infer by analogy how things would have to go for $p = 4$, 5, etc. (When specializing to the 2^p situation later in this section, the special notation introduced in Section 4.3 makes it possible to treat even large-p situations fairly explicitly.)

Three factor (quadruple subscript) notation

Then for $p = 3$ factors A, B, and C having (respectively) I, J, and K levels, let

y_{ijkl} = the lth observation in the sample from the ith level of A, jth level of B, and kth level of C

For the $I \cdot J \cdot K$ different samples corresponding to the possible combinations of a level of A with one of B and one of C, let

n_{ijk} = the number of observations in the sample from the ith level of A, jth level of B, and kth level of C

$\bar{y}_{ijk}$ = the sample mean of the n_{ijk} observations in the sample from the ith level of A, jth level of B, and kth level of C

s_{ijk} = the sample standard deviation of the n_{ijk} observations in the sample from the ith level of A, jth level of B, and kth level of C

and further continue the dot notations used in Section 4.3 for unweighted averages of the $\bar{y}_{ijk}$. In comparison to the notation of Chapter 7, this amounts to adding two subscripts in order to acknowledge the three-way structure in the samples.

The use of additional subscripts is helpful not only for naming empirical quantities but also for naming theoretical quantities. That is, with

μ_{ijk} = the underlying mean response corresponding to the ith level of A, jth level of B, and kth level of C the standard one-way normal model assumptions can be rewritten as

Three-way model statement

$$y_{ijkl} = \mu_{ijk} + \epsilon_{ijkl} \tag{8.11}$$

where the ϵ_{ijkl} terms are iid normal random variables with mean 0 and variance σ^2. Formula (8.11) could be called the **three-way (normal) model equation** because it recognizes the special organization of the $I \cdot J \cdot K$ samples according to combinations of levels of the three factors. But beyond this, it says no more or less than the one-way model equation from Section 7.1.

The initial objects of inference in three-way factorial analyses are linear combinations of theoretical means μ_{ijk}, analogous to the fitted effects of Section 4.3. Thus, it is necessary to carefully define the theoretical or underlying main effects, 2-factor interactions, and 3-factor interactions for a three-way factorial study. In the definitions that follow, a dot appearing as a subscript will (as usual) be understood to indicate that an average has been taken over all levels of the factor corresponding to the dotted subscript. Consider first main effects. Parallel to Definition 7 in Chapter 4 (page 182) for fitted main effects is a definition of theoretical main effects.

Definition 3

In a three-way complete factorial study with factors A, B, and C, the **main effect of factor A at its *i*th level** is

$$\alpha_i = \mu_{i..} - \mu_{...}$$

the **main effect of factor B at its *j*th level** is

$$\beta_j = \mu_{.j.} - \mu_{...}$$

and the **main effect of factor C at its *k*th level** is

$$\gamma_k = \mu_{..k} - \mu_{...}$$

These main effects measure how (when averaged over all combinations of levels of the other factors) underlying mean responses change from level to level of the factor in question. Definition 3 has the algebraic consequences that

$$\sum_{i=1}^{I} \alpha_i = 0, \quad \sum_{j=1}^{J} \beta_j = 0, \quad \text{and} \quad \sum_{k=1}^{K} \gamma_k = 0$$

The theoretical counterpart of Definition 8 in Chapter 4 is a definition of theoretical 2-factor interactions.

Definition 4

In a three-way complete factorial study with factors A, B, and C, the **2-factor interaction of factor A at its *i*th level and factor B at its *j*th level** is

$$\alpha\beta_{ij} = \mu_{ij.} - (\mu_{...} + \alpha_i + \beta_j)$$

the **2-factor interaction of A at its *i*th level and C at its *k*th level** is

$$\alpha\gamma_{ik} = \mu_{i.k} - (\mu_{...} + \alpha_i + \gamma_k)$$

and the **2-factor interaction of B at its *j*th level and C at its *k*th level** is

$$\beta\gamma_{jk} = \mu_{.jk} - (\mu_{...} + \beta_j + \gamma_k)$$

Like their empirical counterparts defined in Section 4.3, the 2-factor interactions in a three-way study are measures of lack of parallelism on two-way plots of means obtained by averaging out over levels of the “other” factor. And it is an algebraic consequence of the form of Definition 4 that

$$\sum_{i=1}^{I} \alpha\beta_{ij} = \sum_{j=1}^{J} \alpha\beta_{ij} = 0, \qquad \sum_{i=1}^{I} \alpha\gamma_{ik} = \sum_{k=1}^{K} \alpha\gamma_{ik} = 0$$

and

$$\sum_{j=1}^{J} \beta\gamma_{jk} = \sum_{k=1}^{K} \beta\gamma_{jk} = 0$$

Finally, there is the matter of three-way interactions in a three-way factorial study. Direct analogy with the meaning of fitted three-way interactions given as Definition 9 in Chapter 4 (page 183) gives the following:

Definition 5

In a three-way complete factorial study with factors A, B, and C, the **3-factor interaction of factor A at its *i*th level, factor B at its *j*th level, and factor C at its *k*th level** is

$$\alpha\beta\gamma_{ijk} = \mu_{ijk} - (\mu_{...} + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk})$$

Like their fitted counterparts, the (theoretical) 3-factor interactions are measures of patterns in the μ_{ijk} not describable in terms of the factors acting separately or in pairs. Or differently put, they measure how much what one would call the AB interactions at a single level of C change from level to level of C. And, like the fitted 3-factor interactions defined in Section 4.3, the theoretical 3-factor interactions defined here sum to 0 over levels of any one of the factors. That is,

$$\sum_{i=1}^{I} \alpha\beta\gamma_{ijk} = \sum_{j=1}^{J} \alpha\beta\gamma_{ijk} = \sum_{k=1}^{K} \alpha\beta\gamma_{ijk} = 0$$

Factorial effects are L's, fitted effects are corresponding $\hat{L}$'s

The fundamental fact that makes inference for the factorial effects defined in Definitions 3, 4, and 5 possible is that they are particular linear combinations of the means μ_{ijk} (L's from Section 7.2). And the fitted effects from Section 4.3 are the corresponding linear combinations of the sample means $\bar{y}_{ijk}$ ($\hat{L}$'s from Section 7.2). So at least in theory, to make confidence intervals for the factorial effects, one needs only to figure out exactly what coefficients are applied to each of the means and use formula (7.20) of Section 7.2.

Example 5

Finding Coefficients on Means for a Factorial Effect in a Three-Way Factorial

Consider a hypothetical example in which A appears at $I = 2$ levels, B at $J = 2$ levels, and C at $K = 3$ levels. For the sake of illustration, consider how you would make a confidence interval for $\alpha\gamma_{23}$. By Definitions 3 and 4,

$$\begin{aligned}
\alpha\gamma_{23} &= \mu_{2.3} - (\mu_{...} + \alpha_2 + \gamma_3) \\
&= \mu_{2.3} - (\mu_{2..} + \mu_{..3} - \mu_{...}) \\
&= \frac{1}{2}(\mu_{213} + \mu_{223}) - \frac{1}{6}(\mu_{211} + \mu_{221} + \mu_{212} + \mu_{222} + \mu_{213} + \mu_{223}) \\
&\quad - \frac{1}{4}(\mu_{113} + \mu_{213} + \mu_{123} + \mu_{223}) + \frac{1}{12}(\mu_{111} + \mu_{211} + \cdots + \mu_{223}) \\
&= \frac{1}{6}\mu_{213} + \frac{1}{6}\mu_{223} - \frac{1}{12}\mu_{211} - \frac{1}{12}\mu_{221} - \frac{1}{12}\mu_{212} - \frac{1}{12}\mu_{222} \\
&\quad - \frac{1}{6}\mu_{113} - \frac{1}{6}\mu_{123} + \frac{1}{12}\mu_{111} + \frac{1}{12}\mu_{121} + \frac{1}{12}\mu_{112} + \frac{1}{12}\mu_{122}
\end{aligned}$$

so the "$\sum c_i^2/n_i$" applicable to estimating $\alpha\gamma_{23}$ via formula (7.20) of Section 7.2 is

$$\begin{aligned}
\sum \frac{c_{ijk}^2}{n_{ijk}} &= \left(\frac{1}{6}\right)^2 \left(\frac{1}{n_{213}} + \frac{1}{n_{223}} + \frac{1}{n_{113}} + \frac{1}{n_{123}}\right) \\
&\quad + \left(\frac{1}{12}\right)^2 \left(\frac{1}{n_{211}} + \frac{1}{n_{221}} + \frac{1}{n_{212}} + \frac{1}{n_{222}} + \frac{1}{n_{111}} + \frac{1}{n_{121}} + \frac{1}{n_{112}} + \frac{1}{n_{122}}\right)
\end{aligned}$$

and using this expression, endpoints for a confidence interval for $\alpha\gamma_{23}$ are

$$ac_{23} \pm t s_{\mathrm{P}} \sqrt{\sum \frac{c_{ijk}^2}{n_{ijk}}}$$

It is possible to work out (unpleasant) general formulas for the "$\sum c_i^2/n_i$" terms for factorial effects in arbitrary p-way factorials and implement them in computer software. It is not consistent with the purposes of this book to lay those out here. However, in the special case of 2^p factorials, there is no difficulty in describing how to make confidence intervals for the effects or in carrying out a fairly complete analysis of all of these "by hand" for p as large as even 4 or 5. This is because the 2^p case of the general p-way factorial structure allows three important simplifications. First, for any factorial effect in a 2^p factorial, the coefficients "c_i" applied to the means to produce the effect are all $\pm\frac{1}{2^p}$. So the "$\sum c_i^2/n_i$" term needed to make a confidence interval for any effect in a 2^p factorial is

Coefficients applied to means to produce 2^p factorial effects are all $\pm\frac{1}{2^p}$

$$\left(\pm\frac{1}{2^p}\right)^2\left(\frac{1}{n_{(1)}}+\frac{1}{n_{\mathrm{a}}}+\frac{1}{n_{\mathrm{b}}}+\frac{1}{n_{\mathrm{ab}}}+\cdots\right)$$

where the subscripts (1), a, b, ab, etc. refer to the combination-naming convention for 2^p factorials introduced in Section 4.3.

So let E stand for a generic effect in a 2^p factorial (a particular kind of L from Section 7.2) and $\hat{E}$ be the corresponding fitted effect (the corresponding $\hat{L}$ from Section 7.2). Then endpoints of an individual two-sided confidence interval for E are

Individual confidence limits for an effect in a 2^p factorial

$$\boxed{\hat{E} \pm t s_{\mathrm{P}} \frac{1}{2^p}\sqrt{\frac{1}{n_{(1)}}+\frac{1}{n_{\mathrm{a}}}+\frac{1}{n_{\mathrm{b}}}+\frac{1}{n_{\mathrm{ab}}}+\cdots}} \qquad \mathbf{(8.12)}$$

where the associated confidence is the probability that the t distribution with $\nu = n - r = n - 2^p$ degrees of freedom assigns to the interval between $-t$ and t. The usual device of using only one endpoint from formula (8.12) and halving the unconfidence produces a one-sided confidence interval for the effect. And in balanced-data situations where all sample sizes are equal to m, formula (8.12) can be written even more simply as

Balanced data confidence limits for an effect in a 2^p factorial

$$\boxed{\hat{E} \pm t \frac{s_{\mathrm{P}}}{\sqrt{m2^p}}} \qquad \mathbf{(8.13)}$$

Estimating one 2^p effect of a given type is enough

There is a second simplification of the general p-way factorial situation afforded in the 2^p case. Because of the way factorial effects sum to 0 over levels of any factor involved, estimating *one* effect of each type is sufficient to completely describe a 2^p factorial. For example, since in a 2^p factorial,

$$\alpha\beta_{11} = -\alpha\beta_{21} = -\alpha\beta_{12} = \alpha\beta_{22}$$

it is necessary to estimate only one AB interaction to have detailed what is known about 2-factor interactions of A and B. There is no need to labor in finding separate estimates of $\alpha\beta_{11}$, $\alpha\beta_{12}$, $\alpha\beta_{21}$, and $\alpha\beta_{22}$. Appropriate sign changes on an estimate of $\alpha\beta_{22}$ suffice to cover the matter.

The third important fact making analysis of 2^p factorial effects so tractable is the existence of the Yates algorithm. As demonstrated in Example 9 of Chapter 4, it is really quite simple to use the algorithm to mechanically generate one fitted effect of each type for a given 2^p data set: those effects corresponding to the high levels of all factors.

Example 4 *(continued)*

Consider again the metal working power requirement study. Agreeing to (arbitrarily) name tool type 2, the 30° tool bevel angle, and the interrupted cut type as the "high" levels of (respectively) factors A, B, and C, the eight combinations of the three factors are listed in Table 8.9 in Yates standard order. Taking the sample means from that table in the order listed, the Yates algorithm can be applied to produce the fitted effects for the high levels of all factors, as in Table 8.10.

Recall that for the data of Table 8.8, $m = 4$ and $s_{\text{P}} = 1.492$ mm with $24(= 32 - 2^3)$ associated degrees of freedom. So one has (from formula (8.13)) that for (say) individual 90% confidence, the factorial effects in this example can be estimated with two-sided intervals having endpoints

$$\hat{E} \pm 1.711\frac{1.492}{\sqrt{4 \cdot 2^3}}$$

Table 8.10
The Yates Algorithm Applied to the Means in Table 8.9

Combination	$\bar{y}$	Cycle 1	Cycle 2	Cycle 3	Cycle 3 ÷ 8
(1)	28.250	55.625	115.125	222.375	$27.7969 = \bar{y}_{...}$
a	27.375	59.500	107.250	−2.375	$-.2969 = a_2$
b	29.875	52.375	−1.125	6.375	$.7969 = b_2$
ab	29.625	54.875	−1.250	.625	$.0781 = ab_{22}$
c	26.500	−.875	3.875	−7.875	$-.9844 = c_2$
ac	25.875	−.250	2.500	−.125	$-.0156 = ac_{22}$
bc	27.750	−.625	.625	−1.375	$-.1719 = bc_{22}$
abc	27.125	−.625	0.000	−.625	$-.0781 = abc_{222}$

that is,

$$\hat{E} \pm .45$$

Then, comparing the fitted effects in the last column of Table 8.10 to the $\pm.45$ value, note that only the main effects of Tool Bevel Angle (factor B) and Type of Cut (factor C) are statistically detectable. And for example, it appears that running the machining process at the high level of factor B (the 30° bevel angle) produces a dynamometer reading that is on average between approximately

$$2(.80 - .45) = .7 \text{ mm} \quad \text{and} \quad 2(.80 + .45) = 2.5 \text{ mm}$$

The difference between main effects at high and low levels of a factor is twice the effect

higher than when the process is run at the low level of factor B (the 15° bevel angle). (The difference between B main effects at the high and low levels of B is $\beta_2 - \beta_1 = \beta_2 - (-\beta_2) = 2\beta_2$, hence the multiplication by 2 of the endpoints of the confidence interval for β_2.)

8.2.3 2^p Studies Without Replication and the Normal-Plotting of Fitted Effects

The use of formula (8.12) or (8.13) in judging the detectability of 2^p factorial effects is an extremely practical and effective method. But it depends for its applicability on there being replication somewhere in the data set. One must have a pooled sample standard deviation s_P. Unfortunately, it is not uncommon that poorly informed people do unreplicated 2^p factorial studies. Although such studies should be avoided whenever possible, various methods of analysis have been suggested for them. The most popular one follows from a very clever line of reasoning due originally to Cuthbert Daniel.

Daniel's idea was to invoke a principle of **effect sparsity**. He reasoned that in many real engineering systems, the effects of only a relatively few factors are the primary determiners of system mean response. Thus, in terms of the 2^p factorial effects used here, a relatively few of $\alpha_2, \beta_2, \alpha\beta_{22}, \gamma_2, \alpha\gamma_{22}, \ldots$, etc., often dominate the rest (are much larger in absolute value than the majority). In turn, this would imply that often among the fitted effects $a_2, b_2, ab_{22}, c_2, ac_{22}, \ldots$, etc., there are a few with sizable means, and the others have means that are (relatively speaking) near 0. Daniel's idea for identifying those cases where a few effects dominate the rest was to normal-plot the fitted effects for the "all high treatment" combination (obtained, for example, by use of the Yates algorithm). When a few plot in positions much more extreme than would be predicted from putting a line through the majority of the points, they are identified as the likely principal determiners of system behavior. (Actually, Daniel originally proposed making a **half normal plot** of the absolute values of the fitted effects. This was to eliminate any visual effect of the somewhat arbitrary naming of one level of each factor as the high level. For several reasons, among them simplicity, this presentation will use the full normal plot modification of Daniel's method. The idea of half normal plotting is considered further in Chapter Exercise 9.)

Example 6
(Example 12, Chapter 4, revisited—page 195)

Identifying Detectable Effects in an Unreplicated 2^4 Factorial Drill Advance Rate Study

Section 4.4 discussed an example of an unreplicated 2^4 factorial experiment taken from Daniel's *Applications of Statistics to Industrial Experimentation*. There the effects of the four factors

Factor A	Load
Factor B	Flow Rate
Factor C	Rotational Speed
Factor D	Type of Mud

on the logarithm of an advance rate of a small stone drill were considered. (The raw data are in Table 4.24.) The Yates algorithm applied to the $16 = 2^4$ observed log advance rates produced the following fitted effects:

$$
\begin{array}{llll}
\bar{y}_{....} = 1.5977 & & & \\
a_2 = .0650 & b_2 = .2900 & c_2 = .5772 & d_2 = .1633 \\
ab_{22} = -.0172 & ac_{22} = .0052 & ad_{22} = .0334 & \\
bc_{22} = -.0251 & bd_{22} = -.0075 & cd_{22} = .0491 & \\
abc_{222} = .0052 & abd_{222} = .0261 & acd_{222} = .0266 & \\
bcd_{222} = -.0173 & abcd_{2222} = .0193 & &
\end{array}
$$

Figure 8.6 is a normal plot of the 15 fitted effects a_2 through $abcd_{2222}$.

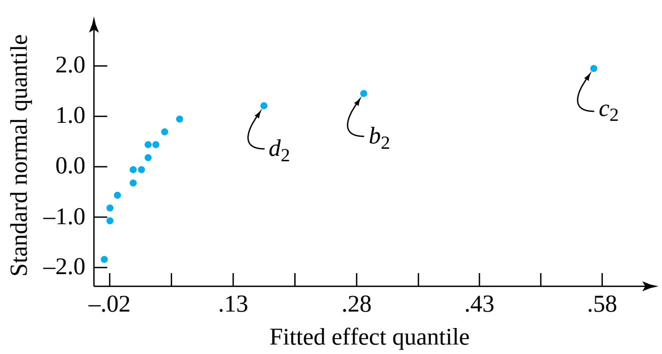

Figure 8.6 Normal plot of the fitted effects for Daniel's drill advance rate study

Applying Daniel's reasoning, it is obvious that the points corresponding to the C, B, and D main effects plot off any sensible line established through the bulk of the plotted points. So it becomes natural to think that these main effects are detectably larger than the other effects, and therefore distinguishable from experimental error even if the others are not. Thus, it seems that drill behavior is potentially describable in terms of the (separate) action of the factors Rotational Speed, Flow Rate, and Mud Type.

The plotted fitted effects concern the natural logarithm of advance rate. So the fact that $c_2 = .5772$ says that changing from the low level of rotational speed to the high level produces roughly an increase of $2(.5772) \approx 1.15$ in the natural log of the advance rate—i.e., increases the advance rate by a factor of $e^{1.15} \approx 3.2$.

Interpreting a normal plot of fitted effects

Example 6 is one in which the normal plotting clearly identifies a few effects as larger than the others. However, a normal plot of fitted effects sometimes has a fairly straight-line appearance. When this happens, the message is that the fitted effects are potentially explainable as resulting from background variation. And it is risky to make real-world engineering decisions based on fitted effects that haven't been definitively established as representing consistent system reactions to changes in level of the corresponding factors. A linear normal plot of fitted effects from an unreplicated 2^p study says that more data are needed.

This normal-plotting device has been introduced primarily as a tool for analyzing data lacking any replication. However, the method is useful even in cases where there is some replication and s_{P} can therefore be calculated and formula (8.12) or (8.13) used to judge the detectability of the various factorial effects. Some practice making and using such plots will show that the process often amounts to a helpful kind of "data fondling." Many times, a bit of thought makes it possible to trace an unusual pattern on such a plot back to a previously unnoticed peculiarity in the data.

As an example, consider what a normal plot of fitted effects would point out about the following eight hypothetical sample means.

$$\begin{aligned} \bar{y}_{(1)} &= 95 & \qquad \bar{y}_{\text{c}} &= 145 \\ \bar{y}_{\text{a}} &= 101 & \bar{y}_{\text{ac}} &= 103 \\ \bar{y}_{\text{b}} &= 106 & \bar{y}_{\text{bc}} &= 107 \\ \bar{y}_{\text{ab}} &= 106 & \bar{y}_{\text{abc}} &= 97 \end{aligned}$$

This is an exaggerated example of a phenomenon that sometimes occurs less blatantly in practice. $2^p - 1$ of the sample means are more or less comparable, while one of the means is clearly different. When this occurs (unless the unusual mean corresponds to the "all high treatment" combination), a normal plot of fitted effects roughly like the one in Figure 8.7 will follow. About half the fitted effects will be large positively and the other half large negatively. (When the unusual mean is the one corresponding to the "all high" combination, the fitted effects will all have the same sign.)

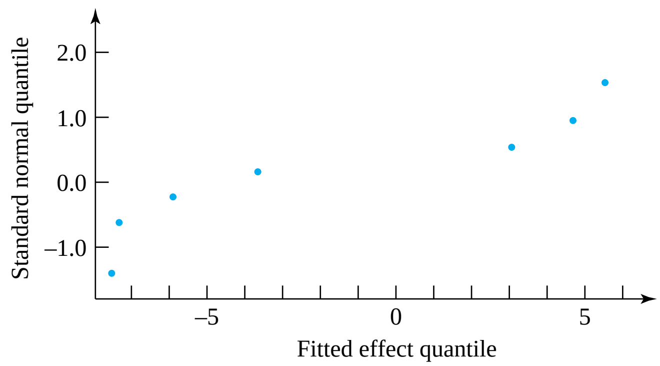

Figure 8.7 Normal plot of fitted effects for eight hypothetical means

8.2.4 Fitting and Checking Simplified Models in Balanced 2^p Factorial Studies and a Corresponding Variance Estimate (*Optional*)

When beginning the analysis of a 2^p factorial, one hopes that a simplified p-way model involving only a few main effects and/or low-order interactions will be adequate to describe it. Analyses based on formulas (8.12) or (8.13) or normal-plotting are ways of identifying such potential descriptions of special p-way structure. Once a potential simplification of the 2^p analog of model (8.11) has been identified, it is often of interest to go beyond that identification to

1. the fitting and checking (residual analysis) of the simplified model, and even to
2. the making of formal inferences under the restricted/simplified model assumptions.

When a 2^p factorial data set is *balanced*, the model fitting, checking, and subsequent interval-oriented inference is straightforward.

With balanced 2^p factorial data, producing least squares fitted values is no more difficult than adding together (with appropriate signs) desired fitted effects and the grand sample mean. Or equivalently and more efficiently, the reverse Yates algorithm can be used.

Example 4 (*continued*)

In the power requirement study and the data of Table 8.8, only the B and C main effects seem detectably nonzero. So it is reasonable to think of the simplified version of model (8.11),

$$y_{ijkl} = \mu_{...} + \beta_j + \gamma_k + \epsilon_{ijkl} \qquad \textbf{(8.14)}$$

for possible use in describing dynamometer readings. From Table 8.10, the fitted version of $\mu_{..}$ is $\bar{y}_{..} = 27.7969$, the fitted version of β_2 is $b_2 = .7969$, and the fitted version of γ_2 is $c_2 = -.9844$. Then, simply adding together appropriate signed versions of the fitted effects, for the four possible combinations of j and k, produces the corresponding fitted responses in Table 8.11. So for example, as long as the 15° bevel angle (low level of B) and a continuous cut (low level of C) are being considered, a fitted dynamometer reading of about 27.98 is appropriate under the simplified model (8.14).

Table 8.11
Fitted Responses for a "B and C Main Effects Only" Description of Power Requirement

j	k	b_j	c_k	$\hat{y} = \bar{y}_{...} + b_j + c_k$
1	1	$-.7969$	.9844	27.9844
2	1	.7969	.9844	29.5782
1	2	$-.7969$	$-.9844$	26.0156
2	2	.7969	$-.9844$	27.6094

Example 6 (*continued*)

Having identified the C, B, and D main effects as detectably larger than the A main effect or any of the interactions in the drill advance rate study, it is natural to consider fitting the model

$$y_{ijkl} = \mu_{...} + \beta_j + \gamma_k + \delta_l + \epsilon_{ijkl} \quad \textbf{(8.15)}$$

to the logarithms of the unreplicated 2^4 factorial data of Table 4.24. (Note that even though $p = 4$ factors are involved here, five subscripts are not required, since a subscript is not needed to differentiate between multiple members of the 2^4 different samples in this unreplicated context. y_{ijkl} is the single observation at the ith level of A, jth level of B, kth level of C, and lth level of D.) Since the drill advance rate data are balanced (all sample sizes are $m = 1$), the fitted effects given earlier (calculated without reference to the simplified model) serve as fitted effects under model (8.15). And fitted responses under model (8.15) are obtainable by simple addition and subtraction using those.

Since there are eight different combinations of j, k, and l, eight different linear combinations of $\bar{y}_{....}$, b_2, c_2, and d_2 are required. While these could be treated one at time, it is more efficient to generate them all at once using the reverse Yates algorithm (from Section 4.3) as in Table 8.12. From Table 8.12 it is evident, for example, that the fitted mean responses for combinations bcd and abcd ($\hat{y}_{\text{bcd}}$ and $\hat{y}_{\text{abcd}}$) are both 2.6282.

Example 6 (*continued*)

Table 8.12
The Reverse Yates Algorithm Used to Fit the "B, C, and D Main Effects" Model to Daniel's Data

Fitted Effect	Value	Cycle 1	Cycle 2	Cycle 3	Cycle 4 ($\hat{y}$)
$abcd_{2222}$	0	0	0	.1633	2.6282
bcd_{222}	0	0	.1633	2.4649	2.6282
acd_{222}	0	0	.5772	.1633	2.0482
cd_{22}	0	.1633	1.8877	2.4649	2.0482
abd_{222}	0	0	0	.1633	1.4738
bd_{22}	0	.5772	.1633	1.8849	1.4738
ad_{22}	0	.2900	.5772	.1633	.8938
d_2	.1633	1.5977	1.8877	1.8849	.8938
abc_{222}	0	0	0	.1633	2.3016
bc_{22}	0	0	.1633	1.3105	2.3016
ac_{22}	0	0	.5772	.1633	1.7216
c_2	.5772	.1633	1.3077	1.3105	1.7216
ab_{22}	0	0	0	.1633	1.1472
b_2	.2900	.5772	.1633	.7305	1.1472
a_2	0	.2900	.5772	.1633	.5672
$\bar{y}_{...}$	1.5977	1.5977	1.3077	.7305	.5672

Fitted means derived as in these examples lead in the usual way to residuals, R^2 values, and plots for checking on the reasonableness of simplified versions of the general 2^p version of model (8.11). In addition, corresponding to simplified or *reduced* models like (8.14) or (8.15), there are what will here be called **few-effects** s^2 values. When $m > 1$, these can be compared to s_{P}^2 as another means of investigating the reasonableness of the corresponding models.

Definition 6

In a balanced complete 2^p factorial study, if a reduced or simplified model involving u different effects (including the grand mean) has corresponding fitted values $\hat{y}$ and thus residuals $y - \hat{y}$, the quantity

$$s_{\text{FE}}^2 = \frac{1}{m2^p - u} \sum (y - \hat{y})^2 \tag{8.16}$$

will be called a **few-effects sample variance**. Associated with it are $\nu = m2^p - u$ degrees of freedom.

The quantity (8.16) represents an estimator of the basic background variance whenever the corresponding simplified/reduced/few-effects model is an adequate description of the study. When it is not, s_{FE} will tend to overestimate σ. So comparing s_{FE} to s_{P} is a way of investigating the appropriateness of that description.

It is not obvious at this point, but there is a helpful alternative way to calculate the value of s_{FE}^2 given in formula (8.16). It turns out that

An alternative formula for a few effects sample variance

$$s_{\mathrm{FE}}^2 = \frac{1}{m2^p - u}\left[SSTot - m2^p \sum \hat{E}^2\right] \tag{8.17}$$

where the sum is over the squares of the $u - 1$ fitted effects corresponding to those main effects and interactions appearing in the reduced model equation, and (as always) $SSTot = \sum (y - \bar{y})^2 = (n-1)s^2$.

Example 4 *(continued)*

Residuals for the power requirement data based on the full model (8.11) are obtained by subtracting sample means in Table 8.9 from observations in Table 8.8. Under the reduced model (8.14), however, the fitted values in Table 8.11 are appropriate for producing residuals. The fitted means and residuals for a "B and C main effects only" description of this 2^3 data set are given in Table 8.13. Figure 8.8 is a normal plot of these residuals, and Figure 8.9 is a plot of the residuals against the fitted values.

If there is anything remarkable in these plots, it is that Figure 8.9 contains a hint that smaller mean response has associated with it smaller response variability. In fact, looking back at Table 8.13, it is easy to see that the two smallest fitted means correspond to the high level of C (i.e., interrupted cuts). That is, the hint of change in response variation shown in Figure 8.9 is the same phenomenon related

Table 8.13
Residuals for the "B and C Main Effects Only" Model of Power Requirement

Combination	$\hat{y}$	Residuals $(y - \hat{y})$
(1)	27.9844	1.0156, −1.4844, 2.5156, −.9844
a	27.9844	.0156, .5156, .0156, −2.9844
b	29.5782	−1.0782, −1.0782, .4218, 2.9218
ab	29.5782	−.0782, 2.4218, −.5782, −1.5782
c	26.0156	1.9844, −1.0156, .4844, .4844
ac	26.0156	−1.5156, −1.0156, 1.9844, −.0156
bc	27.6094	−.6094, 1.3906, −.1094, −.1094
abc	27.6094	−.1094, .3906, −.6094, −1.6094

Example 4
(*continued*)

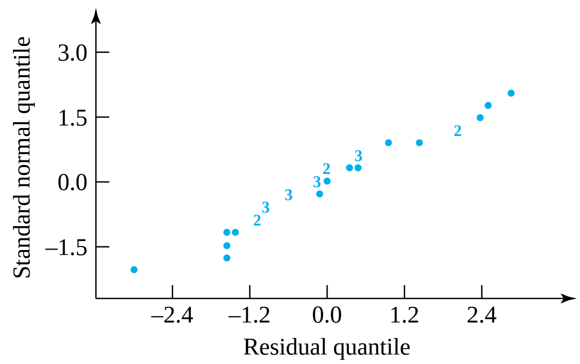

Figure 8.8 Normal plot of residuals for the power requirement study (B and C main effects only)

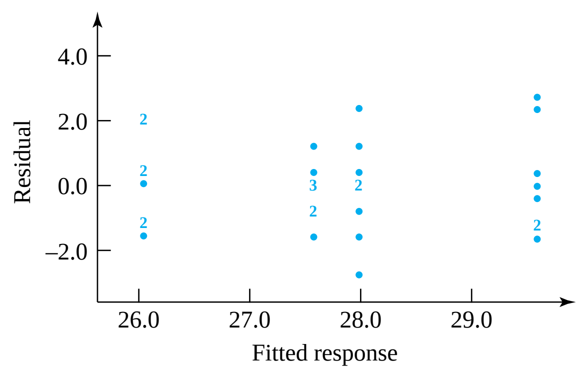

Figure 8.9 Plot of residuals versus fitted power requirements (B and C main effects only)

to cut type that was discussed when these data were first introduced. It appears that power requirements for interrupted cuts may be slightly more consistent than for continuous cuts. But on the whole, there is little in the two figures to invalidate model (8.14) as at least a rough-and-ready description of the mechanism behind the data of Table 8.8.

For the power requirement data,

$$SSTot = (n-1)s^2 = 108.93$$

Then, since $s_P^2 = 2.226$, the one-way ANOVA identity (7.49, 7.50, or 7.51) of Section 7.4 says that

$$SSTr = SSTot - SSE = 108.93 - 24(2.226) = 55.51$$

so R^2 corresponding to the general or "full" model (8.11) is (as in equations (7.52) or (7.53))

$$R^2 = \frac{SStr}{SStot} = \frac{55.51}{108.93} = .51$$

On the other hand, it is possible to verify that for the simplified model (8.14), squaring and summing the residuals in Table 8.13 gives

$$SSE = \sum (y - \hat{y})^2 = 57.60$$

(Recall Definition 6 in Chapter 7 for SSE.) So for the "B and C main effects only" description of dynamometer readings,

$$R^2 = \frac{SStot - SSE}{SStot} = \frac{108.93 - 57.60}{108.93} = .47$$

Thus, although at best only about 51% of the raw variation in dynamometer readings will be accounted for, fitting the simple model (8.14) will account for nearly all of that potentially assignable variation. So from this point of view as well, model (8.14) seems attractive as a description of power requirement.

Note that formulas (8.16) and (8.17) imply that for balanced 2^p factorial data, fitting reduced models gives

$$\sum (y - \hat{y})^2 = SSTot - m2^p \sum \hat{E}^2$$

So it is not surprising that using the $b_2 = .7969$ and $c_2 = -.9844$ figures from before,

$$\begin{aligned} SSTot - m2^p \sum \hat{E}^2 &= 108.93 - 4 \cdot 2^3 \cdot \left((.7969)^2 + (-.9844)^2\right) \\ &= 108.93 - 51.33 \\ &= 57.60 \end{aligned}$$

which is the value of $\sum (y - \hat{y})^2$ just used in finding R^2 for the reduced model. From formula (8.16) or (8.17), it is then clear that (corresponding to reduced model (8.14))

$$s_{\text{FE}}^2 = \frac{1}{4 \cdot 2^3 - 3}(57.60) = 1.986$$

so

$$s_{\text{FE}} = \sqrt{1.986} = 1.409 \text{ mm}$$

Example 4 (*continued*)

which agrees closely with $s_{\mathrm{P}} = 1.492$. Once again on this account, description (8.14) seems quite workable.

Example 6 (*continued*)

Table 8.14 contains the log advance rates, fitted values, and residuals for Daniel's unreplicated 2^4 example. (The raw data were given in Table 4.24, and it is the few-effects model (8.15) that is under consideration.)

The reader can verify by plotting that the residuals in Table 8.14 are not in any way remarkable. Further, it is possible to check that

$$SSTot = \sum(y - \bar{y})^2 = 7.2774$$

and

$$SSE = \sum(y - \hat{y})^2 = .1736$$

So (as indicated earlier in Example 12 in Chapter 4) for the use of model (8.15),

$$R^2 = \frac{SStot - SSE}{SStot} = \frac{7.2774 - .1736}{7.2774} = .976$$

Table 8.14
Responses, Fitted Values, and Residuals for the "B, C, and D Main Effects" Model and Daniel's Drill Advance Rate Data

Combination	y, ln(advance rate)	$\hat{y}$	$e = y - \hat{y}$
(1)	.5188	.5672	−.0484
a	.6831	.5672	.1159
b	1.1878	1.1472	.0406
ab	1.2355	1.1472	.0883
c	1.6054	1.7216	−.1162
ac	1.7405	1.7216	.0189
bc	2.2996	2.3016	−.0020
abc	2.2050	2.3016	−.0966
d	.7275	.8938	−.1663
ad	.8920	.8938	−.0018
bd	1.4085	1.4738	−.0653
abd	1.5107	1.4738	.0369
cd	2.0503	2.0482	.0021
acd	2.2439	2.0482	.1957
bcd	2.4639	2.6282	−.1643
abcd	2.7912	2.6282	.1630

Since there is no replication in this data set, fitting the 4-factor version of the general model (8.11) would give a perfect fit, R^2 equal to 1.000, all residuals equal to 0, and no value of s_{P}^2. Thus, there is really nothing to judge $R^2 = .976$ against in relative terms. But even in absolute terms it appears that the "B, C, and D main effects only" model for log advance rate fits the data well.

An estimate of the variability of log advance rates for a fixed combination of factor levels derived under the assumptions of model (8.15), is (from formula (8.16))

$$s_{\text{FE}} = \sqrt{\frac{1}{1 \cdot 2^4 - 4}(.1736)} = .120$$

As noted, there's no s_{P} to compare this to, but it is at least consistent with the kind of variation in y seen in Table 8.14 when responses are compared for pairs of combinations that (like combinations b and ab) differ only in level of the factor A.

8.2.5 Confidence Intervals for Balanced 2^p Studies under Few-Effects Models (*Optional*)

Since the basic p-way factorial model is just a rewritten version of the one-way normal model from Chapter 7, the confidence interval methods of that chapter can all see application in p-way factorial studies. But when a simplified/few-effects model is appropriate, sharper real-world engineering conclusions can usually be had by using methods based on the simplified model than by applying the general methods of Chapter 7. And for balanced 2^p studies, it is possible to write down simple, explicit formulas for several useful forms of interval-oriented inference.

As a first example of what is possible under a few-effects model in a balanced 2^p factorial study, consider the estimation of a particular mean response. For balanced data, the 2^p fitted effects (including the grand mean) that come out of the Yates algorithm are independent normal variables with means equal to the corresponding underlying effects and variances $\sigma^2/m2^p$. So, if a simplified version of model (8.11) involving u effects (including the overall mean) is appropriate, a fitted response $\hat{y}$ has mean equal to the corresponding underlying mean, and

$$\text{Var}\,\hat{y} = u\frac{\sigma^2}{m2^p}$$

It should then be plausible that under a few-effects model in a balanced 2^p factorial study, a two-sided interval with endpoints

Balanced data individual confidence limits for a mean repsonse under a simplified model

$$\boxed{\hat{y} \pm t s_{\text{FE}}\sqrt{\frac{u}{m2^p}}} \qquad \textbf{(8.18)}$$

may be used as an individual confidence interval for the corresponding mean response. The associated confidence is the probability that the t distribution with $\nu = m2^p - u$ degrees of freedom assigns to the interval between $-t$ and t. And a one-sided confidence interval for the mean response can be obtained in the usual way, by employing only one of the endpoints indicated in formula (8.18) and appropriately adjusting the confidence level.

Example 4 (*continued*)

Consider estimating the mean dynamometer reading corresponding to a 15° bevel angle and interrupted cut using the "B and C main effects only" description of Miller's power requirement study. (These are the conditions that appear to produce the smallest mean power requirement.) Using (for example) 95% confidence, a fitted value of 26.02 from Table 8.11, and $s_{\text{FE}} = 1.409$ mm possessing $\nu = 4 \cdot 2^3 - 3 = 29$ associated degrees of freedom in formula (8.18), leads to a two-sided interval with endpoints

$$26.02 \pm 2.045(1.409)\sqrt{\frac{3}{4 \cdot 2^3}}$$

that is, endpoints

$$26.02 \text{ mm } \pm .88 \text{ mm} \tag{8.19}$$

that is,

$$25.14 \text{ mm} \quad \text{and} \quad 26.90 \text{ mm}$$

In contrast to this interval, consider what the method of Section 7.2 provides for a 95% confidence interval for the mean reading for tool type 1, a 15° bevel angle, and interrupted cuts. Since $s_{\text{P}} = 1.492$ with $\nu = 24$ associated degrees of freedom, and (from Table 8.9) $\bar{y}_{\text{c}} = 26.50$, formula (7.14) of Section 7.2 produces a two-sided confidence interval for μ_{c} with endpoints

$$26.50 \pm 2.064(1.492)\frac{1}{\sqrt{4}}$$

that is,

$$26.50 \text{ mm } \pm 1.54 \text{ mm} \tag{8.20}$$

A major practical difference between intervals (8.19) and (8.20) is the apparent increase in precision provided by interval (8.19), due in numerical terms primarily to the "extra" $\sqrt{3/8}$ factor present in the first plus-or-minus calculation but not in the second. However, it must be remembered that the extra precision is bought at the price of the use of model (8.14) and the consequent use of all observations in the generation of $\hat{y}_c$ (rather than only the observations from the single sample corresponding to combination c).

A second balanced-data confidence interval method based on a few-effects simplification of the general 2^p model is that for estimating the effects included in the model. It comes about by replacing s_P in formula (8.13) with s_{FE} and appropriately adjusting the degrees of freedom associated with the t quantile. That is, under a few-effects model in a 2^p study with balanced data, a two-sided individual confidence interval for an effect included in the model is

Balanced data individual confidence limits for a 2^p effect under a simplified model

$$\hat{E} \pm t\frac{s_{FE}}{\sqrt{m2^p}} \qquad \textbf{(8.21)}$$

where $\hat{E}$ is the corresponding fitted effect and the confidence associated with the interval is the probability that the t distribution with $\nu = m2^p - u$ degrees of freedom assigns to the interval between $-t$ and t. One-sided intervals are made from formula (8.21) in the usual way.

Unlike formula (8.13), formula (8.21) can be used in studies where $m = 1$. This makes it possible to attach precision figures to estimated effects in unreplicated factorial studies, provided one is willing to base them on a reduced or simplified model.

Example 6 (*continued*)

Consider again Daniel's drill advance rate study and, for example, the effect of the high level of rotational speed on the natural logarithm of advance rate. Under the "B, C, and D main effects only" description of log advance rate, $s_{FE} = .120$ with $\nu = 1 \cdot 2^4 - 4 = 12$ associated degrees of freedom. Also, $c_2 = .5772$. Then (for example) using a 95% confidence level, from formula (8.21), a two-sided interval for γ_2 under the simplified model has endpoints

$$.5772 \pm 2.179\frac{.120}{\sqrt{1 \cdot 2^4}}$$

that is,

$$.5772 \pm .0654$$

that is,

$$.5118 \quad \text{and} \quad .6426$$

Example 6 (*continued*)

This in turn translates (via multiplication by 2, since $\gamma_2 - \gamma_1 = 2\gamma_2$) to an increase of between

$$1.0236 \quad \text{and} \quad 1.2852$$

in average log advance rate as one moves from the low level of rotational speed to the high level. And upon exponentiation, a multiplication of median advance rate by a factor between

$$2.78 \quad \text{and} \quad 3.62$$

is indicated as one moves between levels of rotational speed. (A normal mean is also the distribution's median, and under a transformation the median of the transformed values is the transformation applied to the median. So the inference about the mean logged rate can be translated to one about the median rate. However, since the mean of transformed values is not in general the transformed mean, the interval obtained by exponentiation unfortunately does not apply to the mean advance rate.)

There are other ways to use the reduced model ideas discussed here. For example, a simplified model for responses can be used to produce prediction and tolerance intervals for individuals. Section 8.3 of Vardeman's *Statistics for Engineering Problem Solving* is one place to find an exposition of these additional methods.

Section 2 Exercises

1. Consider again the situation of Exercise 2 of Section 4.3.
   (a) For the logged responses, make individual 95% confidence intervals for the effects corresponding to the high levels of all three factors. Which effects are statistically detectable?
   (b) Fit an appropriate few-effects model suggested by your work in (a) to these data. Compare the corresponding value of s_{FE} to the value of s_{P}.
   (c) Compare a two-sided individual 95% confidence interval for the mean (logged) response for combination (1) made using the fitted few-effects model to one based on the methods of Section 7.2.
2. Chapter Exercise 9 in Chapter 4 concerns the making of Dual In-line Packages and the number of pullouts produced on such devices under 2^4 different combinations of manufacturing conditions. Return to that exercise, and if you have not already done so, use the Yates algorithm and compute fitted 2^4 factorial effects for the data set.
   (a) Use normal-plotting to identify statistically detectable effects here.
   (b) Based on your analysis from (a), postulate a possible few-effects model for this situation. Use the reverse Yates algorithm to fit such a model to these data. Use the fitted values to compute residuals. Normal-plot these and plot them against levels of each of the four factors, looking for obvious problems with the model.
   (c) Based on your few-effects model, make a recommendation for the future making of these devices. Give a 95% two-sided confidence interval (based on the few-effects model) for the mean pullouts you expect to experience if your advice is followed.
3. A classic unreplicated 2^4 factorial study, used as an example in *Experimental Statistics* (NBS Handbook

#91) by M. G. Natrella, concerns flame tests of fire-retardant treatments for cloth. The factors and levels used in the study were

A	Fabric Tested	sateen (−) vs. monk's cloth (+)
B	Treatment	X (−) vs. Y (+)
C	Laundering Condition	before (−) vs. after (+)
D	Direction of Test	warp (−) vs. fill (+)

The response variable, y, is the inches burned on a standard-size sample in the flame test. The data reported by Natrella follow:

Combination	y	Combination	y
(1)	4.2	d	4.0
a	3.1	ad	3.0
b	4.5	bd	5.0
ab	2.9	abd	2.5
c	3.9	cd	4.0
ac	2.8	acd	2.5
bc	4.6	bcd	5.0
abc	3.2	abcd	2.3

(a) Use the (four-cycle) Yates algorithm and compute the fitted 2^4 factorial effects for the study.

(b) Make either a normal plot or a half normal plot using the fitted effects from part (a). What subject-matter interpretation of the data is suggested by the plot? (See Chapter Exercise 9 regarding half normal-plotting.)

(c) Natrella's original analysis of these data produced the conclusion that both the A main effects and the AB two-factor interactions are statistically detectable and of practical importance. We (based on a plot like the one asked for in (b)) are inclined to doubt that the data are really adequate to detect the AB interaction. But for the sake of example, temporarily accept the conclusion of Natrella's analysis. What does it say in practical terms about the fire-retardant treating of cloth? (How would you explain the results to a clothing manufacturer?)

8.3 Standard Fractions of Two-Level Factorials, Part I: $\frac{1}{2}$ Fractions

The notion of a fractional factorial data structure was first introduced in Section 1.2. But as yet, this text has done little to indicate either how such a structure might be chosen or how analysis of fractional factorial data might proceed. The delay is a reflection of the subtle nature of these topics rather than any lack of importance. Indeed, fractional factorial experimentation and analysis is one of the most important tools in the modern engineer's kit. This is especially true where many factors potentially affect a response and there is little a priori knowledge about the relative impacts of these factors.

This section and the next treat the (standard) 2^{p-q} fractional factorials—the class of fractional factorials for which advantageous methods of data collection and analysis can be presented most easily and completely. These structures, involving $\frac{1}{2^q}$ of all possible combinations of levels of p two-level factors, are among the most useful fractional factorial designs for application in engineering experimentation. In addition, they clearly illustrate the general issues that arise any time only a fraction of a complete factorial set of factor-level combinations can be included in a multifactor study.

This section begins with some general qualitative remarks about fractional factorial experimentation. The standard $\frac{1}{2}$ fractions of 2^p studies (the 2^{p-1} fractional factorials) are then discussed in detail. The section covers in turn (1) the proper choice of such fractions, (2) the resultant aliasing or confounding patterns, and (3) corresponding methods of data analysis. The section closes with a few remarks about qualitative issues, addressed to the practical use of 2^{p-1} designs.

8.3.1 General Observations about Fractional Factorial Studies

In many of the physical systems engineers work on, there are many factors potentially affecting a response y. In such cases, even when the number of levels considered for each factor is only two, there are a huge number of different combinations of levels of the factors to consider. For instance, if $p = 10$ factors are considered, even when limiting attention to only two levels of each factor, at least $2^{10} = 1{,}024$ data points must be collected in order to complete a full factorial study. In most engineering contexts, restrictions on time and other resources would make a study of that size infeasible. One could try to guess which few factors are most important in determining the response and do a smaller complete factorial study on those factors (holding the levels of the remaining factors fixed). But there is obviously a risk of guessing wrong and therefore failing to discover the real pattern of how factors affect the response.

A superior alternative is to conduct the investigation in at least two stages. A relatively small screening study (or several of them), intended to identify those factors most likely influencing the response, can be done first. This can be followed up with a more detailed study (or studies) in those variables. It is in the initial screening phase of such a program that fractions of 2^p studies are most appropriate. Tools such as full factorials are appropriate for the later stage (or stages) of study.

Once the reality of resource limitations leads to consideration of fractional factorial experimentation, several qualitative points become clear. For one, there is no way to learn as much from a fraction of a full factorial study as from the full factorial itself. (There is no Santa Claus who for the price of eight observations will give as much information as can be obtained from 16.) Fractional factorial experiments inevitably leave some ambiguity in the interpretation of their results. Through careful planning of exactly which fraction of a full factorial to use, the object is to hold the ambiguity to a minimum and to make sure it is of a type that is most tolerable. Not all fractions of a given size from a particular full factorial study have the same potential for producing useful information.

Example 7 Choosing Half of a 2^2 Factorial Study

As a completely artificial but instructive example of the preceding points, suppose that two factors A and B each have two levels (low and high) and that instead of conducting a full 2^2 factorial study, data at only $\frac{1}{2}$ of the four possible combinations will be collected

(1), a, b, and ab

If (1) is chosen as one of the two combinations to be studied, two of the three possible choices of the other combination can easily be eliminated from consideration. The possibility of studying the combinations

(1) and a

is no good, since in both cases the factor B is held at its low level. Therefore, no information at all would be obtained on B's impact on the response. Similarly, the possibility of studying the combinations

(1) and b

can be eliminated, since no information would be obtained on factor A's impact on the response. So that leaves only the set of combinations

(1) and ab

as a $\frac{1}{2}$ fraction of the full 2^2 factorial that is at all sensible (if combination (1) is to be included). Similar reasoning eliminates all other pairs of combinations from potential use except the pair

a and b

But now notice that any experiment that includes only combinations

(1) and ab

or combinations

a and b

must inevitably produce somewhat ambiguous results. Since one moves from combination (1) to combination ab (or from a to b) by changing levels of *both* factors, if a large difference in response is observed, it will not be clear whether the difference is due to A or due to B.

At least in qualitative terms, such is the nature of all fractional factorial studies. Although very poor choices of experimental combinations may be avoided, some level of ambiguity must be accepted as the price for not conducting a full factorial.

Example 8

Half of a Hypothetical 2^3 Factorial

As a second hypothetical but instructive example of the issues that must be dealt with in fractional factorial experimentation, consider a system whose behavior is governed principally by the levels of three factors: A, B, and C. (For the sake of concreteness, suppose that A is a temperature, B is a pressure, and C is a catalyst type, and that the effects of these on the yield y of a chemical process are under consideration.) Suppose further that in a 2^3 study of this system, the factorial effects on an underlying mean response μ are given by

$$\begin{array}{llll} \mu_{...} = 10, & \alpha_2 = 3, & \beta_2 = 1, & \gamma_2 = 2, \\ \alpha\beta_{22} = 2, & \alpha\gamma_{22} = 0, & \beta\gamma_{22} = 0, & \alpha\beta\gamma_{222} = 0 \end{array}$$

Either through the use of the reverse Yates algorithm or otherwise, it is possible to verify that corresponding to these effects are then the eight combination means

$$\begin{array}{llll} \mu_{(1)} = 6, & \mu_{\text{a}} = 8, & \mu_{\text{b}} = 4, & \mu_{\text{ab}} = 14, \\ \mu_{\text{c}} = 10, & \mu_{\text{ac}} = 12, & \mu_{\text{bc}} = 8, & \mu_{\text{abc}} = 18 \end{array}$$

Now imagine that for some reason, only four of the eight combinations of levels of A, B, and C will be included in a study of this system, namely the combinations

a, b, c, and abc

Suppose further that the background noise is negligible, so that observations for a given treatment combination are essentially equal to the corresponding underlying mean. Then one essentially knows the values of

$$\mu_{\text{a}} = 8, \qquad \mu_{\text{b}} = 4, \qquad \mu_{\text{c}} = 10, \qquad \mu_{\text{abc}} = 18$$

Figure 8.10 shows the complete set of eight combination means laid out on a cube plot, with the four observed means circled.

As a sidelight, note the admirable symmetry possessed by the four circled corners on Figure 8.10. Each face of the cube has two circled corners (both levels of all factors appear twice in the choice of treatment combinations). Each edge has one circled corner (each combination of all pairs of factors appears once). And collapsing the cube in any one of the three possible directions (left to right, top to bottom, or front to back) gives a full factorial set of four combinations. (Ignoring the level of any one of A, B, or C in the four combinations a, b, c, and abc gives a full factorial in the other two factors.)

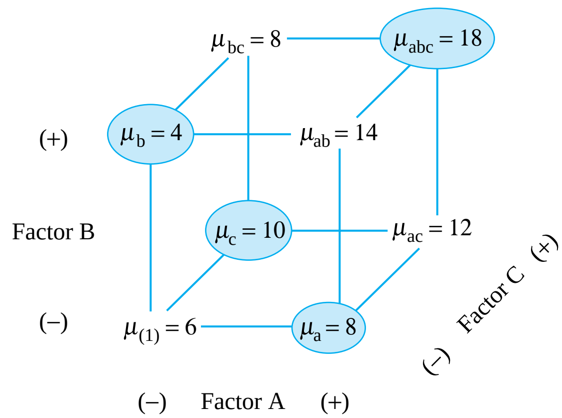

Figure 8.10 2^3 hypothetical means, with four known means circled

Now consider what an engineer possessing only the values of μ_a, μ_b, μ_c, and μ_{abc} might be led to conclude about the system. In particular, begin with the matter of evaluating an A main effect. Definition 3 says that

$$\alpha_2 = \mu_{2..} - \mu_{...}$$
$$= \left(\begin{array}{l}\text{the average of all four mean}\\ \text{responses where A is at its}\\ \text{second or high level}\end{array}\right) - \left(\begin{array}{l}\text{the grand average of all}\\ \text{eight mean responses}\end{array}\right)$$

which can be thought of as the right-face average minus the grand average for the cube in Figure 8.10. Armed only with the four means μ_a, μ_b, μ_c, and μ_{abc} (the four circled corners on Figure 8.10), it is not possible to compute α_2. But what might be done is to make a calculation similar to the one that produces α_2 using only the available means. That is,

$$\alpha_2^* = \text{a “}\tfrac{1}{2}\text{ fraction A main effect”}$$
$$= \left(\begin{array}{l}\text{the average of the available}\\ \text{two means where A is at its}\\ \text{high level}\end{array}\right) - \left(\begin{array}{l}\text{the grand average of the}\\ \text{available four means}\end{array}\right)$$
$$= \frac{1}{2}(\mu_a + \mu_{abc}) - \frac{1}{4}(\mu_a + \mu_b + \mu_c + \mu_{abc})$$
$$= \frac{1}{2}(8 + 18) - \frac{1}{4}(8 + 4 + 10 + 18)$$
$$= 13 - 10$$
$$= 3$$

Example 8 *(continued)*

And, amazingly enough, $\alpha_2^* = \alpha_2$ here.

It appears that using only four combinations, as much can be learned about the A main effect as if all eight combination means were in hand! This is too good to be true in general, as is illustrated by a parallel calculation for a C main effect.

$$\begin{aligned}\gamma_2^* &= \text{a “}\tfrac{1}{2}\text{ fraction C main effect”}\\ &= \left(\begin{array}{l}\text{the average of the two}\\ \text{available means where}\\ \text{C is at its high level}\end{array}\right) - \left(\begin{array}{l}\text{the grand average of the}\\ \text{four available means}\end{array}\right)\\ &= \frac{1}{2}(\mu_{\text{c}} + \mu_{\text{abc}}) - \frac{1}{4}(\mu_{\text{a}} + \mu_{\text{b}} + \mu_{\text{c}} + \mu_{\text{abc}})\\ &= 4\end{aligned}$$

while this hypothetical example began with $\gamma_2 = 2$. Here, the $\frac{1}{2}$ fraction calculation gives something quite different from the full factorial calculation.

The key to understanding how one can apparently get something for nothing in the case of the A main effects in this example, but cannot do so in the case of the C main effects, is to know that (in general) for this $\frac{1}{2}$ fraction,

$$\alpha_2^* = \alpha_2 + \beta\gamma_{22}$$

and

$$\gamma_2^* = \gamma_2 + \alpha\beta_{22}$$

Since this numerical example began with $\beta\gamma_{22} = 0$, one is “fortunate”—it turns out numerically that $\alpha_2^* = \alpha_2$. On the other hand, since $\alpha\beta_{22} = 2 \neq 0$, one is “unfortunate”—it turns out numerically that $\gamma_2^* = \gamma_2 + 2 \neq \gamma_2$.

Relationships like these for α_2^* and γ_2^* hold for all $\frac{1}{2}$ fraction versions of the full factorial effects. These relationships detail the nature of the ambiguity inherent in the use of the $\frac{1}{2}$ fraction of the full 2^3 factorial set of combinations. Essentially, based on data from four out of eight possible combinations, one will be unable to distinguish between certain pairs of effects, such as the A main effect and BC 2-factor interaction pair here.

8.3.2 Choice of Standard $\frac{1}{2}$ Fractions of 2^p Studies

Three fundamental issues in the use of a fractional factorial

The use of standard 2^{p-q} fractional factorial data structures depends on having answers for the following three basic questions:

1. How is $\frac{1}{2^q}$ of 2^p possible combinations of factor levels to include in a study rationally chosen?

2. How is the pattern of ambiguities implied by a given choice of 2^{p-q} combinations determined?

3. How is data analysis done for a particular choice of 2^{p-q} combinations?

These questions will be answered in this section for the case of $\frac{1}{2}$ fractions (2^{p-1} fractional factorials) and for general q in the next section.

Prescription for a best half fraction of a 2^p factorial

In order to arrive at what is in some sense a best possible choice of $\frac{1}{2}$ of 2^p combinations of levels of p factors, do the following. For the first $p-1$ factors, write out all 2^{p-1} possible combinations of these factors. By multiplying plus and minus signs (thinking of multiplying plus and minus 1's) corresponding to levels of the first factors, then arrive at a set of plus and minus signs that can be used to prescribe how to choose levels for the last factor (to be used in combination with the indicated levels of the first $p-1$ factors).

Example 9

A 2^{5-1} Chemical Process Experiment

In his article "Experimenting with a Large Number of Variables" (*ASQC Technical Supplement Experiments in Industry*, 1985), R. Snee discusses a successful 2^{5-1} experiment on a chemical process, where the response of interest, y, was a coded color index of the product. The factors studied and their levels are as in Table 8.15.

The standard recommendation for choosing a $\frac{1}{2}$ fraction was followed in Snee's study. Table 8.16 shows an appropriate set of 16 lines of plus and minus signs for generating the $\frac{1}{2} \cdot 32 = 16$ combinations included in Snee's study. The first four columns of this table specify levels of factors A, B, C, and D for the $16 = 2^4$ possible combinations of levels of these factors (written in Yates standard order). (The first line, for example, indicates the low level of all of these first four factors.) The last column of this table is obtained by multiplying the first four plus or minus signs (plus or minus 1's) in a given row. It is this last column that can be used to determine how to choose a level of factor E for use when the factors A through D are at the levels indicated in the first four columns.

Table 8.15
Five Chemical Process Variables and Their Experimental Levels

Factor	Process Variable	Factor Levels
A	Solvent/Reactant	low (−) vs. high (+)
B	Catalyst/Reactant	.025 (−) vs. .035 (+)
C	Temperature	150°C (−) vs. 160°C (+)
D	Reactant Purity	92% (−) vs. 96% (+)
E	pH of Reactant	8.0 (−) vs. 8.7 (+)

Example 9 (*continued*)

Table 8.16
Signs for Specifying a Standard 2^{5-1} Fractional Factorial

A	B	C	D	ABCD Product
−	−	−	−	+
+	−	−	−	−
−	+	−	−	−
+	+	−	−	+
−	−	+	−	−
+	−	+	−	+
−	+	+	−	+
+	+	+	−	−
−	−	−	+	−
+	−	−	+	+
−	+	−	+	+
+	+	−	+	−
−	−	+	+	+
+	−	+	+	−
−	+	+	+	−
+	+	+	+	+

In Snee's study, the signs in the ABCD Product column were used without modification to specify levels of E. The corresponding treatment combination names (written in the same order as in Table 8.16) and the data reported by Snee are given in Table 8.17. Notice that the 16 combinations listed in Table 8.17 are $\frac{1}{2}$ of the $2^5 = 32$ possible combinations of levels of these five factors. (They are those 16 that have an odd number of factors appearing at their high levels).

Example 10

A 2^{5-1} Agricultural Engineering Study

The article "An Application of Fractional Factorial Experimental Designs" by Mary Kilgo (*Quality Engineering*, 1988) provides an interesting complement to the previous example. In one part of an agricultural engineering study concerned with the use of carbon dioxide at very high pressures to extract oil from peanuts, the effects of five factors on a percent yield variable y were studied in a 2^{5-1} fractional factorial experiment. The five factors and their levels (as named in Kilgo's article) are given in Table 8.18.

Interestingly enough, rather than studying the 16 combinations obtainable using the final column of Table 8.16 directly, Kilgo *switched* all of the signs in the ABCD product column before assigning levels of E. This leads to the use of "the other" 16 out of 32 possible combinations (those having an even number of

Table 8.17
16 Combinations and Observed Color Indices in Snee's 2^{5-1} Study (*Example 9*)

Combination	Color Index, y
e	−.63
a	2.51
b	−2.68
abe	−1.66
c	2.06
ace	1.22
bce	−2.09
abc	1.93
d	6.79
ade	6.47
bde	3.45
abd	5.68
cde	5.22
acd	9.38
bcd	4.30
abcde	4.05

Table 8.18
Five Peanut Processing Variables and Their Experimental Levels

Factor	Process Variable	Factor Levels
A	Pressure	415 bars (−) vs. 550 bars (+)
B	Temperature	25°C (−) vs. 95°C (+)
C	Peanut Moisture	5% (−) vs. 15% (+)
D	Flow Rate	40 1/min (−) vs. 60 1/min (+)
E	Average Particle Size	1.28 mm (−) vs. 4.05 mm (+)

factors appearing at their high levels). The 16 combinations studied and corresponding responses reported by Kilgo are given in Table 8.19 in the same order for factors A through D as in Table 8.16.

The difference between the combinations listed in Tables 8.17 and 8.19 deserves some thought. As Kilgo named the factor levels, the two lists of combinations are quite different. But verify that if she had made the slightly less natural but nevertheless permissible choice to call the 4.05 mm level of factor E the low (−) level

Table 8.19
16 Combinations and Observed Yields in Kilgo's 2^{5-1} Study (*Example 10*)

Combination	Yield, y (%)
(1)	63
ae	21
be	36
ab	99
ce	24
ac	66
bc	71
abce	54
de	23
ad	74
bd	80
abde	33
cd	63
acde	21
bcde	44
abcd	96

and the 1.28 mm level the high (+) level, the names of the physical combinations actually studied would be exactly those in Table 8.17 rather than those in Table 8.19.

The point here is that due to the rather arbitrary nature of how one chooses to name high and low levels of two factors, the names of different physical combinations are themselves to some extent arbitrary. In choosing fractional factorials, one chooses some particular naming convention and then has the freedom to choose levels of the last factor (or factors for $q > 1$ cases) by either using the product column(s) directly or after switching signs. The decision whether or not to switch signs *does affect* exactly which physical combinations will be run and thus how the data should be interpreted in the subject-matter context. But generally, the different possible choices (to switch or not switch signs) are a priori equally attractive. For systems that happen to have relatively simple structure, all possible results of these arbitrary choices typically lead to similar engineering conclusions. When systems turn out to have complicated structures, the whole notion of fractional factorial experimentation loses its appeal. Different arbitrary choices lead to different perceptions of system behavior, none of which (usually) correctly portrays the complicated real situation.

Fractional factorials fully reveal system structure only for simple cases

8.3.3 Aliasing in the Standard $\frac{1}{2}$ Fractions

Once a $\frac{1}{2}$ fraction of a 2^p study is chosen, the next issue is determining the nature of the ambiguities that must arise from its use. For 2^{p-1} data structures of the type described here, one can begin with a kind of statement of how the fractional factorial plan was derived and through a **system of formal multiplication** arrive at an understanding of which (full) factorial effects cannot be separated on the basis of the fractional factorial data. Some terminology is given next, in the form of a definition.

Definition 7

When it is only possible to estimate the sum (or difference) of two or more (full) factorial effects on the basis of data from a fractional factorial, those effects are said to be **aliased** or **confounded** and are sometimes called **aliases**. In this text, the phrase **alias structure** of a fractional factorial plan will mean a complete specification of all sets of aliased effects.

As an example of the use of this terminology, return to Example 8. There, it is possible only to estimate $\alpha_2 + \beta\gamma_{22}$, not either of α_2 or $\beta\gamma_{22}$ individually. So the A main effect is confounded with (or aliased with) the BC 2-factor interaction.

The way the system of formal multiplication works for detailing the alias structure of one of the recommended 2^{p-1} factorials is as follows. One begins by writing

Generator for a standard half fraction of a 2^p factorial

$$\left(\begin{array}{l}\text{the name of the}\\ \text{last factor}\end{array}\right) \leftrightarrow \pm \left(\begin{array}{l}\text{the product of names of}\\ \text{the first } p-1 \text{ factors}\end{array}\right) \qquad \textbf{(8.22)}$$

where the plus or minus sign is determined by whether the signs were left alone or switched in the specification of levels of the last factor. The double arrow in expression (8.22) will be read as "is aliased with." And since expression (8.22) really says how the fractional factorial under consideration was chosen, expression (8.22) will be called the plan's **generator**. The generator (8.22) for a 2^{p-1} plan says that the (high level) main effect of the last factor will be aliased with plus or minus the (all factors at their high levels) $p-1$ factor interaction of the first $p-1$ factors.

Example 9 (*continued*)

In Snee's 2^{5-1} study, the generator

$$\text{E} \leftrightarrow \text{ABCD}$$

was used. Therefore the (high level) E main effect is aliased with the (all high levels) ABCD 4-factor interaction. That is, only $\epsilon_2 + \alpha\beta\gamma\delta_{2222}$ can be estimated based on the $\frac{1}{2}$ fraction data, not either of its summands individually.

Example 10 (*continued*)

In Kilgo's 2^{5-1} study, the generator

$$\text{E} \leftrightarrow -\text{ABCD}$$

Example 10 (continued)

was used. The (high level) E main effect is aliased with minus the (all high levels) ABCD 4-factor interaction. That is, only $\epsilon_2 - \alpha\beta\gamma\delta_{2222}$ can be estimated based on the $\frac{1}{2}$ fraction data, not either of the terms individually.

Conventions for the system of formal multiplication

The entire alias structure for a $\frac{1}{2}$ fraction follows from the generator (8.22) by multiplying both sides of the expression by various factor names, using two special conventions. These are that any letter multiplied by itself produces the symbol "I" and that any letter multiplied by "I" is that letter again. Applying the first of these conventions to expression (8.22), both sides of the expression may be multiplied by the name of the last factor to produce the relation

Defining relation for a standard half fraction of a 2^p factorial

$$\text{I} \leftrightarrow \pm \text{ the product of names of all } p \text{ factors} \tag{8.23}$$

Expression (8.23) means that the grand mean is aliased with plus or minus the (all factors at their high level) p-factor interaction. There is further special terminology for an expression like that in display (8.23).

Definition 8

The list of all aliases of the grand mean for a 2^{p-q} fractional factorial is called the **defining relation** for the design.

By first translating a generator (or generators in the case of $q > 1$) into a defining relation and then multiplying through the defining relation by a product of letters corresponding to an effect of interest, one can identify all aliases of that effect.

Example 9 (continued)

In Snee's 2^{5-1} experiment, the generator was

$$\text{E} \leftrightarrow \text{ABCD}$$

When multiplied through by E, this gives the experiment's defining relation

$$\text{I} \leftrightarrow \text{ABCDE} \tag{8.24}$$

which indicates that the grand mean $\mu_{.....}$ is aliased with the 5-factor interaction $\alpha\beta\gamma\delta\epsilon_{22222}$. Then, for example, multiplying through defining relation (8.24) by the product AC produces the relationship

$$\text{AC} \leftrightarrow \text{BDE}$$

Thus, the AC 2-factor interaction is aliased with the BDE 3-factor interaction. In fact, the entire alias structure for the Snee study can be summarized in terms of the aliasing of 16 different pairs of effects. These are indicated in Table 8.20, which was developed by using the defining relation (8.24) to find successively (in Yates order) the aliases of all effects involving only factors A, B, C, and D. Table 8.20 shows that main effects are confounded with 4-factor interactions and 2-factor interactions with 3-factor interactions. This degree of ambiguity is as mild as is possible in a 2^{5-1} study.

Table 8.20
The Complete Alias Structure for Snee's 2^{5-1} Study

I ↔ ABCDE	D ↔ ABCE
A ↔ BCDE	AD ↔ BCE
B ↔ ACDE	BD ↔ ACE
AB ↔ CDE	ABD ↔ CE
C ↔ ABDE	CD ↔ ABE
AC ↔ BDE	ACD ↔ BE
BC ↔ ADE	BCD ↔ AE
ABC ↔ DE	ABCD ↔ E

Example 10 *(continued)*

In Kilgo's peanut oil extraction study, since the generator is E ↔ −ABCD, the defining relation is I ↔ −ABCDE, and the alias structure is that given in Table 8.20, except that a minus sign should be inserted on one side or the other of every row of the table. So, for example, $\alpha\beta_{22} - \gamma\delta\epsilon_{222}$ may be estimated based on Kilgo's data, but neither $\alpha\beta_{22}$ nor $\gamma\delta\epsilon_{222}$ separately.

8.3.4 Data Analysis for 2^{p-1} Fractional Factorials

Once the alias structure of a 2^{p-1} fractional factorial is understood, the question of how to analyze data from such a study has a simple answer.

1. Temporarily ignore the last factor and compute the estimated or fitted "effects."
2. Somehow judge the statistical significance and apparent real importance of the "effects" computed for the complete factorial in $p-1$ two-level factors. (Where some replication is available, the judging of statistical significance can be done through the use of confidence intervals. Where all 2^{p-1} samples are of size 1, the device of normal-plotting fitted "effects" is standard.)
3. Finally, seek a plausible simple interpretation of the important fitted "effects," recognizing that they are estimates not of the effects in the first $p-1$ factors alone, but of *those effects plus their aliases*.

Example 9 *(continued)*

Consider the analysis of Snee's data, listed in Table 8.17 in Yates standard order for factors A, B, C, and D (ignoring the existence of factor E). Then, according to the prescription for analysis just given, the first step is to use the Yates algorithm (for four factors) on the data. These calculations are summarized in Table 8.21.

Each entry in the final column of Table 8.21 gives the name of the effect that the corresponding numerical value in the "Cycle 4 ÷ 16" column would be estimating if factor E weren't present, *plus* the alias of that effect. The numbers in the next-to-last column must be interpreted in light of the fact that they are estimating sums of 2^5 factorial effects.

Since there is no replication indicated in Table 8.17, only normal-plotting fitted (sums of) effects is available to identify those that are distinguishable from noise. Figure 8.11 is a normal plot of the last 15 entries of the Cycle 4 ÷ 16 column of Table 8.21. (Since in most contexts one is a priori willing to grant that the overall mean response is other than 0, the estimate of it plus its alias(es) is rarely included in such a plot.)

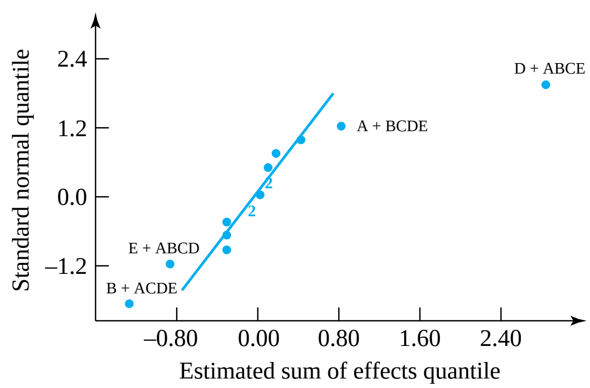

Figure 8.11 Normal plot of estimated sums of effects in Snee's 2^{5-1} study

Depending upon how the line is drawn through the small estimated (sums of) effects in Figure 8.11, the estimates corresponding to D + ABCE, and possibly B + ACDE, E + ABCD, and A + BCDE as well, are seen to be distinguishable in magnitude from the others. (The line in Figure 8.11 has been drawn in keeping with the view that there are four statistically detectable sums of effects, primarily because a half normal plot of the absolute values of the estimates—not included here—supports that view.) If one adopts the view that there are indeed four detectable (sums of) effects indicated by Figure 8.11, it is clear that the simplest possible interpretation of this outcome is that the four large estimates are each reflecting primarily the corresponding main effects (and not the aliased 4-factor

Table 8.21
The Yates Algorithm for a 2^4 Factorial Applied to Snee's 2^{5-1} Data

y	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 4 ÷ 16	Sum Estimated
−.63	1.88	−2.46	.66	46.00	2.875	$\mu_{.....} + \alpha\beta\gamma\delta\epsilon_{22222}$
2.51	−4.34	3.12	45.34	13.16	.823	$\alpha_2 + \beta\gamma\delta\epsilon_{2222}$
−2.68	3.28	22.39	7.34	−20.04	−1.253	$\beta_2 + \alpha\gamma\delta\epsilon_{2222}$
−1.66	−.16	22.95	5.82	.88	.055	$\alpha\beta_{22} + \gamma\delta\epsilon_{222}$
2.06	13.26	4.16	−9.66	6.14	.384	$\gamma_2 + \alpha\beta\delta\epsilon_{2222}$
1.22	9.13	3.18	−10.38	1.02	.064	$\alpha\gamma_{22} + \beta\delta\epsilon_{222}$
−2.09	14.60	1.91	2.74	.66	.041	$\beta\gamma_{22} + \alpha\delta\epsilon_{222}$
1.93	8.35	3.91	−1.86	.02	.001	$\alpha\beta\gamma_{222} + \delta\epsilon_{22}$
6.79	3.14	−6.22	5.58	44.68	2.793	$\delta_2 + \alpha\beta\gamma\epsilon_{2222}$
6.47	1.02	−3.44	.56	−1.52	−.095	$\alpha\delta_{22} + \beta\gamma\epsilon_{222}$
3.45	−.84	−4.13	−.98	−.72	−.045	$\beta\delta_{22} + \alpha\gamma\epsilon_{222}$
5.68	4.02	−6.25	2.00	−4.60	−.288	$\alpha\beta\delta_{222} + \gamma\epsilon_{22}$
5.22	−.32	−2.12	2.78	−5.02	−.314	$\gamma\delta_{22} + \alpha\beta\epsilon_{222}$
9.38	2.23	4.86	−2.12	2.98	.186	$\alpha\gamma\delta_{222} + \beta\epsilon_{22}$
4.30	4.16	2.55	6.98	−4.90	−.306	$\beta\gamma\delta_{222} + \alpha\epsilon_{22}$
4.05	−.25	−4.41	−6.96	−13.94	−.871	$\alpha\beta\gamma\delta_{2222} + \epsilon_2$

interactions). That is, a tentative (because of the incomplete nature of fractional factorial information) description of the chemical process is that D (reactant purity), B (catalyst/reactant), A (solvent/reactant), and E (pH of reactant) main effects are (in that order) the principal determinants of product color. Depending on the engineering objectives for product color index y, this tentative description of the system could have several possible interpretations. If large y were desirable, the high levels of A and D and low levels of B and E appear most attractive. If small y were desirable, the situation would be reversed. But in fact, Snee's study was done not to figure out how to maximize or minimize y, but rather to determine how to reduce variation in y. The engineering implications of the "D, B, A, and E main effects only" system description are thus to focus attention on the need to control variation first in level of factor D (reactant purity), then in level of factor B (catalyst/reactant), then in level of factor A (solvent/reactant), and finally in level of factor E (pH of reactant).

Example 10 *(continued)*

Verify that for Kilgo's data in Table 8.19, use of the (four-cycle) Yates algorithm on the data as listed (in standard order for factors A, B, C, and D, ignoring factor E) produces the estimated (differences of) effects given in Table 8.22.

Example 10 *(continued)*

Table 8.22
Estimated Differences of 2^5 Factorial Effects from Kilgo's 2^{5-1} Study

Value	Difference Estimated	Value	Difference Estimated
54.3	$\mu_{.....} - \alpha\beta\gamma\delta\epsilon_{22222}$	0.0	$\delta_2 - \alpha\beta\gamma\epsilon_{2222}$
3.8	$\alpha_2 - \beta\gamma\delta\epsilon_{2222}$	−2.0	$\alpha\delta_{22} - \beta\gamma\epsilon_{222}$
9.9	$\beta_2 - \alpha\gamma\delta\epsilon_{2222}$	−.9	$\beta\delta_{22} - \alpha\gamma\epsilon_{222}$
2.6	$\alpha\beta_{22} - \gamma\delta\epsilon_{222}$	−3.1	$\alpha\beta\delta_{222} - \gamma\epsilon_{22}$
.6	$\gamma_2 - \alpha\beta\delta\epsilon_{2222}$	1.1	$\gamma\delta_{22} - \alpha\beta\epsilon_{222}$
.6	$\alpha\gamma_{22} - \beta\delta\epsilon_{222}$	.1	$\alpha\gamma\delta_{222} - \beta\epsilon_{22}$
1.5	$\beta\gamma_{22} - \alpha\delta\epsilon_{222}$	3.5	$\beta\gamma\delta_{222} - \alpha\epsilon_{22}$
1.8	$\alpha\beta\gamma_{222} - \delta\epsilon_{22}$	22.3	$\alpha\beta\gamma\delta_{2222} - \epsilon_2$

The last 15 of these estimated differences are normal-plotted in Figure 8.12. It is evident from the figure that the two estimated (differences of) effects corresponding to

$$\beta_2 - \alpha\gamma\delta\epsilon_{2222} \quad \text{and} \quad \alpha\beta\gamma\delta_{2222} - \epsilon_2$$

are significantly larger than the other 13 estimates. The simplest possible interpretation of this outcome is that the two large estimates are each reflecting primarily the corresponding main effects (not the aliased 4-factor interactions). That is, a tentative description of the oil extraction process is that average particle size (factor E) and temperature (factor B), acting more or less separately, are the principle determinants of yield. This is an example where the ultimate engineering objective is to maximize response and the two large estimates are both positive. So, for best yield one would prefer the high level of B (95°C temperature) and

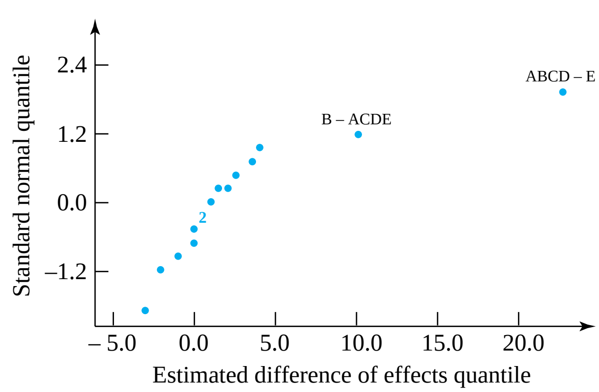

Figure 8.12 Normal plot of estimated differences of effects in Kilgo's 2^{5-1} study

low level of E (1.28 mm particle size). ($-\epsilon_2$ is apparently positive, and since $\epsilon_1 = -\epsilon_2$, the superiority of the low level of E is indicated.)

8.3.5 Some Additional Comments

The next section treats general $\frac{1}{2^q}$ fractions of 2^p factorials. But before closing this discussion of the special case of $q = 1$, several issues deserve comment. The first concerns the range of statistical methods that will be provided here for use with fractional factorials. The data analysis methods presented in this section and the next are confined to those for the identification of potential "few effects" descriptions of a p-factor situation. (For example, we do not go on to issues of inference under such a reduced model.) This stance is consistent with the fact that fractional factorials are primarily screening devices, useful for gaining some idea about which of many factors might be important. They are typically not suited (at least without additional data collection) to serve as the basis for detailed modeling of a response. The insights they provide must be seen as tentative and as steps along a path of learning about what factors influence a response.

A second matter regards the sense in which the $\frac{1}{2}$ fractions recommended here are the best ones possible. Other $\frac{1}{2}$ fractions could be developed (essentially by using a product column of signs derived from levels of fewer than all $p - 1$ of the first factors to assign levels of the last one). But the alias structures associated with those alternatives are less attractive than the ones encountered in this section. That is, here main effects have been aliased with $p - 1$ factor interactions, 2-factor interactions with $p - 2$-factor interactions, and so on. Any other $\frac{1}{2}$ fractions fundamentally different from the ones discussed here would have main effects aliased with interactions of $p - 2$ or less factors. They would thus be more likely to produce data incapable of separating important effects. The "l order effects aliased with $p - l$ order effects" structure of this section is simply the best one can do with a 2^{p-1} fractional factorial.

The last matter for discussion concerns what directions a follow-up investigation might take in order to resolve ambiguities left after a 2^{p-1} study is completed. Sometimes several different simple descriptions of system structure remain equally plausible after analysis of an initial $\frac{1}{2}$ fraction of a full factorial study. One approach to resolving these is to complete the factorial and "run the other $\frac{1}{2}$ fraction."

Example 11 A 2^{4-1} Fabric Tenacity Study Followed Up by a Second 2^{4-1} Study

Researchers Johnson, Clapp, and Baqai, in "Understanding the Effect of Confounding in Design of Experiments: A Case Study in High Speed Weaving" (*Quality Engineering*, 1989), discuss a study done to evaluate the effects of four two-level factors on a measure of woven fabric tenacity. The factors that were studied are indicated in Table 8.23.

Example 11 (*continued*)

Table 8.23
Four Weaving Process Variables and Their Experimental Levels

Factor	Weaving Process Variable	Factor Levels
A	Side of Cloth (l. to r.)	nozzle side (−) vs. opposite side (+)
B	Yarn Type	air spun (−) vs. ring spun (+)
C	Pick Density	35 ppi (−) vs. 50 ppi (+)
D	Air Pressure	30 psi (−) vs. 45 psi (+)

Factor A reflects the left-to-right location on the fabric width from which a tested sample is taken. Factor C reflects a count of yarns per inch inserted in the cloth, top to bottom, during weaving. Factor D reflects the air pressure used to propel the yarn across the fabric width during weaving.

Initially, a replicated 2^{4-1} study was done using the generator D $\leftrightarrow$ ABC. $m = 5$ pieces of cloth were tested for each of the eight different factor-level combinations studied. The resulting mean fabric tenacities $\bar{y}$, expressed in terms of strength per unit linear density, are given in Table 8.24. Although it is not absolutely clear in the article, it also appears that pooling the eight s^2 values from the $\frac{1}{2}$ fraction gave $s_{\text{P}} \approx 1.16$.

Apply the (three-cycle) Yates algorithm to the means listed in Table 8.24 (in the order given) and verify that the estimated sums of effects corresponding to the means in Table 8.24 are those given in Table 8.25.

Temporarily ignoring the existence of factor D, confidence intervals based on these estimates can be made using the $m = 5$ and $p = 3$ version of formula (8.13) from Section 8.2. That is, using 95% two-sided individual confidence intervals, since $\nu = 8(5 - 1) = 32$ degrees of freedom are associated with s_{P}, a precision of roughly

$$\pm\frac{(2.04)(1.16)}{\sqrt{5 \cdot 8}} = \pm .375$$

should be associated with each of the estimates in Table 8.25. By this standard, the estimates corresponding to the A + BCD, AB + CD, C + ABD, and BC + AD

Table 8.24
Eight Sample Means from a 2^{4-1} Fabric Tenacity Experiment

Combination	$\bar{y}$ (g/den.)	Combination	$\bar{y}$ (g/den.)
(1)	24.50	cd	25.68
ad	22.05	ac	24.51
bd	24.52	bc	24.68
ab	25.00	abcd	24.23

Table 8.25
Estimated Sums of 2^4 Effects in a 2^{4-1} Fabric Weaving Experiment

Estimate	Sum of Effects Estimated
24.396	$\mu_{....} + \alpha\beta\gamma\delta_{2222}$
−.449	$\alpha_2 + \beta\gamma\delta_{222}$
.211	$\beta_2 + \alpha\gamma\delta_{222}$
.456	$\alpha\beta_{22} + \gamma\delta_{22}$
.379	$\gamma_2 + \alpha\beta\delta_{222}$
.044	$\alpha\gamma_{22} + \beta\delta_{22}$
−.531	$\beta\gamma_{22} + \alpha\delta_{22}$
−.276	$\alpha\beta\gamma_{222} + \delta_2$

sums are statistically significant. Two reasonably plausible and equally simple tentative interpretations of this outcome are

1. There are detectable A and C main effects and detectable 2-factor interactions of A with B and D.
2. There are detectable A and C main effects and detectable 2-factor interactions of C with B and D.

(For that matter, there are others that you may well find as plausible as these two.)

In any case, the ambiguities left by the collection of the data summarized in Table 8.24 were unacceptable. To remedy the situation, the authors subsequently completed the 2^4 factorial study by collecting data from the other eight combinations defined by the generator D $\leftrightarrow$ −ABC. The means they obtained are given in Table 8.26.

One should honestly consider (and hopefully eliminate) the possibility that there is a systematic difference between the values in Table 8.24 and in Table 8.26 as a result of some unknown factor or factors that changed in the time lapse between the collection of the first block of observations and the second block. If

Table 8.26
Eight More Sample Means from a Second 2^{4-1} Fabric Tenacity Study

Combination	$\bar{y}$	Combination	$\bar{y}$
d	23.73	c	24.63
a	23.55	acd	25.78
b	25.98	bcd	24.10
abd	23.64	abc	23.93

Example 11 (*continued*)

that possibility can be eliminated, it would make sense to put together the two data sets, treat them as a single full 2^4 factorial data set, and employ the methods of Section 8.2 in their analysis. (Some repetition of a combination or combinations included in the first study phase—e.g., the center point of the design—would have been advisable to allow at least a cursory check on the possibility of a systematic block effect.)

Johnson, Clapp, and Baqai don't say explicitly what sample sizes were used to produce the $\bar{y}$'s in Table 8.26. (Presumably, $m = 5$ was used.) Nor do they give a value for s_P based on all 2^4 samples, so it is not possible to give a complete analysis of the full factorial data à la Section 8.2. But it is possible to note what results from the use of the Yates algorithm with the full factorial set of $\bar{y}$'s. This is summarized in Table 8.27.

Table 8.27
Fitted Effects from the Full 2^4 Factorial Fabric Tenacity Study

Effect	Estimate	Effect	Estimate
$\mu_{....}$	$\bar{y}_{....} = 24.407$	δ_2	$d_2 = -.191$
α_2	$a_2 = -.321$	$\alpha\delta_{22}$	$ad_{22} = .029$
β_2	$b_2 = .103$	$\beta\delta_{22}$	$bd_{22} = -.197$
$\alpha\beta_{22}$	$ab_{22} = .011$	$\alpha\beta\delta_{222}$	$abd_{222} = .093$
γ_2	$c_2 = .286$	$\gamma\delta_{22}$	$cd_{22} = .446$
$\alpha\gamma_{22}$	$ac_{22} = .241$	$\alpha\gamma\delta_{222}$	$acd_{222} = .108$
$\beta\gamma_{22}$	$bc_{22} = -.561$	$\beta\gamma\delta_{222}$	$bcd_{222} = -.128$
$\alpha\beta\gamma_{222}$	$abc_{222} = -.086$	$\alpha\beta\gamma\delta_{2222}$	$abcd_{2222} = -.011$

The statistical significance of the entries of Table 8.27 will not be judged here. But note that the picture of fabric tenacity given by the fitted effects in this table is somewhat more complicated than either of the tentative descriptions derived from the original 2^{4-1} study. The fitted effects, listed in order of decreasing absolute value, are

$$\text{BC, CD, A, C, AC, BD, D, } \ldots \text{, etc.}$$

Although tentative description (2) (page 609) accounts for the first four of these, the A and C main effects indicated in Table 8.27 are not really as large as one might have guessed looking only at Table 8.25. Further, the AC 2-factor interaction appears from Table 8.27 to be nearly as large as the C main effect. This is obscured in the original 2^{4-1} fractional factorial because the AC 2-factor interaction is aliased with an apparently fairly large BD 2-factor interaction *of opposite sign*.

Ultimately, this example is one of a fairly complicated system of effects. It admirably illustrates the difficulties and even errors of interpretation that can arise when only fractional factorial data are available for use in studying such systems.

In conclusion, it should be said that when a 2^{p-1} fractional factorial seems to leave only very mild ambiguities of interpretation, it can be possible to resolve those with the use of only a *few* additional data points (rather than requiring the addition of the entire other $\frac{1}{2}$ fraction of combinations). But this is a more advanced topic than is sensibly discussed here. The interested reader can refer to Chapter 14 of Daniel's *Applications of Statistics to Industrial Experimentation* for an illuminating discussion of this matter.

Section 3 Exercises

1. In a 2^{5-1} study with defining relation I $\leftrightarrow$ ABCDE, it is possible for both the A main effect and the BCDE 4-factor interaction to be of large magnitude but for both of them to go undetected. How might this quite easily happen?

2. The paper "How to Optimize and Control the Wire Binding Process: Part I" by Scheaffer and Levine (*Solid State Technology*, November 1990) contains the results of a 2^{5-1} fractional factorial experiment with additional repeated center point, run in an effort to determine how to improve the operation of a K&S Model 1484 XQ wire bonder. The generator E $\leftrightarrow$ ABCD was used in setting up the 2^{5-1} part of the experiment involving the factors and levels indicated in the accompanying table.

Factor A	Constant Velocity	.6 in./sec (−) vs. 1.2 in./sec (+)
Factor B	Temperature	150°C (−) vs. 200°C (+)
Factor C	Bond Force	80 g (−) vs. 120 g (+)
Factor D	Ultrasonic Power	120 mW (−) vs. 200 mW (+)
Factor E	Bond Time	10 ms (−) vs. 20 ms (+)

The response variable, y, was a force (in grams) required to pull wire bonds made on the machine under a particular combination of levels of the factors. (Each y was actually an average of the pull forces required on a 30 lead test sample.) The responses from the 2^{5-1} part of the study were as follows:

Combination	y	Combination	y
e	8.5	d	5.8
a	7.9	ade	8.0
b	7.7	bde	7.8
abe	8.7	abd	8.7
c	9.0	cde	6.9
ace	9.2	acd	8.5
bce	8.6	bcd	8.6
abc	9.5	abcde	8.3

In addition, three runs were made at a constant velocity of .9 in./sec, a temperature of 175°C, a bond force of 100 g, a power of 160 mW, and a bond time of 15 ms. These produced y values of 8.1, 8.6, and 8.1.

(a) Place the 16 observations from the 2^{5-1} part of the experiment in Yates standard order as regards levels of factors A through D. Use the four-cycle Yates algorithm to compute fitted sums of 2^5 effects. Identify what sum of effects each of these estimates. (For example, the first estimates $\mu_{.....} + \alpha\beta\gamma\delta\epsilon_{22222}$.)

(b) The three center points can be thought of as providing a pooled sample variance here. You may verify that $s_{\text{P}} = .29$. If one then wishes to make confidence intervals for the sums of effects, it is possible to use the $m = 1$, $p = 4$, and

$\nu = 2$ version of formula (8.13) of Section 8.2. What is the plus-or-minus value that comes from this program, for individual 95% two-sided confidence intervals? Using this value, which of the fitted sums of effects would you judge to be statistically detectable? Does this list suggest to you any particularly simple/intuitive description of how bond strength depends on the levels of the five factors?

(c) Based on your analysis from (b), if you had to guess what levels of the factors A, C, and D should be used for high bond strength, what would you recommend? If the CE + ABD fitted sum reflects primarily the CE 2-factor interaction, what level of E then seems best? Which of the combinations actually observed had these levels of factors A, C, D, and E? How does its response compare to the others?

3. Return to the fire retardant flame test study of Exercise 3 of Section 8.2. The original study, summarized in that exercise, was a full 2^4 factorial study.

(a) If you have not done so previously, use the (four-cycle) Yates algorithm and compute the fitted 2^4 factorial effects for the study. Normal-plot these. What subject-matter interpretation of the data is suggested by the normal plot?

Now suppose that instead of a full factorial study, only the $\frac{1}{2}$ fraction with generator D $\leftrightarrow$ ABC had been conducted.

(b) Which 8 of the 16 treatment combinations would have been run? List these combinations in Yates standard order as regards factors A, B, and C and use the (three-cycle) Yates algorithm to compute the 8 estimated sums of effects that it is possible to derive from these 8 treatment combinations. Verify that each of these 8 estimates is the sum of two of your fitted effects from part (a). (For example, you should find that the first estimated sum here is $\bar{y}_{....} + abcd_{2222}$ from part (a).)

(c) Normal-plot the last 7 of the estimated sums from (b). Interpret this plot. If you had only the data from this 2^{4-1} fractional factorial, would your subject-matter conclusions be the same as those reached in part (a), based on the full 2^4 data set?

8.4 Standard Fractions of Two-Level Factorials Part II: General 2^{p-q} Studies

Section 8.3 began the study of fractional factorials with the $\frac{1}{2}$ fractions of 2^p factorials, considering in turn the issues of (1) choice, (2) determination of the corresponding alias structure, and (3) data analysis. The approaches used to treat 2^{p-1} studies extend naturally to the smaller $\frac{1}{2^q}$ fractions of 2^p factorials for $q > 1$.

This section first shows how the ideas of Section 8.3 are generalized to cover the general 2^{p-q} situation. Then it considers the notion of design resolution and its implications for comparing alternative possible 2^{p-q} plans. Next an introduction is given to how the 2^{p-q} ideas can be employed where a blocking variable (or variables) dictate the use of a number of blocks equal to a power of 2. The section concludes with some comments regarding wise use of this 2^{p-q} material.

8.4.1 Using 2^{p-q} Fractional Factorials

The recommended method of choosing a $\frac{1}{2}$ fraction of a 2^p factorial uses a column of signs developed as products of plus and minus signs for *all* of the first $p - 1$ factors. The key to understanding how the ideas of the previous section generalize to $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$, etc. fractions of 2^p studies is to realize that there are several possible similar columns that could be developed using only *some* of the first $p-1$ factors. When moving from $\frac{1}{2}$ fractions to $\frac{1}{2^q}$ fractions of 2^p factorials, one makes use of such columns in assigning levels of the last q factors and then develops and uses an alias structure consistent with the choice of columns.

For example, first consider the situation for cases where $p-q=3$—that is, where $2^3=8$ different combinations of levels of p two-level factors are going to be included in a study. A table of signs specifying all eight possible combinations of levels of the first three factors A, B, and C, with four additional columns made up as the possible products of the first three columns, is given in Table 8.28.

Choosing a 2^{p-q} fractional factorial with $p-q=3$

The final column of Table 8.28 can be used to choose levels of factor D for a best possible 2^{4-1} fractional factorial study. But it is also true that two or more of the product columns in Table 8.28 can be used to choose levels of several additional factors (beyond the first three). If this is done, one winds up with a fractional factorial that can be understood in the same ways it is possible to make sense of the standard 2^{p-1} data structures discussed in Section 8.3.

Table 8.28
Signs for Specifying all Eight Combinations of Three Two-Level Factors and Four Sets of Products of Those Signs

A	B	C	AB Product	AC Product	BC Product	ABC Product
−	−	−	+	+	+	−
+	−	−	−	−	+	+
−	+	−	−	+	−	+
+	+	−	+	−	−	−
−	−	+	+	−	−	+
+	−	+	−	+	−	−
−	+	+	−	−	+	−
+	+	+	+	+	+	+

Example 12

A 2^{6-3} Propellant Slurry Study

The text *Probability and Statistics for Engineers and Scientists*, by Walpole and Myers, contains an interesting 2^{6-3} fractional factorial data set taken originally from the *Proceedings of the 10th Conference on the Design of Experiments in Army Research, Development and Testing* (ARO-D Report 65-3). The study investigated the effects of six two-level factors on X-ray intensity ratios associated with a particular component of propellant mixtures in X-ray fluorescent analyses of propellant slurry. Factors A, B, C, and D represent the concentrations (at low and high levels) of four propellant components. Factors E and F represent the weights (also at low and high levels) of fine and coarse particles present.

Example 12
(continued)

Eight different combinations of levels of factors A, B, C, D, E, and F were each tested twice for X-ray intensity ratio, y. The eight combinations actually included in the study can be thought of as follows. Using the columns of Table 8.28, levels of factor D were chosen using the signs in the ABC product column directly; levels of factor E were chosen by reversing the signs in the BC product column; and levels of factor F were chosen by reversing the signs of the AC product column. Verify that such a prescription implies that the eight combinations included in the study (written down in Yates order for factors A, B, and C) were as displayed in Table 8.29. The eight combinations indicated in Table 8.29 are, of course, $\frac{1}{8}$ of the 64 different possible combinations of levels of the six factors.

Table 8.29
Combinations Included in the 2^{6-3} Propellant Slurry Study

A	B	C	F	E	D	Combination Name
$-$	$-$	$-$	$-$	$-$	$-$	(1)
$+$	$-$	$-$	$+$	$-$	$+$	adf
$-$	$+$	$-$	$-$	$+$	$+$	bde
$+$	$+$	$-$	$+$	$+$	$-$	abef
$-$	$-$	$+$	$+$	$+$	$+$	cdef
$+$	$-$	$+$	$-$	$+$	$-$	ace
$-$	$+$	$+$	$+$	$-$	$-$	bcf
$+$	$+$	$+$	$-$	$-$	$+$	abcd

The development of 2^{p-q} fractional factorials has been illustrated with eight-combination (i.e., $p - q = 3$) plans. But it should be obvious that there are 16-row, 32-row, 64-row, . . . , etc. versions of Table 8.28. Using any of these, one can assign levels for the last q factors according to signs in product columns and end up with a $\frac{1}{2^q}$ fraction of a full 2^p factorial plan. When this is done, the 2^p factorial effects are aliased in 2^{p-q} groups of 2^q effects each. The determination of this alias structure can be made by using q **generators** to develop a defining relation for the fractional factorial. A general definition of the notion of generators for a 2^{p-q} fractional factorial is next.

Determining the alias structure of a 2^{p-q} factorial

Definition 9

When a 2^{p-q} fractional factorial comes about by assigning levels of each of the "last" q factors based on a different column of products of signs for the "first" $p - q$ factors, the q different relationships

$$\left(\begin{array}{l}\text{the name of an}\\ \text{additional factor}\end{array}\right) \leftrightarrow \pm \left(\begin{array}{l}\text{a product of names of some}\\ \text{of the first } p-q \text{ factors}\end{array}\right)$$

corresponding to how the combinations are chosen are called **generators** of the plan.

Each generator can be translated into a statement with I on the left side and then taken individually, multiplied in pairs, multiplied in triples, and so on until the whole **defining relation** is developed. (See again Definition 8, page 602, for the meaning of this term.) In doing so, use can be made of the convention that minus any letter times minus that letter is I.

Example 12 *(continued)*

In the Army propellant example, the $q = 3$ generators that led to the combinations in Table 8.29 were

$$\mathrm{D} \leftrightarrow \mathrm{ABC}$$
$$\mathrm{E} \leftrightarrow -\mathrm{BC}$$
$$\mathrm{F} \leftrightarrow -\mathrm{AC}$$

Multiplying through by the left sides of these, one obtains the three relationships

$$\mathrm{I} \leftrightarrow \mathrm{ABCD} \tag{8.25}$$
$$\mathrm{I} \leftrightarrow -\mathrm{BCE} \tag{8.26}$$
$$\mathrm{I} \leftrightarrow -\mathrm{ACF} \tag{8.27}$$

But in light of the conventions of formal multiplication, if $\mathrm{I} \leftrightarrow \mathrm{ABCD}$ and $\mathrm{I} \leftrightarrow -\mathrm{BCE}$, it should also be the case that

$$\mathrm{I} \leftrightarrow (\mathrm{ABCD}) \cdot (-\mathrm{BCE})$$

that is,

$$\mathrm{I} \leftrightarrow -\mathrm{ADE}$$

Similarly, using relationships (8.25) and (8.27), one obtains

$$\mathrm{I} \leftrightarrow -\mathrm{BDF}$$

using relationships (8.26) and (8.27), one obtains

$$\mathrm{I} \leftrightarrow \mathrm{ABEF}$$

Example 12 (continued)

and finally, using all three relationships (8.25), (8.26), and (8.27), one has

$$\mathrm{I} \leftrightarrow \mathrm{CDEF}$$

Combining all of this, the complete defining relation for this 2^{6-3} study is

$$\begin{aligned} &\mathrm{I} \leftrightarrow \mathrm{ABCD} \leftrightarrow -\mathrm{BCE} \leftrightarrow -\mathrm{ACF} \leftrightarrow \\ &-\mathrm{ADE} \leftrightarrow -\mathrm{BDF} \leftrightarrow \mathrm{ABEF} \leftrightarrow \mathrm{CDEF} \end{aligned} \tag{8.28}$$

Defining relation (8.28) is rather formidable, but it tells the whole truth about what can be learned based on the $\frac{1}{8}$ of 64 possible combinations of six two-level factors. Relation (8.28) specifies all effects that will be aliased with the grand mean. Appropriately multiplying through expression (8.28) gives all aliases of any effect of interest. For example, multiplying through relation (8.28) by A gives

$$\mathrm{A} \leftrightarrow \mathrm{BCD} \leftrightarrow -\mathrm{ABCE} \leftrightarrow -\mathrm{CF} \leftrightarrow -\mathrm{DE} \leftrightarrow -\mathrm{ABDF} \leftrightarrow \mathrm{BEF} \leftrightarrow \mathrm{ACDEF}$$

and for example, the (high level) A main effect will be indistinguishable from minus the (all high levels) CF 2-factor interaction.

Data analysis for a 2^{p-q} study

With a 2^{p-q} fractional factorial's defining relation in hand, the analysis of data proceeds exactly as indicated earlier for $\frac{1}{2}$ fractions. It is necessary to

1. compute estimates of (sums and differences of) effects ignoring the last q factors,
2. judge their statistical detectability using confidence interval or normal plotting methods, and then
3. seek a plausible tentative interpretation of the important estimates in light of the alias structure.

Example 12 (continued)

In the Army propellant study, $m = 2$ trials for each of the 2^{6-3} combinations listed in Table 8.29 gave $s_{\mathrm{P}}^2 = .02005$ and the sample averages listed in Table 8.30.

Temporarily ignoring all but the ("first") three factors A, B, and C (since the levels of D, E and F were derived or generated from the levels of A, B and C), the (three-cycle) Yates algorithm can be used on the sample means, as shown in Table 8.31. Remember that the estimates in the next-to-last column of Table 8.31 must be interpreted in light of the alias structure for the original experimental plan. So for example, since (both from the original generators and

Table 8.30
Eight Sample Means from the 2^{6-3} Propellant Slurry Study

Combination	$\bar{y}$	Combination	$\bar{y}$
(1)	1.1214	cdef	.9285
adf	1.0712	ace	1.1635
bde	.9415	bcf	.9561
abef	1.1240	abcd	.9039

from relation (8.28)) one knows that D $\leftrightarrow$ ABC, the $-.0650$ value on the last line of Table 8.31 is estimating

$$\alpha\beta\gamma_{222} + \delta_2 \pm \text{(six other effects)}$$

So if one were expecting a large main effect of factor D, one would expect it to be evident in the $-.0650$ value.

Since a value of s_{P} is available here, there is no need to resort to normal-plotting to judge the statistical detectability of the values coming out of the Yates algorithm. Instead (still temporarily calculating as if only the first three factors were present) one can make confidence intervals based on the estimates, by employing the $\nu = 8 = 16 - 8$, $m = 2$, and $p = 3$ version of formula (8.13) from Section 8.2. That is, using 95% two-sided individual confidence intervals, a precision of

$$\pm 2.306 \frac{\sqrt{.02005}}{\sqrt{2 \cdot 2^3}} = \pm .0817$$

should be attached to each of the estimates in Table 8.31. By this standard, none of the estimates from the propellant study are clearly different from 0. For

Table 8.31
The Yates Algorithm for a 2^3 Factorial Applied to the 2^{6-3} Propellant Data

$\bar{y}$	Cycle 1	Cycle 2	Cycle 3	Cycle 3 $\div$ 8	Sum Estimated
1.1214	2.1926	4.2581	8.2101	1.0263	$\mu_{....}$ + aliases
1.0712	2.0655	3.9520	.3151	.0394	α_2 + aliases
.9415	2.0920	.1323	$-.3591$	$-.0449$	β_2 + aliases
1.1240	1.8600	.1828	$-.0545$	$-.0068$	$\alpha\beta_{22}$ + aliases
.9285	$-.0502$	$-.1271$	$-.3061$	$-.0383$	γ_2 + aliases
1.1635	.1825	$-.2320$	.0505	.0063	$\alpha\gamma_{22}$ + aliases
.9561	.2350	.2327	$-.1049$	$-.0131$	$\beta\gamma_{22}$ + aliases
.9039	$-.0522$	$-.2872$	$-.5199$	$-.0650$	$\alpha\beta\gamma_{222}$ + aliases

Example 12 (*continued*)

engineering purposes, the bottom line is that more data are needed before even the most tentative conclusions about system behavior should be made.

Example 13

A 2^{5-2} Catalyst Development Experiment

Hansen and Best, in their paper "How to Pick a Winner" (presented at the 1986 annual meeting of the American Statistical Association), described several industrial experiments conducted in a research program aimed at the development of an effective catalyst for producing ethyleneamines by the amination of monoethanolamine. One of these was a partially replicated 2^{5-2} fractional factorial study in which the response variable, y, was percent water produced during the reaction period. The five two-level experimental factors were as in Table 8.32. (The T-372 support was an alpha-alumina support and the T-869 support was a silica alumina support.)

The fractional factorial described by Hansen and Best has ($q = 2$) generators D $\leftrightarrow$ ABC and E $\leftrightarrow$ BC. The resulting defining relation (involving $2^2 = 4$ strings of letters) is then

$$\text{I} \leftrightarrow \text{ABCD} \leftrightarrow \text{BCE} \leftrightarrow \text{ADE}$$

where the fact that the ADE 3-factor interaction is aliased with the grand mean can be seen by multiplying together ABCD and BCE, which (from the generators) themselves represent effects aliased with the grand mean. Here one sees that effects will be aliased together in eight groups of four.

The data reported by Hansen and Best, and some corresponding summary statistics, are given in Table 8.33. The pooled sample variance derived from the values in Table 8.33 is

$$\begin{aligned} s_{\text{P}}^2 &= \frac{(3-1)(2.543) + (2-1)(2.163) + (2-1)(.238)}{(3-1) + (2-1) + (2-1)} \\ &= 1.872 \end{aligned}$$

Table 8.32
Five Catalysis Variables and Their Experimental Levels

Factor	Process Variable	Levels
A	Ni/Re Ratio	2/1 (−) vs. 20/1 (+)
B	Precipitant	$(NH_4)_2CO_3$ (−) vs. none (+)
C	Calcining Temperature	300°C (−) vs. 500°C (+)
D	Reduction Temperature	300°C (−) vs. 500°C (+)
E	Support Used	T-372 (−) vs. T-869 (+)

Table 8.33
Data from a 2^{5-2} Catalyst Study and Corresponding Sample Means and Variances

Combination	% Water Produced, y	$\bar{y}$	s^2
e	8.70, 11.60, 9.00	9.7670	2.543
ade	26.80	26.800	—
bd	24.88	24.880	—
ab	33.15	33.150	—
cd	28.90, 30.980	29.940	2.163
ac	30.20	30.200	—
bce	8.00, 8.69	8.345	.238
abcde	29.30	29.300	—

with $\nu = (3-1) + (2-1) + (2-1) = 4$ associated degrees of freedom. The corresponding pooled sample standard deviation is

$$\sqrt{s_{\text{P}}^2} = \sqrt{1.872} = 1.368$$

So temporarily ignoring the existence of factors D and E, it is possible to use the $p = 3$ version of formula (8.12) to derive precisions to attach to the estimates (of sums of 2^5 factorial effects) that result from the use of the Yates algorithm on the sample means in Table 8.33. That is, for 95% two-sided individual confidence intervals, precisions of

$$\pm 2.776(1.368)\frac{1}{2^3}\sqrt{\frac{1}{3}+\frac{1}{1}+\frac{1}{1}+\frac{1}{1}+\frac{1}{2}+\frac{1}{1}+\frac{1}{2}+\frac{1}{1}}$$

that is,

$$\pm 1.195\% \text{ water}$$

can be attached to the estimates.

The reader can verify that the (three-cycle) Yates algorithm applied to the means in Table 8.33 gives the estimates in Table 8.34. Identifying those estimates in Table 8.34 whose magnitudes make them statistically detectable according to a criterion of ± 1.195, there are (in order of decreasing magnitude)

$$\alpha_2 + \beta\gamma\delta_{222} + \alpha\beta\gamma\epsilon_{2222} + \delta\epsilon_{22} \text{ estimated as } 5.815$$
$$\beta\gamma_{22} + \alpha\delta_{22} + \epsilon_2 + \alpha\beta\gamma\delta\epsilon_{22222} \text{ estimated as } -5.495$$
$$\alpha\beta\gamma_{222} + \delta_2 + \alpha\epsilon_{22} + \beta\gamma\delta\epsilon_{2222} \text{ estimated as } 3.682$$
$$\alpha\beta_{22} + \gamma\delta_{22} + \alpha\gamma\epsilon_{222} + \beta\delta\epsilon_{222} \text{ estimated as } 1.492$$

Example 13
(continued)

Table 8.34
Estimates of Sums of Effects for the Catalyst Study

Sum of Effects Estimated	Estimate
grand mean + aliases	24.048
A + aliases	5.815
B + aliases	−.129
AB + aliases	1.492
C + aliases	.399
AC + aliases	−.511
BC + aliases	−5.495
ABC + aliases	3.682

The simplest possible tentative interpretation of the first two of these results is that the A and E main effects are large enough to see above the background variation. What to make of the third, given the first two, is not so clear. The large 3.682 estimate can equally simply be tentatively attributed to a D main effect or to an AE 2-factor interaction. (Interestingly, Hansen and Best reported that subsequent experimentation was done with the purpose of determining the importance of the D main effect, and indeed, the importance of this factor in determining y was established.)

Exactly what to make of the fourth statistically significant estimate is even less clear. It is therefore comforting that, although big enough to be detectable, it is less than half the size of the third largest estimate. In the particular real situation, the authors seem to have found an "A, E, and D main effects only" description of y useful in subsequent work with the chemical system.

The reader may have noticed that the possibilities discussed in the previous example do not even exhaust the plausible interpretations of the fact that three estimated sums of effects are especially large. For example, "large DE 2-factor interactions and large D and E main effects" is yet another alternative possibility. This ambiguity serves to again emphasize the tentative nature of conclusions that can be drawn on the basis of small fractions of full factorials. And it also underlines the absolute necessity of subject-matter expertise and follow-up study in sorting out the possibilities in a real problem. There is simply no synthetic way to tell which of various simple alternative explanations suggested by a fractional factorial analysis is the right one.

8.4.2 Design Resolution

The results of five different real applications of 2^{p-q} plans have been discussed in Examples 9, 10, 11, 12, and 13. From them, it should be clear how important it is to

Good choice of a fractional factorial

have the simplest alias structure possible when it comes time to interpret the results of a fractional factorial study. The object is to have low-order effects (like main effects and 2-factor interactions) aliased not with other low-order effects, but rather only with high-order effects (many-factor interactions). It is the defining relation that governs how the 2^p factorial effects are divided up into groups of aliases. If there are only long products of factor names appearing in the defining relation, low-order effects are aliased only with high-order effects. On the other hand, if there are short products of factor names appearing, there will be low-order effects aliased with other low-order effects. As a kind of measure of quality of a 2^{p-q} plan, it is thus common to adopt the following notion of **design resolution**.

Definition 10

The **resolution** of a 2^{p-q} fractional factorial plan is the number of letters in the shortest product appearing in its defining relation.

In general, when contemplating the use of a 2^{p-q} design, one wants the largest resolution possible for a given investment in 2^{p-q} combinations. Not all choices of generators give the same resolution. In Section 8.3, the prescription given for the $\frac{1}{2}$ fractions was intended to give 2^{p-1} fractional factorials of resolution p (the largest resolution possible). For general 2^{p-q} studies, one must be a bit careful in choosing generators. What seems like the most obvious choice need not be the best in terms of resolution.

Example 14

Resolution 4 in a 2^{6-2} Study

Consider planning a 2^{6-2} study—that is, a study including 16 out of 64 possible combinations of levels of factors A, B, C, D, E, and F. A rather natural choice of two generators for such a study is

$$E \leftrightarrow ABCD$$
$$F \leftrightarrow ABC$$

The corresponding defining relation is

$$I \leftrightarrow ABCDE \leftrightarrow ABCF \leftrightarrow DEF$$

The resulting design is of resolution 3, and there are some main effects aliased with (only) 2-factor interactions.

On the other hand, the perhaps slightly less natural choice of generators

$$E \leftrightarrow BCD$$
$$F \leftrightarrow ABC$$

Example 14 *(continued)*

has defining relation

$$\mathrm{I} \leftrightarrow \mathrm{BCDE} \leftrightarrow \mathrm{ABCF} \leftrightarrow \mathrm{ADEF}$$

and is of resolution 4. No main effect is aliased with any interaction of order less than 3. This second choice is better than the first in terms of resolution.

Table 8.35 indicates what is possible in terms of resolution for various numbers of factors and combinations for a 2^{p-q} fractional factorial. The table was derived from a more detailed one on page 410 of *Statistics for Experimenters* by Box, Hunter, and Hunter, which gives not only the best resolutions possible but also generators for designs achieving those resolutions. The more limited information in Table 8.35 is sufficient for most purposes. Once one is sure what is possible, it is usually relatively painless to do the trial-and-error work needed to produce a plan of highest possible resolution. And it is probably worth doing as an exercise, to help one consider the pros and cons of various choices of generators for a given set of real factors.

Table 8.35 has no entries in the "8 combinations" row for more than 7 factors. If the table were extended beyond 11 factors, there would be no entries in the "16 samples" row beyond 15 factors, no entries in the "32 samples" row beyond 31 factors, etc. The reason for this should be obvious. For 8 combinations, there are only 7 columns total to use in Table 8.28. Corresponding tables for 16 combinations would have only 15 columns total, for 32 combinations only 31 columns total, etc.

As they have been described here, 2^{p-q} fractional factorials can be used to study at most $2^t - 1$ factors in 2^t samples. The cases of 7 factors in 8 combinations, 15 factors in 16 combinations, 31 factors in 32 combinations, etc. represent a kind of extreme situation where a maximum number of factors is studied (at the price of creating a worst possible alias structure) in a given number of combinations. For the case of $p = 7$ factors in 8 combinations, effects are aliased in $2^{7-4} = 8$ groups of $2^4 = 16$; for the case of $p = 15$ factors in 16 combinations, the effects are aliased in $2^{15-11} = 16$ groups of $2^{11} = 2{,}048$; etc. These extreme cases of $2^t - 1$ factors in 2^t combinations are sometimes called **saturated fractional factorials**. They have very complicated alias structures and can support only the most tentative of conclusions.

Table 8.35
Best Resolutions Possible for Various Numbers of Combinations in a 2^{p-q} Study

Number of Combinations (2^{p-q})	Number of Factors (p): 4	5	6	7	8	9	10	11
8	4	3	3	3	—	—	—	—
16		5	4	4	4	3	3	3
32			6	4	4	4	4	4
64				7	5	4	4	4
128					8	6	5	5

Example 15

A 16-Run 15-Factor Process Development Study

The article "What Every Technologist Should Know About Experimental Design" by C. Hendrix (*Chemtech*, 1979) includes the results from an unreplicated 16-run (saturated) 15-factor experiment. The response, y, was a measure of cold crack resistance for an industrial product. Experimental factors and levels were as listed in Table 8.36.

Table 8.36
15 Process Variables and Their Experimental Levels

Factor	Process Variable	Levels
A	Coating Roll Temperature	115° (−) vs. 125° (+)
B	Solvent	Recycled (−) vs. Refined (+)
C	Polymer X-12 Preheat	No (−) vs. Yes (+)
D	Web Type	LX-14 (−) vs. LB-17 (+)
E	Coating Roll Tension	30 (−) vs. 40 (+)
F	Number of Chill Rolls	1 (−) vs. 2 (+)
G	Drying Roll Temperature	75° (−) vs. 80° (+)
H	Humidity of Air Feed to Dryer	75% (−) vs. 90% (+)
J	Feed Air to Dryer Preheat	No (−) vs. Yes (+)
K	Dibutylfutile in Formula	12% (−) vs. 15% (+)
L	Surfactant in Formula	.5% (−) vs. 1% (+)
M	Dispersant in Formula	.1% (−) vs. 2% (+)
N	Wetting Agent in Formula	1.5% (−) vs. 2.5% (+)
O	Time Lapse Before Coating Web	10 min (−) vs. 30 min (+)
P	Mixer Agitation Speed	100 RPM (−) vs. 250 RPM (+)

The experimental plan used was defined by the $q = 11$ generators

$$\begin{aligned}&\text{E} \leftrightarrow \text{ABCD},\ \text{F} \leftrightarrow \text{BCD},\ \text{G} \leftrightarrow \text{ACD},\ \text{H} \leftrightarrow \text{ABC},\ \text{J} \leftrightarrow \text{ABD},\ \text{K} \leftrightarrow \text{CD},\\ &\text{L} \leftrightarrow \text{BD},\ \text{M} \leftrightarrow \text{AD},\ \text{N} \leftrightarrow \text{BC},\ \text{O} \leftrightarrow \text{AC},\quad \text{and}\quad \text{P} \leftrightarrow \text{AB}\end{aligned}$$

The combinations actually run and the cold crack resistances observed are given in Table 8.37.

Ignoring all factors but A, B, C, and D, the combinations listed in Table 8.37 are in Yates standard order and are therefore ready for use in finding estimates of sums of effects. Table 8.38 shows the results of using the (four-cycle) Yates algorithm on the 16 observations listed in Table 8.37. A normal plot of the last 15 of these estimates is shown in Figure 8.13. It is clear from the figure that the two corresponding to B + *aliases* and F + *aliases* are detectably larger than the rest.

Example 15
(continued)

Table 8.37
16 Experimental Combinations and Measured Cold Crack Resistances

Combination	y	Combination	y
eklmnop	14.8	dfgjnop	17.8
aghjkln	16.3	adefhmn	18.9
bfhjkmo	23.5	bdeghlo	23.1
abefgkp	23.9	abdjlmp	21.8
cfghlmp	19.6	cdehjkp	16.6
acefjlo	18.6	acdgkmo	16.7
bcegjmn	22.3	bcdfkln	23.5
abchnop	22.2	abcdefghjklmnop	24.9

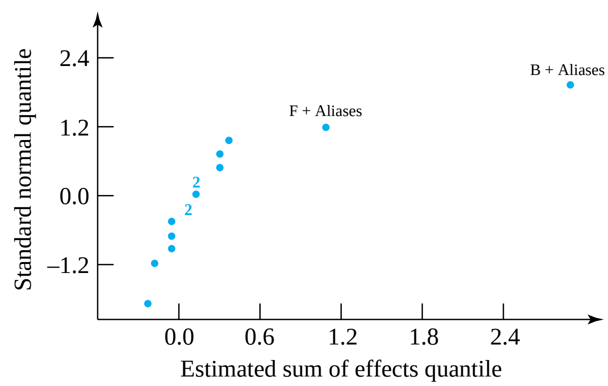

Figure 8.13 Normal plot of estimated sums of effects in the 2^{15-11} process development study

It is not feasible to write out the whole defining relation for this 2^{15-11} study. Effects are aliased in $2^{p-q} = 2^{15-11} = 16$ groups of $2^q = 2^{11} = 2{,}048$. In particular (though it would certainly be convenient if the 2.87 estimate in Table 8.38 could be thought of as essentially representing β_2), β_2 has 2,047 aliases, some of them as simple as 2-factor interactions. By the same token, it would certainly be convenient if the small estimates in Table 8.38 were indicating that all summands of the sums of effects they represent were small. But the possibility of cancellation in the summation must not be overlooked.

The point is that only the most tentative description of this system should be drawn from even this very simple "two large estimates" outcome. The data in Table 8.37 hint at the primary importance of factors B and F in determining cold crack resistance, but the case is hardly airtight. There is a suggestion of a direction for further experimentation and discussion with process experts but certainly no detailed map of the countryside where one is going.

Table 8.38
Estimates of Sums of Effects for the 2^{15-11} Process Development Study

Sum of Effects Represented	Estimate
grand mean + aliases	20.28
A + aliases	.13
B + aliases	2.87
P + aliases (including AB)	−.08
C + aliases	.27
O + aliases (including AC)	−.08
N + aliases (including BC)	−.19
H + aliases (including ABC)	.36
D + aliases	.13
M + aliases (including AD)	.03
L + aliases (including BD)	.04
J + aliases (including ABD)	−.06
K + aliases (including CD)	−.26
G + aliases (including ACD)	.29
F + aliases (including BCD)	1.06
E + aliases (including ABCD)	.11

One thing that can be said fairly conclusively on the basis of this study is that the analysis points out what is in retrospect obvious in Table 8.37. Consistent with the "B + *aliases* and F + *aliases* sums are positive and large" story told in Figure 8.13, the largest four values of y listed in Table 8.37 correspond to combinations where both B and F are at their high levels.

8.4.3 Two-Level Factorials and Fractional Factorials in Blocks (*Optional*)

A somewhat specialized but occasionally useful adaptation of the 2^{p-q} material presented here has to do with the analysis of full or fractional two-level factorial studies run in complete or incomplete blocks. When the number of blocks under consideration is itself a power of 2, clever use of the methods developed in this chapter can guide the choice of which combinations to place in incomplete blocks, as well as the analysis of data from both incomplete and complete block studies.

The basic idea used is to formally represent one 2^t-level factor "Blocks" as t "extra" two-level factors. One lets combinations of levels of these extra factors define the blocks into which combinations of levels of the factors of interest are placed. In data analysis, effects involving only the extra factors as Block main effects and effects involving both the extra factors and the factors of interest are recognized as Block×Treatment interactions. In carrying out this program, it is fairly common (though not necessarily safe) to operate as if the Block×Treatment interactions were all negligible. How choice and analysis of blocked 2^{p-q} studies proceed will be illustrated with a series of three examples that are variations on Example 11.

Example 16
(Example 11 revisited)

A 2^4 Fabric Tenacity Study Run in Two Blocks

In the weaving study of Johnson, Clapp, and Baqai, four factors—A, B, C, and D—were studied. The discussion in Section 8.3 described how the authors originally ran a replicated 2^{4-1} fractional factorial with defining relation I ↔ ABCD. This was followed up later with a second 2^{4-1} fractional factorial having defining relation I ↔ −ABCD, thus completing the full 2^4 factorial. However, since the study of the two $\frac{1}{2}$ fractions was separated in time, it is sensible to think of the two parts of the study as different *blocks*—that is, to think of a fifth two-level factor (say, E) representing the time of observation.

How then to use the formal multiplication idea to understand the alias structure? Notice that there are 16 different samples and five factors for consideration. This suggests that somehow (at least in formal terms) this situation might be thought of as a 2^{5-1} data structure. Further, the two formal expressions

$$\text{I} \leftrightarrow \text{ABCD} \tag{8.29}$$

$$\text{I} \leftrightarrow -\text{ABCD} \tag{8.30}$$

define the two sets of 8 out of 16 ABCD combinations actually run. These result from a formal expression like

$$\text{I} \leftrightarrow \text{ABCDE} \tag{8.31}$$

where E can be thought of as contributing either the plus or the minus signs in expressions (8.29) and (8.30). If one calls block 1 (the first set of 8 samples) the high level of E, expression (8.31) leads to exactly the I ↔ ABCD $\frac{1}{2}$-fraction of 2^4 combinations of A, B, C, and D for use as block 1. And the I ↔ −ABCD $\frac{1}{2}$-fraction for use as block 2. This can be seen in Table 8.39.

With factor E designating block number, the two columns of Table 8.39 taken together designate the I ↔ ABCDE $\frac{1}{2}$-fraction of 2^5 A, B, C, D, and E combinations. And (ignoring the e) the first column of Table 8.39 designates the I ↔ ABCD $\frac{1}{2}$-fraction of 2^4 A, B, C, and D combinations, while the second designates the I ↔ −ABCD $\frac{1}{2}$-fraction of 2^4 A, B, C, and D combinations.

Once it is clear that the Johnson, Clapp, and Baqai study can be thought of in terms of expression (8.31) with the two-level blocking factor E, it is also clear how any *block effects* will show up during data analysis. One temporarily ignores the blocks and uses the Yates algorithm to compute fitted 2^4 factorial effects. It is then necessary to remember, for example, that the fitted ABCD 4-factor interaction reflects not only $\alpha\beta\gamma\delta_{2222}$ but any block main effects as well.

Table 8.39
A 2^{5-1} Fractional Factorial or a 2^4 Factorial in Two Blocks

Block 1	Block 2
e	a
abe	b
ace	c
bce	abc
ade	d
bde	abc
cde	acd
abcde	bcd

And for example, any 2-factor interaction of A and blocks will be reflected in the fitted BCD 3-factor interaction. Of course, if all interactions with blocks are negligible, all fitted effects except that for the ABCD 4-factor interaction would indeed represent the appropriate 2^4 factorial effects.

Example 17

A 2^4 Factorial Run in Four Blocks

For the sake of illustration, suppose that Johnson, Clapp, and Baqai had a priori planned to conduct a full 2^4 factorial set of ABCD combinations in four incomplete blocks (of four combinations each). Consider how those blocks might have been chosen and how subsequent data analysis would have proceeded.

The one four-level factor Blocks can here be thought of in terms of the combinations of two extra two-level factors, which can be designated as E and F. In order to accommodate the original four factors and these two additional ones in 16 ABCDEF combinations, one must choose a 2^{6-2} design by specifying two generators. The choices

$$\mathrm{E} \leftrightarrow \mathrm{BCD} \tag{8.32}$$

$$\mathrm{F} \leftrightarrow \mathrm{ABC} \tag{8.33}$$

leading to the defining relation

$$\mathrm{I} \leftrightarrow \mathrm{BCDE} \leftrightarrow \mathrm{ABCF} \leftrightarrow \mathrm{ADEF} \tag{8.34}$$

will be used here. Table 8.40 indicates the 16 combinations of levels of factors A through F prescribed by the generators (8.32) and (8.33).

Example 17 *(continued)*

The four different combinations of levels of E and F ((1), e, f, and ef) can be thought as designating in which block a given ABCD combination should appear. So generators (8.32) and (8.33) prescribe the division of the full 2^4 factorial (in the factors A through D) into the blocks indicated in Table 8.40 and Table 8.41.

As always, the defining relation (given here in display (8.34)) describes how effects are aliased. Table 8.42 indicates the aliases of each of the 2^4 factorial effects, obtained by multiplying through relation (8.34) by the various combinations of the letters A, B, C, and D. Notice from Table 8.42 that the BCD and ABC 3-factor interactions are aliased with block main effects. So is the AD 2-factor

Table 8.40
16 Combinations of Levels of A through F

A	B	C	D	E	F	Block Prescribed by Levels of E and F
−	−	−	−	−	−	1
+	−	−	−	−	+	3
−	+	−	−	+	+	4
+	+	−	−	+	−	2
−	−	+	−	+	+	4
+	−	+	−	+	−	2
−	+	+	−	−	−	1
+	+	+	−	−	+	3
−	−	−	+	+	−	2
+	−	−	+	+	+	4
−	+	−	+	−	+	3
+	+	−	+	−	−	1
−	−	+	+	−	+	3
+	−	+	+	−	−	1
−	+	+	+	+	−	2
+	+	+	+	+	+	4

Table 8.41
A 2^4 Factorial in Four Blocks
(from a 2^{6-2} Fractional Factorial)

Block 1	Block 2	Block 3	Block 4
(1)	ab	a	b
bc	ac	abc	c
abd	d	bd	ad
acd	bcd	cd	abcd

Table 8.42
Aliases of the 2^4 Factorial Effects When Run in Four Blocks Prescribed by Generators (8.32) and (8.33)

I ↔ BCDE ↔ ABCF ↔ ADEF
A ↔ ABCDE ↔ BCF ↔ DEF
B ↔ CDE ↔ ACF ↔ ABDEF
AB ↔ ACDE ↔ CF ↔ BDEF
C ↔ BDE ↔ ABF ↔ ACDEF
AC ↔ ABDE ↔ BF ↔ CDEF
BC ↔ DE ↔ AF ↔ ABCDEF
ABC ↔ ADE ↔ F ↔ BCDEF
D ↔ BCE ↔ ABCDF ↔ AEF
AD ↔ ABCE ↔ BCDF ↔ EF
BD ↔ CE ↔ ACDF ↔ ABEF
ABD ↔ ACE ↔ CDF ↔ BEF
CD ↔ BE ↔ ABDF ↔ ACEF
ACD ↔ ABE ↔ BDF ↔ CEF
BCD ↔ E ↔ ADF ↔ ABCEF
ABCD ↔ AE ↔ DF ↔ BCEF

interaction, since one of its aliases is EF, which involves only the two-level extra factors E and F used to represent the four-level factor Blocks. On the other hand, if interactions with Blocks are negligible, it is *only* these three of the 2^4 factorial effects that are aliased with other possibly nonnegligible effects. (For any other of the 2^4 factorial effects, each alias involves letters both from the group A, B, C, and D and also from the group E and F—and is therefore some kind of Block × Treatment interaction.)

Analysis of data from a plan like that in Table 8.41 would proceed as indicated repeatedly in this chapter. The Yates algorithm applied to sample means listed in Yates standard order for factors A, B, C, and D produces estimates that are interpreted in light of the alias structure laid out in Table 8.42.

Example 18

A 2^{4-1} Fractional Factorial Run in Four Blocks

As a final variant on the 4-factor weaving example, consider how the original $\frac{1}{2}$ fraction of the 2^4 factorial might itself have been run in four incomplete blocks of two combinations. (Imagine that for some reason, only two combinations could be prepared on any single day and that there was some fear of Day effects related to environmental changes, instrument drift, etc.)

Example 18 (*continued*)

Only eight combinations are to be chosen. In doing so, one needs to account for the four experimental factors A, B, C, and D and two extras E and F, which can be used to represent the four-level factor Blocks. Starting with the first three experimental factors A, B, and C (three of them because $2^3 = 8$), one needs to choose three generators. The original 2^{4-1} study had generator

$$D \leftrightarrow ABC$$

so it is natural to begin there. For the sake of example, consider also the generators

$$E \leftrightarrow BC$$
$$F \leftrightarrow AC$$

These give the defining relation

$$I \leftrightarrow ABCD \leftrightarrow BCE \leftrightarrow ACF \leftrightarrow ADE \leftrightarrow BDF \leftrightarrow ABEF \leftrightarrow CDEF \qquad \textbf{(8.35)}$$

and the prescribed set of combinations listed in Table 8.43. (The four different combinations of levels of E and F ((1), e, f, and ef) designate in which block a given ABCD combination from the $\frac{1}{2}$ fraction should appear.)

Table 8.43
A 2^{6-3} Fractional Factorial or a 2^{4-1} Fractional Factorial in Four Blocks

Block 1	Block 2	Block 3	Block 4
ab	ade	bdf	ef
cd	bce	acf	abcdef

Some experimenting with relation (8.35) will show that *all* 2-factor interactions of the four original experimental factors A, B, C, and D are aliased not only with other 2-factor interactions of experimental factors but also with Block main effects. Thus, any systematic block-to-block changes would further confuse one's perception of 2-factor interactions of the experimental factors. But at least the main effects of A, B, C, and D are not aliased with Block main effects.

Examples 16 through 18 all treat situations where blocks are *incomplete*—in the sense that they don't each contain every combination of the experimental factors studied. **Complete block plans** with 2^t blocks can also be developed and analyzed through the use of t "extra" two-level factors to represent the single (2^t-level) factor Blocks. The path to be followed is by now worn enough through use in this chapter that further examples will not be included. But the reader should have no trouble figuring out, for example, how to analyze a full 2^4 factorial that is run completely once in each of two blocks, or even how to analyze a standard 2^{4-1} fractional factorial that is run completely once in each of four blocks.

8.4.4 Some Additional Comments

This 2^{p-q} fractional factorial material is fascinating, and extremely useful when used with a proper understanding of both its power and its limitations. However, an engineer who tries to use it in a cookbook fashion will usually wind up frustrated and disillusioned. The implications of aliasing must be thoroughly understood for successful use of the material. And a clear understanding of these implications will work to keep the engineer from routinely trying to study many factors based on very small amounts of data in a one-shot experimentation mode.

Engineers newly introduced to fractional factorial experimentation sometimes try to routinely draw final engineering conclusions about multifactor systems based on as few as eight data points. The folly of such a method of operation should be apparent. Economy of experimental effort involves not just collecting a small amount of data on a multifactor system, but rather collecting the minimum amount sufficient for a practically useful and reliable understanding of system behavior. Just a few expensive engineering errors, traceable to naive and overzealous use of fractional factorial experimentation, will easily negate any supposed savings generated by overly frugal data collection.

Choice of experiment size

Although several 8-combination plans have been used as examples in this section, such designs are often too small to provide much information on the behavior of real engineering systems. Typically, 2^{p-q} studies with $p-q \geq 4$ are recommended as far more likely to lead to a satisfactory understanding of system behavior.

It has been said several times that when intelligently used as factor-screening tools, 2^{p-q} fractional factorial studies will usually be followed up with more complete experimentation, such as a larger fraction or a complete factorial (often in a reduced set of factors). It is also true that techniques exist for choosing a relatively small second fraction in such a way as to resolve certain particular types of ambiguities of interpretation that can remain after the analysis of an initial fractional factorial. The interested reader can refer to Section 12.5 of *Statistics for Experimenters* by Box, Hunter, and Hunter for discussions of how to choose an additional fraction to “dealias” a particular main effect and all its associated interactions or to “dealias” all main effects.

Section 4 Exercises

1. What are the advantages and disadvantages of fractional factorial experimentation in comparison to factorial experimentation?
2. Under what circumstances can one hope to be successful experimenting with (say) 12 factors in (say) 16 experimental runs (i.e., based on 16 data points)?
3. What is the principle of “sparsity of effects” and how can it be used in the analysis of unreplicated 2^p and 2^{p-q} experiments?
4. In a 7-factor study, only 32 different combinations of levels of (two-level factors) A, B, C, D, E, F, and

G will be included, at least initially. The generators F $\leftrightarrow$ ABCD and G $\leftrightarrow$ ABCE will be used to choose the 32 combinations to include in the study.

(a) Write out the whole defining relation for the experiment that is contemplated here.

(b) Based on your answer to part (a), what effects will be aliased with the C main effect in the experiment that is being planned?

(c) When running the experiment, what levels of factors F and G are used when all of A, B, C, D, and E are at their low levels? What levels of factors F and G are used when A, B, and C are at their high levels and D and E are at their low levels?

(d) Suppose that after listing the data (observed y's) in Yates standard order as regards factors A, B, C, D, and E, you use the Yates algorithm to compute 32 fitted sums of effects. Suppose further that the fitted values appearing on the A + aliases, ABCD + aliases, and BCD + aliases rows of the Yates computations are the only ones judged to be of both statistical significance and practical importance. What is the simplest possible interpretation of this result?

5. In a 2^{5-2} study, where four sample sizes are 1 and four sample sizes are 2, $s_{\text{P}} = 5$. If 90% two-sided confidence limits are going to be used to judge the statistical detectability of sums of effects, what plus-or-minus value will be used?

6. Consider planning, executing, and analyzing the results of a 2^{6-2} fractional factorial experiment based on the two generators E $\leftrightarrow$ ABC and F $\leftrightarrow$ BCD.

(a) Write out the defining relation (i.e., the whole list of aliases of the grand mean) for such a plan.

(b) When running the experiment, what levels of factors E and F are used when all of A, B, C, and D are at their low levels? When A is at its high level but B, C, and D are at their low levels?

(c) Suppose that $m = 3$ data points from each of the 16 combinations of levels of factors (specified by the generators) give a value of $s_{\text{P}} \approx 2.00$. If individual 90% two-sided confidence intervals are to be made to judge the statistical significance of the estimated (sums of) effects, what is the value of the plus-or-minus part of each of those intervals?

Chapter 8 Exercises

1. Return to the situation of Chapter Exercise 4 of Chapter 4. That exercise concerns some unreplicated 2^3 factorial data taken from a study of the mechanical properties of a polymer. If you have not already done so, use the Yates algorithm to compute fitted 2^3 factorial effects for the data given in that exercise. Then make a normal plot of the seven fitted effects $a_2, b_2, \ldots, abc_{222}$ as a means of judging the statistical detectability of the various effects on impact strength. Interpret this plot.

2. Chapter Exercise 5 in Chapter 4 concerns a 2^3 study of mechanical pencil lead strength done by Timp and M-Sidek. Return to that exercise, and if you have not already done so, use the Yates algorithm to compute fitted 2^3 effects for the logged data.

(a) Compute s_{P} for the logged data. Individual confidence intervals for the theoretical 2^3 effects are of the form $\hat{E} \pm \Delta$. Find Δ if 95% individual two-sided intervals are of interest.

(b) Based on your value from part (a), which of the factorial effects are statistically detectable? Considering only those effects that are both statistically detectable and large enough to have a material impact on the breaking strength, interpret the results of the students' experiment. (For example, if the A main effect is judged to be both detectable and of practical importance, what does moving from the .3 diameter to the .7 diameter do to the breaking strength? Remember to translate back from the log scale when making these interpretations.)

(c) Use the reverse Yates algorithm to produce fitted $\ln(y)$ values for a few-effects model corresponding to your answer to (b). Use the fitted values to compute residuals (still on the log scale). Normal-plot these and plot them against levels of each of the three factors and against the fitted values, looking for obvious problems with the few-effects model.

(d) Based on your few-effects model, give a 95% two-sided confidence interval for the mean $\ln(y)$ that would be produced by the abc treatment combination. By exponentiating the endpoints of this interval, give a 95% two-sided confidence interval for the median number of clips required to break a piece of lead under this set of conditions.

3. The following are the weights recorded by $I = 3$ different students when weighing the same nominally 5 g mass with $J = 2$ different scales $m = 2$ times apiece. (They are part of the much larger data set given in Chapter Exercise 5 of Chapter 3.)

	Scale 1	Scale 2
Student 1	5.03, 5.02	5.07, 5.09
Student 2	5.03, 5.01	5.02, 5.07
Student 3	5.06, 5.00	5.10, 5.08

Corresponding fitted factorial effects are: $a_1 = .00417$, $a_2 = -.01583$, $a_3 = .01167$, $b_1 = -.02333$, $b_2 = .02333$, $ab_{11} = -.00417$, $ab_{12} = .00417$, $ab_{21} = .01083$, $ab_{22} = -.01083$, $ab_{31} = -.00667$, and $ab_{32} = .00667$. Further, a pooled standard deviation is $s_P = .02483$.

(a) To enhance an interaction plot of sample means with error bars derived from 95% two-sided individual confidence limits for the mean weights, what plus-or-minus value would be used to make those error bars? Make such a plot and discuss the likely statistical detectability of the interactions.

(b) Individual 95% two-sided confidence limits for the interactions $\alpha\beta_{ij}$ are of the form $ab_{ij} \pm \Delta$. Find Δ here. Based on this, are the interactions statistically detectable?

(c) Compare the Student main effects using individual 95% two-sided confidence intervals.

(d) Compare the Student main effects using simultaneous 95% two-sided confidence intervals.

4. The oil viscosity study of Dunnwald, Post, and Kilcoin (referred to in Chapter Exercise 8 of Chapter 7) was actually a 3×4 full factorial study. Some summary statistics for the entire data set are recorded in the accompanying tables. Summarized are $m = 10$ measurements of the viscosities of each of four different weights of three different brands of motor oil at room temperature. Units are seconds required for a ball to drop a particular distance through the oil.

	10W30	SAE 30
Brand M	$\bar{y}_{11} = 1.385$ $s_{11} = .091$	$\bar{y}_{12} = 2.066$ $s_{12} = .097$
Brand C	$\bar{y}_{21} = 1.319$ $s_{21} = .088$	$\bar{y}_{22} = 2.002$ $s_{22} = .089$
Brand H	$\bar{y}_{31} = 1.344$ $s_{31} = .066$	$\bar{y}_{32} = 2.049$ $s_{32} = .089$

	10W40	20W50
Brand M	$\bar{y}_{13} = 1.414$ $s_{13} = .150$	$\bar{y}_{14} = 4.498$ $s_{14} = .204$
Brand C	$\bar{y}_{23} = 1.415$ $s_{23} = .115$	$\bar{y}_{24} = 4.662$ $s_{24} = .151$
Brand H	$\bar{y}_{33} = 1.544$ $s_{33} = .068$	$\bar{y}_{34} = 4.549$ $s_{34} = .171$

(a) Find the pooled sample standard deviation here. What are the associated degrees of freedom?

(b) Make an interaction plot of sample means. Enhance this plot by adding error bars derived from 99% individual confidence intervals for the cell means. Does it appear that there are important and statistically detectable interactions here?

(c) If the Tukey method is used to find simultaneous 95% two-sided confidence intervals for all differences in Brand main effects, the intervals produced are of the form $\bar{y}_{i.} - \bar{y}_{i'.} \pm \Delta$. Find Δ.

(d) If the Tukey method is used to find simultaneous 95% two-sided confidence intervals for all differences in Weight main effects, the intervals produced are of the form $\bar{y}_{.j} - \bar{y}_{.j'} \pm \Delta$. Find Δ.

(e) Based on your answers to (c) and (d), would you say that there are statistically detectable Brand and/or Weight main effects on viscosity?

(f) We strongly suspect that the "$m = 10$" viscosity measurements made for each of the 12 Brand/Weight combinations were made on oil from a single quart of that type of oil. If this is the case, s_P, the baseline measure of variability, is measuring only the variability associated with experimental technique (not, for example, from quart to quart of a given type of oil). One might thus argue that the real-world inferences to be made, properly speaking, extend only to the particular quarts used in the study. Discuss how these interpretations (of s_P and the extent of statistically based inferences) would be different if in fact the students used different quarts of oil in producing the "$m = 10$" different viscosity measurements in each cell.

5. The article "Effect of Temperature on the Early-Age Properties of Type I, Type III and Type I/Fly Ash Concretes" by N. Gardner (*ACI Materials Journal*, 1990) contains summary statistics for a very large study of the properties of several concretes under a variety of curing conditions. The accompanying tables present some of the statistics from that paper. Given here are the sample means and standard deviations of 14-day compressive strengths for $m = 5$ specimens of Type I cement/fly ash concrete for all possible combinations of $I = 2$ water-cement ratios and $J = 4$ curing temperatures. The units are MPa.

	0°C	10°C
.55 Water/Cement Ratio	$\bar{y}_{11} = 28.99$ $s_{11} = .91$	$\bar{y}_{12} = 30.24$ $s_{12} = 1.26$
.35 Water/Cement Ratio	$\bar{y}_{21} = 38.70$ $s_{21} = .77$	$\bar{y}_{22} = 36.16$ $s_{22} = 1.92$

	20°C	30°C
.55 Water/Cement Ratio	$\bar{y}_{13} = 33.99$ $s_{13} = 1.85$	$\bar{y}_{14} = 36.02$ $s_{14} = .93$
.35 Water/Cement Ratio	$\bar{y}_{23} = 40.18$ $s_{23} = 2.86$	$\bar{y}_{24} = 42.36$ $s_{24} = 1.35$

(a) Find the pooled sample standard deviation here. What are the associated degrees of freedom?

(b) Make an interaction plot of sample means. Enhance this plot by adding error bars derived from simultaneous 95% confidence intervals for the cell means. Does it appear that there are important and statistically detectable interactions here? What practical implications would this have for a cold-climate civil engineer?

(c) Compute the fitted factorial effects from the eight sample means.

(d) If one wished to make individual 95% confidence intervals for the Ratio × Temperature interactions $\alpha\beta_{ij}$, these would be of the form $ab_{ij} \pm \Delta$, for an appropriate value of Δ. Find this Δ. Based on this value, do you judge any of the interactions to be statistically detectable?

6. The same article referred to in Exercise 5 reported summary statistics (similar to the ones for Type I cement/fly ash concrete) for the 14-day compressive strengths of Type III cement concrete. These are shown in the accompanying tables.

	0°C	10°C
.55 Water/Cement Ratio	$\bar{y}_{11} = 47.82$ $s_{11} = 4.03$	$\bar{y}_{12} = 42.75$ $s_{12} = 2.96$
.35 Water/Cement Ratio	$\bar{y}_{21} = 42.14$ $s_{21} = 2.64$	$\bar{y}_{22} = 36.72$ $s_{22} = 3.03$

	20°C	30°C
.55 Water/Cement Ratio	$\bar{y}_{13} = 42.38$ $s_{13} = 2.62$	$\bar{y}_{14} = 43.45$ $s_{14} = 1.80$
.35 Water/Cement Ratio	$\bar{y}_{23} = 36.72$ $s_{23} = 1.51$	$\bar{y}_{24} = 37.70$ $s_{24} = .89$

(a) Find the pooled sample standard deviation here. What are the associated degrees of freedom? ($m = 5$, as in Exercise 5.)

(b) Make an interaction plot of sample means useful for investigating the size of Ratio × Temperature interactions. Enhance this plot by adding error bars derived from simultaneous 95% confidence intervals for the cell means. Does it appear that there are important and statistically detectable interactions here? What practical implications would this have for a cold-climate civil engineer?

(c) Compute the fitted factorial effects from the eight sample means.

(d) If one wished to make individual 95% confidence intervals for the Ratio × Temperature interactions $\alpha\beta_{ij}$, these would be of the form $ab_{ij} \pm \Delta$, for an appropriate value of Δ. Find this Δ. Based on this value, do you judge any of the interactions to be statistically detectable?

(e) Give and interpret a 90% confidence interval for the difference in water/cement ratio main effects, $\alpha_2 - \alpha_1$. How would this be of practical use to a cold-climate civil engineer?

7. Suppose that in the context of Exercises 5 and 6, you judge that for the Type I cement/fly ash concrete there are important Ratio × Temperature interactions, but that for the Type III cement concrete there are not important Ratio × Temperature interactions. Taking the whole data set from both exercises together (both concrete types), would there be important (3-factor) Type × Ratio × Temperature interactions? Explain.

8. The ISU M.S. thesis, "An Accelerated Engine Test for Crankshaft and Bearing Compatibility," by P. Honan, discusses an industrial experiment run to investigate the effects of three factors on the wear of engine bearings. The factors and levels shown here were used in a 100-hour, 20-step engine probe test.

A	Crankshaft Material	cast nodular iron (−) vs. forged steel (+)
B	Bearing Material	aluminum (−) vs. copper/lead (+)
C	Debris Added to Oil	none (−) vs. 5.5 g SAE fine dust every 25 hours (+)

Two response variables were measured:

$$y_1 = \text{rod journal wear } (\mu\text{m})$$
$$y_2 = \text{main journal wear } (\mu\text{m})$$

The values of y_1 and y_2 reported by Honan are as follows.

Combination	y_1	y_2	Combination	y_1	y_2
(1)	2.7	5.6	c	3.1	3.2
a	.9	1.4	ac	18.6	27.3
b	3.0	7.1	bc	2.5	6.0
ab	1.1	1.6	abc	60.3	99.7

(a) Use the Yates algorithm and compute the fitted effects of the three experimental factors on both the rod and main bearing wear figures.

(b) Because there was no replication in this relatively expensive industrial experiment, there is no real option for judging the statistical significance of the 2^3 factorial effects except the use of normal-plotting. Make normal plots of the seven fitted effects, $a_2, b_2, \ldots, abc_{222}$ for both response variables. Do these identify one or two of the 2^3 factorial effects as clearly larger than the others? How hopeful are you that there is a simple, intuitively appealing few-effects description of the effects of factors A, B, and C on y_1 and y_2?

(c) Your normal plots from (b) ought to each have an interesting gap in the middle of the plot. Explain the origin of both that gap and the fact that all of your fitted effects should be positive, in terms of the relative magnitudes of the responses listed. (How, for example, does the response for combination abc enter into the calculation of the various fitted effects?)

(d) One simple way to describe the outcomes obtained in this study is as having *one very big response and one moderately big response*. Is there much chance that this pattern in y_1 and y_2 is in fact due only to random variation (i.e., that none of the factors have any effect here)? Make a normal plot of the raw y_1 values and one for the raw y_2 values to support your answer.

9. There is a certain degree of arbitrariness in the choice to use signs on the fitted effects corresponding to the "all high treatment" combination when normal-plotting fitted 2^p factorial effects. This can be eliminated by probability plotting the absolute values of the fitted effects and using not standard normal quantiles but rather quantiles for the distribution of the absolute value of a standard normal random variable. This notion is called *half normal-plotting the absolute fitted effects*, since the probability density of the absolute value of a standard normal variable looks like the right half of the standard normal density (multiplied by 2). The half normal quantiles are related to the standard normal quantiles via

$$Q(p) = Q_z\left(\frac{1+p}{2}\right)$$

and one interprets a half normal plot in essentially the same way that a normal plot is interpreted. That is, one thinks of the smaller plotted values as establishing a pattern of random-looking variation and identifies any of the larger values plotting off a line on the plot established by the small values as detectably larger than the others.

(a) Redo part (a) of Exercise 2 of Section 8.2 using a half normal plot of the absolute values of the fitted effects. (Your ith plotted point will have a horizontal coordinate equal to the ith smallest absolute fitted effect and a vertical coordinate equal to the $p = \frac{i-.5}{15}$ half normal quantile.) Are the conclusions about the statistical detectability of effects here the same as those you reached in Exercise 2 of Section 8.2?

(b) Redo Exercise 1 here using a half normal plot of the absolute values of the fitted effects. (Your ith plotted point will have a horizontal coordinate equal to the ith smallest absolute fitted effect and a vertical coordinate equal to the $p = \frac{i-.5}{7}$ half normal quantile.) Are the conclusions about the statistical detectability of effects here the same as those you reached in Exercise 1?

10. The text *Engineering Statistics* by Hogg and Ledolter contains an account (due originally to R. Snee) of a partially replicated 2^3 factorial industrial experiment. Under investigation were the effects of the following factors and levels on the percentage impurity, y, in a chemical product:

A	Polymer Type	standard (−) vs. new (but expensive) (+)
B	Polymer Concentration	.01% (−) vs. .04% (+)
C	Amount of an Additive	2 lb (−) vs. 12 lb (+)

The data that were obtained are as follows:

Combination	y (%)	Combination	y (%)
(1)	1.0	c	.9, .7
a	1.0, 1.2	ac	1.1
b	.2	bc	.2, .3
ab	.5	abc	.5

(a) Compute the fitted 2^3 factorial effects corresponding to the "all high treatment" combination.
(b) Compute the pooled sample standard deviation, s_{P}.
(c) Use your value of s_{P} from (b) and find the plus-or-minus part of 90% individual two-sided confidence limits for the 2^3 factorial effects.
(d) Based on your calculation in (c), which of the effects do you judge to be detectable in this 2^3 study?
(e) Write a paragraph or two for your engineering manager, summarizing the results of this experiment and making recommendations for the future running of this process. (Remember that you want low y and, all else being equal, low production cost.)

11. The article "Use of Factorial Designs in the Development of Lighting Products" by J. Scheesley (*Experiments in Industry: Design, Analysis and Interpretation of Results*, American Society for Quality Control, 1985) discusses a large industrial experiment intended to compare the use of two different types of lead wire in the manufacture of incandescent light bulbs under a variety of plant circumstances. The primary response variable in the study was

y = average number of leads missed per hour (because of misfeeds into automatic assembly equipment)

which was measured and recorded on the basis of eight-hour shifts. Consider here only part of the original data, which may be thought of as having replicated 2^4 factorial structure. That is, consider the following factors and levels:

A	Lead Type	standard (−) vs. new (+)
B	Plant	1 (−) vs. 2 (+)
C	Machine Type	standard (−) vs. high speed (+)
D	Shift	1st (−) vs. 2nd (+)

$m = 4$ values of y (each requiring an eight-hour shift to produce) for each combination of levels of factors A, B, C, and D gave the accompanying $\bar{y}$ and s^2 values.

Combination	$\bar{y}$	s^2	Combination	$\bar{y}$	s^2
(1)	28.4	97.6	d	36.8	146.4
a	21.9	15.1	ad	19.2	24.8
b	20.2	5.1	bd	19.9	5.7
ab	14.3	61.1	abd	22.5	22.5
c	30.4	43.5	cd	25.5	53.4
ac	25.1	96.2	acd	21.5	56.6
bc	38.2	100.8	bcd	22.0	10.4
abc	12.8	23.6	abcd	22.5	123.8

(a) Compute the pooled sample standard deviation. What does it measure in the present context? (Variability in hour-to-hour missed lead counts? Variability in shift-to-shift missed lead per hour figures?)
(b) Use the Yates algorithm and compute the fitted 2^4 factorial effects.
(c) Which of the effects are statistically detectable here? (Use individual two-sided 98% confidence limits for the effects to make this determination.) Is there a simple interpretation of this set of effects?
(d) Would you be willing to say, on the basis of your analysis in (a) through (c), that the new lead type will provide an overall reduction in the number of missed leads? Explain.
(e) Would you be willing to say, on the basis of your analysis in (a) through (c), that a switch to the new lead type will provide a reduction in missed leads for every set of plant circumstances? Explain.

12. DeBlieck, Rohach, Topf, and Wilcox conducted a replicated 3×3 factorial study of the uniaxial force required to buckle household cans. A single brand of cola cans, a single brand of beer cans, and a single brand of soup cans were used in the study. The cans were prepared by bringing them to 0°C, 22°C, or 200°C before testing. The forces required to buckle each of $m = 3$ cans for the nine different Can Type/Temperature combinations follow.

Can Type	Temperature	Force Required, y (lb)
cola	0°C	174, 306, 192
cola	22°C	150, 188, 125
cola	200°C	200, 198, 204
beer	0°C	234, 246, 300
beer	22°C	204, 339, 254
beer	200°C	414, 200, 286
soup	0°C	570, 704, 632
soup	22°C	667, 593, 647
soup	200°C	600, 620, 596

(a) Make an interaction plot of the nine combination sample means. Enhance it with error bars derived using 98% individual two-sided confidence intervals.

(b) Compute the fitted main effects and interactions from the nine combination sample means. Use these to make individual 98% confidence intervals for all of the main effects and interactions in this 3×3 factorial study. What do these indicate about the detectability of the various effects?

(c) Use Tukey's method for simultaneous comparison of main effects and give simultaneous 99% confidence intervals for all differences in Can Type main effects. Then use the same method and give simultaneous 99% confidence intervals for all differences in Temperature main effects.

13. Consider again the 2^4 factorial data set in Chapter Exercise 20 of Chapter 4. (Paper airplane flight distances collected by K. Fellows were studied there.) As a means of making the evaluation of which of the fitted effects produced by the Yates algorithm appear to be detectable, normal-plot the fitted effects. Interpret the plot.

14. Boston, Franzen, and Hoefer conducted a 2×3 factorial study of the strengths of rubber bands. Two different brands of bands were studied. From both companies, bands of three different widths were used. For each Brand/Width combination, the strengths of $m = 5$ bands of length 18–20 cm were determined by loading the bands till failure. Some summary statistics from the study are presented in the accompanying table.

Factor B *Width*	Factor A *Brand*	
1 narrow (<2 mm)	1	$\bar{y}_{11} = 2.811$ kg, $s_{11} = .0453$ kg
	2	$\bar{y}_{21} = 2.459$ kg, $s_{21} = .4697$ kg
2 medium (3.5 mm)	1	$\bar{y}_{12} = 4.164$ kg, $s_{12} = .2490$ kg
	2	$\bar{y}_{22} = 4.111$ kg, $s_{22} = .1030$ kg
3 wide (5.5 mm)	1	$\bar{y}_{13} = 8.001$ kg, $s_{13} = .8556$ kg
	2	$\bar{y}_{23} = 6.346$ kg, $s_{23} = .1924$ kg

(a) Compute s_{P} for the rubber band strength data. What is this supposed to measure?

(b) Make an interaction plot of sample means. Use error bars for the means calculated from 95% two-sided individual confidence limits. (Make use of your value of s_{P}.)

(c) Based on your plot from (b), which factorial effects appear to be distinguishable from background noise? (Brand main effects? Width main effects? Brand $\times$ Width interactions?)

(d) Compute all of the fitted factorial effects for the strength data. (Find the a_i's, the b_j's, and the ab_{ij}'s defined in Section 4.3.)

(e) To find individual 95% confidence intervals for the interactions $\alpha\beta_{ij}$, intervals of the form

$ab_{ij} \pm \Delta$ are appropriate. Find Δ. Based on this value, are there statistically detectable interactions here? How does this conclusion compare with your more qualitative answer to part (c)?

(f) To compare Width main effects, confidence intervals for the differences $\beta_j - \beta_{j'}$ are in order. Find individual 95% two-sided confidence intervals for $\beta_1 - \beta_2$, $\beta_1 - \beta_3$, and $\beta_2 - \beta_3$. Based on these, are there statistically detectable Width main effects here? How does this compare with your answer to part (c)?

(g) Redo part (f), this time using simultaneous 95% two-sided confidence intervals.

15. In Section 8.3, you were advised to choose $\frac{1}{2}$ fractions of 2^p factorials by using the generator

last factor $\leftrightarrow$ product of all other factors

For example, this means that in choosing $\frac{1}{2}$ of 2^4 possible combinations of levels of factors A, B, C, and D, you were advised to use the generator D $\leftrightarrow$ ABC. There are other possibilities. For example, you could use the generator D $\leftrightarrow$ AB.

(a) Using this alternative plan (specified by D $\leftrightarrow$ AB), what eight different combinations of factor levels would be run? (Use the standard naming convention, listing for each of the eight sets of experimental conditions to be run those factors appearing at their high levels.)

(b) For the alternative plan specified by D $\leftrightarrow$ AB, list all eight pairs of effects of factors A, B, C, and D that would be aliased. (You may, if you wish, list eight sums of the effects $\mu_{....}$, α_2, β_2, $\alpha\beta_{22}$, γ_2, . . . etc. that can be estimated.)

(c) Suppose that in an analysis of data from an experiment run according to the alternative plan (with D $\leftrightarrow$ AB), the Yates algorithm is used with $\bar{y}$'s listed according to Yates standard order for factors A, B, and C. Give four equally plausible interpretations of the eventuality that the first four lines of the Yates calculations produce large estimated sums of effects (in comparison to the other four, for example).

(d) Why might it be well argued that the choice D $\leftrightarrow$ ABC is superior to the choice D $\leftrightarrow$ AB?

16. $p = 5$ factors A, B, C, D, and E are to be studied in a 2^{5-2} fractional factorial study. The two generators D $\leftrightarrow$ AB and E $\leftrightarrow$ AC are to be used in choosing the eight ABCDE combinations to be included in the study.

(a) Give the list of eight different combinations of levels of the factors that will be included in the study. (Use the convention of naming, for each sample, those factors that should be set at their high levels.)

(b) Give the list of all effects aliased with the A main effect if this experimental plan is adopted.

17. The following are eight sample means listed in Yates standard order (left to right), considering levels of three two-level factors A, B, and C:

70, 61, 72, 59, 68, 64, 69, 69

(a) Use the Yates algorithm here to compute eight estimates of effects from the sample means.

(b) Temporarily suppose that no value for s_{P} is available. Make a plot appropriate to identifying those estimates from (a) that are likely to represent something more than background noise. Based on the appearance of your plot, which if any of the estimated effects are clearly representing something more than background noise?

(c) As it turned out, $s_{\text{P}} = .9$, based on $m = 2$ observations at each of the eight different sets of conditions. Based on 95% individual two-sided confidence intervals for the underlying effects estimated from the eight $\bar{y}$'s, which estimated effects are clearly representing something other than background noise? (If confidence intervals $\hat{E} \pm \Delta$ were to be made, show the calculation of Δ and state which estimated effects are clearly representing more than noise.)

Still considering the eight sample means, henceforth suppose that by some criteria, only the estimates ending up on the first, second, and sixth lines of the Yates calculations are considered to be both statistically detectable and of practical importance.

(d) If in fact the eight $\bar{y}$'s came from a (4-factor) 2^{4-1} experiment with generator D $\leftrightarrow$ ABC, how would one typically interpret the result that the first, second, and sixth lines of the Yates calculations (for means in standard order for factors A, B, and C) give statistically detectable and practically important values?

(e) If in fact the eight $\bar{y}$'s came from a (5-factor) 2^{5-2} experiment with generators D $\leftrightarrow$ ABC and E $\leftrightarrow$ AC, how would one typically interpret the result that the first, second, and sixth lines of the Yates calculations (for means in standard order for factors A, B, and C) give statistically detectable and practically important values?

18. A production engineer who wishes to study six two-level factors in eight experimental runs decides to use the generators D $\leftrightarrow$ AB, E $\leftrightarrow$ AC, and F $\leftrightarrow$ BC in planning a 2^{6-3} fractional factorial experiment.

(a) What eight combinations of levels of the six factors will be run? (Name them using the usual convention of prescribing for each run which of the factors will appear at their high levels.)

(b) What seven other effects will be aliased with the A main effect in the engineer's study?

19. The article "Going Beyond Main-Effect Plots" by Kenett and Vogel (*Quality Progress*, 1991) outlines the results of a 2^{5-1} fractional factorial industrial experiment concerned with the improvement of the operation of a wave soldering machine. The effects of the five factors Conveyor Speed (A), Preheat Temperature (B), Solder Temperature (C), Conveyor Angle (D), and Flux Concentration (E) on the variable

y = the number of faults per 100 solder joints (computed from inspection of 12 circuit boards)

were studied. (The actual levels of the factors employed were not given in the article.) The combinations studied and the values of y that resulted are given next.

Combination	y	Combination	y
(1)	.037	de	.351
a	.040	ade	.360
b	.014	bde	.329
ab	.042	abde	.173
ce	.063	cd	.372
ace	.100	acd	.184
bce	.067	bcd	.158
abce	.026	abcd	.131

Kenett and Vogel were apparently called in after the fact of experimentation to help analyze this nonstandard $\frac{1}{2}$ fraction of the full 2^5 factorial. The recommendations of Section 8.3 were not followed in choosing which 16 of the 32 possible combinations of levels of factors A through E to include in the wave soldering study. In fact, the generator E $\leftrightarrow$ −CD was apparently employed.

(a) Verify that the combinations listed above are in fact those prescribed by the relationship E $\leftrightarrow$ −CD. (For example, with all of A through D at their low levels, note that the low level of E is indicated by multiplying minus signs for C and D by another minus sign. Thus, combination (1) is one of the 16 prescribed by the generator.)

(b) Write the defining relation for the experiment. What is the resolution of the design chosen by the authors? What resolution does the standard choice of $\frac{1}{2}$ fraction provide? Unless there were some unspecified extenuating circumstances that dictated the choice of $\frac{1}{2}$ fraction, why does it seem to be an unwise one?

(c) Write out the 16 different differences of effects that can be estimated based on the data given. (For example, one of these is $\mu_{.....} - \gamma\delta\epsilon_{222}$, another is $\alpha_2 - \alpha\gamma\delta\epsilon_{2222}$, etc.)

(d) Notice that the combinations listed here are in Yates standard order as regards levels of factors A through D. Use the four-cycle Yates algorithm and find the fitted differences of effects. Normal-plot these and identify any statistically detectable differences. Notice that by virtue of the choice of $\frac{1}{2}$ fraction made by the engineers, the most obviously statistically significant difference is that of a main effect and a 2-factor interaction.

20. The article "Robust Design: A Cost-Effective Method for Improving Manufacturing Processes" by Kacker and Shoemaker (*AT&T Technical Journal*, 1986) discusses the use of a 2^{8-4} fractional factorial experiment in the improvement of the performance of a step in an integrated circuit fabrication process. The initial step in fabricating silicon wafers for IC devices is to grow an epitaxial layer of sufficient (and, ideally, uniform) thickness on polished wafers. The engineers involved in running this part of the production process considered the effects of eight factors (listed in the accompanying table) on the properties of the deposited epitaxial layer.

Factor A	Arsenic Flow Rate	55% (−) vs. 59% (+)
Factor B	Deposition Temperature	1210°C (−) vs. 1220°C (+)
Factor C	Code of Wafers	668G4 (−) vs. 678G4 (+)
Factor D	Susceptor Rotation	continuous (−) vs. oscillating (+)
Factor E	Deposition Time	high (−) vs. low (+)
Factor F	HC1 Etch Temperature	1180°C (−) vs. 1215°C (+)
Factor G	HC1 Flow Rate	10% (−) vs. 14% (+)
Factor H	Nozzle Position	2 (−) vs. 6 (+)

A batch of 14 wafers is processed at one time, and the experimenters measured thickness at five locations on each of the wafers processed during one experimental run. These $14 \times 5 = 70$ measurements from each run of the process were then reduced to two response variables:

y_1 = the mean of the 70 thickness measurements

y_2 = the logarithm of the variance of the 70 thickness measurements

y_2 is a measure of uniformity of the epitaxial thickness, and y_1 is (clearly) a measure of the magnitude of the thickness. The authors reported results from the experiment as shown in the accompanying table.

Combination	y_1 (μm)	y_2
(1)	14.821	−.4425
afgh	14.888	−1.1989
begh	14.037	−1.4307
abef	13.880	−.6505
cefh	14.165	−1.4230
aceg	13.860	−.4969
bcfg	14.757	−.3267
abch	14.921	−.6270
defg	13.972	−.3467
adeh	14.032	−.8563
bdfh	14.843	−.4369
abdg	14.415	−.3131
cdgh	14.878	−.6154
acdf	14.932	−.2292
bcde	13.907	−.1190
abcdefgh	13.914	−.8625

It is possible to verify that the combinations listed here come from the use of the four generators E ↔ BCD, F ↔ ACD, G ↔ ABD, and H ↔ ABC.

(a) Write out the whole defining relation for this experiment. (The grand mean will have 15 aliases.) What is the resolution of the design?

(b) Consider first the response y_2, the measure of uniformity of the epitaxial layer. Use the Yates algorithm and normal- and/or half normal-plotting (see Exercise 9) to identify statistically detectable fitted sums of effects. Suppose that only the two largest (in magnitude) of these are judged to be both statistically significant and of practical importance. What is suggested about how levels of the factors might henceforth be set in order to minimize y_2? From the limited description of the process above, does it appear that these settings require any extra manufacturing expense?

(c) Turn now to the response y_1. Again use the Yates algorithm and normal- and/or half normal-plotting to identify statistically detectable sums of effects. Which of the factors seems to be most important in determining the average epitaxial thickness? In fact, the target thickness for this deposition process was 14.5 μm. Does it appear that by appropriately choosing a level of this variable it may be possible to get the mean thickness on target? Explain. (As it turns out, the thought process outlined here allowed the engineers to significantly reduce the variability in epitaxial thickness while getting the mean on target, improving on previously standard process operating methods.)

21. Arndt, Cahill, and Hovey worked with a plastics manufacturer and experimented on an extrusion process. They conducted a 2^{6-2} fractional factorial study with some partial "replication" (the reason for the quote marks will be discussed later). The experimental factors in their study were as follows:

Factor A	Bulk Density, a measure of the weight per unit volume of the raw material used
Factor B	Moisture, the amount of water added to the raw material mix
Factor C	Crammer Current, the amperage supplied to the crammer-auger
Factor D	Extruder Screw Speed
Factor E	Front-End Temperature, a temperature controlled by heaters on the front end of the extruder
Factor F	Back-End Temperature, a temperature controlled by heaters on the back end of the extruder

Physically low and high levels of these factors were identified. Using the two generators E $\leftrightarrow$ AB and F $\leftrightarrow$ AC, 16 different combinations of levels of the factors were chosen for inclusion in a plant experiment, where the response of primary interest was the output of the extrusion process in terms of pounds of useful product per hour. A coded version of the data the students obtained is given in the accompanying table. (The data have been rescaled by subtracting a particular value and dividing by another so as to disguise the original responses without destroying their basic structure. You may think of these values as output measured in numbers of some undisclosed units above an undisclosed baseline value.)

Combination	y
ef	13.99
a	6.76
bf	20.71
abe	11.11, 11.13
ce	19.61
acf	15.73
bc	23.45
abcef	20.00
def	24.94
ad	24.03, 25.03
bdf	24.97
abde	24.29
cde	24.94, 25.21
acdf	24.32, 24.48
bcd	30.00
abcdef	33.08

(a) The students who planned this experiment hadn't been exposed to the concept of design resolution. What does Table 8.35 indicate is the best possible resolution for a 2^{6-2} fractional factorial experiment? What is the resolution of the one that the students planned? Why would they have been better off with a different plan than the one specified by the generators E$\leftrightarrow$ AB and F $\leftrightarrow$ AC?

(b) Find a choice of generators E $\leftrightarrow$ (some product of letters A through D) and F $\leftrightarrow$ (some other product of letters A through D) that provides maximum resolution for a 2^{6-2} experiment.

(c) The combinations here are listed in Yates standard order as regards factors A through D. Compute $\bar{y}$'s and then use the (four-cycle) Yates algorithm and compute 16 estimated sums of 2^6 factorial effects.

(d) When the extrusion process is operating, many pieces of product can be produced in an hour, but the entire data collection process leading to the data here took over eight hours. (Note, for example, that changing temperatures on industrial equipment requires time for parts to heat up or cool down, changing formulas of raw material means that one must let one batch clear the system, etc.) The repeat observations above were obtained from *two consecutive pieces of product*, made minutes apart, without any change in the extruder setup in between their manufacture. With this in mind, discuss why a pooled standard deviation based on these four "samples of size 2" is quite likely to underrepresent the level of "baseline" variability in the output of this process under a fixed combination of levels of factors A through F. Argue that it would have been extremely valuable to have (for example) rerun one or more of the combinations tested early in the study again late in the study.

(e) Use the pooled sample standard deviation from the repeat observations and compute (using the $p = 4$ version of formula (8.12) in Section 8.2) the plus-or-minus part of 90% two-sided confidence limits for the 16 sums of effects estimated in part (c), acting as if the value of s_P were a legitimate estimate of background variability. Which sums of effects are statistically detectable by this standard? How do you interpret this in light of the information in part (d)?

(f) As an alternative to the analysis in part (e), make a normal plot of the last 15 of the 16 estimated sums of effects you computed in part (c). Which sums of effects appear to be statistically detectable? What is the simplest interpretation of your findings in the context of the industrial problem? (What has been learned about how to run the extruding process?)

(g) Briefly discuss where to go from here if it is your job to optimize the extrusion process (maximize y). What data would you collect next, and what would you be planning to do with them?

22. The article "The Successful Use of the Taguchi Method to Increase Manufacturing Process Capability" by S. Shina (*Quality Engineering*, 1991) discusses the use of a 2^{8-3} fractional factorial experiment to improve the operation of a wave soldering process for through-hole printed circuit boards. The experimental factors and levels studied were as shown in the accompanying table.

Factor A	Preheat Temperature	180° (−) vs. 220° (+)
Factor B	Solder Wave height	.250 (−) vs. .400 (+)
Factor C	Wave Temperature	490° (−) vs. 510° (+)
Factor D	Conveyor Angle	5.0 (−) vs. 6.1 (+)
Factor E	Flux Type	A857 (−) vs. K192 (+)
Factor F	Direction of Boards	0 (−) vs. 90 (+)
Factor G	Wave Width	2.25 (−) vs. 3.00 (+)
Factor H	Conveyor Speed	3.5 (−) vs. 6.0 (+)

The generators F $\leftrightarrow$ −CD, G $\leftrightarrow$ −AD, and H $\leftrightarrow$ −ABCD were used to pick 32 different combinations of levels of these factors to run. For each combination, four special test printed circuit boards were soldered, and the lead shorts per board, y_1, and touch shorts per board, y_2, were counted, giving the accompanying data. (The data here and on page 644 are exactly as given in the article, and we have no explanation for the fact that some of the numbers do not seem to have come from division of a raw count by 4.)

Combination	y_1	y_2
(1)	6.00	13.00
agh	10.00	26.00
bh	10.00	12.00
abg	8.50	14.00
cfh	1.50	18.75
acfg	.25	16.25
bcf	1.75	25.75
abcfgh	4.25	18.50
dfgh	6.50	6.50

(*continued*)

Combination	y_1	y_2
adf	.75	.00
bdfg	3.50	1.00
abdfh	3.25	6.50
cdg	6.00	7.25
acdh	9.50	11.25
bcdgh	6.25	10.00
abcd	6.75	12.50
e	20.00	29.25
aegh	16.50	31.25
beh	17.25	28.75
abeg	19.50	41.25
cefh	9.67	21.33
acefg	2.00	10.75
bcef	5.67	28.67
abcefgh	3.75	35.75
defgh	6.00	22.70
adef	7.30	25.70
bdefg	8.70	30.00
abdefh	9.00	29.70
cdeg	19.30	32.70
acdeh	26.70	25.70
bcdegh	17.70	45.30
abcde	10.30	37.00

(a) Verify that the 32 combinations of levels of the factors A through H listed here are those that are prescribed by the choice of generators. (For each combination of levels of the factors A through E, determine what levels of F, G, and H are prescribed by the generators and check that such a combination is listed.)

(b) Use the generators given here and write out the whole defining relation for this study. (You will end with I aliased with seven other strings of letters.) What is the resolution of the design used in this study? According to Table 8.35, what was possible in terms of resolution for a 2^{8-3} study? Could the engineers in charge here have done better at containing the ambiguity that unavoidably follows from use of a fractional factorial study?

(c) Note that the 32 combinations of the 8 factors above are listed in Yates standard order as regards Factors A through E (ignoring F, G, and H). By some means (using a statistical analysis package like MINITAB, implementing spreadsheet calculations, or doing the 5-cycle Yates algorithm "by hand") find the estimated sums of effects for the response y_1. Normal-plot the last 31 of these. You should find that the largest of these would be the CD 2-factor interaction, the E main effect, and the CDE 3-factor interaction if only 5 factors were involved (instead of 8). These are all positive and clearly larger in magnitude than the other estimates. If possible, give a simple interpretation of this in light of the alias structure specified by the defining relation you found in part (b).

(d) Now find and normal-plot the estimated sums of effects for the response y_2. (Normal-plot 31 estimates.) You should find the estimate corresponding to the E main effect plus aliases to be positive, larger in magnitude than the rest, and detectably nonzero.

(e) In light of your answers to (c) and (d), the signs of the fitted linear combinations of effects, and a desire to reduce both y_1 and y_2 to the minimum values possible, what combination of levels of the factors do you tentatively recommend here? Is the combination of levels that you see as promising one that is among the 32 tested? If it is not, how would you recommend proceeding in the real manufacturing scenario? (Would you, for example, order that any permanent process changes necessary to the use of your promising combination be adopted immediately?)

The original article reported a decrease in solder defects by nearly a factor of 10 in this process as a result of what was learned from this experiment.

23. In the situation of Exercise 22, the 32 different combinations of levels of factors A through H were run in the order listed. In fact, the first 16 runs were made by one shift of workers, and the last 16 were made by a second shift.

(a) In light of the material in Chapter 2 on experiment planning and the formal notion of confounding, what risk of a serious logical flaw did the engineers run in the execution of their experiment? (How would possible shift-to-shift differences show up in the data from an experiment run like this? One of the main things learned from the experiment was that factor E was very important. Did the engineers run the risk of clouding their view of this important fact?) Explain.

(b) Devise an alternative plan that could have been used to collect data in the situation of Exercise 22 without completely confounding the effects of Flux and Shift. Continue to use the 32 combinations of the original factors listed in Exercise 22, but give a better assignment of 16 of them to each shift. (*Hint:* Think of Shift as a ninth factor, pick a sensible generator, and use it to put half of the 32 combinations in each shift. There are a variety of possibilities here.)

(c) Discuss in qualitative terms how you would do data analysis if your suggestion in (b) were to be followed.

24. The article "Computer Control of a Butane Hydrogenolysis Reactor" by Tremblay and Wright (*The Canadian Journal of Chemical Engineering*, 1974) contains an interesting data set concerned with the effects of $p = 3$ process variables on the performance of a chemical reactor. The factors and their levels were as follows:

Factor A	Total Feed Flow (cc/sec at STP)	50 (−) vs. 180 (+)
Factor B	Reactor Wall Temperature (°F)	470 (−) vs. 520 (+)
Factor C	Feed Ratio (Hydrogen/Butane)	4 (−) vs. 8 (+)

The data had to be collected over a four-day period, and two combinations of the levels of factors A, B, and C above were run each day along with a center point—a data point with Total Feed Flow 115, Reactor Wall Temperature 495, and Feed Ratio 6. The response variable was

$$y = \text{percent conversion of butane}$$

and the data in the accompanying table were collected.

Day	Feed Flow	Wall Temp.	Feed Ratio	Combination	y
1	115	495	6	—	78
1	50	470	4	(1)	99
1	180	520	8	abc	87
2	50	520	4	b	98
2	180	470	8	ac	18
2	115	495	6	—	87
3	50	520	8	bc	95
3	180	470	4	a	59
3	115	495	6	—	90
4	50	470	8	c	76
4	180	520	4	ab	92
4	115	495	6	—	89

(a) Suppose that to begin with, you ignore the fact that these data were collected over a period of four days and simply treat the data as a complete 2^3 factorial augmented with a repeated center point. Analyze these data using the methods of this chapter. (Compute s_P from four center points. Use the Yates algorithm and the eight corner points to compute fitted 2^3 factorial effects. Then judge the statistical significance of these using appropriate 95% two-sided confidence limits based on s_P.) Is any simple interpretation of the experimental results in terms of factorial effects obvious?

According to the authors, there was the possibility of "process drift" during the period of experimentation. The one-per-day center points were added to the 2^3 factorial at least in part to provide some check on that possibility, and the allocation of two ABC combinations to each day was very carefully done in order to try to minimize the possible confounding introduced by any Day/Block effects. The rest of this problem considers analyses that might be performed on the experimenters' data in recognition of the possibility of process drift.

(b) Plot the four center points against the number of the day on which they were collected. What possibility is at least suggested by your plot? Would the plot be particularly troubling if your experience with this reactor told you that a standard deviation of around 5(%) was to be expected for values of y from consecutive runs of the reactor under fixed operating conditions on a given day? Would the plot be troubling if your experience with this reactor told you that a standard deviation of around 1(%) was to be expected for values of y from consecutive runs of the reactor under fixed operating conditions on a given day?

(c) The four-level factor Day can be formally thought of in terms of two extra two-level factors—say, D and E. Consider the choice of generators D $\leftrightarrow$ AB and E $\leftrightarrow$ BC for a 2^{5-2} fractional factorial. Verify that the eight combinations of levels of A through E prescribed by these generators divide the eight possible combinations of levels of A through C up into the four groups of two corresponding to the four days of experimentation. (To begin with, note that both A low, B low, C low and A high, B high, C high correspond to D high and E high. That is, the first level of Day can be thought of as the D high and E high combination.)

(d) The choice of generators in (c) produces the defining relation I $\leftrightarrow$ ABD $\leftrightarrow$ BCE $\leftrightarrow$ ACDE. Write out, on the basis of this defining relation, the list of eight groups of aliased 2^5 factorial effects. Any effect involving factors A, B, or C with either of the letters D (δ) or E (ϵ) in its name represents some kind of interaction with Days. Explain what it means for there to be no interactions with Days. Make out a list of eight smaller groups of aliased effects that are appropriate supposing that there are no interactions with Days.

(e) Allowing for the possibility of Day (Block) effects, it does not make sense to use the center points to compute s_{P}. However, one might normal-plot (or half normal-plot) the fitted effects from (a). Do so. Interpret your plot, supposing that there were no interactions with Days in the reactor study. How do your conclusions differ (if at all) from those in (a)?

(f) One possible way of dealing with the possibility of Day effects in this particular study is to use the center point on each day as a sort of baseline and express each other response as a deviation from that baseline. (If on day i there is a Day effect γ_i, and on day i the mean response for any combination of levels of factors A through C is $\mu_{\text{comb}} + \gamma_i$, the mean of the difference $y_{\text{comb}} - y_{\text{center}}$ is $\mu_{\text{comb}} - \mu_{\text{center}}$; one can therefore hope to see 2^3 factorial effects uncontaminated by additive Day effects using such differences in place of the original responses.) For each of the four days, subtract the response at the center point from the other two responses and apply the Yates algorithm to the eight differences. Normal-plot the fitted effects on the (difference from the center point mean) response. Is there any substantial difference between the result of this analysis and that for the others suggested in this problem?

25. The article "Including Residual Analysis in Designed Experiments: Case Studies" by W. H. Collins and C. B. Collins (*Quality Engineering*, 1994) contains discussions of several machining experiments concerned with surface finish. Given here are the factors and levels studied in (part of) one of those experiments on a particular lathe.

	Factor	Levels
A	Speed	2500 RPM ($-$) vs. 4500 RPM ($+$)
B	Feed	.003 in/rev ($-$) vs. .009 in/rev ($+$)
C	Tool Condition	New ($-$) vs. Used (after 250 parts) ($+$)

$m = 2$ parts were turned on the lathe for each of the 2^3 different combinations of levels of the 3 factors, and surface finish measurements, y, were made on these. (y is a measurement of the vertical distance traveled by a probe as it moves horizontally across a particular 1 inch section of the part.) Next are some summary statistics from the experiment.

Combination	$\bar{y}$	s	Combination	$\bar{y}$	s
(1)	33.0	0.0	c	35.5	6.4
a	45.5	7.8	ac	44.0	7.1
b	222.5	4.9	bc	216.5	6.4
ab	241.5	4.9	abc	216.5	0.7

(a) Find s_{P} and its degrees of freedom. What does this quantity intend to measure?
(b) 95% individual two-sided confidence limits for the mean surface finish measurement for a part turned under a given set of conditions are of the form $\bar{y}_{ijk} \pm \Delta$. Based on the value of s_{P} found above, find Δ.
(c) Would you say that the mean surface finish measurements for parts of types "(1)" and "a" are detectably different? Why or why not? (Show appropriate calculations.)
(d) 95% individual two-sided individual confidence limits for the 2^3 factorial effects in this study are of the form $\hat{E} \pm \Delta$. Find Δ.
(e) Compute the 2^3 factorial fitted effects for the "all high" combination (abc).
(f) Based on your answers to parts (d) and (e), which of the main effects and/or interactions do you judge to be statistically detectable? Explain.
(g) Give the practical implications of your answer to part (f). (How do you suggest running the lathe if small y and minimum machining cost are desirable?)
(h) Suppose you were to judge only the B main effect to be both statistically detectable and of practical importance in this study. What surface finish value would you then predict for a part made at a 2500 RPM speed and a .009 in/rev feed rate using a new tool?

26. Below are 2^4 factorial data for two response variables taken from the article "Chemical Vapor Deposition of Tungsten Step Coverage and Thickness Uniformity Experiments" by J. Chang (*Thin Solid Films*, 1992). The experiment concerned the blanket chemical vapor deposition of tungsten in the manufacture of integrated circuit chips. The factors studied were as follows:

A	Chamber Pressure	8 (−) vs. 9 (+)
B	H_2 Flow	500 (−) vs. 1000 (+)
C	SiH_4 Flow	15 (−) vs. 25 (+)
D	WF_6 Flow	50 (−) vs. 60 (+)

The pressure is measured in Torr and the flows are measured in standard cm^3/min. The response variable y_1 is the "percent step coverage," 100 times the ratio of tungsten film thickness at the top of the side wall to the bottom of the side wall (large is good). The response variable y_2 is an "average sheet resistance" (measured in mΩ).

Combination	y_1	y_2	Combination	y_1	y_2
(1)	73	646	d	83	666
a	60	623	ad	80	597
b	77	714	bd	100	718
ab	90	643	abd	85	661
c	67	360	cd	77	304
ac	78	359	acd	90	309
bc	100	335	bcd	70	360
abc	77	318	abcd	75	318

(a) Make a normal plot of the 15 fitted effects $a_2, b_2, \ldots, abcd_{2222}$ as a means of judging the statistical detectability of the effects on the response, y_1. Interpret this plot and say what is indicated about producing good "percent step coverage."
(b) Repeat part (a) for the response variable y_2.

Now suppose that instead of a full factorial study, only the half fraction with defining relation $D \leftrightarrow ABC$ had been conducted.

(c) Which 8 of the 16 treatment combinations would have been run? List these combinations in Yates standard order as regards factors A, B, and C and use the (3-cycle Yates algorithm) to compute the 8 estimated sums of effects that it is possible to derive from these 8 treatment combinations for response y_2. Verify that each of these 8 estimates is the sum of two of your fitted effects from part (b). (For example, you should find that the first estimated sum here is $\bar{y}_{....} + abcd_{2222}$ from part (b).)

(d) Normal-plot the last 7 of the estimated sums from (c). Interpret this plot. If you had only the data from this 2^{4-1} fractional factorial, would your subject-matter conclusions be the same as those reached in part (b), based on the full 2^4 data set?

27. An engineer wishes to study seven experimental factors, A, B, C, D, E, F and G, each at 2 levels, using only 16 combinations of factor levels. He plans initially to use generators E $\leftrightarrow$ ABCD, F $\leftrightarrow$ ABC, and G $\leftrightarrow$ BCD.

(a) With this initial choice of generators, what 16 combinations of levels of the seven factors will be run?

(b) In a 2^{7-3} fractional factorial, each effect is aliased with 7 other effects. Starting from the engineer's choice of generators, find the defining relation for his study. (You will need not only to consider products of pairs but also a product of a triple.)

(c) An alternative choice of generators is E $\leftrightarrow$ ABC, F $\leftrightarrow$ BCD, G $\leftrightarrow$ ABD. This choice yields the defining relation

$$\mathrm{I} \leftrightarrow \mathrm{ABCE} \leftrightarrow \mathrm{BCDF} \leftrightarrow \mathrm{ABDG}$$
$$\leftrightarrow \mathrm{ADEF} \leftrightarrow \mathrm{CDEG} \leftrightarrow \mathrm{ACFG} \leftrightarrow \mathrm{BEFG}$$

Which is preferable, the defining relation in part (b), or the one here? Why?

28. The article "Establishing Optimum Process Levels of Suspending Agents for a Suspension Product" by A. Gupta (*Quality Engineering*, 1997–1998) discussed an unreplicated fractional factorial experiment. The experimental factors and their levels in the study were:

A	Method of Preparation	Usual (−) vs. Modified (+)
B	Sugar Content	50% (−) vs. 60% (+)
C	Antibiotic Level	8% (−) vs. 16% (+)
D	Aerosol	.4% (−) vs. .6% (+)
E	CMC	.2% (−) vs. .4% (+)

The response variable was

y = separated clear volume (%) for a suspension of antibiotic after 45 days

and the manufacturer hoped to find a way to make y small. The experimenters failed to follow the recommendation in Section 8.3 for choosing a best half fraction of the factorial and used the generator E $\leftrightarrow$ ABC (instead of the better one E $\leftrightarrow$ ABCD).

(a) In what sense was the experimental plan used in the study inferior to the one prescribed in Section 8.3? (How is the one from Section 8.3 "better"?)

The Yates algorithm applied to the 16 responses given in the paper produced the 16 fitted sums of effects:

mean + *alias* = 37.563	D + *alias* = −7.437
A + *alias* = .187	AD + *alias* = .937
B + *alias* = 2.437	BD + *alias* = .678
AB + *alias* = .312	ABD + *alias* = .812
C + *alias* = −1.062	CD + *alias* = 1.438
AC + *alias* = .312	ACD + *alias* = .062
BC + *alias* = −1.187	BCD + *alias* = .062
ABC + *alias* = −2.063	ABCD + *alias* = −.062

(a) Make a normal plot of the last 15 of these fitted sums.

(b) If you had to guess (based on the results of this experiment) the order of the magnitudes of the five main effects (A, B, C, D and E) from smallest to largest, what would you guess? Explain.

(c) Based on the normal plot in (b), which sums of effects do you judge to be statistically detectable? Explain.

(d) Based on your answers to (c) and (d), how do you suggest that suspensions of this antibiotic be made in order to produce small y? What mean y do you predict if your recommendations are followed?

(e) Actually, the company that ran this study planned to make suspensions using both high and low levels of antibiotic (factor C). Does your answer to (d) suggest that the company needs to use different product formulations for the two levels of antibiotic? Explain.

29. The paper "Achieving a Target Value for a Manufacturing Process," by Eibl, Kess, and Pukelsheim (*Journal of Quality Technology*, 1992) describes a series of experiments intended to guide the adjustment of a paint coating process. The first of these was a 2^{6-3} fractional factorial study. The experimental factors studied were as follows (exact levels of these factors are not given in the paper, presumably due to corporate security considerations):

A	Tube Height	low (−) vs. high (+)
B	Tube Width	low (−) vs. high (+)
C	Paint Viscosity	low (−) vs. high (+)
D	Belt Speed	low (−) vs. high (+)
E	Pump Pressure	low (−) vs. high (+)
F	Heating Temperature	low (−) vs. high (+)

The response variable was a paint coating thickness measurement, y, whose units are mm. $m = 4$ workpieces were painted and measured for each of the $r = 8$ combinations of levels of the factors studied. The $r = 8$ samples of size $m = 4$ produced a value of $s_{\text{P}} = .118$ mm.

(a) Suppose that you wish to attach a precision to one of the $r = 8$ sample means obtained in this study. This can be done using 95% two-sided confidence limits of the form $\bar{y} \pm \Delta$. Find Δ.

(b) Following are the mean thicknesses measured for the combinations studied, listed in Yates standard order as regards levels of factors A, B, and C. Use the Yates algorithm and find eight estimated (sums of) effects.

A	B	C	$\bar{y}$
−	−	−	.98
+	−	−	1.58
−	+	−	1.13
+	+	−	1.74
−	−	+	1.49
+	−	+	.84
−	+	+	2.18
+	+	+	1.45

(c) Two-sided confidence limits based on the estimated (sums of) effects calculated in part (b) are of the form $\hat{E} \pm \Delta$. Find Δ if (individual) 95% confidence is desired.

(d) Based on your answer to (c), list those estimates from part (b) that represent statistically detectable (sums of) effects.

In fact, the experimental plan used by the investigators had generators D $\leftrightarrow$ AC, E $\leftrightarrow$ BC, and F $\leftrightarrow$ ABC.

(e) Specify the combinations (of levels of the experimental factors A, B, C, D, E and F) that were included in the experiment.

(f) Write out the whole defining relation for this study. (You will need to consider here not only products of pairs but a product of a triple as well. The grand mean is aliased with seven other effects.)

(g) In light of your answers to part (d) and the aliasing pattern here, what is the simplest possible potential interpretation of the results of this experiment?