CHAPTER 9

Split-Unit, Split-Lot, and Related Designs

In this chapter we consider designs that do not have a single error structure. We cover a split-unit design (better known as a split-plot design), a split-lot design, and other similar designs such as a strip-plot design. There has been much research activity regarding these various types of designs during the past 10 years. Although designs such as a split-plot design have not been accorded much space in many books on experimental design, Bisgaard (2000) stated that split-plot designs play a key role in the industrial application of factorial experiments and their use is much more prevalent than the literature on design of experiments in engineering would suggest. Box, Hunter, and Hunter (2005, p. 336) quoted Cuthbert Daniel, famous statistician, author, and consultant of some years ago, as having stated "All industrial experiments are split-plot experiments" although they recognize that may be a slight exaggeration. Somewhat similarly, Langhans, Goos, and Vandebroek (2005) stated that split-plot designs and their properties receive much less attention in the chemometrics literature than the designs receive in the general statistical literature. Robust parameter designs were discussed in Chapter 8 and Bingham and Sitter (2003) discussed the use of fractional factorial split-plot (FFSP) designs in robust parameter experiments. (Fractional factorial split-plot designs are discussed in Section 9.1.3.) Emptage, Hudson-Curtis, and Sen (2003) discussed the treatment of microarrays as split-plot experiments, with that application area possibly also underutilized, although this is a relatively new field. (Microarray experiments are discussed in Section 13.21.)

Goos and Vandebroek (2004) showed that split-plot designs will often outperform completely randomized designs in terms of D-efficiency and G-efficiency, and Goos and Vandebroek (2003) discussed the construction of D-optimal split-plot designs. As with other types of designs, the response variable is assumed to have a normal distribution, but Robinson, Myers, and Montgomery (2004) discussed the analysis of data from industrial split-plot experiments when the response variable does not have a normal distribution.

Modern Experimental Design By Thomas P. Ryan

We will use the terms "split unit" and "split plot" interchangeably in the sections that follow, with the choice between them for discussing a particular type of design determined in part by the terminology that has been used in the literature.

9.1 SPLIT-UNIT DESIGN

This design has for decades been referred to as a *split-plot* design; this was the original name given to it as this design along with many designs were originally used in agricultural applications and a "plot" was a plot of land. That is, a small plot of land was literally "split," so that a "whole plot" was split into two or more pieces, which were called subplots. These designs are now used heavily in industry rather than mainly in agriculture, however, sometimes even unknowingly when the error structure is not understood. Therefore, following Mead (1988), Giesbrecht and Gumpertz (2004), and Ramírez (2004), it seems more appropriate to call the design a *split-unit* design, while realizing that there are still agricultural experiments being performed for which the term "split-plot design" is of course appropriate.

Another reason for eschewing the term "plot" in discussing a split-unit design in general and its variations is that a normal probability plot is often used in analyzing the data from experiments using these and other designs, and that plot is a graph, not something that can be "split." Thus, it seems desirable to avoid the use of the word to mean two different things, especially when discussing the analysis of data from a designed experiment.

Despite these presumably persuasive arguments for a change in terminology, we will still refer to a split-plot design at times in this chapter, especially in discussing journal articles in which the term is used in the title of the article and/or in the article contents. Suggestions for permanent names for the designs presented in this chapter are given in Section 9.4. In this chapter we will often use WP to denote "whole plot" and SP to denote "split-plot."

Consider a very simple example of two fixed factors, each at two levels. We will first assume that all four combinations of the factor levels are feasible, and the level changes of each factor can be easily made. If the four combinations are run in random order, we have a 2^2 design. For example, the order in which the treatment combinations might be run could be A_1B_2, A_2B_1, A_2B_2, and A_1B_1.

Now assume that although each factor is suspected of having a significant effect on the response, one of the two factors is definitely of secondary interest. Also, assume that this factor is hard to change. One of the levels of this factor is randomly selected and then used in combination with each of the two levels of the other factor, which are also randomly selected. Then this process is repeated for the second level of the first factor. Thus, the order in which the treatment combinations are run could be as follows, assuming that factor A is the hard-to-change factor: A_2B_2, A_2B_1, A_1B_1, and A_1B_2.

Notice that this last sequence of treatment combinations and the one given previously *could* of course be the same, but the data would still have to be analyzed differently because of the restricted randomization in the second case. That is, there are only eight possible sequences of treatment combinations with the restriction, whereas there are 24 possible sequences without the restriction.

With the restriction, is this a randomized block design (RCB) since we seem to be blocking on the first factor? Recall from Section 3.1 that the experimental units are considered to be more homogeneous within blocks than between blocks for the RCB design. Recall also that a blocking variable is considered to be an extraneous factor and that blocks are generally considered to be random. Thus, although the experimental layout coincides with the layout that would be used in a randomized block design, we clearly cannot analyze the data as having come from such a design because the conditions are not the same.

Because of the restricted randomization, this also is not a 2^2 design. Then what is left? Although the description of the scenario did not suggest that a plot or unit was literally split (as in splitting a plot of land), the data would be logically analyzed in that manner because of the manner in which the experiment was conducted.

Recall from Appendix C to Chapter 4 that in a 2^k design each effect is estimated with the same precision. This does not happen with a split-plot design as sub-plot factors are generally estimated with greater precision than are whole-plot factors. This should be intuitively apparent for an agricultural experiment in which sub-plots would be strips of land within a whole plot, as the sub-plots would certainly be more homogeneous in terms of land fertility than are the whole plots, especially if the whole plots are large and so the distance between the centers of the two plots is large.

In addition to being quite intuitive, at least for agricultural experiments, it can be shown mathematically that the variance of any whole-plot effect estimate must exceed the variance of any subplot effect estimate for any 2^k design. Bisgaard and de Pinho (2004) state that whole-plot factors and their interactions have a variance of $\frac{4}{N}(2^q\sigma_1^2 + \sigma_0^2)$, whereas subplot factors and their interactions have a variance of $\frac{4}{N}(\sigma_0^2)$. Here σ_1^2 and σ_0^2 denote the whole plot and subplot error variance, respectively.

Sub-plot effects will not necessarily be estimated with greater efficiency in industrial experiments or various other types of experiments, however. It is obvious from the expressions of the variances that this could not happen if σ_0^2 and σ_1^2 were estimated individually using the same data, but this does not happen as a sum of squares estimates a mean square, which is a linear combination of variance components for whole-plot factors and interactions. Thus, the data are used in one way to estimate $\frac{4}{N}(\sigma_0^2)$ and in another way to estimate $\frac{4}{N}(2^q\sigma_1^2 + \sigma_0^2)$.

This issue of whole-plot effect and sub-plot effect variance estimates is addressed by Giesbrecht and Gumpertz (2004, p. 169), who point out that the estimate of the split-plot error may exceed the estimate of the whole-plot error, with opinions differing as to how to proceed when that occurs. Their opinion is to proceed as if the assumed model were valid, with that model containing the whole-plot factor(s), the subplot factor(s), the interaction(s) between them, and the two error terms.

To illustrate the analysis for both cases (complete randomization and restricted randomization for the whole-plot factor), we will consider the following simple example.

Example 9.1

We will assume that factor A is the hard-to-change factor and factor B is not hard to change, with the experiment being such that material (e.g., a board) is divided into two pieces and the two levels of factor *A* applied to the two pieces, one level to each piece. Then the pieces are further subdivided and each of the two levels of factor *B* are applied to the subdivided pieces. Three pieces of the original length (e.g., three full boards) are used. The data are given below.

A	B	Observations (replications)		
1	1	2.5	2.4	2.6
	2	2.7	2.6	2.5
2	1	2.3	2.3	2.4
	2	2.7	2.7	2.8

If the data are improperly analyzed as a 2^2 design with three replications, the results are as follows.

Analysis of Variance for Y

Source	DF	SS	MS	F	P
A	1	0.0008	0.0008	0.13	0.733
B	1	0.1875	0.1875	28.13	0.001
A*B	1	0.0675	0.0675	10.13	0.013
Error	8	0.0533	0.0067		
Total	11	0.3092			

The proper analysis of the data as having come from a split-plot design is not easily achieved. The following statement by Box (1995–1996) needs some clarification: "The numerical calculation of the analysis of variance for split-plot experiments—computation of the degrees of freedom, sums of squares and mean squares—is the same as for any other design and can be performed by any of the many computer programs now available." Presumably, what was meant was that the *basic methods* for computing sums of squares, and so on are the same as for other designs. Other designs, such as full factorial and fractional factorial designs, do not have two or more error terms, however, so the *direct production* of ANOVA tables is not possible with most statistical software.

As Potcner and Kowalski (2004) explained, however, statistical packages can be tricked into performing the correct analysis by assuming a nested model and forcing a nested model analysis. This requires, of course, that the software package being used has nested design capability but unfortunately it is either very difficult or impossible to do the analysis with most software packages, as explained in Section 7.2.

Thus, the analysis of data from a split-plot design can be very difficult. Here, we will take a more straightforward but slightly cumbersome approach and discuss how the analysis can be performed by starting with the analysis without the randomization restriction and simply decomposing the error term into the two appropriate error terms: the whole-plot error and the subplot error. The other sums of squares in the ANOVA are the same as those assuming a completely randomized design.

In decomposing the error term given above, we have to make a decision as to whether or not to isolate the replication factor. Montgomery (1996, p. 524), for example, does so as the replicates were run over days, so it would be appropriate to treat these as blocks. Potcner and Kowalski (2004) in a similar example do not do so. We will initially follow the approach of Potcner and Kowalski (2004), especially since replications was not isolated in the above analysis and we will assume that there is no compelling reason for them to be treated as blocks.

This means that the whole-plot error will have degrees of freedom determined as follows. The degrees of freedom will be the interaction degrees of freedom, replicates $\times A$ since A is the whole-plot factor, plus the replicates degrees of freedom. That is, $2 \times 1 + 2 = 4$. The subplot error degrees of freedom will then be $12 - 4 = 8$.

The corresponding error sums of squares are not difficult to compute by hand, especially since only one has to be computed directly, with the other obtained by subtraction after the "wrong analysis" (i.e., the one given above) is computer generated to produce the error sum of squares that is to be decomposed. The sum of squares for replications is computed the same way that the treatment sum of squares is computed in one-way ANOVA and the replicates $\times A$ sum of squares is computed as the sum of squares for the "cells," with each cell formed by taking a replicate and a level of A (so there are six cells), minus the sum of the squares for the replications and the sum of squares of the whole-plot factor, A. Applied to this example we have

$$\text{SS(replicates} \times A) = \text{SS (cells)} - \text{SS(replications)} - \text{SS}(A)$$

with

$$\begin{aligned}
\text{SS(cells)} &= (2.5 + 2.7)^2/2 + (2.4 + 2.6)^2/2 \\
&\quad + (2.6{+}2.5)^2/2{+}(2.3{+}2.7)^2/2{+}(2.3{+}2.7)^2/2{+}(2.4{+}2.8)^2/2 \\
&\quad - (\text{sum of all observations})^2/12 \\
&= 77.5450 - 77.5208 \\
&= 0.0242 \\
\text{SS(replications)} &= (2.5 + 2.7 + 2.3 + 2.7)^2/4 + (2.4 + 2.6 + 2.3 + 2.7)^2/4 \\
&\quad + (2.6 + 2.5 + 2.4 + 2.8)^2/4 - (\text{sum of all observations})^2/12 \\
&= 77.5325 - 77.5208 \\
&= 0.0117
\end{aligned}$$

Therefore,

$$\begin{aligned}
\text{SS(replicates} \times A) &= \text{SS(cells)} - \text{SS(replications)} - \text{SS}(A) \\
&= 0.0242 - 0.0117 - 0.0008 \\
&= 0.0117
\end{aligned}$$

The latter would be the whole-plot error sum of squares if replications were one of the components of the ANOVA table. Since we are not using that approach for this example, we add back SS(replications) to obtain SS(whole-plot error) = 0.0234.

We now have what we need to begin constructing the ANOVA table, which leads to

Analysis of Variance for Y

Source	DF	SS	MS	F	P
A	1	0.0008	0.0008	0.138	0.729
Error (WP)	4	0.0234	0.0058	*	*
B	1	0.1875	0.1875	25.00	0.008
A*B	1	0.0675	0.0675	9.00	0.040
Error (SB)	4	0.0299	0.0075	*	*
Total	11	0.3091			

A somewhat different picture emerges when the data are analyzed correctly. Whereas the A*B interaction had a p-value of .013 when the randomization restriction was ignored, now the p-value is much closer to .05, so the evidence that the interaction effect exists is not as strong.

The difference in the conclusions drawn with the wrong analysis and the conclusions made with the proper analysis can be much greater than the difference in this example. This was illustrated by Potcner and Kowalski (2004) who showed that a significant main effect in the complete randomization analysis can become a non-significant whole-plot main effect when the split-plot analysis is performed, and a non-significant main effect in the complete randomization analysis can become a significant subplot main effect when the split-plot analysis is performed. This was illustrated by their first example.

Similarly, Lucas and Hazel (1997) compared a completely randomized design and a split-plot design for the same experimental situation. This provides a tutorial on the use of split-plot designs. Lucas (1999) states that in his experience the most common situation that requires the use of a split-plot design is when there is a single hard-to-change factor. Czitrom (1997) gave an example of such an experiment, and the reader is asked to analyze the data in Exercise 9.1. Goos and Vandebroek (2001) gave an algorithm for constructing D-optimal split-plot designs (optimal designs are covered in Section 13.7), and Goos (2002) discussed optimal split-plot designs, including designs with hard-to-change factors.

Hard-to-change factors were discussed extensively in Section 4.19. Even though the industrial use of split-plot designs motivated by the recognition of hard-to-change factors may seem to have been motivated by research articles during the past 10 years, Daniel (1976, p. 270) discussed the use of a split-plot design with one hard-to-change factor and one factor that is not hard to change. Thus, the idea of using a split-plot design when there is at least one hard-to-change factor is not of recent origin.

One of the best known nonagricultural applications of a split-plot design was given by Box and Jones (1992), which is incorrectly described by Miller (1997) as a strip-plot experiment. (In Section 9.3 we try to clarify somewhat the fine difference between the designs presented in this chapter.)

9.1.1 Split-Plot Mirror Image Pairs Designs

There are many ways in which a split-plot design could be constructed because we can view this, analogous to the discussion in Section 8.2, as a "product design." That is, there is a separate design for the whole-plot factors and one for the subplot factors. One possibility is a *split-plot mirror image pairs design* (SPMIP), which is a design such that there are two subplot runs for each whole-plot run that are mirror images. Of course "mirror image" means two, so we are talking about two-level designs, at least for the subplot factors. These designs have been considered by, in particular, Tyssedal, Kulahci, and Bisgaard (2005) and Tyssedal and Kulahci (2005).

Of course the simplest such design would be a design with a single subplot factor and two levels of that factor: -1 and $+1$. For more complex designs, it is useful to view the general form of the design matrix, given by Tyssedal and Kulahci (2005) for example, in partitioned form as

$$\begin{bmatrix} \boldsymbol{W} & \boldsymbol{S} \\ \boldsymbol{W} & -\boldsymbol{S} \end{bmatrix}$$

Clearly, we need only focus on the form of $[\boldsymbol{W} \quad \boldsymbol{S}]$ since the other parts of the full partitioned design matrix are defined from this submatrix. Many possibilities exist for the form of $[\boldsymbol{W} \quad \boldsymbol{S}]$, including a Plackett–Burman design. Tyssedal and Kulahci (2005) used such a design in illustrating their procedure for computing effect estimates for SPMIP designs.

9.1.2 Split-Unit Designs in Industry

Although the split-plot design originated in agriculture, as stated in Section 9.1, it is now commonly used in industry because of the cost savings relative to other types of designs (Bisgaard and Sutherland, 2003–2004).

In this section we look at a couple of industrial applications of split-unit experiments.

Example 9.2

Bisgaard, Fuller, and Barrios (1996) presented a split-plot experiment in which the surface of a security paper was modified via plasma treatment to make it more susceptible to ink. There were four factors that were believed to affect plasma creation—pressure, power, gas flow rate, and type of gas—in addition to the paper type. A 2^5 design was ruled out for the following reason. The creation of plasma requires a vacuum, with this being done by placing a sample of security paper in a reactor and pumping out the air. This process takes approximately half an hour, so running a 2^5 design would require about 16 hours. To reduce the labor and the time required to run the experiment, the decision was made to place two types of papers in the reactor at the same time.

The paper type is then the subplot factor and the other four factors are run as a 2^4 experiment. The data are given below, with factor E of course being the subplot factor.

				E	
A	B	C	D	−	+
−	−	−	−	48.6	57.0
+	−	−	−	41.2	38.2
−	+	−	−	55.8	62.9
+	+	−	−	53.5	51.3
−	−	+	−	37.6	43.5
+	−	+	−	47.2	44.8
−	+	+	−	47.2	54.6
+	+	+	−	48.7	44.4
−	−	−	+	5.0	18.1
+	−	−	+	56.8	56.2
−	+	−	+	25.6	33.0
+	+	−	+	41.8	37.8
−	−	+	+	13.3	23.7
+	−	+	+	47.5	43.2
−	+	+	+	11.3	23.9
+	+	+	+	49.5	48.2

Since this is an unreplicated design, there is not an error sum of squares to decompose into two components. Instead, two normal probability plots will be used. We will also analyze these data "by hand," since statistical software in general is inadequate to the task. If we erroneously analyze the data as having come from a 2^5 design, we obtain the following results if we assume that all three-factor and higher-order interactions are not real.

Estimated Effects and Coefficients for Y (coded units)

Term	Effect	Coef	SE Coef	T	P
A	11.825	5.913	1.088	5.44	0.000
B	4.225	2.112	1.088	1.94	0.070
C	-3.388	-1.694	1.088	-1.56	0.139
D	-15.100	-7.550	1.088	-6.94	0.000
E	3.137	1.569	1.088	1.44	0.169
A*B	-4.212	-2.106	1.088	-1.94	0.071
A*C	2.975	1.488	1.088	1.37	0.190
A*D	16.563	8.281	1.088	7.61	0.000
A*E	-5.900	-2.950	1.088	-2.71	0.015
B*C	-0.850	-0.425	1.088	-0.39	0.701
B*D	-3.313	-1.656	1.088	-1.52	0.147
B*E	-0.300	-0.150	1.088	-0.14	0.892
C*D	1.675	0.837	1.088	0.77	0.453
C*E	-0.138	-0.069	1.088	-0.06	0.950
D*E	1.025	0.512	1.088	0.47	0.644

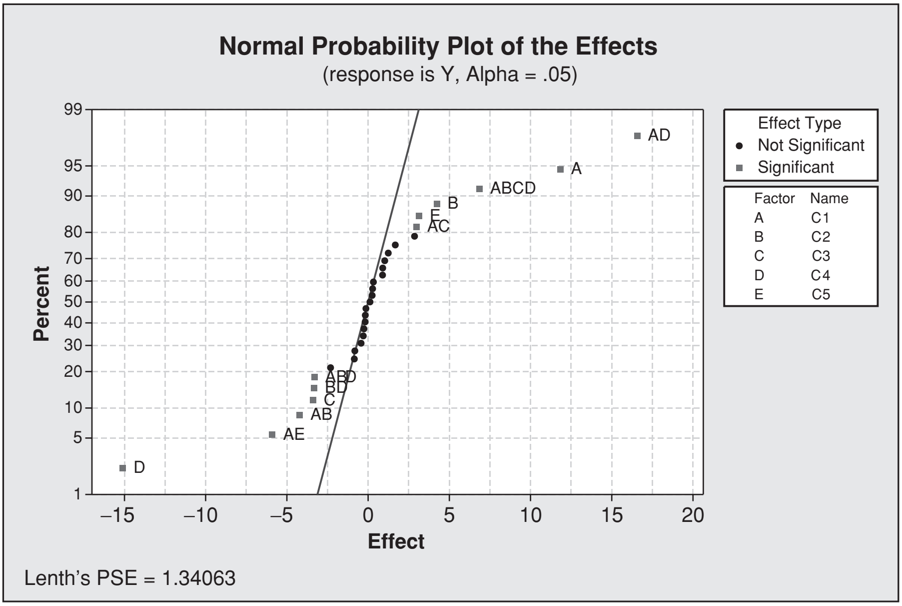

Figure 9.1 "Improper" normal probability plot of effect estimates.

We see that the A, D, and AD effects appear to be real, with the B and AB interactions perhaps deserving some attention. If we construct a normal probability plot of the effect estimates, as shown in Figure 9.1, we obtain a completely different message.

Here we see that 12 of the 31 effects are declared significant using Lenth's procedure with $\alpha = .05$, including a four-factor interaction. The significance of the latter, in particular, should tell us that something is wrong. The problem is that when the data are analyzed improperly in this manner, there is a mixing of the whole-plot error term and the subplot error term. Of course with an unreplicated design, there is a conceptual mixing since the two error terms cannot be computed. Since the subplot factor is estimated with greater precision than the whole-plot factors, too many whole-plot effects may be identified as real when any pseudo-error term is computed in conjunction with the probability plot.

Since we have an unreplicated design, the simplest and best approach would be to assume that any interaction effects could be real and construct the two appropriate normal probability plots, which would be a plot of the whole-plot effects and a plot of the subplot main effect and interactions between the subplot factor and the whole-plot factors. Bisgaard et al. (1996) did just that and obtained rough estimates of the standard errors of a whole-plot effect estimate and a subplot effect estimate from the slopes of the lines fit through the fitted points. The whole-plot standard error was estimated at 7 and the subplot error at 1.

Notice that all but two of the 12 effects labeled significant in Figure 9.1 are whole-plot effects, and also notice that the Lenth estimate of σ is 1.34, which leads to an effect standard error that is far too small for the whole-plot effects. With this in mind and remembering the standard error estimate for subplot effects, we might guess from Figure 9.1 that the real effects are A, D, AD, E, and AE, and in fact this is what Bisgaard et al. (1996) obtained from their two separate plots. (It is worth noting that there must be more than one or two whole-plot factors in order to construct a normal probability plot of the whole-plot effects, as there must be enough plotted points to have a high probability of distinguishing real effects from the effects that are not real. Other methods have been proposed for small designs, but as demonstrated by Goos, Langhans, and Vandebroek (2006), the methods don't work very well.)

There is a large number of published industrial applications of split-plot designs that have been performed; other such examples include Bjerke, Aastveit, Stroup, Kirkhus, and Naes (2004), and Gregory and Taam (1996), with alternative methods of analysis presented for each example.

The Bjerke et al. (2004) article is a case study but there were certain complexities and complications that rendered it somewhat unsuitable as a case study here. It will be discussed briefly, however, and the reader is urged to study the article as an example of how experiments and the subsequent data analysis are not always straightforward. A split-plot structure resulted, according to the authors, because the production samples for a storing experiment were divided into subunits and there were two subplot storing factors: time (7, 14, and 38 weeks) and temperature (4 and 21°C).

This requires some thought, however, because when measurements are made over time, we would generally not regard time as a factor. (Obviously, it is not a factor whose "levels" can be varied at random.) Instead, it would be more appropriate to view the time factor as producing repeated measurements, for which the observations, for a fixed combination of the whole-plot factors, should be autocorrelated. This violates the error structure for a split-plot design, for which the whole-plot errors and split-plot errors are assumed to be independent.

The manufacturer decided to use only the measurements taken at the colder of the two temperatures for analysis because that temperature most closely resembled the real-life storage of mayonnaise.

Therefore, the temperature subplot factor was essentially removed from the analysis and the time subplot factor seems to be actually a repeated measurement. This would mean that there really isn't a split-plot structure. The authors seemed to be aware of this, although they did not state it strongly. They did, however, describe two other analysis approaches: a mixed model approach and a robustness approach. Again, the reader is urged to study the article, which serves as a good example of data analysis complexities at the model-determination stage. Although it is often said that data from a well-designed experiment will practically analyze itself (i.e., be very easy to analyze), that won't always be true.

Gregory and Taam (1996) discussed the use and interpretation of a split-plot experiment to determine significant factors in the development of fracture-resistant automobile windshields. The objective of the experiment was to determine factors that would make the windshields more resistant to fracture caused by the impact of small rocks or pebbles. Three product factors were investigated: glass thickness (two levels), paint type used at the edges of the inner surface (two levels), and windshield coating (three types). All combinations of these factors were tested at eight locations of each windshield produced. Because of manufacturing restrictions, it was necessary to perform certain steps of the process for entire sets of the possible combinations of factors (e.g., all windshields made from the thicker glass were cut at one time). The authors appropriately referred to this as restricted randomization since not all steps in the process were performed in a completely randomized order. Each windshield manufactured was the whole-plot unit and each of the locations tested on each windshield was the subplot unit. Because of the restricted randomization, the authors considered thickness to be a blocking variable, and paint and coating as crossed factors randomized within each glass thickness.

In addition to the analysis of the data as a split-plot experiment, two alternative methods of analysis were suggested. The first option was to use all eight locations as repeated observations for each of the windshields assuming that there is a restriction on randomization of the testing sequence. The second approach was to treat each of the eight observations associated with test sites as a single multivariate response. Results obtained by the first alternative were similar to those obtained by the original interpretation as a split-plot experiment. Those obtained by the multivariate approach were not similar.

These two discussions illustrate that the analysis of data from a split-plot experiment is not necessarily clear-cut.

9.1.3 Split-Unit Designs with Fractional Factorials

The split-unit design concept can be applied when there are more than two factors. Bingham, Schoen, and Sitter (2004), Bingham and Sitter (1999b, 2001), and Huang, Chen, and Voelkel (1998) discussed and illustrated the application of the split-unit design concept to fractional factorials, with Bingham and Sitter (2001) providing an actual industrial example and certain theoretical results (Bingham and Sitter, 1999a) regarding the impact of randomization restrictions on the choice of fractional factorial split-unit designs. Loeppky and Sitter (2002) discussed methods of analyzing data from these designs. Kulahci, Ramírez, and Tobias (2006) question the use of the minimum aberration criterion for these designs and instead suggest the use of the maximum number of clear two-factor interactions criterion recommended for fractional factorials in general by Wu and Hamada (2000).

The notation used in some of the literature for fractional factorial split-unit designs is $2^{(n_1+n_2)-(k_1+k_2)}$, with n_1 and n_2 denoting the number of whole-plot and subplot factors, respectively, and k_1 and k_2 denoting the number of whole-plot and subplot fractional generators, respectively (Bingham and Sitter, 1999a, b; Huang et al., 1998). The choice of notation is somewhat unfortunate because (1) here $n_1 + n_2 = n$ denotes the total number of factors, whereas it generally denotes the number of experimental runs, and (2) the standard notation for a two-level fractional factorial design is 2^{k-p}, so if something is to be "split," it would have been better to split k and p than to split n and k. Bisgaard (2000) used different notation as the designs were represented by $2^{(k-p)-(q-r)}$, and gave many examples of such designs. This is better notation than others have used.

Fractional factorial split-plot designs, designated as FFSP designs in the literature, are constructed similar to the way that fractional factorial designs are constructed, as generators are chosen for the whole-plot factors and for the subplot factors. For example, Bingham and Sitter (1999a) illustrated one way to construct a $2^{(3+3)-(1+1)}$ design by letting $C = AB$ be the generator for the third WP factor and $F = ABDE$ be the generator for the third SP factor, with A, B, and C representing the WP factors and D, E, and F denoting the SP factors. (A generator for a WP factor may contain only WP factors, whereas a SP generator may contain both SP and WP factors.) Whole plots are fixed, so using SP factors as WP generators would destroy the split-plot nature of the design.

When we put these together, we have the defining relation given by $I = ABC = ABDEF = CDEF$, with $CDEF$ being the generalized interaction of the first two components of the defining relation, as in a fractional factorial design. This is analogous to how a 2^{6-2} design is constructed as one starts with a 2^4 design and then selects generators for factors E and F. The difference here is that there is a category for WP factors and a category for SP factors and that distinction must be used in constructing the design. For example, we could not use two SP factors as generators. The reader is asked to explain why this is true in Exercise 9.4.

Since factorial effects are estimated with differing precision when a split-unit arrangement is employed, decisions must be made as to which effects are going to be estimated with the higher precision.

Example 9.3

This was illustrated in the example given by Bingham and Sitter (2001). In that example, a wood products company was interested in investigating the factors that affect the swelling properties of a wood product. The goal of the experiment was to determine optimum process settings that would minimize the swelling of the product after it had been saturated with water and allowed to dry. The fabrication process consisted of two stages, with each batch of wood and additives subdivided to form several sub-batches.

There were five factors that were believed to affect the response variable at the first stage, the mixing stage, and three factors that were thought to be influential in the processing stage. Thus, there were eight factors to be examined; a 2^{8-p} design could not be used because the levels of the five factors in the first stage were set when the second stage commenced because of operational restrictions.

If a $2^{(5+3)-(1+1)}$ FFSP design is to be used, it is necessary to choose one WP generator and one SP generator. Letting A–E denote the WP factors and F–H denote the SP factors, an obvious choice for the generator of E is $E = ABCD$. *If* the generator for H consisted only of SP factors, the generator would be $H = FG$, which, as pointed out by Bingham and Sitter (2001), would be equivalent to constructing a 2^{5-1} design in the WP factors and a 2^{3-1} design in the SP factors. This would confound main effects with interactions among the SP factors, however, so the design would be only resolution III for that group of factors.

We can obviously do much better than this if we let the generator of H be a function of both SP and WP factors. The design recommended by Bingham and Sitter (2001) had $H = ABFG$ so that the defining relation was $I = ABCDE = ABFGH = CDEFGH$, which is obviously resolution V because the shortest word length is 5.

Operationally, the design would be run the same way as any split-plot design, with the runs involving SP factors randomized.

A problem with various methods that had been given for constructing FFSP designs is that they can sometimes result in too many subplots per whole plot or too few whole plots, as noted by Bingham et al. (2004). They gave a method for FFSP designs that is a solution to this problem.

9.1.4 Blocking Split-Plot Designs

Just as it may not be possible to make all the runs in a factorial design or a fractional factorial design under the same set of conditions, it may not be possible to do so with a split-unit design either. McLeod and Brewster (2004) considered the blocking of FFSP designs, with their work motivated by a practical problem.

Specifically, a company was experiencing problems with one of its chrome-plating processes, as excessive pitting and cracking, poor adhesion, and uneven deposition and adhesion across a chrome-plated, complex-shaped part were observed when the part was being plated. A screening experiment was subsequently planned to identify key factors affecting the quality of the process. Six factors were identified, three of which were hard to vary. The latter became the split-plot factors when a split-plot design was used. (There were actually two additional sub-subplot factors that were used in the experiment, but the authors elected not to discuss them so as to focus on the blocking aspect.) Two levels were to be used for each factor and there were multiple responses, including pits and cracks.

The experiment involved a rectifier and different levels of the SP factors could be used with each rectifier. There was only one tank (bath) available for the experiment, and although the tank had four rectifiers, two of them had to be used for another experiment. Consequently, only two parts could be plated each day, suggesting the use of a 2^{3-2} design for the SP factors. Instead, the decision was made to change one of the SP factors to a WP factor so as to simplify the design problem. Of course this meant that one of the hard-to-change factors would not be treated specially. The experiment was to run for 16 days in the authors' modified example (the actual experiment ran for 20 days), which were regarded as four 4-day weeks and it was desirable to block the experiment by week. With four whole-plot factors there was thus a 2^4 design that could be run in four blocks at the whole-plot level, and this is what was done. The design that was used was thus a $2^{(4+2)-(0+1)}$ FFSP design run in four blocks.

As McLeod and Brewster (2004) pointed out, however, blocking a FFSP is not simple because there is fractionation at both the whole-plot and subplot levels, although that was not the case with this design since the whole-plot design was not fractionated. Let A, B, and C designate the original WP factors and let P denote the SP factor that was converted to a WP factor, leaving q and r as the two *SP* factors. The generator $r = ABCPq$ was used to generate the second SP factor. (Recall from Chapter 5 that in two-level fractional factorial designs, factors are created from columns of the base design, which is a full factorial. Here the base design at the subplot level is a 2^1 design.) The blocking generators were, in the authors' notation, $\beta_1 = ABC$ and $\beta_2 = ABP$, which of course also confounds the product, CP.

After the experiment was run, the authors questioned whether or not the design was optimal. They presented three distinct methods of constructing the blocks and provided a lengthy discussion. Their conclusion was that the design used was not at all bad, but they provided what they considered to be a better approach to constructing the design. See their paper for the details.

9.1.5 Split-Unit Plackett–Burman Designs

Split-plot Plackett–Burman designs have been discussed and illustrated to a moderate extent in the literature, including Kulahci and Bisgaard (2005) and Tyssedal and Kulahci (2005). Of course such designs are alternatives to split-plot fractional factorial designs in the same way that Plackett–Burman designs are alternatives to fractional factorial designs.

9.1.6 Examples of Split-Plot Designs for Hard-to-Change Factors

In this section we examine some actual split-plot experiments with hard-to-change factors, with the split-plot arrangement necessitated by the existence of a hard-to-change factor. This is now a common use of split-plot designs and outside of agricultural use is certainly the motivation for the use of these designs.

Example 9.4

An application of this type of design is described at the Stat-Ease Web site: www.statease.com/pubs/morton.pdf. Morton Powder Coatings of Reading, PA, turned to experimental design after some of its customers began detecting defects in one of the company's leading coating products. Consequently, a team of researchers was formed to address the problem. The researchers suspected that the problem was in the formulation of the powder. They had the use of the customer line for only two shifts, and thus needed to collect data on as many different combinations of factor levels as possible within the allotted time. Five factors were selected for an experiment; three of these were components of the powder: level of a catalyst, amount of a certain proprietary additive, and the coarseness of the powder. The other two factors were line speed and oven temperature, which could be changed. The three factors relating to the powder coating formula could not be easily changed, however.

The researchers used a split-plot design, with the hard-to-change factors comprising the whole plots and the easy-to-change factors comprising the subplots. The way that the experiment would be conducted would be to randomly select a level for each of the whole-plot factors and then randomly vary the levels of the subplot factors.

Notice the relationship between this prescription for carrying out the experiment and the way an experiment is performed when a randomized block design is used. With the latter, the levels of a factor are randomly used within each block. With a split-plot design, we might more or less view each set of levels of the whole-plot factors as constituting a "block," with the sub-plot factors randomly varied within each "block." Then the levels of the whole-plot treatments are changed and the process is repeated. (Of course a split-plot design may be blocked, literally, as was discussed in Section 9.1.4.)

Example 9.5

Another experiment that used a split-plot design because there was at least one hard-to-change factor was the experiment described in Kowalski, Landman, and Simpson (2003). The experimental unit was a NASCAR Winston Cup race car. These cars are of course quite expensive, highly valued, and owned by different racing teams, so experimentation with multiple cars would not be feasible.

The overall objective, of course, is to make a car go as fast as possible. In recent years, many race car teams have constructed the layout for experiments, using an ad hoc approach that does not always produce useful results, despite a considerable investment of time and money.

Four factors were used in the experiment and there were four response variables that each measured aerodynamic efficiency. The four factors were: front car height, rear car height, yaw angle (the angle that the car centerline makes with the air stream), and coverage of the radiator grille (tape). The first two factors were hard to change, thus precluding the use of a factorial design in the four factors. Instead, a 2^2 design was constructed for the hard-to-change factors, with the four points replicated and a centerpoint added, for a total of nine points. Another 2^2 design with a centerpoint was constructed and used in conjunction with each of these 9 points, producing a total of 45 points. (Notice that the construction of the design relates to the idea of an assembly experiment, which was discussed in Section 5.15, and it also relates to a product array, which was discussed in Section 8.3.)

In general, factors should be reset before each experimental run, even those factors that are reset to the same level that was used in the previous run (see the discussion of this in Section 4.20), which means that the front car height and rear car height would have to be set 45 times if 45 design points were used. It would take about 35 minutes for the resets for each design point, however, so just the resets would take over 26 hours. Instead, the entire experiment took only 9 hours to perform because the car heights were reset just nine times.

In split-plot terminology, the two car heights are the WP factors and the other two factors are the SP factors.

We won't pursue this further since the data were not given, but we note that the interaction effect of the whole-plot factors was greater than the main effect of the front height factor, with both effects being significant. Thus, there is a problem with the prediction equation that was stated "can be used to predict the front coefficient of lift for various settings of front height, rear height, tape, and yaw." Specifically, the coefficient for front height, which is 0.0059, is not a particularly good representation of the effect of front height at each of the two levels. The coefficient of rear height is also shaky for a similar reason.

Unfortunately, as stated by Reece (2003, p. 327), very few software packages can generate a split-unit (split-plot) design directly.

9.1.7 Split-Split-Plot Designs

A split-plot design could be converted into a split-split-plot design by adding an additional split to accommodate a third factor and such a design could be split further, if desired. Gumpertz and Brownie (1993) are discussed in the literature as having given an example of a split-split-plot design with time as the third factor because repeated measurements were used, but such measurements will be correlated whereas these designs are assumed to have independent errors at each level of splitting. Thus, as in Example 9.2, it might seem inappropriate to say that a split-split-plot design is used when time is claimed to be the third factor with repeated measurements made over time. Milliken (2004) does discuss and illustrate such a design, however, although the repeated measurements aspect does complicate the analysis. The starting point is to initially analyze the data *as if* the split-plot assumptions hold, with those assumptions being that all the effect estimators are identically and independently normally distributed. A covariance structure for the effect associated with the repeated measurements must then be selected, and Milliken (2004) and Milliken and Johnson (2001) suggest starting with a small set of candidate covariance structures and selecting the simplest one.

For an unreplicated design, three normal probability plots would have to be constructed, analogous to the two normal probability plots that are needed for the split-plot design. Therefore, the same consideration must be made as with a split-plot design; namely, is there a sufficient number of effects to permit the construction of each plot? When replications are used, consideration must be given to whether there is a sufficient number of degrees of freedom for each of the error terms to provide sufficient power for each test. Of course, this same consideration must be made for a split-plot design.

9.2 SPLIT-LOT DESIGN

Another type of "split" design is a split-lot design, which is used in manufacturing processes when a product is formed in two or more process stages. These designs were invented by Mee and Bates (1998); a more recent source is Butler (2004).

A unique feature of these designs is that each factor is used in one and only one processing stage, with multiple factors used at each stage and the design at each stage having a split-plot structure. For example, as illustrated by Butler (2004), a fractional factorial design such as a 2^{9-3} design might be used with three of the nine factors used at each of three stages. A somewhat extreme example relative to the number of stages would be nine stages. Mee and Bates (1998) considered a 2^{9-3} split-lot design with one factor used at each of the nine stages. This is not just a matter of splitting the design, however, as a split-plot structure is used *at each stage*, as stated previously. Of course this raises the question of what the design should be called.

Taguchi (1987) has termed the design a multiway split-unit design, which in some ways is a better term as the "split" designs discussed in this chapter are for two groups, such as whole plots and subplots, whereas the type of design discussed in this section can have far more "groups."

9.2.1 Strip-Plot Design

A strip-plot design is one which is applied to a multistage process. It is equivalent to a split-lot design when there are two processing stages, so it is a special case of a split-lot design. Miller (1997) proposed a method of constructing strip-plot designs for fractional factorials and mixed factorials, and gave as an illustrative example a laundry experiment that consisted of washing in the first stage and drying in the second stage. As with a split-plot design, however, we can question whether or not the term "plot" should be used, since technically there is no "plot" that is being "stripped." There is a splitting of sorts that occurs, however, as what would be a single error term if a fractional factorial design were used in a single stage becomes "split" into two or three components.

As Giesbrecht and Gumpertz (2004, p. 176) point out, these designs are also referred to as *strip-block* or *split-block* designs, either of which may be a better term since blocking is involved. A SAS macro for analyzing data from a strip-plot experiment is given at http://64.233.161.104/search?q=cache:-PuFuIroQAoJ:home.nc.rr.com/schabenb/Strip-Plot.html+strip-plot++design&hl=en.

Example 9.6

Miller (1997) states that the primary motivation for using a strip-plot design is that more treatment combinations can be investigated for the same amount of experimental resources and uses a (somewhat disguised) actual laundry experiment for illustration. That experiment was performed to investigate wrinkling of clothes that are washed and dried, and to obviously determine the best way to perform the two-stage operation so as to minimize wrinkling while performing a small experiment in terms of resources expended. Six factors that represented washing conditions and four factors that represented drying conditions were to be used in the experiment, with each factor at two levels.

If this were a single-stage experiment, one obvious possibility would be a 2^{10-p} design for a suitably chosen value of p. It was decided, however, to use two 4×4 strip plots, each arranged as a 4×4 Latin square with columns representing dryers and the rows representing washers. This might at first seem to be a peculiar design layout because the number of washing conditions is not the same as the number of drying conditions. One thing is apparent: There will be 32 treatment combinations used, which is far less than $2^{10} = 1024$, so some fractionization must occur.

The experiment was performed by using a 2^{6-3} design for the washers and a 2^{4-1} design used for the dryers, with the eight treatment combinations for each split up into two blocks of four. When these are crossed in the 4×4 Latin square layout, there are 16 treatment combinations in each block, for a total of 32 treatment combinations.

An obvious question to ask at this point is "Since 32 runs are to be made, why not just use a 2^{10-5} design?" This would be fine if 32 identical washers and 32 identical dryers were available, or if only a single washer and a single dryer were available and the experiment could be conducted over a long period of time. The point is that the use of such a design would ignore the resources that are available—the four washers and the four dryers.

Since a fractional factorial was used for the washers and for the dryers, the fractions should be constructed in an optimal manner. That is, using the notation of Miller (1997) and letting lowercase letters denote the dryer factors and capital letters denote the washer factors, the defining relation for the fraction for the dryers should be $I = abcd$, with various equally good choices available for the defining relation for the washer fraction.

Milliken, Shi, Mendicino, and Vasudev (1998) described the application of a strip-plot design to a two-step process and explained that a mixed model is an appropriate analysis for such a design. They provided SAS code for performing the analysis. That article serves as somewhat of a tutorial for the construction of both split-plot designs and strip-plot designs.

Milliken (2004) also discussed and illustrated a strip-plot design, with PROC Mixed code for SAS used for the analysis since the model contained both fixed effects and random effects. An example of a strip-plot design, including a diagram that shows the layout of the design, is also given in Section 5.5.5 of the *e-Handbook of Statistical Methods* (Croarkin and Tobias, 2002).

Unfortunately, as stated by Reece (2003, p. 327), very few software packages can generate strip-plot designs.

9.2.1.1 Applications of Strip-Block (Strip-Plot) Designs

Although a "strip-block" design is the same as a strip-plot design, as stated in the preceding section, in this section we describe applications of the design when it has been described in the literature as a strip-block design.

These designs have been used in agricultural applications since the late 1930s, but there have been relatively few industrial applications. As with the other types of designs discussed in this chapter, the factors to be studied are divided into two (or more) groups so as to enable the experiment to be conducted more efficiently than would be the case if a factorial design without grouping had been used.

An example given by Vivacqua and de Pinho (2004) will illustrate the basic idea. (The same example is given by Vivacqua, Bisgaard, and Steudel (2002), which is available without restrictions at http://qsr.section.informs.org/download/paper1_Carla.pdf.) A battery company, Rayovac, invited a team of consultants to work on a problem with one of its products that was costing the company over \$154,000 in losses per year. The first step taken in the project was to construct a detailed flowchart of the specific process involved and of the company in general.

Six factors were selected for use in the experiment; four were associated with the assembly process and two were thought to possibly have an effect on the open-circuit voltage. The ultimate objective was to determine the levels of the process variables that would result in the production of high-quality battery cells. A 2^6 design was out of the question because each curing cycle required 5 days, so 64 runs would require $64 \times 5 = 320$ days. Instead, a strip-block design was used. There were 2000 batteries used for each treatment combination and the experimental design was a 2^4 design in the assembly factors and a 2^2 design in the curing factors. The experiment was performed by first randomizing the 16 runs and the assignment of the 16 treatment combinations to 16 sets of 2000 batteries. Then each lot was split into four sublots and each of these was assigned to one of the four curing conditions. All 16 subplots assigned to the same curing condition were then processed simultaneously.

Vivacqua and de Pinho (2004) make the important point that a normal probability plot of all the effects in a strip-block design cannot be constructed because there are different error terms. (Of course the variance of the plotted effect estimates must be the same when a normal probability plot is constructed, as has been emphasized previously.) Instead, a separate plot would have to be constructed for all effects that have the same error term. Of course this also applies to any design for which there are multiple error terms. In their example there were three separate error terms, which were for the assembly effects, the curing effects, and the interactions between the assembly factors and the curing factors. This would necessitate the use of three distinct normal probability plots, provided that there were enough effect estimates in each of the three groups to justify the construction of each plot. That is obviously not the case in this application, however, because there are only three curing effects: the two curing factors and the interaction between them. Consequently, Vivacqua and de Pinho (2004) used normal probability plots only for the assembly effects and the interaction effects.

One of the results of the experiment was that the defective rate was reduced by 80 percent (of what the old defective rate was) to a defective rate of approximately 1 percent.

Vivacqua et al. (2002) additionally considered a modification of this design scenario, assuming that only eight sublots could be accommodated in the storage room rather than 16. This would necessitate the use of a 2^{4-1} design for the assembly factors. (The use of fractional factorials in strip-block designs was discussed by Miller, 1997.) This design would be of resolution IV because the 2^{4-1} design is resolution IV. Vivacqua et al. (2002) discussed a way of improving the resolution of this design by using a *post-fractionated design* approach. If the interaction of the four assembly factors ($ABCD$) is aliased with the interaction of the curing factors (ED), a resolution VI design results since the defining relation is $I = ABCDEF$. Vivacqua et al. (2002) termed this a "postfractionated design of order 1," since it is a $1/2$ fraction of the original design, with again eight sublots used for each of the four curing conditions.

The authors also illustrated the construction of a post-fractionated design of order 2, which would have to be used if the storage room could accommodate only four lots. That results in a design of resolution IV. (The authors presented tables of these various designs and that would be helpful here, but the article is available on the Internet with no restrictions, so readers are referred to those tables of the design layouts.)

The authors also presented a general framework for the construction of post-fractionated strip-block designs and readers are referred to their article for details. They also presented the analysis framework for the analysis of data from such designs, pointing out, in particular, that there are four error strata rather than three with a strip-block design with post-fractionazation, with the fourth stratum being the post-fraction stratum. They provide a rule that allows the effects to be placed in the appropriate categories without the need to compute variances for all of the effect estimates to determine the category into which each effect should be placed, as has been done in other papers. See their paper for details.

9.3 COMMONALITIES AND DIFFERENCES BETWEEN THESE DESIGNS

Until the last 10 years or so, the only "split" design discussed in the literature was a split-plot design. Now, with the various related designs to select from, experimenters need a good understanding of these designs and their relationships and differences in order to select an appropriate design for a given situation. Therefore, in this section we'll examine these relationships.

With what has commonly been referred to as a split-plot design, a whole plot is divided into at least two subplots and subplot treatments are assigned randomly to each subplot for each whole plot, with the whole-plot treatment combinations assigned to the whole plots.

We might think of a strip-block design as somewhat the opposite or reverse of a split-plot design, as with the former the complete set of treatment combinations of the "whole-plot factors" would all be run for each treatment combination of the subplot factors, then all the treatment combinations would be run for the next treatment combination of the subplot factors, and so on until all the runs have been made for all treatment combinations of the whole-plot factors for the last treatment combination of the subplot factors.

A split-lot design is specifically for multiple process stages, with a split-plot design used in each stage. When there are two stages, the design is equivalent to a strip-plot design, which has also been termed a split-block design.

Better understanding as well as more efficient use of these designs might be facilitated by the adoption of names for the designs that allow them to be better distinguished and by using only one name for each design. For example, a split-plot design might be called simply a split-unit design and a split-lot design might be more appropriately termed a multiway split-unit design. A strip-plot design could be called a strip-block design, which of course is one of the other names for the design anyway. See also Guseo (2000) for a review and comparison of these designs.

9.4 SOFTWARE

As one might suspect, the construction and analysis of some of the designs given in this chapter can be somewhat cumbersome and most software do not have this capability. The SAS 9.1 ADX Interface for Design of Experiments *can* be used to construct full factorial and FFSP designs, however, with the analysis of data from such designs performed using a mixed model approach (see Chapter 11 of http://support.sas.com/documentation/onlinedoc/91pdf/sasdoc_91/qc_gs_7304.pdf.)

The list of software that cannot handle split-plot designs is long and includes the following. A split-unit design cannot be constructed and analyzed directly in MINITAB, as discussed in http://www.scimag.com/ShowPR.aspx?PUBCODE=030&ACCT=3000039460&ISSUE=0306&RELTYPE=PR&ORIGRELTYPE=FE&PRODCODE=00000000&PRODLETT=A. That article does discuss in a general way how MINITAB can be used in a somewhat manual manner to overcome the problem, however, and is an interesting application of a split-plot design.

Similarly, data from a split-plot design cannot be analyzed directly with Design-Expert. The following excerpts from the User's Guide for Design-Expert 6 (the current version is 7) are relevant:

> The analysis of a split-plot design is tricky, even for statisticians. It can be done in Design-Expert by properly designating effects in specific ways for subsequent analysis of variance. Proceed if you dare! (pp. 4–11)
>
> The analysis of a split-plot design must be done somewhat manually. (pp. 4–13)

Because of the latter, they state that two separate ANOVAs must be run: one for the whole-plot treatment(s) and interactions of those factors, and the other one for the subplot treatment(s) and the interaction(s) between the whole-plot treatment(s) and the subplot treatment(s), as well as interactions between the subplot factors. This is illustrated for Version 6 of their software at http://www.statease.com/x6ug/DX04-Factorial-General.pdf and is also illustrated by Whitcomb and Kraber (manuscript) in their company case study at www.statease.com/pubs/pcr_via_split-plot.pdf.

D. o. E. Fusion Pro provides the capability to design split-plot experiments and analyze the resultant data, and in fact the user is given the option of selecting a split-plot structure on the first screen display when specifying a design. Reece (2003, p. 335) noted in reference to the software's split-plot design capability, "While this investigation didn't formally investigate this function, this ability is unusual among packages such as this." RS/Discover has similar capability as Reece (2003, p. 364) states ". . . it can design split-plot and strip plot designs for the knowledgeable user."

Because split-plot designs and their variations are becoming more popular, motivated to a considerable extent by the work of Box and Jones (1992), it would be helpful if more software companies would give some attention to incorporating split-plot designs into their software, and perhaps include some or all of the other types of designs discussed in this chapter.

9.5 SUMMARY

Basic, introductory information on split-plot designs is given in Box (1995–1996), Box and Jones (2000), Kowalski (2002), Kowalski and Potcner (2003), and Potcner and Kowalski (2004), and the reader is referred to these articles for additional sources that are easily readable. Other introductory-level sources of information on split-plot designs include Gardner and Cawse (2003), with information on the variants presented of this basic design described in the references that have been cited, as well as Vivacqua and Bisgaard (2004). It is best that split-plot designs have more than one or two whole plot factors since the identification of real whole plot effects is difficult when there are few whole plots.

REFERENCES

Bingham, D. R. and R. R. Sitter (1999a). Some theoretical results for fractional factorial split-plot designs. *Annals of Statistics*, **27**, 1240–1255.

Bingham, D. R. and R. R. Sitter (1999b). Minimum aberration fractional factorial split-plot designs. *Technometrics*, **41**, 62–70.

Bingham, D. R. and R. R. Sitter (2001). Design issues in fractional factorial split-plot experiments. *Journal of Quality Technology*, **33**(1), 2–15.

Bingham, D. R. and R. R. Sitter (2003). Fractional factorial split-plot designs for robust parameter experiments. *Technometrics*, **45**, 80–89.

Bingham, D. R., E. D. Schoen, and R. R. Sitter (2004). Designing fractional factorial split-plot experiments with few whole-plot factors. *Applied Statistics*, **53**(2), 325–339.

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

Bisgaard, S. and A. L. S. de Pinho (2004). The error structure of split-plot designs. *Quality Engineering*, **16**(4), 671–675.

Bisgaard, S. and M. Sutherland (2003-4). Split plot experiments: Taguchi's Ina tile experiment reanalyzed. *Quality Engineering*, **16**(1), 157–164.

Bisgaard, S., H. T. Fuller, and E. Barrios (1996). Two level factorials run as split plot experiments. *Quality Engineering*, **8**(4), 705–708.

Bjerke, F., A. H. Aastveit, W. W. Stroup, B. Kirkhus, and T. Naes (2004). Design and analysis of storing experiments: A case study. *Quality Engineering*, **16**(4), 591–611.

Box, G. E. P. (1995–1996). Split-plot experiments. *Quality Engineering*, **8**(3), 515–520.

Box, G. E. P. and S. Jones (2000). Split-plots for robust product and process experimentation. *Quality Engineering*, **13**(1), 127–134.

Box, G. E. P. and S. Jones (1992). Split-plot designs for robust product experimentation. *Journal of Applied Statistics*, **19**, 3–26. (This is available as Report No. 61, Center for Quality and Productivity Improvement, University of Wisconsin-Madison (see http://www.engr.wisc.edu/centers/cqpi).

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

Butler, N. A. (2004). Construction of two-level split-lot fractional factorial designs for multi-stage processes. *Technometrics*, **46**, 445–451.

Croarkin, C. and P. Tobias, eds. (2002). *NIST/SEMATECH e-Handbook of Statistical Methods* (www.itl.nist.gov/div898/handbook), a joint effort of the National Institute of Standards and Technology and International SEMATECH.

Czitrom, V. (1997). Introduction to design of experiments. In *Statistical Case Studies for Industrial Process Improvement*, Chap. 14 (V. Czitrom and P. D. Spagon, eds.). Philadelphia: Society for Industrial and Applied Mathematics; Alexandria, VA: American Statistical Association.

Daniel, C. (1976). *Applications of Statistics to Industrial Experimentation*. New York: Wiley.

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

Gardner, M. M. and J. N. Cawse (2003). Split-plot designs. In *Experimental Design for Combinatorial and High Throughput Materials Development*, Chap. 8. Hoboken, NJ: Wiley.

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

Goos, P. (2002). *The Optimal Design of Blocked and Split-Plot Experiments*. New York: Springer.

Goos, P. and M. Vandebroek (2001). Optimal split-plot designs. *Journal of Quality Technology*, **33**, 436–450.

Goos, P. and M. Vandebroek (2003). D-optimal split-plot designs with given numbers of whole plots. *Technometrics*, **45**, 235–245.

Goos, P. and M. Vandebroek (2004). Outperforming completely randomized designs. *Journal of Quality Technology*, **36**(1), 12–26.

Goos, P., I. Langhans, and M. Vandebroek (2006). Practical inference from industrial split-plot designs. *Journal of Quality Technology*, **38**(2), 162–179.

Gregory, W. L. and W. Taam (1996). A split-plot experiment in an automobile windshield fracture resistance test. *Quality and Reliability Engineering International*, **12**, 79–87.

Gumpertz, M. L. and C. Brownie (1993). Repeated measures in randomized block and split-plot experiments. *Canadian Journal of Forest Research*, **23**, 625–639.

Guseo, R. (2000). Split and strip-plot configurations of two-level fractional factorials: A review. *Journal of the Italian Statistical Society*, **9**, 85–86.

Huang, P., D. Chen, and J. Voelkel (1998). Minimum aberration two-level split-plot designs. *Technometrics*, **40**, 314–326.

Kahn, W. and C. Baczkowski (1997). Factors which affect the number of aerosol particles released by clean room operators. In *Statistical Case Studies for Industrial Process Improvement*, Chap. 10 (V. Czitrom and P. D. Spagon, eds.). Philadelphia: Society for Industrial and Applied Mathematics; Alexandria, VA: American Statistical Association.

Kowalski, S. M. (2002). 24 run split-plot experiments for robust parameter design. *Journal of Quality Technology*, **34**(4), 399–410.

Kowalski, S. M. and K. J. Potcner (2003). How to recognize a split-plot experiment. *Quality Progress*, November, 60–66. (This paper can be read at http://www.minitab.com/resources/Articles/RecognizeSplitPlot.pdf.)

Kowalski, S., D. Landman, and J. Simpson (2003). Design of experiments enhances race car performance. *Scientific Computing and Instrumentation*, June (available at http://www.scamag.com/ShowPR.aspx?PUBCODE=030&ACCT=3000039460&ISSUE=0306&RELTYPE=PR&ORIGRELTYPE=FE&PRODCODE=00000000&PRODLETT=A)

Kulahci, M. and S. Bisgaard (2005). The use of Plackett–Burman designs to construct split-plot designs. *Technometrics*, **47**(4), 495–501.

Kulahci, M., J. G. Ramírez, and R. Tobias (2006). Split-plot fractional designs: Is minimum aberration enough? *Journal of Quality Technology*, **38**, 56–64.

Langhans, I., P. Goos, and M. Vandebroek (2005). Identifying effects under a split-plot design structure. *Journal of Chemometrics*, **19**, 5–15.

Loeppky, J. L. and R. R. Sitter (2002). Analyzing unreplicated blocked or split-plot fractional factorial designs. *Journal of Quality Technology*, **34**(3), 229–243.

Lucas, J. M. and M. C. Hazel (1997). Running experiments with multiple error terms: How an experiment is run is important. In *ASQ Annual Quality Congress Transactions*, pp. 283–296. Milwaukee, WI: American Society for Quality. (This article is available online to ASQ members at www.asq.org/members/news/aqc/51_1997/10549.pdf.)

Lucas, J. M. (1999). Comparing randomization and a random run order in experimental design. *Chemical and Process Industries Division News*, Summer, **7**, 10–14. (This article is available online at http://www.asq.org/forums/cpi/newsletters/summer_1999.pdf.)

McLeod, R. G. and J. F. Brewster (2004). The design of blocked fractional factorial split-plot experiments. *Technometrics*, **46**, 135–146.

Mead, R. (1988). *The Design of Experiments: Statistical Principles for Practical Application*. Cambridge, UK: Cambridge University Press.

Mee, R. W. and R. L. Bates (1998). Split-lot designs: Experiments for multistage batch processes. *Technometrics*, **40**(2), 127–140.

Miller, A. (1997). Strip-plot configurations of fractional factorials. *Technometrics*, **39**(2), 153–160.

Milliken, G. A. (2004). Mixed models and repeated measures: Some illustrative examples. In *Handbook of Statistics*, Vol. 22, Chap. 5 (R. Khattree and C. R. Rao, eds.). Amsterdam, The Netherlands: Elsevier Science B. V.

Milliken, G. A. and D. E. Johnson (2001). *Analysis of Messy Data, Vol. 3: Covariance Models*. London: CRC Press/Chapman-Hall.

Milliken, G. A., X. Shi, M. Mendicino, and P. K. Vasudev (1998). Strip-plot design for two-step processes. *Quality and Reliability Engineering International*, **14**, 197–210.

Montgomery, D. C. (1996). *Design and Analysis of Experiments*, 4th ed. New York: Wiley.

Potcner, K. J. and S. M. Kowalski (2004). How to analyze a split-plot experiment. *Quality Progress*, December, 67–74. (This article can be read at http://www.minitab.com/resources/Articles/AnalyzeSplitPlot.pdf.)

Ramírez, J. G. (2004). To split or not to split: Do we have a choice? *SPES/Q&P Newsletter*, **11**(2), pp. 10–12.

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

Robinson, T. J., R. H. Myers, and D. C. Montgomery (2004). Analysis considerations in industrial split-plot experiments when the responses are non-normal. *Journal of Quality Technology*, **36**, 180–192.

Taguchi, G. (1987). *System of Experimental Design*, Vols. 1 and 2. White Plains, NY: UNIPUB/Kraus International.

Tyssedal, J., M. Kulahci, and S. Bisgaard (2005). Split-plot designs with mirror image pairs as subplots. (submitted).

Tyssedal, J. and M. Kulahci (2005). Analysis of split-plot designs with mirror image pairs as sub-plots. *Quality and Reliability Engineering International*, **21**, 539–551.

Vivacqua, C. A. and S. Bisgaard (2004). Strip-block experiments for process improvement and robustness. *Quality Engineering*, **16**(3), 495–500.

Vivacqua, C. A. and A. L. S. de Pinho (2004). On the path to Six Sigma through DOE. *Annual Quality Transactions* (available to ASQ members at http://www.asq.org/members/news/aqc/58_2004/20116.pdf)

Vivacqua, C. A., S. Bisgaard, and H. J. Steudel (2002). Using strip-block designs as an alternative to reduce costs of experimentation in robust product design and multi-stage processes. Paper presented at *INFORMS meeting*. Available without restrictions at http://qsr.section.informs.org/download/paper1_Carla.pdf)

Whitcomb, P. and S. Kraber (2005). PCR process optimized via split-plot DOE. Stat-Ease case study. (www.statease.com/pubs/pcr_via_split-plot.pdf.)

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

EXERCISES

9.1 Czitrom (1997) gave an example of an experiment for which a split-plot design was used because there was one hard-to-change factor, which was temperature. The data, from a 2^3 design with a centerpoint, are given below.

Run	Temperature (°C)	Pressure	Argon flow	Tungsten deposition rate	Tungsten non-uniformity (%)	Stress
1	440	0.8	0	265	4.44	10.73
2	440	0.8	300	329	8.37	10.55
3	440	4.0	0	989	4.48	9.71
4	440	4.0	300	1019	7.89	9.77
5	500	0.8	0	612	6.39	8.35
6	500	0.8	300	757	8.92	7.73
7	500	4.0	0	2236	4.44	7.55
8	500	4.0	300	2389	7.83	7.48
9	470	2.4	150	1048	7.53	8.59

Notice that this list of the runs, which is in the same order as in Czitrom (1997), is in Yates order. When the experiment was performed, the actual order of the runs was 4, 2, 1, 3, 9, 7, 5, and 8. Analyze the data, using the second response variable (only), and determine the significant effects, if any. If possible, compare your result with that given by Czitrom (1997).

9.2 Kahn and Baczkowski (1997, references) presented a case study that is worth discussing not only in the context of split-plot designs because that is what was used, but also in regard to the interpretation of data from designed experiments, in general. The title of their article "Factors which affect the number of aerosol particles released by clean room operators" explains the objective of the experimentation that is described in their (detailed) case study. The authors stated that previous experimentation had "not achieved adequately reproducible results" because of the failure to identify certain large variance components. The experiment that was described in the case study had eight factors: garment type, garment size, person, day, time, garment, location, and protocol. Five of the six factors whose levels defined the tests that were performed had two levels and the other factor had three levels. Thus, 96 tests were performed and since they were performed at two workstation levels (near hand level and near foot level) in addition to three (protocol) activities (deep knee bends, marching and slapping, and reaching), a total of 576 particle measurements were taken.

This was not a factorial arrangement because the six measurements that resulted from the six combinations of levels of the last two factors were made on each test; that is, the levels of the first six factors were fixed when the six measurements were obtained (together). Because there was no randomization in obtaining the measurements for the combinations of levels of the last two factors, the first six factors constitute WP factors and the last two factors are the SP factors.

The authors initially collapsed the data over the 96 tests and performed a graphical analysis of the two SP factors, which amounts to analyzing data from a replicated 2×3 design. The authors then proceeded to what they termed the "advanced analysis." They pointed out that there are "several possible ways to analyze a split-plot design," and chose to use two separate ANOVA "runs" (their terminology): one for the whole-plot factors and one for the split-plot factors and split-plot-by-whole-plot interactions. (Note that this is contrary to what was discussed in Section 9.4 regarding Design-Expert.)

In their split-plot analysis, the authors' ANOVA table showed 10 of the 12 effects having a p-value of .0001, including a three-factor interaction. Although they did not provide effect estimates, they did state how such estimates would be computed from the ANOVA results and concluded that most of the effects with p-values of .0001 were not of practical interest, apparently reaching this conclusion because certain effects were "... physically small with respect to the other effects...."

(a) Since we would expect most studied effects, and certainly most interactions, to not be significant (see, e.g., the discussion in Section 4.10), what action would you take if you were presented with such results?

(b) Near the end of their case study, the authors pointed out that, for each of the 96 tests, they did not perform each of the three activity protocols twice, once for the workstation-level measurement and once for the hand-level measurement. Rather, each protocol was performed once and then there were two measurements, one for each level. Does this mean that the design was not a split-plot design and could this possibly explain the fact that almost all of the effects were significant in the split-plot analysis? (You are welcome to read the authors' explanations of these matters.)

9.3 Potcner and Kowalski (2004, see references) used an example to illustrate the analysis of data from an experiment in which a split-plot design was used. (The paper is available at http://www.minitab.com/resources/Articles/AnalyzeSplitPlot.pdf.) Read the article and use appropriate software, such as D. o. E. Fusion Pro to perform your analysis. Does your output agree with the results given by the authors (as it certainly should)? Do you agree with the conclusions drawn by the authors?

9.4 Consider an FFSP design with four WP factors and two SP factors. Let the SP factors be E and F, with A, B, C, and D denoting the WP factors. Show why the design could not be constructed by using the generators $E = ABD$ and $F = BCD$.

9.5 The data from the famous cake mix experiment given by Box and Jones (1992) are given below, using notation somewhat different from that used by Box and Jones, with the response variable being a measure of how good each cake mix tasted.

F	S	E	(1)	a	b	ab
-1	-1	-1	1.1	1.4	1.0	2.9
1	-1	-1	1.8	5.1	2.8	6.1
-1	1	-1	1.7	1.6	1.9	2.1
1	1	-1	3.9	3.7	4.0	4.4
-1	-1	1	1.9	3.8	2.6	4.7
1	-1	1	4.4	6.4	6.2	6.6
-1	1	1	1.6	2.1	2.3	1.9
1	1	1	4.9	5.5	5.2	5.7

The first three factors, Flour (F), Shortening (S), and Egg Powder (E) constitute a 2^3 design in the whole-plot factors; the other two factors (A and B, say) are environmental variables arranged in a subplot whose four treatment combinations are indicated; that is, this is a split-plot design. More specifically, there were eight separate cake mixes and there were four different environmental conditions under which the cakes were made, and these were different baking times and temperatures. Analyze the data and state your conclusions, thinking in part about what one wants to accomplish (as in Chapter 8) regarding robust design and making a product that is robust to the environmental conditions, such as overcooking in this case.

9.6 Box (1995–1996, references) gave the following data from an experiment with a split-plot design, with the data listed as it appears in that publication, which reflects the randomization that was used. (The data and the analysis of it are also in Chapter 9 of Box et al. (2005, references).) The objective of the experiment was to improve the corrisive resistance of steel bars by applying a surface coating and then baking the bars in a furnace for a specified time. Four coatings and three temperatures were used, and each temperature was used twice. The bars with each of the four coatings were randomly positioned in the furnace for each temperature level; the way that the temperatures were run will be discussed later. Here the four coatings are denoted by 1, 2, 3, and 4. The response variable is corrosive resistance and the response values are given in parentheses below.

Temperature	Coatings			
360°	2	3	1	4
	(73)	(83)	(67)	(89)
370°	1	3	4	2
	(65)	(87)	(86)	(91)
380°	3	1	2	4
	(147)	(155)	(127)	(212)
380°	4	3	2	1
	(153)	(90)	(100)	(108)
370°	4	1	3	2
	(150)	(140)	(121)	(142)
360°	1	4	2	3
	(33)	(54)	(8)	(46)

Notice the progression of temperatures, which would suggest that temperature was not run in random order. Indeed that was the case but we want to compare the coatings, not the temperatures, and the nonrandom temperature sequence does not affect those comparisons.

Analyze the data and draw a conclusion regarding the coatings. Specifically, is there a coatings effect? If so, what coating would you recommend for use? Is there a specific temperature setting that you would recommend with that coating setting if you feel that one coating is better than the others? Explain. (You may wish to compare your conclusions with those given by the author. In addition to the stated reference, the technical report, which is #131, can be downloaded at www.engr.wisc.edu/centers/cqpi.)

9.7 Explain the difference between a split-plot design and a strip-block design.

9.8 Taguchi (1987, pp. 445–456) gave an application of a $2^{13-9} \times 2^{3-1}$ split-plot design to an experiment on washing and carding of wool. How many effects would be plotted in a normal probability plot analysis of SP factors and their interactions? How many effects would be plotted in the corresponding analysis of WP factors?

9.9 Using the notation of Bisgaard (2000), give three different combinations of values of k, p, q, and r in a $2^{(k-p)-(q-r)}$ design for which you would recommend that a normal probability plot of SP factors and their interactions not be constructed. Would you also suggest that a normal probability plot of WP factors not be constructed for these three combinations? Explain. If you would suggest that the plots be constructed, give three combinations for which you would advise against the construction of the normal plot.

9.10 Consider a split-plot design with one WP factor and one SP factor. There are four levels of the WP factor and two replications of these are used. There are three levels of the SP factor. How many degrees of freedom will there be for the SP error?

9.11 Assume that a split-split-plot design is to be used and there are two WP factors, two first-level SP factors and one second-level SP factor, with all factors having two levels. Would you recommend that a Plackett–Burman design be considered? Why, or why not?

9.12 An inappropriate analysis of data from what was actually a split-plot design with three replications of the whole-plot factor revealed the SP factor to be significant and the WP factor to not be significant, although the factors were simply labeled as A and B, and an analysis of the "factorial design" was made under the assumption of complete randomization. A proper analysis of the data as having come from a split-plot arrangement revealed that the WP factor was significant. What must have been the general relationship between the magnitude of the error term in the factorial design analysis and the magnitude of the WP and SP error terms in the split-plot analysis?

9.13 It is important to recognize a split-plot data structure when it exists, as was emphasized in the chapter. It is also important, however, although less frequently stressed, to recognize when such a structure does *not* exist. The following is a message posted on an Internet message board several years ago, with some necessary editing.

> We are planning an experiment to test if plants produced by selfing are less fit than those produced by outcrossing. We have plants from three different alpine valleys, picked randomly among all the possible valleys. In each valley, we have a number of individuals, also picked at random. Seeds from these individuals were brought back to the green house and sawned. When they flowered, five of the flowers were selfed and five were outcrossed. The number of seeds produced by each flower was then recorded.
>
> Would you all agree that valleys and individuals are random factors? If so, would you also agree that the data should be analysed as a split-plot design?

Explain to this person why the data should not be analyzed as having come from a split-plot design.

9.14 Consider Example 9.1. Assume that after the experiment was run and the data were collected, a discovery was made that suggests the data from the second and third replications should not be used because of problems with the experiment. What will the whole-plot error sum of squares be if computed using just the four numbers from the first replicate? Will it be necessary or at least desirable to estimate the whole-plot standard error from a normal probability plot? Explain.