19 Split-Plot Designs

19.1 Introduction

Split-plot designs are needed when the levels of some treatment factors are more difficult to change during the experiment than those of others. The designs have a nested blocking structure. In a block design, the experimental units are nested within the blocks, and a separate random assignment of units to treatments is made within each block. In a split-plot design, the experimental units are called *split plots*, and are nested within *whole plots*, which themselves may or may not be nested within blocks.

The split plots within each whole plot are assigned at random to the levels of one or more of the treatment factors. The levels of other treatment factors are assigned to whole plots and remain constant for all split plots within a whole plot. Typically, these will be the factors whose levels are difficult to change, and the effects of their levels will be less precisely compared than those assigned to the split plots.

In Sect. 19.2 we show an example of an experiment designed as a split-plot design, together with a typical model for this type of design. The analysis of split-plot designs is discussed in Sect. 19.3 and illustrated via a second experiment. Designs with an extra level of nesting (split-split-plot designs) are briefly described in Sect. 19.4, and the issue of confounding treatment contrasts is introduced in Sect. 19.5. Section 19.6 introduces an experiment using a split-plot design without blocking. Section 19.7 introduces an experiment planned as a split-plot design with blocking, but that can also be viewed as a split-split-plot design. The use of SAS and R for analysis of split-plot designs is illustrated in Sects. 19.8 and 19.9, respectively.

19.2 Designs and Models

When a factorial experiment is run as a completely randomized design or a randomized complete block design, the levels of all the factors generally have to be changed frequently during the course of the experiment. For example, in Block I of the design in Table 13.13, p. 452, we see that as the experiment progressed on day 1, the level of the first factor had to be changed from 1 to 0 to 1 to 0 to 1, and the level of the second factor had to be changed from 0 to 1 to 0 to 1 to 0. The levels of the third and fourth factors also had to be changed four times. In most experiments this is no particular problem, but sometimes the level of one of the factors is *not* particularly easy to change.

An experiment is described by Munro in his 1986 University of Southampton dissertation on the effect of lighting conditions (factor A) and the speed of a rotating drum (factor B) on a subject's ability

A. Dean et al., *Design and Analysis of Experiments*,
Springer Texts in Statistics, DOI 10.1007/978-3-319-52250-0_19

Table 19.1 Part of a split-plot design for the rotating drum experiment

	Whole plot 1 (Session 1)		Whole plot 2 (Session 2)	
Block (Subject)	Level of A (Lighting)	Levels of B (Speed)	Level of A (Lighting)	Levels of B (Speed)
I	0	0 3 1 2	1	1 0 2 3
II	0	1 0 2 3	1	2 1 3 0
III	1	2 1 3 0	0	3 2 0 1
:	:	: : : :	:	: : : :
s	0	0 1 3 2	1	2 3 1 0

to focus on the center of the drum. In this experiment, it was easy to change the speed of rotation by the turn of a dial. The lighting conditions, however, took time to set up, and Munro wished to change these as seldom as possible. He therefore asked each subject to view all the rotation speeds (in a randomized order) under one set of lighting conditions during one session and return for a second session with different lighting conditions at a later date. Part of a possible design is shown in Table 19.1.

A whole plot is defined by a session for a particular subject. A split plot is defined by a time slot nested in a particular session for a particular subject. The two whole plots (sessions) within each block are assigned at random to the levels of one factor (A), and the four split plots (time slots) within each whole plot are assigned at random to the levels of the other factor (B).

If we look at the design in Table 19.1 and ignore factor B, we see that the levels of A are assigned according to a randomized complete block design, where the s subjects play the role of blocks, the 2 whole plots per block play the role of experimental units, and the 2 levels of A are assigned at random to the 2 whole plots within each block. Assuming no block $\times A$ interaction, the difference in the two levels of A could be analyzed like any randomized complete block design, using the whole-plot totals as the observations.

If we now look at the levels of B, they have also been assigned according to a randomized complete block design, but this time, the whole plots play the role of the blocks, and the four split plots nested within each whole plot are assigned to the four levels of B.

The analysis of the split-plot design is divided into two parts, reflecting this nested blocking system, each part with its own error. Analysis of the main effect of A involves comparisons of responses from split plots in different whole plots, whereas analysis of the main effect of B and the AB interaction involve comparisons of responses from split plots within the same whole plots.

In general, split plots within a whole plot will be more similar than split plots in different whole plots. Consequently, within-whole-plot comparisons will generally be more precise than between-whole-plot comparisons. So, in the rotating drum experiment, the main effect of B and the AB interaction will very likely be more precisely estimated than the main effect of A.

Ignoring the effects of the treatment factors for the moment, the response could be modeled as

$$Y_{hpq} = \mu + \theta_h + \epsilon^W_{p(h)} + \epsilon^S_{q(hp)}\,,$$

where θ_h is the effect of the hth block, $\epsilon^W_{p(h)}$ is the effect of the pth whole plot nested within the hth block, and $\epsilon^S_{q(hp)}$ is the effect of the qth split plot nested within the pth whole plot in the hth block. We model the whole-plot and split-plot effects, and possibly the block effects, as random effects that are independent and normally distributed with mean 0 and variances σ^2_W, σ^2_S, and σ^2_θ, respectively.

Now, suppose that the levels of factor A are assigned to the whole plots, and in the hth block, the pth whole plot receives the uth assignment of the ith level of A ($i = 1, \ldots, a$; $u = 1, \ldots, \ell$). Also, suppose that the levels of factor B are assigned to the split plots, and in the (hp)th whole plot, the qth split plot receives the tth assignment of the jth level of B ($j = 1, \ldots, b$; $t = 1, \ldots, m$). Then the model includes the effects of A, B and AB, and p is replaced by iu and q is replaced by jt, as follows:

$$\begin{aligned} Y_{hiujt} &= \mu + \theta_h + \alpha_i + \epsilon^W_{iu(h)} \\ &\quad + \beta_j + (\alpha\beta)_{ij} + \epsilon^S_{jt(hiu)}\,, \end{aligned} \tag{19.2.1}$$
$$\epsilon^W_{iu(h)} \sim N(0, \sigma^2_W)\,, \quad \epsilon^S_{jt(hiu)} \sim N(0, \sigma^2_S)\,,$$
$$\epsilon^W_{iu(h)}\text{'s and } \epsilon^S_{jt(hiu)}\text{'s are all mutually independent,}$$
$$h = 1, \ldots, s;\quad i = 1, \ldots, a;\quad u = 1, \ldots, \ell;\quad j = 1, \ldots, b;\quad t = 1, \ldots, m\,,$$

where θ_h is the effect of the hth block, α_i is the effect of the ith level of factor A measured on the whole plots, the random variables $\epsilon^W_{iu(h)}$ represent the random effects of the whole plots, β_j is the effect of the jth level of B measured on the split plots, $(\alpha\beta)_{ij}$ is the interaction effect of A at level i and B at level j, and the random variables $\epsilon^S_{jt(hiu)}$ represent the random effects of the split plots.

The model (19.2.1) has been written on two lines to emphasize the two different parts of the design. In the design of Table 19.1, each level of A appears exactly once per block and each level of B appears exactly once per whole plot, so we may drop the subscripts u and t, and the model for this design becomes

$$\begin{aligned} Y_{hij} &= \mu + \theta_h + \alpha_i + \epsilon^W_{i(h)} \\ &\quad + \beta_j + (\alpha\beta)_{ij} + \epsilon^S_{j(hi)}\,, \end{aligned} \tag{19.2.2}$$
$$\epsilon^W_{i(h)} \sim N(0, \sigma^2_W)\,, \quad \epsilon^S_{j(hi)} \sim N(0, \sigma^2_S)\,,$$
$$\epsilon^W_{i(h)}\text{'s and } \epsilon^S_{j(hi)}\text{'s are all mutually independent,}$$
$$h = 1, \ldots, s;\quad i = 1, \ldots, a;\quad j = 1, \ldots, b.$$

In some experiments the whole plot is the largest unit, which is equivalent to there being only one block ($s = 1$). In this case, model (19.2.1) becomes simpler, since the block effect θ_h and all subscripts h are omitted:

$$\begin{aligned} Y_{iujt} &= \mu + \alpha_i + \epsilon^W_{iu} \\ &\quad + \beta_j + (\alpha\beta)_{ij} + \epsilon^S_{jt(iu)}\,, \end{aligned} \tag{19.2.3}$$
$$\epsilon^W_{iu} \sim N(0, \sigma^2_W)\,, \quad \epsilon^S_{jt(iu)} \sim N(0, \sigma^2_S)\,,$$
$$\epsilon^W_{iu}\text{'s and } \epsilon^S_{jt(iu)}\text{'s are all mutually independent,}$$
$$i = 1, \ldots, a;\quad u = 1, \ldots, \ell;\quad j = 1, \ldots, b;\quad t = 1, \ldots, m.$$

For unequal sample sizes, the ranges of the subscripts would be modified in models (19.2.1) and (19.2.3).

19.3 Analysis of a Split-Plot Design with Complete Blocks

In this section we consider only the case of equal sample sizes and randomized complete block designs for each of the treatment factors. There are then s blocks, each of which is divided into a whole plots, and each of these is subdivided into b split plots, giving a total of sab observations. Model (19.2.2) is used, and the degrees of freedom and sums of squares are calculated according to the rules in Chap. 18,

Table 19.2 Outline analysis of variance table for the rotating drum split-plot design

Source of variation	Degrees of freedom	Sum of squares	Mean square	Ratio
Block (Subjects)	$s-1$	$ss\theta$	–	–
A (Lighting)	$a-1$	ssA	msA	msA/msE_W
Whole-plot error	$(s-1)(a-1)$	ssE_W	msE_W	
Whole-plot total	$sa-1$	ssW	–	–
B (Speed)	$b-1$	ssB	msB	msB/msE_S
AB	$(a-1)(b-1)$	$ss(AB)$	$ms(AB)$	$ms(AB)/msE_S$
Split-plot error	$a(b-1)(s-1)$	ssE_S	msE_S	
Total	$abs-1$	$sstot$		

Computational formulae

$$ss\theta = ab\Sigma_h \overline{y}^2_{h..} - sab\overline{y}^2_{...} \qquad ssW = b\Sigma_h \Sigma_i \overline{y}^2_{hi.} - sab\overline{y}^2_{...}$$

$$ssA = sb\Sigma_i \overline{y}^2_{.i.} - sab\overline{y}^2_{...} \qquad ssB = sa\Sigma_j \overline{y}^2_{..j} - sab\overline{y}^2_{...}$$

$$ssE_W = ssW - ss\theta - ssA \qquad ss(AB) = s\Sigma_i \Sigma_j \overline{y}^2_{.ij} - sb\Sigma_i \overline{y}^2_{.i.}$$

$$sstot = \Sigma_h \Sigma_i \Sigma_j y^2_{hij} - sab\overline{y}^2_{....} \qquad - sa\Sigma_j \overline{y}^2_{..j} + sab\overline{y}^2_{...}$$

$$ssE_S = sstot - ssW - ssB - ss(AB)$$

as shown in the following two subsections. The analysis of variance is outlined in Table 19.2 and, for this setting, is appropriate whether the blocks are fixed or random. The analysis of a split plot design with incomplete blocks is illustrated in Sects. 19.8.4 and 19.9.4, using the SAS and R software, respectively.

19.3.1 Split-Plot Analysis

Consider first the *split-plot analysis*, which is that part of the analysis (shown in the bottom half of the analysis of variance Table 19.2) that is based on the observations arising from the split plots within whole plots. There are $sab-1$ total degrees of freedom, and the total sum of squares is

$$sstot = \sum_h \sum_i \sum_j y^2_{hij} - sab\overline{y}^2_{...} \,. \tag{19.3.4}$$

The b levels of factor B are assigned to the split plots within each whole plot according to a randomized complete block design. The sa whole plots are playing the role of sa blocks, so there are $sa-1$ whole-plot degrees of freedom, and the whole-plot-total sum of squares is

$$ssW = b\sum_h \sum_i \overline{y}^2_{hi.} - sab\overline{y}^2_{...} \,. \tag{19.3.5}$$

Due to the fact that all levels of B are observed in every whole plot as in a randomized complete block design, the sum of squares for B needs no adjustment for whole plots, and is given by

$$ssB = sa\sum_j \overline{y}^2_{..j} - sab\overline{y}^2_{...} \tag{19.3.6}$$

corresponding to $b-1$ degrees of freedom. The interaction between the factors A and B is also calculated as part of the split-plot analysis. Again, due to the complete block structure of both the whole-plot design and the split-plot design, the interaction sum of squares needs no adjustment for blocks. The number of interaction degrees of freedom is $(a-1)(b-1) = ab - a - b + 1$, and the sum of squares is

$$ss(AB) = s\sum_i\sum_j \overline{y}^2_{.ij} - sb\sum_i \overline{y}^2_{.i.} - sa\sum_j \overline{y}^2_{..j} + sab\overline{y}^2_{...}\,. \quad (19.3.7)$$

Since there are b split plots nested within the sa whole plots, there are, in total, $sa(b-1)$ split-plot degrees of freedom. Of these, $b-1$ are used to measure the main effect of B, and $(a-1)(b-1)$ are used to measure the AB interaction, leaving

$$sa(b-1) - (b-1) - (a-1)(b-1) = a(s-1)(b-1)$$

degrees of freedom for error. Equivalently, this can be obtained by subtraction of the whole-plot, B, and AB degrees of freedom from the total

$$(sab-1) - (sa-1) - (b-1) - (a-1)(b-1) = a(s-1)(b-1)\,.$$

The split-plot error sum of squares can also be calculated by subtraction:

$$ssE_S = sstot - ssW - ssB - ss(AB)\,. \quad (19.3.8)$$

The split-plot error mean square $msE_S = ssE_S/[a(s-1)(b-1)]$ is used as the error estimate in testing hypotheses and calculating confidence intervals for contrasts in B and AB. Notice that we cannot compare the levels of factor A on the split plots, since within each whole plot the level of A is held constant. The A contrasts are, in fact, confounded with whole plots.

The sums of squares (19.3.4)–(19.3.8) and their associated degrees of freedom are summarized in the bottom half of the analysis of variance table shown in Table 19.2.

19.3.2 Whole-Plot Analysis

We now move on to the *whole-plot analysis*, which is the part of the analysis based on comparisons of whole-plot totals. The levels of A are assigned to the whole plots within blocks according to a randomized complete block design, and so the sum of squares for A needs no block adjustment. There are $a-1$ degrees of freedom for A, so the sum of squares is given by

$$ssA = sb\sum_i \overline{y}^2_{.i.} - sab\overline{y}^2_{...}\,. \quad (19.3.9)$$

There are $s-1$ degrees of freedom for blocks, giving a block sum of squares of

$$ss\theta = ab\sum_h \overline{y}^2_{h..} - sab\overline{y}^2_{...}\,. \quad (19.3.10)$$

There are a whole plots nested within each of the s blocks, so there are, in total, $s(a-1)$ whole-plot degrees of freedom. Of these, $a-1$ are used to measure the effects of A leaving $(s-1)(a-1)$ degrees of freedom for whole-plot error. Equivalently, this can be obtained by the subtraction of the block and A degrees of freedom from the whole-plot total degrees of freedom

$$(sa - 1) - (s - 1) - (a - 1) = (s - 1)(a - 1)\,.$$

Similarly, the whole-plot error sum of squares, which is used in testing hypotheses and calculating confidence intervals for contrasts in factor A, is obtained by subtraction:

$$ssE_W = ssW - ss\theta - ssA\,. \tag{19.3.11}$$

The sums of squares (19.3.9)–(19.3.11) and their corresponding degrees of freedom are summarized in the top half of Table 19.2 (p. 706).

If the whole plot is the largest unit—namely, if there is no blocking of the whole plots—then the sum of squares for blocks in Table 19.2 effectively gets pooled into the whole-plot error sum of squares. Then the latter, still used in testing hypotheses and calculating confidence intervals for contrasts in factor A, is given by

$$ssE_W = ssW - ssA\,, \tag{19.3.12}$$

with $(sa - 1) - (a - 1) = a(s - 1)$ degrees of freedom.

19.3.3 Contrasts Within and Between Whole Plots

The formulae for the least squares estimates of the main effect and interaction treatment contrasts are similar to those given by rule 10 of Sect. 7.3 for fixed effects, since no block adjustments are needed. Thus

$$\begin{aligned} \sum_i c_i \hat{\alpha}_i^* &= \sum_i c_i \overline{y}_{.i.}\,, \\ \sum_j d_j \hat{\beta}_j^* &= \sum_j d_j \overline{y}_{..j}\,, \\ \sum_i \sum_j k_{ij} \widehat{(\alpha\beta)}_{ij} &= \sum_i \sum_j k_{ij} \overline{y}_{.ij}\,, \end{aligned} \tag{19.3.13}$$

where $\sum_i c_i = 0$, $\sum_j d_j = 0$, and $\sum_i k_{ij} = \sum_j k_{ij} = 0$.

Consider the consequences of confounding factor A with whole plots. Using model (19.2.2), the corresponding main-effect-of-A contrast estimator can be expressed as

$$\sum_i c_i \hat{\alpha}_i^* = \sum_i c_i \overline{Y}_{.i.} = \sum_i c_i (\alpha_i^* + \overline{\epsilon}^W_{i(.)} + \overline{\epsilon}^S_{.(.i)}).$$

Consequently, this estimator has mean $\sum_i c_i \alpha_i^*$ and variance $\sum_i (c_i^2/sb)(b\sigma_W^2 + \sigma_S^2)$, where $(b\sigma_W^2 + \sigma_S^2)$ replaces σ^2 in rule 11 of Sect. 7.3 for fixed effects models. So, although main effects of A are confounded with whole plots, because the whole plot effects are random with mean zero, main effects of A are estimable but with larger variance reflecting whole plot variability in addition to split plot variability.

For the estimates in Eq. (19.3.13), the corresponding estimated variances reflect whether the contrasts are measured in terms of whole-plot differences (as for contrasts in the levels of A) or split-plot (within whole-plot) differences (as for contrasts in B or AB). The former use the whole-plot error mean square, and the latter use the split-plot error mean square as follows.

$$\widehat{\mathrm{Var}}\left(\sum_i c_i \hat{\alpha}_i^*\right) = \sum_i \frac{c_i^2}{sb}\, msE_W\,, \tag{19.3.14}$$

$$\widehat{\mathrm{Var}}\left(\sum_j d_j \hat{\beta}_j^*\right) = \sum_j \frac{d_j^2}{sa}\, msE_S\,,$$

$$\widehat{\mathrm{Var}}\left(\sum_i \sum_j k_{ij} \widehat{(\alpha\beta)}_{ij}\right) = \sum_i \sum_j \frac{k_{ij}^2}{s}\, msE_S\,.$$

For main effect and interaction contrasts, the methods of multiple comparison of Bonferroni, Scheffé, Tukey, and Dunnett can be used as usual (incorporating either the whole-plot or split-plot error mean square as above). Inferences for other contrasts in the treatment effects $\tau_{ij} = \alpha_i + \beta_j + (\alpha\beta)_{ij}$, such as all pairwise comparisons, are more complicated and are discussed later in this chapter.

19.3.4 A Real Experiment—Oats Experiment

An experiment on the yield of three varieties of oats (factor A) and four different levels of manure (factor B) was described by F. Yates in his 1935 paper *Complex Experiments*. The experimental area was divided into $s = 6$ blocks. Each of these was then subdivided into $a = 3$ whole plots. The varieties of oat were sown on the whole plots according to a randomized complete block design (so that every variety appeared in every block exactly once). Each whole plot was then divided into $b = 4$ split plots, and the levels of manure were applied to the split plots according to a randomized complete block design (so that every level of B appeared in every whole plot exactly once). The design, after randomization, is shown in Table 19.3, together with the yields in quarter pounds. Model (19.2.2) was used.

Analysis of Variance—Oats Experiment

Using the formulae (19.3.4)–(19.3.11), we obtain the sums of squares shown in Table 19.4. To test, at significance level $\alpha = 0.01$, the hypothesis H_0^{AB} that the interaction between variety and manure level is negligible against the alternative hypothesis that the interaction is not negligible, we reject H_0^{AB} if

$$\frac{ms(AB)}{msE_S} = \frac{53.63}{177.08} = 0.30 > F_{6,45,0.01}\,.$$

Since $F_{6,45,0.01} \approx 3.2$, we do not reject H_0^{AB}, and we conclude that the interaction is negligible.

The hypothesis H_0^B of no difference in yield due to the different levels of manure (averaged over variety) is also tested using the split-plot error mean square as the denominator. We reject H_0^B in favor of the alternative hypothesis, that the manure levels do affect yield of oats, if

$$\frac{msB}{msE_S} = \frac{6673.50}{177.08} = 37.69 > F_{3,45,0.01}\,.$$

Since $F_{3,45,0.01} \approx 4.3$, we conclude that these four manure levels have different effects on the yield of the oat varieties tested.

Factor A is measured on the whole plots, so the whole-plot error is used as the denominator of the test statistic. We reject the hypothesis H_0^A of no difference in the average yields of the different varieties averaged over the manure levels if

Table 19.3 Split-plot design and yields (in quarter lb) for the oats experiment

Block	Level of A	Level of B (yield)	Block	Level of A	Level of B (yield)
I	2	3 (156) 2 (118) 1 (140) 0 (105)	II	2	2 (109) 3 (99) 0 (63) 1 (70)
	0	0 (111) 1 (130) 3 (174) 2 (157)		1	0 (80) 2 (94) 3 (126) 1 (82)
	1	0 (117) 1 (114) 2 (161) 3 (141)		0	1 (90) 2 (100) 3 (116) 0 (62)
III	2	2 (104) 0 (70) 1 (89) 3 (117)	IV	1	3 (96) 0 (60) 2 (89) 1 (102)
	0	3 (122) 0 (74) 1 (89) 2 (81)		0	2 (112) 3 (86) 0 (68) 1 (64)
	1	1 (103) 0 (64) 2 (132) 3 (133)		2	2 (132) 3 (124) 1 (129) 0 (89)
V	1	1 (108) 2 (126) 3 (149) 0 (70)	VI	0	2 (118) 0 (53) 3 (113) 1 (74)
	2	3 (144) 1 (124) 2 (121) 0 (96)		1	3 (104) 2 (86) 0 (89) 1 (82)
	0	0 (61) 3 (100) 1 (91) 2 (97)		2	0 (97) 1 (99) 2 (119) 3 (121)

Source Yates (1935).

$$\frac{msA}{msE_W} = \frac{893.18}{601.33} = 1.49 > F_{2,10,0.01}\,.$$

Since $F_{2,10,0.01} = 7.56$, there is no evidence to conclude a difference in average yields of the three varieties of oats.

The same conclusions could be reached from the p-values in Table 19.4.

Multiple Comparisons—Oats Experiment

Suppose that level 0 of A was the currently used variety and that level 0 of B was the usual level of manure, and suppose that two-sided treatment-versus-control intervals had been required for both A and B at an overall level of 98% (that is, 99% for each set of Dunnett intervals).

Since both the split-plot and whole-plot designs are randomized complete block designs, the least squares estimates of the treatment contrasts are given by the formula in (19.3.13), so

$$\begin{aligned} \hat{\alpha}_0^* - \hat{\alpha}_1^* &= \overline{y}_{.0.} - \overline{y}_{.1.} = -6.875\,, \\ \hat{\alpha}_0^* - \hat{\alpha}_2^* &= \overline{y}_{.0.} - \overline{y}_{.2.} = -12.167\,, \end{aligned} \qquad \begin{aligned} \hat{\beta}_0^* - \hat{\beta}_1^* &= \overline{y}_{..0} - \overline{y}_{..1} = -19.500\,, \\ \hat{\beta}_0^* - \hat{\beta}_2^* &= \overline{y}_{..0} - \overline{y}_{..2} = -34.833\,, \\ \hat{\beta}_0^* - \hat{\beta}_3^* &= \overline{y}_{..0} - \overline{y}_{..3} = -44.000\,, \end{aligned}$$

Table 19.4 Analysis of variance for the oats split-plot experiment

Source of variation	Degrees of freedom	Sum of square	Mean square	Ratio	p-value
Blocks	5	15875.28	3175.06	–	
A (variety)	2	1786.36	893.18	1.49	0.2724
Whole-plot error	10	6013.31	601.33		
Whole-plot total	17	23674.94	1392.64		
B (manure)	3	20020.50	6673.50	37.69	0.0001
AB	6	321.75	53.63	0.30	0.9322
Split-plot error	45	7968.75	177.08		
Total	71	51985.94			

where $\overline{y}_{.i.}$ is an average over the $b = 4$ split plots within the $s = 6$ whole plots (one per block) on which level i of A is measured. Similarly, $\overline{y}_{..j}$ is an average over the $sa = 18$ split plots (one per whole plot) on which level j of B is measured. The confidence intervals are obtained from (19.3.13) and (19.3.14) as follows:

$$\begin{aligned}
\alpha_0 - \alpha_1 &\in \left((\overline{y}_{.0.} - \overline{y}_{.1.}) \pm t^{(0.5)}_{2,10,0.01}\sqrt{\frac{2}{24} msE_W} \right) \\
&= \left(-6.875 \pm 3.53\sqrt{\frac{2}{24}(601.33)} \right) \\
&= (-6.875 \pm 24.99) = (-31.87, 18.12)\,, \\
\alpha_0 - \alpha_2 &= (-37.16, 18.12)\,, \\
\beta_0 - \beta_1 &\in \left((\overline{y}_{..0} - \overline{y}_{..1}) \pm t^{(0.5)}_{3,45,0.01}\sqrt{\frac{2}{18} msE_S} \right) \\
&= \left(-19.5 \pm 3.09\sqrt{\frac{2}{18}(177.03)} \right) \\
&= (-19.5 \pm 13.70) = (-33.20, -5.79)\,, \\
\beta_0 - \beta_2 &\in (-48.54, -21.13)\,, \\
\beta_0 - \beta_3 &\in (-57.70, -30.30)\,.
\end{aligned}$$

It is clear that the treatment-versus-control comparisons for the factor B manure levels are made more precisely ($msd = 13.70$) than those for the factor A oat varieties ($msd = 24.99$). This is primarily due to the much smaller error variance estimate, $msE_S < msE_W$, which reflects the fact that split plots within a whole plot are generally more similar than split plots in different whole plots. There are also more degrees of freedom associated with the split-plot error than with the whole-plot error, which also helps to reduce the minimum significant difference. Comparisons for factor B are more precise, despite the fact that the means $\overline{y}_{.i.}$ for factor A involve more observations.

19.4 Split-Split-Plot Designs

In the split-plot designs illustrated in Sect. 19.2, the factor A contrasts were confounded with the whole-plot contrasts, so that the main effect of A was assessed against the whole-plot variability, while the main effect of B and the AB interaction were assessed against the split-plot variability. It is possible

Table 19.5 Part of a split-split-plot design for the rotating drum experiment

			Split-plot 1 First half session		Split-plot 2 Second half session	
Block (Subject)	Whole-plot (Session)	Level of A (Light)	Level of C (Direction)	Levels of B (Speed)	Level of C (Direction)	Levels of B (Speed)
I	1	0	1	0 3 1 2	0	2 1 3 0
	2	1	0	1 0 2 3	1	3 2 0 1
II	1	1	1	1 0 2 3	0	0 3 1 2
	2	0	0	2 1 3 0	1	3 2 0 1
:	:	:	:	: : : :	:	: : : :

Table 19.6 Analysis of variance for a split-split-plot design

Source of variation	Degrees of freedom	Mean square	Ratio
Blocks (subjects)	$s-1$	$ms\theta$	
A (lighting)	$a-1$	msA	msA/msE_W
Whole-plot error	$(s-1)(a-1)$	msE_W	
Whole-plot total	$sa-1$	msW	
C (direction)	$c-1$	msC	msC/msE_S
AC	$(a-1)(c-1)$	$ms(AC)$	$ms(AC)/msE_S$
Split-plot error	$a(s-1)(c-1)$	msE_S	
Split-plot total	$sac-1$	msE_S	
B	$b-1$	msB	msB/msE_{SS}
AB	$(a-1)(b-1)$	$ms(AB)$	$ms(AB)/msE_{SS}$
CB	$(c-1)(b-1)$	$ms(CB)$	$ms(CB)/msE_{SS}$
ACB	$(a-1)(c-1)(b-1)$	$ms(ACB)$	$ms(ACB)/msE_{SS}$
Split-split-plot error	$ac(s-1)(b-1)$	msE_{SS}	
Total	$sacb-1$	$mstot$	

to extend this idea, and to divide the split plots into split split plots on which are assigned the levels of a third factor.

For example, in the drum rotation experiment described in Sect. 19.2, the experimenter used a third factor C, the direction of rotation of the drum. A possible design for the experiment would be to ask each subject to be present at two sessions (whole plots) with a different lighting condition (A) at each session. In the first half of a session (split plot), set the direction of rotation (C), and run through each speed (B) in a random order (split split plots), changing the direction of rotation in the second half of the session. The design would then appear as in Table 19.5.

The model and analysis of variance table would have three parts, one for the whole plots nested within blocks together with the factor A effect, one for the split plots nested within whole plots together with the factor C effect and the AC interaction, and one for the split split plots nested within split plots together with the factor B effect and the other interactions, as shown in model (19.4.15):

$$\begin{aligned} Y_{hijk} = \mu &+ \theta_h + \alpha_i + \epsilon^W_{i(h)} \\ &+ \gamma_j + (\alpha\gamma)_{ij} + \epsilon^S_{j(hi)} \\ &+ \beta_k + (\alpha\beta)_{ik} + (\gamma\beta)_{jk} + (\alpha\gamma\beta)_{ijk} + \epsilon^{SS}_{k(hij)}\,. \end{aligned} \tag{19.4.15}$$

The analysis of variance table, shown in Table 19.6, has three sections, reflecting the three parts of the model, and is illustrated through a real experiment in Sect. 19.7.3.

19.5 Split-Plot Confounding

If there are a number of different factors involved in a split-plot design, the size of the whole plots may not be large enough to allow a randomized complete block design to be used for the split-plot factors. We can obtain smaller blocks by confounding one or more interaction contrasts as we did for the single-replicate designs in Chap. 13. For example, suppose a two-replicate 2^4 experiment is to be conducted for treatment factors A, B, C, and D, and for practical reasons, the observations are to be divided into eight whole plots of size four. Suppose that the levels of A are sufficiently difficult to change that it is decided to change the level only after each whole plot of four observations is taken. The A contrasts are confounded with the whole plots, since each whole plot is assigned only one level of A.

Now, only four of the eight combinations of factors B, C, and D can be taken in any one whole plot. Thus, ignoring factor A, a design with $b = 8$ whole plots of size $k = 4$ and $v = 8$ treatment labels is required. An incomplete block design, such as a cyclic design, would be a possible choice. However, since a split-plot design is a complex design to analyze, it is better to select a repeated single-replicate design, so that we know exactly what is confounded with the whole plots. If we choose to confound BCD with the whole plots, as well as A, then $ABCD$ will also be confounded. The single-replicate design that confounds A, BCD, and $ABCD$ is obtained from the equations

$$\begin{aligned} a_1 \qquad\qquad &= 0 \text{ or } 1 \bmod 2\,, \\ a_2 + a_3 + a_4 &= 0 \text{ or } 1 \bmod 2\,. \end{aligned}$$

If we repeat this single-replicate design twice, we obtain the split-plot plan shown in Table 19.7. Before this plan can be used, the eight whole plots would need to be randomly ordered, and the four split plots within each whole plot would need to be randomly ordered. An outline analysis of variance table is shown in Table 19.8.

19.6 A Real Experiment—UAV Experiment

Sriram Mahadevan (2009) conducted three experiments at Wright State University to evaluate the performance of a semi-automated computer display system designed to support a human operator's ability to monitor and control the complex dynamic operation of multiple unmanned aerial vehicles (UAVs) when the UAVs are involved in multiple combat-related tasks. One of his experiments was conducted to examine the effects of different visual tools on the task performance efficiency and situation awareness of a single operator at a computer display to monitor and control dual-task scenarios involving UAVs.

The experiment involved 16 subjects (factor W) and $a = 2$ cue conditions (factor A), with eight of the 16 participants utilizing a baseline user interface with basic visual tools ($A = 1$), and with

Table 19.7 A split-plot confounded 2^4 experiment in 8 whole plots of size 4

Whole plot	A	Levels of B, C, D on the split plots				Whole plot	A	Levels of B, C, D on the split plots			
I	0	000	011	101	110	II	1	001	010	100	111
III	0	001	010	100	111	IV	1	000	011	101	110
V	0	000	011	101	110	VI	1	001	010	100	111
VII	0	001	010	100	111	VIII	1	000	011	101	110

Table 19.8 Outline analysis of variance table for a split-plot confounded 2^4 experiment in 8 whole plots of size 4

Source of variation	Degrees of freedom	Mean square	Ratio
A	1	msA	msA/msE_W
BCD	1	$ms(BCD)$	$ms(BCD)/msE_W$
$ABCD$	1	$ms(ABCD)$	$ms(ABCD)/msE_W$
Whole-plot error	4	msE_W	
Whole-plot total	7	msW	
B	1	msB	msB/msE_S
C	1	msC	msC/msE_S
D	1	msD	msD/msE_S
BC	1	$ms(BC)$	$ms(BC)/msE_S$
BD	1	$ms(BD)$	$ms(BD)/msE_S$
CD	1	$ms(CD)$	$ms(CD)/msE_S$
AB	1	$ms(AB)$	$ms(AB)/msE_S$
AC	1	$ms(AC)$	$ms(AC)/msE_S$
AD	1	$ms(AD)$	$ms(AD)/msE_S$
ABC	1	$ms(ABC)$	$msABC/msE_S$
ABD	1	$ms(ABD)$	$ms(ABD)/msE_S$
ACD	1	$ms(ACD)$	$ms(ACD)/msE_S$
Split-plot error	12	msE_S	
Total	31		

the other eight participants utilizing an advanced user interface involving more advanced visual tools ($A = 2$). Each subject ran through eight trials—one at each combination of task similarity (factor B, with $b = 2$ levels) and task complexity (factor C, with $c = 4$ levels). Task similarity depended on whether the primary and secondary tasks (which were done concurrently) were similar (coded 1), each task being a suppression-of-enemy-air-defenses (SEAD) mission, or dissimilar (coded 2), the primary and secondary tasks being SEAD and reconnaissance missions, respectively. Task complexity corresponds to the number of UAVs in the scenario, the levels being: simple-simple if both the primary and secondary tasks each involves two UAVs; simple-complex if the primary and secondary tasks involve two and four UAVs, respectively; complex-simple if the primary and secondary tasks involve four and two UAVs, respectively; and complex-complex if both the primary and secondary tasks each involves four UAVs. These levels were coded 1, 2, 3, 4, respectively.

This may be viewed as a split-plot design, with subjects serving as the whole plots, and with the eight trials per subject serving as the split plots. Main effects of cue are comparisons between subjects (i.e. between whole plots), since subjects are nested within cue levels, whereas main effects of B and C and all interactions are comparisons within whole plots.

One of the response variables measured was the time taken to perform situation awareness perception tasks in the primary task, yielding the data shown in Table 19.9. The model, including all treatment effects, is as follows.

Table 19.9 Split-plot design and times (in seconds) for the UAV experiment: situation awareness perception in the primary task

B (Similarity)		1				2			
C (Complexity)		1	2	3	4	1	2	3	4
A (Cue)	*W* (Subject)								
1	1	29	28	49	46	36	35	42	48
	2	26	26	53	42	35	40	46	44
	3	33	25	45	56	34	36	40	40
	4	27	28	44	47	36	32	50	54
	5	31	26	47	48	33	36	44	60
	6	28	27	51	41	31	34	38	43
	7	34	28	57	44	34	32	49	35
	8	25	35	50	54	35	50	45	48
2	1	15	13	16	18	21	18	20	21
	2	12	11	19	22	18	19	25	22
	3	14	13	21	19	25	23	21	23
	4	18	16	17	20	24	27	25	24
	5	15	15	18	21	19	20	21	21
	6	13	16	19	17	21	22	23	25
	7	17	18	20	16	22	23	27	24
	8	14	14	16	19	20	21	22	22

Source Mahadevan (2009). Copyright © 2009 Sriram Mahadevan. Reprinted with permission

$$\begin{aligned} Y_{iujk} = \mu + \alpha_i + \epsilon^W_{iu} \\ + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon^S_{jk(iu)}\,, \\ \epsilon^W_{iu} \sim N(0, \sigma^2_W)\,, \quad \epsilon^S_{jk(iu)} \sim N(0, \sigma^2_S)\,, \\ \epsilon^W_{iu}\text{'s and } \epsilon^S_{jk(iu)}\text{'s are all mutually independent,} \\ i = 1, 2; \quad u = 1, \ldots, 8; \quad j = 1, 2; \quad k = 1, 2, 3, 4\,, \end{aligned} \tag{19.6.16}$$

This is analogous to model (19.2.1), except: (i) this model excludes block effects ($s = 1$); (ii) it includes the factorial effects γ_k, $(\alpha\gamma)_{ik}$, $(\beta\gamma)_{jk}$, and $(\alpha\beta\gamma)_{ijk}$ associated with factor C, all estimated as split-plot differences; and (iii) there is no replication of treatments within whole plots ($m = 1$).

19.6.1 Analysis of Variance

The analysis of variance table for the UAV experiment, given in Table 19.10, is similar to the one in Table 19.4 for the oats experiment, except for the following. There are no block effects, so the whole-plot error is obtained by subtraction as

$$ssE_W = ssW - ssA\,,$$

with the corresponding degrees of freedom adjusted accordingly. Also, more treatment effects are involved in the split-plot analysis so the error sum of squares, obtained by subtraction of other effect sums of squares from the total sum of squares, is

Table 19.10 Analysis of variance for UAV experiment

Source of variation	Degrees of freedom	Sum of square	Mean square	Ratio	p-value
A (cue)	1	12880.13	12880.13	693.30	0.0001
Whole-plot error	14	260.09	18.58		
Whole-plot total	15	13140.22	876.01		
B (similarity)	1	457.53	457.53	34.57	0.0001
C (complexity)	3	2470.03	823.34	62.22	0.0001
AB	1	98.00	98.00	7.41	0.0077
AC	3	1177.94	392.65	29.67	0.0001
BC	3	351.28	117.09	8.85	0.0001
ABC	3	153.31	51.10	3.86	0.0117
Split-plot error	98	1296.91	13.23		
Total	127	19145.22			

$$ssE_S = sstot - ssW - ssB - ssC - ss(AB) - ss(AC) - ss(BC) - ss(ABC)\,.$$

Factor A is tested relative to whole-plot error, while the other treatment effects are tested relative to the split-plot error. As one can see from the p-values in Table 19.10, all treatment main and interaction effects are significant at the individual $\alpha = .01$ levels, except for the ABC interaction which has a p-value slightly larger than .01.

19.6.2 Multiple Comparisons

The experimenter was interested in the main effect of cue, averaged over the other factors; that is,

$$\alpha_1^* - \alpha_2^* = [\alpha_1 + \overline{(\alpha\beta)}_{1.} + \overline{(\alpha\gamma)}_{1.} + \overline{(\alpha\beta\gamma)}_{1..}] - [\alpha_2 + \overline{(\alpha\beta)}_{2.} + \overline{(\alpha\gamma)}_{2.} + \overline{(\alpha\beta\gamma)}_{2..}],$$

comparing the effects of the two cue conditions, averaging over combinations of levels of similarity and complexity. Paralleling the discussion in Sect. 19.3: the least squares estimate of the main effect of cue is

$$\hat{\alpha}_1^* - \hat{\alpha}_2^* = \overline{y}_{1...} - \overline{y}_{2...} = 39.4531 - 19.3906 = 20.0625 \text{ seconds;}$$

also, $\text{Var}(\overline{Y}_{1...} - \overline{Y}_{2...}) = \text{Var}[(\overline{\epsilon}^W_{1.} + \overline{\epsilon}^S_{..(1.)}) - (\overline{\epsilon}^W_{2.} + \overline{\epsilon}^S_{..(2.)})] = 2(\sigma^2_W/8 + \sigma^2_S/64) = \sigma^2_W/4 + \sigma^2_S/32$, and this is estimated by $msE_W/32$. So, with 99% confidence,

$$\begin{aligned}\alpha_1^* - \alpha_2^* &\in \left((\overline{y}_{.1..} - \overline{y}_{.2..}) \pm t_{14,0.005}\sqrt{msE_W/32}\right)\\ &= \left(20.0625 \pm 2.977\sqrt{18.58/32}\right)\\ &= (20.0625 \pm 2.2684) = (17.7941,\ 22.3309)\,.\end{aligned}$$

Hence, with 99% confidence, the main effect of cue is between 17.7941 and 22.3309 seconds so, on average, operators using the enhanced user interface spend between 17.7941 and 22.3309 seconds less time to perform the requested situation awareness perception tasks in the primary task than do operators using the baseline user interface.

Given the significant main effect of cue, coupled with the p-value of 0.0117 for the ABC interaction, the experimenter decided to also investigate the simple pairwise differences comparing the effect of cue at each of the eight combinations of the other two factors. For each such combination jk of BC, the estimator of the simple pairwise difference

$$[\alpha_1 + (\alpha\beta)_{1j} + (\alpha\gamma)_{1k} + (\alpha\beta\gamma)_{2jk}] - [\alpha_2 + (\alpha\beta)_{2j} + (\alpha\gamma)_{2k} + (\alpha\beta\gamma)_{2jk}] \tag{19.6.17}$$

is $\overline{Y}_{1.jk} - \overline{Y}_{2.jk}$, with variance

$$\text{Var}[(\overline{\epsilon}^W_{1.} + \overline{\epsilon}^S_{jk(1.)}) - (\overline{\epsilon}^W_{2.} + \overline{\epsilon}^S_{jk(2.)})] = 2(\sigma^2_W/8 + \sigma^2_S/8) = (\sigma^2_W + \sigma^2_S)/4.$$

These simple pairwise differences are neither within- nor between-whole-plot comparisons, so a composite variance estimator is needed. Now, $E[MSE_S] = \sigma^2_S$. Also, one can show that $E[MSE_W] = 8\sigma^2_W + \sigma^2_S$, (by rule 17 of Sect. 17.8.1, there being 8 observations on each whole plot). So, the variance estimate is

$$(msE_W + 7msE_S)/32 = (18.578 + 7(13.234))/32 \approx 3.4755,$$

with estimated standard error 1.8643 obtained as the square root. The degrees of freedom associated with this variance estimate is computed using Satterthwaite's approximation as follows:

$$df = \frac{(msE_W + 7msE_S)^2}{(msE_W)^2/14 + (7msE_S)^2/98} = \frac{12,369}{24.6533 + 87.5659} \approx 110\,.$$

So for example, an individual 99% confidence interval for (19.6.17) has minimum significant difference $msd = (t_{110,0.005})(1.8643) = (2.621)(1.8643) \approx 4.8863$. The corresponding simple pairwise difference estimates are as follows,

jk	11	12	13	14	21	22	23	24
$\overline{y}_{.1jk} - \overline{y}_{.2jk}$	14.375	13.375	31.250	28.250	13.000	15.250	21.250	23.750

where levels 1 and 2 of B are similar and dissimilar, respectively, and levels 1–4 of C are simple-simple, simple-complex, complex-simple, and complex-complex, respectively. Since each of these simple pairwise difference estimates exceeds the minimum significant difference 4.8863 and is positive, the enhanced user interface provides significant mean time reductions in performing the requested situation awareness perception tasks in the primary task for each combination of task similarity and task complexity. The overall significance level for testing these eight differences would have been at most 0.08, had these comparisons been preplanned. While they perhaps were not, one can also show that each of these comparisons is significantly nonzero with a p-value of less than 0.0001 when tested individually, and so one can feel rather confident that the effects are real.

19.7 A Real Experiment—Mobile Computing Field Study

Mary Mc. Wesler (2001) conducted a field experiment to study the effective use of mobile computing devices for real-time navigation and situation awareness, with applications in the military domain. One goal of the experiment was to study the effects of display type and visual presentation format on navigational performance, taking information garnered from laboratory studies and putting it to the test in more realistic field studies. Two display types (factor A) were studied: a handheld, headdown display; and a helmet mounted display, facilitating headup usage. Three visual presentation formats (factor B) were studied: second person egocentric; birds' eye view egocentric; and birds' eye view exocentric. Each experimental run consisted of a subject attempting to navigate, quickly and accurately, a prescribed route or path through unfamiliar, moderate, lightly wooded terrain in the dark of night. Response variables of interest included: elapsed time, distance traveled, and root mean square error of distance off of the prescribed path. The latter two response variables were based on GPS data.

The experiment involved 12 subjects and 24 days, with each subject participating for two days. There were various constraints on the experiment. For example, on any given day it was feasible to make three runs with one subject, with subjects and days both considered possible sources of variation. Involving each subject for two days provided a single replicate of the 2×3 combinations of A and B with each subject, allowing all treatment effects to be within-subject comparisons. However, with only three runs per day, it was decided to fix display type but vary visual presentation format each day, so visual presentation format comparisons are within days, but display type comparisons are between days. Correspondingly, each subject only needed to be trained on one display type each day. In short, the experimenter used a split-plot design—the 12 subjects constitute blocks, the 24 days constitute whole plots with two whole plots per block, and the three runs per day constitute split plots.

The experiment also involved paths and run order—additional nuisance sources of variation the experimenter chose to control in the experiment and adjust for in the analysis. Each subject navigated each of three paths frontward and backward for six paths in total, each path being roughly square in shape with north, south, east, and west legs.

Twelve subjects was enough to design the experiment so treatment effects were not confounded with subject, path, or run order, though display type was confounded with days (i.e. whole plots). The design construction is discussed in Sect. 19.7.4. The resulting root mean square error (*RMSE*) data are in Table 19.11. Prior to analysis of the data, it was discovered that one of the subjects traversed one of the paths in the reverse direction. This was deemed to adversely affect some response variables, making the original data inappropriate to use in the evaluation. As such, one of the 72 planned observations is listed as missing.

The model used by the experimenter for the data analysis is as follows.

$$\begin{aligned} Y_{dhijqr} = \mu &+ \theta_h + \alpha_i + (\phi_2)_{q_2} + (\rho_2)_{r_2} + \epsilon^W_{d(h)} \\ &+ \beta_j + (\alpha\beta)_{ij} + (\phi_3)_{q_3} + (\phi_2\phi_3)_{q_2 q_3} \\ &+ (\rho_3)_{r_3} + (\rho_2\rho_3)_{r_2 r_3} + \epsilon^S_{ijq_3r_3(dhq_2r_2)} , \end{aligned} \tag{19.7.18}$$

$$\theta_h \sim N(0, \sigma^2_\theta) , \quad \epsilon^W_{d(h)} \sim N(0, \sigma^2_W) , \quad \epsilon^S_{ijq_3r_3(dhq_2r_2)} \sim N(0, \sigma^2_S) ,$$

$$\theta_h \text{'s, } \epsilon^W_{d(h)} \text{'s and } \epsilon^S_{ijq_3r_3(dhq_2r_2)} \text{'s are all mutually independent,}$$

$$d = 1, 2; \quad h = 1, \ldots, 12; \quad i = 1, 2; \quad j = 1, 2, 3;$$

$$q = 3(q_3 - 1) + 2(q_2 - 1) + 1, \text{ for } q_2 = 1, 2, \quad q_3 = 1, 2, 3;$$

$$r = 3(r_3 - 1) + 2(r_2 - 1) + 1, \text{ for } r_2 = 1, 2, \quad r_3 = 1, 2, 3.$$

In the model, θ_h denotes subject (i.e. block) effects, α_i and β_j denote the display type and visual presentation format effects, respectively, $\epsilon^W_{d(h)}$ denotes day (i.e. whole plot) effects, and $\epsilon^S_{ijq_3r_3(dhq_2r_2)}$ denotes the split-plot effects.

Table 19.11 Mobile computing field study: root mean square error (*RMSE*) data for each subject (Subj), day, run order, and path-treatment combination (*PAB*), with one observation missing

Subj	Day 1						Day 2					
	Run 1		Run 2		Run 3		Run 4		Run 5		Run 6	
	PAB	*RMSE*	*PAB*	*RMSE*	*PAB*	*RMSE*	*PAB*	*RMSE*	*PAB*	*RMSE*	*PAB*	*RMSE*
1	111	49.321	212	24.386	313	37.680	421	34.291	522	32.053	623	43.121
2	213	45.469	311	24.224	112	29.063	523	32.020	621	33.478	422	36.942
3	312	20.378	113	47.680	211	39.962	622	19.876	423	37.471	521	27.410
4	121	50.088	222	19.010	323	28.749	411	41.306	512	21.924	613	41.469
5	223	27.452	321	46.204	122	42.188	513	28.919	611	24.599	412	28.361
6	322	73.913	123	59.856	221	73.709	612	32.315	413	80.809	511	36.911
7	421	45.506	522	40.380	623	37.135	111	54.271	212	26.685	313	26.076
8	523	29.368	621	34.627	422	27.021	213	33.799	311	43.408	112	40.724
9	622	19.946	423	36.251	521	29.726	312	26.041	113	53.162	211	27.546
10	411	77.787	512	66.956	613	51.114	121	67.176	222	38.774	323	48.602
11	513	26.584	611	41.038	412	47.183	223	37.914	321	30.026	122	44.577
12	612	31.994	413	35.671	511	30.146	322	.	123	58.481	221	62.659

Source Wesler (2001). Copyright © 2001 Mary Mc. Wesler. Research was performed under U.S. Army Research Laboratory, Federated Laboratory Research Consortium (DAAL01-96-0003) directed by Mr. Bernie Corona. Reprinted with permission

Furthermore, construction of the design involved various pseudofactors. For example, the factor path (P) has been represented by pseudofactors: path direction P_2 at two levels, corresponding to using a path frontwards ($P_2 = 0$, for paths 1–3) or backwards ($P_2 = 1$, for paths 4–6); and P_3 at three levels, corresponding to the three paths used. The corresponding effects of P_2, P_3 and P_2P_3 are modeled by the parameters $(\phi_2)_{q_2}$, $(\phi_3)_{q_3}$, and $(\phi_2\phi_3)_{q_2q_3}$, respectively. Likewise, the factor run order (O) has been represented by pseudofactors: O_2 at two levels, corresponding to runs on day 1 ($O_2 = 0$, for runs 1–3) or day 2 ($O_2 = 1$, for runs 4–6); and O_3 at three levels, corresponding to the three runs on a given day. The effects corresponding to O_2, O_3 and O_2O_3 are modeled by the parameters $(\rho_2)_{r_2}$, $(\rho_3)_{r_3}$, and $(\rho_2\rho_3)_{r_2r_3}$, respectively.

Since each subject makes the first three runs on one day and the other three on a second day, one degree of freedom for run order corresponding to O_2 is confounded with days (i.e. whole plots). Also, since paths 1–3 are always used on one day for each subject and paths 4–6 on the other day, the one degree of freedom for path direction corresponding to P_2 is also confounded with days.

Consequently, the between-whole-plot comparisons consist of: main effects of display type A, subject main effects, the path direction effect P_2, and the run order effect O_2. Within-whole-plot comparisons include: main effects of display format B, AB interaction effects, the path effects P_3 and P_2P_3, and the run order effects O_3 and O_2O_3.

19.7.1 Analysis of Variance

Recall, one observation is missing since one of the subjects traversed one of the paths in the wrong direction from what was intended. Consequently, the resulting 71-run experiment deviates modestly from the balanced 72-run experiment that was planned, complicating the data analysis, since the standard formulas for a balanced design are no longer applicable. Direct analysis of the 71 observations collected is an option, and it will be illustrated using the SAS and R software packages in Sects. 19.8.4

Table 19.12 Analysis of variance for the mobile computing field study

Source of variation	Degrees of freedom	Sum of square	Mean square	Ratio	p-value
Subject	11	6230.92	566.45	2.92	0.0594
Order (O_2)	1	30.48	30.48	0.16	0.7012
Path (P_2)	1	379.34	379.34	1.95	0.1957
A (display)	1	47.95	47.95	0.25	0.6311
Whole-plot error (day)	9	1747.52	194.17	2.17	0.0483
Whole-plot total	23	8436.21			
Order (O_3, O_2O_3)	4	56.56	14.14	0.16	0.9581
Path (P_3, P_2P_3)	4	1941.84	485.46	5.42	0.0016
B (VPF)	2	806.44	403.22	4.51	0.0179
AB	2	166.48	83.24	0.93	0.4038
Split-plot error	36	3222.00	89.50		
Total	71	14629.54			

and 19.9.4, respectively. Another option, illustrated here, is the following traditional remedy involving the estimation of missing values.

Given the missing observation, a standard approach is to fit the intended model using the 71 observations collected, estimate the missing value, then analyze the data as a 72-run experiment including the estimated value as if not missing, except making the following accommodation. With one observation missing, there is one fewer degrees of freedom for the analysis. In comparing the results when using software to fit model (19.7.18) with the 71 and 72 observations, respectively, the analysis of variance for the 71-run fit (not shown here) yields one fewer degrees of freedom for split-plot error—namely, 35 rather than 36. So, the same effects are estimable in each case, but the lost observation results in one fewer split-plot error degrees of freedom. Consequently, in analyzing the 72-run data set including the estimated missing value, the number of split-plot error degrees of freedom should be reduced by one. That said, the reduction from 36 to 35 split-plot error degrees of freedom has negligible impact on the analysis, as corresponding critical values are nearly the same with this many error degrees of freedom, so such an adjustment is not always made in the analyses presented here.

Recall, the experimental plan was a split-plot design, with a single response on each independent variable for each of the six treatment combinations per subject. The subjects served as blocks in the design, days as whole plots, and the 72 runs as split plots. The analysis of variance is given in Table 19.12 and is analogous to the one in Table 19.4 (p. 710) for the oats split-plot experiment.

The experimenter planned to test each treatment effect using a 5% significance level. As one can see from the p-values in Table 19.12, the interaction between display type and visual presentation format is not significant ($p = 0.4038$) using $\alpha = 0.05$, nor is the main effect of display type ($p = 0.6311$), but the main effect of visual presentation format is significant ($p = 0.0179$) at the 5% level.

As one can also see from the p-values, there are significant effects of path (i.e. P_3 and P_2P_3 collectively) but not of P_2, so the three paths used had significant differences either themselves or interacting with path direction, but the path direction was not significant. Also, the subject and day effects were modestly significant with $p = 0.0594$ and $p = 0.0483$, respectively, but order effects were not significant.

19.7.2 Multiple Comparisons

For each treatment effect found to be significant in the analysis of variance, the experimenter followed up with multiple comparisons using simultaneous 95% confidence intervals. Since only the main effect of visual presentation format was significant, Tukey's method was used to compare the three formats.

The estimated pairwise comparisons are:

$$\begin{aligned}\hat{\beta}_1^* - \hat{\beta}_2^* &= \overline{y}_{...1..} - \overline{y}_{...2..} = 7.792\,,\\ \hat{\beta}_3^* - \hat{\beta}_2^* &= \overline{y}_{...3..} - \overline{y}_{...2..} = 6.102\,,\\ \hat{\beta}_1^* - \hat{\beta}_3^* &= \overline{y}_{...1..} - \overline{y}_{...3..} = 1.690\,.\end{aligned}$$

Since the corresponding F-test uses the split-plot mean squared error as the denominator and the 72-run design is balanced, the estimated standard error of each pairwise comparison also utilizes the split-plot mean squared error. Also, since each treatment sample mean is the average of 24 observations, the variance estimate of each pairwise contrast is $2msE_S/24 = 89.5/12 \approx 7.4583$. So, the minimum significant difference for each comparison is $(q_{3,35,0.05}/\sqrt{2})\sqrt{7.4583} \approx 6.69$, for $q_{3,35,0.05} = 3.465$, with the split-plot error degrees of freedom having been reduced by one to 35 because of the missing value. Hence, visual presentation format 1 (second person egocentric) has a significantly larger mean *RMSE* from the intended path than does visual presentation format 2 (birds' eye view egocentric), indicating that format 2 is preferable. The other two pairwise comparisons are not significantly different, though format 3 also does substantially worse than format 2. The reader may verify for example that the minimum significant difference would only be 5.89 for simultaneous 90% confidence intervals, in which case visual presentation format 2 would be significantly better than both of the other formats.

19.7.3 Analysis as a Split-Split-Plot Design

The original experimental plan was a split-plot design, with a single response on each independent variable for each of the six treatment combinations per subject. The subjects served as blocks in the design, days as whole plots, and the 72 runs as split plots. The split-plot analysis was provided in Sects. 19.7.1 and 19.7.2.

It subsequently became of interest to study whether the effectiveness of the display types (factor A) or visual presentation formats (factor B) depended on the direction being navigated. One could consider utilizing the data for this purpose because, as noted previously, each path was roughly square in shape with north, south, east, and west legs, (coded 1–4, respectively). Furthermore, data was available on the independent variables for each of the four legs of each run. It was ultimately decided to analyze the data as a split-split-plot design, including leg direction as a factor, treating each leg of a run as a split-split-plot. In that regard, one should note that leg order may also have an effect and may be confounded with leg direction. Nonetheless, it was decided to include leg direction rather than leg order as a factor, since previous studies indicate that leg direction is a stronger independent variable, noting for example that participants in another study had difficulty cognitively transposing information from a fixed north-up display while traversing a south-bound leg of a test course. For this reason, visual display format by leg direction interaction effects were also included in the model.

Adding leg direction factor D with effects δ_t ($t = 1, \ldots, 4$), visual presentation format by leg direction interaction effects $(\beta\delta)_{jt}$, and split-split-plot errors $\epsilon^{SS}_{t(dhijqr)}$ to the split-plot design model (19.7.18), the split-split-plot design model is as follows.

$$\begin{aligned} Y_{dhijqrt} = \mu + \theta_h + \alpha_i + (\phi_2)_{q_2} + (\rho_2)_{r_2} + \epsilon^W_{d(h)} \\ + \beta_j + (\alpha\beta)_{ij} + (\phi_3)_{q_3} + (\phi_2\phi_3)_{q_2q_3} \\ + (\rho_3)_{r_3} + (\rho_2\rho_3)_{r_2r_3} + \epsilon^S_{ijq_3r_3(dhq_2r_2)} \\ + \delta_t + (\beta\delta)_{jt} + \epsilon^{SS}_{t(dhijqr)}\,, \end{aligned} \tag{19.7.19}$$

$$\theta_h \sim N(0, \sigma^2_\theta)\,, \quad \epsilon^W_{d(h)} \sim N(0, \sigma^2_W)\,,$$

$$\epsilon^S_{ijq_3r_3(dhq_2r_2)} \sim N(0, \sigma^2_S)\,, \quad \epsilon^{SS}_{t(dhijqr)} \sim N(0, \sigma^2_{SS})$$

$$\theta_h\text{'s},\ \epsilon^W_{d(h)}\text{'s},\ \epsilon^S_{ijq_3r_3(dhq_2r_2)}\text{'s},\ \epsilon^{SS}_{t(dhijqr)}\text{'s are all mutually independent,}$$

$$d = 1, 2;\quad h = 1, \ldots, 12;\quad i = 1, 2;\quad j = 1, 2, 3;$$

$$q = 3(q_3 - 1) + 2(q_2 - 1) + 1,\ \text{for } q_2 = 1, 2,\quad q_3 = 1, 2, 3;$$

$$r = 3(r_3 - 1) + 2(r_2 - 1) + 1,\ \text{for } r_2 = 1, 2,\quad r_3 = 1, 2, 3;$$

$$t = 1, 2, 3, 4.$$

The data for the split-split-plot design consist of the *RMSE* values for each of the four legs of each run. The data, shown in Table 19.13, now include four missing observations, since the one missing run of the split-plot design includes four legs. Estimating these four missing values, the analysis of variance is provided in Table 19.14. As the reader may verify, if the model is fit without the estimated missing values, one loses one degree of freedom for split-plot error and three for split-split-plot error, but adjusting the corresponding degrees of freedom from 36 to 35 and from 207 to 204, respectively, would have minimal impact on any analysis. The first two sections of Table 19.14 duplicate information given in Table 19.12 for the split-plot analysis—the degrees of freedom, ratio, and p-value is the same for each source so conclusions are unchanged, though the sum of squares and mean squares are different since the data are observations on legs of runs, not runs. The visual presentation format by leg direction interaction is not significant, but leg direction is significant with $p = 0.0117 < 0.05$. Hence, the experimenter compared the leg direction effects using simultaneous 95% confidence intervals obtained via Tukey's method.

The estimated pairwise comparisons are:

$$\begin{aligned} \hat\delta^*_2 - \hat\delta^*_1 &= \overline{y}_{......2} - \overline{y}_{......1} = 2.8645\,, & \hat\delta^*_4 - \hat\delta^*_1 &= \overline{y}_{......4} - \overline{y}_{......1} = 1.5666\,, \\ \hat\delta^*_2 - \hat\delta^*_3 &= \overline{y}_{......2} - \overline{y}_{......3} = 2.0217\,, & \hat\delta^*_4 - \hat\delta^*_3 &= \overline{y}_{......4} - \overline{y}_{......3} = 0.7239\,, \\ \hat\delta^*_2 - \hat\delta^*_4 &= \overline{y}_{......2} - \overline{y}_{......4} = 1.2979\,, & \hat\delta^*_3 - \hat\delta^*_1 &= \overline{y}_{......3} - \overline{y}_{......1} = 0.8427\,. \end{aligned}$$

Since the corresponding F-test uses the (split-split-plot) mean square error as the denominator and the 288-leg design is balanced, the estimated standard error of each pairwise comparison also utilizes the mean squared error. Also, since each treatment sample mean is the average of 72 of the 288 observations, the variance estimate of each pairwise contrast is $2(msE)/72 = 2(28.19)/72 \approx 0.7831$. So, the minimum significant difference for each comparison is $(q_{4,204,0.05}/\sqrt{2})\sqrt{0.7831} \approx 2.27$, for $q_{3,204,0.05} \approx q_{3,\infty,0.05} = 3.63$, with the split-plot error degrees of freedom having been reduced by three to 204 because of the missing values. Hence, mean *RMSE* is significantly greater for leg direction 2 (south) than for leg direction 1 (north). This provides some credence to the observation in another study that participants had difficulty cognitively transposing information from a fixed north-up display while traversing a south-bound leg of a test course.

Table 19.13 Mobile computing field study split-split-plot design: root mean square error (*RMSE*) data for each subject (Subj), leg, day, run order, and path-treatment combination (*PAB*), with four observations missing

Subj	Leg	Day 1						Day 2					
		Run 1		Run 2		Run 3		Run 4		Run 5		Run 6	
		PAB	*RMSE*	*PAB*	*RMSE*	*PAB*	*RMSE*	*PAB*	*RMSE*	*PAB*	*RMSE*	*PAB*	*RMSE*
1	1	111	6.313	212	3.653	313	6.762	421	6.223	522	6.888	623	7.506
	2	111	25.032	212	4.259	313	4.269	421	5.621	522	9.165	623	12.035
	3	111	7.600	212	7.011	313	11.870	421	13.733	522	8.129	623	17.760
	4	111	10.376	212	9.464	313	14.780	421	8.713	522	7.871	623	5.820
2	1	213	8.574	311	6.332	112	5.313	523	5.832	621	3.601	422	5.813
	2	213	7.244	311	5.257	112	12.927	523	7.441	621	6.756	422	7.498
	3	213	19.585	311	10.759	112	3.666	523	8.659	621	4.837	422	10.979
	4	213	10.066	311	1.877	112	7.157	523	10.088	621	18.283	422	12.651
3	1	312	8.700	113	10.037	211	11.989	622	3.607	423	10.548	521	9.446
	2	312	5.436	113	15.766	211	6.458	622	5.892	423	7.323	521	4.318
	3	312	4.101	113	9.985	211	11.384	622	3.577	423	6.705	521	5.304
	4	312	2.142	113	11.892	211	10.130	622	6.800	423	12.895	521	8.343
4	1	121	3.700	222	4.378	323	4.348	411	18.534	512	3.393	613	4.472
	2	121	26.995	222	4.065	323	10.347	411	4.280	512	4.250	613	13.711
	3	121	6.061	222	5.764	323	2.645	411	9.041	512	7.794	613	9.561
	4	121	13.332	222	4.804	323	11.409	411	9.450	512	6.487	613	13.724
5	1	223	7.924	321	8.801	122	6.337	513	3.435	611	5.669	412	7.954
	2	223	2.563	321	16.128	122	21.415	513	4.446	611	8.044	412	6.689
	3	223	6.068	321	4.969	122	4.171	513	4.716	611	5.759	412	5.960
	4	223	10.897	321	16.305	122	10.265	513	16.321	611	5.127	412	7.758
6	1	322	22.194	123	11.316	221	11.540	612	3.791	413	15.781	511	5.236
	2	322	13.209	123	24.914	221	18.154	612	9.139	413	20.223	511	3.422
	3	322	13.197	123	9.051	221	34.058	612	7.538	413	28.214	511	4.137
	4	322	25.314	123	14.575	221	9.957	612	11.847	413	16.591	511	24.117
7	1	421	15.819	522	17.896	623	6.257	111	2.320	212	6.519	313	4.058
	2	421	6.846	522	7.257	623	8.781	111	26.503	212	6.613	313	5.936
	3	421	14.657	522	7.117	623	13.149	111	21.167	212	8.820	313	11.339
	4	421	8.184	522	8.110	623	8.948	111	4.280	212	4.733	313	4.743
8	1	523	11.039	621	3.718	422	7.356	213	3.837	311	5.856	112	6.410
	2	523	7.973	621	18.952	422	9.132	213	3.759	311	16.997	112	26.575
	3	523	6.242	621	5.573	422	4.332	213	9.298	311	8.671	112	2.342
	4	523	4.115	621	6.384	422	6.200	213	16.905	311	11.883	112	5.397
9	1	622	4.371	423	8.725	521	7.110	312	4.352	113	7.832	211	7.286
	2	622	4.846	423	11.287	521	6.264	312	6.641	113	31.707	211	6.662
	3	622	6.666	423	10.870	521	9.541	312	6.450	113	1.774	211	9.589
	4	622	4.063	423	5.369	521	6.810	312	8.597	113	11.848	211	4.008
10	1	411	22.287	512	22.921	613	15.016	121	26.405	222	7.143	323	10.467
	2	411	19.107	512	12.482	613	5.827	121	13.756	222	11.160	323	21.939
	3	411	15.018	512	15.430	613	18.707	121	11.025	222	12.443	323	5.404
	4	411	21.375	512	16.123	613	11.564	121	15.990	222	8.029	323	10.791
11	1	513	5.883	611	6.531	412	10.834	223	9.284	321	6.103	122	3.570
	2	513	7.288	611	15.533	412	8.584	223	10.828	321	12.479	122	24.502
	3	513	3.260	611	8.494	412	18.250	223	9.102	321	6.008	122	5.115
	4	513	10.154	611	10.479	412	9.515	223	8.701	321	5.436	122	11.389
12	1	612	5.826	413	12.041	511	8.747	322	.	123	9.794	221	16.560
	2	612	10.782	413	8.086	511	8.952	322	.	123	23.941	221	17.369
	3	612	7.168	413	5.918	511	1.080	322	.	123	20.566	221	11.995
	4	612	8.218	413	9.626	511	11.367	322	.	123	4.180	221	16.735

Source Wesler (2001). Copyright © 2001 Mary Mc. Wesler. Research was performed under U.S. Army Research Laboratory, Federated Laboratory Research Consortium (DAAL01-96-0003) directed by Mr. Bernie Corona. Reprinted with permission

Table 19.14 Analysis of variance for the mobile computing field study

Source of variation	Degrees offreedom	Sum ofsquare	Meansquare	Ratio	p-value
Subject	11	1557.78	141.62	2.92	0.0594
O2	1	7.62	7.62	0.16	0.7011
P2	1	94.84	94.84	1.95	0.1957
Display	1	11.99	11.99	0.25	0.6311
Whole-plot error (day)	9	436.86	48.54	2.17	0.0483
Whole-plot total	23	2109.09			
Order	4	14.15	3.54	0.16	0.9581
Path	4	485.43	121.36	5.42	0.0016
VPF	2	201.60	100.80	4.51	0.0179
VPF*Display	2	41.62	20.81	0.93	0.4038
Split-plot error	36	805.47	22.37	0.79	0.7934
Split-plot total	71	3657.36			
LegDir	3	317.98	105.99	3.76	0.0117
VPF*LegDir	6	34.85	5.81	0.21	0.9746
Split-split-plot error	207	5835.85	28.19		
Total	287	9846.03			

19.7.4 Design Construction

In this section we discuss construction of the split-plot design used in the mobile computing field study. The design used (Table 19.11, p. 719) is a fractional factorial design with some day effects confounded with blocks (subjects). The design construction illustrated here utilizes the techniques of Chaps. 13 and 14, including the use of pseudofactors introduced in Sect. 14.3. The first step in the design construction is to determine an appropriate fraction of the non-blocking factors, then some effects (and their aliases)—some day effects in this case—are chosen to be confounded with blocks. Before constructing the fraction, here are some preliminary considerations.

Recall, the basic experiment was 2×3 in the treatment factors A (display type) and B (visual presentation format), the purpose being accurate comparison of the treatment effects. Multiple subjects were to be recruited for the study based on appropriate protocols. It was pragmatically reasonable to have each subject make three runs per day and to involve each subject for two days. This would allow each subject to be assigned each of the six treatment combinations once, providing a balanced design with treatment comparisons as within-subject comparisons, good for precision. Since day-to-day variation was anticipated to be relatively small and simplicity of the experimental design is desirable, it was decided to confound main effects of A with days. Viewing subjects as blocks, days as whole plots, and runs as split plots, levels of A would be assigned to whole plots and levels of B to split plots.

There remains the question of how many days $2h$ to include in the experiment, (or equivalently, how many subjects, h). Consider this question from the viewpoint of design construction rather than as a sample size question, to see how small an appropriate design can be given the circumstances. For a 2×3 experiment, standard confounding techniques can be applied if the number of subjects is a power of two times a power of three, $h = 2^m 3^n$ say, so represent the subjects factor S by pseudofactors S_{2i} ($i = 1, \ldots, m$) each at two levels and pseudofactors S_{3j} ($j = 1, \ldots, n$) each at three levels, where the number of each remains to be determined. The other nuisance sources of variation are days D, with two days for each subject, and run order O and path P each at six levels. Let O_2, O_3, P_2 and P_3 be pseudofactors for run order and path, with the subscript indicating the number of levels. Likewise, denote pseudofactors for days by D_{2i} $(i = 0, 1, \ldots, m)$ each at two levels and D_{3j} $(j = 1, 2, \ldots, n)$ each at three levels. So, factors at two levels are A_2, the D_{2i}, O_2, P_2, and the S_{2i}, each with levels 0 and 1, say. Similarly, factors at three levels are B_3, the D_{3j}, O_3, P_3, and the S_{3j}, each with levels 0, 1 and 2, say.

In the resulting design, take the treatment factor levels to be: $A = A_2 + 1$ and $B = B_3 + 1$. Furthermore, let order $O = 3 \times O_2 + O_3 + 1$, so the levels of O_2 distinguish a subject's three runs on their first day (i.e. $O = 1, 2, 3$) from the subject's three runs on their second day (i.e. $O = 4, 5, 6$), and the levels of O_3 distinguish the three runs on each day. Similarly, let path $P = 3 \times P_2 + P_3 + 1$, with the understanding that the levels of P_2 distinguish whether a path is run frontwards ($P_2 = 0$, say) or backwards ($P_2 = 1$, say), and the levels of P_3 correspond to the three physical paths. Similar relationships between factors and pseudofactors for days and subjects will be provided later.

The design is a fractional factorial block design. Following the approach introduced in Sect. 15.5, the first step is to determine an appropriate fractional factorial design in the non-blocking factors, then determine what effects to confound with blocks. That said, subjects will not participate on the same day, so comparisons between subjects will necessarily be comparisons between days, so the corresponding between-days comparisons will be confounded with subjects. Let the levels of D_{20} represent the two days each subject participates, so D_{20} will not be confounded with subjects, but anticipate that the remaining days pseudofactors will be confounded with subjects.

The experiment includes one observation for each combination of day and O_3. Thus, one can view the day pseudofactors D_{2i} $(i = 0, \ldots, m)$ and D_{3j} $(j = 1, \ldots, n)$ and the order pseudofactor O_3 as *free variables* in constructing the fractional factorial design, with all other pseudofactors defined in relation to them using defining relations.

Consider first the remaining 2-level pseudofactors—A_2, O_2, and P_2—in relation to free variables D_{2i}. Run order is necessarily related to days, since each subject makes their first three runs on their first day and their last three runs on their second day. This corresponds to the constraint $O_2 = D_{20}$, or equivalently $I_2 = D_{20}O_2$.

To avoid aliasing the effects of A_2 with either order or path, the experimenter chose to let $A_2 = D_{21} + D_{22} + O_2 \pmod 2$ and $P_2 = A_2 + D_{22} \pmod 2$, so the defining relation for the 2-level pseudofactors is

$$
\begin{aligned}
I_2 &= \underline{D_{20}O_2} &&= \underline{A_2D_{21}D_{22}O_2} &&= A_2D_{20}D_{21}D_{22} \\
= \underline{A_2D_{22}P_2} &= A_2D_{20}D_{22}O_2P_2 &&= D_{21}O_2P_2 &&= D_{20}D_{21}P_2\,,
\end{aligned}
$$

with defining relation generators underlined, the other terms being generalized interactions.

With these choices used by the experimenter, using only three 2-level day pseudofactors D_{20}, D_{21} and D_{22}, the fraction of the 2-level pseudofactors is a 1/8 fraction with eight runs involving eight days. Hence, as will be seen, the number of days in the full experiment is a multiple of eight. Technically, the fraction for the 2-level pseudofactors is not of resolution III, because day effects D_{2i} are aliased with the effects of three factors: A, since $A_2 = D_{20}D_{21}D_{22}$; order, since $O_2 = D_{20}$; and path, since $P_2 = D_{20}D_{21}$. However, the day effects are random, modeling variability but with mean zero. So, the effects of A, O_2 and P_2, being "aliased" with the random effects of days, are comparisons between days (i.e. between whole plots), so these effects are estimable, but estimation and testing of these effects will involve the whole-plot mean squared error.

Now consider the remaining pseudofactors at three levels—namely, B_3 and P_3—to be defined in relation to the free variables O_3 and the D_{3j}. Four 3-level factors are necessary and sufficient to obtain a resolution III fraction, so one 3-level days pseudofactor D_{31} will suffice. The experimenter chose to let $B_3 = 2D_{31} + O_3 \pmod 3$ and $P_3 = D_{31} + O_3 \pmod 3$. Thus, the defining relation for the resulting 1/9 fraction for the 3-level pseudofactors is

$$\begin{aligned} I_3 \quad &= \underline{B_3 D_{31} O_3^2} = B_3^2 D_{31}^2 O_3 \\ = \underline{D_{31} O_3 P_3^2} &= B_3 D_{31}^2 P_3^2 = B_3^2 O_3^2 P_3^2 \\ = D_{31}^2 O_3^2 P_3 &= B_3 O_3 P_3 \quad = B_3^2 D_{31} P_3 , \end{aligned}$$

involving the contrasts $(B; B^2)$, $(B_3 D_{31} O_3^2; B_3^2 D_{31}^2 O_3)$, $(D_{31} O_3 P_3^2; D_{31}^2 O_3^2 P_3)$, $(B_3 D_{31}^2 P_3^2; B_3^2 D_{31} P_3)$, and $(B_3 O_3 P_3;\ B_3^2 O_3^2 P_3^2)$.

Since the effects of B_3, O_3 and P_3 are not aliased with other main effects, and in particular not with day effects, they are estimable within-day comparisons under a main-effects model, free of day-to-day (i.e. whole-plot-to-whole-plot) variation, as desired.

With these choices used by the experimenter, using only one 3-level day pseudofactor D_{31}, the fraction of the 3-level pseudofactors is a 1/9 resolution III fraction involving nine runs, including three days and three runs per day (corresponding to the nine combinations of the free variables D_{31} and O_3).

The 2-level and 3-level fractions are combined by pairing each of the eight 2-level pseudofactor combinations in the 2-level fraction with each of the nine 3-level pseudofactor combinations in the 3-level fraction, providing a *composite fractional factorial design* with 72 pseudofactor combinations or runs, 24 days, six run orders, and six paths. The corresponding *composite defining relation* includes 72 terms, including each product of one of the eight terms from the defining relation of the 2-level fraction with one of the nine terms from the defining relation of the 3-level fraction. In short, letting $I_6 = I_2 \times I_3$, and treating I_2 and I_3 as multiplicative identities, the composite defining relation is as follows.

$$\begin{array}{c|cccc} \times & I_2 & = D_{20} O_2 & = \ldots = & D_{20} D_{21} P_2 \\ \hline I_3 & I_6 & = D_{20} O_2 & = \ldots = & D_{20} D_{21} P_2 \\ = B_3 D_{31} O_3^2 & = B_3 D_{31} O_3^2 & = D_{20} O_2 B_3 D_{31} O_3^2 & = \ldots = & D_{20} D_{21} P_2 B_3 D_{31} O_3^2 \\ = B_3^2 D_{31}^2 O_3 & = B_3^2 D_{31}^2 O_3 & = D_{20} O_2 B_3^2 D_{31}^2 O_3 & = \ldots = & D_{20} D_{21} P_2 B_3^2 D_{31}^2 O_3 \\ = \quad \vdots & = \quad \vdots & = \quad \vdots & = \vdots = & \vdots \\ = B_3^2 D_{31} P_3 & = B_3^2 D_{31} P_3 & = D_{20} O_2 B_3^2 D_{31} P_3 & = \ldots = & D_{20} D_{21} P_2 B_3^2 D_{31} P_3 \end{array}$$

This composite defining relation includes each of the terms from the 2-level and 3-level defining relations, indicating for example that days are still aliased with the effects of A, O_2 and P_2 in the composite fraction, as was the case in the 2-level fraction, so A and one degree of freedom for each of order and path remain between-day comparisons in the composite fraction. However, the remaining terms in the composite defining relation are the generalized interactions that include terms other than I_2 and I_3 from each of the 2-level and 3-level defining relations, so they tend to be of higher order. Consequently, one may observe for example that AB effects are *not* aliased with day effects, so they are within-day (i.e. within-whole-plot) comparisons. The same is true of the order and path effects except O_2 and P_2, so order and path each have four degrees of freedom that are within-day comparisons, in addition to each having one degree of freedom that is a between-day (i.e. between-whole-plot) comparison.

Finally, the block design is obtained by confounding the day effects D_{21}, D_{22} and D_{31} with subjects, which also confounds the corresponding generalized interactions $D_{21} D_{22}$, $D_{21} D_{31}$, $D_{22} D_{31}$ and $D_{21} D_{22} D_{31}$, confounding 11 degrees of freedom for days with subjects. The *composite design* involves 12 subjects and 24 days, with three runs per day, so 72 runs in total. The resulting design is shown in Table 19.11, with factor levels in the table obtain as follows: $D = D_{20} + 1$ and $\text{Subj} = 6S_{21} + 3S_{22} + S_{31} + 1$.

In summary, 12 subjects is enough to design the experiment so treatment effects are not aliased with subject, path, or run order, though effects of A (display type) are aliased with days (whole plots), as is one degree of freedom each for order and path. One might say that the design is of resolution III in the fixed effects but of resolution II when random effects are considered.

The experimenter determined that this was a large enough experiment, with each pairwise comparison of visual presentation formats being estimable with variance $\sigma_S^2/12$, and with the pairwise comparison of display types being estimable with variance $(3\sigma_W^2 + \sigma_S^2)/18$.

If a larger design was needed, one could be obtain by inclusion of additional free variables D_{2i} $(i > 2)$ or D_{3j} $(j > 1)$ and corresponding additional subject pseudofactors, coupled with an appropriate choice of defining relations to determine all non-free variables in terms of the free variables. For example, to have two replicates of the design obtained above, simply include the additional free variable D_{23} to double the number of days, include the pseudofactor S_{23} to double the number of subjects, impose the same constraints as above, and confound D_{23} with subjects. Alternatively, one could seek a better fraction—namely, either one of higher resolution or, short of that, a fraction with less aliasing together of lower order effects of interest—by inclusion of additional free variables D_{2i} or D_{3j} and corresponding subject pseudofactors, coupled with some better choice of defining relations that presumably would involve these additional free variables.

The approach used above to obtain a blocked fractional factorial experiment involved two steps: first obtaining a fraction of the non-blocking factors, then confounding some effects with blocks. Another equivalent approach is simply to obtain an appropriate fraction of all of the factor combinations, including the blocking factors as non-free variables. For example, one could obtain the same blocked fractional factorial design in one step by imposing the same defining relations as above plus the relations $S_{21} = D_{21}$, $S_{22} = D_{22}$, and $S_{31} = D_{31}$.

19.8 Using SAS Software

In this section, we illustrate use of the SAS software for analyzing split-plot designs. The analysis of variance approach can be used for balanced data using `PROC GLM` or `PROC MIXED`, as illustrated for analysis of the oats experiment in Sect. 19.8.1. In Sect. 19.8.2, the UAV experiment is briefly revisited to illustrate the analysis of simple contrasts using the `SLICE` statement in `PROC MIXED`. In Sect. 19.8.3, we introduce a restricted maximum likelihood approach to analysis of split plot designs using `PROC MIXED`, comparing it to the analysis of variance approach and `PROC GLM` in the context of the UAV switch experiment—a new example involving a negative variance component estimate. In Sect. 19.8.4, using a subset of the oats data yielding a balanced incomplete block design for the whole-plots factor, `PROC MIXED` is used to illustrate the recovery of inter-block information. Finally, in Sect. 19.8.5, analysis of unbalanced data is illustrated using `PROC MIXED` and the restricted maximum likelihood approach for the mobile computing field study, excluding the missing observation from analysis of the split plot design.

19.8.1 The Analysis of Variance Approach

The *analysis of variance approach* to the analysis of mixed models involves: (i) fitting the model by ordinary least squares, treating the random effects as fixed; and (ii) using the expected mean squares to determine unbiased estimates of the variance components. In analyzing a balanced split-plot design by the analysis of variance approach, care needs to be taken with `PROC GLM` or `PROC MIXED` in order to obtain the two separate parts of the analysis of variance table, corresponding to

Table 19.15 SAS program illustrating type I and III analyses—the oats split-plot experiment

```
DATA OAT;
  INPUT BLOCK WP A B  Y;
  LINES;
  1   1  2 3 156
  1   1  2 2 118
  1   1  2 1 140
  1   1  2 0 105
  1   2  0 0 111
  :   :  : :  :
  6   3  2 3 121
;
*** Analysis of variance: type I;
PROC GLM;
  CLASS BLOCK A B WP;
  MODEL Y = BLOCK A WP(BLOCK)  B A*B / E1;
  RANDOM BLOCK WP(BLOCK) / TEST;
  MEANS A / DUNNETT('0') CLDIFF ALPHA = 0.01;
  * The following statement fails to generate output due to
    nonestimability;
  LSMEANS A / ADJUST = DUNNETT ALPHA = 0.01 CL PDIFF = CONTROL('0')
              E = WP(BLOCK);
  LSMEANS B / ADJUST = DUNNETT ALPHA = 0.01 CL PDIFF = CONTROL('0');
PROC MIXED METHOD = TYPE1;
  CLASS BLOCK A B WP;
  MODEL Y = A B A*B / DDFM = SAT;
  RANDOM BLOCK WP(BLOCK);
  LSMEANS A B / PDIFF = CONTROL('0', '0')
                CL DIFF ADJUST = DUNNETT ALPHA = 0.01;
*** Analysis of variance: type III;
PROC GLM;
  CLASS BLOCK A B;
  MODEL  Y = BLOCK A BLOCK*A  B A*B;
  RANDOM BLOCK A*BLOCK / TEST;
  LSMEANS A / ADJUST = DUNNETT ALPHA = 0.01 CL PDIFF = CONTROL('0')
              E = BLOCK*A;
  LSMEANS B / ADJUST = DUNNETT ALPHA = 0.01 CL PDIFF = CONTROL('0');
PROC MIXED METHOD = TYPE3;
  CLASS BLOCK A B;
  MODEL Y = A B A*B / DDFM = SAT;
  RANDOM BLOCK A*BLOCK;
  LSMEANS A B / PDIFF = CONTROL('0', '0')
                CL DIFF ADJUST = DUNNETT ALPHA = 0.01;
```

between-whole-plots and within-whole-plots comparisons. We illustrate two different methods of producing an analysis of variance table similar to the one in Table 19.4 (p. 710) presented for the oats experiment in Sect. 19.3.4. Table 19.15 contains SAS code illustrating the two methods—namely, the type I and type III analyses—using the data of the oats experiment.

Type I Analysis

One approach to analysis of a split plot design is a *type I analysis*, utilizing the type I (sequential) sums of squares and corresponding type I expected mean squares. In the first call of `PROC GLM` in Table 19.15, the option `E1` in the model statement requests a type I analysis. Correspondingly, the sources of variation need to be entered into the model in the same order as in model (19.2.2). The first three terms yield the whole-plot analysis. The whole-plot error term $\epsilon^W_{i(h)}$ is represented by `WP(BLOCK)`—the nested effect of whole plots within blocks—since it accounts for any whole plot comparisons not yet modeled. The remaining terms provide the split-plot analysis. No term is needed to represent the split-plot error term $\epsilon^S_{j(hi)}$, since this plays the role of the usual error variable, and the corresponding error sum of squares is automatically calculated by the SAS software. The type I analysis is necessary because, with inclusion of `WP(BLOCK)` in the model, the type III analysis would yield $SSA = 0$ with zero degrees of freedom for A. Otherwise, the type I and type III sums of squares would be the same. Inclusion of the statement

```
RANDOM BLOCK WP(BLOCK) / TEST;
```

ensures that the correct denominators are used for all of the hypothesis tests.

The output is shown in Fig. 19.1. The type I sums of squares are listed, but the F-statistics and p-values are incorrect for testing the effects of blocks and A, since SAS software uses the split-plot error mean square MSE throughout. The expected mean squares are listed, and the reader can verify that the error estimate for A differs from those of B and AB. The correct hypothesis tests are listed in the bottom half of the output.

One would like to use the `LSMEANS` statement to obtain standard multiple comparisons procedures. For example, Dunnett's procedure for comparing each level of A with control level 0 and each level of B with control level 0 is requested via the `LSMEANS` statements shown in Table 19.15. Unfortunately, this generates no output for A, because `PROC GLM` treats all effects as fixed, including the random effects, so concludes that main effects of A are not estimable. While we have generally avoided use of the `MEANS` statement for multiple comparisons, since it uses means rather than least squares means, the two are equivalent in this case, since both the whole-plot design and the split-plot design are randomized complete block designs. As such, the `MEANS` statement can be used to obtain standard multiple comparison procedures for A, while the `LSMEANS` statement cannot. For example, Dunnett's procedure for comparing each level of A with control level 0 and each level of B with control level 0 are obtained via the `MEANS` statement and the second `LSMEANS` statement, respectively, shown in Table 19.15. For the `MEANS` statement, SAS software will use the split-plot term MSE to estimate standard errors unless told to do otherwise, which would be correct for the B contrasts but not for the A contrasts. To use the whole-plot error mean square for the A contrasts, we include the option `E = WP(BLOCK)`. Partial output is shown in Fig. 19.2. For both factors, note that the SAS software has relabeled the levels starting at 1 rather than 0.

The first call of `PROC MIXED` in Table 19.15 would generate the same correct information, including only the correct F-tests for the analysis of variance, and automatically computing correct standard errors for contrasts. The option `METHOD = TYPE1` calls for a type I analysis. In `PROC MIXED`, all fixed effects are entered into the model before the random effects, but that is not problematic in this case.

Type III Analysis

The second approach to analysis of a split plot design is a *type III analysis*, utilizing the type III sums of squares and corresponding expected mean squares. This approach makes use of the fact that the whole-plot error sum of squares uses the same collective degrees of freedom as any block by whole-plot-treatment interactions deemed negligible. In the oats experiment, with only one whole-plot factor, A, and with the block $\times A$ interaction deemed negligible, the block $\times A$ effect represents whole-plot error in the model. In the second call of `PROC GLM` in Table 19.15, the same sources of variation are entered into the SAS model as in model (19.2.2), but the whole plot error $\epsilon^W_{i(h)}$ is replaced by `BLOCK*A`—the negligible block $\times$ whole-plot-treatment interaction. For the type III analysis, the

The GLM Procedure
Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	26	44017.19444	1692.96902	9.56	<.0001
Error	45	7968.75000	177.08333		
Corrected Total	71	51985.94444			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
BLOCK	5	15875.27778	3175.05556	17.93	<.0001
A	2	1786.36111	893.18056	5.04	0.0106
WP(BLOCK)	10	6013.30556	601.33056	3.40	0.0023
B	3	20020.50000	6673.50000	37.69	<.0001
A*B	6	321.75000	53.62500	0.30	0.9322

Source	Type I Expected Mean Square
BLOCK	Var(Error) + 4 Var(WP(BLOCK)) + 12 Var(BLOCK)
A	Var(Error) + 4 Var(WP(BLOCK)) + Q(A,A*B)
WP(BLOCK)	Var(Error) + 4 Var(WP(BLOCK))
B	Var(Error) + Q(B,A*B)
A*B	Var(Error) + Q(A*B)

Tests of Hypotheses for Mixed Model Analysis of Variance

	Source	DF	Type I SS	Mean Square	F Value	Pr > F
	BLOCK	5	15875	3175.055556	5.28	0.0124
*	A	2	1786.361111	893.180556	1.49	0.2724
	Error: MS(WP(BLOCK))	10	6013.305556	601.330556		

* This test assumes one or more other fixed effects are zero.

	Source	DF	Type I SS	Mean Square	F Value	Pr > F
	WP(BLOCK)	10	6013.305556	601.330556	3.40	0.0023
*	B	3	20021	6673.500000	37.69	<.0001
	A*B	6	321.750000	53.625000	0.30	0.9322
	Error: MS(Error)	45	7968.750000	177.083333		

* This test assumes one or more other fixed effects are zero.

Fig. 19.1 Type I analysis of variance—the oats split-plot experiment

Least Squares Means
Adjustment for Multiple Comparisons: Dunnett

Standard Errors and Probabilities Calculated Using the Type III MS for BLOCK*A as an Error Term

A	Y LSMEAN	99% Confidence Limits	
0	97.625000	81.761076	113.488924
1	104.500000	88.636076	120.363924
2	109.791667	93.927743	125.655591

Least Squares Means for Effect A				
i	j	Difference Between Means	Simultaneous 99% Confidence Limits for LSMean(i)-LSMean(j)	
2	1	6.875000	-18.122835	31.872835
3	1	12.166667	-12.831168	37.164502

B	Y LSMEAN	99% Confidence Limits	
0	79.388889	70.952864	87.824914
1	98.888889	90.452864	107.324914
2	114.222222	105.786197	122.658247
3	123.388889	114.952864	131.824914

Least Squares Means for Effect B				
i	j	Difference Between Means	Simultaneous 99% Confidence Limits for LSMean(i)-LSMean(j)	
2	1	19.500000	5.878974	33.121026
3	1	34.833333	21.212307	48.454360
4	1	44.000000	30.378974	57.621026

Fig. 19.2 Multiple comparisons—the oats split-plot experiment

order of entry of terms into the SAS model does not matter. The SAS output is not shown, but is identical to that in Fig. 19.1 with `WP(BLOCK)` replaced by `BLOCK*A`, and `Type I Expected Mean Square` replaced by `Type III Expected Mean Square`. The second call of `PROC MIXED` in Table 19.15 would generate the same information, but with only the correct F-tests for the analysis of variance, the option `METHOD = TYPE3` calling for a type III analysis.

More generally, suppose the experiment included two whole-plot factors A and B and one split-plot factor C, with the AB-combinations assigned to whole plots within blocks constituting a randomized complete block design, and with the levels of C assigned to split plots within whole plots constituting a randomized complete block design. Then, assuming all block by whole-plot-treatment interactions are negligible, the following SAS statements would provide the analogous type III analysis.

Table 19.16 SAS program illustrating analysis of simple contrasts—UAV experiment

```
DATA UAV;
 INPUT A W B C TIME;
 LINES;
 1 1 1 1 29
 1 1 1 2 28
 1 1 1 3 49
 1 1 1 4 46
 1 1 2 1 36
 : : : : ::
 2 8 2 4 22
;
PROC MIXED METHOD = TYPE3;
 CLASS A W B C;
 MODEL TIME = A | B | C / DDFM = SAT;
 RANDOM W(A);
 LSMEANS A / CL DIFF ALPHA = 0.01;
 * Compares levels of A at each B*C combination;
 SLICE A*B*C / SLICEBY = B*C CL DIFF ALPHA = 0.01;
;
PROC GLM;
 CLASS A W B C;
 MODEL TIME = A | B | C W(A);
 RANDOM W(A) / TEST;
 MEANS A / T CLDIFF E = W(A) ALPHA = 0.01;
 * need PDIFF, else E = mse erroneously;
 LSMEANS A / CL PDIFF E = W(A) ALPHA = 0.01;
```

```
PROC GLM;
  CLASS BLOCK A B C;
  MODEL  Y = BLOCK A B A*B BLOCK*A*B  C A*C B*C A*B*C;
  RANDOM BLOCK BLOCK*A*B / TEST;
  LSMEANS A B / ADJUST = DUNNETT ALPHA = 0.01 CL
                PDIFF = CONTROL('0', '0') E = BLOCK*A*B;
  LSMEANS C / ADJUST = DUNNETT ALPHA = 0.01 CL PDIFF = CONTROL('0');
```

In this case, the whole-plot error sum of squares uses the collective degrees of freedom of the block $\times A$, block $\times B$, and block $\times A \times B$ interactions. These are collected into the single term BLOCK*A*B by including it in the model but leaving out the terms BLOCK*A and BLOCK*B.

19.8.2 Simple Contrasts

In this section, we briefly revisit the UAV experiment of Sect. 19.6 to illustrate the analysis of simple contrasts using the SLICE statement in PROC MIXED. The SAS program in Table 19.16 shows how either PROC MIXED or PROC GLM can be used to generate the analyses provided in Sect. 19.6. For example, both procedures generate the analysis of variance in Table 19.10, as well as the 99% confidence interval for the main effect of cue (A).

PROC MIXED generally provides more options and functionality for multiple comparisons. For example, the SLICE statement illustrated in the SAS program generates the output in Table 19.17, which includes the information on simple contrasts for cue that was provided in Sect. 19.6.2 (p. 717), comparing the effects of the two levels of cue (A) at each of the eight BC combinations. The standard

Table 19.17 SAS output for simple contrasts—UAV experiment

The SAS System
The Mixed Procedure

Simple Differences of A*B*C Least Squares Means

Slice	A	_A	Estimate	Standard Error	DF	t Value	Pr>|t|	Alpha	Lower	Upper
B 1 C 1	1	2	14.3750	1.8643	110.2	7.71	<.0001	0.01	9.4885	19.2615
B 1 C 2	1	2	13.3750	1.8643	110.2	7.17	<.0001	0.01	8.4885	18.2615
B 1 C 3	1	2	31.2500	1.8643	110.2	16.76	<.0001	0.01	26.3635	36.1365
B 1 C 4	1	2	28.2500	1.8643	110.2	15.15	<.0001	0.01	23.3635	33.1365
B 2 C 1	1	2	13.0000	1.8643	110.2	6.97	<.0001	0.01	8.1135	17.8865
B 2 C 2	1	2	15.2500	1.8643	110.2	8.18	<.0001	0.01	10.3635	20.1365
B 2 C 3	1	2	21.2500	1.8643	110.2	11.40	<.0001	0.01	16.3635	26.1365
B 2 C 4	1	2	23.7500	1.8643	110.2	12.74	<.0001	0.01	18.8635	28.6365

errors involve composite variance estimates, and Satterthwaite's approximation is used as a consequence of the denominator degrees of freedom option `DDFM = SAT`. The `SLICE` statement uses much of the same syntax as the `LSMEANS` statement in `PROC MIXED`. Analogous results do not seem to be available from `PROC GLM`.

19.8.3 The Restricted Maximum Likelihood Approach

In this section, we introduce a restricted maximum likelihood (ReML) approach to analysis of split plot designs using `PROC MIXED`, comparing it to the analysis of variance approach in the context of a new example—UAV switch experiment. The two approaches yield the same results if (i) the design is balanced and (ii) all variance component estimates are either positive or allowed to be negative. The restricted maximum likelihood approach is generally preferable for (i) the estimation and testing of variance components and (ii) the analysis of fixed effects given sufficiently unbalanced data. The UAV switch experiment provides an interesting comparison of the approaches, because the design is balanced but a variance component estimate is negative if unconstrained.

The analysis of variance approach was illustrated in Sects. 17.10.2, 18.5, 19.8.1 and 19.8.2 for balanced data using both `PROC MIXED` and `PROC GLM`. Indeed, the analysis of variance approach can be reasonable and appropriate for the analysis of balanced data, as the fixed effect estimates are then best linear unbiased estimates, and the variance component estimates are then minimum variance unbiased estimates under normality. That said, variance component estimates can be negative even for balanced data. This may be reasonable for analysis of fixed effects, but it seems problematic if one is interested in inferences on variance components. Moreover, for unbalanced data, the fixed effects estimates can be inefficient, and the properties of the ANOVA-based variance component estimates are not well understood.

As an alternative to the analysis of variance approach, consider the use of maximum likelihood estimates, which generally have good large-sample properties. Under appropriate regularity conditions, the maximum likelihood estimator $\hat{\theta}$ of an estimable parameter θ is asymptotically $N(\theta, \text{CRLB})$, where CRLB is the Cramer–Rao lower bound for the variance of an unbiased estimator. Mind you, for the regularity conditions to hold, the estimate must be a solution to the likelihood equations—not obtained as a boundary condition—so these asymptotic properties do not apply for example to a variance component estimator when the estimate is constrained to be zero because the solution to the likelihood equations is negative. Likewise, given a factor with random effects but with few levels observed in an experiment, one should be skeptical of the asymptotic properties of the corresponding variance component estimate. That said, the asymptotic properties of MLEs should usually be reasonably applicable for the analysis of fixed effects unless an experiment is quite small. We will utilize restricted maximum likelihood estimation—a special case of maximum likelihood estimation.

The *restricted maximum likelihood approach* to fitting a mixed model involves two steps: (i) estimating the variance components by restricted maximum likelihood; then (ii) estimating the fixed effects by *estimated generalized least squares*—namely, treating the variance component estimates as true values, then computing the maximum likelihood estimates of the fixed effects, or equivalently, the *generalized least squares estimates*. To compute restricted maximum likelihood (ReML) estimates of the variance components, the original data is essentially replaced by a maximal set of linearly independent contrasts in the data each with mean zero (i.e. data contrasts the distributions of which do not involve the fixed effects), then the likelihood function of the contrasts is maximized. (Equivalently, one can estimate the fixed effects by ordinary least squares, then maximize the likelihood function of the residuals, the joint distribution of which does not involve the fixed effects. Hence, the restricted maximum likelihood approach is also known as *residual maximum likelihood*.) The ReML estimates of the variance components may be preferable to the usual maximum likelihood estimates, because the ReML estimates are the same as the ANOVA-based estimates if the data are balanced and all variance component estimates are positive. Also, restricted maximum likelihood adjusts in some sense for fixed effects, so often provides unbiased estimates of variance components. A new experiment will be introduced to illustrate the restricted maximum likelihood approach to analysis of split plot designs.

UAV Switch Experiment

Mahadevan (2009) conducted three experiments to evaluate the performance of a semi-automated computer display system designed to support a human operator's ability to monitor and control the complex dynamic operation of multiple unmanned aerial vehicles (UAVs) when the UAVs are involved in multiple combat-related tasks. His third experiment involved: 16 subjects (factor W); two alert techniques—a visual alert ($A = 1$) and an audio-visual alert ($A = 2$); and two levels of task complexity—a simple primary task coupled with a simple secondary task ($B = 1$), and a complex primary task coupled with a simple secondary task ($B = 2$). The experiment was a 2×2 split-plot design, with subjects as whole plots and two trials per subject as split-plots, with each level of A assigned to half of the subjects, and with both levels of B observed once on each subject. Hence, subject is nested within alert type. For each trial, each subject, while working on the primary task, was warned of the secondary task using one of the two alert techniques. One of the response variables measured was the time taken to switch to the secondary task following the alert, and the experimenter was interested in the effect of the nature of the alert (A) on the mean time to switch from the primary to the secondary task. The data are shown in Table 19.18. The model used by the experimenter is as follows.

$$\begin{gathered}
\begin{aligned} Y_{iuj} = \mu + \alpha_i &+ \epsilon^W_{iu} \\ &+ \beta_j + (\alpha\beta)_{ij} + \epsilon^S_{j(iu)}\,, \end{aligned} \\
\epsilon^W_{iu} \sim N(0, \sigma^2_W)\,, \quad \epsilon^S_{j(iu)} \sim N(0, \sigma^2_S)\,, \\
\epsilon^W_{iu}\text{'s and } \epsilon^S_{j(iu)}\text{'s are all mutually independent,} \\
i = 1, 2; \quad u = 1, \ldots, 8; \quad j = 1, 2.
\end{gathered} \tag{19.8.20}$$

Table 19.18 Time taken (in seconds) to switch to secondary task for the UAV switch experiment

A (Alert Type)	B (Complexity)	W (Subject)							
		1	2	3	4	5	6	7	8
1	1	6	5	5	5	5	7	5	6
	2	16	22	16	20	12	18	16	14
2	1	7	6	5	6	5	6	4	4
	2	6	7	6	8	6	6	7	6

Source Mahadevan (2009). Copyright © 2009 Sriram Mahadevan. Reprinted with permission

The SAS program for the UAV switch experiment is shown in Table 19.19. The first calls of `PROC GLM` and `PROC MIXED` each fit model (19.8.20), and some of the corresponding output is shown in Fig. 19.3. The random effects term `W(A)` in each SAS model represents the whole plot errors ϵ^{W}_{iu} in model (19.8.20), so $\sigma^2_{\texttt{W(A)}}$ in the SAS output corresponds to σ^2_W in our model. Also, `msE` in SAS output corresponds to msE_S in our formulae. `PROC MIXED` is by default fitting the model by ReML but, because of the `NOBOUND` option, the variance component estimates are not constrained to be positive. In fact, the whole-plots variance component estimate, $\hat{\sigma}^2_{\texttt{W(A)}} = -0.03571$, is negative, matching the estimate one could compute by hand from the `PROC GLM` output—namely, $\hat{\sigma}^2_{\texttt{W(A)}} = [\texttt{ms(W(A))} - \texttt{msE}]/2 \approx -0.036$. Because the data are balanced and the ReML estimates are allowed to be negative, the two procedures provide the same estimates of fixed effects and variance components, the same test results for fixed effects, and the same results for multiple comparisons (not shown).

Consider in comparison the output generated by the second calls of these procedures, shown in Fig. 19.4. `PROC MIXED` is still fitting model (19.8.20), but the restricted maximum likelihood estimate of $\sigma^2_{\texttt{W(A)}}$, now bounded, is zero. Correspondingly, the whole-plots term is effectively removed from the model, with the 14 degrees of freedom associated with it essentially pooled with the 14 degrees of freedom for error. Hence, there are now 28 degrees of freedom for error in the analysis, the (residual) error variance estimate of 3.1205 being the average of `ms(W(A))` and `msE` generated by the first call of `PROC GLM`. These results match those at the bottom of Fig. 19.4, generated by the second call of `PROC GLM`, for which the whole-plots term has been removed from the model for sake of illustration. While the multiple comparisons output generated by these procedure calls is not shown, the reader may confirm that the results would also match, as they do for the F-tests, with both procedures using 28 error degrees of freedom for all contrast standard errors.

The above information begs the question, "What is the correct analysis?" The analyses provided by the first calls of `PROC GLM` and `PROC MIXED` are arguably correct, except one should use the multiple comparisons results from `PROC MIXED` in any instances where `PROC GLM` uses incorrect standard errors. Otherwise, the SAS procedures correctly implement the planned statistical analyses using the proposed model. If the model is reasonable and appropriate, then the statistical results are valid, even if a simpler model with $\sigma^2_{\texttt{W(A)}} = 0$ suggested by the data would also be correct.

It is an open problem whether the analyses generated by the second calls of `PROC MIXED` or `PROC GLM` are correct—namely, whether they control error rates for any preplanned analysis, assuming the originally posed model (19.8.20) is correct. It seems *improper* for example to fit the first model using `PROC GLM`, see that the estimate $\hat{\sigma}^2_{\texttt{W(A)}}$ is negative, so change the model by removing the whole-plots term from the model, then fit the reduced model as in the second call of `PROC GLM` and use these results for the analysis. Control of error rates is an open problem when one uses the data to determine the model then uses the model to analyze the same data. Still, it is interesting that the second call of `PROC MIXED` fits and conducts the analysis under the assumption that model (19.8.20) is correct and, while it so happens that $\hat{\sigma}^2_{\texttt{W(A)}} = 0$, the originally postulated model is used for the analysis. If this approach could be shown to control error rates, then this would be the preferred analysis, since there are

Table 19.19 SAS program for the UAV switch experiment

```
DATA UAV3;
 INPUT A W B TIME;
 LINES;
 1 1 1  6
 1 1 2 16
 1 2 1  5
 1 2 2 22
 1 3 1  5
 : : :  :
 2 8 2  6
;
PROC GLM;
 TITLE 'Proc GLM: full model';
 CLASS A W B;
 MODEL TIME = A | B W(A);
 RANDOM W(A) / TEST;
 LSMEANS A / CL PDIFF E = W(A); * need PDIFF, else E = mse;
 LSMEANS A*B; * SAS would erroneously use MSE for CLs;
PROC MIXED NOBOUND;
 TITLE 'Proc Mixed NoBound';
 CLASS A W B;
 MODEL TIME = A | B / DDFM = SAT;
 RANDOM W(A);
 LSMEANS A A*B / CL DIFF;
PROC MIXED;
 TITLE 'Proc Mixed (Bounded)';
 CLASS A W B;
 MODEL TIME = A | B / DDFM = SAT;
 RANDOM W(A);
 LSMEANS A A*B / CL DIFF;
* Dropping W(A) from the Proc GLM model;
PROC GLM;
 TITLE 'Proc GLM: W(A) dropped from model';
 CLASS A B;
 MODEL TIME = A | B;
 LSMEANS A A*B / CL PDIFF;
```

benefits associated with essentially pooling `ms(W(A))` and `msE`, including for example increased power for some tests, as well as the availability of a common variance estimator for multiple comparisons of AB-treatment combinations.

Given any uncertainty regarding whether the latter approach (corresponding to the second calls) controls error rates, our formal (rather than exploratory) approach to data analysis is to advocate that the experimenter use the analyses generated by the first calls of the procedures, since this approach is known to control error rates for any preplanned analyses if the proposed model (19.8.20) is correct.

19.8.4 Recovery of Inter-block Information

The oats experiment, introduced in Sect. 19.3.4, involves a split-plot design with complete blocks at both the whole- and split-plots levels—namely, the levels of A assigned to whole plots within blocks

Fig. 19.3 Output for the UAV switch experiment: Procs GLM and Mixed, First Calls

Proc GLM: full model

The GLM Procedure

Dependent Variable: TIME

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	769.7812500	45.2812500	14.35	<.0001
Error	14	44.1875000	3.1562500		
Corrected Total	31	813.9687500			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	1	215.2812500	215.2812500	(see test below)	
B	1	306.2812500	306.2812500	97.04	<.0001
A*B	1	205.0312500	205.0312500	64.96	<.0001
W(A)	14	43.1875000	3.0848214	0.98	0.5168

Source	Type III Expected Mean Square
A	Var(Error) + 2 Var(W(A)) + Q(A,A*B)
B	Var(Error) + Q(B,A*B)
A*B	Var(Error) + Q(A*B)
W(A)	Var(Error) + 2 Var(W(A))

Tests of Hypotheses for Mixed Model Analysis of Variance

	Source	DF	Type III SS	Mean Square	F Value	Pr > F
*	A	1	215.281250	215.281250	69.79	<.0001
	Error: MS(W(A))	14	43.187500	3.084821		

* This test assumes one or more other fixed effects are zero.

Proc Mixed NoBound

The Mixed Procedure

Covariance Parameter Estimates

Cov Parm	Estimate
W(A)	-0.03571
Residual	3.1563

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
A	1	14	69.79	<.0001
B	1	14	97.04	<.0001
A*B	1	14	64.96	<.0001

constitute a randomized complete block design, and the levels of B assigned to split plots within whole plots constitute a randomized complete block design. With this dual complete block structure, the data analysis is the same whether the block effects are modeled as fixed (as in Sect. 19.3.4) or random (as in Sect. 19.8.4). Such would not be the case if for example the split-plot design involves incomplete blocks at the whole-plots level, as illustrated in this section.

Consider again the oats experiment and corresponding data in Table 19.3 (p. 710), but suppose one only has the data for levels 1 and 2 of A in blocks 3 and 5, for levels 0 and 2 of A in blocks 1 and 4, and for levels 0 and 1 of A in blocks 2 and 6. For this subset of the data, the levels of A assigned to whole plots in blocks constitute a balanced incomplete block design with blocks of size 2. The SAS program in Table 19.20 provides several approaches to the analysis of these data.

Fig. 19.4 Output for the UAV switch experiment: Procs GLM and Mixed, Second Calls

Results Viewer - sashtml.htm

Proc Mixed (Bounded)

The Mixed Procedure

Covariance Parameter Estimates	
Cov Parm	Estimate
W(A)	0
Residual	3.1205

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
A	1	28	68.99	<.0001
B	1	28	98.15	<.0001
A*B	1	28	65.70	<.0001

Proc GLM: W(A) dropped from model

The GLM Procedure

Dependent Variable: TIME

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	726.5937500	242.1979167	77.61	<.0001
Error	28	87.3750000	3.1205357		
Corrected Total	31	813.9687500			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	1	215.2812500	215.2812500	68.99	<.0001
B	1	306.2812500	306.2812500	98.15	<.0001
A*B	1	205.0312500	205.0312500	65.70	<.0001

In the first call of `PROC MIXED`, the ANOVA approach is used to conduct a type 3 analysis, and block effects are modeled as fixed. Figure 19.5 contains the corresponding output. Because the block effects are modeled as fixed, the estimates of main-effect-of-A contrasts are *intra-block estimates*—namely, each is composed (by summing over blocks) of within-block contrasts of observations—as is necessary for the fixed block effects to cancel out to yield unbiased estimates. Analogously, for testing for main effects of A, the numerator of the F-statistic is the mean square for A adjusted for block effects, which can be computed from the sum of squares of an appropriate set of such intra-block estimates of main-effect-of-A contrasts. The corresponding data analysis is called the *intra-block analysis*. The second call of `PROC MIXED` uses the restricted maximum likelihood approach but provides the same intra-block analysis, since the block effects are again modeled as fixed. The output (not shown) would include the same `Type 3 Tests of Fixed Effects` and `Differences of Least Squares Means` as in Fig. 19.5. For sake of comparison with subsequent analyses, note that the pairwise comparisons with the control for A each have estimated standard error 11.2883 with 4 degrees of freedom.

Table 19.20 SAS program illustrating the intra-block analysis and also the recovery of inter-block information—the oats split-plot experiment

```
DATA OATBIBD;
 INPUT BLOCK WP A B  Y;
 IF A = 0 AND (BLOCK = 3 OR BLOCK = 5) THEN DELETE;
 IF A = 1 AND (BLOCK = 1 OR BLOCK = 4) THEN DELETE;
 IF A = 2 AND (BLOCK = 2 OR BLOCK = 6) THEN DELETE;
 LINES;
  1 1  2 3  156
  1 1  2 2  118
  1 1  2 1  140
  1 1  2 0  105
  1 2  0 0  111
  : :  : :    :
  6 3  2 3  121
;
*** Intra-block analyses;
PROC MIXED METHOD = TYPE3;
 TITLE 'ANOVA Approach, Type 3, Fixed Block Effects';
 CLASS BLOCK A B;
 MODEL Y = BLOCK A B A*B / DDFM = SAT;
 RANDOM A*BLOCK; * Random whole-plot effects;
 LSMEANS A / PDIFF = CONTROL('0') CL DIFF ADJUST = DUNNETT ALPHA = 0.01;
PROC MIXED METHOD = ReML NOBOUND;
 TITLE 'ReML Approach, Fixed Block Effects';
 CLASS BLOCK A B WP;
 MODEL Y = BLOCK A B A*B / DDFM = SAT;
 RANDOM WP(BLOCK);
 LSMEANS A / PDIFF = CONTROL('0') CL DIFF ADJUST = DUNNETT ALPHA = 0.01;
*** Analyses recovering inter-block information;
PROC MIXED METHOD = TYPE3;
 TITLE 'ANOVA Approach, Type 3, Random Block Effects';
 CLASS BLOCK A B;
 MODEL Y = A B A*B / DDFM = SAT;
 RANDOM BLOCK A*BLOCK;
 LSMEANS A / PDIFF = CONTROL('0') CL DIFF ADJUST = DUNNETT ALPHA = 0.01;
PROC MIXED METHOD = ReML NOBOUND;
 TITLE 'ReML Approach, Random Block Effects';
 CLASS BLOCK A B WP;
 MODEL Y = A B A*B / DDFM = SAT;
 RANDOM BLOCK WP(BLOCK);
 LSMEANS A / PDIFF = CONTROL('0') CL DIFF ADJUST = DUNNETT ALPHA = 0.01;
```

In the third call of `PROC MIXED`, the ANOVA approach is used to conduct a type 3 analysis, but the block effects are modeled as independent random effects with mean zero and variance σ_θ^2. With random block effects, one can obtain unbiased ordinary least squares estimates of main-effect-of-A contrasts using contrasts of block totals (the sum total of the observations in each block), and these so-called *inter-block estimates* are in addition to and independent of the intra-block estimates. So, for any main-effect-of-A contrast, any fixed weighted average of the corresponding intra- and inter-block estimates provides an unbiased estimate of the contrast, and the best (minimum variance) estimate would be obtained when each weight is inversely proportional to the variance of the corresponding estimate. Unfortunately, this best weighting is unknown, because the variance of the intra-block estimate is a

ANOVA Approach, Type 3, Fixed Block Effects

The Mixed Procedure

Type 3 Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	Expected Mean Square
BLOCK	5	10237	2047.308333	Var(Residual) + 4 Var(BLOCK*A) + Q(BLOCK)
A	2	376.541667	188.270833	Var(Residual) + 4 Var(BLOCK*A) + Q(A,A*B)
B	3	15966	5321.916667	Var(Residual) + Q(B,A*B)
A*B	6	1361.500000	226.916667	Var(Residual) + Q(A*B)
BLOCK*A	4	3058.208333	764.552083	Var(Residual) + 4 Var(BLOCK*A)
Residual	27	4197.750000	155.472222	Var(Residual)

Source	Error Term	Error DF	F Value	Pr > F
BLOCK	MS(BLOCK*A)	4	2.68	0.1806
A	MS(BLOCK*A)	4	0.25	0.7928
B	MS(Residual)	27	34.23	<.0001
A*B	MS(Residual)	27	1.46	0.2294
BLOCK*A	MS(Residual)	27	4.92	0.0041
Residual	.			

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
BLOCK	5	4	2.68	0.1806
A	2	4	0.25	0.7928
B	3	27	34.23	<.0001
A*B	6	27	1.46	0.2294

Differences of Least Squares Means								
Effect	A	_A	Estimate	Standard Error	DF	Adjustment	Adj Lower	Adj Upper
A	1	0	6.0417	11.2883	4	Dunnett-Hsu	-54.5061	66.5894
A	2	0	7.4583	11.2883	4	Dunnett-Hsu	-53.0894	68.0061

Fig. 19.5 SAS output illustrating the intra-block analysis—the oats split-plot experiment

multiple of $\sigma^2 + 4\sigma_W^2$, the variance of the inter-block estimate is a multiple of $\sigma^2 + 4\sigma_W^2 + 8\sigma_\theta^2$, and these variances are unknown. A common approach is to use the estimated variances to determine the weights, but random weights would make the resulting analyses approximate. If the weights are well chosen, the resulting combined estimate is better (has smaller variance) than the intra-block estimate, due to recovery of inter-block information.

The output corresponding to the third call of `PROC MIXED` is in Fig. 19.6. Because a type 3 analysis is requested, the model is initially fit by ordinary least squares, providing a corresponding `Type 3 Analysis of Variance` identical to that generated by the first call of `PROC MIXED`, again with effects of A adjusted for block effects. However, this ANOVA information is then used to estimate the variance components, then the variance component estimates are in turn used to re-estimate the fixed

Fig. 19.6 SAS output illustrating the recovery of inter-block information—the oats split-plot experiment; ANOVA approach

ANOVA Approach, Type 3, Random Block Effects

The Mixed Procedure

Covariance Parameter Estimates	
Cov Parm	Estimate
BLOCK	178.16
BLOCK*A	152.27
Residual	155.47

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
A	2	5	0.48	0.6422
B	3	27	34.23	<.0001
A*B	6	27	1.46	0.2294

Differences of Least Squares Means								
Effect	A	_A	Estimate	Standard Error	DF	Adjustment	Adj Lower	Adj Upper
A	1	0	3.5224	10.6837	5	Dunnett-Hsu	-45.8986	52.9434
A	2	0	10.3425	10.6837	5	Dunnett-Hsu	-39.0785	59.7635

effects by estimated generalized least squares. Using this revised information, SAS software provides updated `Type 3 Tests of Fixed Effects`, including an approximate F test for A with 5 rather than 4 denominator degrees of freedom, and revised estimates of the pairwise comparisons with the control for A, each now with estimated standard error 10.6837 with 5 degrees of freedom. The smaller standard error and increased number of degrees of freedom are both helpful consequences of the recovered inter-block information, increasing test power and decreasing confidence interval width.

In the fourth call of `PROC MIXED`, the restricted maximum likelihood approach is used for the analysis, with the block effects again modeled as independent random effects with mean zero and variance σ_θ^2. The output corresponding to the third call of `PROC MIXED` is in Fig. 19.7. Under this approach, ReML estimates of the variance components are computed, then these are used to estimate the fixed effects by generalized least squares. Based on these estimates, SAS software provides `Type 3 Tests of Fixed Effects`, including an approximate F test for A with 4.95 denominator degrees of freedom, and estimates of the pairwise comparisons with the control for A, each with estimated standard error 10.7095 with 4.95 degrees of freedom. These values are slightly worse than the results of the third call of `PROC MIXED`, but still much better than for the intra-block analysis generated by the first two calls. Clearly there is a helpful recovery of inter-block information.

Turning attention to the split-plot comparisons, note that the F-tests for B, and likewise for AB, are the same in each analysis. This is the case because a randomized complete block design was used for the assignment of levels of B to split plots in whole plots (serving as blocks). Consequently, for each main effect of B and AB-interaction contrast, the corresponding contrast in treatment means is a within-whole-plot contrast, free of whole-plot or block effects, so no adjustment is needed.

Fig. 19.7 SAS output illustrating the recovery of inter-block information—the oats split-plot experiment; ReML approach

ReML Approach, Random Block Effects

The Mixed Procedure

Covariance Parameter Estimates	
Cov Parm	Estimate
BLOCK	178.31
WP(BLOCK)	153.25
Residual	155.47

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
A	2	4.95	0.48	0.6432
B	3	27	34.23	<.0001
A*B	6	27	1.46	0.2294

Differences of Least Squares Means								
Effect	A	_A	Estimate	Standard Error	DF	Adjustment	Adj Lower	Adj Upper
A	1	0	3.5161	10.7095	4.95	Dunnett-Hsu	-46.2998	53.3319
A	2	0	10.3497	10.7095	4.95	Dunnett-Hsu	-39.4661	60.1656

In the above example, recovery of the inter-block information provided better information for analysis of the whole-plots factor A. For example, for each treatment-versus-control contrast estimate for A, recovery of inter-block information provided a tighter confidence interval, due to both a smaller standard error and more error degrees of freedom. In planning such an experiment, the experimenter should determine in advance how the data will be analyzed. While one could plan on conducting the intra-block analysis, recovery of the inter-block information should usually be beneficial.

19.8.5 ReML and Unbalanced Data

In this section, we illustrate the restricted maximum likelihood approach using `PROC MIXED` for analysis of an unbalance split-plot design, revisiting the mobile computing field study experiment introduced in Sect. 19.7. Recall, while the planned experiment was nicely balanced, it was discovered prior to analysis of the data that one of the subjects traversed one of the paths in the reverse direction, making the corresponding observation inappropriate to use in the analysis. As such, one of the 72 planned observations (given in Table 19.11, p. 719) is listed as missing. In Sects. 19.7.1 and 19.7.2, an approximate analysis of the split-plot design was provided by estimating the missing value and treating it as observed, using standard formulas for analysis of the balanced design via the analysis of variance approach, but adjusting degrees of freedom appropriately. With one observation lost, the data are no longer balanced, making the data analysis conceptually and computationally more difficult. However, SAS `PROC MIXED` handles this situation nicely, as we shall illustrate.

Table 19.21 SAS program for the mobile computing field study experiment

```
DATA MCFS71;
 INPUT SUBJ DAY ORDER O2 O3 PATH P2 P3 A B Y;
 LINES;
  1  1  1  0  0  1  0  0  1  1  49.321
  1  1  2  0  1  2  0  1  1  2  24.386
  1  1  3  0  2  3  0  2  1  3  37.680
  :  :  :  :  :  :  :  :  :  :     :
 12  1  3  0  2  5  1  1  1  1  30.146
 12  2  4  1  0  3  0  2  2  2     .
 12  2  5  1  1  1  0  0  2  3  58.481
 12  2  6  1  2  2  0  1  2  1  62.659
;
PROC MIXED NOBOUND;
 TITLE 'Analysis of RMSE without missing value';
 CLASS SUBJ DAY O2 O3 P2 P3 A B;
 MODEL Y = O2 P2 A  O3 O2*O3 P3 P2*P3 B A*B / DDFM = SAT;
 RANDOM SUBJ DAY(SUBJ);
 LSMEANS A B / CL DIFF ALPHA = 0.05;
* To get consolidated tests for order and path;
PROC MIXED NOBOUND;
 CLASS SUBJ DAY ORDER PATH A B;
 MODEL Y = ORDER PATH A B A*B / DDFM = SAT;
 RANDOM SUBJ DAY(SUBJ);
```

The SAS program for analysis of the 71-observation split-plot design via the restricted maximum likelihood approach is shown in Table 19.21, with the corresponding output in Fig. 19.8. The mixed model for the analysis was provided in Eq. (19.7.18), p. 718. Using `PROC MIXED`, the three variance components are (by default) estimated by restricted maximum likelihood. In this case, the three estimates are all positive. Given these variance component estimates, the fixed effects are then estimated by generalized least squares.

`PROC MIXED` provides Type III F-tests for each fixed effect—Type III in the sense that the tests are based on estimates of effects and corresponding variability obtained by fitting the full model. The option `DDFM = SAT` in the `MODEL` statement causes use of a general Satterthwaite approximation for the denominator degrees of freedom, equivalent to Satterthwaite's approximation for a balanced design. Readers are referred to SAS documentation for details. Regarding the two treatment factors, only the effects of visual presentation format (B) are significant, with $p = 0.0166$. Concerning the nuisance factors path and order, these involved both within- and between-whole-plot comparisons before the lost observation caused design imbalance. For each of these factors, the first call of `PROC MIXED` provides tests of each of three pseudofactor components, including for example tests for P_2, P_3 and P_2P_3 for path. The second call of `PROC MIXED` provides a consolidated test for each of these nuisance factors. It is not surprising that path has significant effects, and these are attributable to the three distinct paths used as distinguished by P_3.

The `LSMEANS` statement generated the pairwise `Differences of Least Squares Means` output for each of factors A and B, including individual t-tests and individual 95% confidence intervals for each pairwise comparison. These results are similar to those obtained in Sect. 19.7.2, where the approach used was to estimate the missing value and analyze the balanced design, though there Tukey's method was used. Based on the individual 95% confidence intervals provided here, level 2 of B—the

Analysis of RMSE without missing value

The Mixed Procedure

Covariance Parameter Estimates	
Cov Parm	Estimate
SUBJ	63.9909
DAY(SUBJ)	28.1379
Residual	92.9959

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
O2	1	8.5	0.25	0.6316
P2	1	8.5	1.87	0.2064
A	1	8.5	0.19	0.6768
O3	2	34.6	0.04	0.9581
O2*O3	2	34.6	0.37	0.6927
P3	2	34.6	10.62	0.0003
P2*P3	2	34.6	0.14	0.8718
B	2	34.6	4.62	0.0166
A*B	2	34.6	0.79	0.4634

Differences of Least Squares Means													
Effect	A	B	_A	_B	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
A	1		2		-1.3630	3.1587	8.5	-0.43	0.6768	0.05	-8.5734	5.8474
B		1		2	8.1958	2.8325	34.8	2.89	0.0065	0.05	2.4442	13.9475
B		1		3	1.6903	2.7838	34.4	0.61	0.5477	0.05	-3.9648	7.3454
B		2		3	-6.5055	2.8325	34.8	-2.30	0.0278	0.05	-12.2572	-0.7539

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
ORDER	5	31.1	0.21	0.9555
PATH	5	31.1	4.70	0.0026

Fig. 19.8 Output for the mobile computing field study experiment

birds' eye view egocentric visual presentation format—has a significantly smaller *RMSE* from the intended path than do the other two visual presentation formats. It is interesting that the denominator degrees of freedom is not constant for the factor *B* pairwise comparisons, due to the imbalance created by the lost observation. This indicates that the pairwise comparisons do not all utilize the same variance estimator. As a consequence, the Bonferroni method should be used if multiple comparisons for *B* are of interest. These comparisons could be generated by adding the option `ADJUST = BON` to the `LSMEANS` statement. Replacing the keyword `BON` by `DUNNETT`, `SCHEFFE` or `TUKEY` would yield the corresponding multiple comparisons method, though these latter methods assume availability of a common variance estimator so are not applicable here.

One could also generate pairwise comparisons for the *AB* combinations by including `A*B` in the list of effects in the `LSMEANS` statement. However, only the Bonferroni method of multiple comparisons would be applicable, since there would not be a common variance estimator for all of these pairwise comparisons. Also, the corresponding variance estimators would generally be composite variance estimators, in which case Satterthwaite's approximation would be used.

19.9 Using R Software

In this section, we illustrate use of the R software for analyzing split-plot designs. The analysis of variance approach can be used for balanced data using `aov`, as illustrated for analysis of the oats experiment in Sect. 19.9.1. In Sect. 19.9.2, the UAV experiment is briefly revisited to illustrate the analysis of simple contrasts by conditioning on other factor level combinations, given a model fit using the `aov` function. In Sect. 19.9.3, we introduce a restricted maximum likelihood approach to analysis of split plot designs using the `lmer` function of the `lme4` package, comparing it to the analysis of variance approach using the `aov` function in the context of UAV switch experiment—a new example involving a negative variance component estimate. In Sect. 19.9.4, the recovery of inter-block information is illustrated, using a subset of the oats data yielding a balanced incomplete block design for the whole-plots factor, and using the `lmer` function to fit the mixed model by restricted maximum likelihood. Finally, in Sect. 19.9.5, analysis of unbalanced data is illustrated using the `lmer` function and restricted maximum likelihood estimation for the mobile computing field study, excluding the missing observation from analysis of the split plot design.

19.9.1 The Analysis of Variance Approach

The *analysis of variance approach* to the analysis of mixed models involves: (i) fitting the model by ordinary least squares, treating the random effects as fixed; and (ii) using the expected mean squares to determine unbiased estimates of the variance components. In analyzing a balanced split-plot design by the analysis of variance approach, the `aov` function automatically provides correct F tests of fixed effects, taking into account the two separate parts of the analysis of variance table, corresponding to between-whole-plots and within-whole-plots comparisons. Table 19.22 contains R code and output, providing an analysis of variance table similar to the one in Table 19.4 (p. 710) presented for the oats experiment in Sect. 19.3.4.

The approach used by `aov` in this analysis of a split plot design is a *type III analysis*, utilizing the type III sums of squares and corresponding expected mean squares. In the call of `aov` in Table 19.22, the model explicitly includes all sources of variation as in model (19.2.2) except the split-plot error term $\epsilon^S_{j(hi)}$, which plays the role of the usual error variable. The whole-plot error term $\epsilon^W_{i(h)}$ is represented by `fWP:fBlock`—the random effects of whole plots nested within blocks—entered into the model via the `Error` function used to designate random effects. The term `fBlock` is likewise included in the `Error` function to designate the block effects as random.

Table 19.23 continues the R program and output of Table 19.22, illustrating the use of `lsmeans` to implement standard multiple comparisons procedures. For example, Dunnett's procedure for comparing each level of A with control level 0 and each level of B with control level 0 is requested via the `lsmeans` and `summary(contrast`... statements shown in Table 19.23. The syntax `ref=1` designates the first level of a factor as the reference level or control, 0 being the first level of both A and B. R automatically uses the correct standard error estimates, using the split-plot term *MSE* to estimate standard errors for the B contrasts, while using the whole-plot error mean square represented by `fBlock:fWP` for the A contrasts.

Table 19.22 R program and output illustrating type 3 analysis of variance, including tests for fixed effects—the oats split-plot experiment

```
> oats.data = read.table("data/oats.txt", header=T)
> oats.data = within(oats.data, {fBlock = factor(Block);
+             fWP = factor(WP); fA = factor(A); fB = factor(B) })
> head(oats.data, 3)
  Block WP A B   y fB fA fWP fBlock
1     1  1 2 3 156  3  2   1      1
2     1  1 2 2 118  2  2   1      1
3     1  1 2 1 140  1  2   1      1

> # Least squares ANOVA
> # Set contrast options for correct lsmeans and contrasts
> options(contrasts = c("contr.sum", "contr.poly"))
> model1 = aov(y ~ fA + fB + fA:fB + Error(fBlock + fWP:fBlock),
+              data=oats.data)
> summary(model1)

Error: fBlock
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  5  15875    3175

Error: fBlock:fWP
          Df Sum Sq Mean Sq F value Pr(>F)
fA         2   1786     893    1.49   0.27
Residuals 10   6013     601

Error: Within
          Df Sum Sq Mean Sq F value  Pr(>F)
fB         3  20021    6674    37.7 2.5e-12
fA:fB      6    322      54     0.3    0.93
Residuals 45   7969     177
```

19.9.2 Simple Contrasts

In this section, we briefly revisit the UAV experiment of Sect. 19.6 to illustrate the analysis of simple contrasts by conditioning on other factor level combinations. The R program in Table 19.24 shows how to generate the analyses provided in Sect. 19.6 and includes selected output. The `aov` and `summary` functions generate the analysis of variance in Table 19.10 (not shown here). The `lsmeans` and `summary(contrast...)` functions are used to generate the 99% confidence interval for the main effect of cue (A). Then they are used a second time to generate information for the simple contrasts for cue, analogous to what was provided in Sect. 19.6.2 (p. 716), comparing the effects of the two levels of cue (A) at each of the eight BC combinations. For the simple contrasts, the syntax "`fA | fC:fB`" says to compare the levels of A given each combination of levels of B and C. The corresponding standard errors involve composite variance estimates, and Satterthwaite's approximation is used by R by default. Output for three of the eight simple contrasts is in the bottom of Table 19.24.

Table 19.23 Multiple comparisons—the oats split-plot experiment

```
> # Multiple comparisons: Dunnett's method
> library(lsmeans)
> lsmA = lsmeans(model1, ~ fA)
> set.seed(21531957)
> summary(contrast(lsmA, method="trt.vs.ctrl", adjust="mvt", ref=1),
+         infer=c(T,T), level=0.99, side="two-sided")

 contrast estimate     SE df lower.CL upper.CL t.ratio p.value
 1 - 0       6.875 7.0789 10  -18.123   31.873   0.971  0.5423
 2 - 0      12.167 7.0789 10  -12.831   37.164   1.719  0.1973

Results are averaged over the levels of: fB
Confidence level used: 0.99
Conf-level adjustment: mvt method for 2 estimates
P value adjustment: mvt method for 2 tests

> lsmB = lsmeans(model1, ~ fB)
> set.seed(21531957)
> summary(contrast(lsmB, method="trt.vs.ctrl", adjust="mvt", ref=1),
+         infer=c(T,T), level=0.99, side="two-sided")

 contrast estimate     SE df lower.CL upper.CL t.ratio p.value
 1 - 0      19.500 4.4358 45   5.8789   33.121   4.396  0.0002
 2 - 0      34.833 4.4358 45  21.2122   48.454   7.853  <.0001
 3 - 0      44.000 4.4358 45  30.3789   57.621   9.919  <.0001

Results are averaged over the levels of: fA
Confidence level used: 0.99
Conf-level adjustment: mvt method for 3 estimates
P value adjustment: mvt method for 3 tests
```

19.9.3 The Restricted Maximum Likelihood Approach

In this section, we introduce a restricted maximum likelihood (ReML) approach to analysis of split plot designs using the `lmer` function of the `lme4` package, comparing the restricted maximum likelihood approach to the analysis of variance approach in the context of a new example—UAV switch experiment. The two approaches yield the same results if (i) the design is balanced and (ii) all variance component estimates are either positive or allowed to be negative, though the `lmer` function does not allow variance component estimates to be negative. The restricted maximum likelihood approach is generally preferable for (i) the estimation and testing of variance components and (ii) the analysis of fixed effects given sufficiently unbalanced data. The UAV switch experiment provides an interesting comparison of the approaches, because the design is balanced but a variance component estimate is negative if unconstrained.

The analysis of variance approach was illustrated in Sects. 17.11.2, 18.6, 19.9.1 and 19.9.2 for balanced data using the `aov` function. Indeed, the analysis of variance approach can be reasonable and appropriate for the analysis of balanced data, as the fixed effect estimates are then best linear unbiased estimates, and the variance component estimates are then minimum variance unbiased estimates under normality. That said, variance component estimates can be negative even for balanced data. This may be reasonable for analysis of fixed effects, but it seems problematic if one is interested in inferences

Table 19.24 R program and selected output illustrating analysis of simple contrasts—UAV experiment

```
> uav.data = read.table("data/uav.txt", header=T)
> uav.data = within(uav.data, {fA = factor(A); fW = factor(W);
+                               fB = factor(B); fC = factor(C) })
> head(uav.data, 3)
A W B C Time fC fB fW fA
1 1 1 1 1   29  1  1  1  1
2 1 1 1 2   28  2  1  1  1
3 1 1 1 3   49  3  1  1  1

> # Least squares ANOVA
> # Set contrast options for correct lsmeans and contrasts
> options(contrasts = c("contr.sum", "contr.poly"))
> model1 = aov(Time ~ fA*fB*fC + Error(fA:fW), data=uav.data)
> summary(model1)

> # Compare 2 levels of fA pairwise
> library(lsmeans)
> lsmA = lsmeans(model1, ~ fA)
> summary(contrast(lsmA, method="pairwise"), infer=c(T,T))

contrast estimate      SE df lower.CL upper.CL t.ratio p.value
1 - 2      20.062 0.76195 14   18.428   21.697   26.33  <.0001

Results are averaged over the levels of: fB, fC
Confidence level used: 0.95

> # Compare 2 levels of fA pairwise given each fB:fC combination
> library(lsmeans)
> lsmAgBC = lsmeans(model1, ~ fA | fC:fB)
> summary(contrast(lsmAgBC, method="pairwise", adjust="none"),
+          infer=c(T,T), level=0.99)

fC = 1, fB = 1:
contrast estimate     SE     df lower.CL upper.CL t.ratio p.value
1 - 2      14.375 1.8643 110.22   9.4885   19.262   7.711  <.0001

fC = 2, fB = 1:
contrast estimate     SE     df lower.CL upper.CL t.ratio p.value
1 - 2      13.375 1.8643 110.22   8.4885   18.262   7.174  <.0001
...
fC = 4, fB = 2:
contrast estimate     SE     df lower.CL upper.CL t.ratio p.value
1 - 2      23.750 1.8643 110.22  18.8635   28.637  12.740  <.0001

Confidence level used: 0.99
```

on variance components. Moreover, for unbalanced data, the fixed effects estimates can be inefficient, and the properties of the ANOVA-based variance component estimates are not well understood.

As an alternative to the analysis of variance approach, consider the use of maximum likelihood estimates, which generally have good large-sample properties. Under appropriate regularity conditions, the maximum likelihood estimator $\hat{\theta}$ of an estimable parameter θ is asymptotically $N(\theta, \text{CRLB})$, where CRLB is the Cramer–Rao lower bound for the variance of an unbiased estimator. Mind you, for the regularity conditions to hold, the estimate must be a solution to the likelihood equations—not obtained as a boundary condition—so these asymptotic properties do not apply for example to a variance component estimator when the estimate is constrained to be zero because the solution to the likelihood equations is negative. Likewise, given a factor with random effects but with few levels observed in an experiment, one should be skeptical of the asymptotic properties of the corresponding variance component estimate. That said, the asymptotic properties of MLEs should usually be reasonably applicable for the analysis of fixed effects unless an experiment is quite small. We will utilize restricted maximum likelihood estimation—a special case of maximum likelihood estimation.

The *restricted maximum likelihood approach* to fitting a mixed model involves two steps: (i) estimating the variance components by restricted maximum likelihood; then (ii) estimating the fixed effects by *estimated generalized least squares*—namely, treating the variance component estimates as true values, then computing the maximum likelihood estimates of the fixed effects, or equivalently, the *generalized least squares estimates*. To compute restricted maximum likelihood (ReML) estimates of the variance components, the original data is essentially replaced by a maximal set of linearly independent contrasts in the data each with mean zero (i.e. data contrasts the distributions of which do not involve the fixed effects), then the likelihood function of the contrasts is maximized. (Equivalently, one can estimate the fixed effects by ordinary least squares, then maximize the likelihood function of the residuals, the joint distribution of which does not involve the fixed effects. Hence, the restricted maximum likelihood approach is also known as *residual maximum likelihood*.) The ReML estimates of the variance components may be preferable to the usual maximum likelihood estimates, because the ReML estimates are the same as the ANOVA-based estimates if the data are balanced and all variance component estimates are positive. Also, restricted maximum likelihood adjusts in some sense for fixed effects, so often provides unbiased estimates of variance components. A new experiment will be introduced to illustrate the restricted maximum likelihood approach to analysis of split plot designs.

UAV Switch Experiment

Mahadevan (2009) conducted three experiments to evaluate the performance of a semi-automated computer display system designed to support a human operator's ability to monitor and control the complex dynamic operation of multiple unmanned aerial vehicles (UAVs) when the UAVs are involved in multiple combat-related tasks. His third experiment involved: 16 subjects (factor W); two alert techniques—a visual alert ($A = 1$) and an audio-visual alert ($A = 2$); and two levels of task complexity—a simple primary task coupled with a simple secondary task ($B = 1$), and a complex primary task coupled with a simple secondary task ($B = 2$). The experiment was a 2×2 split-plot design, with subjects as whole plots and two trials per subject as split-plots, with each level of A assigned to half of the subjects, and with both levels of B observed once on each subject. Hence, subject is nested within alert type. For each trial, each subject, while working on the primary task, was warned of the secondary task using one of the two alert techniques. One of the response variables measured was the time taken to switch to the secondary task following the alert, and the experimenter was interested in the effect of the nature of the alert (A) on the mean time to switch from the primary to the secondary task. The data are shown in Table 19.25. The model used by the experimenter is as follows.

Table 19.25 Time taken (in seconds) to switch to secondary task for the UAV switch experiment

A (Alert Type)	B (Complexity)	W (Subject)							
		1	2	3	4	5	6	7	8
1	1	6	5	5	5	5	7	5	6
	2	16	22	16	20	12	18	16	14
2	1	7	6	5	6	5	6	4	4
	2	6	7	6	8	6	6	7	6

Source Mahadevan (2009). Copyright © 2009 Sriram Mahadevan. Reprinted with permission

$$
\begin{aligned}
Y_{iuj} = \mu + \alpha_i + \epsilon^W_{iu} & \\
+ \beta_j + (\alpha\beta)_{ij} + \epsilon^S_{j(iu)}\,, & \\
\epsilon^W_{iu} \sim N(0, \sigma^2_W)\,, \quad \epsilon^S_{j(iu)} \sim N(0, \sigma^2_S)\,, & \\
\epsilon^W_{iu}\text{'s and } \epsilon^S_{j(iu)}\text{'s are all mutually independent,} & \\
i = 1, 2; \quad u = 1, \ldots, 8; \quad j = 1, 2. &
\end{aligned}
\tag{19.9.21}
$$

The R program and selected output for the UAV switch experiment is shown in Tables 19.26 and 19.27. Table 19.26 illustrates the analysis of variance approach, which is recommended for analysis of fixed effects given a balanced design—the case here. The model term `Error(fA:fW)` in the `aov` function models the whole plot errors ϵ^W_{iu} as random whole plots effects $W(A)$, and the corresponding mean square provides the denominator for testing main effects of A, as appropriate. R code for multiple comparisons is shown, but not the corresponding output.

For sake of comparison, the restricted maximum likelihood approach is illustrated in the top of Table 19.27. The `lmer` function fits a linear mixed effects model by restricted maximum likelihood. This function is part of the `lme4` package, but loading it via the `lmerTest` package adds the p-values to the analysis of variance table. For the model specified in the `lmer` statement, the term `(1|fA:fW)` causes inclusion of random $W(A)$ effects in the model, representing the whole plot errors ϵ^W_{iu}. The corresponding variance component estimate, bounded to be nonnegative, is essentially zero, ($\hat{\sigma}^2_W = 3.88 \times 10^{-19}$ per output from the `summary` command). Consequently, the restricted maximum likelihood approach effectively removes this term from the model, pooling its degrees of freedom with that for (split plot) error, yielding 28 error degrees of freedom rather than 14 each for $W(A)$ and error as in the analysis of variance approach. This is confirmed by the analysis of variance provided at the bottom of Table 19.27, obtained by removing whole plot effects from the model. These two analyses provide exactly the same F tests for the fixed effects. They also generate the same multiple comparisons results (not shown), with both procedures using 28 error degrees of freedom for all contrast standard errors.

The analysis in Table 19.26 is correct and should be used. It correctly implements the planned statistical analyses using the proposed model. If the model is reasonable and appropriate, then the statistical results are valid, even if a simpler model with $\sigma^2_W = 0$ suggested by the data would also be correct.

Table 19.26 R program and selected output for the UAV switch experiment: analysis of variance approach

```
> uav3.data = read.table("data/uav3.txt", header=T)
> uav3.data = within(uav3.data,
+                     {fA = factor(A); fW = factor(W); fB = factor(B) })
> head(uav3.data, 3)

  A W B Time fB fW fA
1 1 1 1    6  1  1  1
2 1 1 2   16  2  1  1
3 1 2 1    5  1  2  1

> # ANOVA: full model
> # Set contrast options for correct lsmeans and contrasts
> options(contrasts = c("contr.sum", "contr.poly"))
> model1 = aov(Time ~ fA + fB + fA:fB + Error(fA:fW), data=uav3.data)
> summary(model1)

Error: fA:fW
          Df Sum Sq Mean Sq F value  Pr(>F)
fA         1  215.3   215.3    69.8 8.2e-07
Residuals 14   43.2     3.1

Error: Within
          Df Sum Sq Mean Sq F value  Pr(>F)
fB         1  306.3   306.3      97 1.1e-07
fA:fB      1  205.0   205.0      65 1.3e-06
Residuals 14   44.2     3.2

> # Multiple comparisons
> library(lsmeans)
> lsmA.aov = lsmeans(model1, ~ fA)
> summary(contrast(lsmA.aov, method="pairwise"),
+         infer=c(T,T), side="two-sided")
> lsmAB.aov = lsmeans(model1, ~ fA:fB)
> summary(contrast(lsmAB.aov, method="pairwise", adjust="tukey"),
+         infer=c(T,T), side="two-sided")
```

It is an open problem whether the analyses provided in Table 19.27 are strictly correct—namely, whether they control error rates for any preplanned analysis, assuming the originally posed model (19.9.21) is correct. It seems *improper* for example if one were to fit the original full model by least squares, see that the estimate $\hat{\sigma}^2_W$ is negative, so change the model by removing the whole-plots term from the model, then fit the reduced model and use these results for the analysis. Control of error rates is an open problem when one uses the data to determine the model then uses the model to analyze the same data. Still, it is interesting that the restricted maximum likelihood approach fits and conducts the analysis under the assumption that model (19.9.21) is correct and, while it so happens that $\hat{\sigma}^2_W = 0$, the originally postulated model is used for the analysis. If this approach could be shown to control error rates, then this would seem to be the preferred analysis, since effectively setting $\hat{\sigma}^2_W = 0$

Table 19.27 R program and selected output for the UAV switch experiment, continued: restricted maximum likelihood approach, and reduced ANOVA model

```
> # ReML: full model
> library(lmerTest)
> model2 = lmer(Time ~ fA + fB + fA:fB + (1|fA:fW), data=uav3.data)
> anova(model2)

Analysis of Variance Table of type III  with  Satterthwaite
approximation for degrees of freedom
      Sum Sq Mean Sq NumDF DenDF F.value  Pr(>F)
fA       215     215     1    28    69.0 4.9e-09
fB       306     306     1    28    98.2 1.2e-10
fA:fB    205     205     1    28    65.7 8.0e-09

> summary(model2)

...
Random effects:
 Groups   Name        Variance Std.Dev.
 fA:fW    (Intercept) 3.88e-19 6.23e-10
 Residual             3.12e+00 1.77e+00
...

> # Multiple comparisons
> # Detach then reload lsmeans, to avoid issues from its masking by lmerTest
> detach("package:lsmeans", unload=TRUE)
> library(lsmeans)
> lsmA2 = lsmeans(model2, ~ fA)
> summary(contrast(lsmA2, method="pairwise"),
+         infer=c(T,T), side="two-sided")
> lsmAB.reml = lsmeans(model2, ~ fA:fB)
> summary(contrast(lsmAB.reml, method="pairwise", adjust="tukey"),
+         infer=c(T,T), side="two-sided")

> # ANOVA: reduced model--without fW
> model3 = aov(Time ~ fA + fB + fA:fB, data=uav3.data)
> summary(model3)

            Df  Sum Sq Mean Sq F value  Pr(>F)
fA           1   215.3   215.3    69.0 4.9e-09
fB           1   306.3   306.3    98.2 1.2e-10
fA:fB        1   205.0   205.0    65.7 8.0e-09
Residuals   28    87.4     3.1

> # Multiple comparisons
> lsmA.aov2 = lsmeans(model3, ~ fA)
> summary(contrast(lsmA.aov2, method="pairwise"),
+         infer=c(T,T), side="two-sided")
> lsmAB.aov2 = lsmeans(model3, ~ fA:fB)
> summary(contrast(lsmAB.aov2, method="pairwise", adjust="tukey"),
+         infer=c(T,T), side="two-sided")
```

causes pooling of *ms(W(A))* and *msE* (i.e. msE_W and msE_S) to estimate σ_S^2 with more degrees of freedom. One benefit should be increased power for some tests—especially tests of effects compared between whole plots. It would also provide a common variance estimator for multiple comparisons of *AB*-treatment combinations, making most if not all methods of multiple comparisons applicable. Moreover, the restricted maximum likelihood approach becomes preferable to the analysis of variance approach for unbalanced designs.

19.9.4 Recovery of Inter-block Information

The oats experiment, introduced in Sect. 19.3.4, involves a split-plot design with complete blocks at both the whole- and split-plots levels—namely, the levels of A assigned to whole plots within blocks constitute a randomized complete block design, and the levels of B assigned to split plots within whole plots constitute a randomized complete block design. With this dual complete block structure, the data analysis is the same whether the block effects are modeled as fixed (as in Sect. 19.3.4) or random (as in Sect. 19.9.4). Such would not be the case if for example the split-plot design involves incomplete blocks at the whole-plots level, as illustrated in this section.

Consider again the oats experiment and corresponding data in Table 19.3 (p. 710), but suppose one only has the data for levels 1 and 2 of A in blocks 3 and 5, for levels 0 and 2 of A in blocks 1 and 4, and for levels 0 and 1 of A in blocks 2 and 6. For this subset of the data, the levels of A assigned to whole plots in blocks constitute a balanced incomplete block design with blocks of size 2. The R program and output in Tables 19.28 and 19.29 provides two approaches to the analysis of these data.

In Table 19.28, the analysis of variance approach is used and block effects are modeled as fixed. This is a Type I analysis. Since block effects are modeled as fixed, unbiased estimates of main-effect-of-A contrasts must be *intra-block estimates*—namely, each is composed (by summing over blocks) of within-block contrasts of observations—as is necessary for the fixed block effects to cancel out to yield unbiased estimates. Analogously, for testing for main effects of A, the numerator of the F-statistic is the mean square for A adjusted for block effects, which can be computed from the sum of squares of an appropriate set of such intra-block estimates of main-effect-of-A contrasts. The corresponding data analysis is called the *intra-block analysis*. For sake of comparison with subsequent analyses, note that the pairwise comparisons with the control for A each have estimated standard error 11.288 with 4 degrees of freedom. The same results would be obtained by the restricted maximum likelihood approach using the following R code, except Type III tests would be provided, so the sum of squares for blocks would be adjusted for A.

```
library(lmerTest)
model2 = lmer(y ~ fBlock + fA + fB + fA:fB + (1|fWP:fBlock),
              data=oats2.data)
anova(model2)
```

The program and output are continued in Table 19.29, where the restricted maximum likelihood approach is used, but this time block effects are modeled as random. With random block effects, one can obtain unbiased estimates of main-effect-of-A contrasts using contrasts of block totals (the sum total of the observations in each block), and these so-called *inter-block estimates* are in addition to and independent of the intra-block estimates used in the first analysis. So, for any main-effect-of-A contrast, any fixed weighted average of the corresponding intra- and inter-block estimates provides an unbiased estimate of the contrast, and the best (minimum variance) estimate would be obtained when each weight is inversely proportional to the variance of the corresponding estimate. Unfortunately, this best weighting is unknown because, for any main-effect-of-A contrast, the variances of the intra- and inter-block estimates are different and depend on the unknown variance components. A common

Table 19.28 R program and output illustrating the intra-block analysis—the oats split-plot experiment

```
> oats.data = read.table("data/oats.txt", header=T)
> oats.data = within(oats.data, {fBlock = factor(Block);
+               fWP = factor(WP); fA = factor(A); fB = factor(B) })
> # Reduce data so level of A constitute a BIBD
> oats2.data = subset(oats.data,
+                       !(fA == 0 & (fBlock == 3 | fBlock == 5))
+                       & !(fA == 1 & (fBlock == 1 | fBlock == 4))
+                       & !(fA == 2 & (fBlock == 2 | fBlock == 6)) )
>
> # Intra-block analysis: ANOVA, fixed block effects
> options(contrasts = c("contr.sum", "contr.poly")) # For correct lsmeans
> model1 = aov(y ~ fBlock + fA + fB + fA:fB + Error(fWP:fBlock),
+              data=oats2.data)
> summary(model1)

Error: fWP:fBlock
          Df Sum Sq Mean Sq F value Pr(>F)
fBlock     5  12064    2413    3.16   0.14
fA         2    377     188    0.25   0.79
Residuals  4   3058     765

Error: Within
          Df Sum Sq Mean Sq F value  Pr(>F)
fB         3  15966    5322   34.23 2.4e-09
fA:fB      6   1361     227    1.46    0.23
Residuals 27   4198     155

> # Dunnett's Method for A: ANOVA approach
> library(lsmeans)
> lsmA1 = lsmeans(model1, ~ fA)
> set.seed(19831957)
> summary(contrast(lsmA1, method="trt.vs.ctrl", adjust="mvt"),
+          infer=c(T,T), level=0.99)

 contrast estimate     SE df lower.CL upper.CL t.ratio p.value
 1 - 0      6.0417 11.288  4  -54.508   66.592   0.535  0.8228
 2 - 0      7.4583 11.288  4  -53.092   68.008   0.661  0.7503

Confidence level used: 0.99
Conf-level adjustment: mvt method for 2 estimates
P value adjustment: mvt method for 2 tests
```

approach is to use the estimated variances to determine the weights, but random weights would make the resulting analyses approximate. If the weights are well chosen, the resulting combined estimate is better (has smaller variance) than the intra-block estimate, due to recovery of inter-block information.

Consider now the output in Table 19.29, obtained by recovering and using this inter-block information. For factor A, the F-test and corresponding p values have changed, as have the results of Dunnett's method comparing levels of A to the control level. Recovery of the inter-block information ought to provide contrast estimates that are more accurate, and indeed the standard errors for the treatment-versus-control comparisons are smaller than in the inter-block analysis. Also, the variance estimate for each main-effect-of-A contrast is essentially a composite variance estimate, so Satterthwaite's method is used to compute the number of denominator degrees of freedom (i.e. 4.95) for the F test of A,

Table 19.29 R program and output illustrating the recovery of inter-block information—the oats split-plot experiment

```
> # Recovering inter-block information: ReML, random block effects
> model3 = lmer(y ~ fA + fB + fA:fB + (1|fBlock) + (1|fWP:fBlock),
+               data=oats2.data)
> anova(model3)

Analysis of Variance Table of type III  with  Satterthwaite
approximation for degrees of freedom
      Sum Sq Mean Sq NumDF DenDF F.value  Pr(>F)
fA       150      75     2  4.95     0.5    0.64
fB     15966    5322     3 27.00    34.2 2.4e-09
fA:fB   1362     227     6 27.00     1.5    0.23

> # Dunnett's Method for A
> lsmA3 = lsmeans(model3, ~ fA)
> set.seed(19831957)
> summary(contrast(lsmA3, method="trt.vs.ctrl", adjust="mvt"),
+         infer=c(T,T), level=0.99)

 contrast estimate    SE   df lower.CL upper.CL t.ratio p.value
 1 - 0       3.516 10.71 4.95  -46.042   53.074   0.328  0.9257
 2 - 0      10.350 10.71 4.95  -39.208   59.908   0.966  0.5638
```

which is also equal to the number of degrees of freedom associated with the standard error for each main effect of A contrast. For these pairwise comparisons, each standard error is smaller with more degrees of freedom than for the intra-block analysis, so there is a helpful recovery of the inter-block information.

The restricted maximum likelihood approach was used above to recover the inter-block information for main-effect-of-A contrasts. One may not be able to accomplish this directly in R using the analysis of variance approach. Consider the following code, for example.

```
model4 = aov(y ~ fA + fB + fA:fB + Error(fBlock + fWP:fBlock),
             data=oats2.data)
summary(model4)
```

The `Error` function, used to enter random effects in the model, causes separate analyses to be conducted for each error strata. In this example, the error strata correspond to: comparisons within whole plots, that involve split-plot error; comparisons between whole plots within blocks, that involve split-plot and whole-plot error; and comparisons between blocks, that involve split-plot error, whole-plot error, and variation due to random block effects. The intra-block and inter-block estimates of main-effect-of-A contrasts involve two different error strata, so the above code using the analysis of variance approach generates two F tests for A—the intra-block test in Table 19.28 and an inter-block test (not shown)—but no composite test.

Turning attention to the split-plot comparisons, note that the F-tests for B, and likewise for AB, are the same in each analysis. This is the case because a randomized complete block design was used for the assignment of levels of B to split plots in whole plots (serving as blocks). Consequently, for each main effect of B and AB-interaction contrast, the corresponding contrast in treatment means is a within-whole-plot contrast, free of whole-plot or block effects, so no adjustment is needed.

In the above example, recovery of the inter-block information provided better information for analysis of the whole-plots factor A. For example, for each treatment-versus-control contrast estimate for A, recovery of inter-block information provided a tighter confidence interval, due to having both a smaller standard error and more error degrees of freedom. In planning such an experiment, the experimenter should determine in advance how the data will be analyzed. While one could plan on conducting the intra-block analysis, recovery of the inter-block information should usually be beneficial.

19.9.5 ReML and Unbalanced Data

In this section, we illustrate the restricted maximum likelihood approach using the linear mixed effects function `lmer` for analysis of an unbalance split-plot design, revisiting the mobile computing field study experiment introduced in Sect. 19.7. Recall, while the planned experiment was nicely balanced, it was discovered prior to analysis of the data that one of the subjects traversed one of the paths in the reverse direction, making the corresponding observation inappropriate to use in the analysis. As such, one of the 72 planned observations (given in Table 19.11, p. 719) is listed as missing. In Sects. 19.7.1 and 19.7.2, an approximate analysis of the split-plot design was provided by estimating the missing value and treating it as observed, using standard formulas for analysis of the balanced design via the analysis of variance approach, but adjusting degrees of freedom appropriately. With one observation lost, the data are no longer balanced, making the data analysis conceptually and computationally more difficult. However, the R function `lmer` handles this situation nicely, as we shall illustrate.

The R program and output for analysis of the 71-observation split-plot design via the restricted maximum likelihood approach is shown in Tables 19.30 and 19.31. The mixed model for the analysis, provided in Eq. (19.7.18), p. 718, was fit in Table 19.30 using the `lmer` function of the `lme4` (and `lmerTest`) package. The three variance components are estimated by restricted maximum likelihood. In this case, the three estimates are all positive, as shown in selected output from the `summary` command. Given these variance component estimates, the fixed effects are then estimated by generalized least squares.

The `anova` function then provides Type III F-tests for each fixed effect—Type III in the sense that the tests are based on estimates of effects and corresponding variability obtained by fitting the full model. By default, Satterthwaite's approximation is used to compute the denominator degrees of freedom for each F test, analogous to Satterthwaite's approximation for a balanced design. Regarding the two treatment factors, only the effects of visual presentation format (B) are significant, with $p = 0.01660$. Concerning the nuisance factors path and order, these involved both within- and between-whole-plot comparisons before the lost observation caused design imbalance. For each of these factors, tests are provided for each of three pseudofactor components, including for example tests for P_2, P_3 and P_2P_3 for path. At the top of Table 19.31, a second call of the `lmer` and `anova` functions fits a model using factors rather than pseudo-factors for order and path, providing a consolidated F-test for each of these nuisance factors. It is not surprising that path has significant effects, and these are attributable to the three distinct paths used as distinguished by P_3.

At the bottom of Table 19.31, the `lsmeans` function has generated the pairwise comparisons for each of factors A and B, including individual t-tests and individual 95% confidence intervals for each pairwise comparison. These results are similar to those obtained in Sect. 19.7.2, where the approach used was to estimate the missing value and analyze the balanced design, though there Tukey's method was used. Based on the individual 95% confidence intervals provided here, level 2 of B—the birds' eye view egocentric visual presentation format—has a significantly smaller *RMSE* from the intended path than do the other two visual presentation formats. It is interesting that the denominator degrees of freedom is not constant for the factor B pairwise comparisons, due to the imbalance created by the lost observation. This indicates that the pairwise comparisons do not all utilize the same variance estimator, though presumably nearly so. As a consequence, the Bonferroni method should perhaps be used if multiple comparisons for B are of interest.

Table 19.30 R program and output for the mobile computing field study experiment

```
> MCFS71.data = read.table("data/MCFS71.txt", header=T)
> head(MCFS71.data, 3)

  Subj Day Order O2 O3 Path P2 P3 A B      y
1    1   1     1  0  0    1  0  0 1 1 49.321
2    1   1     2  0  1    2  0  1 1 2 24.386
3    1   1     3  0  2    3  0  2 1 3 37.680

> MCFS71.data = within(MCFS71.data, {fSubj = factor(Subj);
+        fDay = factor(Day); fOrder = factor(Order); fO2 = factor(O2);
+        fO3 = factor(O3); fPath = factor(Path); fP2 = factor(P2);
+        fP3 = factor(P3); fA = factor(A); fB = factor(B) })

> # ReML
> library(lmerTest)
> model1 = lmer(y ~ fO2 + fP2 + fA + fO3 + fO2:fO3 + fP3 + fP2:fP3
+                   + fB + fA:fB + (1|fSubj) + (1|fSubj:fDay),
+                   data=MCFS71.data)
> summary(model1)

...
Random effects:
 Groups     Name        Variance Std.Dev.
 fSubj:fDay (Intercept) 28.2     5.31
 fSubj      (Intercept) 64.0     8.00
 Residual               93.0     9.64
...

> anova(model1)

Analysis of Variance Table of type III  with  Satterthwaite
approximation for degrees of freedom
        Sum Sq Mean Sq NumDF DenDF F.value  Pr(>F)
fO2         23      23     1   8.5    0.25 0.63172
fP2        174     174     1   8.5    1.87 0.20643
fA          17      17     1   8.5    0.19 0.67688
fO3          8       4     2  34.6    0.04 0.95807
fP3       1976     988     2  34.6   10.63 0.00025
fB         860     430     2  34.6    4.62 0.01660
fO2:fO3     69      35     2  34.6    0.37 0.69270
fP2:fP3     26      13     2  34.6    0.14 0.87176
fA:fB      146      73     2  34.6    0.79 0.46336
```

One could also generate pairwise comparisons for the *AB* combinations by using `A:B` as the effect in an `lsmeans` statement, as follows.

Table 19.31 R program and output, continued, for the mobile computing field study experiment

```
> # Consolidating tests for Order and Path
> model2 = lmer(y ~ fOrder + fPath + fA + fB + fA:fB + (1|fSubj)
+                        + (1|fSubj:fDay), data=MCFS71.data)
> anova(model2)

Analysis of Variance Table of type III  with  Satterthwaite
approximation for degrees of freedom
       Sum Sq Mean Sq NumDF DenDF F.value Pr(>F)
fOrder     98      20     5  31.1    0.21 0.9555
fPath    2184     437     5  31.1    4.70 0.0026
fA         17      17     1   8.5    0.19 0.6769
fB        860     430     2  34.6    4.62 0.0166
fA:fB     146      73     2  34.6    0.79 0.4634


> # Multiple comparisons
> library(lsmeans)
> lsmA = lsmeans(model1, ~ fA)
> summary(contrast(lsmA, method="pairwise", adjust="none"),
+         infer=c(T,T), level=0.95)

 contrast estimate     SE   df lower.CL upper.CL t.ratio p.value
 1 - 2     -1.3631 3.1606 8.98  -8.5149   5.7887  -0.431  0.6764

> lsmB = lsmeans(model1, ~ fB)
> summary(contrast(lsmB, method="pairwise", adjust="none"),
+         infer=c(T,T), level=0.95)

 contrast estimate     SE    df lower.CL upper.CL t.ratio p.value
 1 - 2      8.1957 2.8360 35.39   2.4406 13.95082   2.890  0.0065
 1 - 3      1.6903 2.7837 35.02  -3.9608  7.34143   0.607  0.5476
 2 - 3     -6.5054 2.8360 35.39 -12.2605 -0.75034  -2.294  0.0278
```

```
lsmAB = lsmeans(model1, ~ fA:fB)
summary(contrast(lsmAB, method="pairwise", adjust="bon"),
        infer=c(T,T), level=0.95)
```

However, only the Bonferroni method of multiple comparisons would be applicable, since these comparisons involve a mix of with-and between-day comparisons, even before an observation was lost, so there would not be a common variance estimator for all of these pairwise comparisons. Also, the corresponding variance estimators would generally be composite variance estimators, in which case Satterthwaite's approximation would be used.

Exercises

1. **Drug experiment**

   An experiment designed as a split-plot design was described by W.M. Wooding in the *Journal of Quality Technology* in 1973. The experiment concerned the evaluation of eight drugs (factor A at $a = 8$ levels) for the treatment of arthritis. A second factor was the dose of the drug (factor B at $b = 2$ levels), and the third factor was the length of time (factor C at $c = 2$ levels) that a measurement was taken after injection by a substance known to cause an inflammatory reaction. The

Table 19.32 Fluid (milliliters) in pleural cavity for the drug experiment

Block I										
Whole Plot	Dose B	Time C	Drug A 1	2	3	4	5	6	7	8
1	1	1	5.7	8.6	6.9	6.6	6.7	7.4	5.7	6.7
2	1	2	8.4	9.6	9.3	11.1	12.5	8.7	9.3	9.5
3	2	1	5.1	7.2	6.8	6.4	6.6	8.7	6.7	7.0
4	2	2	7.3	8.7	7.9	6.9	8.9	9.5	8.3	11.3
Block II										
Whole Plot	Dose B	Time C	Drug A 1	2	3	4	5	6	7	8
5	1	1	5.8	6.8	7.0	8.5	7.8	7.3	6.4	8.5
6	1	2	9.1	10.8	6.9	12.2	9.9	10.4	10.6	10.5
7	2	1	5.4	7.9	8.0	6.4	8.4	7.1	6.4	7.2
8	2	2	5.3	10.4	8.2	8.1	10.9	9.8	8.4	14.6

Source Wooding (1973). Reprinted with Permission from Journal of Quality Technology © 1973 ASQ, www.asq.org

experimental units used in the study were $n = 64$ rats. The response was the amount of fluid (in milliliters) measured in the pleural cavity of an animal after having been administered a particular treatment combination.

In many pharmacological studies, time of day has an effect on the response due to changing laboratory conditions, etc. Consequently, the experiment was divided into blocks, whole plots and split plots. The blocks were of size 32, each set of 32 observations being measured on a single day. Each treatment combination was measured once per day. Each day was then subdivided into 4 whole plots of size 8, where the eight measurements within a whole plot were taken fairly close together in time.

Since the effect of the drug (A) was of primary importance, and since the effects of B and C were of interest only in the form of an interaction with A, the main effects of B and C and the BC interaction were confounded with the whole plots. The data are shown in Table 19.32, and the experimenter used the logarithms of the data in his analysis. Notice that the design for A on the split plots is a randomized block design, and the design for the BC combinations on the whole plots is also a randomized block design.

(a) Write out a model for this experiment.
(b) Calculate an analysis of variance table using the logarithms of the data. Distinguish between the effects measured on the whole plots and those measured on the split plots. Identify the whole-plot error and split-plot error.
(c) Test any hypotheses of interest and state your conclusions clearly.
(d) Examine interaction plots of any important interactions. Calculate a set of 95% confidence intervals for the differences between pairs of drugs. State your conclusions.

2. **Fishing line experiment**

The fishing line experiment was run by C. Reynolds, B. Grunden, and K. Taylor in 1996 in order to compare the strengths of two brands of fishing line exposed to two different levels of stress. Two different reels of fishing line were purchased for each of the two brands, and sections of line were cut from each reel. Thus the reels were automatically assigned to the levels of factor A (Brand), and constituted the four whole plots. There were no blocks in this experiment. The split plots constituted sixteen sections of line, four cut from each of the four reels (that is, 16 split plots in total, 4 per whole plot). The split plots were randomly assigned to two different stress levels (stressed, “S”;

Table 19.33 Strength of line for the fishing line experiment

Whole plot (Reel)	*A* (Brand)	Level of *B* (weight)			
1	1	N (6.70)	S (6.40)	S (7.20)	N (7.00)
2	2	S (8.10)	S (8.90)	N (8.00)	N (6.10)
3	2	S (8.00)	S (8.00)	N (8.75)	N (8.50)
4	1	N (8.50)	S (9.50)	N (9.70)	S (9.40)

nonstressed, "N") so that each stress level was assigned two split plots per whole plot. The stress was induced by hanging a brick from the assigned section of line for 14 hours. Although this did not precisely mimic the stress induced during fishing, it was still expected to give some information about the strength of the lines under stress.
The strength test was accomplished by hanging a bucket on the end of the line, which was suspended from a beam. The bucket was gradually filled with water through a small hole in the lid until the line broke. The data are the resulting weights of water to the nearest 0.01 lb and are shown in Table 19.33.

(a) Write down a model for this experiment.
(b) Construct an analysis of variance table and test the hypotheses that you think are of interest. State your conclusions.
(c) If you were to repeat this experiment, suggest ways in which you would try to improve it.

3. **Cigarette experiment**

The cigarette experiment was run by J. Edwards, H. Hwang, S. Jamison, J. Kindelberger, and J. Steinbugl in 1996 in order to determine the factors that affect the length of time that a cigarette will burn. There were three factors of interest:

- "Tar" (factor *A*) at two levels, "regular" and "ultra-light,"
- "Brand" (factor *B*) at two levels, "name brand" and "generic brand" (coded 1 and 2),
- "Age" (factor *C*) at three levels, "fresh," "24 hour air exposure," "48 hour air exposure."

The cigarettes were to be burned in whole plots of size six. This was to help with the difficulty of recording burning times and to help keep the amount of smoke in the room at a reasonable level.

Table 19.34 Burning times for the cigarette experiment

Whole plot (Time)	*A* (Tar)	Levels of *BC* (Burning times in seconds)					
1	1	22 (301)	11 (326)	23 (260)	13 (290)	12 (312)	21 (292)
2	2	11 (329)	12 (331)	13 (285)	21 (306)	22 (258)	23 (276)
3	2	22 (290)	11 (380)	12 (335)	13 (309)	23 (243)	21 (334)
4	2	11 (321)	21 (337)	23 (275)	12 (316)	13 (307)	22 (250)
5	2	22 (308)	11 (345)	21 (307)	23 (288)	13 (321)	12 (330)
6	1	11 (344)	23 (283)	21 (281)	22 (261)	13 (307)	12 (292)
7	1	21 (274)	13 (310)	12 (304)	22 (279)	23 (277)	11 (330)
8	1	13 (302)	12 (325)	22 (301)	11 (338)	23 (270)	21 (297)
9	2	12 (323)	13 (334)	23 (265)	11 (326)	22 (269)	21 (297)
10	1	23 (309)	13 (314)	22 (259)	11 (344)	21 (310)	12 (322)

There were ten whole plots, and these were assigned at random to the tar levels so that each tar level was assigned five whole plots.
The six split plots (time slots) in each whole plot were assigned at random to the six brand/age treatment combinations. Marks were made across the seam of each cigarette at a given distance apart. Each cigarette was lit at the beginning of its allotted time slot, and the time taken to burn between the two marks was recorded. The data are shown in Table 19.34.

(a) Write down a model for this experiment.
(b) Construct an analysis of variance table and test the hypotheses that you think are of interest. State your conclusions.
(c) Use Tukey's method to examine the pairwise differences in the effects on burning time of the six BC treatment combinations (averaged over tar levels).
(d) Examine the linear and quadratic trends of burning time due to different ages, for each brand separately.
(e) State your conclusions about the experiment, including your choice of overall significance levels and overall confidence levels.

4. **Injection molding experiment, continued**

The injection molding experiment, introduced in Exercise 10 of Chap. 15, was run in order to examine the effect of six factors on the shrinkage of a part produced by injection molding. The six factors were injection velocity (factor A), cooling time (factor B), barrel zone temperature (factor C), mold temperature (factor D), hold pressure (factor E), and back pressure (factor F). Each factor had two levels coded 0 and 1. The treatment combinations, which are shown in Table 15.56, p. 555, were not completely randomized. The levels of factor D were time-consuming to change, so for the first four observations D was held at its low level and the combinations of the other factors were randomly ordered. For the last four observations, D was held at its high level and the combinations of the other factors were randomly ordered. Thus we can think of this design as having two whole plots assigned to the two levels of D, and having four split plots nested within each whole plot assigned at random to combinations of levels of A, B, C, E, and F. Table 19.35 contains the corresponding split plot design and mean lengths. Let σ_W^2 and σ_S^2 denote the variances of the random whole-plot and split-plot effects, respectively.

(a) Which fixed effects can be estimated? Compute their estimates.
(b) Which variance components can be estimated? Compute their estimates.
(c) Is it possible to analyze this experiment in a sensible way? If so, present your results.

Table 19.35 Mean lengths for the injection molding experiment

Whole plot	Split plot	A	B	C	D	E	F	Mean length
1	1	0	0	0	0	0	0	2
	2	0	1	1	0	0	1	46
	3	1	0	1	0	1	0	108
	4	1	1	0	0	1	1	128
2	1	0	0	0	1	1	1	73
	2	0	1	1	1	1	0	105
	3	1	0	1	1	0	1	53
	4	1	1	0	1	0	0	58

Table 19.36 Split-plot design and times (in seconds) for the UAV experiment: situation awareness comprehension

B (Similarity)		1				2			
C (Complexity)		1	2	3	4	1	2	3	4
A (Cue)	W (Subject)								
1	1	20	21	27	32	23	21	30	28
	2	24	19	33	35	24	23	32	30
	3	19	22	28	33	21	20	34	31
	4	30	27	37	30	26	24	31	33
	5	25	24	35	38	28	27	33	30
	6	22	20	40	41	26	25	29	37
	7	28	23	34	36	28	26	37	34
	8	21	26	33	34	24	28	34	40
2	1	8	9	11	9	7	7	8	8
	2	10	8	13	13	12	8	18	10
	3	11	9	15	14	9	10	15	18
	4	8	9	12	18	8	12	8	9
	5	7	10	17	13	12	12	12	17
	6	9	7	13	12	10	7	12	10
	7	11	12	10	11	13	8	10	8
	8	10	11	15	13	7	13	9	8

Source Mahadevan (2009).

5. **UAV experiment, continued**

The UAV experiment, introduced in Sect. 19.6, was run using a $2^2 \times 4$ split-plot design with 16 subjects (W) as whole plots, with two levels of cue condition (A) assigned to whole-plots, and with 2×4 combinations of levels of task similarity (B) and task complexity (C) assigned to split plots. Another response variable measured by the experimenter was the time taken to perform situation awareness comprehension tasks in the primary task, yielding the data shown in Table 19.36. Using model (19.6.16), p. 715, conduct the following analyses.

(a) Construct an analysis of variance table. For each main effect and interaction in the treatment factors, determining which effects are significant at the 1% level.
(b) Construct a 95% confidence interval for the main effect of cue, and interpret the results.
(c) If the three-factor interaction is significant, then compare the effect of cue at each combination of the other two factors. Otherwise, if the factor cue interacts significantly with either of the other treatment factors, then compare the effect of cue at each level of each factor with which cue interacts significantly. Use individual 99% confidence intervals.

6. **UAV switch experiment, continued**

The UAV switch experiment, introduced in Sect. 19.8.3, was run using a 2×2 split-plot design with 16 subjects as whole plots, with two levels of alert type (A) assigned to whole-plots, and with two levels of complexity (B) assigned to split plots. Main effects and interactions for A and B were all significant, but the primary interest is in comparing the effects of the levels of A. Use the approach analogous to the first call of `PROC GLM` in Table 19.19 to do the following.

(a) Construct a 95% confidence interval for the main effect of alert type, and interpret the results.
(b) For each level of complexity, estimate the difference in effects of the two levels of alert type. Determine the variance of each of the two corresponding pairwise comparison estimators,

Table 19.37 Alert detection times (in seconds) for UAV switch experiment

A (Alert type)	B (Complexity)	W (Subject)							
		1	2	3	4	5	6	7	8
1	1	5	6	4	6	5	5	5	6
	2	10	10	6	8	6	6	6	6
2	1	6	5	4	5	5	4	6	6
	2	5	6	4	5	5	4	6	5

Source Mahadevan (2009).

compute the corresponding estimated standard errors, and use Satterthwaite's approximation to compute the approximate number of degrees of freedom associated with each estimated standard error.

(c) Using the results of part (b), construct individual 99% confidence intervals for the difference in effects of the two levels of alert type for each level of complexity. Interpret the results.

7. **UAV switch experiment, continued**

The UAV switch experiment, introduced in Sect. 19.8.3, was run using a 2×2 split-plot design with 16 subjects as whole plots, with two levels of alert type (A) assigned to whole-plots, and with two levels of complexity (B) assigned to split plots. Another response variable measured by the experimenter was the alert detection time (in seconds), yielding the data shown in Table 19.37. Using model (19.8.20), p. 734, conduct the following analyses.

(a) Construct an analysis of variance table. For each main effect and interaction in the treatment factors, determining which effects are significant at the 1% level.
(b) Construct a 95% confidence interval for the main effect of alert type, and interpret the results.
(c) For each level of complexity, compare the effects of the two levels of alert type, using individual 99% confidence intervals.

8. **Mobile Computing Field Study, continued**

The Mobile Computing Field Study, introduced in Sect. 19.7, was run using a 2×3 split-plot design with 12 subjects as blocks, 24 days as whole plots, three runs per day as split plots, two display types (A) assigned to whole-plots, three visual presentation formats (B) assigned to split-plots, six runs per subject, and utilizing six paths. Another response variable measured by the experimenter was the time (in minutes) to navigate the path, yielding the data shown in Table 19.38. Using model (19.7.18), p. 718, conduct the following analyses, estimating the missing value and utilizing it in the analysis.

(a) Estimate the missing value.
(b) Construct an analysis of variance table analogous to Table 19.12. For each main effect and interaction in the treatment factors, determining which effects are significant at the 1% level.
(c) Compare the visual presentation formats using simultaneous 99% confidence intervals and Tukey's method. Interpret the results.

Table 19.38 Mobile computing field study: time (in minutes) for each subject (Subj), day, run order, and path-treatment combination (*PAB*), with one observation missing

Subj	Day 1						Day 2					
	Run 1		Run 2		Run 3		Run 4		Run 5		Run 6	
	PAB	Time	*PAB*	Time	*PAB*	Time	*PAB*	Time	*PAB*	Time	*PAB*	Time
1	111	15.633	212	11.717	313	13.550	421	11.900	522	11.817	623	16.517
2	213	13.767	311	13.867	112	11.233	523	10.533	621	11.966	422	10.633
3	312	11.150	113	13.167	211	12.633	622	9.367	423	10.350	521	9.750
4	121	11.867	222	8.600	323	8.700	411	10.400	512	9.117	613	10.067
5	223	10.183	321	11.600	122	8.350	513	16.667	611	9.783	412	7.600
6	322	7.650	123	9.667	221	12.417	612	11.400	413	14.900	511	10.350
7	421	9.233	522	8.700	623	12.083	111	10.700	212	7.833	313	9.050
8	523	11.130	621	8.717	422	7.967	213	8.867	311	8.917	112	6.667
9	622	12.133	423	12.800	521	12.433	312	11.767	113	12.367	211	12.983
10	411	18.150	512	10.483	613	13.370	121	14.100	222	10.217	323	11.583
11	513	10.883	611	15.117	412	10.833	223	16.783	321	15.333	122	12.483
12	612	9.567	413	16.450	511	11.217	322	.	123	10.567	221	10.517

Source Wesler (2001). Copyright © 2001 Mary Mc. Wesler. Research was performed under U.S. Army Research Laboratory, Federated Laboratory Research Consortium (DAAL01-96-0003) directed by Mr. Bernie Corona. Reprinted with permission

9. **Mobile Computing Field Study, continued**

The Mobile Computing Field Study was introduced in Sect. 19.7, and data on an additional response variable, time, was provided in Exercise 8. Using model (19.7.18), p. 718, conduct the following analyses of the response variable time, using only the 71 observations available—namely, without estimating and using the missing value. If using SAS software, you may adapt the program in Table 19.21; if using R, you may adapt the program in Tables 19.30 and 19.31.

(a) Provide the resulting variance component estimates.
(b) Test for significance of each treatment factor main effect and interaction. Report the observed significance level of each test, and interpret the results.
(c) Using an appropriate method of multiple comparisons, construct simultaneous 99% confidence intervals for pairwise comparison of the visual presentation formats (B). Interpret the results.