12 Designs with Two Blocking Factors

12.1 Introduction

In Chaps. 10 and 11, we discussed designs for experiments involving a single system of blocks. However, as we saw in the randomized complete block design of Table 10.1, p. 308, a block label can represent a combination of levels of several factors. The design in Table 10.1 was presented as having six blocks—the six block labels being the six combinations of levels of the factors "run of the oven" and "shelf." When a block design is used in this way, the $b - 1$ degrees of freedom for the block effects include not only those degrees of freedom for the effects of the two factors, but also for their interaction.

In this chapter, we look at designs for experiments that involve two blocking factors that *do not* interact. When the blocking factor interactions can be omitted from the model, fewer observations are needed, and the experiment can be designed with only one observation per combination of levels of the blocking factors. The plan of the design is then often written as an array (that is, a table) with the levels of one blocking factor providing the row headings and those of the other providing the column headings. These designs are called *row–column designs*, and the two sets of blocks are called row blocks and column blocks or, more simply, rows and columns. Such designs are described in Sect. 12.2, where we restrict discussion to the case of exactly one treatment label allocated to each cell of the table. The randomization procedure needed for row–column designs is given in Sect. 12.2.1.

In Sect. 12.2.2, Latin square designs are described. These are the two-dimensional counterparts of randomized complete block designs since the row blocks are complete blocks and the column blocks are also complete blocks. Latin square designs require that the numbers of levels of both blocking factors be the same as (or a multiple of) the number of treatments. Section 12.2.3 concerns Youden designs, in which the column blocks form a randomized block design and the row blocks form a balanced incomplete block design (or vice versa). Cyclic and other row–column designs are discussed in Sect. 12.2.4.

The standard row–column–treatment model for row–column designs is shown in Sect. 12.3, together with an overview of the analysis of variance and confidence intervals for general row–column designs. The analysis becomes greatly simplified for both Latin square designs and Youden designs, and this is shown in Sects. 12.4 and 12.5, respectively. Checks on the model assumptions are described briefly in Sect. 12.6, and an extension of the model to cover factorial experiments in row-column designs is given in Sect. 12.7. All row–column designs can be analyzed using statistical software (see Sects. 12.8 and 12.9 for analysis by SAS and R software).

A. Dean et al., *Design and Analysis of Experiments*,
Springer Texts in Statistics, DOI 10.1007/978-3-319-52250-0_12

Table 12.1 A row–column experimental plan with $b = 7$, $c = 4$, $v = 7$, $r = 4$

		Column blocks			
		I	II	III	IV
Row blocks	I	1	3	6	7
	II	2	4	7	1
	III	3	5	1	2
	IV	4	6	2	3
	V	5	7	3	4
	VI	6	1	4	5
	VII	7	2	5	6

Some experiments are designed with more than two blocking factors. Block designs with a single system of blocks can still be used where each block represents some combination of the levels of three or more blocking factors, but designs with more than two blocking factors and a single observation per combination of their levels are not discussed in this book.

12.2 Design Issues

12.2.1 Selection and Randomization of Row–Column Designs

In most row–column designs and, in particular, in all Latin square designs and Youden designs, all treatment contrasts are estimable. However, estimability is not as easy to verify for general row-column designs as it was for incomplete block designs (Sect. 11.2.3) since it cannot be deduced from the row blocks and column blocks separately. One way to check that a miscellaneous row–column design is suitable for a planned experiment is to enter a set of hypothetical data for the design into a computer package and see whether the required contrasts are estimable.

Once the numbers of rows, columns, and treatments have been selected, there are two stages to designing an experiment with two blocking factors and one observation at each combination of their levels. The first design stage is to choose an experimental plan. As an example, the plan in Table 12.1 has two sets of blocks corresponding to two blocking factors—one with 7 levels represented by the $b = 7$ rows, and the other with 4 levels represented by the $c = 4$ columns. There are $v = 7$ treatment labels, each appearing $r = 4$ times in the design, and at most once per row and per column. The second design stage is the random assignment of the labels in the design to the levels of the treatment factors and blocking factors in the experiment, as follows:

(i) The row block labels in the design are randomly assigned to the levels of the first blocking factor.
(ii) The column block labels in the design are randomly assigned to the levels of the second blocking factor.
(iii) The treatment labels in the design are randomly assigned to the levels of the treatment factor.

Since there is only one experimental unit in each cell, there is no need for random assignment of experimental units to treatment labels within a cell.

Table 12.2 Latin squares with $b = c = v = 4$

Square 1	Square 2	Square 3	Square 4
A B C D	A B C D	A B C D	A B C D
B C D A	B A D C	B A D C	B D A C
C D A B	C D A B	C D B A	C A D B
D A B C	D C B A	D C A B	D C B A

12.2.2 Latin Square Designs

A $v \times v$ *Latin square* is an arrangement of v Latin letters into a $v \times v$ array (a table with v rows and v columns) in such a way that each letter occurs once in each row and once in each column. For example, the following 3×3 array is a 3×3 Latin square:

$$\begin{array}{ccc} A & B & C \\ B & C & A \\ C & A & B \end{array}$$

A *Latin square design* has v treatment labels, and v^2 experimental units arranged in v row blocks and v column blocks, where experimental units within each row block are alike, units within each column block are alike, and units not in the same row block or column block are substantially different. The experimental plan of the design is a $v \times v$ Latin square. Randomization of row block, column block, and treatment labels in the plan is carried out as in Sect. 12.2.1.

If we look only at the row blocks of a Latin square design, ignoring the column blocks, we have a randomized complete block design, and if we look at the column blocks alone, ignoring the row blocks, we also have a randomized complete block design. Each level of the treatment factor is observed $r = v$ times—once in each row block and once in each column block.

The 3×3 Latin square shown above is a "standard, cyclic Latin square." A Latin square is a *standard Latin square* if the letters in the first row and in the first column are in alphabetical order, and it is *cyclic* if the letters in each row can be obtained from those in the previous row by cycling the letters in alphabetical order (cycling back to letter A after the vth letter).

There is only one standard 3×3 Latin square, but there are four standard 4×4 Latin squares, and these are shown in Table 12.2. The first square is the cyclic standard Latin square. A standard cyclic Latin square exists for any number of treatments.

An example of a 6×6 Latin square design was shown in Table 2.5 (p. 19). It was a design that was considered for the cotton-spinning experiment. The row blocks represented the different machines with their attendant operators, and the column blocks represented the different days over which the experiment was to be run. The treatment labels were the combinations of degrees of twist and flyer used on the cotton-spinning machines. After careful consideration, the experimenters decided not to use this design, since it required the same six machines to be available over the same six days, and this could not be guaranteed in the factory setting.

Latin square designs are often used in experiments involving subjects, especially where the subjects are allocated a sequence of treatments over time and where the time effect is thought to have a major effect on the response. For example, in an experiment to compare the effects of v drugs, the rows of the Latin square might correspond to v subjects to whom the drugs are given, and the columns might correspond to v time periods, with each subject receiving one drug during each time period. An experiment of this type, in which each subject receives a sequence of treatments, is called a *crossover experiment*.

In a crossover experiment, it is possible that a treatment will continue to have some effect on the response when another treatment is administered to the same subject in a subsequent time period. Such an effect is called a *carryover effect* or *residual effect* and must be accounted for in the design and analysis of the experiment. This is outside the scope of this book, but for further information, see Ratkowsky et al. (1993) or Jones and Kenward (2003). We will consider experiments in which either there is no carryover effect or in which the gap between the administration of one treatment and the next is sufficient to allow the carryover effect to diminish (and, hopefully, to disappear). The "gap" is called a *washout period*.

To design the experiment, a square is selected from the list of standard squares and the randomization procedure of Sect. 12.2.1 is performed. Thus for an experiment with $b = 4$ levels of the row blocking factor, $c = 4$ levels of the column blocking factor, and $v = 4$ levels of the treatment factor, one of the four standard squares of Table 12.2 would be selected. For larger squares, we have not provided a list of standard squares, since there are so many. However, it is straightforward to write down the standard cyclic Latin square and perform the randomization procedure on this.

Although we have not given a procedure which selects a Latin square at random from the set of all possible Latin squares, the above procedure should be sufficient for the occasional experiment. Fisher and Yates (1973) give more information on this, as well as a complete list of standard squares for $v = 5$ and 6, and selected squares for $v = 7$–12.

Replication of Latin Squares

A design based on a single Latin square has $r = v$ observations on each treatment, which may not be adequate for an analysis of the treatment effects. One way of increasing the number of observations is to piece together a number, s say, of $v \times v$ Latin squares. We will call such a design an *s-replicate Latin square*. Use of an s-replicate Latin square requires the column (or row) blocks to be of size vs. For example, stacking two 3×3 Latin squares, one above the other, we can obtain two possible 2-replicate Latin squares as in Table 12.3. For either plan, the number of observations per treatment is now $r = 6 = 2v$ rather than only $r = 3 = v$. Plan 2 is preferable to Plan 1 because the row blocks consist of each possible ordering of the three treatments (and this will remain true even after randomization).

Suppose we were to use Plan 2 for an experiment to compare the efficacy of $v = 3$ drugs, where there are two blocking factors, say "subjects"—the people to whom the drugs will be administered, and "time period"—the time during which each subject receives a single drug and a response is measured. If rows correspond to subjects, then $b = 6$ subjects would be required over $c = 3$ time periods, but if columns correspond to subjects, then each of $c = 3$ subjects would stay in the study for $b = 6$ periods. In practice, in drug studies, subjects are rarely used for more than 4 time periods, since the subject drop-out rate tends to be high after this length of time.

An s-replicate Latin square may allow some column-treatment interaction contrasts to be measured (after adjusting for row block effects as in Chap. 11). However, these will often not form a full set of

Table 12.3 Two 2-replicate Latin squares for $v = 3$, $b = 6$, $c = 3$

Plan 1			Plan 2		
A	B	C	A	B	C
B	C	A	B	C	A
C	A	B	C	A	B
A	B	C	A	C	B
B	C	A	B	A	C
C	A	B	C	B	A

Table 12.4 Two Youden squares for $v = 4$, $b = 4$, $c = 3$

Plan 3			Plan 4		
A	B	C	A	C	B
B	C	D	B	D	C
C	D	A	C	A	D
D	A	B	D	B	A

$(v-1)^2$ orthogonal contrasts. For $v = 3$, Plan 2 does allow the adjusted column-treatment interaction to be measured based on the full set of $(v-1)^2 = 4$ degrees of freedom, but Plan 1 does not.

12.2.3 Youden Designs

A $v \times c$ *Youden square* is an arrangement of v Latin letters into a $v \times c$ array (with $c < v$) in such a way that each letter occurs once in each column and at most once in each row. In addition, every pair of treatments occurs in the same number of rows, so the columns form complete blocks and the rows form a balanced incomplete block design.

Notice that a Youden square, written in this way, is not, in fact, square! We have defined a Youden square to have more rows than columns, but the array could be turned so that there are more columns than rows. Plans 3 and 4 in Table 12.4 are examples of 4×3 Youden squares with $v = 4$ treatment labels, with Plan 3 being a standard Youden square.

In general, a *Youden design* has v treatment labels and vc experimental units arranged in $b = v$ row blocks and $c < v$ column blocks, where experimental units within each row block are alike, experimental units within each column block are alike, and experimental units not in the same row block or column block are substantially different. The experimental plan is a $v \times c$ Youden square, and randomization of row, column, and treatment labels is carried out as in Sect. 12.2.1. Each level of the treatment factor is observed $r = c$ times—once in each column and at most once in each row.

The plans of Table 12.4 are both *cyclic Youden designs* prior to randomization. In each case, the row blocks form a cyclic balanced incomplete block design with every pair of treatment labels occurring in $\lambda = 2$ rows. Any cyclic balanced incomplete block design (with the full set of v blocks) can be used as a cyclic Youden design. The design in Table 12.1 is also a cyclic Youden design, having $b = v = 7$ row blocks and $c = 4$ column blocks and every pair of treatments occurring in $\lambda = 2$ row blocks. Youden designs with $c = v - 1$ column blocks can be obtained by deleting any column from a $v \times v$ Latin square design.

Replication of Youden Squares

We can obtain $r = cs$ observations per treatment by stacking s Youden squares one above another (giving $b = vs$ row blocks and $c(<v)$ column blocks). This results in large column-block sizes but may allow for estimation of some column$\times$treatment interaction contrasts after adjusting for row block effects (although usually not the full set of $(c-1)(v-1)$ contrasts). The row blocks still form a balanced incomplete block design, and the column blocks still form a complete block design.

Suppose $v = 7$ drugs are to be compared and that a number of subjects are each to be given a sequence of 4 of the drugs over $c = 4$ time periods, with washout periods in between to avoid carryover effects. The experimental plan in Table 12.1, which uses 7 subjects, would be suitable for the experiment if $r = 4$ observations per drug is deemed adequate. Otherwise, several copies of this design could be stacked. Or column permutations could be made before stacking (cf. stacking the two plans in Table 12.4). If, for example, $s = 3$ copies of the Table 12.1 design are pieced together, the resulting 3–replicate Youden design would require 21 subjects, 4 time periods, and would have $r = cs = 12$

Table 12.5 Incomplete-row complete-column designs with $v = b = 8$ and $r = c = 3$

	Plan 5				Plan 6		
	I	II	III		I	II	III
I	1	2	4	I	1	2	7
II	2	3	5	II	2	4	1
III	3	4	6	III	3	7	6
IV	4	5	7	IV	4	8	2
V	5	6	8	V	5	6	8
VI	6	7	1	VI	6	3	5
VII	7	8	2	VII	7	1	3
VIII	8	1	3	VIII	8	5	4

observations per treatment, with $b = 21$, $c = 4$, $v = 7$, and $\lambda = 6$. The design would be randomized before use as described in Sect. 12.2.1.

12.2.4 Cyclic and Other Row–Column Designs

Any arrangement of v treatment labels into b rows and c columns can be used as a row–column design for an experiment with two blocking factors. As with incomplete block designs, some row–column designs are better than others. The better designs (in terms of average length of confidence intervals for pairwise comparisons) have every pair of treatment labels occurring the same, or nearly the same, number of times in the row blocks and also in the column blocks. This is satisfied by Latin square designs, Youden designs, and also some cyclic designs.

A *cyclic row–column design with complete column blocks* is a row–column design in which the row blocks form a cyclic block design and the column blocks are complete blocks. For example, the experimental plan in Table 12.1 is a cyclic row–column design. The class of cyclic row–column designs is very large, and a design with $b = vs$ rows can always be found when a Youden design does not exist. For example, consider an experiment for comparing $v = 8$ treatments when there are two blocking factors having $b = 8$ and $c = 3$ levels, respectively (so $r = c = 3$). There does not exist a Youden design for $b = v = 8$ and $c = 3$, because there does not exist a balanced incomplete block design for 8 treatments in 8 blocks of size $c = 3$. (Notice that (11.3.1) gives $\lambda = r(c - 1)/(v - 1) = 6/7$ which is not an integer.) A cyclic design that may be suitable for the experiment is given as Plan 5 in Table 12.5. The design is obtained by cycling the labels in the first row block. Treatment pairs (1, 5), (2, 6), (3, 7), and (4, 8) never occur together in a row block, but all other pairs of treatments occur in exactly one row block, so this should be a reasonably good design. The design should be randomized via the procedure in Sect. 12.2.1 before it is used.

Plan 6 of Table 12.5, which is neither a Youden design nor a cyclic design, could also be used, but we might guess that it is not quite as good for pairwise comparisons. Treatment 1, for example, appears in one block with each of treatments 3 and 4, in two blocks with 2 and 7, and not at all with treatments 5, 6, or 8. Thus, the treatments are not quite so evenly distributed in the row blocks of Plan 6 as they are of Plan 5. Nevertheless, Plan 6 was used for the exercise bike experiment described below, and will be analyzed using the SAS and R software in Sects. 12.8 and 12.9, respectively.

Example 12.2.1 `Exercise bicycle experiment`

Yuedong Wang, Dong Xiang, and Yili Lu conducted an experiment in 1992 at the University of Wisconsin to investigate the effects of exercise on pulse rate. The exercise was performed on a stationary

Table 12.6 Row–column design showing treatments and data for the exercise bicycle experiment

Day	Subject					
	Lu		Wang		Xiang	
1	112	45	212	25	221	18
2	222	27	211	20	212	32
3	212	40	222	23	112	28
4	221	17	112	32	122	24
5	211	30	111	36	222	20
6	122	29	221	13	121	20
7	111	34	121	18	211	25
8	121	21	122	22	111	34

bicycle that included both foot pedals and hand bars. The experiment involved three treatment factors each at two levels. These were "time duration of exercise" with levels 1 and 3 min (coded 1, 2), "exercise speed" with levels 40 and 60 rpm (coded 1, 2), and "pedal type" with levels foot pedal and hand bars (coded 1, 2).

The three experimenters served as the $b = 3$ subjects in the experiment, since they were interested in the effects of the different exercises on themselves rather than on a large population. Had the latter been the case, then the subjects would have been selected at random from the population of interest and regarded as "random effects" (see Chap. 17).

Each subject was assigned a different exercise on each of 8 different days. To minimize any carryover effect, there was a training period prior to the experiment, and there was at least one day of rest between observations. The subject's pulse was taken immediately after completion of the exercise, and the response variable was the time, in seconds, for 50 heart beats.

Plan 6 of Table 12.5 was used. The design, after randomization, and the corresponding data are shown in Table 12.6. The rows were randomly assigned to days in the order V, III, VI, VIII, VII, IV, I, II. The columns were randomly assigned to subjects in the order I = Lu, II = Wang, III = Xiang. The treatment labels were randomly assigned to treatment combinations in the order

1	5	2	4	7	6	8	3
111	112	121	122	211	212	221	222

The data for this experiment are analyzed by computer software in Sects. 12.8 and 12.9. □

12.3 Analysis of Row–Column Designs

12.3.1 Model for a Row–Column Design

For a row–column design with b rows and c columns, the row–column–treatment model is

$$Y_{hqi} = \mu + \theta_h + \phi_q + \tau_i + \epsilon_{hqi}\,, \tag{12.3.1}$$
$$\epsilon_{hqi} \sim N(0, \sigma^2)\,,$$
$$\epsilon_{hqi}\text{'s are mutually independent}\,,$$
$$h = 1, \ldots, b; \quad q = 1, \ldots, c; \quad i = 1, \ldots, v; \quad (h, q, i) \text{ in the design,}$$

Table 12.7 Analysis of variance table for a connected row–column design with no interactions

Source of variation	Degrees of freedom	Sum of squares	Mean square	Ratio
Rows (unadj)	$b-1$	$ss\theta$	–	–
Columns (unadj)	$c-1$	$ss\phi$	–	–
Treatments(adj)	$v-1$	ssT_{adj}	msT_{adj}	msT_{adj}/msE
Error	$bc-b-c-v+2$	ssE	msE	
Total	$bc-1$	$sstot$		

Formulae

$Q_i = T_i - \frac{1}{c}\Sigma_{h=1}^{b} n_{h.i}B_h - \frac{1}{b}\sum_{q=1}^{c} n_{.qi}C_q + \frac{r}{bc}G$	$ssT_{\text{adj}} = \Sigma_{i=1}^{v} Q_i\hat{\tau}_i$
$sstot = \Sigma_{h=1}^{b}\Sigma_{q=1}^{c}\Sigma_{i=1}^{v} n_{hqi}y_{hqi}^2 - \frac{1}{bc}G^2$	$ss\theta = \frac{1}{c}\Sigma_{h=1}^{b} B_h^2 - \frac{1}{bc}G^2$
$ssE = sstot - ss\theta - ss\phi - ssT_{\text{adj}}$	$ss\phi = \frac{1}{b}\Sigma_{q=1}^{c} C_q^2 - \frac{1}{bc}G^2$
$T_i = \Sigma_{h=1}^{b}\Sigma_{q=1}^{c} n_{hqi}y_{hqi}$	$B_h = \Sigma_{q=1}^{c}\Sigma_{i=1}^{v} n_{hqi}y_{hqi}$
$C_q = \Sigma_{h=1}^{b}\Sigma_{i=1}^{v} n_{hqi}y_{hqi}$	$G = \Sigma_{h=1}^{b}\Sigma_{q=1}^{c}\Sigma_{i=1}^{v} n_{hqi}y_{hqi}$

where θ_h, ϕ_q, and τ_i denote the effects on the response of row block h, column block q, and treatment i, respectively. The term "(h, q, i) in the design" means that the model is valid for whichever treatment i is observed in the cell defined by the hth row block and qth column block. The model includes the usual assumptions that the error terms ϵ_{hqi} are independent and normally distributed with constant variance. It also includes the assumptions of no interaction between the row and column blocking factors nor between these and the treatment factors.

12.3.2 Analysis of Variance for a Row–Column Design

We first give an overview of the analysis for connected row–column designs with no interactions. This is followed by the specific analyses for Latin square designs and Youden designs in Sects. 12.4 and 12.5, respectively. Analysis of general row–column designs can be done by computer, and this is illustrated via the SAS and R computer packages in Sects. 12.8 and 12.9.

For row–column designs with b row blocks, c column blocks, and r observations on each of v treatments with the row–column–treatment model (12.3.1), the top portion of Table 12.7 shows the form of the analysis of variance obtained when the row and column factors are entered into the model before the treatment factor. The (unadjusted) sum of squares $ss\theta$ for row blocks, and the (unadjusted) sum of squares $ss\phi$ for column blocks are calculated in a similar way to the block sum of squares in a complete block design; that is

$$ss\theta = \frac{1}{c}\sum_{h=1}^{b} B_h^2 - \frac{1}{bc}G^2 \quad \text{and} \quad ss\phi = \frac{1}{b}\sum_{q=1}^{c} C_q^2 - \frac{1}{bc}G^2$$

where B_h is the total of the observations in the hth row block, C_q is the total of the observations in the qth column block, and G is the grand total of all the observations.

The sum of squares for treatments adjusted for rows and columns (i.e. adjusted for the fact that not all treatments may be in every row block and every column block, and some blocks are better than others) can be shown to be

$$ssT_{\text{adj}} = \Sigma_{i=1}^{v} Q_i \hat{\tau}_i \text{ with } Q_i = T_i - \frac{1}{c}\sum_{h=1}^{b} n_{h.i}B_h - \frac{1}{b}\sum_{q=1}^{c} n_{.qi}C_q + \frac{r}{bc}G\,, \tag{12.3.2}$$

where $n_{h.i}$ is the number of times that treatment i appears in row block h, and and $n_{.qi}$ is the number of times that treatment i appears in column block q. In general, it is not easy to obtain least squares solutions by hand for the treatment parameters τ_i in row–column designs except for the cases of Latin square and Youden designs. Least squares solutions for these special cases are given in Sects. 12.4 and 12.5, respectively. The analysis of a more complicated row–column design is illustrated using the SAS and R computer packages in Sects. 12.8 and 12.9, respectively.

In general, the sum of squares for error for the row–column–treatment model (12.3.1) is obtained by subtraction as usual

$$ssE = sstot - ss\theta - ss\phi - ssT_{\text{adj}}\,, \tag{12.3.3}$$

where

$$sstot \;=\; \sum_h \sum_q \sum_i n_{hqi} y_{hqi}^2 - \frac{1}{bc}G^2\,,$$

and $n_{hqi} = 1$ if treatment label i is allocated to the combination of row block h and column block q and zero otherwise. As in any connected block design, the numbers of degrees of freedom for rows, columns and treatments are $b-1$, $c-1$ and $v-1$, respectively. The number of degrees of freedom for error is obtained by subtraction:

$$df = (bc-1)-(b-1)-(c-1)-(v-1) \;=\; bc-b-c-v+2\,, \tag{12.3.4}$$

where bc is the total number of observations.

A test of the null hypothesis $H_0^T : \{\tau_1 = \tau_2 = \cdots \tau_v\}$ against the general alternative hypothesis H_A^T :{at least two of the τ_i's differ} is given by the decision rule

$$\text{reject } H_0^T \text{ if } msT_{\text{adj}}/msE > F_{v-1,bc-b-c-v+2,\alpha} \tag{12.3.5}$$

for some chosen significance level α. The test (12.3.5) is most conveniently set out in an analysis of variance table, as in the top portion of Table 12.7. The bottom portion of the table shows the specific formulae.

12.3.3 Confidence Intervals and Multiple Comparisons

The multiple-comparison methods of Bonferroni and Scheffé can be applied for any row–column design. Simultaneous $100(1-\alpha)\%$ confidence intervals for treatment contrasts $\Sigma_i d_i \hat{\tau}_i$ are of the form

$$\left(\sum_i d_i \hat{\tau}_i \pm w\sqrt{\widehat{\text{Var}}(\sum_i d_i \hat{\tau}_i)}\right)\,, \tag{12.3.6}$$

where the critical coefficient w is

$$w_B = t_{bc-b-c-v+2,\alpha/2m} \text{ or } w_S = \sqrt{(v-1)F_{v-1,bc-b-c-v+2,\alpha}}\,,$$

for the Bonferroni or Scheffé methods, respectively. Tukey and Dunnett methods of pairwise comparisons can be used in the special cases of Latin square designs and Youden designs (Sects. 12.4.2 and 12.5.2).

12.4 Analysis of Latin Square Designs

12.4.1 Analysis of Variance for Latin Square Designs

Table 12.7, p. 406, shows the analysis of variance table for the row–column–treatment model (12.3.1) for any row–column design. In the table, T_i, B_h, C_q, and G are, respectively, the sum of the observations on the ith level of the treatment factor, the hth level of the row blocking factor, the qth level of the column blocking factor, and the grand total of all the observations. The constant n_{hqi} is equal to 1 if treatment i is observed in the combination of row block h and column block q, and n_{hqi} is equal to zero otherwise. For an s-replicate Latin square design, every treatment appears once in every row block and s times in every column block, so there are $b = vs$ rows, $c = v$ columns, and we have $n_{h.i} = 1$, $n_{.qi} = s$. It can be shown that, for Latin square designs, least squares solutions for the treatment parameters τ_i in model (12.3.1) are given by

$$\hat{\tau}_i = \frac{1}{vs} Q_i\,, \quad \text{with } Q_i \;=\; T_i - \frac{1}{c}\sum_h n_{h.i}B_h \;=\; T_i - \frac{1}{v}G\,. \tag{12.4.7}$$

Thus, for an s-replicate Latin square design, the treatment sum of squares (12.3.2) become

$$ssT_{\text{adj}} \;=\; \sum_{i=1}^{v} Q_i\hat{\tau}_i \;=\; \frac{1}{vs}\sum_{i=1}^{v}\left(T_i - \frac{1}{v}G\right)^2\,,$$

and we see that the "adjusted" treatment sum of squares for a Latin square design is actually *not* adjusted for row-block effects nor for column-block effects. This is because every treatment is observed the same number of times in every row block and in every column block (even though only bc of the bcv row–column–treatment combinations are actually observed). The computational formula for ssT_{adj} is obtained by expanding the terms in parentheses to give

$$ssT_{\text{adj}} \;=\; \frac{1}{vs}\sum_i T_i^2 - \frac{1}{v^2 s}G^2\,.$$

The error sum of squares and degrees of freedom are obtained by subtraction, as in (12.3.3). For an s-replicate Latin square design with $b = vs$ rows and $c = v$ columns, we have error degrees of freedom equal to

$$df = bc - b - c - v + 2 \;=\; v^2 s - vs - 2v + 2 \;=\; (vs - 2)(v - 1)\,. \tag{12.4.8}$$

The test for equality of treatment effects compares the ratio of the treatment and error mean squares with the corresponding value from the F-distribution, in the usual way (see Example 12.4.1). We will not utilize a test for the hypothesis of negligible row-block or of negligible column-block effects. However, we conclude that the current experiment has benefited from the use of the row (or column) blocking factor if the row (or column) mean square exceeds the mean square for error.

Table 12.8 Unrandomized 2-replicate Latin Square (Plan 2, Table 12.3), and possible randomization for the dairy cow experiment (with data in parentheses)

Row	Column			Cow	Period		
	1	2	3		1	2	3
1	A	B	C	6	2 (63)	3 (101)	1 (1)
2	B	C	A	4	3 (76)	1 (86)	2 (46)
3	C	A	B	2	3 (86)	2 (109)	1 (39)
4	A	C	B	5	1 (35)	2 (75)	3 (34)
5	B	A	C	1	2 (25)	1 (38)	3 (15)
6	C	B	A	3	1 (72)	3 (124)	2 (27)

Source Data adapted from *Experimental Designs*, Second Edition, by Cochran and Cox (1957), Copyright © 1957, John Wiley & Sons, New York.

Example 12.4.1 `Dairy cow experiment`

Cochran and Cox (1957, p. 135) described an experiment that studied the effects of three diets on the milk production of dairy cows. The $v = 3$ diets (levels of a treatment factor) consisted of roughage (level 1), limited grain (level 2), and full grain (level 3). A crossover experiment was used, with each of the $b = 6$ cows being fed each diet in turn, each for a six-week period. The response variable was the milk yield for each cow during each of the $c = 3$ periods. Data were collected using a randomization of the 2-replicate Plan 2 Latin square in Table 12.3, p. 403, with $v = 3$ treatment labels, $b = vs = 6$ rows, $c = 3$ columns, and with $v = 3$ treatments each observed $r = vs = 6$ times. A possible randomization and the data collected are shown in Table 12.8.

Provided that each measurement on milk yield has been taken after a cow has been on a new diet for a sufficiently long period of time to allow the effect of the previous diet to "wash out" of the system, we need not be concerned with carryover effects. For this experiment, a model without carryover effects describes the data fairly well.

Table 12.9 shows the analysis of variance table. To test the null hypothesis H_0^T that all three diets have the same effect, the decision rule is

$$\text{reject } H_0^T \text{ if } msT/msE > F_{2,8,.01} = 8.65\,,$$

with a probability of $\alpha = 0.01$ of making a Type I error. Since $msT/msE = 1138.39/103.06 = 11.05 > 8.65$, we reject the null hypothesis at the $\alpha = 0.01$ significance level and conclude that the diets do not all have the same effect.

To evaluate whether or not blocking was worthwhile, observe that the mean squares for periods and cows are both considerably larger than the error mean square. Thus, inclusion of both blocking factors in the experiment was beneficial in reducing the experimental error. (Since this is a Latin square design, no adjustment is needed to the row and column sums of squares to draw this conclusion.) □

12.4.2 Confidence Intervals for Latin Square Designs

All treatment contrasts are estimable in Latin square designs. Using the row–column–treatment model (12.3.1), the least squares estimate of a treatment contrast $\Sigma d_i\tau_i$, with $\Sigma d_i = 0$, is obtained from (12.4.7) as

$$\sum d_i\hat{\tau}_i = \frac{1}{vs}\sum d_iT_i\,, \qquad (12.4.9)$$

Table 12.9 Analysis of variance table for the dairy cow experiment

Source of variation	Degrees of freedom	Sum of squares	Mean square	Ratio	p-value
Cow (row)	5	5781.11	1156.22	–	–
Period (column)	2	11480.11	5740.06	–	–
Diets (treatment)	2	2276.78	1138.39	11.05	0.0050
Error	8	824.44	103.06		
Total	17	20362.44			

and, since T_i is the sum of vs observations, the corresponding variance is

$$\text{Var}\left(\sum d_i \hat{\tau}_i\right) = \sum d_i^2 \left(\frac{\sigma^2}{vs}\right), \tag{12.4.10}$$

The multiple-comparison methods of Bonferroni, Scheffé, Tukey, and Dunnett can be used for s-replicate Latin square designs. Formulae for confidence intervals for treatment contrasts $\Sigma d_i \tau_i$ are of the form

$$\left(\frac{1}{vs}\Sigma_i d_i T_i \;\pm\; w\sqrt{msE\left(\frac{\Sigma d_i^2}{vs}\right)}\right), \tag{12.4.11}$$

where the appropriate critical coefficients w for the four methods are,

$$w_B = t_{(vs-2)(v-1),\alpha/2m}\ ;\ \ w_S = \sqrt{(v-1)F_{v-1,(vs-2)(v-1),\alpha}}\ ;$$
$$w_T = q_{v,(vs-2)(v-1),\alpha}/\sqrt{2}\ ;\ \ w_{D2} = |t|^{(0.5)}_{v-1,(vs-2)(v-1),\alpha}\ .$$

Example 12.4.2 `Dairy cow experiment, continued`

The dairy cow experiment was described in Example 12.4.1, p. 409. Tukey's method of all pairwise comparisons can be applied to determine which diets, if any, are significantly different from any of the others. The treatment sample means, which can be computed from the data in Table 12.8, are

$$\frac{1}{vs}T_1 = 45.167,\quad \frac{1}{vs}T_2 = 57.500,\quad \frac{1}{vs}T_3 = 72.667\,,$$

so that the least squares estimates of the pairwise differences are

$$\hat{\tau}_2 - \hat{\tau}_1 = 57.500 - 45.167 = 12.333\,,$$
$$\hat{\tau}_3 - \hat{\tau}_1 = 72.667 - 45.167 = 27.500\,,$$
$$\hat{\tau}_3 - \hat{\tau}_2 = 72.667 - 57.500 = 15.167\,.$$

The value of the error mean square is $msE = 103.06$ from the analysis of variance table, Table 12.9. The error degrees of freedom are given by (12.4.8) as

$$df = (vs-2)(v-1) = (2\times 3-1)(3-1) = 8\,.$$

Using a simultaneous confidence level of 95%, we obtain the critical coefficient for Tukey's method as $w_T = q_{3,8,.05}/\sqrt{2} = 4.04/\sqrt{2}$. Hence, using (12.4.11), the minimum significant difference for pairwise differences is

$$msd = (4.04/\sqrt{2})\sqrt{msE(2/6)} = 16.743\,.$$

The simultaneous 95% confidence intervals are therefore

$$\begin{aligned}
\tau_3 - \tau_1 &\in (27.500 \ \pm \ 16.743) \ = \ (\ 10.58, 44.24)\,,\\
\tau_2 - \tau_1 &\in (12.333 \ \pm \ 16.743) \ = \ (-4.41, 29.08)\,,\\
\tau_3 - \tau_2 &\in (15.167 \ \pm \ 16.743) \ = \ (-1.58, 31.91)\,.
\end{aligned}$$

From these intervals, we can deduce that at overall significance level $\alpha = 0.05$, the full grain diet (level 3) results in a mean yield of 10.58–44.24 units higher than the roughage diet (level 1), but the limited grain diet (level 2) is not significantly different from either of the other two. □

12.4.3 How Many Observations?

The formula for determining the sample size needed to achieve a power $\pi(\Delta)$ of detecting a difference Δ in the treatment effects for given v, α, and σ^2, using a Latin square design, is the same as (10.6.13) for a randomized complete block design, but with b replaced by v (since $r = vs$ rather than $r = bs$). Thus, after simplification of the formula, we need to find s to satisfy

$$s \ \geq \ \frac{2\sigma^2\phi^2}{\Delta^2}\,. \tag{12.4.12}$$

Alternatively, the confidence interval formula (12.4.11) can be used to calculate the sample sizes needed for achieving confidence intervals of a desired width (see Example 12.4.3).

Example 12.4.3 `Sample-size calculation for an s-replicate Latin square design`

Consider an experiment to compare $v = 3$ computer keyboard designs with respect to the time taken to type an article of given length. Typists and time periods are the two blocking factors, with each of $b = 3s$ typists using each of the keyboards in a sequence of $c = 3$ time periods. An s-replicate Latin square design will be used, with $r = 3s$ observations per keyboard layout (treatment), and with sufficient time between observations to prevent a carryover effect.

With 3 treatments, there are 3 pairwise comparisons. Using Tukey's method of multiple comparisons and a simultaneous confidence level of 95%, suppose that the experimenters want confidence intervals with half-width (minimum significant difference) of 10 min or less. The error standard deviation is expected to be at most 15 min (a variance of at most 225 min^2). The error degrees of freedom, as given in (12.4.8), are

$$df = (vs - 2)(v - 1) = 2(3s - 2) = 6s - 4\,.$$

Then, using the confidence interval formula (12.4.11) for Tukey's method of pairwise comparisons, the minimum significant difference is

$$msd \approx \frac{q_{3,6s-4,.05}}{\sqrt{2}}\sqrt{\frac{225 \times 2}{vs}} = q_{3,6s-4,.05}\sqrt{\frac{225}{3s}}\,.$$

To obtain $msd \le 10$, we require $0.75q^2_{3,6s-4,.05} \le s$. Sample size is then computed by trial and error as follows:

s	$6s-4$	$q_{3,6s-4,.05}$	$0.75q^2_{3,6s-4,.05}$	Required
	∞	3.31	8.22	$s \ge 9$
9	50	3.42	8.77	$s = 9$

So, $s = 9$, and a 9-replicate Latin square would be needed, giving $msd \approx 3.42\sqrt{\frac{225}{27}} = 9.87$ min and requiring $b = vs = 27$ typists and $r = vs = 27$ observations per keyboard layout. □

12.5 Analysis of Youden Designs

12.5.1 Analysis of Variance for Youden Designs

The s-replicate Youden design with v treatments has complete column blocks of size $b = vs$ and row blocks forming a balanced incomplete block design with blocks of size c. Each treatment is observed $r = cs$ times. For the row–column–treatment model (12.3.1) with no interactions, a solution to the normal equations can be shown to be similar to the solution (11.4.9), p. 359, for a balanced incomplete block design; that is,

$$\hat{\tau}_i = \frac{c}{\lambda v} Q_i\,, \text{ where } Q_i = T_i - \frac{1}{c}\sum_h n_{h.i} B_h\,, \tag{12.5.13}$$

for $i = 1, \ldots, v$, where $\lambda = r(c-1)/(v-1)$.

As in the analysis of a balanced incomplete block design, the estimators of the treatment effects in a Youden design are not independent of the estimators of the row-block effects. As a result, the treatment-effect estimators must be adjusted for row blocks. However, since the column blocks form a randomized complete block design, no adjustment is needed for these. Table 12.7, p. 406, shows the analysis of variance table for a row–column design with no interactions. Using (12.5.13), the adjusted treatment sum of squares given in the table becomes

$$ssT_{\text{adj}} = \sum_{i=1}^{v} Q_i \hat{\tau}_i = \frac{c}{\lambda v}\sum_{i=1}^{v} Q_i^2\,,$$

and, because of the complete column blocks, this is exactly the same formula as for a balanced incomplete block design with blocks of size $k = c$ (see Table 11.7, p. 358).

To test the null hypothesis H_0^T of no treatment differences, at significance level α, the decision rule is

$$\text{reject } H_0^T \text{ if } \frac{msT_{\text{adj}}}{msE} > F_{v-1,df,\alpha}\,,$$

where, as given in Table 12.7, the number of error degrees of freedom for an s-replicate Youden square is

$$\begin{aligned} df &= bc - b - c - v + 2 \\ &= vsc - vs - c - v + 2 \\ &= (vs-1)(c-1) - (v-1)\,. \end{aligned} \tag{12.5.14}$$

12.5.2 Confidence Intervals for Youden Designs

All treatment contrasts $\Sigma d_i \tau_i$ are estimable in all Youden designs. The least squares estimate of the contrast $\Sigma d_i \tau_i$ is

$$\sum_i d_i \hat{\tau}_i = \frac{c}{\lambda v} \sum_i d_i Q_i = \left(\frac{v-1}{vs(c-1)}\right) \sum_i d_i Q_i ,$$

since $\lambda = r(c-1)/(v-1)$ as for a balanced incomplete block design with $k = c$. The variance of the corresponding estimator is

$$\mathrm{Var}(\sum_i d_i \hat{\tau}_i) = \left(\frac{c}{\lambda v}\right) \sum_i d_i^2 \sigma^2 = \left(\frac{v-1}{vs(c-1)}\right) \sum_i d_i^2 \sigma^2 . \tag{12.5.15}$$

Confidence interval formulae for treatment contrasts $\Sigma d_i \tau_i$ are the same as those for a balanced incomplete block design with block size $k = c$; that is,

$$\left(\frac{v-1}{vs(c-1)}\right) \sum_i d_i Q_i \ \pm\ w \sqrt{msE \left(\frac{v-1}{vs(c-1)}\right) \sum d_i^2} , \tag{12.5.16}$$

where w, as usual, is the critical coefficient for the Bonferroni, Scheffé, Tukey, or Dunnett method, given by

$$w_B = t_{(vs-1)(c-1)-(v-1),\alpha/2m} \ ; \ w_S = \sqrt{(v-1)F_{v-1,(vs-1)(c-1)-(v-1),\alpha}} \ ;$$
$$w_T = q_{v,(vs-1)(c-1)-(v-1),\alpha}/\sqrt{2} \ ; \ w_{D2} = |t|^{(0.5)}_{v-1,(vs-1)(c-1)-(v-1),\alpha} \ .$$

12.5.3 How Many Observations?

The methods of sample-size calculation for an s-replicate Youden square are analogous to those for computing sample sizes for a balanced incomplete block design. To calculate the number of observations $r = cs$ per treatment that are required to achieve a power $\pi(\Delta)$ of detecting a difference Δ in the treatment effects for given v, α, and σ^2, the power tables in Appendix A.7 can be used in the same way as for a balanced incomplete block design (Sect. 11.6). Thus, we need to find s satisfying

$$cs \geq \frac{2v\sigma^2\phi^2}{\Delta^2}\left[\frac{c(v-1)}{v(c-1)}\right] ,$$

which reduces to

$$s \geq \frac{2\sigma^2\phi^2(v-1)}{\Delta^2(c-1)} .$$

Alternatively, the confidence interval formula (12.5.16) can be used to calculate the sample sizes needed for achieving confidence intervals of a desired width (see Example 12.5.1).

Example 12.5.1

Suppose an experiment is run to compare six paint additive formulations (levels 2–7 of the treatment factor) with a standard "control" formulation (level 1) with respect to the drying time (in minutes). The paint is sprayed through $c = 4$ different nozzles so that $c = 4$ paints can be sprayed simultaneously. A total of $b = 7s$ panels will each be painted with strips of 4 of the 7 paints (using a standard formulation between each test panel to clean the nozzles and create a washout period). The error variability is expected to be at most 25 min^2 (standard deviation at most 5 min), and simultaneous 90% confidence intervals with half-width of at most 3.5 min are required for the six treatment-versus-control contrasts.

A Youden square with 7 rows, 4 columns, and 7 treatments is shown in Table 12.1. Suppose that s copies (or column randomizations) of this basic square are to be stacked, giving an s-replicate Youden design with the same number of observations on the experimental and control treatments. The number of error degrees of freedom, given in (12.5.14), is then

$$df = (vs - 1)(c - 1) - (v - 1) = (7s - 1)(4 - 1) - (7 - 1) = 21s - 9\,.$$

Using Dunnett's method of multiple comparisons for treatment versus control, the minimum significant difference is

$$msd = w_{D2}\sqrt{msE \times \left(\frac{(v-1)}{vs(c-1)}\right) \times 2}\,,$$

so we require

$$msd \approx w_{D2}\sqrt{\frac{(25)(6)(2)}{(7)s(3)}} \le 3.5\,,$$

that is,

$$w_{D2}^2 \le 0.8575s\,,$$

where $w_{D2} = |t|^{(.5)}_{6,21s-9,.1}$. Using Table A.10 for the values of $w_{D2} = |t|^{(.5)}_{6,21s-9,.1}$, we have

s	$21s - 9$	$w_{D2} = \|t\|^{(.5)}_{6,21s-9,.1}$	w_{D2}^2	$0.8575s$	Action
20	411	2.30	5.29	17.15	Decrease s
6	117	2.32	5.38	5.15	Increase s
7	138	2.32	5.38	6.00	Decrease s

So we see that $s = 7$ stacked Youden squares would be adequate, requiring $b = vs = 49$ panels and giving $r = sc = 28$ observations on each of the $v = 7$ paint formulations, both experimental and control. If 49 panels are not available for the experiment, then the experimenter would have to be satisfied with wider confidence intervals. □

12.6 Checking the Assumptions on the Model

The error assumptions on the row–column–treatment model (12.3.1) can be checked by plotting the standardized residuals against the run order, the predicted values $\hat{y}_{ijk}$, the levels of the row blocking factor, the levels of the column blocking factor, the levels of the treatment factor, and the normal scores.

The data collected from a row–column design can be examined by plotting the adjusted observations against the treatment labels. The observations are adjusted by subtracting the least squares estimates for row blocks and column blocks (which are adjusted for treatments):

$$y^*_{hqi} = y_{hqi} - (\hat{\theta}_h - \hat{\overline{\theta}}_.) - (\hat{\phi}_q - \hat{\overline{\phi}}_.)\,. \tag{12.6.17}$$

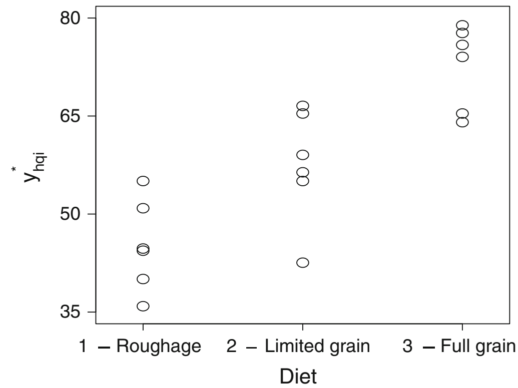

Fig. 12.1 Adjusted data for the dairy cow experiment

For the s-replicate Latin square design with $b = vs$ row blocks and $c = v$ column blocks, and the row–column–treatment model (12.3.1), the row-block and column-block effect estimators are independent of treatment effects, and (12.6.17) becomes

$$y^*_{hqi} = y_{hqi} - \left(\frac{1}{vs}B_h - \frac{1}{v^2 s}G\right) - \left(\frac{1}{v}C_q - \frac{1}{v^2 s}G\right) .$$

For other designs, the adjusted observations can be calculated by computer, as shown in Sects. 12.8 and 12.9. Since the variability due to the blocking factors has been extracted from the adjusted observations, the data plots will exhibit less variability than really exists.

Example 12.6.1 `Dairy cow experiment, continued`

For the dairy cow experiment, which was run as a Latin square design, the adjusted observations are plotted against treatment labels in Fig. 12.1. The plot shows how milk yield tended to increase as the quality of the diet improves from roughage (level 1) to limited grain (level 2) to full grain (level 3). The plot is consistent with the results of Example 12.4.1, where simultaneous confidence intervals (with an overall 95% confidence level) showed a significant difference in diets 1 and 3, but were unable to distinguish between diets 1 and 2 and between diets 2 and 3. $\square$

The row–column–treatment model (12.3.1) assumes that the two blocking factors do not interact with each other nor with the treatment factor. If interactions are present, the error variance estimate will be inflated, decreasing the powers of hypothesis tests and widening confidence intervals for treatment contrasts. When the column blocks are complete blocks, the column–treatment interaction can be checked by plotting the row-adjusted observations (as in Chap. 11) against treatments, with plot labels being the column-block labels. Interaction is indicated by nonparallel lines. The row–column interaction can be investigated in the same way using the treatment-adjusted observations and plotting against row labels. The row–treatment interactions can only be investigated if the row blocks are also complete blocks as in Latin square designs.

12.7 Factorial Experiments in Row–Column Designs

The exercise bike experiment of Example 12.2.1, p. 404, was a factorial experiment with three treatment factors having two levels each. These were "time duration of exercise" (1 and 3 min, coded as 1 and 2), "exercise speed" (40 and 60 rpm, coded as 1 and 2), and "pedal type" (foot pedal and hand bars, coded as 1 and 2). The data were shown in Table 12.6.

If we use the row–column–treatment model with three-digit labels for the treatment combinations, then we can write the model as

$$Y_{hqijk} = \mu + \theta_h + \phi_q + \tau_{ijk} + \epsilon_{hqijk} \,,$$

where θ_h is effect of day h, ϕ_q is the effect of subject q, and τ_{ijk} is the effect of treatment combination ijk. If we now rewrite τ_{ijk} in terms of its constituent main effects and interactions, we obtain the following form of the row–column–treatment model:

$$\begin{aligned} Y_{hqijk} &= \mu + \theta_h + \phi_q + \alpha_i + \beta_j + \gamma_k \\ &\quad + (\alpha\beta)_{ij} + (\alpha\gamma)_{jk} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{hqijk} \,, \\ h &= 1, \ldots, 8\,;\;\; q = 1, 2, 3\,;\;\; i = 1, 2\,;\;\; j = 1, 2\,;\;\; k = 1, 2\,; \end{aligned}$$

where α_i is the effect of the ith duration, β_j is the effect of the jth speed, and γ_k is the effect of the kth pedal type, and the other terms represent interactions between the treatment factors. The analysis of this experiment by the programs SAS and R software is shown in Sects. 12.8 and 12.9, respectively, for both forms of the model. If some of the treatment interactions are thought to be negligible, they can be dropped from the latter model.

12.8 Using SAS Software

In this section, a sample program is given to illustrate the analysis of row–column designs using the SAS software. The program uses the data of the exercise bicycle experiment, which was described in Example 12.2.1, p. 404. The design is a cyclic row–column design with $v = 8$ treatment labels representing the eight treatment combinations shown in Table 12.6, and with $b = 8$ row blocks representing days and $c = 3$ column blocks representing subject. The treatment combinations were combinations of the levels of the three treatment factors "time duration of exercise," "exercise speed," and "pedal type."

A SAS program for analyzing this experiment is shown in Table 12.10. The data are entered into a data set called `BIKE`, using `DAY` and `SUBJECT` as the two blocking factors, and using `DURAT`, `SPEED`, and `PEDAL` as the three treatment factors, and `PULSE` for the response variable "pulse rate." The combinations of the levels of the three treatment factors are recoded as levels of a factor `TRTMT`.

The `MODEL` statement in the first `GLM` procedure causes generation of the analysis of variance table shown in Fig. 12.2. The blocking factors are entered into the model before the treatment factor, so the treatment sum of squares is adjusted for block effects whether one looks at the Type I or Type III sums of squares. The row and column effects are independent of each other, since there is one observation at each combination of levels of the row and column blocking factors. Consequently, the Type I sums of squares reproduce the analysis of variance table, Table 12.7, with unadjusted block effects and adjusted treatment effects. In this particular experiment, the column blocks (subjects) are complete blocks, and

Table 12.10 SAS program for analysis of a row–column design—Exercise bicycle experiment

```
DATA BIKE;
  INPUT  DAY SUBJECT  PULSE  DURAT$ SPEED$ PEDAL$;
  TRTMT = trim(DURAT)||trim(SPEED)||trim(PEDAL);
  LINES;
   1 1  45  1 1 2
   1 2  25  2 1 2
   1 3  18  2 2 1
   2 1  27  2 2 2
   : :   :  : : :
   8 3  34  1 1 1
;
PROC PRINT;
*    row-column-treatment model;
PROC GLM;
  CLASS DAY SUBJECT TRTMT;
  MODEL PULSE = DAY SUBJECT TRTMT / SOLUTION;
  OUTPUT OUT=RESIDS PREDICTED=PRED RESIDUAL=Z;
  ESTIMATE 'DURATION DIFF'  TRTMT -1 -1 -1 -1  1  1  1  1 / DIVISOR=4;
  ESTIMATE 'SPEED DIFF'     TRTMT -1 -1  1  1 -1 -1  1  1 / DIVISOR=4;
  ESTIMATE 'PEDAL DIFF'     TRTMT -1  1 -1  1 -1  1 -1  1 / DIVISOR=4;
* Standardize residuals and compute normal scores;
PROC STANDARD STD=1.0; VAR Z;
PROC RANK NORMAL=BLOM; VAR Z; RANKS NSCORE;
* Residual plots can now be generated using PROC SGPLOT;
```

so the treatment sum of squares is actually only adjusted for row-block (day) effects. The Type III sums of squares show the row-block (day) effects adjusted for the treatment effects.

The `ESTIMATE` statements request the SAS software to calculate the information needed to calculate three confidence intervals. The three selected contrasts are the main-effect contrasts comparing the effect on pulse rate of the two levels of each treatment factor (averaging over any interaction that might be present). The output, shown in Fig. 12.3, gives the least squares estimates of the contrasts and their associated standard errors. From this information, confidence intervals can be calculated in the usual way. The contrast estimates are adjusted for the incomplete row blocks.

12.8.1 Factorial Model

To write the row–column–treatment model in terms of factorial treatment combinations, the model statement in Table 12.10 is replaced by

```
MODEL PULSE = DAY SUBJECT DURAT SPEED PEDAL DURAT*SPEED
              DURAT*PEDAL SPEED*PEDAL DURAT*SPEED*PEDAL;
```

The Type III sums of squares are shown in Fig. 12.4, and we can see that the sums of squares for the main effects and interactions of the three treatment factors do not add to the Type III treatment sum of squares in Fig. 12.2 due to the individual adjustments for block effects and for the other treatment factor effects.

Fig. 12.2 Analysis of variance for a row–column design—Exercise bicycle experiment

Results Viewer - SAS Output

The GLM Procedure
Dependent Variable: PULSE

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	1257.678571	78.604911	3.28	0.0588
Error	7	167.946429	23.992347		
Corrected Total	23	1425.625000			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
DAY	7	202.2916667	28.8988095	1.20	0.4062
SUBJECT	2	201.0000000	100.5000000	4.19	0.0636
TRTMT	7	854.3869048	122.0552721	5.09	0.0238

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DAY	7	23.7202381	3.3886054	0.14	0.9903
SUBJECT	2	201.0000000	100.5000000	4.19	0.0636
TRTMT	7	854.3869048	122.0552721	5.09	0.0238

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	19.86607143	B	5.59247154	3.55	0.0093
DAY 1	3.99107143	B	5.43709471	0.73	0.4868
DAY 2	1.56250000	B	5.43709471	0.29	0.7821
DAY 3	2.57142857	B	5.55403486	0.46	0.6574
DAY 4	0.80357143	B	4.94173873	0.16	0.8754
DAY 5	2.51785714	B	4.94173873	0.51	0.6261
DAY 6	1.63392857	B	4.39085009	0.37	0.7208
DAY 7	0.20535714	B	4.39085009	0.05	0.9640
DAY 8	0.00000000	B	.	.	.
SUBJECT 1	5.25000000	B	2.44909917	2.14	0.0693
SUBJECT 2	-1.50000000	B	2.44909917	-0.61	0.5596
SUBJECT 3	0.00000000	B	.	.	.
TRTMT 111	12.64285714	B	4.94173873	2.56	0.0376
TRTMT 221	-7.25892857	B	5.43709471	-1.34	0.2236
TRTMT 222	0.00000000	B	.	.	.

Note: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

Fig. 12.3 Output from the `ESTIMATE` statement—Exercise bicycle experiment

Parameter	Estimate	Standard Error	t Value	Pr > |t|
DURATION DIFF	-5.3437500	2.71854736	-1.97	0.0901
SPEED DIFF	-10.3214286	2.47086937	-4.18	0.0042
PEDAL DIFF	4.3080357	2.71854736	1.58	0.1571

Fig. 12.4 Analysis of variance for a row–column design—Exercise bicycle experiment

Results Viewer - sashtml.htm

The GLM Procedure
Dependent Variable: PULSE

Source	DF	Type III SS	Mean Square	F Value	Pr > F
DAY	7	23.7202381	3.3886054	0.14	0.9903
SUBJECT	2	201.0000000	100.5000000	4.19	0.0636
DURAT	1	92.7024457	92.7024457	3.86	0.0901
SPEED	1	418.6516291	418.6516291	17.45	0.0042
PEDAL	1	60.2500647	60.2500647	2.51	0.1571
DURAT*SPEED	1	7.2858135	7.2858135	0.30	0.5987
DURAT*PEDAL	1	21.6694862	21.6694862	0.90	0.3736
SPEED*PEDAL	1	11.5766693	11.5766693	0.48	0.5097
DURAT*SPEED*PEDAL	1	7.7142857	7.7142857	0.32	0.5884

The ESTIMATE statements for the factorial model become

```
ESTIMATE 'DURATION DIFF'  DURAT -1 1;
ESTIMATE 'SPEED DIFF'     SPEED -1 1;
ESTIMATE 'FOOT/HAND DIFF' PEDAL -1 1;
```

and give output identical to that of Fig. 12.3.

12.8.2 Plots

The statements needed for calculating the standardized residuals and normal scores are also shown in Table 12.10. Residual plots (not shown here) can then be generated as in Sects. 5.8.1 and 6.8.3.

The SOLUTION option in the MODEL statement causes a set of least squares solutions to the normal equations to be printed. These are subsequently used to calculate the adjusted data values needed for examining the data.

For obtaining the adjusted observations, the statements in Table 12.11 would be added to the program in Table 12.10 for a second and third run of the program. The data are adjusted using (12.6.17); that is,

$$y^*_{hqi} = y_{hqi} - (\hat{\theta}_h - \hat{\bar{\theta}}_.) - (\hat{\phi}_q - \hat{\bar{\phi}}_.)\,,$$

where $\hat{\theta}_h$ and $\hat{\phi}_q$ are obtained from the output of the SOLUTION option shown in Fig. 12.2. The second run of the program takes as input the values of $\hat{\theta}_h$ and $\hat{\phi}_q$ from the first run of the program and calculates $\hat{\bar{\theta}}_. = \Sigma\hat{\theta}_h/3 = 1.25$ and $\hat{\bar{\phi}}_. = \Sigma\hat{\phi}_q/8 = 1.66$, which are then copied by hand into the statements for the third run of the program. The symbols @@ allow the input to be entered with more than one record per line.

In the third run of the program, the data set BIKE5 is created as a copy of the data set BIKE. This is done to create the new variable YADJ, which contains the values of PULSE, adjusted for the row and column effects. The SGPLOT procedure is used to plot the adjusted observations by treatment. The SAS plot is analogous to that in Fig. 12.1 (p. 415) and is not shown here.

Table 12.11 SAS code for calculating and plotting the adjusted observations—Exercise bicycle experiment

```
* Add the following code for the second run of the program;
DATA BIKE3; * input subject effect estimates from first run;
  INPUT SUBJECT SBJHAT; @@;
  LINES;
   1  5.25     2 -1.50     3  0.00
PROC MEANS MEAN; * print average of the subject effect estimates;
  VAR SBJHAT;
DATA BIKE4; * input day effect estimates from first run;
  INPUT DAY DHAT @@;
  LINES;
   1 3.9911     2 1.5625     3 2.5714     4 0.8036
   5 2.5179     6 1.6334     7 0.2054     8 0.0000
PROC MEANS MEAN; * print average of the day effect estimates;
  VAR DHAT;
* Add the following code for the third run;
* Adjust data for subject and day effects, then plot adjusted data;
DATA BIKE5; SET BIKE;
  IF SUBJECT = 1 THEN YADJ = PULSE-(5.25-1.25);
    ELSE IF SUBJECT = 2 THEN YADJ = PULSE-(-1.50-1.25);
      ELSE IF SUBJECT = 3 THEN YADJ = PULSE-(0.00-1.25);
  IF DAY = 1 THEN YADJ = YADJ-(3.9911-1.660);
    ELSE IF DAY = 2 THEN YADJ = YADJ-(1.5625-1.6607);
      ELSE IF DAY = 3 THEN YADJ = YADJ-(2.5714-1.6607);
        ELSE IF DAY = 4 THEN YADJ = YADJ-(0.8036-1.6607);
          ELSE IF DAY = 5 THEN YADJ = YADJ-(2.5179-1.6607);
            ELSE IF DAY = 6 THEN YADJ = YADJ-(1.6334-1.6607);
              ELSE IF DAY = 7 THEN YADJ = YADJ-(0.2054-1.6607);
                ELSE IF DAY = 8 THEN YADJ = YADJ-(0.0000-1.6607);
PROC SGPLOT;
  SCATTER X = TRTMT Y = YADJ;
  X AXIS LABEL = 'Treatment'; REFLINE 0/ AXIS = X;
  Y AXIS LABEL = 'Adjusted Observations'; REFLINE 0/ AXIS = Y;
```

12.9 Using R Software

In this section, a sample program is given to illustrate the analysis of row–column designs using the R software. The program uses the data of the exercise bicycle experiment, which was described in Example 12.2.1, p. 404. The design is a cyclic row–column design with $v = 8$ treatment labels representing the eight treatment combinations shown in Table 12.6, and with $b = 8$ row blocks representing days and $c = 3$ column blocks representing subject. The treatment combinations were combinations of the levels of the three treatment factors "time duration of exercise," "exercise speed," and "pedal type."

An R program for analyzing this experiment is shown in Table 12.12. The data are entered into a data set called `bike.data`, using `fDay` and `fSubject` as the two blocking factors, and using `fDurat`, `fSpeed`, and `fPedal` as the three treatment factors, and `Pulse` for the response variable "pulse rate." The combinations of the levels of the three treatment factors are recoded as levels of a factor `fTrtmt`.

In the second block of code, the linear models function `lm` is used to fit the row–column–treatment model (12.3.1), then corresponding `anova` and `drop1` functions generate the type 1 and type 3

Table 12.12 R program for analysis of a row–column design—Exercise bicycle experiment

```
bike.data = read.table("data/exercise.bicycle.txt", header=T)
head(bike.data, 3)

       Day Subject Durat Speed Pedal Pulse Trtmt
     1  1       1     1     1     2    45   112
     2  2       1     2     2     2    27   222
     3  3       1     2     1     2    40   212

# Create factor variables
bike.data = within(bike.data,
  {fDay=factor(Day); fSubject=factor(Subject); fDurat=factor(Durat);
   fSpeed=factor(Speed); fPedal=factor(Pedal); fTrtmt=factor(Trtmt) })

# ANOVA: treatment combinations
modelTC = lm(Pulse ~ fDay + fSubject + fTrtmt, data=bike.data)
anova(modelTC)
drop1(modelTC, ~., test = "F")
# Treatment contrasts
library(lsmeans)
lsmTrtmt = lsmeans(modelTC, ~ fTrtmt)
summary(contrast(lsmTrtmt,
        list(DurationDiff=c(-1,-1,-1,-1, 1, 1, 1, 1)/4,
                SpeedDiff=c(-1,-1, 1, 1,-1,-1, 1, 1)/4,
                PedalDiff=c(-1, 1,-1, 1,-1, 1,-1, 1)/4)),
        infer=c(T,T))

# Compute variables for residual plots
bike.data = within(bike.data, {
   # Compute predicted, residual, and standardized residual values
   ypred = fitted(modelTC); e = resid(modelTC); z = e/sd(e);
   # Compute Blom's normal scores
   n = length(e); q = rank(e); nscore = qnorm((q-0.375)/(n+0.25))
   # Compute vbl TLevel with levels 1:8 for equispaced plotting vs TC's
   TLevel = 4*(Durat-1) + 2*(Speed-1) + (Pedal-1) })
# Residual plots
plot(z ~ TLevel, data=bike.data, xaxt="n", xlab="Treatment")
  axis(1, at=bike.data$TLevel, labels=bike.data$fTrtmt)
plot(z ~ ypred + Day + Subject + nscore, data=bike.data)
```

analysis of variance tables, respectively, shown at the top of Table 12.13. The blocking factors are entered into the model before the treatment factor, so the treatment sum of squares is adjusted for block effects whether one looks at the Type I or Type III sums of squares. The row and column effects are independent of each other, since there is one observation at each combination of levels of the row and column blocking factors. Consequently, the Type I sums of squares reproduce the analysis of variance table, Table 12.7, with unadjusted block effects and adjusted treatment effects. In this particular experiment, the column blocks (subjects) are complete blocks, and so the treatment sum of squares is actually only adjusted for row-block (day) effects. The Type I sums of squares show the row-block (day) effects adjusted for the treatment effects.

The `summary` and `contrast` functions of the `lsmeans` package are coupled to calculate estimates, standard errors, tests and confidence for the three main-effect contrasts comparing the effect

Table 12.13 Analysis of variance and contrast estimation for a row–column design—Exercise bicycle experiment

```
> # ANOVA: treatment combinations
> modelTC = lm(Pulse ~ fDay + fSubject + fTrtmt, data=bike.data)
> anova(modelTC)

Analysis of Variance Table

Response: Pulse
          Df Sum Sq Mean Sq F value Pr(>F)
fDay       7    202    28.9    1.20  0.406
fSubject   2    201   100.5    4.19  0.064
fTrtmt     7    854   122.1    5.09  0.024
Residuals  7    168    24.0

> drop1(modelTC, ~., test = "F")

Single term deletions

Model:
Pulse ~ fDay + fSubject + fTrtmt
         Df Sum of Sq  RSS   AIC F value Pr(>F)
<none>                 168  80.7
fDay      7        24  192  69.9    0.14  0.990
fSubject  2       201  369  95.6    4.19  0.064
fTrtmt    7       854 1022 110.0    5.09  0.024

> # Treatment contrasts
> library(lsmeans)
> lsmTrtmt = lsmeans(modelTC, ~ fTrtmt)
> summary(contrast(lsmTrtmt,
+                  list(DurationDiff=c(-1,-1,-1,-1, 1, 1, 1, 1)/4,
+                       SpeedDiff=c(-1,-1, 1, 1,-1,-1, 1, 1)/4,
+                       PedalDiff=c(-1, 1,-1, 1,-1, 1,-1, 1)/4)),
+          infer=c(T,T))

 contrast      estimate     SE df lower.CL upper.CL t.ratio p.value
 DurationDiff   -5.3438 2.7185  7 -11.7721   1.0846  -1.966  0.0901
 SpeedDiff     -10.3214 2.4709  7 -16.1641  -4.4788  -4.177  0.0042
 PedalDiff       4.3080 2.7185  7  -2.1203  10.7364   1.585  0.1571

Results are averaged over the levels of: fDay, fSubject
Confidence level used: 0.95
```

on pulse rate of the two levels of each treatment factor (averaging over any interaction that might be present). The corresponding output is shown in the bottom of Table 12.13.

Predicted values and residuals are available, and the residuals can be standardized and plotted as in Sects. 5.9.1 and 6.9.3. Sample code is provided at the bottom of Table 12.12.

Table 12.14 Analysis of variance for a row–column design—Exercise bicycle experiment

```
> # ANOVA: factorial effects
> modelFE = lm(Pulse ~ fDay + fSubject + fDurat*fSpeed*fPedal,
+              data=bike.data)
> anova(modelFE)
> drop1(modelFE, ~., test = "F")

Single term deletions

Model:
Pulse ~ fDay + fSubject + fDurat * fSpeed * fPedal
                    Df Sum of Sq RSS   AIC F value Pr(>F)
<none>                           168  80.7
fDay                 7      23.7 192  69.9    0.14  0.990
fSubject             2     201.0 369  95.6    4.19  0.064
fDurat               1     129.2 297  92.4    5.38  0.053
fSpeed               1     269.1 437 101.6   11.22  0.012
fPedal               1       1.1 169  78.9    0.05  0.833
fDurat:fSpeed        1      15.0 183  80.7    0.63  0.455
fDurat:fPedal        1      36.5 204  83.4    1.52  0.257
fSpeed:fPedal        1      24.3 192  81.9    1.01  0.348
fDurat:fSpeed:fPedal 1       7.7 176  79.8    0.32  0.588
```

12.9.1 Factorial Model

To write the row–column–treatment model in terms of factorial effects, the linear models statement in Table 12.12 is replaced by

```
modelFE = lm(Pulse ~ fDay + fSubject + fDurat*fSpeed*fPedal,
             data=bike.data)
```

The Type III sums of squares are shown in Table 12.14, and we can see that the sums of squares for the main effects and interactions of the three treatment factors do not add to the Type III treatment sum of squares in Table 12.13 due to the individual adjustments for block effects and for the other treatment factor effects.

The statements to estimate contrasts for the factorial model become

```
lsmDurat = lsmeans(modelFE, ~ fDurat)
summary(contrast(lsmDurat,  list(DurationDiff=c(-1, 1))), infer=c(T,T))
lsmSpeed = lsmeans(modelFE, ~ fSpeed)
summary(contrast(lsmSpeed, list(SpeedDiff=c(-1, 1))), infer=c(T,T))
lsmPedal = lsmeans(modelFE, ~ fPedal)
summary(contrast(lsmPedal, list(PedalDiff=c(-1, 1))), infer=c(T,T))
```

and give output identical to that of Table 12.13.

Table 12.15 R code for calculating and plotting the adjusted observations—Exercise bicycle experiment

```
bike.data = read.table("data/exercise.bicycle.txt", header=T)
# Create factor variables, plus numeric variable TLevel for plotting
bike.data = within(bike.data,
    {fDay=factor(Day); fSubject=factor(Subject); fDurat=factor(Durat);
     fSpeed=factor(Speed); fPedal=factor(Pedal); fTrtmt=factor(Trtmt);
     # Vbl TLevel with treatment levels 1:8 for equispaced plotting;
     TLevel = 4*(Durat-1) + 2*(Speed-1) + (Pedal-1) })

# ANOVA: treatment combinations
modelTC = lm(Pulse ~ fDay + fSubject + fTrtmt, data=bike.data)

# Plotting data adjusted for row and column effects
modelTC$coefficients # Display all model coefficient estimates

   (Intercept)       fDay2       fDay3       fDay4       fDay5       fDay6
       41.7500     -2.4286     -1.4196     -3.1875     -1.4732     -2.3571

         fDay7       fDay8   fSubject2   fSubject3  fTrtmt112  fTrtmt121
       -3.7857     -3.9911     -6.7500     -5.2500    -1.2143   -14.7054

     fTrtmt122   fTrtmt211   fTrtmt212   fTrtmt221  fTrtmt222
       -9.5714    -10.1875     -4.1339    -19.9018   -12.6429

thetahat = c(0, modelTC$coefficients[2:8]) # Row effect estimates
thetamean = mean(thetahat) # Mean row effect estimate
phihat = c(0, modelTC$coefficients[9:10]) # Column effect estimates
phimean = mean(phihat) # Mean column effect estimate
# Compute adjusted y-values
bike.data$yadj = ( bike.data$Pulse
                      - (thetahat[bike.data$Day] - thetamean)
                      - (phihat[bike.data$Subject] - phimean) )
# Plot adjusted response versus treatment
plot(yadj ~ TLevel, data=bike.data, xaxt="n", xlab="Treatment",
     ylab="y Adjusted")
axis(1, at=bike.data$TLevel, labels=bike.data$fTrtmt)
```

12.9.2 Plots

The statements needed for calculating the standardized residuals and normal scores and for generating the residual plots are also shown in Table 12.12. These are as discussed in Sects. 5.9.1 and 6.9.3. The `plot` statement is used to generate the usual residual plots (not shown here).

The R program in Table 12.15 illustrates how to compute the adjusted observations—namely, the observations adjusted for row and column effect estimates—and how to plot the adjusted observations versus treatment. The data are adjusted using (12.6.17); that is,

$$y^*_{hqi} = y_{hqi} - (\hat{\theta}_h - \hat{\overline{\theta}}_.) - (\hat{\phi}_q - \hat{\overline{\phi}}_.)\,.$$

Toward this end, having saved the fitted row-column design model as `modelTC`, the statement `modelTC$coefficients` displays a set of least squares solutions to the normal equations. These are shown for example in Table 12.15. The row and column effect estimates $\hat{\theta}_h$ and $\hat{\phi}_q$ are obtained from the column `modelTC$coefficient` of displayed least squares estimates by specifying the appropriate entries, and the mean of each collection of estimates is computed. Given this information, the adjusted observations are computed.

The `plot` function is used to plot the adjusted observations versus treatment. The adjusted observations are actually plotted against treatment levels 1–8, so the treatment combinations are equally spaced on the x-axis of the plot, but these equispaced levels are replaced by the treatment combination labels 111–222. The R plot is analogous to that in Fig. 12.1 (p. 415) and is not shown here.

Exercises

1. **Randomization**

   (a) Randomize the plan in Table 12.1 (p. 400) so that it can be used for an experiment with seven subjects, each being assigned a sequence of four out of a possible seven antihistamines over four time periods.
   (b) Discuss whether or not one would need to include a carryover effect in the model, or whether this could be avoided through the design of the experiment.

2. **Latin squares**

   (a) Show that there is only one standard 3×3 Latin square. (Hint: Given the letters in the first row and the first column, show that there is only one way to complete the Latin square.)
   (b) Show that there are exactly four standard 4×4 Latin squares.

3. **Sample sizes**

   Consider an experiment to compare 4 degrees of twist in a cotton-spinning experiment with respect to the number of breaks per 100 pounds. A replicated Latin square design is to be used, with time periods and machines being the row and column blocking factors.

   (a) Determine the number s of Latin squares and the number r of observations per degree of twist to include in the experiment if each interval in a simultaneous set of 95% confidence intervals for all pairwise comparisons is to have a minimum significant difference (half-width) of 5 breaks per 100 pounds. The error standard deviation is thought to be at most 6 breaks per 100 pounds. Investigate both the Tukey and the Bonferroni methods.
   (b) Discuss how the resulting design would be randomized.

4. **Youden design randomization**

   (a) Find a Youden square (plan of treatment labels in rows and columns) for 5 treatments in 5 rows and 4 columns.
   (b) Randomize the design found in part (a), assigning the rows to 5 different drying temperatures, the columns to 4 different paint nozzles, and the treatment labels to 5 different paint formulations.

Table 12.16 Latin square design showing treatments and data for the video game experiment

		Day									
		1		2		3		4		5	
Time order	1	1	94	3	100	4	98	2	101	5	112
	2	3	103	2	111	1	51	5	110	4	90
	3	4	114	1	75	5	94	3	85	2	107
	4	5	100	4	74	2	70	1	93	3	106
	5	2	106	5	95	3	81	4	90	1	73

5. **Row–column design randomization**

Consider an experiment to compare 5 protocols with respect to a resting metabolism rate measurement. A row–column design is to be used, blocking on subjects and time periods. Since subjects prefer to stay in a study for a short length of time, only 3 time periods will be used, with each subject assigned a different protocol in each of the 3 time periods. For 10 subjects, the following experimental plan with 10 rows and 3 columns could be used:

	Column				Column		
Row	I	II	III	Row	I	II	III
I	1	2	3	VI	1	2	4
II	2	3	4	VII	2	3	5
III	3	4	5	VIII	3	4	1
IV	4	5	1	IX	4	5	2
V	5	1	2	X	5	1	3

(a) Does this experimental plan have treatments evenly distributed across rows and columns? Explain what you mean by "evenly distributed."
(b) Determine the number of replicates s of this experimental plan, and the corresponding number of observations r per protocol to include in the experiment if each interval in a set of simultaneous 95% confidence intervals for all pairwise comparisons is to have a minimum significant difference (half-width) of 150 units. The error standard deviation is thought to be at most 250 units. Investigate both the Tukey and the Bonferroni methods.
(c) Discuss how the resulting design would be randomized.

6. **Video game experiment**

Professor Robert Wardrop, of the University of Wisconsin, conducted an experiment in 1991 to evaluate in which of five sound modes he best played a certain video game. The first three sound modes corresponded to three different types of background music, as well as game sounds expected to enhance play. The fourth mode had game sounds but no background music. The fifth mode had no music or game sounds. Denote these sound modes by the treatment factor levels 1–5, respectively.

The experimenter observed that the game required no warm up, that boredom and fatigue would be a factor after 4–6 games, and that his performance varied considerably on a day-to-day basis. Hence, he used a Latin square design, with the two blocking factors being "day" and "time order of the game." The response measured was the game score, with higher scores being better. The design and resulting data are given in Table 12.16.

(a) Write down a possible model for these data and check the model assumptions. If the assumptions appear to be approximately satisfied, then answer parts (b)–(f).
(b) Plot the adjusted data and discuss the plot.
(c) Complete an analysis of variance table.
(d) Evaluate whether blocking was effective.
(e) Construct simultaneous 95% confidence intervals for all pairwise comparisons, as well as the "music versus no music" contrast

$$\frac{1}{3}(\tau_1 + \tau_2 + \tau_3) - \frac{1}{2}(\tau_4 + \tau_5)$$

and the "game sound versus no game sound" contrast

$$\frac{1}{4}(\tau_1 + \tau_2 + \tau_3 + \tau_4) - \tau_5 \,.$$

(f) What are your conclusions from this experiment? Which sound mode(s) should Professor Wardrop use?

7. **Video game experiment, continued**

Suppose that in the video game experiment of Exercise 6, Professor Wardop had run out of time and that only the first four days of data had been collected. The design would then have been a Youden design. Repeat parts (c), (e), and (f) of Exercise 6. Do your conclusions remain the same? Is this what you expected? Why or why not?

8. **Air freshener experiment**

A. Cunningham and N. O' Connor (1968, *European Journal of Marketing* 2, 147–151) conducted a two-replicate Latin square design to compare the effects of four price-and-display treatments on the sales of a brand of air fresheners. Treatments 1–3 corresponded to high, middle, and low prices, respectively, and each had an extra display. Treatment 4 corresponded to the middle price and no extra display. The response variable was the unit sales for a one-week period. The experiment involved two blocking factors defined by stores ($c = 8$ levels) and one-week periods ($b = 4$ levels). The design and data are given in Table 12.17.

(a) Factors such as product location and shelf stocking could affect sales. Discuss how these factors might be controlled in such an experiment.
(b) Check the model assumptions.
(c) Plot the adjusted data and comment on the results.
(d) Complete an analysis of variance table.

Table 12.17 Latin square design and data for the air freshener experiment

Week	Store															
	1		2		3		4		5		6		7		8	
1	2	31	1	23	3	12	4	3	1	10	3	30	2	23	4	14
2	1	19	4	16	2	14	3	4	2	21	4	25	3	17	1	14
3	4	15	3	30	1	12	2	6	3	12	1	47	4	5	2	3
4	3	16	2	27	4	5	1	11	4	12	2	38	1	13	3	6

Source Data adapted from Cunningham and O'Connor, European Journal of Marketing, Copyright © Emerald Group Publishing Limited

(e) Evaluate whether blocking was effective.
(f) Test for equality of treatment effects using a 5% significance level.
(g) Construct simultaneous 95% confidence intervals for all pairwise comparisons of the treatments. What would you recommend for the sales of air fresheners if the results of this experiment are still valid today?

9. **Air freshener experiment, continued**

Suppose that the air freshener experiment of Exercise 8 had to be stopped after only 3 weeks. The resulting design would then be a replicated Youden design. Repeat parts (d)–(g) of Exercise 8. Do your conclusions remain the same? Is this what you expected? Why or why not?

10. **Quantity perception experiment**

An experiment was run in 1996 by M. Gbenado, A. Veress, L. Heimenz, J. Monroe, and S. Yu to investigate the effect of color on the perception of quantity. Subjects were recruited at random from the student population. A number of small candies of a specific color were tipped onto a flat tray. A subject was allowed to view the tray for 3 sec and then asked to make a guess as to the number of candies on the tray. The response was the difference between the actual number of candies on the tray and the number guessed by the subject.

The treatment factors of interest were "actual number of candies on the tray" and "color." The selected levels were 17, 29, and 41 for the treatment factor "number", and yellow, orange, brown for the factor "color." The experimenters decided that each subject should view all nine treatment combinations, and they based their design on 9×9 Latin squares.

We consider only part of the original study, constituting a 2-replicate Latin square, as shown in

Table 12.18 2-replicate Latin square design and data in parentheses for the quantity perception experiment—data are 'true number' minus 'guessed number'

Subj	Time Order								
	1	2	3	4	5	6	7	8	9
1	23 (4)	22 (−3)	11 (0)	12 (−3)	32 (−1)	13 (−3)	31 (−6)	33 (−9)	21 (−1)
2	12 (2)	31 (16)	32 (21)	21 (9)	22 (4)	33 (16)	23 (9)	11 (2)	13 (2)
3	21 (4)	23 (−1)	12 (7)	13 (−5)	33 (−1)	11 (−13)	32 (−9)	31 (−19)	22 (−16)
4	32 (21)	12 (4)	22 (3)	31 (11)	13 (0)	21 (4)	11 (0)	23 (4)	33 (11)
5	31 (7)	11 (−2)	21 (2)	33 (3)	12 (−3)	23 (3)	13 (−4)	22 (−5)	32 (−7)
6	11 (3)	33 (7)	31 (14)	23 (11)	21 (12)	32 (17)	22 (10)	13 (5)	12 (0)
7	22 (11)	21 (14)	13 (0)	11 (1)	31 (16)	12 (1)	33 (13)	32 (14)	23 (7)
8	13 (7)	32 (16)	33 (16)	22 (4)	23 (4)	31 (16)	21 (14)	12 (7)	11 (2)
9	33 (21)	13 (2)	23 (10)	32 (24)	11 (6)	22 (13)	12 (2)	21 (8)	31 (20)
10	33 (16)	31 (20)	22 (6)	21 (6)	11 (7)	23 (6)	12 (2)	13 (3)	32 (14)
11	12 (2)	22 (4)	32 (11)	13 (2)	21 (9)	33 (1)	23 (4)	31 (−4)	11 (7)
12	13 (−4)	23 (−11)	33 (−4)	11 (−3)	22 (4)	31 (11)	21 (−1)	32 (1)	12 (−3)
13	21 (4)	12 (−1)	11 (2)	32 (11)	31 (11)	13 (−3)	33 (1)	22 (−1)	23 (11)
14	22 (2)	13 (−7)	12 (−9)	33 (8)	32 (−2)	11 (−6)	31 (−9)	23 (4)	21 (2)
15	31 (21)	32 (21)	23 (14)	22 (14)	12 (4)	21 (16)	13 (0)	11 (5)	33 (11)
16	11 (2)	21 (9)	31 (21)	12 (6)	23 (9)	32 (18)	22 (9)	33 (16)	13 (2)
17	32 (6)	33 (6)	21 (−1)	23 (−1)	13 (2)	22 (4)	11 (−3)	12 (−1)	31 (6)
18	23 (4)	11 (2)	13 (2)	31 (11)	33 (6)	12 (2)	32 (6)	21 (−1)	22 (4)

Table 12.18. Subjects represent the row blocks, and time order represents the column blocks. The treatment combinations have been coded as follows:

1 = (17, yellow)	2 = (17, orange)	3 = (17, brown)
4 = (29, yellow)	5 = (29, orange)	6 = (29, brown)
7 = (41, yellow)	8 = (41, orange)	9 = (41, brown)

(a) The experiment was conducted in a busy hallway at The Ohio State University. Subjects were recruited from the population of noncolorblind students walking past the table. Recruited subjects were not allowed to view the experiment in progress with previous subjects, but they were paid for their participation in candies. Discuss whether or not the subjects in this study are likely to be representative of some larger population of subjects. Are the conclusions of the study likely to be relevant to noncolorblind people in general?

(b) Fit a model that includes the effects of the two blocking factors, "subject" and "time-order", the treatment effect, and the treatment×time-order interaction. By looking at a computer calculated analysis of variance table, verify that the interaction (adjusted for subjects) can be measured with the full set of $(v-1)^2 = 64$ degrees of freedom.

(c) For the model that includes treatment×time-order interaction, check whether the residuals are approximately normally distributed and whether they have approximately the same variance for each treatment. Do you prefer to use the original response variable "guessed number" or the transformed response "square root of guessed number" or "(true number − guessed number)/(true number)" or some other transformation?

(d) Present an analysis of variance table and test any hypotheses that you think are of interest. State your conclusions.

(e) Rewrite your model in terms of main effects and interactions of the two treatment factors. Redo your analysis of variance table. What can you conclude from the experiment?

(f) If you were to plan a followup experiment, what would you wish to study? Write up a checklist for such an experiment.

11. **Caffeine experiment**

An experiment was run by Lisa Carpinello in 2001 to study the effects of caffeine on a subject's blood pressure readings. Two treatment conditions were studied; subject abstaining from caffeine for two hours in advance of the blood pressure readings (level 0), and subject consuming 12 ounces of coffee with one tablespoon of non-dairy creamer one hour before the readings (level 1).
The investigator measured the subject's blood pressure at 9am and 2pm each day for eight days, with the subject sitting quietly for five minutes before the readings were taken. The ordered pairs of treatments were randomly assigned to the days in such a way that on four of the days, the treatment order was level 0 at 9am, level 1 at 2pm, and, on the other four days, the treatment order was reversed (i.e. 1 at 9am, 0 at 2pm). The design can be represented as a randomized 4-replicate

Table 12.19 Latin square design and systolic blood pressure readings (mm Hg) for the caffeine experiment

	Time: 9am	Time: 2pm		Time: 9am	Time: 2pm
Day 1	1 121	0 118	Day 5	0 126	1 130
Day 2	1 123	0 119	Day 6	1 129	0 123
Day 3	0 111	1 117	Day 7	1 127	0 118
Day 4	0 120	1 129	Day 8	0 121	1 130

Latin square design with rows representing days and columns time of day. The design and data are shown in Table 12.19.

(a) Plot the data and comment on the results.
(b) Write down a model for this experiment and check the assumptions on your model.
(c) Complete an analysis of variance table and evaluate whether blocking was effective.
(d) Test for equality of treatment effects using a 5% significance level.
(e) Construct a 95% confidence interval for comparing the treatments, and interpret the results.

12. **Golf driver experiment**

An experiment was conducted by Dale Meyer in 2001 to compare four different golf clubs (drivers) in terms of the distance golf balls travel when hit. Two expensive drivers (coded 1 and 2) and two cheaper ones (coded 3 and 4) were used. Four golfers each hit four shots in each of four rounds, each golfer using each of the four drivers once per round, and balancing for the order in which each driver was used. A different Latin square was used as the design in each round. The final design is shown in Table 12.20 with driver as the treatment factor, and rows and columns defined by golfer and order. Notice that, not only do the combinations of golfer/order/driver constitute a Latin square in each round, but so do combinations of round/order/driver for each golfer, and so do combinations of round/golfer/driver for each order. So the design is extremely well balanced and this allows the golfer×driver interaction to measured without adjusting for order or round. The experimenter was interested in comparing the driver effects and the golfer×driver interaction. No other interaction effects were anticipated. If the golfer×driver interaction were to be significant, then the experimenter also wanted to compare the driver effects separately for each golfer. For each observation, a golf ball was hit in the golf simulator of a golf store, and the simulated distances of the shots (in yards) were recorded and are shown in Table 12.20.

Table 12.20 Driver-distance (yards) pairs for the golf driver experiment

Round	Golfer	Order							
		1		2		3		4	
1	1	1	175	2	90	3	176	4	102
1	2	2	157	1	149	4	166	3	140
1	3	3	196	4	182	1	199	2	158
1	4	4	147	3	216	2	197	1	134
2	1	2	147	1	191	4	137	3	165
2	2	1	124	2	135	3	171	4	79
2	3	4	180	3	195	2	206	1	184
2	4	3	167	4	161	1	143	2	121
3	1	3	187	4	166	1	176	2	181
3	2	4	135	3	123	2	168	1	131
3	3	1	197	2	165	3	200	4	171
3	4	2	216	1	158	4	181	3	180
4	1	4	176	3	192	2	193	1	154
4	2	3	176	4	125	1	95	2	187
4	3	2	150	1	188	4	201	3	192
4	4	1	162	2	170	3	214	4	123

(a) Suggest a model for this experiment. By looking at computer calculated interaction degrees of freedom and the adjusted and unadjusted sums of squares, verify that the golfer×driver interaction can be measured based on the full set of $(v - 1)^2 = 9$ degrees of freedom, and without adjusting for order or round. Test whether the golfer×driver interaction is significantly different from zero using $\alpha = 0.01$.

(b) If there is no significant golfer×driver interaction, compare the driver effects (averaged over golfer, order and round) using simultaneous 99% confidence intervals. Which method of multiple comparisons did you use and why? What can you conclude?

(c) Compare the pairwise effects of the drivers for each golfer, using individual 99% confidence intervals. Which method of multiple comparisons did you use and why? State your overall confidence level and interpret your results.