Incomplete Block Designs

11

11.1 Introduction

When an experiment needs to be run in blocks but, for practical reasons, the block size cannot be a multiple of the number of treatments, then a complete block design (Chap. 10) cannot be used—an *incomplete block design* needs to be used instead. The incomplete block designs discussed in this chapter have block size smaller than the number of treatments, but larger block sizes can be obtained by adding one or more complete set of treatments to every block.

In Sect. 11.2, we discuss basic design issues of block size, randomization and estimability. Then, in Sect. 11.3, three useful and efficient types of incomplete block designs (balanced incomplete block designs, group divisible designs, and cyclic designs) are introduced. Analysis of incomplete block designs is described in Sect. 11.4, including some specific formulae for balanced incomplete block designs and group divisible designs. In Sect. 11.5, we describe and analyze an experiment that was designed as a cyclic group divisible design. Sample-size calculations are discussed in Sect. 11.6, and factorial experiments in incomplete block designs are considered in Sect. 11.7. In general, since not every combination of treatment and block is observed, incomplete block designs are most easily analyzed using computer software. Illustrations are given in Sect. 11.8 by SAS software and in Sect. 11.9 by R.

11.2 Design Issues

11.2.1 Block Sizes

Block sizes are dictated by the availability of groups of similar experimental units. For example, in the breathalyzer experiment examined in Sect. 10.3.1, p. 306, the block size was chosen to be $k = 5$. This choice was made because the pilot experiment indicated that experimental conditions were fairly stable over a time span of five observations taken close together, and also because five observations could be taken by a single technician in a shift. In other experiments, the block size may be limited by the capacity of the experimental equipment, the availability of similar raw material, the length of time that a subject will agree to remain in the study, the number of observations that can be taken by an experimenter before fatigue becomes a problem, and so on. Such restrictions on the block size may result in the blocks being too small for every treatment to be observed the same number of times in every block. The breathalyzer experiment required the comparison of $v = 36$ treatment combinations (twelve different

A. Dean et al., *Design and Analysis of Experiments*,
Springer Texts in Statistics, DOI 10.1007/978-3-319-52250-0_11

Table 11.1 An incomplete block design with $b = 8$, $k = 3$, $v = 8$, $r = 3$

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

alcohol concentrations combined with three air-intake ports) of which only five would be observed per block. Skill was then needed in selecting the best design that would still allow all treatment contrasts to be estimable with high precision.

11.2.2 Design Plans and Randomization

All the designs that we discuss in this chapter are equireplicate; that is, every treatment (or treatment combination) is observed r times in the experiment. These tend to be the most commonly used designs, although nonequireplicate designs are occasionally used in practice.

We use the symbol n_{ih} to denote the number of times that treatment i is observed in block h. In general, it is better to observe as many different treatments as possible in a block, since this tends to decrease the average variance of the contrast estimators. Therefore, when the block size is smaller than the number of treatments, each treatment should usually be observed either once or not at all in a block. Such block designs are called *binary,* and every n_{ih} is either 0 or 1. For most purposes, the best binary designs are those in which pairs of treatments occur together in the same number (or nearly the same number) of blocks. These designs give rise to equal (or nearly equal) lengths of confidence intervals for pairwise comparisons of treatment effects.

There are three stages in designing an experiment with incomplete blocks. The first stage is to obtain as even a distribution as possible of treatment labels within the blocks. This results in an *experimental plan*. The plan in Table 11.1, for example, shows a design with $b = 8$ blocks (labeled I, II, ..., VIII) each of size $k = 3$, which can be used for an experiment with $v = 8$ treatments (labeled $1, \ldots, 8$) each observed $r = 3$ times. The treatment labels are evenly distributed in the sense that no label appears more than once per block and pairs of labels appear together in a block either once or not at all, which is "as equal as possible".

The experimental plan is often called the "design," even though it is not ready for use until the random assignments have been made. There are three steps to the randomization procedure, as follows.

(i) Randomly assign the block labels in the plan to the levels of the blocking factor(s).
(ii) Randomly assign the experimental units in a block to those treatment labels allocated to that block.
(iii) Randomly assign the treatment labels in the plan to the actual levels of the treatment factor.

The randomization procedure is illustrated in the following example.

Example 11.2.1 `Metal alloy experiment`

Suppose an experiment is to be run to compare $v = 7$ compositions of a metal alloy in terms of tensile strength. Further, suppose that only three observations can be taken per day, and that the experiment must be completed within seven days. It may be thought advisable to divide the experiment into blocks, with each day representing a block, since different technicians may work on the experiment on different

Table 11.2 Randomization of an incomplete block design

Block label	Unrandomized design	Block label	Day	Design after step (i)	Design after step (ii)
I	1 2 4	VI	1	6 7 2	2 7 6
II	2 3 5	II	2	2 3 5	5 3 2
III	3 4 6	III	3	3 4 6	4 3 6
IV	4 5 7	I	4	1 2 4	2 1 4
V	5 6 1	V	5	5 6 1	1 5 6
VI	6 7 2	IV	6	4 5 7	4 5 7
VII	7 1 3	VII	7	7 1 3	7 3 1

days and the laboratory temperature may vary from day to day. Thus, an incomplete block design with $b = 7$ blocks of size $k = 3$ and with $v = 7$ treatment labels is needed. The plan shown in the first two columns of Table 11.2 is of the correct size. It is binary, with every treatment appearing 0 or 1 times per block and $r = 3$ times in total. Also, all pairs of treatments occur together in a block exactly once, so the treatment labels are evenly distributed over the blocks. Randomization now proceeds in three steps.

Step (i): The block labels need to be randomly assigned to the 7 days. Suppose we obtain the following pairs of random digits from a random number generator or from Table A.1 and associate them with the blocks:

Random digits:	71	36	65	93	92	02	97
Block labels:	I	II	III	IV	V	VI	VII

Then, sorting the random numbers into ascending order, the blocks of the plan are assigned to the seven days as in columns 3–5 of Table 11.2.

Step (ii): Now we randomly assign time slots within each day to the treatment labels. Again, using pairs of random digits either from a random number generator or from where we left off in Table A.1, we associate the random digits with the treatment labels as we illustrate here for the first three days:

Day:	Day 1			Day 2			Day 3		
Block:	(Block VI)			(Block II)			(Block III)		
Random digits:	50	29	03	65	34	30	74	56	88
Treatment labels:	6	7	2	2	3	5	3	4	6

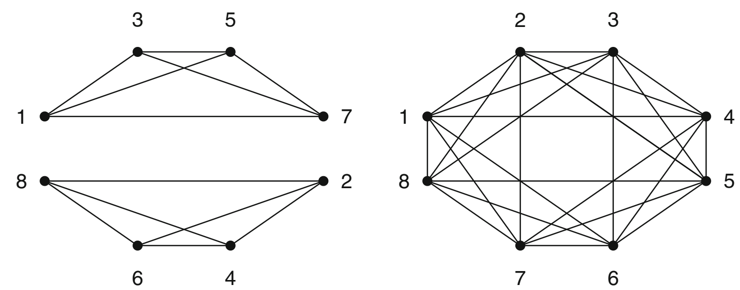
Sorting the random numbers into ascending order for each day separately gives the treatment label order 2, 7, 6 for day 1, and 5, 3, 2 for day 2, and 4, 3, 6 for day 3, and so on. The design after step (ii) is shown in the last column in Table 11.2. A third set of random digits is now required to associate the treatment numbers in the plan with the 7 compositions of metal alloy. □

11.2.3 Estimation of Contrasts

The importance of selecting an experimental plan with an even distribution of treatment labels within the blocks, such as those in Tables 11.1 and 11.2, is to ensure that all treatment contrasts are estimable and that pairwise comparison estimators have similar variances. The plan shown in Table 11.3 is poor. Although all treatments appear $r = 3$ times in the design, some pairs of treatment labels (such as 1 and 3) occur together in two blocks, while other pairs (such as 1 and 2) never appear together. Worse

Table 11.3 A disconnected incomplete block design with $b = 8$, $k = 3$, $v = 8$, $r = 3$

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



(a) Disconnected design of Table 11.3 **(b)** Connected design in Table 11.1

Fig. 11.1 Connectivity graphs to check connectedness of designs

still, is that some blocks contain all the even-numbered treatment labels, and the other blocks contain all the odd-numbered labels. The result is that every pairwise comparison between an even-numbered and an odd-numbered treatment is not estimable. The design is said to be *disconnected*.

Disconnectedness can be illustrated through a *connectivity graph* as follows. Draw a point for each treatment and then draw a line between every two treatments that occur together in any block of the design. The connectivity graph for the disconnected design in Table 11.3 is shown in Fig. 11.1a. Notice that the graph falls into two pieces. There is no line between any of the odd-labeled treatments and the even-labeled treatments.

A design is *connected* if every treatment can be reached from every other treatment via lines in the connectivity graph. The connectivity graph for the connected design in Table 11.1 is shown in Fig. 11.1b and it can be verified that there is a path between every pair of treatments. For example, although treatments 1 and 5 never occur together in a block and so are not connected by a line, there is nevertheless a path from 1 to 4 to 5. *All contrasts in the treatment effects are estimable in a design if and only if the design is connected*. The connectivity graph therefore provides a simple means of checking estimability.

Although disconnected designs will be useful in Chap. 13 for single-replicate ($r = 1$) factorial experiments arranged in blocks, they need never be used for experiments with at least two observations per treatment. All balanced incomplete block designs are connected, and so are most group divisible designs and cyclic designs. These three types of design are described next.

11.3 Some Special Incomplete Block Designs

11.3.1 Balanced Incomplete Block Designs

A *balanced incomplete block design* is a design with v treatment labels, each occurring r times, and with bk experimental units grouped into b blocks of size $k < v$ in such a way that the units within a

Table 11.4 A balanced incomplete block design with $v = 8$, $r = 7$, $b = 14$, $k = 4$, $\lambda = 3$

Block	Treatments				Block	Treatments			
I	1	2	3	4	VIII	2	3	5	8
II	1	2	5	6	IX	2	3	6	7
III	1	2	7	8	X	2	4	5	7
IV	1	3	5	7	XI	2	4	6	8
V	1	3	6	8	XII	3	4	5	6
VI	1	4	5	8	XIII	3	4	7	8
VII	1	4	6	7	XIV	5	6	7	8

block are alike and units in different blocks are substantially different. The plan of the design satisfies the following conditions:

(i) Each treatment label appears either once or not at all in a block (that is, the design is binary).
(ii) Each pair of labels appears together in λ blocks, where λ is a fixed integer.

Block design randomization is carried out as illustrated in Sect. 11.2.2.

All balanced incomplete block designs have the desirable properties that all treatment contrasts are estimable and all pairwise comparisons of treatment effects are estimated with the same variance so that their confidence intervals are all the same length. Balanced incomplete block designs also tend to give the shortest confidence intervals on the average for any large number of contrasts. For these reasons, the balanced incomplete block design is a popular choice among experimenters. The main drawback is that such designs exist only for some choices of v, k, b, and r.

The design in Table 11.2 is a balanced incomplete block design for $v = 7$ treatments and $b = 7$ blocks of size $k = 3$. It can be seen that conditions (i) and (ii) are satisfied, with every pair of labels appearing together in exactly $\lambda = 1$ block. A second example of a balanced incomplete block design, prior to randomization, is shown in Table 11.4 for $v = 8$ treatments in $b = 14$ blocks of size $k = 4$. Again, conditions (i) and (ii) are satisfied, this time with $\lambda = 3$.

We can verify that the design in Table 11.1 (p. 350) with $v = b = 8$, $r = k = 3$ is not a balanced incomplete block design. Label 2, for example, appears in one block with each of labels 1, 3, 4, 5, 7, and 8 but never with label 6. The following simple argument shows that no balanced incomplete block design can possibly exist for this size of experiment. In a balanced incomplete block design with $v = b = 8$, $r = k = 3$, label 2, for example, must appear in $r = 3$ blocks in the design, and in each block there are $k-1 = 2$ other labels. So label 2 must appear in a block with a total of $r(k-1) = 6$ other treatment labels. Consequently, if label 2 were to appear λ times with each of the other $v-1 = 7$ labels, then 7λ would have to be equal to $r(k-1) = 6$. This would require that $\lambda = 6/7 = r(k-1)/(v-1)$. Since λ is not an integer, a balanced incomplete block design of this size cannot exist. However, for the size of design in Table 11.4, λ is an integer since $\lambda = r(k-1)/(v-1) = 7(3)/(7) = 3$.

There are three necessary conditions for the existence of a balanced incomplete block design, all of which are easy to check. These are

$$\begin{aligned} vr &= bk\,, \\ r(k-1) &= \lambda(v-1)\,, \\ b &\geq v\,. \end{aligned} \tag{11.3.1}$$

The first condition is satisfied by all block designs with equal replication and equal block sizes. The second condition is obtained by the argument above, and the third condition is called Fisher's inequality. Although the three necessary conditions can be used to verify that a balanced incomplete block design of a given size may exist, they do not guarantee its existence. Lists of balanced incomplete block designs can be found in Cochran and Cox (1957, chapter 11) and Fisher and Yates (1973), or can be obtained by some computer packages (see, for example, `PROC OPTEX` in the SAS software, described in Sect. 11.8.1 and the `ibd` package in R, Sect. 11.9.1).

11.3.2 Group Divisible Designs

A *group divisible design* is a design with $v = g\ell$ treatment labels (for some integers $g > 1$ and $\ell > 1$), each occurring r times, and bk experimental units grouped into b blocks of size $k < v$ in such a way that the units within a block are alike and units in different blocks are substantially different. The plan of the design satisfies the following conditions:

(i) The $v = g\ell$ treatment labels are divided into g groups of ℓ labels—any two labels within a group are called *first associates* and any two labels in different groups are called *second associates*.
(ii) Each treatment label appears either once or not at all in a block (that is, the design is binary).
(iii) Each pair of first associates appears together in λ_1 blocks.
(iv) Each pair of second associates appears together in λ_2 blocks.

Block design randomization is carried out as in Sect. 11.2.2.

It will be seen in Sect. 11.4.5 that the values of λ_1 and λ_2 govern the lengths of confidence intervals for treatment contrasts. Generally, it is preferable to have λ_1 and λ_2 as close as possible, which ensures that the confidence intervals of pairwise comparisons are of similar lengths. Group divisible designs with λ_1 and λ_2 differing by one are usually regarded as the best choice of incomplete block design when no balanced incomplete block design exists.

An example of a group divisible design (prior to randomization) is the experimental plan shown in Table 11.1, p. 350. It has the following $g = 4$ groups of $\ell = 2$ labels:

$$(1, 5), \quad (2, 6), \quad (3, 7), \quad (4, 8).$$

Labels in the same group (first associates) never appear together in a block, so $\lambda_1 = 0$. Labels in different groups (second associates) appear together in one block, so $\lambda_2 = 1$.

A second example is given by the experimental plan in Table 11.5, and it has $g = 4$ groups of $\ell = 3$ labels:

$$(1, 2, 3), \quad (4, 5, 6), \quad (7, 8, 9), \quad (10, 11, 12),$$

and $\lambda_1 = 3$, $\lambda_2 = 1$ (which is not ideal since λ_1 and λ_2 differ by more than 1).

There are four necessary conditions for the existence of a group divisible design with chosen values of $v = g\ell, b, k, r$ namely,

$$\begin{aligned} g\ell r &= bk\,, \\ r(k-1) &= \lambda_1(\ell-1) + \lambda_2\ell(g-1)\,, \\ r &\geq \lambda_1\,, \\ rk &\geq \lambda_2 v\,, \end{aligned}$$

Table 11.5 A group divisible design with $v = 12$, $r = 3$, $b = 6$, $k = 6$, $\lambda_1 = 3$, $\lambda_2 = 1$

Block	Treatments					
I	1	2	3	4	5	6
II	1	2	3	7	8	9
III	1	2	3	10	11	12
IV	4	5	6	7	8	9
V	4	5	6	10	11	12
VI	7	8	9	10	11	12

for integers λ_1 and λ_2.

All group divisible designs with $\lambda_2 = 0$ should be avoided, since not all of the treatment contrasts are estimable. (It can be verified that the disconnected design of Table 11.3 is a group divisible design with groups (1, 3, 5, 7) and (2, 4, 6, 8) and with $\lambda_1 = 2$ and $\lambda_2 = 0$.) Lists of group divisible designs are given by Clatworthy (1973) and in the more recent references listed by Sinha (1991). Good block designs (which may or may not be group divisible) can be obtained by some computer packages (see, for example, `PROC OPTEX` in SAS software, Sect. 11.8.1 and the `ibd` package in R, Sect. 11.9.1).

11.3.3 Cyclic Designs

A *cyclic design* is a design with v treatment labels, each occurring r times, and with bk experimental units grouped into $b = v$ blocks of size $k < v$ in such a way that the units within a block are alike and units in different blocks are substantially different. The experimental plan, using treatment labels $1, 2, \ldots, v$, can be obtained as follows:

(i) The first block, called the *initial block*, consists of a selection of k distinct treatment labels.
(ii) The second block is obtained from the initial block by *cycling the treatment labels*—that is, by replacing treatment label 1 with 2, 2 with 3, ..., $v - 1$ with v, and v with 1. The third block is obtained from the second block by cycling the treatment labels once more, and so on until the vth block is reached.

Block design randomization is carried out as in Sect. 11.2.2.

The group divisible design in Table 11.1 is also a cyclic design and has initial block (1, 3, 8). There are three cyclic designs in Table 11.6 all with block size $k = 4$. The first two have initial block (1, 2, 3, 6), but one has $v = 7$ treatment labels and the other has $v = 6$. The third design has initial block (1, 2, 3, 4) and $v = 7$. The first design is also a balanced incomplete block design with $\lambda = 2$. The second design has pairs of treatments occurring together in either $\lambda_1 = 2$ or $\lambda_2 = 3$ blocks, which results in only two possible lengths of confidence intervals for pairwise comparisons, but it is not a group divisible design since the treatment labels cannot be divided into groups of first associates. The third design is less good since pairs of treatments occur in $\lambda_1 = 1$ or $\lambda_2 = 2$ or $\lambda_3 = 3$ blocks, resulting in three different lengths of confidence interval for pairwise comparisons.

A cyclic design can have as many as $v/2$ different values of λ_i, yielding as many as $v/2$ different lengths of confidence intervals for pairwise comparisons of treatment effects. Again, if no balanced incomplete block design exists, the best designs are usually regarded as those with two values of λ_i which differ by one.

Table 11.6 Cyclic designs with $k = 4$ generated by (1, 2, 3, 6) for $v = 7$ and $v = 6$, and generated by (1, 2, 3, 4) for $v = 7$

Design 1 $v = 7$		Design 2 $v = 6$		Design 3 $v = 7$	
Block	Treatments	Block	Treatments	Block	Treatments
1	1 2 3 6	1	1 2 3 6	1	1 2 3 4
2	2 3 4 7	2	2 3 4 1	2	2 3 4 5
3	3 4 5 1	3	3 4 5 2	3	3 4 5 6
4	4 5 6 2	4	4 5 6 3	4	4 5 6 7
5	5 6 7 3	5	5 6 1 4	5	5 6 7 1
6	6 7 1 4	6	6 1 2 5	6	6 7 1 2
7	7 1 2 5			7	7 1 2 3

Some cyclic designs have duplicate blocks, such as that with $v = 8$ and initial block (1, 4, 5, 8). These designs are useful when fewer than v blocks are required, since duplicate blocks can be ignored. Otherwise, designs with distinct blocks are usually better. Lists of cyclic designs are given by John et al. (1972), John (1981), and Lamacraft and Hall (1982).

11.4 Analysis of General Incomplete Block Designs

11.4.1 Contrast Estimators and Multiple Comparisons

The standard block–treatment model for the observation on treatment i in block h in a binary incomplete block design is

$$\begin{gathered} Y_{hi} = \mu + \theta_h + \tau_i + \epsilon_{hi} , \\ \epsilon_{hi} \sim N(0, \sigma^2) , \\ \epsilon_{hi}\text{'s are mutually independent} , \\ h = 1, \ldots, b\,; \quad i = 1, \ldots, v\,; \quad (h, i) \text{ in the design.} \end{gathered} \tag{11.4.2}$$

The model, which assumes no block–treatment interaction, is almost identical to block–treatment model (10.4.1) for the randomized block design. The only difference is the phrase "(h, i) in the design", which means that the model is applicable only to those combinations of block h and treatment i that are actually observed. The phrase serves as a reminder that not all treatments are observed in each block.

For every experiment, the assumptions on the model should be checked. However, when blocks do not contain every treatment, it is difficult to check the assumption of no block–treatment interaction by plotting the data block by block, as was recommended for complete block designs in Sect. 10.7. Thus, it is preferable that an incomplete block design be used only when there are good reasons for believing that treatment differences do not depend on the level of the blocking factor(s).

The least squares estimators for the treatment parameters in the model for an incomplete block design must include an adjustment for blocks, since some treatments may be observed in "better" blocks than others. This means that the least squares estimator for the pairwise comparison $\tau_p - \tau_i$ is *not* the *unadjusted estimator* $\overline{Y}_{.p} - \overline{Y}_{.i}$ as it would be for a randomized complete block design. For example, if metal alloys 2 and 7 were to be compared via the balanced incomplete block design in the last column of Table 11.2, we see that alloy 2 is observed on days 1, 2, and 4, and alloy 7 is observed on days 1, 6, and 7. If we were to use $\overline{Y}_{.2} - \overline{Y}_{.7}$ to estimate $\tau_2 - \tau_7$, it would be biased, since

$$\begin{aligned} E[\overline{Y}_{.2} - \overline{Y}_{.7}] &= E[\frac{1}{3}(Y_{12} + Y_{22} + Y_{42}) - \frac{1}{3}(Y_{17} + Y_{67} + Y_{77})] \\ &= \frac{1}{3}(3\mu + \theta_1 + \theta_2 + \theta_4 + 3\tau_2) - \frac{1}{3}(3\mu + \theta_1 + \theta_6 + \theta_7 + 3\tau_7) \\ &= (\tau_2 - \tau_7) + \frac{1}{3}(\theta_2 + \theta_4 - \theta_6 - \theta_7) \\ &\neq (\tau_2 - \tau_7). \end{aligned}$$

If the experimental conditions were to change over the course of the experiment in such a way that observations on the first few days tended to be higher than observations on the last few days, then θ_2 and θ_4 would be larger than θ_6 and θ_7. If the two alloys do not differ in their tensile strengths, then $\tau_2 = \tau_7$, but the above calculation shows that $\overline{Y}_{.2} - \overline{Y}_{.7}$ would nevertheless be expected to be large. This could cause the experimenter to conclude erroneously that alloy 2 was stronger than alloy 7. Consequently, any estimator for $\tau_2 - \tau_7$ must contain an adjustment for the days on which the alloys were observed.

A general formula for a set of least squares solutions for the parameters τ_i in the block–treatment model (11.4.2) adjusted for block differences can be shown to be

$$r(k-1)\hat{\tau}_i - \sum_{p \neq i} \lambda_{pi}\hat{\tau}_p = kQ_i\,, \qquad \text{for } i = 1, \ldots, v, \tag{11.4.3}$$

when $\sum_i \hat{\tau}_i = 0$, $\sum_h \hat{\theta}_h = 0$ are used as the added equations (similar to Sect. 3.4.3). Here, λ_{pi} is the number of blocks containing both treatments p and i, and a formula for calculating Q_i will be given in (11.4.7) in the next section. However, except in special cases, such as that of the balanced incomplete block design, expression (11.4.3) is difficult to solve for the individual τ_i and is usually left for statistical software to calculate. Although the individual τ_i are not uniquely estimable, all contrasts $\sum_i c_i\tau_i$ are estimable if the design is connected, (see Sect. 11.2.3).

The Bonferroni and Scheffé methods of multiple comparisons can be used for simultaneous confidence intervals of estimable contrasts in all incomplete block designs. The method of Tukey is applicable for balanced incomplete block designs, and it is believed to be conservative (true value of α smaller than stated) for other incomplete block designs, but this has not yet been proven. Dunnett's method can be used in balanced incomplete block designs but not in other incomplete block designs without modification to our tables. For each method, the formula for a set of $100(1-\alpha)\%$ simultaneous intervals is

$$\sum_{i=1}^{v} c_i\tau_i \in \left(\sum_{i=1}^{v} c_i\hat{\tau}_i \pm w \sqrt{\widehat{\text{Var}}\left(\sum_{i=1}^{v} c_i\hat{\tau}_i \right)} \right) \tag{11.4.4}$$

exactly as in Sect. 4.4. The correct least squares estimate $\sum c_i\hat{\tau}_i$, as well as the estimated variance $\widehat{\text{Var}}(\sum c_i\hat{\tau}_i)$ and the error degrees of freedom can be obtained from computer software for the design being used.

Table 11.7 Analysis of variance table for a binary incomplete block design with b blocks of size k, and v treatment labels appearing r times

Source of variation	Degrees of freedom	Sum of squares	Mean square	Ratio
Blocks (adj)	$b-1$	$ss\theta_{\text{adj}}$	$ms\theta_{\text{adj}}$	–
Blocks (unadj)	$b-1$	$ss\theta$	–	–
Treatments (adj)	$v-1$	ssT_{adj}	msT_{adj}	$\frac{msT_{\text{adj}}}{msE}$
Error	$bk-b-v+1$	ssE	msE	
Total	$bk-1$	$sstot$		

Formulae

$ss\theta = \sum_{h=1}^{b} B_h^2/k - G^2/(bk)$	$ssE = sstot - ss\theta - ssT_{\text{adj}}$
$ssT_{\text{adj}} = \sum_{i=1}^{v} Q_i\hat{\tau}_i$	$sstot = \sum_{h=1}^{b}\sum_{i=1}^{v} n_{hi}y_{hi}^2 - G^2/(bk)$
$Q_i = T_i - \sum_{h=1}^{b} n_{hi}B_h/k$.	$ss\theta_{\text{adj}} = sstot - ssE - \left(\sum_{i=1}^{v} T_i^2/r - G^2/(bk)\right)$

11.4.2 Analysis of Variance

For a connected incomplete block design with block–treatment model (11.4.2), the four rows in the center section of Table 11.7 show an outline of the analysis of variance obtained when the block factor is entered into the model before the treatment factor; the formulae are listed in the bottom section of the table. The unadjusted sum of squares for blocks, $ss\theta$, is calculated in a similar way to the block sum of squares in a complete block design; that is

$$ss\theta = \sum_{h=1}^{b} B_h^2/k - G^2/(bk)\,, \tag{11.4.5}$$

where B_h is the sum of all observations in block h, and G represents the "grand total" of all the observations. The *sum of squares for treatments adjusted for blocks* (i.e. adjusted for the fact that not all treatments are in every block and some blocks are better than others) is

$$ssT_{\text{adj}} = \sum_{i=1}^{v} Q_i\hat{\tau}_i\,, \tag{11.4.6}$$

where $\hat{\tau}_i$ is the least squares solution for τ_i obtained from (11.4.3), and Q_i is the ith *adjusted treatment total*; that is,

$$Q_i = T_i - \frac{1}{k}\sum_{h=1}^{b} n_{hi}B_h\,, \tag{11.4.7}$$

where n_{hi} is 1 if treatment i is observed in block h and zero otherwise, B_h is defined above, and T_i is the sum of all observations on treatment i. We could write the three quantities T_i, B_h and G as $y_{.i}$ and $y_{h.}$ and $y_{..}$ in the usual way. The reason for changing notation is as a reminder that some of the y_{hi} are not actually observed and the quantities T_i, B_h and G are more accurately written as

$$T_i = \sum_{h=1}^{b} n_{hi}y_{hi}\,,\quad B_h = \sum_{i=1}^{v} n_{hi}y_{hi}\,,\quad \text{and}\quad G = \sum_{h=1}^{b}\sum_{i=1}^{v} n_{hi}y_{hi}\,.$$

The sum of squares for error, ssE, is

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

where $sstot$ is defined as usual as

$$sstot = \sum_{h=1}^{b} \sum_{i=1}^{v} n_{hi} y_{hi}^2 - G^2/(bk) .$$

Also as usual, the number of degrees of freedom for treatments is $v - 1$, the number of degrees of freedom for blocks is $b - 1$, and the number of degrees of freedom for error can be obtained by subtraction as

$$df = (n - 1) - (b - 1) - (v - 1) = bk - b - v + 1 , \tag{11.4.8}$$

where $n = bk$ is the total number of observations.

A test of H_0^T :{all τ_i are equal} against H_A^T :{at least two of the τ_i's differ} is given by the decision rule

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

for some chosen significance level α, where $msT_{\text{adj}} = ssT_{\text{adj}}/(v-1)$, and $msE = ssE/(bk-b-v+1)$.

If evaluation of blocking for the purpose of planning future experiments is required, the quantity $ss\theta$ in (11.4.5) *is not the correct value* to use. It has not been adjusted for the fact that every block does not contain an observation on every treatment. In order to evaluate blocks, we would need the *adjusted block sum of squares*, $ss\theta_{\text{adj}}$, whose formula is listed as the last entry in Table 11.7. Some computer packages will give this value under the heading "adjusted" or "Type III" sum of squares. If the program does not automatically generate the adjusted value, it can be obtained from a "sequential sum of squares" by entering treatments in the model before blocks.

11.4.3 Analysis of Balanced Incomplete Block Designs

The set of least squares solutions given in (11.4.3) for the treatment parameters τ_i in the block–treatment model (11.4.2) for the balanced incomplete block design have a simple form:

$$\hat{\tau}_i = \frac{k}{\lambda v} Q_i , \quad \text{for} \quad i = 1, \ldots, v , \tag{11.4.9}$$

where λ is the number of times that every pair of treatments occurs together in a block, and Q_i is the adjusted treatment total, given in (11.4.7). Thus, the sum of squares for treatments adjusted for blocks in (11.4.6) becomes

$$ssT_{\text{adj}} = \sum_{i=1}^{v} \frac{k}{\lambda v} Q_i^2 , \tag{11.4.10}$$

and the least squares estimator of contrast $\sum c_i \tau_i$ is

$$\sum_{i=1}^{v} c_i \hat{\tau}_i = \frac{k}{\lambda v} \sum_{i=1}^{v} c_i Q_i . \tag{11.4.11}$$

It can be shown that the corresponding variance can be calculated as

$$\text{Var}\left(\sum_{i=1}^{v} c_i \hat{\tau}_i\right) = \sum_{i=1}^{v} c_i^2 \left(\frac{k}{\lambda v}\right) \sigma^2 . \qquad (11.4.12)$$

The Bonferroni, Scheffé, Tukey, and Dunnett methods of multiple comparisons can all be used for balanced incomplete block designs with degrees of freedom for error equal to $df = bk - b - v + 1$. The general formula (11.4.4) for simultaneous $100(1-\alpha)\%$ confidence intervals for a set of contrasts $\Sigma c_i \tau_i$ becomes

$$\sum_{i=1}^{v} c_i \tau_i \in \left(\frac{k}{\lambda v} \sum_{i=1}^{v} c_i Q_i \pm w \sqrt{\sum_{i=1}^{v} c_i^2 \left(\frac{k}{\lambda v}\right) msE}\right) , \qquad (11.4.13)$$

where the critical coefficients for the four methods are, respectively,

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

For testing m hypotheses of the general form $H_0 : \sum c_i \tau_i = 0$ against the corresponding alternative hypotheses $H_a : \sum c_i \tau_i \neq 0$, at overall significance level α, the decision rule using Bonferroni's method for preplanned contrasts is

$$\text{reject } H_0 \text{ if } \frac{ssc_{\text{adj}}}{msE} > F_{1,bk-b-v+1,\alpha/m} , \qquad (11.4.14)$$

and the decision rule using Scheffé's method is

$$\text{reject } H_0 \text{ if } \frac{ssc_{\text{adj}}}{msE} > (v-1)F_{1,bk-b-v+1,\alpha} , \qquad (11.4.15)$$

where

$$\frac{ssc_{\text{adj}}}{msE} = \frac{(\sum c_i \hat{\tau}_i)^2}{\left(\frac{k}{\lambda v}\right)\; (\sum c_i^2)\; msE} = \frac{k(\sum c_i Q_i)^2}{\lambda v(\sum c_i^2) msE} , \qquad (11.4.16)$$

(cf. Sections 4.3.3 and 6.7.2).

As for the case of equireplicate completely randomized designs, two contrasts $\Sigma c_i \tau_i$ and $\Sigma d_i \tau_i$ are orthogonal in a balanced incomplete block design if $\Sigma c_i d_i = 0$. The adjusted treatment sum of squares can then be written as a sum of adjusted contrast sums of squares for a complete set of $(v-1)$ orthogonal contrasts. An example is given in the following section.

11.4.4 A Real Experiment—Detergent Experiment

An experiment to compare dishwashing detergent formulations was described by P. W. M. John in the journal *Technometrics* in 1961. The experiment involved three base detergents and an additive. Detergent I was observed with 3, 2, 1, and 0 parts of the additive, giving four treatments, which we will code 1, 2, 3, and 4. Likewise, Detergent II was observed with 3, 2, 1, and 0 parts of the additive, giving an additional four treatments, which we will code 5, 6, 7, and 8. The standard detergent (Detergent III) with no additive served as a control treatment, which we will code as 9.

The experiment took place in a location where three sinks were available. Three people took part in the experiment and were instructed to wash plates at a common rate. An observation was the number of plates washed in a sink before the detergent foam disappeared. A block consisted of three observations, one per sink. The refilling of the three sinks with water and detergent constituted the beginning of a new block. The amount of soil on the plates prior to washing was held constant. Differences between blocks were due to differences in the common washing rates, the water temperature, the experimenter fatigue, etc.

A design was required with blocks of size $k = 3$ and $v = 9$ treatment labels. A balanced incomplete block design was selected with $b = 12$ blocks giving $r = bk/v = 4$ observations per treatment and every pair of treatment labels occurring in $\lambda = r(k-1)/(v-1) = 1$ block. We have shown a possible randomization of this design in Table 11.8 together with the data from the original article. The positions within a block show the allocations of the three basins to treatments. The observations are plotted against treatment in Fig. 11.2, ignoring the block from which the observation was collected.

Since each pair of treatments occurs together in only one block ($\lambda = 1$), a graphical approach for the evaluation of block–treatment interaction cannot be used. However, it appears from Fig. 11.2 that block differences, block–treatment interaction effects, and random error variability must all be rather small compared with the large detergent differences. A plot of the adjusted data is described below.

Table 11.8 Design and number of plates washed for the detergent experiment

Block	Treatments			Plates washed		
1	3	8	4	13	20	7
2	4	9	2	6	29	17
3	3	6	9	15	23	31
4	9	5	1	31	26	20
5	2	7	6	16	21	23
6	6	5	4	23	26	6
7	9	8	7	28	19	21
8	7	1	4	20	20	7
9	6	8	1	24	19	20
10	5	8	2	26	19	17
11	5	3	7	24	14	21
12	3	2	1	11	17	19

Source John (1961).

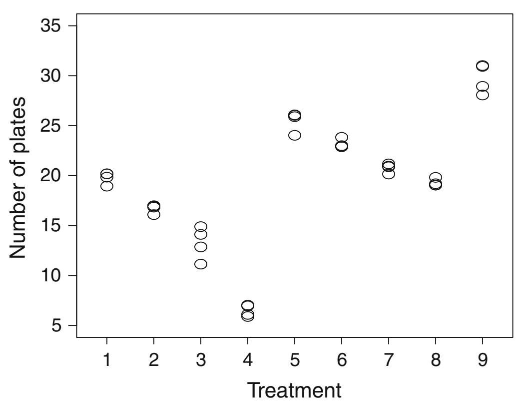


Fig. 11.2 Data plot for the detergent experiment, ignoring block levels

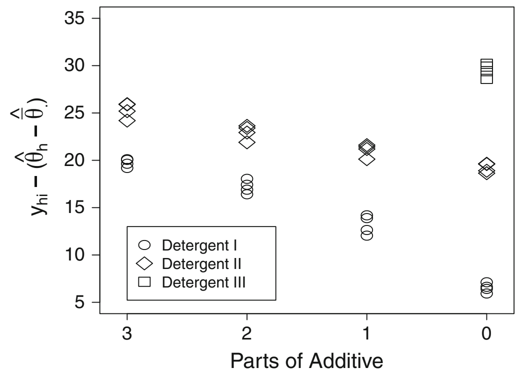


Fig. 11.3 Plot of adjusted observations for the detergent experiment

Block–treatment model (11.4.2) for an incomplete block design was fitted to the data. The residual plots might lead us to question some of the error assumptions, but there are only 4 observations per treatment and these are all from different blocks, so it is difficult to make a proper assessment. We will proceed with the standard analysis for a balanced incomplete block design, recognizing that the stated significance levels and confidence levels are only approximate.

Plotting the Data Adjusted for Block Effects

In this detergent experiment, the treatment differences are fairly clear from the plot of the raw data in Fig. 11.2. However, if the block effects had been substantial, such a plot of the raw data could have painted a muddled picture. In such cases, the picture can be substantially improved by adjusting each observation for the block effects before plotting. The observation y_{hi} is adjusted for the block effects as follows,

$$y^*_{hi} = y_{hi} - (\hat{\theta}_h - \widehat{\overline{\theta}}_.)\,,$$

where $(\hat{\theta}_h - \widehat{\overline{\theta}}_.)$ is the least squares estimate of $(\theta_h - \overline{\theta}_.)$. A SAS program that adjusts the observations for block effects and plots the adjusted observations is given in Sect. 11.8.3 and a corresponding R program in Sect. 11.9.3. It should be noted that since the variability due to block effects has been extracted, a plot of the adjusted observations will appear to exhibit less variability than really exists.

For this particular data set, the block differences are very small, so a plot of the adjusted data would provide information similar to that provided by the plot of the raw data in Fig. 11.2. In Fig. 11.3, the observations adjusted for blocks are plotted against "parts of additive" for each base detergent. It appears that the washing power decreases almost linearly as the amount of additive is decreased and also that the standard detergent is superior to the two test detergents.

Analysis

The analysis of variance table, given in Table 11.9, shows the treatment sum of squares and its decomposition into sums of squares for eight orthogonal contrasts. These contrasts are the linear, quadratic, and cubic trends for each of detergents I and II (as the amount of additive decreases), together with the "I Versus II" contrast that compares the effects of detergents I and II averaged over the levels of the additive, and the "control Versus others" contrast comparing the effect of the control detergent and the average effect of the other eight treatments. For example, the linear trend contrast for detergent I is $-3\tau_1 - \tau_2 + \tau_3 + 3\tau_4$, where the contrast coefficients are obtained from Table A.2. The contrast comparing detergents I and II is the difference of averages contrast

Table 11.9 Analysis of variance table for the detergent experiment

Source of variation	Degrees of freedom	Sum of squares	Mean square	Ratio	p-value
Blocks (adj)	11	10.06	0.91	–	–
Blocks (unadj)	11	412.75	–	–	–
Treatments(adj)	8	1086.81	135.85	164.85	0.0001
I linear	1	286.02	286.02	347.08	0.0001
I quadratic	1	12.68	12.68	15.38	0.0012
I cubic	1	0.22	0.22	0.27	0.6092
II linear	1	61.34	61.34	74.44	0.0001
II quadratic	1	0.15	0.15	0.18	0.6772
II cubic	1	0.03	0.03	0.04	0.8520
I vs II	1	381.34	381.34	462.75	0.0001
Control vs others	1	345.04	345.04	418.70	0.0001
Error	16	13.19	0.82		
Total	35	1512.75			

$$\frac{1}{4}(\tau_1 + \tau_2 + \tau_3 + \tau_4) - \frac{1}{4}(\tau_5 + \tau_6 + \tau_7 + \tau_8)\,,$$

and the contrast comparing the control detergent with the others is the difference of averages contrast

$$\tau_9 - \frac{1}{8}(\tau_1 + \tau_2 + \tau_3 + \tau_4 + \tau_5 + \tau_6 + \tau_7 + \tau_8)\,.$$

A set of simultaneous 99% confidence intervals for all treatment contrasts using Scheffé's method of multiple comparisons is given by (11.4.13) with $w_S = \sqrt{8F_{8,16,.01}}$, and $F_{8,16,.01} = 3.89$, $k = 3$, $\lambda = 1$ and $v = 9$. Using the data shown in Table 11.8, we have treatment totals and grand total

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	G
79	67	53	26	102	93	83	77	119	699

and block totals

B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}	B_{11}	B_{12}
40	52	69	77	60	55	68	47	63	62	59	47

Then, from (11.4.7), the first adjusted treatment total is

$$Q_1 = T_1 - \frac{1}{k}[B_4 + B_8 + B_9 + B_{12}] = 79 - \frac{1}{3}[234] = 1.0\,.$$

The other adjusted treatment totals are calculated similarly, giving

Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8	Q_9
1.00	−6.67	−18.67	−38.67	17.67	10.67	5.00	−0.67	30.33

and since $k/(\lambda v) = 3/9$, the least squares estimate of the contrast $\Sigma c_i \tau_i$ given by (11.4.11) is $\Sigma c_i \hat{\tau}_i = \Sigma c_i Q_i / 3$ $(i = 1, 2, \ldots, 9)$. For example, the least squares estimate for the "control versus others" contrast is

$$\sum_{i=1}^{9} c_i\hat{\tau}_i = \hat{\tau}_9 - \frac{1}{8}\sum_{i=1}^{8}\hat{\tau}_i = \frac{1}{3}\left(Q_9 - \frac{1}{8}\sum_{i=1}^{8} Q_i\right) = 11.375\,,$$

with associated estimated variance

$$\widehat{\text{Var}}\left(\sum_{i=1}^{9} c_i\hat{\tau}_i\right) = \sum_{i=1}^{9} c_i^2\left(\frac{k}{\lambda v}\right) msE = \left(1+\frac{8}{64}\right)\left(\frac{3}{9}\right)(0.82) = 0.3075\,,$$

where $msE = 0.82$ is obtained from Table 11.9. Using Scheffé's method of multiple comparisons at overall level 99%, a confidence interval for the control versus others contrast is then

$$11.375 \pm \sqrt{8F_{8,16,0.01}}\sqrt{0.3075} = 11.375 \pm 3.093 = (8.282,\, 14.468),$$

showing that the control detergent washed between 8.3 and 14.5 more plates than the other detergents on average.

The sum of squares for treatments adjusted for blocks is obtained from (11.4.6), p. 358, as

$$ssT_{\text{adj}} = \frac{k}{\lambda v}\sum Q_i^2 = 1086.81\,,$$

and since

$$\frac{msT_{\text{adj}}}{msE} = \frac{1086.81/8}{0.82} = 164.85 > F_{8,16,0.01} = 3.89\,,$$

we reject the hypothesis of no treatment differences.

The eight orthogonal contrasts can be tested simultaneously using the method of Scheffé. For example, the confidence interval for the "control versus others" contrast calculated above as part of a 99% simultaneous set of intervals does not contain zero, so the hypothesis that the control treatment does not differ from the others would be rejected. The overall significance level for all such tests would be $\alpha = 0.01$. Equivalently, the contrasts can be tested by the Scheffé method using the decision rule (11.4.15), p. 360; that is,

$$\text{reject } H_0 : \sum_{i=1}^{v} c_i\tau_i = 0 \text{ if } \frac{ssc_{\text{adj}}}{msE} > 8F_{8,df,.01} = 31.12\,.$$

The ratios ssc_{adj}/msE are provided in Table 11.9. Comparing their values with 31.12, we see that the linear trends are significantly different from zero for each of the base detergents I and II, as are the comparison of detergents I and II on average and the comparison of the control detergent with the average effects of the other 8 treatments. From significance of the linear trend contrasts, coupled with the direction of the trends, one can conclude that detergents I and II are better with larger amounts of additive.

We cannot use the unadjusted block sum of squares to evaluate the usefulness of blocking. We would need to calculate the adjusted block sum of squares as in Table 11.7, p. 358. Using the values in Table 11.9, this is

$$ss\theta_{\text{adj}} = 1512.75 - 13.19 - \left(\frac{1}{4}(79^2 + 67^2 + \cdots + 119^2) - \frac{1}{36}699^2\right) = 10.06\,.$$

So the adjusted block mean square is $ms\theta_{\text{adj}} = 10.06/11 = 0.91$, which is not much larger than the error mean square, $msE = 0.82$, so the blocking did not help with increasing the power of the hypothesis tests. Nevertheless, it was natural to design this experiment as a block design, and the creation of blocks was a wise precaution against changing experimental conditions.

11.4.5 Analysis of Group Divisible Designs

Group divisible designs were described in Sect. 11.3.2, p. 354, and illustrations were shown in Tables 11.1 and 11.5. The least squares solution (obtained with added equations $\sum_i \hat{\tau}_i = 0, \sum_h \hat{\theta}_h = 0$) for the treatment parameters τ_i adjusted for blocks can be shown to be

$$\hat{\tau}_i = \left[\frac{k}{(r(k-1)+\lambda_1)v\lambda_2}\right] \times \left[(v\lambda_2 + (\lambda_1 - \lambda_2))Q_i + (\lambda_1 - \lambda_2)\sum_{(1)} Q_p\right], \tag{11.4.17}$$

where Q_i is the adjusted treatment total as in (11.4.7), p. 358, and where $\sum_{(1)} Q_p$ denotes the sum of the Q_p corresponding to the treatment labels that are the first associates of treatment label i.

The variance of the least squares estimator $\Sigma c_i \hat{\tau}_i$ of an estimable contrast $\Sigma c_i \tau_i$ is

$$\text{Var}\left(\sum_{i=1}^{v} c_i \hat{\tau}_i\right) = \sum_{i=1}^{v} c_i^2 \,\text{Var}(\hat{\tau}_i) + 2\sum_{i=1}^{v-1}\sum_{p=i+1}^{v} c_i c_p \,\text{Cov}(\hat{\tau}_i, \hat{\tau}_p),$$

where, for a group divisible design,

$$\text{Var}(\hat{\tau}_i) = \frac{k[v\lambda_2 + (\lambda_1 - \lambda_2)]}{v\lambda_2[v\lambda_2 + l(\lambda_1 - \lambda_2)]}\sigma^2$$

and

$$\text{Cov}(\hat{\tau}_i, \hat{\tau}_p) = \begin{cases} \frac{k(\lambda_1-\lambda_2)\sigma^2}{v\lambda_2[v\lambda_2+l(\lambda_1-\lambda_2)]}, & \text{if } i \text{ and } p \text{ are first associates,} \\ 0, & \text{if } i \text{ and } p \text{ are second associates.} \end{cases}$$

Using these quantities, the variance of the least squares estimator of the pairwise comparison $\tau_i - \tau_p$ becomes

$$\text{Var}(\hat{\tau}_i - \hat{\tau}_p) = \begin{cases} \frac{2k\sigma^2}{[v\lambda_2+l(\lambda_1-\lambda_2)]}, & \text{if } i \text{ and } p \text{ are first associates,} \\ \frac{2k[v\lambda_2+(\lambda_1-\lambda_2)]\sigma^2}{v\lambda_2[v\lambda_2+l(\lambda_1-\lambda_2)]}, & \text{if } i \text{ and } p \text{ are second associates.} \end{cases}$$

If λ_1 and λ_2 are as close in value as possible, then the variances for the pairwise comparisons will be as close as possible.

The Bonferroni and Scheffé methods of multiple comparisons can be used for group divisible designs. The Tukey method is believed to be conservative (true value of α smaller than stated). The Dunnett method is not available using our tables, since the critical values can be used only for designs in which $\text{Cov}(\hat{\tau}_i, \hat{\tau}_p)$ are equal for all i and p. However, Dunnett intervals can be obtained from many computer packages (see, for example, Fig 11.7, p. 380, and Table 11.24, p. 387). The analysis of variance table for the group divisible design is that given in Table 11.7, p. 358, with $\hat{\tau}_i$ as in (11.4.17) above. An example of an experiment designed as a group divisible design is discussed in Sect. 11.5. SAS and R programs are illustrated in Sects. 11.8 and 11.9 that can be used to analyze any group divisible design.

11.4.6 Analysis of Cyclic Designs

Cyclic designs were described in Sect. 11.3.3, p. 355, and illustrated in Tables 11.1 and 11.6. They are incomplete block designs that may or may not possess the properties of balanced incomplete block designs or group divisible designs. When they do not possess these properties, the least squares solutions $\hat{\tau}_i$ have no simple form and are most easily obtained from computer software. The Bonferroni and Scheffé methods of multiple comparisons can be used for all cyclic designs.

In Sect. 11.5 we reproduce the checklist and analysis of an experiment that was designed as a cyclic group divisible design and, in Sects. 11.8 and 11.9, we illustrate SAS and R computer programs that can be used to analyze any cyclic incomplete block design.

11.5 A Real Experiment—Plasma Experiment

The plasma experiment was run by Ernesto Barrios, Jin Feng, and Richard Kibombo in 1992 in the Engineering Research Center at the University of Wisconsin. The following checklist has been extracted from the experimenters' report. The design used was a cyclic group divisible design. Notice that the experimenters moved step (e) of the checklist forward. They had made a list of all potential sources of variation, but they needed a pilot experiment to help determine which sources they could control and which they could not.

Checklist

(a) **Define the objectives of the experiment**.

In physics, plasma is an ionized gas with essentially equal densities of positive and negative charges. It has long been known that plasma can effect desirable changes in the surface properties of materials.

The purpose of this experiment is to study the effects of different plasma treatments of plastic pipet tips on the capillary action of the pipets. Capillary action concerns the movement of a liquid up the pipet—a small tube. Before a plasma treatment, the capillarity conduct of the tips is too narrow to permit water to move up. Changes in capillary action effected by plasma treatment can be measured by suspending the tip of a vertical pipet into a bed of water and measuring the height of the column of water in the tube.

(e) **Run a pilot experiment**.

At this stage we decided to make a test run to become familiar with the process of setting up and running the experiment, to determine the appropriate treatment factor levels, and to help identify the major sources of variation that could be controlled, and to identify other variables that might affect the response but which could not be controlled.

(b) **Identify all sources of variation.**

From the test run, we determined that pressure and voltage could not both be effectively controlled. More generally, it would be difficult to vary all of the variables initially listed (gas flow rate, type of gas, pressure, voltage, presence or absence of a ground shield, and exposure time of the pipet tips to the ionized gas).

The following factors were potential sources of variation.

- Experimenters. Despite the fact that all of the experimenters were to play certain roles during each run of the experiment, it was noted that most of the variation due to the personnel could be attributed to the person who connects the pipet tips to the gas tube in the ionization chamber and takes the readings of the final response using vernier calipers.
- Room conditions. It was thought that variations in both room temperature and atmospheric pressure could have an effect on response.
- Water purity. If the water used to measure the capillarity has a substantial amount of impurities, especially mineral salts, then the response may be greatly affected, either because of variability in cohesion and adhesion forces of different types of substances, or because of a reaction between the impurities (salts) and the pipet tips.
- Materials. Variability in the quality of both the pipet tips and the gases used is likely to introduce some variation in the response. Within an enclosed room such as a laboratory, the composition of air may vary significantly over time.

Taking into account the results of the pilot run, the following decisions were made.

(i) Treatment factors and their levels.
Scale down the variables of interest to three by keeping both the pressure and voltage constant at 100 mm Torres and 5 volts, respectively, and by keeping the ground shield on. Distilled water will be used to control for impurities in the water. Pipet tips from a single package will be used, so the pipets are more likely to be from the same batch and hence more likely to be homogeneous. The only factors that will make up the various treatment combinations are gas flow rate, type of gas, and exposure time. No attempt will be made to control for variation in the composition or purity of the gases used. (This variation will be subsumed into the error variability).
Set the lower and upper levels of each factor far apart in order to make any (linear) effect more noticeable. Also, we decided to include only 6 of the 8 possible treatment combinations, as shown in Table 11.10.

(ii) Experimental units.
The experimental units are the (combinations of) pipets and time order of observations.

(iii) Blocking factors, noise factors, and covariates.
The two blocking factors are "experimenter" and "day."
(No covariates or noise factors were included.)

(c) **Specify a rule by which to assign the experimental units to the treatments**.

The design will be an incomplete block design with blocks of size three, and three blocks of data will be collected on each of two days. We will use the cyclic design for $v = 6 = b$ and $k = 3 = r$ generated by the treatment labels 1, 4, 5. The labels in the design will be randomly assigned to the six treatment combinations, and the treatments within each block will be randomly ordered.
(The selected cyclic design is shown in Table 11.11 in nonrandomized order so that the cyclic nature of the design can be seen more easily. The design also happens to be a group divisible design (see Exercise 14). The smallest balanced incomplete block design with $v = 6$ and $k = 3$ has $r = 5$ and $b = 10$ and would require more observations.)

(d) **Specify the measurements to be made, the experimental procedure, and the anticipated difficulties**.

The height of the water column will be measured for each pipet. In order to make the measurements as uniform as possible, a device has been constructed consisting of a rectangular sheet of plexiglass with a small hole in which to place the pipet tip. Placing this pipet holder on a water vessel suspends

Table 11.10 The six treatment combinations used in the plasma experiment

Treatment	Factors and levels		
	Type of gas	exposure time (sec)	Gas flow rate (cc/sec)
1	Argon	180	10
2	Air	180	10
3	Argon	180	30
4	Argon	60	30
5	Air	60	30
6	Air	60	10

Table 11.11 Design and data for the plasma experiment

Block	Day	Experimenter	Response (Treatment)		
1	1	Feng	0.482 (1)	0.459 (4)	0.458 (5)
2	1	Barrios	0.464 (2)	0.465 (5)	0.467 (6)
3	1	Kibombo	0.473 (3)	0.472 (6)	0.495 (1)
4	2	Feng	0.283 (4)	0.325 (1)	0.296 (2)
5	2	Barrios	0.410 (5)	0.390 (2)	0.248 (3)
6	2	Kibombo	0.384 (6)	0.239 (3)	0.350 (4)

about 2 mm of the tip of the pipet into the water. After 60 seconds, a mark will be made on the pipet indicating the water level reached. The distance of the mark from the tip of the pipet will be measured using a vernier caliper with tenth of a millimeter precision.
The experimental procedure for each observation is as follows: Place a pipet on the tube through which the plasma will flow, screw in a glass tube, turn on the pump and wait 40 seconds, open the Baratron, open the gas, turn a controller to auto, set the flow to a specified level, turn the pressure controller to auto and set the level, set the voltage, time the treatment, turn off flow and shut off the gas, set the pressure to open, wait until the pressure is less than 20, turn off the Baratron, turn off the pump, unscrew the glass tube, then (wearing a glove) take out the pipet, place the pipet in water (using the device for this purpose), and mark the height of the water column, then go on to the next observation.
Anticipated difficulties: Differences in the way people would mark or measure the water column heights would cause variation. Running the experiment consistently.

(f) **Specify the model**.
(The standard block–treatment model (11.4.2), p. 356, for an incomplete block design was specified.)

(g) **Outline the analysis**.
An analysis of variance test for equality of the treatment effects will be performed. Then confidence intervals for all pairwise comparisons will be obtained, with a simultaneous 95% confidence level using the method of Scheffé. Model assumptions will be evaluated.

(h) **Calculate the number of observations to be taken**.
(The number of observations was limited by the time available.)

(i) **Review the above decisions. Revise if necessary**.
(No revisions were made at this stage.)

Table 11.12 Analysis of variance table for the plasma experiment

Source of variation	Degrees of freedom	Sum of squares	Mean square	Ratio	p-value
Blocks (adj)	5	0.0805	0.0161	–	–
Blocks	5	0.0992	–	–	–
Treatments (adj)	5	0.0196	0.0039	2.99	0.0932
Error	7	0.0092	0.0013		
Total	17	0.1279			

Table 11.13 Analysis of variance for the plasma experiment—day one only

Source	Degrees of freedom	Sum of squares	Mean square	Ratio	p-value
Blocks (adj)	2	0.0001213	0.0000607	364.00	–
Block	2	0.0004029	–	–	–
Treatments (adj)	5	0.0007112	0.0001422	853.40	0.026
Error	1	0.0000002	0.0000002		
Total	8	0.0011142			

Results of the Experiment

During the experiment, an unexpected event occurred. A little tube through which the gas passes was broken, allowing for some leaking of gas. We realized this after our first day's runs and tried to fix this problem the next day, using tape, as a new tube was unavailable. As can be seen from the results, given in Table 11.11, the responses from the last nine runs, corresponding to the second day, were consistently smaller than those from the first nine runs. This underscores the advantage of using time as a blocking factor.

Data Analysis

The analysis of variance is given in Table 11.12. It was obtained via a SAS computer program similar to the one in Table 11.18 in Sect. 11.8. The adjusted block mean square, $ms\theta_{\text{adj}} = 0.0161$, is twelve times larger than the error mean square, so blocking was certainly worthwhile.

The ratio $msT_{\text{adj}}/msE = 2.99$ does not exceed the critical value $F_{5,7,.05} = 3.97$ for testing equality of the treatment effects at the 5% significance level (equivalently, the p-value is greater than 0.05). Based on this result, examination of any individual treatment contrasts may seem unwarranted. However, the broken tube discovered after the first day of runs is an important consideration in this experiment. It is quite possible that the treatments are not the same on day one as on day two, since the broken tube may change the gas flow rate or the type of gas to which the pipet is exposed. So, one must ask the question, "Is there anything to be salvaged from this experiment?"

If the broken tube has in fact changed the treatment effects, and if the breakage occurred after the first day's runs, then it might be most useful to analyze the data for day one or each day separately. The design for each day is no longer a cyclic design or a group divisible design, but it is a connected incomplete block designs and can still be analyzed by computer (see Sects. 11.8 and 11.9).

If a test of the null hypothesis H_0^T of equal treatment effects is conducted separately for each day's data, it can be verified that H_0^T would not be rejected at the 5% significance level for the data collected on day two but would be rejected for the data of day one.

The analysis of variance for day one is shown in Table 11.13. The test ratio is 853.40—which is larger than $F_{5,1,.05} = 230$. The mean square for blocks adjusted for treatments is 0.0000607, which

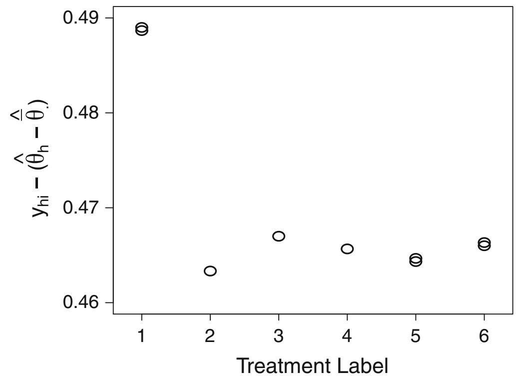


Fig. 11.4 Plasma data adjusted for block effects—day one only

Table 11.14 Pairwise comparisons for the plasma experiment using the Scheffé method and confidence level 95%—day one only

i, p	$\hat{\tau}_i - \hat{\tau}_p$	$\sqrt{\widehat{\text{Var}}(\hat{\tau}_i - \hat{\tau}_p)}$	*msd*	Significant
1, 2	0.025500	0.000646	0.0219	yes
1, 3	0.021833	0.000553	0.0188	yes
1, 4	0.023167	0.000553	0.0188	yes
1, 5	0.024333	0.000471	0.0160	yes
1, 6	0.022667	0.000471	0.0160	yes
2, 3	−0.003667	0.000745	0.0253	
2, 4	−0.002333	0.000745	0.0253	
2, 5	−0.001167	0.000553	0.0188	
2, 6	−0.002833	0.000553	0.0188	
3, 4	0.001333	0.000745	0.0253	
3, 5	0.002500	0.000646	0.0219	
3, 6	0.000833	0.000553	0.0188	
4, 5	0.001167	0.000553	0.0188	
4, 6	−0.000500	0.000646	0.0219	
5, 6	−0.001667	0.000471	0.0160	

is 364 times larger than *msE*, so blocking was helpful for the observations collected on day one. With only one degree of freedom for error, use of residuals to check model assumptions is of little value. Figure 11.4 shows the day-one observations adjusted for block effects, $y_{hi} - (\hat{\theta}_h - \hat{\bar{\theta}}_.)$ plotted against treatment. It appears that treatment 1 (Argon at 10 cc per second for 180 seconds) is very different from the other treatments.

Table 11.14 contains information for applying Scheffé's method of multiple comparisons to the day-one data, using a simultaneous 95% confidence level. The least squares estimates were obtained using SAS software (Sect. 11.8, p. 380, and Fig. 11.8). It can be seen that $\hat{\tau}_i - \hat{\tau}_p$ is larger than the minimum significant difference for all pairwise comparisons with $i = 1$, but for none of the others. Consequently, the only confidence intervals that do not contain zero are those involving treatment 1. We conclude that based on the first day's data, treatment 1 is significantly better than each of the other 5 treatments and should be investigated further.

11.6 Sample Sizes

Given the number of treatments v and the block size k, how many blocks b are required to achieve confidence intervals of a specified length or a hypothesis test of specified power? Since for most purposes the balanced incomplete block design is the best incomplete block design when it is available, we start by calculating b and the treatment replication $r = bk/v$ for this design. Then if a balanced incomplete block design cannot be found with b and r close to the calculated values, a group divisible, cyclic, or other incomplete block design can be considered. Since balanced incomplete block designs are the most efficient, other incomplete block designs would generally require b and r to be a little larger.

Example 11.6.1 `Sample size to achieve confidence interval length`

Suppose Tukey's method for all pairwise comparisons will be used to analyze an experiment with $v = 5$ treatments and block size $k = 3$. It is thought unlikely that *msE* will be larger than 2.0 units^2. Suppose that the experimenters want the length of simultaneous 95% confidence intervals for pairwise comparisons to be at most 3.0 units (that is, a minimum significant difference of at most 1.5). A balanced incomplete block design will ensure that the interval lengths will all be the same.

Using the fact that $bk = vr$ for a block design, the error degrees of freedom (11.4.8) can be written as

$$df = bk - b - v + 1 = vr - vr/k - v + 1 = (v(k-1)r/k) - (v-1)\,, \qquad (11.6.18)$$

so, here, $df = (10r/3) - 4$. For a balanced incomplete block design, the minimum significant difference for a confidence interval for any pairwise treatment comparison, using Tukey's method with an overall 95% confidence level, is given in (11.4.13), p. 360 and, if we set $\lambda = r(k-1)/(v-1)$ from (11.3.1), p. 353, this becomes

$$msd = (q_{v,df,.05}/\sqrt{2})\sqrt{2\left[\frac{k(v-1)}{rv(k-1)}\right] msE} = q_{5,df,.05}\sqrt{\frac{12}{5r}}\,,$$

with $df = (10r/3) - 4$. For the *msd* to be at most 1.5 units, it is necessary that

$$12q^2/5r \geq 1.5^2; \quad \text{that is,} \quad r \geq 1.0667q^2_{5,df,.05}\,.$$

Trial and error shows that around 17–18 observations per treatment would be needed to satisfy the inequality; that is, 85–90 observations in total, which would require 28–30 blocks of size 3. A balanced incomplete block design exists with $v = 5, k = 3, b = 10, r = 6$ (all possible combinations of five treatments taken three at a time as blocks). Repeating this entire design three times would give a balanced incomplete block design with $r = 18$, which will give a minimum significant difference of about

$$q_{5,56,.05}\sqrt{12/(5 \times 18)} \approx 1.46 < 1.5\,.$$

□

Example 11.6.2 `Sample size to achieve specified power`

Suppose a test of the null hypothesis $H_0 : \{\tau_i \text{ all equal}\}$ is required to detect a difference in the treatment effects of $\Delta = 2$ units with probability 0.95, using significance level $\alpha = 0.05$ for a balanced incomplete block design with $v = 5$ treatments and block size $k = 3$.

The least squares estimator $\hat{\tau}_i - \hat{\tau}_p$ of a pairwise comparison contrast $\tau_i - \tau_p$ for a balanced incomplete block design has variance given by (11.4.12), p. 360, with $\Sigma c_i^2 = 2$. Also, from (11.3.1), p. 353, $\lambda = r(k-1)/(v-1)$ so

$$\text{Var}(\Sigma c_i \hat{\tau}_i) = 2\frac{k}{\lambda v}\sigma^2 = 2\left[\frac{k(v-1)}{rv(k-1)}\right]\sigma^2 . \tag{11.6.19}$$

The number r of observations needed per treatment is calculated via a formula similar to (6.6.48), p. 171, with $a = v$ and with $2\sigma^2/b$ replaced by the variance (11.6.19); that is,

$$r = \frac{2v\sigma^2\phi^2}{\Delta^2}\left[\frac{k(v-1)}{v(k-1)}\right] = \frac{2 \times 5 \times \sigma^2\phi^2}{2^2}\left[\frac{3 \times 4}{5 \times 2}\right] .$$

Suppose that σ^2 is believed to be at most 1.0 unit^2; then $r = 3\phi^2$. The power tables in Appendix A.7 can be used to find ϕ^2. The numerator degrees of freedom are $\nu_1 = v-1$ and the denominator degrees of freedom ν_2 are the error degrees of freedom (11.6.18). So for our example, $\nu_1 = 4$ and $\nu_2 = (10r/3)-4$. Trial and error shows that about $r = 9$ observations per treatment are needed to satisfy the equality, requiring about $b = 15$ $(= vr/k)$ blocks. A balanced incomplete block design exists with $v = 5$, $k = 3$, $b = 10$, $r = 6$ (all possible selections of three treatments taken as blocks). Repeating the entire design twice would give a balanced incomplete block design with $r = 12$, which would give more precision than required. Alternatively, one could use a computer program to seek a different type of incomplete block design with $r = 9$ and $b = 15$ (see Sects. 11.8.1 and 11.9.1, for example). □

To meet the requirements of each of Examples 11.6.1 and 11.6.2, the resulting designs needed to be large. However, in cases when σ^2 is expected to be small and the block size can be large, the required number of blocks may be smaller than a balanced incomplete design or group divisible design can accommodate. Software such as that illustrated in Sects. 11.8.1 and 11.9.1 can be used to find other incomplete block designs that satisfy the requirements of the experiment.

11.7 Factorial Experiments

11.7.1 Factorial Structure

Any incomplete block design can be used for a factorial experiment by taking the treatment labels to represent treatment combinations. The incomplete block designs that are the most suitable for factorial experiments allow the adjusted treatment sum of squares ssT_{adj} to be written as a sum of the adjusted sums of squares for main effects and interactions. Thus, for an experiment with two factors C and D, for example, we would like to have

$$ssT_{\text{adj}} = ssC_{\text{adj}} + ssD_{\text{adj}} + ssCD_{\text{adj}} .$$

Such block designs are said to have *factorial structure*.

One benefit of this property is that the computations for, and interpretation of, the analysis of variance are simplified. A design with factorial structure requires that main-effect and interaction contrast estimates be adjusted only for block effects. In designs without factorial structure, the contrast estimates have to be adjusted not only for blocks but also for contrasts in all the other main effects

Table 11.15 Design and data for the step experiment

Block	Treatment combination					
	11	12	13	21	22	23
1		75	87	84	93	99
2	93	84	96	90	108	
3	99	93	96		123	129
4	99	108	99	99		120
5	99		111	90	129	141
6	129	135		120	147	153

and interactions. Although computer software can handle this adjustment, uncorrelated estimates are much easier to interpret and are, therefore, preferred.

All balanced incomplete block designs have factorial structure, and the features are illustrated in the following example.

Example 11.7.1 `Step experiment`

An experiment was run by S. Guerlain, B. Busam, D. Huland, P. Taige, and M. Pavol in 1993 to investigate the effects on heart rate due to the use of a step machine. The experimenters were interested in checking the theoretical model that says that heart rate should be a function of body mass, step height, and step frequency. The experiment involved the two treatment factors "step height" (factor C) and "step frequency" (factor D). Levels of "step height" were 5.75 and 11.5 inches, coded 1 and 2. "Step frequency" had three equally spaced levels, 14, 21, and 28 steps per minute, coded 1, 2, 3. The response variable was pulse rate in beats per minute.

The experiment used $b = 6$ subjects as blocks, and each subject was measured under $k = 5$ of the $v = 6$ combinations of step height and step frequency. The design was a balanced incomplete block design with blocks corresponding to different combinations of subject, run timer, and pulse measurer. All pairs of treatment combinations appeared together in $\lambda = 4$ blocks. The data are shown in Table 11.15.

Writing the treatment combinations as two-digit codes, the block–treatment model (11.4.2), p. 356, becomes

$$Y_{hij} = \mu + \theta_h + \tau_{ij} + \epsilon_{hij} \,,$$

and a set of least squares solutions for the treatment parameters adjusted for subject are given by (11.4.7) and (11.4.9), p. 358 and 359, with two-digit codes; that is,

$$\hat{\tau}_{ij} = \frac{k}{\lambda v} Q_{ij} = \frac{k}{\lambda v}\left[T_{ij} - \frac{1}{k}\sum_{h=1}^{b} n_{hij} B_h\right] ,$$

where T_{ij} is the total of the $r = 5$ observations on step height i, step frequency j, B_h is the total of the $k = 5$ observations on the hth subject; and n_{hij} is 1 if treatment combination ij is observed for subject h and is zero otherwise. We obtain

$\hat{\tau}_{11}$	$\hat{\tau}_{12}$	$\hat{\tau}_{13}$	$\hat{\tau}_{21}$	$\hat{\tau}_{22}$	$\hat{\tau}_{23}$
−8.125	−7.625	−4.125	−11.375	12.375	18.875

Table 11.16 Analysis of variance for the step experiment

Source of variation	Degrees of freedom	Sum of squares	Mean square	Ratio	p-value
Subject (Block) (adj)	5	6685.05	1337.01	–	–
Subject (Block) (unadj)	5	7400.40	–	–	–
Height (C) (adj)	1	1264.05	1264.05	28.63	0.0001
Frequency (D) (adj)	2	1488.90	744.45	16.86	0.0001
Ht×Freq (CD) (adj)	2	990.90	495.45	11.22	0.0006
Error	19	838.95	44.16		
Total	29	11983.20			

For a balanced incomplete block design, the adjusted sums of squares for the main effects of C (step height) and D (step frequency) and their interaction can be obtained by hand by using the values of $\hat{\tau}_{ij}$ in place of $y_{ij.}$, and $k/(\lambda v)$ in place of r in the formulae (6.4.20), (6.4.22), and (6.4.25), p. 157–159, or, equivalently, in Rule 4, p. 209, which leads to

$$ssC_{\text{adj}} = \frac{\lambda v}{k}\left[\frac{1}{d}\sum_{i=1}^{d}\hat{\tau}_{i.}^2 - \frac{1}{cd}\hat{\tau}_{..}^2\right] \;=\; \left(\frac{\lambda v}{k}\right)\frac{1}{d}\sum_{i=1}^{d}\hat{\tau}_{i.}^2$$

$$[3pt] = \left(\frac{24}{5}\right)\frac{1}{3}(-19.875^2 + 19.875^2) \;=\; 1264.05.$$

The adjusted treatment sums of squares, ssT_{adj}, is calculated from (11.4.10), p. 359, and using (11.4.9) becomes

$$ssT_{\text{adj}} = \sum_{i=1}^{v}\left(\frac{k}{\lambda v}\right)Q_i^2 \;=\; \sum_{i=1}^{v}\left(\frac{\lambda v}{k}\right)\hat{\tau}_i^2 \;=\; \left(\frac{4\times 6}{5}\right)\times 779.9688 \;=\; 3743.85\,.$$

The adjusted sums of squares for the main effects of C and D and their interaction are shown in analysis of variance Table 11.16. It can be verified that

$$ssC_{\text{adj}} + ssD_{\text{adj}} + ssCD_{\text{adj}} = ssT_{\text{adj}}\,.$$

Using the p-values in the analysis of variance Table 11.16, the experimenters rejected the hypothesis of negligible interaction. A plot of the data (not shown) suggests that heart rate increases linearly as the step frequency is increased, but that the linear trend is not the same for the two step heights. The experimenters wanted to examine the average behavior of the two factors, so despite this interaction, they decided to examine the main effects. In Exercise 10, the reader is asked to examine the linear trends at each step height separately.

For simplicity of notation, we now drop the subscript "adj." However, all estimates and sums of squares are adjusted for block effects. The experimenters were interested in examining the linear and quadratic trend contrasts for step frequency, that is,

$$D_{\text{L}} = -\overline{\tau}_{.1} + \overline{\tau}_{.3} \;=\; -\frac{1}{2}(\tau_{11}+\tau_{21}) + \frac{1}{2}(\tau_{13}+\tau_{23})\,,$$

$$D_{\text{Q}} = -\overline{\tau}_{.1} + 2\overline{\tau}_{.2} - \overline{\tau}_{.3} \;=\; -\frac{1}{2}(\tau_{11}+\tau_{21}) + \frac{2}{2}(\tau_{12}+\tau_{22}) - \frac{1}{2}(\tau_{13}+\tau_{23})\,.$$

For the balanced incomplete block design, the least squares estimate for a contrast $\sum c_{ij}\tau_{ij}$ and its associated variance are given by (11.4.11) and (11.4.12), p. 359–360; that is,

$$\frac{k}{\lambda v}\sum_i\sum_j c_{ij}Q_{ij} \quad \text{and} \quad \sum_i\sum_j c_{ij}^2\left(\frac{k}{\lambda v}\right)\sigma^2\,,$$

respectively. Using these formulae, we find that the least squares estimates of the linear and quadratic trend contrasts for step frequency (adjusted for subjects) are

$$\hat{D}_{\mathrm{L}} = 17.125 \quad \text{and} \quad \hat{D}_{\mathrm{Q}} = -7.125\,.$$

The linear trend is positive, suggesting that the average pulse rate increases as the step frequency increases, and the quadratic trend is negative, suggesting that the increase in pulse rate is greater from 14 to 21 steps per minute than it is from 21 to 28 steps per minute. The null hypotheses $H_0^L : \{D_{\mathrm{L}} = 0\}$ and $H_0^Q : \{D_{\mathrm{Q}} = 0\}$ should be tested to check whether the perceived trends are significantly different from zero. The variances of the contrast estimators are

$$\left(\frac{4}{4}\right)\left(\frac{5}{24}\right)\sigma^2 \;=\; 0.2083\sigma^2 \quad \text{and} \quad \left(\frac{12}{4}\right)\left(\frac{5}{24}\right)\sigma^2 \;=\; 0.625\sigma^2\,,$$

respectively.

The contrast sum of squares for testing the null hypothesis $H_0^L : \{D_{\mathrm{L}} = 0\}$ is obtained from (11.4.16), p. 360, as

$$ss(D_{\mathrm{L}}) \;=\; \frac{(\hat{D}_{\mathrm{L}})^2}{\sum\sum c_{ij}^2\left(\frac{k}{\lambda v}\right)} \;=\; \frac{17.125^2}{0.2083} \;=\; 1407.675\,,$$

and the contrast sum of squares for testing $H_0^Q : \{D_{\mathrm{Q}} = 0\}$ is

$$ss(D_{\mathrm{Q}}) \;=\; \frac{(\hat{D}_{\mathrm{Q}})^2}{\sum\sum c_{ij}^2\left(\frac{k}{\lambda v}\right)} \;=\; \frac{(-7.125)^2}{0.625} \;=\; 81.225.$$

The linear and quadratic contrasts are orthogonal in a balanced incomplete block design even after adjusting for blocks, and we can now verify that indeed, $ssD = ss(D_{\mathrm{L}}) + ss(D_{\mathrm{Q}})$,

To test the null hypotheses H_0^L and H_0^Q against their respective alternative hypotheses that the null hypothesis is false, we compare each of $ss(D_{\mathrm{L}})/msE = 31.88$ and $ss(D_{\mathrm{Q}})/msE = 1.84$ with $2F_{2,19,.01} = 7.04$ for Scheffé's method and an overall level of $\alpha = 0.01$. We conclude that the quadratic trend is negligible, but there is a nonnegligible linear trend in the heart rate as the stepping frequency increases (averaged over step height). □

Table 11.17 SAS program for generation of an efficient incomplete block design

```
* Using proc optex to search for an efficient block design with v = 7,
b = 7, k = 3;
DATA CANDIDATE;
  DO TREATMNT = 1 to 7;
    OUTPUT;
  END;

PROC OPTEX DATA = CANDIDATE SEED = 72145;
  CLASS TREATMNT;
  MODEL TREATMNT;
  * For 7 blocks of size 3;
  BLOCKS STRUCTURE = (7)3;
  EXAMINE DESIGN;
```

11.8 Using SAS Software

11.8.1 Generation of Efficient Block Designs

`PROC OPTEX` within the SAS software allows one to search for efficient incomplete block designs. Although one cannot specify the type of design to be generated, the software will search for the design that gives the smallest confidence region for all contrasts using the Scheffé method of multiple comparisons. If a balanced incomplete block design exists, it will usually be found by `PROC OPTEX`.

The first set of lines in the SAS program in Table 11.17 specify that there are 7 treatments and these are stored in a dataset called `CANDIDATE`. Then, in the second set of lines, `PROC OPTEX` is run for a block structure "`(b)k`". In Table 11.17, there are $b = 7$ blocks of size $k = 3$. Since the program is not guaranteed to find the optimal design, it makes 10 independent searches (–this number can be changed by the user). The use of a particular "`SEED =`" always starts the search at the same point, so that the same set of designs is obtained. This should be removed when starting a new experiment so that random starts of the search are made. The designs found using the seed given in Table 11.17 are listed in the first part of the output in Fig. 11.5. Among the designs found, the one which gives the smallest confidence region for all contrasts is the one with the largest value under the heading "Treatment D-efficiency". This will often coincide with the design with the largest "Treatment A-efficiency" which has the shortest average length of confidence intervals for pairwise comparisons. To search specifically for the best design under A-efficiency, we insert the statement `GENERATE CRITERION = A;` after the `BLOCKS STRUCTURE` statement, and the designs will then be rank ordered by the A-efficiency.

The designs found in the 10 searches may or may not be exactly the same, but those listed in Fig. 11.5 are equally good as measured by their efficiencies. The command `EXAMINE DESIGN` prints out the best design (the one at the top of the list). It can be seen in Fig. 11.5 that, for this design, Block I (listed as the first three "points") consists of treatments 1, 2, 4, while Block II consists of 2, 3, 7, and so on. It can be checked that this is a balanced incomplete block design, although not the same one as that in Table 11.2, p. 351. To examine a design which is not listed as the best, say the third design in the list, replace `EXAMINE DESIGN;` by `EXAMINE NUMBER = 3 DESIGN;`.

The quoted value of "A-efficiency" is the ratio of the average variance of the pairwise comparisons in the design being examined relative to the average variance in a randomized block design with the same

Fig. 11.5 SAS output from `PROC OPTEX`

Results Viewer - SAS Output

The OPTEX Procedure

Design Number	Treatment D-Efficiency	Treatment A-Efficiency	Block Design D-Efficiency
1	77.7778	77.7778	100.0000
2	77.7778	77.7778	100.0000
3	77.7778	77.7778	100.0000
10	77.7778	77.7778	100.0000

Log determinant of the information matrix	1.6759E+01
Average variance of coefficients	6.1224E-02
D-Efficiency	77.7778
A-Efficiency	77.7778

Point Number	Block Number	TRT
1	1	1
2	1	2
3	1	4
4	2	2
5	2	3
6	2	7
7	3	6
18	6	3
19	7	6
20	7	7
21	7	4

value of r and multiplied by 100%. Since, in this example, the design found is a balanced incomplete block design, the average variance is given by (11.4.12), p. 360, and the ratio is

$$\frac{2\sigma^2/r}{2k\sigma^2/\lambda v}100\% = \frac{\lambda v}{rk}100\% = \frac{v(k-1)}{k(v-1)}100\% = 77.777\% \, .$$

The quoted "average D-efficiency" is related to the volume of the confidence region for all contrasts in the design being examined as compared with that for a randomized block design with the same value of r.

Table 11.18 SAS program for analysis of a balanced incomplete block design—detergent experiment

```
DATA ONE;
  INPUT BLOCK TRTMT Y;
  LINES;
   1 3 13
   : :  :
  12 1 19
;
PROC SGPLOT;
  SCATTER X = TRTMT  Y = Y / GROUP = BLOCK MARKERCHAR = BLOCK;

PROC GLM;
  CLASS BLOCK TRTMT;
  MODEL Y = BLOCK TRTMT;
  OUTPUT OUT = RESIDS PREDICTED = PREDY RESIDUALS = E;
  * contrast sums of squares for 8 orthogonal contrasts;
  CONTRAST 'I linear'           TRTMT -3 -1  1  3  0  0  0  0  0;
  CONTRAST 'I quadratic'        TRTMT  1 -1 -1  1  0  0  0  0  0;
  CONTRAST 'I cubic'            TRTMT -1  3 -3  1  0  0  0  0  0;
  CONTRAST 'II linear'          TRTMT  0  0  0  0 -3 -1  1  3  0;
  CONTRAST 'II quadratic'       TRTMT  0  0  0  0  1 -1 -1  1  0;
  CONTRAST 'II cubic'           TRTMT  0  0  0  0 -1  3 -3  1  0;
  CONTRAST 'I vs II'            TRTMT  1  1  1  1 -1 -1 -1 -1  0;
  CONTRAST 'others vs control' TRTMT  1  1  1  1  1  1  1  1 -8;
  * estimation of  treatment versus control contrasts via LSMEANS;
  LSMEANS TRTMT / PDIFF = CONTROL('9') CL ADJUST = DUNNETT;
  * estimation of  treatment versus control contrast via ESTIMATE;
  ESTIMATE 'Det 9-1' TRTMT -1  0  0  0  0  0  0  0  1;
```

For the requirements of Example 11.6.2, if `PROC OPTEX` is run with this same seed for $v = 5$ treatments and $b = 15$ blocks of size $k = 3$, the best design found has pairs of treatments appearing together in either $\lambda_1 = 4$ or $\lambda_2 = 5$ blocks. It consists of the 10-block balanced incomplete block design together with an additional 5 blocks comprising a cyclic incomplete block design. (A different seed, or no specified seed, may result in the additional 5 blocks being a non-cyclic incomplete block design but the best design listed will most likely still have $\lambda_2 = \lambda_1 + 1$).

11.8.2 Analysis of Variance, Contrasts, and Multiple Comparisons

In this section, sample programs are given to illustrate the analysis of incomplete block designs using the SAS software. The programs shown are for the detergent experiment of Sect. 11.4.4 and the plasma experiment of Sect. 11.5, but similar programs can be used to analyze the data collected in any incomplete block design.

Table 11.18 contains the first sample program. The data are entered into a data set called `ONE`, using the variables `BLOCK, TRTMT,` and `Y` for the block, treatment, and response value, respectively. `PROC SGPLOT` is used to plot the observations against treatments, analogous to Fig. 11.2, p. 361, and the legend identifies the block labels by color (the plot is not shown here). The block labels are printed on the plot by inclusion of the command `MARKERCHAR = BLOCK`. In the next section of Table 11.18, `PROC GLM` is used to fit the block–treatment model (11.4.2), generate the analysis of

Fig. 11.6 Partial output from PROC GLM for analysis of an incomplete block design—detergent experiment

Results Viewer - sashtml.htm

The GLM Procedure
Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	19	1499.564815	78.924464	95.77	<.0001
Error	16	13.185185	0.824074		
Corrected Total	35	1512.750000			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
BLOCK	11	412.750000	37.522727	45.53	<.0001
TRTMT	8	1086.814815	135.851852	164.85	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
BLOCK	11	10.064815	0.914983	1.11	0.4127
TRTMT	8	1086.814815	135.851852	164.85	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
I linear	1	286.0166667	286.0166667	347.08	<.0001
I quadratic	1	12.6759259	12.6759259	15.38	0.0012
I cubic	1	0.2240741	0.2240741	0.27	0.6092
II linear	1	61.3407407	61.3407407	74.44	<.0001
II quadratic	1	0.1481481	0.1481481	0.18	0.6772
II cubic	1	0.0296296	0.0296296	0.04	0.8520
I vs II	1	381.3379630	381.3379630	462.75	<.0001
others vs control	1	345.0416667	345.0416667	418.70	<.0001

variance table, and save the predicted values and residuals in the output data set RESIDS. Residuals can be standardized and plotted as in Chap. 6.

Output from PROC GLM is reproduced in Fig. 11.6. Since BLOCK has been entered before TRTMT in the model statement, the sum of squares for treatments adjusted for blocks is listed under Type I (or sequential) sums of squares as well as under the Type III sums of squares. The adjusted block sum of squares is listed under the Type III sums of squares. In order to use the sequential or Type I sums of squares to obtain the adjusted block sum of squares, one would need to rerun the program with TRTMT entered before BLOCK in the model statement. In Table 11.18, the sums of squares corresponding to $v - 1 = 8$ orthogonal treatment contrasts are requested via the CONTRAST statements, and it can be verified from Fig. 11.6 that the contrast sums of squares add to the treatment sum of squares.

Simultaneous confidence intervals for pairwise comparisons can be obtained via the ESTIMATE statements or via LSMEANS with options as discussed in Sect. 6.8.2, p. 180. Tukey, Scheffé and Dunnett methods can all be used for a balanced incomplete block design with code such as:

```
LSMEANS TRTMT / PDIFF=CONTROL('9') CL ADJUST=DUNNETT;
```

Here, the PDIFF=CONTROL('9') option for Dunnett's method specifies that level 9 is the control, as was the case in the detergent experiment. If the designation "('9')" had been omitted, then the

Fig. 11.7 Partial output from ESTIMATE and LSMEANS for an incomplete block design—detergent experiment with detergent 9 as the control treatment

Results Viewer - sashtml.htm

The GLM Procedure
Dependent Variable: Y

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Det 9-1	9.7777778	0.74120356	13.19	<.0001
Det 9-2	12.3333333	0.74120356	16.64	<.0001

Least Squares Means
Adjustment for Multiple Comparisons: Dunnett-Hsu

Least Squares Means for Effect TRTMT				
i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	9	-9.777778	-11.981887	-7.573669
2	9	-12.333333	-14.537442	-10.129224
3	9	-16.333333	-18.537442	-14.129224
4	9	-23.000000	-25.204109	-20.795891
5	9	-4.222222	-6.426331	-2.018113
6	9	-6.555556	-8.759664	-4.351447
7	9	-8.444444	-10.648553	-6.240336
8	9	-10.333333	-12.537442	-8.129224

lowest level would have been taken to be the control treatment by default. Partial output is shown in Fig. 11.7. The first section of the table shows the output from the ESTIMATE statements. The second section of the output gives simultaneous 95% confidence intervals for the treatment-versus-control comparisons using Dunnett's method. A word of warning is in order here. If the treatments had been labeled anything other than $1, 2, \ldots, 9$, at this point SAS software would have relabeled them. For example, if the control treatment had been labeled as 0 and the test treatments as $1, \ldots, 8$, SAS software would have relabeled the control as treatment 1 and the test treatments as $2, \ldots, 9$.

Figure 11.8 shows partial output from PROC GLM in the first part of Table 11.19 for day one data from the plasma experiment (Sect. 11.5), which was a nonstandard incomplete block design. The Type I and Type III sums of squares are shown, together with partial output from the LSMEANS statement

```
LSMEANS TRTMT / PDIFF = ALL CL ADJUST=SCHEFFE;
```

Output from the above LSMEANS statement, combined with standard error estimates generated by an ESTIMATE statement for each pairwise treatment contrast like the following one for $\tau_1 - \tau_2$, were used to compile Table 11.14 (p. 370):

```
ESTIMATE 'T1-T2' TRTMT 1 -1 0 0 0 0;
```

Fig. 11.8 Partial output from PROC GLM and LSMEANS in a SAS program for analyzing an incomplete block design—plasma experiment, day one

Results Viewer - SAS Output

The GLM Procedure
Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	0.00111406	0.00015915	954.90	0.0249
Error	1	0.00000017	0.00000017		
Corrected Total	8	0.00111422			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
BLOCK	2	0.00040289	0.00020144	1208.67	0.0203
TRTMT	5	0.00071117	0.00014223	853.40	0.0260

Source	DF	Type III SS	Mean Square	F Value	Pr > F
BLOCK	2	0.00012133	0.00006067	364.00	0.0370
TRTMT	5	0.00071117	0.00014223	853.40	0.0260

Least Squares Means
Adjustment for Multiple Comparisons: Scheffe

Least Squares Means for Effect TRTMT				
i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	0.025500	0.003602	0.047398
1	3	0.021833	0.003081	0.040585
1	4	0.023167	0.004415	0.041919
1	5	0.024333	0.008342	0.040325
1	6	0.022667	0.006675	0.038658
2	3	-0.003667	-0.028952	0.021618
2	4	-0.002333	-0.027618	0.022952

11.8.3 Plots

Table 11.19 contains a sample SAS program illustrating how to plot the data adjusted for blocks against the treatment labels, using the day one data of the plasma experiment, (first three blocks of Table 11.11, p. 368). The program, as written, must be run in three passes. In successive passes, information generated by earlier passes must be added as input in later parts of the program. First, the data are entered into a data set called ONE. Since the block effect estimates are needed to adjust the observations, the option SOLUTION is included in the MODEL statement of PROC GLM. This causes a (nonunique) solution to the normal equations for $\hat{\mu}$, $\hat{\tau}_i$, and $\hat{\theta}_h$ to be printed. The solutions will all be labeled "B" for "biased," meaning that the corresponding parameters are not individually estimable (see Fig. 10.8, p. 330, for example).

The least squares solutions $\hat{\theta}_h$ are then entered (as "BHAT") into the data set TWO by the user in the second run of the program. PROC MEANS is used to compute and print the average value $\hat{\bar{\theta}}_.$. Finally,

Table 11.19 SAS program to plot data adjusted for block effects—plasma experiment, day one only

```
* This program requires 3 runs, adding more information in each run;
DATA ONE;
  INPUT BLOCK TRTMT Y;
  LINES;
   1  4  0.459
   1  5  0.467
   :  :    :
   3  3  0.473
;
* Get block effect estimates;
PROC GLM;
  CLASS BLOCK TRTMT;
  MODEL Y = BLOCK TRTMT / SOLUTION;
  LSMEANS TRTMT / PDIFF = ALL CL ADJUST=SCHEFFE;
PROC SORT;  BY BLOCK;
* Add the following code for the second run;
* values BHAT are solutions for block parameters from first run;
DATA TWO;
  INPUT BLOCK BHAT;
  LINES;
        1     -.0126666667
        2     -.0053333333
        3     0.0000000000
;
PROC MEANS MEAN; * print average of BHAT values;
  VAR BHAT;
* Add the following code for the third run;
* The number -0.006 below is average BHAT calculated in second run;
DATA THREE;
  MERGE ONE TWO;
  BY BLOCK;
  Y_ADJ = Y - (BHAT - (-0.006));
PROC SGPLOT data = THREE;
  SCATTER X = TRTMT Y = Y_ADJ / GROUP = BLOCK;
```

in the third run of the program, the block-effect estimates and their average value are used to adjust the data values. The adjusted values are then plotted against treatment. The SAS plot is not shown here, but it is similar to the plot in Fig. 11.4 (p. 370).

11.9 Using R Software

11.9.1 Generating Efficient Incomplete Block Designs

The R function `bibd` from the package `ibd` can be used to construct balanced incomplete block designs when they exist. The first few lines of the program in Table 11.20 install and load the `ibd` package, then ask the program to search for a balanced incomplete block design with $v = 7$ treatments each appearing $r = 3$ times, $b = 7$ blocks of size $k = 3$, and pairs of treatments appearing together in $\lambda = 1$ block. In general, the `bibd` function checks that the necessary conditions for an existence of a

Table 11.20 R code and output for bibd with $v = 7$ treatments, $b = 7$ blocks of size $k = 3$

```
> # install.packages("ibd")
> library(ibd)
> bibd(v = 7, r=3, b = 7, k = 3, lambda = 1)

$design
     [,1] [,2] [,3]
[1,]    1    3    6
[2,]    1    4    5
[3,]    5    6    7
[4,]    1    2    7
[5,]    3    4    7
[6,]    2    4    6
[7,]    2    3    5

$Aeff
[1] 0.9999999

$Deff
[1] 0.9999999
```

bibd are satisfied. If so, it will either provide the design or respond with “design not found”. For the design size in Table 11.20, the design does exist and is shown under the heading `$design`. The seven blocks of the design are given by the seven rows of three treatment labels. The quoted `$Aeff` is the ratio of the average variance of the pairwise comparisons in this incomplete block design relative to the average variance (11.4.12) in a balanced incomplete block design with the same values of v and k. The quoted `$Deff` is related to the volume of the confidence region for all contrasts in the design being examined relative to a balanced incomplete block design with the same values of v and k (which may not actually exist in practice). Since the design found here is, itself, a balanced incomplete block design, A-eff and D-eff are given as 1 (approximately).

If no balanced incomplete block design exists, as for the requirements of Example 11.6.2, with $v = 5$ treatments and $b = 15$ blocks of size $k = 3$, then function `ibd`, whose inputs are v, b, and k, can be used to construct an incomplete block design. For this example, the best design found is shown in Table 11.21 and has “A.Efficiency = 0.9975” and “D.Efficiency= 0.9987”, where these efficiencies are calculated the same way as `$Aeff` and `$Deff` in the `bibd` function. It consists of a balanced incomplete block design with 10 blocks, together with an additional 5 blocks comprising an incomplete block design with $\lambda_2 = \lambda_1 + 1$. The entire 15-block design has pairs of treatments appearing together in either $\lambda_1 = 4$ or $\lambda_2 = 5$ blocks. For a binary design, the values of λ_i are listed as the off-diagonal elements in the array called `$conc.mat` so, for example, the first row of this array shows that treatment 1 appears in 4 blocks with treatments 2 and 3, and in 5 blocks with treatments 4 and 5. The diagonal elements of the array are the values of r.

The `ibd` function prints out the design with the highest D.Efficiency that it finds in 5 independent searches. If the design found has low efficiency, `ibd` may be able to obtain an improved design if the option `ntrial=n` is added into the `ibd` statement, where `n` is some integer larger than the default value of 5; for example,

$$\texttt{ibd(v = 5, b = 15, k = 3, ntrial = 20)}.$$

Table 11.21 R code and output for an incomplete block design with $v = 5$ treatments, $b = 15$ blocks of size $k = 3$

```
> library(ibd)
> ibd(v=5, b=15, k=3)

$design
      [,1] [,2] [,3]
 [1,]    1    2    3
 [2,]    1    4    5
 [3,]    1    4    5
 [4,]    3    4    5
 [5,]    1    3    5
 [6,]    1    3    4
 [7,]    2    3    4
 [8,]    1    3    4
 [9,]    2    4    5
[10,]    2    3    4
[11,]    1    2    5
[12,]    1    2    5
[13,]    1    2    4
[14,]    2    3    5
[15,]    2    3    5

$conc.mat
     [,1] [,2] [,3] [,4] [,5]
[1,]    9    4    4    5    5
[2,]    4    9    5    4    5
[3,]    4    5    9    5    4
[4,]    5    4    5    9    4
[5,]    5    5    4    4    9

$A.Efficiency
[1] 0.9975309

$D.Efficiency
[1] 0.9987647
```

11.9.2 Analysis of Variance, Contrasts, and Multiple Comparisons

In this section, sample programs are given to illustrate the analysis of incomplete block designs using the R software. The programs shown are for the detergent experiment of Sect. 11.4.4 and the plasma experiment of Sect. 11.5, but similar programs can be used to analyze the data collected in any incomplete block design.

Table 11.22, p. 385, contains the first sample program. The data are read into the data set `detrgnt.data`, using the variables `Block`, `Trtmt`, and `y` for the block, treatment, and response value, respectively. Corresponding factor variables `fBlock` ad `fTrtmt` are added to the data set. In the second block of code, the `plot` function is used to plot the observations against treatments, analogous to Fig. 11.2, p. 361, but using block labels as the plotting legend (the plot is not shown here).

In the third block of code, the linear models function `lm` is used to fit the block–treatment model (11.4.2), then corresponding `anova` and `drop1` functions generate the Type I (or sequential)

Table 11.22 R program for analysis of a balanced incomplete block design—detergent experiment

```
detrgnt.data = read.table("data/detergent.txt", header=T)
head(detrgnt.data)
detrgnt.data = within(detrgnt.data,
  {fBlock = factor(Block); fTrtmt = factor(Trtmt) })

# Plot y vs Trtmt using block level as plotting symbol.
plot(y ~ Trtmt, data=detrgnt.data, xaxt="n", type="n") # Suppress x-axis, pts
  axis(1, at=seq(1,9,1)) # x-axis labels 1:9
  text(y ~ Trtmt, Block, cex=0.75, data=detrgnt.data) # Plot y*Trtmt=Block
  mtext("Block=1,...,12", side=3, adj=1, line=1) # Margin text, TopRt, line 1

# Analysis of variance
model1 = lm(y ~ fBlock + fTrtmt, data=detrgnt.data)
anova(model1)
drop1(model1, ~., test="F")

# Contrast estimates
library(lsmeans)
lsmTrtmt = lsmeans(model1, ~ fTrtmt)
cntrsts = summary(contrast(lsmTrtmt,
                            list(I.linear=c(-3,-1, 1, 3, 0, 0, 0, 0, 0),
                                   I.quad=c( 1,-1,-1, 1, 0, 0, 0, 0, 0),
                                  I.cubic=c(-1, 3,-3, 1, 0, 0, 0, 0, 0),
                                II.linear=c( 0, 0, 0, 0,-3,-1, 1, 3, 0),
                                  II.quad=c( 0, 0, 0, 0, 1,-1,-1, 1, 0),
                                 II.cubic=c( 0, 0, 0, 0,-1, 3,-3, 1, 0),
                                  I.vs.II=c( 1, 1, 1, 1,-1,-1,-1,-1, 0),
                              Trt.vs.Ctrl=c( 1, 1, 1, 1, 1, 1, 1, 1,-8))),
                    infer=c(T,T))
# Compute and include contrast sums of squares: ss=t^2*mse
mse = anova(model1)[3,3]; mse
cntrsts = cbind(cntrsts, ss=cntrsts[,"t.ratio"]^2*mse)
options(width=75, digits=3, scipen=2)
cntrsts
options(ooptions)

# Dunnett's method
summary(contrast(lsmTrtmt, method="trt.vs.ctrl", adjust="mvt", ref=9),
        infer=c(T,T))
```

and Type III analysis of variance tables, respectively. Predicted values and residuals are available, and the residuals can be standardized and plotted as in Sect. 6.2.3.

The analysis of variance output from `anova` and `drop1` is reproduced in Table 11.23, p. 386. Since `fBlock` has been entered before `fTrtmt` in the model statement, the sum of squares for treatments adjusted for blocks are obtained as both Type I and Type III sums of squares. The adjusted block sum of squares is listed under the Type III sums of squares whereas, to use the sequential or Type I sums of squares, one would need to rerun the program with `fTrtmt` entered before `fBlock` in the model statement. In the center of Table 11.22, the `summary(contrast(lsmTrtmt`... statement is used to list eight orthogonal treatment contrasts, producing corresponding least squares estimates, standard errors, two-sided t-tests, and individual 95% confidence intervals, as shown in the bottom part

Table 11.23 Selected ANOVA and contrast output for analysis of an incomplete block design—detergent experiment

```
> anova(model1)

Analysis of Variance Table
Response: y
          Df  Sum Sq Mean Sq F value   Pr(>F)
fBlock    11     413    37.5    45.5 6.0e-10
fTrtmt     8    1087   135.9   164.8 6.8e-14
Residuals 16      13     0.8

> drop1(model1, ~., test="F")

Single term deletions
Model:
y ~ fBlock + fTrtmt
       Df Sum of Sq  RSS   AIC F value  Pr(>F)
<none>                13   3.8
fBlock 11        10   23   2.3    1.11    0.41
fTrtmt  8      1087 1100 147.1  164.85 6.8e-14

> cntrsts

     contrast estimate   SE df lower.CL upper.CL t.ratio  p.value       ss
1    I.linear  -43.667 2.34 16   -48.64   -38.70 -18.630 2.85e-12 286.0167
2      I.quad   -4.111 1.05 16    -6.33    -1.89  -3.922 1.22e-03  12.6759
3     I.cubic   -1.222 2.34 16    -6.19     3.75  -0.521 6.09e-01   0.2241
4   II.linear  -20.222 2.34 16   -25.19   -15.25  -8.628 2.05e-07  61.3407
5     II.quad    0.444 1.05 16    -1.78     2.67   0.424 6.77e-01   0.1481
6    II.cubic   -0.444 2.34 16    -5.41     4.52  -0.190 8.52e-01   0.0296
7     I.vs.II  -31.889 1.48 16   -35.03   -28.75 -21.512 3.10e-13 381.3380
8 Trt.vs.Ctrl  -91.000 4.45 16  -100.43   -81.57 -20.462 6.73e-13 345.0417
```

of Table 11.23. The sum of squares for each contrast is subsequently computed from its corresponding t-ratio as `ss =(t-ratio)`$^2 \times$ *mse*, which follows since `(t-ratio)`$^2 =$ `ss`/*mse* from (4.3.16) p. 78, where *mse* is the mean squared error from the `Residuals` line in the analysis of variance table.

Simultaneous confidence intervals for pairwise comparisons can be obtained via the `lsmeans` function as discussed in Sect. 6.9.2, p. 187. For example, Dunnett's method is invoked by the following statements in the bottom of Table 11.22, p. 385.

```
lsmTrtmt = lsmeans(model1, ~ fTrtmt)
summary(contrast(lsmTrtmt, method="trt.vs.ctrl", adjust="mvt", ref=9),
   infer=c(T,T))
```

The `ref=9` option specifies that the 9th level (coincidentally also labeled `"9"`) is the control, as was the case in the detergent experiment. If this designation had been omitted, then the lowest level would have been taken to be the control treatment by default. Table 11.24, p. 387, contains the resulting output, including the estimate of each treatment-versus-control comparison $\tau_i - \tau_9$, as well as the corresponding standard error, simultaneous 95% confidence limits by Dunnett's method, test statistic, and p value for simultaneous tests for whether or not each of the treatment-versus-control comparisons $\tau_i - \tau_9$ is zero, or equivalently whether $\tau_i = \tau_9$, using Dunnett's method.

Table 11.24 Dunnett's method output for an incomplete block design—detergent experiment with detergent 9 as the control treatment

```
> # Dunnett's method
> summary(contrast(lsmT, method="trt.vs.ctrl", adjust="mvt", ref=9),
+         infer=c(T,T))

 contrast estimate     SE df lower.CL upper.CL t.ratio p.value
 1 - 9     -9.7778 0.7412 16 -11.9824  -7.5732 -13.192  <.0001
 2 - 9    -12.3333 0.7412 16 -14.5379 -10.1287 -16.640  <.0001
 3 - 9    -16.3333 0.7412 16 -18.5379 -14.1287 -22.036  <.0001
 4 - 9    -23.0000 0.7412 16 -25.2046 -20.7954 -31.031  <.0001
 5 - 9     -4.2222 0.7412 16  -6.4268  -2.0176  -5.696  0.0002
 6 - 9     -6.5556 0.7412 16  -8.7601  -4.3510  -8.844  <.0001
 7 - 9     -8.4444 0.7412 16 -10.6490  -6.2399 -11.393  <.0001
 8 - 9    -10.3333 0.7412 16 -12.5379  -8.1287 -13.941  <.0001

Results are averaged over the levels of: fB
Confidence level used: 0.95
Confidence-level adjustment: mvt method for 8 estimates
P value adjustment: mvt method for 8 tests
```

Table 11.25, p. 388, shows an R program and partial output for analysis of the day one data from the plasma experiment (Table 11.11, p. 368), which used a nonstandard incomplete block design. After reading the entire set of data, `data/plasma.txt`, and defining the block and treatment factors, the `subset` function is used to create a data set `day1.data` containing only day one data. The Type I and type 3 sums of squares generated by `anova` and `drop1`, respectively, are shown at the top of the table. The bottom of the table contains partial output from the `lsmeans` and `summary(contrast` statements, comparing treatment effects pairwise via Scheffé's method. This is one way to obtain information analogous to that in Table 11.14 (p. 370).

11.9.3 Plots

Table 11.26, p. 389, contains a sample R program illustrating how to plot the data adjusted for blocks against the treatment labels, using the day one data of the plasma experiment, (first three rows of Table 11.11, p. 368). The `subset` function is used to create a data set `day1.data` containing only day one data. Having fit the linear model by the `lm` function and saved the results as `model2`, the least squares estimates of the model parameters are available as `model2$coefficients`. If one were to display these, one would see that the second and third coefficient estimates are the block effect estimates $\hat{\theta}_2 \approx 0.0073333$ and $\hat{\theta}_3 \approx 0.0126667$, but there is no estimate listed for the first block effect, so $\hat{\theta}_1 = 0$. These three block effect estimates are saved in the column `thetahat`, and the mean estimate $\hat{\bar{\theta}}_.$ is saved as `avgest`. These values are used to compute adjusted observations, `yadj` $= y - (\hat{\theta}_h - \hat{\bar{\theta}}_.)$, which are then plotted against treatment. The R plot is not shown here, but it is similar to the plot in Fig. 11.4 (p. 370).

Table 11.25 R program and partial output from anova, drop1, and lsmeans, analyzing an incomplete block design—plasma experiment

```
> plasma.data = read.table("data/plasma.txt", header=T)
> plasma.data = within(plasma.data,
+   {fBlock = factor(Block); fTrtmt = factor(Trtmt) })
> head(plasma.data, 3)
> # Analysis of day 1 data
> day1.data = subset(plasma.data, Day==1) # Keep data with "Day==1"
> model2 = lm(y ~ fBlock + fTrtmt, data=day1.data)
> anova(model2)

Analysis of Variance Table
Response: y
          Df   Sum Sq   Mean Sq F value Pr(>F)
fBlock     2 0.000403 0.0002014    1209  0.020
fTrtmt     5 0.000711 0.0001422     853  0.026
Residuals  1 0.000000 0.0000002

> drop1(model2, ~., test="F")

Single term deletions
Model:
y ~ fBlock + fTrtmt
       Df Sum of Sq      RSS    AIC F value Pr(>F)
<none>              0.000000 -144.2
fBlock  2  0.000121 0.000121  -88.9     364  0.037
fTrtmt  5  0.000711 0.000711  -79.0     853  0.026

> # Scheffe's method
> library(lsmeans)
> lsmTrtmt = lsmeans(model2, ~ fTrtmt)
> summary(contrast(lsmTrtmt, method="pairwise", adjust="Scheffe"),
+         infer=c(T,T))

 contrast    estimate         SE df   lower.CL upper.CL t.ratio p.value
 1 - 2     0.02550000 0.00064550  1  0.0036024 0.047398  39.504  0.0429
 1 - 3     0.02183333 0.00055277  1  0.0030814 0.040585  39.498  0.0430
 1 - 4     0.02316667 0.00055277  1  0.0044147 0.041919  41.910  0.0405
   :         :          :          :   :         :          :       :
 5 - 6    -0.00166667 0.00047140  1 -0.0176584 0.014325  -3.536  0.4451

Results are averaged over the levels of: fBlock
Confidence level used: 0.95
Confidence-level adjustment: scheffe method for a family of 6 estimates
P value adjustment: scheffe method for a family of 6 tests
```

Table 11.26 R program to plot data adjusted for block effects—plasma experiment, day one only

```
# Create data set of day 1 data
plasma.data = read.table("data/plasma.txt", header=T)
plasma.data = within(plasma.data,
  {fBlock = factor(Block); fTrtmt = factor(Trtmt) })
day1.data = subset(plasma.data, Day==1) # Keep data with "Day==1"

# Fit model to day 1 data
model2 = lm(y ~ fBlock + fTrtmt, data=day1.data)

# Create columns of block effect estimates, and compute the avg estimate
model2$coefficients # Display all model coefficient estimates
thetahat = c(0, model2$coefficients[2:3]) # Create col of block effect ests
avgest = mean(thetahat) # Compute average block effect estimate

# Compute adjusted y-values: yadj = y - (thetahat.h - thetahat.bar)
day1.data$yadj = day1.data$y - (thetahat[day1.data$Block] - avgest)

# Plotting day 1 data adjusted for block effects
plot(yadj ~ Trtmt, data=day1.data, xlab="Treatment", ylab="y Adjusted")
```

Table 11.27 Three incomplete block designs

Design I		Design II		Design III	
Block	Treatments	Block	Treatments	Block	Treatments
1	1 2	1	1 2 3	1	1 2 6
2	1 3	2	4 5 6	2	3 4 5
3	1 4	3	7 8 9	3	2 6 8
4	2 3	4	1 4 7	4	4 5 7
5	2 4	5	2 5 8	5	1 6 8
6	3 4	6	3 6 9	6	3 5 7
		7	1 5 9	7	1 2 8
		8	2 6 7	8	3 4 7
		9	3 4 8		

Exercises

1. **Connectedness and estimability**
   (a) For each of the three block designs in Table 11.27, draw the connectivity graph for the design, and determine whether the design is connected.
   (b) If the design is connected, determine whether or not it is a balanced incomplete block design.
   (c) For designs II and III, determine graphically whether or not $\tau_1 - \tau_5$ and $\tau_1 - \tau_6$ are estimable.
   (d) For design III, use expected values to show that $\tau_1 - \tau_8$ is estimable.

2. **Connectedness**

   (a) Determine whether or not the cyclic design with initial block $(1, 3, 5)$ is a connected design if $v = 8$ or $v = 9$.
   (b) Determine whether or not the cyclic design with initial block $(1, 4, 7)$ is a connected design if $v = 8$ or $v = 9$.

3. **Randomization**

   Conduct a block design randomization for design II in Table 11.27.

4. **Cyclic designs**

   Determine whether or not the cyclic design obtained from each initial block below is a balanced incomplete block design or a group divisible design or neither.

   (a) Initial block: $1, 3, 4$; $v = 7$.
   (b) Initial block: $1, 2, 4, 8$; $v = 8$.
   (c) Initial block: $1, 2, 4$; $v = 5$.

5. **Balanced incomplete block design**

   In the following questions, consider an experiment to compare $v = 7$ treatments in blocks of size $k = 5$.

   (a) Show that, for this experiment, a necessary condition for a balanced incomplete block design to exist is that r is a multiple of 5 and b is a multiple of 7.
   (b) Show that r must be at least 15.
   (c) Taking all possible combinations of five treatments from seven gives a balanced incomplete block design with $r = 15$. Calculate the number of blocks that must be in this design.

6. **Sample sizes**

   Consider an experiment to compare 7 treatments in blocks of size 5, with an anticipated error variance of at most 30 squared units.

   (a) Assuming that a balanced incomplete block design will be used, how many observations would be needed for the minimum significant difference to be about 50 units for a pairwise comparison using Tukey's method and a 95% simultaneous confidence level?
   (b) Repeat part (a) for a minimum significant difference of 25 units.
   (c) Repeat part (a) using Dunnett's method for treatment versus control comparisons.

7. **Least squares estimator, detergent experiment, continued**

   Consider the balanced incomplete block design in Table 11.8, p. 361, used for the detergent experiment in Sect. 11.4.4.

   (a) Show that the least squares estimator $\hat{\tau}_4 - \hat{\tau}_6$ is an unbiased estimator of $\tau_4 - \tau_6$ under the block–treatment model (11.4.2), p. 356 (that is, show that $E[\hat{\tau}_4 - \hat{\tau}_6] = \tau_4 - \tau_6$).
   (b) Calculate a confidence interval for $\tau_4 - \tau_6$ as part of a 95% set of simultaneous confidence intervals for pairwise comparisons, using Tukey's method.

Table 11.28 Percentage rust observed for the rust experiment

Temperature (°F)	Block									
	I	II	III	IV	V	VI	VII	VIII	IX	X
50	12	19				20	10	21	19	
55	18		33		19		18		18	24
60	24	36		35	39	22				28
65			39	45		43	34	42		31
70		45	52	55	48			50	43	

8. **Rust experiment**

The rust experiment investigated the effect of temperature on the percentage of surface area of a metal sheet exhibiting rust after a given length of time exposed to certain weathering conditions. Five temperatures were examined in the experiment, but only three could be examined at any one time under identical experimental conditions. A balanced incomplete block design was used, formed from two cyclic designs with initial blocks (1, 2, 3) and (1, 2, 4). The data and design are shown in Table 11.28.

(a) Write down a model for this experiment and test the hypothesis of no difference in the effects of the temperature on the percentage of rust, against the alternative hypothesis that at least two temperatures differ. Use a significance level of 0.01.
(b) What does "a significance level of 0.01" in part (a) mean?
(c) Was blocking worthwhile in this experiment?
(d) Give a formula for a 95% set of simultaneous confidence intervals for pairwise comparisons among the temperatures in this experiment. Calculate, by hand, the interval comparing temperatures 5 and 4 (that is, $\tau_5 - \tau_4$) for illustration. Compare your answer with that obtained from your computer output.
(e) Test the hypothesis that there is no linear trend in the percentage of rust as the temperature increases.

9. **Balanced incomplete block design**

An experiment is to be run to compare the effects of four different formulations of a drug to relieve an allergy. In a pilot experiment, four subjects are to be used, and each is to be given a sequence of three of the four drugs. The measurements are the number of minutes that the subject appears to be free of allergy symptoms. Suppose the design shown in Table 11.29 is selected for the experiment.

(a) Check that this design is a balanced incomplete block design. (Show what you are checking.)
(b) Show a randomization of this design, explaining the steps of your randomization.
(c) The experiment was run as described, and the block–treatment model (11.4.2), p. 356, was used to analyze it. Some information for the analysis is shown in Table 11.29. Using this information and (11.4.7), p. 358, show that Q_1, Q_2, Q_3, Q_4 are −79.333, 81.667, −158.667, 156.333, respectively.
(d) Using the information in part (c), give a confidence interval for $\tau_3 - \tau_2$ assuming that it is part of a set of 95% Tukey confidence intervals.
(e) Test the hypothesis that there is no difference between the effects of the drugs.

Table 11.29 Balanced incomplete block design of Exercise 9 and partial information for its analysis

Block	Levels of treatment factor			Block totals	Treatment totals
I	1	2	3	$B_1 = 417$	$T_1 = 385$
II	1	2	4	$B_2 = 507$	$T_2 = 582$
III	1	3	4	$B_3 = 469$	$T_3 = 329$
IV	2	3	4	$B_4 = 577$	$T_4 = 674$
				$\overline{y}_{..} = 164.1667$	$msE = 3.683$

(f) The typical model and analysis for a balanced incomplete block design is that of parts (c)–(e). Do you think this is a reasonable model and analysis for the experiment described? Why or why not? (Hint: think about the terms in, and assumptions on, the model.)

10. **Step experiment, continued**

The step experiment was described in Example 11.7.1 and the data are shown in Table 11.15, p. 373.

(a) Prepare a plot of the treatment averages and examine the linear trends in the heart rate due to step frequency at each level of step height.
(b) Fit a block–treatment model to the data with $v = 6$ treatments representing the six treatment combinations.
(c) Estimate the linear trends in the heart rate due to step frequency at each level of step height separately, and calculate confidence intervals for these.
(d) Write down a contrast that compares the linear trends in part (c) and test the hypothesis that the linear trends are the same against the alternative hypothesis that they are different.

11. **Beef experiment**

Cochran and Cox (1957) describe an experiment that was run to compare the effects of cold storage on the tenderness of beef roasts. Six periods of storage (0, 1, 2, 4, 9, and 18 days) were tested and coded 1–6. It was believed that roasts from similar positions on the two sides of the animal would be similar, and therefore the experiment was run in $b = 15$ blocks of size $k = 2$. The response y_{hi} from treatment i in block h is the tenderness score. The maximum score is 40, indicating very tender beef. The design and responses are shown in Table 11.30.

(a) What is the value of λ for this balanced incomplete block design?
(b) What benefit do you think the experimenters expected to gain by using a block design instead of a completely randomized design?
(c) Calculate the least squares estimate of $\tau_6 - \tau_1$ and its corresponding variance.
(d) Calculate a confidence interval for $\tau_6 - \tau_1$ as though it were part of a set of 95% confidence intervals using Tukey's method of multiple comparisons.
(e) Calculate a confidence interval for the difference of averages contrast

$$\frac{1}{3}(\tau_4 + \tau_5 + \tau_6) - \frac{1}{3}(\tau_1 + \tau_2 + \tau_3),$$

Table 11.30 Design and data for the beef experiment

Block	Treatment					
	1	2	3	4	5	6
I	7	17				
II			26	25		
III					33	29
IV	17		27			
V		23			27	
VI				29		30
VII	10			25		
VIII		26				37
IX			24		26	
X	25				40	
XI		25		34		
XII			34			32
XIII	11					27
XIV		24	21			
XV				26	32	

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

assuming that you want the overall level of this interval together with the intervals in part (d) to be at least 94%. What does your interval tell you about storage time and tenderness of beef?

12. **Balanced incomplete block design**

(a) Explain under what circumstances you would choose to use a block design rather than a completely randomized design.
(b) For a balanced incomplete block design, why is it incorrect to estimate the difference in the effects of treatments i and p as $\overline{y}_{.i} - \overline{y}_{.p}$? What is the formula for the correct least squares estimate and its corresponding variance if the design of Table 11.2, p. 351, is used?
(c) Suppose that the design of Table 11.2 is used for an experiment with $v = 7$ treatments and $b = 3$ blocks of size $k = 3$, and suppose that

$$\hat{\tau}_1 = 31.6,\ \hat{\tau}_2 = 22.4,\ \hat{\tau}_3 = 17.9,\ \hat{\tau}_4 = 21.5,\ \hat{\tau}_5 = 30.2,\ \hat{\tau}_6 = 18.6,\ \hat{\tau}_7 = 25.5\,.$$

Calculate a set of 99% confidence intervals for comparing the effect of treatment 1 with the effects of the other treatments. State your conclusions.

13. **Lithium bioavailability experiment**

An experiment described by W. J. Westlake (*Biometrics*, 1974) used a balanced incomplete block design to compare the amount of four formulations of lithium carbonate that were present in the blood serum of subjects several hours after administration. The four levels of the factor "formulation" consisted of 300 mg capsule (level 1), 250 mg capsule with different inert ingredients (level 2), 450mg delayed release capsule (level 3), 300 mg in solution (level 4). Level 4 is the standard (control) treatment. Table 11.31 shows the amounts of lithium carbonate in the blood serum after 3 hours, measured in mEq (milliequivalent which is a measure of the quantity of ions

Table 11.31 Data for the lithum bioavailability experiment, showing treatment factor level, and response in parenthesis

Subject					
1	2	3	4	5	6
1 (.467)	4 (.450)	3 (.200)	2 (.520)	4 (.600)	4 (.467)
2 (.440)	3 (.156)	1 (.533)	3 (.222)	1 (.467)	2 (.520)
Subject					
7	8	9	10	11	12
2 (.600)	2 (.640)	3 (.156)	1 (.533)	1 (.433)	3 (.156)
1 (.467)	4 (.400)	4 (.467)	4 (.433)	3 (.267)	2 (.520)

Source Data adapted from Westlake, Biometric, 1974, International Biometric Society

in an electrolyte fluid). There were 12 subjects, each of whom was given two formulations (one week apart).

(a) Verify that the design in Table 11.31 is a balanced incomplete block design.
(b) Plot the observations on the treatments adjusted for the block (subject) effects), and explain what the plot shows.
(c) Test, at level .01, the hypothesis that the four formulations have the same effect on the amount of lithium carbonate in the blood serum after 3 hours.
(d) Calculate a set of 99% simultaneous confidence intervals for the pairwise comparisons of the formulations and state your conclusion.
(e) Calculate a set of 99% simultaneous confidence intervals for the comparisons of the three test formulations with the control (level 4). State your conclusion.

14. **Plasma experiment, continued**

In the plasma experiment of Sect. 11.5, the experimenters used the cyclic design for six treatments generated by 1, 4, 5.

(a) Show that the cyclic design is also a group divisible design by determining the groups and the values of λ_1 and λ_2.
(b) The design for day one of the experiment consisted of the following three blocks of size three: $(1, 4, 5)$; $(2, 5, 6)$; $(3, 6, 1)$. Show that this design is connected by drawing the connectivity graph of the design.

15. **Group divisible design**

(a) Show that the design in Table 11.32 with $v = 12$ treatments and $b = 9$ blocks of size $k = 4$ is a group divisible design with $g = 4$ groups of size $\ell = 3$. Find the groups and calculate λ_1 and λ_2.
(b) Using the values of λ_1 and λ_2 in part (a), show that the requirements in Sect. 11.3.2, p. 354, for a group divisible design are met.
(c) Fit model (11.4.2), p. 356, to these data and check the assumptions on the error variables by plotting the residuals.
(d) Treatment 1 is the control treatment. Using the formulae in Sect. 11.4.5, p. 365, calculate the two possible variances of $\hat{\tau}_1 - \hat{\tau}_p$, $p = 2, \ldots, 12$.

Table 11.32 Data for the group divisible design

Block	Treatments (Response)			
1	2 (4.5)	4 (8.7)	10 (19.1)	11 (21.2)
2	5 (11.8)	7 (13.6)	9 (18.2)	11 (22.2)
3	1 (4.8)	8 (17.3)	11 (24.0)	12 (24.8)
4	4 (10.8)	6 (14.5)	7 (16.6)	8 (18.1)
5	2 (7.9)	3 (9.7)	8 (19.1)	9 (20.6)
6	2 (8.7)	5 (14.4)	6 (16.2)	12 (27.7)
7	1 (7.5)	6 (16.9)	9 (20.7)	10 (23.5)
8	3 (11.3)	7 (19.8)	10 (25.5)	12 (29.0)
9	1 (9.1)	3 (9.8)	4 (12.9)	5 (16.9)

(e) Calculate the two possible lengths of 99% confidence intervals for treatment versus control contrasts using Scheffé's method, and verify your calculations using a computer program.

(f) What can you conclude about the control treatment in this experiment?

(g) Using (4.2.4), p. 73, find the linear trend coefficients for 12 equally spaced treatment levels. If the treatment code corresponds to milliliters of added sugar to a product, and the response is a taste score, test the hypothesis that the linear trend in score as the amount of sugar increases is negligible. Use a significance level of $\alpha = 0.01$ and a two-sided alternative hyothesis.

16. **Fabric stain experiment**

The experiment was run by M. Nelson, T. Blake, D. Sullivan, A. Kemme and J. Jia in 2008 to examine how best to remove a stain caused by black waterproof drawing ink. There were three factors under consideration as follows. Factor A was type of cloth (1 = cotton, 2 = polyester), Factor B was type of pre-treatment (1 = gel stain remover, 2 = antibacterial dish soap, 3 = no pre-treatment), and Factor C was time between staining and washing (0, 1, 2 days, coded as levels 1, 2, 3). The experiment was run in several blocks; in some blocks the washing machine was set to the hot wash setting, and the others to the cold wash setting.
Thirty six pieces of fabric of each type were prepared of approximately the same size. The amount of ink applied to each piece of cloth was held constant, as was the amount of detergent used in each wash. The stained fabric samples were evaluated by a single experimenter using a "Gray Scale" (cf. Chapter 10, Exercise 14) both before and after the wash. The differences between the pairs of evaluations formed the responses and are shown in Table 11.33. The order of the samples was randomized and unknown to the evaluator.

(a) Show that, for this design, pairs of treatments occur together in λ_1 or λ_2 or λ_3 blocks, where these values are different. (This means that there are three different lengths of confidence intervals for treatment contrasts). Thus, argue that the design cannot be a balanced incomplete block design or a group divisible design.

(b) For the standard block–treatment model (11.4.2), p. 356, with $b = 6$ blocks and $v = 18$ treatments, write down the contrast coefficient lists for the following contrasts:

   (i) the comparison of treating the stain immediately as compared with the average of the 1 and 2 day delays,

Table 11.33 Responses for the fabric stain experiment

Block	Treatments (Response)					
1	1 (35)	6 (35)	8 (20)	10 (80)	15 (65)	17 (10)
2	2 (30)	4 (25)	9 (10)	11 (75)	13 (60)	18 (15)
3	3 (30)	5 (30)	7 (30)	12 (75)	14 (60)	16 (15)
4	1 (20)	5 (30)	9 (15)	10 (80)	14 (65)	18 (20)
5	2 (20)	6 (15)	7 (10)	11 (80)	15 (65)	16 (15)
6	3 (20)	4 (15)	8 (20)	12 (70)	13 (50)	17 (15)

(ii) the comparison of no pre-treatment versus the average of the two types of pre-treatment,
(iii) the comparison of a 1 or 2 day delay for each level of pre-treatment separately.

(c) Prepare an analysis of variance table and test the hypothesis that each contrast in part (b) is negligible. Use an overall significance level of not more than 0.1. State your conclusions.
(d) Using Scheffé's method of multiple comparisons at confidence level 95%, state which pairs of treatments are significantly different and give the confidence intervals for these.
(e) From Scheffé's multiple comparisons calculated in part (d), verify that the confidence intervals for $\tau_1 - \tau_p$, for $p = 2, \ldots, 18$, are of three different lengths, and that the shorter lengths correspond to larger values of λ_i.
(f) Using a factorial model for the treatment effects, prepare an analysis of variance table. State your conclusions and, repeat part (c) for the first two contrasts in (b). Use the mapping of treatment labels to levels of A, B, C as:

$$\begin{array}{llllll} 1 = 111 & 2 = 112 & 3 = 113 & 4 = 121 & 5 = 122 & 6 = 123 \\ 7 = 131 & 8 = 132 & 9 = 133 & 10 = 211 & 11 = 212 & 12 = 213 \\ 13 = 221 & 14 = 222 & 15 = 223 & 16 = 231 & 17 = 232 & 18 = 233 \end{array}$$

(g) Does this design have "factorial structure" as defined in Sect. 11.7.1?

17. **Air rifle experiment**

This is a dangerous experiment that should not be copied! It requires proper facilities and expert safety supervision.

An experiment was run in 1995 by C.-Y. Li, D. Ranson, T. Schneider, T. Walsh, and P.-J. Zhang to examine the accuracy of an air rifle shooting at a target. The two treatment factors of interest were the projectile type (factor A at levels 1 and 2) and the number of pumps of the rifle (factor B, 2, 6, and 10 pumps, coded 1, 2, 3). The paper covering the target had to be changed after every four observations, and since there were $v = 6$ treatment combinations, coded 11, 12, 13, 21, 22, 23), an incomplete block design was selected.
Two copies of the following incomplete block design (called a generalized cyclic design) were used:

Table 11.34 Data for the air rifle experiment

Block	Treatment combination (Response)			
I	22 (2.24)	23 (6.02)	12 (11.40)	11 (26.91)
II	13 (7.07)	22 (8.49)	21 (19.72)	11 (24.21)
III	12 (10.63)	23 (6.32)	21 (9.06)	13 (29.15)
IV	11 (11.05)	22 (6.32)	23 (7.21)	13 (23.02)
V	23 (6.71)	22 (15.65)	12 (11.40)	11 (23.02)
VI	21 (17.89)	12 (8.60)	22 (10.20)	13 (10.05)
VII	11 (18.38)	13 (11.18)	23 (11.31)	22 (11.70)
VIII	22 (1.00)	11 (30.87)	21 (20.10)	13 (17.03)
IX	11 (18.03)	21 (8.25)	23 (6.08)	12 (19.24)
X	21 (15.81)	13 (2.24)	12 (17.09)	22 (7.28)
XI	23 (8.60)	21 (15.13)	12 (14.42)	11 (25.32)
XII	13 (8.49)	12 (14.32)	23 (11.66)	21 (17.72)

Block					
I	II	III	IV	V	VI
11	12	13	21	22	23
21	22	23	11	12	13
13	11	12	23	21	22
22	23	21	12	13	11

The total of 12 blocks were randomly ordered, as were the treatment combinations within each block. The data, shown in Table 11.34, are distances from the center of the target measured in millimeters.

(a) Write down a suitable model for this experiment.
(b) Check that the assumptions on your model are satisfied.
(c) The experimenters expected to see a difference in accuracy of the two projectile types. Using a computer package, analyze the data and determine whether or not this was the case.
(d) For each projectile type separately, examine the linear and quadratic trends in the effects of the number of pumps. State your conclusions.