STATISTICAL METHODS IN BIOLOGY

Design and Analysis of Experiments and Regression

S. J. Welham

Rothamsted Research, Harpenden, UK

S. A. Gezan

University of Florida, USA
(formerly Rothamsted Research, Harpenden, UK)

S. J. Clark

Rothamsted Research, Harpenden, UK

A. Mead

Rothamsted Research, Harpenden, UK
(formerly Horticulture Research International, Wellesbourne, UK & University of Warwick, UK)

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20140703

International Standard Book Number-13: 978-1-4398-9805-5 (eBook - PDF)

3

Principles for Designing Experiments

This chapter presents the basic concepts that are required to construct designs to address directly and efficiently the aims of a biological experiment. We first discuss the choice of treatments and materials; treatments should be determined by the aims and the materials should be chosen according to the frame of reference for the experiment (Section 3.1). These two components must then be combined to produce an appropriate design. A good design takes proper consideration of three statistical principles: replication (Section 3.1.1), randomization (Section 3.1.2) and blocking (Section 3.1.3), to reduce bias and maximize the precision of treatment comparisons. We describe the structure of a design with respect to underlying factors using a symbolic form (Section 3.2). Many experiments will use one of the wide and flexible family of standard designs, such as the completely randomized design (CRD) (Section 3.3.1), the randomized complete block design (RCBD) (Section 3.3.2), the Latin square (LS) design (Section 3.3.3), the split-plot (SP) design (Section 3.3.4) or the balanced incomplete block design (BIBD) (Section 3.3.5). Once the structure of the design has been determined, a properly randomized layout can be generated with statistical software (Section 3.3.6).

3.1 Key Principles

As described in Chapter 1, an experimental study investigates the relationship between an outcome and one or more conditions that are manipulated by the researcher. Before considering the appropriate design for any experiment, it is important to be clear about its aims, which are usually associated with one or more scientific questions or hypotheses to be tested. Examples of such questions might be

- Is any reduction in disease infection achieved with a new 'resistant' variety compared with a standard 'control' variety?
- How do plant metabolites respond to increasing drought stress at different stages of development?
- Which chemicals, of several under study, show insecticidal activity?
- How is yield related to plant spacing, and does this relationship vary between varieties?

The aims of the experiment should be well defined to make it easy to assess whether the chosen treatments are sufficient to achieve them. In this context, the term **treatments** is used to describe the set of different experimental conditions to be tested, for example, varieties, nitrogen rates, or chemical compounds, or, more usually, combinations of several such classifying variables. Control treatments – either positive or

negative controls – may be used to provide a baseline, or to verify that the experiment has worked as expected. A negative control usually corresponds to a 'null' treatment, and a positive control usually corresponds to a standard treatment with a known effect. This is discussed further in Section 8.5. In addition to defining the experimental treatments, the experimental units must be chosen. The **experimental unit** for a treatment is defined as 'the smallest division of the experimental material such that any two units may receive different treatments in the actual experiment' (Cox, 1992). For some treatments, this may be larger than the size of unit on which individual observations are recorded (sometimes called the **observational** or **measurement unit**), and may occur at a range of scales, as in the following examples.

- *An area of land on a farm*. A field trial typically has numerous small plots, and experimental treatments are applied to the individual plots. The experimental unit is the plot, and the measurement unit may be either the plot (e.g. yield) or sub-samples from the plot area (e.g. individual plant measurements).
- *Individual soil samples taken from a field*. In the context of a field trial with treatments applied to plots, if a single soil sample is taken from each field plot for processing in the lab, then the soil sample becomes the experimental unit. If multiple soil samples are taken from each plot, then the experimental units are the sets of samples from each plot.
- *Pots, each containing three plants*. If experimental treatments, such as soil nutrient content, are applied to whole pots, then the pot is the experimental unit. The measurement unit may be either the whole pot (e.g. combined biomass) or individual plants.
- *Different leaves from an individual plant*. In the investigation of the response of plants to aphid attack, clip cages with or without aphids might be attached to individual leaves within a plant. The experimental unit is then the individual leaf.
- *Samples of RNA extracted from different plants*. Investigation of gene expression often involves the application of different treatments to individual plants, followed by extraction of RNA from each plant. The experimental unit for further study is then the RNA taken from an individual plant.
- *A batch of 10 insects in a Petri dish*. Experiments on small insects are often done on groups of insects kept together in dishes (or cages), with treatments applied to the dishes. The experimental unit is then the dish. The measurement unit may be the dish, via a summary of insect behaviour such as percent survival, or the individual insects.

Recall from Section 1.2 that the experimental units are considered to consist of a sample from a wider population for which inferences can be made, and that this population should be identified according to the frame of reference for the experiment. For example, if RNA samples for microarray work are taken from only a single plant, then conclusions regarding gene expression in the plant population cannot safely be made without further experimentation, because variation between different plants would be expected. If samples are taken from several randomly selected plants, then variation between plants can be accounted for, and inferences can be applied to the wider population. A similar situation occurs when an experiment is established in a single site, or in a single year, or on a single variety, as there can be no certainty that results can be safely extrapolated to wider circumstances. This issue is especially relevant to field trials, where the chance peculiarities of a

single environment can produce anomalous results; for this reason, many journals will not publish the results of field trials that have not been repeated over several sites or seasons or both. It is therefore important to recognize the frame of reference implied by the choice of experimental units, so that appropriate conclusions can be drawn from the results.

Although we wish to have a representative sample of experimental units, we can get more precise estimates of differences between treatments by making comparisons across similar units. We can deal with this apparent contradiction by using sets of reasonably homogeneous experimental units, and then repeating the comparison across a wider range of circumstances. In some cases, the experimental units may have some intrinsic structure, introducing some heterogeneity between groups of more homogeneous units. This structure should be incorporated in both the choice of experimental units for treatment application and in the statistical analysis. For example, experimental materials may be arranged as plants within pots within trays, giving three structural levels, and we expect different levels of variation within each of these levels. Depending on practical considerations and the aims of the experiment, it may be appropriate to apply treatments at any of these levels and a statistical analysis should account for this structure.

The choice of experimental units and treatments should be made separately: units are chosen according to the appropriate frame of reference for the experiment, and treatments are chosen to enable the hypotheses to be tested. A good design then matches the treatments with the units so that the treatment differences can be estimated without bias (i.e. without systematic over- or under-estimation) and as precisely as possible (i.e. to minimize uncertainty in the results). Our main tool to avoid experimental bias is randomization, i.e. the random allocation of treatments to experimental units, and experimental precision can be improved by the use of proper replication and blocking (terms we discuss in more detail below).

First, it is helpful to recall the role of the underlying unit-to-unit variation in biological experimentation. It is well known that biological individuals vary in any given characteristic or response. The amount of variation may depend on several factors, such as differing genetic backgrounds and environmental effects, but some variation is always present. This natural variation may be inflated by uncertainty introduced by the measurement process in cases where exact measurement is not possible (also known as measurement error). This combined background variation is a potential cause of both bias and uncertainty in experimental results. For example, if two treatments are each applied to one plant only, it is not possible to assess whether any difference in the measured response is due to treatment differences or natural plant-to-plant variation. Statistical design and analysis aim to distinguish, quantify and, subsequently, compare variation between treatments (signal) with background variation (noise). A large signal:noise ratio indicates that substantive treatment differences are present. A small signal:noise ratio indicates that any apparent treatment differences could be explained by the background variation in the system, and therefore cannot confidently be attributed to treatment effects. Proper identification and control of background variation is thus an essential aim of any statistical design.

A good design considers each of the three basic principles: replication, randomization and blocking. **Replication** is the process of applying each treatment to more than one experimental unit, so the number of replicates of a treatment is the number of independent experimental units to which each treatment is applied. **Randomization** means the random allocation of treatments to experimental units and is used to ensure the fair assessment of treatments without bias. For this reason, it can be regarded as an insurance against potential unknown differences between units, and it should be used whenever possible. In some circumstances, it may be possible to identify or construct groups of experimental units

expected to have similar responses in the absence of any treatment effects. This process is known as **blocking**. A **block** is a subset of the experimental material within which experimental units are expected to be homogeneous, with more heterogeneity expected between experimental units in different blocks. In the analysis of experimental data, variation due to blocks can be separated from background variation and, if there are differences among blocks, this separation will increase the precision of treatment comparisons by reducing the estimate of the unit-to-unit background variability. An appropriate use of these three design principles will give confidence that any treatment differences observed are real and not due to some chance combination of circumstances, and will also enable the maximum amount of information to be obtained from the available resources. We now discuss each of these principles in more detail.

3.1.1 Replication

The natural background variation among experimental units means that it is necessary to replicate the application of each treatment to several experimental units. This replication serves two important purposes. First, by repeating each treatment on several experimental units, we get a more reliable estimate of the effect of each treatment. Second, and possibly more importantly, the replicated observations provide an estimate of the background variation between units, which we can use to assess whether treatments differ and to indicate the precision associated with the estimated treatment effect. Usually each treatment will be replicated an equal number of times, but in circumstances where particular treatments are of greater interest, it may be advantageous to have increased replication for those treatments. Conversely, reduced replication may be used where resources for particular treatments are either scarce or expensive, for example, seed for a new breeding line.

To illustrate some issues regarding replication, consider an experiment to compare two pesticide treatments (a standard and a new formulation) applied to six insect-net cages, each cage containing 10 aphids. The new formulation is applied to three cages selected at random, with the standard formulation applied to the remaining three cages. The replication of each treatment in this experiment is only three, even though 30 aphids have been treated with each pesticide and even if the measurement unit is the individual aphid, because each treatment is applied to a cage of aphids and so this is the experimental unit. The individual aphids here are an example of pseudo-replicates. **Pseudo-replication** describes the situation in which multiple measurements are taken from each experimental unit. This can be a very useful experimental technique, but must be properly incorporated into any statistical analysis, which may otherwise produce an incorrect estimate of the between-unit variability (usually too small), possibly leading to incorrect conclusions about the importance of treatment effects. Pseudo-replication usually causes problems when the smallest level of experimental material (i.e. the measurement unit, here, the aphid) is wrongly identified as the experimental unit in a statistical analysis, in place of the level at which treatments were actually applied (here, the cage). As a rule of thumb, replication needs to occur at the level at which the treatments have been applied to be considered 'real'. Consider the following examples of designs for experiments.

- Twelve pots, each containing four plants at the three-leaf stage, with six treatments (A–F) each applied to two of the pots with the allocation made at random (Figure 3.1). Treatments were applied to pots, and so the pot is the experimental unit and the replication of each treatment is two. Measurements from individual plants are pseudo-replicates.

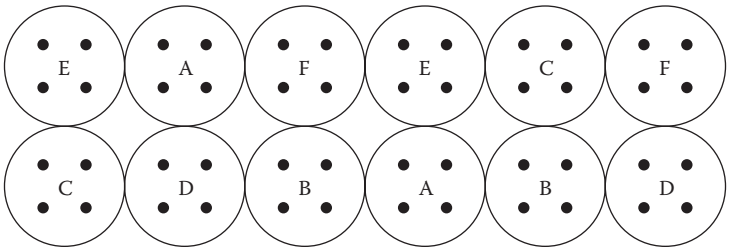

FIGURE 3.1
Design for an experiment with four plants (•) in each of 12 pots with treatments (A–F) applied to pots.

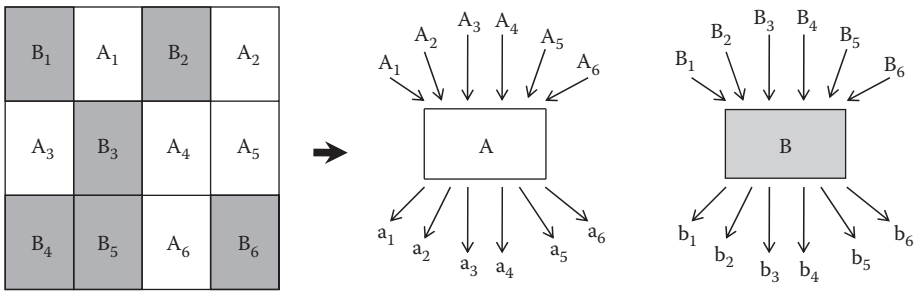

FIGURE 3.2
A two-stage design. The first stage (left) is a field trial with 12 plots and two treatments (A and B). A soil sample is taken from each plot and labelled by its treatment and replicate (A_1–A_6 and B_1–B_6). At the second stage, samples from each treatment are mixed together (bulked) then sub-sampled. Sub-samples from each bulked sample are labelled by lower-case letters and sample number and then measured.

- A field experiment consisting of 12 plots, with two treatments (A and B) each applied to six of the plots selected at random (Figure 3.2). One soil sample was taken per plot, and labelled by the treatment and replicate number, giving samples A_1 … A_6 and B_1 … B_6. The soil samples from the six replicate plots for each treatment (e.g. A_1 … A_6) were bulked (combined) together and mixed thoroughly and six sub-samples taken and measured. These sub-samples were labelled as a_1 … a_6 and b_1 … b_6. In this case, although treatments were originally applied to plots, at the analysis stage there is only a single replicate for each treatment because samples from independent plots have been bulked and the sub-samples are not independent. The sub-samples are pseudo-replicates and give information on the homogeneity of the bulked sample rather than on the variation between plots. Ideally, samples from each plot should have been kept separate, giving six true replicates for each treatment.
- Two controlled environment (CE) cabinets, one at 10°C and one at 20°C, each containing eight seed trays, with two different watering regimes (A and B) each applied to four trays chosen at random within each cabinet (Figure 3.3). Both temperature and watering regime are considered as treatments here. The experimental units for watering regime are the seed trays and each watering regime is applied to eight independent seed trays, giving replication of eight. The experimental units for temperature are the cabinets, and the temperature treatments are unreplicated. To achieve replication of temperature, it would be necessary to use another two cabinets, or to repeat the study under the same controlled conditions with a new randomization of both factors.

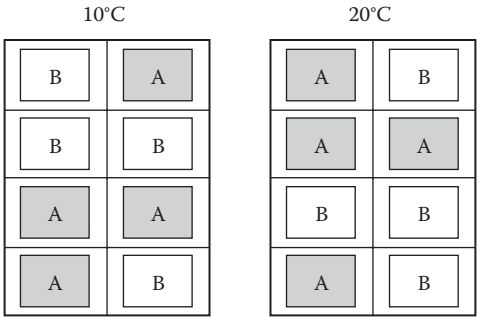

FIGURE 3.3
Design for an experiment with eight trays, with two watering regimes (A and B) applied within each of two CE rooms, with each room operating at a different temperature (10°C or 20°C).

The last of these examples shows that experimental units can occur at several different levels within a structure, and may differ between treatment factors. Therefore, one level of structure may represent pseudo-replicates for one type of treatment and real replicates for another.

It is also important to draw the distinction between technical and biological replication. **Technical replication** occurs when several measurements are taken from the same biological material, while **biological replication** occurs when measurements are taken from several independent biological subjects. The use of adequate biological replication is required to make inferences valid for the population from which the samples were obtained, rather than for a single individual. Technical replication is always pseudo-replication, but biological replication may correspond to either pseudo-replication or true replication, depending on the context. It is clearly important to recognize when measurements are pseudo-replicates that do not increase treatment replication. Note however, that technical replication can be useful in increasing precision where measurement is seriously subject to error, provided that it is properly accounted for in the statistical analysis. We give an example of an analysis accounting for pseudo-replication in Section 7.5.

3.1.2 Randomization

Randomization is required to ensure the fair allocation of treatments to units to guard against bias, and to cope with the natural variation between experimental units. In the simplest case, randomization requires that each permutation of the set of treatments has an equal probability of occurring, so that (for equal replication) every experimental unit has an equal chance of receiving any treatment. Hence, each treatment is equally likely to be applied to 'good' units as to 'bad' units. Where randomization has not taken place, there will always be a question about possible bias in the experiment.

To obtain a proper randomization for a given design, a method is required for assigning treatments to experimental units at random. We use the convention that treatments get assigned to experimental units, though the opposite approach can also be used. So, for example, for an experiment comprising two treatments each replicated six times, we might write A and B on six pieces of paper each, to represent the six replicates of each treatment, and put these in a bag and draw (without replacement and without looking) to obtain a sequence which allocates treatments to units 1 … 12. Alternative approaches could use six tokens of each of two different colours, or six playing cards from each of two different

suits, in the same way. Random number tables can also be used for allocating treatments to units, though care is needed to define the protocol where too many repeats of a treatment occur in the random sequence. However, most randomizations are now done via statistical software. The mechanics of the process are unimportant as long as the property of equal probability for each permutation of treatments is preserved – this is discussed further in Example 3.1. Randomization ensures that any (possibly unconscious) bias of the experimenter (e.g. a tendency to assign the biggest plants to their favoured treatment) is avoided and that any unknown differences between the units are unlikely to consistently favour particular treatments.

To reinforce objectivity in some areas of research, particularly medical research, trials are carried out as either single- or double-blind trials. In a single-blind trial, the subject does not know which treatment has been allocated, while in a double-blind trial, neither the subject nor the investigator knows which treatment has been applied. In plant science, the perception of the subject is not considered relevant. However, the perception of the investigator measuring or assessing the experimental material could be influenced (possibly unconsciously) by their expectation of the applied treatment. It is therefore good practice to make experimental measurements (especially subjective assessments) without knowledge of the treatment allocation as far as possible. For example, in field trials, this can sometimes be achieved by the use of a field plan with plot numbers marked but not treatments.

Randomization leads to estimates of treatment differences that are unbiased when considered across the whole set of possible randomizations. However, this does not guarantee that any individual randomization will produce unbiased results. For example, all instances of one treatment may be assigned to larger plants by chance. For this reason, experimental units should be chosen to be as homogeneous as possible while still being representative of the population of interest. Selecting homogeneous units has the added advantage of reducing the background variation or noise. Where it is not possible to select a completely homogenous set of experimental units, the units need to be grouped into sets (blocks) of more homogeneous experimental units to avoid potential bias (see Section 3.1.3). Sometimes, even where the units are thought to be homogeneous, a randomization can give cause for concern. For example, for 12 pots arranged in a line with two treatments (labelled A and B) each replicated six times, consider the following randomization

A A A A A A B B B B B B.

This particular randomization does not look random, but can occur (with probability 1 in 924). If we are not happy to accept this randomization, it is probably an indication that we do not consider the experimental units to be completely homogeneous, so that some sort of blocking is needed.

EXAMPLE 3.1: RANDOMIZATION

The efficacy of a new pesticide is to be tested in the field with 15 plots of size 5 m $\times$ 10 m arranged in a 3×5 array. Five plots will be sprayed with the pesticide and 10 will be untreated (controls) for comparison. In this case, extra replication of the control treatment is used to obtain a good estimate of background variability and because the new pesticide is available in only small amounts. We evaluate two methods of determining a randomization for this experiment and consider whether each of the methods gives a valid randomization, i.e. with equal probability for each permutation of treatment effects.

TABLE 3.1

Experimental Layout Achieved Using Randomization by Playing Cards for a Field Trial with 15 Plots (Numbered 1–15) and Two Treatments: A Pesticide Treatment (Labelled P) with Five Replicates and a Control Treatment (Labelled C) with 10 Replicates (Example 3.1)

1	2	3	4	5
P	P	C	P	C
6	7	8	9	10
C	C	P	C	P
11	12	13	14	15
C	C	C	C	C

First, we use a pack of cards. We might take 15 cards: five red cards to represent the pesticide treatment and 10 black cards to represent the control. We shuffle the cards (randomization), then deal them out in shuffled order to allocate treatments to plots (in order 1–15) to get, for example, the randomization shown in Table 3.1.

Is this a valid randomization? Let us consider the process. Assuming a fair shuffle, when we pick the first card we have a probability of 5/15 of picking the pesticide treatment for the first plot (and 10/15 of picking the control). These probabilities change as the allocation proceeds. For example, if the first plot is allocated the pesticide treatment, when we pick the second card we have a probability of 4/14 of the second plot also being allocated the pesticide treatment (as we have four red cards out of 14 left), and so on. With this method, the probability of plots 1–5 all being allocated the pesticide treatment is 1/3003 (= $5/15 \times 4/14 \times 3/13 \times 2/12 \times 1/11$). This is the same, for example, as the probability of plots 11–15 all being allocated the pesticide treatment (equivalent to the probability of plots 1–10 being allocated the control, i.e. $10/15 \times 9/14 \times \ldots \times 2/7 \times 1/6 = 1/3003$). There are, in fact, 3003 possible permutations of five pesticide-treated plots and 10 control plots, calculated as the factorial function of 15 (the number of plots) divided by the product of the factorial function of five (the number of pesticide-treated plots) and the factorial function of 10 (the number of control plots):

$$\begin{aligned}\frac{15!}{5! \times 10!} &= \frac{15 \times 14 \times 13 \times \ldots \times 2 \times 1}{(5 \times 4 \times 3 \times 2 \times 1) \times (10 \times 9 \times \ldots \times 2 \times 1)} \\ &= \frac{15 \times 14 \times 13 \times 12 \times 11}{5 \times 4 \times 3 \times 2 \times 1} \\ &= 3003\,.\end{aligned}$$

With this randomization approach, each of these permutations is equally likely.

An alternative, and perhaps at first sight simpler, approach is to toss a coin to construct the randomization, with heads corresponding to the pesticide treatment, and tails corresponding to the control. Working through the plots one by one, we toss the coin once for each plot, allocating the pesticide treatment to the plot if it comes up heads (subject to a maximum of five pesticide plots), and allocating the control treatment to the plot if it comes up tails (subject to a maximum of 10 control plots). Is this a valid randomization? Let us again consider the process. We have a probability of 1/2 (assuming a fair coin) of allocating pesticide to the first plot. The second coin toss takes no account of the allocation for the first plot, so again we have probability 1/2 of pesticide being allocated to the second plot, and so on. With this method, the probability of plots 1–5 being allocated the pesticide treatment is 1/32 (= $1/2 \times 1/2 \times 1/2 \times 1/2 \times 1/2$). In contrast, the probability of plots 11–15 being allocated the pesticide treatment (i.e. plots 1–10 being allocated the

control) is $(1/2)^{10}$ (= 1/1024). The probabilities of these different permutations are obviously not the same, and the probability of any particular permutation depends on how we number the plots!

It is clear that the two processes are not equivalent. The 'coin tossing' approach gives an invalid randomization, with different permutations having different probabilities. By contrast, the 'card shuffling' approach associates the same probability with each permutation, and therefore provides a valid randomization. This example illustrates some of the issues that must be considered when you derive a randomization scheme, and that are automatically accounted for by statistical software.

Some other examples of randomization are presented in Figures 3.1 to 3.3 and in Section 3.3. Note that if an experiment is repeated, then a new randomization should be generated each time the design is used; it is not statistically valid (or sensible) to generate a single randomization and then to use it repeatedly.

3.1.3 Blocking

It is desirable for the set of experimental units that are used to compare treatments to be reasonably uniform (homogeneous) in their natural response, as this decreases our estimate of the background variation, thus increasing precision and the potential for the experiment to detect small treatment differences. So, if the experimental units are intrinsically diverse (heterogeneous), then the experiment is likely to be insensitive. Further, as noted in Section 3.1.2, using a set of homogeneous units increases the chances of a fair comparison between treatments. However, it is not always possible to obtain a sufficient number of homogeneous experimental units for a whole experiment and, even if it is possible, it might not be desirable if it means restricting the frame of reference for the experiment. In such cases, it might be possible to identify groups of experimental units such that the units within each group are reasonably homogeneous, but with different underlying responses between groups. These groups of units can then be considered as blocks within the design. Blocking the units in this way potentially increases the precision of an experiment, as comparisons between treatments within blocks are made against a more uniform background. In this sense, blocking is said to be used for the control of variation, and for this reason is also known as **local control**.

The term block originated in agricultural experiments, where a block corresponded to a set of contiguous field plots; however, the specification of blocking can take more general forms, including the recognition of any physical structure present in the experiment. We often use the term 'block' as synonymous with 'structure'. Blocks may therefore be defined according to proximity of units in space (e.g. neighbouring plots), proximity of units in time (e.g. units measured in the same day or hour), units with similar physical characteristics (e.g. size of plant, age of insect), or logistical factors (e.g. machine, technician). Note that the number of units per block should ideally be determined by consideration of the uniformity or structure of experimental units and not by what is convenient in relation to the number of experimental treatments.

Consider the following examples of types and causes of heterogeneity among experimental units, which can be addressed by the use of blocking.

- *Field characteristics.* A slope, or fertility or pH trend across a field, or local pest problems (e.g. pigeons next to woodland) may be present. Blocks are usually formed from sets of contiguous plots that are expected to be similar in as many respects as possible. Occasionally, blocks may be formed from non-contiguous plots with

similar properties, for example, soil pH, but, in such cases, the other spatial characteristics need to be reasonably homogeneous.

- *Glasshouse characteristics*. Differential shade or temperature due to positioning with respect to walls and doors are common in glasshouses. Blocks are usually formed from sets of trays or pots placed close together and hence in similar environmental conditions.
- *Time of measurement*. Some experiments may be processed over a lengthy period, and time of measurement may have a systematic effect on results. In the laboratory, there may be a limit to the number of samples that can be measured in one batch, and equipment may give slightly different readings on different days. In either case, a set of units processed within the same time period can be considered as a block.
- *Investigator*. For subjective measures, such as visual scores, individuals often perceive pre-determined scores differently. However, even in more objective situations, for example, an investigator following a standard protocol, the use of subtly different procedures can lead to systematic differences in results. If several different investigators are scoring material or carrying out a laboratory process, then it makes sense to regard each person as a block.
- *Batches of chemical, of plants, or of other organisms such as insects*. Again, if there is any possibility of (even small) differences between batches, then batches should be considered as blocks.
- *General structure*. There will often be a natural structure in experimental material. For example, trays of plants may be held on shelves within a CE cabinet. Conditions are more similar for plants within the same tray, for trays on the same shelf, and for shelves within the same cabinet (in the case of several cabinets), so all of these levels of structure should be considered as possible blocking factors.

In each of the examples above, information on the causes of heterogeneity is used to define blocks of reasonably homogeneous units and treatments can then be assigned at random to units within blocks. Note that each block might not be able to contain the full set of treatments (see Section 3.3.5), and that all blocks might not even contain the same numbers of experimental units. The randomization process needs to take account of the structure of the blocks, so that each treatment has the same probability of being applied to any unit within each block. If there are large differences between blocks, this also ensures a fairer allocation of treatments to units, as each treatment will occur within several (often all) blocks. For this reason, blocking can be seen as a set of restrictions on the randomization of treatments to the experimental units. We consider this in more detail for specific designs later (Section 3.3). Note that although blocking is generally intended to increase the precision of treatment comparisons where groups of heterogeneous units are present, the precision may decrease if too much blocking is used where there is no heterogeneity.

3.2 Forms of Experimental Structure

To successfully design, and later analyse, an experiment, it is necessary to identify all components of the experiment, i.e. both the treatments imposed and the structure of the units. In Section 1.3, we partitioned the systematic part of our mathematical model into two

components: the explanatory component describes the treatments present and the structural component describes the blocking, or other structure, of the experimental units. We describe both components using factors which label the different groups present. Often, several factors are required to describe each component fully. For example, in a CE experiment where trays were placed on shelves within cabinets, we need factors to label each of the cabinets, shelves and trays to fully describe the structure. Similarly, a set of experimental treatments may be constructed from an underlying set of treatment factors. A factorial treatment structure consists of all possible experimental treatments constructed by taking one level from each of a set of treatment factors; this gives a particularly efficient form of design and is discussed further in Sections 8.2 and 8.3.

We write our model components using a symbolic notation similar to that commonly used in statistical software. To use this notation effectively, we first need to understand two different types of relationships between factors, namely nested and crossed structures.

Nested structures are used to describe hierarchical relationships. These most often occur within the structural component, but also occasionally within the explanatory component (e.g. see Section 8.4). A nested structure describes the situation where multiple units at one structural level are entirely contained within units at a higher level, and there is no direct relationship between units with the same label at the lower level. For example, consider an experiment with different treatments to be applied to four leaves (factor Leaf, with four levels) within each of 10 plants (factor Plant, with 10 levels). Leaves within plants are the experimental units, and we consider the Leaf factor to be nested within the Plant factor, written symbolically as Plant/Leaf. In this hierarchical structure, there is no association between leaves with the same label across plants, for example there is no association between leaf 1 on plant 1 and leaf 1 on plant 2. The / **(forward slash) operator** is used to indicate a nested relationship. In fact, this operator generates two separate model terms, as

$$\text{Plant/Leaf} = \text{Plant} + \text{Plant.Leaf}$$

The first term consists of the Plant factor alone, and labels each of the 10 individual plants. In the second term, the . **(dot) operator** generates all combinations of levels of the two factors, in this case labelling the 40 individual leaves. These two terms label the units within the two levels of the design.

Crossed structures occur when two factors are used to classify experimental units both independently and simultaneously. This type of structure occurs frequently within both the explanatory and structural components of the model. For example, consider a laboratory experiment to examine an extraction procedure in which three different filtering methods (factor Filter, with three levels) are tested with four different reagents (factor Reagent, with four levels), giving 12 experimental treatments in total. Both factors act simultaneously, and the crossed structure can be written as Filter*Reagent. In a crossed structure, there is an association between units with the same level of either factor. The * **(star) operator** indicates a crossed relationship and again generates several model terms, as

$$\text{Filter*Reagent} = \text{Filter} + \text{Reagent} + \text{Filter.Reagent}$$

The first two terms are the individual factors, and the third term labels all combinations of the two factors, here the 12 individual treatments. The interpretation of these terms is discussed further in Section 8.2.

In the structural component of the model, the terms generated describe different levels of the design at which variation may occur; these different levels are known as **strata**. For example, the crossed structure of a rectangular layout, Row*Column, generates three strata (Row, Column and Row.Column) and the nested structure Plant/Leaf generates two strata (Plant and Plant.Leaf).

In general, either of the model components may contain nested or crossed relationships or both. Examples 3.2 and 3.3 describe some specific situations in the context of the structural component; in Chapter 8, we consider examples in the context of the explanatory component.

EXAMPLE 3.2: NESTED STRUCTURAL FACTORS

An experiment is set up with two identical CE rooms with three trays, each containing six pots, within each room (see Figure 3.4), with the potential to allocate different treatments at both the tray and pot levels.

The CE rooms can be considered as the highest level of structure, with trays within rooms as the middle level, and pots within trays as the lowest level. This nested structure thus has multiple units at any one level (e.g. trays) completely contained within each unit at the level above (e.g. rooms). We can verify that this is a nested structure by noting that there is no association between tray 1 in room 1 and tray 1 in room 2, and similarly no association between pots with the same label in different trays. The structural factors can be denoted as Room (two levels), Tray (trays labelled within rooms, three levels) and Pot (pots labelled within trays, six levels) and we can write this as

Structural component: Room/Tray/Pot

which can be expanded as

Structural component: Room + Room.Tray + Room.Tray.Pot

The three terms label the three strata, or levels of the hierarchy: the individual rooms (term Room), the individual trays (term Room.Tray) and the individual pots (term Room.Tray.Pot). The appropriate experimental unit for the application of any treatment must then be decided as a separate exercise.

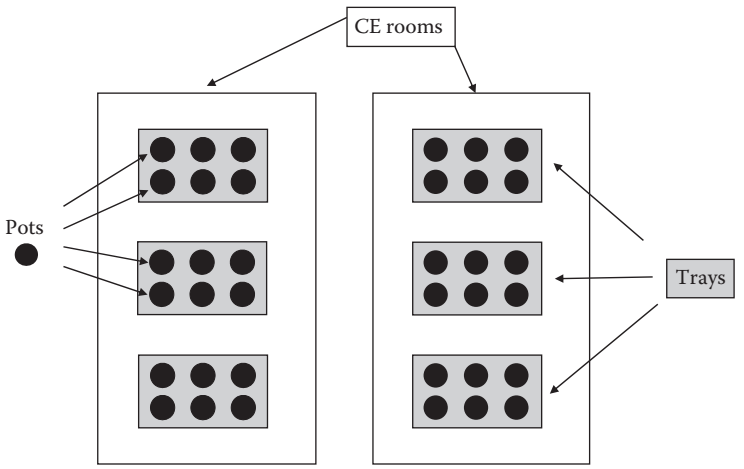

FIGURE 3.4
A nested structure with six pots per tray and three trays per CE room (Example 3.2).

EXAMPLE 3.3: CROSSED AND NESTED STRUCTURAL FACTORS

Consider an experiment where a large set of plant samples are to be processed by a machine. There are two machines that could be used (factor Machine, with two levels) and two scientists available to do the work (factor Scientist, also with two levels). We might want to allow for potential differences in results both between scientists and between machines. If we want to separate the effects of the different scientists from the effects of the two machines, then both scientists must use both machines. So an appropriate design might allocate four sets of samples to be processed by the four machine-by-scientist combinations. Each of the machine-by-scientist combinations can be considered as a block (see Figure 3.5).

Because there may be an association between samples processed either by the same machine or by the same scientist, this is a crossed relationship. The structure of the four blocks can then be written with our symbolic notation as

Machine*Scientist

which can be expanded as

Machine + Scientist + Machine.Scientist

These three terms, or strata, describe an overall effect for each machine, an overall effect for each scientist, and a combined effect for each machine-by-scientist combination. There is no association across samples processed by different machine-by-scientist combinations, and so samples can be considered to be nested within these blocks. The full structure can thus be written as

Structural component: (Machine*Scientist)/Sample

and expanded as

Structural component: Machine + Scientist + Machine.Scientist
+ Machine.Scientist.Sample

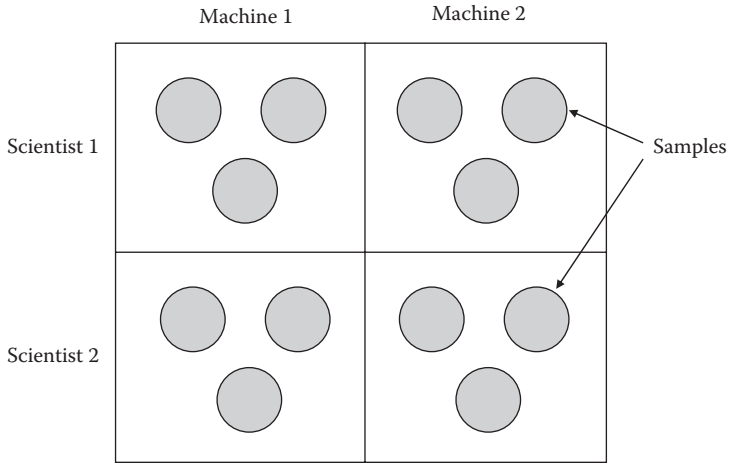

FIGURE 3.5
A crossed and nested structure with three samples within each of four machine-by-scientist combinations (Example 3.3).

So the full structure has four strata: the three described above, plus a fourth that labels the full set of individual samples. One advantage of using the crossed structure at the higher level is that it allows us to establish whether differences between scientists and machines are present, and to estimate their relative size and potential impact on the results. This information can be useful in designing future experiments although, in practice, we would require several repeats of this structure to get reliable estimates of the different sources of variation. An alternative structure for this experiment might combine (or, in statistical terminology, **confound**) the scientist and machine effects, so that each scientist uses just one of the two machines. In this case, the effects of the two scientists and the effects of the two machines would not be separable. Whether this is important depends on the information required from the experiment: this confounded blocking still achieves the main aim of separating block variation from the background variation, but does not allow us to compare the relative influence of the scientists and machines used in the process.

The symbolic notation for model formulae that we have used here was introduced by Wilkinson and Rogers (1973) and is used within GenStat. Unfortunately, conventions for model specification differ somewhat between statistical packages. For example, the R package uses the : (colon) operator where we have used the . (dot) operator. Details can be found by consulting software documentation.

To define a model fully, we need to specify in full both the explanatory and structural components. For analysis, we also need to define the response variable to be analysed. In the examples above, and throughout this book, we have included the individual units within the structural model. This is not strictly necessary, as the individual units obviously correspond to individual observations. However, we believe that retention of information on the full design structure as well as the treatment factors from an experiment is good practice. For example, unless we retain full information on the design, we cannot use all of the diagnostic procedures described in Chapter 5; we cannot plot residuals on the experimental layout if we do not know where each unit was placed. If a few pots in a corner of a glasshouse behave differently to the rest, the cause may be obvious when observations are plotted on the experimental layout but not when examined by treatment classification. Unfortunately, it is not common practice to record this information, for example, we have been unable to determine the full experimental layout for many of the examples in this book. Where this is the case, we use dummy structural factors, for example, we use factor DPot, to arbitrarily label pots, with the D prefix indicating a dummy factor. Treatment factors can sometimes be used as dummy structural factors, but we believe that this practice is confusing and prefer to avoid it. We discuss this further at the end of Section 7.3.

Finally, we mention two concepts often used to describe the structure of designs: orthogonality and balance. Two factors are said to be **orthogonal** if the estimated effects for each factor are the same regardless of whether the other term is included or not in the model. A more rigorous mathematical definition of orthogonality is beyond the scope of this book (but details can be found in Bailey, 2008). Most of the designs that we consider in this book are orthogonal, and the concept and consequences of non-orthogonality are discussed in Chapter 11. The concept of **balance** refers to information on treatment differences. In the simplest case of an unstructured set of experiment units, a design is balanced if the precision of all treatment comparisons is equal. For a structured set of units, a design is balanced if the precision of all treatment comparisons is equal within each stratum. Most of the designs that we consider in this book are balanced, and the complications introduced by unbalanced designs are discussed within Chapters 11 and 16.

3.3 Common Forms of Design for Experiments

There are many types of statistical design for experiments, which differ from one another in their complexity and in their statistical properties. In the following sections, we describe briefly some common designs illustrated using a simple treatment structure (more complex treatment structures are considered in Chapter 8).

3.3.1 The Completely Randomized Design

The **completely randomized design** (CRD) is the simplest form of design and is appropriate if the experimental units are unstructured and homogeneous, so that there is no need for any form of blocking. The random allocation of treatments to experimental units is not constrained in any way, so that each treatment is equally likely to be allocated to each unit. This is the only case in which we omit the structural component from our model, as this comprises only a single factor which indexes each observation.

EXAMPLE 3.4: CALCIUM POT TRIAL*

An experiment was devised to investigate the effect of differences in soil calcium on the root growth of plants. The experimental material consisted of 20 pots, each containing one plant, arranged in a grid with four rows and five columns, under uniform controlled conditions. The treatments comprised four relative concentrations of calcium (A = 1, B = 5, C = 10, D = 20). Each treatment was applied to five pots selected at random to give a CRD. The layout for this design is shown in Table 3.2.

The main advantages of this design are that it is easy to set up and has a simple form of analysis. It is also flexible, as the statistical analysis is still simple if the replication varies between treatments or if data are missing for some units. The CRD also provides maximal information on the background unit-to-unit variation, as none of the between-unit information is used to assess blocking. However, this is also a weakness of the design if heterogeneity among units is present, as this heterogeneity will inflate the background variation and decrease the precision of estimates of treatment differences. The statistical analysis for this design is presented in Chapter 4.

TABLE 3.2

Randomization for the Calcium Pot Trial, with Pot Numbers (1–20), as a CRD with Four Treatments Labelled A–D, Each with Five Replicates (Example 3.4)

1	2	3	4	5
D	A	B	C	D
6	7	8	9	10
A	D	A	C	B
11	12	13	14	15
A	C	A	D	C
16	17	18	19	20
B	D	C	B	B

3.3.2 The Randomized Complete Block Design

The **randomized complete block design** (RCBD) is the simplest design that includes blocking and is probably the most frequently used design. In this design, the number of experimental units in each block must be the same as the number of treatments. Within each block, treatments are then randomly assigned to experimental units with a different randomization for each block. The design is called **complete** because all treatments occur within each block. If we use factors Block to label the blocks, and Unit to label the units within each block, then this is a nested structure with two strata, represented using our symbolic notation as

Structural component: Block/Unit

or written in expanded form as

Structural component: Block + Block.Unit

EXAMPLE 3.5: POTATO YIELDS*

An experiment was devised to investigate the effects of four different types of fungicides (labelled F1, F2, F3, F4) on the yield of potatoes in field plots. An untreated control treatment (labelled Control) was also included to give a baseline comparison. In the field designated for the trial, heterogeneity was thought to be present at large scales, but suitable blocks of five field plots could be identified and so a RCBD with four blocks each of five plots could be used. The randomized layout is shown in Table 3.3. The structural component is written as

Structural component: Block/Plot

The RCBD is popular (and useful) because it includes some blocking to deal with heterogeneity between experimental units, while still being straightforward to manage and with a simple statistical analysis. Because each treatment occurs once in each block, this design is both orthogonal (treatments are orthogonal to blocks) and balanced (all treatment comparisons are made with equal precision). Details of statistical analysis for this design are presented in Chapter 7. A weakness of the design is that the block size must be equal to the number of treatments, and so the RCBD may be inefficient if the natural block size, as determined by the experimental material, is smaller than the number of treatments. The RCBD will also be inefficient if two independent sources of background heterogeneity are present. We introduce appropriate designs for these situations (the balanced incomplete block design and the Latin square design, respectively) below.

TABLE 3.3

Randomization for the Potato Yields Trial as a RCBD with Five Treatments in Four Blocks of Five Plots (Example 3.5)

	Plot 1	**Plot 2**	**Plot 3**	**Plot 4**	**Plot 5**
Block 1	F3	Control	F2	F1	F4
Block 2	F2	Control	F3	F4	F1
Block 3	Control	F2	F3	F4	F1
Block 4	F3	F2	F1	Control	F4

3.3.3 The Latin Square Design

The **Latin square** (LS) **design** is useful where patterns of heterogeneity are associated with two crossed structural factors with the same numbers of levels. Because this design was originally used for square layouts in field trials, the structural factors are often called Row and Column, corresponding to the spatial arrangement of the rows and columns of the layout, respectively. However, these factors often correspond to non-spatial factors, such as time of day and observer. Using our symbolic notation, we write the crossed structure as

Structural component: Row*Column

or written in expanded form as

Structural component: Row + Column + Row.Column

This structure has three strata. The number of treatments must be equal to the numbers of rows and columns, and the treatment allocation is such that each treatment appears exactly once in each row and once in each column (see Figure 3.6a).

Construction of a Latin square is more complex than for the RCBD, as the three-way inter-relationship between rows, columns and treatments must be preserved. Tables of standard Latin squares have been published for small numbers of treatments (e.g. see Cochran and Cox, 1957, or Fisher and Yates, 1963), but statistical software can be used to obtain Latin squares for any number of treatments. To generate a randomization, one standard square of the right size is first selected at random. The order of the columns is then randomized, followed by the order of the rows (as illustrated in Figure 3.6). This randomization preserves the structure of the design while giving a very large number of possible squares, and thus avoiding bias.

EXAMPLE 3.6: LUPIN TRIAL

An experiment was devised to investigate the effects of soil type and water availability on the growth of lupins. The experiment was to be done with pots on a bench in a glasshouse, with a systematic trend running along the bench (left–right) as a result of a temperature gradient, and across the bench (up–down) because of differing light levels. The rows and columns within the array of pots were therefore considered as blocking factors with a crossed structure, and a LS design is appropriate. Four treatments, labelled CL, CH, SL and SH, representing different combinations of soil type (clay or

(a)

	C1	C2	C3	C4
R1	A	B	C	D
R2	B	C	D	A
R3	C	D	A	B
R4	D	A	B	C

(b)

	C4	C2	C1	C3
R1	D	B	A	C
R2	A	C	B	D
R3	B	D	C	A
R4	C	A	D	B

(c)

	C4	C2	C1	C3
R2	A	C	B	D
R3	B	D	C	A
R4	C	A	D	B
R1	D	B	A	C

FIGURE 3.6
Randomization of a LS design. Rows, columns and treatments are labelled R1–R4, C1–C4 and A–D, respectively. (a) Start with a standard LS design, then (b) randomize the order of the columns and (c) finally randomize the order of the rows.

TABLE 3.4

Randomization for the Lupin Trial as a Latin Square Design with Four Treatments Labelled CH, CL, SH and SL (Example 3.6)

	Column 1	**Column 2**	**Column 3**	**Column 4**
Row 1	CH	SL	CL	SH
Row 2	CL	SH	CH	SL
Row 3	SH	CH	SL	CL
Row 4	SL	CL	SH	CH

sand) and the amount of water supplied (low or high) were used. A randomized layout for a LS design for this experiment is shown in Table 3.4. It is straightforward to verify that each treatment can be found once in each row and once in each column and that each row or column contains all four treatments.

The main disadvantage of the LS design is the restriction that the number of rows, columns and treatments must all be equal. This is discussed further in Section 9.1, where some extensions of the LS design are also described.

3.3.4 The Split-Plot Design

The **split-plot** (SP) **design** has a nested structure, and is used in the case where (at least) two treatment factors are present, with the levels of one treatment factor having to be applied to large experimental units while the levels of another treatment factor can be applied to smaller units. Here we consider a standard form of the SP design with two treatment factors, A and B, with a crossed structure, and a nested structural component with three strata. The highest level of structure corresponds to complete replicates of the set of treatments, and we denote this level using the factor Block. Each block is then divided into a number of whole plots (factor WPlot), with levels of treatment factor A randomized to the whole plots separately within each block. Finally, each whole plot is divided into a number of subplots (factor Subplot), and the levels of factor B are randomized onto subplots within each whole plot. This design can be represented in symbolic form as

Explanatory component: A*B
Structural component: Block/WPlot/Subplot

EXAMPLE 3.7: WEED COMPETITION EXPERIMENT

A field trial was set up to study the competitive effects of three different weed species in winter wheat under different levels of water stress. Variation in water stress was provided by the presence or absence of irrigation, which could be applied only to large areas of land whereas the weed species could be applied to small plots. A SP design was therefore deemed suitable, with the two irrigation treatments (factor Irrigation, with two levels) applied to whole plots (factor WholePlot, with two levels). Each whole plot was split into four subplots (factor Subplot, with four levels), and a pre-determined population of each weed species (*Alopecurus myosuroides* (black-grass), *Galium aparine* (cleavers) and *Stellaria media* (chickweed), abbreviated as Am, Ga and Sm, respectively) was sown in one of these four subplots. The remaining subplot within each whole plot had no weed seeds added, and it was used as a control. Factor Species was used to label the

TABLE 3.5

Randomization for the Weed Competition Experiment as a Split-Plot Design with Two Whole-Plot Treatments (Irrigated, Highlighted in Grey, and Non-Irrigated) and Four Subplot Treatments (Weed Species Am, Sm, Ga or Control, –) (Example 3.7)

	Block 1		Block 2		Block 3		Block 4	
Whole plot 1	1 –	2 Am	1 –	2 Ga	1 Sm	2 Ga	1 Am	2 –
	3 Sm	4 Ga	3 Sm	4 Am	3 –	4 Am	3 Ga	4 Sm
Whole plot 2	1 –	2 Sm	1 Am	2 Ga	1 Am	2 Sm	1 Ga	2 Sm
	3 Ga	4 Am	3 –	4 Sm	3 Ga	4 –	3 Am	4 –

four weed treatments, i.e. the three added populations and control. This structure was repeated another three times, giving four blocks (factor Block, with four levels), with a different randomization in each block, as shown in Table 3.5. The model for this design can be written in symbolic form as

Explanatory component: Irrigation*Species
Structural component: Block/WholePlot/SubPlot

The statistical analysis for, drawbacks of and variations on this design are discussed in Section 9.2.

3.3.5 The Balanced Incomplete Block Design

The **balanced incomplete block design** (BIBD) can be useful when there is only one blocking factor but the number of units per block is smaller than the number of treatments. In this case, each block can contain only a subset of the treatments, and designs with this property are known as **incomplete block designs**. A BIBD has the additional characteristic of balance, which requires that all treatment comparisons have equal precision. This is achieved if the treatments have equal replication and each pair of treatments occurs together within a block exactly the same number of times over the whole experiment. If we again use factor Block to label the blocks, and factor Unit to label the units within blocks, then this design has the same nested blocking structure as the RCBD, represented as

Structural component: Block/Unit

Construction of a BIBD is more complex than for the RCBD, as the balanced inter-relationship between blocks and treatments must be preserved. Tables of standard BIBDs have been published (e.g. see Cochran and Cox, 1957, or Fisher and Yates, 1963) and can be used to generate a BIBD. These designs are also available in many statistical packages. The first step in construction is to choose a standard design with the right block size and number of treatments. The standard layout is then randomized first by randomization of the order of the blocks, and then randomization of the order of the treatments present within each block.

TABLE 3.6

Randomization for Grain Protein Content Experiment as a BIBD with Six Treatments Labelled A–F in Six Sessions (Blocks), Each Containing Five Samples (Units) (Example 3.8)

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Session 1	C	F	E	B	A
Session 2	C	E	F	D	A
Session 3	E	D	F	C	B
Session 4	A	C	B	E	D
Session 5	D	B	A	C	F
Session 6	E	F	A	D	B

EXAMPLE 3.8: GRAIN PROTEIN CONTENT*

An experiment was devised to evaluate the grain protein content for six different varieties of pea (labelled A, B, C, D, E and F). Five independent samples of grain were available for each variety. Only five samples (factor Sample, five levels) could be assessed within a session (factor Session, six levels), with possible heterogeneity between sessions, so a BIBD with six blocks (corresponding to sessions), each containing five units (corresponding to samples) was used. The structural component was specified as

Structural component: Session/Sample

A randomized plan for this design is shown in Table 3.6. For this design each variety is replicated five times, as each appears in five of the six sessions, and any pair of varieties is present together in four sessions, for example, varieties E and F are both present in sessions 1, 2, 3 and 6.

One drawback of BIBDs is that the range of available designs is fairly limited: it is not always possible to construct a BIBD for a given number of treatments, number of blocks and block size (number of units per block). More details are given in Section 9.3.

3.3.6 Generating a Randomized Design

Once a design has been chosen, and the numbers of treatments and replicates have been defined, then a randomized layout or plan for the experiment can be generated. Most general statistical software (including GenStat, R and SAS) have some facilities for generating standard designs, including most of those considered in this book. For non-standard designs (including some BIBDs), more specialist software, such as CycDesigN (see http:/www.vsni.co.uk/software/cycdesign) must be used.

EXERCISES

3.1 Suppose that you are planning an experiment to investigate the impact of nutrient deprivation on plant metabolites. You have four different nutrient levels to test, obtained by applying appropriate nutrients to four sub-samples from a single bag of base compost. The resulting volume of each nutrient level is sufficient for four seed trays (i.e. 16 seed trays in total), and six plants will be grown in each seed tray. To achieve the required growing conditions, a small CE cabinet will be used. The cabinet has four shelves, and you can fit four seed trays on each shelf in a 2×2

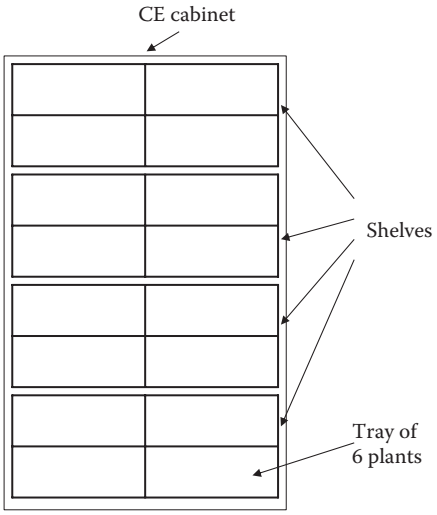

FIGURE 3.7
Structure of a CE cabinet to be used for an experiment to investigate the impact of nutrient deprivation on plant metabolites (Exercise 3.1).

arrangement (Figure 3.7). Although the cabinet is supposed to provide a uniform environment, a technician suggests that light levels may vary between the shelves, and that this might affect plant growth. When they reach the required growth stage, the six plants from each seed tray will be bulked and processed together to form a single sample to be read by a machine. Your colleague tells you that the machine shows some drift over time, but that readings should stay stable across a set of up to six samples.

How might you design this experiment to obtain an unbiased assessment of differences between the four nutrient levels? Consider and discuss the different factors which might affect your choice of design and produce a candidate design. You should consider both stages of the experiment and the following issues:

- What is the experimental unit for the nutrient treatments?
- What are the sources of heterogeneity in the experimental process?
- How might you deal with this heterogeneity?
- How would you allocate the treatments to the experimental units?
- What replication do you have for each treatment?
- What are the advantages/disadvantages of your design?

How would you modify your design if

a. A temperature gradient was discovered between the front and back of the shelves
b. You want to include a CO_2 treatment that can only be applied to a whole CE cabinet and you obtain sufficient resources for 32 trays (eight for each nutrient level)

3.2 Identify the experimental unit, the replication for each treatment and whether pseudo-replication is present in the following experiments.

a. A pot experiment with 12 circular pots in a 2×6 array, in a uniform environment. Each pot contains four plants at the three-leaf stage, and each of four treatments (labelled A–D) were applied at random to one plant per pot as shown in Figure 3.8.

b. A field experiment with 12 homogeneous rectangular plots in a 3×4 grid. Two treatments (labelled A and B) were applied at random to six plots each (Figure 3.9). At harvest, 25 plants are to be sampled per plot, and the plants from each plot will be processed as a single batch for measurement.

c. The field experiment described in part (b) (Figure 3.9) with the height of 25 individual plants per plot measured and recorded *in situ* at 4-weekly intervals from tillering until harvest.

3.3 Four replicates of each of four treatments, labelled A–D, are to be applied at random to batches of aphids in 16 Petri dishes laid out in a 2×8 array (Figure 3.10). The environment is thought to be homogeneous. Use a pack of playing cards to determine an appropriate randomization for this experiment.

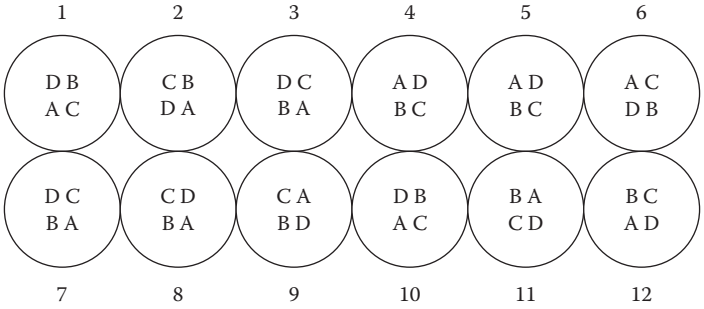

FIGURE 3.8
Experimental layout for a pot experiment with 12 pots and four treatments, labelled A–D, applied to plants within pots. Letters denote the positions of the plants and the treatment applied (Exercise 3.2a).

B	A	B	A
A	B	A	A
B	A	B	B

FIGURE 3.9
Experimental layout for a field experiment with plots in a 3×4 grid, showing the allocation of treatments (A or B) to plots (Exercises 3.2b and c).

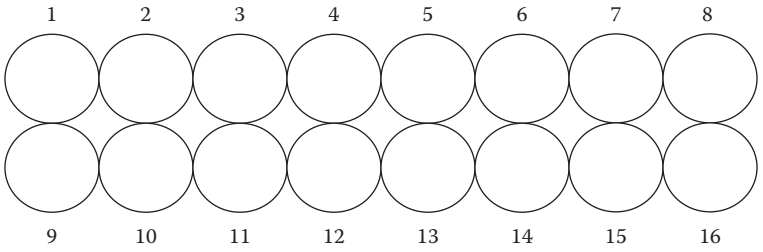

FIGURE 3.10
Layout of 16 numbered Petri dishes for an aphid experiment using four replicates of four treatments (Exercise 3.3).

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	20
21	22	23	24

East

FIGURE 3.11
Layout (with numbered plots) for an experiment testing six treatments in a field with a pH gradient running from west to east (Exercise 3.4).

3.4 Four novel herbicides (labelled A–D) are to be compared with a commercial product (labelled P) and a hand-weeded control (labelled H) in a field trial, giving six treatments in total. The field available can accommodate 24 plots in an array of four columns running north to south, by six rows running west to east (Figure 3.11). The field has a known pH gradient running west to east, i.e. along rows, which may affect crop growth. Produce a RCBD which accounts for this gradient with a randomized allocation of treatments to plots using (a) a pack of playing cards, and (b) a standard six-sided die.

3.5 The efficacy of six synthetic insect pheromones is to be tested in the field. Traps are baited with a single pheromone, deployed at dusk, left out overnight, then retrieved the next morning and the insect catches recorded. There is sufficient material to bait six traps with each pheromone.

a. Consider how you might use a RCBD for this experiment if only six traps are available at any one time and all six traps will be placed in the same field, but a different field will be used each night. Are any structural factors confounded? What are the assumptions of this design? Write down the structural component for this design.

b. How would you change your design if the same trap locations were to be used each night and the positions could not be considered homogeneous? Write down the structural component for this design.

c. How might you modify your design if 18 traps are available at any one time? Under which conditions would designs based on CRD, RCBD or LS arrangements be preferable?

d. What design might you use if only four pheromones are to be tested, with six traps available at any one time, and enough material for nine replicates of each pheromone?

3.6 The effect of temperature on the transmission of a virus by five aphid species is to be investigated. Three small growth chambers are available and three temperatures will be tested. The temperature for each chamber can be set and then applies to the whole chamber, and each chamber can hold five plants in individual pots. One aphid will be placed onto each plant using a clip cage. Forty-five plants and 15 aphids of each species are available. Assuming that chambers (and positions within chambers) can be considered homogeneous, suggest a design to test the effects of temperature and aphid species. What are the experimental units for each factor? Produce a randomized design for this experiment and write down the explanatory and structural components for the design. If you suspected that there were systematic differences between chambers, how would you modify your design? Write down the structural component for this new design.

3.7 A field experiment was set up to investigate how invertebrate abundance is affected by the spatial structure and species composition of weed patches (Smith, 2007). Small weed patches were formed from three pots of plants in a tray. Species composition was varied by using different numbers of mayweed (M) or thistle (T) plants in the patch, i.e. 3M, 2M + 1T, 1M + 2T or 3T. Spatial structure was varied by changing the distance between patches (12 or 6 m). Five blocks of two whole plots were set up, with the two spacings allocated at random to whole plots within blocks. Each whole plot contained 16 patches laid out in a 4×4 array with the designated spacing, with patches allocated to four replicates of each of the four species compositions according to a LS design. A different randomization was used within each whole plot. Write down the explanatory and structural components for this design.

3.8 A glasshouse experiment to compare the effect of two nutrition regimes on the growth of three wheat varieties was set up as a RCBD with 12 blocks of six pots each, as shown in Figure 3.12. The treatments comprise the six combinations of nutrition regime (labelled N1, N2) and variety (labelled V1–V3). The blocks accommodate an expected temperature gradient running from the door to the

	1	2	3	4	5	6 N2 V2	7	8	9	10	11	12
	13	14	15	16	17	18 N1 V2	19	20	21	22	23	24
	25	26	27	28	29	30 N2 V3	31	32	33	34	35	36
Far end (warmer)	37	38	39	40	41	42 N1 V1	43	44	45	46	47	48
	49	50	51	52	53	54 N1 V3	55	56	57	58	59	60
	61	62	63	64	65	66 N2 V1	67	68	69	70	71	72

Door (cooler)

FIGURE 3.12
Layout of pots (labelled 1–72) as a RCBD in a greenhouse experiment to compare the effects of two nutrition regimes (N1 and N2) on the growth of three wheat varieties (V1, V2 and V3). Blocks (columns) contain six pots. One block shows treatment labels in addition to pot numbers (Exercise 3.8).

far end of the glasshouse. Several characteristics of each plant, including height and number of leaves, are to be recorded every week. Suggest acceptable protocols for recording data if

a. You are the only person available to take the measurements

b. There are two people available to take the measurements

Which protocols would be unacceptable and why?