Chapter 1

Introduction

1.1 Prologue

The modern foundations of design of experiments were laid by R. A. Fisher during the early part of the 20th century. Since then, this area has seen a phenomenal growth. Design of experiments has for long been an integral part of almost all scientific investigations and continues to be so. It has therefore played a fundamental role in statistical practice and research. Statistical training also has always emphasized the role of design of experiments in extracting correct information and making valid inference on the underlying problem and thus, design of experiments is an essential component of most statistics curricula.

While designing an experiment, the principles of randomization, replication and local control are of vital importance. These principles were first enunciated by Fisher while planning agricultural experiments. It was observed by Fisher that a completely random allocation of treatments to the experimental units, leading to a completely randomized design, eliminates bias in assessing treatment differences.

In certain experimental situations, there may be systematic variations present among the experimental units. For example, in a field experiment, the experimental units are typically plots of land. In such an experiment, there may be a fertility gradient present such that plots on the same fertility level are more homogeneous than those which are at different fertility levels. In experiments with piglets as experimental units, it is very plausible that piglets belonging to the same litter are genetically closer to each other (being born to the same pair of parents) than those belonging to different litters. Similarly, in experiments with livestock, different breeds (or, different ages) might be involved and animals belonging to the same breed are expected to be more alike than the ones belonging to different breeds. In the context of clinical trials with patients forming the experimental units, the trial may be conducted at different centers (mainly to get enough number of observations) and patients from the same center may be more alike than those from different centers due to differences in treatment practices and/or management procedures followed at different centers.

The above examples, which are merely illustrative and by no means exhaustive, demonstrate that in many situations there is a systematic variation among the experimental units. In such situations, use of a completely randomized design is not appropriate. Rather, one should take advantage of the *a priori* information about this systematic variation while designing the experiment in the sense that this information should be used while designing to eliminate the effect of such variability. The impact of this effort will be reflected in a reduced error, thereby increasing the sensitivity of the experiment. The above considerations led to the notion of local control or *blocking*. The groups of relatively homogeneous experimental units are called *blocks*. When the blocking is done according to one attribute, we get a *block design*. In a block design, the treatments are applied randomly to the experimental units within a block, the randomized allocation of treatments to experimental units within a block being done independently in each block.

The simplest among the block designs is the randomized complete block design. In such a design, each block is required to have as many experimental units as the number of treatments, i.e., the block size is equal to the number of treatments. However, it is not always possible to adopt a randomized complete block design in every experimental situation. Firstly, if one assumes that the intra-block variance is directly dependent on the block size, then adoption of a design with blocks of small sizes is preferable over one which has large block sizes. This restricts the use of randomized complete block designs in situations where the number of treatments is large. For example, in agronomic experiments, the experimenter generally chooses a block of size 10-12 and if this is accepted, then one cannot adopt a randomized complete block design in situations where say 20 treatments are to be compared. Furthermore, in many experimental situations, the block size is determined by the nature of the experiment. For example, with some experiments in psychology, it is quite common to consider the two members of a twin pair as experimental units of a block. In that case, clearly a randomized complete block design cannot be prescribed if the number of treatments is larger than two. Similarly, it is reasonable to take litter-mates (of say mice) as units of a block and litter size may not be adequate to accommodate all the treatments under test.

The few examples considered above clearly show that in many situations, one cannot adopt a randomized complete block design and thus, there is a need to look for designs where not all treatments appear in each block. Such designs are termed as *incomplete block designs*. The present book deals with block designs in general and their analysis, with special emphasis on certain important classes of incomplete block designs. The terms block design and incomplete block design are used interchangeably whenever there is no scope for confusion.

A reasonable amount of familiarity with basic notions of vector spaces and the algebra of matrices is assumed throughout and one may refer e.g., to Bapat (2000) for details on these aspects. We also assume a background of linear statistical models and of the general area of design of experiments at an advanced undergraduate level. Excellent accounts of the general area of design of experiments and its applications are available e.g., in Cox (1958), Hinkelmann and Kempthorne (1994), Dean and Voss (1999), Wu and Hamada (2000) and Bailey (2008).

1.2 Outline of the Book

The book has five more chapters followed by an appendix. In Chapter 2, the discussion is initiated by describing the intra-block analysis of an arbitrary block design. Balancing in incomplete block designs is considered next in Section 2.3 of this chapter. The two notions of balance, viz., variance- and efficiency-balance are reviewed. The analysis of incomplete block designs with recovery of inter-block information is discussed in Section 2.4. Finally, in Section 2.5, the notion of efficiency factor of an incomplete block design is briefly studied.

Balanced designs are considered in Chapter 3. The most important of the balanced designs are the classical balanced incomplete block (BIB) designs. Such designs are still found useful in designing experiments in diverse fields and newer applications of these designs, e.g., in visual cryptography, have been found in recent years (see e.g., Bose and Mukerjee (2006), Adhikary, Bose, Kumar and Roy (2007) and the references cited therein). We initiate the discussion in this chapter by considering some properties of BIB designs in Section 3.2. The analysis of BIB designs is briefly considered in Section 3.3. Some results on construction and existence of BIB designs are presented in Section 3.4. Generalizations of BIB designs are considered in the next section. The BIB designs are the only designs in the class of binary, equireplicate and proper designs that are both variance- and efficiency-balanced; however, it is possible to find other variance- and efficiency-balanced designs if one expands the class of designs to non-binary, non-equireplicate or non-proper designs. The construction methods of variance- and efficiency-balanced designs with possibly unequal replications and unequal block sizes are briefly reviewed in Section 3.6. Properties and construction of nested BIB designs are discussed briefly in Section 3.7.

Partially balanced designs are the subject matter of Chapter 4. Among the partially balanced designs, the partially balanced incomplete block (PBIB) designs are the most studied ones and continue to be used in actual applications. These are therefore covered at some length in Sections 4.2–4.6. PBIB designs are formally introduced in Section 4.2 via the notion of an association scheme. The algebra of association matrices is briefly discussed in Section 4.3. Designs with two or more associate classes as also the analysis of PBIB designs are discussed in Sections 4.4–4.6. In Sections 4.7–4.11, some other partially balanced designs which are not necessarily PBIB designs are covered. These include lattice, cyclic, linked block, C designs and α designs.

In Chapters 3 and 4, incomplete block designs are studied for situations where all the treatments are on equal footing and thus, the interest is mainly on elementary treatment contrasts or, more generally, on a complete set of orthonormal treatment contrasts. However, in practice there are situations where the interest lies in inference on contrasts of special types. Such situations arise typically, e.g., in factorial experiments and biological assays. Incomplete block designs for such experiments are considered in Chapter 5. Specifically, incomplete block designs for factorial experiments (Section 5.2), biological assays (Section 5.3), test-control experiments (Section 5.4) and diallel cross experiments (Section 5.5) are covered. Finally, in Section 5.6, results on incomplete block designs that are robust against an outlier and against missing data are reviewed. Some aspects of trend-free block designs are also covered in this section.

In Chapter 6, optimality aspects of some incomplete block designs are discussed. Different optimality criteria are introduced in Section 6.2. Important results on optimality of proper incomplete block designs for inference on a complete set of orthonormal treatment contrasts are reviewed in Section 6.3. Optimal designs for making inferences on contrasts among several test treatments and a control are discussed in Section 6.4. Optimality of designs for parallel line assays, considered in Chapter 5 are reviewed in Section 6.5. In the last section (Section 6.6), optimal incomplete block designs for diallel crosses are considered.

The Appendix consists of four sections. Some results in linear algebra that are used throughout the book are given in Section A.1. In Section A.2, some basic results in linear statistical models are summarized. Section A.3 describes some essential facts about finite (Galois) fields. In Section A.4, basic ideas and results from finite projective and Euclidean geometries are reviewed.