PART I

OVERTURE

1

Introduction

1.1 Why a statistical theory of design?

The need to develop statistical theory for designing experiments stems, like the need for statistical analysis of numerical information, from the inherent variability of experimental results. In the physical sciences, this variability is frequently small and, when thinking of experiments at school in physics and chemistry, it is usual to think of 'the correct result' from an experiment. However, practical experience of such experiments makes it obvious that the results are, to a limited extent, variable, this variation arising as much from the complexities of the measurement procedure as from the inherent variability of experimental material. As the complexity of the experiment increases, and the differences of interest become relatively smaller, then the precision of the experiment becomes more important. An important area of experimentation within the physical sciences, where precision of results and hence the statistical design of experiments are important, is the control of industrial chemical processes.

Whereas the physical sciences are thought of as exact, it is quite obvious that biological sciences are not. Most experiments on plants or animals use many plants or animals because it is clear that the variation between two plants, or between two animals, is very large. It is impossible, for example, to predict quantitatively the characteristics of one plant from the corresponding characteristics of another plant of the same species, age and origin.

Thus, no medical research worker would make confident claims for the efficacy of a new drug merely because a single patient responded well to the drug. In the field of market research, no newspaper would publish an opinion poll based on interviews with only two people, but would require a sample of at least 500, together with information about the method of selection of the sample. In a drug trial, the sample of patients would normally be less than 500, possibly between 20 and 100. In psychological experiments, the number of subjects used might be only eight to 12. In agricultural experiments, there may be 20 to 100 plots of land, each with a crop grown on it. In a laboratory experiment hundreds of plants may be treated and examined individually. Or just six cows may be examined while undergoing various diets, with measurements taken frequently and in great detail.

The size of an experiment will vary according to the type of experimental method and the objective of the experiment. One of the important statistical ideas of experimental design is the choice of the size of an experiment. Another is the control of the use of experimental material. It is of little value to use large numbers of patients in the comparison of two drugs, if all the patients given one drug are male, aged between 20 and 30, and all the patients given the other drug are female, aged 50 to 65. Any reasonably sceptical person would doubt claims made about the relative merits of the two drugs from such a trial. This example may seem trivially obvious, but the scientific literature in medicine and many other disciplines shows that many examples of badly planned (or unplanned) experiments occur.

And this is just the beginning of statistical design theory. From avoiding foolish experiment, we can go on to plan improvements in precision for experiments. We can consider the choice of experiments as part of research strategy and can, for example, discuss the relative merits of many small experiments or a few large experiments. We can consider how to design experiments when our experimental material is generally heterogeneous, but includes groups of similar experimental units. Thus, if we are considering the effects of applying different chemicals on the properties of different geological materials, then these may be influenced by the environment from which they are taken, as well as by the chemical treatment applied. However, we may have only two or three samples from some environments, but as many as ten samples from other environments; how then do we decide which chemicals to apply to different samples so that we can compare six different chemical treatments?

1.2 History, computers and mathematics

If we consider the history of experimental design, then most of the developments have been in biological disciplines, in particular in agriculture, and also in medicine and psychology. There is therefore an inevitable agricultural bias to any discussion of experimental design. Most of the important principles of experimental design were developed in the 1920s and 1930s by R. A. Fisher. The practical manifestation of these principles was very much influenced by the calculating capacity then available. Had the computational facilities which we now enjoy been available when the main theory of experimental design was being developed then, I believe, the whole subject of design would have developed very differently. Whether or not this belief is valid, it is certainly true that a view of experimental design today must differ from that of 50, or even 30, years ago. The principles have not changed, but the principles are often forgotten, and only the practical manifestation of the principles retained; these practical applications do require rethinking.

The influence of the computer is one stimulus to reassessing experimental design. Another cause for concern in the development of experimental design is the tendency for increasingly formal mathematical ideas to supplant the statistical ideas. Thus the fact that a particularly elegant piece of mathematics can be used to demonstrate the existence of groups of designs, allocating treatments to blocks of units in a particular way, begs the statistical question of whether such designs would ever be practically useful.

Although myself originally a mathematician, I believe that the presentation of statistical design theory has been quite unnecessarily mathematical, and I shall hope to demonstrate the important ideas of statistical design without excessive mathematical encumbrance. The language of statistical theory, like that of physics, is mathematical and there will be sections of the book where those with a mathematical education beyond school level will find a use for their mathematical expertise. However, even in these sections, which I believe should be included because they will improve the understanding of statistical theory of the readers able to appreciate the mathematical demonstrations, there are intuitive explanations of the theory at a less advanced mathematical level.

1.3 The influence of analysis on design

To write a book solely about the theory of experimental design, excluding all mention of the analysis of data, would be impossible. Any experimenter must know how he intends to analyse his experimental data before he designs his experiment to yield the data. If not, how can he know whether the form of information which he collects can be used to answer the questions which prompted him to do an experiment?

Thus, consider again the medical trial to compare two drugs. Suppose the experimenter failed to think about the analysis and argued that one of the drugs was well known, while the other was not; in the controlled experiment to compare them, there are available 40 patients. Since a lot is already known about drug A, let it be given to one patient, and let drug B be given to the other 39. When the data on the response to the drugs is obtained, the natural analysis is to compare the mean responses to the drugs, and to consider the difference $(\bar{x}_A - \bar{x}_B)$. To test the strength of evidence for a real difference in the effects of the two drugs, we need the standard error of $(\bar{x}_A - \bar{x}_B)$, which will be:

$$\sigma(1/1+1/39)^{1/2} = 1.013\sigma.$$

However, if the experimenter had considered this analysis of data before designing the experiment, he would have realised that the effectiveness of his experiment depended on making the standard error of $(\bar{x}_A - \bar{x}_B)$ within his experiment as small as possible (to use the previous knowledge about the effects of drug A requires that the experimental conditions are identical to those in which the previous experience was gained). This is achieved by allocating 20 patients to drug A and 20 to drug B. This gives a standard error of

$$\sigma(1/20+1/20)^{1/2} = 0.316\sigma,$$

giving an improved precision by a factor of more than three. Of course, it would not be necessary to consider the standard error formally to guess that equal allocation of patients gives the most precise answer. In a sense, it would be intuitively surprising if any unequal division were to be more efficient than the equal division. Prior thought of what is to be done with the results of the experiment should lead to avoiding a foolish design.

Nevertheless, although analysis is integral to any consideration of design, this book is about the design of experiments, and will not be concerned with the analysis of data, except insofar as this is essential to the understanding of the design principles discussed. The general theory of the analysis of data from designed experiments is developed in Chapter 4, and particular examples of this general theory will appear in later chapters for some examples of experimental designs. In Chapter 5, the developments in computing which allow the analysis by a general computer program of any designed experiment are described. The joint implication of these two chapters is clear; not only is it possible to derive the algebraic formulae necessary to define the calculations required for the analysis of data from any experiment, but the computer programs should make it possible for an experimenter, without the mathematical skills necessary to derive the algebraic formulae, to understand the statistical basis for the interpretation of the calculations. The need, in the earlier development of statistical methods, to be able to derive the algebraic form of analysis has made it easier to discuss some of the important design principles we shall discuss later. In this way, the earlier lack of advanced computing power may be regarded as a blessing, though there must be many erstwhile technical assistants to statisticians who would find that hard to accept! A caveat to the advice to use the general computer programs to analyse all experimental data must be to remember that much of our understanding of design theory was stimulated by algebraic necessity because of the lack of computers. Consequently, we should search for increased understanding both through the use of computers and through algebraic manipulation.

1.4 Separate consideration of units and treatments

Throughout this book, I shall emphasise the need to think separately about the properties of the experimental units available for the experiment, and about the choice of experimental treatments. Far too often, the two aspects are confused at an early stage, and an inefficient or useless experiment results. Thus, an experimenter considering a comparison of the effects of different diets on the growth of rats may observe that the rats available for experimentation come from litters with between five and ten animals per litter. Having heard about the randomised block design, he decides that he must use litters as blocks, and therefore will have blocks of five units. He further decides that he should consequently have five treatments, and chooses his treatments on this premise, rather than considering how many treatments are required for the objectives of his experiments.

Similarly, a microbiologist investigating the growth of salmonella as affected by five different temperatures, four different media and two independent different chemical additives, may decide that a factorial set of 80 treatments is necessary. Recognising the possibility of day-to-day variation and also knowing about the randomised block design he tries to squeeze the preparation of 80 units into a single day, where only 40 can be efficiently managed, and consequently suffers a much higher error variance for his observations than would normally be expected.

In both of these examples, which though plausible are fictitious so far as my personal experience extends, I believe that separate consideration of properties of units and choice of treatments, together with an appreciation of the possibilities of modern statistical design, would lead to better experiments. Thus, in the rat growth experiment, on considering the experimental material available, the experimenter might recognise that each litter forms a natural block, and that he has one block of five units, one of six units, two of seven units, and one of ten units; each block could be reduced in size while still retaining its natural block quality. For treatments, he has a control and three additions to the basic diet, each of which he wishes to try at two concentrations; a total of seven treatments. I believe that it is quite simple to construct a sensible design to compare seven treatments in blocks of five, six, seven, seven and ten units, respectively, and this problem will be discussed further in Chapter 7.

For the microbiologist, again the solution of his problem requires only the application of standard statistical design theory, though because the problem does not fit into any neat mathematical classification, the reader of most books on experimental design might be forgiven for thinking that the problem was impossible or at best required very complex mathematical theory. The problem of using blocks of 40 units with a set of 80 treatments in a $5 \times 4 \times 2 \times 2$ factorial structure is a very simple example of confounding to be met again in Chapter 15.

In summary, to achieve good experimental design, the experimenter should think first about the experimental units, and should make sure that he understands the likely patterns of variation. He should also consider the set of treatments appropriate to the objectives of his experiment. If the set of treatments fits simply with the structure of units using a standard design, then the experiment is already designed. If not, then a design should be constructed to fit the requirements of units and treatments, either by the experimenter alone or with the assistance of a statistician. The natural pattern of units and treatments should not be deformed in a Procrustean fashion to fit a standard design. It may perhaps reassure the reader to point out that, for most experiments (80 or 90%), a standard design will be appropriate. To counterbalance the note of reassurance, it is obvious from the experimental scientific literature that frequently an inappropriate simple standard design has been used when an appropriate but slightly more complex standard design exists and should have been used.

2

Elementary ideas of blocking: the randomised block design

2.1 Controlling variation in experimental units

In an experiment to compare different treatments, each treatment must be applied to several different units. This is because the response from different units varies, even if the units are treated identically. If each treatment is applied to only a single unit, the results will be ambiguous, in that we will not be able to distinguish whether a difference between the response from two units is caused by the different treatments, or is simply due to the inherent differences between the units.

The simplest experimental design is that in which, in order to compare t treatments, treatment 1 is applied to n_1 units, treatment 2 to n_2 units, and treatment t to n_t units. In many experiments the numbers of units per treatment, $n_1, n_2, \ldots, n_t$, will be equal, but this is not necessary, or even always desirable. Some treatments may be of greater importance than others, in which case more information will be needed about them, and this will be achieved by the use of more units for these treatments. This design, in which the only recognisable difference between units is the treatments which are applied to those units, is called a *completely randomised design.*

With several units for each treatment the ambiguity which occurs when each treatment has only a single unit is not eliminated. One treatment might be 'lucky' in the selection of units to which it is to be applied. There is no foolproof method of overcoming the vagaries of allocating treatments to units, since the responses from the units if the same treatment were applied to all units are not known before the experiment. However, most experimenters have some idea about which units are likely to behave similarly, and such ideas can be used to control the allocation of treatments to units in an attempt to make the allocation more 'fair'. Essentially each group of 'similar' units should include roughly equal

Cambridge University Press
978-0-521-28762-3 - The Design of Experiments: Statistical Principles for Practical Applications
R. Mead
Excerpt
More information

numbers of units for each treatment. This control is called *blocking*, and the simplest, and by far the most frequently used, design resulting from the idea of blocking is the randomised block design, sometimes called the *randomised complete block design*.

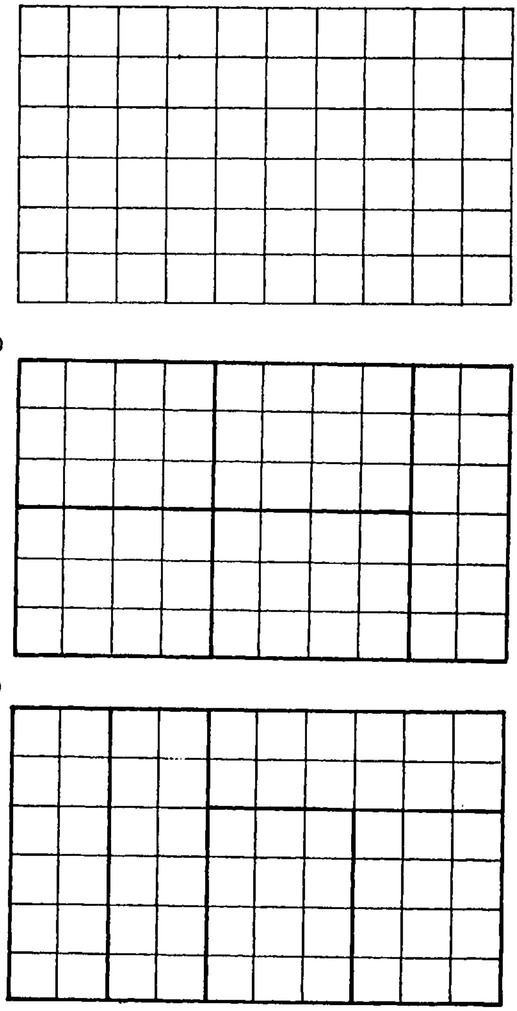
The term blocking arises from the introduction of the idea of control in agricultural crop experiments. Here, the experimental unit is a small area of land, or plot, typically between 2 m² and 20 m², depending on the crop and on the treatments. The set of units or plots for an experiment might be situated as in Figure 2.1(a). In general, we might expect that adjacent plots

Figure 2.1. Possible blocking plans (b) and (c) for a set of 60 plots.

(a)

(b)

(c)



would behave more similarly than those further apart. Of course, the experimenter might have more specific knowledge about the plots than this. For example, there might be a slope from east to west, so that the plots down the east side might be expected to behave similarly, and differently from those on the west side. However, either from general or particular considerations, if the plots are grouped together in sets of, say, 12, so that the plots in a set might be expected to behave similarly, then the resulting grouping will usually look like Figure 2.1(b) or (c). Such groupings are naturally called blocks of plots, and the term has become standard for groupings of units in disciplines other than agronomy.

The patterns of variation on which blocking systems can be based are very numerous, and always the specialist knowledge of the experimenter about his experimental material must be the most important source of information in determining blocks of units. Some examples of blocking systems are discussed here, to help the experimenter to think about the choice of blocks.

In experiments with plants in pots in glasshouses or laboratories, where different treatments are applied to different pots, geographical blocks of similarly situated pots will usually prove useful, the major controllable source of variation between yields from different pots being nearness to the edge of the room or to the door or to other features. In experiments with perennial crops, such as fruit trees, the experimental unit will often be an individual tree, and geographical blocks may again be useful. Possibly more useful blocks can be based on yields of the trees in previous years, or more generally based on the previous histories of the trees. When the experimental unit is an animal, then blocks can be based on genetic similarity, leading to the use of complete litters as blocks, or on the weights of the animals prior to the experiment, or on previous history of the animals.

In medical trials, where the experimental unit is usually an individual person, the number of classifications which can be used as the basis for blocking is large. These include the age of the person, sex, height or weight, social class, medical history, racial characteristics or time at which the person becomes available for the trial. Many of the same classifications are relevant for psychological experiments. Often, in these experiments, the experimental unit will be a short time period in a person's life. If several observations are made for different periods for the same person, with different stress treatments being applied during the different periods, then the experimental unit may be defined to be a short period within a person's life, and the set of units for one person constitutes a sensible block. A possible complication is that time might be thought to be an important