CHAPTER 2

Some Key Assumptions

2.1 INTRODUCTION

In many experiments several types of observation are made on each experimental unit. For example in comparing varieties of sugar beet, yield of roots, yield of tops, yield of sugar, and possibly plant number would be measured, as well as perhaps observations on the incidence of disease, the frequency of bolting, and the chemical analysis of the sugar. In comparing methods of spinning wool yarn, it would be common to measure the yarn irregularity, the yarn strength, and the end-breakage rate in spinning, as well as possibly making tests on fabric woven from the yarns. In a preliminary account it is, however, convenient to suppose that only one observation is made on each experimental unit. This observation may be derived by calculation from a number of experimental readings. For example, measures of yarn irregularity are often obtained by computing a so-called coefficient of variation from a trace showing the changes in thickness along the length of the yarn. Again, in learning experiments in experimental psychology, one observation for analysis is usually a measure of the rate of learning. This is derived from the raw data which consist, for example, of a record of success or failure at each attempt at the experimental task.

The following assumption, or some simple modification of it, underlies the use of most of the designs described in this book. The observation obtained when a particular treatment is applied to a particular experimental unit is assumed to be

$$\begin{pmatrix}\text{a quantity depending}\\ \text{only on the}\\ \text{particular unit}\end{pmatrix} + \begin{pmatrix}\text{a quantity depending}\\ \text{on the treatment}\\ \text{used}\end{pmatrix} \tag{1}$$

and to be unaffected by the particular assignment of treatments to the other units. This can be put more vividly as follows. Denote the alternative treatments by the letters $T_1, \ldots, T_t$; then it is assumed that the observation obtained on any unit when, say, T_1 is applied differs from

the observation that would have been obtained had, say, T_2 been applied by a constant, $a_1 - a_2$. There are constants $a_1, \ldots, a_t$, one for each treatment, and the object of the experiment is to estimate differences such as $a_1 - a_2$; we call such differences the *true treatment effects*.

The essential points about this assumption are that

(*a*) the treatment term in (1) adds on to the unit term rather than, for example, multiplying;

(*b*) the treatment effects are constant;

(*c*) the observation on one unit is unaffected by the treatment applied to other units,

and these three points will be discussed separately in the subsequent sections.

The assumption is particularly important if a full statistical analysis is to be made of the observations. It is, however, still required even if the experiment is analyzed just by calculating simple averages, in the sense that a gross departure from the assumption will affect the whole qualitative interpretation of the results. It is usually possible to check the assumptions to a certain extent from the data, but never possible to avoid completely making some assumption or other. Too much attention should not be paid to the details of the following sections at a first reading.

2.2 ADDITIVITY

The first consequence of the additive law (1) is that the difference between two treatments, say T_1 and T_2, is usually* appropriately estimated by

$$\begin{pmatrix}\text{mean of all observations}\\ \text{on } T_1\end{pmatrix} - \begin{pmatrix}\text{mean of all observations}\\ \text{on } T_2\end{pmatrix}. \tag{2}$$

If the treatment effects and uncontrolled variations are relatively small any functional law for combining unit and treatment terms would be equivalent to the additive law (1) to a first approximation. In other cases, however, it may be worth considering whether some other form may not be more appropriate. The most important alternative form is multiplicative, replacing expression (1) by

$$\begin{pmatrix}\text{a quantity depending}\\ \text{only on the}\\ \text{particular unit}\end{pmatrix} \times \begin{pmatrix}\text{a quantity depending}\\ \text{on the treatment}\\ \text{used}\end{pmatrix}. \tag{3}$$

If this is the appropriate form, we work with the logarithms of the original observations. Since $\log(xy) = \log x + \log y$, equation (3) is thereby converted into form (1).

* The exceptions are incomplete block designs (Chapter 11) and certain types of confounded design (Chapter 12).

Example 2.1. In Example 1.4 we discussed briefly a simple comparative assay for measuring the potency of a drug by comparison with a standard. A natural working hypothesis is that any dose x of the experimental drug is equivalent in all relevant respects to a dose ρx of the standard, where ρ is the potency of the drug with respect to the standard and is constant. Or equivalently the tolerance of an animal (say in mg) for the experimental drug is $1/\rho$ times what it would have been with the standard. This is of the form (3) and is reduced to (1) by working with log tolerances rather than with tolerances.

Example 2.2. Consider a field trial comparing the effect of a number of alternative treatments on the incidence of a certain disease. The treatments are applied one to each plot and after a suitable time the disease is measured, say by counting the number of diseased plants out of one hundred on each plot. It is reasonable to expect that if the proportion diseased varies appreciably over the whole experiment, the difference between proportion diseased for two treatments will be rather greater when the level of disease is fairly high than when it is low. However if the level of disease is very high, it may be that all treatments are ineffective so that differences between treatments decrease again.

At any rate there seems to be no general reason for expecting a constant additive effect for one treatment as compared with another. There are several ways of proceeding. If the experiment is divided into sections within each of which the natural level of disease is fairly constant, it would be reasonable to estimate treatment differences separately for each section. Then by comparing the estimates with the overall level of disease for the section, the change, if any, of treatment effects with level of disease could be assessed. This is probably the best procedure, if it can be used; it amounts to allowing the data to determine the appropriate scale of measurement. Alternatively, if the proportion diseased varied, say from 5 to 50 per cent, it might be reasonable to assume constant proportional differences, and therefore to take logarithms. Or, occasionally, more complicated assumptions might seem justifiable, such as that the treatments have a constant effect on the probit of the proportion diseased. [The probit is a quantity derived by a particular mathematical transformation of a proportion (Goulden, 1952, p. 395).]

Example 2.3. A rather similar example concerns feeding or management trials on pigs. Suppose that two treatments A and B are under comparison and that at the end of the experiment the pigs are examined by a judge and a total score out of 100 assigned to each pig. Then, because of the upper limit to the scale, the following might happen: a pig which would have scored 50 with treatment A would score 70 with treatment B, but a very good pig, which would have scored 85 with treatment A, would score 90 if given B. That is, we are measuring on a scale on which the treatment effects are not additive. A conventional way of attempting to deal with this is to work not with the total score x but with $\log[(x + \frac{1}{2})/(100\frac{1}{2} - x)]$, which should often nullify the restriction at the upper and lower ends of the scale. For two values a certain distance apart and near the top, or bottom, of the scale differ much more after transformation than do two values initially the same distance apart but near the center of the scale.

In all these examples the comparison of treatments by the mean difference (2) is valid in the narrow sense that this will estimate the average treatment difference over the units used in the experiment, i.e., the mean observation that would have been obtained if T_1 had been applied to all units minus the corresponding mean for T_2. But if the assumption (1) does not hold, this difference is rather an artificial quantity. Thus in Example 2.1 the difference in mean tolerances depends on the particular animals, and if these vary appreciably in tolerance from laboratory to laboratory a comparison of mean tolerances would not be independent of laboratories. Moreover the mean difference, even if it was reproducible, would not have the simple physical interpretation of the difference in mean log tolerance, which estimates $\log \rho$.

Again, to take an extreme case, suppose that the experiment in Example 2.2 happened to fall into two roughly equal parts:

(*a*) with an average proportion diseased of 10 per cent, T_1 giving 8 per cent and T_2 12 per cent on the average;

(*b*) with an average proportion diseased of 50 per cent, T_1 giving 40 per cent and T_2 60 per cent on the average.

Then an averaging of the proportions diseased would give a difference between T_2 and T_1 of $36 - 24 = 12$ per cent. But this is clearly an artificial figure that depends on the particular incidence of disease encountered in the experiment; it is in this case much more revealing to say that T_1 gives a proportion $\frac{2}{3}$ of that corresponding to T_2.

Of course this is an extreme and oversimplified example, but it has been discussed to emphasize that the importance of the additive assumption is not essentially connected with details of statistical technique. However it would often happen that, if the experiment falls into sections with different treatment effects, the amount and distribution of the uncontrolled variation would be different in the different sections. A full statistical analysis will involve differential weighting of the sections; this will not be considered here.

Fortunately the complications that we have been considering are frequently unimportant because, as remarked above, if the variations involved are relatively small, the additive law (1), the multiplicative law (3), and other similar laws are nearly equivalent. In many applications it is probably enough to consider which of (3) and (1) is likely to be the more appropriate and to take or not take logarithms accordingly.

2.3 CONSTANCY OF TREATMENT EFFECTS

In the previous section we discussed the assumption that the observations are measured on a scale on which the effect of treatments is represented by the *addition* of appropriate quantities rather than by some other functional law, such as multiplication. In this section we in effect continue that discussion by considering other ways in which the treatment effects can fail to be constant.

First note that an additional completely random component added to the treatment term in formula (1) is indistinguishable from a random component added to the first, or unit, term and so can be disregarded, provided that the distribution of the random component is the same for all treatments. This possibility will not be discussed further. We shall deal in detail with what happens when the treatment effects depend on the value of some supplementary measurement that can be made on each unit.

Example 2.4. Suppose that it is required to compare two alternative processes A and B for extracting a product P from a raw material containing P in small quantities. The experimental units are different batches of raw material and the observation is the yield y of product. A supplementary observation x is also made by obtaining for each batch before processing, an estimate of the percentage of P in the batch. Then it might happen that the difference between the processes depends on the amount of P, e.g., A may work relatively much better when the raw material is rich in P. Information that this was so might not only be important in deciding what practical action to take, but also might throw some light on the fundamental reasons for process differences. Further the information might help to link the results with previous work in which, perhaps, the content of P in the raw material was systematically different.

A comparison of the mean value of y for those units receiving process A with the corresponding mean for process B would, with correct design, always estimate the mean process difference over the raw material used in the experiment. Although this would usually be of some interest, it is clear from the previous paragraph that such an overall difference may be only a partial description of the difference between the processes. Unless there is good prior reason for expecting the process difference to be constant, the data would therefore be analyzed by plotting y against x, distinguishing between the results for the two processes. This graphical analysis would be supplemented, if necessary, by appropriate statistical calculations, such as the fitting of regression lines. Attention would be paid to any change with x in the random variation of y.

Another way of dealing with the results of this experiment would be to work with y/x, which is proportional to the fraction of P in the raw material that is extracted in processing; if the difference between treatments in the ratio were expected to be constant, this would be the natural thing to do. However the general remarks on the constancy of treatment effects would still be relevant.

This example illustrates the use of a supplementary observation to examine whether a treatment difference is constant. A further use of supplementary observations is to increase precision, and this will be considered in detail in Chapter 4.

Example 2.5. Jellinek (1946) has described an experiment to compare three drugs A, B, C for the relief of headaches, with a pharmacologically inactive control D. Each subject used each drug for two weeks and one of the observations was the success rate, i.e., the number of headaches relieved divided by the number of headaches treated in the two-week period. Precautions, which need not be gone into here, were taken to remove any effect of the order in which the drugs were used. The first line of the table shows the mean success rates averaged over all subjects. They suggest that A, B, C are not appreciably different and all have appreciably higher success rates than D.

TABLE 2.1

MEAN SUCCESS RATES

	A	B	C	D
All subjects	0.84	0.80	0.80	0.52
Subjects not responding to D	0.88	0.67	0.77	0
Subjects responding to D	0.82	0.87	0.82	0.86

However the subjects fell quite sharply into two groups, those on whom D had no effect and those who did respond to D. The second and third lines of the table show the corresponding mean success rates. For subjects that do respond to D, the four drugs have practically the same success rates, whereas for those who do not respond to D, A has a higher success rate than C and a much higher rate than B. Comparisons based on averages for all subjects are thus quite misleading. The difference between the two groups in the response to the drugs is possibly due to a difference in type of headache.

In this example the response to D is used to divide the experimental units in a way similar to that in which the supplementary observations were used in Example 2.4.

The general conclusion to be drawn from these examples is the desirability of being able to detect variations in the treatment effects if these are likely to be important. This means making supplementary observations where appropriate and, in other cases, assigning the treatments to the units in such a way that the variations may be detected. Methods for doing this will be discussed later. In most of the book it will, however, be assumed, in accordance with (1), that the treatment effects are constant.

2.4 INTERFERENCE BETWEEN DIFFERENT UNITS

The last aspect of the assumption (1) to need discussion is the requirement that the observation on one unit should be unaffected by the particular assignment of treatments to the other units, i.e., that there is no "interference" between different units. In many experiments the different units are physically distinct and the assumption is automatically satisfied. If, however, the same object is used as a unit several times, or if different units are in physical contact, difficulties can arise and these will now be illustrated by some examples.

Example 2.6. In the textile process called carding, an entangled mass of fibers is passed over rotating cylinders carrying teeth, which straighten the fibers.

Consider an experiment to investigate the effect of various amounts of oil applied to the raw material. The treatments are, say, four percentages of oil and the experimental units are batches of raw material. Now when a batch with a high oil content is carded, some of the oil remains on the teeth, so that the following batch, or at any rate the part of it carded first, receives in effect a larger amount of oil than its nominal treatment implies. In other words the observation on any unit is likely to depend not only on the treatment applied to that unit but also on the treatment applied to the preceding unit and even, in certain cases, on the unit two before.

One way of avoiding this difficulty is to follow each experimental batch by a control batch sufficiently large to restore the amount of oil to a standard value or, alternatively, to use large experimental batches and to make observations only on the latter part of each batch, which is unlikely to be affected by the preceding treatment. However, both these procedures, and particularly the first, would very often not be economical ways of arranging the experiment. Instead it may be preferable to accept the overlap of the treatment effects and to deal with it in the design and analysis of the experiment. This is possible provided that it is reasonable to introduce a simple modification of (1), such as that the observation on any unit is

$$\begin{pmatrix}\text{a quantity depending}\\ \text{only on the}\\ \text{unit}\end{pmatrix} + \begin{pmatrix}\text{a quantity depending}\\ \text{on the treatment}\\ \text{used}\end{pmatrix} + \begin{pmatrix}\text{a quantity depending}\\ \text{on the treatment}\\ \text{applied to the}\\ \text{preceding unit}\end{pmatrix}. \quad (4)$$

This is plausible in the present example, provided that the oil contents investigated do not vary over too wide a range. If (4) is accepted it is natural to arrange that each treatment follows each other treatment (or each treatment) the same number of times. Then the systematic change caused by following the highest oil content affects all treatments equally. Such designs are discussed in Chapter 13.

Example 2.7. Similar problems arise in investigating the effect of different diets on the milk yield of cattle. If each animal is fed on a constant diet there is no difficulty, but it would often be preferable to change over the diets in the course of the experiment and, if possible, to use each diet once on each animal. This would eliminate the effect of systematic differences between animals.

Thus, with three diets, one animal might receive diet A for the first two weeks, diet B for the second two weeks, and diet C for the third. The main observation to be analyzed would be the milk yield determined as the average of two or three days' yield at the end of each two-week period. By thus taking observations at the end of each experimental period it would be hoped that a value would be obtained characteristic only of the treatment applied during the period; however, the overlap of the treatment effects might still occur and then difficulties like those of the preceding example would arise and in particular the assumption (4) might again be reasonable. It would also be necessary to ensure that for a group of animals each treatment occurred equally frequently in each period.

The interference between different units in the above examples can be coped with because it is of a simple form. Often, however, it is better to go to some trouble to arrange that the different units are isolated, rather than to allow interference and to attempt to deal with it by a more subtle design. For example, in agricultural field trials, guard rows are left between the different plots. Again, in an experiment in which some plots are inoculated with virus-carrying aphids, while other plots are untreated, it would be essential not only to leave substantial space between treated and untreated plots but also, as far as is possible, to check that there is no direct transmission of disease from one plot to another.

Competition may arise within an experimental unit, but this causes no difficulty provided that it is representative of the conditions under investigation. For example, in a poultry feeding trial, each unit might consist of a number of birds kept together and feeding in common. If the food is limited, large healthy birds may gain at the expense of others. However, this will not invalidate the assumption of no interference *between* different groups of birds, which is involved in (1).

In experimental psychology it is frequently required to use the same subject as an experimental unit several times. In this field, however, it often happens that the effect of one treatment on the subsequent observations is not represented by anything as simple as the addition of single constants as in equation (4). Babington Smith (1951) has described experiments on the "Muller–Lyer" illusion, which suggest that responses are dependent in a rather complicated way on the whole sequence of situations that have preceded them. Welford et al. (1950), in some experiments on fatigue in aircrew, noted that subjects who first met a task when tired continued to do it badly when fresh, whereas those who first met it fresh went on doing it well when tired. Other similar effects have been reported in the literature. In such cases either a special hypothesis to replace (4) must be set up appropriate to the problem, or the treatments must be taken as whole sequences of stimuli. These experiments are mentioned here to emphasize that the simple law (4) may not be adequate.

In the remainder of the book it will be assumed, unless explicitly stated otherwise, that interference between different units is absent. If it is suspected that such interference may arise, as when the same object is used as an experimental unit more than once, or when different units are in physical contact, either experimental precautions should be taken to prevent the interference or special allowances should be made in the design and analysis of the experiment.

SUMMARY

In most cases we estimate treatment differences by averaging observations over the whole experiment. There are three points to be watched if this is done, namely (*a*) that the observations should be analyzed on a scale on which the treatment *differences* are relevant;

(*b*) that either only average treatment effects are required, or that the treatment effects are constant. Special precautions should be taken if the treatment effects are expected to depend in an important way on the value of some supplementary observation, or to be different for different groups of units;

(*c*) that the observation obtained on one unit should not be affected by the treatment applied to other units.

In the ordinary way the second and third complications are assumed absent, but if it is suspected that they may arise, they should be allowed for both in the design and in the analysis of the experiment.

REFERENCES

Babington Smith, B. (1951). On some difficulties encountered in the use of factorial designs and analysis of variance with psychological experiments. *Brit. J. Psychol.*, **42**, 250.

Goulden, C. H. (1952). *Methods of statistical analysis.* 2nd ed. New York: Wiley.

Jellinek, E. M. (1946). Clinical tests on comparative effectiveness of analgesic drugs. *Biometrics*, **2**, 87.

Welford, A. T., R. A. Brown, and J. E. Gabb. (1950). Two experiments on fatigue as affecting skilled performance in civilian aircrew. *Brit. J. Psychol.*, **40**, 195.