CHAPTER 13

Cross-Over Designs

13.1 INTRODUCTION

If most of the uncontrolled variation in an experiment is due to qualitative variations in the external conditions under which the experimental units respond, the technique of balancing by randomized blocks, Latin squares, and related designs is likely to be effective. Examples are those agricultural field trials in which local variations of fertility predominate, analytical work where variations between observers, sets of apparatus and between days are the largest sources of error, and animal experiments in which systematic differences between animals from different litters account for an appreciable proportion of the total variation.

On the other hand, if most of the variation arises from the peculiarities of individual experimental subjects, balancing into blocks on the basis of obvious spatial, temporal, or similar groupings is unlikely to be satisfactory, at any rate by itself. For example, in many types of experiment on human or animal subjects, very substantial variations may remain even after grouping on obvious features such as age, sex, etc.

One procedure in such cases is to characterize the individual experimental subjects by one or more skilfully chosen concomitant observations, these then being used as a basis either for blocking or for the calculation of adjusted treatment means of the type described in Chapter 4. Another method is artificially to divide the subjects into sections, each section then being regarded as a separate unit, the original subject forming a block. Thus in § 3.2 we mentioned the device of dividing a clover plant into two halves by cutting along the tap root, the experiment being set out in blocks of two units each, using incomplete block methods or confounding when appropriate. Example 11.2, in which different areas on a cow are used as sites for injections, illustrates the same sort of idea; in this case incomplete block designs were used.

No new problems arise if the different sections into which the subject is divided respond independently of one another, in the sense that the observation obtained on one section does not depend on the treatment

allotted to other sections. This is certainly true in the first example just mentioned, since after subdivision the two parts of the plant are quite separate; the only special consideration here is as to whether conclusions from divided plants of this sort will apply to ordinary plants.

Sometimes, however, although it is not practicable to divide the original experimental units into independent sections in this way, it is possible to use each subject (plant, animal, etc.) as an experimental unit on several occasions. This will usually eliminate the effect of much of the variation between different individuals. For example, as mentioned in § 2.4, a nutritional experiment comparing the effect of different diets on the milk yield of cows is often best arranged by having each cow fed on a sequence of diets rather than by regarding a cow as an experimental unit and keeping to a fixed diet for each cow. The special problems that this sort of design raises are connected with the possibility that the effect of a treatment applied in one period may extend into subsequent periods and hence that one of the assumptions underlying the preceding discussions, see § 2.4, is untrue. That is, the observation obtained on one experimental unit, i.e., on one individual in one period, may depend in part on the treatment applied to other units, i.e., to the same individual in preceding periods.

Arrangements in which different treatments are applied to the same subject in different periods are called *cross-over* (or change-over) *designs* and in this chapter we consider some of the special problems that they raise.

13.2 EXPERIMENTS WITHOUT CARRY-OVER EFFECTS

It may happen in using a cross-over design that it can reasonably be assumed that the complication mentioned in the previous section does not occur, i.e., that each treatment has no effect in periods subsequent to the one in which it is applied. For example, in the experiment on cows described above it might be decided to separate each experimental period by a period in which a standard treatment is applied, these further periods being sufficiently long for any effect of the earlier treatments to have dissipated. The disadvantage of this is, of course, that, if the total time for which a subject can be under observation is fixed, the number of experimental units that can be formed from each individual is reduced and the precision per individual is lowered.

If the absence of carry-over can be assumed, no really fresh problems of design arise. It would usually be reasonable to expect some time trend and hence to use Latin squares, or related designs, to balance out simultaneously variations between individuals (subjects, animals, etc.) and between times. Thus in an experiment comparing the effect on

milk yield of three diets A, B, and C a section of the experiment would consist of a randomized 3 × 3 Latin square such as

	Experimental Period		
	1	2	3
Cow 1	*B*	*A*	*C*
2	*C*	*B*	*A*
3	*A*	*C*	*B*

The three cows in each group would be chosen to have, so far as possible, lactation curves of similar slopes over the period of the experiment, the whole experiment being built up of a series of such squares. In this type of experiment the number of occasions on which each animal can be used is severely limited so that Youden squares, lattice squares or double confounding may be needed if a considerable number of treatments are to be compared.

A somewhat different situation arises if each subject can be used a large number of times, or even indefinitely. This is so in some psychological experiments, and in certain bioassay procedures, for example in a histamine assay technique described by Schild (1942); see for statistical discussion of these bioassays Finney and Outhwaite (1956). Another example is the industrial experiment of Example 2.6, where a possible carry-over effect of an oiling treatment might, under certain circumstances, be assumed negligible and where a large number of observations can be obtained from a single set of machinery. In these cases sufficient precision may be obtained from one, or at any rate a small number, of individuals.

We then have the problem of arranging say one rather long sequence of treatments in suitable order. This can usually be done satisfactorily by the method of randomized blocks. That is, the periods are divided into fairly short sections and each section used as a randomized block. Thus with four treatments we might have

$$B\,C\,A\,D \mid A\,C\,B\,D \mid D\,A\,B\,C \mid \ldots$$

Incomplete block techniques can be used where appropriate. If the experiment is a small one, and if the time variation over the period of the experiment is likely to be a smooth trend, the special type of design described in § 14.2 may be more appropriate than the method of randomized blocks.

Suppose that the absence of carry-over effects has been ensured by the method, mentioned above, of including intermediate sections of sufficient length in which a standard treatment is applied to the individuals. The experiment then leads to valid estimates of the effect of differences

between the treatments in the system investigated, i.e., under conditions in which treatments are being changed frequently. Often, however, our practical interest is in what would happen if treatments were applied continuously to experimental subjects, e.g., in comparing the milk yield of cows fed continuously on diet A with the yield that would have been obtained with continuous feeding on diet B. Thus the danger has to be watched that additional precision may be attained by the cross-over method, only at the cost of distorting the comparisons required.

13.3 CROSS-OVER DESIGNS WITH A LIMITED NUMBER OF PERIODS PER INDIVIDUAL

In the experiments on milk yield discussed above, not more than three or four treatments can usually be applied to each animal, and it is necessary to have a number of animals in the experiment in order to get a satisfactory arrangement. This is quite a common situation and so we consider first designs in which several individuals are used simultaneously.

We must introduce an assumption concerning the effect that a treatment applied in one period may have in subsequent periods. The simplest such assumption is the one mentioned in § 2.4 that the observation obtained on an individual in a particular period is

$$\begin{pmatrix}\text{a quantity depending}\\ \text{only on the individual-}\\ \text{period combination}\\ \text{and independent of}\\ \text{the treatments}\end{pmatrix} + \begin{pmatrix}\text{a quantity}\\ \text{depending on}\\ \text{the treatment}\\ \text{applied in}\\ \text{that period}\end{pmatrix} + \begin{pmatrix}\text{a quantity}\\ \text{depending on}\\ \text{the treatment}\\ \text{applied in the}\\ \text{preceding period}\end{pmatrix}$$

Thus each treatment is characterized by two quantities, one expressing its *direct effect* in the period in which it is applied, the other giving its *residual effect* in the following period. In this case a natural design is one in which each treatment follows each other treatment the same number of times. At the same time we shall want a Latin square design to ensure that each treatment occurs equally often in each period and on each subject.

Williams (1949) has given suitable arrangements, obtained as follows: Suppose first that the number of treatments n is even. Write down as the first row the numbers

$$1 \quad 2 \quad n \quad 3 \quad n-1 \quad 4 \ldots$$

in which the sequence $1, n, n-1, \ldots$ alternates with the sequence $2, 3, 4, \ldots$. Thus with $n = 6$, the first row is 1 2 6 3 5 4.

The remaining rows of the square are now obtained from the first by

successive additions of 1, using the rule that numbers above 6 are to have 6 subtracted from them. The final square for $n = 6$ is thus

1	2	6	3	5	4
2	3	1	4	6	5
3	4	2	5	1	6
4	5	3	6	2	1
5	6	4	1	3	2
6	1	5	2	4	3

The important property of this square, which is a consequence of the particular choice of the first row, is that not only is it a Latin square, but that also each treatment follows each other treatment just once. Thus treatment five follows treatment three in the first row, treatment six in the second row, and so on.

To use the design, groups of six subjects, likely to show similar time trends and residual effects, are taken. Each group is assigned to one such square and each subject assigned randomly to a row of the square. The six numbers in the row determine the treatments to be applied to the subject in the six periods, i.e., rows correspond to subjects, columns to periods.

To obtain similar designs when the number of treatments is odd, it is necessary to consider pairs of squares simultaneously. For one square the first row is taken to be

$$1 \quad 2 \quad n \quad 3 \quad n-1 \quad 4 \quad n-2 \ldots$$

and for the other square of the pair, the first row is this reversed. Thus with $n = 5$, the first rows are

	1	2	5	3	4
and	4	3	5	2	1

so that the full squares are

1	2	5	3	4		4	3	5	2	1
2	3	1	4	5		5	4	1	3	2
3	4	2	5	1		1	5	2	4	3
4	5	3	1	2		2	1	3	5	4
5	1	4	2	3		3	2	4	1	5

These have the property that each treatment follows each other treatment just twice.

To use the squares, make rows correspond to subjects and columns to

periods. Thus with five treatments, a multiple of ten subjects is necessary and each subject must be capable of receiving five different treatments. The ten subjects in each pair of squares should be likely to have similar residual effects, and so far as possible the period effects should be constant within each square of five subjects.

The general property of these designs is that each treatment follows each treatment except itself the same number of times. Therefore the mean of all observations on, say, treatment 1 is influenced by the residual effects of all treatments except the first. This implies that the difference between the mean observations on two treatments is not, as it stands, an estimate of the appropriate true direct treatment effect, if different residual effects are present. Thus if treatment 1 has a large positive residual effect and the other treatments do not, the mean observation on treatment 1 is depressed relative to the other treatments. This can be corrected by the calculation of adjustments analogous to those used for balanced incomplete designs.

Example 13.1. Williams (1949) has given the following example. Samples of pulp suspension at varying concentrations were beaten in a Lampén mill to determine the effect of concentration on the properties of the resulting sheets. Observations of the condition of the mill after each beating indicated that certain concentrations of pulp had an effect on the mill which might affect the next beating. Hence a design balanced for residual effects was used. With six treatments, six runs, and six periods per run, the design and observations (burst factors) of Table 13.1(*a*) were obtained: the rows of the 6 × 6 square given in the text above have been randomized.

The formulas for estimating the treatment and residual effects, and for finding the precision of the estimates, are rather complicated and will not be given here. Full accounts are given in Williams's paper and by Cochran and Cox* (1957, § 4.6*a*).

The effect of the process of adjustment can be judged from Table 13.1(*b*), which gives the unadjusted and adjusted direct and residual effects. For instance the unadjusted mean for T_5 is just the mean of the six observations on this treatment and is 57.75. After adjustment for residual effects, this becomes 57.98. The adjustments to the direct effects are quite small in this example; they would, of course, be greater if large residual effects were present, and they also tend to be greater in smaller squares. In this particular example both direct and residual effects are statistically significant.

If it had not been necessary to apply adjustments, the standard error of the difference in direct effect between two treatments would have been that for the difference of two means each of six observations. In fact, the standard error is slightly greater than this because of random errors in the adjustments that are applied to correct for the presence of residual effects.

Various modifications to the design of Table 13.1 may be worth considering. In some situations the sum of the direct and residual effects

* In the second edition only.

associated with a treatment is of interest for estimating the response that would be obtained if the treatment were applied continuously.* The

TABLE 13.1

A 6 × 6 EXPERIMENT WITH CARRY-OVER EFFECTS

(*a*) *Plan and Observations*

	Period					
	1	2	3	4	5	6
Run 1	T_3:56.7	T_6:53.8	T_2:54.4	T_5:54.4	T_4:58.9	T_1:54.5
2	T_5:58.5	T_3:60.2	T_4:61.3	T_6:54.4	T_1:59.1	T_2:59.8
3	T_1:55.7	T_4:60.7	T_5:56.7	T_2:59.9	T_6:56.6	T_3:59.6
4	T_2:57.3	T_1:57.7	T_6:55.2	T_4:58.1	T_3:60.2	T_5:60.2
5	T_6:53.7	T_5:57.1	T_1:59.2	T_3:58.9	T_2:58.9	T_4:59.6
6	T_4:58.1	T_2:55.7	T_3:58.9	T_1:56.6	T_5:59.6	T_6:57.5

(*b*) *Unadjusted and Adjusted Effects*

	Direct Effects		Residual Effects	
	Unadjusted	Adjusted	Unadjusted	Adjusted
T_1	57.13	57.20	0.92	0.37
T_2	57.67	57.62	−0.48	−0.28
T_3	59.08	59.19	0.24	0.65
T_4	59.45	59.23	−1.62	−1.33
T_5	57.75	57.98	1.22	1.40
T_6	55.20	55.06	−0.26	−0.82

efficiency of estimation of these combined effects is increased by adding a further period at the end for each subject in which the final treatment is repeated. Another method of achieving this is to add a final period in which a uniform control treatment is applied to all subjects. This will be particularly suitable if the residual effects are of intrinsic interest. A final possibility is to add a preliminary period in which the same treatment is applied as in the first experimental period, but in which no observation is made. If the main expense is in making the observation, rather than in applying the treatments, this preliminary period will increase the efficiency with which the treatment effects are estimated.

More complicated designs of this form are needed if the residual effect is suspected to extend for more than one period. It is then natural to look for Latin squares in which each treatment follows each ordered *pair* of

* This assumes that the residual effect when, say, treatment A follows itself is the same as when A is followed by a different treatment, and this may not be true.

other treatments the same number of times. In a further paper Williams (1950) has described such designs; their analysis is rather complicated.

In experiments of the type we are considering, the number of observations that can be made on each subject is severely limited. Hence if a considerable number of treatments is involved, the use of confounding and of balanced incomplete block and related designs is natural. The development of such designs, allowing for the complication of residual treatment effects, has been considered by Patterson (1951).

Finney (1956) has given a careful account of the various types of assumption on which the analysis of this sort of experiment can be based. Patterson (1950) has described a method of analysis that is particularly appropriate when the main difference between subjects is a variation in the slope of their response curves in time.

13.4 DESIGNS WITH A LARGE NUMBER OF OBSERVATIONS PER SUBJECT

Instead of there being a number of subjects with each treatment occurring at most once on each subject, it may happen that there is only one subject, or perhaps two, but that a large number of treatment applications may be made on it. Problems similar to those discussed in § 13.3 will still arise, if there is the possibility of a carry-over of treatment effects from one period to the next.

One example is the textile oiling experiment of Example 2.6, where one set of machinery (subject) is used and where the whole experiment is specified by the sequence of treatments which this subject receives. Another is the bioassay technique (Schild, 1942) mentioned earlier in this chapter, and a third is concerned with the local depression of the milk yield of a cow following injection with insulin, one cow being used and injections of varying amounts and kinds following one another in a long sequence.

One procedure in such cases is to divide the experiment into several separate sections, with a gap between them, to call each section a "subject" and then to apply the methods of § 13.3. Example 13.2 has been set out in this way. Often, however, this is not the best approach, since if the experiment is planned as a single sequence, information about the carry-over effects of treatments is supplied by all observations except the first, so that a single sequence design tends to have higher precision than the design of § 13.3 with the same number of treatment applications.

Finney and Outhwaite (1956) and Sampford (1957) have discussed suitable designs and their papers should be consulted for details.

13.5 SOME OTHER POSSIBILITIES

In the previous sections, we have assumed that any carry-over of treatment effects from one period to another persists for one period, or perhaps two, but is definitely limited in extent. Sometimes other forms of carry-over of treatment effects may seem natural and in such cases an appropriate design has to be found either intuitively or occasionally by theoretical analysis.

For example, it may be suspected that the first treatment that a subject receives has a substantial effect on all its remaining responses, but that there is no other carry-over.

Or, it may be thought that there is a characteristic effect associated with each treatment application, depending in some fairly simple way on the number of times that particular treatment has been applied to the subject before. Pearce (1957) has considered yet another possibility, namely that the units are divided into sets and that each treatment has a direct effect on the unit to which it is applied and an equal carry-over on all units in the same set. In all cases such as these, the general method is to set up a mathematical formula to represent the observations and then, if possible, to construct a design that will allow the quantities in this mathematical expression to be estimated as simply and precisely as possible. Specialist advice may be necessary to do this.

In rotation experiments, treatments are applied in sequence to a subject, e.g., a plot, and there may be a carry-over effect of one treatment into subsequent periods; in these experiments, however, the treatments are applied in definite predetermined sequences and it is the response of the subject to the sequences of treatments that is of interest, rather than the response to the individual treatments. Rotation experiments raise some specialized problems and will not be considered here; see Cochran (1939).

Another possibility is that one may be interested in the result of applying a treatment continuously to a subject for a considerable period. Here, the simplest method is to hold the treatment constant for each subject, and to construct from the observations on each subject (*a*) a measure of the average response of the subject (the mean of all observations on the subject), and (*b*) a measure of the rate of increase or decrease with time of the observation (linear regression coefficient on time). These are then analyzed separately. Stevens (1949) has given an interesting example of an experiment on a perennial crop, coffee, where from yearly observations on each plot he constructed measures of both (*a*) and (*b*), and also a measure of the amplitude of the twoyearly periodic variation in yield. The effect of the treatments on these three aspects of the yield pattern was then analyzed and interpreted separately.

SUMMARY

In situations where a substantial portion of the uncontrolled variation arises from the peculiarities of the individual physical objects (subjects, animals, etc.) that form the experimental units, one method of increasing precision is to use each object as an experimental unit several times. That is we arrange, for example, that each animal receives several treatments, rather than being kept on the same treatment throughout.

Latin squares are a natural design to use in such cases. Sometimes, however, there is the complication that the observation obtained in one period depends on the treatments applied to the object in previous periods, as well as on the current treatment. Special Latin squares are useful when there is such a carry-over effect of treatments.

There are some difficulties connected with these designs. Although a substantial increase in precision may be obtained, the treatment effects when each treatment is applied only for a short period may not be the same as those when each treatment is applied for a long period. If it is the latter that are of interest, the increase in precision will have been attained by answering the wrong question. A second difficulty is that the full analysis of observations, when carry-over effects are present, is a bit complicated.

REFERENCES

Cochran, W. G. (1939). Long-term agricultural experiments. *J. R. Statist. Soc. Suppl.*, **6**, 104.

——— and G. M. Cox. (1957). *Experimental designs.* 2nd ed. New York: Wiley.

Finney, D. J. (1956). Cross-over designs in bioassay. *Proc. Roy. Soc.* B, **145**, 42.

——— and A. D. Outhwaite. (1956). Serially balanced sequences in bioassay. *Proc. Roy. Soc.* B, **145**, 493.

Patterson, H. D. (1950). The analysis of change-over trials. *J. Agric. Sci.*, **40**, 375.

——— (1951). Change-over trials. *J. R. Statist. Soc.* B, **13**, 256.

Pearce, S. C. (1957). Experimenting with organisms as blocks. *Biometrika*, **44**, 141.

Sampford, M. R. (1957). Serially balanced designs. *J. R. Statist. Soc.* B, **19**, 286.

Schild, H. O. (1942). A method of conducting a biological assay on a preparation giving gradual responses illustrated by the estimation of histamine. *J. Physiol.*, **101**, 115.

Stevens, W. L. (1949). Analise estatistica do ensaio de variedades de cafe. *Bragantia*, **9**, 103.

Williams, E. J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Austr. J. Sci. Res.* A, **2**, 149.

——— (1950). Experimental designs balanced for pairs of residual effects. *Austr. J. Sci. Res.* A, **3**, 351.