# IELD EXPERIMENTATION WITH FRUIT TREES AND OTHER PERENNIAL PLANTS

by

S. C. Pearce



*Technical Communication No. 23 of the Commonwealth Bureau of Horticulture and Plantation Crops East Malling, Maidstone, Kent, England*

C·A·B

COMMONWEALTH AGRICULTURAL BUREAUX

*Price* 10s. *or* $1.40

*Enquiries on the subject matter should be sent to the author, those on the work of the Commonwealth Bureau of Horticulture and Plantation Crops to the Director of the Bureau.*

*Orders and requests for lists of Agricultural Science publications should be sent to Commonwealth Agricultural Bureaux, Central Sales Branch, Farnham Royal, Bucks, England.*

# FIELD EXPERIMENTATION WITH FRUIT TREES AND OTHER PERENNIAL PLANTS

By

S. C. PEARCE

East Malling Research Station, Maidstone, Kent, England

*Technical Communication No. 23 of the*
*Commonwealth Bureau of Horticulture and Plantation Crops*
*East Malling, Maidstone, Kent*

COMMONWEALTH AGRICULTURAL BUREAUX

First published in March, 1953, by the Commonwealth Agricultural Bureaux, Farnham Royal, Bucks, England. 10s. or $1.40.

# NOTE

The Bureau is extremely grateful to those who have contributed to this Bulletin. First and foremost we thank Dr. Pearce for his efficiency and timeliness of production, for his long-suffering tolerance of our maundering inquiries on the meaning of certain passages, and his willingness to make concessions to our bucolic dopiness by substituting statistical words of one syllable for more erudite terms less intelligible to the multitude.

Secondly, we thank our Official Correspondents all over the world, and Dr. T. Eden and other experts for their much appreciated and useful comments on the synopsis. This was circulated in May, 1952, and some 32 replies were received.

Lastly we are grateful to Mr. T. N. Hoblyn for his support, friendly criticism and foreword.

D. AKENHEAD.

Commonwealth Bureau,
East Malling.

# FOREWORD

The publication of R. A. Fisher's *Statistical Methods for Research Workers* in 1925 was the occasion of something very like a revolution in methods of experimental design. But though the new methods, involving such ideas as randomization of plots, factorial design and the application of the analysis of variance in elucidation of the results, came quickly into use with annual crops, notably, of course, at Rothamsted, there was at first little experience of their use with perennials—indeed, bearing in mind the peculiar difficulties of experiments with long-lived, often very variable crops, such innovations were regarded by some research workers with considerable suspicion.

For some years, however, R. G. Hatton and his co-workers had been steadily developing a technique for experiments with deciduous fruit trees at East Malling and had accumulated much information on the variation of such material, all of which was invaluable as a basis for future experimentation. Gradually, in a small way, the methods of design advocated by Fisher were coming into use at East Malling, though, as yet, there were few results with which to demonstrate their applicability or greater accuracy.

In 1930, one year after its establishment, the Imperial Bureau of Fruit Production arranged a Conference of all those within the Empire interested in horticultural science. This Conference was attended by many research workers interested in a wide variety of horticultural crops, deciduous, subtropical and tropical, many of them facing a field almost wholly unexplored. Their interest in the technique of experiments and in the application of statistical methods in their researches was then freely expressed, and in response to their demands the Bureau set about the collection of all the information then available on field experiments with fruit trees. The result was a Technical Communication published in 1931, the writer of this foreword being the author.

The said writer devoted much of that publication to technique, but when he considers the immense advances in methods of experimental design that have been made in the twenty-two years that have elapsed, now set forth by Dr. Pearce, he wonders how he had the temerity to recommend the comparatively simple designs and methods of analysis then described. Nevertheless, that early enterprise on the part of the Bureau now appears to be amply justified and, if the hope of very much more accurate field experiments has perhaps not been fully realized, this is due more to the obstinate variability of long lived plants than to the inadequacy of the methods then recommended.

During the last 20 years the whole field of research has widened, and many of the problems now under investigation are very complex. With these modern developments the statisticians have tried to keep pace, improving old methods and devising fresh ones to meet new demands. In the chapters which follow Dr. Pearce has collected and summarized statistical advance in the field of perennial crops.

Twenty-five years experience of fruit experiments have convinced the writer that in the long run simple designs are generally to be preferred to complex ones; nevertheless in some circumstances the latter must be used and Dr. Pearce describes them for use when needed.

Finally, a word of warning to those who, without full understanding, would attempt the statistical analysis of long term experiments by rule of thumb. By all means summarize and examine your results carefully. But, if a statistician is available, ask his advice both in planning and analysis; it will nearly always pay in the end.

T. N. HOBLYN.

East Malling,
*January*, 1953.

# AUTHOR'S PREFACE

Before putting forward this publication, it may be as well to explain its purpose and how it came to be written. The writer has spent some years as a member of the Statistics Section at East Malling applying statistical methods to problems of experimentation with perennial plants and, as might be expected, he has developed a number of ideas and learned a number of techniques, which, it was thought, might be worth introducing to experimenters generally, and this he has tried to do. The work is, in fact, addressed to the practical experimenter and not to the mathematician, who will find little that is novel.

Nevertheless it does not set out to provide an elementary grounding in statistics. This average experimenter has probably had a brief course in statistics as part of his training and has been forced to learn more by the nature of his work. Consequently, he does not need to be told what a standard error is or instructed in the meaning of an *F*-test, but he does want to know why things are done and how various statistical techniques relate to his work, though others do not. For that reason some attention has been given to mentioning the occasions upon which a test may validly be used and giving warning of the occasions upon which it may not.

Because of this desire to make clear the strategy of statistical methods, little has been said about tactics, which for this purpose means the actual handling of data. The writer has been requested by many correspondents to append to each design a worked example and a detailed description of the computing method by which it was obtained. This he has preferred not to do. For one thing, it has been done before and done better than anything he could hope to achieve: for another, he believes that in computing results no less than in designing trials there is a long view of what it is about and a short view of what to do next, and he has tried to help those who want to take the long view. Computing is not really difficult if the student will only get away from rule-of-thumb methods set out in a book and try to see what the figures—sums of squares, mean squares, variance ratios, significant differences and the rest—mean in terms of the experiment and its purpose. It is hoped that the four appendices will help horticultural experimenters to achieve this end.

One difficulty in writing has been to arrange the material both for those who want a text-book to be read through from beginning to end and those who want a book for reference. The proffered solution is to arrange the 66 Sections in a convenient order for continuous reading, but to include frequent references to other Sections, whether earlier or later, for the sake of those who will be reading parts out of their context. These references can be ignored by the systematic reader. If anyone is rash enough to attempt the book without previous statistical experience, he should interpolate Appendices I and II between Chapters 1 and 2, Appendix III between Chapters 2 and 3 and Appendix IV between Chapters 3 and 4.

It remains only to thank those who have helped in this publication. First of all, the staff of the Commonwealth Bureau of Horticulture and Plantation Crops have given invaluable help in many directions. Then, correspondents all over the world have sent in their suggestions based on experience with a diversity of crops and have often brought recommendations firmly down to earth by citing special local difficulties. Finally, the writer must thank his friends and colleagues of the Statistics Section at East Malling, both past and present, for the help and advice they have given, especially Mr. T. N. Hoblyn, who has also contributed a foreword. This book is the expression not just of the experiences of the author whose name appears on the title-page, but of a continuing tradition in which he is glad to have played a part.

# TABLE OF CONTENTS

# LIST OF FIGURES



# TABLES

# FIELD EXPERIMENTATION WITH FRUIT TREES AND OTHER PERENNIAL PLANTS

## CHAPTER 1

## INTRODUCTION

" I heard the Duke of Albemarle's chaplain make a mighty simple sermon : among other things reproaching the imperfection of human learning, he cried : ' All our physicians cannot tell what an ague is, and all our arithmetique is not able to number the days of a man', which, God knows, is not the fault of arithmetique, but that our understandings reach not the thing."

Samuel Pepys, ' Diary ' (Entry for 5.xi.1665).

### 10. The Rise of Statistics

During the past fifty years the research worker in agriculture and horticulture has seen many changes, not the least of which has been his own transformation from a practical farmer with a bent for science to a scientist with a concern for farming. For his new role he has had to acquire many new techniques and, of them all, statistics is to many the most puzzling and, it sometimes seems, the least relevant. For this, the statisticians are themselves partly to blame, for too often they have been content to evolve formulae without thinking what their conclusions mean in the field, though the uncritical acceptance of statistical methods by some research workers has not helped their more cautious colleagues in taking a balanced view of the matter. However, at least as far as the English speaking world is concerned, it does now seem to be accepted that statistics has its valued place in research, though there remains the task of finding the best synthesis of field worker and mathematician [cf. 191].*

In this respect horticulturists have been fortunate, for one of the first major statistical publications in their field, appearing in 1931, was T. N. Hoblyn's " Field Experiments in Horticulture " [78], written by one who knew the practical difficulties and who advocated statistical methods while recognizing the importance of non-statistical considerations and appreciating the special needs of different species. For twenty years, as long as it remained in print, this was a standard work, some of the more modern developments being discussed in a joint paper by Mr. Hoblyn and the present writer [119], published in 1948 and intended as a continuation of the earlier publication. A new edition is now called for. In form the present work has little relation to the earlier, but its author would like to believe that the practical spirit is retained, though the problems dealt with are so widely different.

### 11. The Purpose of Field Trials

The object of an experiment is to obtain data that will elucidate an obscurity or decide a course of action. Consequently a good experimental design can result only from a precise appreciation of the problem to be solved and no refinement of technique can be a substitute for clear-headedness, however valuable it may be in addition. This communication will set out in some detail those statistical techniques that the author has found useful in fruit tree

* Unless otherwise stated, numbers in square brackets indicate the references in the Bibliography.

experimentation, but he would emphasize from the start that nothing he can write will diminish by one iota the thoughtfulness and insight needed to conceive and conclude a successful research programme. His aim is to provide tools and not to decide the task.

In the past it was considered sufficient if a trial evaluated the mean performance resulting from each treatment; but, since Professor Fisher revolutionized the approach, it has been required that each trial should estimate its own accuracy as well. The conditions for achieving this additional aim, namely replication and randomization, did not immediately commend themselves to practical experimenters, for they saw in the first the occasion of much work, and in the second an apparent refusal to minimize positional variation by balancing. That these objections have some substance is undoubtedly true, but it is now generally conceded that these modern methods have won their place because they provide measures not only of treatment means but also of the reliance to be placed upon them.

### 12. The Idea of Significance

Fundamental to these methods is the idea of *significance*, which is so often misunderstood that a note may be helpful, though excellent accounts are to be found elsewhere [*e.g.*, 64*a*]. Suppose a difference is reported to be significant at the level, $P = 0{\cdot}01$. This means that the reader has to choose between two incompatible alternatives—*either* to accept that the difference was in fact due to the treatments, *or* to assert that the result is just a coincidence, the figures happening to fall out like that by chance, the probability of such a coincidence being one in a hundred (0·01). In fact, a significance level informs a would-be sceptic of the probability of such results arising if there really is nothing in the experimenter's claim that they were brought about by his treatments.

To clear up some misconceptions, the statement does not mean that there is a probability of 0·99 that the difference was in fact due to the treatments, for such a statement would have to depend upon something that could not have been investigated, if indeed it exists at all, namely, the inherent probability of such treatments proving effective. Nor does a significance level prove anything conclusively. It calls upon the reader to make a choice and tells him the low objective probability of one of the alternatives before him as a reason for choosing the other, the probability of which is indeterminate.

The reader's choice will depend both upon the treatment and the preconceptions of his own mind. Suppose the treated trees have given a crop greater than that of the controls and it is this difference that is significant at the 1 in 100 level. If the treatment was an invocation and ritual dance to the goddess Pomona, a modern would doubtless prefer to believe in the coincidence, but not if the treatment had been the provision of sulphate of ammonia; yet to the ancient world contrary choices would perhaps have appeared equally obvious. In fact, although the significance level is itself objectively determined, there is a subjective element in its interpretation. Cogent reasons or collateral evidence in favour of a treatment effect will reduce the level needed to carry conviction, just as contrary evidence will make belief more difficult.

In general, if $P$ is less than 0·05, the evidence in favour of a genuine treatment effect is considered strong enough to be worth publishing. If $P$ is less than 0·01 the effect is pretty well established, while with $P$ less than 0·001 the evidence is considered very strong. By a convention that could well be more widely followed, the first of these levels is indicated by one asterisk, the second by two and the third by three.

There is a tendency at the present time to decry this concept of significance by saying, quite truly, that reliance should not be placed upon any single trial but only upon a series. With annual species this contention is very reasonable, for it is easy to repeat an experiment in a later season and results may well prove different. The same applies, for example, to spraying trials with fruit trees, but with cultural and varietal trials on perennials the investigation will often take an inconveniently long time in any case and the initial trial does itself cover a range of seasonal conditions. This emphasis on significance is not intended in any way to deprecate the use of modern designs to assign a standard error to the experimental results.

### 13. Auxiliaries to Field Trials

Such a field trial is not by any means the only way of conducting research and for some purposes it is unsatisfactory, chiefly on account of its being restricted to one locality. There is much to be said for a chain of sub-stations to confirm the conclusions reached at a main research station, though, with the kind of crops here being considered, an investigation takes so many years to complete that any postponement of the final decision is to be avoided if possible. Some very simple issues, such as the performance of a new variety, can well be investigated in this way, the main research centre conducting its trial concurrently with the sub-centres. In this connection, interested growers, farm institutes and schools with demonstration gardens can all be asked to play a part.

Yet another method of research is the survey, of which there are two kinds. In the one, the aim is to obtain factual information, *e.g.*, to find out what crops are actually being grown on soils of a certain type, and this has little relation to the field trial. In the other, the aim is to find factors associated with high cropping, disease or other characteristics. This kind of survey is subject to two pitfalls, both avoidable given good planning. One arises from possible causes being themselves associated. For example, suppose it were found that wide planted orchards crop better than close planted ones, would this indicate a genuine effect of spacing or a tendency for good growers to leave enough room for spray machinery ? The other is the risk of spurious correlations arising when studying many possible causes. Given a series of data and a column of random numbers there is one chance in twenty of finding a significant correlation ($P = 0{\cdot}05$) between the two. Given twenty columns of random numbers it is to be expected that one of them would show significance at that level. Consequently if a search is conducted for attributes of farms that will show a correlation with the yields obtained, success must almost certainly result if the search goes on long enough. It is not, of course, suggested that these difficulties are overlooked by competent investigators, but they do show that a survey is not as easy of interpretation as novices sometimes suppose. The method can be valuable and, in its application to fruit trees, has been used by Wilcox [176].

The scope of this publication includes only field trials, but these other methods are mentioned to set the matter in perspective. Field trials, though they cannot answer every question, provide an exceedingly valuable method of research, especially when they are undertaken in conjunction with laboratory investigations of a more fundamental nature and with these other methods of research, which are valuable auxiliaries.

### 14. Characteristics of Perennial Plants

It may well be asked in what respect fruit trees and perennial plants in general are so different from other crops as to need special statistical methods. First of all, of course, they live longer and are therefore more susceptible to mishaps, so greater caution is needed in making plans about them. Also, they are in general larger than annuals and are of greater interest as individuals. Thus, the crop of a single cacao tree can readily be measured and considered, but wheat is usually studied only in aggregates of several hundred plants, if not of several thousands. This concentration on the individual means also that there is a further source of variation to be considered. The single cacao tree may be large or small from factors that reside in the tree itself, but an equal area of wheat is made up of many plants, some large and some small, whose varying sizes compensate for one other, leaving only positional factors such as soil to be considered. With the perennial, on the other hand, positional variation is rarely of sole importance.

To take these points one by one, the greater caution needed with perennials militates against plans that will be ruined if some trees die, while it bids the experimenter bear in mind the time when his trial is finished but the trees will remain—ready, if foresight has been exercised, for a further trial. Also, the greater size of plants means that the experimenter cannot divide up his experimental area as he would often like to. If it measures 180 feet by 120 feet and his trees ought to be planted 20 feet apart and he proposes to use one-tree plots, then he must have 54 plots, nine by six. With smaller plants he could have divided up his land with much more freedom.

The final point concerns the sources of variation. This is a matter on which much has been written, the main conclusions being, to say the least, conflicting. It is, however, clear that the experimenter with perennial plants must not proceed as if all his variation were positional as his colleagues working with annuals may justifiably do—a fact pointed out by Hatton as long ago as 1931 [75]. Some of it almost always comes from the plants themselves. Recent work at East Malling suggests that the proportion of total variation arising from non-positional sources is greatest—

(1) in young trees that have not long been subject to the effect of their planting position.

(2) with trees of variable genetical make-up.

Consequently, it is a mistake to pay attention solely to the control of positional variation as is rightly done with short-lived plants.

These considerations suggest that perennial plants pose a number of distinctive, statistical questions that need to be answered. They also suggest the approach in certain border-line cases. For example, is sugar cane, which though perennial is herbaceous, to be treated statistically as a perennial or an annual? The writer should here make it plain that he can claim no experience with the crop and there is not a lot of guidance in the literature, but it is to be expected that sugar cane would behave sometimes like one and sometimes like the other. Thus, first of all, it is exposed to more hazards than short term annuals and one set of plants might be used for successive trials. It is true that damage to a plot of cane in one year, say by a lorry running into it, is likely to be repaired in the next year when the canes grow up again; but there will always be the gap in the accumulated crops over any period of years that includes the mishap. Next, there is size of plant. In the area it occupies, a cane stool is small and a plot contains many of them, measurements being made on an aggregate of plants and not on individuals. Consequently, positional variation is likely to be more important than that inherent in the material and devices to control it may well be worth-while. Also, the experimental area may be divided up into plots more or less as the investigator chooses. In fact, the crop may be expected to behave like a perennial in respect of needing foresight in preparing for a change of treatments and in being ready for missing plots, but like an annual in respect of a need to control positional variation. The report of a recent conference [25] suggests that these points are in fact in the minds of experimenters.

It might also be asked whether the fact that sugar cane is grown for something other than its fruit does not call for further modification of technique. There must, of course, be very different records taken with crops such as sugar, tea, rubber and wattle, but the bases of trial design do not appear to be different in any important way.

### 15. Choice of Experimental Plants

Naturally the trees of an experiment should be chosen so as to be as similar as possible, and this can be done only if the investigator has a good idea where his variation is coming from. As was mentioned in the last Section, with perennials there is a good deal of variation inherent in the plants themselves and this should be minimized as far as possible.

With most species genetical variation can be important. Sometimes clonal material is available and this is the ideal, at other times inbred seedlings are the best possible. It sometimes happens that the source of material has to be considered, whether on account of viruses or possible sports in clones or on account of strains of seedlings. In any experiment the source of each plant should be known and, if it is not practicable to use plants from only one source, either each block should be made up in this way (see Section 50), or a pseudo-variate must be used to eliminate possible differences (see Section 81).

Most experimenters take some pains to standardize the initial size of the trees they plant, though, for some species at least, all plants initially botanically complete end up much the same size. Thus Rogers found this to be so for strawberries [139], while several investigators have reached this opinion for apples [22, 50, 114, 147]. With citrus, on the other hand, it was found by Webber [168] that initial size differences persisted at least for some years, while Lutz

[91] found the same for pecans. Although large differences in size are undesirable at planting, it is open to question whether the attention given to this point is not, for some species at any rate, excessive. Extreme differences are, however, to be avoided.

In a trial of rootstocks or varieties it would in any event be wrong to select all plants within constant limits of size, because this leads to some varieties being represented only by their larger plants and others only by their smaller. Either (*a*) selection must be made from all available plants, or (*b*) each variety must be allowed to set its own standard, selection being made, say, only from those trees within 10% of the varietal mean.

Care should always be taken to exclude from the experiment trees that are misshapen or damaged by pests. With some species defective graft unions especially must be looked for. Also, as Bradford [21] has pointed out, at quite an early age some important characteristics of a tree, *e.g.*, crutch angle, are already determined and these may well affect ultimate development. Quite apart from size grading, already referred to, trees must be excluded that are too small to give the shape of tree desired or are in some other way going to be formed only with difficulty.

When all unsuitable plants have been excluded, the final choice should be made at random. With trees the allocation to field positions should also be at random.

## 16. The Place of Statistics

In deciding upon a design it should be remembered that statistical considerations, though very important, are never paramount. Whatever design is used, it must be thoroughly practicable, for it is no use producing a scheme on paper that will give endless trouble in the field. If trees are going to need 24 feet to grow in, it is no use planting them 20 feet apart to allow of sufficient replication. Pollinators and shade trees are often a nuisance, but if they are needed they must be provided, and in sufficient number to be effective. Statistical requirements are sometimes given such weight that results are based on abnormal trees growing in abnormal conditions, and this is a bad mistake.

Neither may statistical considerations be allowed to modify the treatments. If clean cultivation on the farm means ploughing, the experiment must not test hand digging. Again, a knapsack spray on a small tree might have a different effect from commercial spraying in an established orchard, while a broadcast application of fertilizer is not the same as drilling it in. Of course, some trials of a fundamental nature are better conducted using one leaf in a Petri dish than a whole plantation; but in trials that purport to test farm operations there should be no substitutes unacceptable to practical men.

Also it should be borne in mind that mathematical concepts are introduced only as a means of defining and investigating horticultural questions and never for their own sake. It is, of course, true that mathematical ideas can rarely be adequately expressed in words; but where a question has been translated from the language of the botanist into the language of the mathematician, the answer must be of a sort that can be translated back again. It is no good coming to the conclusion that the cubic effect of this interacts with the parabolic effect of something else, the data being in square-root transformation, and supposing that the answer will without further explanation help an investigator who wants to know how much superphosphate he should give when bushes are planted 800 to the acre.

Along with this, the trial must be statistically valid and reasonably sensitive to treatment effects. This requirement can no more be relaxed than the others and it is sometimes a severe test of the designer's skill to reconcile them all. In particular, care is needed in avoiding the attitude that a good statistician can analyse anything, and consequently almost any design will prove valid if only good enough advice can be obtained eventually. Such confidence can easily be misplaced, especially if the randomization has been restricted to meet practical ends. Some of the more complicated designs were evolved in the first place expressly to make the reconciliation of statistical and practical considerations possible and these can be very useful, but it is better to avoid complexity whenever this can be done. Above all, it is foolish to indulge in the sort of statistical virtuosity that supposes any complex design to be better than any simple one and the laying out of a trial to be an exhibition of knowledge. The true aim of the designer should be to obtain the desired information as easily as possible, so complexity of analysis should really go into the debit side of the account, though even so it can sometimes be well worth while.

To sum up, a design must be satisfactory in several respects. Before any trial is proceeded with it must be possible to answer "Yes" to three questions: Is it practicable? Is it statistically sound? Are the treatments really what they purport to be? A failure in any of these respects is disastrous.

[From this point it will be assumed that the reader is conversant with the general features of an analysis of variance for an orthogonal design, whether factorial or not. Such analyses are described in Appendix I and the first part of Appendix II.]

# CHAPTER 2

# SOME SIMPLE DESIGNS

" And as in races it is not the large stride or high lift that makes the speed; so in business, the keeping close to the matter, and not taking of it too much at once, procureth dispatch."

Francis Bacon, Essay XXV, ' Of Dispatch'.

## 20. Complete Randomization

The simplest of all possible designs is one in which no attempt is made to allow for position, the treatments being allocated to plots completely at random, as in Figure I. This is one of the oldest of modern designs [63*a*]. It will at once be objected that this ignoring of positional

| | | | |
|---|---|---|---|
| B | A | A | B |
| A | B | A | A |
| B | B | B | A |

FIGURE I.—Design with complete randomization.

effects is very dangerous for there is nothing to prevent the plots of one treatment grouping themselves together at one end. This is, of course, quite possible though unlikely, the lack of positional control being reflected rather in a larger value for the error variation ; but since with long-lived plants positional variation is not usually as important as it is with annuals and since there is the certainty of more error degrees of freedom to be set against this possible loss, the argument is not conclusive. Indeed, when there are few treatments to be compared in a small trial, complete randomization is often very useful. In such instances, the increase in error variation can hardly be large because the total area of the trial is small anyway, while the gain in degrees of freedom is of especial value when plots are few. Thus, in the trial in Figure I, the value of $F$ would have to exceed 4·96 for the treatment differences to be judged significant ($P = 0{\cdot}05$), whereas this figure would have been 6·61 had randomized blocks been used instead.

Among other advantages of complete randomization may be mentioned the ease with which measurement may be made of the variation between plots having the same treatment, a matter that is often of great importance as Section 73 will show. When there are no blocks to complicate the issue, it is necessary only to work out the standard error of the plot-to-plot variation for each treatment separately to see if they are all about the same. It will also be noticed that the analysis of results is not made much more difficult by some treatments being represented more frequently than others, though certainly equal replication is the ideal. If the reader does not see the method at once, he can derive it from Appendix I.

For those who are not prepared to risk all plots of one treatment coming together, but who see the advantage of conserving error degrees of freedom, a good compromise is to use blocks fewer in number than the replicates, as has been done in Figure II. Here the experimental area has been split into two blocks each containing three plots of A chosen at random and three of B. The value of $F$ needed for significance is now 5·12, there being one degree of freedom for blocks, one for treatments and nine for error, making eleven in all, so the loss of sensitivity as compared with complete randomization is not serious. The design is of the so-called "orthogonal" type, so the method of analysis given in Appendix I is suitable.

| | | | |
|---|---|---|---|
| A | A | B | A |
| B | B | B | A |
| A | B | A | B |

FIGURE II.—Design in which the number of replicates exceeds the number of blocks.

## 21. Randomized Blocks

To design an experiment in randomized blocks, the total area is divided into as many blocks as there are to be replicates and each block is divided into as many plots as there are to be treatments. Within each block the treatments are assigned to the plots at random, each treatment occurring once and only once in each block, as in Figure III. This also is a long-established method [63*a*]. It is not essential for the blocks to adjoin one another.

With perennial species this is quite the most usual design and it will be profitable to dwell upon its advantages, for these illustrate very well some fundamental points of trial design with long-lived plants. The advantages have also been discussed by Rigney [131].

First of all, it is an *orthogonal* design. The plots may be classified in two ways, by blocks and by treatments. By orthogonality is meant that if the plots are grouped according to their blocks, each group will be made up in the same way in respect of treatments (each occurring

| | | | | | |
|---|---|---|---|---|---|
| C | D | F | E | B | D |
| A | E | B | F | A | C |
| A | F | B | A | C | E |
| E | C | D | D | B | F |

FIGURE III.—Four randomized blocks of six treatments.

once in Figure III, but each occurring three times in Figure II).* Equally, if they are grouped according to treatments each group will be made up in the same way in respect of representation of blocks. In fact, the trials shown in Figures II and III may both be described as based on *orthogonal two-way classifications*. As a consequence, the simple method for analysing results set out in Appendix I may be used, and the significant differences (see Section 72) between treatment means will be a minimum for the degree of replication employed. In general, all computing is very much simplified.

The next thing to notice is that the design is *robust* [119]. In the course of years, accidents are bound to happen to a trial in a way that is unusual with one that lasts only a short time. Thus, some of the treatments may lead eventually to death or abnormality of the trees, while disease may cause havoc in one or more of the blocks and whole plots may be lost. Even if all the trees remain, some of the treatments may prove disappointing so that the experimenter is no longer interested in them. In such circumstances, the robustness of randomized blocks is very useful, for whole blocks and whole treatments may be omitted from the analysis of results without disturbing the basic design, which remains an orthogonal two-way classification. Also, the effect of missing plots is by no means serious (see Sections 83 and 84). Of course, the loss of data must diminish the sensitivity of the trial, but a wise experimenter draws up his design in the first place with something in hand knowing that losses are possible.

A third good property is *flexibility* [119], that is to say, the potentiality of being available for a further set of treatments when the original set is no longer of interest. A fuller discussion of the subject will be left until Section 32, but it may be remarked here that randomized blocks are perhaps the most flexible of all designs.

Against these advantages it can rightly be urged that a trial in randomized blocks, when treatments are numerous, gives rise to a large experimental error on account of the positional variation being uncontrolled within the large area of a block. For this reason, many designs have been suggested in which each block shall contain only some of the treatments and so be smaller. Some of these will be discussed later in Sections 31 and 35; but since with perennial plants much of the variation is inherent in the trees themselves, too much attention can easily be given to the control of positional variation when this is rarely of paramount importance. Nevertheless, large blocks are to be avoided if possible.

## 22. Latin Squares

In a Latin square the plots are disposed in rows and columns and the treatments are applied so that each occurs once and only once in each row and in each column. For this to be possible, the rows, columns and treatments must all be equally numerous. Figure IV will serve as an illustration. Latin squares have been studied since the 18th century [55] and have long been used as an experimental design [63*b*].

It will readily be seen that the design is based on an orthogonal three-way classification of rows, columns and treatments. Thus all the columns, the plots of each being considered as a group, are constituted alike in respect of the representation of rows and treatments. Similarly, the rows and treatments are each constituted alike in respect of the remaining classifications. It follows that the method for analysing results set out in Appendix I may validly be used.

Latin squares have one notable advantage over randomized blocks in often permitting the trees on the outside of the experimental area to be used equally with those inside. Suppose that the four faces of the trial are quite different, one perhaps facing a wood, another being exposed to the prevailing wind without shelter and so on. These differences are associated with rows and columns, the effects of which are to be eliminated anyway. Consequently there is no need to waste land with external guard rows as would be necessary with most other designs. It is assumed only that each headland is homogeneous, *i.e.*, the wood or whatever it may be

* It should here be explained that no satisfactory definition of orthogonality can be given except in terms of the ideas set out in Appendix III. Strictly an orthogonal design is one that gives rise to such a set of parametric equations that, when once the general parameter is known and given a suitable choice of the equations of constraint, each parametric value is given separately by a single equation.

extends the whole length of the headland, and that the effect on a corner tree is the sum of the effects of the two headlands meeting there.

| | | | | |
|---|---|---|---|---|
| D | B | E | C | A |
| A | D | C | B | E |
| B | E | D | A | C |
| E | C | A | D | B |
| C | A | B | E | D |

FIGURE IV.—Example of a $5 \times 5$ Latin square.

Latin squares are fairly robust. Thus, the loss of data can be dealt with quite easily (see Section 83), as can the loss of a complete row, column or treatment. The latter case has been studied in some detail [184], the method of analysing results being also derivable from Appendix III.

Latin squares are more flexible than is sometimes supposed, but this will be considered in the next Section. They do, however, suffer from the disadvantage of being useful for only a limited number of treatments, say four to seven. Smaller squares do not give enough error degrees of freedom, though a $3 \times 3$ square used in duplicate or triplicate (see Section 24) often provides a good design; while in larger squares the control of positional effects is attempted by rows and columns that are so long and narrow as to be virtually ineffective. The plots themselves need not be square. The chief restriction upon the use of Latin squares is the requirement that rows, columns and treatments must all be equal in number.

## 23. The Changing of Treatments in a Latin Square

It should first be emphasized that a change of treatments in the way to be described is permissible only where there is no likelihood of the two sets of treatments interacting. Thus, if the original treatments have left some trees growing well and others nearly stationary in growth, it is not to be expected that they will all react alike to a subsequent set of pruning treatments and it would be most unwise to effect such a change of purpose. Of course, if a period were to elapse during which the trees were treated alike or if efforts had been made to even them up by reversing treatments, the position might become very different. On the other hand, the fact that the original treatments might be expected to leave a permanent residual effect *independent of the treatment following* would be of no importance because the differences would be eliminated in the analysis of variance.

This subject has been studied exhaustively for squares of size $4 \times 4$, $5 \times 5$ and $6 \times 6$ with conclusions important for all who design trials with long-lived plants. Before these results can be presented, however, it is necessary to explain what is meant by a "transformation set".

To take an example, all $4 \times 4$ squares can be generated from one or other of these two squares,

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| I. | A | B | C | D | II. | A | B | C | D |
| | B | A | D | C | | B | A | D | C |
| | C | D | B | A | | C | D | A | B |
| | D | C | A | B | | D | C | B | A, |

by permutation of rows, columns and letters. Now it happens that squares generated from the second, *i.e.*, those of Transformation Set II, are much more flexible than the others [57]. Further, restriction of randomization to secure a square from a chosen transformation set does not bias the tests of significance [64*b*]. Consequently, an experimenter with perennial plants who foresees the possibility, however remote it may be, of losing interest in his original set of treatments and wanting to add a new set is advised to abandon the usual method of selecting Latin squares [66*e*]. Instead he should take Square II, permute its rows at random, then permute at random the columns of the resulting square and then assign the letters at random to his four treatments.

Supposing he does so, what alternatives has he if he does want to change ? Writing small letters for the new treatments and capitals for the original ones, the most interesting possibility is a Graeco-Latin square, thus :

| | | | |
|---|---|---|---|
| Aa | Bb | Cc | Dd |
| Bd | Ac | Db | Ca |
| Cb | Da | Ad | Bc |
| Dc | Cd | Ba | Ab. |

Each new treatment occurs *once* in each row, in each column and upon each original treatment. There are *four* such treatments, so the "partition" will be said to be of the type $(1^4)$. If the scheme had been

| | | | |
|---|---|---|---|
| Aa | Ba | Cc | Dd |
| Bc | Ad | Da | Ca |
| Cd | Dc | Aa | Ba |
| Da | Ca | Bd | Ac, |

*two* new treatments (c and d) occur *once* on each of the original classes and *one* (a) occurs *twice*. The partition is thus of the type $(1^2, 2^1)$, usually written $(1^2, 2)$.

The $4 \times 4$ squares of Transformation Set II permit partitions of all possible types, namely, $(1^4)$, $(1^2, 2)$, $(2^2)$ and $(1, 3)$.

Similarly, if $5 \times 5$ squares are chosen from those of Transformation Set II, *i.e.*, by permutation of rows, columns and letters starting from

| | | | | |
|---|---|---|---|---|
| A | B | C | D | E |
| B | C | E | A | D |
| C | E | D | B | A |
| D | A | B | E | C |
| E | D | A | C | B, |

they will permit of all possible types of partition [58] namely, $(1^5)$, $(1^3, 2)$, $(1^2, 3)$, $(1, 2^2)$, $(2, 3)$ and $(1, 4)$.

Unfortunately there is no $6 \times 6$ Latin square that will permit of a partition of type $(1^6)$, the most useful kind [65]. However, those of Transformation Set X* permit of all the other possibilities [57, 59, 60], namely, $(1^4, 2)$, $(1^3, 3)$, $(1^2, 2^2)$, $(1^2, 4)$, $(1, 2, 3)$, $(1, 5)$, $(2^3)$, $(2, 4)$ and $(3^2)$. Such squares are those generated by permutation of rows, columns and letters, starting from

| | | | | | |
|---|---|---|---|---|---|
| A | B | C | D | E | F |
| B | C | F | A | D | E |
| C | F | B | E | A | D |
| D | A | E | B | F | C |
| E | D | A | F | C | B |
| F | E | D | C | B | A. |

Apart from the work of Norton [101] on the $7 \times 7$ squares, there is little known about larger squares than those considered here. However, Fisher and Yates [66*f*] cite "orthogonal squares" of size $8 \times 8$ and $9 \times 9$, which may be used as a starting-point for randomization to give transformation sets capable of all possible partitions including Graeco-Latin squares. A $7 \times 7$ square giving a transformation set of maximum flexibility is the following:

| | | | | | | |
|---|---|---|---|---|---|---|
| A | B | C | D | E | F | G |
| B | F | E | G | C | A | D |
| C | D | A | E | B | G | F |
| D | C | G | A | F | E | B |
| E | G | B | F | A | D | C |
| F | A | D | C | G | B | E |
| G | E | F | B | D | C | A. |

All the designs considered here are derived from orthogonal four-way classifications (rows, columns, original treatments and new treatments). Consequently they cause little difficulty in the analysis of data, the method given in Appendix I being sufficient. Simplest of all are the Graeco-Latin squares, *i.e.*, designs of the type $(1^n)$, in which all the new treatments are equally replicated.

The randomization of these designs, if it is to be done correctly, requires some thought. If a Graeco-Latin square, to take an example, is written down and then its rows and columns are permuted at random, the two sets of treatments being allocated at random to their respective sets of letters, the resulting analysis of variance gives unbiased tests in respect of both original and new treatments. The problem here is to convert an existing Latin square into a Graeco-Latin square or similar design. Merely to fill in the letters for the new treatments as well as may be is not enough, because it might lead to an unconscious grouping or dispersal of them over the experimental area. Instead, it would seem better to return to the basic square from

* Strictly there are 22 transformation sets of $6 \times 6$ squares, but, as a matter of convenience, conjugates are here ignored. Numbering thus follows that given by Fisher and Yates [65, 66*e*].

which the existing trial had been derived and to add the new treatments to that. Then it only remains to re-enact the permutations previously gone through to produce again the original design but this time with a further classification added, finally allocating the new treatments at random to the letters representing them. In order to be able to do this conveniently, it is a good plan to keep a note how each Latin square was derived from the basic square.

## 24. Multiple Latin Squares, Tied or Otherwise

By a multiple Latin square is meant two or more Latin squares, not necessarily adjacent, having the same treatments and analysed together. Sometimes it is possible for them to stand side by side and for the rows to run through, and it is here suggested that this state of affairs

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| D | A | C | B | C | D | B | A |
| B | C | A | D | D | C | A | B |
| C | B | D | A | B | A | C | D |
| A | D | B | C | A | B | D | C |

FIGURE V.—Double $4 \times 4$ Latin square (no tying).

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| B | B | D | A | C | C | A | D |
| A | D | B | B | D | A | C | C |
| C | A | C | D | B | D | B | A |
| D | C | A | C | A | B | D | B |

FIGURE VI.—Double $4 \times 4$ Latin square (tied).

is well described by saying they are " tied ". Thus in Figure VI the squares are tied but not in Figure V, the difference being that the columns are all randomized together in the one case but in two sets of four in the other. It will be noticed that each Latin square is of Transformation Set II, as recommended in the last Section.

In the main, these designs have the same advantages and disadvantages as simple Latin squares, the important difference being the greater number of replicates. As to tying, if the squares do not lie side by side, this is impossible ; but generally it is advantageous on account of the additional degrees of freedom for error. Such designs are orthogonal and the method of analysis of Appendix I applies without modification when the squares are tied. In the other case, analysis is rather more tricky, because the columns are randomized within the squares, which are only part of the whole experimental area. The correct analysis (for $J$ untied $K \times K$ squares) is set out thus :

| *Source of variation* | *Degrees of freedom* |
|---|---|
| Squares . . . . . . . . | $J - 1$ |
| Rows within squares . . . . . | $J(K - 1)$ |
| Columns within squares . . . . | $J(K - 1)$ |
| Treatments . . . . . . . | $K - 1$ |
| Interaction, Treatments × Squares . . | $(J - 1)(K - 1)$ |
| Error, by difference . . . . . | $J(K - 1)(K - 2)$ |
| Total . . . . . . . | $JK^2 - 1$ |

Two points to notice are these : The rows and columns are compared within squares, so their sums of squares are obtained by subtracting the squares term (not the correction term) from the rows or columns term (see Appendix I). Also, if the squares are close together, it is most unlikely that squares and treatments will interact (*i.e.*, the treatments can be expected to act relative to one another in the same way in all squares) and the interaction can be merged with error. Incidentally, the error sum of squares can be checked by working out the error for each Latin square separately. Their sum should equal the error obtained for the untied multiple square.

## 25. Split-plots

In essence, a split-plot design is one in which the plots of one experiment are used as the blocks of another [187*c*]. Thus, a cover crop trial might be laid down with four-tree plots and these might be used as blocks for a trial of four pruning treatments, one tree of each plot being assigned at random to each of the pruning treatments. This process of splitting can be continued almost indefinitely. Thus, the plots of the first trial having been split into sub-plots for a second set of treatments, these sub-plots, if large enough, can be further split into sub$^2$-plots for yet a third set of treatments and these sub$^2$-plots can be again split into sub$^3$-plots and so on.

With long-lived plants it is sometimes advisable to regard the various year's results not just as a series of figures to be added together to give an aggregate result over a period but as subjects of individual study. One of the best ways of doing this is to regard years as a further set of treatments split on to the smallest sort of plot laid out in the field. In doing this, however, it should be clear that the variability of the plant material is about the same in all years, otherwise difficulties will arise from the error being heterogeneous (see Section 73). The subject is discussed more fully in Section 74.

Split-plots are used for three main reasons. For one thing, they enable emphasis to be put where it is needed. Thus, an experimenter might want to test out a range of six orange scion varieties and might want to use a range of four contrasting rootstocks, not because he wants to study the rootstocks themselves—their general characteristics being well known—but because he wants to detect any interaction there may be of stock and scion. If he lays down a series of randomized blocks, each of 24 trees, one for each combination of stock and scion, his control of positional variation is likely to be poor on account of the large block size. Instead, he might prefer to divide blocks of 24 trees into four plots of six, one plot for each rootstock, and to divide each such plot into six one-tree sub-plots for the scion varieties. The varieties are now compared in one-tree units within areas of six trees, instead of areas of 24 trees as formerly, so their comparison is improved. The rootstocks, on the other hand, are applied to a certain number of six-tree plots instead of six times as many dispersed trees, so their comparison is worse. In fact, emphasis has been placed where it is wanted. Again, however, the warning must be given that with perennial plants there is a large component of variation not derived from position, so the emphasis that can be given is not as great as with some other species ; while, as will appear below, split-plot designs have certain technical disadvantages.

Another reason for using split-plots is convenience in the field. In Figure VII is shown the lay-out of a trial that illustrates this very well. It was part of an investigation at East Malling to discover a control measure for cane-spot in loganberries. Some thought that a

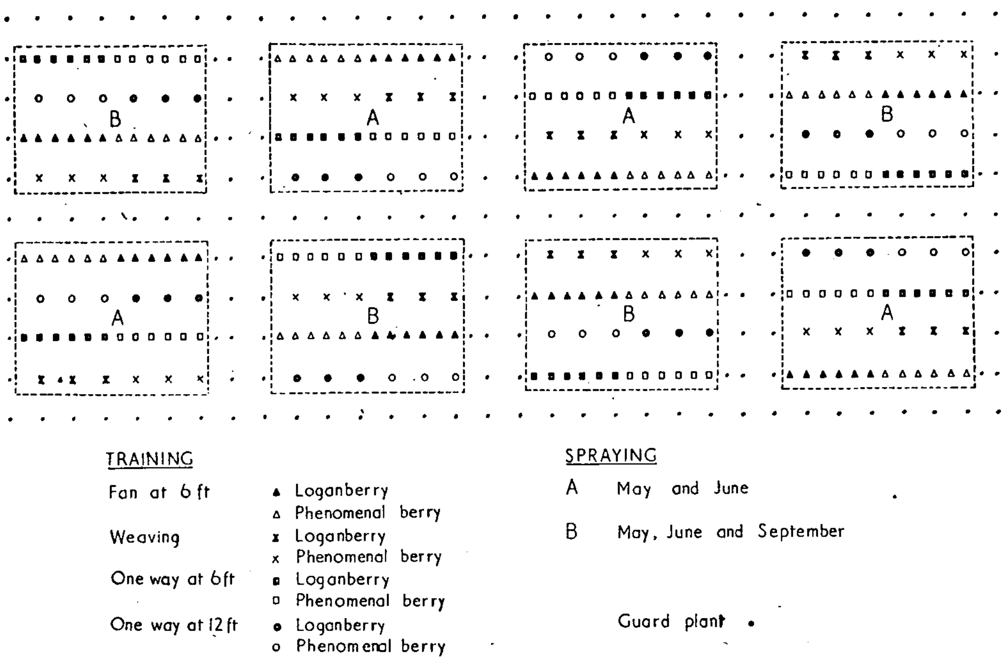



FIGURE VII.—Design with plots doubly split to facilitate field operations.

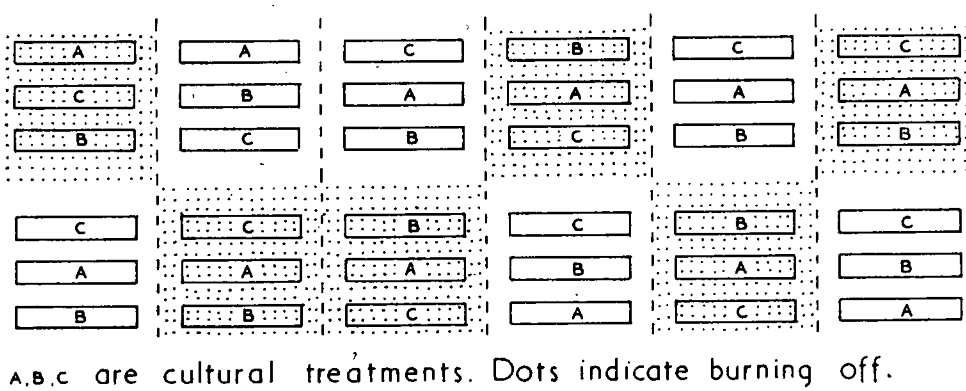



FIGURE VIII.—Split-plot design formed by addition of a further set of treatments.

remedy could be found by spraying and others by seeking a better training method, while others thought that Phenomenal berry might be much more resistant. Now, spraying must be done on large areas; training can be applied only to whole rows on account of the wire-work needed, but varieties can be planted as desired. Hence, it was decided to adopt a double splitting with sprays on the plots, training methods on the long, narrow sub-plots, and varieties on the sub²-plots, as shown in Figure VII. Incidentally, this trial was designed before the principle of randomization had been fully accepted. It is of interest to note that, if emphasis had been the deciding factor in designing the trial, the splitting would have gone the other way, with varieties on the plots, training methods on the sub-plots and spraying on the sub²-plots.

A third occasion for splitting plots is the addition of a new set of treatments, the original set remaining of interest. Thus, the East Malling strawberry trial illustrated in Figure VIII was originally intended to compare three cultural methods, A, B and C, but the question arose later whether or not to burn off the plants. This was investigated by grouping the blocks of cultural treatments into pairs and burning off one but not the other, thus converting the trial into one in split-plots, as shown. (Actually, there was a restriction of randomization here also—quite a valid one this time provided it is taken into account when analysing the data—since in each row of the original blocks, three were burnt and three not.) Another way to use split-plots for the addition of treatments is to divide the original plots instead of grouping the original blocks.

Analysing the data from split-plot designs can be awkward, though the basic method is nearly the same as that of Appendix I. To take the trial in Figure VIII as an example, the first thing is to work out the analysis for the plots, thus (plot values are the *sum* of the three sub-plot values, not their mean.):

| *Source of variation* | *Degrees of freedom* |
|---|---|
| Rows | 1 |
| Columns | 5 |
| Burning *v.* Not burning | 1 |
| Error i (by difference) | 4 |
| Total between plots | 11 |

This presents no difficulty at all, the one thing to remember being that all the sums of squares are one-third what they would have been in the absence of splitting, because each plot value is obtained by adding three observations. The error sum of squares could have been obtained independently by the formula given in Appendix I for orthogonal three-way classifications (rows, columns and burning treatments).

The analysis of the sub-plots also presents no difficulty. The "blocks" are the "plots" of the first analysis, so this line can be taken over and everything is as for a randomized block design except for an interaction of burning and cultural treatments, which in the absence of splitting would have been lacking. The second analysis reads therefore like this:

| *Source of variation* | *Degrees of freedom* |
|---|---|
| Blocks (taken over) | 11 |
| Cultural treatments | 2 |
| Interaction | 2 |
| Error ii (by difference) | 20 |
| Total between sub-plots | 35 |

The error can be checked in two ways. In one, a combined error is obtained by regarding the design as a whole as an orthogonal three-way classification* (rows, columns and combina-

* If the main plots had been in a two-way classification, this sub-plot classification also would have been two-way.

tions of the two kinds of treatment) and then subtracting Error i to leave Error ii. The other way is to think of Error ii as having two components, thus: A randomised block experiment has been carried out in six blocks, each of three treatments, on burnt plants to give an error with 10 degrees of freedom. A similar trial has been carried out on unburnt plants to give another error with 10 degrees of freedom. Addition of these two error components will give Error ii.

In designing trials with split-plots, it is important to arrange that each separate analysis has enough degrees of freedom in the error (see Section 54). Certain special difficulties in the analysis of split-plot designs will be examined later in Sections 71 and 72. In the event of data being incomplete or otherwise unsatisfactory, the methods to be described in Sections 83 and 85 can be applied more simply if it is remembered that the error of the second analysis is easily divisible into parts, as explained at the end of the last paragraph. Thus, for example, if an observation is missing, a value can be fitted to minimize the error component affected without regard to the other components constituting the Error ii.

### 26. The Criss-cross Design of Cochran and Cox

Sometimes it is a help if treatments can be applied in strips terminating at a headland. This is especially true of cultivation treatments, which frequently cause difficulty if there is nowhere for the tractor and implements to turn. A design has recently been suggested [35*b*], the criss-cross design of Cochran and Cox, which appears to the author very well suited to the comparison of two sets of such treatments in one experiment, though he has not yet had an opportunity to use it in practice.

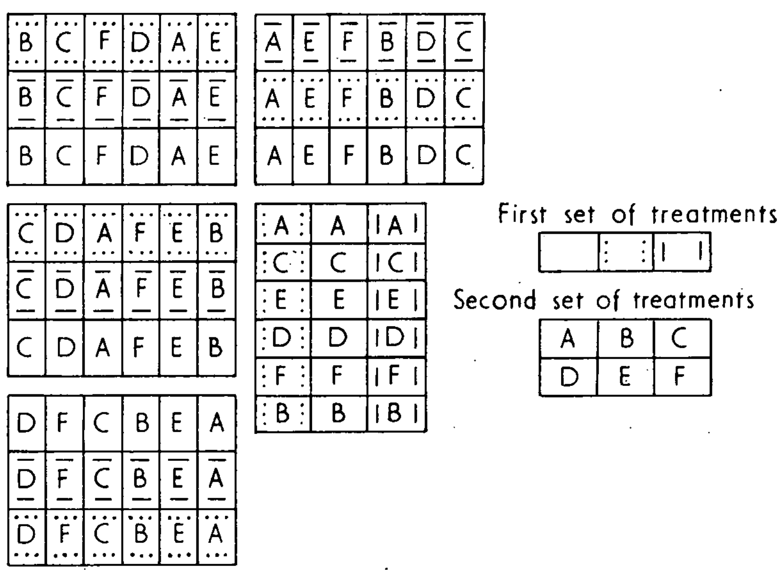



FIGURE IX.—Criss-cross design of Cochran and Cox.

In this design, illustrated in Figure IX, the experimental area is split up into a number of rectangles, $J$ in number, which are not necessarily oriented alike and not necessarily contiguous, so alleys may be left between them if need be. Suppose that there are $K$ treatments in one set and $L$ in the other, then each rectangle must be divided into $KL$ plots, $K$ by $L$, so that one set of treatments can be randomized on to the columns and the other on to the rows of the rectangles.

The analysis of data from such a design is quite easy. There are in all $JKL$ plots, arising from all combinations of $J$ rectangles, $K$ "first treatments" and $L$ "second treatments". Consequently, the $(JKL - 1)$ degrees of freedom can be split up just as in a factorial group of treatments and the sum of squares likewise (see Appendix II), with the following result:

| *Source of variation* | *Degrees of freedom* |
|---|---|
| a. Rectangles . . . . . . . . . . . | $J - 1$ |
| b. Treatments I . . . . . . . . . . | $K - 1$ |
| c. Treatments II . . . . . . . . . . | $L - 1$ |
| d. Treatments I × Treatments II . . . . . . | $(K - 1)(L - 1)$ |
| e. Treatments I × Rectangles . . . . . . . | $(J - 1)(K - 1)$ |
| f. Treatments II × Rectangles . . . . . . . | $(J - 1)(L - 1)$ |
| g. Treatments I × Treatments II × Rectangles . . . | $(J - 1)(K - 1)(L - 1)$ |
| Total . . . . . . . . . . . | $JKL - 1$ |

Each of the sources of variation, b, c and d, may be tested for significance by using its interaction with rectangles as error, *i.e.*, by using e, f and g respectively. There are thus three errors in an experiment of this kind.

### 27. Factorial Designs

In all the preceding work except the last two Sections, it has been assumed that the treatments form a single series, say, of five varieties or of four pruning methods, but this is not always so. Thus, an experimenter might want to try out his pruning methods on a range of varieties, or he might want to investigate several manurial treatments on both clean cultivated soil and on cover crops.

In such instances it is highly desirable that all treatment combinations should be represented. Thus, if it is a question of which pruning method to use in conjunction with various cover crops, it might well be argued that a dwarfing method of pruning is not likely to be of any use with a vigorous cover crop that is itself likely to inhibit tree growth. Nevertheless, it is as well to include the combination, as otherwise the method of Appendix II is not available. This insistence upon a complete set of combinations can become a fetish; but, even so, instances where it may wisely be abated are not common.

A note on nomenclature may here be useful. If, for example, three levels of nitrogen are used in conjunction with three kinds of potash on two varieties of rootstock, there are three *factors* (nitrogen, potash and rootstocks), two of them at three *levels* and the other at two. This can be expressed as a design with $2.3^2$ or 18 *treatment combinations*.

The chief advantage of including two factors together in one trial lies in the opportunity so given to find out if they interact, *i.e.*, if the effect of one factor is the same at all levels of the other. Even if no interaction is likely, a factorial design is sometimes useful as a means of economy, especially when the levels of the factors are few. Thus, it may be proposed to compare spraying against a disease with no spraying, and for this 16 trees, say, will be needed. In another trial it is intended to compare two methods of injecting nutrient solutions and again 16 trees are set aside for use. If the nutrients are unlikely to affect the incidence of disease or the disease to modify the effects of the injection, it would be more economical to use four replicates of the four treatment combinations, thus using only one set of trees. Despite these advantages, factorial designs should not be adopted without good cause.

A method of analysing data from factorially designed trials is set out in Appendix II. The meanings to be ascribed to main effects and interactions have been discussed by Finney [62] and by Williams [179] for the less simple cases.

[From this point it will be assumed that the reader knows how to analyse data from any non-factorial design, a matter discussed in Appendix III.]

# CHAPTER 3

# SOME MORE COMPLICATED DESIGNS

" A mighty maze ! but not without a plan."

Alexander Pope, ' Essay on Man', i, 6.

## 30. Non-orthogonal Designs in General

The designs considered above have all been orthogonal, though not always in the sense of the rather restricted description given in Section 21. Before proceding further, a fuller definition may be attempted :

Two classifications are mutually orthogonal if the various groups of plots formed by one classification are composed of the same proportionate number of plots of the other. Where all pairs of classifications are mutually orthogonal, the whole design is said to be orthogonal.*

| | | | | | |
|---|---|---|---|---|---|
| A | C | C | C | B | B |
| B | C | B | A | C | C |
| C | C | C | B | C | C |
| B | A | B | C | A | B |

FIGURE X.—Example of an orthogonal design.

Thus in the trial shown in Figure X there are two classifications, namely, blocks and treatments, so if these are orthogonal so also is the whole design. First, it will be seen that each block is made up in the same way, one-sixth of its plots receiving Treatment A, one-third Treatment B and one-half Treatment C. Further, the plots of each treatment are all divided in the same way, one-quarter occurring in each block. The design is therefore orthogonal and may be analysed by the method given in Appendix I.

The basic advantage of orthogonality is simply the independent way in which the effects of the various classifications may be sorted out. Thus, in Figure X, if one of the blocks happens to be situated on bad soil, the same proportion of plots will be affected in each treatment, while if one of the treatments proves more effective than the others, normally all the blocks will be affected to about the same degree. With a non-orthogonal design (for example, that in Figure XI) this is not so. Thus, Treatment A in that plan does not occur in Block III, so if a block

* It should again be stated that the only satisfactory definition of orthogonality is in terms of the parametric equations (see footnote on page 9).

is on worse soil than the average some of the treatment means will need adjustment, but not others. Likewise, if a treatment proves different from the rest, some but not all of the block means will need to be adjusted.

| | | | | | | |
|---|---|---|---|---|---|---|
| A | F | G | B | E | A | D |
| G | A | F | E | F | B | B |
| C | D | C | D | C | E | A |
| E | G | B | G | D | F | C |
| I | II | III | IV | V | VI | VII |

FIGURE XI.—Example of a balanced incomplete block design.

In practice, non-orthogonality introduces a number of complications. For one thing, the analysis of variance must follow the general method of Appendix III and not the simplified method of Appendix I, though for many particular designs abbreviated methods of analysis have been devised. For another, treatment means require adjustment on account of the treatments not being distributed evenly over the other classifications. In factorial trials this can be especially awkward. One consequence of this adjustment is that the treatment means thus obtained are subject to more error than in an orthogonal design, because the adjusting quantity is derived from the data and therefore is itself imperfectly known. This need to adjust means leads to a further disadvantage, for it makes impossible any preliminary assessment of the results until the analysis has proceeded at least as far as the solution of the parametric equations. With an orthogonal design, on the other hand, a few simple means are all that is required. Another complication is that corrections by covariance, the fitting of missing plot values and the overcoming of mishaps generally is a more difficult business, yet with perennial plants the need for such techniques arises more commonly than with annuals. Nevertheless, in special circumstances, non-orthogonal designs can be useful, if not indispensable.

## 31. Balanced Incomplete Blocks

This useful design is of interest in showing clearly the principle of "balance" which is desirable whenever orthogonality is unattainable. Thus, the trial in Figure XI is laid out in seven blocks ($b = 7$) but these are "incomplete", only four ($k = 4$) treatments out of seven ($v = 7$) occurring in each. Examination will show that each treatment *occurs* an equal number of times in the trial. Thus treatment A occurs four times ($r = 4$), in Blocks I, II, VI and VII, as does Treatment B in Blocks III, IV, VI and VII. Also, each pair of treatments *concurs* an equal number of times in the trial. For example, Treatments A and B concur together in Blocks VI and VII, and Treatments A and C likewise concur twice, in Blocks I and VII ($\lambda = 2$). It is this equalizing both of occurrences and concurrences that is the essence of balancing. Possible designs have been listed by Fisher and Yates [66*g*]. It may be noted that $\lambda$ always equals $r(k-1)/(v-1)$ and $bk$ always equals $rv$.

As has been mentioned, non-orthogonal designs always lose something in accuracy on account of the adjustments needed to the treatment means. For balanced incomplete blocks, the efficiency is $v\lambda/rk$, or 0·875 in the present instance. This means that the four replicates of the trial in Figure XI would give as much information as ($4 \times 0{\cdot}875$) or $3\frac{1}{2}$ replicates in randomized blocks.

The usual reason given for accepting this loss of information is that it will be recouped by the smaller size of block diminishing uncontrolled positional variation, but this argument will not impress the experimenter with trees if he believes that a good part of his experimental error is due to his plant material rather than to position. With trees, however, the design is sometimes useful, because blocks of large plants cannot be adapted in shape and size to the experimental area as can blocks of small ones. Thus, in Figure XII, which represents an application of Figure XI to a trial for controlling red spider on apple trees, there is no way of keeping the chosen plot size and fitting in four compact blocks of seven plots for a randomized block design, especially when a large degree of guarding at the north end was insisted upon to keep the sprays off some vegetables to be harvested for human consumption soon after the application of treatments. On such occasions, when the block size is restricted, balanced incomplete blocks provide a useful resource.

One occasion for restriction of block size is the use of the guard rows of an existing trial for a subsidiary experiment. Thus, in Figure XIII it is supposed that plots of an extant experiment consist of four trees arranged in a row, separated by single guards. The guard trees indicated by × fall into groups of four, all the trees of a group being treated alike as far as the main trial is concerned, though each of them has a different treatment on its two sides. If the treatments of the proposed subsidiary trial are unlikely to interact with those already being applied, these groups of four trees make a natural system of blocks for the new experiment, but what if there are to be more than four treatments? In such a difficulty, balanced incomplete blocks may provide an excellent solution. Of course, no treatments should be applied to the guard rows that might disturb the original trial.

The method of analysis given in Appendix III may be used or, alternatively, the shorter method given by Yates [186]. Where the efficiency is low, "recovery of inter-block information", a process described also by Yates [188], may be worth while; but hardly when the efficiency exceeds 0·9, which usually it does.

In randomized blocks, $k = v$ and $b = r = \lambda$, making the efficiency 1·0, a result which is obviously correct. Although the word "randomized" does not appear in the name of the design, the allocation of treatments to plots within a block should be as rigorously at random as when blocks are complete. Also, since the blocks do not all contain the same treatments, it is advisable, having drawn up a basic scheme, to randomize the blocks before use, as has been done in deriving Figure XII from Figure XI.

## 32. Other Designs of Latin Square Type

In Section 22 it was explained how a Latin square could, as it were, consume its own headland effect by eliminating variation between rows and columns, but it was mentioned also that Latin squares themselves are not often used for this purpose because of the restriction that rows, columns and treatments must all be equal in number. Now that non-orthogonal designs are being considered, it is pertinent to ask if any of them have this same property.

The problem is, as a matter of fact, closely allied to the changing of treatments in a randomized block design. In both, the investigator finds himself with a two-way orthogonal classification, whether of rows and columns or of blocks and original treatments: in both, he wants to add a third classification so as to give optimum accuracy to the comparison of the new classes. As with the Latin squares (see Section 23), it is assumed that the new set of treatments is unlikely to interact with those originally applied. It may be a matter for surprise that randomized blocks, which are the simpler design, are so much more complicated to change over to a new investigation than are Latin squares, but it should be remembered that the rows, columns and treatments of a Latin square are necessarily all equal in number, but it is only coincidence if the blocks are as numerous as the treatments in a randomized block design. Where this does

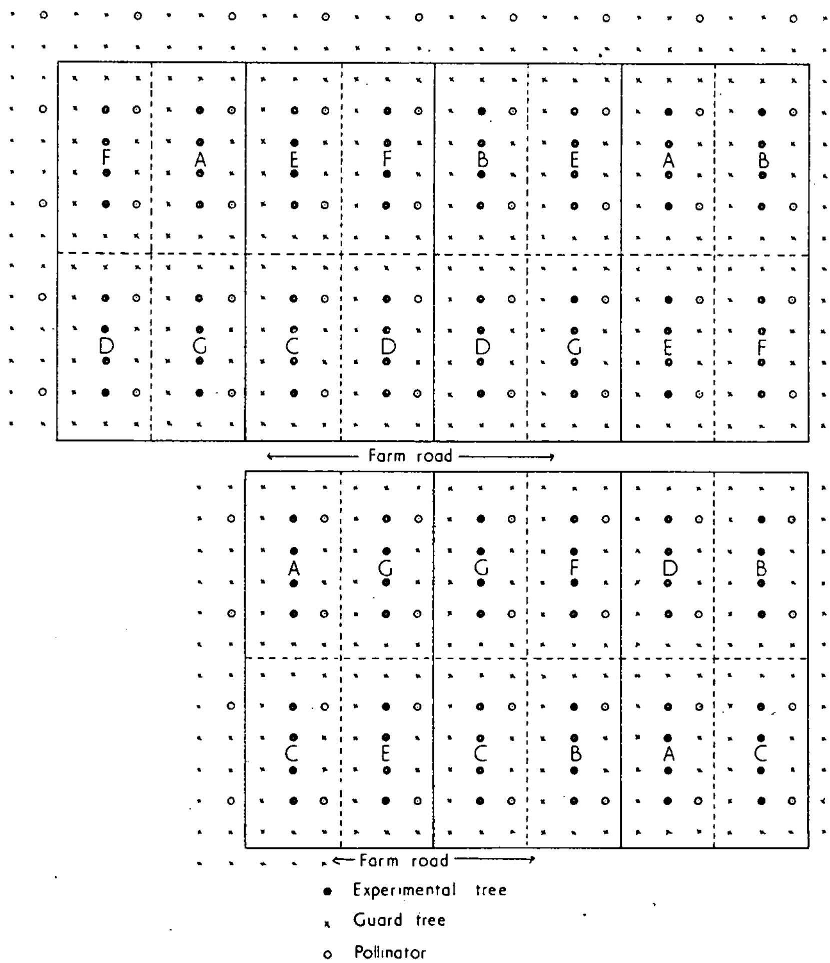



FIGURE XII.—Plan of spraying trial with seven treatments laid out in balanced incomplete blocks.

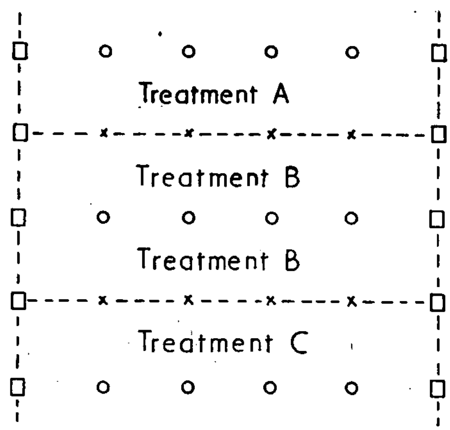



FIGURE XIII.—Use of guard trees for a subsidiary trial.

happen to be true, it is quite easy to effect a change, namely, by regarding blocks as rows, original treatments as columns and using a Latin square.

All the other designs to be suggested are non-orthogonal and in all the data can be analysed by the method of Appendix III. Where other references are given, they are to abbreviated methods suitable only to the particular design. Most of these references discuss the efficiency of the design they refer to.

One possibility is to use a Youden square [194]. In this design, the treatments are orthogonal to the rows and balanced in the columns, as in Figure XIV, which is only Figure XI

| | | | | | | |
|---|---|---|---|---|---|---|
| G | F | B | E | D | A | C |
| C | A | G | B | E | F | D |
| A | D | C | G | F | E | B |
| E | G | F | D | C | B | A |

FIGURE XIV.—Example of a Youden square.

| | | | | | |
|---|---|---|---|---|---|
| B | A | C | D | B | A |
| D | A | B | A | C | C |
| A | C | D | B | C | D |
| C | B | A | A | D | D |
| C | C | B | B | D | A |
| B | D | D | C | A | B |

FIGURE XV.—Example of a balanced three-way classification.

modified to ensure that each treatment occurs once in each row. Equally the treatments may be orthogonal to the columns and balanced in the rows, the difference being only one of nomenclature. Writing blocks for columns and original treatments for rows, this lay-out of Figure XIV would serve equally well for seven new treatments applied to seven blocks each of four original treatments. It would serve no less well if there had been four blocks each of seven original treatments.

Again, it is possible to use a balanced three-way classification [120], illustrated in Figure XV, in which each row and each column contains one, or in some trials more than one, complete replicate of the treatments and a certain number of additional treatments, which are balanced in both rows and columns.

Again, a Latin square can be taken and a Youden square added by its side, the columns being randomized together as in tied Latin squares [35c], a design illustrated in Figure XVI.

| | | | | | | |
|---|---|---|---|---|---|---|
| B | B | A | C | D | C | D |
| C | A | B | D | C | B | A |
| D | C | C | A | A | D | B |
| A | D | D | B | B | A | C |
| L | Y | L | L | Y | Y | L |

Columns marked L form a Latin square,
those marked Y a Youden square.

FIGURE XVI.—Example of a Latin square and a Youden square tied together.

For that matter, it would be quite permissible to place several Latin squares and one Youden square side by side and to tie them all by randomization of columns in one group.

Other possibilities are the Latin square with a row and column added or a column added and a row missing [118], the added rows and columns each containing a complete replicate of treatments in random order. In the former of these designs, which are illustrated respectively in Figures XVII and XVIII, it is necessary to select one of the treatments at random to occupy the plot where the added row and column cross. A similar design is the Latin square with a row and a column missing [184], of which an example is given in Figure XIX.

Attention has also been given to the Latin square with several rows and columns missing [192], but without discovery of any new designs of high utility.

Some other designs that have been suggested are special cases of those already mentioned, namely, the Latin square with a column missing [184], and tied multiple Latin squares with a column missing [35c]. Also, if it be remembered that a single replicate arranged as a column is strictly within the definition of a Youden square, there are the Latin square with a column added (see Figure XX) and tied multiple Latin squares with a column added.

Nothing has been said here about the degree of randomization needed with these lay-outs ; but to make certain, if there is any doubt in the experimenter's mind, it is as well to construct a lay-out conforming to the description given and then to permute its rows and columns at

X

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | C | D | B | E | A | E | |
| X | A | E | D | A | C | B | X |
| | D | E | A | A | B | C | |
| | A | C | D | C | E | B | |
| | B | A | E | D | C | D | |
| | E | B | C | B | D | A | |

X

X . . . X indicates additional row and column.

FIGURE XVII.—Example of a Latin square with a row and a column added.

X

| | | | | | |
|---|---|---|---|---|---|
| D | A | B | B | C | E |
| C | B | E | D | E | A |
| E | D | A | C | A | B |
| A | C | D | E | B | D |
| ~~B~~ | ~~E~~ | ~~C~~ | ~~A~~ | ~~D~~ | ~~C~~ |

X

X---X indicates additional column

FIGURE XVIII.—Example of a Latin square with a column added and a row omitted.

random and finally to allocate the letters at random to the treatments. Generally this will be "overdoing it", but it is impossible in these cases to randomize too much and disastrous to randomize too little.

| | | | | | |
|---|---|---|---|---|---|
| F | E | C | B | D | ~~A~~ |
| E | C | A | D | F | ~~B~~ |
| C | F | B | A | E | ~~D~~ |
| D | A | E | F | B | ~~C~~ |
| B | D | F | C | A | ~~E~~ |
| ~~A~~ | ~~B~~ | ~~D~~ | ~~E~~ | ~~C~~ | ~~F~~ |

FIGURE XIX.—Example of a Latin square with a row and a column omitted.

| | X | | | |
|---|---|---|---|---|
| C | C | B | D | A |
| D | A | A | C | B |
| A | B | D | B | C |
| B | D | C | A | D |
| | X | | | |

X . . . X additional column.

FIGURE XX.—Example of a Latin square with an additional column.

## 33. Substituted Plots

Before field trials of new varieties of fruit trees can be initiated, the varieties are usually multiplied as rapidly as possible from the original seedling. This process may be easier with one variety than another, so, rather than postpone the trial, it may be necessary to start with the varieties in unequal supply. Plainly it would be unwise to leave gaps in the plantation, because this would affect neighbouring trees and it would be a pity to fill in the gaps with trees of which no use is to be made. A good solution is to replace some of the plots of the variety in short supply by additional plots of a variety already in the trial, as has been done in the plan shown in Figure XXI, illustrating a design in randomized blocks with one substituted plot. The trial was for the purpose of comparing at East Malling some Australian cherry rootstock varieties. In this trial material was very limited and it would not have been reasonable to have reduced replication any further.

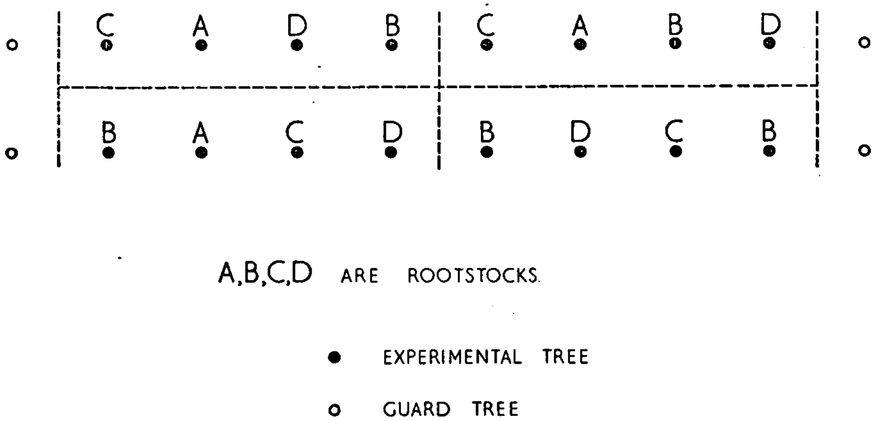



FIGURE XXI.—Example of a trial with a substituted plot.

The method of analysis has been given in full [113] for randomized blocks. No one appears to have studied the question specifically for Latin squares, but a general mathematical study of the subject has recently appeared [74]. In any event the method of Appendix III is always available.

### 34. Change-over Designs

With some treatments, such as manuring and pruning, the residual effect can be as important as the direct effect observed in the season of application and some device for investigating this is very desirable. It is odd to have to record an apparent neglect of the whole question among workers in horticulture, but in animal experimentation the position is different and mention will here be made of their methods, without going into much detail, in the hope that they will prove of value with plants.

The basic method is to use orthogonal Latin squares as set out by Fisher and Yates [66*f*]. Thus, suppose there are four treatments, three squares are needed thus :

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | B | C | D | : | A | B | C | D | : | A | B | C | D |
| B | A | D | C | : | C | D | A | B | : | D | C | B | A |
| C | D | A | B | : | D | C | B | A | : | B | A | D | C |
| D | C | B | A | : | B | A | D | C | : | C | D | A | B |

Let the columns represent twelve plots, arranged in three blocks of four, and let the rows represent four successive periods. It will be seen that each treatment eventually occurs on each plot and occurs an equal number of times in each period. Further, each succession of two treatments, *i.e.*, A followed by B, D followed by C, occurs an equal number of times in the trial.

Methods of analysing data will not be dealt with here, though these should be studied before embarking on the design. They have, however, been described in some detail elsewhere [34, 90, 107, 108, 178].

### 35. Confounding

Many useful accounts have been given of confounding [e.g., 61, 187*a*], the purpose of which is to obtain smaller blocks by sacrificing information about one or more of the effects and thus to improve the comparison of the rest. For example, given an N, P, K manurial trial with all factors at two levels, it might be thought worth while to sacrifice the three-factor interaction in order to reduce the number of plots in a block from eight to four and thus to improve the accuracy of comparison of the effects left, *i.e.*, the three main effects and the two-factor interactions. Obviously the value of this device depends upon the efficiency of reduced block size as a means of reducing error. Where it can be expected to work, the confounding may be advantageous. With trees, however, this cannot be relied upon.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| X | WXYZ | O | XY | XZ | WXY | WZ | Z |
| Z | WY | WYZ | XZ | W | WYZ | X | WXYZ |
| WX | WZ | W | WXY | WXZ | O | Y | XYZ |
| XYZ | Y | YZ | WXZ | YZ | XY | WY | WX |
| I | | II | | III | | IV | |

FIGURE XXII.—Example of a design with $2^4$ treatments in blocks of $2^3$ plots.

Further uses for confounding arise in the need to fit in sets of treatments to blocks of predetermined size (this is similar to the problem described in Section 31) and in carrying out trials with "tentative" treatments (see Section 60). It also enables a block to correspond with a day's work, for example in plucking tea or tapping rubber [53] or in picking or spraying fruit (see Section 50).

The most important case of confounding in practice is that when all factors are at two levels. Suppose, for example, that it is desired to carry out a manurial trial on the presence and absence of five elements, A, B, C, D and E, and suppose further that the interaction A $\times$ B $\times$ C is to be sacrificed in order to reduce the block size from 32 plots to 16. Then the elements can be considered in two groups—those involved in the confounded interaction (*i.e.*, A, B and C) and those not (*i.e.*, D and E). The first group can form in combination eight treatments in all, of which half involve the presence of an *odd* number of elements (*i.e.*, A, B, C, ABC) and the others the presence of an *even* number (*i.e.*, O, BC, AC, AB). The second group give rise to four treatment combinations (*i.e.*, O, D, E, DE). Combining the second group with the two halves of the first group will give two sets of 16 treatments, thus :

*1st Set* : A, AD, AE, ADE, B, BD, BE, BDE, C, CD, CE, CDE, ABC, ABCD, ABCE, ABCDE.

*2nd Set* : O, D, E, DE, BC, BCD, BCE, BCDE, AC, ACD, ACE, ACDE, AB, ABD, ABE, ABDE.

If the First Set are randomized on to the 16 plots of half the blocks, chosen at random, and the Second Set are randomized on to the 16 plots of the remaining blocks, a design will be produced in which the interaction, $A \times B \times C$, is " confounded ", *i.e.*, it has been sacrificed in the interests of block size. Of course, given five factors some less important interaction than that of $A \times B \times C$ would usually be chosen for confounding, but this example has been given to illustrate the method. For further illustration, Figure XXII shows a trial of four factors in which the interaction $X \times Y \times Z$ has been confounded between Blocks I and IV on the one hand and Blocks II and III on the other.

The analysis of data for such a trial is not really difficult, for the method of Appendix I may be used despite the design being, in the strict sense of the definition already given, non-orthogonal. The blocks require more degrees of freedom than in a randomized block design, because there are more of them, and the confounded interaction does not appear at all. Thus, the analysis of the trial in Figure XXII would be thus :

| *Source of variation.* | *Degrees of freedom.* |
|---|---|
| Blocks | 3 |
| W, X, Y and Z | 1 each |
| $W \times X$, $W \times Y$, $W \times Z$, $X \times Y$, $X \times Z$ and $Y \times Z$ | 1 each |
| $W \times X \times Y$, $W \times X \times Z$ and $W \times Y \times Z$ (*not* $X \times Y \times Z$) | 1 each |
| $W \times X \times Y \times Z$ | 1 |
| Error (by difference) | 14 |
| Total | 31 |

The error may be checked both as to its degrees of freedom and its sum of squares by working out the results as if the design were of two randomized blocks each of 16 plots and, from the error thus given, subtracting the new figure for blocks, adding the old figure for blocks and that for the confounded interaction. Thus, in the given example, if randomized blocks had been used, there would have been 15 degrees of freedom in the error. With the confounding this becomes $15 - 3 + 1 + 1 = 14$, which is correct, and similarly with the sums of squares.

By confounding three interactions the block size may be reduced to one-quarter of its original size. Thus, in Figure XXII, $X \times Y \times Z$ is confounded between Blocks I and IV as against II and III. If now an additional factor, V, is added so as to confound $V \times W \times X$ between I and II on the one hand and III and IV on the other, it will be found that $V \times W \times Y \times Z$ has become confounded with the third possible pairing of blocks, namely, I and III against II and IV. This interaction, $V \times W \times Y \times Z$, results from putting together the other two confounded interactions to give $X \times Y \times Z \times V \times W \times X$ and eliminating anything, in this case only X, that occurs twice.

A further device is that of " partial confounding " [187*b*] in which the blocks are in pairs with a different interaction confounded in each pair. When this is used, each partially confounded interaction is evaluated in the analysis of variance on the basis of the blocks in which it is not confounded.

If a main effect is confounded, the result is a split-plot design. Further confounding with split-plots has been discussed elsewhere [120, 187*c*].

## 36. Interlaced Blocks

When designing a trial it is often difficult to gauge how large the trees will grow, so some means of thinning is desirable to relieve possible overcrowding. Trials with many trees to a plot present no problem because half the trees can be removed without damage to the design, but with one-tree plots some special provision is called for, such as the interlacing of blocks.

This simply means that tree positions are divided into temporary and permanent, some blocks being confined to one and some to the other, a device that is legitimate where randomization in all blocks is unrestricted, as it should be. No modification of the statistical analysis is called for. Figure XXIII shows a peach variety trial in which this scheme was used. It is not available where plots need to be guarded.

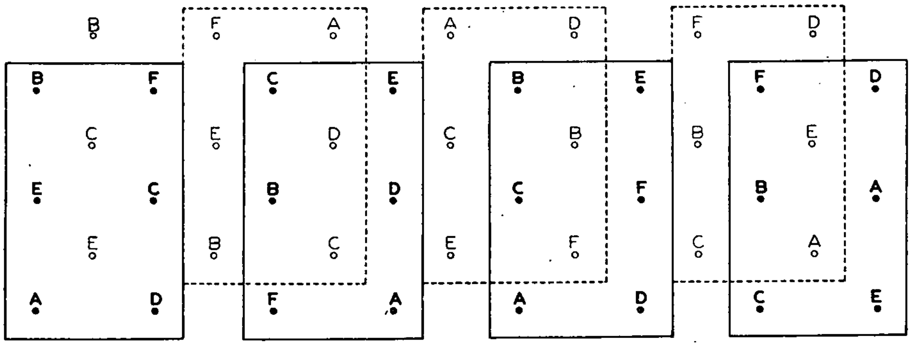
When the number of treatments is a multiple of four, it is often a useful device to form them into four groups designated O, A, B and AB and to take the blocks in pairs. On Blocks I and II, say, let the main effect A be confounded between temporary and permanent trees, the treatments applied to the temporary trees of Block I being assigned to the permanent trees of Block II. Similarly, the main effect of B might be confounded on Blocks III and IV and the interaction on V and VI. If there are yet more blocks, some of the effects can be repeated, while if there are fewer, some need not be confounded at all. Now, when the trial is first planted, the main effect of A will be partially confounded between Blocks IP and IT and between Blocks IIP and IIT. When thinning has taken place, the same effect is partially confounded between Blocks IP and IIP. An example with twelve treatments is presented in Figure XXIV, treatments 1–3 forming Group O; 4–6 A; 7–9, B, and 10–12, AB.

The same device is sometimes valuable in raspberry variety trials. On account of the tendency of some varieties to send up suckers, known in England as "spawn," some distance from the parent plant, such trials have to be kept under frequent observation so that contamination of varieties can be prevented; but the task is made much more difficult if the randomization places side by side or end to end plots of two varieties with indistinguishable spawn. To prevent this, it is enough to arrange that whenever two such varieties are to be compared, they shall always be in the same group. Then, by dividing the plots of each block into sub-blocks on the diagonals like the permanent and temporary trees of a plantation and by partially confounding as already described, no pair of indistinguishable varieties can ever occur on plots side by side or end to end. It is necessary, incidentally, to keep fairly wide cultivated alleys (usually across the rows) between the blocks, for otherwise there is nothing to prevent two adjacent plots in *different* blocks giving similar spawn. The whole subject has been discussed in some detail by Taylor [157]. Figure XXIV illustrates this use of partial confounding no less than the problem for which it was introduced, the only difference being that P and T now indicate the plots of sub-blocks and not permanent and temporary trees. It is not necessary, however, to confound each effect in two blocks, one being sufficient.

## 37. Some Disadvantages of Complex Designs

Some writers have recommended the use of more complicated designs to control positional variation, giving as their reason the great expense incurred anyway in conducting a long-term trial, and the consequent need to economize as far as possible in the number of trees to an experiment. The present writer appreciates the point of the argument, but believes that the case for complexity is by no means unassailable. Indeed, he is himself a whole-hearted supporter of simplicity.

In the first place, for the reasons given in Section 14, it is by no means certain that smaller blocks do in fact give a worth while reduction in experimental error with perennial species. Sometimes they do, but usually they do not. The expectation of success depends both upon the magnitude and nature of the positional effects to be controlled and also upon the extent of the variation inherent in the plant material. Thus, an experimenter working with variable non-clonal material like coconuts or oil-palms probably has most of his variation coming from the plants, and even a complete elimination of the relatively small residuum coming from position in the field might benefit him very little.



A, B, C, D, E, F REPRESENT VARIETIES

• PERMANENT TREES PLANTED 23 FEET BY 21

∘ TEMPORARY TREES AT CENTRES OF RESULTING RECTANGLES

—— BOUNDARIES OF PERMANENT BLOCKS ------- BOUNDARIES OF TEMPORARY BLOCKS

FIGURE XXIII.—Example of a trial with permanent and temporary trees in interlaced blocks.

| I | | II | | III | | IV | |
|---|---|---|---|---|---|---|---|
| 1 P | 5 T | 4 P | 3 T | 11 P | 4 T | 4 P | 10 T |
| 6 T | 8 P | 9 T | 12 P | 9 T | 3 P | 3 T | 6 P |
| 9 P | 11 T | 11 P | 2 T | 10 P | 6 T | 7 P | 2 T |
| 12 T | 2 P | 7 T | 6 P | 5 T | 12 P | 11 T | 9 P |
| 3 P | 10 T | 5 P | 8 T | 1 P | 7 T | 5 P | 12 T |
| 4 T | 7 P | 1 T | 10 P | 8 T | 2 P | 1 T | 8 P |

P and T indicate permanent and temporary trees

FIGURE XXIV.—Interlaced blocks with partial confounding to permit thinning of a trial with one-tree plots.

For another, in a long-term trial mishaps are much more likely than in a trial with annuals. Whole treatments may fail, some of the plots will almost certainly be lost and corrections by covariance are common; so that even designs simple in the first place become complicated, while trials initially complex become ultimately unmanageable. In addition, trees often outlive the investigation for which they were planted and it is a great advantage to be able to use them for a further and quite different investigation.

Nevertheless, these more complex designs have their place. Thus, it does sometimes happen that a better control of positional variation is to be expected from using smaller blocks, and the trial is unlikely to last long enough for many mishaps to occur. Figure XII illustrates such a trial. Again, after a set of trees has fulfilled its original purpose, a further and more complicated trial is sometimes applied to it, using the methods of Sections 23 or 32. This makes for complexity, but with established trees many of the mishaps will already have occurred, so the investigator may feel that further deaths are unlikely. Decisions of this sort, naturally, depend upon the species and its principal diseases. With some, such as plum and cherry, losses where they occur mostly come soon after planting. Further, with older trees, the positional effects will have been operating for some time and their elimination is all the more likely to be worth while. In general, complicated designs are of greatest value in short-term trials on established trees.

In fairness, it should be recorded that at least a few writers speak in encouraging terms of their experiences with complicated designs for perennial species. One, for example, used a $5 \times 5$ lattice square with strawberries [131], another a lattice design with sugar cane [88]. It is true that neither of these was a long-term project, but a trial of irrigation and soil management of citrus using plaid squares has been described [68] and is plainly intended to go on for some time. Further, it has given useful results [69, 70].

## 38. Nonce-designs

Every experimenter has had his occasional bright idea for conducting an experiment and has thought about using some scheme that certainly did not come out of any text-book. When people use words made up on the spur of the moment to meet an unusual situation, etymologists

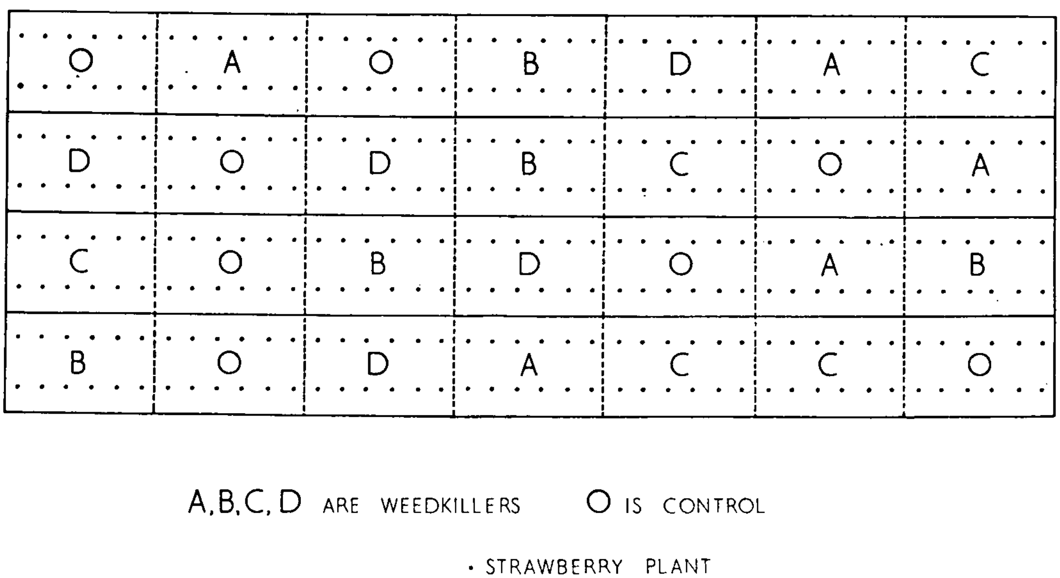



FIGURE XXV.—Example of a non-orthogonal design with unequal replication, blocks arranged horizontally.

speak of " nonce-words " ; equally one can speak of " nonce-designs," which are used once for a particular occasion with no idea of their being used again or generalized into anything wider.

Thus, a few years ago a trial was designed at East Malling in which there were to be four randomized blocks, each of seven plots, for the comparison of five weed-killers against an untreated control, which was to have two plots in each block. At the last moment, when all the plots had been pegged out, one of the weed-killers was held up in transport and the design in Figure XXV was substituted. It belongs apparently to no recognized system of lay-out, unless indeed it is regarded as a particular member of the class of design considered by Rao [129], but anyone who cares to study it will find that it is, for its particular purpose at any rate, an admirable design. Incidentally, the long narrow blocks were chosen on account of the initial distribution of weeds.

Nevertheless, the invention of new lay-outs is a tricky matter and most nonce-designs prove, on mathematical examination, to have some defect. The commonest faults are undue restriction of randomization leading to bias in the error or an inability to divide up the treatment degrees of freedom in any satisfactory manner, but there are others. An experimenter without mathematical training is well advised to keep to designs vouched for by those able to judge. To take an example, a design known as the semi-Latin square* has been " invented " again and again, among others by the present writer, though it is known to give a biased error [183]. The fact that a means has recently been devised for setting limits to the effect of bias [92] does not really alter the point.

This is not to say that there are no new designs to be discovered. Most new systems of lay-out start as nonce-designs and it is later realized that the idea is worth repeating. Even so, the introduction of a new design should not take place until the soundness of the mathematical basis is clearly established, and this is usually beyond the powers of the ordinary experimenter.

* By a " semi-Latin square " is meant a lay-out in which there are twice as many columns as rows, each treatment occurring once in each row and once in each pair of columns, *e.g.*,

| | | | | | |
|---|---|---|---|---|---|
| A | B | C | D | E | F |
| E | D | A | F | B | C |
| C | F | E | B | D | A. |

[From this point it will be assumed that the reader is conversant with the analysis of covariance, at least for orthogonal designs. If he is not, he is advised to read Appendix IV.]

# CHAPTER 4

# CALIBRATION

"I do not know whether there is anything peculiarly exciting in the air of this particular part of Hertfordshire, but the number of engagements that go on seems to me considerably above the proper average that statistics have laid down for our guidance."

Oscar Wilde. 'The Importance of Being Earnest.' Act III. (Lady Bracknell.

## 40. The Idea of Calibration

As has already been emphasized, the variation of perennial plants comes from two main sources, position in the field and factors inherent in the plants as individuals. The last two chapters have been concerned with the allocation of treatments to plots and thus with the control of positional effects, but is there nothing to be done about the second source of variation ?

If some trees are inherently better croppers or more vigorous than others, the fact, if it is to be of use, should be ascertained before applying treatments. After the event it is no good trying to explain treatment differences by suggesting that a certain treatment happened to be on better trees, for that is to invade the province of the significance test ; but the position is different if there was prior reason for expecting one set of trees to do well. Then, if they are undistinguished, the treatment may be judged a poor one. It will be adjudged good only if the trees do excellently. That is to say, the investigator starts his trial with a forecast based on statistical arguments as a guide to what he can expect from his plots and he judges his actual figures in the light of these expected ones. The future performance of trees naturally depends upon future weather, so he cannot say "over the next four years I shall expect to get 300 lb. of nuts from this plot and 250 lb. from the next", but he can often say, "this plot will give 120% of the crop of the next", and that is good enough. Such forecasting has been termed *calibration* [121]. The idea is very old [23, 63c] and has had widespread application.

Where there has been a change of treatments (see Sections 23 and 32) it is sometimes possible to take useful calibrating measurements just before the new set of treatments is applied. Where the experiment is newly planted, some writers have thought that trees can be calibrated by previously growing some annual crop, such as wheat, on the experimental area and harvesting each of the future plots separately so as to gauge their individual fertilities. There may be something in the idea, but the present writer has yet to hear of an instance where it has worked.

To calibrate, the trees must be planted and left without differential treatments for a time, and this is often difficult if not impossible. Thus, in a variety trial the treatments are necessarily applied not merely at planting, but earlier while the plants are still in the nursery. With ungrafted material, indeed, the differential treatments have always been present. Some treatments, again, have to be applied early if they are to operate at all. Supposing that a preliminary period is possible, care is needed in applying a uniform treatment that will not interfere with the comparison of future differential ones. Thus, if trees are irrigated during the preliminary period their roots may well be nearer the surface than if they had been left unirrigated, and this may affect the response to subsequent treatments. For this reason, treatments during the calibrating period need to be chosen carefully so as to produce trees typical of the region, and interpretation of the subsequent data must be made in the light of the history of the trees.

It should perhaps be emphasized that calibration is not a cause of delay in securing experimental results, but rather a means of saving time. As will appear in Sections 43, 44 and 45, it is possible to lay out an orchard for calibration in such a way that the types of trial for which it can be used are scarcely restricted at all. It follows that, if a research station can maintain a supply of such calibrated trees, an experiment can be started at once without waiting for suitable trees to be raised and planted with a further delay waiting for the trees to come into bearing.

The calibrating measurements having been made, the best way of using them is by the analysis of covariance, a technique explained in Appendix IV.

## 41. Covariance and Some Possible Alternatives

While many would agree with the idea of giving trees a trial period in which to discover their potentialities, there does not appear to be the same measure of agreement about the use to be made of the calibrating measurements.

Some use their observations in order to form groups which are then used as blocks instead of trees chosen for proximity. This is open to two objections. First, such a grouping must always leave some initial variation within blocks, whereas the analysis of covariance can allow for all variation. Secondly, it leaves no means of removing positional variation and, though this is not always of first importance, it may as well be eliminated as far as possible.

Others do not modify their lay-out on the basis of their calibrating observations, but evade the analysis of covariance by assuming some relationship between performance in the two periods. For example, instead of working with crop during the experimental period adjusted by covariance on crop during the calibrating period, they work with the difference between the two, the assumption being that, without the application of treatments, each plot would have cropped the same in the two periods. Even if the relationship is exactly what it is assumed to be, nothing has been gained except one degree of freedom in the error. If the assumed relationship does not hold, the results can be most misleading. The analysis of covariance has the great advantage of making a minimum of assumptions ; so, if the calibrating measurements turn out to be irrelevant, they have little effect on the conclusions.

If, as sometimes happens, it is not clear which of two alternative calibrating sets of measurements is to be preferred, it is necessary to compute one analysis of covariance with two independent variates. It is wrong to work out an analysis for each variate and then choose between them.

## 42. Choice of Calibrating Variate

Sometimes, of course, the choice is obvious. The trees in their preliminary period have, for example, succumbed to a disease that will affect their future performance. Plainly, some record of the disease is needed, to be used either alone or in conjunction with other measurements. Usually, the choice is not so easy and in such circumstances it is a good rule to adjust like by like, *i.e.*, if it is proposed to study crop in the trial, to measure crop in the calibrating period also and so on. Also, especially with species having a biennial tendency in growth and cropping, the results of a single year are not of much use, unless the biennial phenomena are so pronounced that the crop in the " on " year in effect represents a two-year period.

Where trunk girth can be measured, it sometimes sums up surprisingly well results that otherwise could be obtained, if at all, only by several years' observations [114, 162, 175]. Since it can be measured at any time, it is useful where uncalibrated trees are taken over and it is desired to start an experiment forthwith.

The numbers of shoots of blueberries and visual ratings in grape vines have been recommended as having the same property [132].

Generally, however, an experimenter has to base his decision on experience aided by botanical insight. It is not obvious that strawberry crops in the second year can be calibrated by weight of blossom in the first year, but this appears to be so [160]. Some empirical results are set out in Table I and there are doubtless many more to be found. Where two variates are suggested, the numbers (1), (2) etc. are used to indicate the order of preference.

## 43. Lay-out of Plantation for Calibration

When laying out a plantation for calibration it often happens that its ultimate use cannot be foreseen. Thus, it may eventually be used for a trial of soil management, which

TABLE I.—EXAMPLES OF VARIATES RECOMMENDED FOR CALIBRATION.

| *Species.* | *Reference.* | *Characteristic to be calibrated.* | *Recommended variate and other conclusions.* |
|---|---|---|---|
| Apples | Collison and Harlan [36] | Crop in five-year period | Crop in preceding five-year period. |
| | Pearce [114] | Crop in two-, four- and six-year periods | (1) Trunk girth at time of applying treatments. Area of cross-section about as good. (2) Total crop before application of treatments. |
| | Sudds and Anthony [155] | Crop (period unspecified) | Area of cross-section of trunk better than girth. |
| | Wilcox [175] | Crop in two-year period | (Trunk girth)$^{1.5}$. This is not strictly a calibration method because girth was taken in the middle of the cropping period. |
| | Yeager and Latimer [193] | Crop, principally in two-year periods | (1) Previous crop. (2) Trunk girth. |
| Apricots | Reed [130] | Crop in a single season | Girth increment during preceding season. |
| Cacao (regular spacing) | Cheesman and Pound [24] | Pod number in a single season | Previous crop. In general, the longer the calibrating period the better. A very large crop in the calibrating period is as unfortunate as a very small one. The adjustments are especially valuable with trees not all the same age. |
| Cacao (irregular spacing) | Pearce and Thom [123] | Pod number in two- and four-year periods | (1) Pod number in previous two-year period. Little advantage in taking four years. (2) Number of trees to a plot. |
| Cloves | Tidbury [162] | Crop in four-year period | Trunk girth multiplied by measure of canopy. Trunk girth was calculated from branch girths three feet above ground level. |
| Oil Palms | Webster [169] | Yield in one- and two-year periods | Previous yields. Biennial bearing can cause difficulty. |
| Oranges | Parker [104] | Crop in four-year period | Past crops (provided recent). Size records ineffective. |
| | Parker and Batchelor [105] | Crop in one-year period | Previous crops. Early crops not representative of future performance. |
| Pecans | Lutz [91] | Crop in twelve-year period from planting | Area of cross-section of trunk at planting. |
| | Sharpe and Blackmon [145] | Crop in single season | (1) Crop in previous five years. (2) Trunk girth. Area of cross-section very little different. |
| Rubber | Murray [96] | Yield in one-year period | Past yields (provided recent). |
| Strawberries | Taylor [160] | Crop in single season | (1) Double covariance on height and spread. (2) Weight of blossom in previous season if deblossomed. |
| Tea | Eden [52] | Yield in one-year period | Yield in previous year of same pruning cycle. |

(1), (2), etc., indicate order of preference.

will need large plots with guard trees, or it may be used for a pruning experiment, in which the treatments are likely to be applied to single unguarded trees. Since calibration takes some time, a research station is well advised to keep a stock of trees to hand, already calibrated, ready to be made use of experimentally ; but if this is done the design of these plantations must, to use a term recently introduced [121], be as *adaptable* as possible.

A little consideration will show that some numbers are in themselves more adaptable than others for the number of trees along one side of a rectangular trial. Suppose, for example, that a plantation has 285 trees, arranged 19 by 15. Assigning headland trees to form outside guards, the rest are arranged 17 by 13. Both these numbers being prime, allocation of the trees into blocks and plots is going to prove exceedingly difficult if waste is to be avoided ; but had the trees been arranged 20 by 14, to leave 18 by 12 after the exclusion of guards, the possible divisions are numerous. In fact, 20 and 14 are *numbers of high adaptability*, whereas 19 and 15 are of *low adaptability*. In designing a calibration trial it is often worth going to some lengths to obtain suitable numbers of trees. Thus, given an area 380 feet by 300 and trees usually planted 20 feet square, the natural thing would be to use the numbers of low adaptability referred to above ; but with a little thought it would appear better to plant 19 feet by 21 feet, so as to obtain the numbers of higher adaptability for the dimensions of the plantation. The whole question has been gone into at greater length elsewhere [121].

In designing a calibration trial it is often advisable to include a range of varieties. With some species this is in any case essential for pollination ; but, even where this consideration does not apply, a range does enable subsequent trials to be widely based in their conclusions. The varieties should be selected to represent distinct classes, as different from one another as possible but suitable for growing in the same plantation. Unless there is a strong reason for the contrary, the varieties chosen should be of commercial importance, but sometimes it is not possible to arrange this. Two special types of lay-out, "stripes" and "tiles", have been suggested for use in such experiments.

### 44. Stripe Designs

This lay-out has been described elsewhere [119, 121]. The varieties are planted in parallel lines down the larger dimension of the plantation, these lines being grouped into stripes, each stripe containing one row of each variety, the allocation of varieties to rows within a stripe being systematic and not at random. The stripes are separated by rows made up of varieties chosen so as to ensure good pollination of those in the experiment. A typical example is illustrated on the left-hand side of Figure XXVI, which illustrates ways of forming plots. Stripe designs have the disadvantage of wasting the pollinator rows if guarding is not called for, though the pollinators can readily be used as guard rows if any are needed. Stripe designs have the compensating advantage that the main varieties do not have to pollinate one another and may therefore be chosen freely.

In designing trials in stripes, it is important that the rows should contain trees to a number of high adaptability. Also, the number of stripes should, if possible, be a power of two or three or a product of these numbers, such as 2, 3, 4, 6, 8 or 9 rather than a number like 5, 7 or 10. The aim should be to obtain a useful number of plots whether guards are needed or not, and in general the best numbers are those of the form $2^m.3^n$ where neither $m$ nor $n$ is small. To take an extreme case, 35 plots can only be used for five replicates of seven treatments or seven replicates of five, while 32 plots can be used only for 2, 4 or 8 treatments in 16, 8 or 4 replications ; but 36 are available for 2, 3, 4, 6 or 9 treatments in 18, 12, 9, 6 or 4 replications—a much wider range. The whole question of designing in stripes has been discussed in greater detail elsewhere [121].

### 45. Tile Designs

In these designs, which have been described in some detail elsewhere [117, 121] the idea is to start off in one corner with a pattern of varieties, known as a "tile", and to repeat this pattern both along the rows and across them until the whole experimental area is filled. There

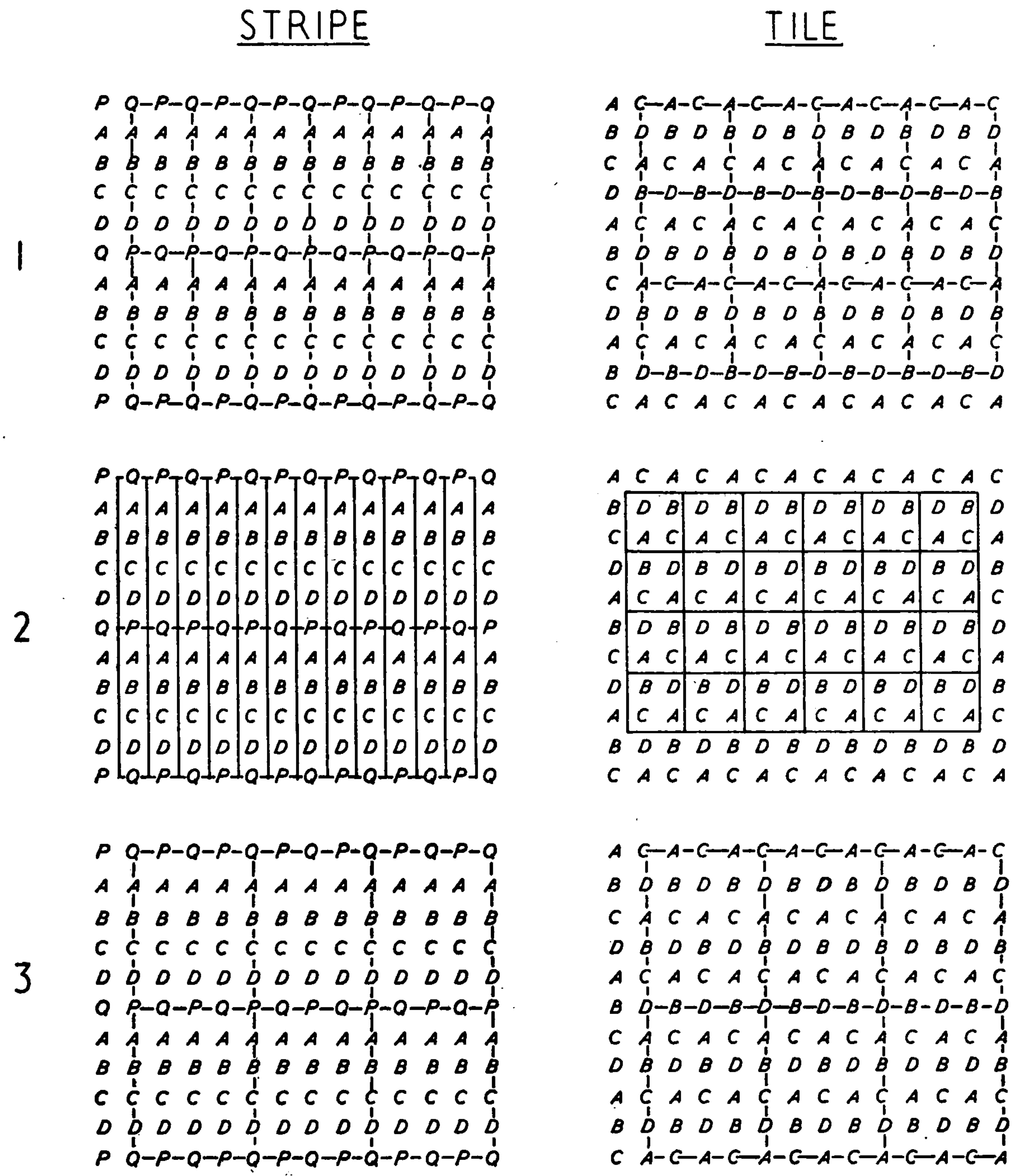



FIGURE XXVI.—Examples of calibration trials laid out in stripes and tiles.

is no objection to parts of tiles to fill in the edges. The designs are so called from the resemblance to laying a floor with tiles of a uniform pattern. An example is set out in the right-hand side of Figure XXVI, from which it will be seen that plots may be formed in several ways, whether guard rows are needed or not. Those shown in the top and middle diagrams are square, but plots four trees long are also possible if arranged vertically.

If there are only two varieties [121], the recommended tile is

A B

B A.

This simply places the two varieties on the diagonal.

If there are three varieties, it is possible to adopt a similar tile [117], thus

A B C

C A B

B C A.

This leads to a very adaptable design, though it is sometimes better to duplicate one variety and to use this [117] instead

A C B C

B C A C.

For four varieties [121], a good tile is

A B C D

C D A B,

the one illustrated in Figure XXVI. While for six [121] a possibility is this

A B C D E F

D E F A B C,

or this

A B C D E F

C D E F A B

E F A B C D.

The design of calibration trials in tiles calls for quite a lot of thought, and the references given will suggest some details of technique for the various types of tile. The basis is always the same, however. The number of trees to the longer headland should be adaptable and the number of plots arising should, if possible, be of the form $2^m.3^n$ whether guards are used or not.

### 46. The Analysis of Data from Stripe and Tile Designs

When the ultimate purpose has been decided and the treatments applied, the analysis of data is somewhat modified by the systematic arrangement of varieties. This need not cause much difficulty, however. Suppose for example, that the design in the top right-hand corner of Figure XXVI had been adopted and that four replications of three treatments had been applied, then the four varieties are on sub-plots each of one tree, split upon main plots, each of four trees. The analysis would ordinarily be like this:

| *Source* | *d.f.* |
|---|---|
| Blocks | 3 |
| Treatments | 2 |
| Error i | 6 |
| Total i | 11 |
| Varieties | 3 |
| Varieties × Treatments | 6 |
| Error ii | 27 |
| Total ii | 47 |

The first part of the analysis will be all right because the treatments will have been applied at random. The remedy for the systematic arrangement of varieties in the second part is to partition Error ii into two components, one vitiated by the lack of randomness and the other unvitiated, the latter being used for investigating the interaction. This may be done by regarding the 48 trees as forming a factorial design of blocks, treatments and varieties and thus working out by the method of Appendix II (a) the sum of squares with 9 degrees of freedom for the interaction of (blocks × varieties) and (b) the sum of squares with 18 degrees of freedom for the three-factor interaction of (blocks × treatments × varieties). It will be found that the sum of these quantities is Error ii. The first component is vitiated by the systematic disposition of varieties but not the other, so the second part of the analysis now reads

| *Source* | *d.f.* |
|---|---|
| Total i | 11 |
| Varieties* | 3 |
| Interaction | 6 |
| Error ii (a)* | 9 |
| (b) | 18 |
| Total ii | 47 |

All these lines need to be worked out to check that everything adds up correctly, but those marked with an asterisk are vitiated and must not be used for any test, the $F$-test for interaction taking place with 6 and 18 degrees of freedom. Because of the systematic arrangement no test for the comparison of varieties is possible, but this is rarely needed. These have been chosen as being of known characteristics and as widely different as possible and it is their interaction with the treatments that is of interest.

### 47. Trials on Commercial Plantations

One of the principal uses of calibration comes when designing a trial on a plantation that had originally been laid down commercially. The careful standardization of material and cultural treatments of an experimental plantation has usually been lacking and injured trees may not have been replanted. In such circumstances the prospect may well appear hopeless. There are, however, usually compensating features. For example, there is only rarely a shortage of trees, so the experimenter can concentrate on getting as much information as he can from each plot without regard to economy of plant material, that is to say, he can use few replicates of large plots (see Section 51) and it is in such circumstances that calibration is commonly most effective.

Of course, he will not be able to calibrate by means of previous performance because this will be unknown, and the grower is unlikely to be willing to co-operate in taking records for some future trial, even if the delay could be tolerated. Nevertheless he can make use of calibrating records like trunk girth, spread and shoot growth. Indeed, the differences between trees are likely to be so large that rough measurements of tree size, usually of little avail, prove quite effective in reducing variability.

Of course, such experimentation is by no means easy despite the help of calibration. There is usually a lack of trained helpers and their apparatus of weighing and recording implements, while distance can make economic use of labour impossible on account of travelling time. Above all, there is the difficulty of adequate supervision. The grower, however co-operative he may be —and some are very co-operative indeed—has almost certainly had no research experience and may do the most alarming things in the best of faith. Thus, while spraying his own trees he may spray the experimental ones as well and cut across the whole object of the experiment; nor is this danger confined to spraying but extends to all routine cultural operations. It follows that the experimenter and the grower must have a clear understanding, and this they will never have if the grower is not visited often for a chat and a look round.

One special source of difficulty centres on picking. A grower wants the best price he can get for his fruit so, when it is ripe or the market is especially favourable, he wants the fruit at once without waiting for careful picking of trees plot by plot followed by lengthy weighing. More trials on commercial holdings have failed on this account, probably, than on any other.

Allied to this difficulty is the question of compensation if some of the trees are rendered less fruitful than they would have been. Human nature being what it is, it would be surprising if grower and experimenter always saw eye to eye in this matter; but a lot of disputes would be avoided if designers would make a point of including a treatment to consist of those cultural operations the grower himself favours. This has several advantages. For one thing, it keeps the trial close to the local practices upon which it is hoped to improve: also, it provides a reliable test of the loss of crop occasioned by any treatment that raises a grievance in the owner of the trees and one that both sides can readily accept as a fair arbiter.

Finally, a grower who co-operates in an experiment must not only be visited often to see he does nothing wrong, but he must be treated as a collaborator and an ally. In particular, he will want to know what results are coming out of the trial and will appreciate a few figures even if they are hedged about with the reservations and caution commended in scientific circles. There is, after all, a certain prestige and reputation for enterprise associated with an experiment —but only if the grower has something to say when his friends ask about it.

It will be seen that this Section does not really fit into any chapter but the reason for it is worth noting. Experiments on outside plantations depend upon two pillars that are very diverse in nature; one is calibration and the other is tact.

### 48. The Method of Papadakis

In this method an expected performance of each plot is derived from that of its neighbours. Thus, if a plot is in the midst of others that are growing poorly its own poor growth is unremarkable, but good growth would strongly suggest that the treatment given to it had been especially effective. The method thus has affinities to calibration in that the performance of each plot is judged in relation to an expected figure, but there are two important differences. For one thing, no preliminary period is called for: for another, the purpose is to control positional effects, not the variation inherent in the plant material.

The method as described here is more general than the original form given by Papadakis [103]. Bartlett [9] has established that the method will not give results seriously in error provided blocks are large, but its value as a technique with perennial species is problematical. It is, however, presented here because experimenters in some parts of the world often have to work with land that is patchy and the method is at its best in such cases. Also, the results of Parker [104] with citrus suggest that there can be considerable similarity between the performances of adjacent trees.

To illustrate the approach, the following figures are taken to represent yields from a group of plots in the corner of a trial, the letters representing treatments. The experimental design is for this purpose immaterial.

| | | | | |
|---|---|---|---|---|
| B, 60; | A, 40; | B, 56; | C, 45; | . . . . . . . |
| C, 51; | C, 48; | A, 42; | B, 57; | . . . . . . . |
| A, 47; | C, 50; | B, 57; | A, 39; | . . . . . . . |
| B, 67; | A, 46; | C, 47; | B, 56; | . . . . . . . |
| . | . | . | . | |
| . | . | . | . | |
| . | . | . | . | |

The next step it to work out treatment means over the trial as a whole. In the present example it will be assumed that these come to 45 for A, 65 for B and 50 for C, then it is evident that this corner of the field is a patch giving low yields, though some parts of it give lower than others. The differences for each plot between actual figures and treatment means are:

| | | | | |
|---|---|---|---|---|
| B, −5; | A, −5; | B, −9; | C, −5; | . . . . . . |
| C, +1; | C, −2; | A, −3; | B, −8; | . . . . . . |
| A, +2; | C, 0; | B, −8; | A, −6; | . . . . . . |
| B, +2; | A, +1; | C, −3; | B, −9; | . . . . . . |
| . | . | . | . | |
| . | . | . | . | |
| . | . | . | . | |

It is now possible to get an idea for each plot how it might be expected to behave on the basis of the performance of its neighbours. The adjusting value for the second plot in the second row is $-7$ ($= -5 + 1 - 3 + 0$), the sum of the differences for its four neighbours, two along the row and two across it. Similarly the adjusting value for the third plot in the third row is $-12$ ($= -3 + 0 - 6 - 3$). This method breaks down for plots on the edge, but the method here is to take two neighbours only and to double their sum. Thus, the first plot in the second row should be given the adjusting value $-6 = 2(-5 + 2)$, the corner plot $-8 = 2(+1 - 5)$ and the second plot in the first row $-28 = 2(-5 - 9)$.

These adjusting values once obtained, it is convenient though not essential to add to each of them some constant number large enough to make them all positive. An analysis of covariance (see Appendix IV) may then be worked out,* the actual figure for each plot being adjusted by

* In carrying out an analysis of covariance it is usual to reduce the degrees of freedom for error and for total by one (see Appendix IV). Bartlett [9] has given reasons for thinking that the association between the dependent and independent variates in this type of analysis is such that this reduction is not enough, and he recommends that two degrees of freedom should be removed instead of one.

the values just derived. The method of Wellman, Thurston and Whaley [172] is similar to that of Papadakis but inferior, one difference being in the method of obtaining the adjusting values. This, in itself, is no criticism of their method. The precise formula whereby these values are derived from the differences from treatment means of nearby plots is not a point of principle—some prefer complicated expressions and others are content with simple ones, such as that suggested here. It is important only that the adjusting values should be derived in some objective manner independent of any preliminary opinions the experimenter may have formed of the probable effect of his treatments. The objection to their method lies in their avoidance of the analysis of covariance. The reason given in Section 41, namely, the need to keep clear of doubtful assumptions, applies with no less force here than in calibration and covariance is therefore essential.

# CHAPTER 5

# BLOCKS, PLOTS AND REPLICATIONS

" . . . —and a most curious country it was . . . the ground between was divided up into squares. . . ."

Lewis Carroll, 'Through the Looking Glass,' Chap. II.

### 50. Size and Shape of Blocks

Since much of the variation on a plantation may be coming from sources other than positional ones, the study of size and shape of blocks is not as important as it is with annual crops, but even so it should not be neglected. The general rules are that blocks should be as small as possible and should be of some compact shape, *i.e.*, square rather than long and narrow ; but exceptions may be made in special cases. The real aim is that there must be as little variation as possible *within* blocks. This means that differences in the field should be associated as far as possible with blocks.

Thus, on very hilly land "contour blocking" may be desirable, trees being assigned to blocks according to their altitude. This is necessary if conditions at the top of the hill are very different on account of depth of soil, exposure or drainage from those at the bottom. In such cases greater uniformity within blocks will be obtained by grouping together trees that are high up and those that are low down rather than by dividing them up into neat squares, each of which may cover several contours. Similarly, if some part of the field has poor drainage or has different soil, a block system that will take account of these differences may well be preferable to one that does not, even though it uses blocks that appear to straggle when marked on a plan. Again, to take an example, an apple rootstock trial at East Malling was being designed on a strip of trees, four rows wide, lying across the prevailing wind. In the event it was laid out in four randomized blocks, each block occupying one row. Thus one block was on an outside row to windward, two were on inside rows and one on an outside row to leeward. Despite the length and narrowness of these blocks, it was considered that this design, by taking account of the principal sources of positional variation, namely, shelter and competition, was better than one in square blocks. The general rule for block shape is derived from the observation that neighbouring trees are usually more alike than trees far apart ; but if there is any reason to think that resemblances are going to follow some other pattern, as in the cases just cited, the blocks should be chosen accordingly.

This example of the apple rootstock trial also illustrates one useful property of blocks, namely, their ability on occasion to remove headland effects by arranging that all plots of a block shall be on the outside or, alternatively, all on the inside. Another example, this time from a cherry trial planned at East Malling but not planted, is illustrated in Figure XXVII, where the effects of three unguarded headlands are eliminated by the blocks. A weakness here is the exposure of the two plots in the bottom corners to two headland effects, only one of which is eliminated, but this is probably not serious and could, if need be, be rectified by the use of a pseudo-variate (see Section 81). As with the designs discussed in Sections 22, 24 and 32, if outside guards are to be omitted it is essential that each headland should be homogeneous in its effect, *i.e.*, there must be no buildings, trees, etc., that affect some of the outside trees on that headland but not others.

It is also desirable that all blocks of a trial should be similar in shape and area, but this is not always possible. Usually they will be of the same area anyway, because each will contain the same number of plots, but exceptions do occur. For example, if the land has obstructions, such as outcrops of rocks or a winding stream, spaces may have to be left that will increase the surface of some blocks but not others. As to shape, there is a temptation sometimes to form most of the blocks in some compact manner and to assign the residue of plots to the last block, no matter how diverse they may be. An example is illustrated on the left-hand side of Figure XXVIII, the right-hand side suggesting a better method.

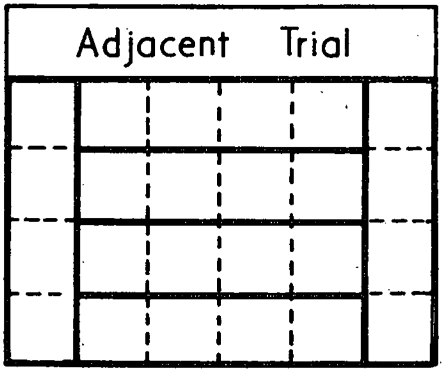



FIGURE XXVII.—A system of blocks to eliminate headland effects.

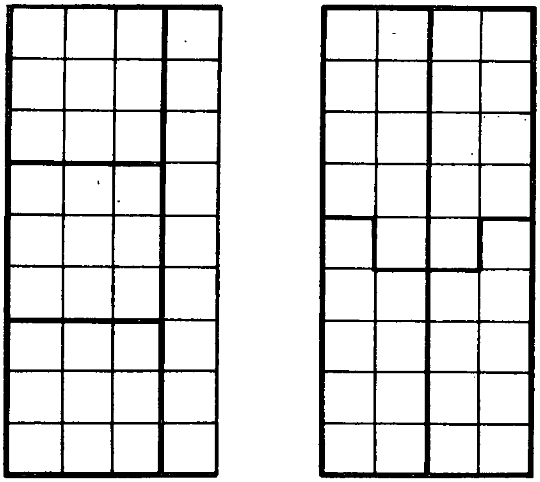


FIGURE XXVIII.—Alternative methods of forming blocks.

A block system having been adopted, good use should be made of it throughout the whole period of the trial. Thus, it may well be that not all the trees can be picked on one day and the weather may prevent an early resumption. In this case it is desirable to stop picking at a block boundary so that any effect of time of picking may be associated with blocks and eliminated with the positional effects. Again, it may be that records are to be made by estimation (see Section 93) and the work must be divided between two or more recorders. Again, it may be necessary to employ several pruners. If each is assigned whole blocks, their personal differences will be eliminated. In the conduct of a trial many differences have to be introduced, but if

## Table II.

### CATALOGUE OF UNIFORMITY TRIALS.

| *Species.* | | *Reference.* | *Recommendation.* | *Comments.* |
|---|---|---|---|---|
| Apples | *A | Batchelor and Reed [12] | 8 trees per plot. | Considers only information per replicate. |
| | *B | Strickland [153] | 8 ,, ,, | ,, ,, ,, Single tree plots thought quite possible. |
| | †C | Pearce and Thom [122] | Single tree plots. | Cases are noted where the recommendation does not apply. |
| | †D | Pearce and Thom [124] | 40 trees per treatment. | Plots taken to be separated by single guard rows. |
| Black currants | *E | Hatton, Grubb and Knight [76] | 50 bushes per treatment. | Method open to serious criticism (see Section 52). |
| Cacao (regular spacing) | F | Cheesman and Pound [24] | 12–18 trees per plot. | It is shown that larger plots give more information per replicate and this is considered to be the largest practicable plot. |
| Cacao (irregular spacing) | G | Jolly [84] | 0·05 acres per plot. | Considers principally information per unit area. |
| | H | Pearce and Thom [123] | 0·15 acres per treatment, <0·025 acres per plot. | ,, ,, ,, ,, Results based on calibrated trees. |
| Coconuts | *I | Joachim [83] | 18–20 palms per plot. | Considers only information per replicate. |
| | J | Pieris and Salgado [126] | 18 palms per plot. | ,, ,, ,, |
| Coffee | K | Gilbert [72, 73] | 20 bushes per plot. | ,, ,, ,, |
| Hops | L | Keller [85] | 5 plants per plot. | ,, ,, per unit area. |
| Lemons | *M | Batchelor and Reed [12] | 8 trees per plot. | ,, ,, per replicate. |
| Oil palms | N | Webster [169] | 16–32 palms per plot. | ,, ,, ,, |
| | O | Ollagnier [102] | 6–12 palms ,, | Considers principally information per unit area. |
| Oranges | *P | Batchelor and Reed [12] | 8 trees per plot | Considers only information per replicate. |
| Peaches | *Q | Strickland [153] | 8 ,, ,, | ,, ,, ,, |
| | R | McHatton [95] | 6 ,, ,, | Method open to serious criticism (see Section 52). |
| | *S | Shah [144] | 4 ,, ,, | Eight-tree plots gave more information per replicate, but not worth additional land needed. |

TABLE II—*continued.*

CATALOGUE OF UNIFORMITY TRIALS.

| *Species.* | | *Reference.* | *Recommendation.* | *Comments.* |
|---|---|---|---|---|
| Pecans | T | Sharpe and Blackmon [145] | Single tree plots. | Considers only information per unit area. |
| Pineapples | *U | Magistad and Farden [93] | Three beds, 75–150 ft. long, only middle bed harvested. | Recommendation only partly based on uniformity data. |
| Raspberries | V | Hoffman [81] | One row, 30 feet long. | Considers only information per replicate and one-row plots. |
| Strawberries | †W | Edgar [54] | 50 plants per plot or rather fewer. | Considers only information per replicate. |
| | †X | Taylor [160] | Plots of two rows each of 12 plants. | Recommendation based on practical convenience as well as data. |
| Tea | – | Prillwitz [127] | No recommendation as to size. | Although sometimes cited for tea, figures actually refer to European field crops. |
| | *Y | Eden [51] | 1/18 acre per plot. | Considers only information per replicate. |
| Vines (Irrigated) | Z | Covas and Christensen [42] | Four vines per plot. | Results not in accord with AA, based on same data. |
| | AA | Christensen [26] | 12–16 vines per plot. | Considers only information per replicate. |
| | BB | Pearce [115] | No general recommendation. | The results of AA discussed from the point of view of information per unit area. |
| Vines (Unirrigated) | CC | Strickland, Forster and Vasey [154] | 8 vines per plot. | Considers only information per replicate. |
| Walnuts | *DD | Batchelor and Reed [12] | 8 trees per plot. | ,, ,, |

* Publishes data of possible value to future investigators.

† Data may be obtained by application to the East Malling Research Station, Maidstone, Kent, England.

they can be associated with the blocks, they will usually do no harm. Equally in a design of Latin square type they should be associated with rows or columns. As a result of such practices the differences due to blocks can be very complex, and in the analysis of data it is not usually worth while to investigate the significance of the block differences. Experimenters sometimes say, "But I want to find out if there are any positional effects." If the blocks have been used in the way recommended as a sort of dust-bin for unwanted differences, dissimilarities between them are not necessarily due to position, though they may be. In order to make them of such a size that they correspond to a day's work in tapping rubber or plucking tea, confounding devices such as incomplete blocks may be necessary.

**51. Size and Shape of Plots**

A great deal of work has been done on the size of plots and this is summarized in Table II. Whether it was all necessary or not is an open question, for the empirical law for soil heterogeneity given by Fairfield Smith [149] sums up much of the early work and provides an answer to most practical problems.

Recent work at East Malling has avoided making use of this law because it has not been clear how well it applies to plants the variation of which is not dominated by positional effects, but it is now clear that the general conclusions apply to perennials no less than to annuals.

These conclusions are that economy of land is best effected by using small plots, but that more information per replicate is obtained from large ones. Thus, if an apple trial is being designed without guard rows and it is intended that the significant difference for crop between treatments shall be about 20% of the general mean, the experimenter would be faced with the choice of using eight replicates of one-tree plots, four replicates of three-tree plots or three replicates of five-tree plots [114]. His decision would depend entirely upon circumstances. If he were planning the trial for a research station with the usual shortage of land but a sufficiency of trained staff, he would choose the smallest size of plot: if he were planning for a grower's plantation with many available trees but no skilled experimental assistance, he would choose the largest. There is no general rule in the matter, and reports of uniformity trials that purport to inform their readers of the "best plot size" for a certain species are merely misleading.

In deciding upon a suitable plot size, the principal consideration is the one just given, *i.e.*, to economize land, use small plots: to economize skilled labour, use large ones. There are, however, a number of subsidiary points to consider. Thus, plots must be sufficiently numerous to give enough error degrees of freedom—say 15 and, if possible, 25. Also, plots must not be so small as to give a standard error greater than 30% of the general mean [49, 94], though this is a matter that can be considered only in the light of past experience. Further, plots should be of such a size that their crops can be conveniently weighed in the field, difficulties being caused by quantities so large as to need numerous weighings and by quantities so small that the diminutive scales required are unusable in anything except a complete calm. Sometimes, too, small quantities cannot be weighed with sufficient accuracy.

Plot size must also be considered in the light of possible losses. Here it should be explained that, in the analysis of data, completely missing plots are manageable provided there are not too many of them: at the other extreme, no irreparable harm is done if one or two plants die in a plot of, say, twelve or more. Difficulty is, however, caused by a plant being lost from a medium-sized plot. Consequently, if losses are expected—and with some species they are inevitable—there is a case for using either large plots that can stand a few losses, or small ones that will be completely wiped out, and for avoiding medium-sized ones in which losses can be very awkward. Admittedly, however, there are exceptions. Thus, at East Malling it is now established practice with plums and cherries to use two-tree plots, but these species are distinctive in that most losses occur early. Consequently, it is possible after quite a short period to decide upon the course of action: either (1) to regard all plots from which one or two trees have been lost as completely missing, (2) to reduce all plots with survivors to one tree by the random elimination of a tree where both have survived and accept a few completely missing plots, or (3) to replant all losses simultaneously and use a pseudo-variate (see Section 81).

Two important general points remain to be noted. It sometimes happens that (1) the treatments are being applied directly to the trees as individuals, no guard rows being needed, and (2) each tree must be recorded separately. In such cases the skilled labour needed is directly proportionate to the number of experimental trees which, in the absence of guards, is directly proportionate to the area of the experiment. That is to say, there is no longer any conflict between the economizing of land and of labour and there is a strong case for using the smallest possible plot—one-tree plots, in fact, unless there is a good reason to the contrary.

The last general point concerns the effect of guard rows. The smaller the plot the better the use of land but the greater the wastage due to guard trees, and it is not at all clear which of these two effects is dominant. That this difficulty has received such little attention is rather surprising in view of the extensive use made of guarded plots, but it is clear that with guard rows small plots should be used with reserve. For apples it appears that single guard rows approximately balance any economy of land effected by reduced plot size so, for singly-guarded plots, information per unit of area is nearly independent of the number of trees to a plot [124]. Incidentally, square plots need fewer guards than long ones of the same area.

All this will help when experience has once been gained with a crop. It will suggest when to make plots smaller and when to make them larger, but it will not help if the investigator is working with the species for the first time. Here the recommendations of Table II may be of great help, since these are given by people who know something of the various crops. Even though some of the trials are open to criticism and though the conclusions reached need not be adhered to as rigidly as the writers sometimes appear to suggest, they do represent a reasonable sort of size for the beginner to adopt. Actually, plot size may be varied to a surprising degree if replication is adjusted with it. Thus, with sugar cane Turner, Warneford and Charter [165] used plots as small as ten square yards, while Paterson and Hanschell [106] used 220 square yards, yet both conducted perfectly satisfactory experiments.

In general, plots should be long and narrow—a conclusion that has been demonstrated empirically for certain annual species [27, 28] and doubtless applies to perennials also. Indeed, it is the counterpart of the rule that blocks should be compact; because whereas blocks should be associated with positional differences, the plots of a block should resemble one another as much as may be. Nevertheless square plots are sometimes more convenient and the loss of uniformity brought about by their use is not often so large as to be serious. It is not essential for plots even to be rectangular and, indeed, the suggestion has been made that they can sometimes most conveniently be arranged in a spiral [82].

Some diseases show a tendency to spread along rows rather than across them. Where this is expected to occur it must have an effect upon the orientation of plots. Despite certain general recommendations that in such cases plots should be as square as possible [54] or should be disposed across the rows [6], the present writer believes that the decision must depend upon the purpose of the trial. If the intention is to test some cultural treatment, the expected disease being no more than a nuisance, it is obviously wise to dispose plots across the rows so that they are all affected to about the same extent if a row does become diseased. If, on the other hand, the disease is itself the object of study, it would appear better to arrange the plots along the rows in order to confine an outbreak to the plot in which it started. Apart from this consideration, the orientation of plots probably matters very little (cf. [28], referring to cotton), and is usually a question of convenience. In practice it is usually easier to run plots along the rows, making use of natural landmarks, such as posts for wiring, to indicate boundaries.

## 52. Review of Uniformity Trials

As explained in the previous Section, the writer has little confidence in uniformity trials as a means of deciding the optimum size of plots, though they are useful for several other purposes. Thus, they do indicate the level of variability that may be expected with a species, though all such indications should be considered critically (see Section 54). Also, uniformity trials are invaluable for suggesting useful calibrating variates (see Section 42), and it is no coincidence that many of the references in Table I turn up again in Table II, which lists the recommendations as to plot size from a range of uniformity trials conducted on perennial species.

It is unfortunate that some of these trials are in themselves open to criticism. Thus, a change of plot-size logically involves a change in area of block as well, but this has been overlooked in several instances (Trials A, B, L, M, P, Q, S, U, V, W, CC and DD in Table II). Again, not only must the area of block be altered, but it must be altered in such a way as to keep the number of plots it contains constant, which was not done in Trials J and N or in Trial Y, where Latin squares were used with varying numbers of treatments. Also, the assumption in working out the results of Trial E that the standard error of a plot varies inversely with the square-root of its area is now known to be mistaken. Exception must also be taken to the method in Trial R, where the conclusion is based on a high correlation between results for the plots as a whole and those for certain selected trees, an argument that overlooks the possibility of trees selected for proximity behaving differently from a selection at random.

With perennial plants a uniformity trial takes so long that it is not usually advisable to hold up experiments until reliable results are available. (The working out of data already to hand is, of course, another matter.) Sometimes a preliminary study of variability is necessary as, for example, in Trial H, when it was desired to determine if an experiment would be possible given certain conditions; but generally it is better to start experimenting at once, using large plots and many replicates, reducing one or another as experience shows it to be permissible. Plot size is better chosen on the basis of the considerations set out in Section 51, but a knowledge of the means of calibration is so important that many more uniformity trials are called for, the information not being so readily obtainable from trials planted primarily for another purpose.

The variability of apples in store has been studied by Strickland [153] and by Hoblyn [79].

## 53. The Use of Guard Rows

Guard rows are of two kinds, *external* and *internal*. The first are used to equalize the competition along the outside edge of the experimental area: the second to prevent the treatment given to one plot affecting its neighbours.

Ideally all trials should be guarded externally, though several designs have been suggested (see Sections 22, 24, 32 and 50) which allow for headlands being unguarded if need be. Of course, this is never permissible if the headland is *heterogeneous* like those in Figure XXIX, but if it is reasonable to assume that there is a constant effect on all the outside trees down one side of the trial, *i.e.*, the headland is *homogeneous*, the elimination of this effect by statistical means instead of a guard row is quite permissible. It should be added that outside trees are more easily lost than inside ones, so especial care is needed in turning tractors, etc., when external guards are omitted.

The assumptions made in eliminating headland effects in this way are two in number: (1) the absence of an interaction between the treatments and the headland effect, *i.e.*, the treatments should give relatively the same effect on inside and outside trees, and (2) that in any row, column or block, each plot has the same proportion of trees on the headland. The special problem of plots in a corner has been discussed previously in Section 22.

Internal guards are needed between plots whenever the treatments are not applied directly to the tree as in manuring [86] and also with treatments like spraying. Even where they are applied directly, as in pruning, guards will be needed if large differences in tree size are likely to be brought about, but a better solution here is often wide spacing, really a form of the guard alleys to be discussed below. For obvious reasons the number of guards should be kept to a minimum and usually a single row will suffice, this row being fertilized or sprayed differently on its two sides (see Figure XXXB). If double guards are used the difficulty must be faced of how to treat the insides of the double row. If no treatment is applied (as in Figure XXXC) the trees will be stunted on account of starvation or be a source of infection through not being sprayed. If a treatment is applied (as in Figure XXXD), the advantage of the double row is largely lost. Nevertheless; double or even triple guard rows are sometimes needed to prevent the spread not of the treatment, but of the disease or pest it is intended to control; though, where this is done, some basic treatment over the whole experimental area is advisable for the sake of the guard trees themselves. With some insects no guard rows of practicable width will prevent spread from a plot where the insect is virtually uncontrolled during the season.

In such cases the basic treatment must provide a reasonably good control of the pest because otherwise the trial may be ruined.

Sometimes, instead of internal guards, an alley is left, either by omitting a row of trees or by increasing the planting distance. As a barrier against pests and diseases, this device is often valuable, while it also permits the erection of screens when spraying. In manurial trials it has the disadvantage of altering the area of land available to each tree. Another device is a deep guard furrow, frequently repaired to cut roots that try to pass it.

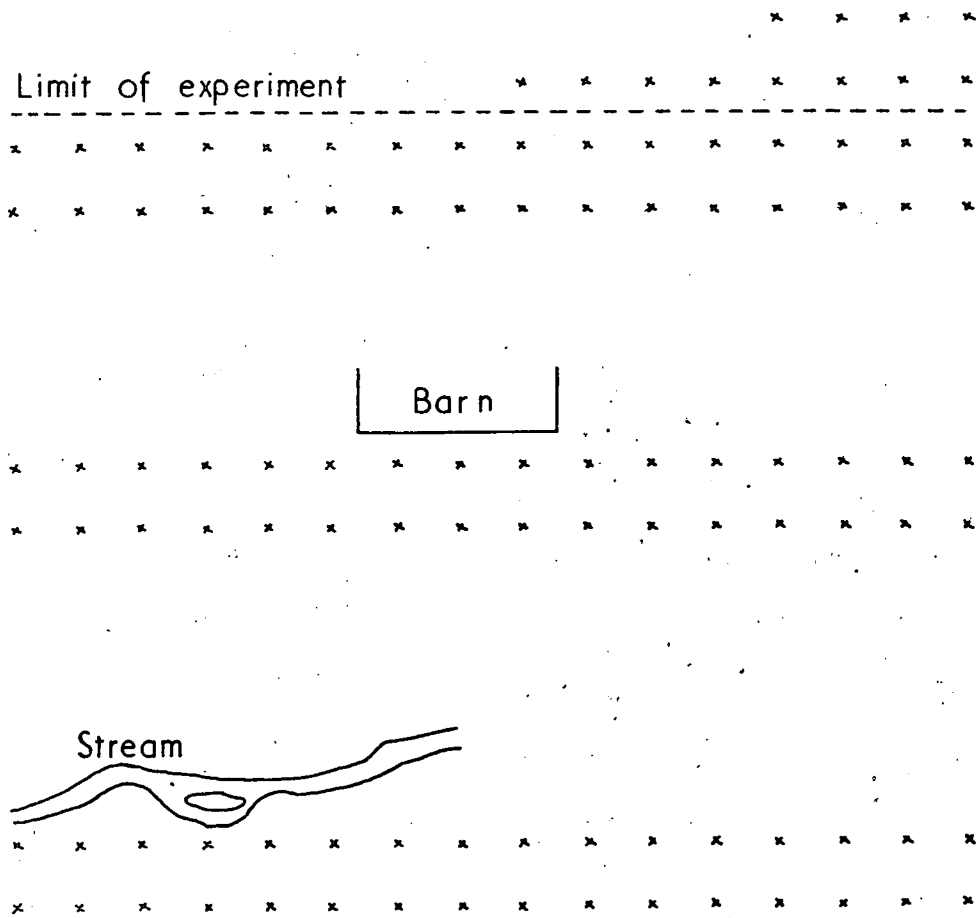



FIGURE XXIX.—Examples of heterogeneous headlands.

### 54. Degree of Replication

Replication serves two purposes: (1) It makes possible an estimate of the residual variability of the experiment by providing error degrees of freedom, and (2) it enhances the estimates of treatment effects, which otherwise would be based on single plots.

It has already been mentioned that replication should be enough to provide, say, 15 degrees of freedom for the error and, if possible, 25 (see Section 51). There is nothing critical about these figures, but study of a table of $t$ [66a] will show the values to rise steeply when degrees of freedom are reduced much below this level, thus confirming the reasonableness of the recommendation.

Although increased replication does lead to better determination of treatment means, it is not wise to base their accuracy on this alone. For one thing, it is often disappointing in its result, *e.g.*, replication has to be increased four-fold to halve the standard error. For another, standardization of experimental material and calibration are usually more effective.

In deciding the degree of replication called for, the possibility of losing plots must be borne in mind, while non-orthogonal designs, which involve adjusting the treatment means, call for more replicates than designs that do not. In such lay-outs the effective replication can be obtained by multiplying the actual figure by the "efficiency factor", which latter quantity is usually set out in the relevant literature or may be calculated by the method given in Appendix III.

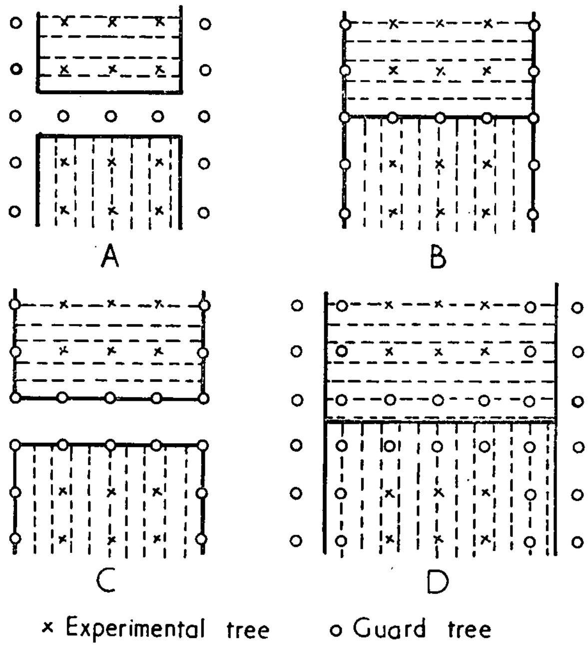



**FIGURE XXX.—Different types of internal guard rows.**

A special problem is presented by factorial designs, which sometimes provide "hidden replication", as it has been called. Thus, if a trial has six replicates of eight treatments, made up of one factor at four levels and another at two, and if there is no interaction, the first factor has in effect twelve replicates and the second has 24. It would be wrong to rely upon this because the inclusion of the two factors in one trial implies a belief that they may interact, but in multifactorial designs with high-order interactions that are not themselves under study but merely arise from the nature of the design, it is often quite reasonable to assume that there will be no last-order interaction and to reduce actual replication accordingly, relying upon the hidden replication to make good the deficiency.

In the last resort, however, the degree of replication must depend upon the variability to be expected, and this is by no means easy to estimate. Several correspondents have requested that this publication should include a table showing the standard errors usual with different species but, valuable though such a table would be, there is no possibility of compiling one because each species presents a study in itself. For one thing, there are many quantities that may be investigated, and a trial able to demonstrate a 20% difference in crop, or a 15% difference in trunk girth, may be unable to show a difference of less than 30% when it comes to a disease.

Again, variability often depends upon local conditions and the cultural treatments applied. Thus, it is not to be expected that citrus grown in irrigated sand will be as variable as that grown in, say, loam with an irregular natural rainfall. Even if these conditions are standardized there remain a number of other factors, such as the duration of the trial and the age of the trees. Apples, for example, judging by experience at East Malling, become less variable in crop as time proceeds, but more variable in trunk girth. Shah [144] confirmed the change in crop variability in peaches, but Jolly [84] found no sign of it with cacao. Also, the size of the plots is very important, and of the blocks also. Finally, in non-clonal material there is the degree of genetical variation, which must depend very much on the source of the plants.

These difficulties are presented, not to argue that no prior forecast of variability is possible, but to warn the reader against facile generalisations. In course of time each research centre builds up a body of local lore that will provide guidance for the future and this is all that is needed, but care must be taken to use it intelligently. The factors given above are all important ones and will suggest when variability may be expected to be higher or lower than usual. For the experimenter who has no local lore to guide him, the best course is to consult the relevant references in Table II. Whatever the value of uniformity trials as a guide to plot size, there can be no doubt of their importance in suggesting the level of variation. Thus a worker with a certain crop reads the original paper and finds that with the six-tree plots that he proposes to use (the reference may recommend 12-tree plots but that does not matter) the previous investigator got a standard error for a plot equal to 25% of the general mean. Now he argues, " My material is every bit as good as that described here, but my blocks will be larger because I have more treatments. Also, my trees are younger and probably, therefore, more variable in crop. My trial will go on for the same period, so I cannot argue that my variability will be less on account of greater duration. Altogether I had better be prepared for an error of 30%." This is the sensible way to go to work, *not* just to take over the previous figure of 25% without critical consideration of all the relevant factors. He should then supplement the information derived from the uniformity trial or trials by noting other published analyses of variance relating to the crop. Admittedly, many experimenters in the tropics have not extensive libraries to hand, but they would all feel that they had failed in their job if they had not read the more important papers in *Tropical Agriculture* and other journals relating to the crops they study. If they would note details of experimental design as well as results, they would be less worried when they came to carry out an experiment themselves. Also, if they would mention their own standard errors when they come to publish, they would help those who come after.

## 55. Formulae for Determining the Number of Replicates

In the literature two methods have been given for determining the replication needed to show a significant difference of a certain amount. In the one, which has many forms, the argument goes like this : The error mean square may be expected to equal $s^2$ (a figure based on past experience as described in the last Section). With an effective replication of $R$, the smallest difference between treatment means that will appear significant is $D = ts\sqrt{\frac{2}{R}}$, where $t$ is derived from tables [66*a*] and depends upon both the level of significance required and the number of degrees of freedom in the error. Equating $D$ to the figure desired will give a value for $R$.

The second method is more general. It will be seen that the above value of $D$ is an average figure, depending as it does on an average figure of $s$. In fact, there is a half chance that in any particular instance $s$ will be above expectation, just as there is a half chance that it will be below. Consequently, if a value of $D$ is selected as representing the sort of sensitivity desired and a value of $R$ adopted accordingly, there is a half chance that variation in $s$ will prevent the selected difference, if it occurs, from being shown significant. Nevertheless, in making a general recommendation, the first method is correctly used as giving an average figure for sensitivity over future trials if the value of $R$ is adopted.

In the second method [35*a*] a difference, $D$, is taken and associated with a probability, $P'$. The investigator says in effect, " If a difference, $D$, should arise, I wish it to be detected on a proportion, $P'$, of occasions, $P'$ being greater than or equal to 0·5." The equation now becomes $D = s\,(t + t')\sqrt{\frac{2}{R}}$, the symbols having the same meaning as before. $t'$, which is new, is obtained from tables in the same way as $t$, the degrees of freedom being those for error and the significance level being $2(1 - P')$. It will be seen that if $P'$ is made equal to 0·5, the second method reduces to the first, $t'$ equalling zero.

Thus, suppose $s^2$ is expected to equal 400 and it is proposed to compare six treatments in randomized blocks. It is desired that if a difference of 25 should arise between treatment means there should be a probability of 0·8 of detecting it at the $P = 0{\cdot}05$ level. Suppose nine replicates were used. This would give 40 degrees of freedom to the error, making $t$ equal to 2·021 (for probability, $P$) and $t'$ 0·851 (for probability, $2[1 - P']$), so the value of $D$ would be $20(2{\cdot}021 + 0{\cdot}851)\sqrt{\frac{2}{9}} = 27{\cdot}1$, and evidently this is not replication enough. Further investigation shows that eleven replicates would be needed to make $D$ less than 25, a result that could have been obtained from the table of Cochran and Cox* [35*a*]. Reference should be made to the same place for the replication needed in one-sided tests, *i.e.*, where it is assumed that a certain treatment cannot be worse than the control, though it might be better (see Section 71).

In deciding upon the number of replications required, it is permissible to take into account any hidden replication arising from high-order interactions assumed to be non-significant.

* These writers use $\sigma$ for $s$, $P$ for $P'$ and $\delta$ for $D$, a notation avoided here, partly from a desire to keep Greek letters for parameters and partly to avoid confusion with $P$ used to indicate a significance level.

# CHAPTER 6

# TRIAL DESIGN IN RELATION TO THE SUBJECT OF INVESTIGATION

" Every purpose is established by counsel."

Proverbs, xx, 18.

## 60. Manurial Trials

There are two broad classes of manurial trials, the designs of which may be termed *definitive* and *tentative*. In the first, the treatments are applied and are kept on to observe their effect. Thus, if symptoms of deficiency appear, no change is made, because the purpose of the trial is to find out the characteristics of plants growing with an excess or a deficiency of certain elements. In the second, the object is to find the optimum manurial programme in a certain district and, consequently, if a treatment leads to deleterious effects, it is promptly eliminated from the trial and perhaps another substituted. The two are so different that an experimenter should be quite clear from the start which sort of trial he is conducting. Whatever the decision, the general comments of Crowther [48] are very valuable where little is known of local conditions.

Both need internal guard rows if they are to go on for any length of time. Usually single guards will suffice, though double ones may be needed if bulky organic manures are being tested of a kind likely to be pulled long distances by cultivation. There is some evidence that guard rows can be dispensed with if the applications are to be made over only a short period, or if the fertilizer is applied to an area considerably smaller than the space bounded by lines half way between the trees. Thus, Sharpe and Winsor [146], working with pecans, fertilized some trees with boron but not others and found no evidence of cross-feeding. Fertilizer was applied in some cases up to 21 feet from the trunks, the trees being at a minimum 50 feet apart. Only one application of boron was made. Again, Lott, Satchell and Hall [89] gave a single application of radio-active phosphorus to grape vines and reached the conclusion that cross-feeding was negligible between rows, which had been cultivated so as to sever roots near the surface, but that double guards were needed within rows. It is, however, clear from the work of Rogers and Vyvyan [140, 141] that, in the case of apples at least, the roots of adjacent trees at commercial planting distances generally overlap to some extent, though this overlapping is limited by competition among the roots themselves, and this means that there must be some cross-feeding. Also, there must be a tendency for roots from a poorly fertilized tree that happen to find themselves in a better fertilized plot to develop unduly and so to accentuate the bad effects of not having guard rows. For coconuts, Salgado [142] considered guard rows essential.

Calibrated trees should, if possible, be used for either sort of trial; but, if the trees are likely to respond to fertilization only when young, this may be impossible. Also, with definitive treatments it is sometimes thought advisable to start applying the fertilizers as soon as the trees are planted, or even before, so as to secure characteristic conditions.

It should perhaps be emphasized that the manurial problems of perennial species are essentially different from those of annuals. Thus, with a short-lived plant it is possible to talk of the optimum application of a certain element and to design a factorial trial with $3^n$ treatments in order to find out the level of fertilization at which further applications will be uneconomic. With a perennial, however, the optimum application is changing all the time as the trees become older so there is no point in trying to determine what it is. Instead the experimenter needs to know (*a*) how to recognize the symptoms of excess or deficiency of any element, for which purpose he needs a trial with definitive treatments, and (*b*) how to apply a given element when he believes it to be needed, a question in the province of a trial with tentative treatments. Consequently, when treatments are definitive, the design will usually be of the type $3^n$, so as to incorporate each element at a low level likely to induce deficiency, a medium level for normal growth and a high one. When treatments are tentative, the design will instead be of the type $2^n$, so as to study if the proposed method of application is indeed effective. Consequently one level should be quite a high one and the other might well be complete absence of the element from the fertilizer programme.

In a tentative design care must be taken to ensure that new factors can be added as new questions arise. In fact, such a trial may well be started with no differential treatments at all, the whole area being given what seems a reasonable programme for the district. First, as defects in this programme appear, new treatments may be added to try to effect a remedy. Then, as soon as the solution is found, the differential treatments can be discontinued and the remedial treatment applied to all plots. The method has been described more fully elsewhere [120]. Usually a $2^n$ design is best because the questions to be answered are of the form "Is more of this element needed?" not "How much should be applied?" and the trial commonly begins with plots to the number of a power of two arranged in four or eight blocks. The example in Figure XXXI will illustrate the possible development of such an experiment. Initially, it may be supposed, no differential treatments are applied, but after a while deficiency symptoms might appear suggesting the need for more Nitrogen. The first diagram accordingly shows half the plots receiving supplementary sulphate of ammonia. Later, the need for more Potash might be suggested but, since with some species a high level of Potash can induce Magnesium deficiency, it might well be thought that two further factors are needed instead of one, and the second diagram shows supplementary sulphate of potash and sulphate of magnesium under test as well as the differential treatment already applied. At a still later stage, it might become apparent that the supplementary Nitrogen was indeed essential, and therefore it would be continued on the plots already receiving it and a higher level applied in the others in an endeavour to eliminate the residual effect. Simultaneously, the suggestion might be made that the trees needed either Iron or Copper, and these two elements might be applied by injection. The result is shown in the third diagram, the brackets round the N, used to indicate application of Nitrogen, showing that only the residual effects are here under consideration. At this stage three interactions are confounded, the number of treatment combinations being now four times the number of plots to a block. It is unwise, incidentally, to confound the interaction of all the factors extant at any one time, because this makes difficult the application of further factors. Also, when the number of treatment combinations equals the number of plots, as in the third diagram, some of the higher order interactions must be used as error. After a time the residual effect of the Nitrogen applications will have been removed, leaving room for a further factor if one should be suggested, a stage represented by the fourth diagram.

In a sense, this approach is a development of factorial design. Many years ago Professor Fisher said that Nature often replies better to a questionnaire than to a question, meaning that a factorial design with its ramification of interactions often secures more information about the effect of a factor than does a simple one. The difficulty with a long-term factorial trial is that some answers are given at once and the subject need not be pursued further: others are not given at all, although the questions are repeated year after year. If a soil already has sufficient potash, and repeated applications of a differential treatment have no effect, the time comes to drop the subject. Perhaps the position may be summed up by saying that Dame Nature is even more informative in a conversation than with a questionnaire. Sometimes the subject can be pursued, sometimes it should be dropped, either because the lady has answered all that has been asked or because she is just not interested and has nothing to say.

Where conducting this sort of trial, it is advisable to have a few additional but comparable trees available for purposes of injection, this method being valuable for diagnosing the nature of mineral deficiencies [133, 134] and thus for suggesting additional treatments. Of course, the establishment of a mineral deficiency by injection is not evidence that soil applications of the element will provide a remedy.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| N | N | O | N | O | O | N | O |
| O | N | N | O | N | N | N | O |
| N | O | O | N | O | N | O | N |
| O | O | O | N | N | O | N | O |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| NKM | NK | KM | N | KM | M | N | O |
| K | N | NK | O | NK | N | NK | KM |
| NM | M | M | NKM | K | NM | M | NKM |
| KM | O | K | NM | NKM | O | NM | K |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| (N)KMC | (N)K | KMIC | (N)I | KMC | MI | (N)C | I |
| KIC | (N)IC | (N)KC | C | (N)KIC | (N) | (N)KI | KM |
| (N)MI | MC | M | (N)KM | K | (N)MC | MIC | (N)KMIC |
| KMI | O | KI | (N)MIC | (N)KMI | IC | (N)M | KC |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| KMC | K | KMIC | I | KMC | MI | C | I |
| KIC | IC | KC | C | KIC | O | KI | KM |
| MI | MC | M | KM | K | MC | MIC | KMIC |
| KMI | O | KI | MIC | KMI | IC | M | KC |

FIGURE XXXI.—Possible successive stages in the evolution of a design with tentative treatments.

## 61. Trials of Cover Crops, Methods of Soil Cultivation and Methods of Irrigation

These types of trial are dealt with together because, in general, each calls for long plots. It is possible, though rather inconvenient, to fertilize individual trees, but there are no means of using implements properly in a restricted area, the chief difficulty lying in giving tractors, mowers, etc., easy access to the individual plots as well as room to turn when they are set to work. It is therefore desirable that the plots should extend right across the experimental area from headland to headland or, if this is not possible, from a headland to a landway up the middle of the trial.

If irrigation methods are to be tested, it is again desirable that plots should be long and narrow unless the sprinkler system is being adopted, in which case the problems resemble rather those of spraying and dusting trials.

Several types of lay-out can be adopted to secure the kind of plots desired. One is to form plots running from headland to headland, and to divide them into smaller sub-plots for the comparison of some other factor. This is perhaps the only case commonly met with where it is reasonable to include two factors together in one experiment even though they are not expected to interact. Large plots have to be used for the first experiment, so the additional factor on the split-plots is studied with no further expenditure for trees and little for additional labour.

Another possibility is to use the criss-cross design of Cochran and Cox (see Section 26), leaving spaces between the rectangles for the turning of implements, supply channels for water or whatever else may be needed.

If long plots cannot be accommodated, the best alternative is to use square ones, with guard alleys between them.

Treatments of this sort are best compared with guard rows because edge effects can be serious. Where guard alleys are used instead, the treatments should be applied some way beyond the outside trees because of these effects.

Where cover crops are not established when trees are young, calibration is easy to carry out and valuable. Irrigation, however, is usually needed, if at all, from the start, so the chances of calibration are here much reduced. It may well be argued that uniform irrigation followed by a trial of different methods prejudices the issues from the beginning, because the root growth favoured by the initial treatment may handicap the trees with certain of the subsequent treatments. This argument would not apply if the trial were of quantities of water. A similar caution applies to trials of methods of soil cultivation.

## 62. Spraying and Dusting Trials

Trials of this sort usually need fairly large plots because of the difficulty of confining the treatments within a limited area. Also, a fairly high standard of guarding is called for because pests and diseases can spread as well as treatments. Consequently, it is often advisable to provide single or double guards to keep the treatments to the desired trees and then to record only the middle trees of the plots thus formed. Some relief from excessive guarding can be obtained, however, by deciding on a likely direction of wind at the time of applying treatments and reducing the number of guard rows running in that direction, but this device depends upon reliable weather forecasts. Alleys or screens or both can be used instead of guard trees and are often preferable. Trials of sprinkler methods of irrigation are essentially similar.

A constant difficulty is the introduction of control plots, which often do no more than declare the obvious while providing a source of infection for the rest of the trial. Unsprayed controls are needed to demonstrate that the pest or disease was indeed present, but it is usually good enough if a few are disposed to leeward of the trial where they are unlikely to harm it. The real control for comparative purposes is the conventional programme of spraying or dusting, which it is hoped to improve. Alternatively, if there is no conventional local practice, the most likely control measure should be used, or it may even be necessary to carry out a simple preliminary test to find some treatment that can safely be included in the trial.

Tentative designs, such as were recommended for use with manurial trials, are valuable here also (see Section 60).

### 63. Trials of Varieties and Rootstocks

In comparing varieties, whether of scions or rootstocks, it will usually be found that the series to be studied falls into groups, such that little interest would attach to a comparison of varieties in different groups, though it is important to select the best within each. Thus, there is no point in comparing a late culinary pear with an early dessert one or a vigorous apple rootstock with a dwarf, but it would be valuable to know which was to be preferred in each class. For this reason some preliminary information is very desirable either by conducting first a simple observational trial or, in the case of rootstocks, by a histological study of roots, a method which, for many species at least, will indicate the vigour to be expected [15]. With rootstocks a preliminary inquiry is especially valuable if large trees are not to be randomized with small. Some initial information is also needed about incompatible stock-scion combinations so that double working can be used as necessary.

This tendency to compare in small groups obviates the need for lattice designs [185, 187*d*, 190], which find extensive application with annual species. These designs, admirable as they are with the crops for which they were evolved, have the major disadvantage with perennials of lacking robustness (see Section 21). An experimenter with annual plants may well test 100 varieties one year knowing that if some of them are unpromising—and some are bound to disappoint—next year his trial can have only 81 or 64; but with a perennial species he cannot do this. At best he can use a robust design, like randomized blocks, in which varieties that prove to be of little interest can be ignored and no further records made upon them.

With new varieties straight from the plant breeder there are reasons for delay other than the need to classify into groups. A variety that has shown up very well in the shelter of an experimental plot may be unable to stand the more rough and ready conditions of a commercial farm. It is accordingly advisable in the observational trial recommended above not to mollycoddle the trees, but, if they appear to need special care, to discard them as unlikely to be acceptable in commerce.

The main trial should not usually be held up on account of certain varieties being in short supply. When a trial is proposed, the propagator is usually given some material and asked to raise so many trees by a certain date. If he does not succeed because certain varieties will not propagate well, those varieties may just as well be discarded as having little commercial future, though if the deficiency is slight, the device of substituted plots (see Section 33) may prove useful.

In a preliminary trial of scion varieties, frameworking [71] on to a uniform series of existing trees is often valuable, and, indeed, sometimes makes a main trial unnecessary. The purpose is to evaluate the main varietal characteristics as quickly as possible.

The best way to go to work is to devise a series of " sieves " by which the number of varieties under study is successively whittled down, an approach that has been discussed in some detail by Hoblyn for cacao [80]. The proportion of varieties to get through to the main trial must depend on circumstances—it will obviously be greater when starting with a number of similar sister seedlings, the result of controlled crosses, than with a number of diverse varieties collected from commerce or sent in by growers—but in general most experimenters appear to pass too many rather than too few. A rule has been given by Yates [189] but it is not easy of application. Where the varieties represent a collection of all those the investigator can find being commercially grown, it is doubtful if more than 5% should call for serious consideration after the conclusion of the preliminary tests and these should by then have been classified into groups as described at the beginning of the Section.

Sometimes the grouping of varieties is obvious, *e.g.*, dessert and culinary, and sometimes it is dictated, at least in part, by the times at which the varieties became available; but often it must to some extent be arbitrary. Thus, suppose a range of rootstocks has been placed in order of expected vigour, then any division into rigidly defined groups must be artificial and, wherever the boundaries are placed, there will be comparisons between groups as important as those within. The method here is to overlap the classes so that some varieties occur in two neighbouring groups.

The method is then quite simple. Suppose the comparison in Group 1 is based on $n_1$ replicates and there are $p$ varieties in common between the two groups, then the standard error for the comparison of the mean of a variety, A, in Group 1 with the mean for the "overlaps" is (cf. Appendix I)

$$\sqrt{\sigma_1^2\left(\frac{1}{n_1}+\frac{1}{pn_1}\right)},$$

where $\sigma_1^2$ is given by the error mean-square of the analysis of variance for Group 1. Similarly, the standard error for the comparison of the mean of a variety, B, in Group 2 with the mean of the "overlaps" is

$$\sqrt{\sigma_2^2\left(\frac{1}{n_2}+\frac{1}{pn_2}\right)},$$

where $n_2$ is the number of replicates in the second trial. To compare A with B, the expression is: (Mean of A in Group 1) − (mean of the overlaps in the trial for Group 1) − (mean of B in Group 2) + (mean of the overlaps in the trial for Group 2), an expression which has the standard error,

$$\sqrt{\sigma_1^2\left(\frac{1}{n_1}+\frac{1}{pn_1}\right)+\sigma_2^2\left(\frac{1}{n_2}+\frac{1}{pn_2}\right)},$$

which may be converted into a significant difference by multiplying by $t$. If the two trials have different numbers of degrees of freedom in their errors, it would be safer to base $t$ on the smaller number.

The difficulty about choosing the number of overlaps is this: If too many are used, the number of treatments to each group will be inflated and precision will fall from lack of replication. It is assumed that the total area available for trial is limited. If too few are used, the comparisons between groups will suffer. Assuming the boundaries between groups to be entirely arbitrary so that comparisons between varieties in adjacent groups are as important as those within groups, the following rules for the number of overlaps may help. The figure given is in every case a maximum, so fractions should be counted to the whole number below.*

(*a*) To divide a series into two groups, first of all divide the varieties as evenly as possible. Let the mean number of varieties per group be $g$, then each should contribute $v = \sqrt{\tfrac{1}{2}g}$ to the other.

Thus, if 26 rootstocks, A–Z in that expected order of vigour, are to be divided into two groups, first divide A–M and N–Z, $g$ being therefore 13·0 and $v$, 2·55. Consequently, two varieties should be contributed by each group to the other to give A–O and L–Z as the final grouping.

(*b*) To divide into three groups, after the preliminary division each outside group should contribute $v$ to the inside one, which should contribute $2v$ to each of the outside ones, where $v = \sqrt{\tfrac{1}{6}g}$.

Thus, the 26 rootstocks would first be divided A–I, J–Q, R–Z, making $g$ equal to 8·7 so $v$ equals 1·20. The groups now become A–K, I–R, P–Z.

(*c*) To divide into four groups, after the preliminary division, each outside group should contribute $v$ to the adjacent inside one, which should contribute $3v$ to the adjacent outside one. Further the two inside ones should contribute $2v$ to each other, $v$ being equal to $\sqrt{\tfrac{1}{12}g}$.

Thus, the 26 rootstocks could first be divided A–G, H–M, N–S, T–Z, making $g$ equal to 6·5 and $v$ equal to 0·74. Thus the groups become A–I, G–N, M–T, R–Z. This is rather an interesting example, because $v$, being less than one, has to be approximated upwards to one instead of downwards as usual.

* These rules, here given for the first time, are based on a simple minimization of the expression for standard error given above, supposing the area of each trial to have been fixed beforehand.

On account of these approximations, it is possible to be left with markedly unequal groups and some adjustment may be needed between their boundaries keeping the number of overlaps unchanged. Also, in the division into four groups it is as well to keep the number of overlaps the same between all pairs of adjacent groups. Thus here, L could be added to the third group or O to the second. The whole grouping should be done with common sense, the rules providing no more than a guide.

It now remains to make the best possible comparison within each group. Sometimes the whole investigation will be repeated at several sites. This is an excellent development, though it is essential for enough replicates to be used at each site for the results to be studied separately.

In variety trials more than in most other kinds the experimenter should consider carefully how to treat his trees in respect of factors not being investigated in the trial. Thus, characteristics of vigour are less masked by light pruning than by heavy: close spacing too will promote disparities in vigour, while the thinning of fruit will often disguise quality differences.

In comparing varieties it should be borne in mind that with different genetical content, differences in the main characteristics of growth and cropping are only to be expected. The question is not "Do these varieties differ from one another?", but "Knowing they must differ, by how much?". Consequently, randomization is chiefly useful to prevent treatment means becoming biased on account of patterns of soil variation. It will be recalled (Sections 44 and 45) that in the designs for calibration trials the varieties were systematically arranged. Nevertheless, this argument should not be taken too far. With different scions, it is to be expected that the shoots of the trees will differ in characters genetically determined; but with different rootstocks it does not follow that the shoots will differ, though the roots must. Consequently, rootstock trials may require significance tests, but trials of scion varieties do not always do so.

Further, the final choice of a variety is often made on a balance of characters rather than any one outstanding feature. The unwary may here be tempted to use discriminant analysis but this is to mistake the problem. The task is not to find and to give weight to those varietal characteristics that are most distinctive but to those that are of most commercial importance. Thus, a variety with excellent flavour and appearance may be rejected because it travels badly and another selected which, though really excelling in nothing, has no notable defects. Consequently, statistical analysis, though valuable in confirming that a certain character is indeed chiefly associated with a certain variety, does not hold the place it does in other sorts of trial.

### 64. Trials of Pruning and Tree Formation

These call for very little comment. They can mostly be carried out using one-tree plots without guard rows and present few difficulties to the designer, though some experienced investigators are of the opinion that guarding is desirable because of the wide differences in height and spread that can be induced by pruning together with the consequent differential competition between the aerial parts of adjacent trees. The effect of shading is also sometimes such as to make guarding necessary. Perhaps the best compromise is to use rather wide spacing between trees, *i.e.*, narrow guard alleys rather than guard rows.

In such trials calibration is sometimes possible, but usually treatments should start as soon as the trees are planted. Apart from pecan trees [91] there seems to be little hope of calibrating without some delay after planting.

One difficulty arises from the great variation in shape of tree and types of growth that can be induced by different methods of pruning and tree formation and the consequent lack of comparative growth records. Thus, peach trees pruned in winter and summer are very difficult to compare in respect of vigour because no shoot measurements made on one set of trees can reasonably be made on the other.

### 65. Trials of Commercial Systems of Fruit-growing, including Spacing Trials

Trials of this sort are exceedingly difficult to carry out. Attention is usually directed to yield in relation to cost, and consequently each plot must be large enough to form an economic unit that can be managed and accounts kept as if it were a separate small-holding. Although in principle this merely calls for a large total experimental area, the practical difficulties are very great. Most information of this sort is in fact obtained from surveys (see Section 13) rather than from field trials.

Nevertheless, such trials have a better chance than most of paying their way and consequently they might well be the subject of co-operative experimentation. Thus, the writer has often thought that an investigation into different methods of growing, *e.g.*, into the best distance between plants or the desirability of putting in filler trees to be removed later, could well be carried out by a group of progressive growers, each of whom would undertake to plant several acres under each system and to keep roughly separated accounts for each parcel of land. The accounting need be no more than would be carried out by any grower who liked to know the sources of his income and expenditure. Although such an investigation would break the rule set out in Section 75 that every site should have enough replicates to give an independent answer, it would permit the analysis of profits as if they were data from a design in randomized blocks and would give information more valuable than that provided by other means. However, such an investigation has yet to be tried in practice.

Spacing trials, alone in this class, can be readily carried out at one research centre. One point to be noted here is the convenience of using planting distances that permit plots to be small yet of even size. Thus, if the proposed distances between rows were 8 ft., 10 ft. and 12 ft. it would be necessary to work with plots 120 ft. wide and containing respectively 15, 12 and 10 rows. If, however, the distances were adjusted to 8 ft., 9·6 ft. and 12 ft., the breadth could be reduced to 48 ft.: if to 7·5 ft., 9·375 ft. and 12·5 ft. the breadth could be only 37·5 ft. and so on. Plots of this sort need a single guard-row between them, which should be placed so that each of its neighbouring rows is spaced from it by the distance under test in that plot. Spacing trials are best carried out on a long, narrow strip of land with all the plots side by side to permit implements running through from one headland to another; but, if several tiers of plots are necessary, wide guard alleys should be left to permit turning, despite the consequent uneconomic use of land.

On account of the differing number of plants per plot, spacing trials can give rise to heterogeneous errors (see Section 73), but this can usually be overcome by one of the methods to be described, namely, partitioning both the treatments sum of squares and the error.

## 66. Trials with Several Kinds of Treatment

The use of factorial designs has led to the question how far several different trials should be combined into one big one. The common-sense answer would seem to be that where there is a reasonable expectation of interactions appearing, the factorial design is desirable. In analysing results, care should be taken to avoid giving too much attention to interactions that appear significant, although they are not under investigation (see Section 71).

Trials of new varieties are usually better not combined with other experiments. It is admittedly tempting to design, say, a manurial trial with large plots split into one-tree subplots for varieties. Although this is admirable for the new varieties, which are thus tested under a range of manurial conditions, it is disastrous for the manurial trial. If this had been carried out on known varieties, selected to give a wide range of characteristics, it would be possible *either* to assert, in the absence of an interaction, that a certain recommendation can be made for all varieties, *or*, if there is one, that it can be made for certain important ones. With new varieties the range may prove too limited to base any generalizations upon it, while an interaction cannot be given the same practical applications. No one wants to know that Seedling 999 is improved by high nitrogen though not Seedling 1001, but which varieties actually being grown may be expected to benefit. On the other hand, new varieties may be combined with another factor in special cases. Some species, for example, show an interaction between variety and method of propagation. Thus, new cacao varieties may well be tested using plants raised from both fan and chupon cuttings.

Nevertheless, it is often a good plan to carry out an investigation using a wide range of standard varieties. This is especially true of rootstock trials and of experiments intended to study methods of pruning and of tree formation. It should be emphasized that this is an application of a general rule that subsidiary factors introduced to extend the applicability of results for the main factor should be *representative* rather than *experimental*.

Although cover crop trials often have a manurial factor added, the interpretation can be rather difficult. Thus, nitrogen added to a grass and clover mixture can kill out the clover, though with low nitrogen the clover may thrive at the expense of the grass. In fact, the treatments do not remain what they purport to be and the resulting interaction can be very complicated.

[To follow the next chapter the reader should be able to divide up his various sums of squares according to any grouping of the treatments in which he may be interested. These methods are explained in Appendix II.]

# CHAPTER 7

# THE ANALYSIS OF RESULTS

" . . . to calculate is not in itself to analyse."

Edgar Allan Poe, 'The murders in the rue Morgue.'

## 70. Some General Considerations

An experimenter with a body of data is usually well advised to study it carefully before starting to analyse it. Quite a simple summary may show that the treatment differences are so small as not to be worth bothering about even if they should be shown significant. Of course, in trials of a fundamental nature the fact that there is an effect at all may be important in itself, but usually a difference is important only if it is large.

This same preliminary study may also suggest which periods in the life of the trial are most worthy of attention. Thus, after trees are grassed down differences may arise that disappear later, and consequently an analysis, to be of use, must be confined to the earlier data. Such concentration of attention is useful, provided it is done objectively and without any tendency to select for study those years in which the results confirm the experimenter's expectations.

Also, before analysing data it should be quite clear what is being investigated. In any test, even if applied to a set of random numbers, there is one chance in twenty of a significant result being obtained by chance, so, if enough tests are carried out, sooner or later something will be found significant and that something may well be spurious. The method of analysis to be adopted should be uniquely designated at the time the experiment is designed and should not be departed from. Even the preliminary examination of data, already recommended, should not lead to modifications in the designated series of tests except to suggest the omission of certain of them as needless.

The only exception to this rule occurs in a series of similar trials. Here, if an effect draws attention to itself in an early trial, it may be tested for in later ones. With annual species this exception is a large one because repeated experiments are customary, but with perennials a decision has usually to be reached on the basis of one experiment, although there may be several as, for example, when fungicides are tested in a number of different seasons.

A further point that needs emphasis is the rôle of the analysis of variance in the interpretation of data. In any trial, it enables the experimenter to judge the accuracy of his results and thus to form an opinion about the adequacy of his techniques and the reliability of the conclusions indicated by the data. Sometimes this is the sole reason for analysis, sometimes it is subsidiary to the need for significance tests. The purpose of these tests is to give an answer " Yes " or " No " to questions that were in the experimenter's mind when he was designing the trial. They should not be used to answer questions posed later after the data have been examined.

It should scarcely be necessary to mention that the unit of a field trial is the plot (or, in split-plot designs, the sub-plot) and therefore data for analysis should first be expressed as either plot totals or plot means. Where there are several trees to a plot, each measured separately, the individual figures are not needed.

When the analysis is finished some discretion is called for in its interpretation, especially in respect of significance levels. It is a great mistake to suppose that treatment differences are either in the black shades of non-significance or the bright light of significance, because really there is a gradation. At the one extreme, the treatment mean square may be less than the error mean square (see Appendix I), *i.e.*, $F$ less than one, in which case there is no evidence at all that the treatments are having any effect; but as $F$ increases above unity the evidence becomes stronger. It is a sensible practice as well as a convenient one to think of significance as a series of steps denoted by one, two or three asterisks as $P$ falls successively below 0·05, 0·01 and 0·001 (see Section 12), because it serves as a reminder that significance is a matter of degree and not an absolute quality, but even these steps can disguise the fact that really there is a continuous slope with no sudden jumps in it at all.

## 71. The Analysis of a Single Experiment

With perennial species it is not usually practicable to do an experiment more than once. It is true that several years' results may be obtained from the same trees but this is not the same thing, for the successive figures from a plot are not independent. Consequently, there is no possibility of confirming doubtful results on a later occasion and some care is needed in interpretation. Also, since there is no question of basing results on the concurrence of several trials, the concept of significance, *i.e.*, the probability of such data as were obtained arising by chance on one particular occasion, loses none of its force.

It is, however, necessary to be clear which effects are really under investigation. Thus, if several factors are included together in order to study their two-factor interactions, there may well be a host of higher order interactions that arise from the experimental design rather than from any concern to investigate them. If there are a lot of them and each is tested separately, some are almost bound to appear significant since each has one chance in twenty of doing so by pure hazard. If, on the other hand, some comprehensive test is carried out first, there is the risk of a significant effect being lost in a swarm of non-significant ones, the whole being non-significant. With repeated trials, if one of these interactions draws attention to itself, it can be looked for on later occasions, but what if there is only one trial?

A similar difficulty arises when the levels of a factor form natural groups. Thus, if an experimenter is studying the action of six fungicides, three derived from copper compounds and three from sulphur, it might well be that there are no differences within either group but a marked difference between them. This last could easily be missed if the five degrees of freedom between fungicides were tested together, the one degree of freedom associated with the difference between groups not being separated from the four degrees of freedom associated with the non-significant differences within the two groups. Again, the position is easier if the trial is going to be repeated, the difference between the groups appearing every time, though the whole is never significant.

There does not appear to be any completely satisfactory answer to these questions, though several have been advanced [*e.g.*, 32, 56, 98, 181]. The practice at East Malling, which is not necessarily better than that elsewhere and may be worse, is to *nominate* right from the beginning those effects that are under study, always to partition the treatments sum of squares (see Appendix II) to isolate these effects and to test each separately. It may be objected that if several effects are nominated, the probability of the trial as a whole showing something significant by chance is greater than one in twenty or whatever the significance level may be; but this objection is really without substance. If there are two factors and each had been tested in separate trials and then the two had been brought together to test their interaction, the probability of something appearing significant by chance in one or other of the three trials is more than one in twenty and the position is not really altered by the three trials being carried out in one operation. (The probability is not the same in the two cases, because in one there are three independent errors and in the other only one, but the principle is the same in both.)

In a factorial trial the usual practice at East Malling is to regard each main effect as nominated (or why was the factor included at all?) and also each two-factor interaction (or why were the two factors included together in one trial?). Higher order interactions are usually unnominated, but exceptions may be made either way.* Each nominated effect is tested separ-

* Sometimes a main effect needs neither to be nominated nor unnominated. Thus, in a pruning trial it might be thought desirable to include a wide range of rootstocks of known characteristics in order to compare the treatments on trees of varying vigour. There would be no point in testing the significance of the rootstock differences because the whole point of having the rootstocks there at all lies in their being known to differ. Indeed, if the rootstocks were planted as the main plots of a split-plot design their comparison might be very insensitive, so much so that there would be little likelihood of a significant main effect, but this would not matter, the interest lying in the interaction.

ately, but the unnominated effects are tested together by adding their degrees of freedom and sums of squares and carrying out an $F$-test on them all. Only if this comprehensive test gives a significant result are the unnominated effects studied separately to find out which is the operative one. In this way the chance of being misled by a random result is kept within bounds. Where there are only three factors, the last order interaction automatically acquires the status of a nominated effect because it is the only one in the comprehensive test. In split-plot designs the comprehensive test is made difficult if the unnominated effects are spread over two or more analyses and thus have different errors. A test does exist [33] but is not of a sort to be undertaken lightly.

In the other case, where the levels of a factor form natural groups, the practice at East Malling is to nominate the groups. Then tests are carried out to see if there are differences *within* any of the groups. If one of them does appear to contain diverse components, it is split according to the differences found. Finally, testing takes place on the differences *between* groups. The computing is illustrated in Appendix II.

Others may prefer systems of their own devising and certainly practice varies a great deal. What is important is that testing should proceed according to a regular plan, drawn up when the trial was designed to answer the questions being asked. This is not to say a research worker should be uncritical. If A interacts with B, B with C and C with A, most people would want to know something about the three-factor interaction—but where there is no chance of securing confirmation of doubtful results the experimenter must come down on one side of the fence or the other and should act so as to be able to rebut any accusation of caprice. In particular, he must not partition and repartition until he finds something. With sufficient ingenuity and determination it is possible to obtain results in this way from almost any trial: the right way is to think what questions are being asked and how they can most sensibly be answered.

Although the analysis of results must not be carried out arbitrarily but in a regular manner, care must be taken to adapt the test to the question being asked. For example, a horticulturist may know that an inorganic compound fertilizer gives beneficial results, but be in doubt whether this is the result of the nitrogen or the potash it is known to contain. Consequently he designs a factorial trial to find out, and the trial might well be identical with that of another horticulturist, who is inquiring if nitrogen and potash bring about any improvement. Despite the identity of designs, the method of interpretation will not be the same, for with one investigator it is an article of faith that there must be a real effect somewhere, while the second is ready to believe that no addition of either element is called for. Thus, if no effects prove significant, the first would still accept as real the one that comes nearest to significance. (The position would be most unsatisfactory for him and further investigation would certainly seem called for, but his choice would be quite reasonable.) It is not unusual for past knowledge to influence the interpretation of results.

Another example lies in the so-called " one-sided " tests where it is believed that Treatment B may be better than A, but can hardly be worse.* For example, B may be two applications of a fungicide and A only one. Again, a modification of technique is called for. The procedure is to work with twice the significance level ($P = 0{\cdot}1$ instead of $0{\cdot}05$, $0{\cdot}02$ instead of $0{\cdot}01$ and so on) and to ignore significant results (half of them, if chance alone determines the differences) that go the wrong way.

These are by no means the only examples that could be given. To sum up, an experimenter with no opportunity to check doubtful results is best guided by some standard procedure, though he should be careful to distinguish the cases to which it will not apply. It should be added that the interpretation of results is still a matter of insight even where the significance levels are objectively determined.

### 72. Significant Differences

It is an old practice to take the standard error of a treatment difference and to derive a " significant difference " from it by multiplication with the quantity $t$ taken from tables [66*a*]

* It is rarely possible to be certain that this is so. Thus, an additional spray may not affect an insect pest but may kill its predators. Again, two applications of a fertilizer are not necessarily better than one.

according to the significance level required and the number of degrees of freedom for error. The computation of significant differences is discussed more fully in Appendices I, III and IV, so this Section will be concerned only with the occasions for their use.

First of all, there is no need to use significant differences if there are only two treatments. In that case the information given by the $F$-test that the treatments do or do not differ is all that is needed and there is nothing to be learnt by further testing. In fact, the tests by $F$ and by significant differences are then exactly equivalent.

The use of significant differences calls for consideration in all cases where there are three or more treatments and an $F$-test has shown that there are differences between them. Here the conventional method, at least as usually understood, is to work out a significant difference between treatment means and to group those treatments which differ by less than this amount. Thus, if the significant difference is 1·3 and the treatment means are

| A | B | C | D |
|---|---|---|---|
| 14·1 | 15·0 | 17·2 | 17·4, |

it would be concluded that A and B form one indistinguishable group, C and D forming another.

This method has recently been attacked by Tukey [164], who appears to have the support of Snedecor (150], the proposed alternative being a combination of significant differences used as a gap test, Nair's straggler test [99] and a repetition of the $F$-test. Keuls [87] also has expressed disagreement with the conventional method, his remedy being an adaptation of Newman's $q$-test [100], otherwise known as the use of "Studentized range." The present writer, who also is not satisfied with significant differences as commonly used, though for quite different reasons, has been in correspondence with both Prof. Tukey and Mr. Keuls, and it is clear that the subject as at present understood gives rise to legitimate differences of opinion, which are not, however, as great as would appear from the literature. Consequently, the following paragraphs should be taken as setting out the personal view of the writer.

Where an $F$-test has shown the treatments to be having a significant effect it would appear reasonable to conclude that all the treatment means are different one from another. This may be seen on statistical grounds alone, because the natural alternative to the hypothesis discredited by the test, namely, that all the treatment parameters are equal (see Appendix III), is that they are values drawn from a continuous distribution, in which case the probability that any two will be equal is zero. It may be seen no less on horticultural grounds. If, for example, it has been shown by an $F$-test that rootstocks affect the leaf colour of the scion and if two trees are on different rootstocks, it is to be expected that they will have different colours of leaves. Again, if it is clear that the form of nitrogen in a fertilizer affects growth, and if two plots have been given their nitrogen in different forms, it is to be expected that growth will be different on the two plots. The next step is not to do any more *testing*, but to *estimate* the difference, whether it be large or small, important or unimportant.

It should be noted that this line of argument does not lead to an abandonment of significant differences but to a modification of their use. If two treatments give means that do not differ significantly, there should be no suggestion that they necessarily have the *same* effect, as in the conventional approach. Usually they are having *similar* effects which the experiment is not sensitive enough to distinguish, but it is none the less salutary for the experimenter to be reminded that *he has not demonstrated* any difference for that particular pair of treatments and it calls upon him to pause for a moment and think, because sometimes two treatments that are different in application are virtually identical in action. Thus, two annual cover crops may be quite different botanically, but if both germinate too late they may be equally ineffective and virtually identical in action. Again, two sets of pruning instructions may look very different, but both may lead to much the same wood being removed from the trees. An experimenter confronted with two treatments that do not differ significantly can well ask himself if they are really as diverse as they appear to be. If, however, his consideration of the matter does not lead to any satisfactory reason for thinking the two treatments to be identical in their action, he should take the difference in their effects to be a real one. It is assumed throughout that the $F$-test has shown a significant effect of the treatments in general.

The grouping of treatments effected in this way should be clearly distinguished from that considered in the preceding Section. There the groups were apparent from the first, and the experiment investigated whether or not they corresponded to differences in tree behaviour. Here there was no initial reason to expect any treatment to behave like one of its rivals rather than another, but the similarity of data has led to a reconsideration of this point.

This division of opinion on the use and meaning of tests has naturally led to diversity of methods in the presentation of the results of statistical analyses. Many laboratories make a practice, when results have been shown significant by an $F$-test, of reporting merely the treatment means and their standard error. This latter figure enables the reader to see how accurate the experiment was and also enables him to work out a significant difference for himself if he wants one. The writer, for his part, would prefer to separate these functions of the standard error by presenting two figures, (*a*) the standard error of a single observation, namely, the square root of the error mean square, to represent accuracy, and (*b*) any significant differences that may be relevant, already worked out. This also has the effect of concentrating attention on the differences between treatments rather than on their absolute performance, which on balance is a good point. However, all these quantities can be calculated very easily one from another, and to dispute which is to be preferred is rather like making it a matter of principle whether to express weights in pounds or grammes. The important thing is that the reader should be given the essential information in the way he will most readily appreciate, taking care that he thoroughly understands what use can and cannot legitimately be made of the quantities cited.

Variety trials present rather a special problem in the use of significance tests, because it may be taken for granted that plant tissues having different genetical contents will differ in respect of most inherited characteristics. It has already been suggested (see Section 63) that significance tests are not very informative in variety trials and this is especially true of significant differences. There may be some doubt if varieties do in fact differ in a certain respect and here an $F$-test can be helpful; but when once it has been shown that they do, there can be little chance of two varieties behaving in an identical manner unless they happen to be synonyms.

Although it is in general wrong to start looking for significant differences before it is clear that the differences as a whole are significant, *i.e.*, before a positive result has been obtained from an $F$-test, an exception arises when certain differences are marked out from the beginning as being of special interest. Thus, suppose that the main purpose of a trial was to study the difference of two treatments, A and B, certain other treatments, C, D and E, being added in a purely exploratory frame of mind. In East Malling practice it would be considered permissible to nominate the difference between A and B and to test it by a significant difference even though the similarity of C, D and E had led to treatments as a whole proving non-significant by the $F$-test. Thus, if the significant difference was 4·8 and the treatment means were

| A | D | E | C | B |
|---|---|---|---|---|
| 42·7 | 46·1 | 46·8 | 48·4 | 49·4, |

the difference between A and B would be accounted significant regardless of the result of the $F$-test.

This, incidentally, provides a good illustration of the idea behind nomination, namely, obviating dilution, as it were, of effects primarily under investigation by others of secondary interest. Thus, if there were no chance to nominate, an experimenter would hesitate to combine two sets of treatments into one factorial design lest the non-significance of one should cover up the significance of the other.

One special case requiring mention concerns the comparison of a control with a range of other treatments, the object being to find if any of the treatments are different from the control rather than if they differ from each other. Examples of such trials are given in Appendices I and III. It can happen that one treatment is markedly different from the control, which closely resembles the rest. Is the effect of this one treatment to be ignored, because the differences between treatments as a whole are not significant? One view is that the comparison of each treatment with the control has been nominated from the start, and a significant difference may be applied even if the $F$-test for treatments has not shown significance. If it be objected that this gives a high chance that one or other of the nominated effects will prove significant even in a blank trial, the writer's reply would be that this always happens in an experiment with several effects under test. If the investigator had designed a series of independent trials, one for the comparison of A with the control, another for B and the control and so on, no objection would have been made. Yet the chance of one or other of the treatments appearing by chance to differ from the control is not very different from the case of the comprehensive trial with all these effects nominated.*

This view is not universally held and Tippett [163] has devised a different test, as also has Nair [99], intended to keep within bounds the chance of one or other of the treatments showing itself significantly better than the control by sheer weight of numbers, as it were. Thus, if there are 20 treatments it is to be expected that one of them will appear significantly different from the control at the level, $P = 0{\cdot}05$, if no special test is used. Neverthless, for the reason given, the present writer would nominate.

In split-plot designs, the computation of significant differences needs a little care. If a factor X has been applied to main plots, the significant difference between means representing the main effect of X should be calculated using the error mean-square and error degrees of freedom for the main plot analysis, *i.e.*, the first analysis. Equally, if a factor Y has been applied to sub-plots, the significant difference for the main effect of Y should be derived from the error of the sub-plot analysis, *i.e.*, the second, and so on. The difficulty comes with the interaction, X $\times$ Y. For the comparison of two means representing treatment combinations having the same level of X but different levels of Y, it is correct to use the error for the sub-plot analysis. If the treatment combinations differ in X, whether they are or are not the same in respect of Y, for a design in randomized blocks the error may be obtained thus [158]:

The appropriate error mean-square is

$$\frac{s_1^2 + (L - 1)\, s_2^2}{L},$$

where $L$ is the number of levels of Y, $s_1^2$, is the main plot error mean-square and $s_2^2$ is the sub-plot error mean-square. To find the number of degrees of freedom, $F$ (which is not in general going to be a whole number), it is first necessary to work out $c$, which equals

$$\frac{s_1^2}{s_1^2 + (L - 1)\, s_2^2},$$

and then $F$ is given by the formula,

$$\frac{1}{F} = \frac{c^2}{f_1} + \frac{(1 - c^2)}{f_2},$$

where $f_1$ and $f_2$ are respectively the degrees of freedom of $s_1^2$ and $s_2^2$. The appropriate expressions in other cases can usually be derived from the work of Satterthwaite [143] or that of Welch and Aspin [4, 5, 170, 171].

### 73. Heterogeneous Errors and Transformations

In the analysis of variance it is assumed that the variability of all data from their expected values is the same irrespective of treatments, but this is often not justified, *e.g.*, if a treatment makes apple trees heavier, their variability in weight will also be increased [109, cf. 136]. One way of overcoming this, available only in randomized block designs, is to partition the error as described by Cochran [31] and well illustrated by Wadleigh and Tharp for a cotton trial [167]. In this method, the treatments sum of squares is partitioned into effects each with a single degree of freedom and the interaction of each effect with blocks worked out (see Appendix II).

* It will not be quite the same because, with the series, the difference between A and the control is independent of the difference between B and the control. In the comprehensive trial this is not so because, if by chance the control gives a low value for comparison with A, it will have a low value to compare with B also.

These interactions will be found to add up to the error sum of squares of the randomized block design. Each effect is then compared by an $F$-test, not with the whole error, but with its own interaction with blocks. This procedure is admirable provided two conditions can be fulfilled, namely, it must be possible without artificiality to partition the treatments into single degree of freedom effects and there must be enough replicates for the interactions to have a reasonable number of degrees of freedom each, since each is to be used as an error on its own account.

Another method, also considered by Cochran, is to use some function of the data that is statistically more acceptable, a process known as "transformation". Thus, in the example given in Appendix II of a trial for the study of methods of soil management, the weight of prunings, $n$, was actually analysed as $2 + \log(n + \frac{3}{8})$, in order to get equal variation within the treatments.

The range of possible transformations is very great. Thus, it sometimes happens that the standard error of a plot is proportionate to the mean of the treatment to which it belongs, as is usually the case when analysing growth measurements, such as trunk girth, length of extension growth and weight of tree [109]. This state of affairs can often be revealed in the following way: For each treatment write down the mean value of its data and also the extreme range of its data, from which table it may appear that the ranges vary more or less with the means. It is then appropriate to analyse, not $x$, the actual data, but $\log x$ [31]. Each plot figure is transformed to its logarithm and the analysis is then done in the usual way.

Sometimes, *e.g.*, with insect counts [177], this transformation is useful but a difficulty arises on account of $\log 0$ being equal to minus infinity. Where $x$ can equal zero and is in any event confined to whole numbers (*i.e.*, it can equal 1, 2, 3, etc., but not 2·4, 3·1, etc.) various alternatives have been suggested, *e.g.*, $\log(x + 1)$ [177] and $\log(2x + 1)$ [111], but the latest research favours $\log(x + \frac{3}{8})$ [3]. Where the data can assume fractional values but are discontinuous (*e.g.*, where plant weights are taken to the nearest quarter pound, the plants being quite small so that the discontinuity is serious), it is still possible to use this approach by first expressing all data in terms of the real unit of measurement, *i.e.*, in quarter pounds, calling $2\frac{3}{4}$, 11 and $3\frac{1}{2}$, 14 and so on.

Quite a different sort of transformation is called for when the ranges appear to vary proportionately to the square roots of the means, as sometimes happens with crop weights, for here $\sqrt{x}$ is the basic transformation [31]. If $x$ is confined to small whole numbers (or its discontinuities can be made to correspond to unit intervals, as described in the paragraph above), the transformations $\sqrt{x + \frac{1}{2}}$ [7] and $\sqrt{x + \frac{3}{8}}$ [3] have both been recommended, but it appears that $(\sqrt{x} + \sqrt{x + 1})$ [67] is better than either.

In dealing with percentages (and it is here assumed that they are such as must lie between 0 and 100) the variability is greatest round about 50 and falls off towards the extremes. If all or nearly all lie between 15 and 85, the error will be so nearly homogeneous that no transformation is needed, but if there are a number of extreme values it is necessary to use *angles of equal information*, $\theta$, [17, 18, 31] such that $\sin^2 \theta = k/n$, where $n$ is the number of specimens examined per plot and $k$ is the number observed to have the characteristic under study. This transformation has been fully tabulated [66*d*]. Where discontinuity is serious, it has been suggested [7] that $k$ should throughout be increased or decreased by $\frac{1}{2}$, whichever is needed to bring it nearer to $\frac{1}{2}n$. Later, however, it was pointed out that it would suffice to transform 0 as if it were $\frac{1}{4}$ and $n$ as if it were $(n - \frac{1}{4})$ and otherwise to use the basic transformation unmodified [8*b*], and this simple device remains very useful. The recommendation has also been made to use $(\theta + \theta')$, where $\sin^2 \theta = \dfrac{k}{n + 1}$ and $\sin^2 \theta' = \dfrac{k + 1}{n + 1}$ [67], but its advantages have not yet been established. With discontinuous percentages that include extreme values, perhaps the best transformation is $\theta$, such that $\sin^2 \theta = \dfrac{k + \frac{3}{8}}{n + \frac{3}{4}}$ [3]. A full discussion of the treatment of data in percentage form has been given by Clark and Leonard [29].

The whole subject of transformations has recently been gone into in great detail by Bartlett [11], whose paper well repays study.

It should be emphasized that transformation, if needed, must take place right at the beginning of the analysis, all fitting of missing plot values, all adjustment by covariance, etc., being done with the transformed variate and not with the original data. At the end, when the conclusions have been reached, it is permissible to "back-transform" the results so as to present them in the original units of measurement, but this is done only to render them more intelligible.

As a result of this process of transformation followed by back-transformation, the means will be rather different from those that would otherwise have been obtained. Thus, to take a simple example, without transformation the mean of the numbers 1, 4, 9, 16 and 25 is 11. Suppose, however, that a square-root transformation is used to give 1, 2, 3, 4 and 5, the mean is now 3, which after back-transformation gives 9. Usually the difference will not be so great because data do not usually vary as much as those given, but logarithmic and square-root transformations always lead to a reduction of the mean, just as angles of equal information usually lead to its moving away from the central value of 50%.

Although transformations make possible a valid analysis, they can be very awkward. For example, although a significant difference can be worked out in the usual way for means of the transformed data, none can be worked out for the treatment means after back-transformation. Brackets can be used to show which differences are not significant (basing these on the result with the transformed data), but that is about all. Again, it is not clear what intelligible meaning is to be given to interactions when the data have been transformed. Fortunately, *log x*, one of the most useful, is free from these strictures, because a significant difference in logarithms can be translated into a significant ratio for the back-transformed values, while the absence of an interaction between A and B means that A has the same *relative* effect whatever the level of B. With data having no transformation, it would mean that the *absolute* effect was constant.

## 74. The Analysis of Several Years' Results

One feature of statistical work with perennial species is the accumulation of successive seasons' results for each plot and it sometimes needs a little thought how to make best use of them.

It should first be commented that there is no need to look for difficulties. Usually it is quite enough to add together all the data from a plot to obtain total crop or total growth, whatever it may be, over a period. With size records, similarly, a figure representing the increment over a period may give all the information required. Trunk girths in particular lend themselves to this treatment. They are usually dealt with in logarithmic transformation and the increment in logarithm of trunk girth is, of course, a measure of rate of growth—a figure of some importance and interest. Usually no more complicated technique need be employed.

If the time taken by the experiment is to be divided into periods, it is best for these to be equal in length and for each to cover an even number of seasons. Most perennial species are to some extent biennial in cropping and growth and, consequently, periods containing an odd number of seasons are rarely comparable one with another.

One useful device for analysing together results for successive periods has already been mentioned (Section 25), namely, regarding the periods as a further factor split upon the smallest plot laid out in the field. Thus, if the trial is already in main plots and sub-plots, the individual results for periods would be regarded as the results of sub²-plots split upon the sub-plots. The difficulty is that variability in a trial is itself varying, and the resulting heterogeneity of the error of the last analysis (see last Section) cannot usually be dealt with by any transformation. Also the years are best regarded as a treatment applied systematically, so the appropriate method of analysis is that given in Section 46.

Rather than use the other method for dealing with a heterogeneous error, namely, partitioning the levels of the factor into effects each with one degree of freedom and partitioning the error accordingly, so that each effect has its own characteristic error (see last Section), there is much to be said for the following method, used with excellent effect by Stevens [152] with coffee. Suppose there are two periods into which the trial has been divided, which may be designated by (i) and (ii). First of all, an analysis can be done on (i) + (ii), which will give a measure of the total yield or growth or disease, as the case may be, over the whole experimental period.

Then a separate analysis can be done on (ii) — (i), representing the build-up of the variate with time. Again, if there are three periods, analysis can usefully be done on (i) + (ii) + (iii), (iii) — (i) and (i) — 2 (ii) + (iii), representing respectively the total over the experiment, the degree of increase during the experiment and the extent to which the middle period gave results above or below the mean of the other two. This last can be important for, if results from the second period are much above the mean, the conclusion is that the quantity, having built up quickly, is later increasing more slowly. Thus, if the figures for crop over three periods are 10, 30, 35, the suggestion is that large further increases in cropping are not to be expected ; whereas figures like 10, 12, 35, give hope of larger crops for the future.*

A special case of this approach is provided by working out an analysis of variance on the regression coefficient of the variate upon time. Thus, suppose crops over four successive periods for a certain plot are 10, 13, 15, 16. Using the ordinary designation of the periods as 1, 2, 3, 4, this gives a regression coefficient of (see Appendix IV)

$$\frac{(10 \times 1) + (13 \times 2) + (15 \times 3) + (16 \times 4) - \frac{1}{4}(10 + 13 + 15 + 16)(1 + 2 + 3 + 4)}{1^2 + 2^2 + 3^2 + 4^2 - \frac{1}{4}(1 + 2 + 3 + 4)^2}.$$

The denominator may be ignored as being the same for all plots. Further, the numerator can be simplified by renumbering the periods in such a way that their numbers are still evenly spaced but add up to zero. Thus, in the present example, renumbering the periods $-3$, $-1$, $+1$, $+3$, gives the expression,

$$(10 \times -3) + (13 \times -1) + (15 \times +1) + (16 \times +3).$$

This is just as good as the regression coefficient for showing the relative degree to which plots are giving an increase in crop as time goes on.

In the investigation of Stevens, already referred to, each plot gave a crop record for ten successive years. First, he analysed the total crop,

$$\text{(i)} + \text{(ii)} + \text{(iii)} + \text{(iv)} + \text{(v)} + \text{(vi)} + \text{(vii)} + \text{(viii)} + \text{(ix)} + \text{(x)}.$$

Then he analysed the difference in cropping between odd and even years to investigate the degree of biennial bearing,

$$\text{(i)} - \text{(ii)} + \text{(iii)} - \text{(iv)} + \text{(v)} - \text{(vi)} + \text{(vii)} - \text{(viii)} + \text{(ix)} - \text{(x)}.$$

Finally, he grouped his years into pairs in order to remove the biennial effect and worked with a regression coefficient, as modified above, to find the extent to which cropping was improving with time, *i.e.*, he analysed

$$-2\,\text{(i)} - 2\,\text{(ii)} - \text{(iii)} - \text{(iv)} + 0\,\text{(v)} + 0\,\text{(vi)} + \text{(vii)} + \text{(viii)} + 2\,\text{(ix)} + 2\,\text{(x)}.$$

In this way he studied the important effects associated with time without artificiality and without making any dubious assumptions about the homogeneity of errors.

Another approach that has been tried at East Malling is to regard each year as providing a different variate and to carry out a multivariate analysis [*e.g.*, 10] with its apparatus of variances and covariances for all variates and combinations of variates. The results, though interesting, were difficult of interpretation and hardly repaid the large amount of computing required. Nevertheless, the method might justify itself in instances where the correlations between successive year's results were high but cut across by treatments.

### 75. Experiments at Several Sites

Here, as in the last section, there is no need to look for difficulties. One cardinal principle is that the trial at each site must be large enough and sensitive enough to give an answer without

* This argument should not be taken too far. Thus, the writer feels dubious about a trial [151] in which conclusions concerning asparagus crops over a long period were in fact based upon yields over the beginning of the period. With most species changes in competition and shelter during their life produce cropping conditions that cannot be forecast so precisely from the early records.

help from outside. Then, if that trial leads to results different from the rest, it can be considered in isolation. Much of the trouble with co-operative experimentation at different sites appears to arise from a neglect of this simple consideration, the search being too often for varieties and treatments that will *on the average* do well, not for those that will succeed *in a specified environment.*

As a formal statistical problem it is easy to see how to test which treatment is best over an area. Sites are chosen at random and an experiment conducted at each. Results are put on to some comparable basis (*e.g.*, crop as pounds per tree or disease as lesions per leaf, etc.) and then written out in a two-way table with sites one way and treatments the other. The whole is analysed as if it were a randomized block design with sites instead of blocks, any conclusions reached about the treatments referring to their mean behaviour over the whole area of the country, province, etc., from which the sites were chosen.

In practice, of course, the sites are not wholly at random, nor is anyone usually interested in the rather academic question of which variety or treatment would be best if only one could be used in the entire area. Sites should, in fact, be chosen so as to be representative of a region and, if Variety A does well in the trial on clay in the South-West, it should be recommended to farmers on clay soils in the South-West despite its failure on the light loams of the North-East. Of course, if an experimenter wishes to know if a variety or treatment is really of value in a specified region, it is a good practice to try it out on a number of sites first before making a recommendation. However, if it fails or is disappointing at some sites, further research is called for to find out why. It will be no consolation to a farmer who has lost 50 lb. of fruit per tree to be told that the regional average has been raised, because for every farmer in his position there are two whose crops have been increased by 40 lb. per tree.

Where it is desirable to analyse together results from several sites, the studies of Cochran [30] and of Roessler and Leach [137] will be found helpful.

# CHAPTER 8

# SOME MISHAPS AND REMEDIES

" . . . of most disastrous chances,
Of moving accidents by flood and field ; "

William Shakespeare, ' Othello,' 1.iii.136–137.

## 80. Incidence of Damage, Disease and Pests

After a trial has been planted, it sometimes happens that some injurious effect is found on the trees and is cutting across the results. Thus, a disease may begin to creep in from one side or the more exposed trees may be damaged by wind. Frost damage may be serious, or harm may have been done by a wide range of animals from red spider mites to elephants.

Before trying to eliminate the effect of such interference it is necessary to consider how it is related to treatments. For example, in a variety trial, a disease may well attack some kinds in preference to others and an elimination of its effect would be unfair to the more resistant varieties. Again, in a pruning trial, branch breakage might well be a consequence of one of the treatments and no allowance should be made for its deleterious effects. In general, with a few possible exceptions, the effect of the damage should be allowed for in only two cases.

(1) It is clear that the incidence of injury is not related to the treatments. Thus, a disease is found in the trees of an experiment on the side adjoining a wood where the pathogen is known to be present. If the patch showing symptoms appears homogeneous, the disease thinning out from a centre, this assumption would appear reasonable. Of course, the only circumstances in which an experimenter can be completely certain that his treatments have not affected the incidence of disease or injury are those in which the trouble was measured before the treatments were applied, and even here the treatments may affect its ultimate development.

(2) The trouble is of such rare occurrence that a treatment is not thought the worse of because it makes the plants susceptible to that kind of damage. Thus, in areas where spring frosts are almost unknown, blossom injury to an early flowering variety might be regarded as a misfortune, irrelevant to any assessment of the value of the variety. Again, quarrying operations may cause dust injury, but it would not be fair to penalize a treatment that proved to be especially susceptible. That is not to say that trials can satisfactorily be carried out in these unusual conditions.

Where allowance for injury is to be made, there are several ways of doing it. One is to leave whole blocks, rows or columns out of all analyses of variance, supposing the design is robust enough to allow of this (see Section 21). Certainly, the surest method of eliminating the effect of some interfering factor is to ignore data from the blocks where it operates, and for this reason the recommendation has already been made (see Section 54) to plant more replicates than will be needed if all the trees remain available for study. Usually, however, this method of allowing for injury is not convenient because the infected area spreads over too many blocks. The remedy then is to measure the injury and adjust the results by the method of covariance (see Appendix IV).

## 81. Subsequent Classification and Pseudo-variates

A somewhat similar problem arises when it is found too late that the trees differ in some way not considered while the trial was being designed. Thus, a set of rootstocks supposedly from the same clone might be found to consist of a mixture of two identifiable clones. Sometimes, too, a guard row has to be removed on account of the needs of an adjacent experiment, thus classifying the remaining trees into "inside" and "outside". The best way of dealing with such mishaps is to associate each tree with a *pseudo-variate* [see. 128], which is 0 if the tree belongs to one class and 1 if it belongs to the other, to add up the tree values to give plot values and to adjust by the analysis of covariance.

Thus, suppose there are four rows each of five trees in an experiment and the north and east headlands become unguarded. Since the two headland effects may well be different, each must have its own pseudo-variate, thus,

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | 1 |

and a double covariance will be needed.

In working out adjusted means, it is usually more sensible to adjust on to a standard value of zero for the independent variate rather than on to its general mean, *i.e.*, to adjust all treatment means to the value of $x$ representing the rootstock that should have been planted or the circumstance of remaining guarded rather than to the mean of what actually occurred. This makes a constant difference to all the adjusted treatment means but does not alter the computations set out in Appendix IV in any other way.

Pseudo-variates can also be used at times to take account of soil variation. Thus, in one raspberry variety trial designed at East Malling difficulty was experienced from a streak of gravel that lay across the area. Plots on good land were assigned the value 0, plots wholly on the streak were assigned the value 1 and a few that lay on the boundary were given values like $\frac{1}{4}$, $\frac{1}{2}$, according to the proportion on gravel. Adjustment by covariance led to a marked improvement in precision.

## 82. Missing Trees and Plants

If a tree dies or has to be removed on account of disease, it is advisable to put another in its place, the new one being known as a *replant*, a *supply* or a *recruit*. The consequences when a *whole* plot is thus lost will be discussed in the next Section, this one dealing with incomplete plots.

If the losses are probably due to the effects of some of the treatments, it would be wrong to make any adjustment except in the circumstances considered in Section 80. That is to say, trees should be replanted as occasion arises and analyses should be carried out on plot totals without adjustment of any kind for the number of gaps or replants, the crop, etc., of the replants being counted in with the other trees.

If the effect of missing plants is to be allowed for, there are two ways of attempting it, but first it should be mentioned that the effect of gaps on adjacent plants depends very much upon the species. Thus, for strawberries the effect, if any, is small [160], but for hops it is appreciable [2]. It has been suggested that it is important for pecan trees also [145].

The best way of allowing for the loss of plants is to use the analysis of covariance, adjustment being necessary both for the number of gaps or replants and for the effect of competition. Two alternative methods suggest themselves:

(1) To work with *plot totals*, ignoring replants, and to adjust upon the number of gaps or replants as the case may be. This is open to some criticism on the ground that the loss of a plant from a plot that is doing well is more serious than from one that is doing poorly, thus vitiating the assumption of the analysis of covariance that the regression coefficient is not affected by the treatments.

(2) To work with *the mean per surviving plant* for each plot, again ignoring replants and again adjusting by the number of gaps or replants. The advantage of this method lies in the number of survivors having already been taken into account, so the experimenter is free to

concentrate on devising an independent variate that shall express the competition between plants. Thus, he may weight gaps by the time since their occurrence, assigning a value 5 to one that appeared five years ago and ignoring one that occurred yesterday. Again, given three gaps in a row he may ignore the middle one as having no effect on the competition of the survivors. Admittedly, he must not become too ambitious and lose himself in a welter of trivialities and he must remain objective, not seeking reasons why his favourite treatment was really best despite all appearances, but he is much freer to devise a useful independent variate. The criticism of the first method applies here also and is very difficult to counter, though it should be remembered that the differences most seriously affected are the large ones, which are obvious anyway.

Generally, it is advisable to replant as soon as possible in order to restore root competition for the adjacent trees ; but, where most losses occur while the trees are young, it is sometimes better to wait until no more are expected and then to replant all the gaps together. The replants are now comparable with one another, so their crops, etc., may be counted in with the plot totals, the difference between them and the original trees being dealt with by a pseudo-variate.

In some crops, *e.g.*, cacao, the number of trees to a plot of standard area can be very variable in commercial holdings. In such instances it is advisable to work with plot totals and to adjust by stand, supposing the latter to have been determined before the treatments were applied. The reason for preferring the first method to the second is that, with no standardized planting distance, attention has to be fixed on results per unit area. This is not really within the province of this Section though it is a closely related subject.

The difficulty with all these methods is their tendency to give heterogeneous errors (see Section 73) that no transformation can deal with. If one plot has 10 trees surviving out of 12 and another only two, it is absurd to suppose that the accuracy of the result for one plot equals that of the other. If too many plants are lost from a plot, it is best regarded as wholly missing.

### 83. Missing Plots

Although the loss of complete plots from an experiment is not a disaster and, with some species, is quite usual, it should be emphasized that *nothing can replace the lost information.* It is a bad mistake to suppose that a few simple modifications in the computing will set all to rights. For one thing, the modifications are not always simple : for another, something must always be lost from the accuracy of treatment comparisons.

The various methods of dealing with incomplete data will now be enumerated, leaving the most usual, namely, the calculation of missing plot values, to the end. Basically, they are all variants on the same method.

(a) *The pseudo-variate method :* In this, each missing plot is assigned some arbitrary figure, it does not matter what, and then each is associated with a pseudo-variate [8*a*, 97]. Thus, if the data were

| *Treatment* | *A* | *B* | *C* | *D* | *E* |
|---|---|---|---|---|---|
| Block I . | 14 | $M_1$ | 19 | 18 | 19 |
| ,, II . | 15 | 14 | 19 | 16 | $M_2$ |
| ,, III . | 14 | 12 | 16 | 15 | 18 |
| ,, IV . | 16 | 17 | 20 | 16 | 21. |

$M_1$ and $M_2$ could be given respectively values like 16 and 18 (it would make no difference to the final result if they were equated to $+$ 1,000 and $-$ 1,000, but it is more convenient to use values within the range given by the other plots) and then a double covariance is carried out on the independent variates

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0. |

The adjusted means are worked out for the independent variates equal to zero, not to their general means.

(b) *The method of Appendix III:* This is not recommended, the solution of the parametric equations being sometimes very awkward. It has, however, the advantage of being clear cut and helping the investigator to see how everything is derived.

(c) *By minimization of the error:* In this method the error is worked out algebraically and minimized by means of the differential calculus. Thus in the example given under (*a*), the error sum of squares may be evaluated thus (see Appendix I): Classification totals are added to the data,

| | | | | | |
|---|---|---|---|---|---|
| 14 | $M_1$ | 19 | 18 | 19 | *70* + $M_1$ |
| 15 | 14 | 19 | 16 | $M_2$ | *64* + $M_2$ |
| 14 | 12 | 16 | 15 | 18 | *75* |
| 16 | 17 | 20 | 16 | 21 | *90* |
| *59* | *43* + $M_1$ | *74* | *65* | *58* + $M_2$ | *299* + $M_1$ + $M_2$. |

to give an error sum of squares equal to

$$\begin{aligned} &[14^2 + M_1^2 + 19^2 + \quad . \; . \; . \quad + 21^2] \\ &- \tfrac{1}{5}\,[(70 + M_1)^2 + (64 + M_2)^2 + 75^2 + 90^2] \\ &- \tfrac{1}{4}\,[59^2 + (43 + M_1)^2 + 74^2 + 65^2 + (58 + M_2)^2] \\ &+ \tfrac{1}{20}\,(299 + M_1 + M_2)^2. \end{aligned}$$

Differentiating by $M_1$ and setting the result equal to zero so as to derive the minimizing value of $M_1$ gives

$$2M_1 - \tfrac{2}{5}\,(70 + M_1) - \tfrac{2}{4}\,(43 + M_1) + \tfrac{2}{20}\,(299 + M_1 + M_2) = 0.$$

Likewise, differentiating by $M_2$,

$$2M_2 - \tfrac{2}{5}\,(64 + M_2) - \tfrac{2}{4}\,(58 + M_2) + \tfrac{2}{20}\,(299 + M_1 + M_2) = 0.$$

Whence,

$$M_1 = 14{\cdot}7, \qquad M_2 = 19{\cdot}3,$$

their "missing plot values". The analysis is then completed using these values for $M_1$ and $M_2$. The evaluation of significant differences will be left to the next Section.

(d) *Evaluation of missing plot values by formula:* For many designs, formulae for missing plot values have been worked out, the book of Cochran and Cox [35] being especially comprehensive in this respect. The earliest publication of such formulae was that of Allan and Wishart [1].* Cornish has given expressions for some of the more complicated designs [38, 39, 40, 41; see also 135], but these are not likely to be of much use with perennial species; while for randomized blocks Baten [13] has given formulae for the cases when two or three plots are missing. For such multiple losses the iterative method of Yates [182] can also be very useful.

Actually, for orthogonal designs there is no need to look up the formula; it can so readily be derived *ab initio*. Thus, to take the example already considered of two plots missing from a design in randomized blocks, it is first necessary to recall the expression given in Appendix I for working out the error sum of squares from the variation terms. For a randomized block, this is Total term − Block term − Treatment term + Correction term. These terms represent

* In this paper there is a slip in the expression given for the Latin square.

respectively the variation of the data, the block means, the treatment means and the general mean from zero. The missing plot values are accordingly given by the equations,

$$M_1 - \frac{43 + M_1}{4} - \frac{70 + M_1}{5} + \frac{299 + M_1 + M_2}{20} = 0$$

$$M_2 - \frac{58 + M_2}{4} - \frac{64 + M_2}{5} + \frac{299 + M_1 + M_2}{20} = 0,$$

*i.e.*, by taking the formula for the error sum of squares and for each missing plot substituting
(i) the missing value for the Total term,
(ii) the incomplete block mean for the Block term,
(iii) the incomplete treatment mean for the Treatment term,
and continuing in the same way, equating the whole to zero.

It should be emphasized that missing plot values are no more than a computing device: they do *not* represent the value the plot would have given had it not been lost.

In an analysis of covariance, if a value is lost from one variate, the corresponding value in the other variate must be regarded as missing also. Computing then usually proceeds using either Method *a* or *d*. If the former is used, the values for the two variates for each missing plot are assigned arbitrarily, and one pseudo-variate added for the gap as in the analysis of variance. If the latter, missing plot values are fitted for each variate separately and the figures worked out by the usual methods.

Where plots are missing, the number of degrees of freedom for both total and error are less than those for the case of complete data by the number of plots missing. Thus, in the example that has been given, the total and error would have respectively 17 and 10 degrees of freedom instead of 19 and 12, as they would have had with complete data. This is true whatever the method used. With (*a*) it arises from the degrees of freedom lost by covariance upon the pseudo-variates; with the others, because the total number of data in the analysis are less than usual.

## 84. Some Consequences of Fitted Values

As has been pointed out elsewhere [182], the loss of data makes a design non-orthogonal, that is to say some treatments are represented on only some of the blocks. If these blocks differ from the rest, the treatments missing from them are compared on a different basis from those present in them. This non-orthogonality matters little where only one or two data are missing, but it can be quite important if many are. The method of working out incomplete data by pseudo-variates, which allows for non-orthogonality, is not usually practicable when many values are missing, nor is Method *b*. With Methods *c* and *d* some allowance has still to be made for this non-orthogonality, though this is not usually done unless more than 10% of the data are missing.

The method resembles that of Appendices III and IV, *i.e.*, an error sum of squares, $S$, is worked out taking treatments into account. Then, a fresh error sum of squares, $S'$, is worked out ignoring treatments. Thus, in the example of the last Section, missing plot values have been derived for the randomized block, namely, 15 and 19, and these lead to an error sum of squares, $S$, of 10·90 with 10 degrees of freedom.

Starting again, ignoring treatments, a one-way classification is left and the missing-plot values are 17·5 and 16·0. Calling these 18 and 16 respectively gives an error sum of squares of 74·00 with 14 degrees of freedom. Accordingly, the full analysis of variance, making allowance for non-orthogonality, is:

| *Source* | *d.f.* | *s.s.* | *m–s.* | *F.* |
|---|---|---|---|---|
| Treatments (by difference) | 4 | 63·10 | 15·78 | 14·48*** |
| Error | 10 | 10·90 | 1·09 | |
| Total | 14 | 74·00 | | |

Another difficulty arises when using missing plot values. The missing plot values, as has been said, are not the values that would have been obtained if the missing observations had been taken, though they are estimates of them. Consequently, treatment means depending in part on missing plot values are not as accurately determined as those based entirely on data actually observed and this should be taken into account when working out significant differences.

For randomized blocks the position is fairly well understood. Suppose two treatments are being compared and give data thus:

(O is an actual observation. M is a missing plot value.)

| *Block* | *I* | *II* | *III* | *IV* | *V* | *VI* | *VII* | *VIII* |
|---|---|---|---|---|---|---|---|---|
| A . | O | O | M | O | O | O | M | O |
| B . | O | M | M | O | O | M | O | O. |

A gives four complete comparisons with B (Block I, IV, V and VIII) and two incomplete ones (II and VI), while in two blocks (III and VII) it gives no comparison at all. Its effective replication in comparison with B may be written $(4 + 2\alpha)$. Likewise, the effective replication of B in comparison with A is $(4 + \alpha)$. What is the value, $\alpha$, to be ascribed to an incomplete comparison? Yates [182] suggested $\frac{1}{2}$, but Taylor [156] has pointed out that $\frac{K-2}{K-1}$ is better, $K$ being the total number of treatments. The significant difference between the means of A and B is thus:

$$\sigma t \sqrt{\frac{1}{4+2\alpha} + \frac{1}{4+\alpha}},$$

where $t$ is derived from tables [66*a*] and $\sigma^2$ is taken to equal the error mean square of the analysis of variance. Recently, Baten [14] has published the exact expressions.

For Latin squares the position is less well understood. Yates [182] has indeed given a rough rule, but it is not known how accurately this applies. Accordingly, the present writer has worked out the following exact solutions for four of the commoner cases. All formulae refer to an incomplete $K \times K$ Latin square, giving an error mean square of $\sigma^2$.

*Case a: One plot missing.*

Significant difference between the mean of the defective treatment and any other:

$$\sigma t \sqrt{\frac{2K^2 - 5K + 4}{K(K-1)(K-2)}}.$$

*Case b: Two plots missing, not having a row, column or treatment in common.*

Significant differences between the mean of a defective treatment and of one not defective:

$$\sigma t \sqrt{\frac{2K^3 - 11K^2 + 23K - 22}{K(K-3)(K^2 - 3K + 4)}}.$$

Significant difference between the means of the two defective treatments:

$$\sigma t \sqrt{\frac{2(K-2)}{K(K-3)}}.$$

*Case c: Two plots missing, having a row or column in common.*

Significant difference between the mean of a defective treatment and of one not defective:

$$\sigma t \sqrt{\frac{2K^2 - 7K + 7}{K(K-2)^2}}.$$

Significant difference between the means of the two defective treatments:

$$\sigma t \sqrt{\frac{2(K-1)}{K(K-2)}}.$$

*Case d* : *Two plots missing, having a treatment in common.*

Significant difference between the mean of the defective treatment and any other:

$$\sigma t \sqrt{\frac{2(K^2-3K+4)}{K(K-2)^2}}.$$

For other designs and other cases of the Latin square, it is always possible to use the method of Appendix III to work out the formula for the significant difference, even though some other method has been used in computing the analysis of variance.

Using pseudo-variates this difficulty with significant differences does not arise, because the increase in the value of the significant difference is given by the usual expression in the analysis of covariance (see Appendix IV).

## 85. Some Other Mishaps

It sometimes happens that the yields of two plots become mixed and it is not possible to say how much came from each, though there is no doubt about the total. It has been found [97] that it is sufficient to proceed as if all the crop came from one plot (it does not matter which) and to correct by covariance upon a pseudo-variate having the values $-1$ and $1$ for the two plots that have been muddled and 0 for all the rest. The method as published is indeed of greater generality and may be used if the yields of several plots have been mixed. Again it should be emphasized that this is not a method for finding out what did happen, but for computing the results as easily as possible. It might be known, for example, that one plot had at least 70 lb. and another at least 50 lb. with 20 lb. in doubt, but this will not prevent the plot values coming out at 65 and 75 lb. respectively after adjustment by covariance, if these values will best minimize the error. Alternative but equivalent solutions of the problem have been worked out [19, 20].

Another occasional mishap is the interchanging of the treatments intended for two plots. In a randomized block design this does not matter if the two plots are in the same block; but if they are in different ones, Treatment A will be occurring twice in one block and not at all in another while Treatment B will reverse the fault. A particular solution has been found for randomized blocks [113], as well as a more general one [74]. The method of Appendix III is in any event always available.

# CHAPTER 9

# THE MEASUREMENT OF PERENNIAL PLANTS

" . . . corroborative detail, intended to give artistic verisimilitude to an otherwise bald and unconvincing narrative."

W. S. Gilbert, 'The Mikado,' Act II (Pooh Bah.)

## 90. General Survey

This is a subject of great complexity and not one upon which any attempt will be made to achieve comprehensiveness. It would, indeed, be exceedingly difficult to do so, partly because so much of the work has been published incidentally in the course of describing other investigations, and partly because a number of the ideas involved need reassessment at the present time, thus making difficult a balanced view of the subject. What follows is frankly little more than a résumé of present East Malling practice and outlook, though it is hoped that the methods described will prove helpful elsewhere.

Before going any further it should be emphasized that no system of measuring and recording trees, however well conceived and carried out, can be a substitute for the intimate and almost personal relationship that should subsist between an experimenter and his trees. Plants are not measured primarily to provide data for statistical analyses but to give *records* that will describe to others what was apparent to the experimenter at the time. The other purpose, that of confirming what has been suspected, by making possible a significance test, though important, is yet secondary.

Initially at East Malling very detailed records were taken as a matter of course. Thus, if someone wanted to know how many blossom trusses there were on a tree, a recorder would be detailed to count them ; while colour records on an apple crop were often taken by grading each individual fruit. These methods have been previously described at some length [78, 138] and remain the ideal, though not one to which many can attain.

The war of 1939–45, together with the increasing size of trees, made it impossible for East Malling to maintain these detailed records and it became necessary to find other and less time-consuming methods. The search is by no means finished and the present writer, who has been concerned in most of the modifications adopted, is not wholly satisfied with some of them ; but they do represent an attempt to build up a body of recording techniques that are both reasonably practicable and scientifically based. For that reason, some account of them may be of interest to others who have to take records with restricted resources.

In general, records may be classified in three ways, quite apart from the characteristics recorded. First of all, they may be *complete*, as when someone counts all the fruit picked off a tree or measures all the shoots, or they may be based on *samples*, as when certain fruit are chemically analysed in order to find the sugar content for the crop as a whole. Actually, many records that appear to be complete are not really so, as for example when an estimate is made of the colour of a crop. If this is done on the trees, the recorder notes the colour of the fruit he can see, which may be all or only some of them ; but if it is done in the boxes, he can certainly only see the top ones. He is, in fact, perhaps without realizing it, using a sampling method. It may be mentioned that any experiment at the best represents a sampling procedure, for no investigator studies all the existing trees (let alone all trees, past, present and future) having a certain treatment but only a group of them. Indeed, it would be possible to write quite a logical statistical text-book discussing "experimental design" only as a small section of the chapter on "small sample theory".*

* The point has some bearing on the assessment of statisticians in biological circles. These reprehensible beings are commonly accused of (1) demanding too many trees per treatment and (2) being ready to base results on samples that are not large enough.

The next classification of records is into *measurements, estimates* and *categories*. By the first is meant the comparison with a numerical scale by means of some instrument such as a pair of scales, a metre rule or a colorimeter: By the second is meant the estimation, usually by eye, of numerical values without actually measuring them, as when recording that 20% of leaves are infected by observing the tree without counting all its leaves: By the last is meant the placing of individuals into groups having a definite order but arbitrarily defined limits, as when the incidence of disease is classified as slight, moderate or heavy. This classification is not perfect and border-line cases sometimes arise. In making estimates it is important to define the limits of each class rather than its mid-point, *i.e.*, 10–20 rather than 15.

The last classification is into *direct* and *indirect* records. Thus, the weight of a tree can be measured directly only at planting, transplanting or grubbing, but treatment differences can sometimes be gauged indirectly from trunk girths. Again, it may be too laborious to

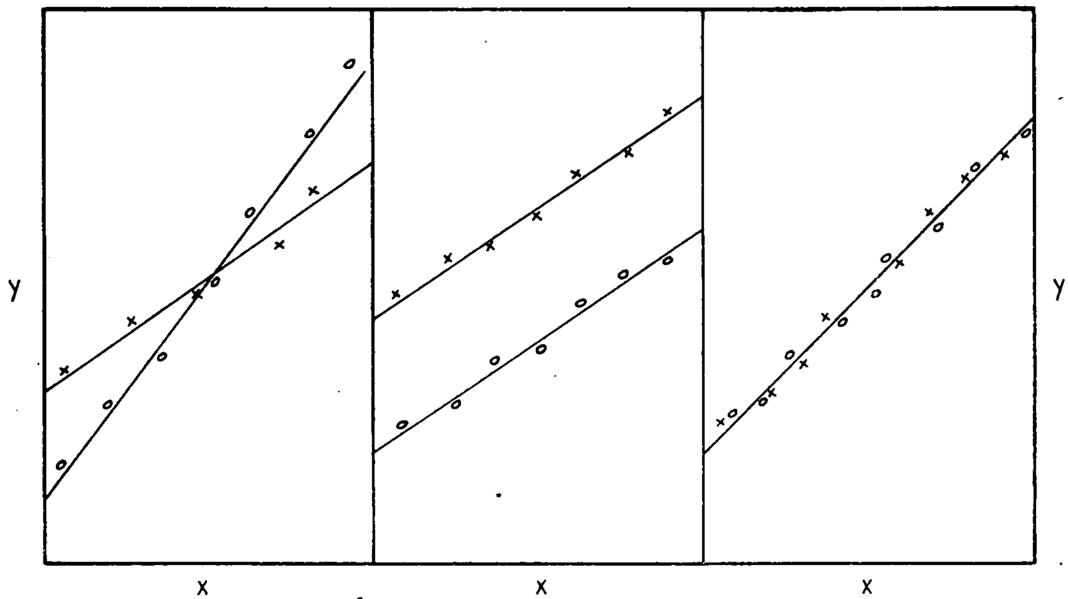



FIGURE XXXII.—Possible relationships between $y$, the quantity estimated, and $x$, its estimator.

measure the length of all new shoots on a tree; but, where the pruning method involves removing a constant proportion of new growth, the weight of prunings may well serve instead.

Plainly, the best records are those that are complete, direct measurements, but this is not usually attainable. The next four Sections will discuss some of the consequences of departing from the ideal.

### 91. A General Consideration with Simplified Methods

In trying to devise methods of recording that shall be both adequate and simple, workers at East Malling have repeatedly run up against a difficulty which, though it figures very little in the literature, is frequently a real one. It is this: a quantity, $x$, obtained from samples or by estimation or in some other way, may be very closely related to $y$, the quantity to be recorded; *but* the relationship may itself be dependent upon the treatments. Figure XXXII illustrates this point. On the left are shown by crosses and circles the values of $y$ for a series of values of $x$ for each of two treatments. An experimenter who tested his recording method on only one treatment might well be led to believe that $x$ is an almost perfect measure of $y$, though really it is most unreliable. If $x$ is *small* and the same for both treatments, so far from the $y$'s being equal, the treatment represented by crosses definitely gives the larger $y$. If, on the other hand, both treatments give the same *large* value of $x$, they again differ in respect of $y$, though in the opposite direction. Since there is no means of deciding *a priori* how the two lines in the diagram do lie in relation to one another, the position is chaotic despite the high correlation between $x$ and $y$ for any particular treatment.

If, as in the middle portion of Figure XXXII, it appears that the lines are at least parallel —not an uncommon state of affairs—it may well be that use can be made of $x$ if some information can be gained on the factors determining the distance between the lines. Ideally, however, the two lines should be identical as in the right-hand portion of the diagram, but this does not always happen.

This difficulty arises at times with almost all methods for simplifying records. To take sampling, for instance, it has been suggested recently that asparagus crops could be measured for only the middle of the season, but this would unduly favour those varieties or treatments of a mid-season tendency, *i.e.*, the relation between crop during the selected period and over the whole season does itself depend on the treatments under investigation. Again, to take estimates, though many people can accurately compare crops of trees when all are similarly treated, they easily go wrong when some of the treatments lead to the fruit being more readily seen, as happens for example, when more red coloration is present, the habit of the tree is more open or defoliation occurs earlier. With indirect measurements the difficulty is very common and usually requires careful investigation before a method can be recommended. The writer has published an account [116] of the position when trunk girth is used as an indirect measure of tree weight in apples, a field where the relations are especially complex. Another example occurs with coffee, where the weight of "cherry" is related to the weight of "clean coffee" in a manner that depends upon soil moisture [125, 161].

It is not, of course, suggested that this objection always applies, though it should always be considered. In about 90% of the published methods for the simple recording of fruit experiments it has been ignored, so some discrimination is needed before taking over a method described by another investigator.

## 92. Sampling Methods

The experimenter who proposes to use sampling methods need not be put off by the fear that he will necessarily introduce inaccuracies into his results. As has already been stressed, an experiment is itself a special form of sampling technique and has itself an error, which arises from the trees selected for each treatment not being entirely representative of the populations to which they belong, being in fact a *sampling error*.

Given a method of sampling that is random, the worst that can happen is that this error will be increased and consequently more replicates will be needed—not in itself a tragedy. Many investigators would prefer sampling from 100 trees to recording 80 in full.

In fact, an experiment as designed has already an error inherent in the trees (here to be written $\sigma_i$), which it would have even if recorded in full. The question is, what happens when the technical error of sampling or, as it is more usually called, the random error (here to be written $\sigma_t$) is added to it ? Contrary to what might be thought, the larger $\sigma_i$ is, the larger $\sigma_t$ may be, *i.e.*, if the trees are already very variable, there is no point in measuring them with extreme accuracy.

For the case of random sampling these two quantities can often be evaluated thus : Take an existing trial and sample each plot twice, add the two sample values together and work out an analysis of variance on their sums. The error mean square will then provide an estimate of $2\,\sigma_t^2 + 4\,\sigma_i^2$. Now take the differences between the duplicate pairs of values, square these differences, add the squares and divide by the number of plots to obtain an estimate of $2\sigma_t^2$. From these two estimates it is possible to obtain values for $\sigma_i^2$ and $\sigma_t^2$. It should be remembered that these results are themselves estimates and subject to error, from which it sometimes happens that $\sigma_i^2$ comes out negative. This can never really be so, and in such instances it is best to conclude that the technical error is large in relation to the inherent, *i.e.*, the sampling method is a bad one. The calculation should in any event be repeated for a range of trials if reliance is to be placed on the conclusion.

It is now possible to calculate the effect of sampling on the number of replicates needed. Supposing that, with complete records, a certain trial would need $R$ replicates to show the desired differences, then, with use of the sampling method, $R\left(1 + \frac{\sigma_t^2}{\sigma_i^2}\right)$ would be needed instead. The aim at East Malling has been to devise methods such that $\sigma_t < \frac{1}{3}\sigma_i$, *i.e.*, for which the technical error of sampling is almost lost in the inherent error, thus calling for no important modification in trial design.

The above work assumes that all selections for the sample have been made at random, *i.e.*, if the method is to take four fruits from each of three bags of similar size, first the bags must be selected at random and then the fruit from within each. In practice, this is rarely done and it is sometimes hard to decide how far a convenient method of selection will suffice. One of the commonest devices is to disperse the sample, *e.g.*, to choose three bags from eleven by taking the 2nd, 6th and 10th in the order of picking, the aim being to take bags, or whatever it may be, at a regular spacing with, as far as possible, a half-space before the first one chosen and another after the last. Again, two branches may be selected on a tree by taking the most northerly and the most southerly.

In general, a systematic sample is rather more accurate than a random one, though its technical error is not so easy of calculation. One way is to take a sample for each plot of an existing trial and then to record the plot in full. The differences between the two results, squared, added and divided by the number of plots give an estimate of the value of $\sigma_t^2$ (not $2\sigma_t^2$ as in the case given above). Now, an analysis of variance on the true values, as given by the complete recording, will obviously give an error mean-square that is an estimate of $\sigma_i^2$. Another way is to carry out the method in duplicate as before using random selections at every point and then to argue that, if dispersal had been used instead, the results would have been, if anything, more favourable to the sampling method. When once $\sigma_t$ and $\sigma_i$ have been evaluated, the rule for increased replication is as before.

A difficulty with non-random samples, whether dispersed or not, is the risk of bias. Thus, the writer has recommended elsewhere [112] that fruit size in apples should be judged from a sample divided between the tops of every third box, which leads to a bias from small apples slipping down between large ones. It is true that the extent of bias can be calculated and allowed for, but the point has to be guarded against. For this reason the better method for gauging the value of a dispersed sample is the first one given, *i.e.*, taking a sample from each plot followed by a complete record, because this provides an easy test for bias. Unfortunately, complete records are sometimes quite impracticable even for a trial of recording techniques. It should be mentioned that a constant bias has no effect on treatment differences.

Even samples that are random at every stage can lead to bias if the first choice is between classes of unequal size. Thus, if raspberry leaves were being selected for chemical analysis by first selecting canes and then an equal number of leaves on each of those canes, the leaves on a small cane would have a better chance of being represented in the final sample than would those from a large one. If they were in some way different, the sample would be biased.

### 93. Estimates and Categories

It is not easy to assess the value of records of this sort, because it depends very much upon the person making them. At the best, they provide a very easy means of showing differences; though at the worst quite large differences can be missed, either because the observer is not good enough, or because he has some private anxiety which inhibits judgment. Some years ago the writer carried out some investigations into the varying abilities of different people to categorize fruit colour in apples, and was surprised to observe how he himself varied from one trial to another, until he noticed that he did best when he had no qualms about the way the organization of his assistants was working out and worst when he feared a muddle. To sum up, if a person, knowing nothing of the treatments and their probable effect, detects differences, there is almost certainly something there; but a negative result may be due to personal factors.

Ideally the recorder does *not* know what the trial is expected to prove or, at all events, does not know which plots are expected to show the differences. However he reacts to such knowledge, he cannot remain unaffected by it. In practice, however, recorders cannot be changed so often and they should be taught to regard trials objectively.

Supposing that the figures produced do show a significant difference, it is not always clear what has in fact been observed. Thus, estimates ostensibly of the number of blossoms may have been chiefly determined by the size of individual blossoms or the earliness of their opening. It is not unusual for observers to have their quite definite statements contradicted by subsequent measurements, but they must have detected differences of some sort.

These sardonic reflections apart, estimation and categorization are very valuable tools in the assessment of experimental results; but, like everything else, they must not be used uncritically. One point that needs attention is the choice of scale. For estimation it is rarely advisable to use more than *ten* grades, and in categorization, where there is no numerical background, more than *five* is usually disastrous. They cannot be kept separate in the observer's mind, and the confusion only obscures what might otherwise be plain.

The writer has a high opinion of two kinds of categories he has used in the past. In the one the recorder makes a preliminary survey of his plots to determine Grade 1, bad; Grade 3, average and Grade 5, good. With these fixed in his mind he adds Grades 2 and 4 for intermediate cases and then works over his trial, grading each plot.

In the second, based on a method evolved by the British Broadcasting Corporation for listener research [148], he ordinarily has only three categories, B+, better than average; B, about average, and B−, worse than average. He also has for emergency use, A, excellent, and C, dreadful; but these are definitely superlatives for exceptional cases. The advantage of this scale lies in its testing the recorder no less than the experimental treatments. An excess of A's and C's marks down the enthusiast with little judgment, while a disproportion of B+'s and B−'s suggests that the observer does not know the average when he sees it. For statistical work the categories need to be translated into some numerical scale and there seems no reason to depart from the simple method of calling A, 5; B+, 4; and so on down to C, 1.

Recorders occasionally complain if they are asked sometimes to use letters and sometimes numbers, but this shows that they have not been instructed properly in the use of the two scales, which are very different in operation and must be kept as distinct as possible in the recorders' minds. In the first, an absence of either Category 1 or 5 after a preliminary survey indicates some misconception, but a recorder who returns no A or C is probably a sensible fellow who can keep his head.

Where several recorders are used it is better for them to work independently and to average their results afterwards rather than to let them come to an agreed result for each plot. Averaging results in this way and using either scale avoids the danger of moderate and balanced observers, who stick to the middle grades, being overborne by dogmatists, who think everything either wonderfully good or astonishingly bad—a great risk where several observers use a scale without clear instructions.

Where a plot contains several trees, it is usually better to grade each tree separately and to average the results, though categorization of whole plots is quite valid.

Altogether, ability to categorize well is not a bad test of character because defects such as assertiveness, timidity or lack of self-confidence all militate against success. In general, experience does not count for a lot unless the observer has frequently checked his results against actual measurements. Indeed, a knowledge of trees can accentuate the difficulty already mentioned of having to decide what an observer actually records. Thus, an experienced horticulturist asked to estimate the crops on a series of trees might observe that one of them had about a bushel more than its neighbour, but he might go on to notice its dark leaves and extensive shoot growth and thus to exaggerate its excellence; whereas a novice, equipped only with a good idea of what a bushel of apples looks like, might estimate the difference correctly. Best of all, however, would be an experienced person who had had opportunities to check the accuracy of his past attempts, *e.g.*, a fruit picker, whose ease of working depends in some measure on knowing exactly how many boxes to take to each tree.

To sum up, success in categorization and estimation depends largely upon the choice of recorder and the clearness and nature of his instructions. With respect to the first, it is best to choose someone who knows his own mind without being dogmatic, who does not know what results the experiment is expected to give, and who is either so new to the job that he will be careful or, if experienced, has been accustomed from time to time to checking his results against actual measurements. Also, he should not be colour-blind. As to the second, he should know clearly what he is supposed to be observing and should have given some thought to distinguishing it from other and similar qualities, and he should understand clearly the scale he is using. Perhaps the worst way of selecting recorders for the job is to send them round with an expert, who discusses each plot in turn,and then to choose the candidate whose results agree most often with the expert's. As a means of selecting "yes-men" it could hardly be bettered, but for the selection of recorders it has no merits at all.

### 94. Indirect Measurements

With indirect measurements more than with any other kind of simplified record the relationship between the quantity actually written down and the true value is likely to be affected by the treatments. Thus, it is often true that the weight of leaves is closely related to the product of their length and breadth, but this would be of no use for a comparison of leaf weight between different varieties where the shape and thickness of the leaves might be expected themselves to vary. Again, raspberries are sometimes trained on a series of equally-spaced horizontal wires and, when this is done, quite a good measure of length of new cane be taken by counting the number of occasions on which a cane and a wire cross. This is all right, provided the average angle at which the canes cross the wires is about the same for all treatments, but the method might not work for the comparison of varieties of different habit. To take a third example, with spur-pruned apple trees there is a close relationship between the length of new wood and the weight of prunings; but, if some rootstocks give thicker shoots than others, this would be of no avail for measuring shoot growth in a rootstock trial.

As has already been mentioned, a detailed study has been made [116] of the relationship between tree weight and trunk girth in apples, and this shows, if nothing else, how cautious an experimenter must be in interpreting indirect measures. Points of the sort described above must be looked for whenever an indirect measurement is proposed.

### 95. Measurements of Size and Growth

Since growth may be defined as change in size, these two groups of records are not very different from one another. What should be noticed, however, is that most woody, dicotyledonous species grow in two ways, (1) by making new shoots, and (2) by thickening those already made. Since the application of treatments almost always upsets the balance between these two forms of growth activity, some measurement should be obtained of each. An exception should perhaps be made in the case of certain soil treatments, where it may well be that additional nutrition affects both forms of activity proportionately, but this has not yet been demonstrated to be generally true.

Measuring the total length of new growth can be exceedingly laborious on a plant of any size, though a useful measure of shoot growth is obtained by selecting certain typical shoots, *e.g.*, those on the outside of the tree and at an angle to the horizontal between 15° and 75° [174] and measuring a sample of them dispersed round the apple tree. This does not give total shoot growth, but it provides a measure of mean growth per shoot. As has already been mentioned, some indication of total shoot growth is obtainable for spur-pruned trees from the weight of prunings, while for peaches the writer has measured every eighth shoot instead of them all. Blake and Hervey [16] thought that some allowance should be made for there being years giving many short shoots and others giving a few long ones and evolved a "New Jersey standard " based on the number of twigs and the number of branches longer than 24 inches for use with peaches. So far as the writer is aware, no wholly satisfactory method is known for measuring shoot growth on a large plant, which is a great pity because the lack of one is serious.

In measuring shoot growth on trees that are pruned, a decision has to be taken whether to do so before or after pruning. If the aim is to measure the activity of the tree, it is better to measure before; but if an indication of future bearing capacity is being sought, it is better to do so after. Given a uniform method of pruning, it usually matters very little either way.

Measurements of thickening, unlike those of extension, are easily made. This can be done by recording the trunk girth at yearly intervals. In general, a measure of the circumference using a tape is better than one of the diameter using calipers. The place of measurement should, if possible, be at least nine or twelve inches from both the graft union and the lowest branch and should be marked, either with paint or nails, so that it can be identified in a year's time.

Because of the ease of taking trunk measurements and the difficulty of recording shoot growth, there is a tendency to use only the first. For trials of varieties, whether of rootstocks or scions, and pruning methods this can be disastrous, because the balance of the two forms of growth activity is so readily upset by such treatments that a measure of one of them provides no information about the other and it is the combination of the two that matters.

Another class of record that is often made relates to size of head of tree and has the advantage that it can be applied to a wide range of species. At East Malling height of head in combination with spread, in two perpendicular directions at 45° to the rows, is commonly measured but is not highly regarded, because the figures obtained depend so much on the weight of crop being borne at the moment. Height and spread can be used separately, the one to indicate the difficulty of picking, the other to suggest a suitable planting distance, and are then very valuable. If they must be used as a measure of tree size, they are better combined into one quantity, as in the investigation of Hedrick and Anthony [77], who assumed the head of an apple tree to be a cone and worked with its volume.

Probably the best of all measures of the size of a plant is its weight, but this can only be recorded at planting, transplanting or grubbing. To consider individual species, with strawberries height and spread of the whole plant are very useful. For plants that are cut back severely annually, such as raspberries, the number and length of new shoots are almost the only records needed and, with that particular species, can sometimes be measured by counting the number of crossings of canes with supporting wires. Palms can sometimes be measured by taking the length and breadth of a sample of leaves. Rigney, Morrow and Lott [132], working with blueberry, used the number of fruiting shoots.

Quite apart from its value as a size record, trunk girth is worth recording every year for a special reason. If a tree dies, it provides a means of looking back and seeing when abnormality first appeared.

### 96. Measurements on Fruit and Blossom

In many investigations of a practical nature measurements of crop weight are of unique importance, because the ultimate end of a fruit grower is to grow fruit. Nevertheless, other fruit records are often equally essential because some indication of quality is needed as well, though the precise nature of these records must depend upon the species and the purpose for which it is being grown. Blossom records are also of value as indicating the plant's potential crop unvitiated by factors such as poor pollination and frost, while time of blossoming is also frequently of interest for species in which interpollination of varieties is a difficulty.

On the main problem of recording crop weight, the tendency at East Malling is to count the number of containers per plot, to measure the mean weight of fruit in a container by sampling, and then to multiply the two together to give crop per plot. The accuracy of this method depends upon the constancy of weight of fruit in a container, for obviously, if all containers were identical in what they held, the method would be perfect. Actually it is necessary to estimate two variabilities: $\sigma_p$, which represents the variation *from one plot to another* of the weight of fruit in a container, as measured by its standard error, and $\sigma_c$, which is the standard error of the weight of contents between containers *from the same plot.* To evaluate these quantities, it is best to take an existing experiment and to select two containers per plot at random and to weigh the contents of each. Now, adding these two figures per plot together and working out an analysis of variance on the *sums* will give an error mean square which may be equated to $4\sigma_p^2 + 2\sigma_c^2$; while if the *differences* are squared, added and divided by the number of plots, the result may be equated to $2\sigma_c^2$. Hence $\sigma_p$ and $\sigma_c$ are readily found, though it is desirable to estimate them from several different experiments in order to obtain results that would be broadly based. If for any trial $\sigma_p^2$ should come out negative, it is best taken to be zero. The whole process is similar to that described in Section 92 for the evaluation of $\sigma_t$ and $\sigma_i$.

Now, suppose that in some future experiment $p$ containers have been weighed from each plot, there being $k$ plots in all. Let $g$ be the mean weight of fruit in all the containers weighed and let $h$ be the mean weight of fruit in the containers weighed from one specific plot, which gave $n$ containers full in all. What is the best way of estimating the weight of crop from that plot? It is assumed that the experimenter already has figures for $\sigma_p$ and $\sigma_c$ derived as described above. He may then proceed in one of four ways discussed elsewhere by the writer [110, 112]. Only two of these methods, namely the first and the third, will be considered here. In the first case he may argue thus: "There can really be very little difference between the containers in different plots (*i.e.*, $\sigma_p$ is small). I will therefore make use of $g$, which is based on $pk$ containers, rather than $h$, which is based on only $k$, and will estimate the crop as $ng$." If he does so (first method) he will obtain an error of sampling ($\sigma_t$ of Section 92) such that

$$\sigma_t^2 = n^2 \left(\frac{k-1}{k}\right) \sigma_p^2 + \frac{n(n + kp - 2p)}{kp} \sigma_c^2.$$

Suppose, however, that he argues rather like this: "It is true that $g$ is more broadly based than $h$, but this is something of a disadvantage if I thereby estimate the crop from this plot from containers which did not come from it and may not be representative of it (*i.e.*, $\sigma_p$ may be large). I will therefore use $nh$ as my estimate." This (the third method) gives a technical error of sampling, such that

$$\sigma_t^2 = \frac{n(n-p)}{p} \sigma_c^2.$$

Other and more complicated methods of computing have been evolved in which $g$ and $h$ are used in combination, but these two are the chief. By the aid of the formulae given it is possible to work out $\sigma_t$ for all the values of $n$ likely to be encountered on a particular plantation and thus to decide which method of computing is to be preferred and to judge if the proposed number of containers to be weighed from each plot is large enough. The rule at East Malling is to ensure that even in the worst circumstances $\sigma_t$ is less than $\frac{1}{3}\sigma_i$, where $\sigma_i$ is the standard error of an experiment as judged from previous experience when complete measurements have been made.

When working out a technique for future use, containers should be chosen truly at random; but when the method is actually being applied there is no objection to dispersing them systematically over the plot either according to the order of filling or so as to include fruit from as many trees as possible. Such dispersal should prove more convenient and might improve accuracy.

Another crop record that is often of interest concerns mean fruit size. This may be made by taking a sample and then either grading each individual fruit or weighing the sample and thus evaluating the mean weight of 100 fruits. The individual records are of greater value because they show the proportion of crop unsaleable on account of size, but the greater labour of taking such detailed records makes the "weight of 100 fruits" of greater general utility. At East Malling sampling methods have been described for apple, pear and plum [112] and cherries [159], depending on the dispersal of the sample between the tops of containers. For apples there is a tendency for the smaller fruits to slip down and miss the sample, so correction by regression is needed to allow for this. On the tree it appears that larger apple fruits are found on the outside than on the inside [173].

Fruit colour is a quantity that needs careful definition before it can be measured. It may mean proportion of surface flushed, the hue of the flush or the hue of the ground-colour. For the proportion of apple surface flushed, it has been found [112] that there is a strong tendency for the fruit to lie in the boxes in layers of similar coloration, so an effective sampling method requires the taking of fruit from a vertical column in each box to disperse it between the layers. The method is to form a hole in one corner of the box, one hand removing fruit while the other holds back the "wall" thus formed. Sometimes it is necessary to take a firm grip on the last fruits to be removed and to pull, thus collapsing the wall; but usually both hands can be withdrawn at the end leaving a well-defined even hole, a quadrant of a circle in cross-section, like a well. The chief skills required are a certain delicacy of touch in the finger-tips and an ability to judge how wide a hole is needed to yield a specified number of fruit, both of which come with practice. The method works well for apples, less well for pears because of their shape and not at all for cherries because of their stalks. Each fruit of the sample is then judged to fall into one of four grades (1, up to a quarter flushed; 2, a quarter to a half; and so on) and a mean grade given. At least 60 apples are needed for each plot, unless the variety is one with a clearly defined flush. Usually, however, the amount of flush can be estimated by eye, as can the two hues; though it is far from easy to judge one without interference from the other two. If East Malling experience is anything to go by, such estimation should always be done on the tree; because the tendency of fruit to form layers of similar colour in the boxes makes unreliable an estimate based on the top layers only, *i.e.*, the ones that can be seen.

Many other records on fruits are sometimes needed, mostly of ripeness and chemical content. The first of these resembles colour in requiring definition. Does it mean softness of tissue, sugar content, or is it just another name for ground-colour? The point is important, for the three are only roughly related, and a difference in one does not demonstrate a difference in either of the others. Records of chemical content plainly must be based on samples and, in the absence of precise knowledge on variations in chemical elements throughout a tree, these samples should be dispersed as much as possible. Perhaps the safest way, when once the fruit is picked, is to take a vertical section as described for fruit colour and then, if this is too large a sample, to select every fifth (or every tenth or every twentieth) fruit in order of withdrawal from the boxes to form a sub-sample for analysis.

Records of blossoming are mostly of two types: (1) amount of blossom and (2) season of blossoming. The first of these, amount of blossom, is usefully made by estimation, 0 representing no blossom, 5 the maximum for a tree of that size, and 1–4 proportionate to 5. For some species, however, it is possible to work along each branch in succession counting inflorescences—a task that is facilitated by wearing spectacles of a colour that will pass light from the blossoms while cutting out that from the leaves.

With some species, *e.g.*, coconuts, the number of female flowers is of importance. With almost any species there are investigations that require records of blossoms set and fruit fall.

Season of blossoming may be recorded by visiting the trees every day and noting (*a*) date of first blossoms open, (*b*) date of full blossom, *i.e.*, 50% open, and (*c*) date of last blossom. For pollination studies some such record is essential. For other purposes, *e.g.*, the comparison of treatment effects in one year, it will usually suffice to visit the trees once and to estimate the proportion of blossoms open on that occasion.

### 97. The Importance of Good Recording

In a series of experiments on a range of subjects many other kinds of record may be needed, while every species presents its own problems. In particular, records of diseases and pests are often essential and sometimes difficult. In the above notes an attempt has been made to indicate principles rather than to describe specific methods. Indeed, in the present state of the subject it would be very unwise to try to do more. It is hoped that others will, in their varying conditions and with their differing species, take up the problem of measurement and help to build up a reliable body of knowledge on a subject that needs much more attention than it does in fact receive. There are many who can draw up a research project that shall probe matters the farmer really needs to know about, and there are others who can design a trial that shall be statistically sound, but the whole scheme must depend upon an ability to measure the experimental trees accurately with the labour available. It is in this respect that research with perennial plants often fails. In particular, it is desirable to record not only the crop but anything that relates to it. Thus, with tea, permanent foliage is of importance as well as crop. Also, it is desirable as far as possible to record individual plants and not groups, but this is not always practicable. It is to be hoped that the future will greatly increase the knowledge of available devices. The past has brought forward a number of ingenious practices ranging from the use of hypsometers for measuring the height of a tree [166] to the training of monkeys to take shoot samples [37]. Perhaps "the best is yet to be."

# NOTES ON APPENDICES

The methods of analysing data to be described are used with only minor modifications in many statistical laboratories the world over. Nevertheless, it has been thought worth while to present them here for several reasons. For one thing, the setting out of these methods of very general application has considerably lightened the structure of Chapters 2 and 3. It is true that most of the designs there discussed have had " short-cut " methods specially evolved for them, and these, which may be found by looking up the references given, are often extremely useful. It does, however, seem better (at least to the present writer, who is something of a heretic in such matters) to learn the general methods *first*, thus getting the approach clear, and to learn the more specialized methods *afterwards*. Certainly, this makes the computor more self-reliant when faced with a new problem.

Another reason for presenting these appendices is the need to indicate how much should be written down and where. In the writer's experience, analyses carried out by beginners suffer from a characteristic source of inaccuracies. Sometimes a figure once obtained is written down, without sufficient notes as to its meaning, in the midst of a host of others, most of which could quite well have been omitted, so that, when the figure comes to be needed, the wrong one is extracted and used. This often occurs side by side with the contrary error of doing too much on scrap paper and too little on the working sheets. (Where a calculating machine is being used, these defects often arise from an insufficient appreciation of what it can usefully do.) The basic rules are four in number :

(1) Indicate clearly what trial the figures refer to, what measurement they represent and in what units.

(2) Write down on the working sheet every figure read off a calculating machine, however unimportant in itself. Most errors arise from bad copying and have no remedy except checking, which is impossible unless all figures copied are kept for future reference.

(3) Always subject to what has just been said, write down as little as possible. What is written down should fit into a standard format so that it can be identified again, when needed, with no risk of mistake. In what follows, the formats used at East Malling are described in some detail, but there is no need to keep to them if they are not liked. What is important is that each laboratory should evolve its own and keep to them strictly, because a neat and uniform way of writing down the steps of a computation can save many mistakes.

(4) Everything should be checked. The best way of doing this is to repeat the whole calculation using a different method. Otherwise the calculation must be repeated using the same method and the question then arises whether the check is not best carried out by some other person. Each alternative has its dangers. If one person does the work and another checks it, the risk of error is at a minimum, provided each retains a sense of responsibility, but there is a strong tendency for each to leave it to the other. On the other hand, most computors are at their best when they know that they are being trusted to get their figures right and will be held personally responsible if any mistakes are found.

It must be emphasized that the Appendices set out all the computations that could arise with data of the form under consideration. In any particular trial some will not be needed.

# APPENDIX I

## A General Method of Analysing Data for Orthogonal Designs

The following example is derived from a trial for the comparison of raspberry varieties carried out by the National Agricultural Advisory Service in Derbyshire. Variety A, a standard commercial sort, is represented twice in each row and in each column of the chosen design while the new varieties, B, C, D and E, are represented once each. The trial being intended to last for only a short time, no restriction of randomization, such as was described in Section 36, was thought necessary. Figures represent yield in ounces per plot for the first year of cropping :

| | | | | | |
|---|---|---|---|---|---|
| B 88, | A 246, | E 174, | A 236, | C 61, | D 83, |
| C 122, | A 165, | A 194, | B 97, | D 118 | E 145, |
| D 132, | E 124, | A 221, | A 127, | B 105, | C 100, |
| A 149, | C 76, | D 96, | E 122, | A 145, | B 68, |
| A 253, | D 106, | B 94, | C 98, | E 151, | A 145, |
| E 195, | B 106, | C 130, | D 93, | A 180, | A 128. |

Usually data should be expressed to three significant* figures, as here. In this example, however, they represent the first crop borne by the plants and may therefore be expected to show a high level of variability. If they do, the third significant figure may prove unnecessary but it is as well to leave it in so as to be on the safe side.

The design is based on an orthogonal three-way classification of rows, columns and varieties, *i.e.*, the plots of a row are always so apportioned that one-sixth of them occur in each column, one-third receive Variety A and one-sixth receive each of the other varieties, whichever row is chosen. The columns are disposed similarly with respect to the rows and varieties. Finally, for each variety, one-sixth of the plots occur in each row and in each column. For orthogonal designs, the recommended method has five stages :

1. *Computation of Totals.*—The design has three classifications, totals being worked out for each one separately. The grand total also is needed and should be obtained independently from each set of classification totals, this triple method of computing it providing a check on the arithmetic.

Rows: I, 888 ; II, 841 ; III, 809 ; IV, 656 ; V, 847 ; VI, 832.
Columns: I, 939 ; II, 823 ; III, 909 ; IV, 773 ; V, 760 ; VI, 669.
Varieties : A, 2189 ; B, 558 ; C, 587 ; D, 628 ; E, 911.
Grand total : 4873.

It appears from the Variety totals that A is obviously cropping better than the others, but the high variability has already been mentioned. It should be noted incidentally that treatments are designated by letters of the alphabet and blocks, rows and columns by roman numerals. This does not matter in itself, but it is all part of a standard format to minimize the chances of making mistakes.

* " Significant " is here not used in its statistical sense, but to indicate the number of digits altering from one datum to another. Thus data like 1744, 2227, 1603, etc., would become 1740, 2230, 1600, etc. Also, data like 12143, 12094, 12233, etc., would become 143, 94, 233, etc., the 12000 being added on again at the end of the computations.

2. *Computation of Variation Terms.*—These terms are worked out, one for the data and one for each set of totals:

(*a*) The data are squared and added to give $T$, known as the "Total Term", which represents the variation of the data from zero, *i.e.*,

$$T = 88^2 + 246^2 + 174^2 + \quad . \; . \; . \quad + 180^2 + 128^2 = 746\ 571{\cdot}00.$$

(*b*) The "Row Term", $T_1$, is obtained by squaring the row totals, dividing each square by the number of data upon which it is based, and adding, *i.e.*,

$$T_1 = \tfrac{1}{6}(888^2 + 841^2 + \quad . \; . \; . \quad + 832^2) = 665\ 045{\cdot}83.$$

The "Column Term", $T_2$, and the "Variety Term", $T_3$, are obtained similarly:

$$T_2 = \tfrac{1}{6}(939^2 + 823^2 + \quad . \; . \; . \quad + 669^2) = 668\ 003{\cdot}50.$$

$$T_3 = \tfrac{1}{12}(2189^2) + \tfrac{1}{6}(558^2 + 587^2 + \quad . \; . \; . \quad +911^2) = 712\ 683{\cdot}08.$$

There are three such terms because the classification is three-way, *i.e.*, (1) rows, (2) columns, and (3) varieties. They represent respectively the variation of the (1) row, (2) column and (3) variety means from zero.

It will be noticed that all these terms have been taken to two more places of decimals than arose from the squaring. Thus, in this instance, data and totals were taken to the nearest whole number. Their squares accordingly exhibit no decimal places, so the terms were taken to two. It will also be noticed that, for accuracy in copying, the digits are divided by hair spaces into groups of three, starting at the decimal point.

(*c*) A "Correction Term", $T_0$, is likewise obtained from the grand total, *i.e.*,

$$T_0 = \tfrac{1}{36}(4873^2) = 659\ 614{\cdot}69.$$

This represents the variation of the general mean from zero.

3. *Computation of the Analysis of Variance.*—First of all it is necessary to work out the total variation in the trial. This sum of squares is $(T - T_0)$, for $T$ represents the variation of the data from zero and $T_0$ that of the general mean from zero, and their difference accordingly represents that of the data from the general mean.* The degrees of freedom for $(T - T_0)$ are 35, $T$ being obtained by the squaring of 36 quantities and $T_0$ by the squaring of one.

By similar arguments, the sums of squares for the variation of row and column means from the general mean are respectively $(T_1 - T_0)$ and $(T_2 - T_0)$ with five degrees of freedom each, while the sum of squares for the differences associated with treatments is $(T_3 - T_0)$ with four degrees of freedom. Finally the residue, which cannot be associated with any of the classifications and is attributed to error, is found by difference. The analysis, in fact, is as follows:

| *Source.* | *d.f.* | *s.s.* | *m–s.* | *F.* |
|---|---|---|---|---|
| Rows | 5 | 5 431·14 | | |
| Columns | 5 | 8 388·81 | | |
| Varieties | 4 | 53 068·39 | 13 267·10 | 13·88*** |
| Error | 21 | 20 067·97 | 955·62 | |
| Total | 35 | 86 956·31 | | |

Before going further, the Error line should be checked as described later. The mean-square column (*m–s.*) is obtained by dividing each sum of squares (*s.s.*) by its degrees of freedom (*d.f.*). Since the residual sources of variation measured by the error are all-pervading in their

*By virtue of the identity

$$\Sigma(x_i - \bar{x})^2 = \Sigma x_i^2 - (\Sigma x_i)^2/n,$$

where $x_i$ is one of a series of $n$ data having a mean of $\bar{x}$. $\Sigma$ denotes summation of all quantities such as the one following the sign.

effect, the first three lines of the analysis should really read under sources, "rows + error", "columns + error" and "varieties + error" respectively; and, if the varieties really have no effect, the third and fourth lines become independent measures of the residual variation and should give concordant values in the mean-square column. Actually, one value is 13·88 times the other, which suggests that the discrepancy between the two is not just fortuitous but is brought about by real differences between the varieties. From the tables of Fisher and Yates* [66b] it appears that with 4 degrees of freedom for the numerator of $F$ ($n_1 = 4$) and 21 for the denominator ($n_2 = 21$), a ratio of mean-squares as high as 2·84 can arise by chance on 5% of occasions, as high as 4·37 on 1% and as high as 6·95 on 0·1%. Hence, the value of $F$ actually observed is so high that, in the absence of genuine varietal differences, it would arise by chance less frequently than once in a thousand trials ($P < 0{\cdot}001$), so three asterisks are placed against it to show the level of significance (see Section 12).

4. *Checking of Error Line in the Analysis.*—One point to be noted in computing analyses of variance is the need to check the error line by working it out in two ways. It was obtained above by difference, and this is all right provided it is remembered that any mistake in any other line will affect the error so calculated. If now the error can be checked, confidence in the whole analysis will be fortified. (It will have been noticed that the descriptions of several designs in this communication have included an independent means of calculating the error, not because it is of most importance but because it can provide a check on all else.)

Assuming the terms to be correct, and they should always be checked and rechecked with the greatest care, the following expressions for the error sum of squares, which are of very general application, will be found useful:

One-way classification (*e.g.*, a completely randomized design):

$$T - T_1.$$

Two-way classification (*e.g.*, randomized blocks),

$$T - T_1 - T_2 + T_0.$$

Three-way classification (*e.g.*, a Latin square),

$$T - T_1 - T_2 - T_3 + 2T_0.$$

Four-way classification (*e.g.*, a Graeco-Latin square),

$$T - T_1 - T_2 - T_3 - T_4 + 3T_0.$$

In general, for an $n$-way classification,

$$T - T_1 - T_2 - \quad \ldots\ldots \quad - T_n + (n - 1)T_0.$$

The degrees of freedom for error also need to be checked. One way of doing this is to use the expressions just given for the sums of squares, making each term equal to the number of squares it comprises. Thus, in the example, the error degrees of freedom should number $[36 - 6 - 6 - 5 + (2 \times 1)] = 21$, which is correct.

Another way is to take the formula from some well-known design and to argue from it, *e.g.*,

| | | |
|---|---|---|
| (i) | Complete randomization of $N$ plots in $K$ classes .. .. | $N - K$. |
| (ii) | $J$ randomized blocks of $K$ treatments .. .. | $(J - 1)(K - 1)$. |
| (iii) | $K \times K$ Latin square with $K$ treatments .. .. | $(K - 1)(K - 2)$. |
| (iv) | $K \times K$ Graeco-Latin square with $K$ treatments .. | $(K - 1)(K - 3)$. |

The nearest design to that of the example under study is the Latin square, which would have $5 \times 4 = 20$ degrees of freedom. There is, however, an extra degree of freedom ascribable to error, arising from the difference between the two sets of plots receiving A, making 21 in all.

* In these tables, $F$ is written $e^{2z}$ and the ratio of mean-squares is termed the "variance ratio". Both systems of nomenclature are in use and it is as well to know both.

5. *Computation of Significant Differences.*—Since it appears that there are differences between the varieties, it may be pertinent to inquire which differ from which and, in particular, which differ from A, the standard variety. As has been explained in Section 72, this inquiry is not always permissible.

If, in an orthogonal design, there are two treatment means, one based on $p$ data and the other on $q$, and they differ by more than

$$t\sqrt{\text{Error mean-square}\left(\frac{1}{p}+\frac{1}{q}\right)},$$

the means are said to differ significantly. The quantity $t$ is obtained from the tables of Fisher and Yates [66*a*] for the level of significance in which the investigator is interested and for the error degrees of freedom.

Thus, in the present instance, the significant difference between the mean yield per plot of Variety A and the mean yield of any other variety is

$$2{\cdot}080\sqrt{955{\cdot}62\left(\frac{1}{6}+\frac{1}{12}\right)} = 32{\cdot}2,$$

where 2·080 is the value of $t$ for $P = 0{\cdot}05$ and 21 degrees of freedom. Similarly, the difference between the means of any two other varieties is

$$2{\cdot}080\sqrt{955{\cdot}62\left(\frac{1}{6}+\frac{1}{6}\right)} = 37{\cdot}1.$$

It is next necessary to write the variety means in ascending order and to put brackets round pairs that are not significantly different, thus:

$$\underbrace{\overset{\text{B}}{93{\cdot}0}\qquad \overset{\text{C}}{97{\cdot}8}\qquad \overset{\text{D}}{104{\cdot}7}}\qquad \underbrace{\overset{\text{E}}{151{\cdot}8}\qquad \overset{\text{A}}{182{\cdot}4}}$$

At this stage of the computation it is as well to work with one decimal place more than in the data, but in presenting results for publication it is sufficient to round off to three significant figures.

It thus appears that the difference between the means of Varieties A and E is a little less than that needed for significance. This means that the difference could have arisen by chance from the operation of the residual sources of variation. In this particular trial, however, the difference is almost certainly genuinely due to varieties because the $F$-test has demonstrated varietal differences and there can be no doubt that E, a new seedling of known history, is different from A, an established commercial sort. The differences between A and the other varieties are all above the amount needed for significance so their genuineness may be regarded as proven.

In this Appendix all the computations have been set out that might arise with this sort of data, but in many instances some can be omitted. Thus, here the differences are due to varieties and it does not need an $F$-test or significant differences to show that different varieties crop differently. The analysis of variance would, however, be of value in showing how accurately the treatment means were estimated. The standard error of a single observation was in fact $\sqrt{955{\cdot}62}$, which equals 30·9, or 22·8% of the general mean, quite a high figure but not unduly so for the first year of cropping.

For many purposes it would be more convenient to express the results in tons per acre rather than ounces per plot. There is no objection to working out a conversion factor and applying it both to the treatment means and the significant difference.

*Format.*—The following suggestions are made for laying out the computations.

Data are most conveniently set out in a two-way table based on two of the classifications of the design, as in the example. Two sets of classification totals can then be entered in the margins and it will be quite obvious what they refer to; but, if this is done, a different coloured ink should be used to avoid confusion. Any other sets of classification totals should then be written in horizontal lines below the data.

When the variation terms are being calculated, they are most conveniently written down one below the other in a standard order to avoid confusion. A useful convention is to start with the Total Term at the top, then to list terms corresponding to positional differences (*i.e.*, rows, columns, blocks, etc.), then terms corresponding to former treatments no longer regarded as important, if any, then the term corresponding to the present treatments with the Correction Term at the bottom of the column. In the example, the order would have been $T$, $T_1$, $T_2$, $T_3$, $T_0$.

When once the terms have been worked out, construction of the analyses presents little difficulty. Significant differences, however, are a prolific source of mistakes, especially with decimal places. Unless a fully-automatic calculating machine is being used, it is as well to write down each stage of the computation, thus:

$$t\sqrt{955{\cdot}62\left(\frac{1}{6}+\frac{1}{12}\right)} = 2{\cdot}080\sqrt{238{\cdot}905} = 2{\cdot}080 \times 15{\cdot}457 = 32{\cdot}2.$$

As has already been said, the precise mannner of laying out the calculation is not of itself nearly as important as conformity to a system. If each analysis is laid out differently, the result is confusion when the trial comes to be written up, in addition to delay and possible mistakes in computation.

# APPENDIX II

## Partitioning of Treatments Sum of Squares in an Orthogonal Design

A treatments sum of squares as worked out by the method of Appendix I may well be made up of several components, some of which may be significant but others not. The question of when it is permissible to partition such a sum of squares into its components has been discussed in Section 71. Supposing partitioning to be permissible, how is it to be done ?

*A Factorial Experiment*

The first case is that of a factorial design. The following analysis was derived from the data representing crops in pounds over a ten-year period for a trial in eight randomized blocks of twenty treatments, made up of all combinations of five varieties of pear (Beurré d'Amanlis, Beurré Hardy, Conference, Fertility and Pitmaston Duchess), pruned either heavily or lightly with either many or few leaders :

| *Source* | *d.f.* | *s.s.* | *m–s.* | *F* |
|---|---|---|---|---|
| Blocks | 7 | 227 350 | | |
| Treatments | 19 | 2 740 648 | 144 245 | 15·86*** |
| Error | 133 | 1 209 797 | 9 096 | |
| Total | 159 | 4 177 795 | | |

Following the recommendation of Appendix I, two places of decimals were worked out for the sums of squares, but are here omitted. The rule does sometimes err on the side of too many places rather than too few.

In order to partition the sum of squares it is necessary to write down totals for all ways in which the data can be classified by some or all of the factors. Writing varieties as V, number of leaders as N and degree of pruning as D, data can be classified (i) by V, N and D, (ii) by V and N, (iii) by V and D, (iv) by N and D, (v) by V alone, (vi) by N or (vii) by D.

It is now time to look at the 20 totals from which the treatments sum of squares, which is to be partitioned, was derived :

| | *Amanlis* | *Hardy* | *Conference* | *Fertility* | *Pitmaston* | |
|---|---|---|---|---|---|---|
| Few Hard | 5 129 | 3 416 | 3 326 | 4 382 | 2 625 | *18 878* |
| Few Light | 5 181 | 4 045 | 5 022 | 4 875 | 3 480 | *22 603* |
| Many Hard | 6 063 | 3 764 | 3 967 | 4 430 | 2 809 | *21 033* |
| Many Light | 6 626 | 4 516 | 4 909 | 5 906 | 3 664 | *25 621* |

They are the " V, N and D " totals, for the data are being classified according to all three classifications, V, N and D. Down the right-hand margin in italics (or different coloured ink) are the " N and D " totals, obtained by adding sideways and thus eliminating classification by varieties, V.

It is next necessary to obtain the other totals given by two classifications, thus :

| | *Amanlis* | *Hardy* | *Conference* | *Fertility* | *Pitmaston* | |
|---|---|---|---|---|---|---|
| Few | 10 310 | 7 461 | 8 348 | 9 257 | 6 105 | *41 481* |
| Many | 12 689 | 8 280 | 8 876 | 10 336 | 6 473 | *46 654* |
| Hard | 11 192 | 7 180 | 7 293 | 8 812 | 5 434 | *39 911* |
| Light | 11 807 | 8 561 | 9 931 | 10 781 | 7 144 | *48 224* |

These figures are obtained by adding the figures of the first table in appropriate vertical pairs. The upper table gives the "V and N" total with the "N" totals in the margin: the lower gives the "V and D" totals with the "D" totals to the right.

It now remains only to add again vertically to obtain the "V" totals with the grand total to the right:

| *Amanlis* | *Hardy* | *Conference* | *Fertility* | *Pitmaston* | |
|---|---|---|---|---|---|
| 22 999 | 15 741 | 17 224 | 19 593 | 12 578 | *88 135* |

The "N" totals were originally obtained by adding the "N and D" totals in pairs vertically, but they can equally well be obtained by adding the "V and N" totals horizontally. Similarly the "D" totals can be derived from the "V and D" totals. The "V" totals can be obtained from both the "V and N" and the "V and D" totals. The grand total can be obtained from each of the classifications by one factor only, *i.e.*, V, N or D alone. In order to check properly, each set of totals should be obtained in as many ways as possible.

The next step is to work out variation terms for each set of totals. Thus $T_{VND}$ equals $\frac{1}{8}(5129^2 + 3416^2 + \ldots + 5906^2 + 3664^2) = 51\,289\,262$. The divisor is eight because each total contains eight data. $T_{VND}$ is plainly the treatment term of the original analysis under another name. All the other terms are derived similarly, *e.g.*,

$$T_{VN} = \tfrac{1}{16}(10\,310^2 + 7\,461^2 + \ldots + 10\,336^2 + 6\,473^2) = 50\,731\,245$$

$$T_{VD} = \tfrac{1}{16}(11\,192^2 + 7\,180^2 + \ldots + 10\,781^2 + 7\,144^2) = 50\,985\,515$$

$$T_{ND} = \tfrac{1}{40}(18\,878^2 + 22\,603^2 + \ldots + 25\,621^2) = 49\,152\,431$$

$$T_{V} = \tfrac{1}{32}(22\,999^2 + 15\,741^2 + \ldots + 12\,578^2) = 50\,484\,093$$

$$T_{N} = \tfrac{1}{80}(41\,481^2 + 46\,654^2) = 48\,715\,863$$

$$T_{D} = \tfrac{1}{80}(39\,911^2 + 48\,224^2) = 48\,980\,526$$

$$T_{0} = \tfrac{1}{160}(88\,135^2) = 48\,548\,614, \text{ the correction term in the original analysis.}$$

The various sums of squares now follow. In general, suppose that the interaction of a set of factors A, B, C, is needed, the method is to write down $T_{ABC}$, to *subtract* all terms representing *one* classification fewer, *i.e.*, $T_{AB}$, $T_{AC}$, $T_{BC}$, to *add* all terms representing two classifications fewer, *i.e.*, $T_A$, $T_B$, $T_C$, to *subtract* all terms representing *three* classifications fewer, *i.e.*, $T_0$, and so on till the correction term is reached.* Here, the sums of squares are:

| | |
|---|---|
| V | $T_V - T_0$ |
| N | $T_N - T_0$ |
| D | $T_D - T_0$ |
| V × N | $T_{VN} - T_V - T_N + T_0$ |
| V × D | $T_{VD} - T_V - T_D + T_0$ |
| N × D | $T_{ND} - T_N - T_D + T_0$ |
| V × N × D | $T_{VND} - T_{VN} - T_{VD} - T_{ND} + T_V + T_N + T_D - T_0$. |

Degrees of freedom are obtained thus: V, N and D have respectively 4, 1 and 1 for obvious reasons. Interactions have degrees of freedom equal to the product of those for the main effects upon which they are based, so the whole partition reads like this:

* There is nothing mysterious about this rule. The total variation due to A and B is $(T_{AB} - T_0)$, of which $(T_A - T_0)$ is due to the main effect of A and $(T_B - T_0)$ to the main effect of B, leaving $(T_{AB} - T_A - T_B + T_0)$ for the interaction of A × B, and similarly for higher order interactions.

| *Source* | *d.f.* | *s.s.* | *m–s.* | *F* |
|---|---|---|---|---|
| V | 4 | 1 935 479 | 483 870 | 53·20*** |
| N | 1 | 167 249 | 167 249 | 18·39*** |
| D | 1 | 431 912 | 431 912 | 47·48*** |
| V × N | 4 | 79 903 | 19 976 | 2·20 |
| V × D | 4 | 69 510 | 17 378 | 1·91 |
| N × D | 1 | 4 656 | 4 656 | <1 |
| V × N × D | 4 | 51 939 | 12 985 | 1·43 |
| Error | 133 | 1 209 797 | 9 096 | |

It should be checked that the degrees of freedom and sums of squares add up to the quantities being partitioned.

From this it is clear that almost all the variation between means for treatment combinations are due to the main effects, *i.e.*, each of the three factors operates to much the same extent whatever the levels of the other two. Although the interactions, V × N and V × D, are not significant, they are nearly so, and there is some evidence that the effect of each pruning factor depends to some extent upon variety, but in relation to the large main effects this is of little importance.

Considering now the main effects, there is no need to work out the significant differences for the effects of N and D; because there are only two levels of each and there can be no doubt which is the level to differ from another. For V, however, which has five levels, the experimenter may wish to inquire which varieties are significantly different from their fellows. Since each varietal mean is based on 32 data, the significant difference is

$$t\sqrt{\text{Error mean-square} \cdot \tfrac{2}{32}} = 47{\cdot}2 \qquad (P = 0{\cdot}05).$$

Hence the results for varieties may be expressed thus:

| *Pitmaston* | *Hardy* | *Conference* | *Fertility* | *Amanlis* |
|---|---|---|---|---|
| 393·1 | 491·9 | 538·3 | 612·3 | 718·7 |

In publication, the decimal place would not be needed. Also the difference between Hardy and Conference is obviously real despite its non-significance.

It will be noticed that in the absence of interactions involving varieties, the effective replication has been increased from eight to 32, an example of "hidden replication" (see Section 54).

*Another Factorial Experiment*

Suppose, however, that interactions do prove significant, then it is necessary to consider each of the factors separately for each level of the others with which it interacts. Thus, in the following experiment investigating the effects of three systems of soil management (Clean cultivation, Grass, Straw mulch), in combination with high and low nitrogen applications on established apple trees, in the first year the following mean figures were obtained for weight of 100 fruit in pounds:

| | *C.C.* | *S.M.* | *G.* |
|---|---|---|---|
| High N | 24·4 | 28·6 | 34·0 |
| Low N | 34·2 | 32·3 | 27·6 |

Neither main effect (*i.e.*, cultural methods or fertilizer) was significant but the interaction was highly so ($P < 0{\cdot}01$). Each of the means given was based on five data, while the error mean-square was 25·0967 with 17 degrees of freedom so the significant difference between them was 6·7. It is now clear that the high nitrogen significantly reduced fruit size on clean cultivated land, though on grass its effect was nearly significant the other way. In the present instance that is very nearly the whole story, but the approach is one that can lead to misleading results and is not therefore recommended.

A particularly effective way of dealing with interactions when, as here, one of the factors has only two levels, is to work with differences, thus:

| | *C.C.* | *S.M.* | *G.* |
|---|---|---|---|
| Effect of N . . | $-9{\cdot}8$ | $-3{\cdot}7$ | $+6{\cdot}4$ |

The significant difference between these figures is $\sqrt{2}$ times the one already derived, *i.e.*, it is $9{\cdot}5 = 6{\cdot}7\sqrt{2}$, so it appears that the effect of nitrogen on grass plots was indeed different from its effect elsewhere, its effects on clean cultivation and straw mulch not differing significantly.

The two examples discussed above represent the two extremes. In the one, main effects alone are significant: in the other, the interaction alone. Usually intermediate cases arise and some combination of the two approaches is called for. The precise point at which an interaction becomes so large as to vitiate the first approach is a matter for judgment, but the investigator who writes down his means and studies them in the light of his analysis of variance and significant differences is not likely to go far wrong. An interesting example of the need for common sense is, as a matter of fact, provided by the pear variety and pruning trial. The interaction of variety and degree of pruning is not indeed significant, but what there is of it appears to have arisen principally from the smaller benefit of light pruning on the cropping of Beurré d'Amanlis as compared with the other varieties. It may well be that light pruning does not really give better crops with this variety, though, on the other hand, as there is no suggestion of an adverse effect, a general recommendation to prune lightly needs no special reservation.

*The Case of Nominated Groups*

Another occasion for partitioning arises from the levels of a factor falling into groups. Thus, if there are six treatments, A–F, and A–D form one natural group and E, F another, it might seem reasonable to partition the five degrees of freedom for treatments, thus:

| | |
|---|---|
| Between groups . . . . | 1 |
| Between (A–D) . . . | 3 |
| Between E and F . . . . | 1. |

Any of these effects with more than one degree of freedom may be partitioned further and, for purposes of computation, this is often the easiest thing to do. Thus, the degrees of freedom within the group A–D could be further partitioned thus:

| | |
|---|---|
| Between (A, B) and (C, D) . . . . | 1 |
| Between A and B . . . . . | 1 |
| Between C and D . . . . . | 1. |

Alternatively, they could have been divided:

| | |
|---|---|
| Between A and (B–D) . . . . | 1 |
| Between (B–D) . . . . . . | 2, |

which can then be developed into

| | |
|---|---|
| Between A and (B–D) . . . . | 1 |
| Between B and (C, D) . . . . | 1 |
| Between C and D . . . . . | 1. |

The point is that, whatever the final division, it must be derived from successive partitions of grouped degrees of freedom.

For example, a trial of soil management at East Malling has the following treatments :

A. Clean cultivation throughout the year.
B. Autumn weeds following clean cultivation.
C. Oats and tares sown in the autumn after clean cultivation.
D. Natural sward.
E. Sown grass sward.
F. Straw turned in as manure.
G. Farm yard manure.

The first partition is as follows :

| | |
|---|---|
| Between (D, E) and (A, B, C, F, G) . . | 1 |
| Between D and E . . . . . . | 1 |
| Between (A, B, C, F, G) . . . . . | 4. |

This then becomes :

| | |
|---|---|
| Between (D, E) and (A, B, C, F, G) . . | 1 |
| Between D and E . . . . . . | 1 |
| Between (A, B, C) and (F, G) . . . . | 1 |
| Between (A, B and C) . . . . . | 2 |
| Between F and G . . . . . . | 1. |

As a matter of convenience in computing, it is advisable to partition any effects with more than one degree of freedom into components each with only one. This may be done in any way that suggests itself, because the two parts are eventually to be put together again. The division to be adopted here is

| | |
|---|---|
| Between A and (B, C) . . . . | 1 |
| Between B and C . . . . . | 1. |

Also, the first and third lines in the proposed partition (between (D, E) and (A, B, C, F, G) ; between (A, B, C) and (F, G)) can reasonably be combined. The treatments have fallen into three groups (I, Clean cultivation ; II, Swards ; III, Manuring) and these two effects represent respectively II versus (I, III) ; I versus III. Together, they represent a partition of the differences between the three groups.

To introduce some actual data, the following totals for the seven treatments represent weight of prunings over a three-year period from the apple trees planted on these plots. Figures are in pounds transformed to [$log\ (n + \frac{3}{8}) + 2$] (see Section 73). Each represents the sum of 18 data, the factor of soil treatments having been assigned to the sub-plots of a design with doubly split-plots.

| *A* | *B* | *C* | *D* | *E* | *F* | *G* |
|---|---|---|---|---|---|---|
| 40·01 . | 33·15 . | 31·50 . | 15·78 . | 13·92 . | 35·44 . | 36·94 |

These gave rise to an analysis of variance of which the following is an extract :

| *Source* | *d.f.* | *s.s.* | *m–s.* | *F* |
|---|---|---|---|---|
| Treatments . . | 6 . | 36·077 021 . | 6·012 837 . | 52·09*** |
| Error . . . | 24 . | 2·770 229 . | 0·115 426 | |

The first degree of freedom to be isolated relates to the difference between Group II (D and E) on the one hand and Groups I and III (A, B, C, F and G) on the other, *i.e.*, the comparison of two treatments with five. It is now necessary to work out the sum of squares associated with this degree of freedom.

Using square brackets to indicate treatment totals, the first step is to work out

$$5\,[D] + 5\,[E] - 2\,[A] - 2\,[B] - 2\,[C] - 2\,[F] - 2\,[G] = -205{\cdot}58.$$

This should be squared and divided by 18 (the number of data to a total) and by 70 (the sum of the coefficients squared, *i.e.*, $5^2 + 5^2 + (-2)^2 + (-2)^2 + (-2)^2 + (-2)^2 + (-2)^2$. The sum of squares is therefore 33·542 172.

It will be convenient next to consider the other intergroup comparison, *i.e.*, that of A, B, C(I) against F, G(III). The sum of squares is here*

$$(2\,[A] + 2\,[B] + 2\,[C] - 3\,[F] - 3\,[G])^2/18.30 = 0{\cdot}113\,245.$$

Together these make up a sum of squares of 33.655 417 with two degrees of freedom.

The first comparison within a group is that between A and B and C. Between A and (B, C) the sum of squares is $(2\,[A] - [B] - [C])^2/18{\cdot}6 = 2{\cdot}187\,379$, while between B and C it is $([B] - [C])^2/18{\cdot}2 = 0{\cdot}075\,625$. Putting these two components together—it will be recalled that they were separated only as a matter of computing convenience—gives a sum of squares of 2·263 004 with two degrees of freedom.

The other effects are easily derived. The sum of squares between D and E is

$$([D] - [E])^2/18.2 = 0{\cdot}096\,100$$

and between F and G it is similarly 0·062 500.

The partitioned analysis can now be written down.

| *Source* | *d.f.* | *s.s.* | *m–s.* | *F* |
|---|---|---|---|---|
| Between groups | 2 | 33·655 417 | 16·827 709 | 145·79*** |
| Within Group I | 2 | 2·263 004 | 1·131 502 | 9·80*** |
| Within Group II | 1 | 0·096 100 | 0·096 100 | $<1$ |
| Within Group III | 1 | 0·062 500 | 0·062 500 | $<1$ |
| Error | 24 | 2·770 229 | 0·115 426 | |

It is important to add up degrees of freedom and sums of squares to make sure that each sums to the unpartitioned values.

From this analysis it is clear that Groups II and III are homogeneous, *i.e.*, there is no suggestion that Treatment D differs from E, or F from G. Group I, however, is plainly made up of diverse elements which may be distinguished by working out the significant difference between two means, each of 18 data, *i.e.*, 0·234 ($P = 0{\cdot}05$).

This leads to the result :

$$\begin{array}{ccc} A & B & C \\ 2{\cdot}223 & \underbrace{1{\cdot}842 \quad 1{\cdot}750} \end{array}$$

Since B and C do have something in common, they may well give similar results (see Section 72) but it is not possible to be certain.

As to the differences between the groups, which have been shown to be significant, the only one that can be examined is that between II and III because these are the only homogeneous groups. The means are II, 0·825 ; III, 2·011. The significant difference between two means,

* Perhaps this useful notation needs a certain amount of explanation. The full stop, when used to denote multiplication, indicates that the multiplication takes place first and the other operations afterwards. Thus, (a + b × c) can mean either (a + bc) or (a + b)c, but (a + b.c) can mean only (a + bc). When used with numbers the full stop should not be confused with a decimal point, which is above the line.

each based on 36 data, is 0·165 ($P = 0{\cdot}05$); so the difference is evidently far beyond that needed for significance. No generalization should be made about Group I, but it would be quite legitimate to compare individual treatments, or even a group of treatments that were believed to be giving similar results, with one or other of the homogeneous groups using significant differences. Thus Treatments B and C with a combined mean of 1·796 might be compared with Group III, with a mean of 2·011. If this were thought both permissible and desirable, it has already been found that the significant difference between two means, each based on 36 data, is 0·165 ($P = 0{\cdot}05$), so the conclusion would be that the treatments involving autumn sowing, whether artificial or natural, followed by turning in during the spring, do not lead to as high a pruning weight as the manurial treatments.

With this form of partitioning it must be emphasized that a *possible* way of splitting up the degrees of freedom for treatments is not necessarily a *permissible* one. So many ways are available for partitioning a set of treatment effects that, if enough were tried, something would be almost bound to appear significant, even if the data were derived from a table of random numbers. It follows that partitioning should take place only where the groups have been nominated at the time of planning (see Section 71) or there is some other good reason for doing so.

*Partitioning of Interactions*

There remains the problem of partitioning interactions where one or more of the main effects from which they are derived has a nominated comparison. The method can be illustrated by the following example on fictitious data in which it is supposed that one factor has levels, A, B, C and D, and the other has X, Y and Z.

| | *A* | *B* | *C* | *D* | |
|---|---|---|---|---|---|
| X | 14 | 17 | 16 | 17 | *64* |
| Y | 18 | 17 | 15 | 16 | *66* |
| Z | 15 | 19 | 19 | 18 | *71* |
| | *47* | *53* | *50* | *51* | *201* |

These figures will be taken to represent totals for treatments, those in the body of the table being based on four observations.

First it will be supposed that the comparison of A with B, C and D has been nominated as has the comparison of Y with X and Z. The sum of squares for the former will be derived from the quantity $(3.47 - 53 - 50 - 51)$ and the latter from $(-64 + 2.66 - 71)$. Their interaction is similarly derived from

$$(-3.14 + 1.17 + 1.16 + 1.17 + 6.18 - 2.17 - 2.15 - 2.16$$
$$- 3.15 + 1.19 + 1.19 + 1.18) = 31$$

The only difficulty here is with coefficients. The first sum, 14, represents the total data from plots having both A (coefficient, $+3$) and X (coefficient, $-1$) so it has as its own coefficient $(+3)(-1) = -3$ and so on.

The sum of squares is accordingly $31^2/72.4 = 3{\cdot}34$. The divisor 72 is as usual the sum of coefficients squared, *i.e.* $(-3)^2 + 1^2 + \quad . \; . \; . \quad + 1^2$, and the divisor 4 is, of course, the number of observations upon which each of the quantities, 14, 17, 16, . . . , 18, is based.

The interpretation of this sum of squares should be quite easy provided the partitioning of the main effects has made sense, which it should have done. It is supposed that A differs in some respect from B, C and D, and Y in some respect from X and Z. The interaction concerns the degree to which the effect of A in comparison with the rest depends upon the other factor having Level Y on the one hand or X and Z on the other, X and Z having something in common not possessed by Y. Equally it may be considered from the point of view of the effect of A against (B, C and D) on the difference between Y and (X, Z).

In this example it has been supposed that the main effects of both factors require partitioning, but one of them, say the second, might have no nominated comparison. This would happen, for example, if it represented blocks as in the cases considered in Section 73. How in this instance is the sum of squares to be computed ?

There are two ways of going to work. One is to impose some artificial partition on the second factor in order to obtain a number of components that can be put together again afterwards. Thus, here the interaction of $(3A - B - C - D) \times (-X + 2Y - Z)$ is known; it is only necessary to add to it the interaction of $(3A - B - C - D) \times (X - Z)$ to obtain the interaction of $(3A - B - C - D)$ with the comparison of X, Y and Z generally. This second component equals $(3.14 - 17 - 16 - 17 - 3.15 + 19 + 19 + 18)^2/24.4 = 0{\cdot}09$, so the whole is $3{\cdot}34 + 0{\cdot}09 = 3{\cdot}43$.

The other method is to work out $(3A - B - C - D)$ for each level of the other factor. Thus for X it equals $(3.14 - 17 - 16 - 17) = -8$, for Y it equals $+6$ and for Z $-11$. These quantities should be squared and added, the sum being then divided by 12.4 (12 as the sum of coefficients squared, $(3)^2 + (-1)^2 + (-1)^2 + (-1)^2$, and 4 as the number of observations upon which each of the sums is based.) This gives $[(-8)^2 + (+6)^2 + (-11)^2]/12.4 = 4{\cdot}60$. It now remains to subtract the sum of squares for the main effect of the comparison of A with B, C and D, which is $(3.47 - 53 - 50 - 51)^2/12.12 = 1{\cdot}17$, to leave $3{\cdot}43$, the same as before.

# APPENDIX III

## A General Method of Computing the Analysis of Variance whatever the Design

This appendix will set out a method of analysing data that is of very general application, and one that can, with certain extensions, be used for any valid design. Indeed, it will also provide an apparent method of analysis for a number of designs that are not permissible, so its existence must not be used to justify a lay-out of doubtful validity. As far as the present writer can make out, it has never been described before in full detail; but it must be very old and is in widespread use.

The method is extremely simple in conception and serves very well to illustrate what goes on in an analysis of variance. This is often a characteristic of general methods, which, though they can be complicated in application, are usually straightforward in principle. This particular method may be described as tedious but simple.

It will be illustrated on some data from the strawberry weedkiller trial set out in Figure XXV. The purpose of this trial was to apply certain weedkillers of proven efficacy to strawberries to see if any harm resulted. Each of the four blocks contained seven plots, two of which were allocated to the untreated control and one to each of the four weedkillers. This left one plot per block with no treatment allocated and these four plots were assigned to a different weedkiller in each block—to A in Block I, to D in II, to B in III and to C in IV. The following data represent the total spread in inches for twelve strawberry plants per plot in August, approximately two months after the application of treatments in June. First, however, in this instance though not always, a change in notation can usefully be made. In this design, as in many others, there is a correspondence between blocks and treatments, which can be marked by designating each weedkiller not by its original letter but by the number of the block in which it is applied twice. To complete this transfer to a new system the control will be termed Treatment 5. The data are:

| *Block I* | | *Block II* | | *Block III* | | *Block IV* |
|---|---|---|---|---|---|---|
| 4, 107 | . | 1, 136 | . | 3, 118 | . | 5, 173 |
| 1, 166 | . | 5, 146 | . | 1, 117 | . | 4, 95 |
| 2, 133 | . | 4, 104 | . | 5, 176 | . | 4, 109 |
| 3, 166 | . | 3, 152 | . | 2, 132 | . | 1, 130 |
| 5, 177 | . | 2, 119 | . | 3, 139 | . | 2, 103 |
| 1, 163 | . | 5, 164 | . | 5, 186 | . | 5, 185 |
| 5, 190 | . | 2, 132 | . | 4, 103 | . | 3, 147 |

For each plot, the first figure represents the treatment and the second the total spread of the strawberry plants.

Before discussing the details of the method, it may be as well to explain what is meant by a "parameter". Let $\alpha$, known as a "general parameter", represent an average for the quantity under study over all blocks and treatments. It cannot be estimated just by adding together the data and dividing by 28 because some treatments are represented more frequently than others, but the reader will concede that such a quantity can well exist though it is not for the moment apparent how it may be arrived at. Further, let $\beta_1$, a "block parameter", represent the difference made by the plot coming from Block I. In an orthogonal design this would be estimated simply by the difference between the block mean and the general mean, but here the expression will be more complicated. Such a parameter exists for each of the four blocks and, in general, $\beta_j$ will represent the parameter for Block $j$, where $j$ takes values from 1 to 4. Finally, let $\gamma_1$ represent the difference made by the plot receiving Treatment 1. These "treatment parameters" as they are called are likewise not to be evaluated by any simple method because each treatment is represented differently on the blocks. There are five such parameters, which will be written $\gamma_k$ for Treatment $k$, where $k$ can take any value from 1 to 5. No more parameters are needed because the design is based on a two-way classification by blocks and treatments.

If now, a plot in Block $j$ receives Treatment $k$, it is to be expected that the figure for that plot will be $\alpha + \beta_j + \gamma_k$, but examination of this expression will show that many sets of parameters could give equivalent results. Thus, if $\alpha$ were increased by 10, all the $\beta$'s decreased by 6 and all the $\gamma$'s decreased by 4, the final result would be the same for all plots. In order to overcome this lack of definiteness, a convention will be adopted that all the $\beta$'s must sum to zero and so must all the $\gamma$'s, *i.e.*,

$$\beta_1 + \beta_2 + \beta_3 + \beta_4 = 0 \qquad [1]$$

$$\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 + \gamma_5 = 0 \qquad [2].$$

These will be termed "equations of constraint". The numbers in square brackets are reference numbers for use when these equations are appealed to at a later stage.

On account of the error, the above expression, $\alpha + \beta_j + \gamma_k$, will not exactly equal the actual figure, $y$, given by the plot. The deviation of the actual results from those expected over the trial as a whole will be measured by taking all these differences, squaring them and adding the squares. This sum of $(y - \alpha - \beta_j - \gamma_k)^2$ will be termed $S$.

The object of the analysis is to see if it is really necessary to postulate an effect of treatments in order to explain the data (see Section 12), *i.e.*, to see if the data cannot be satisfactorily explained on a basis of the general parameter aided only by block parameters. The next step, therefore, is to start again with parameters $\alpha'$ and $\beta_j'$ (the primes are a reminder that the parametric values may be different from last time), and to evaluate the sum of squares, $S'$, which is the sum of $(y - \alpha' - \beta_j')^2$ over all plots. If there is not really any need to postulate the existence of the $\gamma$'s, a measure of error based on $S'$ should be no greater than one based on $S$. Such is a brief explanation of the approach. It may not satisfy the mathematician, but it will, perhaps, help some readers to see their way through the algebra. The computations have six stages:

1. *Computation of Classification Totals.*—These are worked out as for an orthogonal design. Let the sum of data in Block $j$ be $B_j$ and the sum of data from plots receiving Treatment $k$ be $T_k$ and let the grand total be $G$. Then:

$$T_1 = 712,\ T_2 = 619,\ T_3 = 722,\ T_4 = 518,\ T_5 = 1397,$$

$$B_1 = 1102,\ B_2 = 953,\ B_3 = 971,\ B_4 = 942,\ G = 3968.$$

2. *Parametric Equations and their Solution.*—The next step is to write down the parameters corresponding to each classification total, bearing in mind the basic parametric expression for the result from a single plot, namely, $\alpha + \beta_j + \gamma_k$.

Thus, for $B_1$, the total from Block I, the equation is $7\alpha + 7\beta_1 + 2\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 + 2\gamma_5 = B_1$, because there are seven plots, all of which will contribute an $\alpha$ and a $\beta_1$ to the total, while two will contribute a $\gamma_1$, two a $\gamma_5$ and the others will each contribute one of the other $\gamma$'s. On account of equation [2] this may be written more simply as $7\alpha + 7\beta_1 + \gamma_1 + \gamma_5 = B_1$. In this way may be obtained the so-called "normal parametric equations", as follows:

$$7\alpha + 7\beta_1 + \gamma_1 + \gamma_5 = B_1 = 1102 \qquad [3]$$

$$7\alpha + 7\beta_2 + \gamma_2 + \gamma_5 = B_2 = 953 \qquad [4]$$

$$7\alpha + 7\beta_3 + \gamma_3 + \gamma_5 = B_3 = 971 \qquad [5]$$

$$7\alpha + 7\beta_4 + \gamma_4 + \gamma_5 = B_4 = 942 \qquad [6]$$

$$5\alpha + \beta_1 + 5\gamma_1 = T_1 = 712 \quad . \quad . \quad . \quad . \quad [7]$$

$$5\alpha + \beta_2 + 5\gamma_2 = T_2 = 619 \quad . \quad . \quad . \quad . \quad [8]$$

$$5\alpha + \beta_3 + 5\gamma_3 = T_3 = 722 \quad . \quad . \quad . \quad . \quad [9]$$

$$5\alpha + \beta_4 + 5\gamma_4 = T_4 = 518 \quad . \quad . \quad . \quad . \quad [10]$$

$$8\alpha + 8\gamma_5 = T_5 = 1397 \quad . \quad . \quad . \quad . \quad [11]$$

$$28\alpha + 3\gamma_5 = G = 3968 \quad . \quad . \quad . \quad . \quad [12]$$

These equations provide a ready means of evaluating the parameters.

The first step is usually to evaluate $\alpha$ and the parameter for any treatment having a different degree of replication from the rest, in this instance $\gamma_5$. These may readily be obtained from equations [11] and [12], and prove to equal $+ 137{\cdot}765$ and $+ 36{\cdot}860$ respectively. It will be noted that each has been evaluated to three places of decimals more than the original data.

Substituting these values throughout the other parametric equations, it will appear that they can be solved in pairs. If the design had been orthogonal, each could have been solved separately*; while in more complicated designs it may be necessary to take them by larger groups than pairs. Here [3] and [7] lead to the results that $\beta_1 = + 14{\cdot}140$ and $\gamma_1 = + 1{\cdot}807$, [4] and [8] to $\beta_2 = - 5{\cdot}037$ and $\gamma_2 = - 12{\cdot}958$, [5] and [9] to $\beta_3 = - 5{\cdot}419$ and $\gamma_3 = + 7{\cdot}719$ and [6] and [10] to $\beta_4 = - 3{\cdot}684$ and $\gamma_4 = - 33{\cdot}428$. There is no need to check these results, for it suffices to substitute the values thus obtained into the normal equations and the equations of constraint to see if all are satisfied.

3. *Evaluation of S and S'*.—The sum of squares, $S$, is readily obtained by multiplying each classification total by its corresponding parameter and subtracting all such products from the sum of data squared, thus:

$$\begin{aligned} S &= (107^2 + 136^2 + 118^2 + \quad . \;.\; . \quad + 103^2 + 147^2) \\ &\qquad - (14{\cdot}140)\,(1102) - (-\,5{\cdot}037)\,(953) - \quad . \;.\; . \quad - (137{\cdot}765)\,(3968) \\ &= 585\;178{\cdot}00 - 581\;717{\cdot}78 \\ &= 3\;460{\cdot}22. \end{aligned}$$

The sceptic may like to work out $(y - \alpha - \beta_j - \gamma_k)^2$ for each plot and assure himself that all adds up correctly, but this is *not* needed in ordinary computing! Nonetheless, the evaluation of $S$ needs to be done carefully because the minus signs make mistakes easy and there is no ready check available.

The evaluation of $S'$ is much easier because there is now a one-way classification and the question of non-orthogonality does not arise. In fact,

$$\begin{aligned} S' &= (107^2 + 136^2 + 118^2 + \quad . \;.\; . \quad + 103^2 + 147^2) - \tfrac{1}{7}\,(1102^2 + 953^2 + 971^2 + 942^2) \\ &= 585\;178{\cdot}00 - 564\;688{\cdot}29 \\ &= 20\;489{\cdot}71. \end{aligned}$$

If, of course, a non-orthogonal design had been left even when treatments were being ignored, the full procedure would have been needed here too.

The next step is to work out the degrees of freedom. $S$ has 20, because there are 27 degrees of freedom in all between 28 data, three of which are associated with blocks and four with treatments, leaving 20 for error. Another way is to work out the expression, number of plots (= 28) − number of parameters (= 10) + number of equations of constraint (= 2) = 20. This latter way is really the more fundamental, being derived from a theorem that introduces degrees of freedom to the analysis of variance. Similarly, $S'$ has 24 degrees of freedom.

* This indeed is the basic characteristic of orthogonal designs.

4. *Construction of Analysis of Variance.*—The sum of squares, $S$, represents the residual variation with treatment effects removed: $S'$ represents the residual variation including differences brought about by the treatments. In fact, $S$ is the error sum of squares: $S'$ is the error sum of squares together with the treatment sum of squares. The analysis therefore reads like this:

| *Source* | *d.f.* | *s.s.* | *m–s.* | *F* |
|---|---|---|---|---|
| Treatments (by difference) | 4 | 17 029·49 | 4 257·37 | 24·61*** |
| Error | 20 | 3 460·22 | 173·01 | |
| Error + Treatments | 24 | 20 489·71 | | |

5. *Computation of Adjusted Treatment Means.*—This looks difficult but is really quite easy, the adjusted mean for Treatment $k$ being $\alpha + \gamma_k$. The adjusted means for the treatments are, therefore, 1 = A, 139·6; 2 = D, 124·8; 3 = B, 145·5; 4 = C, 104·3 and 5 = 0, 174·6.

6. *Computation of Significant Differences.*—As with an orthogonal design, this step is needed only in certain circumstances. In the present example there are two cases to be considered: (1) between two means each of five data and (2) between one mean of eight and the other of five. Let the significant difference be written $t\theta\sqrt{\text{Error mean-square}}$, then the task is to evaluate $\theta$. This is rather tedious, but when it has been done once there is no need to do it again for the same lay-out, the same quantity being used in all future analyses.

It will be simpler to take a specific example, *i.e.*, the difference between the adjusted means of Treatments 1 and 2. The actual value is $(\gamma_1 - \gamma_2)$, which may be evaluated thus in terms of the data. By subtracting equation [4] from [3] and [8] from [7], it appears that—

$$7(\beta_1 - \beta_2) + (\gamma_1 - \gamma_2) = B_1 - B_2,$$

$$(\beta_1 - \beta_2) + 5(\gamma_1 - \gamma_2) = T_1 - T_2,$$

whence

$$34\,(\gamma_1 - \gamma_2) = 7(T_1 - T_2) - (B_1 - B_2).$$

It is now easiest to make a skeleton plan of the trial and, whenever a datum enters into the above expression, to write its coefficient in the space representing the appropriate plot, thus: Writing + 7 in all the spaces corresponding to Treatment 1, − 7 in all the spaces for Treatment 2, − 1 in all spaces representing plots in Block I and + 1 for Block II leads to—

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1<br>+7 −1 | 1<br>+7 −1 | 2<br>−7 −1 | 3<br>−1 | 4<br>−1 | 5<br>−1 | 5<br>−1 | Block I |
| 1<br>+7 +1 | 2<br>−7 +1 | 2<br>−7 +1 | 3<br>+1 | 4<br>+1 | 5<br>+1 | 5<br>+1 | Block II |
| 1<br>+7 | 2<br>−7 | 3 | 3 | 4 | 5 | 5 | Block III |
| 1<br>+7 | 2<br>−7 | 3 | 4 | 4 | 5 | 5 | Block IV |

Upper numbers indicate treatments.

It thus appears that two data enter into the expression for $34(\gamma_1 - \gamma_2)$ with coefficient $\pm$ 8, four with coefficient $\pm$ 7, four with coefficient $\pm$ 6, eight with coefficient $\pm$ 1, the others not entering at all. It follows that for the comparison under study

$$34^2.\theta^2 = 2.8^2 + 4.7^2 + 4.6^2 + 8.1^2,$$

whence $\theta^2$ equals $\frac{7}{17}$. If the design had been orthogonal, $\theta^2$ would have equalled $\frac{2}{5}$, so the efficiency factor is $\frac{2}{5} \cdot \frac{17}{7} = \frac{34}{35}$, which is fairly good. From the similarity of equations [3–6] and of [7–10], it is clear that the same result would have been reached whichever pair of treatments had been chosen.

The comparison of Treatments 1 and 5, to take an example of the other case, is rather more difficult. From equations [11] and [12] it follows that

$$50\gamma_5 = 7T_5 - 2G \quad . \quad . \quad . \quad . \quad [13]$$

and

$$200\alpha = 8G - 3T_5.$$

Substituting these results in equations [3] and [7], it appears that

$$7\beta_1 + \gamma_1 = B_1 - \frac{7}{200}(8G - 3T_5) - \frac{1}{50}(7T_5 - 2G)$$

and

$$\beta_1 + 5\gamma_1 = T_1 - \frac{1}{40}(8G - 3T_5).$$

Hence,

$$34\gamma_1 = 7T_1 - B_1 - \frac{7}{50}(8G - 3T_5) + \frac{1}{50}(7T_5 - 2G),$$

which is more easily written

$$1700\gamma_1 = 350T_1 + 28T_5 - 50B_1 - 58G.$$

From [13], by simple multiplication, it appears that

$$1700\gamma_5 = 238T_5 - 68G.$$

Hence

$$170(\gamma_1 - \gamma_5) = 35T_1 - 21T_5 - 5B_1 + G.$$

The plan of coefficients now reads—

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1<br>35 −5+1 . | 1<br>35 −5+1 . | 2<br>−5+1 . | 3<br>−5+1 . | 4<br>−5+1 . | 5<br>−21 −5+1 . | 5<br>−21 −5+1 | Block I |
| 1<br>35 +1 . | 2<br>+1 . | 2<br>+1 . | 3<br>+1 . | 4<br>+1 . | 5<br>−21+1 . | 5<br>−21+1 | Block II |
| 1<br>35 +1 . | 2<br>+1 . | 3<br>+1 . | 3<br>+1 . | 4<br>+1 . | 5<br>−21+1 . | 5<br>−21+1 | Block III |
| 1<br>35 +1 . | 2<br>+1 . | 3<br>+1 . | 4<br>+1 . | 4<br>+1 . | 5<br>−21+1 . | 5<br>−21+1 | Block IV |

This leads to the result that

$$170^2 \cdot \theta^2 = 3.36^2 + 2.31^2 + 2.25^2 + 6.20^2 + 3.4^2 + 12.1^2,$$

so $\theta^2 = \frac{28}{85}$. For an orthogonal design, $\theta^2$ would have equalled $\left(\frac{1}{8} + \frac{1}{5}\right) = \frac{13}{40}$, so the efficiency factor is $\frac{13}{40} \cdot \frac{85}{28} = \frac{221}{224}$, which is very good, being nearly unity.

Applying these expressions for the significant differences to the example, it appears that the significant difference between two means of five is

$$t\sqrt{\frac{7}{17}\text{ Error mean-square}} = 2{\cdot}086\sqrt{71{\cdot}239\,412}$$

which is 17·6 ($P = 0{\cdot}05$). Similarly, for the difference of one mean of five and the other of eight, the value needed for significance is

$$t\sqrt{\frac{28}{85}\text{ Error mean-square}} = 2{\cdot}086\sqrt{56{\cdot}991\,529} = 15{\cdot}7\ (P = 0{\cdot}05).$$

The results may therefore be set out thus,

$$\begin{array}{ccccc} C = 4 & D = 2 & A = 1 & B = 3 & O = 5 \\ 104 & 125 & 140 & 145 & 175 \end{array}$$

The conclusion is, in fact, that all the weed-killers harmed the strawberries to some extent, at least by inhibiting spread, some doing so to a significantly greater degree than others. To give a formal demonstration of this last point would involve partitioning the treatments sum of squares, a task that will not be attempted here for a non-orthogonal design.

As with Appendix I, the aim here has been to set out all the computations that might be called for with data such as these rather than those called for in this particular trial. Thus, the object of the experiment being to find out if any weedkillers gave results different from the control, there would be no need to use the $F$-test or the significant difference between two means of five observations, but only the significant difference between one mean of eight and another of five. Some might also be interested in the accuracy as measured by the Error mean-square, but since all the treatments have been shown to differ from the control, the trial was plainly accurate enough for its purpose.

# APPENDIX IV

## The Analysis of Covariance in an Orthogonal Design

In the analysis of covariance it is supposed that the quantity under study is subject to disturbance by some other quantity, which is distributed over the plots in a manner unrelated to treatments. Thus, in an analysis of the weight of fruit in an experiment it might be expected that the crop from a plot depended to some extent upon the size of the trees at the beginning of the experiment before the treatments were applied, and it might then seem desirable to make allowance for these initial differences.

In such cases the quantity under study is termed the " dependent variate " and will here be denoted by the symbol, $y$. The quantity causing the disturbance is termed the " independent variate " and will be written, $x$. The assumptions that must be justified before undertaking an analysis of covariance do not usually cause much trouble. They may be set out thus :

(i) The independent variate, $x$, is independent of the treatments. Ideally $x$ should be measured before any treatments have been applied. If $x$ is related to the treatments, it must be recognized that the elimination of its effect eliminates also part of the treatment effect (see Section 80).

(ii) The dependent variate, $y$, gives the same error for each treatment (see Section 73), or it must be transformed, if need be, to ensure that this is so. No similar assumption need be made with regard to $x$.

(iii) The relationship between $x$ and $y$ (or the transformed value of $y$) must be of the form, $y = a + bx$, where $a$ and $b$ are constant but unknown numbers. To make sure that this relationship holds it may be necessary to transform $x$, though small departures from this relationship are usually not serious.

(iv) The value of $b$ must be independent of the treatments, *i.e.*, if a change in $x$ produces a certain effect on $y$ for plots receiving a certain treatment, it must produce the same effect for any other treatment. If $x$ or $y$ or both have been transformed, this property should hold for the transformed variates, not the untransformed.

Before proceeding to the details of the analysis of covariance, it may be worth while to explain a little about the association of two variates in the simple case where there is no classification by blocks, rows, columns or treatments but only a single series of paired observations. For the moment it will not matter which is the dependent variate and which the independent. The following figures represent $x$, the minimum air temperature at East Malling for the week preceding the time of writing this appendix, and $y$, the corresponding minimum on grass, both in ° F.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $x$, | 42 | 51 | 56 | 55 | 50 | 50 | 59, | Total 363, | Mean 51·9. |
| $y$, | 40 | 48 | 54 | 52 | 48 | 48 | 55, | Total 345, | Mean 49·3. |

The first step is to work out the sum of data squared for $x$ and to subtract a correction term as in Appendix I, and similarly for $y$. This gives

$$C_{xx} = (42^2 + 51^2 + \quad . \; . \; . \quad + 59^2) - \tfrac{1}{7}(363)^2 = 182{\cdot}86,$$

$$C_{yy} = (40^2 + 48^2 + \quad . \; . \; . \quad + 55^2) - \tfrac{1}{7}(345)^2 = 153{\cdot}43.$$

$C_{xy}$ is then worked out by using the product of $x$ and $y$ in every case instead of $x^2$ or $y^2$, thus :

$$C_{xy} = (42.40 + 51.48 + \quad . \; . \; . \quad + 59.55) - \tfrac{1}{7}(363)(345) = + 166{\cdot}29.$$

The sign is added because $C_{xy}$ can be positive or negative, unlike $C_{xx}$ and $C_{yy}$, which are always positive.

It is now possible to work out the "correlation coefficient" between $x$ and $y$. This measures the degree of association between the two and equals $C_{xy}/\sqrt{C_{xx} \cdot C_{yy}}$, so in the example it equals $+ 0{\cdot}9928$. The value varies from $- 1$ for complete negative association (*i.e.*, an increase in $x$ means a decrease in $y$ and *vice versa*, the value of one exactly determining the value of the other) by way of zero in the case of no association to $+ 1$ in the case of complete positive association (*i.e.*, an increase in one means an increase in the other, the value of one exactly determining the value of the other). This particular value is high and shows that the two variates are associated positively but not quite completely (*i.e.*, the value of one does not quite determine the value of the other, though it nearly does so).

Various other measures of correlation have been worked out for use in special circumstances and, in their application to horticulture, have been considered by Crist [43, 44, 45, 46, 47].

Correlation deals with the closeness of the association between the two variates; it does not say what the relationship is, this being the province of "regression". The regression coefficient of $y$ on $x$ is $C_{xy}/C_{xx} = + 0{\cdot}9094 = b$.* This means that if $x$ changes its value by one, $y$ may be expected to change its value by $b$. Thus, in this example, on the first day, when $x$ differed from its mean by $(42 - 51{\cdot}9)$°F. $= - 9{\cdot}9$°F., it is to be expected that $y$ will differ from its mean by $(- 9{\cdot}9)(+ 0{\cdot}9094)$°F. $= - 9{\cdot}0$°F. This gives $(49{\cdot}3 - 9{\cdot}0)$°F. as the grass minimum temperature to be expected on a day with an air minimum of 42° F. Actually, the figure was 40°F., which is as near as could be expected. The correlation coefficient, though high, was not equal to one, and this gives warning that the grass minimum cannot be calculated exactly from the air minimum, but other factors must operate as well.

Similarly, the regression of $x$ on $y$ is $C_{xy}/C_{yy} = 1{\cdot}0838$. Consequently two lines can be drawn on a graph, one showing the value of $y$ to be expected for each $x$ and the other the value of $x$ to be expected for each $y$. For cases of incomplete association these will not be the same, and care is sometimes needed in choosing the right one in any particular instance.

The significance of correlation coefficients can be judged from tables [66c] or by the following analysis of variance. ($r$ is the correlation coefficient, $n$ is the number of pairs of observations upon which it is based.)

| *Source* | *d.f.* | *s.s.* | *m-s.* | *F* |
|---|---|---|---|---|
| Correlation | $1$ | $r^2$ | $r^2$ | $\dfrac{(n-2)r^2}{(1-r^2)}$ |
| Error | $n-2$ | $1-r^2$ | $(1-r^2)/(n-2)$ | |
| Total | $n-1$ | $1$ | | |

This method is based on an important result, namely, that the variation in $y$ after allowance has been made for $x$ is given by the sum of squares, $C_{yy}(1 - r^2) = C_{yy} - C_{xy}^2/C_{xx}$, with one degree of freedom fewer than $C_{yy}$.

However, with correlations, as with other things, a result can be significant without being important. Unless $r$ lies outside the range $\pm 0{\cdot}7$, the relationship is not close enough to be of much use in estimating one variate from another. Further, a relationship can be significant without being causative. For example, in a recent examination in statistics, candidates were given the figures for $x$, the number of persons certified insane in Britain over a period of years, and $y$, the number of broadcast receiving licences issued, and were asked to comment on the high positive correlation coefficient between the two. It is not recorded if many candidates proffered theories (*a*) that people listened to the radio because they were insane, or (*b*) that they became insane on account of having listened to the radio. Obviously the relationship is not causative, but results from both variates increasing with time, *i.e.*, from common causation. This does not detract in any way from the usefulness of the regression for purposes of estimation. A person who wanted to know how many licences were issued in a certain year

* Some people find this nomenclature a little confusing, the effect of $x$ on $y$ being measured by the regression of $y$ on $x$.

within the period under study would be quite entitled to work it out from the insanity figures, supposing he remembered the regression formula.

The analysis of covariance is a development of regression methods and, like them, has the advantage of making no assumptions as to the exact value of $b$, the change brought about in $y$ by a unit change in $x$. Indeed, the only assumption that has to be made concerning the relationship between $x$ and $y$ is that it is something like a straight line. Thus, it becomes possible to adjust $y$ by $x$ on the basis of the data itself. Given that a tree gave 500 lbs. of fruit over a four-year period of calibration, there is no obvious means of forecasting how much fruit it would give over the six-year period of the subsequent experiment, and any estimate that was made would be regarded with some suspicion by experienced investigators. If regression methods are used, the relationship derived is known to be appropriate to the circumstances. In particular, if it happens that for some reason a useless calibrating variate has been chosen, little harm will have been done, because in that case $b$ comes out to be zero or something small. Accordingly, with the method of covariance it is permissible to " take a chance " on an independent variate that may have something to do with the problem though the experimenter cannot be certain. If it is irrelevant, the resulting adjustments will be negligible: if relevant, something will have been learnt about a further possibility in calibration.

It is now possible to return to the problem of the analysis of covariance. The data in this example are taken from a cover crop experiment with apples at East Malling in which $y$, the crop in pounds over a four-year period, was adjusted by $x$, the crop in bushels over the immediately preceding four-year period during which no differential treatments had been applied. The experiment was designed in randomized blocks.

| | *I* | *II* | *III* | *IV* |
|---|---|---|---|---|
| A | 8·2, 287 | 9·4, 290 | 7·7, 254 | 8·5, 307 |
| B | 8·2, 271 | 6·0, 209 | 9·1, 243 | 10·1, 348 |
| C | 6·8, 234 | 7·0, 210 | 9·7, 286 | 9·9, 371 |
| D | 5·7, 189 | 5·5, 205 | 10·2, 312 | 10·3, 375 |
| E | 6·1, 210 | 7·0, 276 | 8·7, 279 | 8·1, 344 |
| F | 7·6, 222 | 10·1, 301 | 9·0, 238 | 10·5, 357 |

The grouping of the figures is important. If the independent variate is always the left-hand one of a pair and if the spaces between the pairs are well marked, confusion will be minimized. The computations have six stages:

1. *Computation of Classification Totals.*—This is done as for the analysis of variance except that there are two variates to be totalled. It is advisable to write these down in pairs as with the data, the independent variate to the left. If they are written in the margins of the table of data, a different coloured ink should be used. In the example, the totals are—

Blocks: I, 42·6, 1413; II, 45·0, 1491; III, 54·4, 1612; IV, 57·4, 2102.

Treatments: A, 33·8, 1138; B, 33·4, 1071; C, 33·4, 1101; D, 31·7, 1081; E, 29·9, 1109; F, 37·2, 1118.

Grand Total, 199·4, 6618.

2. *Computation of Variation Terms.*—These have to be done in triplicate, once for $x^2$, once for $xy$ and once for $y^2$. The terms for $x^2$ and $y^2$ are worked out exactly as in an analysis of variance of $x$ and $y$. Those for $xy$ are obtained by multiplying together corresponding values of $x$ and $y$ instead of squaring one of them. Thus the Total Term for $xy$ is $(8{\cdot}2.287 + 9{\cdot}4.290 + 7{\cdot}7.254 + \ldots + 10{\cdot}5.357)$ and the Block term is $\frac{1}{6}(42{\cdot}6.1413 + 45{\cdot}0.1491 + 54{\cdot}4.1612 + 57{\cdot}4.2102)$ and so on. The results may be set out thus:

| | $x^2$ | $xy$ | $y^2$ |
|---|---|---|---|
| Total | 1 713·940 0 | 56 649·500 | 1 896 948·00 |
| Blocks | 1 682·313 3 | 55 939·400 | 1 872 766·33 |
| Treatments | 1 664·075 0 | 55 006·400 | 1 825 663·00 |
| Correction | 1 656·681 7 | 54 984·550 | 1 824 913·50 |

It will be noticed that the number of decimal places for the $xy$-term is the mean of those given for the $x^2$- and $y^2$-terms.

3. *Construction of the Analyses of Variance for $x^2$ and $y^2$ and the Analysis of Covariance for $xy$.*—The sums of squares and products are worked out by exactly the same formulae, using exactly the same checks, as are the sums of squares for an analysis of variance. The sole difference to be noted is that a sum of products, unlike a sum of squares, can on occasion be negative. The analyses are as follows:

| *Source* | *d.f.* | $x^2$ | $xy$ | $y^2$ |
|---|---|---|---|---|
| Blocks | 3 | 25·631 6 | + 954·850 | 47 852·83 |
| Treatments | 5 | 7·393 3 | + 21·850 | 749·50 |
| Error | 15 | 24·233 4 | + 688·250 | 23 432·17 |
| Total | 23 | 57·258 3 | + 1664·950 | 72 034·50 |

The figures in the error line after the degrees of freedom are respectively $C_{xx}$, $C_{xy}$ and $C_{yy}$, so $b$, which equals $C_{xy}/C_{xx}$, is + 28·40. This means that an additional bushel of apples from a plot during the calibrating period is associated with an additional 28·40 lbs. of fruit from the plot during the period of the trial, a result that could not have been derived in any other way.

In this trial there was a nominated effect, namely, the difference between Treatment F (clean cultivation) and the others. For $y^2$ the sum of squares is

$$(5.1118 - 1138 - 1071 - 1101 - 1081 - 1109)^2/4.30 = 67{\cdot}50.$$

For $x^2$ it is—

$$(5.37{\cdot}2 - 29{\cdot}9 - 31{\cdot}7 - 33{\cdot}4 - 33{\cdot}4 - 33{\cdot}8)^2/4.30 = 4{\cdot}720\,3.$$

Likewise for $xy$ it is

$$(5.1118 - 1138 - \;\ldots\; - 1109)\,(5.37{\cdot}2 - 29{\cdot}9 - \;\ldots\; - 33{\cdot}8)/4.30 = 17{\cdot}850.$$

The analysis may therefore be written out more fully, thus:

| *Source* | *d.f.* | $x^2$ | $xy$ | $y^2$ |
|---|---|---|---|---|
| Blocks | 3 | 25·631 6 | 954·850 | 47 852·83 |
| F versus A–E | 1 | 4·720 3 | 17·850 | 67·50 |
| Between A–E | 4 | 2·673 0 | 4·000 | 682·00 |
| Error | 15 | 24·233 4 | 688·250 | 23 432·17 |
| Total | 23 | 57·258 3 | 1 664·950 | 72 034·50 |

4. *Construction of the Analysis of Variance for $y$ adjusted by $x$.*—This stage of the computation has a marked resemblance to the method presented in Appendix III. Two estimates of error are worked out, one excluding the effect of the treatments under consideration and the other including it. Thus, working with the error line alone from the above analysis, the sum of squares for $y$ after adjustment by $x$ is $C_{yy} - C_{xy}^2/C_{xx}$, which equals 3 885·26. This, which has 14 degrees of freedom, will be written $S$.

It is possible at this point to break off the main line of analysis to see if the association between the two variates is significantly close. Without adjustment the sum of squares is 23 432·17 with 15 degrees of freedom: with adjustment it is 3 885·26 with 14 degrees of freedom. This leads to the analysis of variance:

| *Source* | *d.f.* | *s.s.* | *m–s.* | *F* |
|---|---|---|---|---|
| Correlation (by difference) | 1 | 19 546·91 | 19 546·91 | 70·43*** |
| Error | 14 | 3 885·26 | 277·52 | |
| Error + Correlation | 15 | 23 432·17 | | |

showing that the reduction of error brought about by adjustment is indeed significant. The test is the same as that given for the significance of a correlation coefficient. In this particular instance there was no call to use it, because it is part of the lore of those who work with apples that these two variates are in fact associated, but if the independent variate had been something of a shot in the dark, a digression at this point might have been worth while to ensure that no further time should be wasted on the independent variate, if it should in fact be irrelevant.

Returning to the main line of the analysis, the next thing is to test if the group of treatments, A, B, C, D and E, is in fact homogeneous (see Section 71). This can be done by including in error the sums of squares and products associated with this effect and seeing if it makes any difference. Accordingly,

$$C_{xx}' = 24{\cdot}233\,4 + 2{\cdot}673\,0 \quad = 26{\cdot}906\,4,$$
$$C_{xy}' = 688{\cdot}250 + 4{\cdot}000 \quad = 692{\cdot}250,$$
$$C_{yy}' = 23\,432{\cdot}17 + 682{\cdot}00 \quad = 24\,114{\cdot}17,$$

each with 19 degrees of freedom. This makes $S' = C_{yy}' - C_{xy}'^2/C_{xx}'$, equal to 6 303·91 with 18 degrees of freedom, leading to the analysis :

| *Source* | *d.f.* | *s.s.* | *m–s.* | *F* |
|---|---|---|---|---|
| Between A–E (by difference) | 4 | 2 418·65 | 604·66 | 2·18 |
| Error | 14 | 3 885·26 | 277·52 | |
| Error + Between A–E | 18 | 6 303·91 | | |

Although the value of $F$ is not significant, it is of such a magnitude as to suggest that there may be differences between the effects of the various cover crops on cropping, but in the absence of definite significance the group is best assumed homogeneous.

The next test is to work out the difference between F and the homogeneous group A–E. The new values of $C_{xx}'$, etc., are

$$C_{xx}' = 24{\cdot}233\,4 + 4{\cdot}720\,3 \quad = 28{\cdot}9537,$$
$$C_{xy}' = 688{\cdot}250 + 17{\cdot}850 \quad = 706{\cdot}100,$$
$$C_{yy}' = 23\,432{\cdot}17 + 67{\cdot}50 \quad = 23\,499{\cdot}67,$$

each with 16 degrees of freedom, which makes $S'$ equal to 6 279·86 with 15 degrees of freedom. This leads to the analysis of variance :

| *Source* | *d.f.* | *s.s.* | *m–s.* | *F* |
|---|---|---|---|---|
| F versus A–E (by difference) | 1 | 2 394·60 | 2 394·60 | 8·63* |
| Error | 14 | 3 885·26 | 277·52 | |
| Error + F versus A–E | 15 | 6 279·86 | | |

From this it is clear that there is a significant difference between the clean cultivated treatment (F) and the various cover crops (A–E).

5. *Adjustment of Treatment Means.*—This is best illustrated by an example. Treatment A gave a mean of 284·5 for $y$ and 8·45 for $x$. The general mean for $x$ is 8·31, so it appears that Treatment A was applied to trees of rather greater fruitfulness than average. Since a change of one in $x$ corresponds to a change of 28·40 in $y$, it follows that the plots of A would have given a mean crop lower by 0·14 (28·40) lbs. had they contained average trees, 0·14 being the adjustment needed in $x$. That is to say, the adjusted mean value for $y$ for this treatment is

$$284{\cdot}5 - 0{\cdot}14\,(28{\cdot}40) = 280{\cdot}5.$$

The other treatment means are all obtained in the same way and are as follows :

B, 266·6 ; C, 274·1 ; D, 281·0 ; E, 300·8 and F. 251·4.

It now appears that F was not really a treatment productive of much crop, but it happened to be on trees of notably high inherent fruitfulness, which caused it to make a better showing on the unadjusted figures than its worth justified.

If the adjusted mean for a group of treatments is wanted, it may be obtained in one of two ways. Either the adjusted means of the component treatments can be added and divided, or the mean can be worked out afresh in the usual way. Thus, in this example, the overall mean for Treatments A–E can be calculated as one-fifth of the total of 280·5, 266·6, 274·1, 281·0 and 300·8, which is 280·6, or it can be obtained direct, *i.e.*, the mean of $y$ for all these treatments is 275·0 and the mean of $x$ is 8·11, the adjusted mean being therefore $275{\cdot}0 + 28{\cdot}40\ (8{\cdot}31 - 8{\cdot}11) = 280{\cdot}7$. The difference arises solely from the rounding off of decimal places and disappears on the results being expressed to three significant figures.

In this trial all adjustments were made on to the mean value of $x$, but this is not always done. Indeed, pseudo-variates are commonly adjusted to a standard value of zero.

6. *Computation of Significant Differences.*—In this particular example significant differences are not required, the conclusions being clear without them. It is nevertheless necessary to explain how to calculate them, because this happy state of affairs does not always obtain.

The general expression for the difference between two adjusted means, one based on $p$ plots and the other on $q$, with a difference of $D$ in their mean values of $x$, is [180].

$$t\sqrt{\left(\frac{1}{p} + \frac{1}{q} + \frac{D^2}{C_{xx}}\right) \text{Error mean-square}}$$

where the error mean-square is that in the analysis of $y$ as adjusted by $x$.

Thus, in this example, the significant difference between the mean of A–E and that of F is calculated thus: The values of $p$ and $q$ are respectively 20 and 4. $D$ equals $8{\cdot}11 - 9{\cdot}30 = -1{\cdot}19$ and $C_{xx}$ has already been evaluated as 24·233 4. The value of $t$ for $P = 0{\cdot}05$ and 14 degrees of freedom is 2·145, while the error mean-square is 277·52. This makes the significant difference 21·4. The actual difference exceeds this, being $280{\cdot}7 - 251{\cdot}4 = 29{\cdot}3$, but the significance of the difference is not really in doubt, having already been demonstrated by the analysis of variance of $y$ adjusted by $x$.

In order to avoid having to work out a fresh significant difference for each treatment comparison, it is usually convenient to work out the extreme significant differences between adjusted treatment means. Thus, in this example, if it had been necessary to work out the significance of all differences between the means of Treatments A–F it would have appeared that the smallest value of $D$ was that between Treatments B and C, which give the same mean value for $x$. The largest value of $D$, on the other hand, is 1·82 between Treatments E and F. Consequently, the smallest significant difference is—

$$t\sqrt{\left(\frac{1}{4} + \frac{1}{4} + \frac{0^2}{24{\cdot}233\,4}\right) \text{Error mean-square}} = 25{\cdot}3,$$

and the largest is

$$t\sqrt{\left(\frac{1}{4} + \frac{1}{4} + \frac{1{\cdot}82^2}{24{\cdot}233\,4}\right) \text{Error mean-square}} = 28{\cdot}5.$$

With this information, namely, that all the significant differences lie somewhere between 25·3 and 28·5, it is possible to put in the brackets without further computation. If there are any doubtful cases, the significant difference can be worked out for that particular value of $D$ to clear up the obscurity.

This example well illustrates the two-fold advantage of adjusting by the analysis of covariance in cases where there is a good degree of correlation between the two variates:

(1) It has led to a substantial reduction of error variation. The error mean square for $y$ before adjustment was 1 562·14, but after adjustment it was 277·52, so the precision of the

trial had been increased to more than five times its former value (1 562·14/277·52 = 5·63), the loss of one degree of freedom from the error being negligible in comparison.

(2) The adjustments to the treatment means have made clear the relative performances of the treatments despite an unlucky randomization that had assigned a disproportionate number of the good trees to the control.

Nevertheless, the analysis of covariance is not a panacea for all ills, its efficacy depending upon a suitable choice of independent variate.

## Double Covariance

It may well happen that an analysis of variance is further complicated by there being two independent variates. Thus, with strawberries it is not unusual to adjust crop after applying treatments, by both height and spread at the time of application. Again, it may be necessary to adjust both by a calibrating variate and a pseudo-variate.

Although a double covariance adjustment is more laborious than a single one, it is not more difficult in conception. The method of computing results is very similar and the successive stages the same. In the following explanation the dependent variate will be called $y$ and the two independent variates, $w$ and $x$. Apart from there being three data per plot instead of two, leading to three sets of sums of squares and three of sums of products, everything is the same until Stage 4, the construction of the analysis of variance of $y$ adjusted by $w$ and $x$. With a double covariance $S$ equals

$$C_{yy} - \frac{C_{ww}C_{xy}^2 + C_{xx}C_{wy}^2 - 2C_{wx}C_{wy}C_{xy}}{C_{ww}C_{xx} - C_{wx}^2}$$

and $S'$ is derived similarly from the values of $C_{yy}'$, $C_{wx}'$ and so on. Each has *two* degrees of freedom fewer than, respectively, $C_{yy}$ and $C_{yy}'$.

At Stage 5 it is necessary to know the regression coefficients of $y$ on $w$ and $x$. The "partial regression coefficient of $y$ on $w$", $x$ remaining unaltered despite its association with $w$, is

$$\frac{C_{xx}C_{wy} - C_{wx}C_{xy}}{C_{ww}C_{xx} - C_{wx}^2} = b_w.$$

Equally, the "partial regression of $y$ on $x$", $w$ remaining unaltered, is

$$\frac{C_{ww}C_{xy} - C_{wx}C_{wy}}{C_{ww}C_{xx} - C_{wx}^2} = b_x.$$

With these quantities known it is possible to adjust on to the standard values of $w$ and $x$. Thus, if $w$ needs to be adjusted by $d_w$ and $x$ by $d_x$ (care is needed here with signs), the adjusted value of $y$ is

$$(\text{unadjusted mean of } y) - b_w d_w - b_x d_x,$$

the adjustment having taken place in two steps.

At the end, in Stage 6, a further modification is needed, this time to work out the significant difference between two treatment means after adjustment, the one being based on $p$ plots and the other on $q$. If the treatments differ by $D_w$ in respect of $w$ and by $D_x$ in respect of $x$, the expression is [180]

$$t\sqrt{\left(\frac{1}{p} + \frac{1}{q} + \frac{D_w^2C_{xx} - 2D_wD_xC_{wx} + D_x^2C_{ww}}{C_{ww}C_{xx} - C_{wx}^2}\right)\text{ Error mean-square},}$$

where the error mean-square is that in the analysis of variance of $y$ adjusted by $w$ and $x$.

It will have been seen from this that the computation of an analysis of variance with a double covariance adjustment calls for a lot of work, but is quite practicable if the effort is necessary.

# BIBLIOGRAPHY AND AUTHOR INDEX

The italic figures at the end of each reference indicate those Sections in which the reference is discussed.


[1] ALLAN, F. E., and WISHART, J. A method of estimating the yield of a missing plot in field experimental work. *J. agric. Sci.*, 1930, **20** : 399–406. *83.*

[2] AMOS, J., and HOBLYN, T. N. Manurial trials with hops. *A. R. East Malling Res. Stat. for 1926–27*, 1928, *II Suppl.*, pp. 165–71. *82.*

[3] ANSCOMBE, F. J. The transformation of Poisson, binomial and negative binomial data. *Biometrika*, 1948, **35** : 246–54. *73* (three times).

[4] ASPIN, A. A. An examination and further development of a formula occurring in the problem of comparing two mean values. *Biometrika*, 1948, **35** : 88–96. *72.*

[5] ASPIN, A. A. Tables for use in comparisons whose accuracy involves two variances, separately estimated. *Biometrika*, 1949, **36** : 290–3. *72.*

[6] BAKER, R. E., and BAKER, G. A. Experimental design for studying resistance of strawberry varieties to verticilium wilt. *Phytopathology*, 1950, **40** : 477–82. *51.*

[7] BARTLETT, M. S. The square root transformation in analysis of variance. *J. R. statist. Soc.*, 1936, *Suppl.* **3** : 68–78. *73* (twice).

[8] BARTLETT, M. S. Some examples of statistical methods of research in agriculture and applied biology. *J. R. statist. Soc.*, 1937, *Suppl.* **4** : 137–83.

[8*a*] *Ibid.*, p. 151. *83.*

[8*b*] *Ibid.*, Footnote p. 168. *73.*

[9] BARTLETT, M. S. The approximate recovery of information from replicated field experiments with large blocks. *J. agric. Sci.*, 1938, **28** : 418–27. *48* (twice).

[10] BARTLETT, M. S. Multivariate analysis. *J. R. statist. Soc.*, 1947, *Suppl.* **9** : 176–97. *74.*

[11] BARTLETT, M. S. The use of transformations. *Biometrics*, 1947, **3** : 39–52. *73.*

[12] BATCHELOR, L. D., and REED, H. S. Relation of the variability of yields of fruit trees to the accuracy of field trials. *J. agric. Res.*, 1918, **12** : 245–85. *Table II* (four times).

[13] BATEN, W. D. Formulas for finding estimates for two and three missing plots in randomized block layouts. *Tech. Bull. Mich. agric. Exp. Stat.*, **165**, 1939. *83.*

[14] BATEN, W. D. Variances of differences between means when there are two missing values in randomized block designs. *Biometrics*, 1952, **8** : 42–50. *84.*

[15] BEAKBANE, A. B. Anatomical structure in relation to rootstock behaviour. In *Proc. 13th int. hort. Congr.* London, 1952, to be published by Royal Horticultural Society Lond., in 1953. *63.*

[16] BLAKE, M. A., and HERVEY, G. W. A standard for estimating the twig growth of one-year-old peach trees. *Bull. N. J. agric. Exp. Stat.*, **475**, 1928. *95.*

[17] BLISS, C. I. The analysis of field experimental data expressed in percentages. [In Russian with English summary.] *Pl. Prot., Leningr.*, 1937, No. 12, pp. 67–77. *73.*

[18] BLISS, C. I. The transformation of percentages for use in the analysis of variance. *Ohio J. Sci.*, 1938, **38** : 9–12. *73.*

[19] BOSE, S. S. Appendix. The estimation of mixed-up yields and their standard errors. *Sankhyā*, 1938, **4** : 112–20. *85.*

[20] BOSE, S. S., and MAHALANOBIS, P. C. On estimating individual yields in the case of mixed-up yields of two or more plots in field experiment. *Sankhyā*, 1938, **4** : 103–11. *85.*

[21] BRADFORD, F. C. Second-year changes in apparent vigor of apple varieties of prospective value as trunk-formers. *Proc. Amer. Soc. hort. Sci.*, 1944, **44** : 215–22. *15.*

[22] BRASE, M. A., and TUKEY, H. B. The relation between size of apple seedling rootstocks and size of orchard tree. *Proc. Amer. Soc. hort. Sci for 1936*, **34** : 298–304. *15.*

[23] CHANDLER, W. H. The trend of research in pomology. *Proc. Amer. Soc. hort. Sci. for 1921*, **18** : 233–40. *40.*

[24] CHEESMAN, E. E., and POUND, F. J. Uniformity trials with cacao. *Trop. Agriculture Trin.*, 1932, **9** : 277–88. *Table I, Table II.*

[25] CHINLOY, T., and INNES, R. R. Some designs used in sugar-cane experimentation in Jamaica, their analyses and interpretation of results. *Proc. Mtg. B.W.I. Sugar Tech., 1950*, pp. 58–72. *14.*

[26] CHRISTENSEN, J. R. Determinación de parcelas experimentales para viñas. *Experimenta Mendoza*, 1948, **1** : 20–5. *Table II.*

[27] CHRISTIDIS, B. G. The importance of the shape of plots in field experimentation. *J. agric. Sci.*, 1931, **21** : 14–37. *51.*

[28] CHRISTIDIS, B. G. Variability of plots of various shapes as affected by plot orientation. *Emp. J. exp. Agric.*, 1939, **7** : 330–42. *51* (twice).

[29] CLARK, A., and LEONARD, W. H. The analysis of variance with special reference to data expressed as percentages. *J. Amer. Soc. Agron.*, 1939, **31** : 55–66. *73.*

[30] COCHRAN, W. G. Problems arising in the analysis of a series of similar experiments. *J. R. statist. Soc.*, 1937, *Suppl.* **4** : 102–18. *75.*

[31] COCHRAN, W. G. Some difficulties in the statistical analysis of replicated experiments. *Emp. J. exp. Agric.*, 1938, **6** : 157–75. *73* (four times).

[32] COCHRAN, W. G. The distribution of the largest of a set of estimated variances as a fraction of their total. *Ann. Eugen., Lond.*, 1941, **11** : 47–52. *71.*

[33] COCHRAN, W. G. Testing a linear relation among variances. *Biometrics*, 1951, **7** : 17–32. *71.*

[34] COCHRAN, W. G. AUTREY, K. M., and CANNON, C. Y. A double change-over design for dairy cattle feeding experiments. *J. Dairy Sci.*, 1941, **24** : 937–51. *34.*

[35] COCHRAN, W. G., and COX, G. M. *Experimental designs.* John Wiley & Sons, Inc., New York, 1950. *83.*

[35*a*] *Ibid.*, Section 2.2. *55* (twice).

[35*b*] *Ibid.*, Section 7.31 and Table 7.12. *26.*

[35*c*] *Ibid.*, Sections 13.31–33. *32* (twice).

[36] COLLISON, R. C., and HARLAN, J. D. Variability and size relations in apple trees. *Tech. Bull. N.Y. St. agric. Exp. Stat.*, **164**, 1930, pp. 38. *Table I.*

[37] CORNER, E. J. H. In *Ann. Rep. Dir. Gdns. Straits Settlements* for 1937, cited by *Kew Bull.*, 1938, p. 306. *97.*

[38] CORNISH, E. A. The estimation of missing values in incomplete block experiments. *Ann. Eugen., Lond.*, 1940, **10** : 112–18. *83.*

[39] CORNISH, E. A. The estimation of missing values in quasi-factorial designs. *Ann. Eugen., Lond.*, 1940, **10** : 137–43. *83.*

[40] CORNISH, E. A. The analysis of quasi-factorial designs with incomplete data. *J. Aust. Inst. agric. Sci.*, 1940, **6** : 31–9. *83.*

[41] CORNISH, E. A. The analysis of quasi-factorial designs with incomplete data. 2. Lattice squares. *J. Aust. Inst. agric. Sci.*, 1941. **7** : 19–26. *83.*

[42] COVAS, G., and CHRISTENSEN, J. R. Determinación del tamaño de parcelas para ensayos comparativos de rendimientos en la vid. *Rev. argent. Agron.*, 1945, **12** : 26–9. *Table II.*

[43] CRIST, J. W. Intraclass correlations for horticultural research. *Proc. Amer. Soc. hort. Sci. for 1938*, 1939, **36** : 347–50. *App. IV.*

[44] CRIST, J. W. Biserial *r* for horticultural research. *Proc. Amer. Soc. hort. Sci. for 1939*, 1940, **37** : 269–71. *App. IV.*

[45] CRIST, J. W. Correlation from ranks, for horticultural research. *Proc. Amer. Soc. hort. Sci.*, 1941, **38** : 593–5. *App. IV.*

[46] Crist, J. W. Tetrachoric correlation for horticultural research. *Proc. Amer. Soc. hort. Sci.*, 1942, **40** : 549–51. *App. IV.*

[47] Crist, J. W. The coefficient of contingency for horticultural research. *Proc. Amer. Soc. hort. Sci.*, 1943, **42** : 484–6. *App. IV.*

[48] Crowther, E. M. Fertilizer experiments in colonial agriculture. Memoranda on colonial fertilizer experiments. II. H.M.S.O., *Colonial* **214**, 1947. *60.*

[49] Desaymard, P. Application des méthodes statistiques de R. A. Fisher aux expériences culturales. *Ann. agron.*, 1939, **9** : 626–57. *51.*

[50] Dorsey, M. J., and Hough, L. F. Relation between seedling vigor and tree vigor in apple hybrids. *Proc. Amer. Soc. hort. Sci.*, 1943, **43** : 106–14. *15.*

[51] Eden, T. Studies in the yield of tea. I. The experimental errors of field experiments with tea. *J. agric. Sci.*, 1931, **21** : 547–73. *Table II.*

[52] Eden, T. Studies in the yield of tea. Pt. III. Field experiments with potash and nitrogen in relation to the pruning cycle. *Emp. J. exp. Agric.*, 1935, **3** : 105–18. *Table I.*

[53] Eden, T. Correspondence with Commonwealth Bureau of Horticulture and Plantation Crops. 1952. *35.*

[54] Edgar, J. L. Strawberry cultivation studies. II. Variability in individual plant size and cropping, with special reference to area and shape of plots for field experiments. *J. Pomol.*, 1938, **16** : 91–100. *51* ; *Table II.*

[55] Euler, L. Recherches sur une nouvelle espèce de quarrés magiques. *Verhandelingen uitgegeven door het Zeeuwsch Genootschap der Wetenschappen te Vlissingen*, 1782, **9** : 85–239. *22.*

[56] Finney, D. J. The joint distribution of variance ratios based on a common error mean square. *Ann. Eugen., Lond.*, 1941, **11** : 136–40. *71.*

[57] Finney, D. J. Some orthogonal properties of the 4 × 4 and 6 × 6 Latin squares. *Ann. Eugen., Lond.*, 1945, **12** : 213–19. *23* (twice).

[58] Finney, D. J. Orthogonal partitions of the 5 × 5 Latin squares. *Ann. Eugen., Lond.*, 1946, **13** : 1–3. *23.*

[59] Finney, D. J. Orthogonal partitions of the 6 × 6 Latin squares. *Ann. Eugen., Lond.*, 1946, **13** : 184–96. *23.*

[60] Finney, D. J. Latin squares of the sixth order. *Experimentia*, 1946, **2** : (seen only in reprint form, page numbers unknown). *23.*

[61] Finney, D. J. The construction of confounding arrangements. *Emp. J. exp. Agric.*, 1947, **15** : 107–12. *35.*

[62] Finney, D. J. Main effects and interactions. *J. Amer. statist. Ass.*, 1948, **43** : 566–71. *27.*

[63] Fisher, R. A. *Statistical methods for research workers.* Oliver & Boyd, Edinburgh, 5th edition, 1934, to 10th edition, 1946.

[63*a*] *Ibid.*, Section 48. *20, 21.*

[63*b*] *Ibid.*, Section 49. *22.*

[63*c*] *Ibid.*, Section 49·1. *40.*

[64] Fisher, R. A. *The design of experiments.* Oliver & Boyd, Edinburgh, 1935.

[64*a*] *Ibid.*, Sections 5–8. *12.*

[64*b*] *Ibid.*, Section 31. *23.*

[65] Fisher, R. A., and Yates, F. The 6 × 6 Latin squares. *Proc. Camb. phil. Soc.*, 1934, **30** : 492–507. *23* (twice).

[66] Fisher, R. A., and Yates, F. *Statistical tables for biological, agricultural and medical research.* Oliver & Boyd, Edinburgh, 1938.

[66*a*] *Ibid.*, Table III. *54, 55, 72, 84* ; *App. I.*

[66*b*] *Ibid.*, Table V. *App. I.*

[66*c*] *Ibid.*, Table VI. *App. IV.*

[66*d*] *Ibid.*, Tables XII and XIII and relevant part of Introduction. *73.*

[66*e*] *Ibid.*, Table XV and relevant part of Introduction. *23* (twice).

[66*f*] *Ibid.*, Table XVI and relevant part of Introduction. *23, 34.*
[66*g*] *Ibid.*, Tables XVII–XIX and relevant part of Introduction. *31.*
[67] FREEMAN, M. F., and TUKEY, J. W. Transformations related to the angular and the square-root. *Princeton Univ., Statist. Res. Group, Memorandum Rep.*, **24**. 1949. *73* (twice).
[68] FRITH, H. J. A factorial field experiment with citrus. Field 466, No. 1. Introduction. *Internal Rep. Irrigation Res. Stat., Griffith, N.S.W.*, **8**, 1949. *37.*
[69] FRITH, H. J. A factorial field trial with citrus, No. 2. Results 1942–48. *Internal Rep. Irrigation Res. Stat., Griffith, N.S.W.*, **9**, 1949. *37.*
[70] FRITH, H. J. A factorial field trial with citrus, No. 3. Results, 1949. *Internal Rep. Irrigation Res. Stat., Griffith, N.S.W.*, **10**, 1950. *37.*
[71] GARNER, R. J., and WALKER, W. F. The frameworking of fruit trees. *Occ. Pap. Bur. Hort. East Malling*, **5** : 1938. *63.*
[72] GILBERT, S. M. Planning field experiments on *Coffea arabica*. *Trop. Agriculture, Trin.*, 1938, **15** : 16–18. *Table II.*
[73] GILBERT, S. M. Plot size in field experiments with *Coffea arabica*. *Trop. Agriculture, Trin.*, 1938, **15** : 52–5. *Table II.*
[74] GRUNDY, P. M. A general technique for the analysis of experiments with incorrectly treated plots. *J. R. statist. Soc., Ser. B*, 1951, **13** : 272–83. *33, 85.*
[75] HATTON, R. G. The elimination of sources of error in field experiments. The standardization of fruit tree rootstocks. *A. R. E. Malling Res. Stat. for 1928–30*, 1931, *II Suppl.*, 13–21. *14.*
[76] HATTON, R. G., GRUBB, N. H., and KNIGHT, R. C. Black currant variety trials. Reliability of results. The variability in cropping of individual black currant bushes as a guide to the suitable size of experimental plots. *J. Pomol.*, 1925, **4** : 200–20. *Table II.*
[77] HEDRICK, U. P., and ANTHONY, R. D. Twenty years of fertilizers in an apple orchard. *Bull. N.Y. agric. Exp. Stat.*, **460**, 1919. *95.*
[78] HOBLYN, T. N. Field experiments in horticulture. *Tech. Commun. Bur. Fruit Prod. East Malling*, **2**, 1931. *10, 90.*
[79] HOBLYN, T. N. A study of the variation in keeping quality of apples in store: as illustrated by the behaviour of the variety McIntosh Red from an Ontario apple orchard. *J. R. statist. Soc.*, 1938, *Suppl.* **5**. 129–61. *52.*
[80] HOBLYN, T. N. The design of field trials with cocoa. *Rep. Proc. Cocoa Res. Conf., Lond.*, 1945, pp. 164–8. *63.*
[81] HOFFMAN, M. B. The use of performance records in laying out a raspberry fertilizer experiment. *Proc. Amer. Soc. hort. Sci. for 1929*, 1930, **26** : 203–7. *Table II.*
[82] HORSFALL, J. G., and RICH, S. Spirally arranged plots in a design for field assay of fungicides. *Bull. Conn. agric. Exp. Stat.*, **530**, 1949. *51.*
[83] JOACHIM, A. W. R. A uniformity trial with coconuts. *Trop. Agriculturist*, 1935, **85** : 198–207. *Table II.*
[84] JOLLY, A. L. Uniformity trials on estate cacao fields in Granada, B.W.I. *Trop. Agriculture, Trin.*, 1942, **19** : 167–74. *54*; *Table II.*
[85] KELLER, K. R. Uniformity trial on hops, *Humulus lupulus* L., for increasing the precision of field experiments. *Agron. J.*, 1949, **41** : 389–92. *Table II.*
[86] KEMMER, E. Ueber die Anordnung von Versuchsparzellen im Obstbau. Reprinted from *Dtsch. Obstb.*, 1942, Bd. **57**. *53.*
[87] KEULS, M. The use of the "Studentized range" in connection with an analysis of variance. *Euphytica*, 1952, **1** : 112–22. *72.*
[88] KNOWLES, W. H. C., and CAMERON, C. Field experiments with sugar cane. *Rep. Sugar Exp. Stat., Dep. Agric. Brit. Guiana for 1948*, 1949. *37.*
[89] LOTT, W. L., SATCHELL, D. F., and HALL, N. S. A tracer-element technique in the study of root extension. *Proc. Amer. Soc. hort. Sci.*, 1950, **55** : 27–34. *60.*

[90] LUCAS, H. L. Bias in estimation of error in change-over trials with dairy cattle. *J. agric. Sci.*, 1951, **41** : 146–8. *34.*

[91] LUTZ, H. The effect of size of young pecan trees on their subsequent growth and yield. *Proc. Amer. Soc. hort. Sci. for 1938*, 1939, **36** : 335–8. *15, 64* ; *Table I.*

[92] MA, R. H., and HARRINGTON, J. B. A study of field experiments of semi-Latin square design. *Sci. Agric.*, 1949, **29** : 241–51. *38.*

[93] MAGISTAD, O. C., and FARDEN, C. A. Experimental error in field experiments with pineapples. *J. Amer. Soc. Agron.*, 1934, **26** : 631–44. *Table II.*

[94] MASSIBOT, J. A. *La technique des essais culturaux et des études d'écologie agricole.* Editions Georges Frère, Tourcoing, 1947. *51.*

[95] MCHATTON, T. H. The comparison of plot size in a peach experiment. *Proc. Amer. Soc. hort. Sci.*, 1947, **49** : 18–20. *Table II.*

[96] MURRAY, R. K. S. The value of a uniformity trial in field experimentation with rubber. *J. agric. Sci.*, 1934, **24** : 177–84. *Table I.*

[97] NAIR, K. R. The application of the technique of analysis of covariance to field experiments with several missing or mixed-up plots. *Sankhyā*, 1940, **4** : 581–8. *83, 85.*

[98] NAIR, K. R. The Studentized form of the extreme mean square test in the analysis of variance. *Biometrika*, 1948, **35** : 16–31. *71.*

[99] NAIR, K. R. The distribution of the extreme deviate from the sample mean and its Studentized form. *Biometrika*, 1948, **35** : 118–44. *72* (twice).

[100] NEWMAN, D. The distribution of range in samples from a normal population expressed in terms of an independent estimate of standard deviation. *Biometrika*, 1939, **31** : 20–30. *72.*

[101] NORTON, H. W. The 7 × 7 squares. *Ann. Eugen., Lond.*, 1939, **9** : 269–307. *23.*

[102] OLLAGNIER, M. Forme, dimension des parcelles et nombre de répétitions dans les essais culturaux sur arachide et sur palmier à huile. *Oléagineux*, 1951, **6** : 707–10. *Table II.*

[103] PAPADAKIS, J. Méthode statistique pour des expériences sur champ. *Bull. Sci. Inst. d'Amélioration des Plantes à Salonique* (Grèce), **23** : 1937. *48.*

[104] PARKER, E. R. Adjustment of yields in an experiment with orange trees. *Proc. Amer. Soc. hort. Sci.*, 1942, **41** : 23–33. *48* ; *Table I.*

[105] PARKER, E. R., and BATCHELOR, L. D. Variation in the yields of fruit trees in relation to the planning of future experiments. *Hilgardia*, 1932, **7** : 81–161. *Table I.*

[106] PATERSON, D. D., and HANSCHELL, D. M. The comparison of four sugar-cane varieties in Trinidad. *Trop. Agriculture, Trin.*, 1938, **15** : 199–201. *51.*

[107] PATTERSON, H. D. The analysis of change-over trials. *J. agric. Sci.*, 1950, **40** : 375–9. *34.*

[108] PATTERSON, H. D. Change-over trials. *J. R. statist. Soc., Ser. B*, 1951, **13** : 256–71. *34.*

[109] PEARCE, S. C. The statistical interpretation of vigour measurements of fruit trees. *J. Pomol.*, 1943, **20** : 111–15. *73* (twice).

[110] PEARCE, S. C. Sampling methods for the measurement of fruit crops. *J. R. statist. Soc.*, 1945, **107** : 117–26. *96.*

[111] PEARCE, S. C. Lognormal distributions. *Nature*, 1945, **156** : 747. *73.*

[112] PEARCE, S. C. The measurement of fruit crops by sampling. *A. R. East Malling Res. Stat. for 1946*, 1947, pp. 77–82. *92, 96* (three times).

[113] PEARCE, S. C. Randomized blocks with interchanged and substituted plots. *J. R. statist. Soc., Ser. B*, 1948, **10** : 252–6. *33, 85.*

[114] PEARCE, S. C. The variability of apple trees. I. The extent of crop variation and its minimization by statistical means. *J. hort. Sci.*, 1949, **25** : 3–9. *15, 42, 51* ; *Table I.*

[115] PEARCE, S. C. The interpretation of uniformity trials. *A. R. East Malling Res. Stat. for 1949*, 1950, pp. 91–2. *Table II.*

[116] PEARCE, S. C. Studies in the measurement of apple trees. I. The use of trunk girth to estimate tree size. *A. R. East Malling Res. Stat. for 1951*, 1952, pp. 101–4. *91, 94.*

[117] PEARCE, S. C. The design of calibration trials with three varieties. *A. R. East Malling Res. Stat. for 1951*, 1952, pp. 105–7. *45* (three times).

[118] PEARCE S. C. Some new designs of Latin square type. *J. R. Statist. Soc., Ser. B*, 1952, **14** : 101–6. *32.*

[119] PEARCE, S. C., and HOBLYN, T. N. A review of experimental design at East Malling, 1919–1947. *A. R. East Malling Res. Stat. for 1947*, 1948, pp. 88–100. *10, 21* (twice), *44.*

[120] PEARCE, S. C., and TAYLOR, J. The changing of treatments in a long-term trial. *J. agric. Sci.*, 1948, **38** : 402–10. *32, 35, 60.*

[121] PEARCE, S. C., and TAYLOR, J. The purposes and design of calibration trials. *A. R. East Malling Res. Stat. for 1949*, 1950, pp. 83–90. *40, 43* (twice), *44* (twice), *45* (four times).

[122] PEARCE, S. C., and THOM, J. M. S. The variability of apple trees. II. The optimum size for unguarded plots. *J. hort. Sci.*, 1951, **26** : 98–108. *Table II.*

[123] PEARCE, S. C., and THOM, J. M. S. A study of plot-size with Nigerian estate cacao. *J. hort. Sci.*, 1951, **26** : 261–7. *Table I, Table II.*

[124] PEARCE, S. C., and THOM, J. M. S. Unpublished paper. *51 ; Table II.*

[125] PEREIRA, H. C. Correspondence with Commonwealth Bureau of Horticulture and Plantation Crops, 1952. *91.*

[126] PIERIS, W. V. D., and SALGADO, M. L. M. Experimental error in field experiments with coconuts. *Trop. Agriculturist*, 1937, **89** : 75–85. *Table II.*

[127] PRILLWITZ, P. M. H. H. Opzet en beoordeeling der resultaten van veldproeven bij de theecultuur. *Arch. Theecult. Ned.-Ind.*, 1929, **3** : 159–210. *Table II.*

[128] QUENOUILLE, M. H. The analysis of covariance and non-orthogonal data. *Biometrics*, 1948, **4** : 240–6. *81.*

[129] RAO, C. R. General methods of analysis for incomplete block designs. *J. Amer. Statist. Ass.*, 1947, **42** : 541–61. *38.*

[130] REED, H. S. Correlations between growth and fruit production of apricots. *Proc. Amer. Soc. hort. Sci. for 1928*, 1929, **25** : 247–9. *Table I.*

[131] RIGNEY, J. A. Some statistical problems confronting horticultural investigators. *Proc. Amer. Soc. hort. Sci.*, 1946, **48** : 351–7. *21, 37.*

[132] RIGNEY, J. A., MORROW, E. B., and LOTT, W. L. A method of controlling experimental error for perennial horticultural crops. *Proc. Amer. Soc. hort. Sci.*, 1949, **54** : 209–12. *42, 95.*

[133] ROACH, W. A. Plant injection for diagnostic and curative purposes. *Tech. Commun. Bur. Hort. East Malling*, **10**, 1938. *60.*

[134] ROACH, W. A., and ROBERTS, W. O. Further work on plant injection for diagnostic and curative purposes. *Tech. Commun. Bur. Hort. East Malling*, **16**, 1945. *60.*

[135] ROBINSON, H. F., and WATSON, G. S. An analysis of simple and triple rectangular lattice designs. *Tech. Bull. N.C. agric. Exp. Stat.*, **88**, 1950. *83.*

[136] ROESSLER, E. B. Valid estimates of variance in the analysis of pooled data. *Proc. Amer. Soc. hort. Sci.*, 1943, **42** : 481–3. *73.*

[137] ROESSLER, E. B., and LEACH, L. D. Analysis of combined data for identical replicated experiments. *Proc. Amer. Soc. hort. Sci.*, 1944, **44** : 323–8. *75.*

[138] ROGERS, W. S. Recording apparatus for horticultural experiments including automatic counting devices. *A. R. East Malling Res. Stat. for 1928, 1929 and 1930*, 1931, *II Suppl.*, pp. 65–73. *90.*

[139] ROGERS, W. S., and EDGAR, J. L. Strawberry cultivation studies. I. The performance of individual plants of clonal families. *J. Pomol.*, 1938, **16** : 63–90. *15.*

[140] ROGERS, W. S., and VYVYAN, M. C. The root sytems of some ten-year-old apples trees on two different root stocks and their relation to tree performance. *A. R. East Malling Res. Stat. for 1926–1927*, 1928, *II Suppl.*, pp. 31–43. *60.*

[141] ROGERS, W. S., and VYVYAN, M. C. Root studies. V. Rootstock and soil effect on apple root systems. *J. Pomol.*, 1934, **12** : 110–50. *60.*

[142] SALGADO, M. L. M. Recent studies on the manuring of coconuts in Ceylon. *Trop. Agriculturist*, 1946, **102** : 149–54. *60.*

[143] SATTERTHWAITE, F. E. An approximate distribution of estimates of variance components. *Biomet. Bull.*, 1946, **2** : 110–14. *72.*

[144] SHAH, S. M. I. Influence of the number of trees per plot on the precision of peach yield trials. *Punjab Fruit J.*, 1950, **14** (47) : 17–22. *54 ; Table II.*

[145] SHARPE, R. H., and BLACKMON, G. H. A study of plot size and experimental design with pecan yield data. *Proc. Amer. Soc. hort. Sci.*, 1950, **56** : 236–41. *82 ; Table I, Table II.*

[146] SHARPE, R. H., and WINSOR, H. W. Cross-feeding and boron placement studies with pecan. *Proc. Amer. Soc. hort. Sci.*, 1951, **57** : 202–6. *60.*

[147] SHAW, J. K. Trunk diameters of young apple trees on clonal stocks. *Proc. Amer. Soc. hort Sci.*, 1942, **40** : 269–71. *15.*

[148] SILVEY, R. J. E. Methods of listener research employed by the British Broadcasting Corporation. *J. R. statist. Soc.*, 1944, **107** : 190–230. *93.*

[149] SMITH, H. FAIRFIELD. An empirical law describing heterogeneity in the yields of agricultural crops. *J. agric. Sci.*, 1938, **28** : 1–23. *51.*

[150] SNEDECOR, G. W. Answer to Query 90. *Biometrics*, 1951, **7** : 299. *72.*

[151] SNEDECOR, G. W., and HABER, E. S. Statistical methods for an incomplete experiment on a perennial crop. *Biomet. Bull.*, 1946, **2** : 61–7. *74.*

[152] STEVENS, W. L. Análise estatistíca do ensaio de variedades de café. *Bragantia*, 1949, **9** : 103–23. *74.*

[153] STRICKLAND, A. G. Error in horticultural experiments. *J. Dep. Agric. Vict.*, 1935, **33** : 408–16. *52 ; Table II* (twice).

[154] STRICKLAND, A. G., FORSTER, H. C., and VASEY, A. J. A vine uniformity trial. *J. Dep. Agric. Vict.*, 1932, **30** : 584–93. *Table II.*

[155] SUDDS, R. H., and ANTHONY, R. D. The correlation of trunk measurements with tree performance in apples. *Proc. Amer. Soc. hort. Sci. for 1928*, 1929, **25** : 244–6. *Table I.*

[156] TAYLOR, J. Errors of treatment comparisons when observations are missing. *Nature*, 1948, **162** : 262. *84.*

[157] TAYLOR, J. A valid restriction of randomization for certain field experiments. *J. agric. Sci.*, 1949, **39** : 303–8. *36.*

[158] TAYLOR, J. The comparison of pairs of treatments in split-plot experiments. *Biometrika*, 1950, **37** : 443–4. *72.*

[159] TAYLOR, J. The estimation of fruit size of cherries by sampling methods. *A. R. East Malling Res. Stat. for 1950*, 1951, pp. 93–9. *96.*

[160] TAYLOR, J. Statistical studies on strawberry crop and vigour measurements. *A. R. East Malling Res. Stat. for 1950*, 1951, pp. 100–7. *42, 82 ; Table I, Table II.*

[161] THOMPSON, J. I. Correspondence with Commonwealth Bureau of Horticulture and Plantation Crops, 1952. *91.*

[162] TIDBURY, G. E. Stem girth as a criterion in field trials with young clove trees. *Emp. J. exp. Agric.*, 1943, **11** : 33–7. *42 ; Table I.*

[163] TIPPETT, L. H. C. *The methods of statistics.* Williams & Norgate Ltd., London. 1931. (*Section 3.5*) *72.*

[164] TUKEY, J. W. Comparing individual treatment means in the analysis of variance. *Biometrics*, 1949, **5** : 99–114. *72.*

[165] TURNER, P. E., WARNEFORD, F. H. S., and CHARTER, C. F. Recent investigations on sugar-cane and sugar-cane soils in Antigua. II. Manurial experiments reaped in 1936. *Trop. Agriculture, Trin.*, 1937, **14** : 150–5, 179–88. *51.*

[166] TURRELL, F. M. Estimating heights of citrus trees. *Proc. Amer. Soc. hort. Sci.*, 1946, **48** : 147–50. *97.*

[167] WADLEIGH, C. H., and THARP, W. H. Factorial design in plant nutrition experiments in the greenhouse. *Bull. Ark. agric. Exp. Stat.*, **401** : 1940. *73.*

[168] WEBBER, H. J. Variation in citrus seedlings and their relation to rootstock selection. *Hilgardia*, 1932, **7** : 1–79. *15.*

[169] WEBSTER, C. C. A note on a uniformity trial with oil palms. *Trop. Agriculture, Trin.*, 1939, **16** : 15–19. *Table I, Table II.*
[170] WELCH, B. L. The generalization of "Student's" problem when several different population variances are involved. *Biometrika*, 1947, **34** : 28–35. *72.*
[171] WELCH, B. L. Further note on Mrs. Aspin's tables and on certain approximations to the tabled function. *Biometrika*, 1949, **36** : 293–6. *72.*
[172] WELLMAN, R. H., THURSTON, H. W., Jr., and WHALEY, F. H. A method for correcting the geographic variation in field experiments. *Contr. Boyce Thompson Inst.*, 1948, **15** : 153–64. *48.*
[173] WHITE, D. G. The size of apples in relation to their location on the tree. *Proc. Amer. Soc. hort. Sci. for 1937*, 1938, **35** : 132–4. *96.*
[174] WILCOX, J. C. Field studies of apple tree growth and fruiting. I. Sampling and measuring terminal shoots. *Sci. Agric.*, 1937, **17** : 563–72. *95.*
[175] WILCOX, J. C. Adjusting apple yields for differences in size of tree. *Sci. Agric.*, 1940–41, **21** : 139–48. *42 ; Table I.*
[176] WILCOX, J. C. Some factors affecting apple yields in the Okanagan Valley. I. Tree size, tree vigour, biennial bearing, and distance of planting. *Sci. Agric.*, 1944, **25** : 189–213. *13.*
[177] WILLIAMS, C. B. The use of logarithms in the interpretation of certain entomological problems. *Ann. appl. Biol.*, 1937, **24** : 404–14. *73* (twice).
[178] WILLIAMS, E. J. Experimental designs balanced for pairs of residual effects. *Aust. J. sci. Res. Ser. A*, 1950, **3** : 351–63. *34.*
[179] WILLIAMS, E. J. The interpretation of interactions in factorial experiments. *Biometrika*, 1952, **39** : 65–81. *27.*
[180] WISHART, J. Tests of significance in analysis of covariance. *J. R. statist. Soc.*, 1936, *Suppl.* **3** : 79–82. *App. IV* (twice).
[181] WISHART, J. Field experiments of factorial design. *J. agric. Sci.*, 1938, **28** : 299–306. *71.*
[182] YATES, F. The analysis of replicated experiments when the field results are incomplete. *Emp. J. exp. Agric.*, 1933, **1** : 129–42. *83, 84* (three times).
[183] YATES, F. Complex experiments. *J. R. statist. Soc.*, 1935, *Suppl.* **2** : 181–247. *38.*
[184] YATES, F. Incomplete Latin squares. *J. agric. Sci.*, 1936, **26** : 301–15. *22, 32* (twice).
[185] YATES, F. A new method of arranging variety trials involving a large number of varieties. *J. agric. Sci.*, 1936, **26** : 424–55. *63.*
[186] YATES, F. Incomplete randomized blocks. *Ann. Eugen., Lond.*, 1936, **7** : 121–40. *31.*
[187] YATES, F. The design and analysis of factorial experiments. *Tech. Commun. Bur. Soil Sci. Rothamsted*, **35**, 1937.
[187*a*] *Ibid.*, pp. 18–35, 42–8, 57–67. *35.*
[187*b*] *Ibid.*, pp. 21–3. *35.*
[187*c*] *Ibid.*, pp. 72–7. *25, 35.*
[187*d*] *Ibid.*, pp. 85–90. *63.*
[188] YATES, F. The recovery of inter-block information in balanced incomplete block designs. *Ann. Eugen., Lond.*, 1940, **10** : 317–25. *31.*
[189] YATES, F. Modern experimental design and its function in plant selection. *Emp. J. exp. Agric.*, 1940, **8** : 223–30. *63.*
[190] YATES, F. Lattice squares. *J. agric. Sci.*, 1940, **30** : 672–87. *63.*
[191] YATES, F. The place of statistics in agricultural research. *Agric. Progr.*, 1946, **21**. (Seen only in reprint form, page numbers unknown.) *10.*
[192] YATES, F., and HALE, R. W. The analysis of Latin squares when two or more rows, columns or treatments are missing. *J. R. statist. Soc.*, 1939, *Suppl.* **6** : 67–79. *32.*
[193] YEAGER, A. F., and LATIMER, L. P. Tree girth and yield as indicators of subsequent apple yield productivity. *Proc. Amer. Soc. hort. Sci. for 1939*, 1940, **37** : 101–5. *Table I.*
[194] YOUDEN, W. J. Use of incomplete block replications in estimating tobacco-mosaic virus. *Contr. Boyce Thompson Inst.*, 1937, **9** : 41–8. *32.*

# SUBJECT INDEX

Primary references are shown in heavy and secondary in ordinary type. References in brackets indicate passages having a bearing on the subject indexed without dealing with it directly.

PRINTED BY
ADLARD AND SON, LIMITED,
BARTHOLOMEW PRESS,
DORKING, SURREY.

# *A List of Publications of the Commonwealth Bureau of Horticulture and Plantation Crops*

"Horticultural Abstracts." A quarterly abstract journal of current literature on progress in the cultivation of horticultural and plantation crops. Issued quarterly since 1931. Annual subscription rates: Direct subscribers (other than trade) in British Commonwealth, 40*s.*, $U.S. or Canadian 5.60. Others 50*s.*, $U.S. or Canadian 7.00. Single copies 15*s.* or $2.10.

| | |
|---|---|
| Author and Subject Index to Volumes I–X (1931–40) | 25*s.* 0*d.* or $3.50 |
| Author and Subject Index to Volumes XI–XV (1941–45) | 35*s.* 0*d.* or $4.90 |
| Author and Subject Index to Volumes XVI–XX (1946–50) | 50*s.* 0*d.* or $7.00 |

## *Technical Communications*

**9.** A REVIEW OF THE LITERATURE ON STOCK-SCION INCOMPATIBILITY IN FRUIT TREES, with particular reference to pome and stone fruits. G. K. Argles. (1937) — 5*s.* 0*d.*

**14.** PROPAGATION BY CUTTINGS AND LAYERS. Recent work and its application, with special reference to pome and stone fruits. R. J. Garner. (1944) — 3*s.* 6*d.*

**15.** SPRING FROST DAMAGE IN ORCHARDS AND ITS POSSIBLE PREVENTION. (1945) — 1*s.* 6*d.*

**16.** FURTHER WORK ON PLANT INJECTION FOR DIAGNOSTIC AND CURATIVE PURPOSES. W. A. Roach and W. O. Roberts. (Reprint from *Journal of Pomology*). (1945) — 1*s.* 6*d.*

**17.** CHEMICAL COMPOSITION OF PLANTS AS AN INDEX OF THEIR NUTRITIONAL STATUS. D. W. Goodall and F. G. Gregory. (1947) — 9*s.* 0*d.*

**18.** FRUIT FALL AND ITS CONTROL BY SYNTHETIC GROWTH SUBSTANCES. M. C. Vyvyan. (1946) — 3*s.* 6*d.*

**19.** SEED PRODUCTION OF EUROPEAN VEGETABLES IN THE TROPICS. A. G. G. Hill. (1948) — 2*s.* 0*d.*

**20.** GROWTH SUBSTANCES AND THEIR PRACTICAL IMPORTANCE IN HORTICULTURE. H. L. Pearse. (1948) — 12*s.* 6*d.*

**21.** RECENT ADVANCES IN FRUIT JUICE PRODUCTION. V. L. S. Charley and others. (1950) — 15*s.* 0*d.*

**22.** SAND AND WATER CULTURE METHODS USED IN THE STUDY OF PLANT NUTRITION. E. J. Hewitt. (1952) — 42*s.* 0*d.* or $6.25

## *Occasional Papers*

**2.** EXPERIMENTAL DATA ON ORCHARD AND SMALL FRUIT MANURING. S. T. Antoshin. (1933) — 1*s.* 0*d.*

**6.** HARICOT BEANS. G. St. Clair Feilden. (1941) — 1*s.* 0*d.*