204

RECENT PROGRESS IN EXPERIMENTAL DESIGN

BY

R. A. FISHER.

1. GENERAL PRINCIPLES.

In surveying the progress of recent years in the application of statistical experience to the planning and design of experiments, I need not go further back than my paper of 1926 " The Arrangement of Field Experiments ", in which I set out the principles which had at that time been recently developed at Rothamsted, and which have since become much more widely understood and applied to experiments both in the field and in the Laboratory.

For many years the practice of replication, that is the repetition of an experimental comparison on different areas of land, or, in general, upon different samples of experimental material, had been gaining ground among critical experimenters. The purposes of replication had, however, been frequently misunderstood, through lack of recognition that two distinct functions were fulfilled: (I) the increase in the precision of the comparisons made, and (II) the estimation of this precision, by means of which alone the experimenter can judge what conclusions it is safe to draw. To be definite, these conclusions must be based on tests of significance, and, since the most widely useful of these tests involve estimates of error, it is essential that the estimates provided by the experiments shall be valid, that is, that the components used in the estimation of error shall be produced by exactly the same system of causes as produces the actual errors of the experiment.

From *L'application du calcul des probabilités*, (*Proceedings of the International Institute of Intellectual Cooperation*, Geneva, 1939), 19-31, (1945).

2. Randomisation.

To ensure the validity of the estimation of error, it is sufficient to supplement replication by a second principle, namely randomisation. We may state the principle briefly by saying that, while it is impossible to compare two varieties, or treatments, in the same year, on exactly the same land, yet they may be compared on random samples of the same land, in such a way that the errors produced by differences in the areas of land on which they are grown shall be correctly estimable from the discrepancies between areas similarly treated, or sown with the same variety.

The efficacy of this principle is easily seen in such a case as that referred to above, in which the whole experimental area is divided into plots, of which so many are assigned at random to each variety. It is, however, customary, in order to diminish the magnitude of the real errors, to eliminate in advance the effects of heterogeneity between certain large areas, such as the blocks, rows or columns of an experiment, by requiring that all treatments shall be applied equally to such areas. It may then be shown that randomisation, subject to such simple restrictions, still exactly fulfils its purpose of providing an estimate of error, after the elimination of the grosser elements, which is validly representative of the actual errors of the comparisons desired.

For example, if five varieties are to be tested, the experiment may consist of a number of compact blocks, each divided into five plots. In each block one plot is assigned at random to each variety. Then it is evident that the errors in the comparisons of the aggregate yields of the different varieties are due solely to such soil heterogeneity as exists between different plots of the same blocks, and will be wholly unaffected by differences between the fertilities of the blocks as wholes. The errors of the comparisons may, therefore, properly be estimated from the discrepancies between the different comparisons made in each block, eliminating entirely from the estimate of error heterogeneity between whole blocks. The precision of the experiment may be increased indefinitely by increasing the number of blocks, even though this requires that the experimental area as a whole shall be made more heterogeneous.

The principle of dividing the errors to which an experimental comparison is exposed into two separable fractions, of which one, representing generally the larger sources of error, is completely eliminated by the methods of experimentation adopted, while, by means of randomisation, errors of the second fraction are scrupulously divided between the comparisons which finally constitute our experimental results and those other comparisons from which our estimate of error is built up, is of very wide and varied application. The great progress of recent years has, indeed, largely consisted in the elaboration of sometimes very intricate schemes appropriate to different types of experimental material, and to the other requirements which the experiment may have to meet. Since, however the notion of randomisation, though simple in itself and completely general in its application, has been occasionally misunderstood, it may be as well first to answer certain objections that have been felt before discussing any of the more elaborate structures which depend upon it.

3. Sampling.

The nature of randomisation may be very well illustrated from its use in the sampling of an agricultural crop, *e.g.*, of a single experimental plot, or of an area given over to commercial production. The technique by which such sampling can be carried out, on a large or a small scale, efficiently and without danger of bias, is of very widespread statistical utility, by no means confined to agriculture. Indeed, during my recent visit to India the best source of official information upon many sociological facts of the utmost importance appeared to lie in the application to whole districts or provinces of methods which at Rothamsted we had developed to grapple only with the complexities of the crop growing in an agricultural field.

The principles of crop sampling may be readily grasped in terms of the hierarchy of subdivisions which it involves. First we may eliminate major elements of heterogeneity by dividing the whole area into sections, or strata, chosen for their apparent homogeneity, each of which is known as a sampling area. This area is conceived as composed of a number of sampling units, each of which may be only a small fraction of the sampling area. Not less than two of these units, chosen at random from the whole, constitutes the sample, and these are examined, measured or analysed independently. Finally, each of these sampling units is, in practice, generally composed of smaller portions known as *ultimate units*, which are the smallest elements which it is convenient to handle separately.

For example, with a cereal crop drilled in rows we may choose as our ultimate unit a length of 25 cms. of a single row. Four such lengths lying side by side constitute a convenient sampling unit on which the quantity of produce can be measured at harvest, or at other times of the year the dry weight, the number of ears, or the number of plants can be ascertained. With a crop consisting of larger plants, such as sugar-beet, the ultimate unit will be a single plant. The sampling unit will be a number of plants, but these need not grow in a compact group. On the contrary, a convenient form of sampling unit consists of every twentieth plant, commencing with a particular individual chosen by lot from the first twenty, and counting every plant of the sampling area in a predetermined order.

It will be noticed that randomisation is not applied within the sampling unit. Its parts are always related to one another in a systematic and predetermined manner. Corresponding with this fact, we may note that no separate determinations need be made, in the process of sampling, upon the parts of which the sampling unit is composed. It is sufficient to count the number of ears on the whole unit, or to determine the sugar content of the mixed pulp of all the plants which constitute it, rather than analysing these separately. Differences within the same sampling unit are not randomised because they are not used in the estimation of sampling error. They are only of interest for the different and preliminary process of discovering what type of sampling unit it is best to use, namely, that by which we can reduce the sampling error to a predetermined level with a minimum of trouble.

The estimate of sampling error is derived solely from the differences observed between different sampling units from the same sampling area. It is to guarantee the validity of this estimate that we choose our sampling units independently and at random from the area which they are to represent. It is for this reason that we never take less than two. If, for example, we had pairs of sampling units from each of forty sampling areas, the sum of the squares of the forty differences in any measurable character will provide a satisfactory estimate of the sampling variance of the total of the forty samples. In practical sampling we always require such an estimate, for only from it can we judge whether the sampling procedure has been sufficiently intensive or, in general, sufficiently accurate for its purpose. It should be noted that the large differences which may exist between different sampling areas contribute nothing to the error of sampling, for these areas are made to contribute proportionately to their extent. Like the differences within sampling units, though for a different reason, they are irrelevant to the estimation of sampling error.

The terminology of this simple procedure has been designed to make clear exactly in what randomisation consists, and in what particular way it aids our researches. One fallacy to be avoided consists in asserting that randomisation is inconsistent with the elimination of the grosser elements of soil heterogeneity, as it would be if we were forced to ignore the distinction between different sampling areas, and to choose by chance varying numbers of samples from them. Equally, it is a fallacy to suppose that it is inconsistent with the choice of a sampling unit of complex structure when this happens to be advantageous, as if we were forced to take the smallest available element as our sampling unit. The object of the complex sampling unit suggested above for sugar-beet is that it shall contain plants grown in a variety of places within the sampling area, the aggregate, or average, of which shall be closely similar to that of the other sampling units which might be chosen. Equally, in comparing a pair of cereal varieties, we may choose for closer comparison to grow them always on pairs of adjacent parallel strips, assigning at random only which shall occupy the first strip and which the second. In this case the experiment will provide a number of independent comparisons between different pairs, the errors of which have been properly randomised. Further, if it were feared that important differences might exist between adjacent parallel strips, by reason of a transverse fertility gradient, we may, if we choose, set up a balanced configuration, or " sandwich " of strips ABBA and confine randomisation to deciding whether for each configuration it shall be ABBA or BAAB. In this case, each complete configuration will provide a single comparison, properly randomised, so that, from a number of such comparisons, a valid estimate of experimental error may be determined.

I have stressed these preliminary and fundamental points, not because they are widely misunderstood by practical experimenters, who, indeed, usually appreciate them thoroughly, but because, in recent years, certain theoretical statisticians, having little contact with practical work, have thought to score an argumentative point by misrepresenting the function and nature of randomisation in experimental technique. The same authors, actuated by a desire to defend the early and inaccurate applications of statistics to experimental data, have undertaken an exposition of the tests of significance, essential to the interpretation of well-planned experiments, with an equal lack of understanding of their logical basis and, therefore, of the questions which they are competent to answer. Thus Neyman has developed an elaborate attack upon the experimental design known as the Latin Square, in ignorance of the fact that the test of significance which he declares to be biassed is a test of the null hypothesis that the varieties compared are *equivalent* in those quantitative characteristics to which the test is applied and is therefore perfectly adapted to its purpose.

4. The Latin Square.

The Latin Square, named after the combinatorial problem studied by Euler, is an arrangement in which the number of replications is equal to the number of treatments to be compared, but in which these replications are not grown separately but in a single compact block of plots having the same number of rows as of columns. The treatments are then so arranged that each treatment appears once in each row and once in each column, as may always be done whatever the number of treatments. To obtain a random arrangement it is not necessary to choose one at random out of all the possible solutions for a square of given size; it is sufficient, starting with any one solution, to rearrange at random first the rows, and then the columns, finally assigning at random the treatments to be used to the letters of the square. This procedure, in effect, chooses a solution at random from among one of the transformation sets into which the solutions may be divided.

The great practical utility and widespread use of the Latin

Square, both in direct experimentation, as described above, and as a basis for more complex designs, has stimulated the study of the transformation sets and of the problem of enumeration which, in the mathematical literature has been somewhat disconnected and desultory. For 6×6 squares there are in all twenty two transformation sets, as defined above, containing numbers varying from 20 to 1080 standard squares, *i.e.*, squares having the letters of the first row and column in a standard order. By permuting the rows and columns each standard square will generate $6!\,5!$ or 86,400 different arrangements. It is, therefore, sufficient and convenient to enumerate the standard squares of each set. Different transformation sets may, however, be connected in another manner, namely by permuting the different categories, rows, columns and letters with each other. The interchange of rows and columns produces what is called the conjugate square and, for squares larger than 5×5, conjugate squares need not belong to the same transformation set; consequently, we have such things as conjugate sets. It is equally possible to interchange the rows and letters, or the columns and letters, to produce what are called adjugate squares, and it is evidently possible for as many as six sets, all of which must contain the same number of standard squares, to be thus related by adjugacy. Actually, among the 22 sets of 6×6 squares there are 5 triplets of adjugate sets, in addition to 7 completely self-adjugate sets. Since the majority of properties of importance are invariant, both to the permutation (*e.g.* of different rows) within each category, and to permutations among the three categories, there may be said to be twelve species of 6×6 Latin Squares, containing from 20 to 3240 standard solutions each.

Among the properties unaffected by these two kinds of transformation is the possibility or impossibility of forming a Graeco-Latin solution, *i.e.*, of adding a fourth category with the same number of classes, such that each class occurs once in each row, once in each column and once with each letter. In none of the twelve species of 6×6 square is this possible, as was conjectured by Euler, and demonstrated by Tarry in 1900, who must also have enumerated the 6×6 squares, for he states their total correctly. He speaks, however, of their being seventeen kinds, evidently through recognising the equivalence of conjugate sets, but not of adjugate sets.

A second group of invariant properties, very serviceable for the recognition and identification of different sets of solutions, consists of the number and distribution of the intercalated 2×2 squares, which most Latin Squares contain. These must appear whenever the permutation partition connecting any pair of rows or columns contains a part 2, as do the partitions (5, 2) and $(3, 2^2)$ of the partible number 7. Consequently, it commonly occurs that such intercalated squares are numerous, and by mapping these it may often be seen at a glance that no rearrangement of the rows, columns or letters can reproduce the square as it stands, or that a few particular permutations may possibly do so. These can then be tried without much labour. Again, intrinsic differences between the categories are at once obvious, and when any two, or all three, of these are apparently equivalent, they may be tested directly. In exploring the 7×7 squares, no two sets have been found, apart from a few which contain no intercalates, the difference between which was not apparent as soon as intercalates were mapped.

In addition to affording the means of recognising a set already known, and of determining the size and structure of the set to which a given square belongs, the intercalates are of further value in that, by reversing them, leaving the other elements of the square unchanged, a square belonging to another set is usually produced. The whole body of sets, or at least all that are connected in this way, may thus be systematically generated by making all possible reversals, or, as is sufficient, all that the symmetry properties of the square in question show to be intrinsically different. H. W. Norton at the Galton Laboratory has recently enumerated the 7×7 squares by this method, finding in all 146 species. The enumeration is not theoretically exhaustive, but, since precautions have been taken beyond those which I can here mention, it is improbable that any solutions have escaped discovery. Norton has also explored the Graeco-Latin sets of size 7.

5. Confounding.

The Latin Square has been especially successful when the number of treatments or varieties to be compared is from 4 to 10, although squares as large as 12×12 have been used successfully. Much larger numbers may, however, be required. This occurs inevitably in plant selection work where, after crossing, the lines to be tested for favourable combinations may run into hundreds, but it also occurs deliberately, as was foreshadowed in 1926, when advantage is taken of the high efficiency and comprehensiveness of *factorial* experiments in which several different factors, manurial, cultural or varietal, are introduced in all possible combinations into the same experiment. The number of different combinations may then be considerable, *e.g.*, 48 or more, and it was in experiments of this type that the method of *confounding* was first introduced. This consists in the use of blocks of smaller size than would be required for a complete replication. Certain of the experimental contrasts within the replication are thus represented by the difference in yield between whole blocks and are liable to much greater experimental error than are the much more numerous contrasts which can be made up within blocks. By choosing, therefore, contrasts deemed to be experimentally unimportant for confounding with plot differences, the contrasts of experimental importance can be given the higher precision attainable by the use of small blocks.

The subject is of great variety and occasionally of some intricacy. I have attempted to clarify its logical principles and to illustrate its more important applications in *The Design of Experiments*, 2nd Ed., 1937.

6. Incomplete Balanced Randomised Blocks.

Some very important developments in this field are due to F. Yates. Yates pointed out that, if a number of variates v were grown in blocks each containing k plots, an exact and appropriate analysis of the results was possible without any serious complexity, provided that each pair of varieties fall equally frequently in the same block. If b blocks are sufficient to provide r replications of each variety, then evidently

$$vr = kb$$

If, also, every pair of varieties come together λ times then

$$r(k-1) = \lambda(v-1)$$

Two equations thus connect the five integral values, and the practical question arises for what integral numbers can the problem be solved for a moderate number of replications.

It is easy to see that, if all possible selections of k varieties were chosen from the v available, we should have

$$b = \frac{v\,!}{k\,!\,(v - k)\,!}$$

whence

$$r = \frac{(v - 1)\,!}{(k - 1)\,!\,(v - k)\,!}$$

and

$$\lambda = \frac{(v - 2)\,!}{(k - 2)\,!\,(v - k)\,!}$$

Generally the expression above for r will be more than the number of replications which can be grown, and a successful arrangement would only be possible if the three expressions contain a high common factor. This is a primary arithmetical requirement, which, however, leaves open the question of the existence of a combinatorial solution. Thus, it might be thought that with $v = 43$, $k = 7$, a solution could be found with 43 blocks and 7 replications, but this is known to be impossible.

The non-existence of a combinatorial solution when the arithmetical conditions are satisfied is, however, exceptional. It can be shown that the number in a block, k, must be less than or equal to the number of replications. Consequently, the arithmetical possibilities up to any finite number of replications can be listed exhaustively, as has been done with replications up to 10. Of these, a dozen cases remain where the existence of a solution is uncertain, but, for the majority, solutions are available which permit of the use of this method for a useful range of numbers of varieties extending, if 10 replications be allowed, up to 91 in blocks of 10.

One of the most useful of this series of solutions is derivable from a study of completely orthogonal squares. As is well known, the Graeco-Latin square may be regarded as two Latin Squares in super-position, these having the property that each letter of one square falls once on each letter of the other. With $s \times s$ squares a group of $s - 1$ Latin Squares may be mutually orthogonal. This possibility is easily realised when s is a prime number, and has been shown to be always realizable when s is a power of a prime. From a complete set of mutually orthogonal squares of side s it is easy to solve the solution of the problem of randomised blocks where $k = r = s + 1$, and $v = b = s^2 + s + 1$. The non-existence of the 6×6 Graeco-Latin Square is, thus, the reason for the non-existence of the incomplete block-solution stated above for 43 varieties in blocks of 7. A second series of incomplete block problems for which $k = s$, $r = s + 1$, $v = s^2$ and $b = s^2 + s$, may also be derived from orthogonal squares. But these two series, though providing immediately useful solutions, are only a first step in the elucidation of the general combinatorial problem.

The immediate practical problem is to obtain any solution. This can then easily be randomised and used in experimentation. Such a first solution is often found by a process of cyclic substitution. In many cases such solutions are unique, as are those based on orthogonal squares up to 7×7, in the sense that, from any one solution, all the others possible may be generated by permuting the varieties. In other cases solutions exist of a variety of types or species not connected by any such permutation. Given any one solution, these may be explored by interchanges among a limited number of blocks analagous to the reversal of intercalates in the exploration of Latin Squares. For example, for arranging 13 varieties in 26 blocks of 3 with 6-fold replication there are two sets of solutions with different relationships of symmetry among the varieties. In examining the arrangements with 15 varieties in 35 blocks of 3 with 7 replications, I find no less than 79 different sets of solutions, of which a few possess the further property, of some experimental interest, that the 35 blocks can be divided into 7 groups of 5, each group containing one complete replication. These constitute the totality of the intrinsically different solutions of a problem which attracted some attention during the nineteenth century, under the name of " Kirkman's schoolgirl problem ".

7. Lattice Designs.

For very large numbers of varieties to be compared, Yates has also developed a system of quasi-factorial arrangements in two-fold or multiple lattices, wherein the number used in a single block is the square, cube-or higher root of the number of varieties. Thus, a recent experiment in forestry genetics, carried out in California, uses blocks of only 9 rows of seedling each from 729 different seed parents. These 729 varieties are arranged diagrammatically in a $9 \times 9 \times 9$ cube, and the 81 rows of 9, parallel to any edge, constitute a single replication of 81 blocks. A set of 3 replications then suffices to include rows of 9 parallel with all edges, and one or more such set of replications constitutes a complete experiment in which, although the precision of all possible comparisons is not exactly equalized, yet it is made sufficiently equal for all practical purposes, and very considerably higher than if an entire replication had to be included in a single block.

When the number of varieties is a square, only two replications are needed to complete a set, having in their blocks the contents of the rows and columns of a diagrammatic square. By using a Latin Square solution a third replication and, with a Graeco-Latin Square, a fourth may be added having the same relation to the first two as these have to each other. If this process is carried so far as to use a completely orthogonal square, it is found that the arrangement has become balanced with respect to the precision of all comparisons, and we have a balanced incomplete block arrangement of the second series referred to above as derivable from orthogonal squares.

A very beautiful and effective application is the arrangement known as the lattice square. The total degrees of freedom among s^2 plots or varieties number $s^2 - 1$. A completely orthogonal square may be regarded as a method of dividing up these into $s + 1$ parts, each of $s - 1$ degrees of freedom. In a Latin square $s - 1$ degrees of freedom are discarded in the elimination of rows or columns, so that, if s is odd, *e.g.*, 11, the 12 parts, each of 10 degrees of freedom, may be assigned to the comparisons between rows and columns in 6 different squares. Thus with only six-fold replication a number as large as 121 different varieties may be tested in a perfectly balanced design with the high precision appropriate to comparisons in a Latin Square. Randomisation in such a case is effected as with the simple Latin Square, by permutation of the rows and columns and by random assignment of the varieties to the symbols used in building up the plan.

In general, it may be observed that the modern tendency is to make experiments larger, more comprehensive, and more intricately planned. This tendency has given an entirely new importance to combinatorial analysis and has thrown up numerous problems to which, in the present state of that subject, no adequate answer can be given. By utilizing what is already known, however, the experimenter has a wealth of resources from which to choose, according to the nature of his material and the aims of his experimental programme.

REFERENCES

FISHER, R. A. (1926). *The Arrangement of Field experiments.* Journal of the Ministry of Agriculture, 33, 503-513.

NEYMAN, J., IWASZKIEWICZ, K. and KORODZIEJCZYK, St. (1935). *Statistical Problems in Agricultural Experimentation.* Journal of the Royal Statistical Society, Supplement to Vol. 2, 107.

EULER, L. (1782). *Recherches sur une nouvelle espèce de quarrés magiques.* Verh. v. h. Zeeuwsch. Genootsch. der Wetensch. Vlissingen, 9, 83-239.

TARRY, G. (1900). *Le problème des trente-six officiers.* C. R. Ass. Fr. Ac. Sc., 29, 170-203.

NORTON, H. W. (1939). *The 7 × 7 Squares.* Ann. Eugen., 9, No. 3.

YATES, F. (1936). *Incomplete Randomised Blocks.* Ann. Eugen., 7, 121-140.

FISHER, R. A. and YATES, F. (1938). *Statistical Tables.* Oliver and Boyd.