Application of Statistics to Field Technique in Agriculture.

By Rao Bahadur M. Vaidyanathan, M.A., L.T.,
Statistician, Imperial Council of Agricultural Research, New Delhi.

Introduction.

APPLICATION of statistics to field technique in agriculture is now assuming an importance and a usefulness which can be compared only to the utility of a tool to a mechanic or of an instrument to a surgeon. Statistics is now an indispensable dissector for an agricultural experimenter to judge the results of his experiments, and the modern agronomist cannot now for a moment dispense with the applications of modern statistical theories for properly designing his experiments and for a valid interpretation of his results. On these aspects of field technique a co-ordinated research is necessary as in every other science, and it seems doubly so in the case of statistics applied to agriculture owing to its varied applications and to the varying conditions under which it is to be applied. The American Society of Agronomy is doing its best to co-ordinate the statistical researches as applied to agriculture, and the Agricultural Research Council in England has been emphasising the need for sound statistical treatments in connection with field experiments. The Imperial Council of Agricultural Research in India is not slow to take to modern statistical ideas, and has been pleading for a correct statistical technique in the case of field experiments in the Provinces and Indian States.

It should, however, be mentioned that statistics is only a means to an end, and that its application to field technique is intended merely to provide standards for comparisons of results from experiments conducted under known conditions. It is in no way intended to create an art by itself and in no way meant to discourage the experimenter from utilising his full knowledge to his best advantage. Even experiments which had not been designed from the point of view of modern statistical ideas could be studied and interpreted, but it should be emphasised that a proper design for an experiment with a view to a valid interpretation of results, would go a great way in strengthening the hands of the experimentalist, and giving him a courage and a conviction in "disentangling the diverse factors that contribute to a joint result". This is just what happens in the study of results of an agricultural experiment, where a number of variations such as soil heterogeneity, varietal effects and manurial effects, operate jointly to produce a single result of, what is known, as the *plot yield*. Under old ideas, a *high* percentage difference in plot yields, say, of two varieties under trial, is a sufficient guarantee that one variety is superior to another, but according to modern ideas while *high* and *low* are purely relative terms, the *significance* of the difference should be based upon a knowledge how far *chance* had operated in bringing out that difference. If the *chance error* is high, even a high difference—say 30 to 40 per cent. difference—may not be *significant*.* This is the experience met with in some of the recent results in agricultural experiments in India, where it was found that a high percentage difference of even 30 per cent. between treatments was found not *significant*. Thus a successful experimenter should try to bring down his random error as low as possible; so that even *small* differences between different factors at work—say differences in yield of different varieties—may on the basis of this error be *significant*. But even more important is the *validity of estimates* of error, which can be secured only by a *suitable design*. The validity of estimate of error depends partly upon whether estimates of other variations such as soil heterogeneity in an agricultural experiment are properly eliminated from our accounts, and partly upon whether our sample of plot yields is a *random* sample of population. A proper design for an experiment is thus the only panacea for ensuring a valid interpretation of results. We have now reached a stage in the progress of field technique, when we could plan even a complex experiment where a number of interrelated factors can be simultaneously studied, and significance deduced.

Principles of Modern Experimental Design.

This takes us to the *three* broad criteria for a satisfactory experimental design, which are now more or less accepted. *Firstly*, as seen already, the error of the experiment

* The exact connotation of *significance* is explained in the subsequent paras.

should be as low as possible giving a maximum precision for the experiment. This can be secured only by a sufficient number of replications. The need for replications in a field trial is now easily recognised, as the experimenter knows by experience that this is the only way of eliminating from his comparisons the effects of soil variation obtaining in the field. The *second* condition for a satisfactory field lay-out is that the error as estimated should be a *valid* estimate. Old plans of lay-out and methods adopted for estimating the error did not aim at separating the portions of the differences due to several causes, and hence led to aggregates which were of only limited application. For example, Mercer and Hall's method of basing the error of an experiment upon plot variations of individual treatments did not discriminate between the variations due to several factors; so also Engledow and Yule's which did not take into account the soil variation. Modern methods of field experimentation are intended to remedy this defect. The *validity of the estimate* for error can, however, be secured by a sufficient number of replications, provided the plots represent random samples of the experimental area. As this condition is generally not satisfied, what is known as *randomisation* is now introduced in all field experiments [by which plots in each block (or row and column) are randomised with respect to treatments under trial]. This ensures, in the language of Fisher, that "the differences utilised in the estimation of error (by which differences between unlike plots are judged) are properly representative of the other errors which produce the actual errors of the experiment". So long as we recognise that a correct interpretation of results depends upon valid estimates for error, the best way for securing such a valid estimate seems to be only by arranging the plots at *random*. Any systematic arrangement of plots cannot secure an unbiassed estimate for error and the old idea of having unlike plots as close as possible deprives us of a useful method for deriving a valid estimate for error. The arguments against *randomisation* sometimes advanced by agricultural workers are discussed in the next para. The third criterion is that the design should be such that all soil differences could be eliminated from the comparisons to be made and equally well from the estimates of error which form the very basis for judging "significance". It is found from experience that increasing the replications beyond a certain limit cannot bring on an additional precision for the experiment, which, therefore, limits the number of replications. To secure the object of eliminating soil differences, replication cannot therefore be a sole remedy, and the modern plan is to limit the area of the experimental field and to demarcate it into "blocks", or "rows and columns"; thus securing "localisation of control", by which it is possible to eliminate soil differences by proper field arrangement, and to increase the precision of the experiment, without unduly increasing the number of replications.

These three broad criteria for a proper design for a field experiment have led to three broad concepts:—*Replication*, *Randomisation* and *Localisation of Control*. Their objects are, briefly mentioned, to secure a valid estimate of error by means of a suitable design and an improved precision for the experiment by taking into account all possible variations such as soil heterogeneity and varietal effects, and by eliminating them from our calculations of error.

RANDOMISATION—POSSIBLE OBJECTIONS AND HOW THEY ARE MET.

While *randomisation* seems thus theoretically a necessity, there has been an acute controversy among the agricultural experimenters in India and elsewhere with regard to the utility of the method. It would be worthwhile at this stage to examine the objections raised against the method, and see how they could be met. A common objection raised is that the randomised method does not show on the field to the eye of an observer the relative differences between treatments; thus while an arrangement A B C, A B C, in several blocks systematically arranged might show the relative differences of treatments, an arrangement like A B C, A C B, C A B...... could not show at a glance these differences. But this is no argument against *randomisation*, and besides there is a confusion in this argument between a *field experiment* and a *demonstration* plot. While in the latter, the results should be demonstrated to the layman to appeal to his eyes, the former is essentially the domain of an experimentalist—for him to ensure the results and to satisfy, before they can be tried on a large field scale for confirmation. There could be no mistaking these two, and it is wrong to presume that one is a substitute for another.

The second argument usually advanced is that mistakes are easily committed in a randomised arrangement owing to untrained labour, which are avoided in a systematic arrangement. This is again no argument at all and the general experience is that Indian labour, if properly trained even for a short period of one season, picks up the details of even complicated plot arrangements so quickly, that it is unfair to advance this point as an argument against randomisation. The third argument against randomisation is that in several of the systematic arrangements recently adopted in the Indian farms, the standard error is found to be very low—even at 3 per cent.—and that therefore there is no need for changing the plan. But then there is no guarantee that this small error is a *valid estimate* of error, which, as we have seen, is a fundamental condition for a valid interpretation of results. Experience has shown that correlations between plots such as existing between high-yielding and low-yielding plots are often marked, and that they vitiate the results sometimes very badly. Thus any systematic arrangement of treatments in a field experiment seems to have no justification whatsoever. It is therefore seen that while the arguments usually advanced against randomisation are mainly from the point of view of practical agricultural considerations, the theoretical aspects of the problem have not entered into them. *Randomisation* in a field experiment seems therefore a necessity and it seems irrevocable till we can find a substitute for it.

PRINCIPLES OF ANALYSES OF VARIANCE AND CO-VARIANCE.

The whole statistical problem connected with the lay-out of a field experiment thus resolves into the possibility of analysis of total variation of plot-yields into those due to the component factors, and the use to which such an analysis can be put for testing the significance. Such a possibility presupposes a suitable design for the experiment —a design adapted to proper calculations and valid separations of those variations. In a randomised block arrangement, for example, each plot-yield may be regarded as in part due to the particular block in which it is situated, as in part due to the particular treatment and as in part due to a residue on account of what is called *error*.

The method of analysis of total variation into component items is effected by what is now generally known as Fisher's "Analysis of Variance." The method of analysis of variance has become classic in the modern theory of statistics, and it is now freely applied not only to agricultural experiments but to economic and biological studies where a number of factors operate to produce a joint result.

The principle of the analysis of variance as applied to data where only two factors cause variation is brought out by the single algebraic identity :—

$$\sum_{1}^{Kn}(x-\bar{x})^2 = K\Sigma(\bar{x}_p-\bar{x})^2+\sum_{1}^{Kn}(x-\bar{x}_p)^2.$$

The formula exemplifies the simple truth that if Kn observations be split up into n groups with K individuals in each group, then the total variation or what is known as the 'total sum of squares' can be split up into (1) the 'sum of squares' due to deviations of the *means* of the groups from the general mean (multiplied by the number in each group), and (2) 'sum of squares' due to variations of the individuals from the means of groups to which they belong. These two causes of variation are known as due to 'between classes' and 'within classes'. The two causes of variation are *independent*, and any individual x in the pth group for example may be expressed as $x=\bar{x}_p+\bar{x}+$ error, the error portion being that left over after assuming the effects of the two independent causes. The relative importance of the two variances† measures the correlation, if any, in the sample; if the variances are equal, the correlation is 0, and if they are not we may express the relationship in terms of r. If the variance of 'between classes' is larger than that due to 'within classes', then the intra-class correlation is +ve and if smaller it is – ve. Apart from finding the existence of any correlation or otherwise in the sample the analysis of the total variation into component factors, as we shall see in the subsequent para, provides us with a test of significance for homogeneity or otherwise of the sample, which is by far the most important use to which 'the analysis of variance' has been put.

Equation (1) can now be extended to the case of three items of variation producing a joint result. Thus in the case when Kn individuals are split up into n groups with

† Variance is 'the sum of squares' divided by the appropriate number of degrees of freedom or independent estimates.

K individuals in each group and with the restriction say that the variation of the rth individuals in the several groups is also to be considered, then the identity becomes :—

$$\sum_{1}^{Kn} (x - \bar{x})^2 = K\Sigma(\bar{x}_p - \bar{x})^2 + n\Sigma(\bar{x}_r - \bar{x})^2 + \Sigma(x - \bar{x}_p - \bar{x}_r + \bar{x})^2.$$

This corresponds to the case of a randomised block arrangement with n blocks, and K plots in each block corresponding to K treatments under trial. The last item corresponds to "error" or "interaction" between the first two items of variations. The whole analysis is easily seen to be a process of fitting constants so that the error variance is least ; that is to say, if the observed plot yield y_{uv} (*i.e.*, of uth block and vth treatment) is considered to be the sum of different effects such that $y_{uv} = K + t_u + b_v +$ error then by summing this for all the plots and by applying the method of least squares for minimising the error the best values for the constants will be :—

K = general mean of all plot-yields.
t_u = the treatment mean.
b_v = the block mean.

The principle of "analysis of variance" can now be extended to any number of simultaneous classifications of different sets of groups or classes. Now defining an n-fold classification as one containing n classes or groups into which a sample can be analysed, a randomised block arrangement, say with 4 treatments, is then a *double four-fold classification* (*double*, because only two items of variance *blocks* and *treatments* enter into calculations) ; similarly a 4×4 Latin Square will be *a triple four-fold classifition*, and so on. It may be of interest to note that in the types of designs which we are dealing with (*i.e.*, orthogonal‡ designs), as one set of effects does not alter the other sets, the constants may be fitted in *simultaneously* or *one after another*, and the sum of squares for the different effects will be the same by either process. Again so long as the design is *orthogonal*, that is to say, so long as the different items of variance can be estimated separately and directly, the total sum of squares will be equal to the total of the sums of squares contributed by individual items. Thus in a Latin Square arrangement the variances of the different items—rows, columns and treatments—may be separately calculated and their sums of squares totalled up will be equal to the total sum of squares. The calculation of the analysis of variance in cases of orthogonal designs has been simplified very much recently, and the easiest method will be to calculate for each item of variance—say block variance in a randomised block arrangement—the sum of the squares of the totals of several blocks divided by the number of plots in each block and to subtract the correction T^2/n (where T is the total of the plot-yields and n the number of plots) ; and the correction will be the same for all items of variance.

Just as the variation of a single variable x could be separated into several items such as those due to "between classes" and "within classes", similarly if pairs of observation of two correlated variables x and y occur in groups, the co-variation of x and y could be separated into different items. Thus in an agricultural field experiment involving blocks and treatments, the plot-yields in any two years may be correlated, and the co-variance may be analysed into items (1) blocks, (2) treatments and (3) error. The co-variation of x and y is of course measured in terms of *mean product*, just as variation is measured in terms of *mean square* ; b, the regression co-efficient, is the ratio of the co-variance § of x and y to the variance of x. To an agricultural experimenter, the chief interest in the "co-variance" lies in its application to the correction of plot-yields in a set of plots in one year on the basis of yields in the same set in the previous year or years. In a field experiment, a knowledge of preliminary yields may help to know *firstly* how the yields in the experimental year are affected in relation to the preliminary yields and *secondly*—which seems more important—how the standard error of the experiment changes. Assuming a linear regression of y on x, where x is the preliminary yield and y the experimental yield, any correction to be made in the yields of any two plots treated alike in the experimental year, should obviously be based on the difference in yields of those plots during the preliminary period ; and assuming a linear regression $y = bx$, b, the co-efficient of regression is the ratio of "co-variance of error in xy

‡ Explanation of 'orthogonal designs' is given in this and succeeding paras.

§ The *co-variance* is the mean product of x and y measured from their means.

analysis" to the "variance of error" in the analysis of preliminary yields. It is interesting to note that the analysis of variance of x and y, and the analysis of co-variance of x and y follow the same procedure in the matter of computation and that from these tables "b" is easily computated, and hence the sum of squares of the adjusted yields from the formula :—

$(y - bx)^2 = y^2 - 2bxy + b^2x^2$. The adjusted yield itself is then $y - b\,(x - \bar{x})$.

In experiments on perennial crops (such as 'tea'), a plan of lay-out is now adopted of what is known as 'equalisation of plot-yields,' by which sets of plots in the several blocks are so chosen that the sum total of yields in the same set during the preliminary period is the same. This method of lay-out combined with an assumption of regression between the preliminary and the experimental yields has given excellent results in the reduction of the standard error, and in the effective comparisons between treatments. But in the case of a few experiments on perennial crops conducted in India the co-variance between preliminary and experimental yields has not given any increased precision for the experiment. It should, however, be pointed out that where treatments themselves have produced differential effects during the preliminary period, the method of co-variance (or the assumption of regression of *error* between the plot-yields during the preliminary and the experimental periods) fails to give a correct perspective for an altered *precision* for the experiment. This is a very important point to be borne in mind in the application of the method for judging the improved precision of experiments on the basis of preliminary trials. But where the preliminary trial is an unbiassed uniformity trial (*i.e.*, subject to the same or no treatment), then the method can be freely applied.

OTHER APPLICATIONS OF THE METHOD OF CO-VARIANCE.

It is possible to apply the method of co-variance to other cases arising in agricultural experiments, where accidental factors come into play such as uneven germination and insect pest, the effects of which cannot be measured. In such cases, we might, for example, correlate the number of plants with the eventual yields, and thus correct for the differences in the plant number in different plots by the method of co-variance. The germination count in different plots will thus be a very important guide in judging the effects of such accidental factors. Another use to which the method of co-variance can be usefully employed in an agricultural experiment is to know what exactly the factor or factors connected with the crop that influence the eventual yields. This will be of great help to understand the different stages of plant growth leading to the yield as the effect of treatments. Thus in the case of cotton, if boll-count should be the deciding factor, we might assume a regression of the yield on boll-count; or if the yield should depend upon the boll-size that will be the factor to be correlated. As another example, in the case of rice or wheat, we know 'tillering' is a very important factor influencing the yield, and we might usefully study by the method of co-variance its effects, at several stages of plant growth, on the yield. Thus the method of 'co-variance' helps to study 'the mechanism' by which the treatments produce their eventual effects. Such intensive studies have been undertaken in some of the Indian farms but the results need to be collated in a broader perspective.

CHOICE OF PROPER STATISTICS.

Once that the variations can be analysed into their component factors on certain valid assumptions, the problem turns out to be one of the study of '*significance*'. Stripped of all technical language the question is :—"If with respect to a *sample* the variance of one is larger than that due to another, can we say that the variation of the first is *significantly* higher than that of the second?" In an agricultural experiment to judge, say, the comparative performances of varieties, what is needed is firstly whether the variance due to varieties is *significantly* larger than that due to error, so that we can say with confidence that our experiment is a success; and secondly whether on the basis of 'error' one variety is *significantly* superior to another. Both these tests depend upon the exact meaning and implication we attach to the expression 'significance'.

The connotation of what is termed as 'significance' or 'significant difference' obviously depends upon what we can expect in the *population* of which our data are a sample. From the *statistic* or *statistics* calculated from our *sample*, can we say that it is a random sample of the original *population?* In other words, a sample of size n is observed, and the problem is whether we could say that the sample is a random sample of the original population. In case the character of the population is known, inferences with regard to the sample are expressed in terms of mathematical probability, but where we should infer from the sample only, the problem is firstly one of estimation of the population and then the probability of occurrence of the sample, which, in the language of Fisher, is a function of the unknown parameters of the population which we are trying to evaluate. Fisher would call this function *likelihood*, and his solution by "the method of maximum likelihood" (explained later) would provide *efficient statistics*. In an agricultural experiment n cannot be large and this adds to the complexity of the problem. The main difficulty, however, is to specify the population in terms of the sample. If the character of the population *can* be assumed,—such an assumption is not always valid,—then it is easier for us to verify by mathematical processes whether the sample is a random sample. The *specification* of the population is by means of parameters based upon *statistic* or *statistics* which are functions of the variables. Thus we may specify a population by $y = a + bx + cx^2$ and so on, where $a, b, c, \ldots\ldots$ depend upon the statistics to be calculated from the observations. The choice of the mathematical expression itself is largely intuitive, and χ^2 test (explained later) will show how far the assumption is justified.

In the problem of estimation, however, we shall have to assume the form of the curve for the population with one or more unknown parameters. Now then with the *sample* values, the first requirement is the choice of the statistics for an estimation of the parametric functions of the population, and the second is to calculate the *chance* for the sample *being a random sample* of the population. While the first requirement involves a suitable choice of the statistic or statistics,—for any number of such statistics will be available for estimating the unknown parameters of the population,—the second requirement is answered by a study of the mathematical law of distribution of the statistic or the parameter evolved out of it, as it varies from sample to sample of a constant size. n, the size of the sample, thus becomes a primary consideration both in the choice of the statistic or statistics and in the evaluation of distributions. Fisher classifies all *statistics* into those *consistent and inconsistent, efficient and inefficient* and *sufficient and insufficient.* In estimating parameters of the population, we could have innumerable statistics from which to estimate them, but the conditions for a proper statistic are:—*firstly*, that it should tend to a fixed value as the size of the sample is continuously increased, or in other words, that it should centre round a fixed *value* with errors or deviations from it distributed in a normal curve; *secondly*, the particular statistic selected should give a very low variance in *large* samples, *i.e.*, lower than those of other statistics which we could possibly think of, and *thirdly*, the statistic selected should be examined for its sufficiency, that is, whether it can supply all information regarding the sample, in which case, even if it does not give a low variance, there is no need for the calculation of other statistics. The first criterion secures *consistency*, the second *efficiency* and the third *sufficiency*. Taking the Arithmetic mean of a sample as an example of the statistic from which to calculate the parameters of the population, we know that it is *consistent*, since in large samples it is distributed in a normal law. But its *efficiency* will depend upon whether other statistics are *not* available also normally distributed as n is increased, but giving a *lower variance* for the purpose of estimating the parameter of the population. Now since the variance falls off inversely with n, the condition for efficiency is that the limiting value of $\frac{1}{nV} \leqslant i$, where i is independent of the estimation used. Thus if the original curve be *normal*, the Arithmetic *mean* is consistent and efficient in estimating a parameter of the curve, but its efficiency is lowered when it is used to estimate say an exponential curve, where other statistics define the parameter of the curve more accurately. Again in *small* samples, the *Arithmetic mean* is sufficient to give complete information of the sample, and though it may *not* be efficient in the sense explained above, it serves the purpose so far as it completely summarises all possible and available information from the sample.

Thus we shall have to choose from a number of *consistent efficient* statistics the most suitable one to deduce the best estimates of the parameters of the population. Fisher's method of maximum likelihood helps a solution of the problem. If θ be the unknown parameter of the population, the method consists in multiplying the logarithm of the expected frequency in each class by the observed number, and summing for all the classes; and solving for θ such that the sum is a maximum. As a simple example, if a, b be observed numbers in two classes so that $a + b = n$, with probabilities of their occurrences say $f(\theta)$ and $1 - f(\theta)$ respectively, then the maximum likelihood solution will give θ for which

$a \log f(\theta) + b \log \{1 - f(\theta)\}$ is a maximum. The positive solution for θ secures a statistic with a low variance.

χ^2, t, z TESTS OF SIGNIFICANCE.

From what has been said, what is needed with respect to a sample of n observations (n not being large) is the deduction of valid tests of *significance*, to know *firstly* whether the sample is a random sample of a homogeneous population, and *secondly* whether the *means* calculated from the sub-samples differ significantly on the basis of analysis of variance. Dealing with 'variance', it is probably legitimate to assume that the original population is *normal*, in which case the scheme of analysis of variance explained already, combined with a knowledge of distributions of the statistic or statistics in random samples helps to arrive at proper tests of significance. In the case of an agricultural experiment, the procedure of analysis into variances due to blocks, treatments and error (or in the case of complex experiments including all interactions of higher orders) helps an understanding of how best the test could be applied.

There are three such tests now in vogue which are easily explained. What is known as χ^2* test (or test of 'goodness of fit') given by Karl Pearson in 1900 is intended to test agreement *between observation* and *hypothesis* where the variates are normally distributed and mutually correlated. It is based upon the distribution:

$$df = K \ \chi^{n-1} e^{-\frac{1}{2}\chi^2} d\chi.$$

'Student' showed in 1908 that the same law holds good in the case of the mean square of a random sample drawn from a normal population. χ^2 test has however been found not effective when either the method of fitting is inefficient, or when negligible values are included in the cells of the sample. But generally speaking, the method has been found to be one of the most powerful tools in modern statistics, which ensures the very first step in all biological studies for verifying observation with any assumed hypothesis. But there is some misunderstanding with regard to the full utility of χ^2 test which seems to have arisen from the confusion sometimes caused in the two independent statements:—

(1) A sample does not differ *significantly* from an assumed population $f(x)$.

(2) A sample is most *likely* to be a *sample* of a population $f(x)$.

While (1) can be tested by χ^2 method, it does *not* however follow from (1) that (2) is true. There might be any number of populations of which the given sample could have been extracted, all satisfying the χ^2 test at the same or particular levels of significance, but only those giving low variances are to be preferred. In other words, χ^2 method is useful only to this extent, that it can safely be employed to test whether the given observations agree with an assumed law or not, but not to test the reverse that the original population should be the one assumed. Thus, for example, it can be employed to test, in genetics, whether there is agreement between observations and Mendelian class frequencies, or in biology *independence* in a four-fold or an n-fold classification assuming the marginal totals to be true. But in either case, unless fresh evidence is adduced, the complete identity of the sample cannot be assured.

't' and 'z' Tests.—The two other tests 't' and 'z' have now become very popular with the agricultural experimenter. In fact, no experimenter now-a-days takes the trouble of enquiring whether conditions necessary for the applications of these tests are fully satisfied; but it is however found that even when those conditions are *not* completely satisfied, they can safely be employed for testing 'significance'. What is known as 't' test is to test whether the observed 't' from a sample, *i.e.*, $\left(\frac{\text{mean}}{\text{S. E. of mean}}\right)$ follows the distribution of 't' from all possible random samples of size n (for varying values of n). 'Student' gave his distribution of 't' in 1908, in his classical paper 'the probable error of the mean'. The utility of Student's distribution is now seen in almost every kind of problem where the significance of any statistic in terms of its standard error

* $\chi^2 = \Sigma x^2/m$ where $m+x$ is the number observed, and m the expected.

has to be tested. 't' tables have been constructed (*e.g.*, Fisher's tables) based upon theoretical distributions, giving the *probability* or *odds* for or against deviations from the observed 't' occurring due to chance from which the significance could be judged. The theoretical distribution of 't', as in the case of those of other statistics, is based on the assumption that the original population is normally distributed, and that a random sample of size n is drawn from it so that the chance of $(x_1, x_2 .., x_n)$ in the interval $(dx_1, dx_2 .., dx_n)$ is given by :—

$$df = \mathrm{K}e^{-\frac{1}{2}\sum_1^n \left(\frac{x_r - x_m}{\sigma}\right)^2} dx_1, dx_2 .., dx_n$$

(where m and σ relate to the population). By a suitable transformation, the distribution of s^2 {*i.e.*, that of 'the variance' of the sample calculated from the expression $\frac{1}{n-1}\Sigma\left(x_\gamma - \bar{x}\right)^2$ is deduced, and similarly that of 't'. We have after transformation :—

$$df = K'e^{\frac{-n(\bar{x}-m)^2}{2\sigma^2}} e^{\frac{-(n-1)s^2}{2\sigma^2}} s^{n-2}\, d\bar{x}ds$$

showing that x and s are independent, so that the distribution of s is :—

$$df = K'\left(\frac{\mathrm{S}}{\sigma}\right)^{n-2} e^{-(n-1)s^2/2\sigma^2} \frac{ds}{\sigma}$$

It can be shown that the mean value of s^2 from all possible samples is σ^2 showing that s is an unbiassed estimate for σ and that by the method of maximum-likelihood the best estimate for σ is s (*i.e.*, giving the smallest sampling variance). But be it noted that s^2 is calculated with $(n-1)$ as divisor in place of n to give an unbiassed estimate for σ^2, which is necessary for the simple reason that we are estimating both the mean and the standard error from the same sample, both deviating from their true values.

What is known as Fisher's 'Z' test is more comprehensive (t test, as we shall see, is only a special case of 'Z' test) and is intended to test whether the given sample is a random sample of a normal population. Fisher's 'Analysis of Variance' combined with the 'Z' test are now a landmark in the theory and practice of statistics applied to agricultural field technique. The principle of 'Z' test is this :—When the total sum of squares is split up into different sums due to a number of items, the variance of each item (*i.e.*, the sum of squares divided by the appropriate degrees of freedom) should be an *unbiassed* estimate of σ^2 of the population. If s_1 and s_2 are two such estimates of samples of a normal population derived respectively from n_1 and n_2 degrees of freedom, the distribution of

$$Z(= \tfrac{1}{2}\log_e s_1/s_2)$$

for varying values of n_1 and n_2 is an efficient statistic, and should help in judging whether our sample is a random sample of a normal population. What are known as 'Z' tables, at 5%* and 1%* levels have been constructed by Fisher on the basis of the distribution of Z, for varying n_1 and n_2. The distribution of 'Z' is based upon the distribution of s_1 and s_2, and is given by :—

$$df = \frac{2n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}}\, \mathrm{T}\left(\frac{n_1+n_2}{2}\right)}{\mathrm{T}\left(\frac{n_1}{2}\right)\mathrm{T}\left(\frac{n_2}{2}\right)} \times \frac{e^{n_1 z}dz}{(n_1e^{2z}+n_2)^{\frac{n_1+n_2}{2}}}$$

s_1 and s_2 should not appreciably differ if they relate to a random sample of a homogeneous normal population, but if they do, it will be indicated by a low value of P, and the sample cannot then be a random one. It may be pointed out that the 't' test is only a special case of 'Z' test, since for $n_1 = 1$ and $n_2 = n$, $Z = \frac{1}{2}\log_e t^2 = \log_e t$. Thus if we take any value of 't' from 't' table for n, $\log_e t$ will be the same as 'Z' from 'Z' table for $n_1 = 1$, and $n_2 = n$.

In an agricultural experiment where the total variance is split up into variances, say, due to blocks, treatments, and error, denoted respectively by s_1, s_2 and s_3, then if $\frac{1}{2}\log_e s_2/s_3$ is greater than 'Z' from the tables, at any level significance, (5% or 1% level is by convention the usual level taken), we infer that the soil is heterogeneous; if $\frac{1}{2}\log_e s_2/s_3$ is similarly greater than the theoretical 'Z', the general effect of treatments is significant. In either case, however, the *sample cannot be a random sample* from a homogeneous normal population. The success or failure of an experiment will depend upon the later criterion, *i.e.*, whether the variance due to treatments is significantly greater than that due to 'error'; if it is not greater, then the 'error' preponderates and no inference is possible from the experiment. *From* the 'Z' test we

---

* 5 per cent. level=chance 1 in 20 ; 1 per cent. level= chance 1 in 100.

proceed to compare the treatment-means by 't' test on the basis of the residual error. If s be the standard deviation per plot and n the number of replications, $s/\sqrt{n}$ is the standard error of treatment-mean and $\sqrt{2}s/\sqrt{n}$ is the standard error of difference of two means; this multiplied by 't' from the tables (at 5% or 1% level of significance) will be the critical difference between the means; if the difference between any two treatment-means exceeds this critical difference the difference between treatments is taken to be *significant*.

Doubts have been raised off and on, both by statisticians and agronomists, firstly about the validity of 'Z' test on the score that the original distribution may or may not be normal for which in any case there is no evidence; and secondly, whether, after establishing that the sample is *not* a random sample from a homogeneous normal population by the 'Z' test, we are justified in accepting the estimates of variances as *valid* estimates. The first objection is equally applicable to distributions of other statistics also, such as s and t, and our justification is that an assumption of a homogeneous normal distribution for the original population is sufficiently valid for all practical purposes, and that any departure from normality does not sufficiently impress upon our final form. With regard to the second point, it can be easily proved that the validity of the estimate for 'error' is *not* affected by any change in our hypothesis and that only the variances due to other factors are affected. But though this may be true, the adequacy of 'Z' and 't' tests is still there, and it is not in any way vitiated by the change in variances under 'blocks' or 'treatments.' It should be remembered that our analysis of variance procedure is not intended so much to *estimate* the variances, as to provide *adequate tests of significance.*

EXPERIMENTAL DESIGNS—MISSING PLOT TECHNIQUE.

Enough has been said to show that for a valid interpretation of results in a field trial a suitable design (combined with a proper method of analysis of the results) is absolutely and fundamentally necessary. Such designs may be classified into *orthogonal* and *non-orthogonal.* Examples of orthogonal types are the usual randomised block method and the Latin Square arrangement of plots where it is possible to estimate *separately* and *directly* the different items of variance; in such cases, the *mean* yield of all the plots receiving the same treatment, or the *mean* yield of block totals provides the best estimate for the treatment or block effects. It is thus possible in a randomised block arrangement to estimate from the treatment means the treatment effects, from the block means the block effects and so on, without either of them affecting the rest in the eventual calculations. In any such design, increase in fertility in one block affects all treatments *alike*, and conversely the effect of a particular treatment influences the yields of all the blocks; blocks and treatments are thus mutually orthogonal. Fisher's procedure of Analysis of Variance is particularly adapted to orthogonal types of experiments, though the procedure of analysis of variance in all cases is only an application of fitting constants which is the general method of analysis in dealing with all designs including even non-orthogonal types. Cases of non-orthogonal types are however unavoidable even in experiments of what are known as *simple* types, *i.e.*, with only one set of factors under trial—say a few varieties to be tested. It is, for example, quite an ordinary occurrence to see a few experimental plots spoiled by accident such as insect pest and flood so that the intrinsic yields of these plots are not known, or again for lack of knowledge of the initial fertility of the plots to get differences in treatment yields very much pronounced. In a recent case of an experiment which came to the notice of the author, not only was the design faulty as both the plot size and block size were abnormal with an insufficient number of replications, but also some plots were found to give abnormal yields. The orthogonality of the design in such cases is so much disturbed, that either the whole experiment should be discarded, or mathematical devices employed to correct for the abnormalities. In such cases the usual method of analysis of variance should be modified to suit each particular case.

In case where only one or two plots are 'missing', the usual procedure of the Analysis of Variance easily helps to calculate the best values for the missing plots. Thus if x, y be the 'missing' values, we have only to calculate algebraically, 'the error variance' by the usual method of analysis (which will involve x and y), and to minimise the variance by differentiating it with respect to x and y, by equating the two differentiated functions to zero, and by solving the equations for x and y to obtain the missing values. The process is the same as fitting constants to give a minimum variance to 'error'. The princple may be extended to the case of any number of missing plots, but it is *not* advisable to carry the process to more than 2 or 3 missing plots. There are, however, two points to be noted in such an analysis :—Firstly, the number of degrees of freedom for 'error' will be the usual number less the number of 'missing' plots, for the simple reason that values of constants have been derived from the known plots only. Secondly, 'the treatment variance' obtained after substituting the missing values is bound to be less—though only slightly ordinarily—than usual, and thus the application of any missing plot formula will show an exaggerated accuracy for the treatment averages. In such cases, a correction has to be applied which will depend upon the analysis of the original values, with 'the error' portion deduced from the calculated values.

COMPLEX EXPERIMENTS—CONFOUNDING.

It is the glory of the recent developments of statistics, that modern field technique aims at testing any number of inter-related factors simultaneously. Thus in the same experiment of a cultivation trial, sowing date, spacing of plants and age of seedling may all be tried together, instead of having three separate experiments with one for each of the factors. This would not only economise time, space, and energy, but would also aim at the truth more accurately than what an experiment with only a single factor could do. In fact, where a number of deliberate factors influence a result, such as sowing, spacing and age of seedling would with respect to the yield, it seems futile to try each of the factors separately. The only satisfactory method is a complex layout involving all the factors with a suitable planning. Such complex experiments may always be *orthogonal*, that is, if for example 3 sowing dates, 4 spacings and 3 ages of seedling should be tested, $3 \times 4 \times 3 = 36$ treatments may all be completely *randomised* in the same block, with say 4 or 5 replications. This is an ideal method for such trials, and is analogous to a simple experiment which we have dealt with already, except for a change in the items to be considered in the eventual analysis of variance; in such cases not only should we consider the main effects but also the interactions between the several factors which may in some cases be appreciable. Thus, in the particular example which we are considering, the different items of variance will be (1) sowing dates, (2) spacings, (3) ages, (4) to (6) interaction between sowing and spacing, that between spacing and age and so on, (7) second order interaction between all these factors, and (8) error. The method of calculation of the several variances is the usual procedure and for arithmetical calculations for working out the interactions, say that due to sowing and spacing involving 12 ultimate treatments, the variance of these 12 treatments *minus* that due to sowing *minus* that due to spacing will give the interaction required. This procedure is, as noted already, the same as fitting constants to represent the several effects and deducing them by the method of least squares.

But difficulties in the conduct of orthogonal complex experiments are experienced to be :—(1) the agricultural difficulties in the arrangement of diverse factors in a single lay-out; for example, where differential irrigation is involved, it brings on lateral seepage from plot to plot; (2) the huge extent of land needed for the experiment, which the experimenter usually finds difficult to secure. In either case, the remedy is found to be to 'confound' the effects by a deliberate plan, and to alter suitably the usual methods of analysis; such a process of 'confounding' will not only economise labour and space but will provide, as Fisher has shown, very efficient tests of significance.

We will consider here two simple methods of 'confounding' which may be usefully adopted by the agricultural experimenter in India. Take the case of an experiment involving two treatments, represented by types A and B, with n_1 and n_2 numbers respectively in each class, so that there are $n_1 \times n_2$ ultimate treatments. A type of layout which meets the first difficulty is to have n_1 sub-blocks in each block, and to have n_2 plots in each sub-block; the n_1, A-treatments, and n_2 B-treatments corresponding to each of n_1 A-treatments, being separately randomised. This is a case of 'confounding' the main effects with a *partial* randomisation only, and the Analysis of Variance should separate the error of the A-treatments from that of B-treatments as shown below :—

A–TREATMENTS.

	Degrees of Freedom
Blocks	.. (K − 1)
A–treatments	.. $n_1 - 1$
Error (A)	.. Kn_1-K-n_{1+1}
Total :—	$Kn_1 - 1$

B–TREATMENTS.

	Degrees of Freedom
From Blocks and A–*treatments*	.. $Kn_1 - 1$
From B–treatments	.. $n_2 - 1$
Interaction between A & B	$(n_1-1)(n_2-1)$
Error (B)	.. $(K-1)(n_1n_2-n_1)$
Grand Total :—	$Kn_1n_2 - 1$

The defect in this method of lay-out however is that the two errors A and B cannot be consistent in the sense, that while A-error is derived from fairly big plots, B-error is deduced from contiguously small-sized plots which would therefore be small. The efficiency of the experiment is thus disturbed, but for all practical purposes the lay-out may be considered satisfactory so long as block size is not unduly large.

(2) A *second method of confounding* which may be usefully adopted meets the second difficulty. Instead of having all the ultimate treatments in the same block, a complete replication may be sacrificed in such a way that each block may be *divided* into sub-blocks and sub-block differences may be confounded with higher-order interactions (such as sowing date × spacing × seedling age), and eventually allowing for the confounding in the analysis of variance. This is on the assumption that higher order interactions are small as compared to the experimental errors, and that instead of *adequate replication* which provides the basis for *error*, these interactions may be substituted. Thus a block containing the treatments *n*, *p*, *k*, *np*, *nk*, *pk*, *npk* and *c* (control) may be split up into

n	*k*
p	*npk*
nk	*np*
pk	*c*

so that second-order interactions are confounded with sub-block differences. In the case of those items which are not confounded, *i.e.* which are orthogonal to one another, the usual procedure of computing the sum of squares will be followed but in the case of those confounded, the general method is *fitting constants* to represent those items and calculating the sum of squares due to fitting. But where the design permits to take account of block differences which confound the interactions, it is easier to compensate for such differences, eventually leaving the confounded degrees of freedom orthogonal with the blocks and also with all the other treatments. Designs of particular types alone will answer these conditions and here it is, where confounding is resorted to, a clear idea of the plan and the procedure of analysis is necessary; otherwise the experimenter will be landing himself in extreme difficulties in the matter of analysis.

In India for field experiments non-orthogonal designs are slowly coming in, in different forms, and research is necessary to explore the full possibilities of such designs with correct methods of analysis. Here indeed the Statistician has his part to play, as indeed on so many other matters connected with the field plot technique.

EXPERIMENTS ON PLOT-TECHNIQUE IN INDIA.

From the Indian experimental data available so far, it is apparent that there has been a lack of uniformity in the conduct of field experiments in the several Provinces and States. Not only are the field experiments sometimes not properly planned, but also they are not carried through for a sufficient number of years to allow for a reasonable weather sampling. The various factors governing the error of an experiment—such as the plan of the experiment (*e.g.*, whether it should be of randomised block type, or a particular Latin Square type), or what should be the suitable *plot size* and *block size* and the *border effect*—should be fully examined under Indian conditions. Experiments to decide these factors are in progress in some of the farms, but the results have not been collated to be of much guidance. The usefulness of complex experiments should be fully explored; so also of confounded experiments which will not only economise labour and time but will provide very efficient tests of significance. "Sampling technique" *i.e. methods of taking samples* from experimental plots such as for physiological study, have not been studied yet with different crops. Again, in the case of *manurial* experiments particularly, we should know the *residual and accumulated effects of manures* for which special planning for experiments is necessary. What are known as "Permanent Manurial Experiments" should be suitably planned if the results should be of any value.

Enough has been said to show that India should evolve her own methods both in the matter of planning, and in the conduct of agricultural field experiments. A co-ordinating agency is of course necessary, and there is ample scope for mutual fellowship between the statistician and the agricultural experimenter.

The Antianæmic Principle of Liver.

By H. B. Sreerangachar, M.Sc., A.I.I.Sc., and M. Sreenivasaya, B.A., F.I.I.Sc.
Department of Biochemistry, Indian Institute of Science, Bangalore.

THE use of liver in the treatment of pernicious anæmia constitutes a striking therapeutic advance of great importance. The idea must have first originated from Whipple and Robschiet Robbins[1] who in their search for blood regenerative foodstuffs found that of all the substances they investigated, liver was most potent as a hæmopoietic material. This discovery led Minot and Murphy[2] in 1926 to make a clinical trial on pernicious anæmia patients and as a result of their classical researches they obtained a remarkable improvement in the blood picture of the treated patients. Since then, there have been a number of investigations supporting their regimen and now we can completely restore the anæmic patients to normal health by administration of liver.

ETIOLOGY OF PERNICIOUS ANÆMIA.

Although the results of Minot and Murphy made it a logical conclusion that pernicious anæmia is a disease due to a dietary deficiency, there were also other theories prevalent to explain its cause. The accumulation of toxins in the body, the infectious disorders in the intestinal flora and the absence of the anti-hæmolytic substance were individually suggested as the causative factors. The exact etiological significance of the defective gastric secretion was first suggested by Fenwick[3] in 1880 and has, since then, been supported by other investigators. Goldhammer[4] has shown that gastric secretion is proportional to the red blood cells and in pernicious anæmia there is a sub-normal amount of gastric secretion also characterised by complete anacidity. Castle and his coworkers[5,6] have shown conclusively that the stomach of a normal human being secretes some enzymic principle which, when allowed to react *in vivo* or *in vitro* with some substance present in the animal proteins of the food, produces the necessary antianæmic factor. The non-occurrence of this reaction in the body is believed to be a defect in the gastric digestion leading to pernicious anæmia. The secretory product is called the intrinsic factor and the substance derived from the food the extrinsic factor. The specific antianæmic principle thus produced is stored in liver from which it is elaborated as required by the bone marrow to produce the normal quota of erythrocytes.

The site and the mode of interaction of these two factors are not known. Their chemical nature is also obscure. The intrinsic factor is believed to be unrelated to either hydrochloric acid, pepsin, rennin or lipase. Klein and Wilkinson[7] have studied this intrinsic factor in considerable detail and have named it enzyme "hæmopoietin". Like most of the enzymes it is destroyed by heat. Griffith[8] has observed that its action is confined to P_H 3·5–5·5. The food factor is, on the other hand, thermostable and is found to be present in beef-muscle, autolysed yeast, rice polishings, eggs and liver. It is not identifiable with any portion of vitamin B complex.[9,10]

Following the earlier papers of Castle and his coworkers, Sergius and Isaac,[11] and Wil-

---

1 Whipple, G. H., and Robschiet Robbins, F. S., *Amer. J. Physiol.*, 1925, **72**, 395.

2 Minot, G. R., and Murphy, W. P., *J. Amer. Med. Assoc.*, 1926, **87**, 470 ; 1927, **89**, 759.

3 Fenwick, S., "On Atrophy of the stomach and on the nervous affections of the digestive organs." J. & A. Churchhill, London, 1880.

4 Goldhammer, S. M., *Proc. Soc. Expt. Biol. Med.*, 1935, **32**, 476.

5 Castle, W. B., and his coworkers, *Am. J. Med. Sci.*, 1920, **178**, 748-764.

6 Castle, W. B., and his coworkers, *Am. J. Med Sci.*, 1930, **180**, 305 ; 1931, **182**, 741.

7 Klein, L., and Wilkinson, J. F., *Biochem. J.*, 1934, **29**, 1684.

8 Griffith, W. J., *Biochem. J*, 1934, **28**, 671.

9 Diehl, F., and Kuhnau, J., *Deutsch. Arch f. Klin. Med.*, 1933, **176**, 149.

10 Lassen, H. C. A., and Lassen, H. K., *Am. J. Med. Sci.*, 1934, **188**, 461.

11 Sturgis, C. C., and Isaac, R., *J. Am. Med. Assoc.*, 1929, **93**, 747.