THE DEVELOPMENT OF FIELD EXPERIMENTS IN AGRICULTURAL RESEARCH—PART II

T. EDEN, M.Sc. (Manc.), A.I.C.,
TEA RESEARCH INSTITUTE OF CEYLON

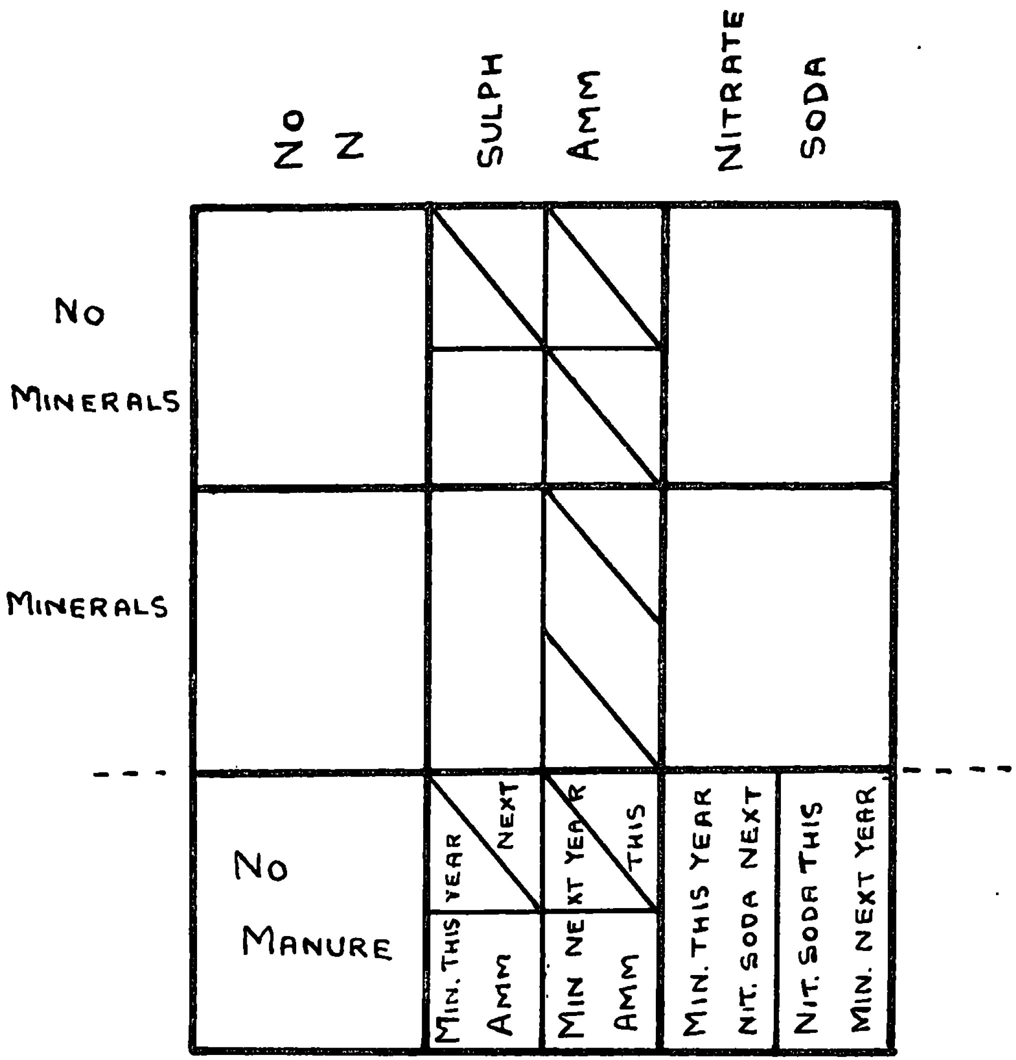
THE PERIOD OF EXPANSION, 1900-1923

THE work of Lawes and Gilbert stimulated interest in agricultural experiments in many parts of the world. In England, after the death of the Harpenden pioneers, the classical work at Rothamsted continued, but no important development resulted until the second decade of the century when more land became available through the help of the Development Commission in 1911. Agricultural colleges and farm institutes entered the experimental field, but with few exceptions, such as at Cockle Park, their efforts were not conceived on the same scale or with the same precision as at Rothamsted. The war, though it provided new problems in crop production, depleted the personnel of agricultural investigators, and thus militated against progress.

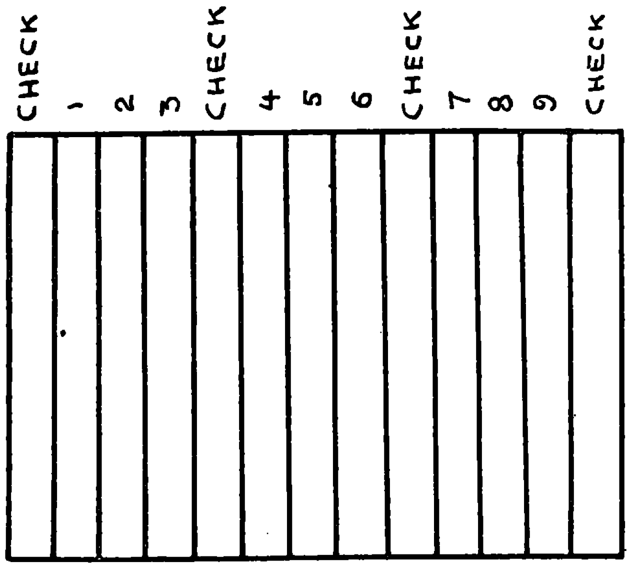
In the United States of America activity was more remarkable. Experiment Stations sprang up in a great many states, and much time and energy was devoted to field experiments. The incentive of the Rothamsted work is undeniable, but the American Stations embarked on a wider sphere of experimentation than had been customary in England. Every conceivable type of cultural and varietal experiment is to be found in the early records of the American Stations, but nowhere is there the same evidence of continuity as at Rothamsted. It was inevitable that in a country that was new in the agricultural sense, where there were many pressing problems, the satisfactory solution of these problems could not keep pace with the continual expansion of agricultural effort. The urgent demand for specific information led to the institution of short time experiments, the results of which were frequently contradictory. The volume of this work can be judged from the fact that, in the 'eighties of last century, the Experimental Station Record was started to give abstract reports of the published results from the many stations already in existence.

By the beginning of the present century it was beginning to be realised that the promise of clear-cut guidance from the multifarious field experiments was not being fulfilled and this gave rise to dissatisfaction with field experimental methods. American commentators of that time are commendably candid about the state of affairs. Piper and Stevenson [15] in a paper on the standardisation of field experiments, in 1909, suggest that though in the past the large sources of error in field experiments had been recognised, the difficulty of such experiments had exercised a retrogressive influence on those whose work it was to carry them out. In short, experimentalists had become unnecessarily careless in their technique. They had reached an impasse, for whilst they admitted the existence of formidable obstacles to accurate experimentation, they were undecided as to what steps to take to mend matters. This was the state of affairs when in 1907 the Journal of the American Society of Agronomy commenced publication, and led the way in the study of more refined methods of experiment.

In the immediately succeeding years, whilst agronomists were casting about for a solution to their difficulties, a number of workers in widely scattered centres were considering a new method of approach, the use of the theory of probability in the interpretation of divergent results in agricultural trials. To whom priority should be given is not clear. At first an approximate methods was used, and Egorov [16] in Russia employed the mean deviation of plot yield as a measure of heterogeneity. He studied reduction in error as plot sizes increased from one to 120 units of approximately 50 square feet (one square sazhen). His results were published in 1909. Wood and Stratton [17] in 1910 gave the first explanatory statement of field errors in England, and this was followed by Hall and Russell [18] in 1911, by Mercer and Hall [19] in the same year, and by continental workers later. Kostecki [20], Fröhlich [21], Frischauf [22], Lehn [23], Gregoire [24], and in Japan, Miyake [25], had all examined the method of least squares in relationship to agricultural experiments, particularly in variety trials, between 1912 and 1916. Of these papers, those of Wood and Stratton and Mercer and Hall had the greatest effect on subsequent progress, the latter to a



Cross-dressing Lay-out
WOBURN
Barred Plots have had Lime



Check-Plot Lay-out

B	A	C	D
C	D	A	B
D	C	B	A
A	B	D	C

4 x 4 Latin Square

large extent because of the practical application of the results through the collaboration of "Student". For the moment the history of these early attempts to apply statistical method is all that concerns us, but Student's work whilst amongst the earliest, had an aspect of modernity that makes it the link between the earlier and the later development in method; and because of its great importance its consideration will occupy a separate section later.

Wood and Stratton, (*loc. cit.*) set out to examine the general magnitude of crop experimental errors and their relation to size. By combining small plots on a uniformity trial of mangolds, they concluded that 1/80 acre represented an optimum size; and that for ranges larger than that there was no correlation between size and error. From 200 pairs of similarly treated plots drawn from crops of wheat, barley, oats, mangolds, swedes, potatoes and seed-hay they assessed the average probable error per plot as 5 per cent. a value which agreed with that for optimum plot size from the detailed mangold experiment. In the light of subsequent work they undoubtedly underestimated the plot variability, for their comparisons were based upon a larger number of plots than it is usual to find in a single agricultural trial. This however affected the permanent value of their contribution to the subject but little, for they had focussed attention on plot errors, and they had laid the foundation of replication as the remedy.

Mercer and Hall re-emphasized the necessity for careful choice of land and further explored useful plot sizes. They were not so fortunate as Wood and Stratton in their diminution of error with increasing size, and they recommended 1/40 acre as the standard optimum size. They too were optimistic in their estimate of ultimate error values, suggesting that a fivefold replication would give errors below two per cent. on land judged to be suitable for experimental purposes. In conjunction with Wood and Stratton they were responsible for the replacement of large or single plots, by small replicated units.

THE CHECK PLOT METHOD

At this time the main contribution from the American Stations was the development of the check plot. There was much activity in varietal work, and the changes that could be rung on the varietal theme were more extensive than with manurial treatments. Faced with a continually expanding list of varieties in cereals, and discontented with the old single plot methods, the experiment stations tried to combine control of heterogeneity with economy of plots by replication of standard check plots. The general principle was to arrange that at intervals of three, four or five plots, a repetition of a standard variety should occur to which all comparisons should be referred. This method of multiple controls was first used in Sweden by Holtsmark and Larsen [26] with check plots every third plot. They used the three nearest controls as the basis for their comparisons and corrected this average by reference to the average of all controls. Leidner [27] and his colleagues also favoured multiple control plots, though they concluded, after trial of this and other methods, that the ideal method had yet to be found. There is evidence of a considerable variety of methods of interpretation. In some cases attention was restricted to the two checks lying on either side of the specific variety. The assumption was made that fertility gradient between check plots was uniform. A simple formula based on this hypothesis allowed the potential fertility of the intermediate plots to be calculated, and this hypothetical plot value was used as a correction.

In other cases the correction for fertility was made more general by taking into account variation over the whole area. For any particular variety the correction was the difference between the nearest check yield and the average check yield. The sign was positive or negative according as the yield of variety was below or above the average check yield [28].

The disadvantages of these methods far outweighed their advantages. In the first place, comparisons could only be made through the appropriate check variety instead of directly. Further, the yields were entirely hypothetical; they represented nothing real. Most serious of all is the objection that after correcting, one has no criterion of the errors of correction. It is far from satisfactory to try and reduce one source of error by introducing yet another, even though the expectation is that the second error will normally be smaller.

Gradually the concept of replication of all varieties made its way into the technique. Pritchard [29] combined the system of alternate checks with a replication of other varieties as high as tenfold. Such a plan besides being wasteful of space takes no account of the fact that the mean check yield from which all corrections and comparisons derive will be estimated with far greater accuracy than the rest of the experimental yields.

When statistical formulae were applied to results from the check plot method, Kiesselbach [30] plainly showed its deficiencies. He investigated the question of optimum spacing of checks using checks alternately, every second, and every third plot.

He found that there was much variation in his calculated errors, and that checks every third plot gave lower errors than when closer spacing was adopted He concluded that these anomalous results invalidated the method, and made a strong plea for replication as the mainstay of experimental accuracy.

THE ROD ROW METHOD

The multiplicity of experimental plots led to an exploration of further means of saving time and labour in the collection of data. Experiments, chiefly associated with the name of Arny, were made to test the efficacy of sampling small areas on standard plots, instead of harvesting whole plot areas. The result was a development of the rod row method in all its various forms (Arny and Garber, [31] Arny and Steinmetz [32], Hayes and Arny [33]).

Arny and Garber took regularly arranged rows, one rod in length, within each plot. They had nine in all, and considered them in groups of 9, 5 and 4. They assessed the coefficient of variability of the data resulting from these groupings, and concluded that the nine row sample provided a sufficiently accurate estimate of yield. Later, using square yard samples, Arny and Steinmetz compared the error of yield calculated from a few combined units with that of the main 1/10 acre plots. Although they found it greater for the square yard aggregates, they presumed that by increasing the number of units in the aggregates they could ultimately obtain errors smaller than those from the whole plots. "Student" [34] pointed out the grossly fallacious reasoning employed in this argument, for, reduced to its logical conclusion, it meant that when all but a fraction of the plot was used the errors would be many times less than using the plot itself. Arny and Steinmetz had assumed that increase of size of plot would reduce error in the proportion of $1/\sqrt{n}$ where n was the number of aggregated units, *i.e.*, that increase in size was as effective as increase in replication. On the contrary, because the variability is not randomly distributed there is a high correlation between adjacent plot yields and the correct reduction is given by the formula $\sqrt{\frac{1+(n-1)r}{n}}$.

where r is the correlation between the square yards in the composite plot.

The whole subject of systematic variation in soil heterogeneity was clarified by the work of Harris [35]. He showed that adjacent plots were more likely to yield alike than similar plots scattered about an area, and for a variety of crops he calculated the correlations, calling them the coefficient of heterogeneity, from the formula

$$r_{p_1 p_2} = \left\{ \frac{S(C_n^2) - S(p^2)/m\ [n(n-1)]}{\sigma_p^2} \right\} - p^2$$

where C_n is the yield of large plots of p units.

m the number of large plots considered

σ_p the σ of the ultimate in units.

S is the summation symbol.

The existence of this correlation is ubiquitous. Maskell [36] has tabulated its value for a number of crops, and for the soil characteristics of moisture content, carbon and nitrogen content. His values range from $+\cdot317$ to $+\cdot830$. Harris' work gravely discredited much of the technique used in connection with rod row estimates of yield, and not until Clapham [37] with the advantages of later researches at his disposal re-investiaged its possibilities were the advantages of sampling yields restored to favour.

BORDER EFFECTS

Although check plot and rod row methods contained inherently grave defects it would be untrue to suggest that their development contributed little to the advance of experimental technique. They threw into high relief the problems of soil heterogeneity, plot shape and plot size, and if for a time no satisfactory synthesis could be built up from the manifold analytical data they produced, still they played an important part in preparing the way. One important contribution was the study of border effects. The general acceptance of small plots made the unrepresentative character of the outside rows a serious matter. With varietal grain trials, overshadowing of short strawed crops by long ones proved to be a source of interference, and the greater root range at the edge of uncropped paths sometimes gave certain varieties special advantages. Arny and Hayes [38] found that by discarding two border rows the rank of their varietal yields was in some cases changed significantly. Arny [39] adopted the system of cropping the alleys, without however including those rows in yield measurements. Where winter wheat was cropped with alleys of spring wheat, the effect was satisfactory, but not when oats or barley was used as the dividing crop. The only satisfactory way of avoiding border effect whether from uncropped alleys, or different rooting capacity of alley crop and main crops was to discard outside rows. In many experiments this meant planting as many guard rows as yield rows. Arny [40] recommended sacrificing three rows.

The trouble with border effects was accentuated by the trend of opinion regarding plot shape. Long narrow plots were convenient for management, but gave a disproportionate perimeter for a given area. They were also favourably looked upon from the point of view of error reduction, though there was some conflict of evidence. Love [41] Spragg [42] and Summerby [43] in their researches into technique supported long plots. Day [44] reached the heart of the problem when he brought in the consideration of the direction of the fertility slope. He preferred a square plot to a long one having its greatest dimension in the direction of least soil variation, and only recommended the strip plot where the trial plots lay side by side in the direction of least fertility gradient. Summerby's (*loc. cit.*) decision that length decreased error more than increase in width is placed under suspicion because he omitted any reference to the fertility slopes he was dealing with. Later with the advantage of a better statistical technique Christidis [45] showed that theoretical consideration in favour of long plots was fully borne out by practical examples, but he did not consider the complications of border effects.

There is much repetition in the American work but its cumulative effect was to emphasise the importance of replication of small plots. This brought in its wake other sources of error not recognised at first but subsequently brought to light and remedied. The fact of soil heterogeneity was readily admitted, and overshadowed every other consideration, but though Mercer and Hall had pointed the way, the technique that had grown out of their work in the second decade of the century was satisfactory neither to the statistician nor the experimentalist.

THE MODERN STATISTICAL PERIOD

The modern statistical period does not fall neatly into a clearly defined section of time. In sketching out the chronological sequence of events, some of these have only a temporary significance, whilst others, not at the time fully appreciated, fall into a truer perspective later. This has been so in the realm of experimental technique, so that in studying the modern period it is more advantageous to regard the subject from the point of view of the unfolding or fruition of a method, without undue adherence to the actual sequence of events.

STATISTICAL METHODS RELATED TO DESIGN OF EXPERIMENTS

What distinguishes this period from its forerunner is the fact that, previously, statistical conceptions had been applied to field experiments as an accessory. Statistical methods had been elaborated outside the sphere of agricultural science. When use was made of them it was as a last stage in the development of the experimental plan. It might be true to say that no experiment was complete without them, and that every agronomist was expected to make use of them, but the statistical procedure was something apart, a veneer that gave a polish, not something that penetrated to the very core.

The new viewpoint puts statistical procedure in an entirely different place. From the very first stages of an experimental project, the statistical method has to be borne in mind; the mathematical process must fit the data that are to be gathered, and the experimental design must be in accordance with the statistical procedure.

There is an element of danger in this development inasmuch as it can easily lead to statistical method becoming an end in itself. Statistics is an ancillary weapon, but under the new auspices it has become a weapon that has a rightful place in a permanent armoury.

The modern viewpoint then demands three things; a valid statistical procedure, a design that conforms to this procedure;

and a criterion of significance of results as a basis of interpretation. It is remarkable that at the time of Mercer and Hall's fundamental researches these three desiderata were in the main fulfilled by Student.

STUDENT'S METHOD

In 1908 Student had taken the first step towards solving the problem of the application of the theory of probability to small samples [46]. What in effect he did was to investigate the standard deviation of a population of which he had only a small sample. The true standard deviation of the population was unknown, but the experimental data provided an estimate of it. "Student" studied the distribution of this estimated standard deviation, and showed that it agreed with the true distribution for those statistics generally used in comparison of means, and Fisher [47] later extended its application to cover the higher statistical derivatives.

Student thus provided a valid statistical procedure and an appropriate criterion of significance (Student's z). In an appendix to Mercer and Hall's paper (*loc. cit.*), he elaborated a field experimental design that was appropriate to his processes.

The classical example of Student's design is an experiment to test the yield capacities of two varieties of wheat. He postulated that these should be sown in the experimental field in alternate strips, allowing an equal number for each variety. Before harvesting, each of these strips was bisected longitudinally, thus doubling the original number of plots. The two outer half strips were discarded, so that from n original plots he produced n—1 comparisons for the estimate of yield differences, in place of the $\frac{n}{2}$ that were available from the original
siderations in favour of long plots was fully borne out by practical
plots. Labelling the varieties A and B the sequence of the half strips became A B B A A B B A. Now considering two adjacent half strips as a unit, this arrangement provided that if there were a steady fertility gradient across the strips neither variety would be favoured. The mean difference between A & B is calculable from the sum of the differences between A and B in each unit separately, and the only effect that systematic cross fertility gradient has on mean differences, is what is encountered within the narrow confines of the double half strip unit.

In calculating the errors of the experiment, the differences between varieties A and B for each unit were tabulated and the standard error of the mean difference evaluated from these data. The striking accomplishment of this design was that mean differences, from which fertility or positional advantages had been largely eliminated, were interpreted with reference to an error from which the disturbing influence of these same positional factors had been consciously removed also. This was the first occasion on which the correlation between nearly adjacent plots was turned to advantage by using it to overcome the fertility gradient that had been such a handicap to the American workers.

Advantage was taken of Student's arrangement for comparing two varieties by Beaven [48] who in his half-drill strip method ingeniously arranged that the drilling of the two varieties in the required order should be done by the same drill. The seed hopper was divided into two compartments served by the same number of coulters, one division for each variety. The sowing of the plots thus became a simple continuous operation, since on turning the drill the order of the varieties was reversed. Beaven used this plan in 1920 and it was adopted as a standard procedure by the National Institute of Agricultural Botany in 1923. In the same year it found in America an advocate in Love [49] who remarks that "this method has been little used in interpreting experimental work. Probably it has been overlooked".

It is convenient to mention here that when Student's work received due recognition, the statistical processes used in his original exposition were frequently divorced from the niceties of his design. The statistical calculation became known as "Student's method" which is somewhat of a misnomer. With popularity came misconception, examples of which will be discussed later.

THE ANALYSIS OF VARIANCE

For a single comparison the "method" is simplicity itself, but in applying it to trials involving a relatively large number of comparisons the complexity is vastly increased. Student [34] extended his investigations to a trial of eight varieties replicated twentyfold in a regular chess board arrangement, and, in order to derive an error common to the whole experiment, he estimated the error of the mean difference for every combination of possible pairs, 28 in all. In seeking to simplify the arithmetical calculation he divided the variation encountered within the experiment into a number of parts due respectively to differences in variety, differences in position of his replicated groups of varieties on variable blocks of land, and finally casual variation or error.

At the same time, and independently, Fisher had seen the possibilities of subdividing variation in the same way. The simplicity of Fisher's formula and working have resulted in his method superseding Student's, so that consideration can be confined to the former's statement of the problem.

The Analysis of Variance, to give it the name that Fisher suggested, postulates that if a number of data are arranged in a number of classes containing an equal number of observations in each class, then the total variance about the general mean is divisible into two parts; that which is represented by sums of squares of deviations of individuals in the groups from their own group mean, and that represented by the deviations of the group means from the general mean.

If there are n^1 classes and each has in it k individuals then from the point of view of sums of squares of deviations the following identity holds good [50].

$$\overset{kn^1}{\underset{1}{S}}(x-\bar{x})^2 = k\overset{n^1}{\underset{1}{S}}(\bar{x}_p-\bar{x})^2 + \overset{kn^1}{\underset{1}{S}}(x-\bar{x}_p)^2$$

S is the summation symbol.

x is any datum.

$\bar{x}$ is the general mean.

$\bar{x}_p$ is the mean of a particular one of the n^1 groups.

From the sums of squares of deviations a mean square or variance is derived by dividing by the number of degrees of freedom, one less than the number in each variance class. This variance is the square of the standard deviation of that particular class.

The analysis may be tabulated as follows:

	Degrees of freedom	Sum of squares	Mean square
Within classes	$n^1(K-1)$	$\overset{kn^1}{\underset{1}{S}}(x-\bar{x}_p)^2$	$\overset{kn^1}{\underset{1}{S}}(x-\bar{x}_p)^2/n^1(k-1)$
Between classes	n^1-1	$k\overset{n^1}{S}(\bar{x}_p-\bar{x})^2$	$k\overset{n^1}{S}(\bar{x}_p-\bar{x})^2/n^1-1$
Total	n^1k-1	$\overset{kn^1}{\underset{1}{S}}(x-\bar{x})^2$	$\overset{kn^1}{\underset{1}{S}}(x-\bar{x})^2/n^1k-1$

All these functions are directly determinable from the original data and the mean square for within classes provides an efficient error from which to assess the reliability of differences between class means. The closer examination of the underlying statistical theory has been carried out by Irwin [51], [52].

The principle is capable of extension. A number of data may be classified in more than one way. For example, in a varietal experiment with crops, such as Student examined, the data may be classified in the form of the eight varietal means, and also in the form of the twenty means of the replicate blocks. There is an each to each correspondence; all varieties occur in all replications; each replication contains all varieties. Reduced to the appropriate standards such an experiment analyses as follows:

	Degrees of freedom	Sums of squares
Between varieties	7	a
Between blocks	19	b
Remainder	133	c
Total	159	d

The remainder, which is in reality an interaction between varieties and blocks, measures the deviations from the assumption that yield of any plot can be expressed as the sum of an intrinsic varietal factor and an intrinsic block fertility factor.

It is therefore a measure of consistency in varietal response, and is the error to which varietal differences can be referred. It can be calculated by difference when the other three values have been calculated directly.

The advantages of this method are many and varied. First in order of importance from the point of view of field experiments is the control of local fertility differences. The arrangement of chess board experiments into blocks each containing a member of all varieties or treatments to be tested, makes *ipso facto* a distinction between the degree of soil variation to which varieties in the same block or in different blocks is subject. If the replicates of a single variety are used as the basis for error evaluation, the resultant error figure will be affected by gross variation from block to block. The comparison of two varieties for which such error figures are available will be further complicated by the fact that using the formula

$$\sigma^2_{A-B} = \sigma^2_{A} + \sigma^2_{B}$$

the independance of the variables is assumed. Now the mean difference between two varieties is calculable from the sum of the individual differences block by block, and these are unaffected by the gross block differences. Consequently, by this method which was at one time commonly used, mean differences, into which block variation does not enter, are judged by reference to an error into which block variation does enter. By taking precautions to plan the experimental layout and the statistical procedure in concert, this anomaly can be avoided. The resultant gain in precision when gross block variance is eliminated before making the estimate of error has been demonstrated by Eden and Fisher [53]. In their experiments, examples were obtained of a more than fourfold increase in precision.

The analysis of variance, used with adequately designed experiments, provides another advantage. When systematic variance due to treatment and position have been eliminated, the remaining variance contributions between parallel plots can be pooled irrespective of the treatment or block from which they ultimately derive This gives a greater number of degrees of freedom for error estimation, and hence increased accuracy; and it provides one error figure relevant to any treatment or varietal comparison included in the experimental design.

One of the first attempts to eliminate positional variance was based upon a design called the Latin Square [54]. In this the number of replications equalled the number of treatments. The actual layout as shown in the accompanying diagram provided that each row and each column contained each of the treatments. Under these circumstances comparison of means of rows and means of columns represented positional variation unaffected by treatments. It was thus possible to eliminate gross positional variance in two directions at right angles to one another, so that the untoward effect of fertility gradient in any direction was largely eliminated from both means and errors.

The analysis of variance has so far been viewed in the light of developments from Student's position. There was one fundamental difference that Fisher insisted on, that within replications, the order of treatments should be random. Where this is impossible as in the Latin Square, a large element of randomness can still be obtained, any given square being a random choice from the total possible number. Yates [55] has enumerated all the possible types of square up to 6 x 6. Tedin [56] has confirmed the necessity of randomisation by showing that systematic or planned arrangements affect error values to a sufficient extent to invalidate their use in subsequent interpretation.

In concluding this exposition of the use of the analysis of variance for control of local soil variation, it is of interest to set out the scheme of Student's original strip experiment and to interpret it in terms of Fisher's analysis.

Let us take two varieties A and B replicated twenty times. Student's procedure was to set down the yields and to analyse them as follows:

		Difference	Deviation rom mean	Deviation squared
A	B	A − B		
a_1	b_1	$a_1 - b_1$	d_1	$d_1{}^2$
a_2	b_2	$a_2 - b_2$	d_2	d^2_2
.	.			
.	.			
a_{20}	b_{20}	$a_{20} - b_{20}$	d_{20}	$d^2{}_{20}$
		Total		Sum
		Mean		

It is seen that there were twenty squares to sum providing nineteen degrees of freedom.

The Fisherian analysis would be of the form

	df.	S.S.
Between varieties A & B	1	a
Between replications	19	b
Error. Interaction $V \times R$.	19	c
Total	39	d

Student's error figure for the S.E. of the mean difference is identical with that derived from the interaction of variety and replication c when the latter is reduced to terms of a difference between two means of twenty. The formulae are

$$\begin{array}{cc} \text{Student} & \text{Fisher} \\ \dfrac{\text{Difference S.S.}}{19 \times 20} = & \dfrac{2 \text{ x Interaction S.S.}}{19 \times 20} \end{array}$$

If Student's strips had been arranged in random order with respect to position of A and B within the replication unit, his error valuation would have provided the first example of a randomised block experiment interpreted on the lines of modern statistical methods.

MISAPPLICATIONS OF STUDENT'S METHOD

In America particularly, there grew up a tendency to believe that Student's method was applicable to all types of data and that the interpretation of field experiments could be based entirely upon examination in that way. This rigid adherence to the method still holds in some modern experiments, and, with the parallel process of the analysis of variance at our disposal, it is possible to point out and resolve some of the difficulties that have cropped up.

Salmon's [57] criticism of the method as used in America is acute and generally reasonable, though his papers on the whole give the impression that he deprecates statistical analyses in general. Arguing from single plots devoted to two different treatments over a period of years, he was the first to suggest that "Student's method" often gave an interpretation in conflict with common-sense. Thus a series of A-B differences might all be positive, and yet because of their marked variability, the variance of the difference might be high enough to discount the reality of the mean difference. The trouble in such cases is that the variance so estimated may be due either to random causes or to differential response of treatments in a number of different years. In the absence of a more complete analysis, to which in any case a single plot experiment does not lend itself, the method must often give negative results, not because it is faulty, but because it is not designed to answer the detailed question that has sometimes been asked of it. In a replicated experiment it would be possible to compare this differential variance (or interaction) between season and treatment with casual errors. In many cases the result is negative because, admitting that seasonal differences in response may be likely, the tenure of the experiment is not long enough to decide with accuracy whether, on the whole, one treatment turns out better than another. This is a difficulty that the original Rothamsted experiments with their long duration completely avoided. Salmon also pointed out that a significant result obtained by this method could be validly accounted for on the assumption that the tract of land relegated to one of the treatments was intrinsically better than that allotted to the other. Again in the absence of replications the matter is left in doubt.

Salmon dealing with experiments of the single plot per treatment type could give no formal demonstration of his contention, but the point is well illustrated by data from the Dutch East Indies on both rubber and tea experiments. Barclay and Grantham [58] have recently shown that, for rubber, monthly yield differences between two treatments when used as a basis for calculating error by "Student's method" give lower error figures than when the error is derived from an analysis of replicates. They rightly say that the fact that an error figure assessed by a method which gives a lower value than an alternative method, is not necessarily an indication of its validity; but they do not show how the two error figures are related. The following example is based upon figures given by Prillwitz [59] for a uniformity trial on tea. Two hypothetical treatments A and B were replicated sixfold in pairs. Monthly yields were taken for a period of six months, and an error calculated on the basis of monthly differences of the totals of the six replications.

The full analysis is as follows:

	df.	S.S.	M.S.
Treatment (Hypothetical)	1	2·7	2·7
Blocks	5	7282·8	1456·56
Months	5	20767·5	4153·50
Treatment × Blocks	5	1910·0	382·00
Treatment × months	5	345·6	69·12
Blocks × months	25	4570·2	182·81
Treatment × blocks × months	25	3346·7	133·87
Total	71	38225·5	

As calculated from the monthly totals the error mean square is 69·12, and represents the interaction of treatment and months. This is a much lower figure than 382·00 which is the interaction of treatment and blocks. The latter is the valid error for comparing treatment differences as a whole, and its use prevents the misinterpretation noted by Salmon, due to there being intrinsic differences between the plot or plots assigned to one treatment, that have nothing to do with treatment. The other figure merely provides a test for the consistency of monthly yields. Bearing in mind the fact that tea is a perennial crop, it is not surprising to find that a given bush or plot behaves similarly in successive months irrespective of treatment. A similar analysis in Eden's [60] data again shows a much smaller time interaction than block interaction.

SEASONAL VARIABILITY

The misuse of statistical interpretation that has just been cited arose from a failure to distinguish between what may be called time or seasonal errors, and those due to soil heterogeneity. Engledow and Yule [61] considered the separation of these effects when the results from a number of varietal trials over a series of years were pooled, or when several sowings in the same year were amalgamated. They do not give a fully worked out example, but the following hypothetical case shows the principle. Supposing that there were three varieties replicated eight times in each of five years (not making use of the same plot sites) a total of 120 plots is given. For the three possible comparisons between the three varieties, (*e.g.*, for varieties A,B,C, these are A – B, A – C, C – A) a sum of squares of deviations is obtained by treating all replicates as of equal standing, irrespective of year, and using the mean for the whole period. This error variance is designated $d\,\sigma^2\,w+g$ and embodies variation due to both weather or season, and ground or soil. A further sum of squares taken from the mean of each year separately and summed over the three years is used for estimating d σ^2 g the variance for ground only.

From the relationship

$$d\sigma^2_w = d\sigma^2_{w+g} - d\sigma^2_g$$

an error for weather is arrived at.

Looking at the problem from the standpoint of the analysis of variance the groups into which the variance divides are:

	df.
Varieties	2
Blocks	35
Years	4
Varieties x years	8
Remainder	70
Total	119

The difference between the *sums of squares* estimated according to the scheme outlined above would give a sum of squares combining differences between years, and interaction between varieties and years. But to compare differences between varieties over the three-year period with an "error" for weather it is appropriate only to use the variance for the variety-year interaction, for, according to the design of the experiment, every variety being tested in each year, the gross effect of year is removed.

Engledow and Yule's work emphasized the fact that an experiment done in one place or one season has a limited inductive basis, and that before the general superiority of a variety or treatment is assumed, treatments must be tried under a number of different climatic conditions. The analysis of variance affords a simple and straightforward way of doing this by means of standard experiments carried out over series of years. The general consistency of varietal behaviour can then be judged in conjunction with the appropriate error variance (variety x year) and further the significance of any apparent differential response of varieties to different climates can be tested by comparing (variety x year) with the higher order interaction of the remainder.

(To be continued).