STATISTICAL AND SOCIAL INQUIRY SOCIETY OF IRELAND.

THE USE OF STATISTICAL METHODS IN AGRICULTURAL EXPERIMENTS.

By F. P. HUSSEY, B.Agr.Sc., B.Sc. (Econ.)

(*Read on Friday, 20th November,* 1936.)

There is nothing new in the idea of agricultural experiments. Modern agriculture is largely the result of successful experiment, and the student of agricultural history finds that the facts of one generation are the fruits of the experiments of previous ones. Our own industrious Department of Agriculture has for over 30 years carried on a country-wide series of experiments on all types of crops and stock and every phase of agricultural effort. It is not generally appreciated what an enormous amount of agricultural experience has thereby been accumulated; few countries can boast that their agriculture is so well or so widely charted. It is not, however, with such schemes of multiple experiments that this paper is intended to deal; it may be remarked, however, in passing that from the broad point of arriving at a result from sheer weight of evidence the simple system of trials generally practised in these experiments is perfectly satisfactory. A comparison of different treatments in side-by-side plots in one hundred different centres is bound to provide at least interesting evidence of their relative merits under differing conditions.

The agriculturist faced with a new problem cannot, however, take this refuge in numbers. His object is to obtain as much and as reliable evidence as possible in the limited amount of land or material at his disposal. He cannot, of course, however imposing his results, extend them to cover any conditions except those in which they were collected. He should, however, ensure that if in these very conditions a result is obtainable, the design of his experiment is such that it can be readily and validly obtained. It is a common disappointment to statisticians to find sets of statistics which, however honestly they may have been collected, fail to yield as complete and as valid results as could have been obtained with the exercise of a little more forethought in their collection. In agricultural experiments the value of the evidence obtained depends on the design of the experiment; the statistical technique by which the results are to be interpreted is governed by the nature of its lay-out. If statistical methods are to be used in the investigation, more thought is necessary before the experiment than after it.

If a farm crop, for instance, be sown in a plot of, say, one-fortieth acre, what value is to be placed on the result of the trial? The yield is influenced by rain, sunshine, sowing, manuring and a number of environmental factors which need not be stressed. Assuming these, however, as standard conditions of the trial—would the same result be achieved if the trial could be repeated in similar conditions? No one familiar either with farming or with farm experiments would expect such a coincidence; there is too large a body of random experimental errors such as faults in measurements, in seeding, damage by pests, birds, animals, likely to affect the precision of the final yield. Some of this error is, of course, remediable, and care in the measurement, treatment, and lay-out of a plot naturally enhances the reliability of the final yield, but there remains a core of error ordinarily incommensurable which must make even the most carefully obtained yield merely an estimate of the " true " yield of the plot. The final figure for any trial plot must indeed be regarded as a sample from an infinite population of yields, obtainable from the plot under the environmental conditions of the trial. The conception of such an infinite population of yields is in itself of little value but it does stress the approximate nature of any individual yield, which approaches the " true " yield according as the random error is lessened. The reduction of this random error is, of course, the first care of the agricultural investigator.

One important source of variation in field experiments calls for separate mention. That is the variation due to soil inequalities. " Soil heterogeneity " as it is generally known is a matter of everyday experience. It is not merely a question of fertility differences in soils of different origins. Soils of the same origin within the same field vary considerably. Definite fertility contours or fertility gradients are to be found in many fields, though the extent of the variation may be small. Regular gradation of fertility would not present an insuperable obstacle to the investigator but unfortunately variation tends to be irregular and even capricious. Uniformity trials to test the nature and extent of soil variation as reflected by the yields of crops, are recommended for any trial ground whose soil character is unknown.

One immediate result of this inequality of the medium in which trials are carried out is that inequalities of soil cannot be eliminated by enlarging the size of trial plots. It is a popular fallacy that experiments carried out in large plots are proportionately more reliable than where small plots are used. Obviously, the greater the size of the plot the greater the room for variation within the plot. To test two treatments on large contiguous plots, no matter how carefully minor errors may have been eliminated, is to widen the chance of disparity between the plots as regards inherent fertility. To widen the class-limits in a frequency distribution is to weaken the accuracy of centring the classes at the mean. Equally, to extend the size of plots under treatment is to lessen the likelihood of the two soil media being of equal fertility. For the benefit of members unacquainted with the subject, I may be forgiven for introducing the classic experiment as regards size of plots carried out by Mercer and Hall (*Jrnl. Agr. Sc.,* 1911. iv. 107-132). These workers harvested carefully the yield of a measured uniform acre of wheat which had been sub-divided into 500 small plots of equal area. By classifying these small plot yields in a frequency distribution a range of variation from 2·7 to 5·2 lb. per plot was exhibited. The standard deviation—the root-mean-square deviation, the usual statistical measure of variation —estimated from the grouped totals of single plots was 11·7 per cent. of the mean of a single plot. By grouping the small plots into large blocks of plots the standard deviation was reduced to 6·3 per cent. for blocks of 10 plots, to 5·7 per cent. for blocks of 20 plots and to 5·1 per cent. for blocks of 50 plots. Reduction of the variation was brought about by increasing the size of plots but the extent of reduction was diminished considerably and the authors arrived at the opinion, now generally accepted, that there is little to be gained in precision by increasing the size of the plot above, say, 1/40 acre. If we regard the yields of the sub-plots as being but variates of the same hypothetical yield, we should have expected the yields of aggregated plots to have had proportionately smaller standard deviations as being larger samples whose variation would be reduced in proportion to the square root of the number of sub-plots aggregated. That of a 50 plot block would, for example, have been expected to be about $\frac{11 \cdot 7}{\sqrt{50}}$ or approximately 1·65 per cent. of the mean. That such a reduction was not effected, must be attributed to the non-independence of individual plots due to the high fertility correlation between neighbouring plots. Had the sub-plots grouped into larger plots been chosen *at random* over the experimental area, a greater reduction in the variation would have been expected. The correlation between contiguous plots, which must obviously be high no matter what the nature of fertility variation, is avoided by randomising the distribution of plots having the same treatment, and a reduction in the error may reasonably be hoped for. The error is not only likely to be reduced, but the scattering process supplies a random sample which is logically necessary if a statistical treatment based on the theory of errors is to be applied in the interpretation of the results.

The repetition involved in scattering plots over the experimental area must tend to level out the random errors, and supply a more reliable estimate of the mean value of the plot yields by reducing the error, whilst it also provides the means whereby the error involved may be calculated. Replication of plots of the same treatment, then, provides the error by which the reliability of the experimental results may be judged; randomisation ensures that the error so found will be a valid one, an assurance which no systematic arrangement of field plots could supply. Actually the basis of the methods of field experiment which it is proposed to treat of in this paper, is the random distribution of replicated plots.

The estimate of error above referred to is derived as in ordinary statistical procedure from the sum of squares of deviations from the mean with a slight but distinctive difference. If n plots of any treatment are taken, the sum of squares is divided not by n but by n-1, known as the number of " degrees of freedom " or the number of independent comparisons that can be made, independence between the n plots not being complete in that they are connected by their mean value. With this divisor, the summed squares yield the variance, whose square root is the standard deviation of the n plot yields. This is, of course, expressed in the same units as the plot yield and is the estimate of the error of the trial. Clearly, the primary aim in any field trial is the reduction of this error with the consequent increase in the reliability of the mean or aggregated yields. As previously explained, such a mean value can itself only be regarded as an estimate of the true mean, and the significance attaching to the estimate is largely governed by the size of its standard error and the validity of the method by which it was obtained.

Now, the field experimenter is not generally interested in the acquisition of a mean as an estimate of an ideal mean. Most of his interests lie in the comparisons of different mean values. But even such comparisons are of little interest unless it is possible to attach some measure of significance to differences in mean yields. Since he works with variable material in a variable medium it must be clear that no estimate for the value of a mean is actually impossible, and hence no difference however great between mean or aggregate yields can be accepted as positive evidence of actual differences, since each is itself capable of supplying estimates actually differing as widely. In short, the problem of comparisons is simply to ascertain whether two or more means are to be regarded as random estimates of the same mean, or whether the probability of their being so, is so small as to be insignificant. The experimenter adopts the "null hypothesis," as Fisher calls it, that any two or more estimates are merely random values of the same variate. It is not possible to prove a positive hypothesis of difference but it is possible to invalidate the null hypothesis to any postulated degree. If the distribution of a variable be normal it is possible to discover the odds against any particular value occurring at random. One, for instance, which would not be likely to be exceeded more than five times in one hundred, taken at random; could be considered as having a probability of 20 to 1 against its occurrence. Hence, if a difference between two means, say, is obtained which would exceed that likely to occur only five times in one hundred if both were to be considered as estimates of the same value, it is legitimate to reckon the odds against their being estimates of the same mean as being twenty-to-one. Normality cannot always be postulated particularly for small samples, but "Student" has discovered the distribution applicable to small samples and tables of the probability integral of this distribution enable the experimenter to find the probability of any particular difference occurring. A probability of 20 to 1 against is the criterion usually adopted, though the significance level of 100 to 1 against is often used when a more rigorous test is required. Differences exceeding those at these levels of probability are pronounced "significant." If, for example, two varieties of a farm crop give yields which can be considered as differing significantly, it may be understood that the two yields are unlikely to be just random values of the same yield, the measure of the unlikelihood being stated. The natural assumption is that the factor in which they differ, namely variety, is responsible for the disparity in yields. What measure of improbability we are prepared to adopt as invalidating the null hypothesis is, of course, arbitrary; however, it is usual, as stated above, to adopt either a five per cent. or one per cent. probability as sufficient.

One of the main advantages of the modern system of agricultural experiment is its capacity to supply a number of different comparisons. In a randomised experiment with a number of different "treatments"—the word is here used in the wide sense of indicating differences in variety, cultivation, seeding, manuring or in any possible particular—we are enabled to arrive at an estimate of the error involved in the experiment. It is in the light of this error that we can compare treatment means or totals with a view to ascertaining if their differences can be regarded as sensibly significant. Clearly the smaller this error is, the more important will these differences appear in proportion to the error. But the error, as has been seen, is composed of a number of elements.

Primarily, there is the error introduced by soil heterogeneity and, secondly, there is what may be termed the "inherent" error of the experiment, that random error which may be contributed to in countless ways and which no precautions however elaborate can completely remove. There may be a third component of the error, the legitimate variation which is produced by real differences between the treatments. If we could remove from the error that part due to soil differences and analyse the remainder into the part due to random error and that attributable to treatment differences, we should be able to examine these latter differences in the light of the inherent or random error segregated by the analysis. The aim of modern experimental design is to enable us to perform this analysis of the variance and to examine our treatment results in the light of a validly acquired measure of the unavoidable error of the experiment.

The elimination of that portion of the error ascribable to soil heterogeneity is effected by what is known as "local control." The two most popular methods are those known as the "Random Block" system and the "Latin Square." A description of a simple case of the former method should make clear the means by which soil heterogeneity is accounted for and later eliminated in the analysis of the variance. A number of blocks of land of equal size and shape are selected and each sub-divided into as many plots as there are treatments to be tested. Each treatment is then represented once, and once only, in each block, the allocation within the block being entirely at random. This random allocation is established separately for each block. The size of both plot and block depends principally on the nature of the crop to be tested, but experience has fixed optimum sizes (within fairly wide limits) for each type of crop. Randomisation ensures that each treatment has an equal chance of being tested on any particular plot and hence equalises the chances of its allotment to a plot of more or less than average fertility. No systematic arrangement of the plots within the block could achieve this and the consequent derivation of a valid estimate of error.

Now, since each block contains exactly the same constituent plots, that is, one plot devoted to each treatment, we should expect that if inherent errors are ignored, the variation between the block totals would solely be due to differences of fertility between the blocks. As the effect of every treatment is felt in each block the differences between block totals must, ignoring random experimental error, be accounted as due to soil differences between the blocks. The variance of the block totals, obtained as described above, may be taken as a measure of the error introduced by soil heterogeneity.

Tabulation of the results by treatments and blocks makes the next step more easy to follow. By arranging the blocks in rows (horizontally) and the treatments in columns (vertically) deviations within treatments can be readily calculated. Suppose 5 treatments, A-E, are being tested in 4 blocks, 1-4.

Block	Treatment					Block Total
	A	B	C	D	E	
1						
2						
3						
4						
Treatment Total						

For treatment A, for instance, we can see the range of yields given by the representative plot in each of the four blocks. Variations between them can be due only to two of the sources of error—soil differences and random error—since all are of the one treatment. By finding the variation within each of the treatments A, B, C, D, E, and summing them we can find the total variation within varieties, that is, the variation due to soil heterogeneity and to random error. Now, the variation due to soil heterogeneity has already been found; the part due to random errors may be found by subtraction. The third part of the error, that due to differences, if any, caused by the different treatments may be found by finding the squared deviations of the variety totals from their mean. The total variation has now been synthesised from the three contributing sources, as may be tested by finding the sum of squares of all the plots ($5 \times 4 = 20$ in the above example) from the general mean.

I am painfully aware that this explanation is far from clear and hasten to supply an example which may perhaps clarify matters. The example chosen is a simple variety test with potatoes carried out at Glasnevin. I have suppressed the more agricultural features of the test as it forms part of an experiment which is not yet concluded; in any event, the actual results are irrelevant to this paper. Nine varieties of main-crop potatoes were compared for yield. It was decided to carry out the experiment in random blocks, with six replications. The six blocks were carefully chosen, the portion of the field on which they were laid out being known not to be subject to marked soil variation. The whole area of the experiment had been under uniform treatment for some years previously in the rotation. Each plot in each block was carefully measured and differed in no apparent way from its neighbours. The nine varieties, A, B, C, D, E, F, G, H, K, were then allocated one to each block, for each block separately, the process of allocation being at random. The results of the experiment are tabulated in Table I. The random arrangement of plots within each block does not, of course, appear in the tables.

Variety	Block						Variety Total
	I	II	III	IV	V	VI	
A ...	25.4	25.0	20.1	22.0	21.6	20.3	134.4
B ...	18.4	20.4	20.7	20.0	20.6	20.2	120.3
C ...	22.4	18.1	18.6	18.9	19.7	21.1	118.8
D ...	18.8	21.4	18.5	16.8	19.0	17.5	112.0
E ...	15.9	16.6	17.5	20.9	18.0	19.2	108.1
F ...	15.5	17.8	18.0	18.7	14.9	16.6	101.5
G ...	17.2	16.5	15.5	17.2	14.7	18.1	99.2
H ...	16.9	13.7	15.4	17.3	15.3	14.2	92.8
K ...	14.1	14.4	15.6	15.5	13.3	14.5	87.4
Block Total ...	164.6	163.9	159.9	167.3	157.1	161.7	974.5

Some results are immediately apparent. The experimental area has been well chosen, soil variation being very small, there being only a maximum range of about 6 per cent. between the block totals (157·1-167·3). There do, however, seem to be marked differences between yields of the same variety in different blocks though a moment's inspection will show that the nature of the difference from block to block is not the same for each variety. Perhaps the most striking feature, however, is the marked differences existing between variety totals. Variety here, of course, is the factor in which the subjects differ and is to be considered as expressing differences in "treatment" between different plots. The whole aim of the experiment is to find to what extent these differences may be considered significant. It will, I think, be granted that even without the arithmetical processes of the analysis of variance, the very design of the experiment would allow us to regard the evidence of the variety totals as worthy of serious consideration.

The arithmetic of the analysis of variance may now be commenced. The squared deviations of the 54 plots from the general mean are first found, in this case 397·994. It may be here remarked that the apparently tedious arithmetic involved can be greatly eased by the use of a simple algebraic formula. Next the sum of squares for the "block" deviations is found from the last line of the table. It amounts to the small total of 7·325 when allowance is made for the fact that each total is itself the sum of 9 plots. The squared deviations between varieties is found from the totals (or their means) in the last column. This amounts to 292·559. We have now dissociated from the total variation that portion due to differences in position (from the block totals) and the much larger portion due to differences in variety. The remainder, 98·110, can only be ascribed to the random error of the experiment. It may also be arrived at, if a check on the work is needed, by summing the squares of the deviations within each of the nine varieties and subtracting the variation due to soil heterogeneity as explained previously.

The analysis of the variation now reads as follows:—*

		Sum of squares.
Between varieties	...	292·559
Within varieties:		
Due to position	...	7·325
„ „ error	...	98·110
	Total	397·994

It remains now to find the mean square or variance due to each of these factors. This is simply found by dividing the above sums by the number of degrees of freedom available to each. There are 54 plots giving 53 independent comparisons, hence, 53 degrees of freedom, 5 of which are contributed by the 6 blocks, 8 by the 9 varieties and the remaining 40 by the error component. The analysis may now be recast:—

* The mathematical basis of the analysis resides in the following formula (J. O. Irwin, *Supplement to the Journal of the Royal Statistical Society*, Vol. I, No. 2, 1934, p. 238):—

$$S(x_{uv}-\bar{x})^2=S(\bar{x}_{.v}-\bar{x})^2+S(\bar{x}_{u.}-\bar{x})^2+S(x_{uv}-\bar{x}_{.v}-\bar{x}_{u.}+\bar{x})^2$$

where x_{uv} is the observation in the uth column and the vth row, $\bar{x}_{u.}$ and $\bar{x}_{.v}$ the respective means of the uth column and the vth row, x the mean of all observations and S indicating summation to all the observations in the sample. The term on the left hand side gives the total sum of squares and the three terms on the right hand side, respectively, the sums of squares due to variations between (1) rows (in this example varieties) (2) columns and (3) "error."

	Degrees of freedom	Sum of squares	Variance (mean sq.)
Blocks	5	7.325	1.465
Varieties ...	8	292.559	36.5698
Error	40	98.110	2.45278
	53	397.994	40.48758

Now, the variance due to variety difference is clearly very important in comparison with that contributed by random errors. Fisher has evolved a simple test (the "z test")* by means of which these variances may be compared. On consulting his table of the distribution of z we find that the difference of these two variances is undoubtedly significant, being far larger than would occur only once in one hundred times if the two variances were to be considered as random values of the same variance. This is, of course, a result which it would scarcely need any statistical analysis to confirm but the test must be made before the variety means can be compared. The square root of the variance due to random error is obviously the standard error to be adopted as that of a single plot. From this may be calculated the standard error of a mean of 6 plots (·64) whilst for purposes of comparison the standard error of the difference of two means is calculated ·904. Approximately twice this latter error (or 3 times that of a mean of 6 plots) may be taken as the significant difference which if exceeded by any comparison of two means points to an effective difference between the two varieties.

The result of the experiment may now be tabulated:—

A	B	C	D	E	F	G	H	K	Gen. Mean	Standard Error
22.4	20.1	19.6	18.7	18.0	16.9	16.5	15.5	14.6	18.05	.64

or stating the yields and error as percentages of the general mean

A	B	C	D	E	F	G	H	K	Mean	Standard Error
124.1	111	108.6	103.4	99.8	93.7	91.6	85.7	80.7	100	3.5

Any two varieties whose yields differ by more than, in this case, about 10 per cent. may here be taken as differing significantly. The nine varieties may, therefore, be classified with some assurance in view of the relatively high differences established between certain varieties. Variety A is outstanding; it is not possible to claim definite superiority for any one of B, C, D amongst themselves, but all three are clearly more productive than the group G-K.

Another simple experimental design has been termed by Fisher the "Latin Square." It allows of a more compact system of plot arrangement and is of most value when the number of treatments to be tested

* With the notation of the previous footnote:—

$$z=\tfrac{1}{2}\log_e \frac{(r-1)S(\bar{x}_{.v}-\bar{x})^2}{S(x_{uv}-\bar{x}_{.v}-\bar{x}_{u.}-\bar{x})^2}$$

where r is the number of columns (blocks).

is from four to eight. It consists essentially of a rectangular group of plots arranged in rows and columns, there being the same number of plots in each row and column as there are treatments to be tested. The plots are again allocated at random with the restriction that each treatment must occur once, and once only, in each row and in each column. Five treatments, for instance, might conceivably be arranged as follows:—

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

It is obvious that complete randomisation in the general sense of the term is not possible with the above condition. Actually, the lay-out of any particular Latin square is selected at random from all the possible Latin squares available for the number of treatments. A simple method of transformation recorded by Yates (*Emp. Jrnl., Exp. Agr.*, 1935. 1. 235) may, however, be applied to any Latin square however derived and the resulting square will be sufficiently free from arrangement to satisfy the most exacting.

Soil heterogeneity is accounted for in two right-angled directions and is eliminated from the mean-square deviations by calculating the variation of the row and column totals separately. Treatment totals are next considered and the relevant sum of squared deviations removed as in the Random Block system. The balance of the total variation serves as a legitimate estimate of the experimental error and may be compared as before with the variation, if any, ascribable to treatment differences. This comparison is carried out exactly as before. Usually a relatively large amount of the error can be attributed to soil variation owing to the two-directional nature of the plot arrangement. Unfortunately this element of the variation is rather expensive in degrees of freedom, twice as many being sacrificed as in the case of a random block system with an equivalent number of blocks. Thus, for a Latin square of 6 treatments having 36 plots and 35 degrees of freedom, there will be six rows and six columns each claiming 5 degrees of freedom. With 5 allotted to treatments only 20 are left for the computation of the error. A random block system of 6 blocks each of 6 plots would remove only 10 degrees of freedom for position and treatments, leaving 25 degrees of freedom for the error component. For this reason Latin squares of 4 or less rows and columns are liable to leave an error component too large to show significance, owing to the low number of degrees of freedom available for dividing into the error sum of squares. This difficulty may, however, be overcome by increasing the number of Latin squares. A square in which only three treatments are under test will provide only two degrees of freedom for the error estimation; the same test with five Latin squares—individually randomised—provides 18 degrees of freedom for error. This method of multiple Latin squares is very suitable where only three or four varieties or treatments are under test, particularly in the case of cereals where relatively small plots are used.

A simple example of a 6×6 Latin square may be of interest. Six varieties of oats, the products of selections from a crossing of two well-known varieties, were on trial. The arrangement selected and the yields per plot of the six varieties (Nos. 1, 2, 3, 4, 5, 6) were :—

4 403	3 470	1 425	6 408	5 396	2 480
3 478	2 436	6 406	5 393	1 422	4 451
5 473	1 451	4 427	2 409	6 415	3 460
2 459	5 453	3 439	1 394	4 448	6 442
6 368	4 360	5 384	3 422	2 437	1 401
1 401	6 374	2 405	4 376	3 447	5 483

Variety totals :—

1	2	3	4	5	6
2,494	2,626	2,716	2,465	2,582	2,413 grammes.

The analysis of variance reads as follows :—*

Due to :	Degrees of Freedom	Sum of Squares	Mean Square
Rows	5	8,758	1,751
Columns	5	9,196	1,839
Varieties	5	10,581	2,116.2=a
Error	20	11,319	565.95=b

Ratio $a/b=3{\cdot}74$

$\frac{1}{2}\log_e 3{\cdot}74={\cdot}6595=z$

For P=·05, $n_1=5$, $n_2=20$, $z={\cdot}4986$

For P=·01, $n_1=5$, $n_2=20$, $z={\cdot}7058$

The test shows significance at the 5 per cent. level; in other words,

* The mathematical basis of the analysis is as follows (Irwin, *op. cit.*, p. 239) :—

$$S(x_{uv}-\bar{x})^2=S(\bar{x}_{u.}-\bar{x})^2+S(\bar{x}_{.v}-\bar{x})^2+S(\bar{x}_t-\bar{x})^2+S(x_{uv}-\bar{x}_{u.}-\bar{x}_{.v}-\bar{x}_t+2\bar{x})^2$$

where $\bar{x}_t$ denotes the "treatment" mean to which the observation x_{uv} belongs and the remaining notation is as given in the footnote to page 7.

the varietal difference would only be expected to occur by chance less than once in 20 trials. The z value was not sufficiently large to show significance at the 1 per cent. level, naturally a more stringent test. The lesser test is, however, generally accepted as being sufficient to warrant the extraction of the error of the experiment. For this value we naturally take that given by the error line in the analysis of variance 565·95. Standard error per plot $=\sqrt{565 \cdot 95}=23 \cdot 8$ grms. or approximately 5·6 per cent. of the mean value of a single plot. Standard error of a group of six plots$=\sqrt{565 \cdot 95 \times 6}=58 \cdot 3$ or 2·3 per cent. of the mean. It is customary to regard a difference equal to about three times the standard error as sufficient to indicate significance. As this paper is merely a general description of methods employed and not a statistical treatise, no excuse is offered for omitting the mathematical principle by which this figure is arrived at. The significant difference is, therefore, about 171 grms.

The table of variety totals may now be arranged :—

Variety	1	2	3	4	5	6
Total	2,492	2,626	2,716	2,465	2,582	grms. 2,413
Diff. from Variety 1 ...	0	+132	+222	−29	+88	−81

The result may be stated as showing that variety 3 is significantly better than the "control"—variety 1—and varieties 4 and 6. Variety 2 shows marked superiority to variety 6 and probably is superior to variety 4. Other comparisons merit the judgment—"not proven."

An interesting feature of the experiment from which the above figures are extracted is that it is one of two replicated Latin squares laid down to test the yield behaviour of these six varieties. Although the other Latin square failed to show significance by the "z" test, only a surprisingly small amount of the error being removed by the row and column deviations, the two squares in combination showed significance at the 1 per cent. level. It may be noted that the more effective Latin square illustrated here did not in itself prove capable of showing this measure of significance. In the combined test, varieties 2 and 3 were shown to be significantly superior to the other four varieties, of which variety 6 enjoyed the distinction of being proved inferior to each of the other five. The effectiveness of the combined experiment may be ascribed to the large share of the error eliminated in the comparison of the two Latin squares themselves, and the increase in the number of degrees of freedom for the error component from 20 to 45 which reduced the error mean square to relatively small proportions.

The practical value of this test, carried out on quite small plots, as is evident from the totals, may be of interest. Variety 1, a popular and widely-grown oat, was one of the two parents used in the cross, with another variety bred by my colleague, Mr. Caffrey, who kindly supplied me with the above figures. Varieties 2-6 were different selections from this cross and it may be understood how the findings of this and similar trials were of considerable value to him as indicating which selections were worth persevering with from the standpoint of yield.

The two examples shown presented one simple question for Nature to answer. In the experiment with potatoes, for example, a straightforward answer is sought to the question of yield differences of varieties. Such varietal differences have been established, but it should clearly be understood that the results of the test apply only to the conditions under which it was carried out. It is conceivable that important alterations in the above results might follow changes in either environmental or cultural factors. In a rainy summer or on a light soil (the soil at Glasnevin is a rather heavy clay-loam) comparable results might not be obtained, and no extension to cover these conditions can justifiably be made from the experiment. Differences in cultivation, such as altering the width of drills, the size of seed or the spacing between sets, might result in a different set of yields and possibly a different order in the ranking of the varieties. If it was desired to test these factors singly, only one factor being altered in each complete experiment, it would necessitate a tedious repetition of experiments. It is, perhaps, the greatest advantage of the modern experimental technique that it enables us to design experiments in which the effects of more than one factor may be investigated. These multiple factor trials enable the investigator to obtain answers to a number of questions within the confines of a single experiment. Many examples of this method could be cited but one somewhat relevant to the points just mentioned will suffice. In the *Journal of Agricultural Science* (1935. 25. 297-313) Bates describes an experiment with potatoes in which four different spacings and three different sizes of seed were used, within the one experiment, with the object of finding the effect of varying these factors on the size of the resulting tubers. A somewhat simpler experiment involving two factors was carried out at Glasnevin on the growing of onions. The commercial growing of this vegetable has lately been receiving attention and it was decided that a small-scale trial of different varieties would be of interest. At the same time some information was sought on the question of time and nature of sowing. An experiment was designed to supply an answer to these questions. Six well-known commercial varieties were selected. These were tested in random blocks. Four blocks of six plots each were carefully prepared and the six varieties were allocated at random for each block. The method of allocation may be worth mentioning. A number of packs of clean playing cards were available. Ten each of aces, twos, threes, fours, fives and sixes were taken, all the 60 cards being of equal size and free from cracks, tears or any feature which might lead to one card being selected rather than another. These were carefully shuffled by an assistant whose proficiency could only have been the result of a mis-spent youth. The six varieties were allotted numbers by placing the names in a hat and noting the order of their withdrawal. The cards were shuffled, cut and reshuffled 24 times, the cut card being reinserted after each cut. The allotment of varieties within the blocks was carefully noted. Each plot was now subdivided into two equal sub-plots. The allocation of the two methods of sowing to be compared—outdoors in spring and under glass in autumn—was made separately for each of the 24 variety plots by tossing a coin. The plots and sub-plots were then sown as the randomisation directed and the whole experiment received the recognised horticultural treatment for the crop during the growing season. To ensure that contiguous plots should not be adversely affected by each other, an unsown area was left between sub-plots and between plots. Further to eliminate this "edge effect" the border row right round each plot was rejected at the time of harvesting. This also precludes such adventitious effects as trampling and mechanical injury from interfering with the final yields. The 48 sub-plots were all harvested on the same day and weighed. The final weighing took place some weeks later when the bulbs were ripened.

The analysis of variance for the experiment was:—

Due to:	Degrees of Freedom	Mean Square
Blocks	3	669.6
Varieties ...	5	2715.0
Error	15	103.9
Time of Sowing	1	4070.1
Interaction ...	5	271.9
Error ...	18	67.0
	47	

'z' for varieties ... ... 1.6287; for $P = .01$ $z = .7582$
'z' for time of sowing ... 2.0533; for $P = .01$ $z = 1.0572$

Definite variety differences have been established and a strongly significant difference between indoor and outdoor sowing. Two other features of the experiment call for comment. Block differences showed significance at the 5 per cent. level, due entirely to unexpectedly low yields in Block 4. "Interaction" is appreciably significant at the 5 per cent. level. This portion of the error is that due to the "interaction" of the two factors under test, being ascribable to variations resulting from the combined effect of variety and time of sowing. It was found, for example, that some of the varieties responded badly to autumn sowing. The set-back to these varieties naturally had a depressing effect both on the totals for autumn-sown plots and on the totals for the varieties affected. This part of the variation could not, however, be associated definitely with either variety or sowing without biassing these factors and this portion of the mean square deviation is eliminated to allow of a more effective comparison between these two factors singly. It is worth noting that this small trial provides 24 comparisons for time of sowing alone. The two error residuals are a feature of this split-plot treatment and are used for the two different comparisons of variety and sowing.

It has been possible to mention only the simplest applications of the modern technique to field experiments. Its application to experiments with live stock has received considerable attention and an appreciable measure of success, though the difficulties involved can be appreciated. The improvement in the "co-variance" method by which trials with more than one variable may be carried out has enormously increased the possibilities of modern experimental methods in all forms of agricultural experiment. A simple statement of the type of problem whose solution is made possible by the co-variance method must suffice. In a cereal trial there is the possibility that yields may be affected by plant number, and where treatments are to be compared the correction of the final yields for this variable is often required. The co-variance method allows for the equalisation of this secondary variable in the comparison of treatment means. It has become a very valuable tool to the agricultural investigator—though rather a difficult one for the non-mathematical investigator to feel entirely comfortable with.

There remains one problem with which every worker is faced and of which some mention must be made. This is the problem of sampling. It is a mistake to imagine that the agricultural investigator straightens his back after sowing, with the comforting thought that he need not revisit his experiment until the harvest. Constant attention to his trial plots during the growing season and in some cases even after harvest is imperative if the points of any trial are to be appreciated. The time of appearance, number of tillers (in corn crops), percentage stand, reaction to weather and disease, time of ripening, are all points of first-rate importance from the farming viewpoint and must be closely studied during the progress of the experiment. Qualitative data on many of these features may be just as interesting as the quantitative data of the final result. The problem of sampling, however, is encountered if it is proposed to substitute quantitative results for approximate opinions on such points. It may not be possible to obtain the requisite data for the whole of each plot or treatment—apart from the tediousness of the task, the crop is not likely to be improved by interim interference. How are representative values to be obtained? Briefly, replicate randomisation is practised to secure a sample which is free from the charge of bias. The main sample from each plot is sub-divided into sampling units, the number of units depending on the nature of the crop. The units themselves may be further sub-divided if it is thought necessary, to give a properly representative sample. The solution of these sampling problems calls for much thought and a sound acquaintance with the habits and nature of the crop under test. In the case of root crops, for instance, where such characters as sugar content, percentage dry matter, or weight of starch is required, it is advisable to select the roots for test by some random method before the actual harvest, as the physical nature of the roots would make it difficult to choose without bias a representative sample from the harvested crop. Sampling problems are a constant worry to the investigator as the necessity for obtaining representative data from relatively small samples, calls for the most meticulous care to prevent any suspicion of bias in their collection.

I cannot help feeling that I have subjected the members of the Society to an unwanted class in arithmetic, and fear that I have invested a simple experimental technique with the appearance of a complex problem in mathematics. May I conclude by stressing the fact that the test is the real object of the experiment and that the arithmetic is merely incidental? It does, however, enable the experimenter to give to his results a sanction which is usually missing from the older haphazard arrangements and, in this respect, merits the consideration of every agriculturist faced with the carrying out of farm experiments. If the agricultural wood tends occasionally to be obscured by the mathematical trees so that the non-mathematical reader of agricultural literature tends to restrict his reading to the summary, it is to be hoped that as the effectiveness of the method becomes more widely known, the arithmetic may lose much of its terrors and the conclusions receive the acceptance which the care in their preparation has merited.

Printed by CAHILL & CO., LTD., *Parkgate Printing Works, Dublin.*

DISCUSSION.

Mr. Brady: I think Mr. Hussey has given us a very clear and interesting account of the value of modern statistical methods to the agricultural investigator. The paper is, I understand, the first of its kind read before this Society, and I trust that it will stimulate a wider interest in a very interesting and important subject. In this country agriculture is our chief industry and agricultural research is a necessity if the industry is to progress. The success of agricultural research depends to a considerable extent on the intelligent use of statistical methods. Hence, I think, there is very good reason why we should pay particular attention to the study of these methods. In plant breeding—work which is of great national importance—the types of experimental layout and the appropriate analytical methods described by Mr. Hussey have played an important part in the production of improved varieties of cereals.

The experimental results quoted in the paper are very interesting. I refer especially to the experiment designed to test the yielding capacities of six cereal varieties. In one Latin Square a varietal difference at the 5 per cent. point was established. The other Latin Square referred to, and which contained the same varieties, showed no varietal difference. When, however, both Latin Squares were combined a varietal difference at the 1 per cent. point was established. This interesting result indicates clearly (1) the need for adequate replication in order to get reliable results, and (2) the increased precision which it is possible to get by increased replication when this is done in accordance with modern methods of layout. Replication, as practised before the introduction of the methods described in the paper, did not necessarily mean that the experimental error would be reduced; because the greater the area taken up with plots, the greater was the chance of encountering soil variations over which one had no control. But now, since it is possible to eliminate the differences in the aggregate of rows, columns, and even entire Latin Squares, it is only the soil variation within these units which contributes to experimental error. Replication, therefore, can be relied on to give increased precision to the experiment, and there is no reason, from a statistical viewpoint, why experiments with similar plans could not be carried out simultaneously at a number of centres throughout the country and the resulting data summarised in a single table showing the analysis of the variance.

With regard to the variety test with potatoes referred to in the paper, I agree that the experimental area has been well chosen since the experimental error is relatively small. From the context, however, one might get an impression which I am sure Mr. Hussey never intended to convey, viz., that the narrowness of the range between block totals gave a criterion of the suitability of the experimental area. This is not so, and it is easy to visualise a set of circumstances where a very badly chosen area would result in very little "between block" variation. It should be the aim of the experimenter to get the greatest possible uniformity within blocks; the extent of the variation in soil fertility between blocks does not matter since it will be eliminated in the analysis.

In my estimation the paper is a very excellent one, and it is with great pleasure that I propose a very hearty vote of thanks to Mr. Hussey.

Mr. Geary, seconding the vote of thanks, said that in most statistical societies it was traditional that there should be one paper in each session completely incomprehensible to the great majority of members; and it was right that our Society, as one of the oldest in the world, should be true to this tradition. It was perhaps not surprising that Mr. Hussey's paper could not be followed in all its details by most members, but at least they could appreciate the very practical importance of the problem and the significance of the results. This was precisely the kind of paper which the Society requires.

In principle the method of approach in sampling problems is nearly always the same: from the necessarily limited data certain statistics are computed; if there is no significant difference in the data, i.e., if the differences between the measurements made could be regarded as due solely to chance, then it is unlikely that these computed statistics will exceed a certain quantity which may be ascertained in advance. We cannot be certain from the limited sample that the data reveal real differences, but we can say something like this: "there may be no real differences (in crop yields, etc.), but the odds against this being so are 100 to 1 against."

The method known as the "analysis of variance," which Mr. Hussey has utilised, is due principally to Professor R. A. Fisher, who in recent years developed certain results of Mr. W. S. Gosset, better known to statisticians the world over under the pseudonym of "Student," who has lived and worked for the past thirty years in Dublin. Mr. Gosset amuses his friends when he says that he is "no mathematician," for his discoveries of the frequency distribution of the variance of normal samples and of the ratio of the mean to the standard deviation (the square root of the variance) are generally regarded as the most important contribution to the theory of statistics in the last half century. It should be emphasised that implicit in the method which Mr. Hussey has used is the assumption that the "universes" of the yield of crops, using the same treatments and generally grown under the same conditions (so that the differences between individual observations are due to the multitude of uncontrollable causes which go by the name of chance), are normally distributed. Quite lately tests have been developed for determining whether samples of given number may reasonably be regarded as having been drawn at random from a normal universe, and it is strongly recommended that before applying the " Student "-Fisher theory these tests,* which are very simple, should be carried out.

Considerable care must be exercised in interpreting the significance of differences between any two varieties when more than two varieties are under test. For instance, if only two varieties are under test, and if the difference in the means is in excess of twice the standard error, then it is safe to say that the difference is significant, because the odds against such a difference being due to chance are something like 20 to 1 against. If, however, there are ten varieties under test, of which the difference between the highest and lowest means is in excess of twice the standard error, a corresponding deduction above the significance of the difference between the varieties concerned is by no means so safe. If, in fact, the ten varieties were identical, the odds against the highest and lowest observations being as much as twice the standard error are only about 2 to 1 against. From such considerations some little doubt attaches to the validity of Mr. Hussey's deductions on page 8.

Miss Beere said that the kind of work Mr. Hussey was doing was of enormous value to the State. Highly trained agriculturists were by no means numerous, statisticians were perhaps even rarer, while the combination of agriculturist and statistician in the one man made him a person with talents of unusual distinction. In America, for example, the value of the agricultural-statistician is fully appreciated and his work is so highly skilled that the Government service require those whom it employs in this capacity to have 3 years' practical farming experience and 4 years in an agricultural college in addition to statistical training. So great indeed is the importance attached in America to the official estimates of the production of the various crops—such as cotton—that the estimates are made in a locked and guarded room, which allows no means of communication with the outside world, until the given moment when the news is flashed to all parts of the country simultaneously.

In the Saorstát agriculture is of first importance, and the raw materials of the principal industries are for the most part the products of agriculture—sugar beet for the manufacture of sugar, barley for the brewing industry, potatoes for industrial alcohol, being a few examples of the kind. Sugar has been manufactured in this country for a decade, but the import of sugar beet seed has amounted to £44,000 in the

* "Moments of the Ratio of the Mean Deviation to the Standard Deviation for Normal Samples." By R. C. Geary. *Biometrika*, Vol. XXVIII. Parts III and IV, December, 1936. Page 295.

present year. And while the production of wheat has increased, the import of wheat seed has risen to £136,000. It was possible, she was informed, to produce these seeds in the country, but presumably success in this direction could only be the result of extensive experiments. She asked Mr. Hussey in how far the experiments in the Albert College were in relation to the production of these items on which the new industries so largely depended.