ON TESTING VARIETIES OF CEREALS.

BY "STUDENT."

Being a Paper read before the Society of Biometricians and Mathematical Statisticians. May 28th, 1923.

OBJECT OF EXPERIMENTS.

THE object of testing varieties of cereals is to find out which will pay the farmer best. This may depend on quality, but in general it is an increase of yield which is profitable, and since yield is very variable from year to year and from farm to farm it is a difficult matter upon which to obtain conclusive evidence.

Yet it is certain that very considerable improvements in yield have been made as the result of replacing the native cereals by improved varieties; as an example of this I may cite the case of Ireland, where varieties of barley have been introduced which were shown by experiment to have an average yield of 15 °/₀ to 20 °/₀ above those which they replaced. This represents, probably, a gain to the country of not less than £250,000 per year. As the cost of experiments from the commencement to the present time cannot have reached £40,000 the money has been well spent.

ORIGIN OF VARIETIES.

In the first place the ordinary cereals, wheat, barley, oats, and so on (maize is not here considered), are all self-fertilized and occur in races broadly distinguished by different botanical characters—Potato Oats, Rivett Wheat, Chevalier Barley, and so forth.

Besides these botanically distinguishable races, it is possible to pick out strains from commercial seed which differ from one another in all kinds of ways: time of ripening, percentage of nitrogen, yield, etc., although botanically the same. Many of these strains have been selected from time to time, certainly from the end of the eighteenth century up to the present time.

Finally there are hybrids, the result of deliberate crossing, and the selection of the best individuals out of the many thousands which may be grown in three generations is one of the more difficult problems with which the plant breeder has to deal, but it is only after he has made his preliminary selection that his hybrids concern the experimenter who is testing varieties.

Owing to the fact of self-fertilization, the various races, strains, and even to a large extent the hybrids, remain practically constant from year to year if once pure seed has been obtained.

Chief Sources of Error.

The peculiar difficulties of the problem lie in the fact that the soil in which the experiments are to be carried out is nowhere really uniform; however little it may vary to the eye, it is found to vary not only from acre to acre but from yard to yard, and even from inch to inch. This variation is anything but random, so that the ordinary formulae for combining errors of observation which are based on randomness are even less applicable than usual.

Next, of course, is the weather: that will hardly affect experiments carried out in the same field in the same year, but experiments carried out in different districts and seasons meet with variations of weather which may produce results quite inconsistent with the experimental error determined at either place. Obviously, the weather needs to be well sampled before drawing general conclusions.

The effects of soil and weather on the yields are far greater than the differences which we have to investigate, and it is because the planning of experiments and their interpretation when completed are not quite straightforward that this paper has been written.

Methods of Operating.

There are, broadly speaking, two methods of operating:

i. On a large enough scale to use the ordinary agricultural implements, ploughs, seed drills, reaping machines, etc.

ii. On quite a small scale with spades and dibblers, and scissors, under a wire net to keep out birds and rabbits.

Taking first the large scale, it has the advantage that the farmer, who always has a healthy contempt for gardening, may pay some attention to the results; he is to this extent right, that large scale conditions cannot be accurately reproduced in a wire cage, and in fact some varieties which have come out well on the small scale have not done as well in the field, though this is not at all common. Large-scale work then, is necessary as a final demonstration, and historically, it was on the large scale that variety experiments were first carried out.

Large Scale Work.

As an instance of large scale work, we may take a series of experiments carried out by the Department of Agriculture in Ireland to find out the best variety to grow in that country.

The experiments lasted six years, *vide* Table I, and during that time seven varieties were tested; only two, however, Archer and Goldthorpe, were carried through from start to finish, as the others were either dropped when they were found to be inferior, or were not among those chosen in the first place. The original seed was ordinary commercial seed, and the plots were two acres in extent. This is very large even for a large scale plot, but it was intended that the produce should form the raw material for further manufacturing experiments. This was a wise precaution, as has been found recently when a barley in other ways among the best was found to be quite unsuitable as malting material.

The produce of the plots was all valued (in those days—1901–1906—values were fairly steady from year to year), and this gives a method of combining yield and quality, but although the quality varied very much from one farm to another, there was generally only a small difference between the quality of different varieties grown on the same farm in the same season. The value of the crop per acre depended chiefly on the yield.

During the six years 193 plots were grown and at different times eighteen farms provided the land. These farms were scattered up and down the barley-growing districts in Ireland. Here, however, we shall deal only with the 51 plots of Archer, and the corresponding 51 plots of Goldthorpe.

The value per acre, then, of the 51 Archer plots varied between 90/- and 234/- with a mean of 178/- and a standard deviation of 33·6 shillings. The value per acre of the Goldthorpe plots varied between 99/- and 230/- with a mean of 166/- and a standard deviation of 33/-. The difference, therefore, was 12/-, and at first sight this hardly appears significant, for had the Archer and Goldthorpe plots been independent, the standard deviation of their difference would have been about 6·5.

This brings us to the first principle of all agricultural experiments, viz., that only comparative values are of any use. If we are told that on a certain farm a new variety of barley produced 30 cwts. to the acre, we admit that the crop is good, but are not much interested. If, in addition, we hear that Archer gave 25 cwts. to the acre on the same farm, we begin to take notice; for it is some evidence as to the value of the new variety, and it is the difference of 5 cwts. to the acre which appeals to us and not the actual yields themselves. In point of fact, of course, the yields in these experiments were not independent. Each Archer has a corresponding Goldthorpe, and by considering the 51 differences, we find that the mean difference between Archer and Goldthorpe has a standard deviation of 3·3 shillings.

This reduction of the S.D. of the mean difference from 6·5 to 3·3 shillings, by considering the individual differences between corresponding pairs, depends of course on the fact that corresponding pairs are highly correlated, so that the last term in the formula

$$\sigma^2_{A-B} = \sigma^2_A + \sigma^2_B - 2r_{AB}\,\sigma_A\sigma_B$$

is by no means negligible. The art of designing all experiments lies even more in arranging matters so that r_{AB} is as large as possible, than in reducing σ^2_A and σ^2_B.

That the conclusion that Archer was better than Goldthorpe was fully justified is shown by the fact that taking the yearly averages Archer beat Goldthorpe every year, while in the individual farms Archer beat Goldthorpe in all but three out of eighteen, and of these one farm was only used one season, and the other two in two seasons. Further, it was discovered during the course of the experiments that the Archer was practically identical with a barley which the Danes called Prentice, which had beaten all others in their long series of experiments. Both Archer and Goldthorpe were, practically speaking, new to Ireland, and they—or some improvement* on them—have now almost entirely driven out the other inferior barleys from most parts of the country.

Such, then, is the sort of error which attaches to large experimental plots, that is to say a standard deviation of about 10–15 °/ₒ for a single comparison, and this is found to be the order of the error in all ordinary large scale work—it does not vary very closely with the size of the plot, provided that the plot be above say one-tenth of an acre, though there may be a slight decrease of error with increase of size.

It follows that although it is quite within the power of any individual farmer to carry out a large scale experiment (and the larger the easier to carry out), it is only by co-operation that enough evidence can be obtained to be of any value. This co-operation can in practice only be arranged by a government department, a large agricultural company, or a farmers' association, and it is government departments that have had most success.

SMALL SCALE WORK.

We may next discuss small scale work, leaving to the end a modification introduced by Dr E. S. Beaven, which combines the advantages of the ordinary large scale with a considerably smaller error. The considerations which led to this modification were derived from experience of small scale technique.

Preliminary Considerations. Before coming to any actual comparison of varieties on the small scale, attention is directed to some preliminary experiments carried out by three different sets of investigators: Stratton and Wood † at Cambridge, Mercer and Hall at Rothamsted ‡, and Montgomery at Nebraska Agricultural Experimental Station §.

The first harvested 9/10th acre of mangolds in 1/1000-acre plots : the second, one acre of wheat in 1/500-acre plots, and an acre of mangolds in 1/200-acre plots: the third two years in succession harvested the same 7/45th acre of wheat in 1/1440-acre plots, and all weighed the produce of each plot; Montgomery determined the percentages of nitrogen as well. All three experiments showed the same thing: that the variation is not random; the yield varies from point to point with an irregular regularity; there is consequently correlation between one plot and its neighbours, and generally there is a tendency for one end of a field to yield more than the other.

This is only what is to be expected from *a priori* considerations; naturally the nearer two plots are together the more likely is the soil and its condition to be

* In particular a hybrid of Archer with Spratt made by Capt. Hunter, Spratt-Archer 37/6, which proved its superiority to Archer and other varieties in "chessboard" trials similar to that detailed below.

† *Journal of Agricultural Science*, Vol. III. p. 417, "The Interpretation of Experimental Results."

‡ *Journal of Agricultural Science*, Vol. IV. p. 107, "The Experimental Error of Field Trials."

§ *Nebr. Agr. Exp. Sta. 25th Ann. Report*, 1910—11, pp. 164—180, "Variation in Yield and Methods of arranging Plots to secure comparative Results"; and *U.S. Dept. Agr. Bur. Plant. Indus. Bul.* 269, "Experiments in Wheat Breeding: Experimental Error in the Nursery and Variation in Nitrogen and Yield."

similar on each of them, and the obvious conclusion may be drawn that the smaller the plots the more exactly can the yield of adjacent plots be compared.

Taking the investigation of Mercer and Hall on the 500 "plots" of wheat, it should be noted that they were only taken as plots at harvest and before cutting formed an unusually uniform area of one acre, part of a much larger field of wheat. The mean yield of grain per plot was 3·95 lbs. with a range of 2·75—5·14, and a standard deviation of ·46 lb., or 11·6 °/$_\circ$ of the mean weight of a plot.

If two adjacent plots were taken as $\frac{1}{250}$ ac. plots the S.D. fell to 10 °/$_\circ$ instead of the 8·2 °/$_\circ$ of random sampling.

If four adjacent plots were taken as $\frac{1}{125}$ ac. plots the S.D. fell to 8·9 °/$_\circ$ instead of the 5·8 °/$_\circ$ of random sampling.

If ten adjacent plots were taken as $\frac{1}{50}$ ac. plots the S.D. fell to *6·3 °/$_\circ$ instead of the 3·7 °/$_\circ$ of random sampling.

If twenty adjacent plots were taken as $\frac{1}{25}$ ac. plots the S.D. fell to *5·7 °/$_\circ$ instead of the 2·6 °/$_\circ$ of random sampling.

If fifty adjacent plots were taken as $\frac{1}{10}$ ac. plots the S.D. fell to *5·1 °/$_\circ$ instead of the 1·6 °/$_\circ$ of random sampling.

The high value of the S.D. of the larger plots compared with that which would have been expected had the aggregation been carried out randomly is due to a similar cause to that which decreased the error of the comparison of Archer and Goldthorpe. There is correlation between the neighbouring small plots which make up the larger plots, so that the last term in the formula

$$\sigma^2_{A+B} = \sigma^2_A + \sigma^2_B + 2r_{AB}\sigma_A\sigma_B$$

is not negligible. This last term is in fact the bridge over a pitfall which has trapped many, including—as will be shown later—the present writer.

In an appendix to Mercer and Hall's paper I pointed out that advantage may be taken of this correlation if we consider the difference between adjacent plots.

Thus we have:

Size of plot	S.D. of single plot as per cent.	Calculated S.D. of difference between random pairs	Actual S.D. of difference between adjacent pairs	Total acreage required to reduce S.D. of a comparison to 1 per cent.
1/500	11·6	16·4	11·2	·50 acre
1/250	10·0	14·1	9·7	·74 "
1/125	8·9	12·6	9·3*	1·37 "
1/50	6·3	8·9	3·7*	1·10 "
1/25	5·7	8·1	3·9*	3·84 "

Except in the case of the 1/125 plots we actually find that the standard deviation of a difference between two plots is less than the standard deviation of a single plot, and that working with 1/500 plots, the standard deviation of a

* The numbers are too few to do much more than indicate the tendency.

TABLE I.

Irish Experimental Barley Plots. Yield and Money Value Per Acre of Archer and Goldthorpe 1901—1906.

Farmer	Place	District	Archer					Goldthorpe				
			Yield		V. P. A.			Yield		V. P. A.		
			Barrels	Stones	£	*s.*	*d.*	Barrels	Stones	£	*s.*	*d.*
1901:												
McCarthy ...	Ballinacurra .	Cork	11	4	9	0	0	7	0	5	2	0
Hawkins ...	Whitegate ...	„	10	3	8	3	0	7	12	6	3	0
Dwan ...	Thurles ...	Central Plain	15	2	11	13	0	13	14	11	0	0
Wolfe ...	Nenagh ...	„	11	0	8	13	0	10	0	8	3	0
1902:												
McCarthy ...	Ballinacurra .	Cork	12	6	8	13	0	11	14	8	11	0
Hawkins ...	Whitegate ...	„	14	0	10	12	0	13	0	10	3	0
Wolfe ...	Nenagh ...	Central Plain	12	2	9	4	0	13	6	10	2	0
Willington...	Birr... ...	„	12	6	9	16	0	9	3	7	6	0
Gorman ...	Enniscorthy .	Wexford	11	5	9	2	0	11	14	9	2	0
Nunn ...	Castlebridge .	„	11	3	8	18	0	11	4	9	0	0
1903:												
McCarthy ...	Ballinacurra .	Cork	6	10	4	13	0	7	4	5	5	0
Hawkins ...	Whitegate ...	„	8	12	7	1	0	7	5	5	19	0
Wolfe ...	Nenagh ...	Central Plain	8	2	5	9	0	8	7	6	11	0
Willington...	Birr... ...	„	9	13	7	9	0	8	0	6	6	0
Gorman ...	Arnestown ...	Wexford	5	5	4	10	0	7	11	5	15	0
Nunn ...	Castlebridge .	„	12	7	9	16	0	9	15	7	16	0
Quinn ...	Carlingford...	Louth	11	12	8	4	0	9	3	7	0	0
Kearney ...	Greenore ...	„	11	3	8	12	0	7	13	5	19	0
1904:												
McCarthy ...	Ballinacurra .	Cork	10	4	7	15	0	11	14	9	8	0
Hawkins ...	Whitegate ...	„	10	11	8	7	0	10	4	8	4	0
Wolfe ...	Nenagh ...	Central Plain	13	3	10	9	0	11	8	9	7	0
Willington...	Birr... ...	„	11	3	8	17	0	11	14	9	4	0
Kelly ...	Portarlington	„	12	1	9	12	0	11	3	9	0	0
Allardyce ...	Monasterevan	„	10	7	8	1	0	10	5	8	7	0
Roche ...	New Ross ...	Wexford	8	2	5	16	0	7	0	5	6	0
Nunn ...	Castlebridge .	„	9	2	7	8	0	6	4	4	19	0
Kearney ...	Carlingford...	Louth	8	0	6	7	0	9	7	7	9	0
Segrave ...	Dunleer ...	„	12	1	9	9	0	11	7	9	7	0
1905:												
McCarthy ...	Ballinacurra .	Cork	12	8	9	8	0	13	1	9	16	0
Hawkins ...	Whitegate ...	„	11	11	9	8	0	11	5	8	14	0
Wolfe ...	Nenagh ...	Central Plain	14	6	10	14	0	15	10	10	3	0
Willington...	Birr... ...	„	14	11	11	14	0	13	8	10	11	0
Luttrell ...	Monasterevan	„	14	8	11	0	0	12	13	9	14	0
Kelly ...	Portarlington	„	12	1	8	19	0	10	8	7	17	0
Matthews ...	Tullamore ...	„	13	12	10	18	0	10	10	8	1	0
Nunn ...	Castlebridge .	Wexford	11	6	7	15	0	11	6	8	0	0
Dooley ...	New Ross ...	„	13	0	10	0	0	13	10	10	12	0
Kearney ...	Carlingford...	Louth	14	6	9	12	0	11	4	7	12	0
Segrave ...	Dunleer ...	„	14	7	11	11	0	12	8	9	19	0
1906:												
McCarthy ...	Ballinacurra .	Cork	9	14	7	6	0	9	11	7	2	0
Hawkins ...	Whitegate ...	„	10	9	7	12	0	8	14	6	9	0
Wolfe ...	Nenagh ...	Central Plain	11	12	8	14	0	8	13	6	11	0
Willington...	Birr... ...	„	10	15	8	0	0	9	15	7	6	0
Luttrell ...	Monasterevan	„	9	10	7	3	0	10	9	7	17	0
Mulhall ...	„	„	12	8	9	6	0	13	14	10	7	0
Matthews ...	Tullamore ...	„	8	14	6	16	0	8	11	6	11	0
Tennant ...	Bagnalstown.	„	15	4	11	7	0	13	14	10	9	0
Nunn ...	Castlebridge .	Wexford	11	10	8	19	0	10	9	7	18	0
Dooley ...	New Ross ...	„	14	3	10	7	0	12	5	9	0	0
Kearney ...	Carlingford .	Louth	11	11	8	9	0	12	12	9	8	0
Segrave ...	Dunleer ...	„	14	6	10	8	0	13	6	9	16	0

NOTE. The Irish barrel of barley contains 16 stones.

DIAGRAM I.

Giving the Yields in Grammes of the Individual Plots of Dr Beaven's No. 1 *Yield Experiment of* 1913.

Direction described as "vertical" ↑

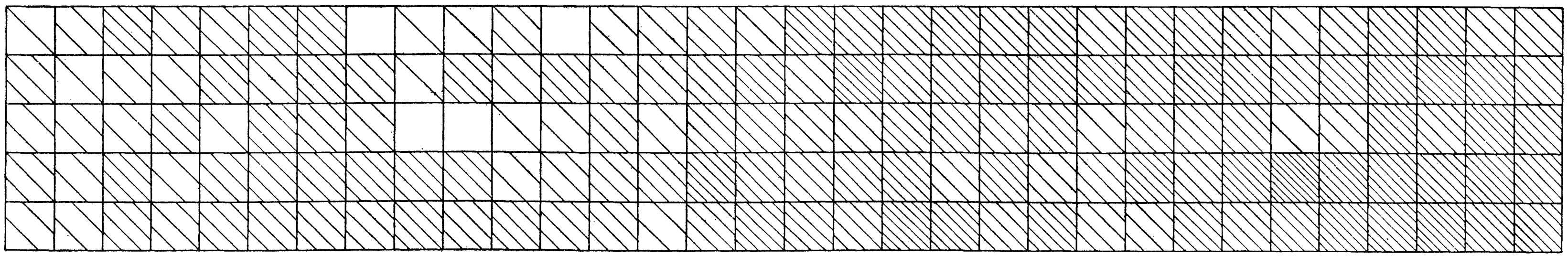
																			*											*	
E 230·1	B 249·3	G 312·2	D 263·4	A 263·4	F 273·4	C 314·9	H 198·0	E 227·9	B 225·7	G 265·4	D 203·0	A 236·9	F 228·0	C 293·8	H 237·6	E 336·6	B 330·2	G 323·5	D 341·0	A 327·0	F 308·1	C 324·9	H 298·7	E 334·8	B 316·6	G 258·2	D 299·8	A 338·9	F 321·0	C 317·2	H 266·8
D 255·9	A 222·6	F 218·7	C 258·2	H 273·3	E 268·2	B 301·1	G 285·7	D 212·0	A 296·7	F 237·7	C 308·8	H 225·0	E 290·4	B 297·5	G 334·4	D 282·6	A 374·3	F 300·4	C 377·9	H 303·0	E 351·5	B 344·7	G 317·7	D 362·4	A 331·8	F 298·7	C 366·4	H 325·6	E 377·0	B 354·7	G 321·8
C 265·6	H 205·0	E 246·7	B 277·4	G 232·3	D 286·0	A 286·1	F 229·3	C 221·2	H 170·7	E 231·8	B 235·5	G 284·4	D 239·7	A 302·2	F 278·8	C 301·2	H 260·3	E 328·0	B 324·1	G 305·9	D 308·7	A 248·0	F 272·7	C 324·2	H 295·6	E 253·5	B 286·7	G 372·6	D 322·3	A 362·9	F 326·3
B 265·9	G 236·7	D 295·8	A 238·3	F 265·0	C 320·9	H 272·9	E 298·7	B 307·9	G 301·3	D 244·2	A 263·1	F 257·5	C 318·2	H 331·7	E 324·0	B 301·7	G 326·8	D 325·5	A 280·2	F 315·0	C 305·9	H 269·2	E 376·4	B 314·2	G 376·0	D 428·4	A 405·7	F 362·9	C 379·9	H 348·6	E 371·3
A 236·5	F 210·4	C 291·1	H 223·9	E 295·6	B 290·5	G 286·1	D 266·2	A 291·6	F 247·3	C 321·4	H 249·9	E 286·5	B 246·6	G 301·0	D 333·0	A 322·0	F 282·4	C 399·3	H 330·8	E 303·9	B 343·3	G 285·2	D 273·5	A 333·4	F 295·9	C 362·6	H 349·8	E 395·3	B 400·6	G 364·9	D 361·8

——Direction described as right——→ and left ←——

A. English Archer. B. 145. C. 145/46. D. Plumage.
E. Early Archer. F. Irish Archer. G. 7 *A*. H. Biffin.

DIAGRAM II.

Plan of Experimental Area, showing Variation in Yield in Different Parts. Each line represents 20 *grammes more yield compared with the mean of the race grown on that plot.*



* Owing to a slip of the draughtsman the vertical rules between these asterisks have been omitted; the omission has no significance.

comparison between the varieties grown on a total area of half an acre is as low as 1 °/₀. On the lines of the 2-acre plots more than half a square mile would have been required. Further, there is every indication that smaller plots would be still more economical of ground.

These have been termed preliminary experiments, and so they are for the purpose of this paper; but in point of fact they followed the practical application of the principle which has just been outlined, and a further step in advance had already been made.

Carrying the principle of maximum contiguity, which he had deduced *a priori*, to its extreme logical limit, Beaven had compared two varieties in his cage by sowing alternate rows. He used a pure line of Archer barley, and one of a variety called "Plumage," which is allied to the Goldthorpe of the Irish experiments. He also grew 1/10th acre of each outside the cage and found that whereas the Archer gave slightly the better yield outside the cage, the cage work gave the yield of Plumage some 20 °/₀ better than the Archer.

He sent me the figures to look at, and I found that so far from the correlation between the yields of adjacent drills being positive, it was significantly negative.

This was quite unexpected at the time (1905), but the explanation was simple, viz., that when a plant of one variety is grown next to one of another variety it is abnormally situated, and is subject to abnormal competition.

In this case the Plumage was a taller barley and shaded the Archer; probably also, it started growth more quickly underground and so annexed more of the soil than its competitor. Anyhow, it was clear that a comparison of adjacent rows, with the possibility of interference of this kind, was useless.

The Square Yard Plot.

To avoid this difficulty, Beaven invented in 1909 the "square yard" plot, which is formed by sowing eight rows six inches apart, four feet long, and with seed two inches apart in the row. This gives in the first place a plot 4 feet by 4 feet; but at harvest the outside rows are rejected and the outside 6 inches at each end of all the other rows, thus leaving the inside square * yard for the measurement of yield free from the competition of other varieties.

So far as I am aware, no one has made any further enquiry as to the most economical size of plot; the square yard plot only utilises for yield determination 9/16 of the experimental area, and to make it smaller would waste still more ground, while the larger the plot the more we depart from the principle of maximum contiguity.

* There has been some controversy in America as to the advisability of testing varieties in alternate rows, but lat→ T. A. Kiesselbach (*Journ. Am. Soc. Agron.* 1919, No. 6, pp. 235—241, "Experimental Error in Field Trials"; pp. 242—247, "Plant competition as a source of error in Field plots") has come to much the same conclusion as Beaven, viz., that although certain varieties *may* not under some circumstances interfere with one another, yet it is dangerous to allow any chance of the experiment being subject to this source of error, and that the only safe thing to do is to surround each experimental area with a border of the variety grown upon it, and to discard this border at harvest.

There are probably not enough data to discover by the calculus the size of plot which will give the minimum probable error per acre, and no one seems to have faced the labour of an experimental determination. At all events, without any further investigation the square yard plot has been adopted as the unit in some six or seven experimental cages in the British Isles.

COMPARISON ON A "CHESSBOARD."

Having adopted the unit, it was a comparatively simple matter to set units of two varieties in a "chess" or "chequer" board: subsequently it was found that more than two varieties could be economically compared at the same time.

To illustrate the problems which arise when we come to compare several varieties grown together on a "chessboard," we may take Beaven's No. 1 Yield Experiment of 1913 *.

In this, twenty plots of each of eight races of barley were grown on a regular system of repetition, and the following observations were made for each plot:

Number of plants.
Number of ears.
Weight of ears.
Weight of straw.

For the purpose of this illustration we need only consider yield of corn, i.e., weight of ears.

The eight races consisted of:

Four Strains of Archer	English Archer ...	Selection made by Beaven.
	7A	
	Irish or Early Archer	Selection made by Capt. H. Hunter, B.Sc., of the Irish Department of Agriculture.
	Irish Archer, No. 5	
	Plumage	A selection made by Beaven which originated in Denmark. Wide-eared barley somewhat like Goldthorpe.

Each of these was, of course, descended from a single seed a few generations back, and

Three Hybrids	145, and 145/46	From a Plumage-Archer cross made by Beaven, the second being a re-selection from the first.
	"Biffin" ...	Selected by the Professor of that name from a Plumage-Archer cross of his own.

In order to simplify the comparison of errors it is best to work as long as possible, not with the standard error but the "variance," or square of the standard error. It has two advantages: (i) that variance can be added or subtracted without the preliminary squaring and subsequent extraction of the square root, and (ii) that the area required to give any required accuracy varies directly with it; in order to give the same error a comparison with a variance of 60 only requires half as much ground as a comparison with a variance of 120.

Further, the variance taken in each case will be the variance of the average of

* *Vide* Diagram I, p. 277.

20 plots or differences between plots, or whatever it may be, and to get this we divide by 19, and not by 20, to correct for the small number.

The following table gives the means and variances of the average of 20 plots for the eight races as follows:

TABLE II.

	Mean weight per plot, grammes	Variance of the average of 20 plots
145/46	318·7	94·7
Early Archer ...	306·5	138·9
7*A*	304·6	80·7
145	300·7	94·9
English Archer ...	297·8	128·8
Plumage	295·2	150·8
Irish Archer, No. 5	276·5	81·7
Biffin	270·8	142·0
Average	296·4	114·1

CORRECTION FOR POSITION *.

There is a great disadvantage in correcting any figures for position, inasmuch as it savours of cooking, and besides the corrected figures do not represent anything real. It is better to arrange in the first place so that no correction is needed.

In the present case the "vertical" arrangement is satisfactory, but as to right and left it is not so. English Archer averages ·2 rows to the left of 145, ·4 to the left of 145/46 and so on, 1·4 rows to the left of Biffin. As the average value per plot of a row is about 3·3 grammes higher than that of the row on its left, it might be thought right to make the following corrections:

145/46 ...	318·7 + 1·0 = 319·7
Early *A* ...	306·5 − 0·3 = 306·2
7*A*	304·6 − 1·7 = 302·9
145	300·7 + 1·7 = 302·4
English *A* ...	297·8 + 2·3 = 300·1
Plumage ...	295·2 + 0·3 = 295·5
Irish *A*, No. 5	276·5 − 1·0 = 275·5
Biffin ...	270·8 − 2·3 = 268·5

* For an elaborate method of Correction for Inequality of Soil, see Pearl, "A Method of Correcting for Soil Heterogeneity in Variety Tests," *Journal of Agricultural Research*, Vol. v. p. 1039.

In this paper Dr Pearl has corrected yield on the analogy of a contingency table. The method, which is probably as good a way as any of correcting for position, seems to me to be open to serious objections. A blot on the paper is the publishing of a "probable error" calculated from four cases without either correcting for the very small number or calling attention to the fact that they are appreciably too low.

The error of a comparison would no doubt be reduced very slightly as it generally is by any operation of this kind.

In any case the order is not altered, and I do not think the correction is worth making; the proper course would have been to reverse the order of the plots half way through so as to compensate for a possible tendency to improve from one end of the experimental area to the other.

VARIANCE IN TABLE II.

With the small numbers in question the variance figures do not differ significantly, but incidentally there is no indication that the hybrids are more variable in yield than the pure lines.

In order to get a clear idea of what these figures mean, let us suppose that a standard error of 1 °/₀ is desired, say 3 grammes, a variance of 9. That would require an area $\frac{114 \cdot 1}{9}$, or 12·7 times as large as the present 20 plots.

If now the plots had been randomly placed, the variance of a comparison between two of the races would have been approximately 228, and about 25 times as much ground as was used would have been required to reduce the standard error of a comparison to 1 °/₀.

In order to give a general idea of the nature of the variability, chiefly due to soil, which has to be regarded as error when we consider the yield of varieties, Diagram II has been prepared in which each 20 grammes of yield above 100 grammes below the average yield of the variety is represented by a diagonal line drawn across the square representing the plot. It will be noticed that the shading grows heavier towards the right of the diagram, and that while it is by no means regular, the correlation between the shading of neighbouring plots is obvious to the eye.

The arrangement of the different races in a chessboard is of course designed to take advantage of this correlation by comparing always neighbouring plots as in the following example which concerns the first pair of races in the table.

Beginning at the left hand of Diagram I, 145/46 is in the middle of the first vertical line, and Early Archer at the top—the former being indicated by the letter C, and the latter by E. The yield of the first is 265·6, and of the second, 230·1. That gives a positive difference of 35·5. The next appearance is in the third line, again a positive difference, this time of 44·4. In the third occurrence the 145/46 is in the fourth line, and the Early Archer in the fifth line, and the difference this time is negative and 37·4, and so on.

The variance of the average of the 20 differences thus obtained is 124·0, very much less than the 233·6, which is the sum of the variances of the averages of the two races.

Now, if there were only two races in the chessboard it would be comparatively straightforward—the standard deviation would be found from the variance, and Sheppard's Tables (or preferably with such small numbers, "Student's") would be used to judge the significance of the mean difference. In point of fact, however, the two races do not stand alone, and the question arises whether it would not be better to take the average variance of all the 28 differences between all the possible pairs of eight races.

Of course it is not likely that all our races would have the same variance, but with our small numbers such differences as there may be are almost certainly swamped by the error of random sampling, which, as pointed out above, will account for the observed values. From that point of view then it is better to average.

Again, all the comparisons are not of equal value: Irish Archer No. 5 is always found exactly on the right of English Archer, while Plumage is either three squares above English Archer or two below and one row to the right, and as will be shown later, there are indications that this is enough to affect the variance. Still it is not a very big thing, and the advantages of using a single figure far outweigh the slight loss of accuracy. I have calculated the 28 variances and they range from 44·1 (English Archer—Irish Archer No. 5) to 192·9 (Early Archer—Plumage), with a mean of 107·9. This is slightly lower than the 114·1, the average variance of the races. In other words, we have gained by chessboarding to the extent that we are as accurate as if we had devoted twice the area to plots randomly arranged.

The calculation of these 28 variances is tedious, but fortunately there is a short cut which gives an identical result.

In the following proof capital subscripts indicate variance directly measurable, which is taken as the mean value of such variance, while small subscripts indicate variance deducible from the observations.

If we suppose the total variance σ^2_t of mn plots (i.e. n groups of one of each of m races subject to the error of random sampling) to be divided into three parts:

(i) that due to the m races if measured without error...σ_r^2;

(ii) that due to the position of the n groups of m races from left to right of the diagram (in this case 20 groups of eight) also measured without error...σ_g^2;

(iii) the casual error, which is the only part subject to random sampling...σ_e^2;

these three parts may be assumed to be independent so that

$$\sigma_t^2 = \sigma_r^2 + \sigma_g^2 + \sigma_e^2;$$

also the variance of the means of the races as we measure them is

$$\sigma_R^2 = \sigma_r^2 + \frac{\sigma_e^2}{n} - \frac{\sigma_e^2}{mn},$$

the last term being due to the fact that we have only mn cases to give us the mean.

Similarly the variance of the means of the groups as we measure them is

$$\sigma_G^2 = \sigma_g^2 + \frac{\sigma_e^2}{m} - \frac{\sigma_e^2}{mn},$$

and the total variance as we measure it is

$$\sigma_T^2 = \sigma_t^2 - \frac{\sigma_e^2}{mn};$$

from which eliminating $\sigma_t^2,\quad \sigma_r^2,\quad \sigma_g^2,$

we find

$$\sigma_e^2 = \frac{mn(\sigma_T^2 - \sigma_R^2 - \sigma_G^2)}{(m-1)(n-1)},$$

and consequently

$$\frac{2\sigma_e^2}{n},$$

which is the variance of a comparison between n groups of two races, is

$$\frac{2m\,(\sigma_T^2 - \sigma_R^2 - \sigma_G^2)}{(m-1)(n-1)}*.$$

* In my first attempt to obtain this formula, I overlooked the $-\frac{\sigma_e^2}{mn}$ in the three equations for σ_R^2, σ_G^2, and σ_T^2. It was only after receiving a letter from Mr R. A. Fisher, who had independently arrived at the correct formula, that I found my mistake. Mr Fisher sent me two proofs, one of which was purely algebraical, proving in his notation the identity

$$\frac{2}{m(m-1)}\,S\,\frac{1}{2(n-1)}\left\{\overset{n}{\underset{1}{S}}(X_{pq}-X_{p'q})^2 - n\,(\overline{X}_p - \overline{X}_{p'})^2\right\}$$

$$= \frac{\overset{n}{\underset{1}{S}}\,\overset{m}{\underset{1}{S}}\,(X-\overline{\overline{X}})^2 - m\,\overset{n}{\underset{1}{S}}\,(\overline{X}_q-\overline{\overline{X}})^2 - n\,\overset{m}{\underset{1}{S}}\,(\overline{X}_p - X)^2}{(m-1)(n-1)};$$

and the other, which he himself prefers, I append below:

"Let there be n trials indicated by suffices $1\ldots,\ q\ldots,\ n$ of each of m varieties similarly indicated by suffices $1\ldots,\ p\ldots,\ m$.

Recognising that not only differences of variety but differences in the conditions of the trials may have affected the yields, we may obtain an estimate of what the variability would be if the conditions of any one trial could be replicated in a number of experiments with the same variety, provided the following simple assumptions hold good. The yield obtained in any experiment is the sum of three quantities, one depending only on the variety; a second, depending only on the 'trial'; and a third, which may be regarded as the 'experimental error' varying independently of variety and trial in a normal distribution about zero with a standard deviation which it is desired to estimate.

To obtain such an estimate we may fit the system of yields X_{pq} with a system of values $A_p + B_q$, choosing the latter so that

$$S\,(X_{pq} - A_p - B_q)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

is a minimum. Any one of the $m+n$ quantities A_p, B_q may be assigned an arbitrary value, and the remaining $m+n-1$ are then determinate: the observed values may therefore differ from those fitted in $(m-1)(n-1)$ degrees of freedom, and the corresponding estimate of the standard deviation ascribable to experimental error will be found by dividing the minimum value of (1) by $(m-1)(n-1)$. Evidently (1) will be a minimum if

$$A_p + B_q = \overline{X}_p + \overline{X}_q - \overline{\overline{X}},$$

where $\overline{X}_p$ is the mean of the values obtained with variety p, $\overline{X}_q$ the mean of the values obtained with trial q, and $\overline{\overline{X}}$ is the general mean.

The actual evaluation is most conveniently carried out in the following form of the analysis of variance:

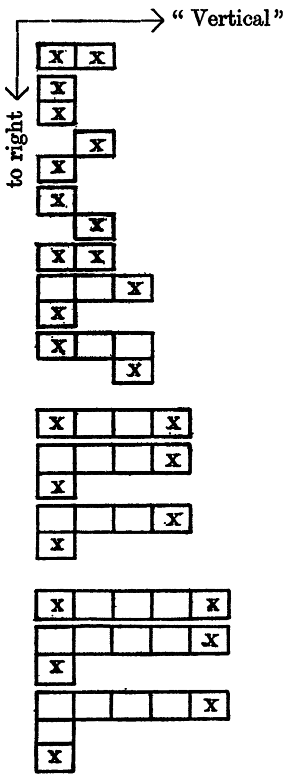
Variance	*Degrees of Freedom*	*Sum of Squares*
(*a*) Due to variety	$m-1$	$n\,\overset{m}{\underset{1}{S}}\,(\overline{X}_p - \overline{\overline{X}})^2$
(*b*) Due to trial	$n-1$	$m\,\overset{n}{\underset{1}{S}}\,(\overline{X}_q - \overline{\overline{X}})^2$
(*c*) Random variation	$(m-1)(n-1)$	$\overset{m}{\underset{1}{S}}\,\overset{n}{\underset{1}{S}}\,(X_{pq} - \overline{X}_p - \overline{X}_q + X)^2$
(*d*) Total	$mn-1$	$\overset{m}{\underset{1}{S}}\,\overset{n}{\underset{1}{S}}\,(X - \overline{\overline{X}})^2$

The sum of squares in line (*c*) being calculated by subtracting the values of lines (*a*) and (*b*) from the

To obtain the variance by this formula is a comparatively simple operation. In this case owing to the fact that I grouped the 160 observations in 10-gramme groups I got 109·3 by the short cut instead of 107·9, but it really should give an identical value.

Taking the square root we get a S.D. of 10·4 grammes or thereabouts for the standard error of a comparison, i.e., a probable error of about 2·4 °/₀. This is probably as near as it is worth while going in any one season, for the experiment

DIAGRAM III.

Position	Distance between centres	Variance	Number of Differences
	4′	113·9	112
	4′	66·5	60
	5·7′	92·9	64
	5·7′	125·5	32
	8′	91·2	72
	9′	101·2	60
	9′	167·7	12
	12′	146·5	40
	12·7′	114·6	48
	14·4′	146·5	8
	16′	132·0	16
	16·5′	131·1	28
	17·9′	94·5	8

total. If either variety or ‘trial’ were without significant effect on the yield, the corresponding mean square would not differ significantly from that of line (*c*). To test the significance of such a difference we may use the fact that the estimates of variance in (*a*), (*b*) and (*c*) are all independent, and when m and n are fairly large the natural logarithm of the mean square has standard deviation $\sqrt{\frac{2}{n_1}}$, where n_1 is the number of degrees of freedom. In comparing two such independent estimates of the mean square, we therefore obtain the difference of their natural logarithms, and assign to it a standard deviation

$$\sqrt{\frac{2}{n_1}+\frac{2}{n_2}}.\text{”}$$

must be repeated several times to sample the weather properly, and cage area is too valuable to expend more than is absolutely necessary on a single experiment.

Before leaving this subject of chessboards, I would like to show in rather more detail that even with such small plots as these, slight differences in the arrangement within the group tend to increase the variance over that due to the ideal juxtaposition.

I have, therefore (see Diagram III, p. 284), separated the various kinds of comparisons and averaged the variance, in each case as that of the average of 20 differences.

The figures are not of course worth a great deal, but there is a marked tendency for the comparisons between the more distant plots to be the less accurate.

For purposes of illustration, I have correlated the distances with the variance for the 13 positions by the Spearman method, and get $\rho = + \cdot 41$.

THE HALF DRILL STRIP METHOD*.

The small scale work with which I have just dealt affords a means of picking out good varieties which can be tested in field trials. The whole eight varieties were tested on about 1/17 acre, sowing about a quarter of a pound of seed for each race. We now proceed to the most accurate method yet devised for field trials by which two varieties are compared on a total area of 5200 square yards, just over an acre, with, in the case which I shall give you, a standard error of $\cdot 63\,^{\circ}/_{\circ}$. Of course, it will not necessarily be as low as this always.

The field is cultivated as usual up to the time of sowing, except that particular care is taken to clean the ground of weeds.

When sowing, the seed box of the drill is divided into two across the middle, and the middle coulter put out of action. The seed of the two varieties is put in the seed box, one on each side of the division. Thus when sowing a drill strip, one half (i.e. 6 or 7 rows) is sown with one variety and the other half with the other. On turning the drill at the end, the next strip is sown so that two half strips of the same variety are next each other, but care is taken to leave an interval between the two drill strips exactly equal to the gap in the middle of each drill strip between the two varieties. It requires careful steering but it can be done.

When the experimental field is sown, we get first a single half drill strip of one variety, then two of the other, then two of the first and so forth, ending with a half drill strip of the first. This ending is necessary in order to discount any fertility slope from one end to the other of the field. The space outside the experimental area should be sown all round with a similar grain, as the outside is naturally abnormal and is more liable to attacks from all kinds of enemies.

At harvest the outside row of each half drill strip next to the other variety is pulled up by hand and discarded to eliminate the "border" effect, and also to facilitate the use of the ordinary reaping machine. If the two varieties do not ripen

* For a full account *vide* "Trials of New Varieties of Cereals," by E. S. Beaven, *Journal of the Ministry of Agriculture*, Vol. XXIX. Nos. 4 and 5, 1922.

together one must be cut by hand when ripe, but if there is so little difference that both can be cut on the same day the reaping machine can be used on both. In either case each half drill strip is cut in such a way that the produce of each 1/500 acre can be tied up in two sheaves separately. In Beaven's case ten such 1/500 acre plots went to each half drill strip.

These sheaves can be weighed on the field, and so we can get the total produce of the field in plots of 1/500 acre and can compare each 1/500 acre with an adjoining one of the other variety.

Two things are to be noted at this point: (1) That without a very great deal of trouble the plots cannot be threshed out separately, but, fortunately, it has so far always been found where the matter has been put to the test that the variability of the yield of grain expressed as a percentage of the grain is less than the variability of the total yield expressed as percentage of total yield. In the Mercer and Hall experiment, the standard errors were 11·6 °/₀ and 11·9 °/₀, and Beaven's experience has been similar. Thus the figure which we obtain for the Standard Error is likely to be in excess of the truth. (2) From a practical point of view it is easier to work with a few half drill strips than a larger number of short ones, but if we depend on the weights of a few drill strips, there is considerable uncertainty about the Standard Error of the result. It was hoped that by determining the Standard Error of the difference between adjacent 1/500 acre plots, we could deduce the standard error of the average of n such differences by the formula $\sigma_a = \frac{\sigma}{\sqrt{n}}$, so that it would be immaterial whether the drill strips were long and few or short and many, as long as altogether there were n pairs of adjacent subplots. Indeed up to the time when I came to write this section, it was believed that this could be done. Beaven showed me his figures before publication, and I did not at the time observe that the formula cannot be used without further investigation, nor, so far as I am aware, has anyone else drawn attention to it. Nevertheless, I think it will be clear from the general considerations which have been advanced throughout the paper that there is a danger that the differences between corresponding constituent plots of a drill strip, even when they are as narrow as these, will tend to be correlated, and the formula $\sigma_a = \frac{\sigma}{\sqrt{n}}$, which requires independence of the individuals which are to be averaged, cannot be used without correction*. That this is so in the particular case which we are considering is made highly probable from the fact

* A fallacy arising from a similar neglect of correlation has come under my notice in some American work, but there the absurdity is more easily demonstrated. In the *Journal of the American Society of Agronomists*, Vol. IX. 1917, p. 138, A. G. McCall proposed that in order to save the trouble of harvesting and weighing 1/10th acre plots a number of square yards should be cut out and harvested separately, the square yards being taken systematically throughout the 1/10th acre plot, and the yield per acre calculated from these square yards. So far, so good, by taking enough square yards the slight loss of accuracy may perhaps be made up by gain in time or feasibility of operating. But in 1919, Arny and Steinmetz, *Journal of the American Society of Agronomists*, Vol XI. pp. 88, 89, applying this method, compared the error of the yield calculated from a few square yards cut from each of a number of 1/10th acre plots with that calculated from the 1/10th acre plots themselves. They found it substantially greater, but, say they, by increasing the number of square yards cut from each 1/10th acre plot to n, we

that the variance, expressed in terms of the percentage of the total weight of C, of the difference between the total produce from A and C is ·664 of the total weight of C when calculated from the 27 differences between adjacent half drill strips, while it is only ·301 when calculated from the 270 differences between adjacent subplots. The two figures should be the same within the error of random sampling, but differ probably by more than twice their S.D.

The results of the 1921 Trial are shown in Tables III and IV, which are taken, with his kind permission and that of the Ministry of Agriculture, from the Supplement to Beaven's paper, and give the weights of the sheaves, on the individual half drill strips, and on 243 of the 270 "plots," which go to make up the half drill strips respectively.

It will be seen that by taking the differences between adjoining half drill strips (or plots) a large part of the error is, as usual, eliminated.

Further, it is obvious that there is a general decrease in fertility as we go from drill strips with low numbers to drill strips with high numbers. It follows that the difference $A - C$ will tend to be greater when C follows A than when A follows C, and since this is always possible, experiments of this nature should always be planned so that there shall be an even number of differences, the series should begin and end with half drill strips of the same variety: in this case we may simply leave out the last drill strip and finish at half drill strip 52.

There is also a curious feature about these figures which can only be put down to some systematic error in technique; namely that when we compare together the adjacent half drill strips of A, that with the higher number always yields higher, although the general fertility runs the other way, and the same is true with regard to C in eight cases out of 13.

Both these kinds of error (that due to the general fertility slope and that due to the different fertility of odd and even half drill strips) are largely eliminated by Beaven's arrangement by which in alternate comparisons A follows C and C follows A and this can be made evident by adopting as unit not the difference between adjacent half drill strips but that between the sum of the two contiguous half drill strips of A and the sum of the two half drill strips of C which enclose them.

can decrease the error in the proportion $\frac{1}{\sqrt{n}}$, and so we can actually determine the yield more accurately by weighing up 10 or 20 square yards than by weighing up the whole half acre. It is rather surprising that they did not realise that there are 484 square yards in 1/10th acre, so that by taking 484 square yards they would be likely to be more accurate than if they took any lesser number and *a fortiori* tremendously more accurate than they would be if they took the same 484 square yards and called it 1/10th acre! Of course their formula also should be $\sigma\sqrt{\frac{1+(n-1)r}{n}}$, where r is the correlation between the yields on the square yards composing 1/10th acre plots, and not $\frac{\sigma}{\sqrt{n}}$.

The same fallacy has been used to extol the "rod row" method of determining yield, i.e., the method of cutting along the drill a row one rod in length to represent the yield of the plot from which it is cut.

TABLE III.

Warminster Field Variety Trial, 1921. *Half Drill Strip Weights, comparing:— Two races of barley, viz. "C" and "A." Area of each half drill strip*=100 *sq. yds. Total area* = 2700 *sq. yds.* = ·56 *acre for each race. Showing total weight of sheaves on each half drill strip.*

Half drill strip		Weight of sheaves on half drill strip		Difference between "A" & "C"	Half drill strip		Weight of sheaves on half drill strip		Difference between "A" & "C"
Number		lb.		lb.	Number		lb.		lb.
"C"	"A"	"C"	"A"	"A" – "C"	"C"	"A"	"C"	"A"	"A" – "C"
1	2	165·4	164·6	−0·8	29	30	160·9	160·2	− 0·7
4	3	159·5	173·4	+13·9	32	31	153·2	164·3	+11·1
5	6	169·3	169·3	—	33	34	144·9	154·3	+ 9·4
8	7	179·8	174·9	− 4·9	36	35	147·7	158·6	+10·9
9	10	172·5	177·6	+ 5·1	37	38	142·4	143·0	+ 0·6
12	11	170·7	182·9	+12·2	40	39	138·7	143·6	+ 4·9
13	14	173·3	167·5	− 5·8	41	42	131·1	143·2	+12·1
16	15	166·1	178·5	+12·4	44	43	141·6	145·3	+ 3·7
17	18	174·5	170·3	− 4·2	45	46	145·0	150·1	+ 5·1
20	19	163·3	176·0	+12·7	48	47	155·4	154·0	− 1·4
21	22	166·0	159·1	− 6·9	49	50	151·1	149·3	− 1·8
24	23	161·2	168·7	+ 7·5	52	51	145·6	149·7	+ 4·1
25	26	169·3	164·2	− 5·1	53	54	146·3	158·5	+12·2
28	27	156·5	167·0	+10·5	Total		4251·3	4368·1	—
					Average		157·5	161·8	+ 4·3
					per cent.		100	102·7	+ 2·7

TABLE IV.

Showing weights of each "plot" (two consecutive sheaves), 9 plots (= 18 sheaves) on each half drill strip. Total number of plots on 27 half drill strips = 27 by 9 = 243 plots of each race (excluding end sheaves which are left out of account in this case in computing "probable error").

Nos. of half drill strips, "C" "A"	"Plot" Weights 2 sheaves. lb. "C"	"A"	A − C lb.	Nos. of half drill strips, "A" "C"	"Plot" Weights 2 sheaves. lb. "A"	"C"	A − C lb.	Nos. of half drill strips, "C" "A"	"Plot" Weights 2 sheaves. lb. "C"	"A"	A − C lb.	Nos. of half drill strips, "A" "C"	"Plot" Weights 2 sheaves. lb. "A"	"C"	A − C lb.
1 2	17·0	17·6 ...	+ 0·6	3 4	17·4	16·3 ...	+ 1·1	5 6	16·4	16·2 ...	− 0·2	7 8	18·1	17·9 ...	+ 0·2
	15·8	16·9 ...	+ 1·1		16·6	16·0 ...	+ 0·6		17·7	16·8 ...	− 0·9		16·8	17·1 ...	− 0·3
	14·2	15·0 ...	+ 0·8		16·5	16·1 ...	+ 0·4		13·0	14·8 ...	+ 1·8		14·8	15·3 ...	− 0·5
	14·2	13·8 ...	− 0·4		15·4	13·2 ...	+ 2·2		14·5	14·4 ...	− 0·1		16·0	15·9 ...	+ 0·1
	16·2	15·8 ...	− 0·4		15·6	13·7 ...	+ 1·9		15·2	15·8 ...	+ 0·6		16·3	16·6 ...	− 0·3
	17·3	17·0 ...	− 0·3		18·9	16·7 ...	+ 2·2		17·7	17·3 ...	− 0·4		17·0	18·3 ...	− 1·3
	18·0	17·1 ...	− 0·9		18·6	15·6 ...	+ 3·0		18·7	17·9 ...	− 0·8		18·5	19·7 ...	− 1·2
	17·8	17·2 ...	− 0·6		17·8	16·9 ...	+ 0·9		17·7	16·5 ...	− 1·2		17·9	18·6 ...	− 0·7
	16·6	17·8 ...	+ 1·2		18·6	17·3 ...	+ 1·3		19·1	18·9 ...	− 0·2		19·3	19·7 ...	− 0·4
	147·1	148·2			155·4	141·8			150·0	148·6			154·7	159·1	
	18·3*	16·4*			18·0*	17·7*			19·3*	20·7*			20·2*	20·7*	
	165·4	164·6 ...	− 0·8		173·4	159·5 ...	+ 13·9		169·3	169·3 ...	—		174·9	179·8 ...	− 4·9
9 10	17·6	16·7 ...	− 0·9	11 12	18·0	16·1 ...	+ 1·9	13 14	16·4	15·7 ...	− 0·9	15 16	18·5	16·4 ...	+ 2·1
	17·2	16·9 ...	− 0·3		16·1	15·6 ...	+ 0·5		16·5	17·1 ...	+ 0·6		15·9	16·1 ...	− 0·2
	16·0	15·8 ...	− 0·2		15·9	14·2 ...	+ 1·7		14·4	13·7 ...	− 0·9		15·9	16·0 ...	− 0·1
	17·0	15·9 ...	− 1·1		16·9	15·6 ...	+ 1·3		14·4	15·0 ...	+ 0·6		16·0	16·0 ...	—
	17·3	17·4 ...	+ 0·1		18·1	16·5 ...	+ 1·6		17·2	15·8 ...	− 1·4		15·6	16·1 ...	− 0·5
	16·0	17·2 ...	+ 1·2		17·6	15·7 ...	+ 1·9		16·5	17·3 ...	+ 0·8		17·4	16·1 ...	+ 1·3
	20·1	19·0 ...	− 1·1		19·4	18·5 ...	+ 0·9		19·6	16·8 ...	− 2·8		19·6	16·7 ...	+ 2·9
	18·6	18·5 ...	− 0·1		18·2	17·9 ...	+ 0·3		18·5	18·6 ...	+ 0·1		20·3	17·4 ...	+ 2·9
	20·5	19·0 ...	− 1·5		19·2	19·9 ...	− 0·7		21·8	17·3 ...	− 4·5		20·1	17·1 ...	+ 3·0
	160·3	156·4			159·4	150·0			155·3	147·3			159·3	147·9	
	12·2*	21·2*			23·5*	20·7*			18·0*	20·2*			19·2*	18·2*	
	172·5	177·6 ...	+ 5·1		182·9	170·7 ...	+ 12·2		173·3	167·5 ...	− 5·8		178·5	166·1 ...	+ 12·4
17 18	18·1	17·1 ...	− 1·0	19 20	16·2	15·7 ...	+ 0·5	21 22	16·3	15·9 ...	− 0·4	23 24	14·4	15·2 ...	− 0·8
	16·6	17·6 ...	+ 1·0		17·7	15·2 ...	+ 2·5		16·5	14·6 ...	− 1·9		16·5	14·7 ...	+ 1·8
	17·1	16·1 ...	− 1·0		16·8	15·4 ...	+ 1·4		16·4	13·6 ...	− 2·8		15·3	16·2 ...	− 0·9
	16·8	16·4 ...	− 0·4		16·6	14·1 ...	+ 2·5		14·7	14·0 ...	− 0·7		14·6	14·2 ...	+ 0·4
	15·8	14·6 ...	− 1·2		16·8	14·6 ...	+ 2·2		13·3	13·5 ...	+ 0·2		14·3	14·6 ...	− 0·3
	16·0	15·3 ...	− 0·7		16·4	17·3 ...	− 0·9		17·1	15·4 ...	− 1·7		16·9	15·3 ...	+ 1·6
	18·1	18·1 ...	—		19·1	17·7 ...	+ 1·4		18·2	18·1 ...	− 0·1		18·7	17·3 ...	+ 1·4
	17·3	18·0 ...	+ 0·7		18·0	16·3 ...	+ 1·7		16·8	17·2 ...	+ 0·4		18·2	16·6 ...	+ 1·6
	18·5	17·3 ...	− 1·2		19·1	16·4 ...	+ 2·7		16·9	16·7 ...	− 0·2		18·7	17·0 ...	+ 1·7
	154·3	150·5			156·7	142·7			146·2	139·0			147·6	141·1	
	20·2*	19·8*			19·3*	20·6*			19·8*	20·1*			21·1*	20·1*	
	174·5	170·3 ...	− 4·2		176·0	163·3 ...	+ 12·7		166·0	159·1 ...	− 6·9		168·7	161·2 ...	+ 7·5
25 26	15·5	14·1 ...	− 1·4	27 28	14·7	14·6 ...	+ 0·1	29 30	14·7	14·0 ...	− 0·7	31 32	14·2	13·5 ...	+ 0·7
	13·7	16·7 ...	+ 3·0		15·6	15·3 ...	+ 0·3		14·2	13·6 ...	− 0·6		15·3	13·6 ...	+ 1·7
	16·0	14·5 ...	− 1·5		14·8	15·3 ...	− 0·5		14·8	14·5 ...	− 0·3		14·4	13·5 ...	+ 0·9
	14·1	15·7 ...	+ 1·6		14·2	13·7 ...	+ 1·5		15·7	14·8 ...	− 0·9		13·3	12·0 ...	+ 1·3
	16·6	16·9 ...	+ 0·3		15·4	15·8 ...	− 0·4		16·9	15·9 ...	− 1·0		17·1	17·3 ...	− 0·2
	17·2	15·2 ...	− 2·0		18·8	15·8 ...	+ 3·0		16·4	16·5 ...	+ 0·1		16·8	14·7 ...	+ 2·1
	19·4	18·3 ...	− 1·1		18·3	17·0 ...	+ 1·3		17·0	16·8 ...	− 0·2		18·9	17·3 ...	+ 1·6
	18·0	17·1 ...	− 0·9		19·2	15·7 ...	+ 3·5		16·7	18·0 ...	+ 1·3		19·0	15·9 ...	+ 3·1
	18·6	16·8 ...	− 1·8		17·5	16·6 ...	+ 0·9		16·5	17·5 ...	+ 1·0		17·2	17·8 ...	− 0·6
	149·1	145·3			148·5	139·8			142·9	141·6			146·2	135·6	
	20·2*	18·9*			18·5*	16·7*			18·0*	18·6*			18·1*	17·6*	
	169·3	164·2 ...	− 5·1		167·0	156·5 ...	+ 10·5		160·9	160·2 ...	− 0·7		164·3	153·2 ...	+ 11·1
33 34	13·5	14·8 ...	+ 1·3	35 36	15·0	13·1 ...	+ 1·9	37 38	12·6	11·4 ...	− 1·2	39 40	11·8	11·4 ...	+ 0·4
	10·7	12·5 ...	+ 1·8		13·7	11·3 ...	+ 2·4		11·4	12·9 ...	+ 1·5		11·9	11·1 ...	+ 0·8
	12·8	12·7 ...	− 0·1		12·6	13·5 ...	− 0·9		11·9	11·9 ...	—		12·3	12·9 ...	− 0·6
	12·6	14·0 ...	+ 1·4		14·3	14·5 ...	− 0·2		15·4	14·3 ...	− 1·1		12·8	13·6 ...	− 0·8
	15·3	16·3 ...	+ 1·0		16·4	16·8 ...	− 0·4		15·1	16·7 ...	+ 1·6		16·6	13·1 ...	+ 3·5
	15·3	15·9 ...	+ 0·6		17·4	12·6 ...	+ 4·8		14·0	15·8 ...	+ 1·8		14·2	15·3 ...	− 1·1
	17·5	18·1 ...	+ 0·6		15·3	17·4 ...	− 2·1		15·1	13·8 ...	− 1·3		16·8	16·0 ...	+ 0·8
	15·9	17·5 ...	+ 1·6		16·1	14·6 ...	+ 1·5		13·8	14·7 ...	+ 0·9		15·3	14·7 ...	+ 0·6
	14·8	15·1 ...	+ 0·3		17·9	16·3 ...	+ 1·6		16·0	14·7 ...	− 1·3		16·2	13·8 ...	+ 2·4
	128·4	136·9			138·7	130·1			125·3	126·2			127·9	121·9	
	16·5*	17·4*			19·9*	17·6*			17·1*	16·8*			15·7*	16·8*	
	144·9	154·3 ...	+ 9·4		158·6	147·7 ...	+ 10·9		142·4	143·0 ...	+ 0·6		143·6	138·7 ...	+ 4·9

* These figures represent weights of the first and last sheaves on each half drill strip added together, and are excluded in calculating the average weights and also in calculating the "probable error."

TABLE IV (*continued*).

Nos. of half drill strips, "C" "A"	"Plot" Weights 2 sheaves. lb. "C"	"A"	A − C lb.
41 42	10·6	11·4 ...	+ 0·8
	9·7	11·1 ...	+ 1·4
	11·4	10·1 ...	− 1·3
	11·9	10·7 ...	− 1·2
	12·9	13·3 ...	+ 0·4
	13·8	15·9 ...	+ 2·1
	16·5	19·5 ...	+ 3·0
	14·4	16·6 ...	+ 2·2
	14·1	16·2 ...	+ 2·1
	115·3	124·8	
	15·8*	18·4*	
	131·1	143·2 ...	+12·1

Nos. of half drill strips, "A" "C"	"Plot" Weights 2 sheaves. lb. "A"	"C"	A − C lb.
43 44	11·1	11·8 ...	− 0·7
	10·9	12·0 ...	− 1·1
	12·3	11·0 ...	+ 1·3
	12·0	13·6 ...	− 1·6
	14·4	15·6 ...	− 1·2
	15·5	15·0 ...	+ 0·5
	17·3	16·2 ...	+ 1·1
	15·7	15·8 ...	− 0·1
	18·3	14·3 ...	+ 4·0
	127·5	125·3	
	17·8*	16·3*	
	145·3	141·6 ...	+ 3·7

Nos. of half drill strips, "C" "A"	"Plot" Weights 2 sheaves. lb. "C"	"A"	A − C lb.
45 46	11·6	13·2 ...	+ 1·6
	11·7	12·5 ...	+ 0·8
	10·7	12·2 ...	+ 1·5
	13·9	14·5 ...	+ 0·6
	15·8	15·6 ...	+ 0·2
	14·8	16·2 ...	+ 1·4
	16·8	16·1 ...	− 0·7
	17·4	16·0 ...	− 1·4
	16·8	16·2 ...	− 0·6
	129·5	132·5	
	15·5*	17·6*	
	145·0	150·1 ...	+ 5·1

Nos. of half drill strips, "A" "C"	"Plot" Weights 2 sheaves. lb. "A"	"C"	A − C lb.
47 48	11·8	12·9 ...	− 1·1
	13·0	13·4 ...	− 0·4
	13·9	13·9 ...	—
	16·7	14·9 ...	+ 1·8
	15·6	16·3 ...	− 0·7
	16·4	17·0 ...	− 0·6
	15·6	17·0 ...	− 1·4
	17·6	15·7 ...	+ 1·9
	16·3	17·1 ...	− 0·8
	136·9	138·2	
	17·1*	17·2*	
	154·0	155·4 ...	− 1·4

Nos. of half drill strips, "C" "A"	"Plot" Weights 2 sheaves. lb. "C"	"A"	A − C lb.
49 50	13·6	13·8 ...	+ 0·2
	12·5	13·0 ...	+ 0·5
	13·6	13·8 ...	+ 0·2
	14·5	13·0 ...	− 1·5
	15·3	13·9 ...	− 1·4
	15·9	15·6 ...	− 0·3
	16·6	15·6 ...	− 1·0
	16·8	17·1 ...	+ 0·3
	16·4	16·0 ...	− 0·4
	135·2	131·8	
	15·9*	17·5*	
	151·1	149·3 ...	− 1·8

Nos. of half drill strips, "A" "C"	"Plot" Weights 2 sheaves. lb. "A"	"C"	A − C lb.
51 52	14·2	13·5 ...	+ 0·7
	12·8	11·5 ...	+ 1·3
	14·3	13·7 ...	+ 0·6
	13·5	14·2 ...	− 0·7
	14·3	14·1 ...	+ 0·2
	15·3	14·1 ...	+ 1·2
	14·5	16·0 ...	− 1·5
	16·7	15·7 ...	+ 1·0
	16·4	15·6 ...	+ 0·8
	132·0	128·4	
	17·7*	17·2*	
	149·7	145·6 ...	+ 4·1

Nos. of half drill strips, "C" "A"	"Plot" Weights 2 sheaves. lb. "C"	"A"	A − C lb.
53 54	13·8	13·1 ...	− 0·7
	12·3	12·2 ...	− 0·1
	11·8	14·8 ...	+ 3·0
	13·9	13·6 ...	− 0·3
	16·2	15·6 ...	− 0·6
	15·1	16·5 ...	+ 1·4
	15·5	16·0 ...	+ 0·5
	15·4	19·0 ...	+ 3·6
	17·2	16·5 ...	− 0·7
	131·2	137·3	
	15·1*	21·2*	
	146·3	158·5 ...	+12·2

Average weight per "plot" (end sheaves excluded)	15·56	15·88 ...	+0·32
%	100·0	102·1 ...	+2·10

* These figures represent weights of the first and last sheaves on each half drill strip added together, and are excluded in calculating the average weights and also in calculating the "probable error."

This may be described as a "sandwich," and it may be noted that just as there are subplots composing a half drill strip, so there are "sub-sandwiches" which will also tend to eliminate the same errors as the "sandwiches."

The following table gives the differences $A - C$ for the thirteen "sandwiches" composed of half drill strips 1 to 52:

TABLE V.

Half drill strip numbers	$A - C$	Half drill strip numbers	$A - C$
1 to 4	+13·1	29 to 32	+10·4
5 " 8	− 4·9	33 " 36	+20·3
9 " 12	+17·3	37 " 40	+ 5·5
13 " 16	+ 6·6	41 " 44	+15·8
17 " 20	+ 8·5	45 " 48	+ 3·7
21 " 24	+ ·6	49 " 52	+ 2·3
25 " 28	+ 5·4		

The mean $A - C$ for sandwiches is $+ 8{\cdot}05$ and the variance, making allowance for the pitifully small number, is 51·41. This leads to a variance of the difference between the total produce of A and of C expressed in terms of the total weight of C of ·398, intermediate between the ·664 calculated from the half drill strip differences and the ·301 calculated from the subplot differences.

It should be noted at this point that the "sandwich" is a perfectly legitimate device for eliminating errors common to both variants whose difference is to be measured, and that it is only by using it that we can get the true value of the error of the comparison, whereas the subplot difference would really lead to a larger value than ·301 if we had sufficient knowledge to be able to apply the true formula

$$\frac{\sigma^2(1+(n-1)r)}{n}.$$

A similar calculation based on the "sub-sandwiches," i.e. sandwiches one plot in depth, gives a value of the variance ·248 corresponding to the ·398 from the whole sandwiches. The S.D of these to some extent correlated figures is not easy to determine but the difference between them must be of the order of once the S.D. This is not significant but with our small numbers it is not inconsistent with the expected correlation between the sub-sandwiches composing a sandwich. Until a number of experiments have been carried out in several places and the results submitted to analysis, it would be wise to keep the number of drill strips as large as possible and economise in length in spite of the practical difficulties of doing so.

Since the variance calculated from the drill strip sandwiches is subject to a large error of random sampling owing to the necessary paucity of numbers, it is well to calculate also from the "sub-sandwiches" and take the larger of the two in determining the standard error.

It is possible that some of my readers may devise some better method of utilising the weights of the "subplots" than I have been able to do, and I commend the problem to them.

In the present case it is probably better with only thirteen sandwiches to take the standard error of a single sandwich and use "Student's" tables, when the probability that such a large positive difference should occur by chance is found to be ·001. The difference is therefore quite significant. If however it is required to compare the standard error with other experiments, we can say that the most probable value is only ·63 °/ₒ on a total area of about 1 acre.

Other precautions such as correction for moisture, etc., are taken as a matter of course.

CONCLUSIONS.

The chief difficulty of comparing varieties consists in the fact that the differences to be measured are quite small compared with the variations due to soil and weather. While the latter is not within our control, the errors due to the soil may be reduced to reasonable proportions in any one of three ways:

(1) Large plots may be repeated many times. An instance is given of this when in the Irish two-acre experimental plots a difference of 7 °/ₒ in the value per acre was proved with a standard deviation of about 2 °/ₒ in 51 trials, extending over six years.

Undertakings of this magnitude are hardly to be put in hand by any but Departments of State.

(2) Quite small plots of one square yard, surrounded by a border of the same variety as in the square yard, may be grown under a wire cage on a regular system, technically called a "chessboard." An instance of this is given when, in Beaven's No. 1 Yield Experiment of 1913, eight varieties were compared on a total area of about 1/17th acre using about 5 oz. of seed of each variety, with a standard deviation of a comparison in a single year of about $3\frac{1}{2}$°/$_{\circ}$.

The large number of varieties which may be compared at once, and the small area which is required, make this an ideal method of testing new varieties. On the other hand, a wire cage is not a cornfield, and the varieties found to be best in the cage will always require further testing on the large scale. The method is, however, within the powers of anyone who can build a cage, and has the necessary skill and patience to conduct the experiments.

(3) By means of Beaven's "half drill strip" method, two varieties may be compared on a total area of about one acre in one year with a standard deviation of a comparison of less than 1°/$_{\circ}$. This combines the advantage of growing corn on the large scale with an accuracy almost as great as that of small scale work; and is within the powers of anyone who can combine the necessary knowledge and patience with the control of skilled agricultural labour.

It is shown that methods (2) and (3) depend for their accuracy on the fact that the nearer two plots of ground are situated, the more highly are the yields correlated, so that we are able to increase the effect of the last term of the equation

$$\sigma^2_{A-B} = \sigma^2_A + \sigma^2_B - 2\, r_{AB}\, \sigma_A\, \sigma_B$$

(where A and B are the varieties to be compared) by placing the plots to be compared with one another as near together as possible.

A formula, due to Mr R. A. Fisher, is given for calculating the error of a comparison in a "chessboard" experiment, which may perhaps be found useful elsewhere.

Finally I have to thank Dr Beaven both for allowing me to use his experimental material and for much invaluable assistance in the preparation of the paper.

Addendum.

Since writing the above I have had the advantage of witnessing the harvesting of Dr Beaven's 1923 experiment and of discussing the whole question with him very thoroughly.

He thinks it probable that the whole or a part of the correlation between the yields of the "plots" which together formed a drill strip in the 1921 experiment may have been due to slight differences in area consequent on irregular steering of the seed drill, such as would have been caused by the horses pulling unequally.

Measurements which we made on the stubble of the similar 1923 experiment showed not only that such inaccuracies occur, but also that they can favour one of the varieties.

It is however a fairly easy matter after harvest to measure the total width from the outside drill of one half drill strip to the outside drill of the same variety. This measurement includes the space between the drill strips which is variable owing to the difficulty of steering and is now made in practice across each drill strip in several places.

It is thus possible to estimate accurately the total area occupied by each variety and to make the necessary correction to the total yields.

As however it would hardly be possible to correct the individual drill strips or "plots" which are used for the purpose of calculating the error, that calculated error will be in excess of the truth.

In Dr Beaven's opinion the operation of taking differences has for all practical purposes eliminated the correlation due to the position of the "plots," and in view of the other causes of variation in the differences, numerous and diverse as they are, he still considers it legitimate to treat the differences between the "plots" as if they were random, and to use the formula $\frac{\sigma}{\sqrt{n}}$ in calculating the error of his mean difference. I feel however that a single operation of this nature is hardly likely to eliminate all the correlation and that there is need for further enquiry: if as the result of a number of experiments it is found that the error of the mean difference calculated from the weights of the half drill strips is not significantly greater than that calculated from the "plots," then the latter undoubtedly provide the more accurate data for the calculation of that error, and it will be a matter of indifference whether the drill strips be few and long or short and many.

Meanwhile they should be made as numerous as is consistent with the successful carrying out of the various agricultural operations which are of course made infinitely more difficult and tedious by the necessity of turning horses and machines at the end of each short length.

But whether we use few long or many short strips is not a question of the first importance: in either case the method is without doubt the best that has hitherto been devised for large scale experiments.