HILGARDIA

A Journal of Agricultural Science Published by the California Agricultural Experiment Station

VOLUME 27 SEPTEMBER, 1957 NUMBER 5

IMPLICATIONS OF A UNIFORMITY TRIAL WITH SMALL PLOTS OF WHEAT

G. A. BAKER
and
E. B. ROESSLER

UNIVERSITY OF CALIFORNIA · BERKELEY, CALIFORNIA

A field trial with small plots of wheat, which can be regarded as a uniformity trial, is presented. This trial was examined in a manner similar to that used on other field trials, with strikingly contrasting results. Results of randomization trials in other studies have differed markedly from the conventional published analysis-of-variance tests based on the F tables. This trial, however, although judged poor by experienced investigators because of weed infestation, soil variation, and stand variation, seems to correspond very closely in many important respects to the model and to resulting tests given in the textbooks.

H I L G A R D I A

A Journal of Agricultural Science Published by the California Agricultural Experiment Station

VOL. 27 SEPTEMBER, 1957 No. 5

IMPLICATIONS OF A UNIFORMITY TRIAL WITH SMALL PLOTS OF WHEAT[1]

G. A. BAKER[2] and E. B. ROESSLER[3]

INTRODUCTION

ANALYSES of large-scale yield trials as conventionally conducted sometimes lead to difficulties in interpretation. This fact was pointed out in detail by Riddle and Baker (1944) and by Baker and Briggs (1956) for wheat trials based on small plots. The fact that large plots do not improve the situation very much was shown by Baker, Huberty, and Veihmeyer (1952) with data for a 10-year uniformity trial with barley. Further studies of uniformity trials were made by Baker and Baker (1953) for strawberries, and a theoretical analysis of a simple uniformity trial was published by Baker (1952).

The present data were not designed to give a uniformity trial, but they can be so regarded since the treatment effect they attempted to measure appears to be small or zero. Since an adequate theory of field trials should be based on as extensive an investigation of uniformity trials as possible, and since the number of such trials is small and such studies involve considerable time and expense, it is felt that the data for this trial merit presentation. We shall examine some of the implications of using the yields to detect actual differences if they exist; that is, we shall consider probabilities of errors of the first and second kinds.

The trial consists of sixteen $40' \times 40'$ blocks subdivided into nine plots each. The data were secured in 1939–1940 from White Federation wheat sowed at the rate of 54 pounds per acre. The yields of the plots are given in table 1. The design of the experiment was square with alleys 20 feet wide between blocks. The plots were 10 feet long with two guard rows on each side. The yields given are for cleaned seed.

The trial was originally designed to test the effect of direction of ploughing on yield, but the experiment did not give evidence of any effect; for our present purpose, therefore, the data can be regarded as a uniformity trial.

Morning-glories infested the middle two columns to such an extent that the yields were somewhat depressed. However, the rows do not differ sig-

[1] Received for publication March 13, 1957.

[2] Professor of Mathematics and Statistician in the Experiment Station, Davis.

[3] Professor of Mathematics and Statistician in the Experiment Station, Davis.

nificantly; furthermore, the over-all variation of the yields in the two outside columns of blocks does not differ significantly from the similar variation for the two middle columns of blocks. This situation corresponds to a usual occurrence with field trials where very often some of the sets of blocks differ greatly in yield from other sets. In such cases the cause of the differences in yield may be hidden beneath the surface of the soil but is nonetheless real. The infestation of morning-glories was quite uniform over the blocks affected.

The intra-block variation does not bear any apparent relation to block mean yield.

TABLE 1

PLOT YIELDS IN GRAMS OF CLEAN WHEAT SEEDS FOR A TRIAL OF 16 BLOCKS

40′ × 40′ blocks each divided into 9 plots

3475	5290	4400	2140	1810	2060	2830	2350	2200	5180	5130	5410
2760	4325	3560	1320	1720	2050	2170	2950	3140	4930	5660	5520
3655	3920	3020	1705	1510	2860	2110	2800	2855	5060	4300	4235
4615	4630	2440	1590	2925	3275	4050	3710	2460	4565	4425	4270
4190	5320	3300	3190	3110	3670	3035	3550	2890	5130	5265	4790
4470	4595	3910	1730	2250	3535	2480	2640	2545	4540	4765	4625
4510	3510	4530	1610	2090	2345	4550	2560	2790	5185	5505	5200
3745	2690	4350	1670	2540	1070	4615	4580	3165	4965	4890	5165
2755	3600	4080	1240	1740	2370	4765	3265	2490	4795	4950	5165
2520	3805	4530	1550	2170	4370	4665	3345	3880	3920	4350	4830
3345	5420	4265	1350	1970	2895	4440	3185	3300	3485	4030	4465
3220	3710	3950	1100	1920	2030	4070	3165	2855	3930	3650	3800

1. Blocks plowed north and south alternate with blocks plowed east and west in a checkerboard arrangement. The upper left-hand block was plowed north and south. North is towards the bottom of the page.
2. The middle two columns of blocks were infested with morning-glories. This land was also lower and more adobe-like.

ERRORS OF THE FIRST KIND

An error of the first kind is committed if a significant difference is said to exist when in fact no such difference is present.

The data of table 1 could well be regarded as a randomized block experiment consisting of nine varieties (or treatments) replicated 16 times. One hundred randomized samples of different assignments of dummy varieties were drawn from the data of table 1. It is noted that for 8 and 120 degrees of freedom $F_{.05} = 2.01$ and $F_{.01} = 2.65$ in the conventional analysis-of-variance model which is supposed to apply, approximately at least, to the F values calculated in the manner described. Actually, four observed F values were greater than 2.01 and none was greater than 2.65.

The distribution of F values seemed to be curtailed, that is, lacking in extreme values. In his discussion of uniformity field trials, Baker (1952) pointed out that when differences in fertility levels of subplots are not included in experimental error, such curtailed distributions of F values are possible, and the probability of errors of the second kind may then be greater than expected.

Latin square trials for these data were investigated with similar results. By randomization there are two Latin squares for each of the large blocks of table 1. Thus 32 analyses of variance for 3×3 Latin squares are possible and these were computed. Three F's were significant at the 5 per cent level and none at the 1 per cent level. This agreement is satisfactory as far as errors of the first kind are concerned.

ERRORS OF THE SECOND KIND

An error of the second kind is made if a difference is said not to exist when in fact a difference is present in the tested situation.

To investigate the behavior of errors of the second kind we assigned varietal differences in grams as follows:

I −300, −225, −150, −75, 0, 75, 150, 225, 300
II −600, −450, −300, −150, 0, 150, 300, 450, 600

Tang (1938) defines

$$\phi = \frac{\sqrt{\frac{1}{k}\sum \delta_i^2}}{\sigma/\sqrt{n}}$$

where δ_i is the difference of the i^{th} variety from the mean of all varieties, k is the number of varieties, n is the number of replications, and σ is the standard error of the experiment. In this case we have a randomized block experiment with 16 replications and nine varieties. The standard error of the experiment is 670. Thus, $k = 9$, $n = 16$, $\sigma = 670$. Hence, ϕ for I is 1.16, and ϕ for II is 2.32.

According to Tang's tables the probability of detecting a difference for I is about 0.6 at the 5 per cent level and 0.4 at the 1 per cent level. Actually,

58 samples out of 100 indicated differences at the 5 per cent level and 40 at the 1 per cent level. The difference between the best and worst "variety" in this case is 600 grams, or 17 per cent of the general mean. These figures indicate a better conformity with expectation than was obtained in earlier studies (Baker, 1952; Baker and Baker, 1953).

For the differences for II we would expect that all samples would detect differences, and such was the case.

Suppose we try a little further with nine varieties and three replications. The replications are to be selected at random in sets of three from the sixteen blocks. The varietal differences applied were –1,000, –750, –500, –250, 0, 250, 500, 750, 1,000. Tang's tables indicate that 82 samples should be detected at the 5 per cent level and 47 at the 1 per cent level. Actually 75 samples were judged significant at the 5 per cent level and 56 at the 1 per cent level.

TWO BLOCKS OF TWO SUBPLOTS EACH

We picked 64 sets of four adjacent plots and analyzed them as trials of two randomized blocks with two subplots, each with and without variety differences. The same procedure was followed for the case of two blocks which were considerably separated. There were 62 such sets of this latter kind. Randomization as explained by Baker (1952) gave twice as many in each category, that is, 128 in the first set and 124 in the second. The differences imposed on the varieties were +350 grams and –350 grams. These amounts are about 10 per cent of the over-all general mean (3,509.9 grams). Tang's tables indicate that the probability of detecting such difference is small.

Since the F distribution for one-and-one degrees of freedom is

$$f(F) = 1/(\pi(1+F)F^{1/2}) \quad (1)$$

it is easy to find the cumulative distribution of F as

$$C(F) = (2/\pi) \text{ arc tan } (F^{1/2}) \quad (2)$$

It is found that the cumulative distribution of the F's for Class I of adjacent blocks with no varietal differences is very close to equation (2), while the cumulative distribution of F's for Class II (varietal differences imposed) for adjacent blocks shows a proper deviation from equation (2) and some tendency to indicate the superimposed varietal differences.

In contrast, for separated blocks the real differences are not detected and the Class I F's do not correspond so well to equation (2), since they run behind and then ahead in cumulative frequency, which means that the medium values are more probable than they should be.

CORRELATIONS OF PLOT YIELDS

The correlation of the yields of adjacent plots is 0.77 and for plots one plot apart is 0.68. The first correlation is based on 192 possible pairs and the second correlation on 96 pairs.

Most of this correlation is due to the great differences between yields of whole blocks. The intra-class correlation for the 16 blocks with nine plots each is 0.71. A large intra-class correlation indicates relatively small variation among the individuals of the blocks. In the limiting case where there is no variation within the blocks the intra-class correlation is 1.0.

As a further check on the correlation situation, we took 64 pairs of deviations from the block means, 4 pairs from each block. Starting at the upper left-hand corner, we took the first pair in the first row, the third in the first row with the one below it, the other two in the second row as another pair, and finally the first two in the third row as the fourth pair. Every other one as we went along was called the first of the pair, and the other was the second. The correlation coefficient between the pairs of deviations from the block means chosen in this way is 0.05, verifying the surmise made above.

TABLE 2

THE NUMBER OF TIMES EACH OF THE RANKS 1–9 WAS OBSERVED IN 100 SAMPLES FOR EACH OF THE NINE TRUE RANKS FOR TRIALS OF THREE REPLICATIONS AND NINE VARIETIES

True rank	Observed rank									Average rank
	1	2	3	4	5	6	7	8	9	
1	54	30	12	3	0	1	0	0	0	1.68
2	30	38	15	11	4	1	1	0	0	2.28
3	13	16	25	27	12	5	1	1	0	3.33
4	2	11	25	29	18	10	5	0	0	4.00
5	1	3	15	19	25	25	8	4	0	4.91
6	0	1	6	4	25	25	24	12	3	6.02
7	0	1	2	5	10	22	27	19	14	6.77
8	0	0	0	1	4	7	21	47	20	7.69
9	0	0	0	1	2	4	13	17	63	8.32

RANKING OF VARIETIES

Our study of maintenance of the proper rank was made on the simulated experiments referred to under the section on errors of the second kind for the case where three blocks were picked from the sixteenth at random and then nine varieties were assigned at random to the nine plots. To repeat, the varietal differences applied were –1,000, –750, –500, –250, 0, 250, 500, 750, 1,000. The comparison of the true ranks with the observed ranks is made in table 2. The differences between the average ranks show an interesting bias.

SUMMARY

We have presented a field trial with small plots of wheat that can be regarded as a uniformity trial. The yields on about half of the plots were severely disturbed by an infestation of morning-glories. It should be noted

that the depression of yields for the infested plots may not have been due solely to the prevalence of the invading plants, since these plots were somewhat different topographically (mainly lower) and also differed noticeably in important soil characteristics (heavier or more adobe-like).

This trial was examined in a manner similar to that used on other field trials, with strikingly contrasting results. In other studies it has been found that randomization trials gave results that differed markedly from the conventional published analysis-of-variance tests based on the F tables. However, this trial, which was judged poor by experienced investigators because of weed infestation, soil variation, and stand variation, seems to correspond very closely in many important respects to the model and to resulting tests given in the textbooks.

In particular, the errors of the first and second kinds were controlled as expected on the basis of published normal theory results for both randomized blocks and Latin squares. The correlations of adjacent and separated plots within blocks were as small as could be hoped for.

In addition, some work was done on ranking, which shows how well the proper ranks are maintained in a special case. It is noted that the scale of the average ranks shows a distortion which is not unexpected, since ranks smaller than 1 and larger than 9 are not possible.

LITERATURE CITED

BAKER, G. A.
1952. Uniformity field trials when differences in fertility levels of subplots are not included in experimental error. The Annals of Mathematical Statistics **23**: 289–293.

BAKER, G. A., and R. E. BAKER
1953. Strawberry uniformity yield trials. Biometrics **9**: 412–421.

BAKER, G. A., and F. N. BRIGGS
1950. Yield trials with backcross derived lines of wheat. Annals of the Institute of Statistical Mathematics **2**: 61–67.

BAKER, G. A., M. R. HUBERTY, and F. J. VEIHMEYER
1952. A uniformity trial on unirrigated barley of ten years' duration. Agronomy Journal **44**: 267–270.

RIDDLE, O. C., and G. A. BAKER
1944. Biases encountered in large-scale yield tests. Hilgardia **16**: 1–14.

TANG, P. C.
1938. The power function of the analysis of variance tests with tables and illustrations of their use. Statistical Research Memoirs, Vol. 2, University of London.

4*m*-10,'57(C8452)BEB 141

The journal *Hilgardia* is published at irregular intervals, in volumes of about 600 pages. The number of issues per volume varies.

Subscriptions are not sold. The periodical is sent as published only to libraries, or to institutions in foreign countries having publications to offer in exchange.

You may obtain a single copy of any issue free, as long as the supply lasts; please request by volume and issue number from:

Agricultural Publications
Room 22, Giannini Hall
University of California
Berkeley 4, California

The limit to nonresidents of California is 10 separate issues on a single order. A list of the issues still available will be sent on request.