# STATISTICAL EVALUATION OE AGRICULTURAL FIELD EXPERIMENTS

BY
S.N. BAJPAI
AND
A.K. NIGAM
*I.A.S.R.I., New Delhi-12*
(Received : January, 1979)

## 1. INTRODUCTION

Field experimentation plays a major rôle in any agronomic research programme. A large number of agricultural field experiments are conducted every year all over the world. Looking into the huge investments involved in such agronomic experimental programmes and the far reaching implications of the results derived from such experiments, it becomes desirable to examine the utility of such programmes and suggest corrective measures, if required. However, very little thought appears to have been given to devise a procedure for statistical evaluation of agricultural field experiments.

First attempt in this direction is due to Seth *et al* on a suggestion from D.J. Finney. Seth *et al.* measure the amount of information from an experiment in terms of a score, where the score corresponds to the product of '$f$' and $\sqrt{r}$. $f$ being number of linearly independent contrasts giving information on pure response or simple effects and $r$ is number of replications. Thus a $3\times3$ factorial experiment in $r$ randomized blocks would receive a score of $6\sqrt{r}$ marks for each factor. Also, loss of information from an experiment has been defined as the ratio of difference between the score that would have been obtained and the score actually recorded to the total score that would have been obtained for a single replicate. For example, an experiment with treatments as all combination of

$$\begin{bmatrix} 0 \\ 20 \; Lb \; N+15 \; Lb \; P_20_5 \\ 30 \; Lb \; N+15 \; Lb \; P_20_5 \end{bmatrix} \times \begin{bmatrix} 0 \\ 20 \; Lb \; K_2 \\ 40 \; Lb \; K_2 \end{bmatrix}$$

will entail 50 percent loss of information. An experiment with all possible combination of

$$\begin{bmatrix} & 0 & \\ 20 & Lb & N \\ 40 & Lb & N \end{bmatrix} \times \begin{bmatrix} & 0 & \\ 20 & Lb & P_2 0_5 \\ 40 & Lb & P_2 0_5 \end{bmatrix}$$

will give full information. The method of Seth *et al.* ignores objective of the experiment while extracting the amount of information from the experiment. This results in extraction of incomplete information and also loss of meaningful information. As an example, consider the experimext conducted at Baroda (India) for paddy crop under simple fertilizer trial scheme to study the response of $N$ alone and in presence of other nutrients $P$ and $K$. The eight treatment combinations tried were 000, 100, 200, 010, 110, 210, 220 and 221. The design was incomplete factorial in seven randomised blocks. Following Seth *et al.* each of the factors $N$ and $P$ would receive a score of $4\sqrt{7}$ although, we cannot estimate the parameters $u$, $N_L$, $N_Q$, $N_L P_L$, $N_L P_Q$, $N_Q P_Q$ and $N_L P_L K_L$ (because of the singularity of information matrix), this being prime object of the experiment. Had the 8 treatment combinations 000, 100 200, 021, 110, 210, 220 and 221 been tried, we could have estimated all the above mentioned parameters. But the procedure of Seth *et al*, allocates a score of $2\sqrt{7}$ for $N$ which is less than $4\sqrt{7}$ the seore obtained from a singular design. The authors, therefore, in this paper suggest some indices for measuring the amount of information and the precision obtained from an experiment. It is hoped that these indices prove to be useful in evaluating bulk of the agronomic experimental programme.

## 2. The Proposed Index

Suppose there are $n$ experiments conducted in an agronomic research programme. A group of $n$ experiments with common objective are evaluated by the Index I give by

$$I = \frac{\sum^{n} \sqrt{W_1 W_2}}{n} \times 100$$

where summation extenns over all experiments, $W_1$ and $W_2$ assume values between 0 to 1. $W_1$ is the weight assigned depending upon the design properties and $W_2$ corresponds to the precision with which information is available from the experiment. By design property is meant the selection of design matrix whose properties are to be avaluated on thc basis of characterstics like orthogonality, balance, optimality etc. In many cases, it may suffice to relate this with the efficiency factor of the design.

The precision of the experiment depends upon the sensitivity which can be measured by evaluating the power with which inferences have been drawn or coefficient of variation observed in the experiment depending on the size of experiment.

The overall index for entire research programme is given by

$$\text{I (Over all)} \qquad = \sum_{i=1}^{r} I_i \, q_i \Big/ \sum_{i=1}^{r} q_i$$

where entire research programme is divided into $r$ groups. $I_i$ is research index of $i$-th group and $q_i$ is the amount spent on $i$-th group.

## 3. Determination of weights $W_1$ and $W_2$

As $W_1$ and $W_2$ are related with the design used and the precision of the experiment. We accordingly classify agricultural experiments into 3 broad categories as given below :

(*i*) Comparative type experiments such as varietal trials, cultural trials and disease control trials etc. In such trials, problem of fitting a response surface does not arise.

(*ii*) Multi-factor experiments consisting of treatments which are purely quantitative in nature such as manurial trials, where response surface studies can be made.

(*iii*) Factorial experiments conducted with the object of examining the existence of main effects, interaction etc.

## 4. Determination of $W_1$ and $W_2$ for category I type of experiments

Category I experiments are either conducted in complete block design or in incomplete block design. For complete block design, $W_1$ will assume value 1 these are most efficient designs. For incomplete block designs, $W_1$ will assume the value equal to the efficiency factor of the design. For instance, for a B.I.B.D. with parameters, $(v, b, r, k, \lambda)$ the value of $W_1$ will be $\frac{\lambda v}{rk}$. As B.I.B. designs are known to be optimal in the entire class of binary designs, the use of efficiency factor as $W_1$ is further justified. For P B I B and other similar designs not much work is available on optimality aspect. As such the best course is again to take average efficiency factor as $W_1$.

Determination of the value of $W_2$ would largely depend on the size of the experiment.

For large size experiments it may be sufficient to examine the coefficient of variation for evaluating $W_2$. Johnson and Welch (2) observed that the coefficient of variation of the order 33% or higher is not desirable from the point of power of inference drawn from the experiments. Authors have also studied (a separate paper under communication) the distribution of coefficient of variation and are of the view that if plot size is ·02 hectare and number of plots per block does not exceed 8 (which would rarely occur) then about 80% of the experiments fall in category of 13% or less coefficient of variation for wheat, sugarcane and rice crops. We therefore suggest the following working rule for evaluating $W_3$ for most of food crops.

TABLE 4.1

| *% coefficient of variation* | *13% and below* | *13% to <16%* | *16% to <19%* | *19% to <25%* | *25% to <30%* | *30% to 33%* | *Above 33%* |
|---|---|---|---|---|---|---|---|
| Value of $W_2$ | 1 | .9 | .8 | .7 | .5 | .2 | 0 |

For small or moderate size experiments the value of $W_2$ can be computed by one of the two following procedures. The first procedure is applicable in the circumstances where object to experiment is to establish inequality $t_i<t_j$ or $t_i=t_j$ as the case may be among treatment means irrespective of cost considerations. For such experiments power of all elementary contrasts is calculated by taking normal deviate as

$$\text{normal deviate}=\frac{\text{treatment difference}\times\sqrt{r}}{\sqrt{2}\,\sigma},$$

where $\sigma$ is the estimate of standard error per plot observed in the experiment. If out of $C^n_2$ contrast (where $n$ is number of treatments) $k$ contrasts have power exceeding .5 then $W_2=\frac{k}{n}$ The second procedure is applied when cost components involved in the experiments are known or can be evaluated and object is to recommend some treatment in comparison to control or standard treatment. In such a case, $W_2$ will assume value between 0 and 1 according as none, some or all treatments satisfy the inequality suggested by Finney [1].

(Additional expected gain by the treatment over control or standard treatment)>(additional cost of the treatment)....4.1.

The procedure to work out $W_2$ satisfying the above inequality is as follows. Let $y$ represent additional cost of treatments in terms of yield. Then we calculate the quantity $t_0$ as given below

$$t_0=\frac{y\times\sqrt{r}}{\text{S.E. per plot}}$$

where $r$ is number of replications. Now fixing the farmer's risk at 10% (say) and utilizing non central tables given Johnson and Welch [2] we can find the value of non centrality parameter $\delta$ for error d.f. and having $P(t>t_0)=.90$. It is easy to work out the desired yield which would ensure with 90% confidence that in case treatment yield is equivalent or higher than desired yield it would satisfy the inequality (4·1). It is so because

$$\delta=\frac{(\text{desired yield}-\text{control yield})\times\sqrt{r}}{\text{S.E. per plot}}$$

Further, suppose there are, say, 4 treatments other than control treatment and suppose yield of 3 treatments out of 4 is greater than desired calculated as above than $W_2$ for the experiment would assume value 3/4. If yield of 2 treatments is greater than desired yield than $W_2$ would assume value 2/4 and so on.

We shall consider two examples to elucidate the two procedures mentionod above.

For first procedure we consider an experiment with the object to find out suitable time of application of nitrogen @ 120 kg/ha conducted at Government Agriculture Research Station Hardoi (India) during the year 1968. Experiment was conducted in randomised block design having 5 treatments and replications. The plot size was 4·00×3·20 meter. The detais of treatments, their mean values in kg/ha, S.E. per plot, coefficient of variation (C.V.) are given in Table 4·2.

TABLE 4·2

| *Treatments* | *Average yield kg/hec.* |
|---|---|
| $T_1$=Full dose of $N$ applied at sowing time | 5208 |
| $T_2$=$\frac{2}{3}$ dose at sowing time and $\frac{1}{3}$ at 1st irrigation | 4974 |
| $T_3$=$\frac{3}{4}$ dose of $N$ at sowing time<br>$\frac{1}{4}$ dose of $N$ 1st irrigation | 4974 |
| $T_4$=$\frac{1}{2}$ dose of $N$ at sowing time<br>$\frac{1}{4}$ at 1st irrigation and<br>$\frac{1}{4}$ at flowering stage | 5599 |
| $T_5$=$\frac{1}{3}$ dose of $N$ at sowing time<br>$\frac{1}{3}$ at 1st irrigation and<br>$\frac{1}{3}$ at flowering stage | 5104 |
| S.E. per plot=710·9 C.V.=13·74% | |

The power corresponding to elementary contrasts is given in the Table 4.3.

TABLE 4·3

| *Elementary contrast* | *Normal deviate corresponding to elementary contrast* | *Power (1-β) against alternative hypothesis* $t_i > t_j$ | | |
|---|---|---|---|---|
| | | *for 8 d.f. corresponding to error* | *for 12 d.f. corresponding to error* | *for 16 d.f. corresponding to error* |
| $(T_1-T_2)$ | $\frac{(5208-4974)\times\sqrt{3}}{710{\cdot}9\times\sqrt{2}}$ | 0·29 | 0·54 | 0·67 |
| $(T_1-T_3)$ | $\frac{(5208-4974)\times\sqrt{3}}{710{\cdot}9\times\sqrt{2}}$ | 0·29 | 0·54 | 0·67 |
| $(T_4-T_1)$ | $\frac{(5599-5208)\times\sqrt{3}}{710{\cdot}9\times\sqrt{2}}$ | 0·59 | 0·76 | 0·83 |
| $(T_1-T_5)$ | $\frac{(5298-5174)\times\sqrt{3}}{710{\cdot}9\times\sqrt{2}}$ | 0·00 | 0·32 | 0·52 |
| $(T_2-T_3)$ | 0 | 0·00 | 0·00 | 0·00 |
| $(T_2-T_4)$ | $\frac{(5599-4974)\times\sqrt{3}}{710{\cdot}9\times\sqrt{2}}$ | 0·89 | 0·97 | 0·99 |
| $(T_2-T_5)$ | $\frac{(5104-4974)\times\sqrt{3}}{710{\cdot}9\times\sqrt{2}}$ | 0·08 | 0·30 | 0·62 |
| $(T_3-T_4)$ | $\frac{(5599-4974)+\sqrt{3}}{710{\cdot}9\times\sqrt{2}}$ | 0·89 | 0·97 | 0·99 |
| $(T_3-T_5)$ | $\frac{(5104-4974)\times\sqrt{3}}{710{\cdot}9\times\sqrt{2}}$ | 0·08 | 0·30 | 0·62 |
| $(T_4-T_5)$ | $\frac{(5509-5124)\times\sqrt{3}}{710{\cdot}9\times\sqrt{2}}$ | 0·73 | 0·86 | 0·91 |

Now there are 4 treatments out of $\binom{5}{2}=10$ for which power exceeds ·5, hence $W_2$ would assume value 4/10. Had the experimeter used one more replication and assuming Coeff. of variation to be fixed at 13·74 %, 6 contrasts out of 10 would have the power exceeded by ·5 and $W_2$ would have the value 6/10. With 16 degrees of freedom the power of these 6 elementary contrasts increased to such an extent whereby we could make conclusions with great certainty. Also, this example has further justified our earlier emperical study that for large size experiments, if c.v.$<$13% the $W_2$ component measuring precision should have value 1.

Let us take another example to elucidate second procedure for determining $W_2$ for small and moderate size experiments. The example pertains to a green manuring experiment conducted in randomised block design with 4 replications at Govt. Agriculture Research Station, Khopali, (M.S.), India during 1960, 61 and 62. The object was to study the effect of green manuring on paddy crop and to make recommendation for Konkan area. There were four treatments *viz* : $G_0 = N_0$ manuring, $G_1$=Green manuring by Dhainchas $G_3$=Green manuring by Sun hemp and $G_3$=Green manuring by Sesbania. The results were as under :

TABLE 4·4

| *Treatment/ Year* | *Yield of in kg/hectare* | | | | *Sig.* | *S.E./plot* |
|---|---|---|---|---|---|---|
| | $G_0$ | $G_1$ | $G_2$ | $G_3$ | | |
| 1960 | 1710 | 1745 | 2057 | 2081 | ** | 190·0 |
| 1961 | 774 | 1186 | 824 | 899 | * | 214·2 |
| 1962 | 1923 | 2106 | 1998 | 2067 | * | 109·1 |

Now since the design adopted is R.B.D., $W_1$ would assume value 1 for all the three experiments. Also, the cost of raising green manuring crop can be worked out from certain supplementary information available for the experiment to be equivalent to price of 170 kg. of paddy per hectare for almost all the treatments except some nominal differences in seed cost. As such second procedure is opted for computing $W_2$.

TABLE 4·5

**Computation of non-centrality parameter and desired yield**

| *Year* | *Value of* $t_0$ | *Value of* ($\delta$) | *Values of desired yield in (kg)* |
|---|---|---|---|
| 1960 | $\frac{170 \times \sqrt{4}}{190 \cdot 0}$ | 2·9560 | 1991 |
| 1961 | $\frac{170 \times \sqrt{4}}{214 \cdot 2}$ | 2·8550 | 1080 |
| 1962 | $\frac{170 \times \sqrt{4}}{109 \cdot 1}$ | 4·3738 | 2126 |

There are 3 treatments other than control in each year and we observe that in 1960, yields of only two treatments exceed the desired yield ; similarly in 1961 only one treatment and in 1962 none have larger yields than the desired level. Consequently we would fix $W_2$ as 2/3; 1/3 and 0 respectively for these experiments. If we average research indices for the three years we find

$$I = \frac{\sum^{3} \sqrt{W_1 W_2}}{3} \times 100 = 47\%$$

## 5. Determination of $W_1$ and $W_2$ for Category II type Expts.

Main object of multifactor experiment is to approximate the relationship between a set of factors and the response. Once the appropriate model (response surface) has been explored, the next aim is to exploit the fitted model for determining an optimum combination of level of different factors. A necessary condition for fitting an appropriate response surface is the non-singularity of the information matrix $X'X$, where $X_{N \times p}$ is the design matrix. Thus, all experiments which do not allow estimation of the parameters of the model because of the singularity of the information matrix should have the value of $W_1$ as zero. In case of non singular $X'X$ and only exploration of appropriate model being the objective, $W_1$ should be fixed in accordance with one of the optimality criteria.

Some of the well known optimality criteria are $A$-, $D$-, $E$- and $G$- criteria. For the response surface problem $G$- efficiency or optimality criterion is considered to be most appropriate (Snee and Marquardt, [6]). Thus the value of $W_1$ can be taken as the value of the $G$- efficiency of the design. If the objective includes both the exploration as well as the exploitation of the response surface, then $W_1$ is a composite index given by $\sqrt{W_{11} . W_{12}}$, where $W_{11}$ is to be evaluated as explained above and $W_{12}$ can be evaluated as indicated below.

If the optimum is a feasible optimum $W_{12}$ will receive value$_1$ otherwise it will be zero. By an optimum is meant that combination of factor which is expected to provide maximum response (or profit) and a feasible optimum is the one in the range of levels tried in the experiment. Value of $W_2$ can be determined in similar way as for category I type experiment.

## 6. Determination of $W_1$ and $W_2$ for Category III type experiments

Factorial experiments are conducted with the objective of studying main effects of different factors and their inter-actions. The whole theory of usual factorial experiments revolves round the assumption that with the increase in the order of the effects, its importance declines. That is higher order effects are less important compared to lower order effects, Here again. $W_1$ is composed of two indices $W_{11}$ and $W_{12}$ such that $W_1 = \sqrt{W_{11} \cdot W_{12}}$. Here $W_{11}$ is related to the relative importance of the effects. $W_{11}$ will assume the value zero for all confounded experiments in which lower order effects have been confounded although it was possible to confound higher order interactions; otherwise, $W_{11}$ will be 1. For complete factorials $W_{12}$ will be equal to the ratio of the total per d.f. information recovered for all the effects to the total d.f. for all the effects. For instance, in the $5 \times 3 \times 2$ factorial reported in Sardana and Das [4] in 6 plot blocks and with 4 replications the total loss of information for confounded effects $A$, $AB_L$ and $AB_Q$ is 4. Therefore, the $W_{12}$ will be 25/29: 29 being the total degree of freedom. For fractional factorial $W_{11}$ is again related to optimality criterion as in case of category II type experiments. However, in factorial experiments, trace optimality is considered to be more convenient by various workers. $W_{11}$ should, therefore, be the ratio of the trace of $(X'X)^{-1}$ of the optimal design to that of the design under consideration.

Also, note that $W_{11}$ for fractorial factorial would be zero, if the information matrix of the effects desired to be estimated turn out to be zero. In case of split plot type experiments, however, interaction between sub-plot and main plot is more important than $\delta$ effect of main plots. In such experiments $W_{11}$ may be taken as 1 and $W_{12}$ assumes value between 0 and 1 depending upon the extent of confounding for sub-plot treatments.

The value of $W_2$ is computed as described below. For an $s_1 \times s_2 \times \ldots\ldots \times s_k$ experiment, we wrie

$$W_2 = \frac{\sum_{i=0}^{k-1} n_i \, w_{2i}}{n_i}$$

where $n_i$ are the degrees of freedom associated with the $i$th order effects *viz.*, main-effects, first order interactions and so on and $w_{2i}$ are the weight associated with $i$th order effects. $w_{2i}$ are computed as follows. $w_{2i}$ takes the value zero if the expected magnitude of $i$th order effect can not be detected as significant with the coefficient of variation observed in the experiment otherwise it takes the value 1. Needless to mention that the effect can be detected at 5% probability level if it exceeds roughly twice the observed coefficient of variation.

## 7. Discussion

It has been realised by now that no exact formulae can be put forth for evaluating the values of $W_1$ and $W_2$. This necessitated to categorize different experiments into three broad classes so that more objectivity could be achieved in assigning the values of $W_1$ and $W_2$. It is revealed from the studies carried out at the institute under the projects National Index of Field Experiments' and 'Statistical Evaluation of Field Fxperiments that about 90% of the experiments are covered under these category of experiments. As such it is hoped that the procedure set out in this paper would be useful for evaluating bulk of the agronomic research programmes.

Lastly, a point about the overall research index I. It is clear that I can take values between 0 and 100. Although I=100 would be the ideal situation, it would barely be attainable. A working rule for deciding good, satisfactory and bad index can again be proposed. Value of index lying between 70 and 100 may be considered as an indication of a good index. If index lies between 50 and less than 70 it may be rated satisfactory and an index below 50, a bad index. A poor index is a warning to examine into the values of $W_1$ and $W_2$ and to adopt corrective measures. If $W_1$ is low, there is case for adopting alternative designs as the design used for experimentation was devoid of desirable design properties. Similarly, if $W_2$ is found to be low, then it is indicative of faulty execution of the experiment. For instance, in case of a ranomized block experiment, this could be due to faulty formation of blocks or this could be because of the presence of the block-treatment interaction.


## Summary

In this paper a procedure for statistical evaluation of agricultural field experiments is proposed. The procedure consists of evaluating a composite index as a measure of performance of the experiment,


## References


[1] Finney, D.J. (1959) : *The Economic Efficiency of Experimentation.* Publication of the Mathematical institute of Hungarian Academy of Sciences Vol. IV, FASC 3-4 pp. 203-221.

[2] Johnson, N.L. and Welch B.L. (1939) : Applications of the Non Central t-distribution *Biometrika*, Vol. 31, 1939, pp. 362-389.

[3] Resinkoff and Liberman (1957) : *Tables of the Non-Central 't' distribution.* Stanford University Press, Stanford, California 1957.

[4] Sardana, M.G. and Dass M.N. (1965) : On the construction and Analysis of some confounded Asymmetrical Factorial Designs *Biometriks*, Vol. 21, No. 4 December 1965. pp. 949-955.

[5] Seth, G.R., Sukhatme, B.V. Marutiram, (1958) : A survey of Agronomic research programmes in India. *Ind. J. Agri. Sci.* **28** (3); 409-468.

[6] Snee, R.D. and Marquardt, D.W. (1974) : *Extreme vertics designs for linear mixture models Technometrics*, **16**, pp. 399-408.