JOURNAL

OF THE

Indian Society of Agricultural Statistics

Vol. IX] 1957 [No. 2

SOME CONSIDERATIONS IN THE PLANNING AND ANALYSIS OF FERTILIZER EXPERIMENTS IN CULTIVATORS' FIELDS*

BY UTTAM CHAND AND T. P. ABRAHAM

Indian Council of Agricultural Research

1. INTRODUCTION

THE need for carrying out fertiliser trials in cultivators' fields under actual farming conditions so as to provide a sound basis for making practical recommendations on fertilizer use, is now well recognised. The planning and analysis of such experiments involves a synthesis of some of the basic techniques of experimental design and sample surveys. The main restriction in the designing of such experiments is the fact that in a given cultivator's field not more than 5 or 6 plots should be laid out.

Simple fertilizer trials on randomly selected sites with an unreplicated 3-plot experiment at each site superimposed on the normal practice of the cultivator were conducted in some parts of India on the basis of the recommendations made by Stewart[1] in his report on soil fertility investigations in India. It was soon realised that the scope of these experiments for providing information on fertilizer use could be considerably enlarged by increasing the number of plots from 3 to 5 or 6. Experience gained on the basis of large-scale experiments conducted in cultivators' fields by H. N. Mukherjee in Bihar showed that the number of plots could be increased to this extent without impairing the efficiency of such types of experimental programmes.

---

* Parts of this paper were presented at the annual meetings of the Indian Society of Agricultural Statistics held at New Delhi in 1956.

Five and six plots experiments designed to provide information on a number of aspects of fertilizer use such as optimum dose and relative efficiency of different sources of the same fertilizer were planned at the Indian Council of Agricultural Research and conducted in a large number of Community Project centres under a scheme jointly sponsored by the Government of India and U.S. Technical Co-operation Mission (T.C.M.). Certain new ideas of symmetry and balance on the analogy of incomplete block designs were introduced into the planning of such experiments by Yates and Finney during 1953 when they were working as F.A.O. experts with the Indian Council of Agricultural Research. A detailed account of these experiments as well as some of the practical problems involved in the planning of experiments in cultivators' fields were given by Panse and Sukhatme.[2]

The object of the present paper is to give a systematic analysis for some of the typical designs that could be used on cultivators' fields. Analysis of variance tables for different designs have been given for purposes of completeness though it is recognised that experimental investigations on fertiliser use relate mostly to problems of estimation and not to tests of significance. Some aspects of planning of these experiments so as to provide the necessary background for the methods of analysis presented here have also been discussed. Designs considered here are only illustrative to indicate the type of results that can be obtained under a specified scheme of distribution of these experiments in a given region. The problem of allocation of experiments and the total number of experiments taking cost of experimentation and loss resulting from the application of a recommended fertilizer dose below or above the optimum has also been discussed.

2. Some Aspects of Planning

Since each experiment is required to provide a self-contained demonstration, the cultivators' normal practice should be included to provide a control plot and the remaining treatments superimposed on this normal practice. If the objective of the inquiry is as simple as obtaining a dose response relationship for a single fertilizer, a 3 or 4-plot experiment of the type o, n_1, n_2, n_3 where suffixes indicate levels could be accommodated in a field. In a preliminary investigation on the effects of n, p and k, the best choice could be given by a 4-plot experiment of the type o, n, np, npk. It would be noticed that an outstanding consideration for the appropriate choice of treatments would be the number of plots which can be managed in a cultivator's field.

For example if response curves of two sources of a nitrogenous fertilizer are to be compared on soils not expected to respond to phosphate, a 5-plot experiment o, n_1, n_2, n_1', n_2' where n and n' refer to forms and suffixes indicate levels will be required to obtain a symmetrical comparison between the two types. If, however, we had to make the above comparison in four plots we shall have to split the above set into two sub-sets (i) o, n_1, n_2; n_1', (ii) o, n_1', n_2', n_1. Each sub-set will provide information on the response curve of one form and an interior point from the response curve of the other form and equal number of experiments on the two sub-sets of treatments will be carried out in different fields. Any adequate statistical analysis should take into account the design of the experiment and the analysis of the two non-orthogonal 4-plot arrangements may not present any special mathematical difficulty but is bound to be complicated. Similarly comparison of three sources of a nitrogenous fertilizer might be made by carrying out an equal number of experiments for each of the following three 5-plot sets:

(A) (i) o, n_1, n_2, n_1', n_2',

(ii) o, n_1, n_2, n_1'', n_2'',

(iii) o, n_1', n_2', n_1'', n_2''.

In case we decide to make these comparisons in 4-plot experiments, the above three sub-sets will be further split into the following six 4-plot sets:

(B) (i) o, n_1, n_2, n_1',

(ii) o, n_1, n_2, n_1'',

(iii) o, n_1', n_2', n_1,

(iv) o, n_1', n_2', n_1'',

(v) o, n_1'', n_2'', n_1,

(vi) o, n_1'', n_2'', n_1',

where each sub-set provides information on the response curve of one source of nitrogen and a point nearabout the guessed optimum of the other source. It will be noticed that equal precision for both the levels has not been attempted and a sub-set of the type o, n_1, n_2, n_2' has not been provided for to avoid having too many sub-sets for a given choice of treatments.

When more than one nutrient like nitrogen and phosphorus are involved, all combinations of three levels of nitrogen and three levels of phosphate could be investigated by means of the following 6-plot arrangements:

(C) (i) $o, p_1, n_1, n_2, n_1p_1, n_2p_1,$

(ii) $o, p_2, n_1, n_2, n_1p_2, n_2p_2.$

In case it is not possible to accommodate six plots at a given site, Panse and Sukhatme[2] have suggested the following arrangement involving 3 and 4 plots:

(D) (i) $o, n_1, n_2,$

(ii) $o, p_1, n_1p_1, n_2, p_1,$

(iii) $o, p_2, n_1p_2, n_2p_2.$

This arrangement uses the principle of confounding in which levels of phosphate are confounded with block differences. It is interesting to note that if we omit control plots in the second and third sets above, the analogy of this arrangement with a split plot design with levels of nitrogen in sub-plots is clearly brought out.

3. Comparison of a Single Set of Treatments

Suppose that we wish to try a single set of treatments such as *o, n, np, npk* in a given tract such as a Community Project area. We may select at random r villages and in the i-th selected village n_i fields may be selected at random. A single unreplicated experiment may be carried out in each field. Let $Y_{ijk} = \mu + v_i + f_{ij} + t_k + a_{ik} + b_{ijk}$ where μ denotes the general mean, v_i the effect of the i-th village, f_{ij} the effect of the j-th field in the i-th village, t_k the average effect of the k-th treatment, a_{ik} the interaction of the k-th treatment effect with the i-th village and b_{ijk} the interaction of the k-th treatment effect with the j-th field in the i-th village. With this set up the relevant portion of the analysis of variance for the estimation of error or responses will be as given below:

Table I

Source	d.f.	M.S.	Expected Value
Villages × treatments ..	$3(r-1)$	s_1^2	$\sigma_{ft}^2 + \lambda\sigma_{vt}^2$
Fields within villages × treatments	$3\Sigma(n_i - 1)$	s_2^2	σ_{ft}^2

where

$$\lambda = \frac{1}{r-1}\left(\Sigma\, n_i - \frac{\Sigma\, n_i^2}{\Sigma\, n_i}\right), \quad Ea_{ik}^2 = \sigma_{vt}^2$$

and

$$Eb^2_{ijk} = \sigma^2_{ft}.$$

The variance of a response will be given by

$$V = 2\left[\frac{\sum_1^r n_i^2\,\sigma_{vt}^2}{\left(\sum_1^r n_i\right)^2} + \frac{\sigma_{ft}^2}{\sum_1^r n_i}\right].$$

Estimates of

$$\sigma_{vt}^2 \text{ and } \sigma_{ft}^2$$

can be obtained from the above table of analysis of variance as

$$\hat{\sigma}_{ft}^2 = s_2^2 \text{ and } \sigma_{vt}^2 = \frac{s_1^2 - s_2^2}{\lambda}.$$

A two-stage sampling procedure has been adopted in the above case for the distribution of the experiments. The procedure can be extended to a case when the selection of fields is made in more than two stages.

We may select a random number say m thanas (small administrative units) in each of the n districts and conduct q experiments in randomly selected fields from each thana. This pattern of distribution of experiments was adopted in the case of manurial trials carried out in the Bihar State.

Let y_{ijkl} represent the yield of the plot in the i th district, j th thana, k th experiment and l th treatment. We, therefore, have:

$$Y_{ijkl} = \mu + \delta_i + L_{ij} + \beta_{ijk} + \tau_l + \gamma_{il} + \eta_{ijl} + e_{ijkl}$$

where μ is the general mean, δ_i is the average deviation for district i, τ_l the average deviation for treatment l and L_{ij} and η_{ijl} are the average deviations for the j th thana in the i th district and variations in this from treatment to treatment respectively, γ_{il} is an additional deviation for treatment l common to the whole of the district i and β_{ijk} is a deviation for the particular site of an experiment within a thana, e_{ijkl} includes both the experimental error and field to field variation of the l-th treatment within a thana.

The following analysis of variance relating to interaction variances will be used to estimate different components of variance for finding standard errors of responses.

TABLE II

Source	Degree of Freedom	M.S.	Expected Value
District × treatments	$3(n-1)$	s_1^2	$\sigma^2_{dcft} + q\sigma^2_{dct} + mq\sigma_{dt}^2$
Thanas × treatments	$3(m-1)$	s_2^2	$\sigma^2_{dcft} + q\sigma^2_{dct}$
Fields within thanas × treatments	$3\,nm(q-1)$	s_3^2	σ^2_{dcft}

where

$$E\gamma_{il}^2 = \sigma_{dt}^2, \; E\eta^2_{ijl} = \sigma^2_{dct} \text{ and } E\,e^2_{ijkl} = \sigma^2_{dcft}.$$

It will frequently happen that although an equal number of experiments were planned in each thana, results of all such experiments are not available for analysis due to certain unavoidable reasons; thereby introducing an extra complication in the usual heavy computation. However, for obtaining satisfactory estimates of standard errors it is not necessary that the analysis of variance should be based on all the experiments. Therefore the analysis of variance might as well be performed on a random sample of completed experiments per district taking equal numbers from each thana.

Results of such experiments will be usually classified and grouped according to broader soil types cutting across district borders. Let $R_N X$ denote the response to nitrogen averaged over N experiments belonging to a given soil type of which n_{ij} experiments come from the i-th district and j-th thana. It is readily seen that,

$$V(R_N) = \frac{2\sigma^2_{dcft}}{N} + 2\sigma^2_{dct} \sum_{i,j} \frac{n_{ij}^2}{N^2} + 2\sigma_{dt}^2 \sum_{i} \frac{n_i^2}{N^2}$$

where

$$\sum_j n_{ij} = n_i, \; \sum_i n_i = N.$$

If we wish to compare the response to nitrogen on two soil types, then

$$V(R_N - R_{N'}) = 2\sigma^2_{dcft}\left(\frac{1}{N} + \frac{1}{N'}\right) + 2\sigma^2_{dct} \sum_{ij} \left(\frac{n_{ij}}{N} - \frac{n_{ij}'}{N'}\right)^2$$

$$+ 2\sigma^2_{dt} \sum_{i} \left(\frac{n_i}{N} - \frac{n_i'}{N'}\right)^2$$

where n_{ij}' experiments contribute to the response on the second soil type and where $\sum_j n_{ij}' = n_i'$, $\Sigma n_i' = N'$. Summations here are taken over all thanas in which either of n_{ij}, n_{ij}' is non-zero and over all districts in which either n_i and n_i' is non-zero. Estimates of standard errors may now be easily written down by replacing σ_n^2 by appropriate functions of mean squares in the above formula. For example, s_3^2 will estimate σ^2_{dcft},

$$\frac{(s_2^2 - s_3^2)}{q}$$

will estimate σ^2_{dct} and $(s_1^2 - s_2^2)/mq$ will estimate σ_{dt}^2.

In the foregoing analysis we have assumed homogeneity of different interaction variances. In situations where this is not true, estimates and their standard errors may be built for each thana separately and then combined appropriately to obtain district and State estimates of various responses. Statistical aspects of combination of estimates from different experiments are exhaustively discussed by Cochran[3] and the choice of suitable estimates and their standard errors will be chiefly determined by the nature of the experimental data.

4. Comparison of Qualities and Levels of a Single Fertilizer in 5-Plot Experiments

Case I.—Suppose the three sets marked (A) as given in Section 2 are arranged in a random sample of $3r$ fields in a given region such that r experiments are allocated to each of the sets. It is clear that the treatments are not orthogonal with fields, since only 5 out of the 7 treatments are tried in each field. Estimates of responses or differences in response for different sources can be obtained by combining suitably the estimates from the r experiments of each type. However, a combined analysis of the $3r$ experiments can be carried out without much difficulty; and such an analysis, apart from providing best estimates, facilitates the overall tests of significance. The overall analysis takes a relatively simple form on account of the presence of balance and symmetry in the grouping of treatments into the 3 sets. We shall describe both methods of analysis and compare their relative efficiency.

(*a*) *Combination of estimates from individual types.*—As a first step the analysis of variance of each type is carried out separately giving for a given type the following partitioning of d.f.:

Source of variation	d.f.
Fields	$r - 1$
Treatments	4
Fields × treatments	$4(r-1)$

The interaction mean square pooled over all the 3 types provides an estimate s^2 of error variation. The response to any one of the levels of a fertilizer is obtained from the means of the control and the corresponding treatment plot of the $2r$ experiments where the given fertilizer was applied and its variance will be estimated by s^2/r. To estimate the difference in response to the fertilizers at a given level say $n_1 - n_1'$, we have a direct estimate say R_1 of the response difference from type 1 where they are both tried together with an estimated variance $2s^2/r$. An indirect estimate of the same response difference is obtained by taking

$$\left(n_1 - \frac{o + n_1'' + n_2''}{3}\right)$$

from type 2 and

$$\left(n_1' - \frac{o + n_1'' + n_2''}{3}\right)$$

from type 3 and subtracting the latter from the former. This estimate say R_2 has an estimated variance of $8\,s^2/3r$. The two estimates are uncorrelated and, therefore, the best linear combination of the estimates is given by taking the average of the two estimates weighted inversely to the variance. This estimate is given by:

$$R = \frac{4\,R_1 + 3\,R_2}{7}$$

with an estimated variance of $8\,s^2/7\,r$. Similarly the estimates of differences in response to the other level and qualities can be found.

(*b*) *Analysis under the combined model*—

$$y_{ij} = \mu + b_i + t_j + e_{ij}$$

represent the yield under the j-th treatment in the i-th experiment where without loss of generality the suffix i takes the value 1 to r for fields containing the first set of treatments, $i = r + 1$ to $2r$ for fields containing the second treatments and $i = 2r + 1$ to $3r$ for fields containing the third set of treatments and where the suffix j takes the value 0 to 6 for the treatments o, n_1, n_2, n_1', n_2', n_1'', n_2''. Since the two levels of a given fertilizer always occur together in a set and the different forms are compared symmetrically, the comparisons of levels and forms $\times$ levels are orthogonal to blocks and other effects. Therefore a reparametrization of the treatment parameters as given below will result in some simplification in the estimation and testing of treatment effects. Let

$$t_0 = \theta_0;\ t_1 + t_2 = 2\theta_1,\ t_3 + t_4 = 2\,\theta_2,\ t_5 + t_6 = 2\theta_3$$
$$t_1 - t_2 = 2\phi_1,\ t_3 - t_4 = 2\phi_2,\ t_5 - t_6 = 2\phi_3.$$

We shall estimate the different parameters subject to the restrictions $\sum_{i=1}^{3r} \hat{b}_i$ and $3\hat{\theta}_0 + 4(\hat{\theta}_1 + \hat{\theta}_2 + \hat{\theta}_3) = 0$ where $\hat{b}$ and $\hat{\theta}$ are estimates of the corresponding parameters. The normal equations under this set up can be easily written down. Let G stand for the grand total and T's denote the sum of yields of the plots in which the corresponding treatment occurs and B_i represent the total of the i-th field containing the first set of treatments, etc. We now set

$$G_1 = \sum_{i=1}^{r} B_i\ ;\quad G_2 = \sum_{j=r+1}^{2r} B_j\ ;\quad G_3 = \sum_{k=2r+1}^{3r} B_k\ ;$$

$$Q_0 = T_0 - \frac{G}{5}\ ;\quad Q_1 = T_1 + T_2 - \tfrac{2}{5}(G_1 + G_2);$$

$$Q_2 = T_3 + T_4 - \tfrac{2}{5}(G_1 + G_2);\quad Q_3 = T_5 + T_6 - \tfrac{2}{5}(G_2 + G_3).$$

After simplification of the normal equations we obtain

$$\hat{\theta}_0 = \frac{Q_0}{3r},\quad \hat{\theta}_1 = \frac{1}{16r}\left(5\,Q_1 + \frac{Q_0}{3}\right);\quad \hat{\theta}_2 = \frac{1}{16r}\left(5\,Q_2 + \frac{Q_0}{3}\right);$$

$$\hat{\theta}_3 = \frac{1}{16\,r}\left(5\,Q_3 + \frac{Q_0}{3}\right);$$

$$\hat{\phi}_1 = \frac{P_1}{4r};\quad \hat{\phi}_2 = \frac{P_2}{4r}\ ;\quad \hat{\phi}_3 = \frac{P_3}{4r}$$

where we have set $T_1 - T_2 = P_1;\quad T_3 - T_4 = P_2;\quad T_5 - T_6 = P_3.$ The estimates of treatment means and their variances can now be obtained, *e.g.*,

$$\hat{\mu} + \hat{t}_1 = \frac{G}{15r} + \frac{Q_0}{48r} + \frac{5Q_1}{16r} + \frac{P_1}{4r}$$

with a variance of $13\sigma^2/24r$.

The analysis of variance can be carried out as follows: The total S.S. for treatments is obtained by first finding the sum of squares for all fitted constants which is given by

$$\sum^{3r} \frac{B_i^2}{5} + \frac{5}{16r}\sum_0^3 Q_i^2 + \sum_1^3 \frac{P_i^2}{47}$$

and subtracting from this, the sum of squares obtained by fitting constants for blocks ignoring treatments. The S.S. for treatments obtained in this way will be

$$\frac{5}{16r}\sum_{0}^{3} Q_i^2 + \sum_{1}^{3}\frac{P_i^2}{4r}.$$

The error S.S. can be obtained by subtracting from the total S.S., the S.S. for blocks ignoring treatments + the S.S. for treatments adjusted for blocks. An overall test of significance of the treatments can then be carried out as usual by the F test for treatment mean square/error mean square.

As the factorial set of treatments included are of a qualitative *cum* quantitative nature, the useful tests of significance are on the main effects and interactions, *viz.*, (1) average difference between forms, (2) average difference between levels and (3) interaction of level with forms. Due to the orthogonal property of levels and levels × forms referred to earlier, the S.S. for these items can, therefore, be obtained as in the case of an orthogonal design. For obtaining the S.S. for forms and control *versus* treatments, we use the well-known technique of subtracting the S.S. due to a specified hypothesis from the total unrestricted hypothesis. For forms the hypothesis is $\theta_1 = \theta_2 = \theta_3$ and for control *versus* treatments it is $\theta_0 = 0$. The complete analysis of variance along with the expected values of mean squares is given in Table III.

The efficiency of estimates by the two methods will be compared in Table IV.

It will be seen from Table IV that the relative efficiency of estimates under the combined model is not appreciably different from that obtained from combination of estimates from different sets except under (i). In this case the higher efficiency under the combined model is due to utilizing information from all the control plots, while in the other method, controls from only two of the sets have been used. As the comparison (*a*) between two forms, (*b*) between the two levels of any one form and (*c*) between the two non-zero levels, are orthogonal to the different sets, both the estimates and their variances will be the same in the two cases.

Case II.—It might be considered desirable to allocate all the three sets of treatments to a cluster of fields in a village. To distribute a total of $3r$ experiments in an administrative unit say a tehsil or a Community Project area, one might select r villages at random and

TABLE III

Analysis of Variance

Source	d.f.	S.S.	Expected value of mean square
(1) Fields (ignoring treatments)	$3r-1$	$\sum \frac{B_i^2}{5} - \frac{G^2}{15r}$	
(2) Treatments (eliminating blocks)	6	$\frac{5}{16r}\sum_{i=0}^{3} Q_i^2 + \sum_{1}^{3} \frac{p_i^2}{4r}$	$\sigma^2 + \frac{8r}{5}\sum \frac{(\theta_i - \bar{\theta})^2}{3} + \frac{2r}{3}\sum \phi_i^2$
(*a*) Forms ..	2	$\frac{5}{16r}\, dev^2 (Q_1, Q_2, Q_3)$	$\sigma^2 + \frac{16r}{5}\sum_{i=1}^{3} \frac{(\theta_i - \bar{\theta})^2}{2}$
(*b*) Levels ..	1	$\frac{(\Sigma p_i)^2}{12r}$	$\sigma^2 + 12r\bar{\phi}^2$
(*c*) Forms × levels	2	$\frac{1}{4r}\, dev^2 (p_1, p_2, p_3)$	$\sigma^2 + 4r \frac{\Sigma (\phi_i - \bar{\phi})^2}{2}$
(*d*) Control *versus* rest	1	$\frac{5}{16r} Q_0^2$	$\sigma^2 + \frac{45}{16} r\, \theta_0^2$
(3) Error ..	$12r - 6$	(4) — {(1) + (2)}	σ^2
TOTAL ..	$15r-1$	Σy_{ij}^2 — C.F.	

then select three fields at random in each village. The mathematical model will then be given by

$$Y_{ijk} = \mu + v_i + f_{ij} + t_k + \lambda_{ik} + e_{ijk}$$

where

$$i = 1, \ldots . r; \quad j = 1, 2, 3; \quad k = 0, 1, 2 \ldots . 6$$

where v_i represents the effect of the i-th village, f_{ij} stands for the effect of the j-th field in the i-th village and λ_{ik} corresponds to variation in

TABLE IV

Variance of Estimates

	Combination of estimates from different sets	Estimates under the combined model	Efficiency of estimates under the combined model (%)
(i) Response to a single *versus* double level of any one form	$\frac{\sigma^2}{r}$	$\frac{7\sigma^2}{8r}$	114·67
(ii) Difference between two forms at any specified level	$\frac{8}{7}\frac{\sigma^2}{r}$	$\frac{9\sigma^2}{8r}$	101·69
(iii) Difference between two forms	$\frac{5}{8}\frac{\sigma^2}{r}$	$\frac{5\sigma^2}{8r}$	..
(iv) Difference between the two levels of any one form	$\frac{\sigma^2}{r}$	$\frac{\sigma^2}{r}$	..
(v) Difference between levels	$\frac{\sigma^2}{3r}$	$\frac{\sigma^2}{3r}$	..

the response from village to village. The analysis of this model reduces to the analysis of the three sets of treatments for each village and then combining for all the villages in the manner shown below.

Let F_{ij} denote the yield total for the j-th field in the i-th village. Set $\sum_{j=1}^{3} F_{ij} = V_i$, the total for the i-th village. Further let

$$Q_{is}\ (i = 1, \ldots r;\ s = 0, 1, 2)$$

and

$$P_{ij}\ (i = 1 \ldots r;\ j = 1, 2, 3)$$

denote the Q and P quantities defined under Case I for each village. If we write $\sum_i Q_{is} = Q_s$ and $\sum_i P_{ij} = P_j$ the analysis of the model under discussion works out as given in Table V.

The sum of squares for treatments and treatments × villages may easily be computed by forming two-way tables for villages × Q's and for villages × P's. It will be further noticed that if individual village

TABLE V

Analysis of Variance

Source	d.f.	S.S.	Expected value of mean squares
Village	$r-1$	$\frac{\Sigma V_i^2}{15} - \frac{G^2}{15r}$	
Fields within villages (ignoring treatments)	$2r$	$\sum_i \left(\sum_j \frac{F_{ij}^2}{5} - \frac{V_i^2}{15} \right)$	
Treatments (eliminating blocks)	6	$\frac{5}{16r} \sum_{s=0}^{3} Q_s^2 + \sum_{j=1} \frac{P_j^2}{4r}$	$\sigma^2 + \frac{9}{5} \sigma_\lambda^2 + \frac{8}{5} r \sum_i^3 \frac{(\theta_i - \overline{\theta})^2}{3} + \frac{2}{3} r \sum \phi_i^2$
Treatments × villages	$6(r-1)$	$\frac{5 \sum_i \sum_s Q_{is}^2}{16r} + \sum_j \sum_i P_{ij}^2 - \frac{5 \Sigma Q_s^2}{16r} - \frac{\Sigma P_j^2}{4r}$	$\sigma^2 + \frac{9}{5} \sigma_\lambda^2$
Error	$6r$	By difference	σ^2
TOTAL ..	$15r - 1$		

analysis is available, then error S.S. will be added to obtain the total error S.S. with $6r$ d.f. Moreover the total of $6r$ d.f. for the adjusted S.S. due to treatments for the r villages will be split into 6 d.f. for the S.S. due to treatments and $6(r-1)$ d.f. for the interaction.

5. COMPARISON OF QUALITIES AND LEVELS OF A SINGLE FERTILIZER IN 4-PLOT ARRANGEMENTS

Suppose that the six sets of treatments given in Section 2 and marked as (B) are arranged in a random sample of $6r$ fields in a given region such that r experiments are allocated to each set. The mathematical model will be the same as discussed under Case I of Section 4 where without loss of generality the suffix i takes the value

$1, 2 \ldots r;\ r+1, \ldots 2r;\ 2r+1;\ 3r;\ 4r+1, \ldots 5r;\ 5r+1 \ldots 6r,$ for fields containing the first, second, third, fourth, fifth and sixth set of treatments respectively.

To reparametrize the treatment parameters we set

$$t_0 = \theta_0 \qquad t_2 = \theta + \theta_1$$
$$t_1 = \phi + \phi_1 \qquad t_4 = \theta + \theta_2$$
$$t_3 = \phi + \phi_2 \qquad t_6 = \theta + \theta_3$$
$$t_5 = \phi + \phi_3$$

where $\Sigma \phi_i = 0$ and $\Sigma \theta_i = 0$. We shall estimate the different parameters subject to the additional restriction $\sum_{i=1}^{6r} b_i = 0$ and $\theta_0 + 2\phi + \theta = 0$. Let G_1 to G_6 denote totals of blocks corresponding to different sets of treatments and $Q_0, \ldots Q_6$ the adjusted yield totals. For example,

$$G_1 = \sum_i^r B_i \text{ and } Q_1 = T_1 - \frac{G_1 + G_2 + G_3 + G_5}{4}.$$

Solving the normal equations we obtain,

$$\hat{\theta}_0 = \frac{Q_1}{6r}$$

$$\hat{\phi}_1 = \frac{4}{83r}[6(Q_0 - \bar{Q}_1) + Q_2 - \bar{Q}_2]$$

$$\hat{\phi}_2 = \frac{4}{83r}[6(Q_3 - \bar{Q}_1) + Q_4 - \bar{Q}_2]$$

$$\hat{\phi}_3 = \frac{4}{83r}[6(Q_5 - \bar{Q}_1) + Q_6 - \bar{Q}_2]$$

$$\hat{\theta}_1 = \frac{4}{83r}[14(Q_2 - \bar{Q}_2) + Q_1 - \bar{Q}_1]$$

$$\hat{\theta}_2 = \frac{4}{83r}[14(Q_4 - \bar{Q}_2) + Q_3 - \bar{Q}_1]$$

$$\hat{\theta}_3 = \frac{4}{83r}[14(Q_6 - \bar{Q}_2) + Q_5 - \bar{Q}_1]$$

$$\hat{\phi} = \frac{\bar{Q}_1}{47} \qquad \hat{\theta} = \frac{\bar{Q}_2}{27}$$

where we have set

$$\bar{Q}_1 = \frac{Q_1 + Q_3 + Q_5}{3} \text{ and } \bar{Q}_2 = \frac{Q_2 + Q_4 + Q_6}{3};$$

We can now at once write down the expressions for treatment means along with their variances. For example,

$$\hat{\mu} + \hat{t}_1 = \frac{G}{24r} + \frac{4}{83r}\left[6\left(Q_1 - \bar{Q}_1\right) + \left(Q_2 - \bar{Q}_2\right)\right] + \frac{\bar{Q}_1}{4r}$$

with a variance of $275\sigma^2/996$.

The treatment S.S. which is known to be equal to $\sum_{i=0}^{6} t_i Q_i$ may be split into different component parts by setting the hypothesis

(i) $\phi_1 = -\theta_1$, $\phi_2 = -\theta_2$, $\phi_3 = -\theta_3$ for forms;

(ii) $\phi = \theta$ for levels;

(iii) $\phi_1 = \theta_1$, $\phi_2 = \theta_2$, $\phi_3 = \theta_3$ for levels $\times$ forms; and

(iv) $2\theta_0 = \phi + \theta$ for control *versus* treatments.

The following notations have been used to further simplify the analysis of variance given below:

$Q_1 + Q_2 = Q_{12}$	$7\,Q_1 + 15\,Q_2 = P_{12}$	$5\,Q_1 - 13\,Q_2 = P_{1-2}$
$Q_3 + Q_4 = Q_{34}$	$7\,Q_3 + 15\,Q_4 = P_{34}$	$5\,Q_3 - 13\,Q_4 = P_{3-4}$
$Q_5 + Q_6 = Q_{56}$	$7\,Q_5 + 15\,Q_6 = P_{56}$	$5\,Q_5 - 13\,Q_6 = P_{5-6}$

As for the earlier model, discussed in Section 4, estimates of responses and differences in responses can also be obtained by combining estimates from different sets of experiments as follows:

Response to a level of a given fertilizer is obtained by taking difference in yield of that treatment from the average of the corresponding controls. Difference in response to two forms say n and n' at single level is obtained by taking the weighted average of 3 estimates. The first and third sets give a direct estimate of the difference in response with a variance σ^2/r. The second and fourth sets give an indirect estimate as

$$\left[n_1 - \frac{o + n_1''}{2}\right] - \left[n_1' - \frac{o + n_1''}{2}\right]$$

with variance $3\sigma^2/r$, the quantity $(o + n_1'')/2$ in the first bracket being calculated from the set (ii) and in the second bracket from the set (iv). Another indirect estimate is given by

$$\left[n_1 - \frac{o + n_1'' + n_2''}{3}\right] - \left[n_1' - \frac{o + n_1'' + n_2''}{3}\right]$$

TABLE VI

Analysis of Variance

Source	d.f.	S.S.	Expected value of mean squares
Fields (ignoring treatments)	$6r-1$	$\sum_{i}^{6r} \frac{B_i^2}{r} - \frac{G^2}{24r}$	
Treatments (eliminating blocks)	6	$\frac{Q_0^2}{6r} + \frac{3\bar{Q}_1^2}{4r} + \frac{3\bar{Q}_2^2}{2r} + \frac{4}{83r}[5dev^2\,(Q_1, Q_3, Q_5) + 13dev^2\,(Q_2, Q_4, Q_6) + dev^2\,Q_{12}, Q_{34}, Q_{56})]$.	
Forms ..	2	$\frac{2}{11\times 83r}[dev^2\,(P_{12}, P_{34}, P_{56})]$	$\sigma^2 + \frac{83r}{176}\sum_1^3 (\phi_i - \theta_i)^2$
Levels ..	1	$\frac{1}{4r}(\bar{Q}_1 - 2\bar{Q}_2)^2$, ...	$\sigma^2 + 4r(\phi - \theta)^2$
Forms × levels ..	2	$\frac{2}{83\times 9}\,dev^2\,(P_{1-2}, P_{3-4}, P_{5-6})$	$\sigma^2 + \frac{83r}{144}\sum_1^3 (\phi_i - \theta_i)^2$
Control *versus* rest	1	$\frac{Q_0^2}{6r} + \frac{3\bar{Q}_1}{4r} + \frac{3\bar{Q}_2^2}{2r} - \frac{6}{11r}(\bar{Q}_1 - \bar{Q}_2)^2$	$\sigma^2 + \frac{12r}{11}(2\theta_0 - \theta - \phi)^2$
Error	$18r-6$	By difference	σ^2
TOTAL ..	$24r-1$		

from the sets (v) and (vi) with a variance of

$$\frac{8\sigma^2}{3r}.$$

These three estimates being independent can be combined by taking average, weighted inversely to the corresponding variances. Similarly the difference between the two forms at the double levels say $n_2 - n_2'$ can be estimated by combining the estimates from the sets (i) and (iii); having variance $(8/3)\,\sigma^2/r$ and the estimate from the sets (ii) and (iv) having variance $3\sigma^2/r$. The difference between the two levels of any one form, *e.g.*, $n_1 - n_2$ may be estimated as follows: We add the quantity $(o + n_1')/2$ from the set (i) to $[n_1 - (o + n_1')/2]$ from the set (iii); similarly to

$$\left(\frac{o + n_1''}{2}\right)$$

from set (ii) we add $[n_1 - (o + n_1'')/2]$ from set (v). The mean of these two quantities will have a variance of σ^2/r. The mean of n_1 plots, from sets (i) and (ii) has a variance of $\sigma^2/2r$. These two independent estimates are combined by taking a mean weighted inversely to their variances. The difference of this quantity from the mean of the n_2 plots from sets (i) and (ii) can be easily seen to be free from block effects and has a variance of $5\sigma^2/6r$. The variances of these combined estimates are compared with the variances of estimates obtained from the least square analysis in Table VII. Difference between forms under column 2 in this table is taken as the simple average of the difference in response at each level for the two sources and the correlation between estimates of $n_1 - n_1'$ and $n_2 - n_2'$ has been allowed for in deriving the variance expression.

It will be seen from Table VII that the relative efficiency is higher under the combined model under (iii).

6. Comparison of Levels of Nitrogen and Phosphate

To study the effects of levels of nitrogen and phosphorus a convenient design in 6-plot blocks is the two sets of treatments given under (C) in Section 2. Let a random sample of $2r$ fields be taken in a region and r experiments of each type be allocated at random to $2r$ fields. Let the treatment effects be denoted as follows:

$n_1 \rightarrow \hat{t}_1$	$n_1p_1 \rightarrow \hat{t}_4$	$n_1p_2 \rightarrow \hat{t}_7$
$n_2 \rightarrow \hat{t}_2$	$n_2p_1 \rightarrow \hat{t}_5$	$n_2p_2 \rightarrow \hat{t}_8$
$p_3 \rightarrow \hat{t}_3$	$p_2 \rightarrow \hat{t}_6$	

TABLE VII

Variance of Estimates

	Combination of estimates from different sets	Estimates under the combined model	Efficiency of estimate under the combined model (%)
(i) Difference between two forms at the single level	$\frac{24}{41} \cdot \frac{\sigma^2}{r}$	$\frac{48}{83} \cdot \frac{\sigma^2}{r}$	101·2
(ii) Difference between two forms at the double level	$\frac{24}{17} \cdot \frac{\sigma^2}{r}$	$\frac{112}{83} \cdot \frac{\sigma^2}{r}$	104·6
(iii) Difference between levels of any one form	$\frac{5}{6} \frac{\sigma^2}{r}$	$\frac{275}{4\times 83} \frac{\sigma^2}{r}$	112·7
(iv) Difference between two forms	$\frac{356}{697} \frac{\sigma^2}{r}$	$\frac{44}{83} \frac{\sigma^2}{r}$	103·0

Employing the usual additive model the analysis of the design under the restriction $2(t_0 + t_1 + t_2) + \sum_{i=3}^{8} t_i = 0$ will lead to the following estimates:

$$\hat{t}_0 = \frac{Q_0}{2r}, \quad \hat{t}_1 = \frac{Q_1}{2r}, \quad \hat{t}_2 = \frac{Q_2}{2r}, \quad \hat{t}_3 = \frac{Q_3}{r} + \frac{Q_{345} - Q_{678}}{6r}$$

$$\hat{t}_4 = \frac{Q_4}{r} + \frac{Q_{345} - Q_{678}}{6r}, \quad \hat{t}_5 = \frac{Q_5}{r} + \frac{Q_{345} - Q_{678}}{6r}.$$

$$\hat{t}_6 = \frac{Q_6}{6r} - \frac{Q_{345} - Q_{678}}{6r}, \quad \hat{t}_7 = \frac{Q_7}{6r} - \frac{Q_{345} - Q_{678}}{6r},$$

$$\hat{t}_8 = \frac{Q_8}{6r} - \frac{Q_{345} - Q_{678}}{6r}$$

where the Q functions have the usual meaning, *viz.*, Q_i = total yield for the ith treatment $-\frac{1}{6}$ (sum of the block totals in which this treatment occurs) and $Q_{345} = Q_3 + Q_4 + Q_5$, etc.

The estimates of the average effects to the different levels of nitrogen are given by

$$N_0 = \frac{1}{3r}\left(\frac{Q_0}{2} + Q_3 + Q_6\right) + \frac{G}{12r},$$

$$N_1 = \frac{1}{3r}\left(\frac{Q_1}{2} + Q_4 + Q_7\right) + \frac{G}{12r},$$

$$N_2 = \frac{1}{3r}\left(\frac{Q_2}{2} + Q_5 + Q_8\right) + \frac{G}{12r}.$$

Similarly the average effects for the levels of phosphorus are given by

$$P_0 = \frac{1}{6r} Q_{012} + \frac{G}{12r}, \quad P_1 = \frac{1}{6r}(3Q_{345} - Q_{678}) + \frac{G}{12r}.$$

$$P_2 = \frac{1}{6r}(3Q_{678} - Q_{345}) + \frac{G}{12r}.$$

To obtain the standard errors of estimates we note that

$$V(Q_0) = V(Q_1) = V(Q_2) = \frac{5r}{3}\sigma^2;$$

$$V(Q_3) = V(Q_4) = \dots = V(Q_8) = \frac{5r\sigma^2}{6};$$

$$\text{Cov}(Q_0, Q_1) = \text{Cov}(Q_0, Q_2) = \text{Cov}(Q_1, Q_2) = -\frac{2r\sigma^2}{3};$$

$$\text{Cov}(Q_i, Q_j) = -\frac{r\sigma^2}{6}$$

where

(1) $i = 0, 1, 2$	$j = 3$ to 8	when $i = 3, 4, 5$
(2) $i = 3, 4, 5$ $i \neq j$	$j = 3, 4, 5$ (4) $= 0$	$j = 6, 7, 8$
(3) $i = 6, 7, 8$ $i \neq j$	$j = 6, 7, 8.$	

The variances of the average effects for levels of nitrogen $= 5\sigma^2/18r$ for $P_0 = \sigma^2/6r$ and for P_1 and $P_2 = \sigma^2/2r$. The responses to levels of nitrogen and prosphorus are obtained as $N_1 - N_0$ and $P - P_0$ with variances $5\sigma^2/9r$ and $2\sigma^2/3r$ respectively.

The analysis of variance for testing different sets of degrees of freedom is given on next page.

TABLE VIII

Analysis of Variance

Source	d.f.	S.S.
(1) Between fields	$2r - 1$	
(2) Treatments	8	$\frac{1}{6r}\left[3\sum_{0}^{2} Q_i^2 + 6\sum_{3}^{8} Q_i^2 + (Q_{345} - Q_{678})^2\right]$
(3) Levels of N	2	$\frac{1}{40r}\, dev^2 \left(\frac{Q_0}{2} + Q_3 + Q_6,\ \frac{Q_1}{2} + Q_4 + Q_7,\ \frac{Q_2}{2} + Q_5 + Q_8\right)$
(4) Levels of P	2	$\frac{2}{3r}(Q^2_{345} - Q^2_{678})$
(5) Interaction NP	4	$(2) - (4) - \frac{1}{4r}\, dev^2 (Q_{036},\ Q_{147},\ Q_{258})$
(6) Error ..	$10r - 8$	By difference
TOTAL ..	$12r - 1$	

If the NP interaction is absent, appropriate estimates of the treatment effects and the corresponding analysis of variance will have to be modified in view of the fact that treatments are not orthogonal to the fields. In such a situation the estimates of the average effects of the levels of phosphate remain the same while for levels of nitrogen we have

$$N_0 = \frac{1}{4r} Q_{036} + \frac{G}{12r}, \quad N_1 = \frac{1}{4r} Q_{147} + \frac{G}{12r},$$

$$N_2 = \frac{1}{4r} Q_{258} + \frac{G}{12r} \quad \text{with}$$

$$V(N_0) = V(N_1) = V(N_2) = \frac{\sigma^2}{4r}.$$

The analysis of variance in this case will be obtained as follows:

TABLE IX

Source	d.f.	S.S.
(1) Between fields (ignoring treatments)	$2r - 1$	
(2) Treatments ..	4	$\frac{1}{4r} dev^2 (Q_{036}, Q_{147}, Q_{258}) + \frac{2}{3r}(Q^2_{345} + Q^2_{678})$
(3) Levels of N ..	2	$\frac{1}{4r} dev^2 (Q_{036}, Q_{147}, Q_{258})$
(4) Levels of P ..	2	$\frac{2}{3r} (Q^2_{345} + Q^2_{678})$
(5) Error ..	$10r - 4$	(6) − (1) − (2)
TOTAL ..	$12r - 1$	$\Sigma y^2_{ijk} - C.F.$

In some cases it may not be possible to put 6 plots at a given site in cultivators' fields. Arrangements involving lesser number of plots in a field will have to be worked out. One such 'unequal block' arrangement is given in sets marked (D) in Section 2. Let a random sample of $3r$ fields be taken in a region such that the experiments of each type are allocated at random to the $3r$ fields. Employing the usual additive model and under the linear restrictions to

$$3 + \sum_{1}^{\infty} t_i = 0 \text{ and } 3 \sum_{1}^{r} b_i + 4 \sum_{r+1}^{2r} b_j + \sum_{2r+1}^{3r} b_k = 0$$

we obtain the following estimates of the various treatment effects:

$$\hat{t}_0 = \frac{4\,Q_0 + Q_1 + Q_2}{11r}, \quad \hat{t}_1 = \hat{t}_0 + \frac{2\,Q_1 + Q_2}{r};$$

$$\hat{t}_2 = \hat{t}_0 + \frac{Q_1 + 2\,Q_0}{r}, \quad t_3 = \hat{t}_0 + \frac{2\,Q_3 + Q_4 + Q_5}{r};$$

$$\hat{t}_4 = \hat{t}_0 + \frac{Q_3 + 2\,Q_4 + Q_5}{r}, \quad \hat{t}_5 = \hat{t}_0 + \frac{Q_3 + Q_4 + 2\,Q_5}{r};$$

$$\hat{t}_6 = \hat{t}_0 + \frac{2\,Q_6 + Q_7 + Q_8}{r}. \quad \hat{t}_7 = \hat{t}_0 + \frac{Q_6 + 2\,Q_7 + Q_8}{r};$$

$$\hat{t}_8 = \hat{t}_0 + \frac{Q_6 + Q_7 + 2\,Q_8}{r}.$$

where the Q functions have their usual definition. Variances for several contrasts may now be worked out, *e.g.*, variances of the average responses to single or double level of nitrogen and average response to single or double level of phosphate are equal to $8\sigma^2/27r$ and $14\sigma^2/9r$ respectively. The analysis of variance for this set up can be put down in terms of the Q functions but as in the case of other unequal block arrangements, expressions involved are computationally unattractive and are not being given here.

Another approach which might suggest itself is to omit the control plots in sets (ii) and (iii) and pool the analysis from different sets.

It will then be seen that the resulting set up is similar to a split plot design wherein levels of phosphate form the main plot treatments and levels of nitrogen the sub-plot treatments. The analysis of variance may be carried out by splitting the total variation into (*a*) Between fields within sets, (*b*) Between levels of nitrogen, (*c*) Inter-block response to phosphate, (*d*) Levels of phosphate $\times$ levels of nitrogen, (*e*) Between fields within sets $\times$ treatments. The appropriate error for comparing levels of nitrogen and for the interaction of phosphate with nitrogen will be given by (*e*) because they are within field comparisons. However, the response to phosphate being estimated as inter-field comparison will also involve differences between fields and will, therefore, be tested against (*a*).

Although control plots in the second and third sets have been omitted in the above analysis, they can be used to obtain the intra-block estimate of the response to phosphate alone. The combined intra- and inter-block estimate for the response to phosphate alone may be obtained by setting $y = k\bar{y}_1 + (1 - k)\,\bar{y}_2$ where $\bar{y}_1$ and $\bar{y}_2$ are the intra- and inter-block estimates and where k is chosen such that the proposed estimate has a minimum variance. It turns out that

$$k = \frac{2\sigma_f^2 + \sigma^2}{2\,(\sigma^2 + \sigma_f^2)}$$

and

$$\text{Var}(y) = \frac{2\sigma^2}{r}\left[1 - \frac{1}{4\left(\frac{\sigma_f^2}{\sigma^2} + 1\right)}\right]$$

where σ_f^2 is the true component of variation between fields within the tract. It will be observed from the expression for Var (y) that if σ_f^2 is very large in relation to σ^2 there is no further improvement in the precision of the estimate based on the intra-block comparison. From the data of fertilizer trials conducted in India the ratio of σ_f^2/σ^2 has been found to vary from 1 to 4. With this ratio it will be seen that the gain with recovery of information is small being only of the order of 5 to 10%.

It will be of interest to compare the efficiency of estimates by the split plot and the least square methods. We will have to compare the efficiency separately for nitrogen and phosphate effects. The variance for the estimates of response to levels of nitrogen and phosphate under the least square have been given earlier. The corresponding variances of estimates under the split plot model are $2\sigma^2/3r$ and $2(\sigma_f^2/r + \sigma^2/3r)$. Therefore the relative efficiency of the split plot model for nitrogen is 44%. The efficiency of the response to phosphate will depend upon σ_f^2/σ^2. As mentioned earlier this ratio appears to vary from 1 to 4 and, therefore, the efficiency of phosphate response is low and will vary from 18·0% to 58·0%.

7. Optimum Allocation of Experiments and Total Amount of Experimentation

In the type of experimental programmes discussed in the foregoing sections, it is necessary to determine the total amount of experimentation and its distribution between different units. To be able to do this, we should have information both on the cost of experimentation and the variance components which constitute the error of estimates.

Information on variance components is available from several of the schemes on simple fertiliser trials in cultivators' fields recently carried out in India. These experiments have been generally carried out in randomly chosen fields in randomly selected villages in a given area which has been either a thana (an administrative unit consisting of about 100 square miles) or a community project area or a taluk having an area of about 300 to 400 square miles. The area of a village is about 1 to 2 square miles and consists of about 1,000 to 1,500 fields, In this section, the number of experiments to be carried out in a village and the number of villages for estimating the taluk response with a given margin of error has been determined. Optimum allocation of experiments between fields and villages has also been investigated taking into consideration a simple type of cost function. These results have been extended to district estimates as well. The optimum amount of experimentation taking into account both the cost and the expected losses due to error in the estimate has also been considered.

7·1 *Distribution of experiments between villages and within villages*

Estimates of σ_{vt}^2 (true interaction variance of treatments $\times$ villages) and σ_{ft}^2 (true interaction variance of fields $\times$ treatments) are available from the results of Stewart's Scheme and T.C.M. experiments on paddy and wheat. In the Stewart's Scheme villages were selected at random from taluks in Madras State and in Bihar State from subdivisions which are somewhat larger than taluks, whereas in T.C.M. experiments villages were selected at random from a block of a Community Project area which is smaller in area than a taluk. As such $\hat{\sigma}_{vt}^2$ does not

TABLE X

Estimates of Interaction Variances

Source of data	$\hat{\sigma}_{ft}^2$	$\hat{\sigma}_{vt}^2$	Mean $\bar{X}$ mds. per acre	Coefficient of variability	
				$\frac{\sigma_{ft}}{\bar{x}} \times 100$	$\frac{\hat{\sigma}_{vt}}{\bar{x}} \times 100$
1	2	3	4	5	6
Stewart's Scheme (Paddy)					
Madras (1953–54) ..	10·6	3·1	28·2	11·5	6·2
do. (1954–55) ..	8·2	4·3	31·0	9·2	6·7
Bihar (1952–53) ..	10·1	3·7	23·9	13·3	8·0
do. (1953–54) ..	19·8	3·6	30·5	14·6	6·2
do. (1954–55) ..	8·6	3·9	23·3	12·6	8·5
T.C.M. (1954–55) (Paddy)	11·2	13·4	21·9	15·3	16·7
Average for paddy experiments	11·4	5·3	26·5	12·6	8·7
Stewart's Scheme (Wheat)					
Bihar (1952–53) ..	1·1	1·9	9·6	10·9	14·4
do. (1953–54) ..	1·8	2·8	11·5	11·7	14·5
T.C.M. (1953–54) (Wheat)	5·0	2·7	16·2	13·8	10·1
do. (1954–55) do.	6·9	4·5	15·1	17·4	14·0
Average for wheat experiments	3·7	3·0	13·1	13·4	13·2

represent the same interaction variance in the different schemes. However, it is presumed that the differential coverage of areas is not likely to make any appreciable difference in the value of σ_{vt}^2 and, therefore, these values have been averaged and the same value has been taken both in a thana and a taluk.

The estimates of the true interaction variances σ_{vt}^2 and σ_{ft}^2 were obtained from those villages of a taluk where more than one experiment was conducted. Such values for each year were averaged over ten taluks in Madras State and over five subdivisions in Bihar State. In T.C.M. experiments, averages were taken over six Community Project areas both for paddy (1954–55) and wheat (1954–55) and over four Community Project areas for wheat (1953–54).

If in a taluk m experiments per village are carried out in n villages, the variance of a treatment response is given by

$$\sigma_0^2 = 2\left(\frac{m\sigma_{vt}^2 + \sigma_{ft}^2}{mn}\right).$$

The number of villages and fields for different levels of accuracy as obtained from this formula are given in Table XI.

TABLE XI

Minimum Number of Villages (n) *for Different Levels of Accuracy and Given Number of Fields* (m)

Crop	Paddy			Wheat		
Number of fields (m) / S.E. % (p)	1	2	3	1	2	3
3	52	34	29	79	59	52
5	19	12	10	28	21	19
10	5	3	3	7	5	5

From the above table it will be seen that for both wheat and paddy quite a large number of villages will have to be taken for attaining an accuracy of 3% of the mean yield. As such it may be reasonable to aim at an accuracy of about 5% for the taluk estimate. It will also be seen that in the case of wheat about 50% more experiments will have to be carried out for the same accuracy as that of paddy.

We shall now determine optimum allocation of experiments between villages and fields by minimising the cost of experimentation for a fixed precision of a treatment response.

We shall assume a simple cost function of the type $C=c_0+c_1n+c_2mn$ where c_0 is the overhead cost and c_1 is the cost of including one more village and c_2 is the cost of including one more experiment. The salary of the field assistant for the days of his visit to the village, the cost of his travel and the cost of preparing a list of cultivators growing a particular crop in a village will constitute the village component of the cost (c_1). The field assistant is usually paid a fixed salary of Rs. 150 per mensem inclusive of allowances which works out as Re. 1 per man-hour of his working time. During travel outside his headquarters he is paid an allowance of 2 annas per mile for his journey time. A field assistant will be required to visit a given village four times for (i) selection of cultivators within a village, (ii) laying out an experiment, (iii) making observations on crop growth and (iv) harvesting. Moreover, corresponding to two out of his four visits, *viz.*, at the time of laying out a field experiment and at the time of harvesting, the field assistant will be required to revisit some of the villages to complete these observations on account of the unforeseen absence of the selected cultivators. We shall suppose that such revisits will be confined to 25% of the villages.

As a first approximation to travel costs, we shall assume that the field assistant travels from his thana headquarters and returns back at the end of the day without undertaking any travel between villages. Considering the size of a thana, the distance between the thana headquarters and a village may be taken on the average to be equal to three miles and the field assistanct may take about one hour to cover this distance. Since there will be no travel between the villages, the total distance travelled for the visits mentioned earlier for a selected village is $24 + 12(\cdot 25) = 27$ miles. Furthermore he might be required to spend three man-hours per village for listing the names of cultivators. These two items will cost Rs. 12. Supervision of field assistants and their training will increase those parts of variable costs relating to villages and experiments which arise directly out of the man-hours spent by the field assistants. We might assume roughly 5% intensity of supervision at this stage with supervisory wages as $1\frac{1}{2}$ times the wages of the field assistant which will, therefore, have an effect of increasing the working time wages of the field assistant by 7½%.

Training of the field assistants is likely to be arranged on a regional basis and will involve both travel cost and working time. The average distance from the thana headquarters to the regional headquarters may be taken as 150 miles which will take about 8 hours to cover. The actual training might be assumed to last for 14 working hours. Since it will be a trip away from thana headquarters a special allowance equivalent to wages of 8 man-hours will be given to a field assistant. This will, therefore, have an effect of increasing the working time wages per village by 3% and travel time by 20%. Thus we have $c_1 =$ Rs. 18·25.

Component costs which will increase with the number of experiments in a village might be determined as follows:

(i) Cost of fertilizer per experiment	Rs. 10·00
(ii) Special cost of cultivation, bunding, etc., per experiment	,, 3·00
(iii) Labour cost for harvesting five or six centrally located portions each of size 1/50th of an acre	,, 10·00
(iv) Additional cost of computation, stationery, pegs, etc., per experiment	,, 2·00
(v) Personal time of the assistant:	
(a) For contacting one cultivator and locating one experimental field ..	3 man-hours
(b) Fertilizer application and making the plots during second visit	3 ,,
(c) Crop observations during third visit ..	2 ,,
(d) Supervision during harvesting ..	6 ,,

These items will cost Rs. 39. If we assume 20% intensity of supervision at this stage for the work of the field assistants, it will increase the working time wages by 13% which combined with an additional 3% increase on account of training gives $c_2 =$ Rs. 46·00.

In order to determine the optimum allocation of experiments between villages and within villages, we shall minimise the cost (c) for a fixed margin of error given by

$$\sigma_0^2 = 2\left(\frac{m\sigma_{vt}^2 + \sigma_{ft}^2}{mn}\right).$$

This will give

$$m = \sqrt{\frac{c_1}{c_2}} \cdot \frac{\hat{\sigma}_{ft}}{\hat{\sigma}_{vt}} \text{ and } n = \frac{2}{\sigma_0^2}\left(\hat{\sigma}_{vt}^2 + \frac{\hat{\sigma}_{ft}^2}{m}\right).$$

The value $\hat{\sigma}_{ft}/\hat{\sigma}_{vt}$ is a little greater than unity, and $\sqrt{c_1/c_2}$ is much smaller than unity. Therefore it will not pay to have more than one experiment per village. It might be contested that the assistant will be required to pay more than four visits to the villages in order to be able to exercise more effective supervision. This will have the effect of increasing c_1 but even then c_1/c_2 cannot be greater than 1. The number of villages for $m = 1$ for different levels of accuracy have already been given in Table XI.

7.2. *Determination of minimum number of thanas for estimating the response in a district*

In the case of a three-stage sample design when the thanas are the primary units of selection, the variance of a treatment response for any given year based on l thanas, n villages per thana and m field experiments per village is given by

$$V_o = 2\left(\frac{\sigma_{ct}^2}{l} + \frac{\sigma_{vt}^2}{ln} + \frac{\sigma_{ft}^2}{lmn}\right).$$

where $\hat{\sigma}_{ct}^2$ is the true interaction variance of thanas $\times$ treatments.

Information on σ_{ct}^2 is available from the analysis of Bihar manurial trials which has been carried out at the Statistical Wing of the Indian Council of Agricultural Research (unpublished data). These experiments were carried out in a number of fields selected from a thana. Values of $\hat{\sigma}_{ct}^2$ and $\hat{\sigma}_{vt}^2 + \sigma_{ft}^2$ obtained as the average of seven years are given below:

	$\hat{\sigma}_{ct}^2$	$\hat{\sigma}_{vt}^2 + \hat{\sigma}_{ft}^2$
Paddy ..	3·8	10·6
Wheat ..	2·0	5·9

We could now find out the required number of thanas for determining the district response with a given precision. Estimates of σ_{vt}^2 and σ_{ft}^2 as given in Table X are based on data from other series of experiments and as such the estimate of σ_{ct}^2 obtained here is not strictly comparable. From the values given above it will be seen that $\hat{\sigma}_{ct}^2$ is roughly $\frac{1}{3}(\hat{\sigma}_{vt}^2 + \hat{\sigma}_{ft}^2)$. This relation has been utilised to estimate σ_{ot}^2 on a comparable basis from Table X giving $\hat{\sigma}_{ot}^2 = 5 \cdot 6$ and $2 \cdot 2$ for paddy and wheat respectively. Utilising the value of the variance components, the minimum number of thanas required to be taken for estimating the district response for different levels of accuracy are given in Table XII.

TABLE XII

Minimum Number of Thanas Required for Estimating the District Response (with one experiment per village)

	Paddy S.E. % of the thana response			Wheat S.E. % of the thana response		
	3	5	10	3	5	10
No. of villages / S.E. % of district response	52	19	5	79	28	7
3	9	10	14	15	16	20
5	3	4	5	5	6	7
8	1	1	2	2	2	3

It will be observed that on the average we will require more number of thanas for estimating the district response for wheat than for paddy.

We shall now determine the number of thanas by minimising cost of experimentation for a fixed precision of the treatment response. The cost function may be written as $C = C_0 + lc_1' + lc_1n + lc_2mn$ where C_0 is the overhead cost, c_1' is the cost per thana and c_1 and c_2 have been defined earlier in Section 7.1.

Increasing the number of thanas, besides proportionately increasing the cost per individual thana will bring the extra cost component corresponding to the office expenses at the thana headquarters for a period of one year as in next page.

	Rs.
Office rent at Rs. 10 per month	120
Rent for office furniture at the rate of Rs. 4·5 per month ..	54
Office stationery	12
Balance and tapes	14
Total ..	200

Therefore, $c_1' =$ Rs. 200.

The optimum value of m, n and l are given by

$$m = \sqrt{\frac{c_1}{c_2}} \cdot \frac{\sigma_{ft}}{\sigma_{vt}}\,; \quad n = \sqrt{\frac{c_1'}{c_1}} \cdot \frac{\sigma_{vt}}{\sigma_{ct}}\,.$$

$$l = \frac{2}{V_0^2}\left(\sigma_{ct}^2 + \frac{\sigma_{vt}^2}{n} + \frac{\sigma_{ft}^2}{mn}\right).$$

The above formula gives $m = 1$ and $n = 3$ and if the precision of the district response is 6% then $l = 9$ for paddy.

Experiments will ordinarily be repeated over a number of years (y) to sample different seasonal conditions for the estimation of responses. If fresh thanas are selected every year, the variance of a treatment response over years is given by:

$$V = \frac{2}{y}\left(\sigma_{yt}^2 + \frac{\sigma_{ct}^2}{l} + \frac{\sigma_{vt}^2}{ln} + \frac{\sigma_{ft}^2}{lmn}\right).$$

To be able to get a good estimate of σ_{yt}^2 it will be ideal to repeat an experiment on the same site every year. Since experiments on cultivators' fields are ordinarily repeated in different fields in different villages every year, it becomes exceedingly difficult to get a good estimate of σ_{yt}^2 free from other variations due to change in site of the experiments.

Among the series of trials considered here, only the Bihar manurial experiments have been carried out for a sufficient number of years. Estimate of σ_{yt}^2 has, therefore, been calculated from this series only. The levels of N and P tried in the series were 25 and 50 lb. per acre in some of the years while in the remaining years the levels varied from 30 to 60 lb. per acre. In view of the varying doses tried in different years, the responses have been standardised to 30 N and 30 P in all years and σ_{yt}^2 has been calculated from the standardised responses. The values of mean squares between years (s_y^2) and the simple average of error variance of standardised responses in different years (s^2) are given below:

TABLE XIII

Mean Square between Years and Error Variance of Standardised Responses

Bihar Manurial Experiments (Average of 17 districts over 7 years) (md./acre)²

	Paddy		Wheat	
	Nitrogen	Phosphate	Nitrogen	Phosphate
Mean square between years (s_y^2) ..	1·93	2·20	1·16	1·19
Error (s^2) ..	0·74	0·86	0·54	0·49

The estimates of the true years × treatment component is obtained as the difference $s_y^2 - s^2$. As the mean squares for nitrogen and phosphate are similar in magnitude, the estimate of σ_{yt}^2 has been averaged over N and P giving a value $\hat{\sigma}_{yt}^2 = 1\cdot3$ and $0\cdot7$ for paddy and wheat respectively. These values are nearly $\frac{1}{3}$ of the value of σ_{ct}^2. Estimate of σ_{yt}^2 is made comparable to the estimates of other components by using this ratio in the manner as was adopted for getting the estimate of σ_{ct}^2. Utilising this information we find that with one field per village, 7 villages per thana and 6 thanas per district, the response in a district for paddy can be obtained with a standard error of 4% over a period of six years.

7.3. *Optimum amount of experimentation*

In the previous section, we have investigated into the problem of allocation of experiments from the point of view of determining the response to a given treatment with a certain precision both with and without the consideration of cost. The concept of minimizing cost for a fixed precision or maximizing precision for a fixed cost is usually employed for the study of such problems. It will be, however, much more realistic to determine the optimum amount of experimentation by minimizing the cost of experimentation together with the expected loss due to error in the estimate.[4]

One very important object of experimental programmes on fertilizer use will be to estimate the optimum dressing. The dose will be optimum when the cost of a further small increment in the dose exactly equals the value of the resultant average increment in response. We shall assume that the expected loss due to error in the optimum will be proportional to the variance of the estimate of the optimum dressing.

Moreover, the loss, as a result of application of the optimum dose, will also vary with the area which will receive the fertilizer. If we try two non-zero levels of a fertilizer and use the quadratic response curve $y = a + \gamma x + \delta x^2$ to estimate the optimum dose from mnl experiments distributed among l thanas, n villages per thana and m experiments per village, the variance of the optimum dose is given by:

$$V = \frac{1}{\delta^2}\left[\frac{1}{4\Sigma\xi_1^2} + \frac{(x_0 - \bar{x})^2}{\Sigma\xi_2^2}\right]\left[\frac{\sigma_{ct}^2}{l} + \frac{\sigma_{vt}^2}{nl} + \frac{\sigma_{ft}^2}{mnl}\right],$$

where ξ's are orthogonal polynomials, x_0 is the optimum dose and $\bar{x}$ is the mean of the levels tried.

Let $C = c_0 + lc_1' + lc_1 n + lc_2 mn + L$ where c's have been defined in the previous section and the loss function $L = \lambda AV$ where V is the variance of the optimum dose, A is the total area where estimated optimum dose will be applied and λ is a constant of proportionality to be determined from the nature of the response curve. If we denote the cost of fertilizer by q and the price of produce per unit by p, the loss incurred through using Δx quantity of fertilizer beyond the optimum dose as determined from the quadratic response curve is given by

$$L = A\,[q\Delta x - p\,\{a + \gamma(x_0 + \Delta x) + \delta(x_0 + \Delta x)^2 - a - \gamma x_0 - \delta x_0^2\}] = -p\delta(\Delta x)^2 A$$

where we have used the relation $\gamma + 2\delta x_0 = q/p$.

This gives $\lambda = -p\delta$.

We shall now differentiate the expression given above for cost + loss and obtain

$$m = \sqrt{\frac{c_1}{c_2}\frac{\sigma_{ft}}{\sigma_{vt}}},$$

$$n = \sqrt{\frac{c_1'}{c_1}} \cdot \frac{\sigma_{vt}}{\sigma_{ct}},$$

$$l = \sqrt{\frac{\theta\left(\sigma_{ct}^2 + \frac{\sigma_{vt}^2}{n} + \frac{\sigma_{ft}^2}{mn}\right)}{c_1' + c_1 n + c_2 mn}},$$

where

$$\theta = \frac{\lambda A}{\delta^2}\left[\frac{1}{4\Sigma\,\xi_1^{\,2}} + \frac{(x_0 - \bar{x})^2}{\Sigma\,\xi_2^{\,2}}\right].$$

It is interesting to note that under the additive cost function which we have assumed, the optimum allocation of experiments between villages and within villages turns out to be the same as obtained in the previous Section from considerations of minimizing cost for a fixed precision. It is only the total amount of experimentation as determined by the number of thanas which depends both on the cost and loss function.

Substituting values as obtained in the previous Section for c's and values of $\hat{\sigma}_{ft}$, $\hat{\sigma}_{vt}$ and $\hat{\sigma}_{ct}$ for paddy we have $m = 1$ and $n = 3$. To be able to obtain the value of l we notice that the variance of the optimum depends on the theoretical value of the optimum dose itself. If we assume the optimum dose to be near the mean level of the doses tried, the contribution of the second term in the expression for the variance of the optimum dose may be neglected. We shall take the average value of $\delta = -\cdot 16$ lb. per acre which is based on the study of quadratic response curve relating to experiments in paddy in India[5] and $p =$ Rs. 10 per md. If we assume that the 20% of the area in a district having approximately $\cdot 15$ million acres under paddy will receive fertilizer, then we find that $l = 37$. It means that with this type of allocation we will have to conduct experiments in almost all the thanas of a district.

In the optimum allocation the estimate for thana being based only on three experiments will be of very low accuracy and often it will be of interest to obtain more reliable estimates for a thana. The optimum amount of experimentation in such a situation may be obtained in a similar manner. For example, in the case of paddy for estimating the thana response with 5% S.E., we require 19 villages at the rate of one experiment per village (*vide* Table XI). The optimum amount of experimentation $l = 15$ is obtained from the above formula by putting $m = 1$ and $n = 19$.

The authors wish to thank Dr. V. G. Panse for constant encouragement and the interest shown in the work reported in this paper. Dr. D. J. Finney was kind enough to go through the manuscript and the authors are grateful to him for useful suggestions.

SUMMARY

A systematic analysis of some of the typical experimental designs that could be used in cultivators' fields has been given with special

reference to the relative efficiencies of these designs. The problem of allocation of experiments and the amount of experimentation has also been discussed in the light of the data obtained in various fertilizer trials in cultivators' fields carried out recently in India.

REFERENCES

1. Stewart, A. B. .. *Soil Fertility Investigations in India with special reference to Manuring*, Army Press, Delhi, 1947.
2. Panse, V. G. and Sukhatme, P. V. "Experiments in cultivators' fields," *Jour. Ind. Soc. Agri. Stat.*, 1953, **5**, 144–60.
3. Cochran, W. G. .. "The combination of estimates from different experiments," *Biometrics*, 1954, **10**, 101–29.
4. Yates, F. .. "Principles governing the amount of experimentation in developmental work," *Nature*, July 1952, **171**, 138.
5. ——, Finney, D. J. and Panse, V. G. "The use of fertilizers on food grains," *ICAR Research Series, No.* 1.