Jour. Ind. Soc. Ag. Statistics
53(2), 2000 : 172-181

Estimation of Plot Size in Split-Split Plot Design

Md. Saiful Islam, Kalipada Sen,
Kamal Rahim[1] and Md. Yusuf Ali[2]
Dhaka University, Dhaka, Bangladesh
(Received : July, 1999)

SUMMARY

Estimates of soil heterogeneity are obtained from Split-Split plot design. A weighted regression analysis method is used to determine the soil heterogeneity from the analysis of variance technique of the above design. Soil heterogeneity is the major portion for obtaining optimum plot size. Ratio of the cost is the another portion for the same.

Key words : Soil heterogeneity, Weighted regression co-efficient, Optimum plot size, Split-Split plot design.

1. Introduction

The choice of suitable size of plot is an important factor to be decided in the planning of field experiments. The decision depends on experimental design consideration and practical feasibility. Experimental design considerations are based mainly on the accuracy of the estimates with a given amount of experimental area. Sills and Nienhuis [3] used the unweighted regression coefficient of the logarithm of variance on the logarithm of plot size estimated from Split-Split plot design. Hatheway and Williams [1] used the weighted regression coefficient of the logarithm of variance on the logarithm of plot size estimated from randomized block design. No effort seems to have been made to find out an optimum plot size using weighted regression coefficient considering yield data of the field experiment laid out as Split-Split plot design. The plot size depends on soil heterogeneity and cost of the field experiment.

The objective of the study is to determine optimum plot size from a weighted regression coefficient for measuring yield data considering Split-Split plot design.

1 CSO, BRRI, Dhaka, Bangladesh
2 SSO, BARI, Rajshahi, Bangladesh

2. *Layout and Method*

Let us consider a $r \times q \times p \times s$ Split-Split plot design where we have r replications, q main-plots (treatment A), p sub-plots (treatment B) and s sub-sub plots (treatment C). There are $r \times q \times p \times s$ plots in such a design. Since each replication contains all the treatments (*i.e.,* treatment A, treatment B, treatment C), the replications may be treated as the largest plot. In order to get larger sizes of whole plots q (treatment A), each larger sizes of whole plots should be split plots and each large sizes of split plots p (treatment B) should be Split-Split plots s (treatment C). Thus, a Split-Split plot design analysis can be reconstructed to estimate the variances among the different sizes of plots as under :

(i) variance among replications

(ii) variance among whole plots (main plot/treatment A)

(iii) variance among split-plots (sub-plot/treatment B)

(iv) variance among split-split plots (sub-sub-plot/treatment C)

Also estimates of variances corresponding to different sizes of plots can be obtained from usual analysis of variance in Split-Split plot design.

The analysis of variance for Split-Split plot design is given below

Source	*Degrees of Freedom*	*Mean square*
Replication	$(r-1)$	V_1
Main plot (whole plot/treatment A)	$(q-1)$	
Error (1)	$(r-1)(q-1)$	V_2
Sub plot (split plot/treatment B)	$(p-1)$	
Main plot × sub plot (A × B)	$(q-1)(p-1)$	
Error (2)	$q(r-1)(p-1)$	V_3
Sub-sub plot (split-split plot/treatment C)	$(s-1)$	
Main plot × sub-sub plot (A × C)	$(q-1)(s-1)$	
Sub plot × sub-sub plot (B × C)	$(s-1)(p-1)$	
Main plot × sub-plot × sub-sub plot (A × B × C)	$(q-1)(p-1)(s-1)$	
Error (3)	$qp(r-1)(s-1)$	V_4
Total	$rqps-1$	

Now we obtain the variances of plots of various sizes, reduced to a subplot basis, by V′. Thus, the replications are regarded as the largest plot, its variance V'_1 is equal to the replication mean square as it appears in the analysis of variance, *i.e.*, $V'_1 = V_1$.

Now considering whole plots as the next larger size of plot, the variance between whole plots contains, in addition to the variation to whole plots within replication that removed by the stratification of groups of whole plots into replications in the analysis of variance. Thus the total sum of squares for whole plots over the entire area is $r(q-1)V_2 + (r-1)V_1$ and there are $r \times q$ plots, the mean square is

$$V'_2 = \frac{[r(q-1)V_2 + (r-1)V_1]}{(rq-1)}$$

Similarly, the variance between split plots (sub-plots) over the entire area is

$$V'_3 = \frac{[rq(p-1)V_3 + r(q-1)V_2 + (r-1)V_1]}{(pqr-1)}$$

and the variance between split-split plots (sub-sub plots) over the entire area is

$$V'_4 = \frac{[pqr(s-1)V_4 + rq(p-1)V_3 + r(q-1)V_2 + (r-1)V_1]}{(spqr-1)}$$

The value of V_x are obtained by dividing each value of V′ by the number of units per replication, whole plot (main plot), split plot (sub-plot) and split-split plot (sub-sub plot) thus putting them on a unit basis. Thus the unweighted regression coefficient can be easily obtained by the least square method (Koch and Rigney [2]).

Now, the estimate of weighted regression coefficient is obtained below. V_1, V_2, V_3 and V_4 are independent and their estimated variances are $\frac{2V_1^2}{(r-1)}, \frac{2V_2^2}{(r-1)(q-1)}, \frac{2V_3^2}{q(r-1)(p-1)}$ and $\frac{2V_4^2}{pq(r-1)(s-1)}$ respectively.

Now we determine the variance and covariance of V'_1, V'_2, V'_3 and V'_4 which are linear functions of the former set.

Now the estimated variance of $V'_1 = \frac{2V_1^2}{(r-1)} = \frac{2(r-1)V_1^2}{(r-1)^2}$

Similarly, the estimated variance of

$$V'_2 = \frac{\left[2(r-1)\,V_1^2 + \dfrac{2r^2\,(q-1)\,V_2^2}{(r-1)}\right]}{(rq-1)^2}$$

$$V'_3 = \frac{\left[2(r-1)\,V_1^2 + \dfrac{2r^2\,(q-1)\,V_2^2}{(r-1)} + \dfrac{2qr^2\,(p-1)\,V_3^2}{(r-1)}\right]}{(rqp-1)^2}$$

$$V'_4 = \frac{\left[2(r-1)\,V_1^2 + \dfrac{2r^2\,(q-1)\,V_2^2}{(r-1)} + \dfrac{2qr^2\,(p-1)\,V_3^2}{(r-1)} + \dfrac{2pqr^2\,(s-1)V_4^2}{(r-1)}\right]}{(spqr-1)^2}$$

Now, the estimated covariance of V'_1 and V'_2 is

$$\text{Cov}\,(V'_1, V'_2) = \frac{2(r-1)V_1^2}{(rq-1)(r-1)}$$

Similarly,

$$\text{Cov}\,(V'_1, V'_3) = \frac{2(r-1)V_1^2}{(rqp-1)(r-1)}$$

$$\text{Cov}\,(V'_1, V'_4) = \frac{2(r-1)V_1^2}{(rpqs-1)(r-1)}$$

$$\text{Cov}\,(V'_2, V'_3) = \frac{\left[2(r-1)V_1^2 + \dfrac{2r^2\,(q-1)\,V_2^2}{(r-1)}\right]}{(rq-1)(pqr-1)}$$

$$\text{Cov}\,(V'_2, V'_4) = \frac{\left[2(r-1)V_1^2 + \dfrac{2r^2\,(q-1)\,V_2^2}{(r-1)}\right]}{(qr-1)(spqr-1)}$$

$$\text{Cov}\,(V'_3, V'_4) = \frac{\left[2(r-1)V_1^2 + \dfrac{2r^2\,(q-1)\,V_2^2}{(r-1)} + \dfrac{2qr^2\,(p-1)\,V_3^2}{(r-1)}\right]}{(pqr-1)(spqr-1)}$$

Thus, the variance-covariance matrix is

$$\begin{bmatrix} \frac{D}{(r-1)^2} & \frac{D}{(r-1)(qr-1)} & \frac{D}{(r-1)(pqr-1)} & \frac{D}{(r-1)(spqr-1)} \\ \frac{D}{(r-1)(qr-1)} & \frac{C+D}{(qr-1)^2} & \frac{C+D}{(qr-1)(pqr-1)} & \frac{C+D}{(qr-1)(spqr-1)} \\ \frac{D}{(r-1)(pqr-1)} & \frac{C+D}{(qr-1)(pqr-1)} & \frac{B+C+D}{(pqr-1)^2} & \frac{B+C+D}{(pqr-1)(spqr-1)} \\ \frac{D}{(r-1)(spqr-1)} & \frac{C+D}{(qr-1)(spqr-1)} & \frac{B+C+D}{(pqr-1)(spqr-1)} & \frac{A+B+C+D}{(spqr-1)^2} \end{bmatrix}$$

where $A = \frac{2pqr^2(s-1)V_4^2}{(r-1)}$, $B = \frac{2qr^2(p-1)V_3^2}{(r-1)}$, $C = \frac{2r^2(q-1)V_2^2}{(r-1)}$ and $D = 2(r-1)V_1^2$

Let the inverse of the above variance-covariance matrix be

$$\begin{bmatrix} Z_{11} & Z_{12} & Z_{13} & Z_{14} \\ Z_{21} & Z_{22} & Z_{23} & Z_{24} \\ Z_{31} & Z_{32} & Z_{33} & Z_{34} \\ Z_{41} & Z_{42} & Z_{43} & Z_{44} \end{bmatrix}$$

This can be easily obtained as the above variance-covariance matrix is known. The weights for y_i (= log V'_i) are obtained by multiplying each row and each column of this inverse matrix by the corresponding V'_i.

Thus, the weights matrix (w_{ik}) may be taken as

$$\begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \\ w_{41} & w_{42} & w_{43} & w_{44} \end{bmatrix}$$

where $w_{11} = V_1'^2 Z_{11}$, $w_{12} = V'_1 V'_2 Z_{12}$, etc.

It will be found that the sum of the elements of the weight matrix is equal to half the total number of degrees of freedom for the sums of squares from which variance estimates are derived. If the variances are unaffected by size of plot, then all the available sums of squares are estimates of the same basic variance. The different estimates of the logarithm of the variance derived from different lines of the analysis of variance, are independent and have asymptotic variance equal to twice the reciprocal of the corresponding degrees of freedom. Consequently the information from each is half the degrees of freedom, whence the total information is half the total degrees of freedom (Hatheway and Williams [1]).

Thus for the above split-split plot design experiments, the sum of the weights is

$$\sum_i \sum_k w_{ik} \approx \frac{(spqr - spq)}{2}$$

Let x_1 = area of largest plot (replication), x_2 = area of the larger plot (main plot / whole plot), x_3 = area of the large plot (sub plot/split plot), x_4 = area of the smallest plot (sub-sub plot/ split-split plot)

$$\log(V'_1) = y_1,\ \log(V'_2) = y_2,\ \log(V'_3) = y_3,\ \log(V'_4) = y_4$$

$$\log(x_1) = x'_1,\ \log(x_2) = x'_2,\ \log(x_3) = x'_3,\ \log(x_4) = x'_4$$

Now, the weighted regression coefficient b is (due to Hatheway and Williams [1])

$$b = -\frac{\sum_i \sum_k w_{ik}(y - \bar{y})(x'_i - \bar{x})}{\sum_i \sum_k w_{ik}(x'_i - \bar{x})^2}$$

where $\qquad \bar{x}$ = weighted mean of $x'_i = \dfrac{\sum_i \sum_k w_{ik} x'_i}{\sum_i \sum_k w_{ik}}$

and $\qquad \bar{y}$ = weighted mean of $y_i = \dfrac{\sum_i \sum_k w_{ik} y_i}{\sum_i \sum_k w_{ik}}$

Let $X_i = \sum_k w_{ik} x'_i$ and $Y_i = \sum_k w_{ik} y_i$

Then the sum of squares of x′ is

$$T = \sum_i X_i x'_i - \frac{(\sum_i X_i)^2}{\sum_i \sum_k w_{ik}}$$

The sum of products of y with x′ is

$$U = \sum_i X_i Y_i - \frac{(\sum_i X_i)(\sum_i Y_i)}{\sum_i \sum_k w_{ik}}$$

$$= \sum_i Y_i x'_i - \frac{(\sum_i X_i)(\sum_i Y_i)}{\sum_i \sum_k w_{ik}}$$

Thus the weighted regression coefficient $b = -\frac{U}{T}$ (1)

Following Smith [4], we have

Optimum plot size $X_{opt} = \frac{bk_1}{(1-b)k_2}$

where

b is the soil heterogeneity

k_1 is the over head cost per plot and

k_2 is the cost associated with unit size plot

3. Numerical Illustration from Experimental Data

An experiment was conducted On-farm research station Barind, Rajshahi in 1998. A split-split plot design in randomized block experiment with three replications was used. Main plots were the water regime, which included irrigation and non irrigation. Sub plots were the fertilizer doses which included 0 (kg/ha) and 30 (kg/ha) phosphorus. Sub-sub plots were chickpea genotypes, which included Annigeri, ICC4958 and Barichola2.

Experimental units were 36 m^2 plots (*i.e.*, 9 m long and 4m wide). Grain yield data of chickpea was collected from 6 m^2 plots (*i.e.*, 3m long and 2m wide). Yield data used to calculate unweighted regression coefficient and weighted regression coefficient of soil heterogeneity and the optimum plot size (Smith [4]) calculated as $X_{opt} = \frac{bK_1}{(1-b)K_2}$

Table 1 : The analysis of variance for yield data of chickpea in split-split plot design

Source	Degrees of Freedom	Mean square
Replication	2	9570.04 (V_1)
Main plot (whole plot/treatment A)	1	134.48
Error (1)	2	24290.73 (V_2)
Sub plot (split plot/treatment B)	1	1098.04
Main plot × sub plot (A × B)	1	3220.94
Error (2)	4	6885.04 (V_3)
Sub-sub plot (split-split plot/treatment C)	2	62726.91
Main plot × sub-sub plot (A × C)	2	8993.03
Sub plot × sub-sub plot (B × C)	2	8991.11
Main plot × sub-plot × sub-sub plot (A × B × C)	2	2469.33
Error (3)	16	9955.935 (V_4)
Total	35	

No. of replication = 3, No. of whole plot (main plot) = 2
No. of split plot (sub - plot) = 2 and No. of split-split plot (sub-sub plot) = 3

Table 2 : Estimates unweighted regression coefficient (b) of soil heterogeneity from yield data used in split plot design

Plot size	Calculated variance for entire area	Variance per unit area ($V_{\bar{x}}$)	log $V_{\bar{x}}$ (Y′)	No. units (X)	log X (X′)
Replication	9570.04	797.50	6.68	12	2.48
Main plot	18402.45	3067.08	8.03	6	1.79
Sub plot	12120.23	4040.08	8.30	3	1.10
Sub-sub plot	10636.14	10636.14	9.27	1	0

Thus the unweighted regression coefficient of soil heterogeneity (b) is b = 0.97

Now, using the method proposed in the present paper, we get the weights as elements w_{ik} of the information matrix of y

$$w = \begin{bmatrix} 1.03 & -0.17 & 0 & 0 \\ -0.17 & 5.76 & -7.19 & 0 \\ 0 & -7.19 & 11.66 & -3.48 \\ 0 & 0 & -3.48 & 9.71 \end{bmatrix}$$

Weighted matrix multiplying by logarithm variance per unit area (*i.e.*, $\text{Log}(V_x) = y_i$)

$$Y_i = \sum_k w_{ik} y_i$$

$$Y_1 = 5.51,\ Y_2 = -14.48,\ Y_3 = 6.78,\ Y_4 = 61.13,\ \sum_{i=1} Y_i = 58.94$$

Weighted matrix multiplying by logarithm of the area (*i.e.*, $\log x_i = x'_i$)

$$X_i = \sum_k w_{ik} x'_i$$

$$X_1 = 2.25,\ X_2 = 1.98,\ X_3 = -0.04,\ X_4 = -3.83$$

$$\sum_{i=1} X_i = 0.36,\ \sum X_i x'_i = 9.08,\ \sum_i Y_i x'_i = -4.8$$

Table 3 : Cost estimates in worker-hours for measuring yield of chickpea

Task	k_1 (cost / plots)	k_2 (cost/unit of test area/smallest unit)
Planting plans	1.17	
Seed preparation		0.13
Land preparation		0.65
Planting		0.49
Car of plots		0.81
Notes	0.58	
Harvesting	2.92	
Sampling	0.95	
Statistical analysis	0.78	
Total worker-hours	6.40	2.08

the sum of whose elements is $\sum_i \sum_k w_{ik} = \frac{spq(r-1)}{2} = 12$

$$U = -6.57, \; T = 9.07$$

Thus the weighted regression coefficient of soil heterogeneity (b) is $b = -\frac{U}{T} = 0.72$

We know, $X_{opt} = \frac{bk_1}{(1-b)k_2}$

putting the value of b (weighted), k_1 and k_2 then we get, $X_{opt} \approx 8$ (units)
Optimum plot size $= 8 \times 6 \text{ m}^2$ (unit as 6 m^2) $= 48 \text{ m}^2$

4. Concluding Remarks

Optimum plot size depends on the relative costs per plot and per unit area. It also depends on high or low value of soil heterogeneity. Here in this field experiment largest plot size is taken as 72 m^2 and optimum plot size obtained by using the technique developed in this paper is approximately 48 m^2. This shows that approximately 33% area of the field experiment as well as worker cost can be reduced by the methodology derived here.

REFERENCES

[1] Hatheway, W.H. and Williams, E.J. (1958). Efficient estimation of the relationship between plot size and the variability of crop yields. *Biometrics*, **14**, 207-222.

[2] Koch, E.J. and Rigney, J.A. (1951). A method of estimating optimum plot size from experimental data. *Agronomy Journal*, **43**, 17-21.

[3] Silla, G.R. and Nienhuis, J. (1993). Field plot technique affects snap bean yield evaluation. *J. Amer. Soc. Hort. Sci.*, **118(5)**, 672-674.

[4] Smith, H.F. (1938). An empirical lawdescribing heterogeneity in the yields of agricultural crops. *Journal of Agricultural Science*, **28**, 1-23.