INVESTIGATIONS ON SAMPLING FOR ESTIMATION OF CROP ACREAGES—II

By V. K. Mokashi

Institute of Plant Industry, Indore

Introduction

In an earlier article (Mokashi, 1953), the author dealt with the investigations carried out by him for estimation of crop acreages. The village was treated as a unit of sampling and results of single stage sampling of villages with complete enumeration of selected villages were presented in the article. There are, however, situations in which sub-sampling of village as a primary unit becomes necessary. One of the common problems in large-scale sample surveys is the designing of the sample survey within administrative restrictions as to the facilities available by way of staff for the field operations and time within which the survey has to be completed. Under such restrictions sub-sampling of the selected primary units (villages) is usually an advantageous device. For accurate preharvest forecasts of crop acreages, complete enumeration of villages would require that the patwaris (village accountants concerned) carry out their crop inspection earlier than usual. Since, however, the collection of area statistics is only one of the duties of the patwari, he may not get sufficient time for complete field inspection earlier. It would thus be desirable to lighten the work of the individual patwari by sub-sampling and spreading the sample over a large number of patwaris. For supervision of patwari's work also, sub-sampling offers a more practicable tool. The possibilities of sub-sampling within villages have been, therefore, explored and the results presented below.

The choice of sub-sampling units within a village presents features of some theoretical as well as practical importance. In extensive surveys, the determination of the appropriate type of units or a sub-unit is of primary importance. The time required for the observations of crop acreages being mainly dependent upon time spent in moving from field to field in a village, sub-sampling by selection of random fields for observation is not likely to be economical. Instead of enumerating a single field after having reached its location, it is obviously more economical to examine or enumerate a block of fields situated side by side as the additional time required for this purpose would be negligible. But economy in this respect should be considered in relation to the precision resulting from the use of blocks of different number of fields. Sampling of blocks of fields involves the study of the change in variance between blocks resulting from the change in size of the block, *i.e.*, the number of fields constituting a block. Investigations on this problem have been reported by Fairfield Smith (1938), Mahalanobis (1939), Jessen (1942), and Hansen and Hurwitz (1943). Although investigations were carried out in different fields of inquiry by these workers, the fundamentals of these problems were the same in all these cases. In the present case also, the available data for complete enumeration of agricultural area of individual fields for Akola District of Madhya Pradesh, has facilitated the investigation of the change in variance of blocks of fields.

Method of Selection and Analysis

Akola District consists of six tahsils. Each tahsil was stratified in four size strata, the size being determined by the geographical area of the village. For the present study, four villages were selected at random from each size-*cum*-geographical stratum, making a total sample of 96 villages in the whole district. Maps of selected villages were available. On each map two random points were marked. The procedure of marking was simple. It consisted of selecting at random two numbers, x and y, x for X axis and y for Y axis. The position was fixed by the pair of these numbers, giving the co-ordinates of the random points on the map. The field in which this point was located together with a neighbouring contiguous field so selected as to form a compact area, formed a block of two fields at this point. Then with the same initially selected random points as the base, the block size was increased to three, four and five contiguous fields, together forming a compact area.

Two crops, *viz.*, Cotton and Wheat were taken up for investigations. For Cotton, two variates were considered, namely, (1) Field Cotton area adjusted for differences in field size and (2) Proportion of Cotton area of a field to the total area of the field. The adjustment for the differences in field size was carried out by the usual technique of analysis of co-variance. For Wheat, the variates studied were (1) Field Wheat area and (2) Proportion of field Wheat area. Adjustment for differences in field areas was not carried out as it was anticipated that there would not be a high correlation between field size and field Wheat area. Secondly, from the results of Cotton it was found that for this particular problem, the difference between errors of adjusted estimate and estimated proportion was very little and hence to reduce the unnecessary laborious calculations the

JOURNAL

OF THE

Indian Society of Agricultural Statistics

VOLUME VI

Numbers 1—2

1954

PRINTED AT THE BANGALORE PRESS, BANGALORE CITY

AND

PUBLISHED BY

THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS

POST BOX NO. 310, C/O I.C.A.R., NEW DELHI

CONTENTS OF VOLUME VI

21-56 Printed at The Bangalore Press, Bangalore City, by C. Vasudeva Rao, Superintendent,

adjustment was omitted. Analysis of variance for each variate of both Wheat and Cotton was carried out for each size of the block. The results of the analysis of variance are given in Tables I, II and III. If 'k' is the number of strata, 'n' is the number of villages per stratum in the sample, 'm' the number of blocks selected per village and 'f' the number of fields per block, the analysis of variance can be put down as follows:

Due to	D.F.	Mean Square
Strata	$k - 1$	$M.S_1$
Villages within strata ..	$k(n - 1)$	$M.S_2$
Blocks within villages ..	$kn(m - 1)$	$M.S_3$
Fields within blocks ..	$knm(f - 1)$	$M.S_4$

From the analysis of variance, the estimate of variance of the mean per field can be shown to be

$$\hat{V}_{\bar{x}} = \frac{M.S_2}{kmnf}$$

or

$$V_{\bar{x}} = \frac{\sigma^2_v}{n_1} + \frac{\sigma_b^2}{n_1 m} + \frac{\sigma_f^2}{n_1 mf} \qquad (1)$$

where σ_v^2, σ_b^2 and σ_f^2 represent the true variances 'between villages', 'blocks' and 'fields' respectively. This formula requires that a constant number of sub-units should be selected within each primary unit. Here $n_1 = kn$.

Using formula (1) the estimate of variance can also be written as

$$\hat{V}_{\bar{x}} = \frac{\hat{\sigma}_v^2}{n_1} + \frac{M.S_3}{n_1 mf} \qquad (2)$$

The true variance between villages would be independent of the block size; but the variance of the block means, the quantity $M.S_3/f$ given in (2) is related to block size. The quantity $M.S_3/f$ or the variance of the block means does not behave in the usual manner as in sub-sampling of random fields within villages, where sub-units are independently selected. By grouping the fields into blocks of fields, the intra-class correlation between fields within blocks comes into picture and it is upon the magnitude of this correlation that the block variance depends. The problem that presents itself is to

TABLE I

Analysis of variance of cotton area (adjusted)

Due to	2 fields per block		3 fields per block		4 fields per block		5 fields per block	
	d.f.	Adjusted m.sq.	d.f.	Adjusted m.sq.	d.f.	Adjusted m.sq.	d.f.	Adjusted m.sq.
Between villages ..	72	38·76	72	42·86	72	49·44	72	56·66
Between blocks within villages ..	96	28·47	96	26·81	96	27·87	96	26·28
Between fields within blocks ..	192	28·75	384	26·77	576	27·07	768	24·72

TABLE II

Analysis of variance of proportion of cotton and wheat areas to total area

Due to	2 fields per block			3 fields per block			4 fields per block			5 fields per block		
	d.f.	m.sq.	m.sq.	d.f.	m.sq.	m.sq.	d.f.	m.sq.	m.sq.	d.f.	m.sq.	m.sq.
		Cotton	Wheat		Cotton	Wheat		Cotton	Wheat		Cotton	Wheat
Between villages ..	72	·0836	·019136	72	·1031	·023578	72	·1141	·027301	72	·1386	·031164
Between blocks within villages ..	96	·0602	·017949	96	·0635	·016411	96	·0713	·016938	96	·0737	·017637
Between fields within blocks ..	192	·0685	·015808	384	·0606	·016560	576	·0626	·016111	768	·0612	·013467

TABLE III

Analysis of variance of area under wheat

Due to	2 fields per block		3 fields per block		4 fields per block		5 fields per block	
	d.f.	m.sq.	d.f.	m.sq.	d.f.	m.sq.	d.f.	m.sq.
Between villages ..	72	6·53	72	7·02	72	7·69	72	10·82
Between blocks within villages ..	96	6·30	96	5·02	96	5·62	96	7·48
Between fields within blocks ..	192	5·41	384	5·28	576	4·85	768	5·77

determine the relationship between the variance of the block mean and its size. From this relationship, the accuracy of the sample estimate with varying block sizes can be predicted.

VARIANCE FUNCTION

The general problem is that of estimating the variance of a sample of clusters of any size from a sample of clusters of given sizes. This is possible if we know the relationship between $M.S_3/f$, the variance 'between clusters' and cluster size 'f' say. Jessen (1942) has shown that the variance within clusters can be expressed as

$$\hat{\sigma}_w^2 = Af^b \quad (3)$$

where A and b are constants. Hendricks (1948) has discussed the method of using this relationship to estimate the variance of the cluster or block means. We have a sample of blocks of given size (f), *i.e.*, the number of fields. $\hat{\sigma}_w^2$ is then given by

$$\log A + b \log f = \log \text{(variance within clusters of size } f) \quad (4)$$

If we regard the village as a single cluster of 'F' fields, then extending (3) to this situation, we get

$$\log A + b \log F = \log \text{(variance within blocks of size } F) \quad (5)$$

The estimates of A and b can be obtained from the solution of simultaneous equations (4) and (5). From these estimates, it can be shown according to Hendricks (1948) that the variance of the cluster or block means is given by

$$\hat{V}_B = \frac{AF}{F-f}[(F-1)F^{b-1} - (f-1)f^{b-1}] \quad (6)$$

The equation involves two constants which can be estimated from the variance among fields in a village and variance within blocks for one value of f. We can estimate the population variance between fields within villages from the analysis of blocks of given size. F is the average number of fields per village. This particular procedure was followed in the present study. For determining the values of A and b, blocks of two fields are considered for equation (4) wherever this block size showed positive intra-class correlation between fields within blocks. In other cases, blocks of three fields are considered for equation (4) as they exhibited this correlation. From the four analyses of variance for the four block sizes, we can calculate the actual or the observed value of V_B for each size and for blocks of the same size, using (6), we can estimate the value of V_B. The observed and the expected values of the block variance are given in Table IV. It is seen from Table IV that the agreement between observed and expected

TABLE IV

Observed and theoretical values of variance of the block means

Variates / Size of the block	Proportion of cotton area		Field cotton area adjusted		Proportion of wheat area		Field wheat area	
	Observed	Theoretical	Observed	Theoretical	Observed	Theoretical	Observed	Theoretical
2 fields per block ..	·0301	·0313	14·23	13·50	·008971	·008970	3·15	3·17
3 fields per block ..	·0212	·0211	8·93	9·06	·005470	·006298	1·70	2·24
4 fields per block ..	·0178	·0161	6·96	6·80	·004234	·004932	1·40	1·77
5 fields per block ..	·0147	·0130	5·26	5·45	·003527	·004080	1·49	1·48

values obtained from equation (6) for calculating the variance of the block means can be considered satisfactory.

Thus equation (6) may be regarded as providing a satisfactory graduation of the block variance for the change in its size. This value of the block variance can be substituted in the equation (2) in place of $M.S_3/f$. Equation (2) can then be written as

$$\hat{V}_{\bar{x}} = \frac{\hat{\sigma}_v^2}{n_1} + \frac{AF}{F-f}\,[(F-1)\,F^{b-1} - (f-1)\,f^{b-1}]\cdot\,\frac{1}{n_1 m} \qquad (7)$$

Equation (7) now can be used to determine the number of villages and the number of blocks of given size per village required to provide the estimate of the mean per field with a given precision. Tables V, VI, VII and VIII show this number for estimation of the mean with five per cent. standard error.

TABLE V

Field cotton area adjusted for field size

No. of fields per block	No. of blocks per village		
	2	3	4
	No. of villages		
2	99	76	64
3	76	60	53
4	64	53	47
5	57	48	43

TABLE VI

Proportion of area under cotton

No. of fields per block	No. of blocks per village		
	2	3	4
	No. of villages		
2	97	73	60
3	74	59	51
4	63	51	45
5	56	46	42

TABLE VII

Proportion of area under wheat

No. of fields per block	No. of bolcks per village		
	2	3	4
	No. of villages		
2	758	562	463
3	583	445	376
4	493	385	356
5	436	347	303

TABLE VIII

Field wheat area as a variate

No. of fields per block	No. of blocks per village		
	2	3	4
	No. of villages		
2	1181	856	688
3	894	662	544
4	744	562	469
5	656	500	425

For Cotton, the number of villages required for five per cent. standard error are quite small, but for Wheat the number is relatively much larger as the crop is much variable. The size of the sample for Wheat (Tables VII and VIII) is expected to be obviously larger than that for Cotton (Tables V and VI). This is because of the comparatively lower intensity of Wheat cultivation. A similar conclusion was arrived at by Mokashi (1953) for these two crops for one stage sampling (treating village as a sampling unit at 1st stage). In fact two crops were taken up for investigation in order to verify the existing belief that areas under minor crops could be estimated with a tolerable degree of accuracy only with a relatively larger sample size. Akola District is mainly a cotton tract and Wheat crop is grown only where moisture retention capacity of the soil is more. Consequently the Wheat crop shows greater variation. Sampling on this scale for estimation of Wheat area on a district level is obviously impracticable. If the tolerable degree of accuracy of the estimate for Wheat area can be extended to 10 per cent. standard error, the number of villages for this level of accuracy would get reduced to one-fourth of the number given in the above tables. This number would then be sufficiently small to be manageable. Another point that emerges from these tables is that a small number of blocks with a larger number of fields per block is almost equivalent to a large number of blocks constituted from a smaller number of fields as far as statistical precision is concerned. This is due to the small magnitude of the intra-class correlation between fields within blocks. The former types of blocks are, therefore, to be preferred as being more economical for field inspection. It would appear from the present results that two blocks, each consisting of four fields form an approximately optimum sub-sample per village. There would be little further gain in accuracy by increasing the size of the block beyond four fields.

SUMMARY

Investigations reported in the present article were undertaken with the object of developing the area method of sampling for estimation of crop acreages. The importance of developing efficient methods of sampling for this purpose would be appreciated when it is stated that such sampling would serve three-fold objectives, *viz.*, (1) improving the accuracy of the area statistics in temporarily settled states by introducing sample checks over the patwari's work of complete enumeration, (2) preparing preharvest forecasts of crop acreages and (3) estimation of crop acreages in permanently settled and other areas. For evolving an efficient plan of sampling, knowledge of different components of variability of the material is a necessary prerequisite. The sampling studies in this article were, therefore, made in conjunction with the statistics of complete enumeration of crop areas in the Akola District of Madhya Pradesh for the year 1945–46, taking Cotton and Wheat acreages as examples of crops with a high and relatively lower intensities of cultivation respectively.

The sampling pattern examined was what is known as cluster sampling or block sampling in which blocks of contiguous fields constitute the sub-sampling units within the primary units, here villages. The problem of cluster sampling involved the study of the relationship

between variance and size of the block. Jessen's law was studied in this connection and sample size for a given degree of precision was determined.

It was found that Jessen's law provided a satisfactory relationship between the block variance and its size. As far as statistical precision is concerned, two blocks, each of four fields, should form a sub-sample within a village. Thus in sub-sampling a sample of about 64 villages with two blocks of four fields each per village is likely to provide an estimate of the mean cotton area per field with five per cent. standard error. For the proportion of Cotton area, a sample of about 63 villages with the same sub-sampling rate is expected to provide estimate of cotton area with five per cent. standard error. For the estimation of Wheat area per field, about 750 villages with the same sub-sampling rate might be taken to attain the same precision. For the proportion of Wheat area, 500 villages with the same sub-sampling rate would be the sample size required for the above precision.

Acknowledgments

These investigations were carried out under the guidance of Dr. V. G. Panse, B.Sc., Ph.D. (London), Statistical Adviser, I.C.A.R., New Delhi. The author wishes to express his deep gratitude for the keen interest that he took in guiding him and making valuable suggestions during the course of these investigations.

References

1. Hansen, M. H. and Hurwitz, W. N. .. "On the Theory of Sampling from Finite Populations," *Annals of Math. Stat.*, 1943, **14**, 333.
2. Hendricks, W. A. .. "Mathematics of Sampling," *Virginia Agricultural Experiment Station, Special Tech. Bulletin*, 1948.
3. Jessen, R. J. .. "Statistical Investigation of a Sample Survey for obtaining Farm Facts," *Iowa Agric. Exp. Sta. Res. Bull.*, 1942, **304**.
4. Mahalanobis, P. C. .. "A Sample Survey of the Acreage under Jute in Bengal," *Sankhya*, 1939, **4**, 511.
5. Mokashi, V. K. .. "Investigation on Sampling for Estimation of Crop Acreages—I," *Jour. Ind. Soc. Agri. Stat.*, 1953, **5**, 128.
6. Smith, Fair Field .. "An Empirical Law Governing Soil Heterogeneity," *Jour. Agri. Sc.*, 1938, **28**, 1.