JOURNAL

OF THE

Indian Society of Agricultural Statistics

Vol. VI] 1954 [No. 2

EFFICIENCY OF SAMPLING METHODS IN FOREST SURVEYS

By V. K. Mokashi

Forest Statistician, Bombay State, Poona-1

Introduction

It has been universally recognised that a properly planned sample can provide estimates of the character under study with a requisite degree of accuracy. Modern sampling practice is the outcome of the development of the statistical theory appropriate to the characteristics of the material to be sampled simultaneously with the actual experience gained from the use of the various sampling techniques in large-scale sample surveys. In the past few years, scientific sampling methods have been employed extensively in the planning and conduct of large-scale surveys in the fields of agriculture and economics and information on items of agro-economic interest, such as acreage and production of crops, income from and expenditure on farming, etc., is being collected through such surveys. Sampling is steadily extending its application to many other fields, such as commerce, industry, education, forestry and government administration. With the adoption of planned economy and development of social sciences, sample surveys have assumed increasing importance in these sectors of human activity. In India, sampling methods were developed primarily for collecting agricultural statistics involving estimation of crop yields and acreages and sample surveys have become a routine annual feature in agricultural statistics. As far as forest surveys are concerned, no detailed research is conducted so far for developing sampling methods suitable for different forest conditions in India. The 8th Silvicultural Conference held at Dehra Dun in 1951 has emphasized the development of proper sampling methods under Indian conditions. It is well known that a complete enumeration of forest area for estimation of its growing stock, presents certain difficulties and sometimes the task may practically be impossible. Even when feasible it involves considerable expenditure of time, labour and money. It is, therefore, frequently advantageous to estimate the growing stock in a given forest stand, by examining a sample from it, thus making the procedure of estimation more expeditious and economic. A notable beginning in the development of suitable sampling methods in forest surveys for estimation of timber volume was made by Griffith (1945) who examined many complete and partial enumeration records from different parts of India and discussed the efficiency of sampling methods. Another important advance in this direction was made by Finney (1948) who studied the data used by Griffith in a greater detail and examined the efficiency of systematic strip sampling in relation to stratified random sampling. Finney (1948) showed that systematic sampling was more precise than stratified random sampling but indicated that the difference in precision in the two is not substantial to justify the general recommendation that systematic sampling should always be adopted. He had recommended further collection of evidence in this respect by investigating the efficiency of systematic sampling and stratified random sampling on some other forests. Similar investigations were, therefore, carried out in parts of Bombay forests and the results on the efficiency of systematic and stratified random sampling by strips are discussed in this article.

The necessary prerequisite for any sampling procedure is that the estimate obtained thereby should be unbiased and should have a specified precision. For ensuring unbiased estimates, the selection of the sample should be random (unrestricted or stratified) or systematic with random selection of 1st unit. Precision of the estimate can be obtained from a simple random sample or a stratified random sample with selection of two units from each stratum or block. But a single sample in case of systematic sample or stratified random sample with one unit from each block, fails to provide the precision of the estimate though the latter methods are generally found to be more precise. The problem for investigation is, therefore, whether to adopt stratified random sample with two units per block and obtain the precision of the estimate or sacrifice the knowledge of precision by using systematic sampling or stratified random sampling with one unit per block, assuming the latter to be more precise. Definite recommendation for any particular forest cannot be made without past experience about the efficiency of a particular method but general indication can be given about the likely gain or loss in different methods from the examination of data of similar forests. Complete enumerations on a large forest area are helpful in carrying out such studies. Such data on complete enumeration of Taloda forest in Bombay State were collected for examining the efficiencies of the different sampling methods.

MATERIAL AND METHODS OF ANALYSIS

Cent per cent. enumerations were carried out in Taloda division of Bombay State forests by strips in eight groups one from each felling series. The groups were not contiguous. In each group ten important species of timber were enumerated by strips across the contour. Each strip was two chains in width and data were recorded for one acre unit within each strip. The total area covered is approximately 4,000 acres of land. Data used for this paper were the total volumes of timber, irrespective of the species. A strip was treated as a sampling unit and the mean volume per acre for each strip was used as a variate, the differences in the lengths of the strips being ignored for simplification. This was necessary for comparison with past results. The total number of strips in the whole area is shown below according to series:

Series No.	Total number of strips in the series completely enumerated
1	72
2	48
3	48
4	48
5	56
6	72
7	56
8	48
	448

The mean volume per acre for the whole area is 462·9 c.ft.

The principles of different types of strip sampling have been stated by Finney (1947) and followed by him in actual investigations (Finney, 1948). The same principles are followed in the present investigations. If the forest area consists of N strips in all, a 1 in r simple random sample requires N/r strips selected at random from N strips. Two types of stratified random sampling are considered in this investigation. For an intensity of sampling, 1 in r, the first type of stratification is introduced by subdividing the area into blocks of $2r$ strips and selecting 2 strips at random within each block. The second type of stratification is obtained by dividing the area into blocks of r strips and selecting one strip from each block. Though a single sample in the 2nd type fails to provide the estimate of sampling error, it is expected to provide estimates with greater accuracy as compared with first one. True sampling errors can, however, be available from the analysis of complete enumeration data and comparisons can be made between the two types of stratification from such analysis. Thus three types of random sampling were considered here. Total variance per strip within the area will be denoted by σ_1^2, that within blocks of $2r$ by σ_2^2 and that within blocks of r by σ_3^2. σ_1^2 is independent of the intensity of sampling and σ_2^2 and σ_3^2 will change with the block size. σ_2^2 for any r is equal to σ_3^2 for $2r$. Analysis of variance was carried out by Cochran's method (Cochran, 1939) separately for each of the eight series and weighted averages were calculated by using weights proportional to the number of strips per series. For example, let us take the first series which consists of 72 strips. We can divide this series into blocks, each of 2, 3, 4, etc., strips at a time and carry out analysis of variance as given below:

	Source of variation	d.f.	M. Sq.
A	Between blocks of 2 strips ..	35	..
	Within blocks of 2 strips ..	36	..
B	Between blocks of 4 strips ..	17	..
	Within blocks of 4 strips ..	54	..

and so on. The estimates of σ_3^2 will be provided by the within block mean square and by definition, σ_2^2 for any r is equal to σ_3^2 for $2r$. An examination of the complete enumeration data for the study of systematic sampling is carried out usually by forming the set of all possible 1 in r systematic samples with regular spacing and calculating the variation between the members of this set. Finney (1948) has pointed out two disadvantages of this method. The first disadvantage is that the estimated variance for 1 in r is based on only $(r - 1)$ degrees of freedom. Secondly, if there is a marked trend in volume per unit area from one end to the other end, random selection would cause undue dispersion on the estimates obtained by repeated sampling. Finney (1948) has, therefore, studied systematic sampling with selection of 1st strip at random followed by a centralizing adjustment for the whole sample. Detailed arithmetical technique for this method has been given by Finney (1948). Similar analysis has been carried out for systematic sampling in the present investigations. Each series has been divided into small groups consisting of $(rk + r - 1)$ strips, where 1 in r is the intensity of sampling and k is equal to number of strips in one sample. As the number of strips in each series is not sufficiently large, analysis could be carried out only for $r = 11$ or less and $k = 3$. To give an example, Finney has given the arithmetical technique as follows:

Here k is equal to 4. The mean volume per acre for the strips of Mount Stuart may be denoted by $x_1, x_2, x_3, x_4, \ldots, x_{160}$. Consider systematic sampling at the intensity of 1 in 12 of the first 59 strips. A centralized sample will then be adjusted so as to relate to the position of the strip 30 after removal of the effect of any linear trend in the first 59 values. The adjusted totals for the 12 possible samples are

$$x_{12} + x_{24} + x_{36} + x_{48}$$

$$1/12x_1 + x_{13} + x_{25} + x_{37} + 11/12x_{49}$$

$$2/12x_2 + x_{14} + x_{26} + x_{38} + 10/12x_{50}$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$10/12x_{10} + x_{22} + x_{34} + x_{46} + 2/12x_{58}$$

$$11/12x_{11} + x_{23} + x_{35} + x_{47} + 1/12x_{59}.$$

A sum of squares of deviations may then be calculated for these 12 totals, the results being 817900 (11 degrees of freedom). The calculations are repeated with another group, which may overlap the first, say, strips 11 to 69, using totals

$$x_{22} + x_{34} + x_{46} + x_{58}$$

$$1/12x_{11} + x_{23} + x_{35} + x_{47} + 11/12x_{59}$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$11/12x_{21} + x_{33} + x_{45} + x_{57} + 1/12x_{69}$$

to give another sum of squares with 11 degrees of freedom. So further groups are used.

For the present study variances were obtained separately for each series and were later on combined by weighting in proportion to the number of strips per series.

RESULTS

If $\bar{x}$ is an estimate of mean volume per acre for the whole area, its variance may be shown to be

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right).$$

As $N/n = r$ and r is the same for each block, the estimated variance of the mean volume per acre from a sample of n strips is

$$s^2_{\bar{x}} = \frac{s^2}{n}\left(1 - \frac{1}{r}\right).$$

Comparisons can be obtained for all the methods under consideration, after substituting the corresponding value for s^2.

The method of estimation of variances is already stated. Estimated variance, s_1^2, for unrestricted random sampling from the whole area was

$$s_1^2 = 28019.$$

Estimates of variances for two different modes of stratification with random sampling in each, and systematic sampling at intensities of 1 strip in r strips and r ranging from 2 to 12 were obtained separately for each series as mentioned earlier and weighted variances were calculated for the whole area. The values of variances are given in Table I. The values of s_2^2 for any r are identical with s_3^2 for $2r$.

TABLE I

Estimates of variance per strip for sampling of 1 *in* r
[In units of (cu.ft. per acre)^2]

r	s_2^2	s_3^2	s_4^2
2	5678	5174	4136
3	6068	5294	5163
4	6392	5678	5591
5	..	..	3599
6	6953	6068	4565
7	..	..	5060
8	7311	6392	5156
9	..	..	6155
11	..	..	4631
12	7911	6953	..

It was observed that for the same intensity of sampling the variances of stratified random sample with one strip from each stratum and systematic sampling were consistently lower than that for stratified random sample with two strips within each stratum. Secondly, there was a definite trend between the size of the stratum and the variances or in other words, the variance was seen to change with intensity of sampling. Such a trend is useful in comparing the efficiencies of different methods.

Fig. 1 shows that the regression of log s^2 on log r was approximately linear for the values studied. Such relationships have been reported by Smith (1938), Jessen (1942) and Mahalanobis (1944) and recently by Finney (1948) for forest surveys. The regression lines are shown in Fig. 1 and their equations are as follows:

$$\log S_2^2 = 3\cdot6954 + 0\cdot1871 \times \log r$$

$$\log S_3^2 = 3\cdot6525 + 0\cdot1709 \times \log r$$

$$\log S_4^2 = 3\cdot6185 + 0\cdot0904 \times \log r$$

where S^2 is the smoothed estimate of the corresponding σ^2 obtained from the regression.

From the foregoing results, it is possible to compare the efficiencies of the four methods under consideration, *viz.*, (1) Unrestricted random sampling, (2) Stratified random sampling with two strips per stratum, (3) Stratified random sampling with one strip per stratum of half the size of the original stratum, (4) Systematic sampling. The efficiency of systematic sampling in relation to other types of sampling is given by the inverse ratio of the variances of estimates obtainable at any specified intensity of sampling. The percentage efficiency will be given by

$$\text{Efficiency} = 100 \frac{S^2}{S_4^2}$$

where for S^2, any variance to be compared may be substituted. These values were calculated for the two modes of stratification and are given in Table II.

It was evident that systematic sampling was more efficient in relation to both types of stratified sampling. The second type of stratification was more efficient than the first type. Again stratified and systematic sampling were both more efficient as compared with unrestricted sampling. The increase in efficiency due to systematic sampling over 1st type of stratification for intensities of 1 in 10 to

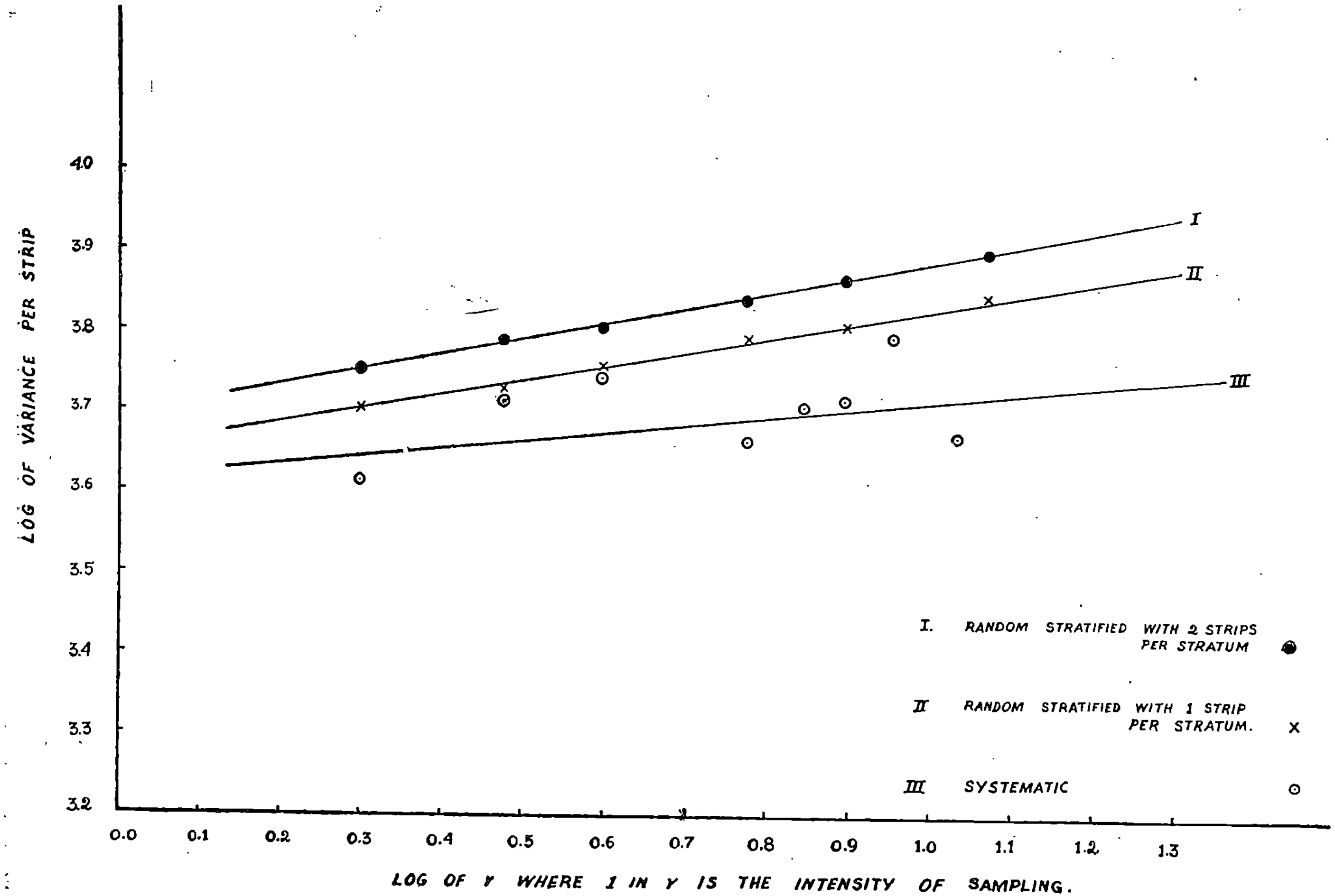

FIG. 1. Showing relationship between log s^2 and log r

TABLE II

Percentage efficiency of Systematic Sampling in relation to two types of Stratified Sampling

Intensity	$100 \frac{S_2^2}{S_4^2}$	$100 \frac{S_3^2}{S_4^2}$
2	127	114
3	133	118
4	136	121
5	139	123
6	142	125
8	146	128
10	149	130
12	152	132
20	159	137

1 in 20 varies from 49 to 59 while the corresponding increase over 2nd type of stratification for the same intensities is of the order of 30 to 37. Thus it was clear that for these intensities of sampling, systematic sampling was more advantageous in comparison to 1st type of stratification while its advantage over 2nd type of stratification was much reduced for the same intensities. The sampling errors per strip may be shown as percentages of the mean volume per acre. Percentage standard error per strip can, therefore, be written as

$$\text{Percentage Standard Error per strip} = \frac{100S\sqrt{(1-1/r)}}{\bar{x}}$$

where the factor $1 - 1/r$ is the adjustment for finite population sampling and $\bar{x}$ is the mean volume per acre. When this percentage *S.E.* per strip is divided by $\sqrt{n}$ (n being the number of sample strips), we obtain the percentage standard error of the sample mean. For comparing the various sampling procedures the standard errors per cent. per strip as given by the formula mentioned above, were plotted against r for each of the series and are shown in Fig. 2. It will be seen that percentage standard errors for the same intensity of sampling were very high for the method of unrestricted sampling, as compared with the remaining three methods of sampling. The results are approximately similar to those obtained by Finney (1948). Methods 3 and 4 did not show a very substantial advantage over method 2 with regard to the percentage standard errors of the mean volume per acre and therefore we may conclude that in practical sampling problems it is desirable to follow method 2 instead of methods 3 and 4, and obtain estimates, the errors of which could be measured from a single sample. Stratified random sampling with

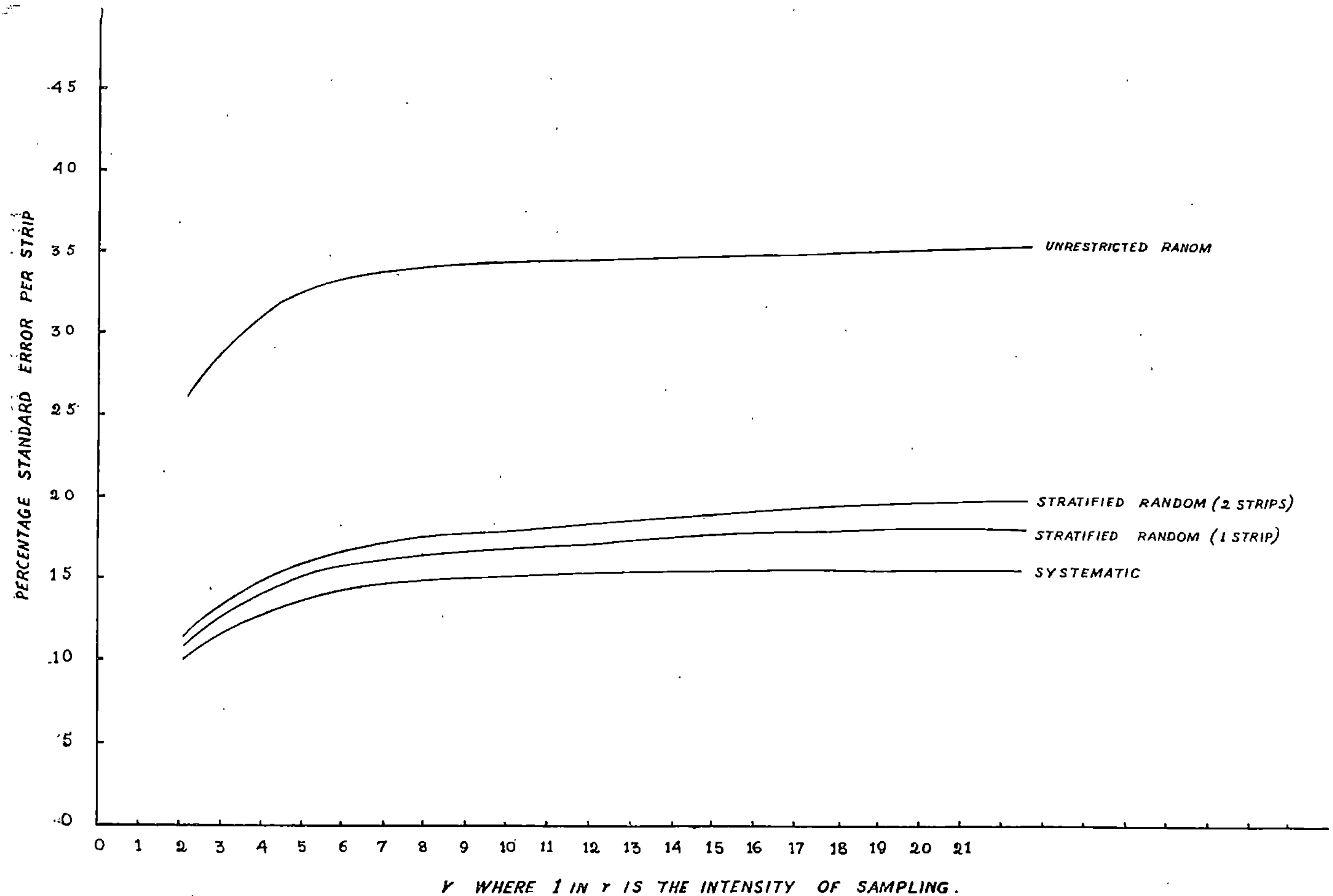

FIG. 2. Showing variation of percentage *S.E.* per strip with the intensity of sampling.

two strips per stratum (method 2) was far more efficient in relation to unrestricted sampling (method 1) and should, therefore, be obviously preferred to the latter.

The choice of the sample size was determined by the standard error of the mean value. The standard errors of the mean were worked out for each method for certain intensities of sampling and results are given in Table III.

TABLE III

Percentage standard errors of mean volume per acre

r	Method 1	Method 2	Method 3	Method 4
5	3·42	1·67	1·57	1·42
8	4·52	2·32	2·17	1·91
10	5·13	2·68	2·50	2·19
12	5·69	3·03	2·82	2·44
20	7·45	4·15	3·85	3·28

Table III shows that in a stratified random sample with two strips per stratum, a sampling intensity of 1 in 5 strips was expected to provide estimates with 3·34% sampling error, at half of this intensity of sampling, *i.e.*, 1 in 10 strips, the error would be 5·36% and for a sampling intensity of 1 in 20 strips, the estimate would be obtained with the error of 8·30%. These errors were considerably smaller than those expected in unrestricted sampling. Stratified random sampling with one strip from each stratum and systematic sampling were subject to lesser errors.

DISCUSSION AND CONCLUSIONS

Griffith (1945) and Hasel (1938) had shown that complete enumerations were not necessary for estimation of the growing stock and that small samples based on scientific principles could provide estimates with a reasonable degree of accuracy. Finney (1948) examined the data of Griffith and Hasel in a greater detail and carried out a comprehensive study of sampling problem in forest enumerations. He also came to the conclusion that complete enumerations can be replaced by a properly planned sampling design for enumerations in forests. In the present investigation also, the author has made a further attempt to examine the results of investigations carried out on the same basis as those of Finney, though in a different forest. The conclusions arrived at by the present author confirm the findings of Finney (1948) that a properly planned sample could provide estimates with a fairly tolerable degree of accuracy. It has been observed that stratified random sampling with two units per stratum was far more efficient than unrestricted sampling but systematic sampling was more efficient than stratified random sampling. The problem to be settled is whether stratified random sampling or systematic sampling should be recommended. As mentioned earlier, stratified random sampling with one unit per stratum or systematic sampling fails to provide the precision of the estimate from the internal evidence of a single sample, while stratified random sample with two strips per stratum is capable of providing the precision of a single sample. The former two methods, though more accurate than the latter one, may not be recommended in actual practice especially because the difference in precisions is not very substantial so that we cannot afford to sacrifice the knowledge of the sampling error attainable with the other method. The practicable method to be recommended appears to be that a given forest should be divided into such blocks that for a given intensity of sampling, two strips are selected at random from each block.

Second feature of this investigation is that there is a definite correlation between neighbouring strips and that within stratum variance is related to the size of the stratum. Such relationship was established in this investigation and was used to compare the efficiency of different sampling procedures.

Summary

The most important problem in forest enumerations is to evolve the best method of sampling which, apart from being efficient, is also convenient in practical application. Finney has discussed the efficiencies of different methods of strip sampling in forest enumerations. Following Finney, the present author has carried out investigations on similar lines.

Data for complete enumerations on Taloda Forests of Bombay State had been employed for studying the efficiencies of the following four methods for estimation of timber volume per acre:

1. Unrestricted sampling by strips.
2. Stratified random sampling with two strips per stratum.
3. Stratified random sampling with one strip per stratum of half the size of stratum used in (2).
4. Systematic strip sampling.

Appropriate variances have been obtained for these methods at sampling intensities varying from 1 in 2 to 1 in 12. Empirical relationship of the form $V = ar^b$ had been established for the latter three methods (here V is the variance and r is the intensity of sampling). This relationship was used to graduate the variance for a given intensity for each of the latter three methods and the variances thus obtained were compared for studying the efficiencies of the different sampling methods.

It has been observed that methods 3 and 4 are more efficient and precise than method 2 which in turn is far more precise and efficient than the method 1. However, the difference between precisions obtainable by methods 3 or 4 and method 2 is not very substantial and the method 2 which is capable of providing estimates of error from a single sample, may therefore be recommended. The disadvantage of the methods 3 and 4 is that, though these are more efficient they fail to provide the precision of the estimate from the internal evidence of a single sample.

It can be concluded from the present investigations that complete enumerations can be safely replaced by sample enumerations based on a properly planned and selected sample. It was observed that in a stratified sample with two strips per stratum, a sampling intensity of 1 in 5 strips is expected to provide estimates of volume per acre with a percentage sampling error of 3·34%; at half the intensity, the error would be 5·36%, and at sampling intensity of 1 in 20, the estimate would be obtained with 8·30% sampling error.

ACKNOWLEDGMENT

The author had the benefit of discussing the results with and receiving valuable suggestions from Dr. V. G. Panse, I.C.A.R., New Delhi, and Dr. K. R. Nair of Forest Research Institute, Dehra Dun, for which thanks are due to them.

REFERENCES

1. Cochran, W. G. .. "The Use of the Analysis of Variance in Enumeration by Sampling," *Jour. Amer. Stat. Assoc.*, 1939, **34.**
2. Finney, D. J. .. "Volume Estimation of Standing Timber by Sampling," *Forestry*, 1947, **21**, 2.
3. ——— .. "Random and Systematic Sampling in Timber Surveys,". *ibid.*, 1948, **22.**
4. Griffith, A. L. .. *The Efficiency of Enumerations*, Forest Research Institute, Dehra Dun, Forest Leaflet No. 85.

5. Hasel, A. A. .. "Sampling Error in Timber Surveys," *Jour. Agric. Res.*, 1938, **56.**

6. Jessen, R. J. .. "Statistical Investigation of a Sample Survey for obtaining Farm acts," F*Iowa Agri. Expt. Sta. Res. Bull.*, 1942.

7. Mahalanobis, P. C. .. "On Large-Scale Sample Surveys," *Phil. Trans. Roy. Soc.*, London, 1944, **231,** 41.

8. Smith, H. F. .. "An Empirical Law describing Heterogeneity in the Yield of Agricultural Crops," *Jour. Agric. Sci.*, 1938, **28.**