ON A TECHNIQUE OF FORMING SAMPLING UNITS

By PAUL JACOB

Indian Statistical Institute

SUMMARY. A method of formation and selection of area units in a progressive manner suggested by Professor D. B. Lahiri is discussed in this note. This method is likely to be of considerable help in cases where well-demarcated area units are not readily available or where identification of an area unit around a randomly located point presents subsequent operational problems. The results of two surveys where this method has been tried out are also given.

1. Introduction

In sample surveys, specially those involving area sampling (e.g. in big plantation areas, forests etc.) the demarcation of a selected sampling unit from a population of overlapping or non-overlapping units presents difficult operational problems, and is subject to personal judgement and discretion at the final stage of the formation or identification of the unit possibly leading to some bias. In this note a technique of formation and selection of units in a progressive manner suggested by Professor D. B. Lahiri is discussed. This technique is likely to be particularly useful in situations where well-defined non-overlapping area units are not readily available and where location of an area unit around a randomly selected point presents operational difficulties leading to possible bias.

2. Lahiri's method of formation and selection of sampling units

The population (for example, say, a paddy field for which the yield is to be estimated) is first divided on the basis of an eye estimation of size (relevant to the characteristic under consideration) into two halves. The demarcating line between the two halves thus obtained will be clearly defined with the elements of units on the demarcating line properly allotted to either of the halves. One of the two halves will then be selected at random. The selected unit will again be divided into well-defined halves and one of them selected at random. This process of division into halves and selection of one at random is continued till a unit of acceptable size is arrived at and this is the ultimate sampling unit selected. The advantage of this method is obvious. The unit is already well demarcated before selection, thereby avoiding chances of any bias involved in demarcation in the conventional method.

It may be noted that this technique does not call for size or any such character of the population for the estimation of the population characteristic. The sample value is inflated by 2^n to obtain the population total where n is the number of splits required to arrive at the ultimate unit. This method is highly useful for populations, like forest areas or big plantations. This can also be successfully applied to sampling from huge stocks of census slips, punched cards etc.

3. TWO CASE-STUDIES

The Indian Statistical Institute adopted this technique (Lahiri's method) in two large scale yield surveys carried out in the recent past (1) Cinchona Survey, 1963, to estimate the quinine quantity in the Russian method cinchona plantations of Nilgiris and Anamallai regions, Madras and (2) Tobacco Survey, 1963-64, to estimate the yield rate of Flue cured Virginia Tobacco in a selected block in Andhra Pradesh.

3.1. *Cinchona Survey, Madras*, 1963. This survey was carried out to obtain an estimate of standing cinchona bark and the yield of quinine sulphate in a particular area of the cinchona plantation. The plantation presented a typical forest-like area with hilly terrain and slopes containing land slides, mountain streams, uncultivated patches with rocks and thorny bushes, staff quarters and coolie lines etc. The area was never properly cadastrally surveyed and only rough sketch maps were available giving only a very notional idea of the area.

Under such a situation, with no proper frames, any conventional method of sampling would have been difficult. Lahiri's method was therefore tried out. The entire area was initially split up into sub-plots of average area of about 10 acres using natural boundaries like streams, footh paths etc. to the extent possible, and by means of artificial markings or peggings made on the ground wherever necessary. Treating these sub-plots as separate strata sampling units were selected from each stratum using the method of splitting. Each sub-plot (stratum) was successively divided into halves and one half selected at random, till a plot containing about 8-10 cinchona plants was obtained. This was the ultimate sampling unit.

In actual practice, two sample units in each stratum were selected by the above procedure, and surveyed by two independent parties of Investigators giving two independent estimates. It may be noted here, that it would have been desirable to have the splitting process for selection of sample units done independently, by the two different parties, so that the samples would be completely independent. But in view of the tremendous field difficulties and cost involved in this, both parties jointly made the division and selection processes till they selected different halves.

It was pointed out earlier that the estimation procedure is very simple. The population total is obtained by inflating the sample value by 2^n where n is the number of splits to arrive at the sample unit. Table 1 gives the estimated yield of dry bark by sub-samples (s.s.) separately for the Nilgiris and Anamallais regions. The standard errors given in col. (7) are built up from the sub-sample differences at the stratum level.

3.2. *Tobacco Survey, Andhra Pradesh*, 1963-64. This survey was carried out in Guntur district of Andhra Pradesh for estimating the acreage and yield rates of Flue cured Virginia (FCV) Tobacco. A three-stage stratified sampling design was adopted with villages and fields as first-and second-stage units and finally areas containing about 49 plants as ultimate units. These ultimate units were obtained by applying Lahiri's method of splitting and selection to each selected field.

TABLE 1. ESTIMATED YIELD OF DRY BARK IN THE JOINT CULTIVATED AREAS OF CINCHONA PLANTATION SEPARATELY FOR NILGIRIS AND ANAMALLAIS REGIONS

region	number of sample units		estimated dry bark in 000 kilogram			percentage error
	s.s. 1	s.s. 2	s.s. 1	s.s. 2	combined	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Nilgiris	77	77	583	438	510±31	6.08
Anamallais	312	312	1229	1138	1184±54	4.56
Nilgiris and Anamallais	389	389	1812	1576	1694±62	3.66

The estimated number of tobacco plants per acre worked out on the basis of returns from Lahiri's samples and the reported area of the fields is given in col. (3) of Table 2 with the percentage error in col. (4). It will be quite interesting to compare the results obtained through these samples with those estimated through the conventional 'area' units in the shape of random squares. Columns (5) and (6) of the table give the estimated number of plants per acre as obtained from a sample of square units of 900 sq.ft. each (three, 30′×30′ squares for each field) and the corresponding percentage error. It can be seen that the new technique has estimated the number of plants with a precision quite comparable with the same in the case of random squares, if we remember that in the former method the size of sample was one unit of about 300 sq. ft. per field while in the case of random squares it was 3 units each of 900 sq. ft. per field.

TABLE 2. MEAN NUMBER OF PLANTS PER ACRE ESTIMATED THROUGH (a) LAHIRI SAMPLES (1) OF APPROXIMATELY 300 SQ. FT. PER FIELD) AND (b) RANDOM SQUARES (3, EACH OF 900 SQ. FT. PER FIELD)

stratum	number of sample fields	number of plants per acre estimated on the basis of			
		Lahiri samples		random squares	
		mean number of plants per acre	percentage error	mean number of plants per acre	percentage error
(1)	(2)	(3)	(4)	(5)	(6)
Northern	42	5485	5.9	5495	9.3
Eastern	42	6175	7.0	5934	4.8
Southern	42	6356	5.6	6667	1.6
Western	42	6193	2.2	6298	2.1
all strata	168	6052	2.9	6098	3.0

As mentioned earlier the estimation in this method of sampling involves only an inflation of the sample value by 2^n. It is desirable that each successive splitting divides the area into two equal parts. Any inequality in splitting is likely to increase the variability but the estimator is unbiased. The accuracy in splitting commensurates with cost and the question therefore is how accurately should these be done in such surveys. These two studies reveal that while for an area of highly irregular terrains like cinchona plantations, the method of splitting on the basis of an eye estimation of area would give estimates with errors within reasonable limits, for a patterned cultivation like tobacco crop the method would be even more satisfactory.

REFERENCES

SENGUPTA, J. M. and JACOB, PAUL (1968) : A technique for yield-estimation surveys in cinchona plantations. *Sankhyā*, Series B, **30**, 129–134.

——— (1968) : Some studies on the estimation of tabacco yield. *Sankhyā*, Series B, **30**, 429–440.

Paper received : November, 1967.

Revised : March, 1969.