SOME EXPERIMENTS WITH DIFFERENT TYPES OF AREA SAMPLING FOR WINTER PADDY IN GIRIDIH, BIHAR: 1945

By J. M. SENGUPTA
Indian Statistical Institute, Calcutta

[The Indian Statistical Institute has been using since 1937 the method of "area-sample" to estimate the area under particular crops. In the standard method developed by the Institute in the two States of Bengal and Bihar during the war, the operational "frame" consisted of two sets of maps: (a) one on the scale of 1 inch = 4 miles for *thanas* (administrative units of area about 150 square miles) on which were demarcated the boundaries of all villages falling within the *thana;* and (b) fairly large scale village maps (16 inches or 8 inches to the mile) on which the boundaries of each plot having a revenue survey number (and hence often called a "survey number") were shown in detail. The sample-units (called "grids") were of square shape and of various sizes from 1 acre to 36 acres. A suitable number of such sample-units (of a specified size) were allocated (in accordance with an appropriate sample-design) to administrative units (like *thanas*) consisting of a group of villages; and were later distributed among villages in a random manner by locating an entry-point at random on the '*thana*' map (the probability of an entry-point falling on any particular village being thus proportional to the area of the village itself). Also, for each entry-point falling on a village, one sample-unit was allocated to the village (a village not having any entry-point being excluded from the sample). In this way it was decided whether any particular village was included in the sample; and, if so, whether it had one grid or two grids or more as its share. For each village included within the sample, the next step was to mark, by a rubber stamp, each grid or sample-unit of the assigned size on the village map (taking care to fix the location of the grid or sample-unit at random on the map). The field investigators went to the villages included in the sample with the village maps (on which the grids or sample-units had been marked as explained above) and identified with the help of the map the plots or survey numbers falling within each grid or sample-unit; and then, by direct physical observation on the ground, recorded what crops were growing on each sample-plot. The proportion of the area under a particular crop within each grid was thus determined; and on the basis of these observed proportions the total area under the given crop was estimated in an appropriate manner.

In addition to the standard method described above, some early experiments had been made by selecting at random one single plot at a time, and some work had been also done with 'line-grids' in which a line of a given length was drawn in a random manner on the village map, and the field investigator marked on the map the position of the segments under each crop. From these observations, it is possible to calculate what 'proportion' of each line-grid (i.e., the linear sample-unit) was under each particular crop, and hence estimate the total area under that crop. J. M. Sengupta has been associated with the crop survey work from the beginning, and he has given in this paper an account of some of the experimental studies made under his leadership in 1945—*P. C. Mahalanobis*].

1. *The material and the object.* A special crop survey experiment was conducted in the year 1945 by the Indian Statistical Institute covering a compact area of 2717 acres, comprising ten villages situated near the Giridih branch of the Institute. The object was to study the relative efficiencies of different sampling units with variations in the method of enumeration for the estimation of acreage under winter paddy.

2. *Field enumeration.* Three different methods using two different types of sample-units were used in this experiment :—

(1) One set of square-shaped grids of sizes 2.25, 4.00, 6.25, 9.00 and 16.00 acres, were marked on cadastral survey maps in which the boundaries of individual

plots are shown.* The two different methods of enumeration employed in using such sampling units were as follows.

(a) the proportion of area occupied by paddy in each individual plot is estimated by physical observation of the crop on the ground and is the standard practice adopted by the Institute for crop surveys in Bengal;

(b) estimation of the proportion of (the total) area occupied by paddy directly for the grid as a whole, without making separate estimation of the crop-proportion in individual plots.

(2) "Line"-grids of different lengths 154, 204, 297, 418 and 528 yards representing the diagonals of the square grids used in method (1).

By physical observation, the actual position of each crop on the diagonal line on the map was marked with a pencil.

3. *Design of the experiments.* The field survey was conducted in three independent rounds, four workers being assigned to each method. After each round, the field records were collected, and each worker was allotted a fresh group of sample-units by rotation. Each worker was to use a different method in the successive rounds, all sample-units being thus independently enumerated by three different methods.

4. *Estimation of the mean proportion of acreage under winter paddy.*

Method 1(a): The proportion of area under paddy was calculated for each grid from the total acreage under paddy contained within the grid, as a sum of products of the size and crop-proportion in individual plots. For this purpose, it was necessary to measure the area of individual plots from the maps by graphical methods.

Method 1(b): Area measurement of individual plots was not necessary in this case, as the crop-proportions were directly estimated by the investigator for the grid as a whole.

Method (2): The intercepts made by individual crop patches on the diagonal line, i.e., the 'line'-grid, were measured correct to one-tenth of an inch on the cadastral maps with the help of a footscale. The total of such contributions to the full length of a 'line'-grid measured in inches, gave the proportion of crop area in individual sample units. The bias due to the borders of the field is assumed to be negligible, since the size of the grids used is very small compared with that of the total field sampled.

* The grids (or sample-units) were marked on the map in the form of a complete set of 5 "nested" units (one of each size) in the following way. A pair of random numbers (say, x and y) were picked up from a table of random numbers; and using the East-West and North-South directions as principal (x- and y-) axes, the entry-point was marked on the map with x and y as the two co-ordinates so that the entry-point was located strictly at random. The set of 5 grids (or sample-units) were then marked on the map using the same entry-point as the South-West corner point of each grid and with the two sides parallel to the East-West and North-South directions. The diagonals were then marked to supply the "line-grids".

5. *Mean proportion of area under paddy.* Mean proportions together with the coefficients of variation have been estimated and are shown in Table 1 below:

TABLE 1. COMPARISON OF DIFFERENT METHODS OF ESTIMATING THE MEAN PROPORTION OF AREA UNDER WINTER PADDY ($n = 80$)

	(1) square grids					(2) 'line'-grids (diagonals of square grids)			t-values	
size of grid		(a) plot by plot enumeration		(b) grid enumerated as a whole		length in yards	mean	c.v.	1(a)—1(b)	1(a)—2
sor.	acres	mean	c.v.	mean	c.v.					
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	2.25	0.2776	131	0.2730	137	154	0.2412	131	0.25	2.06*
2	4.00	0.2739	123	0.2608	136	204	0.2426	133	0.75	1.89
3	6.25	0.2835	114	0.2506	123	253	0.2321	120	2.07*	3.25**
4	9.00	0.2769	111	0.2503	114	297	0.2226	129	1.16	3.60**
5	16.00	0.2676	101	0.2548	110	418	0.2221	129	0.93	3.25**
6	25.00	—	—	—	—	528	0.2314	114	—	—

* Significant at 5% level. ** Significant at 1% level.

Values of '*t*' given in cols. (9) and (10) will show that Method 1(b) has not departed significantly from the Standard Method 1(a) except for the size 6.25 acres, while Method (2) shows significant differences from the Standard Method 1(a) almost for all the sizes. The source of this under-estimating bias has been subsequently explained.

There is also some evidence of the crop-proportions slightly increasing with decreasing sizes of sample-units in Method 1(b) as well as in Method (2). Further investigations are however necessary before this can be confirmed.

Investigators are therefore capable of giving consistent eye-estimates of the proportion under winter paddy for the grid as a whole, but probably under-estimates when the grid size is large.

6. *Coefficients of variation.* It may be noted that the coefficients of variation are roughly of the same order in all the three methods; and diminish, as expected with an increase in the size of the sampling unit. Methods 1(a) and 1(b) refer however to identical sample units while in 1(b) there is an additional element of error which enters in the process of eye-estim. .ng for the grid as a whole. The 'line'-grid in Method (2) however utilises much less information in the sense that it contacts only a fraction of total plots contained in the square grid. This fraction was found to be as low as 35.0% for grids of size 2.25 acres and 29.6% for grids of size 4.00 acres.

7. *Cost function.* A group of plots in a compact formation appears therefore to be relatively less efficient than having the same number of plots in a linear array. There is evidently a considerable amount of intraclass correlation between neighbouring

plots in a field. The 'line'-grid has thus a distinct advantage over a compact group of plots of square shape, so far as the variability is concerned. We have now to weigh the relative costs involved in each of these methods, before a final comparison can be made.

Table 2 gives (a) components of cost preparatory to field work, (b) enumeration cost in the field, (c) laboratory cost up to the stage of computing the crop proportions in individual sample-units. It will be seen that Method 1(b) involves no cost subsequent to field work, while the cost involved in area extraction in Method 1(a) is much larger compared to the time spent in measuring the "crop-length" in line-grids.

TABLE 2. COST OF OPERATIONS IN HOURS PER GRID

size in acres	(1) square grids								(2) line-grids (diagonals of square grids)				
	(a) plot by plot enumeration				(b) grid enumerated as a whole								
	pre-para-tory	field enume-ration	grid propor-tion	total	pre-para-tory	field enume-ration	grid propor-tion	total	length in yards	pre-para-tory	field enume-ration	grid propor-tion	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
2.25	0.112	0.127	0.242	0.481	0.021	0.213	—	0.233	154	0.045	0.078	0.150	0.272
4.00	0.166	0.203	0.387	0.757	0.021	0.332	—	0.352	204	0.049	0.141	0.198	0.387
6.25	0.294	0.381	0.726	1.401	0.021	0.452	—	0.472	263	0.057	0.204	0.246	0.506
9.00	0.385	0.508	0.968	1.861	0.021	0.576	—	0.597	297	0.061	0.263	0.289	0.613
16.00	0.658	0.890	1.694	3.241	0.021	0.727	—	0.748	418	0.073	0.325	0.406	0.804
25.00	—	—	—	—	0.021	—	—	—	528	0.077	0.403	0.513	0.992

Table 3 below gives the total cost of primary work in rupees per grid at Re. 1.00 per hour in the laboratory and at Re. 0.75 per grid in the field,[2] using the three different methods.

Table 3 also gives, in cols. (4), (7) and (11), the product of the variance and the cost or the relative cost per unit of information for each of these methods.[3] It will be seen that Method 1(b) comes out to be the best specially with smaller sizes, while Method (2) comes second in order, Method 1(a) being the costliest. In all the three methods, smaller sizes are found to be more efficient.

[2] These rates give fairly realistic *relative* values of field and laboratory costs in 1945.

[3] For a fuller discussion of the method see P. C. Mahalanobis : On large-scale sample surveys, *Phil. Trans Roy. Soc.*, 231, 329—451.

TABLE 3. COST IN RUPEES PER UNIT OF INFORMATION

size in acres	(1) square grids						(2) 'line'-grids (diagonals of square grids)			
	(a) plot by plot enumeration			(b) grid enumerated as a whole						
	vari-ance	cost in Rs.	(2)×(3)	vari-ance	cost in Rs.	(5)×(6)	length in yards	vari-ance	cost in Rs.	(9)×(10)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
2.25	0.1321	0.4192	0.0593	0.1390	0.1802	0.0250	154	0.1045	0.2528	0.0264
4.00	0.1135	0.7061	0.0801	0.1254	0.2695	0.0338	204	0.1014	0.3522	0.0368
6.25	0.1043	1.3658	0.1362	0.0944	0.3595	0.0339	253	0.0897	0.4551	0.0408
9.00	0.0948	1.7341	0.1644	0.0881	0.4528	0.0399	297	0.0832	0.5460	0.0454
16.00	0.0736	3.0250	0.2226	0.0782	0.5664	0.0443	418	0.0817	0.7225	0.0590
25.00	—	—	—	—	—	—	528	0.0701	0.8916	0.0625

8. *Special advantages of 'line'-grids over square grids or 'area'-units.* The under-estimating 'bias' noticed in 'line'-grids of Method (2), results obviously from ignoring the *ail* area, namely the foot-paths running around each crop plot.

In the usual area survey method, the proportion of area under a crop is enumerated with reference to the cadastral survey plots and thus takes account of the entire geographical area. Appropriate deductions must therefore be made for foot-paths which form a part of the cadastral plot. If, however, a foot-path or a corridor running between two plots happens to possess a cadastral survey number of its own (such as a road) then the question of its allocation to the adjoining crop-plots does not arise. True *ail* or border area thus represents non-cultivated foot-paths which were not separately numbered in a cadastral survey.

The square-grid survey therefore gives the gross crop-area inclusive of the plot-borders not coming under cultivation. The deduction in lieu of *ail* (plot-border) i.e., the proportion of geographical area under *ails* (plot-border) has been estimated as approximately 5% for West Bengal. In the locality where the special 'line'-survey experiments were conducted, the proportion of *ail* (plot-border) area has been found to be of the order of 10 or 12 per cent. This explains the discrepancy between the two estimates based respectively on "line-grids" and "square-grids."

In estimating the total outturn of the crop with square-grids or 'area' units this fact has to be kept in mind. Estimated yield rates must therefore refer to the 'gross' area cultivated, and not to the net area under crop so as not to over-estimate the outturn. It should be noted at this stage that, to obtain a valid estimate of the rate of yield per acre of the *gross* crop-area (inclusive of *ails* or plot-borders) it is necessary that the sample-units for crop-cutting experiments should be located with reference to lines passing through the mid-points of the ails (and not with reference to the physical boundaries of the crop-area itself). In other words each plot should

be considered as extending to the *mid-ail* lines (and not to terminate with the actual boundaries of the crop).

In the special crop-cutting experiments conducted by the Institute, the investigators were instructed to use the mid-ail lines as the axes of reference and to accept all sample-units falling within these boundaries, even if the area towards the mid-ail happens, as it will always be so, to be un-cropped. The results however strongly indicate that such instructions had never been strictly followed. Where the *ails* or foot-paths are of considerable height specially when they slope down gradually, it has been found that references have always been made to the boundary of the actual crop-area i.e., the line along the foot of the *ail*.

It may be explained here that the sample-unit for crop-cutting experiments (that is, the area within which the crop is harvested and weighed to supply an estimate of the rate of yield per area) is usually located in the field by investigator using two random numbers (x and y) and by going x-steps in one direction and then y-steps in a direction at right angles. It would be an improvement if the x and y co-ordinates are measured on the field with a measuring pole graduated in feet instead of going by steps which vary considerably from one individual to another. This method would be more accurate and would be amenable to an objective check.

In a 'line'-survey which does not make use of cadastral survey plots, it is not possible to credit the "true" ail area to the crop-plot to which it properly belongs, because, whether this is a numbered plot or not will not be known. The line-sample estimates therefore do not correspond to the gross crop acreage obtained from an 'area'-grid sample.

The line-sample simply estimates the net crop-area excluding footpaths or any other category of non-cropped area. Thus if a 'line' crosses an *ail* or falls alongside on it, the portion intercepted by this non-cropped land will be automatically allocated to the non-crop category.

The 'line'-sample has therefore several advantages. It avoids (i) the use of maps (except for purposes of location); and (ii) the area measurement of cadastral plots; and (iii) it gives directly an estimate of the *net* crop acreage, not requiring any deduction in lieu of *ails*.

Paper received : February, 1953.