ON A BIAS IN A CROP-CUTTING EXPERIMENT (APPLICATION OF INTEGRAL GEOMETRY TO AREAL SAMPLING PROBLEMS—PART V)

By M. MASUYAMA
Meteorological Research Institute, Tokyo, and
Indian Statistical Institute, Calcutta
and
J. M. SENGUPTA
Indian Statistical Institute, Calcutta

1. In this note we have made use of the data on crop-cutting experiments on winter paddy in Giridih, Bihar (India). Our aim is to point out the bias which arises if we neglect the condition of unbiasedness as is usually done in a sample survey of crop yield, of timber volume, etc. (Grosenbaugh, 1952; Smith and Guttman, 1953; Sengupta, 1954). The condition of unbiasedness which was introduced by one of the present authors is that each movable figure which has at least one point in common with any fixed area in a field T should be contained in T (Masuyama, 1953*a*). Otherwise, we should introduce an extended area T' which contains T and where the condition of unbiasedness holds. Usually we use the minimum extended area which satisfies this condition (Masuyama, 1953*c*). We used a circle as a movable oval. The experiment of which the data has been utilised in this note is as follows.

2. Seven paddy fields were chosen and in each, random points were located with reference to two sets of rectangular coordinates due North-South and East-West. At each point a system of concentric circles with radii 1′, 2′, 4′ and 5′8″ were described with the help of the ISI crop-cutting instrument. A count of the total plants completely inside the circle of 1′ radius and those on the border was taken and so also for three annular spaces between 1′ and 2′, 2′ and 4′, 4′ and 5′8″ and the three borders on het circumference of 2′, 4′ and 5′8″.

In addition to these counts, measurements were also taken of the girths (at the ground level) of the two plants nearest to the centres of the circles.

The following table shows the observed values in cols. (2), (4), and (6) and the estimated values in cols. (3), (5) and (7):

TABLE 1. OBSERVED AND ESTIMATED NUMBERS OF PLANTS ON THE BASIS OF 103 CONCENTRIC CIRCLES OF RADII 1′,2′ AND 5′8″

radius of circle (r)	average number of plants per circle					
	fully within (z)		on the border ($2lr/t$)		total (p)	
	obs.	est.	obs.	est.	obs.	est.
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1′	5.82	6.05	1.35	1.42	7.17	7.47
2′	24.70	24.04	2.87	2.85	27.57	27.49
4′	100.59	100.41	5.97	5.70	106.56	106.11
5′8″	202.67	202.86	8.04	8.07	210.71	210.93

3. Our fundamental equation in this case is under the condition of unbiasedness (Masuyama, 1954)

$$pt = f + lr + \pi r^2, \qquad p = z + 2lr/t, \qquad \dots \quad (3.1)$$

in our notation, where

f: the mean area of plants defined by Masuyama (1954), viz. that of the convex closure of the cross sections of tillers in a bunch at the ground level,

l: the mean perimeter of plants, and

t: the inverse of the mean number of plants per unit square (sq. ft.) (Masuyama 1953*a*, 1953*b*).

4. From the fourth column of the Table 1 we get four estimates of l/t, viz. 0.6750, 0.7175, 0.7403 and 0.7094. Taking the mean of these estimates we put

$$l = 0.7120\, t \qquad \dots \quad (4.1)$$

and accordingly we have from (3.1)

$$(p - 0.7120r)t = f + \pi r^2. \qquad \dots \quad (4.2)$$

As p and r are given in the Table 1, we can get the estimates of t and f by the method of least squares, viz.,

$$1/t = 2.048 \text{ and } f = 0.1621. \qquad \dots \quad (4.3)$$

Thus the mean number of plants per sq. ft. is estimated to be nearly 2.0 in this field.

The mean perimeter is from (4.1)

$$l = 0.3477. \qquad \dots \quad (4.4)$$

5. Now from (3.1) we have (Masuyama, 1954)

$$t(z + 2lr/t) = f + lr + \pi r^2, \qquad \dots \quad (5.1)$$

or

$$z = (f - lr + \pi r^2)/t. \qquad \dots \quad (5.2)$$

Inserting the numerical values (4.3) and (4.4) into (5.2) we get

$$z = 2.048(0.1621 - 0.3477r + \pi r^2),$$

from which the figures in the third column of the Table 1 are computed. The figures in the fifth column of the same table are obtained by (4.1).

The agreement between the theory and the observation is satisfactory, if we take into consideration that the coefficients of variation of p in the Table 1 are nearly 5·6%.

6. In the above case we cannot find the internal discrepancy. However, the estimates of l by the direct measurement for the six density classes, i.e. for the six different $(1/l)$, are smaller than the estimates by the indirect or integral geometrical method described in § 1, except the fifth class.

TABLE 2. ESTIMATES OF l FOR DIFFERENT DENSITY CLASSES

density class	estimate of l	
	direct	indirect
1	4.64	5.18
2	3.74	4.86
3	3.48	4.40
4	3.13	4.24
5	3.33	2.65
6	2.76	2.92

As our fundamental formula (3.1) holds even when the whole field is neither convex nor simply connected nor homogeneous in density of plant, it may be suspected that the condition of unbiasedness has been neglected. That is to say, according to our formula the mean number of plants on the boundary of circle is equal to

$$n = 2lr/t \qquad \dots \quad (6.1)$$

under the condition of unbiasedness. As there is practically no error in r in this experiment, the over-estimation of l means the over-estimation of nt, which arises if we neglect the condition of unbiasedness or the border effect. This bias is negligibly small only when the field is large enough in every direction compared with the double of the diameter of the circular cut, i.e. $4r$.

However, if we assume that the size of a field is 100×200 sq. ft., as was nearly so, and $r = 5' \, 8''$, then the minimum extended area will be

$$T' = 100 \times 200 + 2 \times 11.33(100+200) + 3.14 \times (11.33)^2 = 27,165,$$

whereas the actual size of the field is $T = 100 \times 200 = 20,000$, so that $1/t$ is underestimated by 35% (approximately), when we neglect the border effect.

In this crop-cutting experiment, the sample points were allotted at random uniformly inside the field T but not within the minimal extended area T'. Thus a higher probability of selection was allotted to the central plants.

If the points where some portion of the boundary of the circle lies outside the field are discarded, the situation is still worse.

To remove this bias, as one of the authors stressed in his paper (Masuyama, 1953*a*) and demonstrated experimentally in a seminar in the Indian Statistical Institute in January, 1953, we should use an extended area T' where the minimum distance between the border of T' and that of T should be greater than $2r$. The sample points should be selected at random uniformly in T'.

References

Grosenbaugh, L. R. (1952): Plotless timber estimates. *Journal of Forestry*, **50**, No. 1, 32-37.

Masuyama, M. (1953*a*): A rapid method of estimating basal area in timber survey (Part I). *Sankhyā*, **12**, 291-302.

——— (1953*b*): Analysis of the 1939 model sample survey results from the viewpoint of integral geometry (Part II). *Sankhyā*, **13**, 229-234.

——— (1953*c*): Application of integral geometry to areal sampling problems, (Part III). *Seminar Note*, Indian Statistical Institute.

——— (1954): A note on "Rapid methods of estimating the sum of specified areas in a field of given size." *Rep. Stat. Appl. Res.*, **3**, No. 2, 32.

Sengupta, J. M. (1954): Some experiments with different types of area sampling for winter paddy, in Giridih, Bihar: 1945. *Sankhyā*, **13**, 235-240.

Smith, C. S. and Guttman, L. (1953): Measurement of internal boundaries. *Journal of Metal*, 81-87.

Paper received : January, 1955.