STUDIES IN SAMPLING TECHNIQUE

VIII. Estimation of Incidence of Sugarcane Borers in a Field

BY R. C. ACHARYA, S. K. PRASAD, M. K. GHOSH, B. ACHARYA AND K. L. KHANNA

(*Central Sugarcane Research Station, Pusa*)

Received January 10, 1957

1. INTRODUCTION

TOP, stem and root—the three types of borers are the most damaging pests of sugarcane crop in this part of the country. The intensity of activity of these pests may be measured by percentage length of cane damaged or percentage nodes infested. Khanna and Banerjee[1] have worked out sampling technique for estimation of *length-basis* incidence in a field combined for all the three types of borers, taking incidence per clump as the ultimate variate. In this paper a smapling technique with sample sizes has been evolved for estimation of incidence of each type of borers, the ultimate variate being *node-basis* incidence per cane.

2. MATERIALS AND METHODS

The data of this investigation were collected at Pusa by the staff of Entomology Division of the station during the harvest season of 1951–52 crop. Four plots of sugarcane (each of 72′ × 18′ in size) marked out of fields, two under variety Co. 453 and two under B.O. 11, were completely enumerated in respect of total number of nodes of each cane and number of nodes affected by each type of borers, if any. The percentage incidence figures were calculated after Bartlett's correction and subsequently transformed on angular scale for statistical analysis.[2] Each of the plots had six rows 3′ apart and so the total row length was 432′ which was divided into units of 3′, 6′ and 9′. Each type of division has thrown the finite population of canes into a number of unequal clusters (clusters that contain unequal number of canes). As a measure of homogeneity the statistic δ as defined below has been calculated in respect of each type of clusters:

$$\delta = \frac{\sum_i M_i (\bar{Y}_i - \bar{Y})^2 - \sum_i \sum_j (y_{ij} - \bar{Y})^2}{\sum_i (M_i - 1) \sum_j (y_{ij} - \bar{Y})^2}$$

where,

M_i = number of canes in the i-th cluster.

y_{ij} = j-th cane in the i-th cluster.

$\bar{Y}_i$ = mean of the i-th cluster, and

$\bar{Y}$ = general mean.

The homogeneity index δ will be close to + 1, if the between cluster variance accounts for a large part of the total variance and it will tend towards zero or sometimes negative (above the lower limit of it), if the between cluster variance accounts for a small part of the total variance. Clusters of moderately large size may be efficient sampling units if the δ-value is low positive (or negative) and less efficient when it is high positive. It is evident from Table I furnishing the values of δ in respect of the three types of clusters under examination, that for all types of clusters and borers as well the index is always very small indicating that there is heterogeneity within cluster which, in fact, is a strong support in favour of cluster sampling in this case. As the most moderate and suitable size for practical handling, the study has been confined to unequal clusters of 6′ row length.

TABLE I

Showing values of ' δ '

	Top borer			Stem borer			Root borer		
	3′ unit	6′ unit	9′ unit	3′ unit	6′ unit	9′ unit	3′ unit	6′ unit	9′ unit
Co 453 Plot I	0·0654	0·0546	0·0308	0·1238	0·1279	0·1092	0·0309	0·0990	0·037
Co 453 Plot II	0·0045	0·0343	0·0415	0·0208	0·0295	0·0147	0·0043	−0·0112	0·002
B.O. 11 Plot I	0·0577	0·0529	0·0435	0·0337	0·0518	0·0464	0·0182	0·0351	0·1511
B.O. 11 Plot II	0·0277	0·1088	0·0331	0·0119	0·0576	0·0503	−0·0276	−0·0056	0·0009

Efficiency of two sampling designs, namely, cluster sampling with sub-sampling, without and with stratification has been examined in this paper along with that of the following sample estimates:

(*a*) Mean of cluster means estimate $\bar{Y}_s$, given by,

$$\bar{Y}_s = \frac{1}{n} \sum^{n} \bar{Y}_i$$

and (*b*) Mean of cluster total estimate $\bar{Y}_s'$ given by

$$\bar{Y}_s' = \frac{1}{n\bar{M}} \sum^{n} M_i \bar{Y}_i$$

where,

$\bar{Y}_i$ is the sample mean of the *i*-th cluster,

M_i is the number of elements in the *i*-th cluster,

$\bar{M}$ is the mean number of elements per cluster, and

n is the number of clusters sampled.

The estimate in (*a*) is a biased but consistent estimate. Sukhatme[3] has shown that the magnitude of bias is small if the correlation coefficient between number of elements in the cluster and the cluster mean is small. Estimates of correlation coefficients have been presented in Table II.

TABLE II

Showing correlation coefficients between M_i and mean of clusters

Variety		Plot No.	Top	Stem	Root
Co. 453	..	I	− 0·1006	− 0·0230	0·1258
Co. 453	..	II	− 0·0036	− 0·0738	0·1697
B.O. 11	..	I	0·3545	− 0·3897*	− 0·0361
B.O. 11	..	II	− 0·0164	− 0·4021*	− 0·1168

* Indicates significance at 5% level.

An examination of the table will show that the coefficients are very small in about all the cases suggesting that this estimate, though biased, may be better than the unbiased estimate ($\bar{Y}_s'$) provided its sampling error is also comparatively less. For an examination of this point, thirty 6′-units were selected at random from the population of first stage units in a plot and from each of the selected units 4 random canes were selected. Also for examining efficiency of stratification with rows constituting the strata, 5 random 6′ units were selected from each row with 4 random canes from each unit. Mathematical expressions of estimates of variances for these three cases are given below (after Sukhatme[3]).

(*i*) Variance of the mean of cluster mean estimate:

$$\text{Est. V}(\bar{Y}_s) = \left(\frac{1}{n} - \frac{1}{N}\right) s_b^2 + \frac{1}{nN} \sum^{n} \left(\frac{1}{m_i} - \frac{1}{M_i}\right) s_i^2.$$

Where,

$$s_b^2 = \sum^{n} \frac{(\bar{Y}_i - \bar{Y}_s)^2}{n-1}$$

$$s_i^2 = \frac{1}{m_i - 1} \sum_j (Y_{ij} - \bar{Y}_i)^2$$

N = Total number of 1st stage units in the population.

n = Total number of 1st stage units in the sample.

M_i = Number of 2nd stage units in the *i*-th 1st stage unit in the population.

m_i = Number of 2nd stage units selected as sample from the *i*-th 1st stage unit.

Y_{ij} = *j*-th element in the *i*-th 1st stage unit.

$\bar{Y}_i$ = Sample mean for the *i*-th 1st stage unit.

(*ii*) Variance of the mean of cluster total estimate:

$$\text{Est. V}(\bar{Y}_s') = \left(\frac{1}{n} - \frac{1}{N}\right) s_b'^2 + \frac{1}{nN} \sum^{n} \frac{M_i^2}{\bar{M}^2} \left(\frac{1}{m_i} - \frac{1}{M_i}\right) s_i^2.$$

Where,

$$s_b'^2 = \frac{1}{n-1} \sum^{n} \left(\frac{M_i}{\bar{M}} \bar{Y}_i - \bar{Y}_s'\right)^2$$

$\bar{M}$ = average number of 2nd stage units per 1st stage unit and other symbols are same as in (*i*).

(*iii*) Variance of the unbiassed estimate in stratified two-stage sampling:

$$V(\bar{Y}_w)_s = \sum_t \lambda_t^2 \left\{\left(\frac{1}{n_t} - \frac{1}{N_t}\right) S'^2_{tb} + \frac{1}{n_t N_t} \sum^{n_t} \frac{M_{ti}^2}{\bar{M}^2} \times \left(\frac{1}{m_{ti}} - \frac{1}{M_{ti}}\right) S^2_{ti}\right\}$$

Where,

$$\bar{Y}_w = \sum_t \lambda_t \bar{Y}'_{ts}; \ \lambda_t = \frac{M_{t0}}{M_0}$$

$$= \frac{\text{Total number of 2nd stage units in the } t\text{-th stratum}}{\text{Total number of 2nd stage units in population}}$$

$\bar{Y}'_{ts}$ = Sample mean of t-th stratum

$$= \frac{1}{\bar{M}_t n_t} \sum^{n_t} \frac{M_{ti}}{m_{ti}} \sum^{m_{ti}} Y_{tij}$$

Estimates of S'^2_{tb} and S^2_{ti} are provided by,

$$\text{Est. } S'^2_{tb} = s'^2_{tb} - \frac{1}{n_t} \sum^{n_t} \frac{M^2_{ti}}{\bar{M}_t^2} \left(\frac{1}{m_{ti}} - \frac{1}{M_{ti}}\right) s^2_{ti}$$

and

$$s'^2_{tb} = \frac{1}{n_t - 1} \sum^{n_t} \frac{M_{ti}}{\bar{M}_t} (\bar{Y}_{ti} - \bar{Y}_w)^2$$

$$\text{Est. } S^2_{ti} = s^2_{ti} = \frac{1}{m_{ti} - 1} \sum^{m_{ti}} (Y_{ij} - \bar{Y}_{ti})^2.$$

M_{ti} = Number of 2nd stage units in the i-th 1st stage units of the t-th stratum.

m_{ti} = Number of 2nd stage selected units in the i-th 1st stage units of the t-th stratum.

Y_{tij} = j-th element in the i-th 1st stage unit of t-th stratum.

$\bar{Y}_{ti}$ = Sample mean of the i-th 1st stage unit of t-th stratum.

N_t = Number of 1st stage units in the t-th stratum.

n_t = Number of 1st stage units in the t-th stratum in sample.

$\bar{M}$ = Average number of 2nd stage units per 1st stage unit.

$\bar{M}_t$ = Average number of 2nd stage units per 1st stage in the t-th stratum.

The estimates of variances have been found out and standard errors expressed as percentages of respective means are presented in Table III.

TABLE III

Showing S.E. as percentage of mean

	Co. 453					
	Plot I			Plot II		
	$\bar{y}_s$	$\bar{y}_s'$	$(\bar{y}w)_s$	$\bar{y}_s$	$\bar{y}_s'$	$(\bar{y}w)_s$
Top ..	2·56	6·29	5·30	3·13	6·57	6·89
Stem ..	5·08	5·68	5·48	3·31	4·15	5·11
Root ..	3·54	5·93	6·25	3·36	6·64	7·07
			B.O. 11			
Top ..	3·56	6·85	7·12	2·87	5·86	5·63
Stem ..	4·11	5·76	5·88	3·31	4·00	3·76
Root ..	2·44	5·20	5·46	2·45	4·44	3·08

It is seen that in almost all the cases percentage errors of mean of cluster mean estimate ($\bar{y}_s$) are minimum suggesting that this estimate will be most suitable for all practical purposes. Relative efficiency of stratification is almost nil as will be evident from a scrutiny of Table IV furnishing these values.

TABLE IV

Showing per cent. efficiency of stratification

Variety	Plot No.	Top	Stem	Root
Co. 453 ..	I	141	107	90
Co. 453 ..	II	91	66	88
B.O. 11 ..	I	93	96	91
B.O. 11 ..	II	91	113	107

The probable sampling combinations of clusters × canes have been worked out for different values of m_i in respect of this estimate so that the error of estimation may not exceed 5% and 10% of the mean incidence of borers and are given in Table V. The figures in the parenthesis show the fraction of available canes to be sampled in order to get the estimated incidence at these levels of accuracy.

TABLE V

Showing hypothetical combination of canes × clusters

Variety Co. 453

Plot No. I

Total Number of canes = 1053

$\bar{M} = 14 \cdot 62$

	Error as 5% mean				Error as 10% mean			
	$m=2$	$m=3$	$m=4$	$m=5$	$m=2$	$m=3$	$m=4$	$m=5$
Top ..	41 (7·79)	35 (9·97)	33 (12·54)	32 (15·19)	12 (2·28)	12 (3·42)	11 (4·18)	11 (5·22)
Stem ..	53 (10·07)	43 (12·25)	39 (14·81)	37 (17·57)	15 (2·85)	14 (3·99)	14 (5·32)	13 (6·17)
Root ..	31 (5·89)	26 (7·41)	24 (9·12)	24 (11·40)	8 (1·52)	8 (2·28)	7 (2·65)	7 (3·32)

Co. 453

Plot No. II

Total number of canes = 871

$\bar{M} = 12 \cdot 10$

	$m=2$	$m=3$	$m=4$	$m=5$	$m=2$	$m=3$	$m=4$	$m=5$
Top ..	27 (6·20)	22 (7·58)	21 (9·64)	20 (11·48)	7 (1·61)	6 (2·06)	6 (2·76)	6 (3·44)
Stem ..	35 (8·04)	27 (9·30)	24 (11·02)	23 (13·20)	8 (1·84)	7 (2·41)	7 (3·21)	7 (4·02)
Root ..	26 (5·97)	23 (7·82)	22 (10·10)	21 (12·06)	7 (1·61)	7 (2·41)	7 (3·21)	7 (4·02)

TABLE V (*Contd.*)

B.O. 11

Plot No. I

Total Number of canes = 891

$\overline{M} = 12 \cdot 38$

	Error as 5% mean				Error as 10% mean			
	$m=2$	$m=3$	$m=4$	$m=5$	$m=2$	$m=3$	$m=4$	$m=5$
Top ..	32 (7·18)	27 (9·09)	25 (11·22)	24 (13·47)	8 (1·80)	8 (2·69)	8 (3·59)	7 (3·93)
Stem ..	40 (8·98)	33 (11·11)	30 (13·47)	28 (15·71)	10 (2·24)	10 (3·36)	9 (4·04)	9 (5·05)
Root ..	23 (5·16)	18 (6·06)	16 (7·18)	15 (8·42)	5 (1·12)	5 (1·68)	4 (1·80)	4 (2·24)

B.O. 11

Plot No. II

Total Number of canes = 814

$\overline{M} = 11 \cdot 31$

Top ..	20 (4·91)	17 (6·27)	16 (7·86)	15 (9·21)	5 (1·23)	5 (1·84)	5 (2·46)	5 (3·07)
Stem ..	33 (8·11)	26 (9·58)	23 (11·30)	22 (13·51)	7 (1·72)	7 (2·58)	7 (3·44)	7 (4·30)
Root ..	25 (6·14)	20 (7·37)	18 (8·85)	17 (10·44)	5 (1·23)	5 (1·84)	5 (2·46)	5 (3·07)

3. Discussion of Results

Clusters of 3′, 6′ and 9′ row lengths have been found to be equally heterogeneous indicating suitability of cluster sampling in this case. Correlation coefficients between the number of canes in a cluster and mean value of the cluster are small and non-significant in all the cases except for stem borer incidence in both the plots under B.O. 11. The mean incidences of stem borer in these plots were also fairly high which suggest that for high levels of infestation, the unbiassed estimate, $\bar{y}_s'$, must invariably be taken whereas for lower levels, the biased estimate $\bar{y}_s$ (arithmetic mean of cluster means) with minimum variance may be adopted, the magnitude of bias being very small. Stratification with rows as strata did not generally bring any improvement in the efficiency of the estimates. Sample sizes are highest for stem borer in all the plots pointing to heterogeneity in the distribution of incidence of this borer which had its effect on the variance function. Sampling fractions are minimum for root borer incidence. From the table furnishing sampling combinations it will be seen that a sample of 2 canes per cluster gives the minimum sample size in all the cases, thus providing optimum allocation within a cluster. It will be further seen that with this allocation a sample size not exceeding 10% of the available canes in the plot is needed for an estimate of incidence of all types of borers within an error margin of 5%.

4. Summary of Results

(*i*) Efficiency of some sampling designs has been examined for estimation of incidence of sugarcane borers (Joint basis) in an experimental field.

(*ii*) Cluster sampling (6′ row length forming a cluster) with sub-sampling of canes in the cluster has been found to be the most suitable and efficient for the purpose.

(*iii*) Relative efficiency of stratification with rows as strata is nil in such cases.

(*iv*) The ordinary mean of cluster means, though biased, has the minimum variance and may be adopted only for lower levels of infestation, bias being negligible in these cases.

(*v*) A sample of 2 canes per cluster provides the optimum allocation and percentage total crop to be sampled with this allocation is about 10% for an estimate of incidence of all types of borers within an error margin of 5%.

5. Acknowledgement

The work was carried out as part of the Sugarcane Research Scheme being financed jointly by the State Government of Bihar and the Indian Central Sugarcane Committee to whom grateful thanks are due.

6. References

1. Khanna, K. L. and Banerjee, K. S. "Studies in Sampling Technique. IV. Estimation of borer incidence in Sugarcane," *Proc. Ind. Acad. Sci.*, 1945, **32**, 6.
2. Chandrasekhar, C. .. "A note on inverse Sine Transformation," *Sankhya*, **6**.
3. Sukhatme, P. V. .. *Sampling Survey Methods.*