COMPARISON OF ESTIMATORS OF NUMBER OF CASHEW TREES

By

S.B. Agarwal, B.B.P.S. Goel* and S.S. Pillai*

National Dairy Research Institute, Karnal

(Received : May, 1978)

India is the largest producer of cashewnut, the most useful product of the Cashew tree, in the world. The cashew kernel (Kaju) inside the nut is very palatable, highly nutritive and has a pleasing flavour. It makes a significant contribution to foreign trade and is therefore an important earner of foreign exchange like tea, jute and coffee etc. Not only cashew kernel but all parts of the cashew tree are of considerable economic importance. On account of its high economic value cashew cultivation has received considerable attention in the recent years. Efforts to increase the area under cashew cultivation as also its production have continued. However, such efforts have been hampered considerably for want of reliable statistics on number of trees grown, the proportion of bearing and non-bearing trees, distribution of trees according to age group, average yield per tree/acre etc. For formulating suitable plans for development of cashew cultivation detailed and reliable information on the above aspects is very essential. For developing suitable sampling techniques for collecting such data Indian Agricultural Statistics Research Institute (I.C.A.R.) conducted a pilot sample survey in Andhra Pradesh in 1966-68.

The present study attempts to devise a suitable procedure for estimation of total number of cashew trees by comparing the various available estimators. Further, the gain in efficiency due to pps (probability proportional to size) sampling has also been estimated.

Sampling design

The sampling design adopted for the survey was one of stratified random sampling. Seventeen important cashew growing taluks spread

* Indian Agricultural Statistics Research Institute, New Delhi-12.

over five districts of the state, *viz.*, Srikakulam, Visakhapatnam, East Godavari, Guntur and Nellore which together account for about 90 per cent of the area under cashew in Andhra Pradesh were grouped into five strata on the basis of geographical contiguity. The sample of villages was allocated to strata in proportion to the area under cashewnut in each stratum. From each stratum a specified number of villages was selected with probability proportional to area under cashewnut and with replacement. In each of these villages all cashew trees were completely enumerated. The total sample size was 54 villages selected from a population of 363 villages in all the strata.

The distribution of area under cashew and number of villages in different strata are given in Table 1.

TABLE 1

Distribution of area under cashewnut and number of villages

Stratum	*Area under cashew nut (acres)*	*Number of villages growing cashewnut*	*Number of villages selected*
1	11551	103	15
2	1394	65	3
3	3346	111	9
4	4412	45	21
5	2447	39	6
Overall	23150	363	54

Notation :

Let,

N_h =Total number of villages in the h th stratum.

n_h =Total number of villages selected from the h th stratum.

y_{hi} =Number of cashew trees in the i th selected village from the h th stratum.

x_{1hi} =Geographical area (first auxiliary character) of the i th village in the h th stratum.

x_{2hi} =Garden area (second auxiliary character) of the i th village in the h th stratum.

p_{hi} =Probability of selection of i th unit in the h th stratum,

$$\sum_{i=1}^{N_h} p_{hi}=1$$

p_{hi} $=A_{hi}/A_h$

where A_{hi} is the area reported to be under cashew in the i th village of the h th stratum and

$$A_h=\sum_{i=1}^{N_h} A_{hi}$$

Estimation procedures

1. Simple estimate

An unbiased estimate of total number of trees in h th stratum is given by

$$\hat{Y}_{h\ (pps)}=\frac{1}{n_h} \sum_{i}^{n_h} y_{hi}/p_{hi}$$

and an estimate of its variance by

$$\hat{V}(\hat{Y}_{h\ (pps)})=\frac{1}{n_h\ (n_h-1)}\left[\sum_{i}^{n_h} (y_{hi}/p_{hi})^2-\left(\sum_{i}^{n_h}\frac{y_{hi}}{p_{hi}}\right)^2/n_h\right]$$

2. Ratio estimate

Let $u_{hi}=\dfrac{x_{1hi}}{p_{hi}}$, $v_{hi}=\dfrac{x_{2hi}}{p_{hi}}$ and $z_{hi}=\dfrac{y_{hi}}{p_{hi}}$

Then, the ratio estimate based on geographical area as the auxiliary variate is given by

$$\hat{Y}_{1h}=\frac{\sum_{i}^{n_h} y_{hi}/p_{hi}}{\sum_{i}^{n_h} x_{1hi}/p_{hi}} X_{1h}$$

where X_{1h} is the geographical area of h th stratum. An estimate of its variance is given by

$$\hat{V}(\hat{Y}_{1h}) = \sum_{h=1}^{k} \frac{1}{n_h(n_h-1)} \sum_{i}^{n_h} (z_{hi} - R_{nh} U_{hi})^2$$

where
$$R_{nh} = \frac{\frac{1}{n_h} \sum_{i}^{n_h} z_{hi}}{\frac{1}{n_h} \sum_{i}^{n_h} U_{hi}}$$

Similarly the ratio estimate based on garden area as auxiliary variate can be defined.

3. Combined ratio estimate

The combined ratio estimate based on geographical area is given by

$$(\hat{Y}_{1CR}) = \frac{\sum_{h=1}^{k} \frac{1}{n_h} \sum_{i}^{n_h} z_{hi}}{\sum_{h=1}^{k} \frac{1}{n_h} \sum_{i}^{n_h} u_{hi}} X_1$$

where X_1 is the total geographical area of the population. An estimate of its variance is given by

$$\hat{V}(\hat{Y}_{1CR}) = \sum_{h=1}^{k} \frac{1}{n_h} \left(s^2_{zh} + \hat{R}^2_{nh} \; s^2_{uh} - 2\hat{R}_{nh} \, s_{zh\,uh} \right)$$

$$\text{where } s^2_{zh} = \frac{\sum_{i=1}^{n_h} z^2_{hi} - \left(\sum_{i}^{n_h} z_{hi} \right)^2 / n_h}{(n_h - 1)} \text{ and}$$

s^2_{uh} and s_{zhuh} can be similarly defined. Similarly an estimate of number of cashew trees and that of its variance based on garden area can be defined.

4. Two-variate ratio estimate

Using geographical and garden areas as the two auxiliary variates, a two variate ratio estimate for the h^{th} stratum is given by,

$$(\hat{Y}_{MR})\, h = w_{1h}\hat{Y}_{1h} + w_{2h}\hat{Y}_{2h}$$

where w_{1h} and w_{2h} are the weights to be chosen so that $w_{1h}+w_{2h}=1$, and $\hat{Y}_{1h}$ and $\hat{Y}_{2h}$ are the ratio estimates based on x_1 and x_2. The minimum variance of the estimate is given by,

$$V_{Min}\,(\hat{Y}_{MR})\, h = \frac{1}{n_h}\,\hat{Y}^2{}_{h\,(pps)}\left(\frac{v_{11h}\, v_{22h} - v^2_{12h}}{v_{11h}+v_{22h}-2v_{12h}}\right)$$

where v_{11h}, v_{22h} and v_{12h} are the estimated variances and covariances of ratio estimates for h th stratum defined as,

$$V_{11h} = \frac{1}{n_h}\,\hat{Y}^2_{h\ (pps)}\left(\frac{s^2_{zh}}{\bar{z}^2_h} + \frac{s^2_{uh}}{\bar{u}^2_h} - \frac{2s_{zh\,uh}}{\bar{z}_h\,\bar{u}_h}\right)$$

$$V_{22h} = \frac{1}{n_h}\,\hat{Y}^2_{h\ (pps)}\left(\frac{s^2_{zh}}{\bar{z}^2_h} + \frac{s^2_{vh}}{\bar{v}^2_h} - \frac{2s_{zhvh}}{\bar{z}_h\,\bar{v}_h}\right)$$

$$V_{12h} = \frac{1}{n_h}\,\hat{Y}^2_{\ (pps)}\left(\frac{s^2_{zh}}{\bar{z}^2_h} + \frac{s_{uhvh}}{\bar{u}_h\bar{v}_h} - \frac{s_{zhuh}}{\bar{z}_h\,\bar{u}_h} - \frac{s_{zhvh}}{\bar{z}_h\,\bar{v}_h}\right)$$

Results and discussions

Various estimates *viz.* simple, separate ratio, combined ratio and two-variate ratio of the number of cashew trees as also their percentage standard errors were obtained and are presented in table 2.

It was found that the simple estimate was the most efficient (standard error 10.7 percent). Two-variate ratio estimate was more or less equally efficient with percentage standard error 11.1. The efficiency of combined ratio estimate based on either auxiliary variate was very low, the percentage standard errors being 32.9 (geographical

TABLE 2

Estimate of total number of Cashewnut trees in different strata

Stratum	*Simple Estimate*		*Ratio estimate based on geographical area*		*Ratio estimate based on garden area*		*Two-variate ratio method*			
							Estimate	*% S. E.*	*Weights*	
	Estimate	*% S.E.*	*Estimate*	*% S.E.*	*Estimate*	*% S. E.*			W_1	W_2
1	730,069	16.7	764,856	38.7	769,039	59.9	765,071	18.4	0.95	0.05
2	66,247	7.3	69,943	33.6	69,961	41.2	69,950	10.8	0.61	0.39
3	124,880	15.8	125,743	30.0	125,498	56.1	125,539	20.8	0.17	0.83
4	330,746	20.1	340,061	15.5	339,936	17.1	340,033	6.9	0.78	0.22
5	79,570	19.2	100,392	35.2	88,458	51.7	99,454	31.1	0.92	0.08
Overall	1,331,512	10.7	1,400,995	25.7	1,392,892	32.3	1,400,047	11.1		
Combined ratio estimate			1499007	32.9	1357840	39.8				

area) and 39.8 (garden area). The ratios $\bar{z}_h/\bar{u}_h$ and $\bar{z}_h/\bar{u}_h$ both for geographical and garden area varied widely from stratum to stratum and therefore the combined ratio estimate turned out to be of very low efficiency. These ratios for each stratum are presented in Table 3.

TABLE 3

Ratios for geographical and garden areas

Stratum	*Geographical area* $\bar{z}_h/\bar{u}_h$	*Garden area,* $\bar{z}_h/\bar{v}_h$
1	2.8179	4.8621
2	0.6195	9.3946
3	0.2844	2.0608
4	1.1636	8.0384
5	0.0794	15.1651

Ratios for geographical area ranged from 0.0794 for stratum five to 2.8197 for stratum one. Corresponding figures for garden area ranged from 2.0608 for stratum three to 15.1651 for stratum five.

It may be mentioned here that in stratified sampling, different sampling and estimation strategies can be adopted in different strata. Simple estimate in strata 1, 2, 3, and 5 and two-variate estimate in stratum 4 would have been better. Using these estimators in different strata estimates of total number of cashew trees and its standard errors were obtained. The estimate of total number of trees was 1,340,809, and its standard error 5.0 percent. Thus the estimate obtained in this way is much better than any of the estimates discussed in table 2. The study further demonstrates the advantage of using mixed sampling and estimation strategies in stratified sampling.

Gain in efficiency due to pps sampling

The percentage gain in efficiency due to pps sampling as compared to simple random sampling (srs) with replacement is given by,

$$\frac{\hat{v}(\hat{Y}_{h\,(srs)})-\hat{v}(\hat{Y}_{h\,(pps)})}{\hat{v}(\hat{Y}_{h(pps)})}\times 100$$

where $\hat{v}(\hat{Y}_{h(srs)}) = \frac{1}{n_h}\left(N_h \sum_{1}^{n_h} y^2_{hi}/p_{hi} - n_h \hat{Y}^2_{h(pps)} + \frac{1}{nh}\hat{v}(\hat{Y}_{h\,(pps)}\right)$

The gain in efficiency due to pps sampling for each stratum as compared to srs was worked out for the simple estimate and results are given in table 4.

TABLE 4

Gain (%) in efficiency due to pps sampling

Stratum	*Gain in efficiency* (%)
1	282
2	−434
3	1083
4	—101
5	539
Overall	202

It was observed that the gain in efficiency due to pps sampling was 202 percent. Table 4 would suggest that pps sampling in strata 1,3 and 5 and srs sampling in strata 2 and 4 would have been better.

SUMMARY

Utilising the data collected from a pilot sample survey conducted by Indian Agricultural Statistics Research Institute in important cashewnut growing areas of Andhra Pradesh in 1966-68, the following estimates of number of cashew trees in the area were obtained (*i*) simple estimate (*ii*) separate and combined ratio estimates based on either grographical or garden area as auxiliary variate (*iii*) two-variate ratio estimate using geographical and garden areas as two auxiliary variates.

It was found that the simple estimate was the most efficient. The two-variate ratio estimate was also more or less equally efficient. However, since the ratios $\bar{x}_h/\bar{u}_h$ and $\bar{z}_h/\bar{v}_h$ varied considerably from stratum to stratum. the combined ratio estimate, based on either auxiliary variate was much less efficient. Separate ratio estimate was also much less efficient then simple estimate, possibly on account of a low correlation between number of trees and the auxiliary characters. The gain in efficiency due to pps sampling overall the strata was 202 percent. It was found that pps sampling in strata 1,3 and 5 was more efficient than srs and converse was true in strata 2 and 4.

ACNOWLEDGEMENT

The authors are thankful to the referee for his suggestions in improving the paper.

REFERENCES

[1] Agarwal, S.B. (1969) : "Estimation procedure in Cashewnut Survey" dissertation submitted for award of *M.Sc. degree to P.G. School*, I.A.R.I., New Delhi.

[2] Cochran, W.G. (1963) : "Sampling Techniques" *John Wiley & Sons, Inc.* New York.

[3] Hansen. M.H. Hurwitz W.N. and Madow, W.G. : "Sample Survey Methods and Theory Vols. 1 & 2, *John Wiley & Sons*, New York.

[4] Hansen, M.H. and Hurwitz, W.N. (1943) : On the theory of sampling from finite population. AMS, 14, 333-362.

[5] Murty, M.N. (1967) : Sampling theory and methods.

[6] Olkin, I. (1958) : Multivariate ratio-estimation for finite pop. *Biometrika*, 45, 154-165.

[7] Sukhatme, P.V. and Sukhatme, B.V. (1970. : Sampling Theory of Surveys with Applications, *Indian Society of Agricultural Statistics.*