ESTIMATION OF VEGETABLE PRODUCTION USING PARTIAL HARVEST DATA

By
A.C. Kaistha and B.B.P.S. Goel
I.A.S.R.I, New Delhi
(Received : November, 1979)

Summary

Available sampling techniques fer estimation of production of vegetables involving large number of pickings untail collection of data for all pickings during the entire harvesting period. Besides introducing operational difficulties, this also increases the cost of field work. The double sampling and component sampling approaches discussed in this paper not only overcome these draw backs but also increase the efficiency of the estimates obtained. In double sampliing approach a large sample of fields is first sdlected for recording yield data for one packing/short interval and then a sub-sample of fields is selected for recording complete data. Using these data a double sampling estimate is built up. In component sampling approach in dependent samples of fields are selected and yield for only one picking/inteival is recorded for cach of them. Estimates of average yield for each picking/interval are pooled to obtain estimate of yield overall picking/intervals. Utilizing the data for bean crop from the survey conducted by IASRI in Bangalore district (I971-72), sample sizes of large and sub-samples in double sampling and of independent samples in component sampling approaches have been determined for a given level of precision. The efficiency of these approahces has also been compared with ıbat of simple random sampling.

Introduction

Keeping in view the peculiar nature of vegetable cultivation, Indian Agricultural Statistics Research Institute conducted a few pilot sample surveys for developing a suitable sampling technique for estimating the area under vegetables. their average yield per hectare, total production and studying their cultivation practices. In the pilot sample surveys conducted by IASRI, the estimates of yield were obtained on the basis of crop cutting experiments conducted in a random sample of fields. The data on yield were recorded in respect of all pickings done on such fields. The number of pickings varied from 2 to 20 and the harvesting was spread over a period of 2 to 3 months.

The methodology developed by IASRI (1), (6) and (8) necessitates the availability of whole time field staff during the entire harvesting period for collection of yield data. This results not only in operational difficulties but also increases exhorbitantly the cost of field work. In the present paper an attempt has been made to examine the possibility of sampling the harvestings/harvesting period appropriately for collecting data on yield so that collection of yield data becomes easy and corresponding cost is reduced considerably.

The data collected from the surveys conducted by the IASRI showed that the correlation coefficients between the total yield and yield of individual pickings and also between the total yield and yield of time intervals (as fractions of the total harvesting period) in the selected fields were not only significant but also quite high inmost of the cases. The figures for bean crop are presented in table I.

TABLE 1.

Correlation Coefficient between total yield and Picking wise yield for Bean Crop in 1971-72 in Bangalore District

S. No. of picking	*Season II Jan-March*	*Season III April-June*	*Season IV July-Sept.*
1.	.77	.46	.66
2.	.89	.62	.93
3.	.79	,79	.87
4.	.76	.87	.71
5.	.44	.73	.42

In view of high Correlation Coefficients Double Sampling and Component sampling approaches have been studied for estimating the yield of vegetables.

Double sampling approach

In this approach a large sample of n'' fields is selected for recording the yield in respect of one picking or one time interval during the harvesting period and a sub-sample of n' fields is then retained for recording the yield in respect of all pickings/entire harvesting period. Using the regression coefficient between the total yield and yield in respect of a particular picking/time interval for the n' fields a double sampling estimate of yield is given by

$$\bar{y}_{ld} = \bar{y}_{n'} + \hat{\beta}\,(\bar{y}_{n''} - \bar{y}_{n'}) \qquad \ldots(1.1)$$

where $\hat{\beta} = \dfrac{s_{y^p,\,y}}{s^2_{y^p}}$

y_i^p = is the yield/hectare of *pth* picking for ith field

$\bar{y}^p_{n''}$ = Average yield of *pth* packing based on

n'' fields $= \dfrac{1}{n''} \sum_i^{n''} y_i^p$

$\bar{y}^p_{n'}$ = Average yield of *pth* packing based on n' fields

$$= \frac{1}{n'} \sum_i^{n'} y_i^p$$

$$\bar{y}_{n'} = \sum_{p=1}^{k} \bar{y}^p_{n'}$$

$$s_{y^p,\,y} = \frac{1}{n'-1} \sum_i^{n'} (y_i^p - \bar{y}^p_{n'})\,(y_i - \bar{y}_{n'}) \text{ and}$$

$$s^2_{y^p} = \frac{1}{n'-1} \sum_i^{n'} (y_i^p - \bar{y}^p_{n'})^2$$

Following Sukhatme and Sukhatme (1970) and assuming (*i*) the cost function to be $C_o = n''c + (k-1)\,cn'$ ....(1.2) and (*ii*) that both the samples are selected with equal probability without replacement, where c is the cost of recording yield per experiment per picking, k is the total number of pickings and C_o is the total cost, the optimum variance of estimate (1.1) will reduce to

$$V(\bar{y}_{ld})_{opt} = \frac{c}{C_o} S_y^2 \left(\sqrt{(k-1)(1-\rho^2)} + \rho\right)^2$$

ρ being the correlation coefficient between the total yield and yield of *pth* picking.

Similarly for time intervals of the total harvesting period of the corresponding double sampling estimate of yield is given by

$$\bar{y}_{ld} = \bar{y}_{n}' + \hat{\beta}(\bar{y}^{q}_{n'} - \bar{y}^{q}_{n'}) \qquad \ldots(1.3)$$

where $\hat{\beta}'$, $\bar{y}^{q}_{n}$ and $\bar{y}^{q}_{n''}$ are defined in a way similar to that for $\hat{\beta}$, $\bar{y}^{p}_{n'}$ and $\bar{y}^{p}_{n''}$ for *qth* harvesting interval. The optimum variance of estimate (1.3) is given by

$$V(\bar{y}_{ld})_{opt} = \frac{c'}{C_o} S^2_y \left(\sqrt{(t-1)(1-\rho'^2)} + \rho'\right)^2$$

where ρ' is the correlation coefficient between yield of *qth* interval and total yield. t is the number of time intervals in the harvesting period and S^2_y is given by

$$S^2{}_y = \sum_{i=1}^{N} (y_i - \bar{y}_N)^2 \Big/ _{N-1}$$

where y_i is the total yield in the ith field and $\bar{y}_N = \frac{1}{N}\sum_{i=1}^{N} y_i$

Efficiency of double sampling

Efficiency of double sampling estimate can be compared with usual estimate based on the data of complete harvesting. The estimate of yield based on complete harvesting from a sample of n fields is given by

$$\bar{y}_n = \sum_{p=1}^{k} \bar{y}^{p}_{n} \text{ where } \bar{y}^{p}_{n} = \frac{1}{n}\sum_{i}^{n} y^{p}_{i}$$

Ignoring finite population corrclation factor

$$V(y_n)_{opt} = \frac{S^2 y}{n} \text{ where } n = \frac{C_o}{kc}$$

The relative efficiency of double sampling for a given cost C_0 as compared to usual procedure is given by

$$R.E(\bar{y}_{ld} \text{ over } \bar{y}_n) = \frac{1}{\sqrt{1-\rho^2} + \rho\sqrt{\frac{1}{k-1}}} \text{ for pickings}$$

$$R.E.(\bar{y}'_{ld} \text{ over } \bar{y}_n) = \frac{1}{\sqrt{1-\rho'^2}+\rho'\sqrt{\frac{1}{t-1}}} \text{ for intervals.}$$

Efficiency (percentage) of double sampling relative to usual procedure of Simple Random Sampling for various values of k and t have been worked out and are given in table II.

TABLE II

Efficient (%) of double sampling relative to simple random sampling

ρ or ρ' \ k or t	6	7	8	9
65	91	95	99	102
79	95	100	104	108
75	101	107	112	116
80	109	116	123	128
85	122	131	139	146

In the above table the efficiency of double sampling has been presented for values of k (or t) between 6 and 9. From the data of the pilot sample surveys conducted by the IASRI it was seen that the total number of pickings for most of the vegetable crops was not less than 6. However, for crops like tomato the maximum number of pickings was as large as 20. But as the number of pickings increase beyond 9 the efficiency of double sampling will increase further and the above table can be extended.

Percentage reduction in cost

Alternatively one can work out the percentage reduction in cost using double sampling estimate over the simple estimate for a given level of precision. Let n be the equivalent size of the sample required for recording data of all pickings for a given level of precision. Then it can be easily shown that the percentage reduction in cost using double sampling is given by

$$\left[\left(\frac{n-n'}{n}-\frac{n''-n}{n\times k}\right)\right]\times 100.$$

These percentages were worked out utilizing the data of bean crop for 2nd, 3rd and 4th seasons of first year for the survey conducted in Bangalore Distt. (1971—74) for various levels of precision and are presented in table III.

TABLE III

Percentage reduction in cost for a specified precision of the estimate of average yield of beans using double sampling

Season	*% S.E.*	*Percentage reduction in cost choosing for the large sample*		
		2nd picking	*3rd picking*	*4th picking*
2nd (Jan.—March)	3	39	22	18
	7	39	23	18
	10	41	24	20
3rd (April—June)	3	31	25	39
	7	32	26	40
	10	31	24	38
4th (July—Sept.)	3	52	27	15
	7	52	27	15
	10	50	24	14

It is seen from Tables II and III above that depending upon the values of correlation coefficient and the total number of pickings considerable gain in efficiency or alternatively reduction in cost can be affected in the estimation of yield of vegetables by double sampling method.

Component Sampling

Let n_1, n_2..., n_k be k independent samples of fields. Data in respect of 1st picking are obtained from n_1 fields only, data in respect of 2nd picking from n_2 fields only and so on for the *kth* picking from a sample of n_k fields only. Assuming that all the samples are drawn with equal probability and without replacement the following estimate of yield can be obtained. The average yield of *pth* picking based on n_p fields is given by

$$\bar{y}_{n_p} = \frac{1}{n} \sum_{i}^{n} y_i^p$$

The over all average yield is then given by

$$\sum_{p=1}^{\bar{y}_k} n_p = \sum_{p=1}^{k} \bar{y}_{n_p}$$

its variance is given by

$$V\left(\bar{y}_{k \atop \sum\limits_{p=1} n_p}\right) = \sum_{p=1}^{k} V(\bar{y}_{n_p})$$

where optimum values of n_1, $n_2 \ldots \ldots$, n_k can be obtained by minimising the above variancc for a fixed cost. Let the cost function be

$$C_0 = c \sum_{p=1}^{k} n_p \qquad \ldots(1.4)$$

It can be shown that the optimum values of n_1, $n_2 \ldots \ldots$, n_k are given by

$$n_p = \frac{s_p}{\sum\limits_{p=1}^{k} s_p} \times n \times k$$

where $$s_p^2 = \sum_{i=1}^{N} \frac{(y_i^p - \bar{y}_N)^2}{N-1}$$, N being the total number of fields in the population.

Consequently the optimum variance of the estimate is given by

$$V\left(\bar{y}_{k \atop \sum\limits_{p=1} n_p}\right) \text{opt} = \frac{(\Sigma s_p)^2}{n \times k}$$

The estimate of average yield and its variance can be obtained similarly in case of sampling of harvesting intervals.

Efficiency of Component Sampling

It can be shown that the efficiency of component sampling estimate relative to simplee stimate based on the data of complete harvesting for a fixed cost is given by

$$\frac{V(\bar{y}_n)}{V\left(\bar{y}_{k \atop \sum\limits_{p=1} n_p}\right) \text{opt}} = \frac{S^2 \times k}{\left(\sum\limits_{p=1}^{k} s_p\right)^2}$$

The efficiency of component sampling estimate of yield of bean crop relative to *srs* estimate from the data used for the present study was estimated to be around 350 percent for seasons 2, 3 and 4 respectively.

Percentage Reduction in Cost

Alternatively one can work out the percentage reduction in cost by the use of component sampling for a specified precision. Let for a given precision (Stantard error α%) $n_1, n_2 \ldots n_k$ be the sizes of the component samples of fields and n be the equivalent size of the sample of fields using the data of all pickings. Then it can be shown that the percentage reduction in cost is given by

$$\left(1 - \frac{\sum_{p=1}^{k} n_p}{n \times k} \right) \times 100$$

Again using the data of bean crop the percentage reduction in cost for a specified precision for different seasons and for the different levels of precisions were worked out and are given in table IV.

TABLE IV

Percentage reduction in cost for a specified precision of the estimate of yield of beans using component sampling

Season	*% Standard error* (α)	*% reduction in cost*
2nd (Jan.—March)	5	72
	7	72
	10	73
3rd (April—June)	5	71
	7	71
	10	71
4th (July–Sept.)	5	71
	7	71
	10	71

It is seen that considerable gain in efficiency or reduction in cost can be achieved by using the component sampling approach for estimating the yield of vegetables.

The simple cost functions at 1.2 and 1.4 on pages 114 and 115 have been adopted for the sake of simplicity. In actual practice, however, a small part of the total cost is also incurred on identifying and locating the units and to that extent these cost functions need to be modified. Such modification will not, however, materially disturb the findings of the study. These studies have another important implication in that the field work according to the suggested procedures can be entrusted to the existing field agencies such as VLWs Patwaris or Agricultural Assistants etc. whereas in the usual procedure based on the data of complete harvesting, it is necessary to deploy whole time staff for collecting the basic data. This will further cut down the cost of field work.

References

Annonymous (1973) : 'Pilot Sample Survey to study yield and cultivation practices of vegetables crops in Nasik and Pune districts of Maharashtra States. Research Report—I.A.R.S. New Delhi.

Goel, B.B.P.S. and Singh D. (1970) : 'A Note on component sampling' (Abstract) : *J.I.S.A.S.* Vol. XXIV No. 1.

Goel, B.B.P.S., Raheja S.K. Malhotra, P.C. and Rastogi, V.S. (1978) : 'Methodology for estimating the yield of Cotton base on Partial Harvest Data' (Abstract) *J.I.S.A.S.* Vol. XXVII No. 1.

Kaistha, A.C. (1976) : An applicatiou of Principle Component Analysis estimating total Vegetable Products (Abstract) *J I.S.A.S.* Vol. XXVII No. 2.

Kaistha A.C. (1978) : 'Some Investigations for Estimation of Vegetable Production Using Partial Harvest Data' Diploma Thesis IASRI (Unpublished)

Singh D, Srivastava, A.K. Singh, Padam and Satyal Pal (1976) : *Survey on Vegetables in Bangalore.* An IASRI Publication.

Sukhatm, B.V and Sukhatme P.V. (1970) : *Sampling Theory of Surveys with application* Iowa State University Press.

Sukhatme, B.V., Manwani, A.H. and Bapat S.R. (1973) : Survey on Vegetables in Rural areas of Delhi An I.C.A.R. Publication.