Plant Archives Vol. 24, No. 1, 2024 pp. 1645-1650 e-ISSN:2581-6063 (online), ISSN:0972-5210

Plant Archives

Journal homepage: http://www.plantarchives.org

DOI Url : https://doi.org/10.51470/PLANTARCHIVES.2024.v24.no.1.229

ESTIMATION OF COTTON YIELD USING TWO-PHASE SAMPLING APPROACH

Nitin Varshney[1*], Yogesh Garde[1], Alok Shrivastava[1], Vipul Shinde[2] and Vishal Thorat[3]

[1]Department of Agricultural Statistics, N.M. College of Agriculture, N.A.U., Navsari, Gujarat, India.
[2]Department of Agricultural Engineering, N.M. College of Agriculture, N.A.U., Navsari, Gujarat, India.
[3]Department of Agribusiness Economics and Finance, A.A.B.M.I., N.A.U., Navsari, Gujarat, India.
*Corresponding author E-mail : nitin.caw@nau.in

(Date of Receiving-01-01-2024; Date of Acceptance-12-03-2024)

ABSTRACT

Cotton, a multiple picking crop, is grown in around nine States in India. The existing procedure of estimation of average yield of cotton is based on crop cutting experiment (CCE) approach, which utilizes data on all pickings, is cumbersome and cost prohibitive. The two-phase sampling approach can be gainfully employed in this case by collecting data on picking, which has highest correlation with the total pickings yield on a larger sample and the total pickings yield data on a smaller sample. Accordingly, a stratified two-stage two-phase sampling design has been proposed for selection of representative sample and an appropriate estimation procedure, based on two-phase sampling regression estimator, has been developed for estimation of average yield of cotton at district level. Utilizing the data of survey conducted in the Aurangabad and Amravati district of Maharashtra State and Adilabad and Guntur district of Andhra Pradesh wherein third picking data was collected on a larger sample and total pickings yield data on a smaller sample. An expression for optimum number of villages for larger and smaller samples has been obtained by minimizing cost subject to fixed percentage standard error of the estimates. These have been worked out empirically as well.

Key words : Yield, Cotton, Multiple picking, Sampling methodology, Two-phase sampling, Percentage standard error, Crop Cutting Experiment (CCE).

Introduction

Cotton is a multiple picking crop, which is grown in eleven states in India. Cotton is an important fibre yielding crop of global importance. It is one of the most important cash crops and accounts for around 25% of the total global fibre production. It is also known as the “White Gold” or the “King of Fibres”. India also has the distinction of having the largest area under cotton cultivation in the world *i.e.* about 126.07 lakh hectares. In India, the states of Gujarat (103.84 lakh bales), Maharashtra (83.35 lakh bales) and Telangana (54.44 lakh bales) are the leading cotton producing states having the predominantly tropical wet and dry climate. Cotton crop is harvested in the form of a number of pickings. The total number of pickings may vary from state to state. It varies from 2-3 pickings to 10 pickings. In the raw material consumption basket of the Indian textile industry, the proportion of cotton is around 59%. It plays a major role in sustaining the livelihood of an estimated 5.8 million cotton farmers and 40-50 million people engaged in related activities such as cotton processing and trade.

The existing procedure of estimation of average yield of cotton is based on crop cutting experiment (CCE) conducted under General Crop Estimation Survey (GCES), which utilizes data on all pickings. But this existing procedure is cumbersome and cost prohibitive. The two-phase sampling (double sampling) approach can be gainfully employed in this case by collecting data on picking which has highest correlation with the total picking yield on a larger sample as auxiliary variable and the total picking yield data on smaller sample. Accordingly, a stratified two-stage two phase sampling design is a very effective for the selection of representative sample. An estimation procedure, based on double sampling regression estimator will be used for the estimation of average yield of cotton at district level. This methodology will save cost of the survey significantly and will also be operationally more convenient than GCES procedure. As the traditional system of estimation of crop yield is facing several problems like lack of timely information and reliability of records maintained by the Government agencies. With this point of view, the present investigation will be undertaken with following objectives. To estimate the average yield of cotton at district level using two-phase sampling approach. To compare the results with the traditional methods of estimation of average yield. Optimization of CCE of cotton on the basis of remote sensing and GIS.

Materials and Methods

The data in Table 1 has been collected from the Deputy Director of Agriculture (Extension), Surat, Gujarat.

Methodology

A detailed examination of the data pertaining to Surat district showed that the yield of cotton at different pickings was highly correlated with the total yield. Hence, it may not be statistically proper to collect the data in respect to all the pickings from the same sample of fields since the additional information gained would be marginal. Therefore, it would be desirable to examine the possibility of using other sampling design such as double sampling which may not be only be more efficient, but also operationally more convenient resulting in more reliable data for estimation of crop yield of multiple picking crop like cotton.

Estimation Procedure using two-phase sampling approach under stratified two stagerandom sampling design framework

The estimation procedure for the estimation of the average yield of cotton using two-phasesampling regression approach under stratified two stage sampling design framework is as under:

L Number of strata (mandals/taluka) in a district.

N_h Total number of first stage units (fsu's- villages) in h^{th} stratum (h = 1, 2, …, L).

n'_h Number fsu's selected randomly in h^{th} stratum for observing yield for the p^{th} picking (p = 1, 2, …, P).

n_h Size of sub-sample selected randomly in h^{th} stratum for observing yield for the remaining pickings.

m' Number of second stage units (ssu's- fields) selected for observing yield for the p^{th} picking in i^{th} (i = 1, 2, …, $n'_{h..}$) village of h^{th} stratum.

m Size of sub-sample for observing yield for the remaining picking of these m' ssu's.

$y_{hij}(p)$ Yield of cotton in j^{th} field (j=1, 2,…, m') in h^{th} stratum corresponding to p^{th} picking.

Estimator of average yield corresponding to i^{th} picking, $\bar{y}_{nm}(p)$ for a district is given by

$$\bar{y}_{nm}(p) = \frac{1}{n_0 m}\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m} y_{hij}(p)$$

Where, $n_0 = \sum_{h=1}^{L} n_h$

Table 1 : Taluka wise Average yield of Cotton (Kg/ha) and Number of CCEs.

S. no.	Year	Crop	S. no.	Name of Taluka	CCEs planned	CCEs analyzed	Yield in Kg/ha
1	2020-21	Cotton Irrigated	1	Mangrol	20	20	1529.8
2			2	Umarpada	20	20	1509.6
3		Cotton Unirrigated	1	Mangrol	20	20	1278.6
4			2	Umarpada	20	20	1066.4

Table 2 : Picking wise correlation with total yield in Surat for the year 2020-21.

District	Crop	Taluka	Picking 1	Picking 2	Picking 3	Picking 4	Picking 5
Surat	Cotton Irrigated	Mangrol	0.4514	0.7915	0.5967	0.3423	0.2863
		Umarpada	0.3893	0.6776	0.5323	0.3821	0.2106
	Cotton Unirrigated	Mangrol	0.4229	0.7135	0.5329	0.3129	0.1925
		Umarpada	0.3658	0.6987	0.5758	0.3469	0.2325

An estimator of average yield, $\bar{y}_{nm}$ for a district is given by

$$\bar{y}_{nm} = \sum_{p=1}^{P} \bar{y}_{nm}(p)$$

A double sampling regression estimator for the estimation of average yield of cotton is given by

$$\bar{y}_{ld} = \bar{y}_{nm} + \hat{\beta}[\bar{y}_{n'm'}(p) - \bar{y}_{nm}(p)]$$

Where, $\bar{y}_{n'm'}(p) = \dfrac{1}{n'_0 m'} \sum_{h=1}^{L} \sum_{i=1}^{n'_h} \sum_{j=1}^{m'} y_{hij}(p)$

where $n'_0 = \sum_{h=1}^{L} n'_h$

$$\hat{\beta} = \frac{\left(\frac{1}{n_0} - \frac{1}{n'_0}\right) s_{by(p)y} + \frac{1}{n'_0}\left(\frac{1}{m} - \frac{1}{m'}\right) s_{wy(p)y}}{\left(\frac{1}{n_0} - \frac{1}{n'_0}\right) s^2_{by(p)} + \frac{1}{n'_0}\left(\frac{1}{m} - \frac{1}{m'}\right) s^2_{wy(p)}}$$

where,

$$s_{by(p)y} = \frac{1}{L} \sum_{h=1}^{L} \frac{1}{(n_h - 1)} \sum_{i=1}^{n_h} [\bar{y}_{hi.}(p) - \bar{y}_{h..}(p)](\bar{y}_{hi.} - \bar{y}_{h..})$$

and

$$s_{wy(p)y} = \frac{1}{n_0(m-1)} \sum_{h=1}^{L} \sum_{i=1}^{n_h} \sum_{j=1}^{m} [\bar{y}_{hij}(p) - \bar{y}_{hi.}(p)](\bar{y}_{hij} - \bar{y}_{hi.})$$

Where,

$$\bar{y}_{hi.}(p) = \frac{1}{m} \sum_{j=1}^{m} y_{hij}(p)$$

$$\bar{y}_{h..}(p) = \frac{1}{n_h} \sum_{j=1}^{n_h} \bar{y}_{hi.}(p)$$

$$s^2_{by(p)} = \frac{1}{L} \sum_{h=1}^{L} \frac{1}{(n_h - 1)} \sum_{i=1}^{n_h} [\bar{y}_{hi.}(p) - \bar{y}_{h..}(p)]^2$$

$$s^2_{wy(p)} = \frac{1}{n_0(m-1)} \sum_{h=1}^{L} \sum_{i=1}^{n_h} \sum_{j=1}^{m} [\bar{y}_{hij}(p) - \bar{y}_{hi.}(p))]^2$$

An estimator of $MSE(\bar{y}_{ld})$ is given by

$$\hat{MSE}(\bar{y}_{ld}) = \left\{\left(\frac{1}{n_0} - \frac{1}{N_0}\right) s^2_{by} + \frac{1}{N_0 m} s^2_{wy}\right\}\left(1 - r^{*2}\right)$$

$$+ r^{*2}\left\{\left(\frac{1}{n'_0} - \frac{1}{N_0}\right) s^2_{by} + \left[\frac{1}{N_0 m} - \frac{1}{n'_0}\left(\frac{1}{m} - \frac{1}{m'}\right) s^2_{wy}\right]\right\}$$

Where, $N_0 = L\bar{N}$ and $\bar{N}$ is the harmonic mean of N_{hs}. This is the usual estimate in sub-sampling design.

$$s^2_{by} = \frac{1}{L} \sum_{h=1}^{L} \frac{1}{(n_h - 1)} \sum_{i=1}^{n_h} [\bar{y}_{hi.} - \bar{y}_{h..}]^2$$

$$s^2_{wy} = \frac{1}{n_0(m-1)} \sum_{h=1}^{L} \sum_{i=1}^{n_h} \sum_{j=1}^{m} [\bar{y}_{hij} - \bar{y}_{hi.}]^2$$

$$r^{*2} = \frac{q^2_{y(p)y}}{q_{y(p)y(p)} q_{yy}} \text{ with}$$

$$q_{y(p)y} = \left(\frac{1}{n_0} - \frac{1}{n'_0}\right) s_{by(p)y} + \frac{1}{n_0}\left(\frac{1}{m} - \frac{1}{m'}\right) s_{wy(p)y}$$

$$q_{y(p)y(p)} = \left(\frac{1}{n_0} - \frac{1}{n'_0}\right) s^2_{by(p)} + \frac{1}{n_0}\left(\frac{1}{m} - \frac{1}{m'}\right) s^2_{wy(p)}$$

$$q_{yy} = \left(\frac{1}{n_0} - \frac{1}{n'_0}\right) s^2_{by} + \frac{1}{n_0}\left(\frac{1}{m} - \frac{1}{m'}\right) s^2_{wy}$$

Optimization of CCE using RS & GIS

Kriging is one of several methods that use a limited set of sampled data points to estimate the value of a variable over a continuous spatial field. It differs from simpler methods, such as Inverse Distance Weighted (IDW) Interpolation, Linear Regression, or Gaussian decays in that it uses the spatial correlation between sampled points to interpolate the values in the spatial field: the interpolation is based on the spatial arrangement of the empirical observations, rather than on a presumed model of spatial distribution. So, in the current study we used IDW Interpolation to optimize the CCE, which will help us to find out the yield of the taluka or district level by using limited number of CCEs.

Results and Discussion

Results from Table 3 reveals that 3rd and 2nd picking in Surat district have high correlation with total yield. It is

Table 3 : Comparison of estimates of average yield of cotton (Kg/ha) along with % S.E.using existing procedure and two-phase sampling Procedure.

Existing Procedure (GCES method)										
S. no.	**Year**	**Crop**	**Number of sampled villages**	**S. no.**	**Name of Taluka**	**CCEs planned**	**CCEs analyzed**	**Average yield in Kg/ha**	**% S.E.**	**Total number of picking involved**
1	2020-21	Cotton Irrigated	10	1	Mangrol	20	20	1529.8	5.69	100
2			10	2	Umarpada	20	20	1509.6	5.32	100
3		Cotton Unirrigated	10	1	Mangrol	20	20	1278.6	6.39	100
4			10	2	Umarpada	20	20	1066.4	6.32	100
Two-phase sampling Procedure										
S. no.	**Year**	**Crop**	**Number of sampled villages**	**S. no.**	**Name of Taluka**	**CCEs planned**	**CCEs analyzed**	**Average yield in Kg/ha**	**% S.E.**	**Total number of picking involved**
1	2020-21	Cotton Irrigated	n'=10 (for one picking)n=4 (40% of n') (for remaining picking)	1	Mangrol	20	20	1487.9	5.71	52
2			n'=10 (for one picking)n=4 (40% of n') (for remaining picking)	2	Umarpada	20	20	1478.5	5.65	52
3		Cotton Unirrigated	n'=10 (for one picking)n=4 (40% of n') (for remaining picking)	1	Mangrol	20	20	1210.6	6.59	52
4			4 n'=10 (for one picking)n=4 (40% of n') (for remaining picking)	2	Umarpada	20	20	985.4	6.48	52

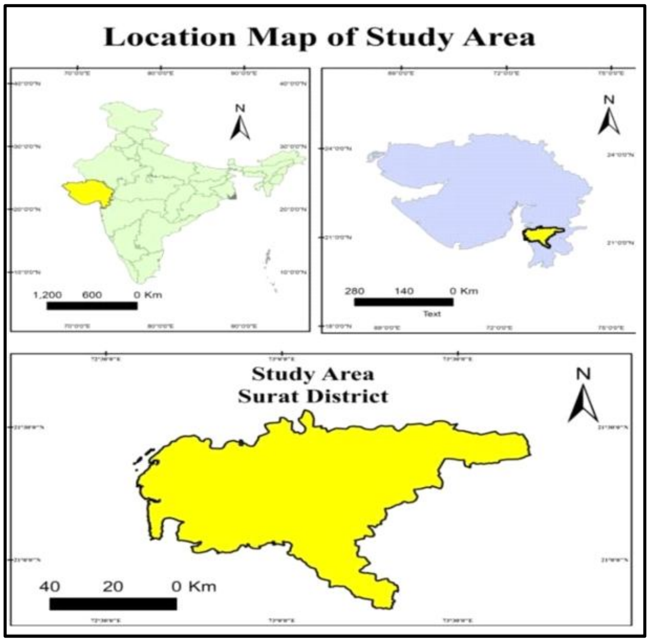

Fig. 1 : Location map of study area.

average yield of cotton along with % S.E. for the year 2020-21 was obtained using two-phase sampling approach under stratified two stage random sampling design framework. In order to compare the results, the estimate along with % S.E. was also obtained using existing GCES procedure. For implementation purpose, the selected CCE villages in the district have been treated as preliminary sample villages (n') for the second picking yield and a subsample of 40% CCE villages (n) were selected by SRSWOR from the preliminary sample villages for the remaining pickings.

A close perusal of the above table reveals that under the double sampling regression procedure involving stratified two stage sampling design framework, the estimates of average yield of cotton was obtained with less than 7% standard error for Surat district which is fairly reliable. The estimates are almost at par with the estimated obtained using existing procedures. Further as the sample size decreases *i.e.* 40%, percentage standard error increases slightly. This alternative method is

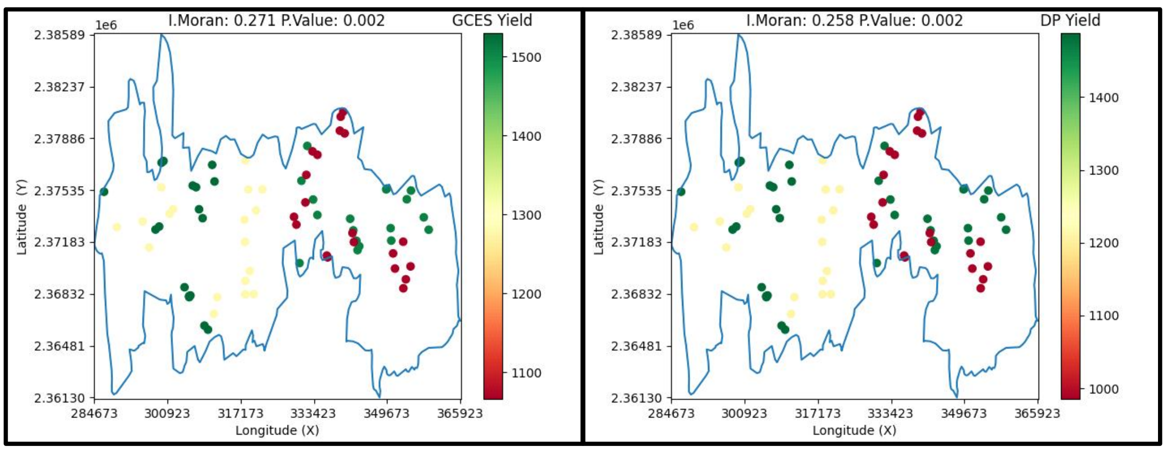

Fig. 2 : Distribution of the CCE plots and their yield.

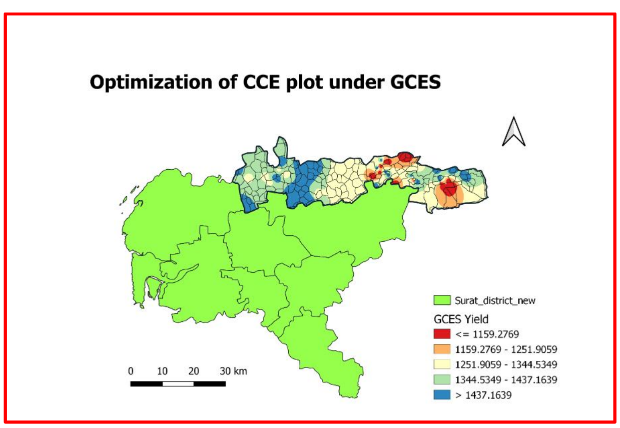

Fig. 3 : Optimization of CCE plot under GCES method.

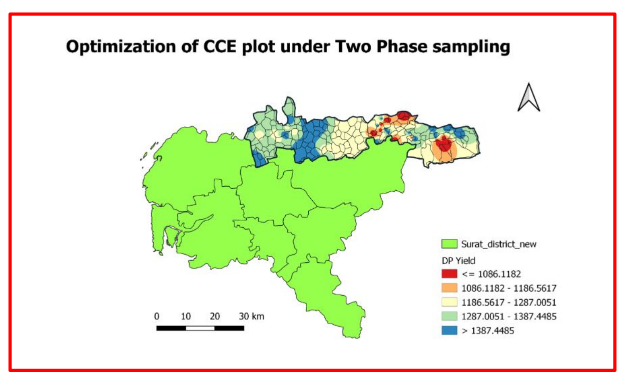

Fig. 4 : Optimization of CCE plot under Two-Phase sampling method.

therefore, desirable to use the yield of picking having highest correlation with total yield as auxiliary variable. It has been observed through the previous studies of Ahmad *et al.* (2013) that generally second and third picking of the cotton crop has the highest correlation with the total yield. Thus, there is an opportunity to use double sampling approach in this situation. The estimate of

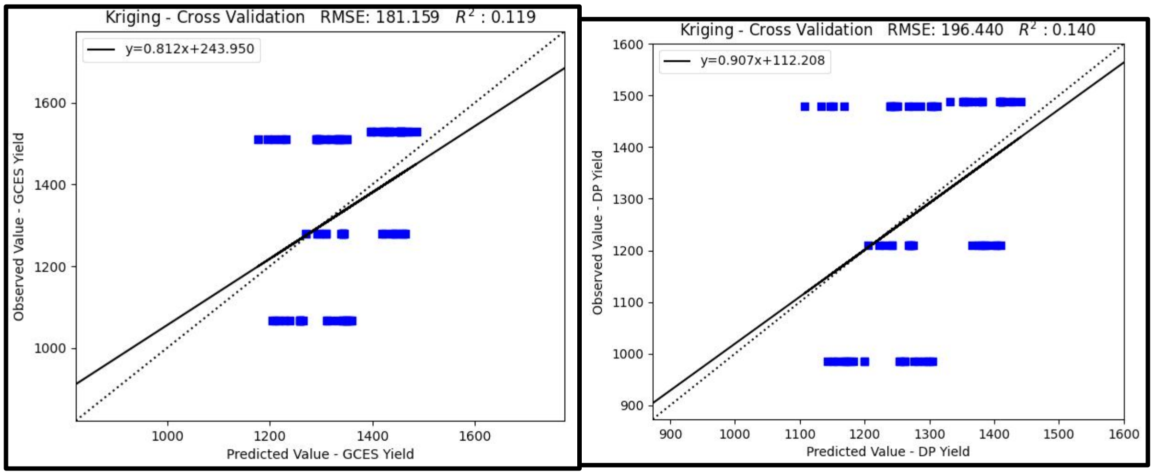

Fig. 5 : Cross-validation of Interpolation.

operationally more convenient than the existing GCES procedure and is expected to reduce the workload of the field staff significantly which in turn will lead to good quality CCE data from these limited number of plots.

Conclusion

The present study has revealed very encouraging results as shown from the analysis of results. The results demonstrated the feasibility of estimating cotton production with limited number of pickings using the two-phase sampling approach. It has been found that the alternative sampling methodology using two-phase sampling regression procedure under stratified two stage sampling design framework can be adopted in all the cotton growing districts of Gujarat. The results revealed that the alternative procedure will not only provide reliable estimates of average yield of cotton but will significantly reduce cost of survey. It will also be operationally more convenient than the GCES procedure (Table 3). Further, the workload of the field staff will be significantly reduced which in turn will lead to good quality CCE data from these limited number of plots. Fig. 2 showed the number of analysed CCE in Mangrol and Umarpada Taluka along with the distribution of yield. Figs. 3 and 4 showing the map of the interpolation and the classification of the yield over the whole area of the Mangrol and Umarpada Taluka. Fig. 5 is showing the cross validation of the interpolation for both the methods. It has been seen that R^2 is increase in Two-Phase sampling method. It is recommended that two-phase sampling regression procedure under stratified two stage sampling design framework will adopt to estimate average cotton yield for getting more reliable and cost-effective estimates of average cotton yield than general crop estimation survey procedure.

Statements and declarations : The authors declare no potential conflict of financial or non-financial interest relevant to this article.

Acknowledgement

The authors are thankful to the Vice-chancellor, Director of Research and Principal, College of Agriculture, Waghai, Navsari, Navsari Agricultural, University for providing the necessary facilities to complete the research and Deputy Director of Agriculture (Extension), Surat for providing necessary data for conducting this research.

References

Ahmad, T., Bhatia V.K., Sud U.C., Rai A. and Sahoo P.M. (2013). *Study to develop an alternative methodology for estimation of cotton production*. Project Report, IASRI publication.

Agricultural Research Data Book (2012, 2013, 2014, 2015, 2016, 2017). IASRI publication.

Ahmad, T., Bathla H.V.L., Rai A., Sahoo P.M., Gupta A.K., Jain V.K. and Mhadgut D.V. (2009). Study to investigate the causesof variation between official and trade estimates of Cottonproduction. Project Report, IASRI publication.

Indian Cotton, A Profile (2010, 2011, 2012). The Cotton Corporation of India Ltd. publication.

Kajale, J. (2006). Study to investigate the causes of variation betweenofficial and trade estimates of Cotton production in Maharashtra. AERC Report., Gokhale Institute of Politics and Economics, Pune.

Panse, V.G., Rajagopalan M. and Pillai S.S. (1966). Estimation of cropyields for smaller areas. *Biometrics*, **22(2)**, 374-384.

Ryerson, R.A., Dobbins R.N. and Thibault C. (1985). Timely crop areaestimates from Landsat. *Photo. Engg. Rem. Sens.*, **51**, 1735-1743.

Sukhatme, B.V. and Koshal R.S. (1959). A contribution to double sampling. *J. Ind. Soc. Agril. Statist.*, **11**, 128-144.

Sukhatme, P.V. and Panse V.G. (1951). Crop surveys in India II. *J. Ind. Soc. Agril. Statist.*, **3**, 97-168.