ESTIMATION OF PADDY CROP IN KOZHIKODE DISTRICT OF KERALA BY THE METHOD OF SMALL AREA ESTIMATION

Girish Babu.M
Assistant Professor, Department of Statistics, Govt. Arts and Science College, Meenchanda, Kozhikode, India.

Abstract
The information on crop area statistics is the backbone of Agricultural statistical system. Reliable and timely information on crop area is of great importance to planners and policy makers for efficient and timely agricultural development and making important decisions with respect to procurement, storage, public distribution, export, import and other related issues. Paddy cultivation was part of the culture of Kerala. Rice is the important cereal and staple food produced and consumed in Kerala. In this study we attempt to estimate the summer paddy production in Kozhikode district of Kerala state. The small area estimation technique was utilized to prepare the estimates and compared the results with the conventional methods. This method was found to be effective when the number of crop cutting experiments in a region was lesser than the desired number of experiments.

Key Words: Small Area Estimation, Direct Estimators, Synthetic Estimators, Regression Methods.

1. INTRODUCTION

India is primarily an agriculture-based country and its economy largely depends upon agriculture and allied sectors. Presently, contribution of agriculture about one third of the national GDP and provides employment to over seventy percent of Indian population in agriculture and allied activities. Therefore, our country's development is mainly depends upon the development of agriculture sector. The information on agricultural production of various crops is very important for planning and allocation of resources to different sectors of agriculture. Agricultural statistics in India have a long tradition.

A number of other organizations like the Institute of Agricultural Research Statistics (IARS) presently named as Indian Agricultural Statistics Research Institute (IASRI), Central Statistical Organization (CSO), Directorate of National Sample Survey, now re-organized as the National Sample Survey Organization (NSSO) and the Indian Statistical Institute (ISI) also participated in the efforts directed towards the improvement of agricultural statistics in different ways. The present system of agricultural statistics generates valuable statistics on a vast number of parameters. Some of the very important statistics are land-use statistics and area under principal crops through the Timely Reporting Scheme (TRS) and also on complete enumeration basis, yield estimates through the General Crop Estimation Surveys (GCES), cost of production estimates, agricultural wages, irrigation statistics etc. The information related to crop area and production plays important role in planning and allocations of resources for development of agriculture sector. The importance of various indicators of crop production statistics along with related schemes are described in the following sections.

2. CROP STATISTICS

The information on crop area statistics is the backbone of Agricultural statistical system. Reliable and timely information on crop area is of great importance to planners and policy makers for efficient and timely agricultural development and making important decisions with respect to procurement, storage, public distribution, export, import and other related issues. India possesses an excellent administrative setup, which has long standing tradition of generating quality information. Most parts of the country are having detailed cadastral survey maps, frequently updated land records and institutions like permanent village reporting agency for providing reliable and continuous data on crop area. However with more emphasis on local area planning, there is further need for crop area with respect to different varieties grown in the area, irrigation availability, the soil type etc. which can go a long way in rapid development of the region

3. YIELD STATISTICS

Crop production estimates are obtained by multiplying yield rates with area sown under the crop. In the past there were two series of estimates i.e. official series and N.S.S. series for yield estimation. The N.S.S. series was mostly confined to cereals only and was discontinued from 1970-71 and more emphasis was laid on the improvement in the quality of data from the official series. The yield estimation in official series was normally done on the basis of normal yield and condition factor. Normally, yield of crops in various states were estimated on the basis of crop cutting experiments, which used to be modified at the time of estimating yield on the basis of actual condition. Under this method production statistics was calculated by

IJMSRR
E- ISSN - 2349-6746
ISSN -2349-6738

finding area under crop and average yield per acre of the particular crop. The average yield was calculated by multiplying normal yield with condition factor.

4. TIMELY REPORTING SURVEY

In the permanently settled states of Kerala, Orissa and West Bengal a scheme similar to TRS was introduced with same objectives of obtaining area estimates based on 20% sample for use of both by Center and States. Here also, it was envisaged that complete enumeration of fields for area figures would be available for all villages over a period of five years as in case of TRS. Area enumeration under TRS has to be completed on priority basis in a random sample of 20% of the villages during each crop season in a state. EARAS was introduced as a sequel to TRS in the non-land record states namely Kerala, Orissa and West Bengal. This scheme provides for setting up whole time agency to cover 20% of villages every year so that all the villages of a state are covered in 5 years. In the sample villages under this scheme, the crop area is to be reported on the basis of complete enumeration.

The EARAS was initiated in Kerala during 1975-76 with 10% sample size of villages, which was increased to 20% in 1977-78. In Orissa, it was introduced during 1976-77 covering 6% of the villages, which was subsequently increased to 10% in 1977-78 and to 20% level in 1981-82. Under this scheme, in Orissa each Community Development Block is considered a stratum. In West Bengal the scheme was implemented from 1981-82 with the coverage of 7% of the villages, which was increased to 10% in 1982-83. In 1985-86 the coverage was extended to 14% and presently the coverage is 18%. Thereafter, the coverage of EARAS in the state was almost of the same order. The method of collection of area statistics in Kerala has undergone a change from 1987-88. The state was divided in 811 Investigator zones. A sample of key plots is selected from each Investigator zone with respect to key plot, a cluster consisting of five sub-survey numbers is formed and land utilization statistics are collected from these 100 clusters of five sub-survey division numbers.

5. SMALL AREA ESTIMATION

The need for crop production estimates for small areas (C.D. blocks, panchayats) has assumed urgency especially after the introduction of NAIS (crop insurance), Expansion of the scale of CCEs to meet this need is almost impossible if NAIS is implemented throughout the country and covers many more crops than at present. An approach other than CCEs has to be sought and the technique of “Small Area Estimation” (SAE) holds out a promising solution. One approach is use of auxiliary information through farmers’ appraisal survey for obtaining crop yield estimate at Block/ Gram Panchayat level for the purpose of estimation of yield rates under NAIS. The procedure involves scaling down the yield estimates from GCES at district/block level to Gram Panchayat (GP) level for insurance purposes. Another approach for estimation at small area level is to make use of remote sensing digital data along with yields data from general crop estimation surveys.

A crucial problem with sample surveys is that the part of samples falling in the specified local area may be too small , resulting in estimates pertaining to that local area with large standard errors as to be of no use. The present crop cutting approach for farming estimates at local level will not be practicable when viewed from the financial and administrative angle. An alternative method which is financially viable and operationally feasible is required to be developed.

Using the state level survey data to produce the estimates at district or further smaller level may end up with very small sample sizes in these domains which results into very unstable estimates of parameters of interest. Hence, we need a special technique to produce estimates for such small domains or small areas. The underlying theory which resolves the problem of small sample sizes is referred to as small area estimation (SAE) in the literature of survey sampling.

Small Area Estimation plays a prominent role in survey sampling due to growing demands for reliable small area statistics from both public and private sectors. Sample surveys, whether conducted by government organizations or by private entities, aim to produce reasonably accurate direct estimators, not only for the characteristics of whole population but also for a variety of subpopulations or domains.

Due to small sample size domain-specific direct estimators provide unacceptably large coefficient of variation. Therefore, it becomes necessary to employ indirect small area estimators that make use of the sample data from related areas or domains through linking models, and thus increase the effective sample size in the small areas. Such estimators can provide significantly smaller coefficient of variation than direct estimators, provided the linking models are valid.

The problem of producing statistics for small areas (districts, subdivisions, municipal areas, blocks consisting of villages) has been drawing increasing attention in the recent times. This is mainly due to the fact that in the present day situation of regional planning, the distribution of central funds are often made on a local or regional basis and may depend on variables such as number of unemployed, the condition of the housing stock ,the use of fertilizers etc.

Various Small Area Estimation methods have been developed in recent past by Gonzalez (1973), Purcell and Kish (1979), Drew et al (1982). Many studies dealing with the small area estimation problem have been discussed by various authors, Ghosh and Rao (1994), Rao (1999). Sing and Sisodia (2011) attempted to propose the estimates of population mean for small areas in longitudinal surveys.

Singh et al (2005) used National Sample Survey (NSS) data for application of Spatio-Temporal Models in Small Area Estimation. Sastry (2003) studied the feasibility of using NSS house hold consumer expenditure survey data for estimation of district poverty estimates.

Srivastava(2009) attempted an Area Level model based small area estimation approach for estimating some of the poverty indicators at district level for Uttar Pradesh using 61st round (2004-05) NSS data .

6. SMALL AREA ESTIMATION METHODOLOGY

Small area typically denotes a subset of the population for which very little information is available from the sample survey. These subsets refer to a small geographic area (e.g., a municipality, a census division, block, tehsil, Gram Panchayat, etc.) or a demographic group (e.g., a specific age-sex-race group of people within a large geographical area) or a cross classification of both. The statistics related to these small areas are often termed as small area statistics.

Traditionally there are two types of small area estimation direct and indirect estimation. The direct small area estimation is based on survey design and include H-T estimator. On the other hand, indirect approaches of small area estimation can be divided into two classes' statistical and economic approaches. The statistical approach is mainly based on different statistical models and techniques. However the economic approach uses techniques such as micro simulation modeling.

Small area model can be classified into two broad types: (i). area level random effect models, which are used when auxiliary information is available only at area level, these relate small area direct estimators to area specific covariates and (ii). nested error unit level regression models, employed originally by Battese et al (1988) for predicting areas under corn and soybean in 12 countries of the state of Lowa in the USA, these models relate the unit values of a study variable to unit specific covariates.

Srivastava et al (2007) were attempted an area level small area model to compare the empirical best linear unbiased predictor estimate and its mean squared error estimates.

7. DIRECT ESTIMATORS

Direct estimators are defined as estimators that use the values of study variables only from the domain under consideration. Direct estimators do not borrow strength in terms of study variable but may or may not borrow strength in terms of auxiliary variables .The commonly used direct estimator is Horvitz- Thompson (H-T) Estimator.

Suppose a linear estimator based on sample weights $\{x_j, j \in s\}$ is used to make inference about population level quantities. Here s, indicates the sample of size n drawn with sampling design p(s) from a population $U = \{1, \dots, N\}$ of size N. Further, if $\pi_j = \sum_{j \in s} p(s)$ are the first order inclusion probabilities then $x_j = \frac{1}{\pi_j}$ defines the design weight of element *j*. Under simple random sampling , $\pi_j = \frac{n}{N}$ and $x_j = \frac{N}{n}$. Let *i* denote restriction to small area *i (i = 1, . . . ,m)*. We assume that the population consists of m non-overlapping domains or small areas U_i each with population of size N_i such that $U = \bigcup_{i=1}^{m} U_i$ and $N = \sum_{i=1}^{m} N_i$. Let s_i be the part of the sample of size n_i that falls in the small area *i* and $n = \sum_{i=1}^{m} n_i$. We denote y_{ij} the value of the j^{th} population unit in small area *i* for the characteristic of interest Y. The population mean of Y in the area i, $\bar{Y}_i = \frac{1}{N_i}\sum_{j\in U_i} y_j$ could be then estimated using the same weights leading to the estimator

$$\hat{\bar{Y}}_i = \frac{\sum_{j\in s_i} x_j y_j}{\sum_{j\in s_i} x_j} \qquad (8.1)$$

or if the population size N_i of the small area *i* is known,

$$\hat{\bar{Y}}_i^{HT} = \frac{1}{N_i}\sum_{j\in U_i} x_j y_j \qquad (8.2)$$

The estimators (8.1) and (8.2) are sometimes referred to as direct estimators of small area *i* mean $\bar{Y}_i$. More precisely , the estimator (8.1) is referred to as Hájek type of the direct estimator, and (8.2) as the Horvitz – Thompson (HT) type of the direct estimator. These names refer to alternative approaches to estimating finite population means in the classical sampling literature, see Cochran (1977) and Särndal et al (1992). Irrespective of which form of direct estimator is used , it is easy to see that its variance can be large when the area sample size n_i is small. For example, under simple random sampling , with no auxiliary information , a direct estimator of the mean of Y for small area *i* ($\bar{Y}_i = \frac{1}{N_i}\sum_{j\in U_i} y_j$) is $\hat{\bar{Y}}_i = \bar{y}_i$,

where $\bar{y}_i = \sum_{s_i} x_j y_j \big/ \sum_{s_i} x_j = \frac{\sum_{j\in s_i} y_j}{n_i}$ is the sample mean of Y in area i, and its variance is $Var_p\left(\hat{\bar{Y}}_i\right) = (1-f_i)\frac{S_i^2}{n_i}$

with $f_i = n_i/N_i$ and $S_i^2 = (N_i-1)^{-1}\sum_{j=1}^{N_i}(y_j-\bar{Y}_i)^2$, $N_i \geq 2$. Here E_p and Var_p denotes the expectation and variance respectively under the design – based approach. An unbiased estimator for S_i^2 is $s_i^2 = (n_i-1)^{-1}\sum_{j\in s_i}(y_j-\bar{y}_i)^2$. Thus , an unbiased estimator for variance is given by $v\left(\hat{\bar{Y}}_i\right) = (1-f_i)\frac{s_i^2}{n_i}$ when N_i is known. For unknown N_i, $f_i = n_i/N_i$ is replaced by $f = n/N$ and then the estimator for variance is $v\left(\hat{\bar{Y}}_i\right) = (1-f)\frac{s_i^2}{n_i}$. It is obvious that for small sample size n_i , the variance will be larger unless the variability of the Y value is sufficiently small.

Suppose that in addition to survey variable Y, values of p-auxiliary variables are also unknown. Let us denote by x_{ij} a *px1* vector of auxiliary variable X for the unit *j* in area *i* . Then with known auxiliary information , a more efficient design-based direct estimator for $\bar{Y}_i$ is the regression estimator defined as

$$\hat{\bar{Y}}_i = \bar{y}_i + (\bar{X}_i - \bar{x}_i)^{\prime}\beta_i \qquad (8.3)$$

where β_i is the vector of regression coefficients in area I, $\bar{x}_i = \frac{\sum_{j\in s_i} x_j}{n_i}$ and $\bar{X}_i = \frac{\sum_{j=1}^{N_i} x_j}{N_i}$ are the sample mean and population mean of auxiliary variable X in the area i respectively. The variance of (8.1) is

$$V\left(\hat{\bar{Y}}_i\right) \approx n_i^{-1}(1-f_i)S_i^2(1-\rho_i^2) = (1-\rho_i^2)Var_p(\hat{\bar{Y}}_i) \qquad (8.4)$$

where ρ_i is the multiple correlation between survey variable Y and auxiliary variable X in area i. An estimate of variance of (8.4) is then $v\left(\hat{\bar{Y}}_i\right) = (1-\hat{\rho}_i^2)(1-f_i)s_i^2\big/n_i$. We note that by use of auxiliary variables, the variance is reduced by the factor $(1-\rho_i^2)$. This indicates that use of good auxiliary information, in the sense of high correlation with survey variable Y, increase the accuracy in SAE. However, the problem with the regression estimator (8.4) is that in practice the regression coefficients β_i are seldom known. Replacing β_i its ordinary least square (OLS) estimates $\hat{\beta}_i$ is not effective because of small sample sizes in each area i.

IJMSRR
E- ISSN - 2349-6746
ISSN -2349-6738

8. INDIRECT ESTIMATORS

When the sample size for each small area is sufficiently large to give reasonably accurate estimates, the direct estimator is the most desirable. As the sources of data are usually sample surveys designed to produce larger or higher level statistics, sample sizes for the smaller areas are usually small. Consequently, the associated variances of these estimators are likely to be unacceptably large.

8.1. Synthetic Estimators

In producing the synthetic estimates for small areas, availability of direct estimates for a set of larger domains of the population is assumed. Appropriate weights or proportions are then applied to these large population domain estimates to obtain the desired small area estimates. This class of estimators implicitly assumes that small areas which are being considered are similar to some larger areas which contain them and for which the reliable direct estimate is available. The term 'synthetic' refers to the fact that an estimator computed from a large domain is used for each of the separate areas comprising that domain, assuming that the areas are 'homogeneous' with respect to the quality that is estimated. Thus, synthetic estimators already borrow information from other 'similar areas'. Thus , synthetic estimators already borrow information from other 'similar areas'.

In synthetic estimation , we assume availability of reliable direct estimates $\ddot{T}_{yg} = \sum_{i=1}^{m} \ddot{T}_{yig}$ for the totals of larger group or class *g(g=1, … ,G)* that encompass the small areas *i(i = 1, … ,m)* for a given survey, where $\ddot{T}_{yg}$is the estimate of population total ($T_{yig} = \sum_{j=1}^{N_{ig}} y_{jig}$) of Y in the (i , g)th cell with population of size N_{ig}. Here y_{jig} is the value of unit j(j = 1, …, N_{ig}) for variable of interest Y in the cell (i,g). From the available estimates for population $\ddot{T}_{yg}$, estimates of population means for group *g* are obtained as $\ddot{\bar{Y}}_g = \frac{\sum_{i=1}^{m} \hat{T}_{yig}}{\sum_{i=1}^{m} N_{ig}}$ = $\ddot{T}_{yg}/N_g$. A suitable auxiliary information available from a census or some other source is used to compute a series of weights or proportions w_{ig} such that$\sum_g w_{ig} = 1$. The weights w_{ig} are then applied to the group means to derive the synthetic estimator for the *i*th small area mean $\bar{Y}_i$ as $\ddot{\bar{Y}}_i = \sum_{j=1}^{G} w_{ig} \ddot{\bar{Y}}_g$. This estimator is referred to as the design – based synthetic estimator, see Gonzales and Hoza (1978). Rao and Choudry (1995) suggested the use of a ratio synthetic estimator. The ratio synthetic estimator for the population total of Y in small area *i* is $\ddot{T}_{y_i} = \bar{R}_i T_{x_i}$. They assumed that area *i* population ratios $R_i = \frac{T_{y_i}}{T_{x_i}} = \sum_{j=1}^{N_i} y_j$ and $T_x = \sum_{i=1}^{N_i} x_i$ respectively being the population total of the characteristic of interest Y and covariate X for the *i*th small area, are homogeneous. Thus, R_i=$R_v = \frac{T_y}{T_x}$, where R_v, T_y and T_x are the values for the whole population. Here R_v is estimated by $\overline{R_v} = \frac{\bar{y}}{\bar{x}}$, where $\bar{y}$ and $\bar{x}$ are the overall sample means. We use a subscript of U to denote the population level quantities.

The design-variance of a synthetic estimator $\ddot{T}_{y_i}$ of the population total of Y in area *i* will be small relative to the *p*- variance of a direct estimator $\ddot{T}_{y_i}{}^{d}$ because it depends on the precision of direct estimators at a large area level. This variance can be estimated using standard design-based methods but it is more difficult to estimate the MSE of $\ddot{T}_{y_i}$ because it is hard to estimate the bias.

9. APPLICATION OF SMALL AREA ESTIMATION

Paddy cultivation was part of the culture of Kerala. Rice is the important cereal and staple food produced and consumed in Kerala. The area and production of summer paddy in Kozhikode district and Kerala state are given in Table.10.1. The area wise contribution of paddy by Kozhikode district is below 1% of the entire state paddy area. Also the contribution of paddy production from this district was only 1437 tonnes and was below 1% of the entire state level production.

IJMSRR
E- ISSN - 2349-6746
ISSN -2349-6738

Table 10.1.Area and Production of Summer Paddy in Kerala and Kozhikode for the Year 2013-14

	AREA (Ha)	PRODUCTION (Tonnes)
KOZHIKODE	563	1437
KERALA	199611	564325

From 2008-09 to 2009-10, area under paddy remain constant as 2.34 lakh Ha. During 2012 – 13 , the area under paddy declined to 1.97 lakh Ha and the production of paddy also declined to 5.08 lakh MT from 5.68 lakh MT in 2011-12. A detailed picture of the trend of area and production of paddy in Kerala and India are shown in table 10.2.

Table 10.2.Area and Production of Paddy in Kerala and India

SL NO	YEAR	AREA (Ha)		PRODUCTION(MT)	
		KERALA	INDIA	KERALA	INDIA
1	2002-03	311000	40410000	689000	75720000
2	2003-04	287000	42496000	570000	88280000
3	2004-05	290000	41665000	667000	85310000
4	2005-06	276000	44258000	630000	91790000
5	2006-07	264000	43810000	642000	93360000
6	2007-08	229000	43900000	528000	96700000
7	2008-09	234000	45600000	590000	99400000
8	2009-10	234000	41920000	598000	89090000
9	2010-11	213187	42560000	522758	95330000
10	2011-12	208160	43970000	568993	102570000
11	2012-13	197277	42410000	508299	104399000

Source : Economics and Statistics Department, Govt of Kerala

The average annual decline in area under paddy during the 8th five year plan was around 22000 Ha, whereas it has come down to an average of 13000 Ha during the 9th five year plan period. The average annual decline in area during the 10th plan was 47200 Ha. In 2007-08 the production of paddy declined from 6.42 lakhs MT to 5.28 lakhs MT. The change in area and production of paddy in the major districts of Kerala during 2009 – 10 to 2013-14 is shown in the table 10.3.

Table 10.3.District-Wise Area and Production of Paddy in Kerala

SL NO	DOSTRICTS	AREA (Ha)				PRODUCTION (Tonnes)			
		2010-11	2011-12	2012-13	2013-14	2010-11	2011-12	2012-13	2013-14
1	Thiruvananthapuram	2919	2395	1816	2001	6923	6139	4096	5326
2	Kollam	3342	2097	1387	1363	7155	4768	2928	3234
3	Pathanamthitta	2986	2802	2280	2467	6628	8989	6041	7554
4	Alapuzha	37060	36251	36195	37403	91325	111980	104593	106866
5	Kottayam	14775	21410	17571	15746	40970	63579	51019	50729
6	Idukki	1819	1264	1176	661	4744	3135	3183	1796
7	Ernakulam	9016	7731	3940	4052	17823	16572	8533	9056
8	Thrissur	20259	21172	23098	22274	53079	62316	67569	66653
9	Palakkad	87511	83998	79201	82896	218155	224413	189229	238065
10	Malappuram	8949	7528	6674	7549	21089	15877	15377	19709
11	Kozhikode	3003	2920	3511	2433	3814	4274	5326	3850

12	Wayanad	11054	8995	10230	11481	27911	23526	28052	30755
13	Kannur	6339	5740	6684	5080	13308	12170	14237	11293
14	Kasarkode	4155	3857	3514	4205	9834	8555	8116	9439
	State	**213187**	**208160**	**197277**	**199611**	**522758**	**568993**	**508299**	**564325**

Source : Economics and Statistics Department , Govt. of Kerala.

In this study we collected the Summer season Crop Cutting Experiment (CCE) paddy yield in Kozhikode district in Kerala during the year 2013-14 which is produced by Economics and statistics department. The details of paddy weight from each CCE conducted during the summer season were collected. Total cropped area in each block was collected from the records of Economics and Statistics department and by field visit. The total number of Crop Cutting Conducted in each block and the average weight of paddy obtained with SD and confidence interval of mean weight with minimum and maximum weights are presented in the table 10.4.

Table 10.4. Zone-Wise Paddy Crop Cutting Statistics

Sl No	Zones	Total cropped area(Acre)	Number of CCE's	Average Production from CCE (Kg)	SD	95% CI	Min	Max
1	KOZHIKODE	555	10	6.87	2.52	5.06-8.67	3.64	11.12
2	CHELANNUR	1158	29	6.39	2.98	5.25-7.52	1.00	14.78
3	KUNNAMAGALAM	1967	32	7.68	2.91	6.63-8.73	2.86	13.16
4	KODUVALLI	637	11	6.65	3.93	4.01-9.28	0.98	13.12
5	BALUSSERY	464	13	6.65	1.28	5.87-7.42	5.38	9.84
6	PERAMBRA	1704	27	8.85	2.57	7.83-9.86	0.61	12.00
7	MELADI	2186	24	8.83	2.85	7.63-10.04	2.43	14.31
8	PANTHALAYANI	1335	19	7.39	3.04	5.92-8.85	1.98	13.72
9	KOYILANDY	793	8	4.64	2.58	2.48-6.80	0.85	10.20
10	VADAKARA	50	4	11.00	1.03	9.35-12.64	9.98	12.28
11	THODANNURR	1270	24	12.60	2.1	11.71-13.48	8.52	17.32
12	KUNNUMMAL	669	8	9.76	2.02	8.08-11.45	6.63	13.23

The variable of interest for which small area estimates are required is yield for paddy crop. We are interested in estimating the average yield at the Zone level. In Kozhikode district there are 57 zones, however supervision, on a subsample, of crop cutting experiments is carried out in 33 zones only. These 33 zones are the small areas for which we are interested in producing the estimates. Establishment of an Agency for Reporting Agricultural Statistics (EARAS) manual of agriculture statistics division of Economics and Statistics department met the maximum number of paddy crop cutting experiment per an investigator zone is 8. In case if any zone does not have 8 cuts in SAE we borrow it from the nearby zones by using Simple Random Sampling (SRS) method. But in direct estimation they took the actual cut and prepare estimate that estimate is not reliable one. Let X_1 is the cultivated area and X_2 is the production in 5m X 5m and Y is the production in cultivated area calculated by using $\frac{X_1 * X_2}{0.62}$. Then estimate the production by using regression-synthetic estimator $\ddot{Y}_{i[s]} = \ddot{X}'_i \ddot{\beta}$. Where X_i is the available auxiliary information in a small area i and $\ddot{\beta}$ is the regression coefficients. From our error calculation we can conclude there is a variation between direct and small area estimation.

Next we cluster 8 areas in to one zone .From now, instead of X_1 we will be using $\overline{X_1}$ which is the mean cultivated area for each zone and instead X_2 we will be using $\overline{X_2}$ which is the mean production in 5mX5m for each zone. Now Y becomes Y', where $Y' = \frac{X_1 * X_2}{0.62}$. We next performed scatter plots of the dependent variable average production against the two independent variables average cultivated area and average production in 5m x 5m. Here Y' and $\overline{X_1}$ are correlated with *r = 0.88* which is positive and highly significant. So $\overline{X_1}$ can be included in our model. Similarly Y' and $\overline{X_2}$ are correlated with r = 0.56 which is positive and significant and so $\overline{X_2}$ can be included in our model. The scatter plots of the variables are shown below,

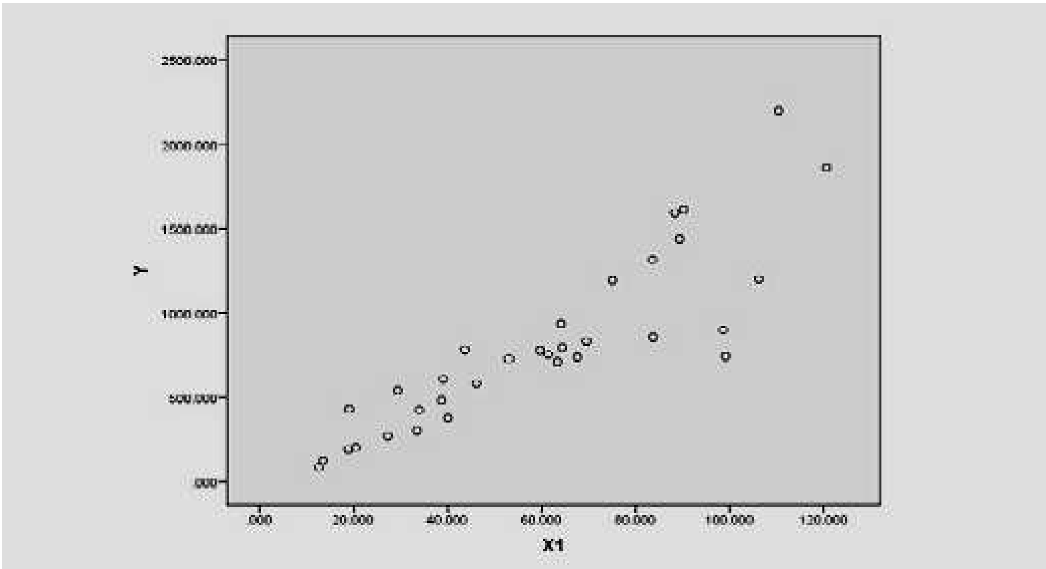

Figure 10.1. Scatter plot of average production against average cultivated area.

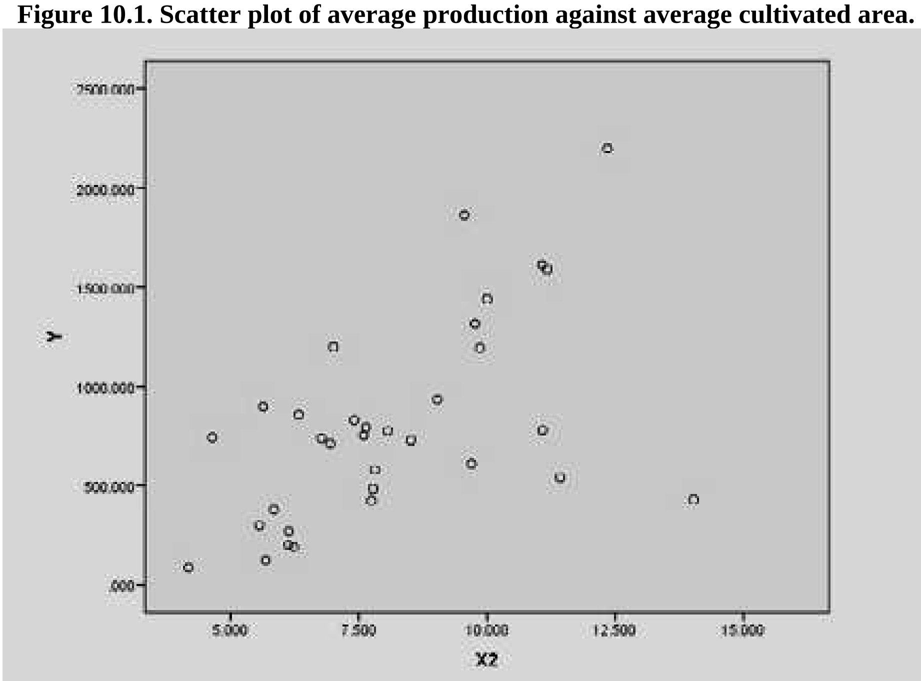

Figure 10.2. Scatter plot of average production against average production in 5m X 5m.

The two scatter plots shows the linear relationships. The higher the value of cultivated area the production is also higher. Also average production in 5m X 5m plot is directly proportional to average production in cultivated area.

So the data can be expressed as the model $Y' = \beta_0' + \beta_1' \overline{X}_1 + \beta_2' \overline{X}_2 + e'$, where $\beta_0', \beta_1', \beta_2'$ are the regression coefficients and e' is the error follows $N(0, \sigma^2)$. From the model we estimated the production $\hat{Y}'$ by using the regression coefficients given below and the estimated production is given in the table 4.6. The normality of the residuals is tested and the figures are given in Figure 4.4 and Figure 4.5.

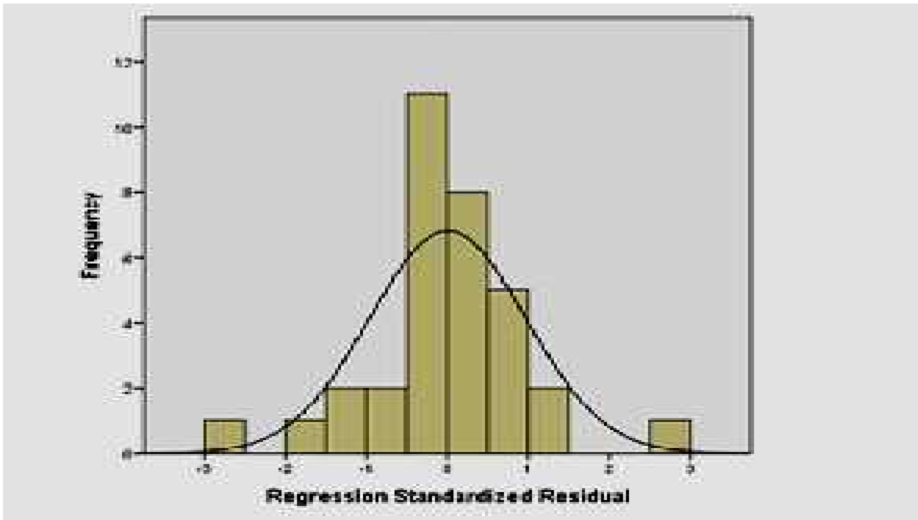

Figure 10.3. Histogram and normality of residuals.

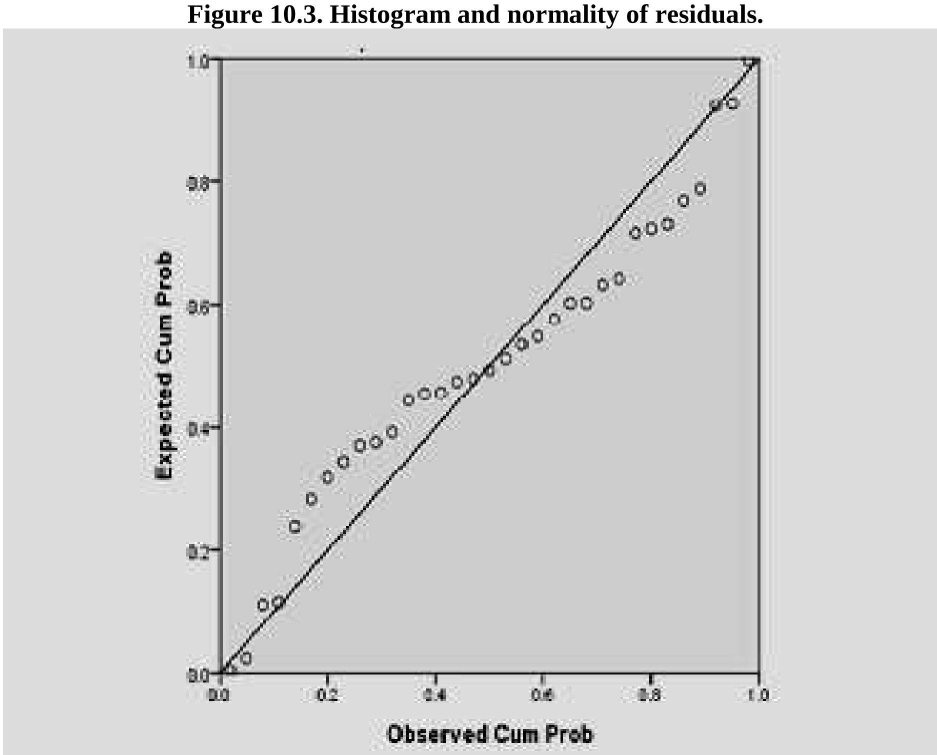

Figure 10.4. P-P plot of residuals.

Block wise and gram Panchayat wise paddy productions of Kozhikode district were computed by direct estimation method and small area estimation technique. The Panchayats which have no representation of paddy as per the records of Economics and Statistics department were excluded for estimation. The panchayats without sufficient number of paddy crop cutting experiments (CCE) were included additional CCE's borrowed from the nearby panchayats by simple random techniques. For preparing the estimates of paddy production in certain panchayats where the numbers of CCE's are less than 8, the CCE details from the nearby panchayats were borrowed by simple random sampling techniques. The linear regression analysis of the cultivated area and the weight of paddy obtained from CCE's were made and the estimates of the regression coefficients were computed. Using this fitted regression line the estimates for Y were obtained. The estimates of paddy production in

IJMSRR
E- ISSN - 2349-6746
ISSN -2349-6738

the selected 33 zones were computed and the differences of these estimates with the conventional estimates were presented in table 10.5. The deviation of these values was also shown in this table.

Table 10.5. Estimated Production of Paddy in the selected Zones

Zones	Cultivated Area(Acre)	Production/CCE(Kg)	Y dash	Y dash cap	Error
1	64.375	7.878	817.978	864.638	46.66
2	64.125	9.034	934.337	896.621	-37.716
3	33.5	5.565	300.757	354.008	53.251
4	12.625	4.181	85.142	14.238	-70.904
5	40	5.845	377.097	455.114	78.017
6	20.375	6.129	201.409	184.459	-16.95
7	38.625	7.778	484.526	494.97	10.444
8	98.625	5.634	896.175	1283.203	387.028
9	88.25	11.171	1590.101	1305.799	-284.302
10	69.5	7.409	830.497	923.166	92.669
11	27.25	6.136	269.698	282.562	12.864
12	61.375	7.593	751.596	813.15	61.554
13	83.125	6.329	854.892	1092.816	237.924
14	46.125	7.813	581.212	602.816	21.604
15	59.625	8.064	775.486	802.729	27.243
16	53	8.518	728.109	722.37	-5.739
17	18.875	6.229	189.625	166.1	-23.525
18	90.125	11.076	1610.076	1329.57	-280.506
19	39	9.698	610.004	559.354	-50.65
20	34	7.748	424.863	428.206	3.343
21	67.625	6.768	738.149	876.753	138.604
22	120.625	9.561	1560.203	1717.178	156.975
23	89.25	10.001	1439.696	1284.054	-155.642
24	63.375	6.941	709.519	821.594	112.075
25	13.5	5.685	123.786	72.94	-50.846
26	106.125	7.003	1198.613	1432.066	233.453
27	75	9.161	1192.893	1076.886	-116.007
28	99.125	4.643	742.238	1259.837	517.599
29	43.625	11.078	779.445	567.635	-211.81
30	29.375	11.426	541.365	475.497	-65.868
31	19	14.024	429.768	407.687	-22.081
32	110.375	12.345	2197.709	1657.867	-539.842
33	83.625	9.764	1316.925	1196.673	-120.252

The difference of estimates of paddy production in Kozhikode district by conventional method and the small area estimation method are very clear from the below figure 4.5.

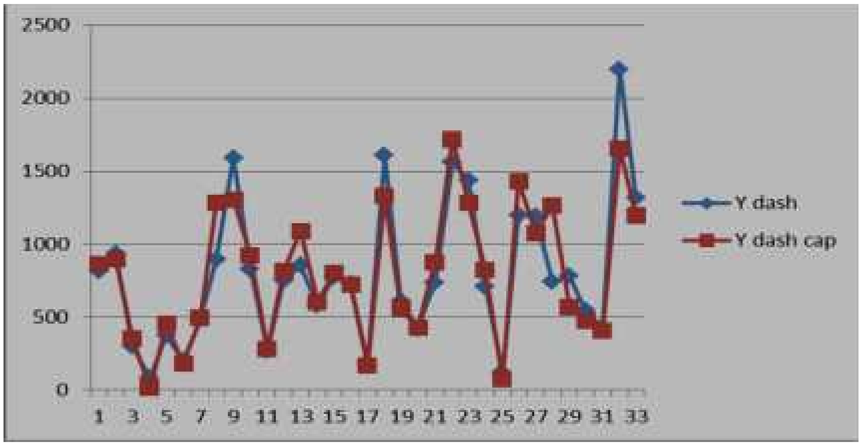

Figure - 10.5. Comparison of Y' and $\hat{Y}'$

10. CONCLUSION

There has been phenomenal growth in the theoretical aspects of SAE techniques. Model based estimation has been a turning point in the growth of SAE methods. Depending on the type of data availability, several types of models and corresponding procedures are available. In recent years applications of these methods are also being made in variety of situations. In the present study we analyzed the summer paddy production of Kozhikode district during 2013-14 agriculture year. The paddy production at gram Panchayat and block Panchayat were analyzed by the conventional methods and by SAE methods and the results were compared. The variation of the results were studied and found that SAE methods can be useful to prepare the paddy estimates in the areas were sufficient CCE's are not available.

REFERENCES

1. Bettese,G.E., Harter,R.M., Fuller,W.A. (1988). An error component model for prediction of country crop areas using survey and satellite data. Journal of American Statistical Association. 95, 1127 - 1142.
2. Cochran,W.G. (1977). Sampling Techniques. John Wiley & Sons., New York.
3. Drew, J.D., Singh,M.P., Choudhury,G.H. (1982). Evaluation of small area estimation techniques for the Canadian Labour Force Survey. Survey Methodology. 8, 17 - 47.
4. Ghosh,M., Rao,J.N.K. (1994). Small Area Estimation; an appraisal. Statistical Science.,9(1),55 - 93.
5. Gonzalez, M.E. (1973). Use and evaluation of synthetic estimates. Proceedings of the Social Statistics Section. American Statistical Association. 33 - 36
6. Gonzalez,M.E., Hoza,C.(1978). Small Area Estimation with Applications to Unemployment and Housing Estimates. Journal of the American Statistical Association. 73, 7-15.
7. Purcell,N.J., Kish,L.(1979). Estimation of small domains.Biometrics.,35, 365 - 384.
8. Rao, J.N.K. (1999). Some recent advances in model based small area estimation. Survey Methodology.,25, 175 - 186.
9. Rao, J.N.K., Choudry,G.H.(1995). Small Area Estimation: Overview and Empirical Study, in B.G. Cox et al (editors).Business Survey Methods, John Wiley and Sons, New York. 527 - 542.
10. Sarndal,C.E., Swensson,B., Wretman,J.(1992). Model Assisted Survey Sampling. Springer -Verlag, New York.
11. Shastry, N.S., (2003). District level poverty estimates - feasibility of using NSS house hold consumer expenditure survey data. Economics and Political Weekely., January 25, 2003.
12. Singh,B.B., Shukla,G.K., Kundu,D. (2005). Spatio-Temporal models in Small Area Estimation. Survey Methods, 31, 183 – 195.
13. Singh,B.B., Sisodia,B.V.S.(2011). Small Area Estimation in longitudinal surveys. Journal of Reliability and Statistical Studies.,4(2), 83 - 91.
14. Srivastava, A.K., (2009). Some aspects of estimating poverty at small area level. Journal of the Indian Society of Agricultural Statistics.,63(1), 1- 23.
15. Srivastava, A.K., Sud,U.C., Chandra,H.(2007). Small area estimation - an application to National Sample Survey data. *Journal of the Indian Society of Agricultural Statistics*.,61(2), 249-254.