© April 2026 | IJIRT | Volume 12 Issue 11 | ISSN: 2349-6002

Small Area Estimation Models in Business Demography:A Case Study

L. Raghavendra[1], P. Vishnu priya[2], K. Sreenivasulu[3], Venkatesh kuncham[4], Kandunuru Vijaya Kumar[5], D.Chandrakesavulu Naidu[6], B. Sarojamma[7]

[1]*Research scholar, Department of Statistics, S V University, Tirupati.*

[2,7]*Department of Statistics, S V University, Tirupati.*

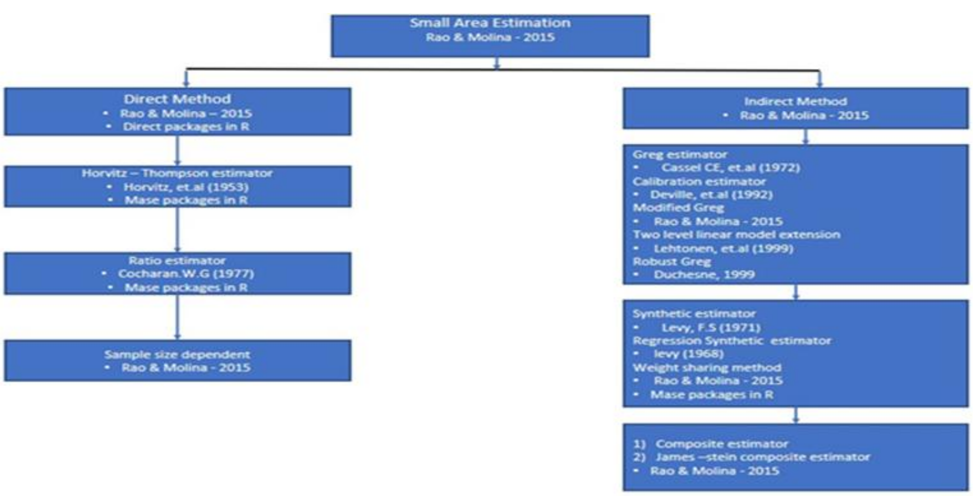
[3]*Department of Statistics, Sri Padmavathi Women's Degree & Pg College, Tirupati.*

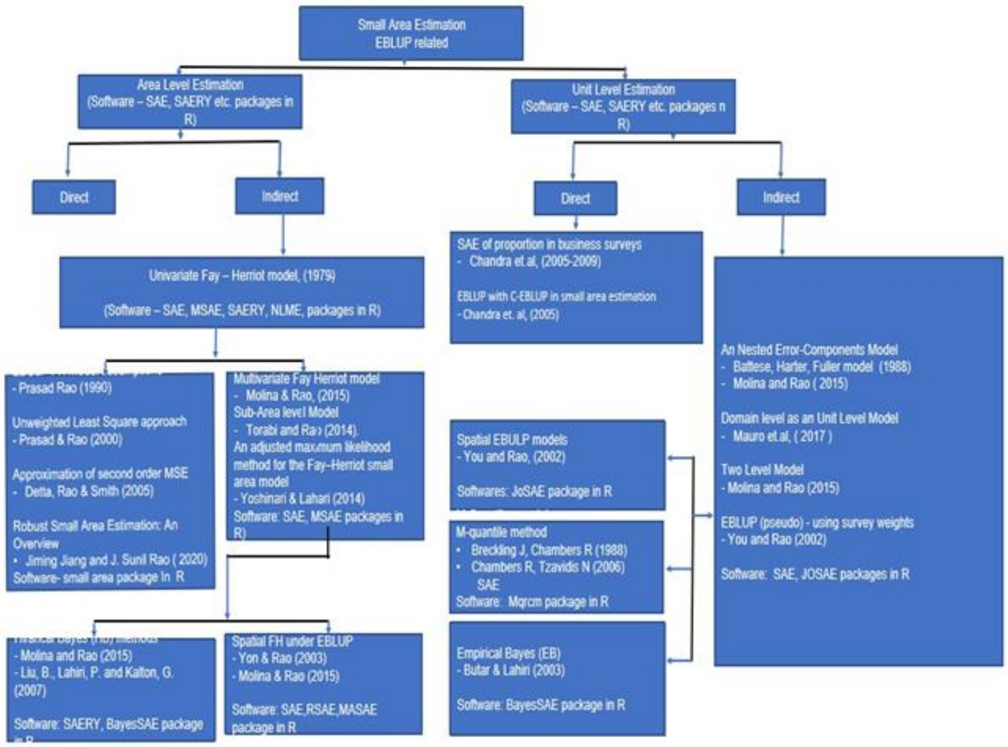
[4,5,6]*Department of Statistics, S.G.S Arts College/ S.V University, Tirupati.*

I. INTRODUCTION-

Small area estimation (SAE) is a survey sampling technique useful to estimate the population counts for small areas (domains or subdomains) or regions from the large-scale surveys or census. The data received for SAEs are borough strength from secondary data (surveys), followed by auxiliary data (administrative records) to obtain reliable estimates. Small area estimates procedures can be used as a ready to fill the gaps between official statistics and local municipalities for data requests for immediate policy making or setting up a business based on population counts. The goal of SAE is to estimate a parameter (counts, averages, rates, ratios) with a minimum of the standard errors and main summary statistics is the average square error (MSE).

Small area estimation or statistics is a widely used technology in survey modeling, due to the growing demand from reliable small area estimates from public and private sectors to find spatial segmentation and obtain domains, sub-domains, regional or national level estimates. The small area models are generally having two types of cases.i.e., general linear models and regression-based modeling. Small area models are first evaluated from Horvitz-Thompson estimator and then used ratio estimator. i.e, they are also called design-based models. The design-based model combined with weights, then it forms generalized regression estimator (GREG) also called model-assisted estimators. Further, GREG estimator was linked with auxiliary data to form regression synthetic estimator. Later, the synthetic-regression estimator was used in unit based and area-based models for further research. The below flow chart gives brief note on small area estimation methodology.





II. METHODOLOGY

Objective-1: The study aimed to evaluate reliable small area direct method estimates and compared with Direct estimates. In this paper, it contains information regarding direct methods of calculating means in small area estimation and different approaches of variance methods. Data wrangling, management, analysis using R for Horvitz-Thompson estimator with its interpretation. Producing state level fact sheets for mean age at marriage, mean age at first birth and represented in error bars using ggplot package. Producing district level fact sheets for mean age at marriage, mean age at first birth and plotting that estimates on map using ggplot package.

Horvitz-Thompson estimator-

The quantity of interest be denoted by Yd for domain d of the population domains could be defined by states or by industries with a certain set of codes. Each sampled unit j has an associated sample weight, denoted wj, which is equal to the inverse of a unit's probability of being selected.

yd=Σwjyj

Variance method-

The Horvitz-Thompson estimator of mean was proposed by Hansen, Horvitz, madow:

$$\frac{\Lambda}{y} = \frac{1}{N}\sum_{J\varepsilon s} y_i \;/\; \Pi_i$$

Where i is the distinct number of units in the sample. The Horvitz-Thompson estimator does not depend on the number of times a unit may be selected. Each distinct unit of the sample is utilized only once.

Variance.

$$v\left(\frac{\Lambda}{y}\right)=\frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N}(\pi I,k \;-\; \pi I \pi k)y_I \;/\; \Pi_i y_k \;/\; \pi_k$$

Data-

small area methods can be adoptable taking account of secondary data. So, we consider National Family and Household Survey Dataset (NFHS).

III. DATA MANAGEMENT-

State Level-

First, we consider the married women (15-49 year) dataset from NFHS survey

- The dataset is in the format of SPSS and we import that dataset into R by using FOREIGN package.
- Then we clean the dataset with plyr, dplyr ,hmisc,janitor,tidyr packages.
- Next, we go through insights and did glimpse and summary statistics.
- Then, we choose the variables for our concern to fulfill the objective.
- Further, we create data frame for states with respondent age at marriage (v118) and respondent age at first birth (v121) from married women dataset.
- Here we include districts in that state to get outcome.
- Later, we calculate sample weights using survey data and samples they chosen from state population. In this dataset the weight they chosen is around 1%.
- Further, we make a data frame with response vector (age at marriage, age at first birth, sample weight).
- We adopt Horvitz- Thompson direct method of estimation to calculate mean age at marriage and mean age at first birth, with respect to variances.
- Finally, we create an error bar for that state level means and variances using ggplot2 package.

IV. DISTRICT LEVEl

- Same like state level data management process, we choose individual districts in those states.
- We consider the required aspects and do summary statistics, glimpse.
- Further, we create data frame for individual districts with respect to states, respondent age at marriage (v118) and respondent age at first birth (v121) from married women dataset.
- In this context, we include Primary Sampling Unit in the Districts to get outcome.
- Later, we calculate sample weights using survey data and samples they had chosen from state population. In this dataset the weight they chosen is around 1%.
- Further, we make a data frame with response vector (age at marriage, age at first birth, sample weight) for district and state.
- We adopt Horvitz- Thompson direct method of estimation to calculate mean age at marriage and mean age at first birth, with respect to variances.
- Finally, we choose administrative map at state that include district map and plotted the outcomes (means) using ggplot2 package.

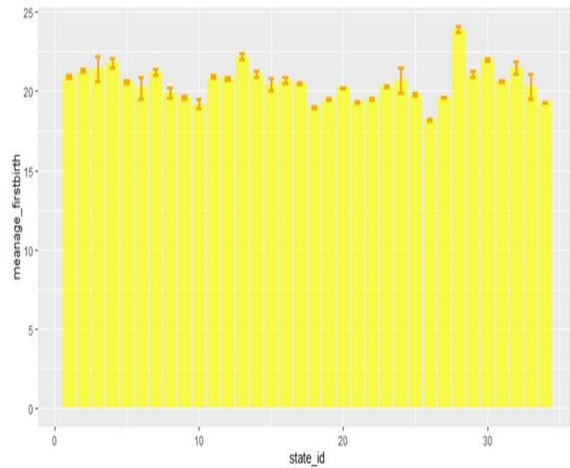
V. RESULTS

state	Mean age at marriage	variance	Mean age at first birth	variance
Jammu Kashmir	19.2	0.1	20.9	0.1
Himachal Pradesh	19.6	0.01	21.3	0.1
Punjab	19.8	0.1	21.4	0.8
Chandigarh	20.3	0.1	21.8	0.3
utharakhand	18.6	0.1	20.6	0.1
Haryana	17.5	0.05	20.2	0.7
Delhi	18.9	0.01	21.2	0.2
Rajasthan	15.6	0.06	19.9	0.3
Uttar Pradesh	15.7	0.06	19.6	0.1
Sikkim	19.6	0.03	20.9	0.3
Bihar	15.2	0.1	19.2	0.1
Arunachal Pradesh	19.4	0.1	20.8	0.1
Manipur	21.3	0.01	22.2	0.2
Mizoram	20.5	0.02	21.1	0.2
Tripura	18.4	0.03	20.4	0.4
Meghalaya	19.6	0.02	20.7	0.
Assam	19.2	0.04	20.5	0.06

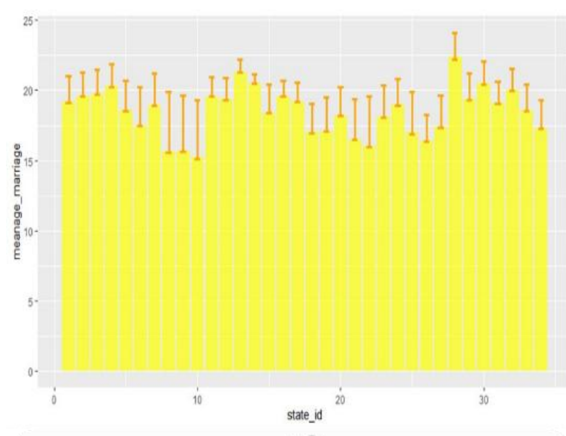
state	Mean age at marriage	variance
West Bengal	17	0.04
Jharkhand	17.1	0.03
Orissa	18.2	9.04
Chattisgarh	16	0.05
Madhya Pradesh	16.5	0.04
Gujrat	18.1	0.04
Daman and diu	19	0.1
Dadar nagar havelli	17	0.1
Andhra Pradesh	16.4	0.04
Karnataka	17.4	0.04
Goa	22.4	0.2
Lakshadweep	19.4	0.1
Kerala	205	0.1
Tamil nadu	19.1	0.04
Pondicherry	20	0.03
Andaman nicobar islands	18.6	0.1
maharastra	17.5	0.03

Mean age at marriage-
We use Hansen- Horvitz (1953) variance estimator to calculate means and variance our variable of interest. The above shows explained that early marriage that take place states in India are Rajasthan, Uttar Pradesh, Madhya Pradesh, Bihar, and mean age of marriages are between 15 to 16 years. Here the variances for those states have not much differ within that region. This shows that the data are more tightly grouped around the average.

Mean age at first birth-
In another context, the late marriages are taken place in goa, Manipur, Pondicherry, Kerala and mean age of marriages are between 20 to 22 years. Here the variances for those states have not much differ within that region. This shows that the data are more tightly grouped around the average. In another side of the table, the early childbearing is taking place most of the states in India like Bihar, Rajasthan, Uttar Pradesh, west Bengal, Chhattisgarh, Andhra Pradesh. and mean age of marriages are between 18 to 20 years. Here the variances for those states have not much differ within that region. This shows that the data are more tightly grouped around the average.



Mean age at marriage-
In this error bar plot shows, the data from goa is mostly differ from all the other regions in india. So, this conclude majority of the data that collected are significantly differ from another. The data from Bihar has shown more early marriages than another regions from India, so this is not conclusive which state have more earlier marriages.



Mean age at first birth-
In the above error bar plot, age at first birth in goa and compared to another states, this shows most of the data collected are significantly differ from another regions. The early bearing started in state of Andhra Pradesh. This shows most of data in that region are differ from another, but it is inconclusive.
In this section we would like to explain about,district level data interpretation using Horvitz-Thompson direct estimates for means.
Mapping of district means on state maps using shape files.

VI. DISTRICT LEVEL

Same as state level data management process, we choose individual districts in those states. We consider the required aspects and do summary statistics, glimpse. Further, we create data frame for individual districts with respect to states, respondent age at marriage (v118) and respondent age at first birth (v121) from married women dataset.

In this context, we include Primary Sampling Unit in the Districts to get outcome.

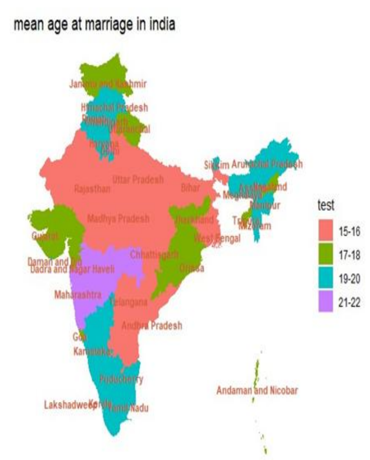
Later, we calculate sample weights using survey data and samples they had chosen from state population. In this dataset the weight they chosen is around 1%.

Further, we make a data frame with response vector (age at marriage, age at first birth, sample weight) for district and state.

We adopt Horvitz- Thompson direct method of estimation to calculate mean age at marriage and mean age at first birth, with respect to variances.

Finally, we choose administrative map at state that include district map and plotted the outcomes (means) using ggplot2 package.

District level mapping of means in states-we use mapping to visualize district level means for that particular states.In this context, we would like to show india map for mean age at marriage, early marriage district and late marriage district.



Objective-2: The study seeks to find out the demographic factors like mean age at marriage and first birth for particular region having UNsampled domain with the combination of auxiliary data. In this chapter, information regarding indirect method of estimating means in small area estimation and also finding their respective mean square error.choosing appropriate indirect methodize., FAY - HERRIOT Model to calculate means and model to be best fit. Interpretation of missing domain level estimates in different regions across india using auxiliary data. producing district level tehsils /CD Blocks fact sheets for mean age at marriage, mean age at first birth and represented in error bars using ggplot2 package.

VII. PROCEDURE-

Direct estimates for sampled domains have to be calculated first by using appropriate small area direct methods. i.e., Horvitz-Thompson estimator/Ratio estimator.

The results that were obtained are in area-level.i.e., domain level.

VIII. FAY-HERRIOT MODEL METHODOLOGY-

Fay-Herriot model is underestimating the mean square error due to unknown or measured error of population quantity and using a non-optimal weighting: the weighting yiv for the contribution of direct estimator yiv is same whether xi is known or estimated. This will increase an error.

This model will help to reduce the bias in the model.

It can be represented as xi is true value for the auxiliary variables for area I, if all components of xi were known we would use the model, with representing the model error and representing the design-based survey error for yi. Since xi maybe measured with error, we substitute and estimator and use instead.

Then, the resulting variable of interest are obtained with the combination of mean square error.

Data-

domain selection-

Tehsils selection is done under the percentage of missing domain information on NFHS-3 Dataset.

domain data coverage in NFHS Dataset

Tehsil name	STATE	Percentage of domain data covered
Kangra	Himachal Pradesh	65%

This district has borders with Punjab, Haryana, Jammu and dharmsala (buddha traditional place).in 2011, the population of Kangra is around 16 lakh s, where male

and female population are almost same with 340k households.
This district has average of 80% literacy rate. The population are distributed evenly with same for both rural and urban areas. There are 21 tehsils /taluk/CD Block in this district.

variable of interest selection-
NFHS-3 data set was adopted. This dataset has variety of compositional sets like household, married women, unmarried women datasets. Married women dataset having women's age while she is married and first birth variables are considered.

auxiliary information-
The variables from socioeconomic survey like Female literacy rate, Per-Capital income for a house hold, Incidence in Poverty have been chosen, and statistical Association tests has done with corresponding variable of interest.
association tests for auxiliary data

Auxiliary variables	Mean age at marriage	Mean age at first birth
Female literacy rate	0.42	0.7
Incidence poverty rate	-0.2	0.05
Per- capital income	0.07	0.2

Synthetic data set creation-
Synthetic Data means Artificially created data with real time attributes and helps to get the information where the data is not readily available, having some restrictions

Software-
R Language have several functions to deal with SAE Methodology. The package saeME provide a brief introduction and functions with regards to small area methodology under the data measured with error. In this paper we use Horowitz-Thompson function from mase package and FH_me function from saeME package.

TEHSIL	=	THE
Mean age at marriage	=	MAM
Mean age at first birth	=	MAFB
Variance for mean age at marriage	=	VMAM
Female literacy rate	=	FLR
Poverty rate	=	PR
Per capita income	=	PCI
Mean square error for female literacy rate	=	MSEFLR
Mean square error for per -capita income	=	MSECI

THE	MAM	MAFB	VMAM	FLR	PR	PCI	MSEFLR	MSECI
palmpur	20.3	22.1	0.2	80.1	25.9	15451	10	125.1
nurpur	19.6	21	0.2	80.7	17.5	13084	10	125.1
dharmasala	19.8	21.2	0.2	80.1	20/7	23844	10	1251
Dera gopipur	20.9	21.4	0.2	83.1	28.6	24100	10	125.1
jawali	20.4	21.3	0.2	80.9	24.7	14590	10	125.1
kangra	19.7	21.2	0.2	82.6	23.4	17486	10	125.1
bajinath	19.4	21	0.2	85.1	31.8	17591	10	125.1
indora	20	21.4	0.2	79	18.6	13295	10	125.1
Nagrota baghwan	19.2	21	0.2	79.2	27.7	19126	10	125.1
shahpur	20.1	22.5	0.2	68.6	25.9	14410	10	125.1
fatehpur	20.5	22.5	0.2	52.6	24.8	16280	10	125.1
jaisinghpur	NA	NA	NA	50	25.2	15550	10	125.1
jwalamukhi	19.2	21.3	0.2	83	25.8	16972	10	125.1
khundion	20.7	21.8	0.2	75.2	23.8	14905	10	125.1
jaswan	NA	NA	NA	82.6	24.1	15451	10	125.1
rakkar	19.7	21.9	0.2	84.1	19.8	16465	10	125.1
baroh	19.7	21.8	0.2	68.9	26.7	15550	10	125.1
dhira	19.4	21.3	0.2	72.6	24.8	17310	10	125.1
thural	NA	NA	NA	71.9	30.6	16985	10	125.1
harchkian	NA	NA	NA	78.4	27.7	17010	10	125.1
multhan	NA	NA	NA	46.4	23.1	17171	10	125.1

IX. RESULTS- MEAN AGE AT MARRIAGE

Number	Representation
HT	Horowitz-Thompson Direct survey estimates
A	Mean age at marriage/first birth with female literacy rate
B	Mean age at marriage/first birth with incidence of poverty rate
C	Mean age at marriage/first birth with per-capital income
D	Mean age at marriage/first birth with female literacy and incidence poverty rate
E	Mean age at marriage/first birth with female literacy and per-capital income
F	Mean age at marriage/first birth with incidence poverty rate and per-capital income
G	Mean age at marriage/first birth with female literacy rate, ncidence of poverty rate and per-capital income

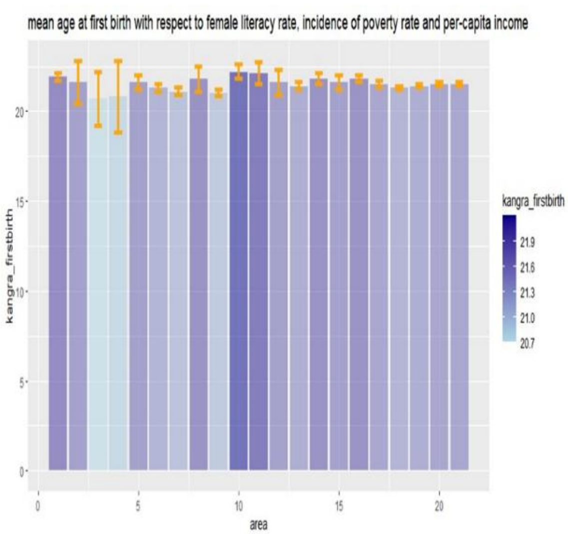
Area	HT	A	B	C	D	E	F	G
palampur	20.3	20.1	20.1	20.4	20.2	20	20.4	20.1
nurpur	19.6	19.6	19.6	19.6	19.4	19.2	19.1	19.5
dharmasala	19.8	19.8	19.8	19.8	19.7	19.8	19.8	19.7
Dera gopipur	20.9	20.5	20.2	20.5	20.1	20.5	20.5	20.5
jawali	20.4	20.2	20.4	20.2	20.2	20.2	20.2	20.2
kangra	19.7	19.7	19.7	19.7	19.7	19.7	19.6	19.1
bajinath	19.4	19.4	19.4	19.4	19.2	19.4	19	19.4
indora	20	20	20	20	20	20	20	20
Nagrota baghwan	19.2	19.4	19.1	19.9	19.2	19.2	19.2	19.2
shahpur	20.1	20	20	20	20.1	20.2	20.5	20
fatehpur	20.5	20.2	20.2	20.2	20.1	20.1	20.3	20
Jaisinghpur*	NA	19	20.1	20	20	19.5	20	19.2
jwalamukhi	19.2	19.4	19.2	20	20	19.1	19	19.1
khundion	20.7	20.4	20.5	201	20	20.2	20.5	20
Jaswan*	NA	19.9	19.9	20	20.1	20.1	20	20.2
rakkar	19.7	19.7	19.7	19.7	19.7	19.7	19.7	19.7
baroh	19.7	19.7	19.7	19.7	19.7	19.7	19.7	19.7
dhira	19.4	19.5	19.4	19.1	19	19.3	19.5	19.2
Thural*	NA	19.8	19.9	19.6	20	19.5	19.8	19
Marchkian*	NA	19.8	19.8	19.6	19.8	19.7	19.7	19.4
Multhan*	NA	19.9	20	19.6	20	19.1	19.6	19.9

MEAN AGE AT FIRST BIRTH-

Area	HT	A	B	C	D	E	F	G
palampur	22.1	21.9	21.5	21.9	22	21.9	21.9	21.9
nurpur	21	21.2	21.4	21.5	21.4	21.2	21.1	21.6
dharmasala	21.2	21.3	21.4	20.7	20.7	21.6	21.3	20.7
Dera gopipur	21.4	21.5	21.5	20.8	20.9	21.5	21.6	20.8
jawali	21.3	21.4	21.5	21.6	21.5	21.2	21.4	21.6
kangra	21.2	21.4	21.5	21.2	21.2	21.3	21.4	21.3
bajinath	21	21.3	21.6	21.1	21.1	20.8	21.4	21.1
indora	21.4	21.4	21.4	21.1	21.7	21.5	21.4	21.8
Nagrota baghwan	21	21.2	21.5	21	21.1	21	21.2	21
shahpur	22.5	21.9	21.5	22.3	22.3	22.2	21.9	22.2
fatehpur	22.5	21.6	21.5	22.1	22.1	22.3	21.6	22.1
Jaisinghpur*	NA	21.5	21.5	21.6	21	21.3	21.5	21.6
jwalamukhi	21.3	21.5	21.5	21.5	21.4	21.4	21.4	21.4
khundion	21.8	21.6	21.6	21.6	21.8	21.8	21.8	21.8
Jaswan*	NA	21.5	21.5	21.6	21.7	21.6	21.5	21.6
rakkar	21.9	21.5	21.8	21.8	21.8	21.8	21.8	21.8
baroh	21.8	21.5	21.5	21.5	21.5	21.5	21.5	21.5
dhira	21.3	21.3	21.3	21.3	21.3	21.3	21.3	21.3
Thural*	NA	21.6	21.6	21.5	21.4	21.8	21.8	21.4
Marchkian*	NA	21.3	21.5	21.5	21.5	20.7	21.4	21.5
Multhan*	NA	21.3	21.5	21.4	21.8	22	21.3	21.5

- Error bar plots for mean age at marriage with respect to Female literacy rate, incidence of poverty rate, per-capital income.
- Mean age at marriage table, dera gopipur is recorded 20.9 years and by using fay-herriot model with the combination of direct estimates (response variable) with poverty and female literacy (Auxiliary variable). i.e., 20.1 years.
- The data obtained for missing domains or unsampled areas are having better precision with very less mean square error.
- In error bar plot, the data from jaisinghpur is not tightly packed, most of the data that were choosen . so, the majority of the data are most significant than other domains.

Area	A	B	C	D
palampur	0.3	0.1	0.3	0.2
nurpur	0.2	0.1	1.2	0.3
dharmasala	0.1	0.2	2	0.1
Dera gopipur	0.1	0.1	2.4	0.1
jawali	0.1	0.3	0.5	0.1
kangra	0.1	0.3	0.2	0.1
bajinath	0.2	0.8	0.2	0.5
indora	0.1	0.1	0.8	0.5
Nagrota baghwan	0.2	0.8	0.3	0.2
shahpur	0.5	0.9	0.4	0.7
fatehpur	0.4	0.1	0.5	1.1
Jaisinghpur*	0.1	0.3	0.2	0.1
jwalamukhi	0.1	0.3	0.2	0.2
khundion	0.2	0.1	0.3	0.3
Jaswan*	0.1	0.2	0.1	0.1
rakkar	0.2	0.2	0.2	0.2
baroh	0.2	0.1	0.2	0.2
dhira	0.1	0.1	0.1	0.1
Thural*	0.1	0.7	0.1	2.1
Marchkian*	0.1	0.1	0.1	1.1
Multhan*	0.1	0.2	0.1	0.8



- Mean age at first birth table, fatehpur is recorded 22.5 years and by using fayherriot model with the combination of direct estimates (response variable) with per-capita income (Auxiliary variable). i.e., 21.5 years.
- The data obtained for missing domains or unsampled areas are having better precision with very less mean square error.
- In error bar plot, the data from dera gopipur is not tightly packed, most of the data that were choosen so, the majority of the data are most significant than other domains.

objective-3: assessing an appropriate small area logistic model for calculating proportions for domains having zero sample sizes.

objective-4: asessing an appropriate small area poisson regression model for calculating counts for domains having zero sample sizes.

- in this section, this contains information regarding small area methods for calculating proportions.
- data selection, management, software

small area poisson regression model-

- The response variable has a Poisson distribution, and assumes the logarithm of its expected value can be modeled by a linear combination of unknown parameters.
- It is also known as log-linear model.
- It is used to predict a dependent variable that consists of count data given one or more independent variables.

Metarnal mortality-

- In India , Metarnal Mortality definition shows that “The death of a woman during pregnancy, at delivery, or soon after delivery”.
- The number of resident maternal deaths within 42 days of pregnancy termination due to complications of pregnancy, childbirth, and the puerperium in a specified geographic area (country, state, county, etc.) divided by total resident live births for the same geographic area for a specified time period, usually a calendar year, multiplied by 100,000.

Logistic regression-

- In some cases, a discrete response variable takes only two possible values, For example, the response variable takes one of only two possible values representing “success” or “failure” , For such cases, the logistic regression model is appropriate.
- Application of logistic regression model in SAE implies that the estimated probability of occurrence of an event in the ith small area and with respect to auxiliary variable.

- The small area popular models like EBLUP are appropriate while deealing this type of non-linear data.
- Dependent variable: - Metarnal Mortality Rate
- Independent variable➢Bathroom availability ➢Drinking water

➢Fuel used for cooking
➢Female Literacy Rate
➢Poverty Rates
➢Mean Age at Marriage
➢Mean Age at First birth
➢Nutrition status
➢Infant Mortality rate
➢Average Females
➢Average Househhold Size
➢Type of House
➢Induced Abortions
➢Pregnancy test
➢Pregnancy Problem-Foetus and High bp
➢Pregnancy Problem-Malaria
➢Delivery Problem-High cost
➢Delivery Problem-Transportation
➢Abortion month
➢Women birth order

X. SOFTWARE-

1) logistic regression method- small area methods for calculating proportions are using Logistic Regression method.

2)Poisson Regression Method- small area methods for calculating Counts using poisson regression method.
Steps involved in calculating the proportions using Logistic Regression model-

- First, consider the variable of interest (Metarnal Mortality Rates direct estimates).
- The State Maternal Mortality in Chhattisgarh for year 2012 is. so, we make it as thereshold.
- we gave coding according to direct estimates. if the value is above. i.e, 1 or else 0.
- we consider different auxiliary data and also created a complex type synthetic dataset.
- The results come in form of proportions.
- we also checked the residual, new predicted threshold, error.

- miss class error-0.12

The error for an initial and predicted are 10%, we can use small area estimates.

- thershold-0.63

The actual thershold for the fit of model is 0.63, so, only 3 districts are having significant profile from another district. the initial thershold is 23.

- sensitivity-0.6specificity-1so, we consider original data Metarnal mortality ratio and fitted a poisson regression model to predict the estimates.
- The above table shows that, most of the districts of direct estimates and predicted estimates are bit differ.
- we observed that, dantewada , damatri shows significant level of difference from the direct estimates. the overall error is 15%
- we noticed that, the danntewada, bastar, damatri are more remote areas, the health workers like asha (ANM) are not much available.This regions have more below poverty people in the state.

	direct estimates Metarnal Mortality Rate	coding according to national estimates	logistic regression estimates coding	predicted proportions	Standard error
	27	1	1	0.89	0.21
	25	1	0	0.17	0.23
	17	0	0	0.06	0.16
	30	1	0	0.47	0.38
	25	1	1	0.76	0.28
a	15	0	0	0.20	0.31
	14	0	0	0.02	0.07
	26	1	1	0.92	0.10
	21	0	0	0.00	0.00
	19	0	0	0.02	0.05
	15	0	0	0.17	0.21
	11	0	0	0.00	0.03
	10	0	0	0.08	0.17
	16	0	0	0.01	0.05
	22	0	0	0.63	0.40
	22	0	0	0.55	0.43

Summary and conclusions: Why We Need Small Estimates Than Direct Estimates-

If n<100, Small area methods are more appropriate than direct estimates. Small area estimation refers to methods of producing sufficiently reliable estimates for geographic areas that are too fine to obtain with precision. Small Area Estimation is statistical technique which is used to provides an analytical framework for improving the level of granularity without necessarily collecting additional data in the field. The data adopted for SAE's are secondary data(surveys) then barrow strength from auxiliary data (administrative records) to get reliable estimates. Advances in data processing capabilities and computational algorithms have long ago made it possible to apply impossible estimation methods. SAEs used to describe relation between explonatory variable and predictor variable.

- The reason behind adopting SAE's techniques is because of poor performance of traditional demographic sample surveys (simple random sampling, systematic sampling etc..) designed for national estimates those which do not provide large enough samples to produce reliable direct estimates for small areas.
- The gradually bit increases sample size and produce better estimates.
- small area estimation handles perfectly for planned and unplanned domain data.
- The direct estimates are constructed based on the survey data and design weights for the area of interest. if the sizes are too finite, this leads to huge variability. whereas in small area estimates uses direct and auxiliary data. they will produce better estimates even for too finer to intreprit.
- Small area can deal the problem that the areas are not computable due to zero sample sizes.
- SAE helps us dealing problems like sampled areas having less responses or we can determine unsampled domains information too, both for planned and unplanned areas.

REFERRENCES

[1] R. Schell, "Estimation in generalized linear models with random effects," Biometrika, vol. 78, pp. 719–727, 1991.

[2] C.-E. Särndal, B. Swensson, and J. H. Wretman, Model Assisted Survey Sampling. New York, NY, USA: Springer-Verlag, 1992.

[3] G. E. Battese, R. M. Harter, and W. A. Fuller, "An error-components model for prediction of county crop areas using survey and satellite data," Journal of the American Statistical Association, vol. 83, pp. 28–36, 1988.

[4] W. Dumouchel, "A Bayesian model and a graphical elicitation procedure for multiple comparisons," in Bayesian Statistics 3, J. M. Bernardo et al., Eds. Oxford, U.K.: Oxford University Press, pp. 127–166, 1987.

[5] R. E. Fay and R. A. Herriot, "Estimates of income for small places: An application of James–Stein procedures to census data," Journal of the American Statistical Association, vol. 74, pp. 269–277, 1979.

[6] "Small area estimation," Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Small_area_estimation

[7] D. Drew, M. P. Singh, and G. H. Choudhry, "Evaluation of small area estimation techniques for the Canadian labor force survey," Survey Methodology, vol. 8, pp. 17–47, 1982.

[8] N. G. N. Prasad and J. N. K. Rao, "On the estimation of mean square error of small area estimators," submitted to Journal of the American Statistical Association, 1987.

[9] J. Foster, J. Greer, and E. Thorbecke, "A class of decomposable poverty measures," Econometrica, vol. 52, pp. 761–766, 1984.

I. Molina and Y. Marhuenda, "sae: An R package for small area estimation," The R Journal, vol. 7, 2015.

[10] M. L. Cohen and E. Sinha, National Patterns of R&D Resources: Future Directions for Content and Methods, 2013.

[11] M. Martinez Jr., "Small area estimation and big data," in Proc. Int. Workshop on SDG Data Disaggregation, Asian Development Bank, Jan. 2019.

[12] S. Arima, G. S. Datta, and B. Liseo, "Accounting for measurement error in covariates in SAE: An overview," in Analysis of Poverty Data by Small Area Methods, M. Pratesi, Ed. Hoboken, NJ, USA: Wiley, 2015.

[13] Baíllo and I. Molina, "Mean-squared errors of small-area estimators under a unit-level multivariate model," Statistics, vol. 43, pp. 553–569, 2009.

[14] M. Bankier, "Power allocation: Determining sample sizes for sub-national areas," The American Statistician, vol. 42, pp. 174–177, 1988.

[15] W. R. Bell, "Accounting for uncertainty about variances in small area estimation," Bulletin of the International Statistical Institute, 1999.

[16] S. Chen, P. Lahiri, and J. N. K. Rao, "Robust mean squared prediction error estimators of

EBLUP under the Fay–Herriot model," unpublished manuscript, 2008.

[17] G. S. Datta, "Model-based approach to small area estimation," in Sample Surveys: Inference and Analysis, vol. 29B, D. Pfeffermann and C. R. Rao, Eds. Amsterdam, Netherlands: North-Holland, pp. 251–288, 2009.

[18] G. S. Datta, J. N. K. Rao, and D. D. Smith, "On measuring the variability of small area estimators under a basic area level model," Biometrika, vol. 92, pp. 183–196, 2005.

[19] G. S. Datta, T. Kubokawa, I. Molina, and J. N. K. Rao, "Estimation of mean squared error of model-based small area estimators," TEST, vol. 20, pp. 367–388, 2011.

[20] R. E. Fay, "Application of multivariate regression to small domain estimation," in Small Area Statistics, R. Platek et al., Eds. New York, NY, USA: Wiley, pp. 91–102, 1987.

[21] R. E. Fay and R. A. Herriot, "Estimation of income from small places: An application of James–Stein procedures to census data," Journal of the American Statistical Association, vol. 74, pp. 269–277, 1979.