



Web-Based Software for Small Area Estimation under Unit Level Model

James Karangwa¹, Anshu Bharadwaj²

¹Electrical and Electronics Engineering, IPRC Gishari-Integrated Polytechnic Regional College, Rwamagana, Rwanda

²Computer Application, ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India

Email: jameskarangwa@gmail.com, Anshu.Bharadwaj@icar.gov.in

How to cite this paper: Karangwa, J. and Bharadwaj, A. (2021) Web-Based Software for Small Area Estimation under Unit Level Model. *Open Access Library Journal*, 8: e7207.

<https://doi.org/10.4236/oalib.1107207>

Received: February 2, 2021

Accepted: April 12, 2021

Published: April 15, 2021

Copyright © 2021 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The purpose of this paper is to produce reasonably accurate direct estimators, not only for the characteristics of whole Population but also for a variety of Subpopulations or domains. Many policymakers and researchers also want to obtain statistics for small domains. These small domains are also called small areas, because the sample size in the area or domain from the Survey is small. Due to small sample size, domain-specific direct estimators provide an acceptably 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 this increases 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. In this paper, a web-based software for small area estimation Under unit level model has been developed But this doesn't include the Unit level effects in the model. This software will help the researchers, academicians, data analysts, Students and other domain groups who have been working in area of the SAE.

Subject Areas

Computer Application and Statistics

Keywords

Small Area Estimation, R Studio, Unit Level Model

1. Introduction

As this is the age of information technology, the data analysis, estimation of various statistics are being done using software specifically developed for partic-

ular method. SAE is a potential area of survey sampling, but no information could be found regarding any existing standard software either stand-alone or web based for small area estimation with unit level model. Few efforts have been made to develop software for SAE under Sample Survey Resource, being developed at ICAR-IASRI. This particular software has been developed for small area estimation for basic area level model. SAE, being a very important area of survey sample, development of a software for Small Area Estimation under Unit Level Model is proposed here to be taken up in this research work. The work resulted in a web-based Tool for Small Area Estimation under Unit Level Model. This software will cater to the researchers, academicians, data analysts, students and other domain groups who have been working in the area of SAE [1] [2].

2. Material and Method

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. Small Area typically denotes a subset of the population for which very little information is available from the sample survey [3]. The statistics related to these small areas are often termed as small area statistics. Due to the increasing demand, survey organizations are faced with producing the small area estimates from existing sample surveys. Unfortunately, sample sizes in small areas tend to be too small, sometimes non-existent, to provide domains specific reliable direct estimates for these small areas. Small area estimation is important in survey analysis when domain (subpopulation) sample sizes are too small to provide adequate precision for direct domain estimators. Popular techniques for small area estimation use implicit or explicit statistical models to indirectly estimate the small area parameters of Interest [4]. Indirect estimation for small areas uses statistical models and auxiliary variables to borrow strength from similar area.

2.1. Material: Unit Level Model

Unit-level models relate the unit values of a study variable to unit-specific auxiliary data. For example, suppose one has a survey of firms that are designed to estimate total wages and salaries paid to workers. Perhaps the survey is designed so that estimates of a specified degree of precision can be made at the state level [5]. After the survey is conducted, one decides to estimate total wages and salaries by industry, but the sample sizes for some industries are so small that the variances of the estimates are unacceptably large. To improve the precision of the estimates, one can use auxiliary data such as firm-level values of gross business income to fit a linear mixed model with industry-specific random effects to improve the efficiency of the estimates [5] [6].

2.2. Method: Software Development Methodology

The system is built using 3-tier architecture (**Figure 1**):

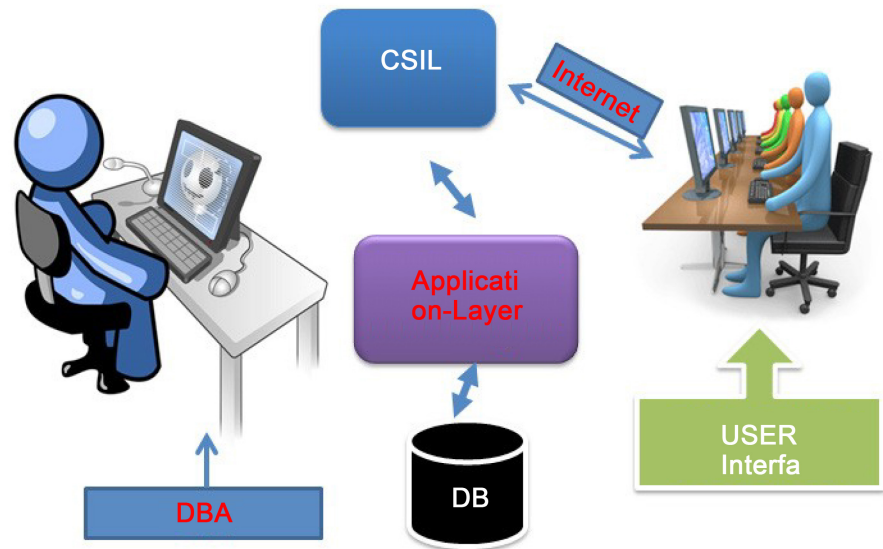


Figure 1. Three tire architecture [7].

- Client Side Interface Layer (CSIL).
- Server Side Application Layer (SSAL).
- Database Layer (DBL).

3. Programming in R STUDIO

R is an open-source environment for data manipulation, statistical analysis, and visualization [8] [9]. It is most convenient to run R within an integrated development environment (IDE), e.g., R Studio. R Studio is a free and open source integrated development environment (IDE) for R. Programming language for statistical computing and graphics. Here, coding and programming are done in R Studio (version 3.3.3) for Small Area Estimation under Unit Level Model (R Package SAE version 1.1). The package includes small area estimation methods based on the basic unit level model called the nested-error linear regression model.

4. JSP AND R Integration

Integration of R Software with JSP by using NetBeans IDE has been done by adding jar files into the libraries [10].

5. Result

In this study Improved Crop Statistics (ICS) and Census Data at District level have been used to find out the small area estimates. During the year 2009-10 in the State of Uttar Pradesh there were 70 districts. However, supervision on a sub-sample of Crop Cutting Experiments (CCEs) work under ICS scheme was carried out in 58 districts only and there is no sample data for the remaining 12 districts. These 12 districts are referred as the out of sample districts. These 70 (58 in sample and 12 out of sample) districts are the small areas for which esti-

mates are produced. In this study, 50 districts are considered as sampled districts for which Small Area Estimation under Unit Level Model has been calculated using R SAE Package. Unit-level models relate the unit values of a study variable to unit-specific auxiliary data. The design of the software presents a Home Page with a Login Panel where Registered users can log in with their authenticated User Id and password. The new users can register from the “Register Here” link below the panel. The Homepage of the software is shown in **Figure 2**.

New client can register through the “Register Here” link given above on the panel. New user have to create their own profile by fulfilling the required fields given on the registration panel like Username, password etc. After successful registration the client or new user will get a message and the username and password will be validated. Then the client can log in from “Go to Login” link through the homepage using valid credentials. After Login client will enter into the software. The client should select and upload a file that may be in text, pdf, csv, xls, ppt or any other image file format (**Figure 3**).

The file should contain the small area estimation data with some missing data values or not depends on which sample you are dealing with (**Figure 4**). The Analysis architecture of the software by clicking the run “Run R script for Analysis”. After the selection of required INPUT you have to click the query for further analysis in order to get the result or the output. The user can get the description of the software by clicking on “ABOUT SOFTWARE” on the links above.



Figure 2. Home page of the software [8].

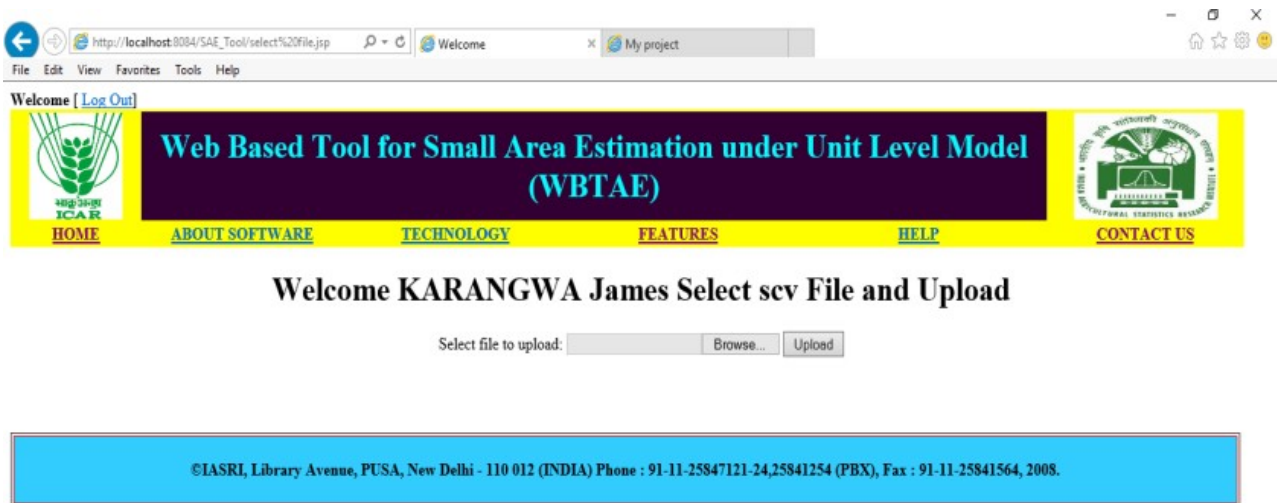


Figure 3. File selection and upload [9] [10].

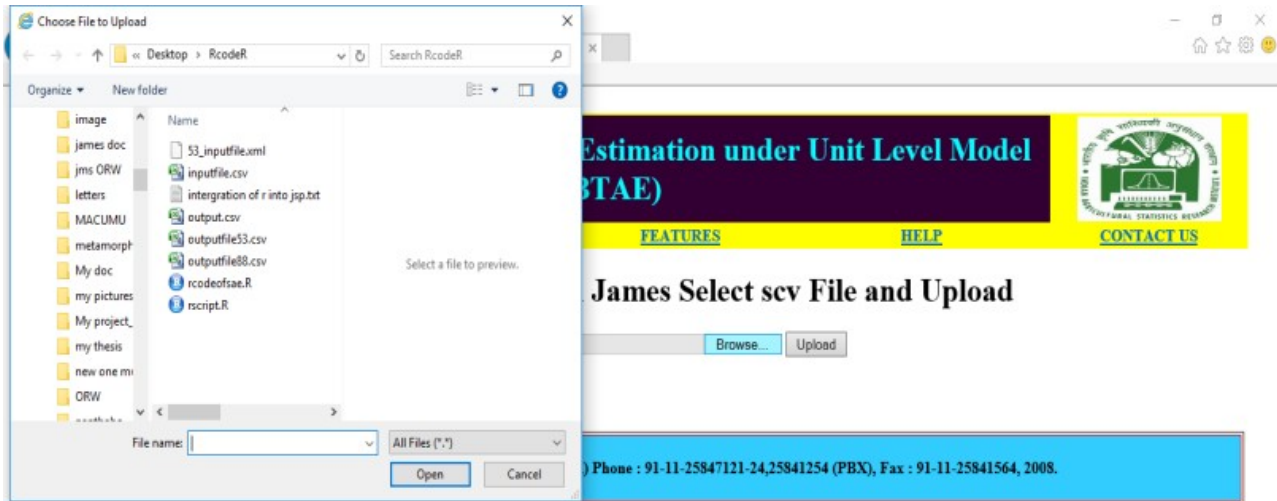


Figure 4. Selected file [9] [11].

The technologies like that have been used to design and develop the software like: HTML, JSP, and MY SQL are described in the link “TECHNOLOGY”. The Features of the software can be seen by clicking the link “FEATURES” on the home page of the software. From the “HELP” link user can get the information about how to use the software.

6. Conclusion

Small Area is 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. Unit level models relate the unit values of a study variable to unit-specific auxiliary data. Many software (web-based or standalone) have been developed for survey data analysis but there have been no software to analyze the data obtained from small area under unit level model. Therefore, this present web-based software has been developed. It has been developed using three-tier architecture. For the testing and validation, the software has used the small area estimation data from ICD and Census Data of Uttar Pradesh. The software is user friendly and will be beneficial to the researchers, students etc. working with small area estimation under unit level model.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Molina, I. and Rao, J.N.K. (2010) Small Area Estimation of Poverty Indicators. *The Canadian Journal of Statistics*, **38**, 369-385. <https://doi.org/10.1002/cjs.10051>
- [2] Prasad, N.G.N. and Rao, J.N.K. (1990) The Estimation of the Mean Squared Error of Small Area Estimators. *Journal of the American Statistical Association*, **85**, 163-171. <https://doi.org/10.1080/01621459.1990.10475320>
- [3] Bennett, S. (1993) The EPI Cluster Sampling Method: A Critical Appraisal. Invited Paper, International Statistical Institute Session, Florence.
- [4] Rao, J.N.K. and Yu, M.Y. (1994) Small Area Estimation by Combining Time-Series and Cross-Sectional Data. *Canadian Journal of Statistics*, **22**, 511-528. <https://doi.org/10.2307/3315407>
- [5] Rao, J.N.K. (1999) Some Recent Advances in Model-Based Small Area Estimation. *Survey Methodology*, **25**, 175-186.
- [6] Rao, J.N.K. (2015) Small Area Estimation. John Wiley & Sons, Ltd., Chichester. <https://doi.org/10.1002/9781118735855>
- [7] <https://www.jinfont.com/resources/bi-defined/3-tier-architecture-complete-overview/>

- [8] RStudio, New Open-Source IDE for R | RStudio Blog (2015-05-01).
<https://blog.rstudio.com/>
- [9] <https://www.studentstutorial.com/java-project/jsp-login-form-using-mysql.php>
- [10] Urbanek, S. (2009) rJava: Low-Level R to Java Interface, R Package Version 0.8-1.
<http://CRAN.R-project.org/package=rJava>
- [11] <https://netbeans.org/kb/docs/web/mysql-webapp.html>