International Journal of Agriculture, Environment and Biotechnology
Citation: *IJAEB*: **17**(03): 601-607, September 2024
DOI: 10.30954/0974-1712.03.2024.4

RESEARCH PAPER

Comparison Between Separate and Combined Type Estimators for Estimating Total Number of In-milk Animals

Anita Sarkar[1,2], Prachi Misra Sahoo[2], Ankit Kumar Singh[1,2*], Tauqueer Ahmad[2], Ankur Biswas[2] and Kaushal Kumar Yadav[1,2]

[1]The Graduate School, ICAR- Indian Agricultural Research Institute, New Delhi, India
[2]ICAR- Indian Agricultural Statistics Research Institute, New Delhi, India

[*]Corresponding author: ankitsinghvns32@gmail.com (**ORCID ID:** 0009-0003-4926-0444)

Paper No. 1174 **Received:** 12-06-2024 **Revised:** 22-08-2024 **Accepted:** 01-09-2024

ABSTRACT

In this article, an effort has been made to compare separate and combined type estimators for estimating the total number of in-milk animals. The Ratio and Regression estimation techniques in stratified random sampling are discussed. An empirical evaluation is conducted to demonstrate the relative performance of the separate and combined type Ratio and Regression estimators for the in-milk animal population using real survey data from Kerala state. The results indicate that the separate type estimator generally outperforms the combined type estimator for both Ratio and Regression estimators across most categories of in-milk animals. Specifically, there is an improvement of more than 10% with the separate type estimator compared to the combined type for the Ratio estimator, and an improvement of over 40% for the Regression estimator.

HIGHLIGHTS

- The study investigated the comparison of separate and combined type estimators.
- A comparison of estimators was conducted to estimate the total population of in-milk animals by using the percentage of standard error.
- Empirical analysis was carried out using real survey data from Kerala state.
- The separate type estimator outperformed the combined type by over 10% for the Ratio estimator and by more than 40% for the Regression estimato.

Keywords: Livestock statistics, Percentage of standard error, Ratio estimator, Regression estimator, Stratified sampling

India is the leading producer of livestock in the world, as per the 2020-2021 livestock statistics, leading in both total livestock population and milk production. Milk is a particularly significant major livestock product (MLP) due to its high nutritional value, widespread availability, and consumption. Often considered a complete food, milk contains all essential nutrients, making it highly important in the MLP sector. India accounts for 22% of global milk production, reinforcing its position as the top milk producer (Department of Animal Husbandry and Dairying; Ministry of Fisheries, Animal husbandry, and Dairying; GOI). For the development, monitoring, and evaluation of various animal husbandry programs, timely access to accurate and current data on key livestock indicators is crucial. Precise estimates of milk production and the number of milking animals are essential for calculating the livestock sector's contribution to GDP and overall economic output. Reliable and efficient statistics are also necessary for making informed policy decisions and determining

How to cite this article: Sarkar, A., Sahoo, P.M., Singh, A.K., Ahmad, T., Biswas, A. and Yadav, K.K. (2024). Comparison Between Separate and Combined Type Estimators for Estimating Total Number of In-milk Animals. *Int. J. Ag. Env. Biotech.*, **17**(03): 601-607.

Source of Support: None; **Conflict of Interest:** None

import export quantities within the livestock sector. Therefore, it is necessary to discuss the enhancement of the estimation procedures for estimating the population of in-milk animals. Numerous scientists (Murthy *et al.* 1975; Goel *et al.* 1983; Singh *et al.* 1993) have contributed to the development and enhancement of livestock statistics through various projects that involved pilot surveys across India at different times. In addition to these efforts, several researchers have devised methodologies for estimating milk production. Singh *et al.* created sampling methodologies for estimating wool, meat, milk, and egg production by conducting surveys in various locations over different years. These pioneers laid the foundation for livestock statistics in India. They conducted surveys for estimating milk production in Gujarat (1958-59), wool production in Rajasthan (1960), and livestock numbers and production in Mysore (1961), Himachal Pradesh (1962), Andhra Pradesh (1963), the Northern region (1969-72), and the Southern region (1971-74). These surveys employed a stratified multistage sampling design, where all the *taluks* in the state were classified into three strata.

In India, four types of cattle namely exotic breeds, crossbred cattle, indigenous breeds, and non-descript cattle and two types of buffalo namely non-descript and indigenous are reared for milk production. The sampling design used to estimate the number of in-milk animals in different categories is stratified cluster sampling. The sampling frame is based on the most recent Livestock Census village list. To estimate the number of animals, the latest livestock census data serves as an auxiliary variable. Auxiliary information can enhance the accuracy of estimators. Techniques like the Ratio estimation method, Regression estimation method *etc.*, are examples of such procedures. Goodman and Hartley (1958) developed an unbiased estimator by modifying traditional ratio estimators (RE) and compared its precision to that of conventional RE. Murthy and Nanjamma (1959) introduced a technique to estimate the bias of an ordinary RE to a specified degree of approximation, using Singh (1965) introduced estimators for calculating the ratio and product of two population parameters, demonstrating that their estimator is more efficient than traditional Ratio and Product estimators. Cochran (1977) provided detailed summaries and discussions on research related to Ratio and Regression estimation methods. Rao (1988) offered a brief discussion on various Ratio estimation methods for single and multistage designs with both equal and unequal probabilities of selection. Singh and Singh (2001) proposed an improved ratio-type estimator for variance using auxiliary information, showing that their estimator is more efficient than standard estimators. Kadilar and Cingi (2007) described several ratio-type estimators and investigated their properties within stratified random sampling, creating different ratio estimates for the total of each stratum. In 2005, Kadilar and Cingi developed equations for the mean square error (MSE) and bias of an estimator of population variance in simple random sampling scheme using an auxiliary variable. Singh and Solanki (2013) proposed efficient Ratio and Product estimators for population mean in stratified random sampling using auxiliary variables.

Milk is the vital factor of income for rural households and have a key role in nutritional security. To support this, government initiatives and programs that focus on boosting milk production, enhancing livestock management, and aiding dairy farmers rely on accurate data on the number of in-milk animals. Among the major livestock products (MLPs), milk is selected for its significance. Traditional methods, such as Ratio and Regression estimators, have been commonly used to estimate in-milk animals; however, their effectiveness can vary depending on the sampling design and the characteristics of the population. Separate and combined type estimators are two widely used approaches within stratified sampling designs. Separate estimators, calculated independently for each stratum, can account for within-stratum variability, potentially leading to more precise estimates. Conversely, combined estimators, which aggregate data across strata, provide a simpler and often more efficient estimation process, particularly in homogeneous populations. However, selecting between these approaches is not always straightforward, as it depends on the populatio's characteristics and the specific goals of the estimation. Thus, this study aims to compare the effectiveness of separate and combined type estimators within a stratified sampling framework for estimating the total number of in-milk animals. The primary objectives are to: estimate the total number of in-milk animals and to provide the estimate of variance; assess the percentage of standard error of separate and combined estimators for Ratio and Regression estimator; and identify the conditions under which each estimator is more efficient, considering the variability across strata. This study offer practical insights into the application of these estimators for policymakers and researchers involved in livestock management and planning.

METHODOLOGY

Ratio method of estimation

The Ratio estimation method utilizes information on auxiliary variable that is highly correlated with the variable under to enhance precision, leading to improvement in estimators, especially in case of linear relationship between *Y* and *X*.

Suppose a variable *X* is auxiliary variable with respect to the variable under study Y, and a paired random sample of (x_i , y_i) {for $i = 1, . . . , n$} of n observations is obtained. The ratio can then be defined as;

$$R = \frac{t_y}{t_x} \quad \ldots(1)$$

where, t_y and t_x are the population total of the variable under study and the auxiliary variable respectively. The corresponding estimator can be written as;

$$\hat{R} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} \quad \ldots(2)$$

The RE can be used to find the estimate the population total (t_y) of N observations and the estimate can be given as;

$$\hat{t}_{yR} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} t_x \quad \ldots(3)$$

Corresponding to the MSE is;

$$MSE(\hat{t}_y) = N^2\left(\frac{1}{n}-\frac{1}{N}\right)\left[S_y^2 + R^2S_x^2 - 2RS_{xy}\right] \quad \ldots(4)$$

where, S_y^2, S_x^2 and S_{xy} are the population variances can be defined as;

$$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2 \ , \ S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2 \text{ and}$$

$$S_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)\times\left(Y_i - \bar{Y}\right).$$

The estimate of mean square error is given as;

$$M\hat{S}E(\hat{t}_y) = N^2\left(\frac{1}{n}-\frac{1}{N}\right)\left[s_y^2 + \hat{R}^2s_x^2 - 2\hat{R}s_{xy}\right] \quad \ldots(5)$$

where, s_y^2, s_x^2 and s_{xy} are the estimates corresponding to S_y^2, S_x^2 and S_{xy} respectively and obtained as;

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 , \ s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \text{ and}$$

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right).$$

Ratio estimator in stratified random sampling

In a stratified sampling setup, let the population *U* of *N* units be divided into *L* strata. The strata sizes are N_h where $h = 1, 2 . . . L$. A sample of size n_h on (x_{hi}, y_{hi}), where $i = 1,2,.., n_h$ is taken from h^{th} stratum using Simple Random Sampling Without Replacement (SRSWOR). To estimate the population parameters, two approaches can be employed: the separate ratio estimator and the combined ratio estimator, which are two distinct approaches for constructing REs in stratified random sampling scheme.

Separate type ratio estimator

To calculate the separate type ratio estimator, (i) apply the ratio estimation method individually within each and every stratum to obtain the RE, assuming the stratum mean is known and ii), combine all these estimates using a weighted arithmetic mean.

Let the population consists of *L* number of strata where h^{th} stratum contains N_h population units and ratio estimate of h^{th} stratum is $\hat{t}_{yhR}$ where $h=1, 2, \ldots L$. Then the separate type RE can be written as;

$$\hat{t}_{ySR} = \sum_{h=1}^{L} \frac{N_h}{N} \hat{t}_{yhR} \qquad \ldots(6)$$

where, $\hat{t}_{yhR} = \frac{\sum_{i=1}^{n_h} y_i}{\sum_{i=1}^{n_h} x_i} t_{xh}$.

Corresponding to variance and estimate of variance are obtained as;

$$MSE(\hat{t}_{ySR}) = \sum_{h=1}^{L} W_h^2 N_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) \left[S_{y_h}^2 + R_h^2 S_{x_h}^2 - 2R_h S_{x_h y_{sh}}\right] \quad \text{and}$$

$$M\hat{S}E(\hat{t}_{ySR}) = \sum_{h=1}^{L} W_h^2 N_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) \left[s_{y_h}^2 + \hat{R}_h^2 s_{x_h}^2 - 2\hat{R}_h s_{x_h y_h}\right] \quad \text{respectively.}$$

Combined type ratio estimator

For the combined type RE, the stratum means of both the study variable and the auxiliary variable are first calculated. The combined estimator is then defined as follows:

$$\hat{t}_{yCR} = \frac{\hat{t}_{yst}}{\hat{t}_{xst}} t_x \qquad \ldots(7)$$

With the variance and estimate of variance are;

$$MSE(\hat{t}_{yCR}) = \sum_{h=1}^{L} W_h^2 N_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) \left[S_{y_h}^2 + R^2 S_{x_h}^2 - 2RS_{x_h y_h}\right] \quad \text{and}$$

$$M\hat{S}E(\hat{t}_{yCR}) = \sum_{h=1}^{L} W_h^2 N_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) \left[s_{y_h}^2 + \hat{R}^2 s_{x_h}^2 - 2\hat{R} s_{x_h y_h}\right] \quad \text{respectively.}$$

Regression method of estimation

The ratio method of estimation uses auxiliary information correlated with the study variable to enhance the precision of estimators, particularly when the regression of Y on X is linear and passes through the origin. However, it is not always necessary for the regression line passes through the origin. In such cases, the Regression type estimator is more suitable for estimating population totals or means.

Assume that $\bar{y}$ and $\bar{x}$ are the sample means based on the n number of samples corresponding to study variable and auxiliary variable and b is the regression coefficient then the regression estimator of population total can be defined as;

$$\hat{t}_{y\text{Reg}} = N\left(\bar{y} + b\left(t_x - \bar{x}\right)\right) \qquad \ldots(8)$$

Mean square error is given as;

$$MSE\left(\hat{t}_{y\text{Reg}}\right) = N^2 \left(\frac{1}{n} - \frac{1}{N}\right)\left[\left(1-\rho^2\right) S_y^2\right] \qquad \ldots(9)$$

where, ρ is the population correlation coefficient and the estimate of MSE is obtained as;

$$M\hat{S}E\left(\hat{t}_{y\text{Reg}}\right) = N^2 \left(\frac{1}{n} - \frac{1}{N}\right)\left[\left(1-r^2\right) s_y^2\right] \qquad \ldots(10)$$

Regression estimator in stratified random sampling

In stratified sampling, there are two approaches to constructing regression estimators for population parameters, akin to the ratio method. These approaches are the separate type regression estimator and the combined type regression estimator.

Separate type regression estimator

Suppose that the population consists of L number of strata where h^{th} stratum contains N_h population units and the Regression estimate ($\hat{t}_{yh\text{Reg}}$) of the population total of h^{th} stratum { h = 1, 2, … L} and $\bar{y}_h$, $\bar{x}_h$ and $\bar{Y}_h$, $\bar{X}_h$ are the sample mean and population mean of h^{th} stratum corresponding to study variable and auxiliary variable. Then the separate type regression estimator of population total can be given as;

$$\hat{t}_{yS\text{Reg}} = \sum_{h=1}^{L} \frac{N_h}{N} \hat{t}_{yh\text{Reg}} \qquad \ldots(11)$$

where, $\hat{t}_{yh\text{Reg}} = N\left[\bar{y}_h + b_h\left(\bar{X}_h - \bar{x}_h\right)\right]$ and $b_h = \frac{s_{hxy}}{s_{hx}^2}$.

The mean square error and its estimate are formulated as;

$$MSE(\hat{t}_{yS\,\mathrm{Reg}}) = N^2 \left[\sum_{h=1}^{L} w_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) \left[S_{hy}^2 + B_h{}^2 S_{hx}^2 - 2B_h S_{hxy} \right] \right] \quad \ldots(12)$$

where, $\hat{t}_{yh\,\mathrm{Reg}} = N\left[\bar{y}_h + b_h\left(\bar{X}_h - \bar{x}_h\right)\right]$ and $b_h = \frac{s_{hxy}}{s_{hx}^2}$

$$MSE(\hat{t}_{yS\,\mathrm{Reg}}) = N^2 \left[\sum_{h=1}^{L} w_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) \left[S_{hy}^2 + B_h{}^2 S_{hx}^2 - 2B_h S_{hxy} \right] \right] \quad \ldots(13)$$

Combine type regression estimator

Then the combine type regression estimator of population total can be obtained as;

$$\hat{t}_{yC\,\mathrm{Reg}} = N\left[\sum_{h=1}^{L} \frac{N_h}{N}\bar{y}_h + b\left(\bar{X} - \sum_{h=1}^{T} \frac{N_h}{N}\bar{x}_h \right) \right] \quad \ldots(14)$$

where, $b = \frac{s_{xy}}{s_x^2}$ and its mean square error and its estimate are given by;

$$MSE(\hat{t}_{yC\,\mathrm{Reg}}) = N^2 \left[\sum_{h=1}^{L} \left(\frac{N_h}{N} \right)^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) \left[S_{hy}^2 + B^2 S_{hx}^2 - 2BS_{hxy} \right] \right] \quad \ldots(15)$$

and

$$M\hat{S}E(\hat{t}_{yC\,\mathrm{Reg}}) = N^2 \left[\sum_{h=1}^{L} \left(\frac{N_h}{N} \right)^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) \left[s_{hy}^2 + b^2 s_{hx}^2 - 2bs_{hxy} \right] \right] \quad \ldots(16)$$

where, $B = \frac{S_{xy}}{S_x^2}$.

Empirical study

To compare the separate and combined types of Ratio and Regression estimators, an empirical analysis is conducted to estimate the total population of in-milk animals. This analysis uses real survey data from Kerala under a project titled "Integrated Sample Survey Solution for Major Livestock Products," funded by the Department of Animal Husbandry and Dairying, GOI, and overseen by the Indian Council of Agricultural Research - Indian Agricultural Statistics Research Institute (ICAR-IASRI).

RESULTS AND DISCUSSION

The comparison of separate and combined type estimators for estimating the total population of in-milk animals was conducted using the percentage of standard error.

The percent standard error (%SE) is computed using the following formula given as;

$$\%SE = \left(\sqrt{M\hat{S}E(\hat{t}_{y\cdot\cdot})} \Big/ \hat{t}_{y\cdot\cdot} \right) \times 100$$

The percentage of standard error (%SE) for the separate and combined type ratio and regression estimators has been calculated for comparison. The results are presented in the table 1.

From the table 1, it is observed that for Indigenous cattle, the RE shows that the separate and combined type estimators have nearly identical % SE values. However, for the Regression estimator, the separate type has a lower % SE compared to the combined type, indicating that the separate type estimator performs better for Indigenous cattle. For Crossbred and Non-descriptive cattle, the separate type estimator consistently has a lower % SE for both the Ratio and Regression estimators, demonstrating that the separate type estimator is significantly better than the combined type. In the case of non-descriptive buffalo, the combined type estimator exhibits a lower % SE for both the Ratio and Regression estimators, suggesting that the combined type performs better. Overall, it can be conclude that the separate type estimator outperforms the combined type estimator for both Ratio and Regression estimators across most categories of in milk animals. There is more than 10% improvements of separate type estimator than combine type estimator in case of RE whereas in case of Regression estimator that is more than 40 %.

Table 1: Category wise %SE of the estimates of total number of in-milk animals in Kerala state

Estimators	Separate Type			Combined Type		
Cattle Indigenous						
Method of estimation	**Estimate of total number**	**Estimate of variance**	**%SE**	**Estimate of total number**	**Estimate of variance**	**%SE**
Ratio	13551	6270970.39	18.48	13590	6320374	18.50
Regression	15451	5778425.26	15.56	10756	5875030	22.53
Cattle Crossbred						
Ratio	725547	924855701.00	4.19	37634	962664316	82.44
Regression	728441	887234843.00	4.09	136168	943706362	22.56
Cattle Non-Descript						
Ratio	9932	5057347.50	22.64	13506	26763041	38.30
Regression	11741	4454132.08	17.97	10004	19402037	44.03
Buffalo Non-Descript						
Ratio	4511	4402630.54	46.51	7487	9750698	41.71
Regression	6251	3841555.09	31.35	7912	4939813	28.09

CONCLUSION

Theoretical analysis suggests that separate type estimators generally outperform combined type estimators when there is considerable variation between strata, as they better account for variability within each stratum. Combined estimators, being simpler, may introduce bias if the assumption of uniformity across strata is not met. Comparing separate and combined type estimators indicates that separate estimators are more robust and effective, particularly in stratified populations with significant variability. On the other hand, combined estimators might be less accurate in heterogeneous populations. Thus, the choice between these estimators should depend on the characteristics of the population and the goals of the estimation. This study evaluates the performance of separate and combined type Ratio and Regression estimators using the percent standard error (%SE) as a statistical measure. The analysis, based on real survey data from Kerala state, demonstrates that separate type estimators generally show greater efficiency for in-milk animals across most categories compared to combined type estimators.

REFERENCES

Cochran, W.G. 1977. *Sampling Techniques*. John Wiley and Sons. New York.

Goel, B.B.P.S., Singh, K.B. and Goyal, J.P. 1983. Pilot sample survey for studying the relative merits of data obtained by actual weighment and those through enquiry for estimation of milk production in the districts of Rohtak (Haryana) and Barabanki (U.P.) during the year 1979-80. Project report. ICAR-IASRI publication, New Delhi.

Goodman, L.A. and Hartley, H.O. 1958. The precision of unbiased ratio-type estimators. *Journal of the American Statistical Association*, **53**(282): 491-508.

Kadilar, C. and Cingi, H. 2003. Ratio estimators in stratified random sampling. *Journal of Mathematical Methods in Biosciences*, **45**(2): 218-225.

Kadilar, C. and Cingi, H. 2007. Improvement in variance estimation in simple random sampling. *Communications in Statistics, Theory and Methods*, **36**(11): 2075-2081.

Murthy, M.N. and Nanjamma, N.S. 1959. Almost unbiased ratio estimates based on interpenetrating sub-sample estimates. *Sankhyā: The Indian Journal of Statistics*, pp. 381-392.

Murthy, V.V.R., Garg, J.N. and Singh, K.B. 1975. Sample survey for estimation of milk production, Himachal Pradesh (1968-69). Project report. ICAR-IASRI publication, New Delhi.

Rao, P.S. 1988. Ratio and regression estimators. *Handbook of Statistics*, **6:** 449-468.

Singh, D., Goel, B.B.P.S., Garg, J.N. and Singh, K.B. 1979. Sampling methodology for estimation of milk production, Southern region (1971-74). Project report. ICAR-IASRI publication, New Delhi.

Singh, D., Goel, B.B.P.S., Maini, J.S. and Goel, J.P. 1979. Sampling methodology for estimation of wool production in Northern region (1969-72) and Southern region (1971-74). Project report. ICAR-IASRI publication, New Delhi.

Singh, D., Rajagopalan, M. and Maini, J.S. 1969. Sample survey for estimation of wool production, Andhra Pradesh (1963-64). Project report. ICAR-IASRI publication, New Delhi.

Singh, D., Rajagopalan, M. and Maini, J.S. 1965. Sample survey for estimation of wool production, Gujarat State (1959-60). Project report. ICAR-IASRI publication, New Delhi.

Singh, D., Rajagopalan, M. and Maini, J.S. 1966. Sample survey for estimation of wool production, Rajasthan (1960-61). Project report. ICAR-IASRI publication, New Delhi.

Singh, D., Rajagopalan, M. and Maini, J.S. 1967. Sample survey for estimation of wool production, Mysore State (1961-62). Project report. ICAR-IASRI publication, New Delhi.

Singh, D., Rajagopalan, M. and Maini, J.S. 1969. Sample survey for estimation of wool production, Himachal Pradesh (1962-63). Project report. ICAR-IASRI publication, New Delhi.

Singh, H.P. and Singh, R. 2001. Improved ratio-type estimator for variance using auxiliary information. *Journal of the Indian Society of Agricultural Statistics*, **54**(3): 276-287.

Singh, H.P. and Solanki, R.S. 2013. Efficient ratio and product estimators in stratified random sampling. *Communications in Statistics-Theory and Methods*, **42**(6): 1008-1023.

Singh, M.P. 1965. On the estimation of ratio and product of the population parameters. *Sankhyā: The Indian Journal of Statistics*, **Series B**: 321-328.

Singh, S., Jain, J.P. and Bhatia, D.K. 1993. Small area estimation of milk production, Project report. ICAR-IASRI publication, New Delhi.