International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Sampling Estimators and Properties of a Good Estimator: A Theoretical and Applied Perspective

***Pranjal Kaser**

*Research Scholar, SoS in Statistics, Pt. RSU Raipur, kaserpranjal@gmail.com

https://doi.org/10.55041/ijsmt.v2i3.343

Cite this Article: Kaser, P. (2026). Sampling Estimators and Properties of a Good Estimator: A Theoretical and Applied Perspective. International Journal of Science, Strategic Management and Technology, *02*(03). https://doi.org/10.55041/ijsmt.v2i3.343

Abstract:

Estimators are essential tools in survey sampling, enabling the approximation of unknown population parameters from limited sample data. Full population surveys are often impractical due to cost, time, and logistical constraints, making estimators such as the sample mean, ratio, regression, and stratified estimators indispensable. This paper reviews the theoretical foundations of these estimators, emphasizing their key properties—unbiasedness, consistency, efficiency, sufficiency, and minimum mean square error (MMSE). Analytical expressions for bias and mean square error (MSE), including the ratio estimator $\left(\hat{Y}R = \bar{y}\frac{X}{\bar{x}}\right)$ with approximate MSE $\left(\frac{1-f}{n}\left(S_y^2 + R^2S_x^2 - 2RSxy\right)\right)$, are presented. Applications across census surveys, agriculture, health, and social sciences are discussed, along with modern developments such as bootstrap and jackknife resampling, Bayesian estimation, and approaches for big data and adaptive sampling. The paper highlights the balance between unbiasedness and efficiency, illustrating how classical and contemporary estimators remain central to reliable statistical inference in increasingly complex survey contexts.

Keywords: Sampling estimator, unbiased estimator, Bias and MSE, Ratio and Regression.

1. Introduction:

Statistical inference plays a central role in scientific research, economics, social sciences, and national surveys. In practice, it is rarely feasible to collect data from the entire population due to cost, time, and logistical constraints. Consequently,

International Journal of Science, Strategic Management and Technology
Volume 02 Issue 03 March-2026 | ISSN: 3108-1762 (Online) | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

researchers rely on sampling techniques, where a smaller subset of the population is observed to make conclusions about population parameters.

Since parameters such as the population mean, variance, or proportion are typically unknown, we require estimators, which are functions of the sample data used to approximate these parameters. The accuracy and reliability of statistical conclusions depend heavily on the properties of the estimators employed.

This paper aims to:

1. Introduce the theoretical foundations of estimators in sampling.
2. Discuss the desirable properties of a good estimator.
3. Analyze common estimators such as sample mean, ratio, regression, and stratified estimators.
4. Derive bias and mean square error (MSE) expressions for selected estimators.
5. Present real-world applications in survey sampling and highlight modern developments.

2. Estimators in Sampling:

2.1 Definition of an Estimator:-

An estimator is a statistic, i.e., a function of the sample observations used to estimate a population parameter. If (θ) is a population parameter, then the estimator is denoted by ($\hat{\theta}$). For example:

Population mean: $[\mu = \frac{1}{N}\sum_{i=1}^{N} Y_i, \quad \hat{\mu} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i]$

Population variance: $[\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (Y_i - \mu)^2, \quad \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n} (y_i - \bar{y})^2]$

2.2 Point and Interval Estimation:-

Point estimator: A point estimator is a single numerical value calculated from a sample to approximate an unknown population parameter. For example, the sample mean ($\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$) is used as a point estimator of the population mean (μ). Point estimation is simple and provides a direct value, but it does not account for the uncertainty inherent in sampling, since different samples may yield different estimates.

Interval estimator: In contrast, an interval estimator provides a range of values, called a confidence interval (CI), within which the population parameter is expected to lie with a given probability. For the population mean (μ), if the population variance (σ^2) is known, the($(1-\alpha)100\%$) confidence interval is:

$$[\bar{y} \pm Z_{\alpha/2}, \frac{\sigma}{\sqrt{n}}]$$

where ($Z_{\alpha/2}$) is the critical value from the standard normal distribution. If (σ^2) is unknown, it is estimated by the sample variance (s^2), and the corresponding confidence interval becomes:

$$[\bar{y} \pm t_{\alpha/2,,n-1}, \frac{s}{\sqrt{n}}]$$

where ($t_{\alpha/2,,n-1}$) is the critical value from the Student's *t*-distribution with (n-1) degrees of freedom. Unlike point estimates, these intervals quantify the precision of estimation and provide insight into the reliability of the result.

3. Properties of a Good Estimator: A good estimator should satisfy the following statistical properties:

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

3.1 Unbiasedness: An estimator ($\hat{\theta}$) is unbiased if: $[E(\hat{\theta}) = \theta]$

Example: The sample mean ($\bar{y}$)is an unbiased estimator of the population mean (μ).

3.2 Consistency: An estimator is consistent if, as sample size increases, it converges in probability to the true parameter:

$$[\lim_{n\to\infty} P\,(|\hat{\theta} - \theta| < \epsilon) = 1, \quad \forall\epsilon > 0]$$

3.3 Efficiency: Among unbiased estimators of a parameter (θ), the most efficient estimator is the one with the smallest variance. If (θ_1) and (θ_2) are two unbiased estimators of (θ), then (θ_1) is said to be more efficient than (θ_2) if

$$[Var(\theta_1) < Var(\theta_2)]$$

The relative efficiency of two estimators is defined as

$$[RE(\theta_1, \theta_2) = \frac{Var(\theta_2)}{Var(\theta_1)}]$$

A higher value of (RE) (closer to 1) indicates that (θ_1) is more efficient. For example, the sample mean ($\bar{y}$) is the Best Linear Unbiased Estimator (BLUE) of the population mean (μ) under the Gauss–Markov theorem.

3.4 Sufficiency: An estimator is sufficient for a parameter if it uses all the information about the parameter present in the sample. Formally, a statistic ($T(x)$)is sufficient for (θ) if the conditional distribution of the sample given ($T(x)$)does not depend on (θ). By the Factorization Theorem, a statistic ($T(x)$)is sufficient for (θ)if the likelihood function can be factorized as:

$[L(\theta;\, x_1,\, x_2,\, \ldots,\, x_n) = g(T(x),\, \theta),\, h(x_1,\, x_2,\, \ldots,\, x_n)]$where ($g$) depends on the sample only through ($T(x)$), and (h) does not depend on θ

For example, if $(X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2))$ with (σ^2)known, the sample mean $(\bar{X})$ is a sufficient statistic for (μ).

3.5 Minimum Mean Square Error (MMSE):-

For biased estimators, the Mean Square Error (MSE) criterion is used:

$$[MSE(\hat{\theta}) = Var(\hat{\theta}) + [Bias(\hat{\theta})]^2]$$

The estimator with the lowest MSE is considered best.

4. Common Estimators in Sampling Theory:

4.1 Sample Mean (SRS):- Under Simple Random Sampling Without Replacement (SRSWOR):

$$[y = \frac{1}{n}\sum_{i=1}^{n} y_i]$$

Unbiased: $(E(\bar{y}) = \mu)$.

Variance: $[Var(\bar{y}) = \frac{1-f}{n}S^2, \quad f = \frac{n}{N}]$

4.2 Ratio Estimator:- Used when auxiliary information (X) is available.

$$[\hat{Y}_R = \bar{y}\frac{X}{\bar{x}}]$$

Bias (approximate): $[Bias(YR) \approx -\frac{1}{n}(\frac{Syx}{\bar{x}} - R\frac{S_x^2}{\bar{x}^2})]$

Useful when (Y) and (X) are positively correlated.

4.3 Regression Estimator:-

$$[\hat{Y}reg = \bar{y} + b(X - \bar{x})]$$

where $(b = \frac{S_{yx}}{S_x^2})$, Generally has smaller MSE than the ratio estimator.

4.4 Stratified Sampling Estimator:- Population is divided into (L) strata, and independent SRS is conducted in each stratum.

$$[\bar{y}st = \sum h = 1^L W_h \bar{\bar{\bar{y}}}_h, \qquad W_h = \frac{N_h}{N}]$$

Variance: $$[Var(\bar{y}st) = \sum h = 1^L W_h^2 \frac{1-f_h}{n_h} S_h^2]$$

5. Bias and Mean Square Error (MSE) of Estimators:- The bias of an estimator measures the difference between the expected value of the estimator and the true population parameter. It is given by:

$$[Bias(\hat{\theta}) = E(\hat{\theta}) - \theta]$$

- If $(Bias(\hat{\theta}) = 0)$, the estimator is unbiased.
- If $(Bias(\hat{\theta}) \neq 0)$, the estimator is biased, though sometimes slightly biased estimators are preferred if they have smaller MSE.

The Mean Square Error (MSE) of an estimator combines both variance and squared bias into a single measure of accuracy:

$$[MSE(\hat{\theta}) = Var(\hat{\theta}) + [Bias(\hat{\theta})]^2]$$

Thus, even a biased estimator can be considered good if it has a small MSE compared to other alternatives.

Example: For Ratio Estimator

Suppose we are interested in estimating the population mean of the study variable (Y) using auxiliary variable (X) (with known population mean (X)). The ratio estimator is defined as:

$$[Y_R = y \cdot \frac{X}{\bar{x}}]$$

International Journal of Science, Strategic Management and Technology
Volume 02 Issue 03 March-2026 | ISSN: 3108-1762 (Online) | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

where ($\bar{y}$) and ($\bar{x}$) are the sample means of (Y) and (X).

Bias (approximate): Using a Taylor series expansion, the bias of the ratio estimator can be shown to be:

$$[Bias(\hat{Y}_{\mathrm{R}}) \approx -\frac{1}{n}(\frac{Syx}{\bar{x}} - R\frac{S_x^2}{\bar{x}^2})]$$

where (S_{yx}) is the covariance between (Y) and (X), (S_x^2) is the variance of (X), and ($R = \frac{\bar{Y}}{\bar{X}}$)

Mean Square Error (MSE):

The approximate MSE of the ratio estimator is given by:

$$\left[MSE(\hat{Y}_{\mathrm{R}}) \approx \frac{1-f}{n}\backslash Big(S_y^2 + R^2S_x^2 - 2RSxy\backslash Big)\right]$$

6. Applications of Estimators in Surveys:

Estimators are widely applied in diverse survey settings to generate reliable conclusions about populations from limited samples. In census surveys, stratified sampling is often employed to estimate population size more efficiently by accounting for regional or demographic variations, with the stratified mean given by

$$[\bar{y}st = \sum h = 1^L W_h \bar{y}\, h, \quad W_h = \frac{N_h}{N}]$$

In agriculture, ratio and regression estimators improve the precision of crop yield estimation by incorporating auxiliary variables such as land area or rainfall; for instance, the ratio estimator is expressed as

$$[\hat{Y}R = \bar{y} \cdot \frac{X}{\bar{x}}]$$

where (X) is the known population mean of the auxiliary variable. In health surveys, stratified cluster sampling is frequently applied to estimate prevalence rates of diseases, with the sample proportion

$$[\hat{P} = \frac{x}{n}]$$

serving as an unbiased estimator of the population prevalence (P). In the social sciences, estimators are crucial for measuring employment levels, literacy rates, and poverty indices, often enhanced through auxiliary data; for example, a regression estimator can be used as

$$[\hat{Y}reg = \bar{y} + b(X - \bar{x}), \quad b = \frac{Syx}{S_x^2}]$$

which reduces sampling error when there is strong correlation between study and auxiliary variables.

7. Challenges and Recent Developments:

In recent years, the field of survey sampling has faced several challenges and witnessed important methodological developments. A key issue is the trade-off between unbiasedness and minimum mean square error (MSE), since an estimator with slight bias but lower variance may be preferable. For example, the ratio estimator $(\hat{Y}R = \bar{y}\frac{X}{\bar{x}})$ is biased, yet often has a smaller MSE than the unbiased sample mean (\bar{y}), as its approximate MSE is given by

$$[MSE(\hat{Y}R) \approx \frac{1-f}{n}(S_y^2 + R^2 S_x^2 - 2RSxy)]$$

Another challenge lies in balancing small-sample properties against asymptotic efficiency; while large-sample theory often ensures consistency and normality, small samples may suffer from skewed distributions and inflated variance. To address this, modern computational methods such as the bootstrap and jackknife are employed: the bootstrap repeatedly resamples the data to estimate variance, e.g.,

while the jackknife systematically omits one observation at a time to estimate and correct bias,

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

$$\left[\widehat{Var}_{jack}(\hat{\theta}) = \frac{1}{B}\sum_{b=1}^{B}\left(\widehat{\theta_b^*} - \bar{\theta}^*\right)^2\right]$$

$$\left[\hat{B}as_{jack}(\hat{\theta}) = (n-1)(\bar{\theta}(\cdot) - \hat{\theta})\right]$$

where $(\bar{\theta}(\cdot))$ is the average of leave-one-out estimates. Furthermore, Bayesian estimators have gained prominence, especially in small-area estimation, where prior information is combined with sample data; for instance, the posterior mean

$$[\hat{\theta}Bayes = E(\theta|\text{data})]$$

minimizes expected squared error under quadratic loss. Finally, the growth of big data and adaptive sampling has pushed researchers to develop estimators that can handle non-probability samples, high-dimensional auxiliary variables, and real-time updating, ensuring that sampling theory remains relevant in modern large-scale applications.

8. Conclusion:

Estimators are fundamental to statistical inference in sample surveys, providing the foundation for drawing reliable conclusions from limited data. An ideal estimator should be unbiased, consistent, efficient, and sufficient; however, in practice, minimizing mean square error often takes precedence over strict unbiasedness. Traditional estimators, such as the sample mean, ratio, and regression estimators, continue to play a central role, while modern approaches—including resampling techniques, Bayesian frameworks, and adaptive sampling—enhance precision and applicability in complex and large-scale survey contexts. Looking forward, ongoing research is exploring the development of estimators that can efficiently handle high-dimensional auxiliary information, non-probability samples, and real-time data streams in big data environments. There is also a growing interest in integrating machine learning methods with classical estimation theory to create hybrid estimators that are both robust and computationally efficient. These developments suggest a dynamic future for estimator theory, where classical principles and modern innovations converge to meet the evolving challenges of survey research.

9. References:

1. Cochran, W. G. (1977): *Sampling Techniques*, John Wile and Sons., New York, III edition.
2. Kish, L. (1965). *Survey Sampling*. John Wiley & Sons.
3. Lehmann, E.L., & Casella, G. (1998). *Theory of Point Estimation*. Springer.
4. Mukhopadhyay, P. (1998): *Theory and methods of survey sampling Prentice*, Hall of India, New Delhi.
5. Raj, D. and Chandak, P. (1998): *Sample survey theory*, Narosa Publishing House, New Delhi.
6. Sarndal, C.E., Swensson, B., & Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer.
7. Sukhatme, P. V. Sukhatme, B. V. Sukhatme, S. and Asok, C. (1992): *Sampling theory of surveys with applications*, lowa State University Press, Ames. Lowa, U. S. A. III revised Edition.