

An Efficient Estimator of Population Mean Using a Poisson Regression-Based Approach: Simulation and Theoretical Efficiency Analysis

Rajesh Singh

Banaras Hindu University

Shobh Nath Tiwari

shobh7285@gmail.com

Banaras Hindu University <https://orcid.org/0009-0006-5930-7973>

Yashveer Singh

Banaras Hindu University

Research Article

Keywords: Poisson regression, Ratio estimators, Mean square error, Bias, Percentage relative efficiency, Auxiliary variable

Posted Date: September 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11160721/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

An Efficient Estimator of Population Mean Using a Poisson Regression-Based Approach: Simulation and Theoretical Efficiency Analysis

Rajesh Singh¹, Shobh Nath Tiwari^{1,*}, Yashveer Singh¹

¹Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi-221005,
Uttar Pradesh, India

Emails: rsinghstat@bhu.ac.in; shobh7285@gmail.com; yashveer.stat.2025@bhu.ac.in

*Corresponding: shobh7285@gmail.com

Abstract

The Poisson regression model is commonly used to model count responses in numerous studies. In this study, an improved Poisson regression-based estimator for the population mean is proposed under simple random sampling (SRS), along with the derivation of its associated mean square error (MSE) formula. The MSE performance of the developed estimator is theoretically evaluated against traditional ratio estimators. In the empirical study, two populations from the environmental health and transportation safety domains were considered, and the proposed estimator outperformed competing estimators. The simulation study also showed similar results. The estimation results validate the improved performance of the suggested estimator over conventional estimators. Moreover, the empirical results supported the theoretical findings of the study.

Keywords: Poisson regression; Ratio estimators; Mean square error; Bias; Percentage relative efficiency; Auxiliary variable.

1. Introduction

In sample survey theory, the use of auxiliary data to enhance estimation accuracy plays a fundamental role. Auxiliary information is frequently used for stratification, determining selection probabilities, and improving estimation efficiency. Estimating the population total or mean is one of the main purposes of most surveys. Several researchers, including Upadhyaya et al. [1], Upadhyaya and Singh [2], Koyuncu [3], Abid et al. [4], Shahzad et al. [5], Zaman and Bulut [6-7], and Zaman [8], have explored the use of auxiliary information to improve population mean estimation. Ratio and regression estimators are commonly preferred for mean estimation when a positive relationship exists between the study and auxiliary variables. To improve the efficiency of population mean estimation, sampling theory frequently incorporates auxiliary information such as the coefficient of variation and kurtosis in ratio estimators. However, the efficiency of classical methods decreases in the presence of outliers or extreme observations. To overcome this issue, Kadilar et al. [9] modified the ratio estimator of Kadilar and Cingi [10] using the Huber-M estimate. Oral and Kadilar [11-12], introduced two methods, namely the modified maximum likelihood and integrated approaches, for mean estimation under SRS. Zaman [8] constructed a new family of robust ratio-type estimators through the combination of the ratio estimators

developed by Zaman and Bulut [6]. Ali et al. [13] introduced a new class of robust regression-type estimators based on Zaman and Bulut [6] for sensitive studies under SRS.

In addition, applying a linear regression model to count data may be inappropriate when the mean is large, as the Poisson distribution tends to approximate the normal distribution. The linear relationship between predicted values and auxiliary variables in the model may result in negative predictions. The assumption of constant variance plays an important role in ensuring the validity of hypothesis testing in linear regression. These assumptions are generally invalid for count data. Therefore, the Poisson regression model is frequently used in applied sciences to model count data. In recent years, Poisson regression-based approaches for population mean estimation have been developed by Shahzad et al. [5], Jana et al. [14], and Raghav et al. [15].

In Poisson regression, the study variable Y_i , where $(y_i = 0, 1, 2, \dots)$, represents the count of events occurring over a specified period and is assumed to follow a Poisson distribution. The sequential procedure underlying the Poisson regression model adopted in the study is graphically illustrated through the flowchart presented in Figure 1.

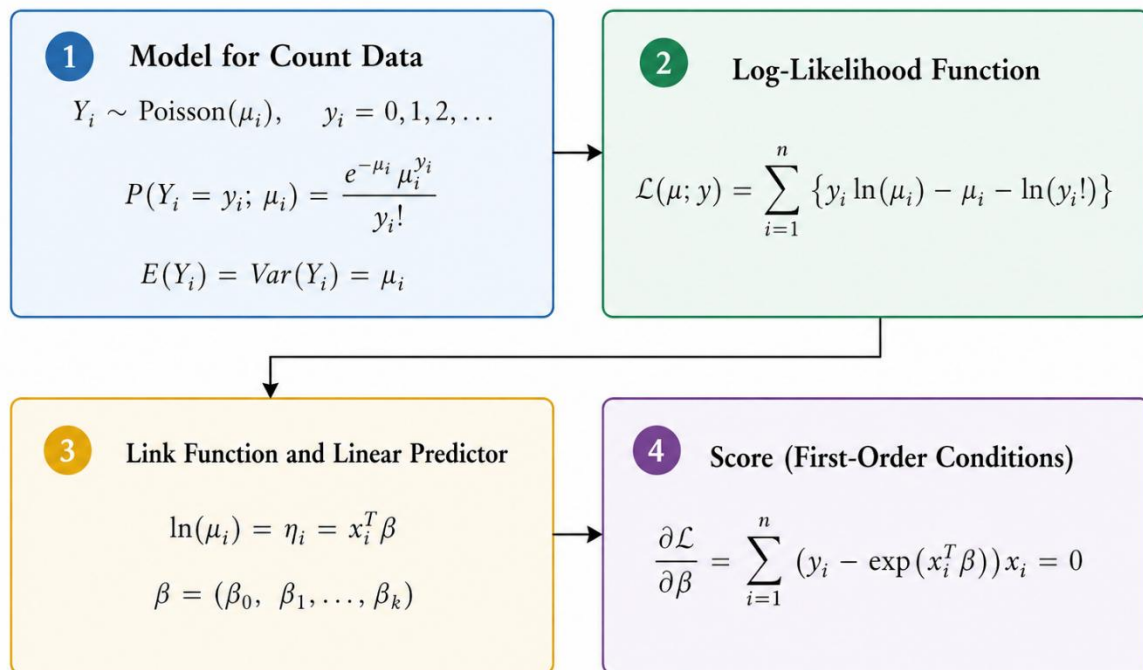


Figure 1: Flowchart of the Poisson Regression Modeling Procedure

To obtain the solution of the k - equations, iterative optimization techniques including the Newton–Raphson and Fisher scoring algorithms are utilized (see Cameron and Trivedi [16], Montgomery et al. [17], and Koç [18]).

2. Review of Existing Estimators

This section briefly reviews some notable estimators and estimator classes that have been proposed in the literature.

The following ratio estimators for the population mean $\bar{Y}$ in SRS were introduced by Kadilar and Cingi [10]:

$$\bar{y}_{KC1} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{\bar{x}} (\bar{X}). \quad (2.1)$$

$$\bar{y}_{KC2} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x} + C_x)} (\bar{X} + C_x). \quad (2.2)$$

$$\bar{y}_{KC3} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x} + \beta_2(x))} (\bar{X} + \beta_2(x)). \quad (2.3)$$

$$\bar{y}_{KC4} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}\beta_2(x) + C_x)} (\bar{X}\beta_2(x) + C_x). \quad (2.4)$$

$$\bar{y}_{KC5} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(\bar{x}C_x + \beta_2(x))} (\bar{X}C_x + \beta_2(x)). \quad (2.5)$$

In this context, C_x and $\beta_2(x)$ represent the population coefficient of variation and population coefficient of kurtosis associated with the auxiliary variable, respectively. Here, $\bar{y}$ and $\bar{x}$ represent the sample means of the study and auxiliary variables, respectively.

The quantity $b = \frac{s_{yx}}{s_x^2}$ refers to the regression coefficient of slope computed by the least squares method. Additionally, s_x^2 is the sample variance of the auxiliary variable, while s_{yx} denotes the covariance between the study and auxiliary variables based on the sample observations.

The MSE expressions of the estimators are derived as follows for $(i = 1, 2, \dots, 5)$:

$$MSE(\bar{y}_{KCi}) = \theta [R_{KCi}^2 S_x^2 + 2BR_{KCi} S_x^2 + B^2 S_x^2 - 2R_{KCi} S_{xy} - 2BS_{xy} + S_y^2]. \quad (2.6)$$

The quantity $B = \frac{s_{xy}}{s_x^2}$ represents the regression coefficient derived using the least squares method, whereas S_x^2 and S_y^2 indicate the population variances of the auxiliary and study variables, respectively. Moreover, $f = \frac{n}{N}$ is defined as the sampling ratio, where n denotes the sample size and N represents the population size and $\theta = \frac{1}{n} - \frac{1}{N}$. The population ratios are calculated as given below:

$$R_{KC1} = R = \frac{\bar{Y}}{\bar{X}}; R_{KC2} = \frac{\bar{Y}}{(\bar{X} + C_x)}; R_{KC3} = \frac{\bar{Y}}{(\bar{X} + \beta_2(x))}; R_{KC4} = \frac{\bar{Y}\beta_2(x)}{(\bar{X}\beta_2(x) + C_x)}; R_{KC5} = \frac{\bar{Y}C_x}{(\bar{X}C_x + \beta_2(x))}.$$

Inspired by the work of Kadilar and Cingi [10], Koç [18] introduced a novel approach to enhance estimation efficiency through Poisson regression-based ratio estimators, given as follows:

$$\bar{y}_1 = \frac{\bar{y} + b_P(\bar{X} - \bar{x})}{\bar{x}} (\bar{X}) \quad (2.7)$$

$$\bar{y}_2 = \frac{\bar{y} + b_P(\bar{X} - \bar{x})}{(\bar{x} + C_x)} (\bar{X} + C_x) \quad (2.8)$$

$$\bar{y}_3 = \frac{\bar{y} + b_P(\bar{X} - \bar{x})}{(\bar{x} + \beta_2(x))} (\bar{X} + \beta_2(x)) \quad (2.9)$$

$$\bar{y}_4 = \frac{\bar{y} + b_P(\bar{X} - \bar{x})}{(\bar{x}\beta_2(x) + C_x)} (\bar{X}\beta_2(x) + C_x) \quad (2.10)$$

$$\bar{y}_5 = \frac{\bar{y} + b_P(\bar{X} - \bar{x})}{(\bar{x}C_x + \beta_2(x))} (\bar{X}C_x + \beta_2(x)) \quad (2.11)$$

Here, b_p refers to the slope coefficient derived from the Poisson regression model such that $E(b_p) = B_p$. The corresponding MSE of $\bar{y}_j$, where $j = 1, 2, \dots, 5$ is expressed as follows:

$$MSE(\bar{y}_j) = \theta [R_{KCi}^2 S_x^2 + 2B_p R_{KCi} S_x^2 + B_p^2 S_x^2 - 2R_{KCi} S_{xy} - 2B_p S_{xy} + S_y^2]. \quad (2.12)$$

Following the frameworks of Zaman and Bulut [7] and Ali et al. [13], the proposed estimator incorporates the Poisson regression-based mean estimator introduced by Shahzad et al. [5].

$$\bar{y}_U = \bar{y} + b_p(\bar{X} - \bar{x}). \quad (2.13)$$

The following expression represents the MSE of the proposed mean estimator:

$$MSE(\bar{y}_U) = \theta [S_y^2 - 2B_p S_{xy} + B_p^2 S_x^2]. \quad (2.14)$$

Following Koç [18], Jana et al. [14] introduced Poisson regression-based ratio-cum-regression estimators for efficient estimation of the population mean under SRSWOR.

$$Z_i = \bar{y} \left[\frac{\alpha_i \bar{X}}{\left\{ \frac{\alpha_i \bar{X} + \beta_i}{\alpha_i \bar{x} + \beta_i} \right\}} + b_p \left\{ 1 - \frac{\bar{x}}{\bar{X}} \right\} \right] \quad (2.15)$$

Here, $\alpha_i \neq 0$, while β_i may represent constants or functions of known parameters of the auxiliary variable X , selected to minimize the MSE of estimator Z_i , $i = 1, 2, \dots, 5$.

The first-order approximate MSE of the estimator Z_i , $i = 1, 2, \dots, 5$ is presented below:

$$MSE(Z_i) = \theta [S_y^2 + R^2(\phi_i^2 + B_p)^2 S_x^2 - 2R(\phi_i^2 + B_p) S_{xy}]. \quad (2.16)$$

Where:

$$\phi_i = \frac{\alpha_i \bar{X}}{\alpha_i \bar{X} + \beta_i}, i = 1, 2, \dots, 5.$$

$$\phi_1 = \frac{\bar{X}}{(\bar{X} + S_x)}; \phi_2 = \frac{C_x \bar{X}}{C_x \bar{X} + S_x}; \phi_3 = \frac{\rho_{yx} C_x \bar{X}}{\{\rho_{yx} C_x \bar{X} + (\frac{1-f}{1+f}) S_x\}}; \phi_4 = \frac{\rho_{yx} \bar{X}}{\{\rho_{yx} \bar{X} + S_x\}}; \phi_5 = \frac{\bar{X}}{\{\bar{X} + (\frac{1-2f}{1+2f}) S_x\}}.$$

3. Proposed Estimator

Inspired by Raghav et al. [15] and the traditional regression estimators employing the Poisson regression approach introduced by Koç [18], we propose an enhanced estimator for the population mean, defined as:

$$Z^* = \left[\alpha \bar{y} \left\{ \frac{1}{2} \left(\frac{\bar{X}}{\bar{x}} + \frac{\bar{x}}{\bar{X}} \right) \right\} + b_p(\bar{X} - \bar{x}) \right] \exp \left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}} \right) \quad (3.1)$$

where, the scalar α consists of unknown constants determined through the minimization of the MSE of the proposed estimator Z^* , whereas b_p retains the same slope coefficient used in Koç [18].

Equation (3.1) can be rewritten in terms of error as follows:

$$Z^* = \left[\alpha \bar{Y}(1 + \epsilon_0) \left\{ \frac{1}{2} ((1 + \epsilon_1)^{-1} + 1 + \epsilon_1) \right\} + b_p(\bar{Z} - \bar{Z}(1 + \epsilon_1)) \right] \exp \left\{ \frac{\bar{Z} - \bar{Z}(1 + \epsilon_1)}{\bar{Z} + \bar{Z}(1 + \epsilon_1)} \right\}. \quad (3.2)$$

Equation (3.2) reduces to the following form after simplification:

$$Z^* - \bar{Y} = \left[\alpha \bar{Y} \left(1 + \epsilon_0 + \frac{\epsilon_1^2}{2} \right) - b_p \bar{Z} \epsilon_1 \right] \exp \left\{ \frac{-\epsilon_1}{2 + \epsilon_1} \right\} - \bar{Y} \quad (3.3)$$

$$Z^* - \bar{Y} = \left[\alpha \bar{Y} \left(1 + \epsilon_0 + \frac{\epsilon_1^2}{2} \right) - b_p \bar{Z} \epsilon_1 \right] \exp \left\{ \frac{-\epsilon_1}{2} \left(1 - \frac{\epsilon_1}{2} + \frac{\epsilon_1^2}{4} \right) \right\} - \bar{Y}. \quad (3.4)$$

Expanding equation (3.4) yields the following expression:

$$Z^* - \bar{Y} = \left[\alpha \bar{Y} \left(1 + \epsilon_0 - \frac{\epsilon_1}{2} + \frac{7\epsilon_1^2}{8} - \frac{\epsilon_0 \epsilon_1}{2} \right) + b_p \bar{Z} \left(\frac{\epsilon_1^2}{2} - \epsilon_1 \right) - \bar{Y} \right] \quad (3.5)$$

The bias can be obtained from equation (3.5) by evaluating expectations as follows:

$$B(Z^*) = \left[\alpha \bar{Y} \left(1 + \frac{7\theta C_z^2}{8} - \frac{\theta \rho_{yz} C_y C_z}{2} \right) + B_p \bar{Z} \left(\frac{\theta C_z^2}{2} \right) - \bar{Y} \right]. \quad (3.6)$$

Squaring both sides of equation (3.5) and ignoring higher-order terms, we obtain:

$$(Z^* - \bar{Y})^2 = \left[\begin{aligned} &\bar{Y}^2 - 2\bar{Y}^2 \alpha - \frac{7}{4} \epsilon_1^2 \bar{Y}^2 \alpha + \epsilon_0 \epsilon_1 \bar{Y}^2 \alpha + \bar{Y}^2 \alpha^2 + 2\epsilon_1^2 \bar{Y}^2 \alpha^2 - 2\epsilon_0 \epsilon_1 \bar{Y}^2 \alpha^2 + \\ &\epsilon_0^2 \bar{Y}^2 \alpha^2 - \epsilon_1^2 \bar{Y} \bar{Z} b_p + 2\epsilon_1^2 \bar{Y} \bar{Z} \alpha b_p - 2\epsilon_0 \epsilon_1 \bar{Y} \bar{Z} \alpha b_p + \epsilon_1^2 \bar{Z}^2 b_p^2 \end{aligned} \right]. \quad (3.7)$$

The expectation of equation (3.7) gives:

$$MSE(Z^*) = \left[\begin{aligned} &\bar{Y}^2 - 2\bar{Y}^2 \alpha - \frac{7}{4} \theta C_z^2 \bar{Y}^2 \alpha + \theta \rho_{yz} C_y C_z \bar{Y}^2 \alpha + \bar{Y}^2 \alpha^2 + 2\theta C_z^2 \bar{Y}^2 \alpha^2 - \\ &2\theta \rho_{yz} C_y C_z \bar{Y}^2 \alpha^2 + \theta C_y^2 \bar{Y}^2 \alpha^2 - \theta C_z^2 \bar{Y} \bar{Z} B_p + 2\theta C_z^2 \bar{Y} \bar{Z} \alpha B_p - \\ &2\theta \rho_{yz} C_y C_z \bar{Y} \bar{Z} \alpha B_p + \theta C_z^2 \bar{Z}^2 B_p^2 \end{aligned} \right]. \quad (3.8)$$

Or

$$MSE(Z^*) = \left[\begin{aligned} &\alpha^2 (\bar{Y}^2 + 2\theta C_z^2 \bar{Y}^2 - 2\theta \rho_{yz} C_y C_z \bar{Y}^2 + \theta C_y^2 \bar{Y}^2) + \\ &\alpha \left(\theta \rho_{yz} C_y C_z \bar{Y}^2 - 2\bar{Y}^2 - \frac{7}{4} \theta C_z^2 \bar{Y}^2 + 2\theta C_z^2 \bar{Y} \bar{Z} B_p - 2\theta \rho_{yz} C_y C_z \bar{Y} \bar{Z} B_p \right) + \\ &\bar{Y}^2 + \theta C_z^2 \bar{Z}^2 B_p^2 - \theta C_z^2 \bar{Y} \bar{Z} B_p \end{aligned} \right] \quad (3.9)$$

$$MSE(Z^*) = [\alpha^2 A + \alpha B + C]. \quad (3.10)$$

where,

$$A = \bar{Y}^2 + 2\theta C_z^2 \bar{Y}^2 - 2\theta \rho_{yz} C_y C_z \bar{Y}^2 + \theta C_y^2 \bar{Y}^2.$$

$$B = \theta \rho_{yz} C_y C_z \bar{Y}^2 - 2\bar{Y}^2 - \frac{7}{4} \theta C_z^2 \bar{Y}^2 + 2\theta C_z^2 \bar{Y} \bar{Z} B_p - 2\theta \rho_{yz} C_y C_z \bar{Y} \bar{Z} B_p.$$

$$C = \bar{Y}^2 + \theta C_z^2 \bar{Z}^2 B_p^2 - \theta C_z^2 \bar{Y} \bar{Z} B_p.$$

Minimizing equation (3.10) yields the optimum values of α as follows:

$$\alpha = \frac{-B}{2A}. \quad (3.11)$$

Replacing α in Equation (3.10) gives the minimum MSE:

$$MSE(t_{st}^*) = C - \frac{B^2}{4A}. \quad (3.12)$$

4. Empirical Study

Population I: This population is based on an air pollution dataset reported in Koç [18]. The data were collected from the Afyon Respiratory Disease Hospital and the Afyon Environmental Department Air Pollution Unit during the period 2006–2010. The weekly number of hospital admissions was considered as the dependent variable Y , while PM10 was taken as the auxiliary variable X .

Population II: The data collected from TUIK for 81 provinces in 2019, as used by Koc [18], were examined by treating traffic accident deaths as the dependent variable Y , while the number of motor vehicles was considered as the auxiliary variable X .

A detailed descriptive summary for each population is reported in Table 1.

Table 1: Summary Statistics for Populations I and II

Values	Population 1	Population 2
N	213	81
n	40	20
Y	4.6760564	82.4074074
X	116.1915	274308
C_y	0.63355516	1.025504149
C_x	0.439846	1.90331
$\beta_2(x)$	-0.6908	35.77509
S_y	2.96253962	84.5091382
S_x	51.10636	522093
S_{xy}	53.51706	38124520.6
ρ_{xy}	0.353	0.864077723
B_l	0.02049002	0.02049002
B_p	0.004	6.131E-07

Table 2: MSE and PRE Values for Populations I

Estimators	MSE	PRE
$\bar{y}_1$	0.1859	100.0000
$\bar{y}_2$	0.1855	100.2049
$\bar{y}_3$	0.1865	99.6729

$\bar{y}_4$	0.1864	99.6996
$\bar{y}_5$	0.1873	99.2471
$\bar{y}_u$	0.1704	109.0978
Z_1	0.1560	119.1551
Z_2	0.1615	115.0634
Z_3	0.1690	110.0186
Z_4	0.1640	113.3082
Z_5	0.1590	116.8705
Z^*	0.1547	120.1844

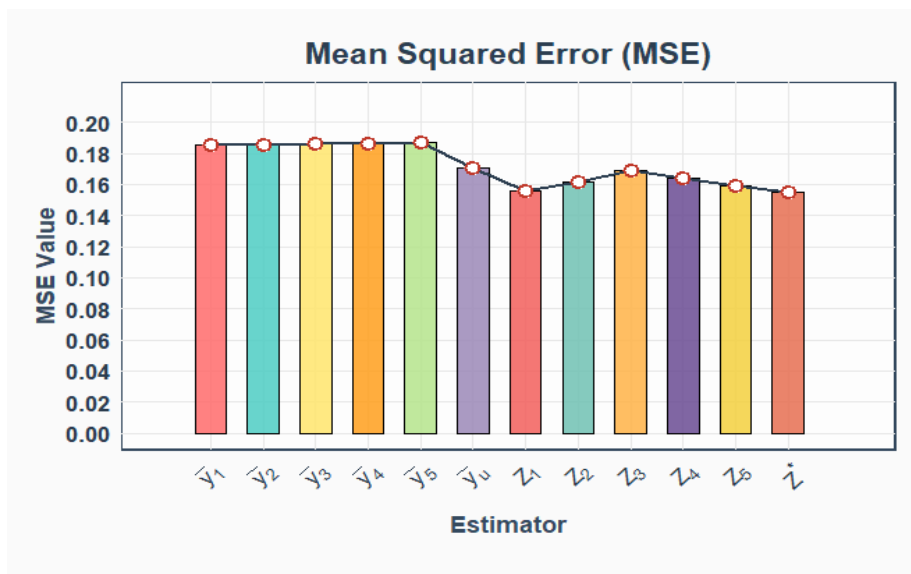


Figure 1: Comparative MSE performance of the estimators for Population I

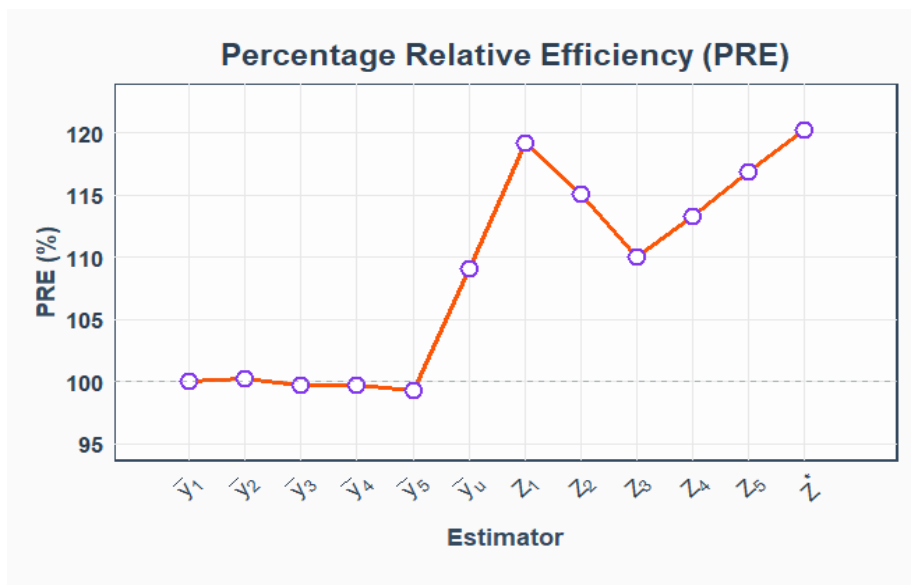


Figure 2: Comparative PRE performance of the estimators for Population I

The calculation of PRE values is carried out using the following expression:

$$PRE = \frac{MSE(\bar{y}_1)}{MSE(Proposed\ estimator)} \times 100.$$

The percentage relative efficiency (PRE) is used to compare the mean squared error (MSE) of the conventional mean estimator ($\bar{y}_1$) with the proposed Poisson regression-based mean estimators.

Table 3: MSE and PRE Values for Populations II

Estimators	MSE	PRE
$\bar{y}_1$	334.7393	100.0000
$\bar{y}_2$	334.7324	100.0021
$\bar{y}_3$	334.6097	100.0387
$\bar{y}_4$	334.7391	100.0001
$\bar{y}_5$	334.6712	100.0203
$\bar{y}_u$	267.1630	125.2940
Z_1	179.6295	186.3499
Z_2	111.1809	301.0761
Z_3	81.2808	411.8306
Z_4	193.6337	172.8724
Z_5	76.6703	436.5956
Z^*	14.4090	2323.1327

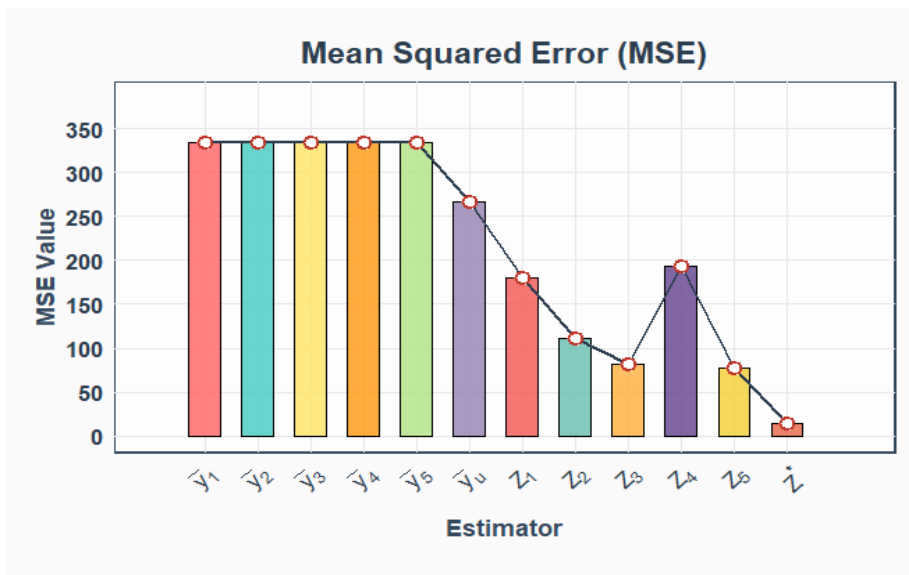


Figure 3: Comparative MSE performance of the estimators for Population II

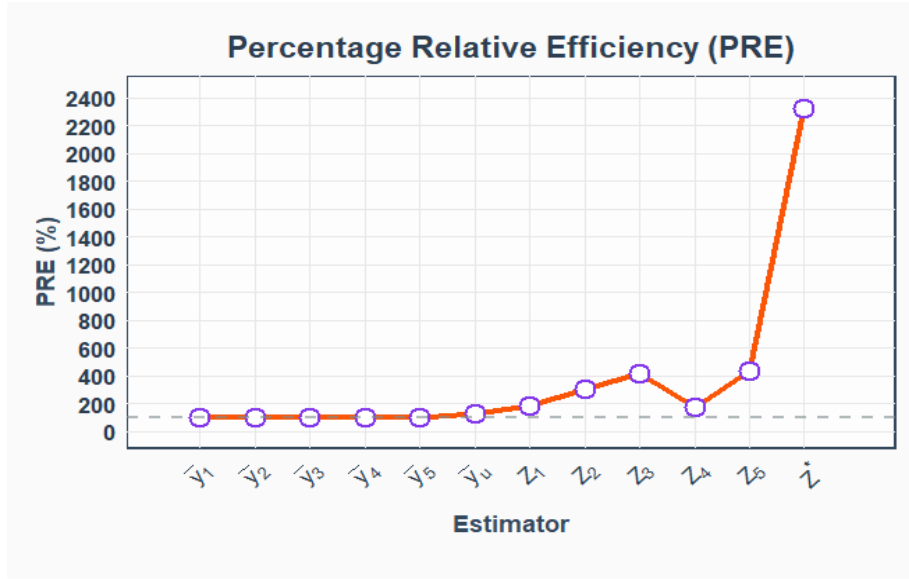


Figure 4: Comparative PRE performance of the estimators for Population II

5. Simulation Study

The simulation study was carried out in accordance with the procedures described by Koç [18] and Raghav et al. [15], utilizing R software to confirm the theoretical findings and assess the performance of both proposed and existing estimators. The data generation process is described as follows:

- A finite population of size $N = 1000$ was generated. The auxiliary variable X was drawn from a uniform distribution, $X \sim U(0, 1)$, and the study variable Y was generated from a Poisson distribution, $Y \sim P(\lambda)$, where

$$\lambda = \exp(\beta_0 + \beta_1 X),$$

with $\beta_0 = 0.1$ and $\beta_1 = 0.75, 0.85$ and 0.95 , respectively.

- From the generated population, samples of size $n = 20, 30, 40$ were selected using simple random sampling without replacement (SRSWOR).
- For each sample, the estimates of the population mean were computed using the proposed estimator as well as existing estimators
- The mean squared error (MSE) of an estimator, say $\hat{\bar{Y}}_i$, such as $\bar{y}_1, Z_1$, etc. (given in Section 3) or Z^* (given in Section 4), was computed as

$$MSE(\hat{\bar{Y}}_i) = \frac{1}{25000} \sum_{i=1}^{25000} (\hat{\bar{Y}}_i - \bar{Y})^2$$

where $\bar{Y}$ denotes the population mean.

- The results for MSE and PRE are summarized in Tables 3–5, with graphical representations of MSE provided in Figures 1–3.

Table 4: Comparative Evaluation of Estimators Using MSE and PRE for $\beta_1 = 0.75$

Estimator	$\beta_1 = 0.75$					
	$n = 20$		$n = 30$		$n = 40$	
	MSE	PRE	MSE	PRE	MSE	PRE
$\bar{y}_1$	0.1036	100.0000	0.0682	100.0000	0.0505	100.0000
$\bar{y}_2$	0.0823	125.9201	0.0543	125.6326	0.0401	125.9600
$\bar{y}_3$	0.0809	127.9810	0.0535	127.5932	0.0395	127.8999
$\bar{y}_4$	0.0855	121.1067	0.0564	120.8858	0.0416	121.1822
$\bar{y}_5$	0.0816	126.8663	0.0539	126.4420	0.0398	126.7501
$\bar{y}_u$	0.0842	123.0043	0.0557	122.5262	0.0411	122.8132
Z_1	0.0855	121.0989	0.0564	121.0095	0.0416	121.2924
Z_2	0.0821	126.1117	0.0542	125.9136	0.0400	126.1860
Z_3	0.0808	128.2522	0.0533	127.8738	0.0394	128.1711
Z_4	0.0808	128.2516	0.0533	127.8967	0.0394	128.1932
Z_5	0.0862	120.1011	0.0571	119.5164	0.0423	119.2607
Z^*	0.0783	132.3153	0.0526	129.5458	0.0393	128.5704

In Table 4, the MSE of all estimators decreases with an increase in sample size (n) from 20 to 40. The estimator Z^* consistently achieves the lowest MSE and highest PRE, indicating superior efficiency, while the proposed estimators outperform the conventional estimator $\bar{y}_1$ across all sample sizes.

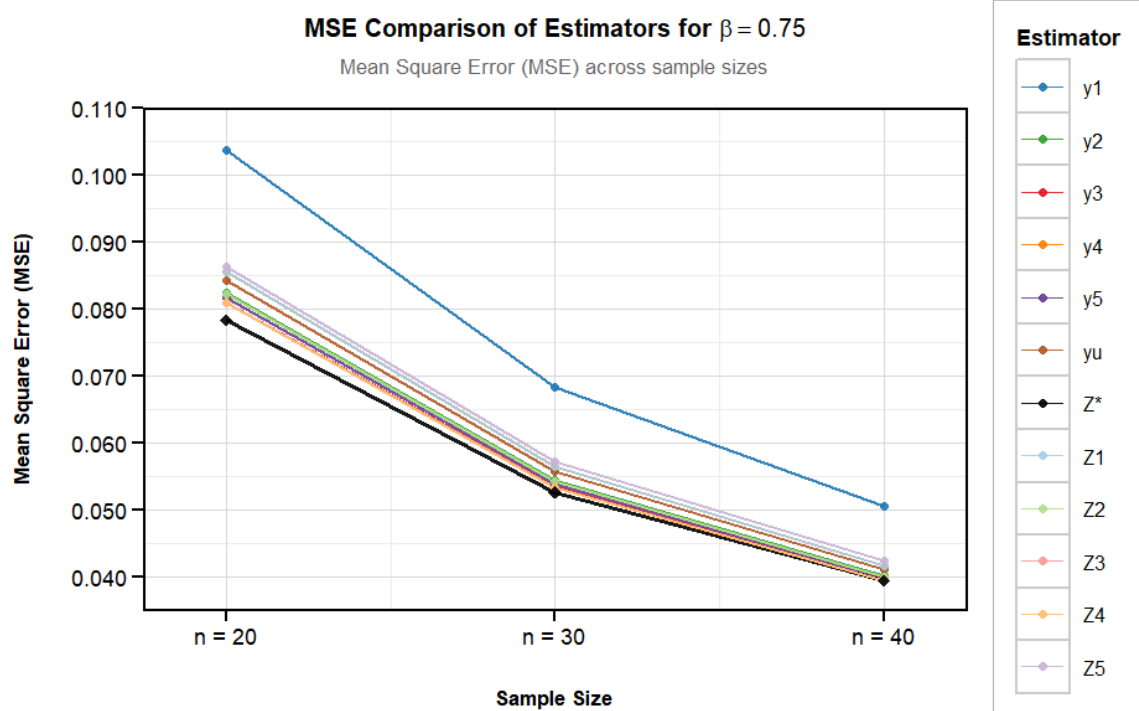


Figure 5: Performance Evaluation of Estimators Based on MSE for $\beta_1 = 0.75$

Figure 5 illustrates that the estimator Z^* consistently attains the lowest MSE, confirming its superior performance across all sample sizes.

Table 5: Comparative Evaluation of Estimators Using MSE and PRE for $\beta_1 = 0.85$

Estimator	$\beta_1 = 0.85$					
	$n = 20$		$n = 30$		$n = 40$	
	MSE	PRE	MSE	PRE	MSE	PRE
$\bar{y}_1$	0.1087	100.0000	0.0713	100.0000	0.0529	100.0000
$\bar{y}_2$	0.0868	125.2956	0.0570	125.0693	0.0422	125.3715
$\bar{y}_3$	0.0862	126.1510	0.0567	125.8290	0.0419	126.1247
$\bar{y}_4$	0.0899	120.9222	0.0591	120.7486	0.0437	121.0240
$\bar{y}_5$	0.0873	124.5821	0.0574	124.2193	0.0424	124.5273
$\bar{y}_u$	0.0906	119.9436	0.0597	119.5150	0.0441	119.8183
Z_1	0.0913	119.0521	0.0599	119.0581	0.0443	119.3451
Z_2	0.0874	124.4093	0.0574	124.3148	0.0424	124.5779
Z_3	0.0857	126.8923	0.0563	126.6105	0.0417	126.8922
Z_4	0.0858	126.6797	0.0564	126.4341	0.0417	126.7252
Z_5	0.0921	117.9985	0.0607	117.4781	0.0451	117.2014
Z^*	0.0827	131.5002	0.0553	128.8933	0.0413	127.9573

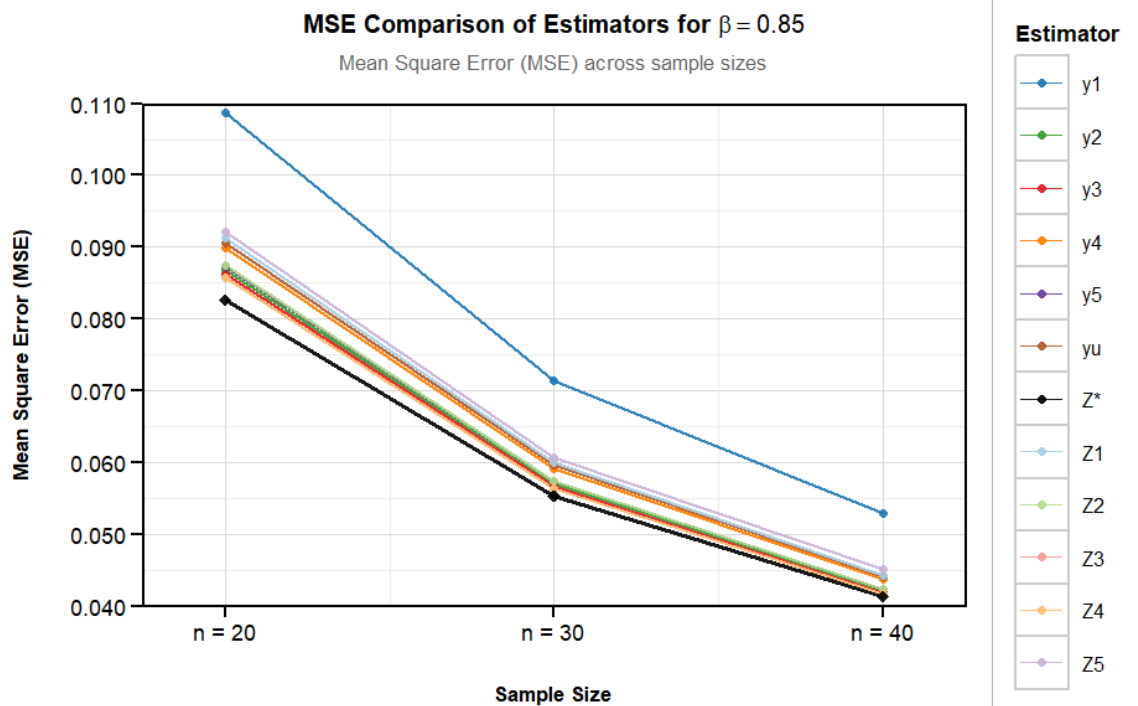


Figure 6: Performance Evaluation of Estimators Based on MSE for $\beta_1 = 0.85$

Table 5 and Figure 6 demonstrate that the estimator Z^* consistently achieves the lowest MSE, confirming its superior efficiency across all sample sizes.

Table 6: Comparative Evaluation of Estimators Using MSE and PRE for $\beta_1 = 0.95$

Estimator	$\beta_1 = 0.95$					
	$n = 20$		$n = 30$		$n = 40$	
	MSE	PRE	MSE	PRE	MSE	PRE
$\bar{y}_1$	0.1140	100.0000	0.0749	100.0000	0.0553	100.0000
$\bar{y}_2$	0.0917	124.4225	0.0603	124.2361	0.0444	124.5892
$\bar{y}_3$	0.0920	123.9068	0.0606	123.6281	0.0446	123.9994
$\bar{y}_4$	0.0946	120.5904	0.0622	120.4466	0.0458	120.7609
$\bar{y}_5$	0.0936	121.8493	0.0616	121.5273	0.0454	121.9190
$\bar{y}_u$	0.0980	116.4082	0.0645	116.0172	0.0475	116.4055
Z_1	0.0978	116.5803	0.0642	116.6320	0.0473	116.9899
Z_2	0.0932	122.3828	0.0612	122.3529	0.0451	122.6842
Z_3	0.0910	125.3166	0.0598	125.1099	0.0441	125.4507
Z_4	0.0914	124.7521	0.0601	124.5877	0.0443	124.9558
Z_5	0.0988	115.4590	0.0651	114.9460	0.0482	114.7040
Z^*	0.0874	130.4650	0.0585	127.9991	0.0435	127.1767

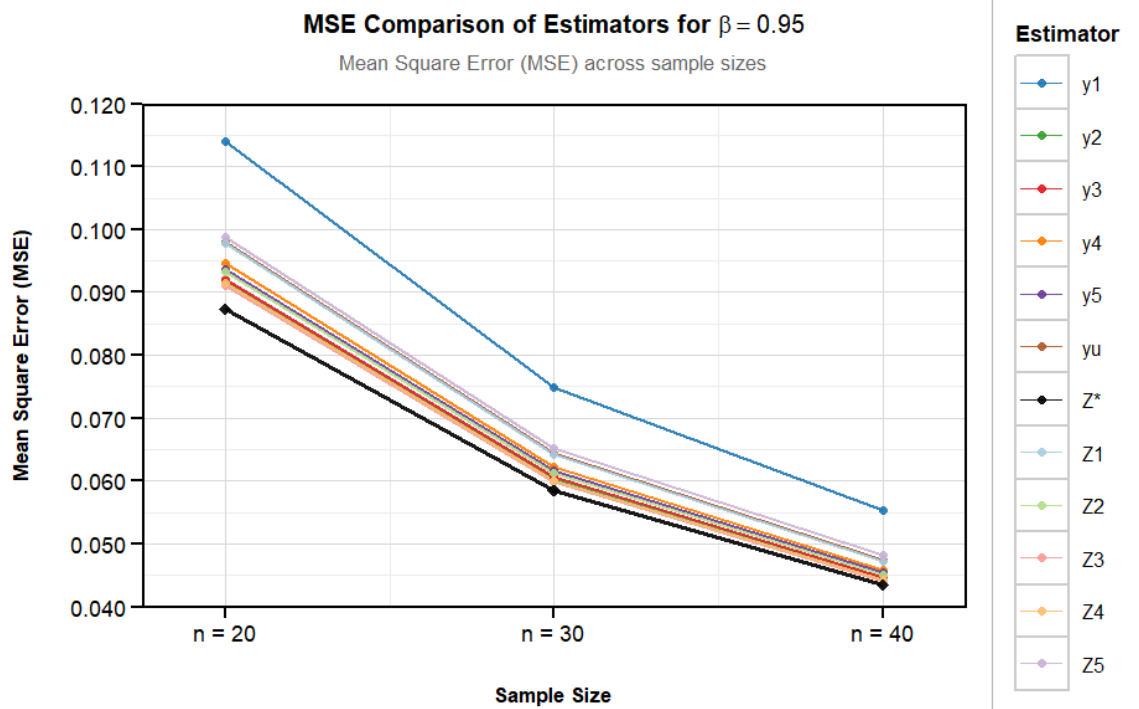


Figure 7: Performance Evaluation of Estimators Based on MSE for $\beta_1 = 0.95$

6. Discussion

The summary statistics of the real datasets are reported in Table 1. The empirical results from two real datasets indicate that the suggested estimator (Z^*) performs better than competing estimators by attaining minimum MSE and improved PRE values, as shown in Tables 2 and 3. The simulation analysis further validated the real data study by producing similar findings. The results in Table 4, 5 and 6 indicate that the proposed estimator achieved superior efficiency compared to Kadilar and Cingi [10] and Koç [18] estimators under all β_1 and sample size (n) combination. Figures 5–7 clearly show the better performance of the proposed estimator over Kadilar and Cingi [10] and Koç [18] estimators at different β_1 values (0.75, 0.85, and 0.95) and sample sizes (20, 30, and 40), with minimum MSE and higher PRE values.

7. Conclusion

This paper introduces an improved population mean estimator utilizing SRS in combination with Poisson regression. The bias and MSE expressions were obtained using first-order approximation. The theoretical results were validated using real and simulated data. The suggested estimator exhibits superior performance compared to the estimators of Kadilar and Cingi [10] and Koç [18]. The graphical presentations of MSE and PRE measures for the considered estimators under Population I and Population II are illustrated in Figures 1–2 and Figures 3–4, respectively. These graphical illustrations provide a clear evaluation of the estimators' performance in the empirical and simulation studies, demonstrating the superior efficiency of the suggested estimator over conventional estimators. The performance of the estimators based on MSE for simulated data with $\beta_1 = 0.75, 0.85$ and 0.95 is presented in Figures 5, 6, and 7, respectively. Therefore, the proposed estimator (Z^*) is recommended over traditional alternatives for count data analysis. Future research may focus on developing estimators for count datasets with excessive zeros through the application of zero-inflated Poisson regression models. Addressing such cases may help improve statistical methods for a wide range of applications.

8. Future works

The present work provides a foundation for extending the proposed methodology to more diverse settings. Future studies may focus on:

- a) To develop more refined Poisson regression-based estimators that address practical issues such as over-dispersion, excess zeros, missing observations, and measurement errors in count data.
- b) Incorporating multiple auxiliary variables within the Poisson regression framework.

- c) Extending the proposed method to other sampling frameworks, such as stratified random sampling, ranked set sampling, and double sampling.
- d) Examining the performance of the estimator under PPS, unequal probability, and adaptive sampling designs.

Data Availability

All the data used for this study can be found inside the manuscript.

Ethical approval

This research does not involve experimentation on Human Participants or Animals.

Funding

The authors declare that no funds or other grants were received for the preparation of this manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Upadhyaya, L. N., Singh, H. P., & Vos, J. W. E. (1985). On the estimation of population means and ratios using supplementary information. *Statistica Neerlandica*, 39(3), 309–318. <https://doi.org/10.1111/j.1467-9574.1985.tb01149.x>.
- [2] Upadhyaya, L. N., & Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal*, 41(5), 627–636.
- [3] Koyuncu, N. (2012). Efficient estimators of population mean using auxiliary attributes. *Applied Mathematics and Computation*, 218(22), 10900–10905.
- [4] Abid, M., Abbas, N., Nazir, H. Z., & Lin, Z. (2016). Enhancing the mean ratio estimators for estimating population mean using non-conventional location parameters. *Revista Colombiana de Estadística*, 39(1), 63–79.
- [5] Shahzad, U., Shahzadi, S., Afshan, N., Al-Noor, N. H., Alilah, D. A., Hanif, M., & Anas, M. M. (2021). Poisson regression-based mean estimator. *Mathematical Problems in Engineering*, 2021, Article 9769029, 1–6.
- [6] Zaman, T., & Bulut, H. (2019). Modified ratio estimators using robust regression methods. *Communications in Statistics – Theory and Methods*, 48(8), 2039–2048. <https://doi.org/10.1080/03610926.2018.1441419>.

- [7] Zaman, T., & Bulut, H. (2020). Modified regression estimators using robust regression methods and covariance matrices in stratified random sampling. *Communications in Statistics – Theory and Methods*, 49(14), 3407–3420. <https://doi.org/10.1080/03610926.2019.1588324>.
- [8] Zaman, T. (2019). Improvement of modified ratio estimators using robust regression methods. *Applied Mathematics and Computation*, 348, 627–631.
- [9] Kadilar, C., Candan, M., & Cingi, H. (2007). Ratio estimators using robust regression. *Hacettepe Journal of Mathematics and Statistics*, 36(2), 181–188.
- [10] Kadilar, C., & Cingi, H. (2004). Ratio estimators in simple random sampling. *Applied Mathematics and Computation*, 151(3), 893–902. [https://doi.org/10.1016/S0096-3003\(03\)00803-8](https://doi.org/10.1016/S0096-3003(03)00803-8).
- [11] Oral, E., & Kadilar, C. (2011). Improved ratio estimators via modified maximum likelihood. *Pakistan Journal of Statistics*, 27(3), 269–282.
- [12] Oral, E., & Kadilar, C. (2011). Robust ratio-type estimators in simple random sampling. *Journal of the Korean Statistical Society*, 40(4), 457–467. <https://doi.org/10.1016/j.jkss.2011.04.001>.
- [13] Ali, N., Ahmad, I., Hanif, M., & Shahzad, U. (2021). Robust-regression-type estimators for improving mean estimation of sensitive variables by using auxiliary information. *Communications in Statistics – Theory and Methods*, 50(4), 979–992. <https://doi.org/10.1080/03610926.2019.1645857>.
- [14] Jana, Z. H. W., Rizvi, S. E. H., Sharma, M., Bhat, M. I. J., & Mushtaq, S. (2022). Modified regression estimators for improving mean estimation: Poisson regression approach. *Pakistan Journal of Statistics and Operation Research*, 18(4), 985–994.
- [15] Raghav, Y. S., Ahmadini, A. A. H., Mahnashi, A. M., & Rather, K. U. I. (2025). Enhancing estimation efficiency with proposed estimator: A comparative analysis of Poisson regression-based mean estimators. *Kuwait Journal of Science*, 52(1), Article 100282. <https://doi.org/10.1016/j.kjs.2024.100282>.
- [16] Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*. Cambridge University Press.
- [17] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2006). *Introduction to linear regression analysis* (4th ed.). John Wiley & Sons.
- [18] Koç, H. (2021). Ratio-type estimators for improving mean estimation using Poisson regression method. *Communications in Statistics – Theory and Methods*, 50(20), 4685–4691. <https://doi.org/10.1080/03610926.2020.1777307>.