Research Article

L-Moments and Calibration-Based Estimators for Variance Parameter

Malik Muhammad Anas,¹ Muhammad Ali,² Ambreen Shafqat,¹ Faisal Shahzad,¹ Kashif Abbass,³ and David Anekeya Alilah⁴

¹Department of Statistics and Financial Mathematics, School of Science, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China

²Department of Mathematics and Statistics, PMAS-Arid Agriculture University Rawalpindi, Rawalpindi 46300, Pakistan

³School of Economics and Management, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China

⁴Department of Mathematics, Masinde Muliro University of Science and Technology, Kakamega, Kenya

Correspondence should be addressed to David Anekeya Alilah; aliladavid2010@gmail.com

Received 23 July 2021; Accepted 3 August 2021; Published 31 August 2021

Academic Editor: Ishfaq Ahmad

Copyright © 2021 Malik Muhammad Anas et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The subject of variance estimation is one of the most important topics in statistics. It has been clarified by many different research studies due to its various applications in the human and natural sciences. Different variance estimators are built based on traditional moments that are especially influenced by the existence of extreme values. In this paper, with the presence of extreme values, we proposed some new calibration estimators for variance based on L-moments under double-stratified random sampling. A simulation study with COVID-19 data is performed to evaluate the efficiency of the proposed estimators. All results indicate that the proposed estimators are often superior and highly efficient compared to the existing traditional estimator.

1. Introduction

Statistics has special importance among the different sciences, whether natural, human, or social, because it deals with human and scientific needs at various levels and in a variety of methods ranging from simplicity describing primary information (statistical data) up to more complex and sophisticated methods of analyzing data related to describing and explaining phenomena or events and their relationship with each other and predicting and controlling them in light of the laws that govern them. That is why statistics came to be viewed as that science that is concerned with studying different life phenomena using a group of statistical means for the purpose of describing and explaining them and establishing expectations about the future behavior of phenomena and directions of their development. Thus, this science becomes an essential tool for reaching correct decisions in describing and studying phenomena through the information available or collected.

The additional information, in the statistical literature, that is attached to each unit or element is referred to as auxiliary (or supplementary, ancillary, concomitant, and supporting) information. There are several types of statistical data or auxiliary information, including data taken from documenting various actual details, statistics gathered from data available in delivery service centers, and surveys. Actually, how this information is used is determined by the type in which it is available. In sampling techniques, the use of this information is very old, and this is due to the pioneers in this field. Watson [1] and Cochran [2] used the auxiliary information in developing estimation techniques that led to high estimation accuracy. Later, Abid et al. [3], Ali et al. [4], Hanif and Shahzad [5], Shahzad et al. [6–8], and Zaman and Bulut [9, 10] have shown important work on the use of auxiliary information in various ways and applications.

The key concern in all sample surveys is to find point estimates for the parameters of interest, and it is equally important to extrapolate variations or discrepancies in these

estimates. For an estimator, the importance of variance estimates is largely due to the fact that it provides a measure of the quality or accuracy of the estimates, enables users to draw specific conclusions, enables statistical agencies to provide data accuracy indicators to users, and is used to calculate confidence intervals [11]. The sampling design that forms the basis of the sample survey is one of the most important considerations deciding the procedure for estimating the variance as well as the number of sampling stages. In this paper, we are interested in estimating the variance with double-stratified random sampling. Stratified sampling is a form of survey sampling in which the population is grouped into a number of different categories (called strata). Then, each stratum is sampled as an independent subpopulation from which individual elements can be selected randomly. In order for each element to have the same probability of being sampled, the strata should not overlap. For more details about estimating the variance and double-stratified random sampling, see the works of Särndal et al. [12] and Al-Omari [13]. The data structure plays a major role in performing the correct analysis of the data by various methods of estimation. The process is greatly facilitated when the data are homogeneous and free of extreme values or outliers. In practice, however, most of the data are not free of such observations that are inconsistent with the rest of the observations and have bad and negative effects on the estimation of variance based on the central moments. One strategy to solve this problem, which provides a robust statistical basis for the data analysis, is to adopt the L-moments [14] which are computed based on a linear set of expected values for the order statistics. In addition, calibration estimation is another popular statistical strategy for improving the accuracy of estimators that are based on the use of auxiliary information to set the original weights of the design. While implementing a set of constraints with additional information, the calibration estimation methodology used calibrated or tuned weights aimed at reducing the measured distance to the original weights. The credit for the use of calibration estimation with survey data was attributed to pioneers Deville and Särndal [15], and then, many works related to the subject matter of estimation were carried out, such as that of Clement [16] and Koyuncu [17]. In the current paper, our objective is to develop some new variance estimators based on the above two strategies, “L-moments and calibration method,” with double-stratified random sampling. The rest of this article is organized as follows. Firstly, the adapted family of estimators is presented in detail. Secondly, L-moments and proposed estimators are given in detail. Then, a simulation study based on data of COVID-19 that is performed to assess the performance of the proposed estimators is given along with its obtained results. Finally, the conclusions are provided.

2. Adapted Estimators

Consider two random variables, “ X as auxiliary variable and Y as study variable,” associated with a finite population Ω of size N , where $\Omega = \{u_1, u_2, \dots, u_n\}$ is stratified into H strata with the stratum h containing N_h units, for $h = 1, 2, \dots, K$,

where $\sum_{j=1}^K N_j = N$. From the stratum h , a simple random sample of size n_h is selected without replacement, such as $\sum_{h=1}^H n_h = n$. Additionally, for X and Y in h^{th} stratum, with $h = 1, 2, \dots, K$ and $i = 1, 2, \dots, N_h$, (x_{hi}, y_{hi}) symbolize the observed values, (S_{xh}^2, S_{yh}^2) and (S_{xh}^2, S_{yh}^2) symbolize the population and sample variances, and $(\beta_{2xh}, \beta_{2yh})$ symbolize the coefficients of skewness. In the light of this stratified sampling design, a new family of estimators [18] was implemented as shown below:

$$F_{sti} = \sum_{h=1}^H \psi_h \frac{s_{yh}^2}{D_{1hi} s_{xh}^2 + D_{2hi}} (D_{1hi} s_{xj}^2 + D_{2hi}), \quad \text{for } i = 1, 2, \dots, 6, \quad (1)$$

with the methodology of the Taylor series, for $h(s_{yh}^2, s_{xh}^2) = F_{sti}$,

$$\text{MSE}(F_{sti}) = \sum_{h=1}^H d_{hi} \sum_h d'_{hi}, \quad (2)$$

where

$$\begin{aligned} d_{hi} &= \left[\frac{\delta h(s_{yh}^2, s_{xh}^2)}{\delta s_{yh}^2} |s_{yh}^2, s_{xh}^2 \frac{\delta h(s_{yh}^2, s_{xh}^2)}{\delta s_{xh}^2} |s_{yh}^2, s_{xh}^2 \right], \\ d_{hi} &= [\psi_h - \psi_h R_{hi}], \quad \text{for } h = 1, 2, \dots, H \text{ and } i = 1, 2, \dots, 6, \\ \Sigma_h &= \begin{bmatrix} V(s_{yh}^2) & \text{Cov}(s_{yh}^2, s_{xh}^2) \\ \text{Cov}(s_{xh}^2, s_{yh}^2) & V(s_{xh}^2) \end{bmatrix}. \end{aligned} \quad (3)$$

After substituting d_{hi} and Σ_h in equation (1), the above expressions of $\text{MSE}(F_{sti})$ will be

$$\text{MSE}(F_{sti}) = \sum_{h=1}^H \psi_h^2 \lambda_h s_{yh}^2 [(\beta_{2yh} - 1) - 2(\theta_h - 1)R_{hi} + (\beta_{2xh} - 1)R_{hi}^2], \quad (4)$$

where

$$\begin{aligned} R_{hi} &= \frac{A_{1hi} s_{yj}^2}{A_{1hi} s_{xh}^2 + A_{2hi}}, \\ \theta_h &= \frac{\mu_{22h}}{\mu_{20h} \mu_{02h}}, \\ \psi_h &= \frac{N_h}{N}. \end{aligned} \quad (5)$$

The adapted family members are provided in Table 1. Hosking and Wallis [18] employed the two robust measures of location “interdecile range (IDR) and midrange (MR),” that are less affected in the presence of extreme values with the conventional second central moment. The formulas of IDR and MR associated with X in stratum h are, respectively, given by $\text{IDR}_{xh} = D_{9h} - D_{1h}$ and $\text{MR}_{xh} = (\max(x_h) + \min(x_h))/2$. Furthermore, ρ_{xh} and C_{xh} represent, respectively, the population correlation coefficient and coefficient of variation associated with X in stratum h .

TABLE 1: Adapted family of estimators.

F_{sti}	A_{1hi}	A_{2hi}
F_{st1}	1	MR_{xh}
F_{st2}	1	IDR_{xh}
F_{st3}	ρ_{xh}	MR_{xh}
F_{st4}	ρ_{xh}	IDR_{xh}
F_{st5}	C_{xh}	MR_{xh}
F_{st6}	C_{xh}	IDR_{xh}

3. L-Moments and Proposed Estimator

Variance is one of the well-known and widely used measures of dispersion. However, it depends on traditional moments that are strongly affected by the presence of extreme values. Hosking [14] provided a solution to this problem, which is the adoption use of L-moments that are determined by linear combinations of the expected values of the order statistics. L-moments have many advantages over traditional moments. The main four advantages are as follows: (i) formulate as linear functions of the data, (ii) suffer less from the effects of sample variability, (iii)

more robust to unusual values “outliers or extreme values” that may contain the data, and (iv) allow for safer inferences from small samples about the underlying probability distribution. For more interesting information about robust regression methods, interested readers are referred to [19–21].

The general population formulas for the first four L -moments associated with X in stratum h are defined as

$$\begin{aligned}
 L_{1xl} &= E(X_{1:1}), \\
 L_{2xl} &= \frac{1}{2}E(X_{2:2} - X_{1:2}), \\
 L_{3xl} &= \frac{1}{3}E(X_{3:3} - 2X_{2:3} + X_{1:3}), \\
 L_{4xl} &= \frac{1}{4}E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}),
 \end{aligned} \tag{6}$$

and the sample formulas corresponding to the above four L -moments can be written as

$$\begin{aligned}
 \hat{L}_{1xl} &= \frac{1}{\binom{n_h}{1}} \sum_{k=1}^{n_h} x_{h(k)}, \\
 \hat{L}_{2xl} &= \frac{1}{2\binom{n_h}{2}} \sum_{k=1}^{n_h} \left\{ \binom{k-1}{1} - \binom{n-k}{1} \right\} x_{h(k)}, \\
 \hat{L}_{3xl} &= \frac{1}{3\binom{n_h}{3}} \sum_{k=1}^{n_h} \left\{ \binom{k-1}{2} - 2\binom{k-1}{1}\binom{n-k}{1} + \binom{n-k}{2} \right\} x_{h(k)}, \\
 \hat{L}_{4xl} &= \frac{1}{4\binom{n_h}{4}} \sum_{k=1}^{n_h} \left\{ \binom{k-1}{3} - 3\binom{k-1}{2}\binom{n-k}{1} + 3\binom{k-1}{1}\binom{n-k}{2} - \binom{n-k}{3} \right\} x_{h(k)},
 \end{aligned} \tag{7}$$

where $(:)$ represents the binomial coefficient and $x_{h(k)}$ is the k^{th} order statistics.

Also, by replacing x with y , we can write the general mathematical forms of L -moments for Y . For further information on L -moments, interested readers may refer to the work of Hosking and Wallis [18].

Now, based on L -moments, some symbolizations for proposed estimators w.r.t. stratum h are given below.

$(\bar{X}_{hl} = L_{1xhl}, \bar{x}_{hl} = \hat{L}_{1xhl})$ and $(\bar{Y}_{hl} = L_{1yhl}, \bar{y}_{hl} = \hat{L}_{1yhl})$ symbolize the X - and Y - population and sample means (L -location).

$(S_{hxl}^2 = L_{2xl}^2, s_{hxl}^2 = \hat{L}_{2xl}^2)$ and $(S_{hyt}^2 = L_{2yt}^2, s_{hyt}^2 = \hat{L}_{2yt}^2)$ symbolize the X - and Y - population and sample variances.

$(S_{hxl} = L_{2xl}, s_{hxl} = \hat{L}_{2xl})$ and $(S_{hyt} = L_{2yt}, s_{hyt} = \hat{L}_{2yt})$ symbolize the X - and Y - population and sample standard deviations (L -scale).

$(C_{hxl} = (L_{2xl}/L_{1xl}), c_{hxl} = (\hat{L}_{2xl}/\hat{L}_{1xl}))$ and $(C_{hyt} = (L_{2yt}/L_{1yt}), \bar{C}_{hyt} = (\hat{L}_{2yt}/\hat{L}_{1yt}))$ symbolize the X - and Y - population and sample coefficients of variation (L -CV).

In this paper, a new calibration estimator of population variance with stratified sampling is considered as follows:

$$G_{st(p)} = \sum_{h=1}^H \vartheta_h^* s_{hyt}^2, \quad (8)$$

where ϑ_h^* are calibrated weights. Using the loss functions of chi-square,

$$L(\vartheta_h^*, \psi_h) = \sum_{h=1}^H \frac{(\vartheta_h^* - \psi_h)^2}{\psi_h \Delta_h}, \quad (9)$$

and the following three calibration constraints

$$\sum_{h=1}^H \vartheta_h^* \bar{x}_{hl} = \sum_{h=1}^H \psi_h \bar{X}_{hl}, \quad (10)$$

$$\sum_{h=1}^H \vartheta_h^* s_{hxl}^2 = \sum_{h=1}^H \psi_h S_{hxl}^2, \quad (11)$$

$$\sum_{h=1}^H \vartheta_h^* c_{hxl} = \sum_{h=1}^H \psi_h C_{hxl}, \quad (12)$$

with the function of Lagrange are given by

$$\begin{aligned} \Omega = & \sum_{h=1}^H \frac{(\vartheta_h^* - \psi_h)^2}{\psi_h \Delta_h} - 2\lambda_1^* \left(\sum_{h=1}^H \vartheta_h^* \bar{x}_{hl} - \sum_{h=1}^H \psi_h \bar{X}_{hl} \right) \\ & - 2\lambda_2^* \left(\sum_{h=1}^H \vartheta_h^* s_{hxl}^2 - \sum_{h=1}^H \psi_h S_{hxl}^2 \right) - 2\lambda_3^* \left(\sum_{h=1}^H \vartheta_h^* c_{hxl} - \sum_{h=1}^H \psi_h C_{hxl} \right). \end{aligned} \quad (13)$$

Now, the process of minimizing the loss functions of chi-square in equation (9) which subject to the three constraints of calibration in equations (10)–(12) provides the weights of calibration with stratified sampling as follows:

$$\vartheta_h^* = \psi_h + \psi_h \Delta_h (\lambda_1^* \bar{x}_{hx} + \lambda_2^* s_{hxl}^2 + \lambda_3^* c_{hxl}). \quad (14)$$

The following system equations are yields from substituting equation (14) in equations (10)–(12):

$$\begin{bmatrix} U_{11} & U_{12} & U_{13} \\ U_{12} & U_{22} & U_{23} \\ U_{13} & U_{23} & U_{33} \end{bmatrix} \begin{bmatrix} \lambda_1^* \\ \lambda_2^* \\ \lambda_3^* \end{bmatrix} = \begin{bmatrix} U_{10} \\ U_{20} \\ U_{30} \end{bmatrix}. \quad (15)$$

Solving system equations in equation (15) w.r.t. λ_s^* , we obtain

$$\begin{aligned} \lambda_1^* &= \frac{(U_{13}U_{23} - U_{12}U_{33})(U_{12}U_{20} - U_{22}U_{10}) - (U_{13}U_{22} - U_{12}U_{23})(U_{12}U_{30} - U_{23}U_{10})}{(U_{12}^2 - U_{11}U_{22})(U_{13}U_{23} - U_{12}U_{33}) - (U_{13}U_{22} - U_{12}U_{23})(U_{12}U_{13} - U_{11}U_{23})}, \\ \lambda_2^* &= \frac{(U_{13}U_{23} - U_{12}U_{33})(U_{12}U_{10} - U_{11}U_{20}) - (U_{12}U_{13} - U_{23}U_{11})(U_{13}U_{20} - U_{12}U_{30})}{(U_{12}^2 - U_{11}U_{22})(U_{13}U_{23} - U_{12}U_{33}) - (U_{13}U_{22} - U_{12}U_{23})(U_{12}U_{13} - U_{11}U_{23})}, \\ \lambda_3^* &= \frac{(U_{12}^2 - U_{11}U_{22})(U_{13}U_{20} - U_{12}U_{30}) - (U_{13}U_{22} - U_{12}U_{23})(U_{12}U_{10} - U_{11}U_{20})}{(U_{12}^2 - U_{11}U_{22})(U_{13}U_{23} - U_{12}U_{33}) - (U_{13}U_{22} - U_{12}U_{23})(U_{12}U_{13} - U_{11}U_{23})}, \end{aligned} \quad (16)$$

where

$$U_{11} = \sum_{h=1}^H \psi_h \Delta_h \bar{x}_{hl}^2,$$

$$U_{22} = \sum_{h=1}^H \psi_h \Delta_h s_{hxl}^4,$$

$$U_{33} = \sum_{h=1}^H \psi_h \Delta_h c_{hxl}^2,$$

$$U_{12} = \sum_{h=1}^H \psi_h \Delta_h \bar{x}_{hl} s_{hxl}^2,$$

$$U_{13} = \sum_{h=1}^H \psi_h \Delta_h \bar{x}_{hl} c_{hxl},$$

$$U_{23} = \sum_{h=1}^H \psi_h \Delta_h s_{hxl}^2 c_{hxl},$$

$$U_{10} = \sum_{h=1}^H \psi_h (\bar{X}_{hl} - \bar{x}_{hl}), \quad (17)$$

$$U_{20} = \sum_{h=1}^H \psi_h (S_{hxl}^2 - s_{hxl}^2),$$

$$U_{30} = \sum_{h=1}^H \psi_h (C_{hxl} - c_{hxl}).$$

Substituting the above values of λ_s^* , $s = 1, 2, 3$, in equation (14) and the resulting in equation (8) by setting $\Delta_h = 1$ yields the proposed estimator of calibration regression for population variance based on stratified double sampling as follows:

$$G_{st(p)} = \sum_{h=1}^H \psi_h s_{hyl}^2 + A_{1h(\alpha)} U_{10} + A_{2h(\alpha)} U_{20} + A_{3h(\alpha)} U_{30}, \quad (18)$$

$$\begin{aligned} A_{1h(\alpha)} &= \frac{\eta_{12} [\eta_{14} (\eta_{22} \eta_{33} - \eta_{23}^2) + \eta_{24} (\eta_{13} \eta_{23} - \eta_{12} \eta_{23}) + \eta_{34} (\eta_{12} \eta_{23} - \eta_{13} \eta_{22})]}{(\eta_{12}^2 - \eta_{11} \eta_{22})(\eta_{13} \eta_{23} - \eta_{12}^2) - (\eta_{13} \eta_{22} - \eta_{12} \eta_{23})(\eta_{12} \eta_{13} - \eta_{11} \eta_{23})}, \\ A_{2h(\alpha)} &= \frac{\eta_{12} [\eta_{14} (\eta_{13} \eta_{23} - \eta_{12} \eta_{33}) + \eta_{24} (\eta_{11} \eta_{33} - \eta_{13}^2) + \eta_{34} (\eta_{12} \eta_{13} - \eta_{11} \eta_{23})]}{(\eta_{12}^2 - \eta_{11} \eta_{22})(\eta_{13} \eta_{23} - \eta_{12}^2) - (\eta_{13} \eta_{22} - \eta_{12} \eta_{23})(\eta_{12} \eta_{13} - \eta_{11} \eta_{23})}, \\ A_{3h(\alpha)} &= \frac{\eta_{12} [\eta_{14} (\eta_{13} \eta_{22} - \eta_{12} \eta_{23}) + \eta_{24} (\eta_{12} \eta_{13} - \eta_{11} \eta_{23}) + \eta_{34} (\eta_{11} \eta_{22} - \eta_{12}^2)]}{(\eta_{12}^2 - \eta_{11} \eta_{22})(\eta_{13} \eta_{23} - \eta_{12}^2) - (\eta_{13} \eta_{22} - \eta_{12} \eta_{23})(\eta_{12} \eta_{13} - \eta_{11} \eta_{23})}, \end{aligned} \quad (19)$$

with

$$\begin{aligned} \eta_{11} &= \sum_{h=1}^H \psi_h \bar{x}_{hl}^2, \\ \eta_{22} &= \sum_{h=1}^H \psi_h s_{hxl}^4, \\ \eta_{33} &= \sum_{h=1}^H \psi_h c_{hxl}^2, \\ \eta_{12} &= \sum_{h=1}^H \psi_h \bar{x}_{hl} s_{hxl}^2, \\ \eta_{13} &= \sum_{h=1}^H \psi_h \bar{x}_{hl} c_{hxl}, \\ \eta_{14} &= \sum_{h=1}^H \psi_h \bar{x}_{hl} s_{hyl}^2, \\ \eta_{23} &= \sum_{h=1}^H \psi_h s_{hxl}^2 c_{hxl}, \\ \eta_{24} &= \sum_{h=1}^H \psi_h s_{hxl}^2 s_{hyl}^2, \\ \eta_{34} &= \sum_{h=1}^H \psi_h c_{hxl} s_{hyl}^2. \end{aligned} \quad (20)$$

4. Simulation Study

In this section, the performance of the proposed estimators will be evaluated by conducting a simulation study based on COVID-19 data.

where the regression coefficients $A_{1h(\alpha)}$, $A_{2h(\alpha)}$, and $A_{3h(\alpha)}$ are, respectively, given by

The coronavirus, which causes COVID-19 disease, was discovered in December 2019 in Wuhan, China, and affected almost all the countries of world. World's population is suffering with disease from last 20 months and the disease is not over yet. Coronaviruses are a broad family of viruses that cause disease in humans and animals.

Around the world, there are seven continents: Asia, Africa, Europe, North America, South America, Australia, and Antarctica. The nation of Asia is the largest continent in the world, followed by Africa, Europe, and North America in the second, third, and fourth rank, respectively. These four continents consist, respectively, of 49, 57, 48, and 39 countries. Furthermore, the total population of the world is 7.8 billion; however, 93.93% of this total population lives on these four continents. That is why we focus on adopting data on COVID-19 in these four main continents (strata), Africa (stratum I), Asia (stratum II), Europe (stratum III), and North America (stratum IV), compiled from the website (<https://www.worldometers.info>). The data are represented by two variables:

X = the total number of infections “total cases” per million

Y = total deaths per million population.

For each stratum, we draw the scatter plot (see Figures 1–4). These plots clearly demonstrate the existence of extreme values and are consequently fitted for our proposed estimator. Some essential characteristics of the COVID-19 data are listed in Table 2.

The simulation steps can be summarized as follows:

Step 1: based on SRSWOR, a random sample of size n_h is selected from stratum h .

FIGURE 1: Scatter plot for $h = 1$. FIGURE 2: Scatter plot for $h = 2$. FIGURE 3: Scatter plot for $h = 3$.

Step 2: through step 1, find the value of mean estimate (say) $\hat{\phi}' = \hat{\alpha}, G_{st(p)}$, where $\hat{\alpha} = F_{sti}, i = 1, 2, \dots, 6$.

Step 3: replicate step 1 and step 2 $R = 5000$ times. Accordingly, we obtain $\hat{\phi}'_1, \hat{\phi}'_2, \dots, \hat{\phi}'_R$.

Step 4: compute the MSE “mean square error” as

$$\text{MSE}(\hat{\phi}') = \frac{1}{R} \sum_{i=1}^R \left(\hat{\phi}' - \overline{\hat{\phi}'} \right)_i^2. \quad (21)$$

Step 5: compute the PRE “percentage relative efficiency” as

FIGURE 4: Scatter plot for $h = 4$.

TABLE 2: Characteristics of COVID-19 dataset.

Stratum I	Stratum II	Stratum III	Stratum IV
$N_1 = 57$	$N_2 = 49$	$N_3 = 48$	$N_4 = 39$
$\bar{X}_1 = 1317.193$	$\bar{X}_2 = 4443.388$	$\bar{X}_3 = 4923.375$	$\bar{X}_4 = 3530.026$
$\bar{Y}_1 = 21.28561$	$\bar{Y}_2 = 42.32857$	$\bar{Y}_3 = 206.9375$	$\bar{Y}_4 = 93.15385$
$S_{x1}^2 = 5280674$	$S_{x2}^2 = 64015965$	$S_{x3}^2 = 16273180$	$S_{x4}^2 = 22925589$
$S_{y1}^2 = 1335.731$	$S_{y2}^2 = 4020.823$	$S_{y3}^2 = 67572.27$	$S_{y4}^2 = 19987.61$
$C_{x1} = 1.744598$	$C_{x2} = 1.800653$	$C_{x3} = 0.8193573$	$C_{x4} = 1.356383$
$C_{y1} = 1.717012$	$C_{y2} = 1.498042$	$C_{y3} = 1.25616$	$C_{y4} = 1.517678$
$\rho_{xy1} = 0.916$	$\rho_{xy2} = 0.491$	$\rho_{xy3} = 0.605$	$\rho_{xy4} = 0.722$
$MR_{x1} = 5917.5$	$MR_{x2} = 20794.5$	$MR_{x3} = 10637$	$MR_{x4} = 9954.5$
$IDR_{x1} = 3666.2$	$IDR_{x2} = 13558.6$	$IDR_{x3} = 7614.4$	$IDR_{x4} = 9109.8$
$n_1 = 14$	$n_2 = 12$	$n_3 = 11$	$n_4 = 9$
$S_{1xl}^2 = 830703.2$	$S_{2xl}^2 = 11100059$	$S_{3xl}^2 = 4221600$	$S_{4xl}^2 = 5084386$
$S_{1yl}^2 = 203.0432$	$S_{2yl}^2 = 865.2452$	$S_{3yl}^2 = 15589.01$	$S_{4yl}^2 = 4371.333$
$C_{1xl} = 0.6919481$	$C_{2xl} = 0.7498052$	$C_{3xl} = 0.4173262$	$C_{4xl} = 0.6387654$
$C_{1yl} = 0.6694344$	$C_{2yl} = 0.6949219$	$C_{3yl} = 0.603351$	$C_{4yl} = 0.7097513$

TABLE 3: PRE of the proposed estimator with respect to adapted estimators.

ϕ'	$G_{st(p)}$
F_{st1}	191.2484
F_{st2}	195.5094
F_{st3}	180.7290
F_{st4}	187.3871
F_{st5}	196.4182
F_{st6}	199.2807

$$PRE(\hat{\phi}') = \frac{MSE(\hat{\alpha})}{MSE(G_{st(p)})} \times 100, \quad (22)$$

which is listed in Table 3.

5. Conclusions

The statistical analysis of the data as a comprehensive description is not complete by determining its shape or by identifying the measure of central tendency that fits it, but rather by determining the degree of spread of observations using an appropriate measure of dispersion.

Variance is one of the well-known and widely used measures of dispersion. Different variance estimators are built based on traditional moments that are especially influenced by the existence of extreme values. In this paper, with the existence of extreme values, we proposed some new estimators for the population variance based on L -moments and calibration strategies under double-stratified random sampling. A simulation study with COVID-19 data for the period January 22, 2020, up to August 23, 2020, is performed to evaluate the efficiency of new estimators. All results of the percentage relative efficiency (PRE) measure indicate that the proposed estimators are often superior and highly efficient ($PRE > 100$) compared to the existing traditional estimator. High PREs of the proposed estimators show the better performance. Furthermore, the proposed estimators F_{st5} and F_{st6} that are built on the coefficient of variation record the highest efficiency among other estimators.

Data Availability

The datasets used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

- [1] D. J. Watson, "The estimation of leaf area in field crops," *The Journal of Agricultural Science*, vol. 27, no. 3, pp. 474–483, 1937.
- [2] W. G. Cochran, "The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce," *The Journal of Agricultural Science*, vol. 30, no. 2, pp. 262–275, 1940.
- [3] M. Abid, N. Abbas, and M. Riaz, "Improved modified ratio estimators of population mean based on deciles," *Chiang Mai Journal of Science*, vol. 43, no. 1, pp. 1311–1323, 2016.
- [4] N. Ali, I. Ahmad, M. Hanif, and U. Shahzad, "Robust-regression-type estimators for improving mean estimation of sensitive variables by using auxiliary information," *Communications in Statistics-Theory and Methods*, vol. 50, no. 4, pp. 979–992, 2019.
- [5] M. Hanif and U. Shahzad, "Estimation of population variance using kernel matrix," *Journal of Statistics and Management Systems*, vol. 22, no. 3, pp. 563–586, 2019.
- [6] U. Shahzad, P. F. Perri, and M. Hanif, "A new class of ratio-type estimators for improving mean estimation of nonsensitive and sensitive variables by using supplementary information," *Communications in Statistics-Simulation and Computation*, vol. 48, no. 9, pp. 2566–2585, 2018.
- [7] U. Shahzad, N. H. Al-Noor, M. Hanif, and I. Sajjad, "An exponential family of median based estimators for mean estimation with simple random sampling scheme," *Communications in Statistics-Theory and Methods*, pp. 1–10, 2020.
- [8] U. Shahzad, N. H. Al-Noor, M. Hanif, I. Sajjad, and M. Muhammad Anas, "Imputation based mean estimators in case of missing data utilizing robust regression and variance-covariance matrices," *Communications in Statistics-Simulation and Computation*, pp. 1–20, 2020.
- [9] T. Zaman and H. Bulut, "Modified ratio estimators using robust regression methods," *Communications in Statistics-Theory and Methods*, vol. 48, no. 8, pp. 2039–2048, 2019.
- [10] T. Zaman and H. Bulut, "Modified regression estimators using robust regression methods and covariance matrices in stratified random sampling," *Communications in Statistics-Theory and Methods*, vol. 49, no. 14, pp. 3407–3420, 2020.
- [11] F. Gagnon, H. Lee, E. Rancourt, and C. E. Särndal, "Estimating the variance of the generalised regression estimator in the presence of imputation for the generalised estimation system," *Proceedings of the Survey Methods Section*, pp. 151–156, Statistical Society of Canada, Ottawa, Canada, 1997.
- [12] C. E. Särndal, B. Swensson, and J. Wretman, *Model Assisted Survey Sampling*, Springer-Verlag, Berlin, Germany, 1992.
- [13] A. I. Al-Omari, "Ratio estimation of the population mean using auxiliary information in simple random sampling and median ranked set sampling," *Statistics & Probability Letters*, vol. 82, no. 11, pp. 1883–1890, 2012.
- [14] J. R. M. Hosking, "L-moments: analysis and estimation of distributions using linear combinations of order statistics," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 52, no. 1, pp. 105–124, 1990.
- [15] J. C. Deville and C. E. Särndal, "Calibration estimators in survey sampling," *Journal of the American Statistical Association*, vol. 87, no. 418, pp. 376–382, 1992.
- [16] E. P. Clement, "Calibration approach separate ratio estimator for population mean in stratified sampling," *International Journal of Modern Mathematical Sciences*, vol. 13, no. 4, pp. 377–384, 2015.
- [17] N. Koyuncu, "Calibration estimator of population mean under stratified ranked set sampling design," *Communications in Statistics-Theory and Methods*, vol. 47, no. 23, pp. 5845–5853, 2018.
- [18] J. R. M. Hosking and J. R. Wallis, *Regional Frequency Analysis: an Approach Based on L-Moments*, Cambridge University Press, Cambridge, UK, 2005.
- [19] M. Subzar, A. Ibrahim Al-Omari, and R. A. Alanzi, "The robust regression methods for estimating of finite population mean based on SRSWOR in case of outliers," *Computers, Materials & Continua*, vol. 65, no. 1, pp. 125–138, 2020.
- [20] I. Almanjahie, A. Ibrahim Al-omari, E. Ekpenyong, and M. Subzar, "Generalized class of mean estimators with known measures for outlier treatment," *Computer Systems Science and Engineering*, vol. 38, no. 1, pp. 1–15, 2021.
- [21] M. Subzar, C. N. Bouza, and A. Ibrahim Al-Omari, "Utilization of different robust regression techniques for estimation of finite population mean in case of presence of outliers through ratio method of estimation," *Revista Investigacion Operacional*, vol. 40, no. 5, pp. 600–609, 2019.