Heliyon 9 (2023) e21427

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Enhancing mean estimators in median ranked set sampling with dual auxiliary information

Randa Alharbi [a], Manahil SidAhmed Mustafa [a], Aned Al Mutairi [b], Mohamed Hussein [c], M. Yusuf [d,*], Assem Elshenawy [e,f], Said G. Nassr [g]

[a] Department of Statistics, Faculty of Science, University of Tabuk, Tabuk 71491, Saudi Arabia
[b] Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[c] Department of Mathematics and Computer Science, Alexandria University, Alexandria 21544, Egypt
[d] Department of Mathematics, Faculty of Science, Helwan University, Cairo 11795, Egypt
[e] Department of Physics and Engineering Mathematics, Faculty of Engineering, Tanta University, Tanta 31733, Egypt
[f] Mathematics and Statistics Department, College of Science, Taibah University, Yanbu, Saudi Arabia
[g] Department of Statistics and Insurance, Faculty of Commerce, Arish University, Al-Arish 45511, Egypt

ARTICLE INFO

Keywords:
Mean estimation
Regression-type estimator
Dual auxiliary information
Median ranked set sampling

ABSTRACT

When measuring the research variable is complicated, expensive, or problematic, median ranked set sampling (MRSS) is often utilized since it is straightforward to rank the components using a low-cost sorting criterion. Using this sampling scheme, many authors considered the problem of population mean estimation with a single auxiliary variable in order to obtain more precised estimators than the traditional ratio type regression estimators. In this article, we extend their ideas based on regression approach using two auxiliary variables and introduce a new regression-type estimator along with its theoretical expression of minimum mean square error (MSE). The suggested estimator's applicability is demonstrated using both simulated and real-world data sets.

1. Introduction

In various practical experiments, especially in the fields of environment, ecology, medicine, and health research, observing the study variable, designated as Y, can often be challenging due to its intricate, expensive, repetitive, or even destructive measurement methods. However, despite the challenges or complexities associated with data collection, ranking the sampled elements can be relatively straightforward without incurring additional expenses. Let's consider this example: The rapid appearance of Calliphoridae flies on a decaying body shortly after death serves as a natural survival mechanism. In their postmortem investigations, forensic entomologists often depend on the larvae of these flies to estimate the time elapsed since death. Once the larvae reach their full size, they immediately stop consuming food. By examining the contents of the insects' intestines, particularly the absence of food in the front part of the intestine during subsequent development, forensic entomologists possess the ability to precisely ascertain the postmortem interval. However, the use of radiographic techniques to evaluate changes in the intestinal contents of maggots poses a significant challenge (Sharma et al., [1]). Meanwhile, evaluating and ordering the length of the larvae is comparatively straightforward as they demonstrate consistent growth throughout their entire life cycle. Another instance arises in a health-related

* Corresponding author.
E-mail addresses: ralharbi@ut.edu.sa (R. Alharbi), msida@ut.edu.sa (M.S. Mustafa), aoalmutairi@pnu.edu.sa (A. Al Mutairi), m.hussein@alexu.edu.eg (M. Hussein), mohammed.yousof@yahoo.com (M. Yusuf), asem.elshennawy@f-eng.tanta.edu.eg (A. Elshenawy), dr.saidstat@gmail.com (S.G. Nassr).

https://doi.org/10.1016/j.heliyon.2023.e21427
Received 2 August 2023; Received in revised form 3 October 2023; Accepted 20 October 2023

Available online 24 October 2023

research, aiming to estimate the average cholesterol level of a population. Instead of performing invasive blood tests on every individual within the sample, individuals can be ranked by their weight, even through visual observation alone. Consequently, a blood sample can be collected from only a limited number of subjects (Shahzad et al., [2]).

Initially, the ranked set sampling (RSS) was introduced by McIntyre [3] as an economical and efficacious sampling method to estimate the mean population of rummage yields and pastures. This sampling plan is appropriate when an auxiliary variable associated with the study variable can be ranked either visually or by any cost free criteria. It can be designed by choosing randomly ϑ samples of ϑ size elements from the population and then ranking all elements within each sample regarding the study variable. For an application of RSS, interested readers may refer to Johnson et al. [4] who considered RSS to develop vegetation research. Cobby et al. [5] considered RSS to investigate grass and grass-clover swards. Chen et al. [6,7] suggested an improved estimation procedure using RSS. Al-Omari [8,9] suggested modified estimation procedures in RSS. Shahzad et al. [10,11] put forward innovative idea in RSS, specifically addressing the issue of sensitivity. Additionally, Shahzad et al. [2] recommended the successive utilization of auxiliary information to estimate the population parameter using RSS. Sabry et al. [12] evaluated the performance of various RSS designs using a hybrid approach. Aljohani et al. [13] conducted a study on RSS, utilizing a modified Kies exponential distribution. Sabry et al. [12] examined parameter estimation of an exponential Pareto distribution in the context of RSS and double RSS designs. Yousef et al. [14] introduced a Bayesian estimation approach using the MCMC method to estimate the system reliability of the inverted ToppLeone distribution in RSS. Akhter et al. [15] developed the generalized Bilal distribution, discussed its properties, and investigated its estimation using RSS.

The modification towards the basic RSS design were suggested by Muttlak [16] i.e. median ranked set sampling (MRSS). The MRSS design that we adopted in this article as follows:

Step 1: Select Ψ_p random samples of size Ψ_p.
Step 2: Rank the elements within each sample via any visual inspection criteria.
Step 3: For odd sample size, choose the $((\Psi_p+1)/2)^{th}$ and $((\Psi_p+2)/2)^{th}$ smallest ordered elements from each set. For even sample size, choose the $(\Psi_p/2)^{th}$ and $((\Psi_p+2)/2)^{th}$ ordered elements from the first and remaining $\Psi_p/2$ sets.
Step 4: Repeat the above steps for achieving desired sample size.

For more details about the MRSS scheme with its steps, modifications and developments see, among others, Al-Omari [17], Koyuncu [18] and Alomair and Shahzad [19]. In sampling theory, when a sufficient positive correlation exists between the interest and auxiliary variables, the traditional ratio estimator is the most communal estimator of the population mean. In the present work, as an extension of ideas provided in references Al-Omari [17] and Koyuncu [18], we are going to propose a new estimator based on using two auxiliary variables.

This article has several sections that make up the remainder. The preliminaries and adapted estimators under MRSS based on two auxiliary variables are briefly introduced in Section 2. In Section 3, a new regression-type estimator along with its theoretical expression of minimum MSE is introduced. Section 4 is dedicated to highlighting the proposed estimator's efficiency. Different numerical investigations are done with an existing estimator based on simulation study and real-life data set. The most relevant conclusions drawn from the obtained results are finally set out in Section 5.

2. Preliminaries and adapted estimators under MRSS with two auxiliary variables

Within this section, we establish the MRSS design that incorporates a single study variable alongside two auxiliary variables. Let's consider a scenario where a median ranked set sample of size Ψ_p is extracted from a finite population denoted as Λ, which comprises N units. Let Y be study variable, X, W be the first and second supplementary variables, respectively.

Let $(X_{i(1)}, W_{i[1]}, Y_{i[1]}), (X_{i(2)}, W_{i[2]}, Y_{i[2]}), \ldots, (X_{i(\Psi_p)}, W_{i[\Psi_p]}, Y_{i[\Psi_p]})$ be the order statistics of $X_{i1}, X_{i2}, \ldots, X_{i\Psi_p}$ and the judgment order of $W_{i1}, W_{i2}, \ldots, W_{i\Psi_p}$; $Y_{i1}, Y_{i2}, \ldots, Y_{i\Psi_p}$ for $(i = 1, 2, \ldots, \Psi_p)$. Here, the symbols () and [] signify a perfect ranking for variable X and imperfect rankings for variables Y and W. We use the designations MRSSO and MRSSE to represent the units measured using MRSS for odd and even sample sizes, respectively.

For odd sample size, $(X_{1(\frac{1+\Psi_p}{2})}, W_{1[\frac{1+\Psi_p}{2}]}, Y_{1[\frac{1+\Psi_p}{2}]})$, $(X_{2(\frac{1+\Psi_p}{2})}, W_{2[\frac{1+\Psi_p}{2}]}, Y_{2[\frac{1+\Psi_p}{2}]}), \ldots, (X_{\Psi_p(\frac{1+\Psi_p}{2})}, W_{\Psi_p[\frac{1+\Psi_p}{2}]}, Y_{\Psi_p[\frac{1+\Psi_p}{2}]})$ denote the observed units by MRSSO. $\bar{x}_{MRSSO} = \frac{1}{\Psi_p}\sum_{i=1}^{\Psi_p} X_{i(\frac{1+\Psi_p}{2})}$, $\bar{w}_{MRSSO} = \frac{1}{\Psi_p}\sum_{i=1}^{\Psi_p} W_{i[\frac{1+\Psi_p}{2}]}$, $\bar{y}_{MRSSO} = \frac{1}{\Psi_p}\sum_{i=1}^{\Psi_p} Y_{i[\frac{1+\Psi_p}{2}]}$ be the MRSS means of X, W, Y, respectively. Further, $Var(\bar{x}_{MRSSO}) = \lambda_{(O)}\sigma^2_{x[\frac{1+\Psi_p}{2}]}$, $Var(\bar{w}_{MRSSO}) = \lambda_{(O)}\sigma^2_{w[\frac{1+\Psi_p}{2}]}$ and $Var(\bar{y}_{MRSSO}) = \lambda_{(O)}\sigma^2_{y[\frac{1+\Psi_p}{2}]}$ be the variances of $\bar{x}_{MRSSO}$, $\bar{w}_{MRSSO}$ and $\bar{y}_{MRSSO}$ respectively, where $\lambda_{(O)} = \dfrac{1}{\Psi_p}$.

For even sample size, $(X_{1(\frac{\Psi_p}{2})}, W_{1[\frac{\Psi_p}{2}]}, Y_{1[\frac{\Psi_p}{2}]})$, $(X_{2(\frac{\Psi_p}{2})}, W_{2[\frac{\Psi_p}{2}]}, Y_{2[\frac{\Psi_p}{2}]})$, $\ldots$, $(X_{\frac{\Psi_p}{2}(\frac{\Psi_p}{2})}, W_{\frac{\Psi_p}{2}[\frac{\Psi_p}{2}]}, Y_{\frac{\Psi_p}{2}[\frac{\Psi_p}{2}]})$, $(X_{\frac{2+\Psi_p}{2}(\frac{2+\Psi_p}{2})}, W_{\frac{2+\Psi_p}{2}[\frac{2+\Psi_p}{2}]}$, $Y_{\frac{2+\Psi_p}{2}[\frac{2+\Psi_p}{2}]})$, $(X_{\frac{4+n}{2}(\frac{2+\Psi_p}{2})}, W_{\frac{4+n}{2}[\frac{2+\Psi_p}{2}]})$, $Y_{\frac{4+n}{2}[\frac{2+\Psi_p}{2}]})$, $\ldots$, $(X_{n(\frac{\Psi_p}{2})}, W_{n[\frac{\Psi_p}{2}]}, Y_{n[\frac{\Psi_p}{2}]})$ denote the observed units by MRSSE. Further

$$\bar{x}_{MRSSE} = \frac{1}{\Psi_p}(\sum_{i=1}^{\frac{\Psi_p}{2}} X_{i(\frac{\Psi_p}{2})} + \sum_{i=\frac{2+\Psi_p}{2}}^{\Psi_p} X_{i(\frac{2+\Psi_p}{2})}),$$

$$\bar{w}_{MRSSE} = \frac{1}{\Psi_p}\left(\sum_{i=1}^{\frac{\Psi_p}{2}} W_{i[\frac{\Psi_p}{2}]} + \sum_{i=\frac{2+\Psi_p}{2}}^{\Psi_p} W_{i[\frac{2+\Psi_p}{2}]}\right),$$

$$\bar{y}_{MRSSE} = \frac{1}{\Psi_p}\left(\sum_{i=1}^{\frac{\Psi_p}{2}} Y_{i[\frac{\Psi_p}{2}]} + \sum_{i=\frac{2+\Psi_p}{2}}^{\Psi_p} Y_{i[\frac{2+\Psi_p}{2}]}\right),$$

be the MRSS mean of X, W and Y respectively. Further, $Var(\bar{x}_{MRSSE}) = \lambda_{(E)}(\sigma^2_{x[\frac{\Psi_p}{2}]} + \sigma^2_{x[\frac{2+\Psi_p}{2}]})$, $Var(\bar{w}_{MRSSE}) = \lambda_{(E)}(\sigma^2_{w[\frac{\Psi_p}{2}]} + \sigma^2_{w[\frac{2+\Psi_p}{2}]})$ and $Var(\bar{y}_{MRSSE}) = \lambda_{(E)}(\sigma^2_{y[\frac{\Psi_p}{2}]} + \sigma^2_{y[\frac{2+\Psi_p}{2}]})$ be the variances of $\bar{x}_{MRSSE}, \bar{w}_{MRSSE}$ and $\bar{y}_{MRSSE}$ respectively, where $\lambda_{(E)} = \frac{1}{2\Psi_p}$.

In order to enhance the efficiency of estimators, it is possible to leverage auxiliary information linked to an auxiliary variable X, which exhibits correlation with the study variable Y. Typically, the mean of X is employed as auxiliary information. However, to further improve the efficiency of estimators, additional auxiliary information concerning variable X, such as the median, coefficient of variation, or correlation coefficient, can also be utilized. Furthermore, when a positive correlation is present between variable Y and variable X, the ratio estimator proves to be effective. There are many realistic circumstances (industrial, economical, biological, and medical) where a positive correlation exists between the two variables, it has been shown that there is a direct positive relationship between (i) the volume of production and the volume of crude oil exports, (ii) the students' grades in Statistics and their grades in Mathematics, (iii) the human body's immunity from the risk of certain diseases and paying attention to fitness, etc.

In this context, we are modifying the conventional regression estimator by incorporating two auxiliary variables within the MRSS framework.

$$\bar{y}_{rg(j)} = \overline{y}_{MRSS(j)} + b_{yx(j)}(\mu_x - \overline{x}_{MRSS(j)}) + b_{yw(j)}(\mu_w - \overline{w}_{MRSS(j)})$$

$$MSE(\bar{y}_{rg(j)}) \cong \begin{cases} \text{if } \Psi_p \text{ is odd,} \\ \lambda_{(O)}\sigma^2_{y[\frac{1+\Psi_p}{2}]}[\rho^2_{xw[\frac{1+\Psi_p}{2}]} + (1-\rho^2_{xw[\frac{1+\Psi_p}{2}]})(1-R^2_{(O)})] \\ \text{if } \Psi_p \text{ is even,} \\ \lambda_{(E)}(\sigma^2_{y[\frac{\Psi_p}{2}]} + \sigma^2_{y[\frac{2+\Psi_p}{2}]})[(\rho^2_{xw[\frac{1+\Psi_p}{2}]} + \rho^2_{xw[\frac{2+\Psi_p}{2}]}) + \{1-(\rho^2_{xw[\frac{1+\Psi_p}{2}]} + \rho^2_{xw[\frac{2+\Psi_p}{2}]})\}(1-R^2_{(E)})] \end{cases}$$

where μ_x, μ_w be the population means and $(b_{xy(j)}, b_{wy(j)})$ representing regression coefficients. Taking into account the selection of odd and even samples, these estimators can be slimmed as follows, with the notation $j = (E, O)$ denoting even and odd sample selection, respectively. Furthermore,

$$R^2_{(O)} = \frac{\rho^2_{xy[\frac{1+\Psi_p}{2}]} + \rho^2_{wy[\frac{1+\Psi_p}{2}]} - 2\rho_{xy[\frac{1+\Psi_p}{2}]}\rho_{wy[\frac{1+\Psi_p}{2}]}\rho_{xw[\frac{1+\Psi_p}{2}]}}{1-\rho^2_{xw[\frac{1+\Psi_p}{2}]}},$$

$$R^2_{(E)} = \frac{(\rho^2_{xy[\frac{\Psi_p}{2}]} + \rho^2_{xy[\frac{2+\Psi_p}{2}]}) + (\rho^2_{wy[\frac{\Psi_p}{2}]} + \rho^2_{wy[\frac{2+\Psi_p}{2}]}) - 2(\rho_{xy[\frac{\Psi_p}{2}]} + \rho_{xy[\frac{2+\Psi_p}{2}]})(\rho_{wy[\frac{\Psi_p}{2}]} + \rho_{wy[\frac{2+\Psi_p}{2}]})(\rho_{xw[\frac{\Psi_p}{2}]} + \rho_{xw[\frac{2+\Psi_p}{2}]})}{1-(\rho^2_{xw[\frac{\Psi_p}{2}]} + \rho^2_{xw[\frac{2+\Psi_p}{2}]})}.$$

3. Proposed estimator

In the context of MRSS (Multivariate Randomized Stratified Sampling), Al-Omari [17] introduced novel estimators of the mean population based on ratios, utilizing the mean, as well as the first and third quartiles of an auxiliary variable. Building upon Al-Omari's idea, Koyuncu [18] extended the concept to introduce regression, exponential, and difference type estimators using a single auxiliary variable in MRSS. In the realm of simple random sampling (SRS), Rao [20] incorporated two tuning parameters to implement a generalized form of the traditional regression estimator. Drawing inspiration from Rao [20], Al-Omari [17], and Koyuncu [18], we propose a regression-type estimator under MRSS that leverages two auxiliary variables, namely, (X, W).

$$\bar{y}_{P(j)} = k_{1(j)}\overline{y}_{MRSS(j)} + k_{2(j)}(\mu_x - \overline{x}_{MRSS(j)}) + k_{3(j)}(\mu_w - \overline{w}_{MRSS(j)})$$

In order to calculate the bias and mean square error (MSE), we can define the following:

$$\varepsilon_{0(j)} = \frac{\overline{y}_{MRSS(j)} - \mu_y}{\mu_y},\ \varepsilon_{1(j)} = \frac{\overline{x}_{MRSS(j)} - \mu_x}{\mu_x},\ \varepsilon_{2(j)} = \frac{\bar{w}_{MRSS(j)} - \mu_w}{\mu_w}.$$

In the case where the sample size Ψ_p is odd, we can express it as follows:

$$E(\varepsilon^2_{0(O)}) = \frac{\lambda_{(O)}}{\mu^2_y}\sigma^2_{y[\frac{1+\Psi_p}{2}]},\ E(\varepsilon^2_{1(O)}) = \frac{\lambda_{(O)}}{\mu^2_x}\sigma^2_{x(\frac{1+\Psi_p}{2})},\ E(\varepsilon^2_{2(O)}) = \frac{\lambda_{(O)}}{\mu^2_w}\sigma^2_{w[\frac{1+\Psi_p}{2}]},$$

$$E(\varepsilon_{0(O)}\varepsilon_{1(O)}) = \frac{\lambda_{(O)}}{\mu_x\mu_y}\sigma_{xy[\frac{1+\Psi_p}{2}]},\ E(\varepsilon_{0(O)}\varepsilon_{2(O)}) = \frac{\lambda_{(O)}}{\mu_w\mu_y}\sigma_{wy[\frac{1+\Psi_p}{2}]},\ E(\varepsilon_{1(O)}\varepsilon_{2(O)}) = \frac{\lambda_{(O)}}{\mu_w\mu_x}\sigma_{wx[\frac{1+\Psi_p}{2}]}.$$

In the case where the sample size Ψ_p is even, we can express it as follows:

$$E(\varepsilon^2_{0(E)}) = \frac{\lambda_{(E)}}{\mu_y^2}(\sigma^2_{y[\frac{\Psi_p}{2}]} + \sigma^2_{y[\frac{2+\Psi_p}{2}]}),\ E(\varepsilon^2_{1(E)}) = \frac{\lambda_{(E)}}{\mu_x^2}(\sigma^2_{x(\frac{\Psi_p}{2})} + \sigma^2_{x(\frac{2+\Psi_p}{2})}),\ E(\varepsilon^2_{2(E)}) = \frac{\lambda_{(E)}}{\mu_w^2}(\sigma^2_{w[\frac{\Psi_p}{2}]} + \sigma^2_{w[\frac{2+\Psi_p}{2}]}),$$

$$E(\varepsilon_{0(E)}\varepsilon_{1(E)}) = \frac{\lambda_{(E)}}{\mu_x\mu_y}(\sigma_{yx[\frac{\Psi_p}{2}]} + \sigma_{yx[\frac{2+\Psi_p}{2}]}),\ E(\varepsilon_{0(E)}\varepsilon_{2(E)}) = \frac{\lambda_{(E)}}{\mu_w\mu_y}(\sigma_{wy[\frac{\Psi_p}{2}]} + \sigma_{wy[\frac{2+\Psi_p}{2}]}),\ E(\varepsilon_{1(E)}\varepsilon_{2(E)}) = \frac{\lambda_{(E)}}{\mu_w\mu_x}(\sigma_{wx[\frac{\Psi_p}{2}]} + \sigma_{wx[\frac{2+\Psi_p}{2}]}).$$

By representing the proposed estimator $\bar{y}_{P(j)}$ in terms of ε's, we obtain the following expression

$$\bar{y}_{P(j)} - \mu_y = k_{1(j)}\mu_y(1+\varepsilon_{0(j)}) - k_{2(j)}\mu_x\varepsilon_{1(j)} - k_{3(j)}\mu_w\varepsilon_{2(j)} - \mu_y \tag{1}$$

Taking square of equation (1) and applying expectation, we get MSE expressions of $\bar{y}_{P1}$. Subsequently, by taking partial derivatives of the MSE and determining the optimal values of weights, denoted as k_1, k_2, and k_3, we can substitute these values into the MSE expressions, resulting in minimum MSE expressions. Briefly, we are providing the expressions of optimum weights and minimum MSE as follows

$$k_1 \cong \begin{cases} \dfrac{\left[-1+\rho^2_{xw[\frac{1+\Psi_p}{2}]}\right]}{\left[-1+\rho^2_{xw[\frac{1+\Psi_p}{2}]} + \lambda_{(O)}(\mu_y^2)^{-1}\sigma^2_{y[\frac{1+\Psi_p}{2}]}\left\{P_{(O)}\right\}\right]} & \text{if } \Psi_p \text{ is odd} \\[2ex] \dfrac{\left[-1+(\rho^2_{xw[\frac{\Psi_p}{2}]} + \rho^2_{xw[\frac{2+\Psi_p}{2}]})\right]}{\left[-1+(\rho^2_{xw[\frac{\Psi_p}{2}]} + \rho^2_{xw[\frac{2+\Psi_p}{2}]}) + \lambda_{(E)}(\mu_y^2)^{-1}(\sigma^2_{y[\frac{\Psi_p}{2}]} + \sigma^2_{y[\frac{2+\Psi_p}{2}]})\left\{P_{(E)}\right\}\right]} & \text{if } \Psi_p \text{ is even} \end{cases}$$

$$k_2 \cong \begin{cases} \dfrac{\sigma_{y[\frac{1+\Psi_p}{2}]}\left[-\rho_{yx[\frac{1+\Psi_p}{2}]} + \rho_{xw[\frac{1+\Psi_p}{2}]}\rho_{yw[\frac{1+\Psi_p}{2}]}\right]}{\sigma_{x(\frac{1+\Psi_p}{2})}\left[-1+\rho^2_{xw[\frac{1+\Psi_p}{2}]} + \lambda_{(O)}(\mu_y^2)^{-1}\sigma^2_{y[\frac{1+\Psi_p}{2}]}\left\{P_{(O)}\right\}\right]} & \text{if } \Psi_p \text{ is odd} \\[2ex] \dfrac{(\sigma_{y[\frac{\Psi_p}{2}]} + \sigma_{y[\frac{2+\Psi_p}{2}]})\left[-(\rho_{yx[\frac{\Psi_p}{2}]} + \rho_{yx[\frac{2+\Psi_p}{2}]}) + (\rho_{xw[\frac{\Psi_p}{2}]} + \rho_{xw[\frac{2+\Psi_p}{2}]})(\rho_{yw[\frac{\Psi_p}{2}]} + \rho_{yw[\frac{2+\Psi_p}{2}]})\right]}{(\sigma_{x(\frac{\Psi_p}{2})} + \sigma_{x(\frac{2+\Psi_p}{2})})\left[-1+(\rho^2_{xw[\frac{\Psi_p}{2}]} + \rho^2_{xw[\frac{2+\Psi_p}{2}]}) + \lambda_{(E)}(\mu_y^2)^{-1}(\sigma^2_{y[\frac{\Psi_p}{2}]} + \sigma^2_{y[\frac{2+\Psi_p}{2}]})\left\{P_{(E)}\right\}\right]} & \text{if } \Psi_p \text{ is even} \end{cases}$$

$$k_3 \cong \begin{cases} \dfrac{\sigma_{y[\frac{1+\Psi_p}{2}]}\left[-\rho_{yw[\frac{1+\Psi_p}{2}]} + \rho_{xw[\frac{1+\Psi_p}{2}]}\rho_{yx[\frac{1+\Psi_p}{2}]}\right]}{\sigma_{w[\frac{1+\Psi_p}{2}]}\left[-1+\rho^2_{xw[\frac{1+\Psi_p}{2}]} + \lambda_{(O)}(\mu_y^2)^{-1}\sigma^2_{y[\frac{1+\Psi_p}{2}]}\left\{P_{(O)}\right\}\right]} & \text{if } \Psi_p \text{ is odd} \\[2ex] \dfrac{(\sigma_{y[\frac{\Psi_p}{2}]} + \sigma_{y[\frac{2+\Psi_p}{2}]})\left[-(\rho_{yw[\frac{\Psi_p}{2}]} + \rho_{yw[\frac{2+\Psi_p}{2}]}) + (\rho_{xw[\frac{\Psi_p}{2}]} + \rho_{xw[\frac{2+\Psi_p}{2}]})(\rho_{yx[\frac{\Psi_p}{2}]} + \rho_{yx[\frac{2+\Psi_p}{2}]})\right]}{(\sigma_{w[\frac{\Psi_p}{2}]} + \sigma_{w[\frac{2+\Psi_p}{2}]})\left[-1+(\rho^2_{xw[\frac{\Psi_p}{2}]} + \rho^2_{xw[\frac{2+\Psi_p}{2}]}) + \lambda_{(E)}(\mu_y^2)^{-1}(\sigma^2_{y[\frac{\Psi_p}{2}]} + \sigma^2_{y[\frac{2+\Psi_p}{2}]})\left\{P_{(E)}\right\}\right]} & \text{if } \Psi_p \text{ is even} \end{cases}$$

where ρ symbols are donating the correlations between different variables as mentioned in their subscripts and

$$P_{(O)} = -1 + \rho^2_{xw[\frac{1+\Psi_p}{2}]} + \rho^2_{yx[\frac{1+\Psi_p}{2}]} + \rho^2_{yw[\frac{1+\Psi_p}{2}]} - 2\rho_{xw[\frac{1+\Psi_p}{2}]}\rho_{yx[\frac{1+\Psi_p}{2}]}\rho_{yw[\frac{1+\Psi_p}{2}]},$$

$$\begin{aligned} P_{(E)} = {} & -1 + (\rho^2_{xw[\frac{\Psi_p}{2}]} + \rho^2_{xw[\frac{2+\Psi_p}{2}]}) + (\rho^2_{yx[\frac{\Psi_p}{2}]} + \rho^2_{yx[\frac{2+\Psi_p}{2}]}) + (\rho^2_{yw[\frac{\Psi_p}{2}]} + \rho^2_{yw[\frac{2+\Psi_p}{2}]}) \\ & - 2(\rho_{xw[\frac{\Psi_p}{2}]} + \rho_{xw[\frac{2+\Psi_p}{2}]})(\rho_{yx[\frac{\Psi_p}{2}]} + \rho_{yx[\frac{2+\Psi_p}{2}]})(\rho_{yw[\frac{\Psi_p}{2}]} + \rho_{yw[\frac{2+\Psi_p}{2}]}). \end{aligned}$$

Table 1
MSE and PRE for BMI data.

$\hat{\theta}$		MSE	PRE		MSE	PRE
$\bar{y}_{rg(j)}$	$\Psi_p = 4$	4.8672	131.2783	$\Psi_p = 5$	3.6744	187.0782
$\bar{y}_{P(j)}$	$\Psi_p = 4$	2.3970	266.5606	$\Psi_p = 5$	2.3579	291.5304
$\bar{y}_{rg(j)}$	$\Psi_p = 6$	5.8679	105.7351	$\Psi_p = 7$	4.6492	144.4548
$\bar{y}_{P(j)}$	$\Psi_p = 6$	2.3330	240.7880	$\Psi_p = 7$	2.2970	292.3741
$\bar{y}_{rg(j)}$	$\Psi_p = 8$	5.5604	118.3864	$\Psi_p = 9$	4.1161	170.4057
$\bar{y}_{P(j)}$	$\Psi_p = 8$	2.2692	290.0866	$\Psi_p = 9$	2.0122	364.1778
$\bar{y}_{rg(j)}$	$\Psi_p = 10$	8.3374	103.9473	$\Psi_p = 11$	4.5006	144.9507
$\bar{y}_{P(j)}$	$\Psi_p = 10$	2.2288	276.6068	$\Psi_p = 11$	2.2121	294.9074
$\bar{y}_{rg(j)}$	$\Psi_p = 12$	4.7780	131.0691	$\Psi_p = 13$	5.1339	115.1614
$\bar{y}_{P(j)}$	$\Psi_p = 12$	2.2106	283.2897	$\Psi_p = 13$	2.1936	269.5280

$$MSE(\bar{y}_{P(j)}) \cong \begin{cases} \text{if } \Psi_p \text{ is odd,} \\ \dfrac{\left[\lambda_{(O)}\sigma^2_{y[\frac{1+\Psi_p}{2}]}(1-R^2_{(O)})\right]}{\left[1+\lambda_{(O)}(\mu_y^2)^{-1}\sigma^2_{y[\frac{1+\Psi_p}{2}]}(1-R^2_{(O)})\right]} \\ \text{if } \Psi_p \text{ is even,} \\ \dfrac{\left[\lambda_{(E)}(\sigma^2_{y[\frac{\Psi_p}{2}]}+\sigma^2_{y[\frac{2+\Psi_p}{2}]})(1-R^2_{(E)})\right]}{\left[1+\lambda_{(E)}(\mu_y^2)^{-1}(\sigma^2_{y[\frac{\Psi_p}{2}]}+\sigma^2_{y[\frac{2+\Psi_p}{2}]})(1-R^2_{(E)})\right]} \end{cases}$$

Finally comparing, $MSE(\bar{y}_{P(j)})$ with $MSE(\bar{y}_{rg(j)})$, we immediately observe that

$$MSE(\bar{y}_{P(j)}) \leq MSE(\bar{y}_{rg(j)}). \tag{2}$$

4. Numerical illustration

In this section, simulation study and real-life application, have been conducted to demonstrate the performance of the proposed estimator over the exiting regression estimator.

4.1. Simulation study

Present sub-section developed to assess the efficiency of the regression and proposed Rao-type estimators, based on simulation study. So for comparisons a multi-variate normal distribution for (Y, X, W) with mean vector $(\bar{Y}, \bar{X}, \bar{W}) = (5.1, 5.1, 5.1)$ and variance-covariance matrix respectively as given by

$$\Sigma = \begin{bmatrix} 11.9 & 2.7 & 2.8 \\ 2.7 & 2.9 & 1.0 \\ 2.8 & 1.0 & 2.9 \end{bmatrix}, \quad \rho_{yx} = 0.4957, \quad \rho_{yw} = 0.5076.$$

From the population, $K = 9000$ samples of even size $(4, 6, 8, 10, 12)$, and odd sizes $(5, 7, 9, 11, 13)$ are selected according to MRSS and for the kth sample, the estimate $(\bar{y}_{rg(j)}, \bar{y}_{P(j)})$ of $\bar{Y}$ is computed. In this way, for each considered estimator, the MSE is obtained as $\text{MSE}(\hat{\theta}) = \sum_{k=1}^{K}(\hat{\theta}^{(k)} - \bar{Y})^2/K$, where $\hat{\theta}^{(k)}$ is denoted by $(\bar{y}_{rg(j)}, \bar{y}_{P(j)})$ estimators. The percentage relative efficiency (PRE) is computed for comparison purposes

$$\text{PRE}(\hat{\theta}) = \frac{Var(\bar{y}_{MRSS(j)})}{\text{MSE}(\hat{\theta}^{(k)})} \times 100.$$

For whole simulation procedure we followed Shahzad et al. [21]. The results of MSE and PRE associated with artificial data are presented numerically in Table 1. It appears clearly that the values of MSE associated with the proposed estimator are smaller than the corresponding ones with the traditional regression estimator. The PRE of the proposed estimator with respect to the corresponding traditional regression estimator based on MRSS is greater than 100. The PRE associated with even samples is greater than odd samples.

5. Real life application

This subsection is based on a real-life dataset aimed at evaluating the effectiveness of traditional regression estimators and the proposed Rao-type regression estimators. To conduct this evaluation, we utilize a health survey dataset prepared by the Turkish

Table 2
MSE and PRE for artificial data.

$\hat{\theta}$		MSE	PRE		MSE	PRE
$\bar{y}_{rg(j)}$	$\Psi_p = 4$	4.747444	228.662354	$\Psi_p = 5$	4.245515	221.540652
$\bar{y}_{P(j)}$	$\Psi_p = 4$	2.578595	420.989627	$\Psi_p = 5$	2.529242	371.872038
$\bar{y}_{rg(j)}$	$\Psi_p = 6$	4.122489	250.131596	$\Psi_p = 7$	3.882289	243.397792
$\bar{y}_{P(j)}$	$\Psi_p = 6$	2.489604	414.188197	$\Psi_p = 7$	2.442237	386.915948
$\bar{y}_{rg(j)}$	$\Psi_p = 8$	3.816635	275.628783	$\Psi_p = 9$	3.670095	241.236521
$\bar{y}_{P(j)}$	$\Psi_p = 8$	2.423363	434.096863	$\Psi_p = 9$	2.381172	371.817250
$\bar{y}_{rg(j)}$	$\Psi_p = 10$	3.715528	338.540145	$\Psi_p = 11$	3.602003	272.024404
$\bar{y}_{P(j)}$	$\Psi_p = 10$	2.396768	524.813287	$\Psi_p = 11$	2.356845	415.739182
$\bar{y}_{rg(j)}$	$\Psi_p = 12$	3.587764	281.505522	$\Psi_p = 13$	3.462109	264.823833
$\bar{y}_{P(j)}$	$\Psi_p = 12$	2.355186	428.830371	$\Psi_p = 13$	2.307687	397.302198

Statistical Institute (TSI), which examines the factors that may influence obesity-related behaviors in Turkey, consisting of data from 800 individuals. This dataset has also been recently explored by Cetin and Koyuncu [22]. In our analysis, the study variable (Y) is the Body Mass Index (BMI), while the weight (X) and age (W) serve as auxiliary variables. We employ a multivariate randomized stratified sampling (MRSS) scheme to select samples of both even sizes (4, 6, 8, 10, 12) and odd sizes (5, 7, 9, 11, 13).

The collected data consist of $N = 800$ observations with $\rho_{xy} = 0.867$, $\rho_{wy} = 0.609$, $\rho_{xw} = 0.483$, $\mu_y = 23.77678$, $\mu_x = 67.55813$, $\mu_w = 30.12625$, $\sigma_x = 13.8396$, $\sigma_y = 4.195531$, $\sigma_w = 11.03823$.

The results of MSE and PRE associated with BMI data are presented numerically in Table 2. Clearly, it appears that the values of MSE associated with the proposed estimator are smaller than the corresponding ones with the traditional regression estimator. The PRE of the proposed estimator with respect to the corresponding traditional regression estimator based on MRSS is greater than 100. The PRE associated with samples of odd size is greater than the corresponding ones with even samples except for $\Psi_p = 13$ compared to $\Psi_p = 12$, where the reverse is true.

Based on the results of numerical illustrations shown in Tables 1 and 2, for all samples under consideration, we noticed that the condition given in equation (2) has been satisfied. So, the performance of the proposed Rao-type estimator always outperforms the performance of the existing and adapted estimators.

6. Conclusion

In this research paper, drawing inspiration from Rao [20], Al-Omari [17], and Koyuncu [18], we introduce a novel Rao regression-type estimator designed to enhance mean estimation in median ranked set sampling (MRSS) using two auxiliary variables. We assess the effectiveness of this proposed estimator by analyzing both simulated and real-life datasets, comprising samples of both even and odd sizes. The real-life dataset, previously studied by Cetin and Koyuncu [22], focused on examining the factors influencing obesity in a population of 800 individuals in Turkey. In our study, we consider the Body Mass Index (BMI) of each individual as the variable of interest, with weight and age serving as auxiliary variables. The results demonstrate that the proposed Rao-type estimator consistently outperforms the traditional ratio regression estimator, as demonstrated in the theoretical comparison section. Based on the findings from both real-life application and simulation illustrations, we conclude that the proposed Rao regression-type estimator is a valuable alternative to the traditional regression estimator. We highly recommend its usage, as it can significantly improve the estimation of the population mean in MRSS when employing two auxiliary variables.

CRediT authorship contribution statement

Randa Alharbi: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Manahil SidAhmed Mustafa:** Conceptualization, Data curation, Formal analysis, Validation, Visualization, Writing – original draft, Writing – review & editing. **Aned Al Mutairi:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Software, Writing – original draft, Writing – review & editing. **Mohamed Hussein:** Conceptualization, Data curation, Formal analysis, Software, Validation, Writing – original draft, Writing – review & editing. **M. Yusuf:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Assem Elshenawy:** Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. **Said G. Nassr:** Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R368), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

References

[1] R. Sharma, R.K. Garg, J.R. Gaur, Various methods for the estimation of the post mortem interval from Calliphoridae: a review, Egypt. J. Forensic Sci. 5 (1) (2015) 1–12.
[2] U. Shahzad, I. Ahmad, E. Oral, M. Hanif, I. Almanjahie, Estimation of the population mean by successive use of an auxiliary variable in median ranked set sampling, Math. Popul. Stud. 28 (3) (2021) 176–199.
[3] G.A. McIntyre, A method for unbiased selective sampling, using ranked sets, Aust. J. Agric. Res. 3 (4) (1952) 385–390.
[4] G.D. Johnson, G.P. Patil, A.K. Sinha, Ranked set sampling for vegetation research, in: Abstracta Botanica, 1993, pp. 87–102.
[5] J.M. Cobby, M.S. Ridout, P.J. Bassett, R.V. Large, An investigation into the use of ranked set sampling on grass and grass-clover swards, Grass Forage Sci. 40 (3) (1985) 257–263.
[6] Z. Chen, Z. Bai, B.K. Sinha, Ranked Set Sampling: Theory and Applications, Springer-Verlag, New York, 2004.
[7] H. Chen, E.A. Stasny, D.A. Wolfe, Improved procedures for estimation of disease prevalence using ranked set sampling, Biom. J. 49 (4) (2007) 530–538.
[8] A.I. Al-Omari, Estimation of mean based on modified robust extreme ranked set sampling, J. Stat. Comput. Simul. 81 (8) (2011) 1055–1066.
[9] A.I. Al-Omari, A new measure of sample entropy of continuous random variable, J. Stat. Theory Pract. 10 (4) (2016) 721–735.
[10] U. Shahzad, M. Hanif, N. Koyuncu, A.V. Garcia Luengo, A new estimator of mean under ranked set sampling alongside the sensitivity issue, J. Stat. Manag. Syst. 21 (8) (2018) 1553–1564.
[11] U. Shahzad, M. Hanif, N. Koyuncu, A.V.G. Luengo, A regression type estimator for mean estimation under ranked set sampling alongside the sensitivity issue, Commun. Fac. Sci. Univ. Ank. Sér. A1 Math. Stat. 68 (2) (2019) 2037–2049.
[12] M.A. Sabry, E.M. Almetwally, H.M. Almongy, G.M. Ibrahim, Assessing the performance of some ranked set sampling designs using hybrid approach, Comput. Mater. Continua 68 (3) (2021) 3737–3753.
[13] H.M. Aljohani, E.M. Almetwally, A.S. Alghamdi, E.H. Hafez, Ranked set sampling with application of modified Kies exponential distribution, Alex. Eng. J. 60 (4) (2021) 4041–4046.
[14] M.M. Yousef, A.S. Hassan, A.H. Al-Nefaie, E.M. Almetwally, H.M. Almongy, Bayesian estimation using MCMC method of system reliability for inverted Topp–Leone distribution based on ranked set sampling, Mathematics 10 (17) (2022) 3122.
[15] Z. Akhter, E.M. Almetwally, C. Chesneau, On the generalized Bilal distribution: some properties and estimation under ranked set sampling, Axioms 11 (4) (2022) 173.
[16] H.A. Muttlak, Median ranked set sampling, J. Appl. Statist. Sc. 6 (1997) 245–255.
[17] A.I. Al-Omari, Ratio estimation of the population mean using auxiliary information in simple random sampling and median ranked set sampling, Stat. Probab. Lett. 82 (11) (2012) 1883–1890.
[18] N. Koyuncu, New difference-cum-ratio and exponential type estimators in median ranked set sampling, Hacet. J. Math. Stat. 45 (1) (2016) 207–225.
[19] A.M. Alomair, U. Shahzad, Optimizing mean estimators with calibrated minimum covariance determinant in median ranked set sampling, Symmetry 15 (8) (2023) 1581.
[20] T.J. Rao, On certail methods of improving ration and regression estimators, Commun. Stat., Theory Methods 20 (10) (1991) 3325–3340.
[21] U. Shahzad, I. Ahmad, I. Almanjahie, N.H. Al-Noor, M. Hanif, Mean estimation using robust quantile regression with two auxiliary variables, Sci. Iran. 30 (3) (2023) 1245–1254.
[22] A.E. Cetin, N. Koyuncu, Estimation of population mean under different stratified ranked set sampling designs with simulation study application to BMI data, Commun. Fac. Sci. Univ. Ank. Sér. A1 Math. Stat. 69 (1) (2020) 560–575.