

Article

Composite Estimators for the Population Mean Under Ranked Set Sampling

Sayed A. Mostafa , Thomas Johnson III and Sanford Jennings

Department of Mathematics & Statistics, North Carolina A&T State University, Greensboro, NC 27411, USA; thjohnson@aggies.ncat.edu (T.S.III); sljennings@aggies.ncat.edu (S.J.)

* Correspondence: sabdelmegeed@ncat.edu

Abstract

Ranked Set Sampling (RSS) is an effective sampling technique, particularly when precise measurement of the study variable is costly or time-consuming, but ranking the units is relatively easy. Estimation of the finite population mean from RSS, with and without auxiliary information, has been widely studied, with estimators such as the RSS sample mean, ratio estimator, and regression estimators receiving considerable attention. While the RSS sample mean does not utilize auxiliary information, the ratio and regression estimators rely heavily on its quality. To address these limitations, this study proposes a shrinkage-type (composite) estimator for the finite population mean. The proposed estimator adaptively combines the RSS sample mean and the ratio estimator, leveraging auxiliary information when it is useful while maintaining robustness when it is not. We derive its statistical properties, including bias, variance, and mean squared error. Simulation studies demonstrate that the proposed estimator can outperform conventional estimators across a range of scenarios. We illustrate the method through a real data application.

Keywords: composite estimators; ranked set sampling; ratio estimator; regression estimator; simple random sampling

MSC: 62D05



Academic Editor: Jiancang Zhuang

Received: 22 July 2025

Revised: 17 September 2025

Accepted: 19 September 2025

Published: 24 September 2025

Citation: Mostafa, S.A.; Johnson, T., III; Jennings, S. Composite Estimators for the Population Mean Under Ranked Set Sampling. *Mathematics* **2025**, *13*, 3071. <https://doi.org/10.3390/math13193071>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Efficient sampling methods are essential in statistical analysis, especially when acquiring data is costly or time-consuming. In such contexts, alternative sampling designs that leverage auxiliary information or partial measurements can offer significant advantages. One such approach is Ranked Set Sampling (RSS), first introduced by McIntyre [1] in the early 1950s to estimate pasture yields in Australia. Since its inception, RSS has gained considerable attention due to its ability to increase estimation efficiency with minimal additional cost or effort, particularly in environments where precise measurement is costly, but ranking units is relatively easy [2]. RSS has been successfully applied in a variety of fields—including agriculture, environmental science, and economics—where it has outperformed Simple Random Sampling (SRS) in estimation efficiency.

To implement RSS, a total of p^2 units are randomly selected from the population and divided into p sets, each containing p units. Within each set, the units are ranked based on visual inspection, expert judgment, concomitant variables, or other means that do not involve precise quantification. The process continues as follows: the smallest unit from the first set is measured, the second smallest unit from the second set is measured, and so forth,

until the largest unit from the p -th set is measured. This procedure constitutes a single cycle of RSS, which can be repeated for m independent cycles to obtain a balanced RSS of size $n = pm$: $\{Y_{[k]l} : k = 1, 2, \dots, p; l = 1, 2, \dots, m\}$, where $Y_{[k]l}$ denotes the k -th judgment order statistic of the units in the k -th set of the l -th cycle. The use of square brackets instead of parentheses signifies that ranking may be subject to error or imperfection. If ranking is perfect, the notation $Y_{(k)l}$ may be used instead.

To illustrate the technique, in a balanced RSS with $p = 4$ and $m = 2$, with ranking based on an auxiliary (or concomitant) variable X , the selection proceeds as follows:

Cycle 1:

$$\begin{aligned} \mathbf{X}_{(1)1} \leq X_{(2)1} \leq X_{(3)1} \leq X_{(4)1} &\implies (X_{(1)1}, Y_{[1]1}) \\ X_{(1)1} \leq \mathbf{X}_{(2)1} \leq X_{(3)1} \leq X_{(4)1} &\implies (X_{(2)1}, Y_{[2]1}) \\ X_{(1)1} \leq X_{(2)1} \leq \mathbf{X}_{(3)1} \leq X_{(4)1} &\implies (X_{(3)1}, Y_{[3]1}) \\ X_{(1)1} \leq X_{(2)1} \leq X_{(3)1} \leq \mathbf{X}_{(4)1} &\implies (X_{(4)1}, Y_{[4]1}) \end{aligned}$$

Cycle 2:

$$\begin{aligned} \mathbf{X}_{(1)2} \leq X_{(2)2} \leq X_{(3)2} \leq X_{(4)2} &\implies (X_{(1)2}, Y_{[1]2}) \\ X_{(1)2} \leq \mathbf{X}_{(2)2} \leq X_{(3)2} \leq X_{(4)2} &\implies (X_{(2)2}, Y_{[2]2}) \\ X_{(1)2} \leq X_{(2)2} \leq \mathbf{X}_{(3)2} \leq X_{(4)2} &\implies (X_{(3)2}, Y_{[3]2}) \\ X_{(1)2} \leq X_{(2)2} \leq X_{(3)2} \leq \mathbf{X}_{(4)2} &\implies (X_{(4)2}, Y_{[4]2}) \end{aligned}$$

producing a balanced RSS:

$$s = \left\{ (X_{(1)1}, Y_{[1]1}), (X_{(2)1}, Y_{[2]1}), (X_{(3)1}, Y_{[3]1}), (X_{(4)1}, Y_{[4]1}), \right. \\ \left. (X_{(1)2}, Y_{[1]2}), (X_{(2)2}, Y_{[2]2}), (X_{(3)2}, Y_{[3]2}), (X_{(4)2}, Y_{[4]2}) \right\}.$$

There is a vast body of literature on statistical inference using RSS, including parametric estimation of population parameters such as the mean (e.g., [3]), variance (e.g., [4]), and quantiles (e.g., [5]), and nonparametric estimation and hypothesis testing (e.g., [6–10]). For a comprehensive review, see [11].

In this paper, we consider the problem of estimating a finite population mean using RSS. Let $U = \{1, 2, \dots, N\}$ denote a finite population of size N , where each unit $i \in U$ is associated with a value y_i of a study variable Y , and a value x_i of an auxiliary variable X . The parameter of interest is the finite population mean:

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i.$$

The goal is to estimate $\bar{Y}$ based on an RSS of size n . Numerous estimators have been proposed for this purpose, including the simple mean of the RSS ([12]), the ratio estimator ([13,14]), and the regression estimator ([15]).

In this study, we propose a shrinkage-type (composite) estimator for estimating the finite population mean in the context of RSS. This estimator builds upon the idea of combining two or more estimators using shrinkage weights, offering a compromise that can outperform the individual components under certain conditions. Specifically, we extend the work of Lui [16], who proposed a composite estimator that blends the simple mean and the ratio estimator under SRS. Lui [16] showed that the composite approach can offer improved performance when certain conditions are met. Their simulation study suggested that the

SRS composite estimator could outperform the simple and ratio estimators individually, especially when the auxiliary variable is moderately to strongly correlated with the study variable. We extend and study this idea under RSS. We provide theoretical results that characterize the performance of the proposed estimator and explore its properties through simulation and real data analysis.

The remainder of this paper is organized as follows. Section 2 reviews the idea of composite estimators under SRS. Section 3 introduces the proposed composite estimator under RSS and derives its theoretical properties. Section 4 presents a simulation study investigating the finite sample performance of the proposed estimator relative to several other estimators. Section 5 illustrates the performance of the proposed estimator on a real dataset. Section 6 concludes the paper with a discussion of key findings and future research directions.

2. Composite Estimators Under SRS

In this section, we review the idea of composite estimators under SRS. The simplest estimator for the finite population mean under SRS is the simple mean estimator defined by [12]:

$$\bar{y}_{\text{SRS}} = \frac{1}{n} \sum_{i=1}^n y_i.$$

The simple mean estimator is known to be unbiased and its variance, or, equivalently, its mean squared error (MSE), is given by [12], ch. 2:

$$\text{MSE}(\bar{y}_{\text{SRS}}) = \text{Var}(\bar{y}_{\text{SRS}}) = \left(1 - \frac{n}{N}\right) \frac{S_y^2}{n},$$

where $S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})^2$ is the variance of the study variable Y .

An alternative estimator that is commonly used to take advantage of available auxiliary information is the ratio estimator, which is defined as follows under SRS [12]:

$$\bar{y}_{\text{rSRS}} = \frac{\bar{y}_{\text{SRS}}}{\bar{x}_{\text{SRS}}} \bar{X}.$$

The approximate variance of the ratio estimator under SRS is given by [12], ch. 6:

$$\begin{aligned} \text{Var}(\bar{y}_{\text{rSRS}}) &\simeq \left(1 - \frac{n}{N}\right) \frac{1}{n} [S_y^2 + R^2 S_x^2 - 2R S_{xy}] \\ &= \left(1 - \frac{n}{N}\right) \frac{1}{n} \bar{Y}^2 [C_y^2 + C_x^2 - 2\rho C_x C_y]. \end{aligned} \quad (1)$$

where $R = \frac{\bar{Y}}{\bar{X}}$, $S_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y})$, $C_z^2 = \frac{S_z^2}{\bar{Z}^2}$ is the squared population coefficient of variation of a generic variable Z , and $\rho = \frac{S_{xy}}{S_x S_y}$ is the population Pearson's correlation coefficient between X and Y . Under SRS, Lui [16] suggested combining the simple mean estimator and the ratio estimator into a composite estimator defined by

$$\bar{y}_{\text{cSRS}} = w \bar{y}_{\text{SRS}} + (1 - w) \bar{y}_{\text{rSRS}},$$

where $0 \leq w \leq 1$ is an unknown weight parameter. Observing the fact that, since $\bar{y}_{\text{SRS}}$ is unbiased,

$$|\text{Bias}(\bar{y}_{\text{cSRS}})| = |\text{E}(\bar{y}_{\text{cSRS}} - \bar{Y})| = |(1 - w)\text{E}(\bar{y}_{\text{rSRS}} - \bar{Y})| \leq |\text{E}(\bar{y}_{\text{rSRS}} - \bar{Y})| = |\text{Bias}(\bar{y}_{\text{rSRS}})|$$

and the fact that the ratio estimator under SRS is asymptotically unbiased, ref. [16] derived the optimal weight by minimizing the variance of the composite estimator:

$$\begin{aligned} w_{\text{SRS}}^* &= \frac{\text{Var}(\bar{y}_{\text{rSRS}}) - \text{Cov}(\bar{y}_{\text{SRS}}, \bar{y}_{\text{rSRS}})}{\text{Var}(\bar{y}_{\text{rSRS}}) + \text{Var}(\bar{y}_{\text{SRS}}) - 2\text{Cov}(\bar{y}_{\text{SRS}}, \bar{y}_{\text{rSRS}})} \\ &= 1 - \rho C_y / C_x, \end{aligned} \quad (2)$$

where $\text{Cov}(U, V) = E[(U - E(U))(V - E(V))]$. According to Lui [16], there are specific circumstances in which the composite estimator surpasses the ratio estimator under SRS. These conditions are related to the correlation between X and Y , as well as their coefficients of variation. In addition, a simulation study was carried out by Lui [16] to showcase the efficiency of the composite estimator compared to the simple mean estimator and the ratio estimator under SRS. In this study, we are building upon the research conducted by [16] and expanding it to the context of RSS. In the following section, we will review the existing estimators for the population mean under RSS and introduce our proposed composite estimator.

3. Composite Estimators Under RSS

3.1. Simple Mean Estimator Under RSS

The unbiased mean estimator under RSS is given by

$$\bar{y}_{\text{RSS}} = \frac{1}{pm} \sum_{k=1}^p \sum_{l=1}^m y_{[k]l}. \quad (3)$$

Under RSS, the MSE of $\bar{y}_{\text{RSS}}$ is given by Samawi and Muttalak [13] as follows:

$$\text{MSE}(\bar{y}_{\text{RSS}}) = \frac{1}{pm} \left(S_y^2 - \frac{1}{p} \sum_{k=1}^p \tau_{y[k]}^2 \right), \quad (4)$$

where $\tau_{y[k]} = (\mu_{y[k]} - \bar{Y})$, with $\mu_{y[k]}$ being the mean of the k -th order statistic of the study variable.

3.2. Ratio Estimator Under RSS

The conventional ratio estimator for the population mean $\bar{Y}$ of the study variable, y , is established as a quotient of two auxiliary variables, namely, the total of the study variable in the sample and the corresponding total of a related variable in the population. This estimator is widely used in various research fields, including business and academia, as it provides a reliable and efficient approach for estimating population means.

The ratio estimator under RSS is given by

$$\bar{y}_{\text{rRSS}} = \frac{\bar{y}_{\text{RSS}}}{\bar{x}_{\text{RSS}}} \bar{X}, \quad (5)$$

where

$$\bar{x}_{\text{RSS}} = \frac{1}{pm} \sum_{k=1}^p \sum_{l=1}^m x_{(k)l}.$$

First-order approximations—assuming a large $n = pm$ —of the bias and MSE of the ratio estimator in Equation (5) are given by Kadilar et al. [14]:

$$\text{Bias}(\bar{y}_{\text{rRSS}}) \simeq \bar{Y} \left[\frac{1}{pm} (C_x^2 - C_{yx}) - (W_x^2 - W_{yx}) \right], \quad (6)$$

$$\text{MSE}(\bar{y}_{\text{rRSS}}) \simeq \frac{1}{pm} \left(S_y^2 - 2RS_{xy} + R^2 S_x^2 \right) - \frac{1}{p^2 m} \left(\sum_{k=1}^p \tau_{y[k]}^2 - 2R \sum_{k=1}^p \tau_{yx(k)} + R^2 \sum_{k=1}^p \tau_{x(k)}^2 \right), \quad (7)$$

where

$$C_{yx} = \rho C_y C_x = \frac{S_{xy}}{\bar{Y}\bar{X}}, \quad W_x^2 = \frac{1}{p^2 m \bar{X}^2} \sum_{k=1}^p \tau_{x(k)}^2, \quad \text{and} \quad W_{yx} = \frac{1}{p^2 m \bar{X}\bar{Y}} \sum_{k=1}^p \tau_{yx(k)},$$

with

$$\tau_{x(k)} = (\mu_{x(k)} - \bar{X}), \quad \tau_{yx(k)} = (\mu_{x(k)} - \bar{X})(\mu_{y[k]} - \bar{Y}), \quad \mu_{x(k)} = E(X_{(k)}), \quad \text{and} \quad \mu_{y[k]} = E(Y_{[k]}).$$

3.3. Proposed Composite Estimator

In this study, we propose the use of a composite estimator, which is defined as a mixture of the simple mean estimator and the ratio estimator, for estimating the finite population mean $\bar{Y}$ under RSS. The estimator is defined as follows:

$$\ddot{y}_{\text{cRSS}} = w\bar{y}_{\text{RSS}} + (1-w)\bar{y}_{\text{rRSS}}, \quad (8)$$

where $0 \leq w \leq 1$ represents an unknown weight parameter. To derive the statistical properties of this estimator, we assume w to be fixed, and hence the use of the dot notation in $\ddot{y}_{\text{cRSS}}$. Later in this section, we will derive the optimal weight and propose a corresponding estimator.

The following theorem and corollary summarize the statistical properties of the composite estimator in (8).

Theorem 1. Under RSS, first-order approximations of the design-bias and variance of the composite estimator $\ddot{y}_{\text{cRSS}}$ are given by

$$\begin{aligned} \text{Bias}(\ddot{y}_{\text{cRSS}}) &= (1-w)\text{Bias}(\bar{y}_{\text{rRSS}}) \\ &\simeq (1-w)\bar{Y} \left[\frac{1}{pm} (C_x^2 - C_{yx}) - (W_x^2 - W_{yx}) \right] \end{aligned} \quad (9)$$

and

$$\text{Var}(\ddot{y}_{\text{cRSS}}) \simeq w^2 A + (1-w)^2 B + 2w(1-w)C, \quad (10)$$

where

$$A := \text{Var}(\bar{y}_{\text{RSS}}) = \frac{1}{pm} \left(S_y^2 - \frac{1}{p} \sum_{k=1}^p \tau_{y[k]}^2 \right), \quad (11)$$

$$\begin{aligned} B &:= \text{Var}(\bar{y}_{\text{rRSS}}) \\ &\simeq \frac{1}{pm} \left(S_y^2 - 2RS_{yx} + R^2 S_x^2 \right) - \frac{1}{p^2 m} \left(\sum_{k=1}^p \tau_{y[k]}^2 - 2R \sum_{k=1}^p \tau_{yx(k)} + R^2 \sum_{k=1}^p \tau_{x(k)}^2 \right) \\ &\quad - \bar{Y}^2 \left[\frac{1}{pm} (C_x^2 - C_{yx}) - (W_x^2 - W_{yx}) \right]^2 \end{aligned} \quad (12)$$

and

$$C := \text{Cov}(\bar{y}_{RSS}, \bar{y}_{rRSS}) = \frac{1}{pm} \left(S_y^2 - \frac{1}{p} \sum_{k=1}^p \tau_{y[k]}^2 \right) - R \frac{1}{pm} \left(S_{yx}^2 - \frac{1}{p} \sum_{k=1}^p \tau_{yx(k)}^2 \right). \quad (13)$$

Proof. We start with the bias statement. First, notice that

$$\begin{aligned} E(\ddot{y}_{cRSS}) &= wE(\bar{y}_{RSS}) + (1-w)E(\bar{y}_{rRSS}) \\ &= w\bar{Y} + (1-w)E(\bar{y}_{rRSS}). \end{aligned}$$

Therefore,

$$\begin{aligned} \text{Bias}(\ddot{y}_{cRSS}) &= E(\ddot{y}_{cRSS}) - \bar{Y} \\ &= w\bar{Y} + (1-w)E(\bar{y}_{rRSS}) - (w\bar{Y} + (1-w)\bar{Y}) \\ &= (1-w)[E(\bar{y}_{rRSS}) - \bar{Y}] \\ &= (1-w)\text{Bias}(\bar{y}_{rRSS}) \\ &\simeq (1-w)\bar{Y} \left[\frac{1}{pm} (C_x^2 - C_{yx}) - (W_x^2 - W_{yx}) \right], \end{aligned}$$

where the last equality follows from Equation (6).

Next, we derive the variance statement. Observe that

$$\begin{aligned} \text{Var}(\ddot{y}_{cRSS}) &= w^2 \text{Var}(\bar{y}_{RSS}) + (1-w)^2 \text{Var}(\bar{y}_{rRSS}) + 2w(1-w) \text{Cov}(\bar{y}_{RSS}, \bar{y}_{rRSS}) \\ &= w^2 A + (1-w)^2 B + 2w(1-w)C. \end{aligned}$$

Now, A in Equation (11) follows directly from Equation (4), whereas B in Equation (12) follows from combining Equations (6) and (7) into $B = \text{Var}(\bar{y}_{rRSS}) = \text{MSE}(\bar{y}_{rRSS}) - \text{Bias}^2(\bar{y}_{rRSS})$. Finally,

$$\begin{aligned} C &= \text{Cov}(\bar{y}_{RSS}, \bar{y}_{rRSS}) \\ &= E[(\bar{y}_{RSS} - \bar{Y})(\bar{y}_{rRSS} - \bar{Y})] \\ &= E\left[(\bar{y}_{RSS} - \bar{Y}) \left(\frac{\bar{y}_{RSS}}{\bar{x}_{RSS}} \bar{X} - \bar{Y} \right)\right] \\ &= E\left[(\bar{y}_{RSS} - \bar{Y}) \frac{\bar{X}}{\bar{x}_{RSS}} \left(\bar{y}_{RSS} - \frac{\bar{Y}}{\bar{X}} \bar{x}_{RSS} \right)\right] \\ &\simeq E[(\bar{y}_{RSS} - \bar{Y}) \cdot 1 \cdot (\bar{y}_{RSS} - R\bar{x}_{RSS})] \\ &= E[(\bar{y}_{RSS} - \bar{Y})(\{\bar{y}_{RSS} - \bar{Y}\} - R\{\bar{x}_{RSS} - \bar{X}\})] \\ &= E[(\bar{y}_{RSS} - \bar{Y})^2] - E[R(\bar{x}_{RSS} - \bar{X})(\bar{y}_{RSS} - \bar{Y})] \\ &= \text{Var}(\bar{y}_{RSS}) - R \text{Cov}(\bar{x}_{RSS}, \bar{y}_{RSS}) \\ &= \frac{1}{pm} \left(S_y^2 - \frac{1}{p} \sum_{k=1}^p \tau_{y[k]}^2 \right) - R \frac{1}{pm} \left(S_{yx}^2 - \frac{1}{p} \sum_{k=1}^p \tau_{yx(k)}^2 \right). \end{aligned}$$

The proof is complete. $\square$

Corollary 1. Under RSS, the first-order approximation of the MSE of the composite estimator $\ddot{y}_{cRSS}$ is given by

$$\begin{aligned}
\text{MSE}(\ddot{y}_{cRSS}) &\simeq w^2 \frac{1}{pm} \left(S_y^2 - \frac{1}{p} \sum_{k=1}^p \tau_{y[k]}^2 \right) \\
&\quad + (1-w)^2 \left[\frac{1}{pm} (S_y^2 - 2RS_{yx} + R^2 S_x^2) \right. \\
&\quad \quad \left. - \frac{1}{p^2 m} \left(\sum_{k=1}^p \tau_{y[k]}^2 - 2R \sum_{k=1}^p \tau_{yx(k)} + R^2 \sum_{k=1}^p \tau_{x(k)}^2 \right) \right] \\
&\quad + 2w(1-w) \left[\frac{1}{pm} \left(S_y^2 - \frac{1}{p} \sum_{k=1}^p \tau_{y[k]}^2 \right) - R \frac{1}{pm} \left(S_{yx}^2 - \frac{1}{p} \sum_{k=1}^p \tau_{yx(k)}^2 \right) \right] \\
&= w^2 A + (1-w)^2 D + 2w(1-w)C,
\end{aligned} \tag{14}$$

where

$$D := \left[\frac{1}{pm} (S_y^2 - 2RS_{yx} + R^2 S_x^2) - \frac{1}{p^2 m} \left(\sum_{k=1}^p \tau_{y[k]}^2 - 2R \sum_{k=1}^p \tau_{yx(k)} + R^2 \sum_{k=1}^p \tau_{x(k)}^2 \right) \right]. \tag{15}$$

3.4. Composite Estimator with Optimal Weight

Let w^* represent the optimal weight that minimizes the MSE of the composite estimator in Equation (14). The proposed composite estimator with optimal weight under RSS is given by

$$\bar{y}_{cRSS}^* = w^* \bar{y}_{RSS} + (1-w^*) \bar{y}_{rRSS}. \tag{16}$$

Theorem 2. The optimal weight w^* that minimizes the (approximate) design-based MSE of the composite estimator under RSS is given by

$$w^* = \frac{D - C}{A + D - 2C}, \tag{17}$$

and the corresponding minimum MSE is given by

$$\text{MSE}(\bar{y}_{cRSS}^*) = \frac{A(D - C)^2 + D(A - C)^2 + 2C(D - C)(A - C)}{(A + D - 2C)^2}, \tag{18}$$

where A , C , and D are given in Equation (11), Equation (13) and Equation (15), respectively.

Proof. To find the weight w that minimizes $\text{MSE} =: f(w)$, we take the derivative of the MSE expression in Equation (14) with respect to w :

$$\begin{aligned}
\frac{d}{dw} \text{MSE}(\ddot{y}_{cRSS}) &= \frac{d}{dw} (w^2 A + (1-w)^2 D + 2w(1-w)C) \\
&= 2wA - 2D + 2Dw + 2C - 4wC \\
&= 2w(A + D - 2C) + 2C - 2D.
\end{aligned}$$

Setting this derivative to zero gives the weight corresponding to the minimum MSE:

$$w^* = \frac{D - C}{A + D - 2C}.$$

Substituting by w^* into Equation (14) gives the minimum MSE in Equation (18) after some basic algebra. $\square$

3.5. Composite Estimator with Estimated Weight

Since the optimal weight w^* is unknown, we define the composite estimator with estimated weight as follows:

$$\bar{y}_{cRSS} = \hat{w}\bar{y}_{RSS} + (1 - \hat{w})\bar{y}_{rRSS}, \quad (19)$$

where the estimated optimal weight is given by

$$\hat{w} = \frac{\hat{D} - \hat{C}}{\hat{A} + \hat{D} - 2\hat{C}}, \quad (20)$$

with

$$\begin{aligned} \hat{A} &= \frac{1}{pm} \left(\hat{S}_y^2 - \frac{1}{p} \sum_{k=1}^p \hat{\tau}_{y[k]}^2 \right), \\ \hat{D} &= \frac{1}{pm} \left(\hat{S}_y^2 - 2\hat{R}\hat{S}_{yx} + \hat{R}^2\hat{S}_x^2 \right) - \frac{1}{p^2m} \left(\sum_{k=1}^p \hat{\tau}_{y[k]}^2 - 2\hat{R} \sum_{k=1}^p \hat{\tau}_{yx(k)} + \hat{R}^2 \sum_{k=1}^p \hat{\tau}_{x(k)}^2 \right), \\ \hat{C} &= \hat{A} - \hat{R} \frac{1}{pm} \left(\hat{S}_{yx} - \frac{1}{p} \sum_{k=1}^p \hat{\tau}_{yx(k)} \right), \end{aligned}$$

with

$$\begin{aligned} \hat{S}_z^2 &= \frac{1}{n-1} \cdot \sum_{i=1}^n (z_i - \bar{z}_{RSS})^2, \\ \hat{S}_{yx} &= \frac{1}{n-1} \cdot \sum_{i=1}^n (y_i - \bar{y}_{RSS})(x_i - \bar{x}_{RSS}), \\ \hat{R} &= \frac{\bar{y}_{RSS}}{\bar{x}_{RSS}}, \\ \hat{\tau}_{x(k)} &= (\hat{\mu}_{x(k)} - \bar{x}_{RSS}), \\ \hat{\tau}_{y[k]} &= (\hat{\mu}_{y[k]} - \bar{y}_{RSS}), \end{aligned}$$

and

$$\hat{\tau}_{yx(k)} = (\hat{\mu}_{x(k)} - \bar{x}_{RSS})(\hat{\mu}_{y[k]} - \bar{y}_{RSS}),$$

for $k = 1, 2, \dots, p$.

4. Simulation Study

In this section, we present the results of our simulation study. We aimed to evaluate the performance of various mean estimators under three different sampling techniques: SRS, RSS with perfect ranking, and RSS with imperfect ranking. Both versions of RSS were implemented using the *RSSampling* package ver. 1.0 in R [17].

4.1. Simulation Settings

4.1.1. Normal Case

To evaluate the effectiveness and robustness of the proposed estimators under normality, we considered a finite population of size $N = 10,000$. The population means of the auxiliary variable X and the study variable Y were both set to 10. The coefficients of variation $CV_x (= C_x)$ and $CV_y (= C_y)$ were varied to reflect moderate-to-high variability. The correlation coefficients between X and Y were set to $\rho = 0.60$ and 0.80 , representing moderate-to-strong positive associations where the ratio and regression estimators are known to perform well. Sample sizes of $n = 15, 30, 45$, and 60 were used to assess estimators' performance under varying sampling rates. Under each of the resulting fi-

nite population and sample size combinations, we drew 10,000 samples using (1) SRS, (2) imperfect-ranking RSS where the ranking is based on the auxiliary variable X , and (3) perfect-ranking RSS where the ranking is based on the study variable Y .

4.1.2. Lognormal Case

To evaluate the estimators' performance under non-normal, skewed data, we generated a finite population of size $N = 10,000$ from the lognormal distribution. The sample size settings were kept the same as those used for the normal case. The population means of the log-transformed variables U and V were set to 0, and their standard deviations were set to 0.5 and 1.5, respectively. The correlation coefficient between U and V was set to $\rho = 0.60$ and 0.80 , capturing moderate-to-strong linear associations on the logarithmic scale. Under each simulation scenario, we again drew 10,000 samples using (1) SRS, (2) imperfect-ranking RSS, and (3) perfect-ranking RSS.

4.2. Estimators

The following five estimators were compared in our simulations:

1. The composite estimator with optimal weight in Equation (16) (denoted by $\bar{y}_c^*$);
2. The composite estimator with estimated weight in Equation (19) (denoted by $\bar{y}_c$);
3. The simple mean estimator in Equation (3) (denoted by $\bar{y}$);
4. The ratio estimator in Equation (5) (denoted by $\bar{y}_r$);
5. The regression estimator (denoted by $\bar{y}_{\text{reg}}$), which is defined under RSS as follows:

$$\bar{y}_{\text{regRSS}} = \bar{y}_{\text{RSS}} + \hat{\beta}_{\text{RSS}}(\bar{X} - \bar{x}_{\text{RSS}}), \quad (21)$$

where

$$\hat{\beta}_{\text{RSS}} = \frac{\sum_{k=1}^p \sum_{l=1}^m (x_{(k)l} - \bar{x}_{\text{RSS}})(y_{[k]l} - \bar{y}_{\text{RSS}})}{\sum_{k=1}^p \sum_{l=1}^m (x_{(k)l} - \bar{x}_{\text{RSS}})^2},$$

and its bias and variance are given in Theorem 1 of [15]. Similarly, under SRS, the regression estimator is defined as follows:

$$\bar{y}_{\text{regSRS}} = \bar{y}_{\text{SRS}} + \hat{\beta}_{\text{SRS}}(\bar{X} - \bar{x}_{\text{SRS}}), \quad (22)$$

where

$$\hat{\beta}_{\text{SRS}} = \frac{\sum_{k=1}^n (x_k - \bar{x}_{\text{SRS}})(y_k - \bar{y}_{\text{SRS}})}{\sum_{k=1}^n (x_k - \bar{x}_{\text{SRS}})^2},$$

and its bias and variance are given by ([12], ch. 7).

4.3. Approximation of the Order Statistics Moments

Note that the calculation of the optimal (or estimated) weight for the composite RSS estimator requires the values (or estimates) of $\mu_{x(k)}$, the first-order moment of the k -th order statistic. For the normal case, an approximation for these moments is given by [18]:

$$\mu_{x(k)} = E[X_{(k)}] = \mu_x + \Phi^{-1}\left(\frac{k - \alpha}{p - 2\alpha + 1}\right)\sigma_x,$$

where $E[X_{(k)}]$ is the expected value of the k -th order statistic from a normal distribution in a sample of size p ; μ_x and σ_x are the mean and the standard deviation of X ; $\Phi^{-1}(\cdot)$ is the inverse of the cumulative distribution function of the standard normal distribution; and α is a constant for correction, typically 0.375. This adjustment improves the accuracy of the approximation, especially for extreme ranks in small samples. A similar approximation was used for $\mu_{y[k]}$. For the estimated weight $\hat{w}$, we use $\hat{\mu}_{x(k)}$, which has the same form

as $\mu_{x(k)}$, but replaces μ_x and σ_x with the sample mean and sample standard deviation, respectively. This is also the case for $\hat{\mu}_{y[k]}$.

The approximation for the lognormal case according to [19] is given by

$$\mu_{x(k)} = \frac{p!}{(k-1)!(p-k)!} \times \int_0^\infty \left(\Phi \left(\frac{\log(z) - \mu_x}{\sigma_x} \right) \right)^{k-1} \left(1 - \Phi \left(\frac{\log(z) - \mu_x}{\sigma_x} \right) \right)^{p-k} \phi \left(\frac{\log(z) - \mu_x}{\sigma_x} \right) dz, \quad (23)$$

where μ_x and σ_x are the mean and the standard deviation of X , and the integral is evaluated using a standard numerical integration method implemented in the 'integrate()' function in R. A similar approximation was used for $\mu_{y[k]}$.

4.4. Performance Metrics

Under each scenario, we calculated the Monte Carlo bias, variance, and MSE for each of the five estimators as follows:

$$\begin{aligned} \text{Bias}(\bar{y}_{\text{est}}) &= \frac{1}{M} \sum_{j=1}^M (\bar{y}_{\text{est}}^j - \bar{Y}), \\ \text{Var}(\bar{y}_{\text{est}}) &= \frac{1}{M} \sum_{j=1}^M \left(\bar{y}_{\text{est}}^j - \frac{1}{M} \sum_{k=1}^M \bar{y}_{\text{est}}^k \right)^2, \\ \text{MSE}(\bar{y}_{\text{est}}) &= \frac{1}{M} \sum_{j=1}^M (\bar{y}_{\text{est}}^j - \bar{Y})^2, \end{aligned}$$

where the number of simulations $M = 10,000$.

4.5. Simulation Results

4.5.1. The Normal Case

In this section, we summarize the simulation results under the normal population scenario. Figures 1–3 display the Monte Carlo distribution of five estimators under SRS, imperfect-ranking RSS, and perfect-ranking RSS, respectively. The MSE results for the five estimators are presented in Figures 4–6. Additional simulation results, specifically bias and variance plots, for the normal case are deferred to Appendix A.1.

First, examining the sampling distributions of the estimators, we observe that in all three sampling designs, the ratio estimator ($\bar{y}_r$) consistently shows the largest interquartile range, longer whiskers, and more outliers, indicating lower efficiency and robustness, and this is especially evident under SRS; see Figure 1. On the other hand, the composite ($\bar{y}_c$ and $\bar{y}_c^*$) and regression ($\bar{y}_{\text{reg}}$) estimators show the shortest interquartile ranges and the least outliers across most scenarios, indicating relatively higher efficacy than other estimators. Although the simple mean estimator ($\bar{y}$) has relatively wide interquartile ranges under SRS and imperfect-ranking RSS, it shows much-improved behavior under perfect-ranking RSS, as it benefits from improved ranking accuracy. As the sample size increases, the precision of all estimators improves, as shown by the tighter interquartile ranges. In general, all estimators demonstrate a good central tendency, as the centers of the boxplots are mostly aligned with the true mean line.

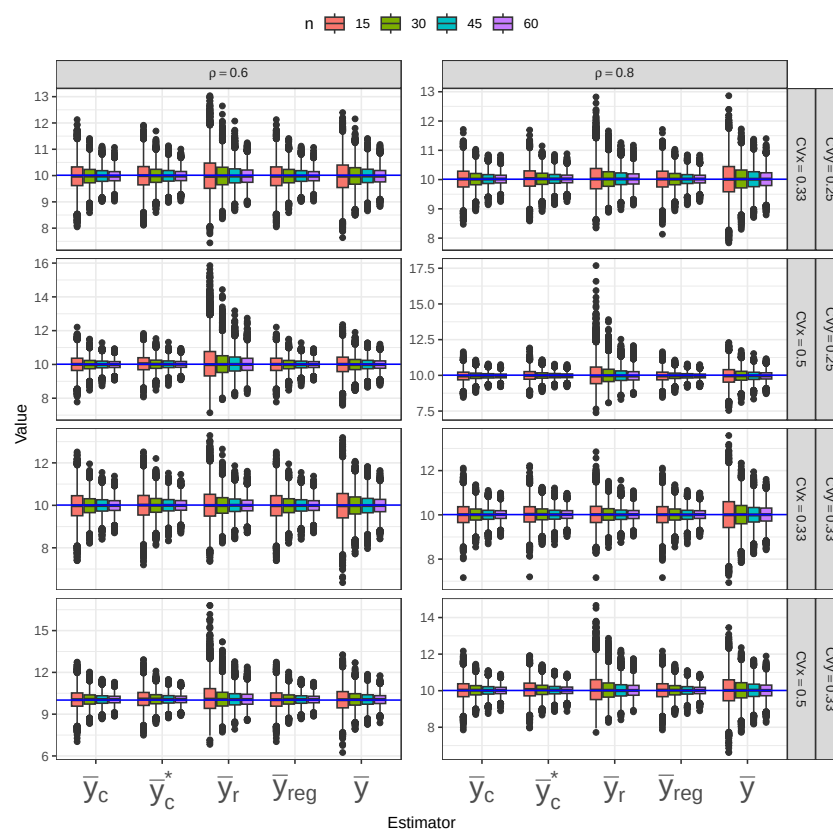


Figure 1. Distribution of five mean estimators under SRS from normal population.

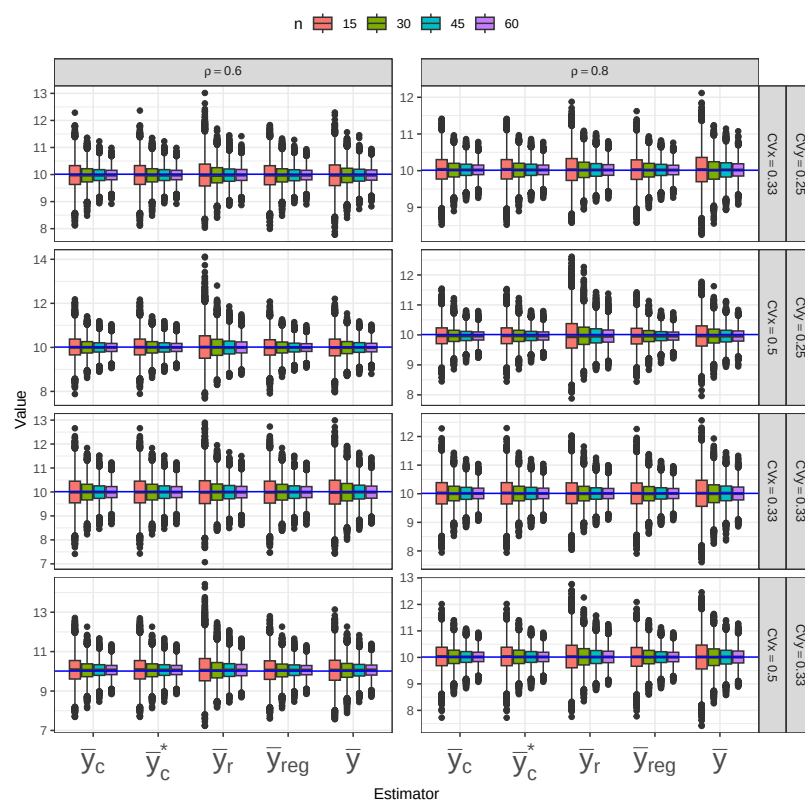


Figure 2. Distribution of five mean estimators under imperfect-ranking RSS (fixing $p = 5$) from normal population.

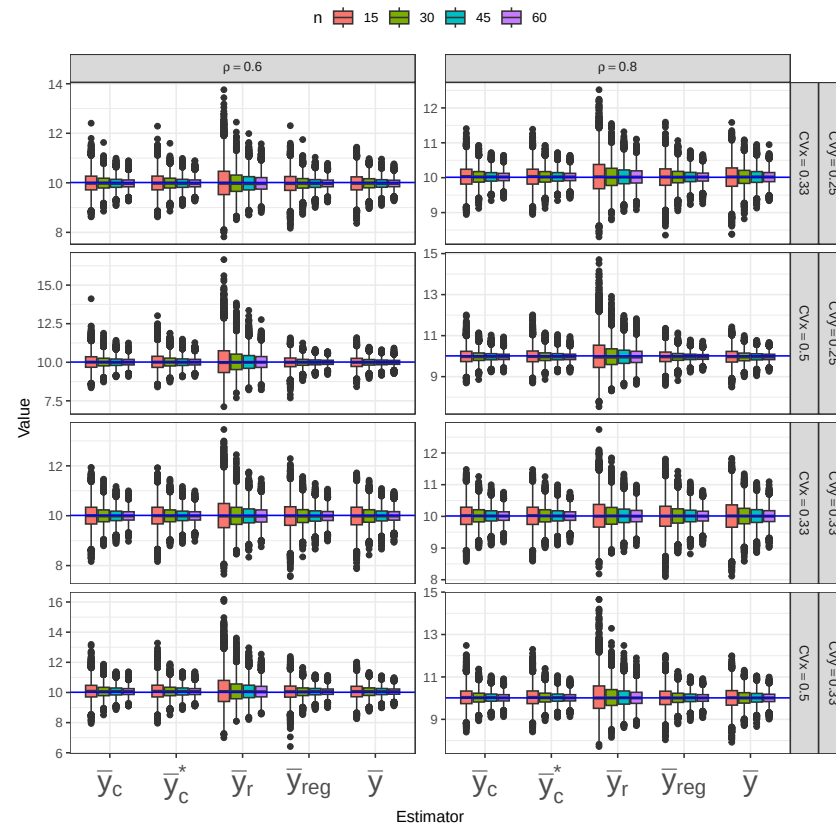


Figure 3. Distribution of five mean estimators under perfect-ranking RSS (fixing $p = 5$) from normal population.

Figures 4–6 illustrate the MSE performance of the five estimators under different sampling designs and parameter settings. Under all three sampling designs, the MSE decreases with increasing sample size for all estimators, supporting the consistency of these estimators. In the SRS setting (Figure 4), the simple mean ($\bar{y}$) and ratio ($\bar{y}_r$) estimators consistently exhibit the highest MSE values across all sample sizes, indicating lower efficiency. In contrast, the composite estimators ($\bar{y}_c$ and $\bar{y}_c^*$) and the regression estimator ($\bar{y}_{reg}$) demonstrate notably lower MSE values, with the two composite estimators performing nearly identically and slightly outperforming the regression estimator. Under the imperfect-ranking RSS design (Figure 5), the simple mean and ratio estimators remain the least efficient across all conditions. The composite and regression estimators show substantial improvements in efficiency, with the two composite estimators again nearly indistinguishable and generally exhibiting the lowest MSEs. Under perfect-ranking RSS (Figure 6), the simple mean estimator shows superior relative performance for the lower-correlation setting ($\rho = 0.6$), especially at low-to-moderate sample sizes. Under this design, the ratio estimator has the highest MSE across all parameter values. The composite estimators consistently outperform other estimators under the high-correlation setting ($\rho = 0.8$), closely followed by the regression estimator.

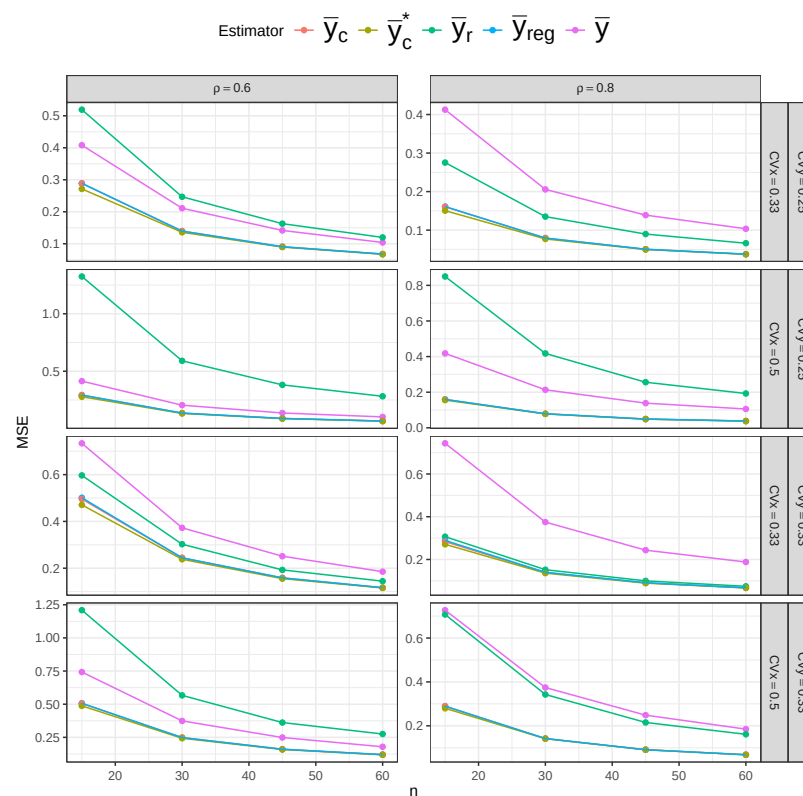


Figure 4. MSE of five mean estimators under SRS from normal population.

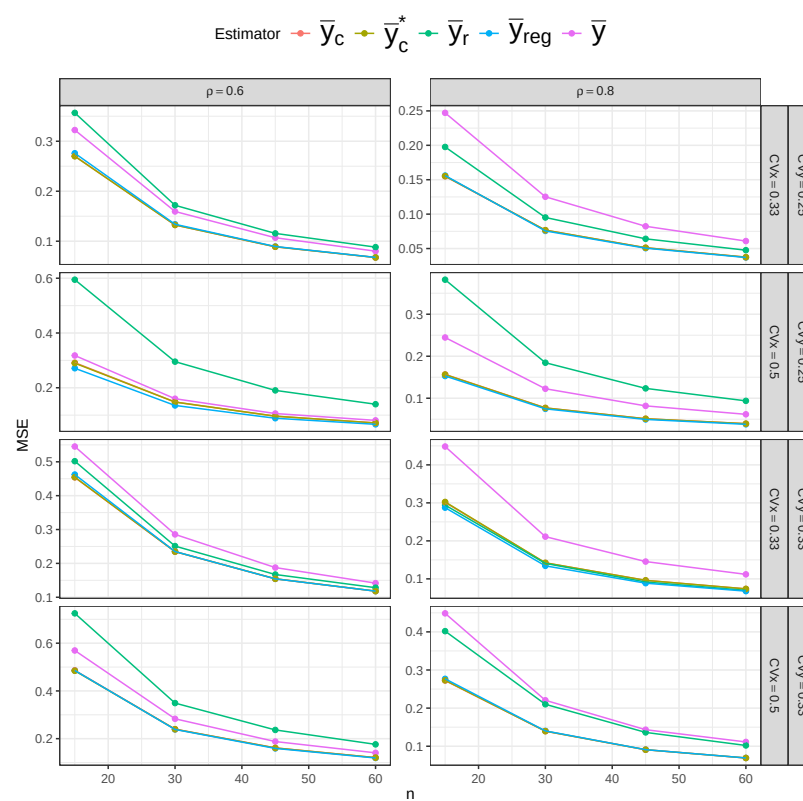


Figure 5. MSE of five mean estimators under imperfect-ranking RSS (fixing $p = 5$) from normal population.

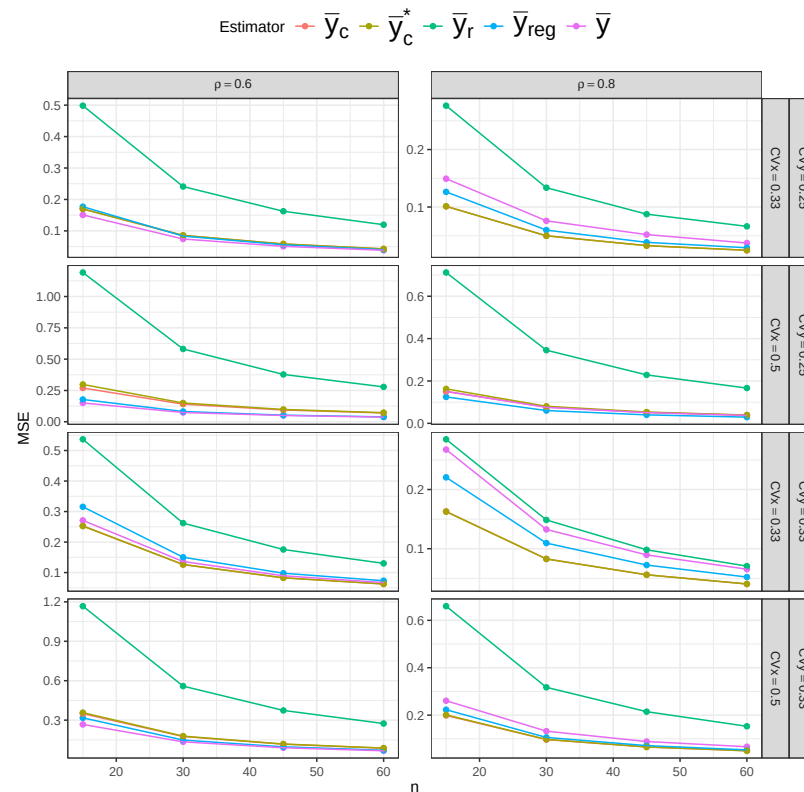


Figure 6. MSE of five mean estimators under perfect-ranking RSS (fixing $p = 5$) from normal population.

4.5.2. The Lognormal Case

Figures 7–9 present the sampling distributions of five mean estimators under lognormal populations across different sampling designs, correlation levels ($\rho = 0.6$ and $\rho = 0.8$), and sample sizes ($n = 15, 30, 45, 60$). In the SRS setting (Figure 7), the composite estimators ($\bar{y}_c$ and $\bar{y}_c^*$) exhibit strong performance, with tight interquartile ranges, relatively symmetric distributions, and limited outliers. The ratio estimator ($\bar{y}_r$) maintains a good central tendency, but with slightly broader variability. In contrast, the regression ($\bar{y}_{reg}$) and simple mean ($\bar{y}$) estimators show wider spreads and more frequent extreme values, especially for smaller sample sizes. Under the imperfect-ranking RSS design (Figure 8), all five estimators show similar behavior, with clear underestimation, as seen from the centers of all the boxplots falling below the true mean line. A similar pattern is observed under perfect-ranking RSS (Figure 9), with the regression estimator showing occasional negative outlying values and the simple mean estimator showing somewhat larger variability under the higher-correlation scenario.

Figures 10–12 illustrate the MSE behavior of the five target estimators across varying sample sizes and correlation levels for lognormal populations. In the SRS setting (Figure 10), the regression ($\bar{y}_{reg}$) estimator consistently achieves the lowest MSE across all sample sizes and both correlation levels, especially under strong correlation ($\rho = 0.8$). The composite ($\bar{y}_c$ and $\bar{y}_c^*$) and ratio ($\bar{y}_r$) estimators follow closely behind, showing comparable MSE performance at moderate correlation ($\rho = 0.6$), but lagging under the stronger-correlation ($\rho = 0.8$) scenario. The simple mean ($\bar{y}$) estimator consistently has the highest MSE values. Figure 11, which corresponds to the imperfect-ranking RSS design, shows a similar overall trend, but with subtle differences. The regression estimator remains dominant, especially under $\rho = 0.8$, showing the lowest MSE across all sample sizes. The ratio estimator follows in superiority, showing the second-lowest MSE values across scenarios.

The composite estimators still perform well, but their advantage is less pronounced under the stronger-correlation and smaller-sample-size scenarios. The simple mean estimator again show relatively higher MSE values, remaining the least efficient. Under perfect ranking (Figure 12), while all estimators have similar MSE curves, the ratio estimator, followed by the composite estimator ($\bar{y}_C^*$), achieves the lowest MSE values at $\rho = 0.6$. When $\rho = 0.8$, the pattern closely replicates that observed under imperfect-ranking RSS.

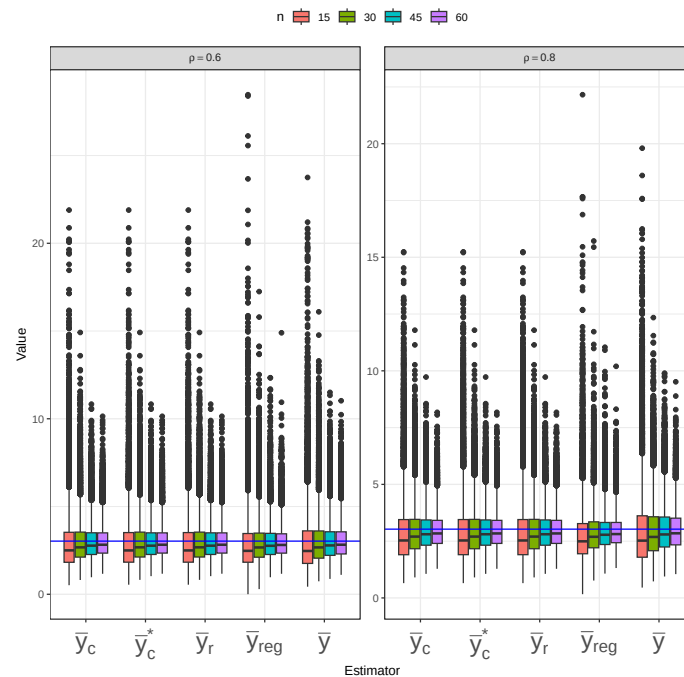


Figure 7. Distribution of five mean estimators under SRS from lognormal population. All estimates with values ≤ 0 or ≥ 30 are excluded.

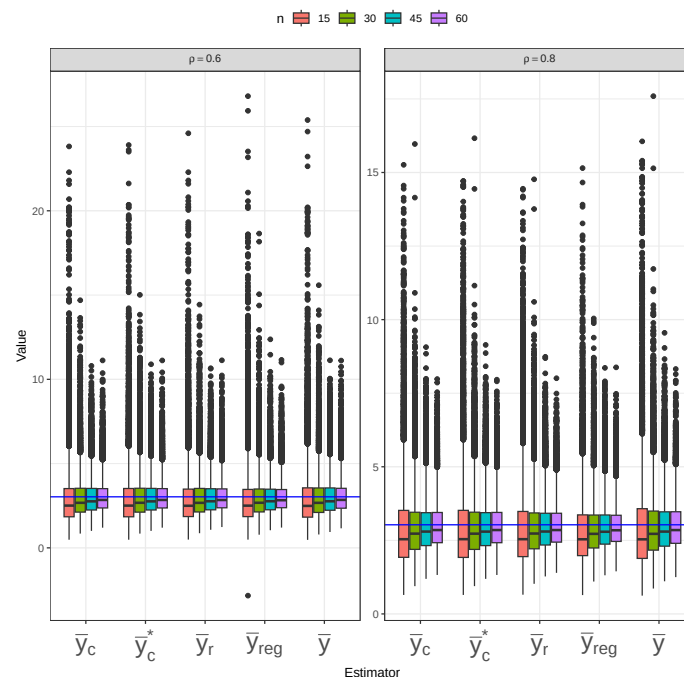


Figure 8. Distribution of five mean estimators under imperfect-ranking RSS (fixing $p = 5$) from lognormal population. All estimates with values ≥ 30 are excluded.

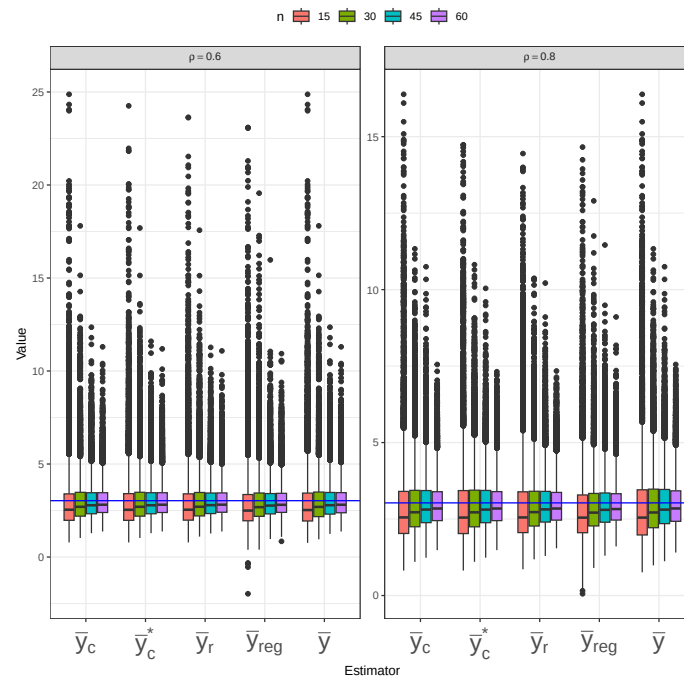


Figure 9. Distribution of five mean estimators under perfect-ranking RSS (fixing $p = 5$) from lognormal population. All estimates with values ≥ 30 are excluded.

Across all figures, the MSE decreases with increasing sample size for all estimators, reflecting improved estimation precision. These results highlight that while the regression estimator offers the best performance overall—particularly in the high-correlation setting—composite and ratio estimators provide competitive alternatives, especially as the ranking accuracy improves.

Additional simulation results, including bias and variance graphs under lognormal data, are located in Appendix A.2.

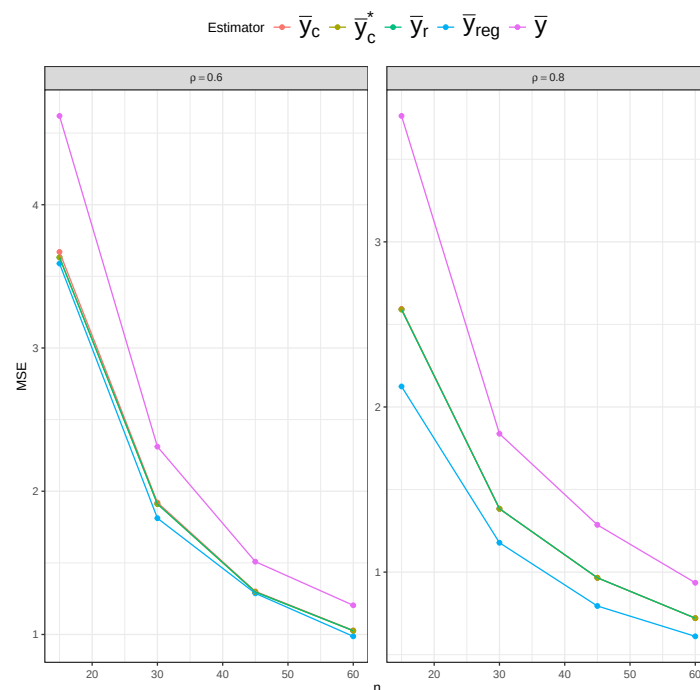


Figure 10. MSE of five mean estimators under SRS from lognormal population.

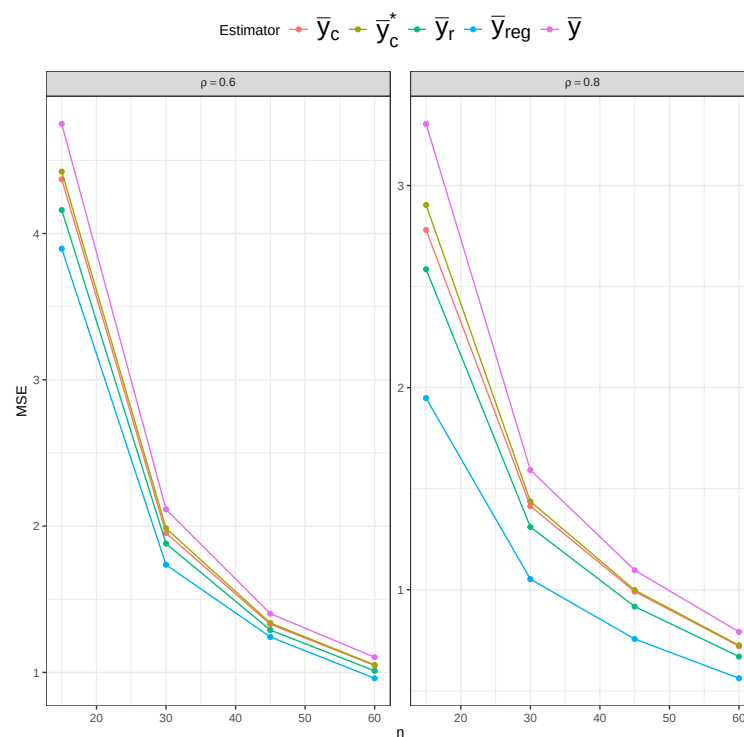


Figure 11. MSE of five mean estimators under imperfect-ranking RSS (fixing $p = 5$) from lognormal population.

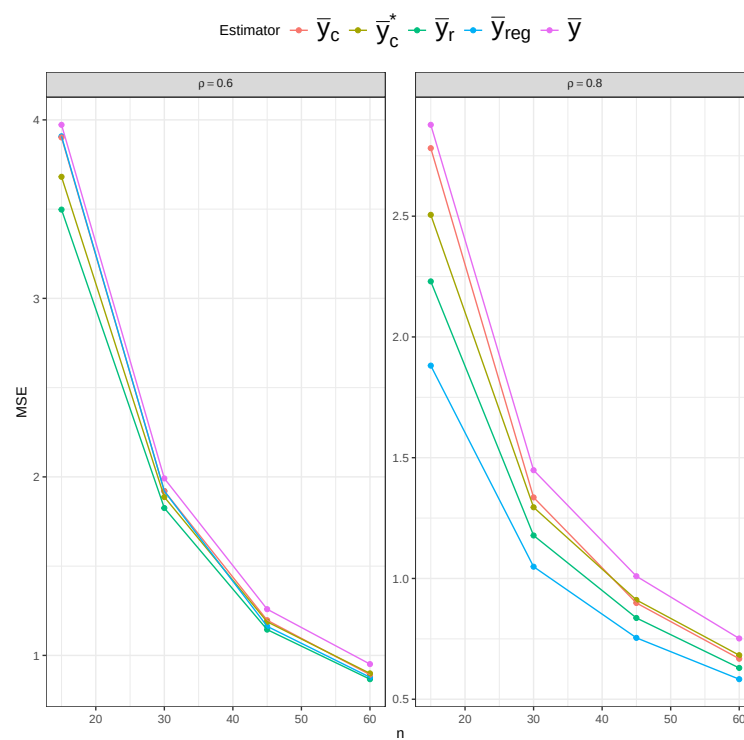


Figure 12. MSE of five mean estimators under perfect-ranking RSS (fixing $p = 5$) from lognormal population.

5. Real Data Application

For the real data application, we use a modified version of the longleaf pine dataset comprising $N = 396$ trees, as analyzed by Jafari Jozani and Johnson [20]. The original dataset can be found in Chen et al. [21] and Platt et al. [22]. In this dataset, the variable X denotes tree diameter at breast height, while Y represents tree height [20]. To remove ties when ranking based on the X variable, a small amount of random noise was added to the X variable: $\varepsilon_x \sim N(0, 0.000001)$, recentered at zero. The Pearson correlation between the transformed X variable and the Y variable was $\rho = 0.91$. Considering the $N = 396$ observations as the finite population, we drew 10,000 samples using each of three sampling designs—SRS, RSS with imperfect ranking (ranking based on X), and RSS with perfect ranking (ranking based on Y)—and computed each of the five estimators considered in the simulation study in the previous section. Goodness-of-fit tests—using the Shapiro–Wilk (‘sw’) option in the ‘gofTest()’ function of the *EnvStats* package ver. 3.1.0 in R [23]—confirmed that both X and Y can be modeled by the lognormal distribution, justifying the application of the lognormal approximation of the moments of sample order statistics (see Equation (23)) under this dataset.

Figures 13–15 present the distribution of five mean estimators under different sampling designs applied to the longleaf pine dataset, with the blue horizontal line indicating the true population mean. In Figure 13 (SRS), the simple mean estimator ($\bar{y}$) exhibits the greatest variance, as its boxplots are widely spread. The ratio ($\bar{y}_r$) and composite ($\bar{y}_c$ and $\bar{y}_c^*$) estimators perform best under SRS, showing less outliers and relatively low variance. The regression ($\bar{y}_{reg}$) estimator follows with slightly higher outlier presence, especially at small sample sizes. All estimators show low-to-no bias, as the boxplots are mostly centered at the true population mean. Similar patterns are observed under imperfect-ranking RSS (Figure 14), but with the composite estimators switching places with the regression estimators. Under perfect-ranking RSS (Figure 15), the simple mean estimator still shows the largest variability across different sample sizes, but it possesses the least bias overall. The ratio and regression estimators again show slightly lower variability than the composite estimators. However, the composite estimators are more centered around the true mean, indicating lower bias than the ratio and regression estimators.

Table 1 displays the MSE behavior of five estimators under the SRS and RSS designs. The bias and variance results are included in Appendix A.3. Across all three sampling designs, the simple mean estimator ($\bar{y}$) exhibits notably higher MSE values than the other estimators across all sample sizes, confirming its inefficiency. However, the impact of the sampling design on the simple mean estimator’s MSE is most evident, as we can clearly see lower MSE values under RSS than under SRS, and under perfect-ranking RSS than under imperfect-ranking RSS. The remaining four estimators have similar MSE values under SRS. The ratio ($\bar{y}_r$) and regression ($\bar{y}_{reg}$) estimators show superior performance under imperfect-ranking RSS across all sample sizes. The superiority of the ratio and regression estimators remains under perfect-ranking RSS, especially for the smaller sample sizes, with the composite estimators ($\bar{y}_c$ and $\bar{y}_c^*$) approaching them for the larger sample sizes ($n = 45, 60$).

Table 1. MSE of five mean estimators under different sampling designs from longleaf pine dataset. For RSS, set size is fixed at $p = 5$.

Sampling Method	Estimator	$n = 15$	$n = 30$	$n = 45$	$n = 60$
SRS	$\bar{y}_c$	42.42	19.68	12.57	8.948
	$\bar{y}_c^*$	40.77	19.53	12.54	8.944
	$\bar{y}_r$	40.77	19.53	12.54	8.944
	$\bar{y}_{reg}$	44.33	19.33	12.17	8.530
	$\bar{y}$	212.71	99.80	63.40	45.346

Table 1. Cont.

Sampling Method	Estimator	$n = 15$	$n = 30$	$n = 45$	$n = 60$
Imperfect-ranking RSS	$\bar{y}_c$	60.31	30.49	20.12	15.389
	$\bar{y}_c^*$	62.30	31.06	20.37	15.534
	$\bar{y}_r$	40.37	19.90	13.08	9.821
	$\bar{y}_{reg}$	41.39	19.44	12.68	9.409
	$\bar{y}$	107.72	53.87	35.48	27.143
Perfect-ranking RSS	$\bar{y}_c$	45.08	24.63	16.93	13.231
	$\bar{y}_c^*$	51.91	26.21	17.55	13.576
	$\bar{y}_r$	36.42	19.36	13.64	11.108
	$\bar{y}_{reg}$	39.10	20.60	14.66	12.164
	$\bar{y}$	93.05	47.09	30.80	23.752

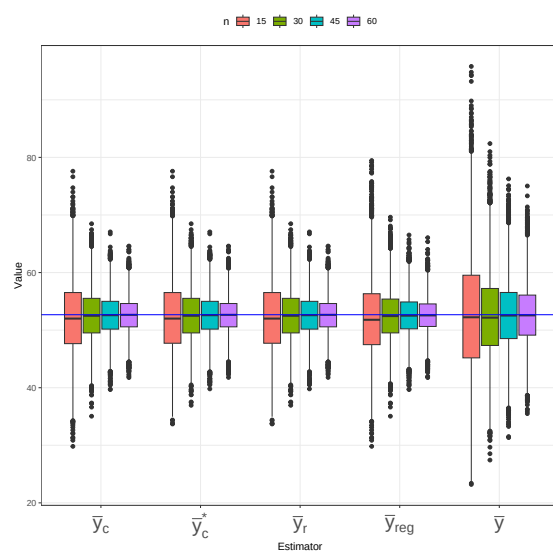
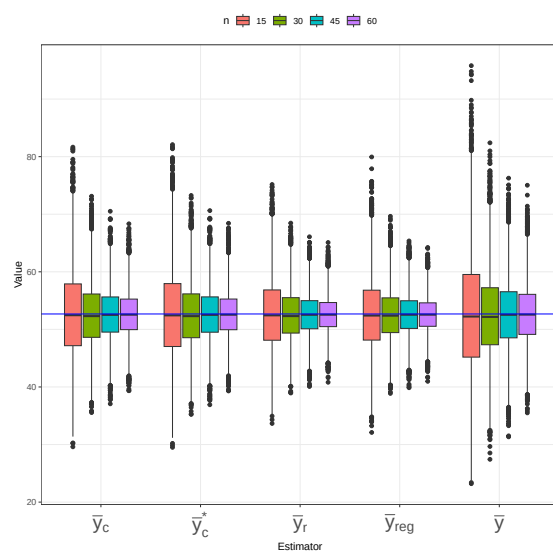


Figure 13. Distribution of five mean estimators under SRS from longleaf pine dataset.

Figure 14. Distribution of five mean estimators under imperfect-ranking RSS (fixing $p = 5$) from longleaf pine dataset.

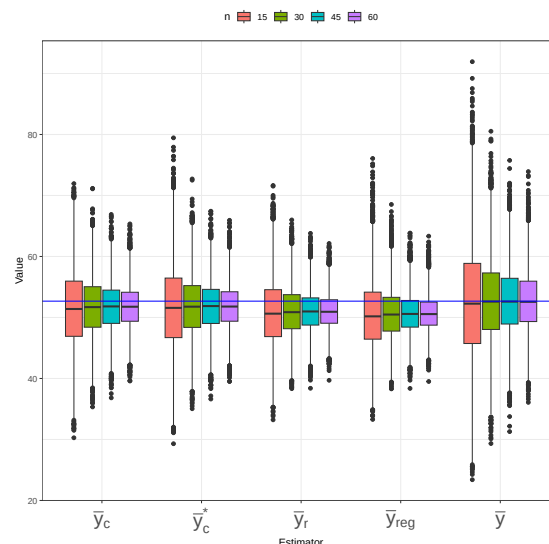


Figure 15. Distribution of five mean estimators under perfect-ranking RSS (fixing $p = 5$) from longleaf pine dataset.

6. Discussion

In this study, we introduced a composite estimator for estimating the finite population mean under RSS. The estimator is defined as a weighted average of the simple mean and the ratio estimator. We derived its MSE and obtained the optimal weight that minimizes the MSE, along with a plug-in estimator for this optimal weight. We evaluated the performance of the composite estimator—both with the known optimal weight and the estimated (plug-in) weight—via Monte Carlo simulations using bivariate normal and lognormal populations, as well as a real finite population of longleaf pine trees. Comparisons were made against the simple mean, ratio, and regression estimators under SRS, imperfect-ranking RSS, and perfect-ranking RSS. The results indicate that the proposed composite estimator is highly efficient, particularly for moderate-to-large sample sizes, even in cases where the simple mean and ratio estimators perform poorly. The regression estimator consistently emerged as a strong competitor across all settings. For all three designs, the performance of the composite, ratio, and regression estimators improved as the correlation between the study and auxiliary variables increased. This trend also benefited the simple mean estimator under imperfect-ranking RSS, as higher correlation improved ranking accuracy. These findings were consistent across normal, lognormal, and real data scenarios. Although the plug-in version of the composite estimator performed well, its efficiency may be further enhanced through refinements in estimating the optimal weight. A key step in this process involves estimating the moments of the order statistics, $\mu_{x(k)}$ and $\mu_{y[k]}$, which we approximated using known formulas for normal and lognormal distributions. Similar approximations exist for other distributions and could be explored in future work. Another promising direction is extending composite estimators to settings with multivariate auxiliary variables, where one or more variables could be used for ranking, estimation, or both. Such generalizations could further enhance the utility and flexibility of composite estimators for survey data.

Author Contributions: Conceptualization, S.A.M. and S.J.; methodology, S.A.M. and S.J.; software, S.A.M., T.J.III, and S.J.; formal analysis, S.A.M. and S.J.; investigation, S.A.M., T.J.III, and S.J.; resources, S.A.M.; data curation, T.J.III; writing—original draft preparation, T.J.III and S.J.; writing—review and editing, S.A.M., T.J.III, and S.J.; supervision, S.A.M.; project administration, S.A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work of Thomas Johnson III was funded by the North Carolina A&T State University Chancellor’s Distinguished Fellowship, a Title III HBGI grant from the U.S. Department of Education.

Data Availability Statement: The original data presented in the study are openly available in [5].

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Additional Simulation Results

Appendix A.1. The Normal Case

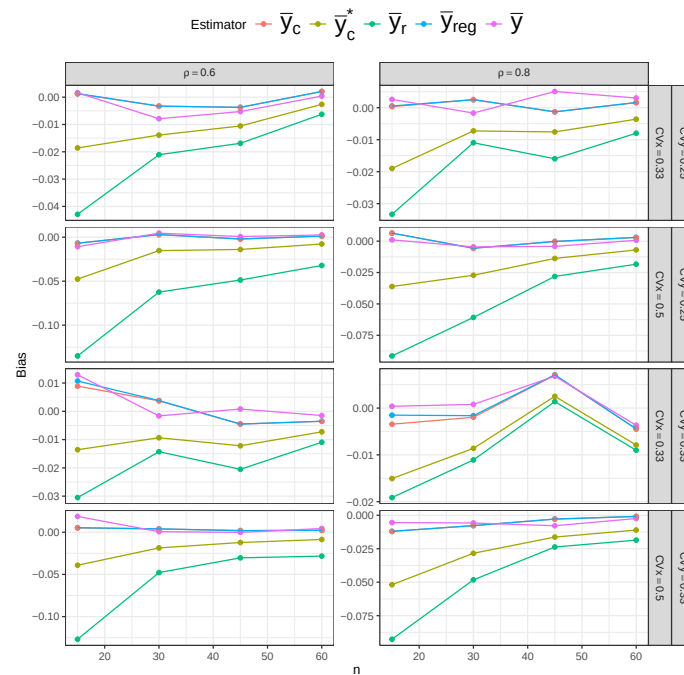


Figure A1. Bias of five mean estimators under SRS from normal population.

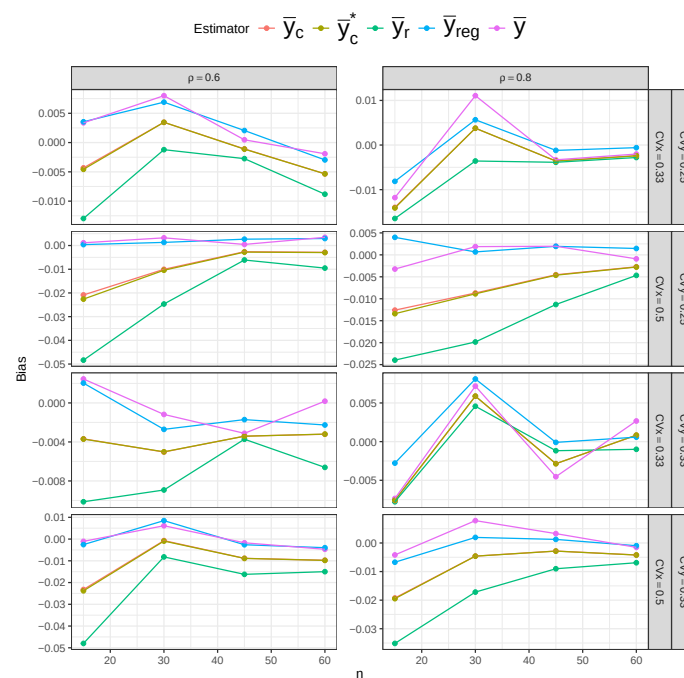


Figure A2. Bias of five mean estimators under imperfect-ranking RSS (fixing $p = 5$) from normal population.

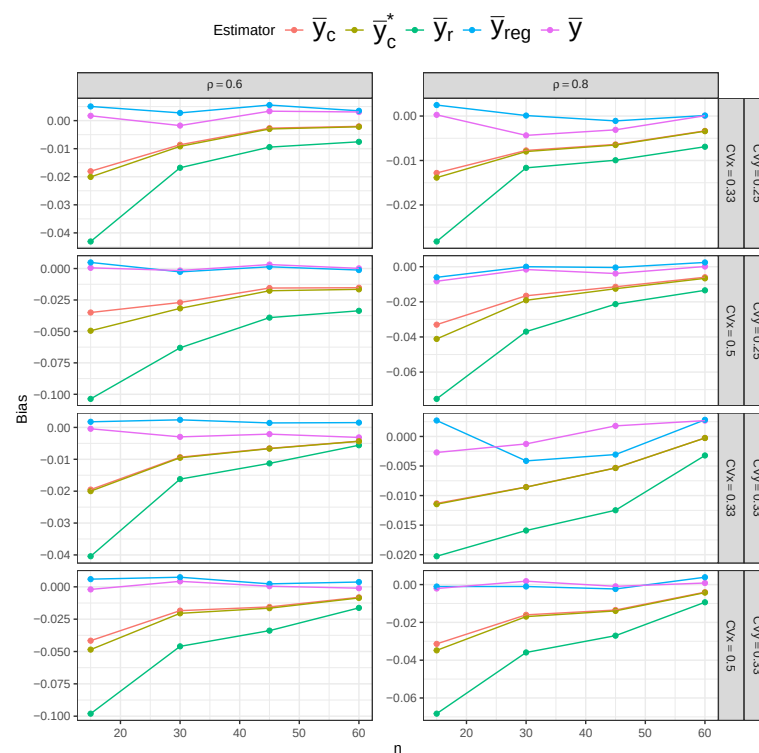


Figure A3. Bias of five mean estimators under perfect-ranking RSS (fixing $p = 5$) from normal population.

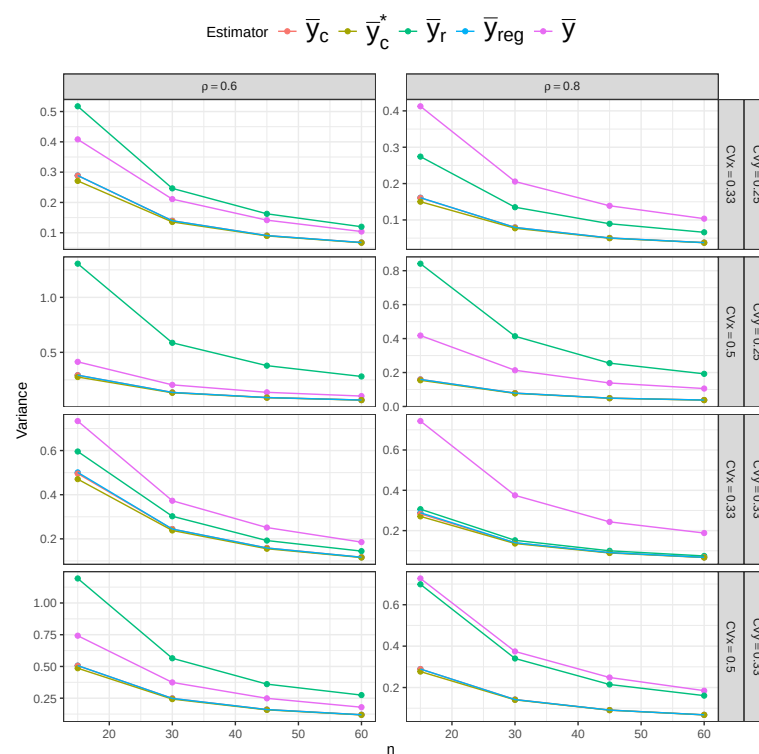


Figure A4. Variance of five mean estimators under SRS from normal population.

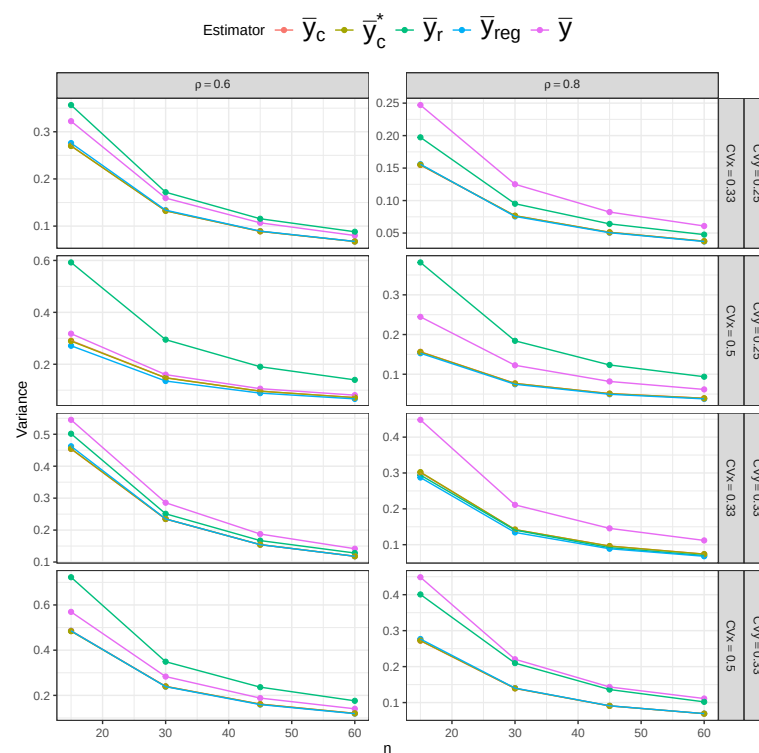


Figure A5. Variance of five mean estimators under imperfect-ranking RSS (fixing $p = 5$) from normal population.

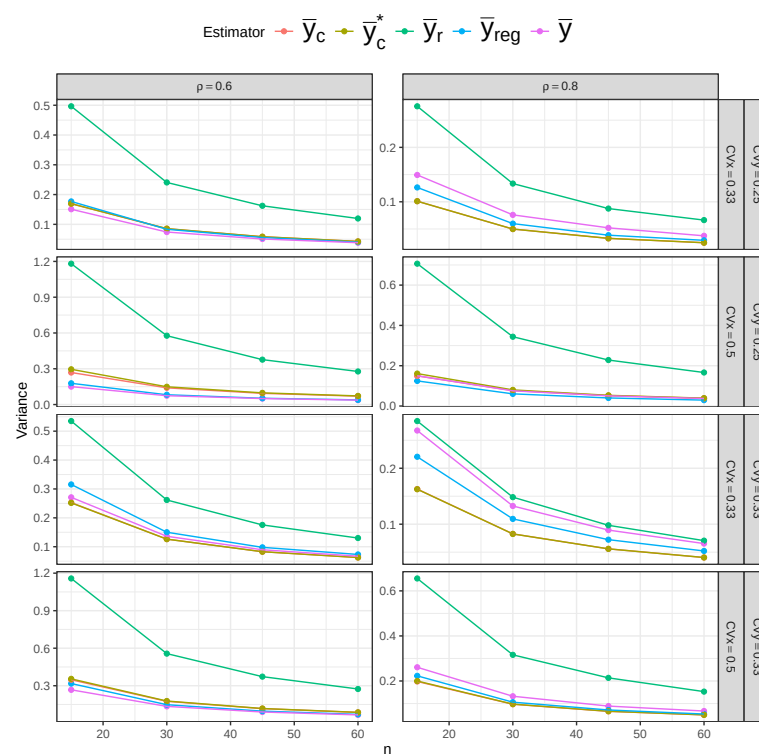


Figure A6. Variance of five mean estimators under perfect-ranking RSS (fixing $p = 5$) from normal population.

Appendix A.2. The Lognormal Case

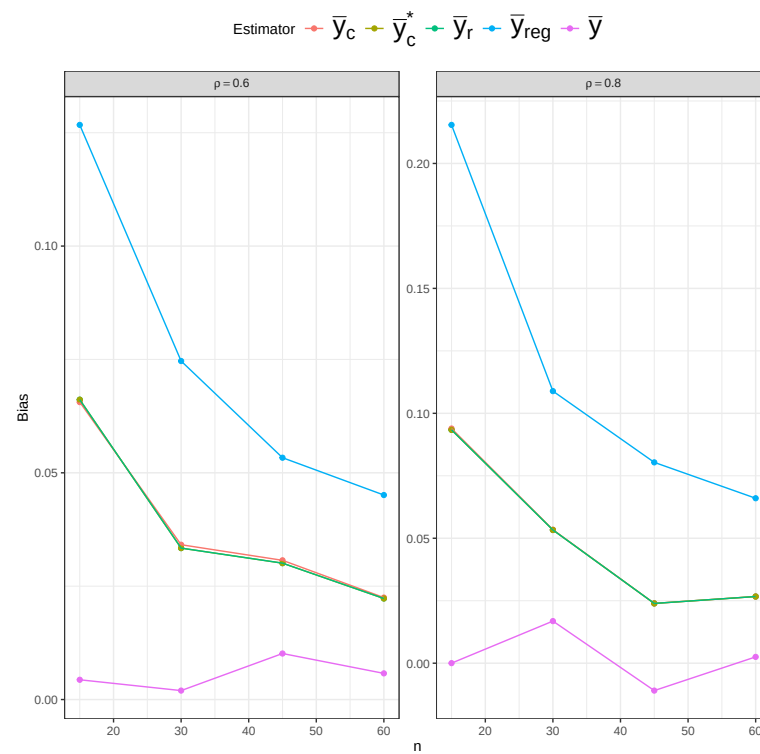
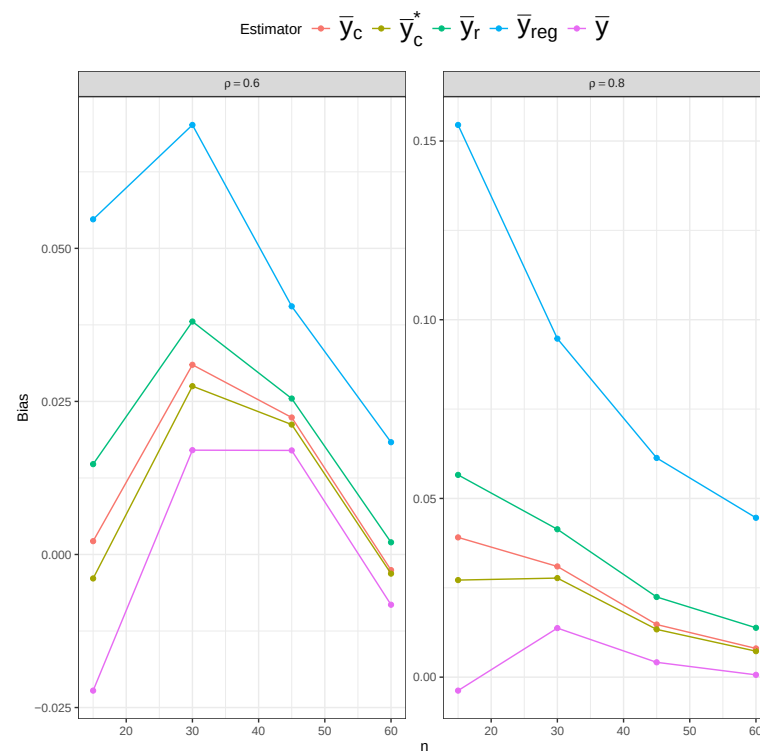


Figure A7. Bias of five mean estimators under SRS from lognormal population.

Figure A8. Bias of five mean estimators under imperfect-ranking RSS (fixing $p = 5$) from lognormal population.

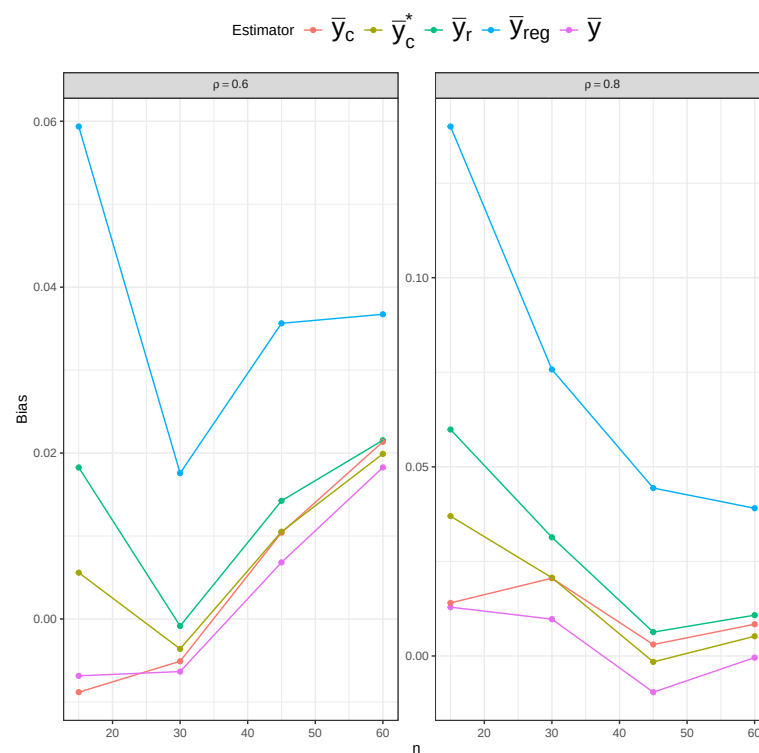


Figure A9. Bias of five mean estimators under perfect-ranking RSS (fixing $p = 5$) from lognormal population.

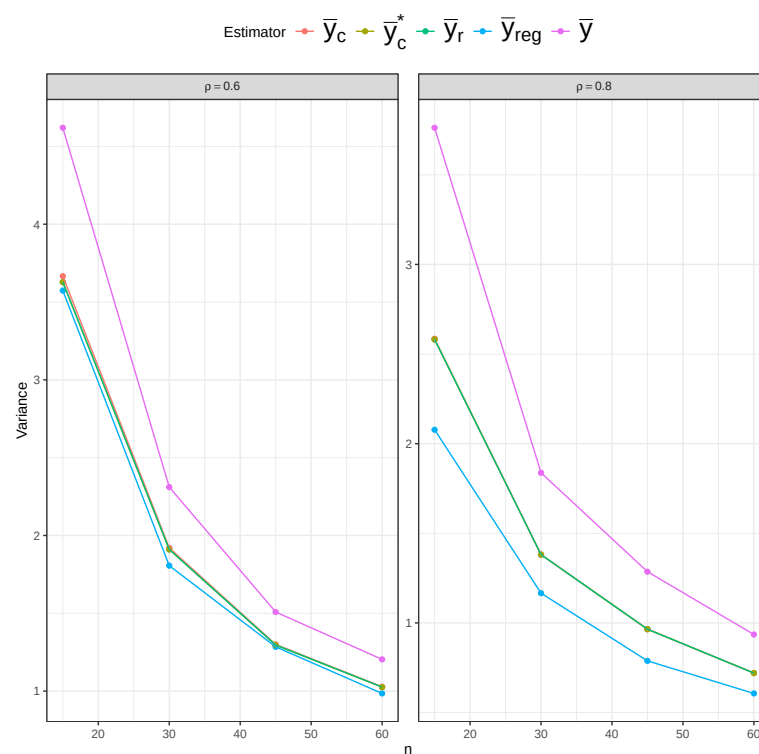


Figure A10. Variance of five mean estimators under SRS from lognormal population.

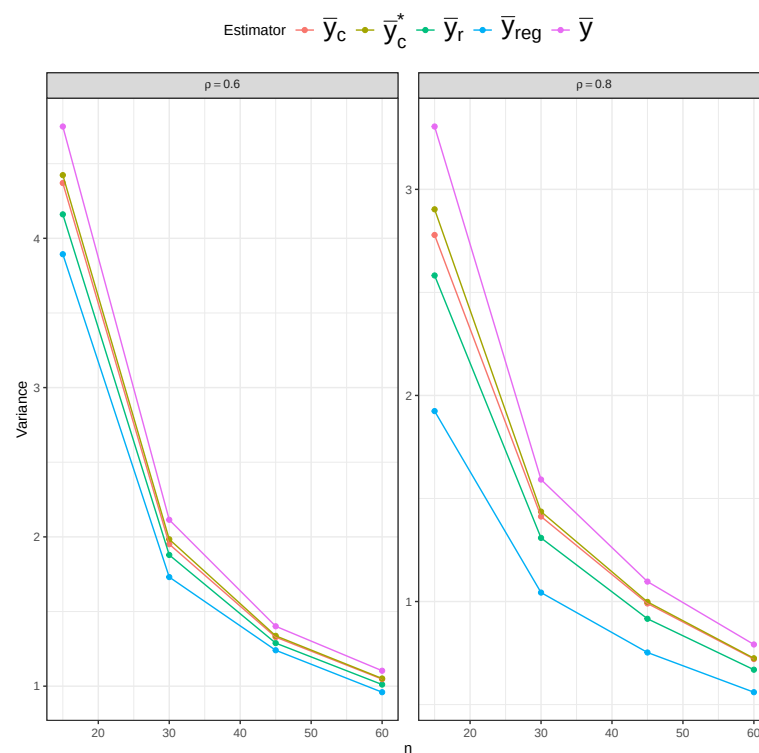


Figure A11. Variance of five mean estimators under imperfect-ranking RSS (fixing $p = 5$) from lognormal population.

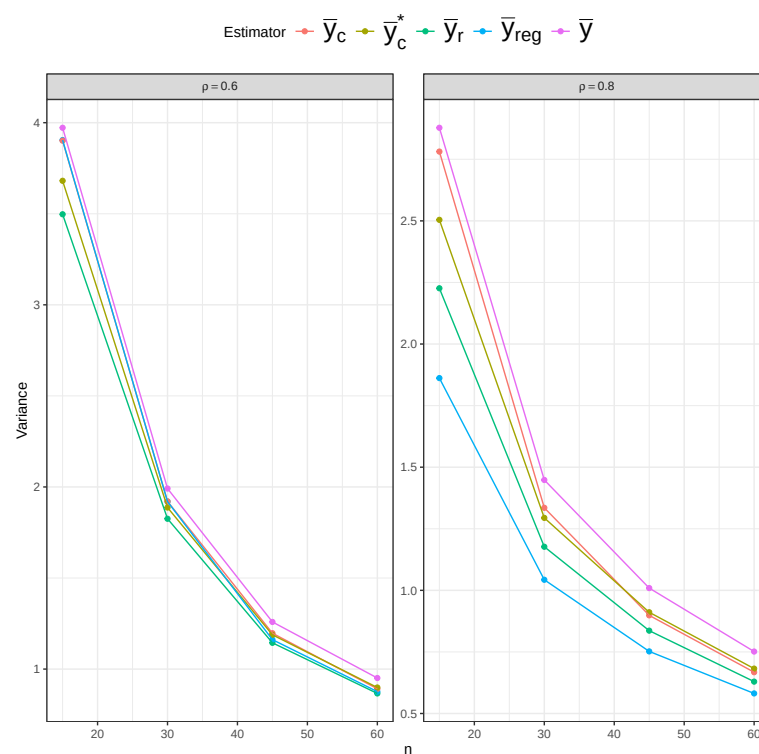


Figure A12. Variance of five mean estimators under perfect-ranking RSS (fixing $p = 5$) from lognormal population.

Appendix A.3. Real Data Application

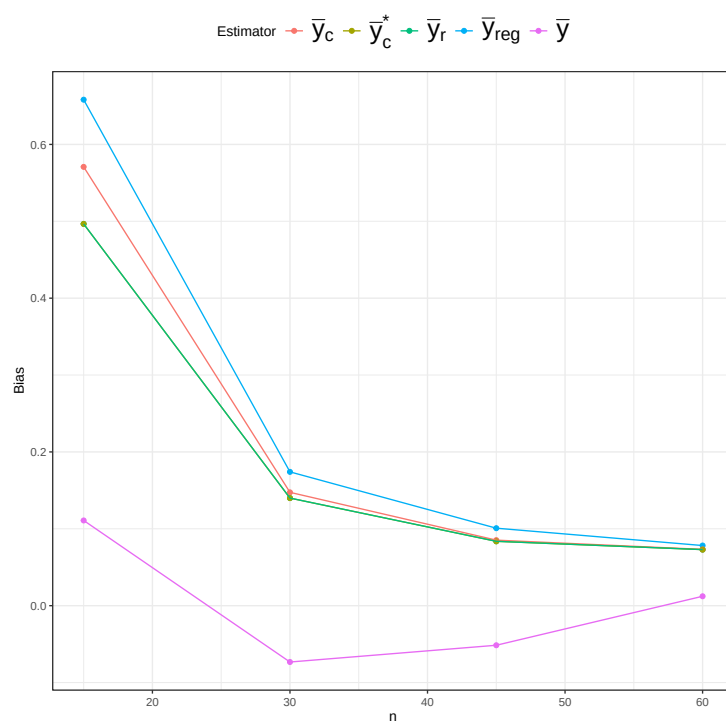


Figure A13. Bias of five mean estimators under SRS from longleaf pine dataset.

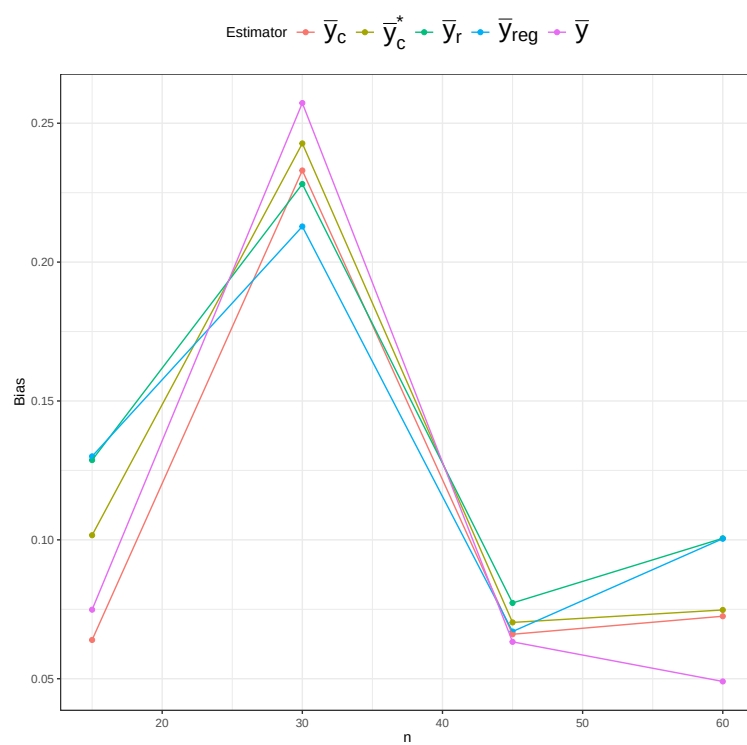


Figure A14. Bias of five mean estimators under imperfect-ranking RSS (fixing $p = 5$) from longleaf pine dataset.

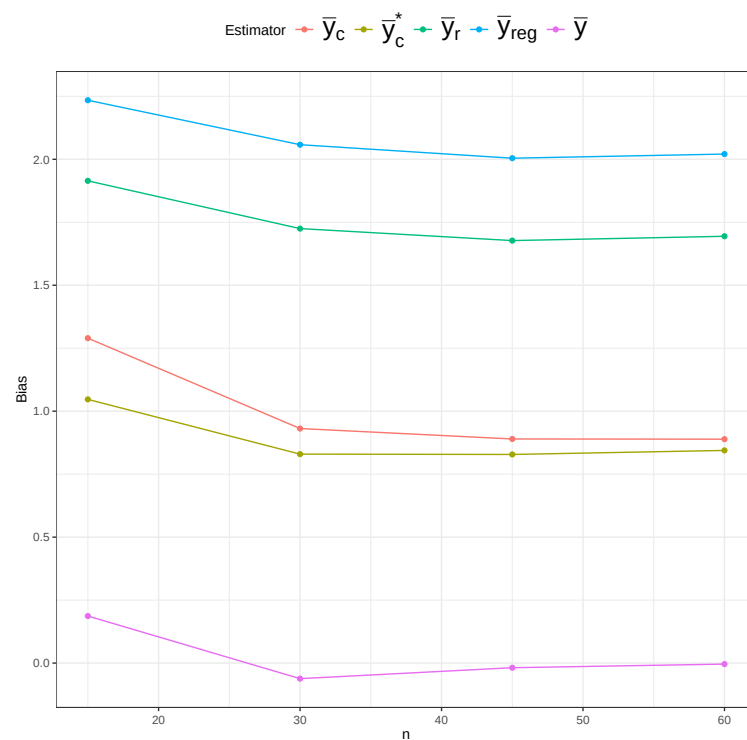


Figure A15. Bias of five mean estimators under perfect-ranking RSS (fixing $p = 5$) from longleaf pine dataset.

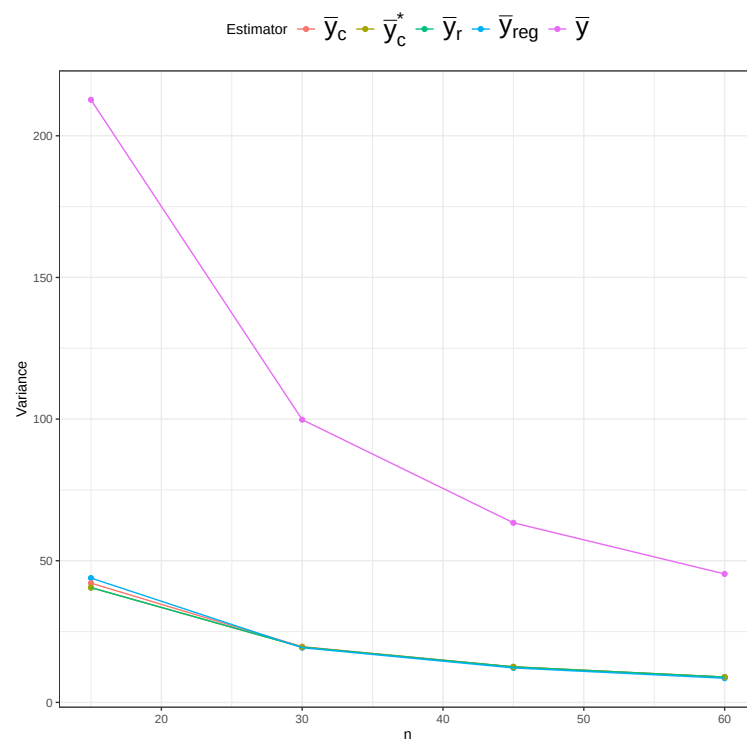


Figure A16. Variance of five mean estimators under SRS from longleaf pine dataset.

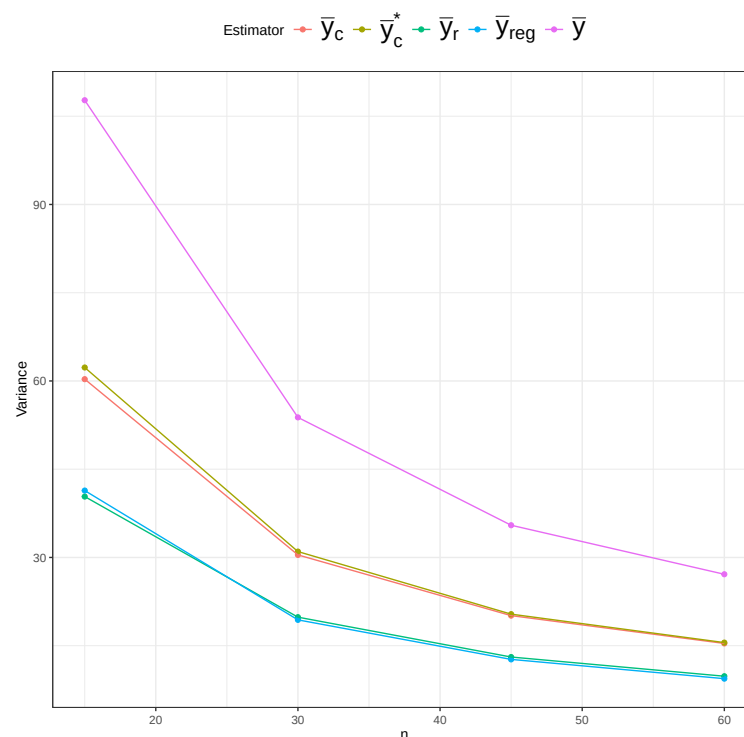


Figure A17. Variance of five mean estimators under imperfect-ranking RSS (fixing $p = 5$) from longleaf pine dataset.

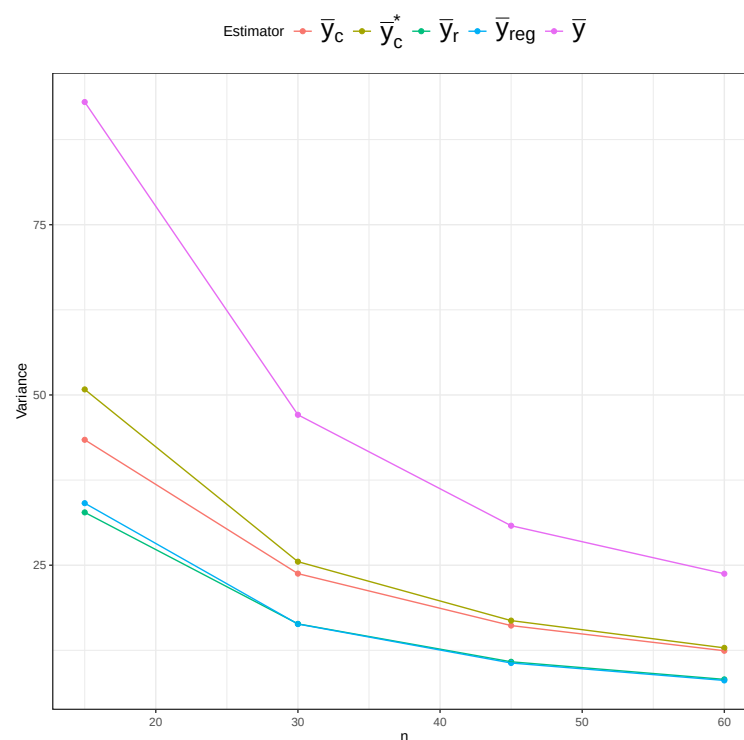


Figure A18. Variance of five mean estimators under perfect-ranking RSS (fixing $p = 5$) from longleaf pine dataset.

References

1. McIntyre, G.A. A method for unbiased selective sampling using ranked sets. *Aust. J. Agric. Res.* **1952**, *3*, 385–390. [\[CrossRef\]](#)
2. Wolfe, D.A. Ranked set sampling: An approach to more efficient data collection. *Stat. Sci.* **2004**, *19*, 636–643. [\[CrossRef\]](#)
3. Takahasi, K.; Wakimoto, K. On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. Stat. Math.* **1968**, *20*, 1–31. [\[CrossRef\]](#)

4. Stokes, S.L. Estimation of Variance Using Judgment Ordered Ranked Set Samples. *Biometrics* **1980**, *36*, 35–42. [[CrossRef](#)]
5. Chen, Z. On ranked-set sample quantiles and their applications. *J. Stat. Plan. Inference* **2000**, *83*, 125–135. [[CrossRef](#)]
6. Stokes, S.L.; Sager, T.W. Characterization of a Ranked-Set Sample with Application to Estimating Distribution Functions. *J. Am. Stat. Assoc.* **1988**, *83*, 374–381. [[CrossRef](#)]
7. Bohn, L.L.; Wolfe, D.A. Nonparametric Two-Sample Procedures for Ranked-Set Samples Data. *J. Am. Stat. Assoc.* **1992**, *87*, 552–561. [[CrossRef](#)]
8. Hettmansperger, T.P. The ranked-set sample sign test. *J. Nonparametric Stat.* **1995**, *4*, 263–270. [[CrossRef](#)]
9. Koti, K.M.; Jogesh Babu, G. Sign test for ranked-set sampling. *Commun. Stat.-Theory Methods* **1996**, *25*, 1617–1630. [[CrossRef](#)]
10. Chen, Z. Density estimation using ranked-set sampling data. *Environ. Ecol. Stat.* **1999**, *6*, 135–146. [[CrossRef](#)]
11. Wolfe, D.A. Ranked set sampling. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 460–466. [[CrossRef](#)]
12. Cochran, W.G. *Sampling Techniques*, 3rd ed.; Wiley: New York, NY, USA, 1977. [[CrossRef](#)]
13. Samawi, H.M.; Muttalak, H.A. Estimation of ratio using rank set sampling. *Biom. J.* **1996**, *38*, 753–764. [[CrossRef](#)]
14. Kadilar, C.; Unyazici, Y.; Cingi, H. Ratio estimator for the population mean using ranked set sampling. *Stat. Pap.* **2009**, *50*, 301–309. [[CrossRef](#)]
15. Philip, L.; Lam, K. Regression estimator in ranked set sampling. *Biometrics* **1997**, *53*, 1070–1080. [[CrossRef](#)]
16. Lui, K.J. Notes on Use of the Composite Estimator: An Improvement of the Ratio Estimator. *J. Off. Stat.* **2020**, *36*, 137–149. [[CrossRef](#)]
17. Sevinc, B.; Cetintav, B.; Esemien, M.; Gurler, S. *RSSampling: Ranked Set Sampling*, R package version 1.0; 2018. Available online: <https://CRAN.R-project.org/package=RSSampling> (accessed on 21 July 2025).
18. Royston, J.P. Algorithm AS 177: Expected Normal Order Statistics (Exact and Approximate). *J. R. Stat. Society. Ser. C (Appl. Stat.)* **1982**, *31*, 161–165. [[CrossRef](#)]
19. Nadarajah, S. Explicit expressions for moments of log normal order statistics. *Econ. Qual. Control* **2008**, *23*, 267–279. [[CrossRef](#)]
20. Jafari Jozani, M.; Johnson, B.C. Design based estimation for ranked set sampling in finite populations. *Environ. Ecol. Stat.* **2011**, *18*, 663–685. [[CrossRef](#)]
21. Chen, Z.; Bai, Z.; Sinha, B.K. *Ranked Set Sampling: Theory and Applications*; Springer: Berlin/Heidelberg, Germany, 2004; Volume 176. [[CrossRef](#)]
22. Platt, W.J.; Evans, G.W.; Rathbun, S.L. The population dynamics of a long-lived conifer (*Pinus palustris*). *Am. Nat.* **1988**, *131*, 491–525. [[CrossRef](#)]
23. Millard, S.P. *EnvStats: An R Package for Environmental Statistics*; Springer: New York, NY, USA, 2013. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.