Franklin Open 8 (2024) 100139

Contents lists available at ScienceDirect

Franklin Open

journal homepage: www.elsevier.com/locate/fraope

Reproducibility of mean estimators under ranked set sampling

Syed Abdul Rehman [a,*], Tahani Coolen-Maturi [b], Frank P.A. Coolen [b], Javid Shabbir [c]

[a] Department of Statistics, Quaid-i-Azam University, Islamabad, 45320, Pakistan
[b] Department of Mathematical Sciences, Durham University, Durham, DH1 3LB, UK
[c] Department of Statistics, University of Wah, Wah Cannt, 47040, Pakistan

ARTICLE INFO

Dataset link: https://archive.ics.uci.edu/dataset/1/abalone

Keywords:
Order statistics
Ranked set sampling
Statistical reproducibility
Mean squared deviation

ABSTRACT

In statistical inferences, the estimation of population parameters using information obtained from a sample is an important method. This involves choosing an appropriate sampling method to collect data. An efficient sampling method used for data collection is Ranked Set Sampling (RSS). In this study, we investigate the reproducibility of four well-known mean estimators under RSS using parametric predictive bootstrapping. These estimators are called conventional, ratio, exponential ratio, and regression estimators. Reproducibility is the ability of a statistical technique to obtain results similar to those based on the original experiment if the experiment is repeated under the same conditions. We conduct a simulation study to compare the reproducibility of mean estimators for varying sample sizes when sampling is based on perfect and imperfect rankings. We consider data on abalone in our simulations to demonstrate real-world applications. This study concludes that the regression estimator is the best reproducible estimator, while the conventional estimator is the worst in this regard.

1. Introduction

Estimating the population parameters from sample data is an important aspect of statistical inferences. This involves selecting a group of individuals from the population who are believed to represent the entire population of interest. Sampling allows researchers to efficiently collect and analyze data from a subset of the population, saving time and resources while estimating reliable results. Although Simple Random Sampling (SRS) is an easy sampling technique that does not require many assumptions, the estimates based on it sometimes lack precision. In order to yield more precise estimates, other sampling strategies, such as Ranked Set Sampling (RSS) are available in the literature. The RSS was initially presented by [1] and is thought to be more suitable in situations where assessing the population units are costly but ranking them is inexpensive. This method improves estimation precision by reducing sampling error. [2] developed an unbiased estimator of the population mean using the RSS method. [3] introduced the use of the concomitant variable to order units of the study variable; for example, one can use the length of the abalone to rank the data on the weights of the abalone. Researchers are currently focused on efficiently estimating the population mean of a study variable under RSS by using the information of the auxiliary variable in the estimation stage. [4] proposed the regression estimator of the population mean under RSS, while [5] proposed the ratio estimator to estimate the population mean of the study variable using the known population mean of the auxiliary variable. The exponential ratio estimator of the population mean under RSS was proposed by [6]. Many other related studies in this regard can be seen in [7–12], and [13]. Previous studies have mainly focused on designing estimators and evaluating them based on relative efficiency and mean square error. To address the methodological limitations, we introduce a new measure, called the reproducibility of estimators. This measure provides a more comprehensive framework for comparing estimators, improving the theoretical foundation of estimation theory.

In the last decade, statistical reproducibility has received the attention of researchers. When estimating population parameters, it refers to the ability to obtain consistently similar estimates when conducting any statistical method, analysis, or experiment several times using the same methods. Reproducing previous findings is essential for statistical methods to enhance evidence. It serves as a primary aspect of scientific research to ensure the reliability and validity of statistical methods and their findings. [14] addressed the topic of the reproducibility of a statistical test and discussed the confusion between reproducibility and the statistical p-value. [15] emphasized how the p-value and reproducibility are different. [16] expresses doubts about drawing meaningful conclusions from a single initial experiment, as the power of the test will be unknown due to the lack of knowledge of the effective sample size. [17] estimated reproducibility by comparing the value of a test statistic calculated based on the actual test with the associated critical value. [18]

* Corresponding author.
E-mail address: rehmanbukhari725@gmail.com (S.A. Rehman).

https://doi.org/10.1016/j.fraope.2024.100139
Received 3 May 2024; Received in revised form 10 June 2024; Accepted 26 July 2024
Available online 29 July 2024

suggested three techniques for estimating reproducibility, including the Bayesian approach. The predictive nature of reproducibility was emphasized by [19], who also relates it to the useful sample size. Multiple definitions of the reproducibility of a statistically significant result were proposed by [20]. He also assessed various reproducibility estimators for the Wilcoxon rank sum test and compared their effectiveness. [21] reproduced approximately 25% significant results in preclinical cancer trials, highlighting concerns about the reproducibility of the experiment in this context.

On the other hand, bootstrapping is a statistical technique that estimates sampling distributions by resampling sample data, unlike conventional approaches that rely on theoretical assumptions. Though many bootstrapping methods are available in the literature, a novel technique for bootstrapping was recently introduced by [22] called parametric predictive bootstrap (PPB). This technique is primarily designed for predictive inference based on parametric models. The PPB allows us to make predictions about future observations, assuming that the underlying distribution has some known parameter(s). This is related to the concept of reproducibility probability (RP) in the traditional frequentist statistical framework, estimated by [23]. Considering the explicitly predictive nature of PPB, reasonable conclusions regarding RP can be drawn. [22] estimated the RP of some parametric tests using the PPB approach. He argued that the explicitly predictive nature of PPB offers an appropriate formulation for the RP inference, as the nature of RP is explicitly predictive as well. He also compared the performance of PP-B for RP with the nonparametric predictive inference bootstrap (NPI-B) method, which also has a predictive nature but does not assume a parametric model.

In this study, we compare the reproducibility of four well-known mean estimators called conventional, ratio, exponential ratio, and regression estimators under RSS employing PPB bootstrapping. Following [24], the reproducibility of an estimator is defined as the probability that, if the RSS sampling is repeated under the same conditions, the mean estimates based on the future sample will be similar to the estimate based on the original sample. Through a comprehensive simulation study, we investigate the reproducibility of these mean estimators for different sample sizes and ranking criteria. The purpose of the study is to contribute to the literature on RSS by examining the reproducibility of mean estimators under RSS using PPB bootstrapping and identifying potential limitations and advantages. The paper begins with an overview of the RSS method in Section 2. Section 3 reviews the mean estimators that involve auxiliary information, while Section 4 explains the concept of statistical reproducibility, highlighting its definition and importance. Section 5 presents the PPB procedure, while Section 6 provides integration of PPB with mean estimators under RSS. In Section 7, a simulation study is conducted to analyze the RP of the mean estimators under RSS. Section 8 presents an application of real-world data, while Section 9 summarizes key findings and conclusions.

2. The ranked set sampling

This method was introduced by [1] and it select a sample in two phases. In the first phase, random sets of units are examined, while in the second phase, a sample of usually small size is selected from the sets of the first phase for estimation. The process of selecting a sample of size m using RSS involves the following steps:

1. **Identification and Assignment:** Identify m^2 units from the population and assign these units to independent m sets of size m.
2. **Ranking in Each Set:** Rank the units within each set. Use visual judgments on the study variable or the order of a closely related auxiliary variable for ranking.
3. **Selection of Order Statistics:** Select the first order statistic from the first set. The second order statistic is selected from the second set. Continue this process until the mth order statistic is selected from the mth set.

This allows us to select a sample of size m. Repetition of this procedure g times yields a final sample of size $n = gm$. The selected sample can be represented as $\{Y_{i(i)j}\}$ such that $i = 1,\dots,m$ and $j = 1,\dots,g$. An estimator of the population mean based on this method was developed by [2] as

$$t_1 = \bar{y}_{rss} = \frac{1}{gm}\sum_{j=1}^{g}\sum_{i=1}^{m} Y_{i(i)j}. \tag{1}$$

where $Y_{i(i)j}$ shows the ith order statistics in the ith set of the jth RSS cycle. The estimator t_1 is called conventional mean estimator under RSS and it is unbiased while its variance is given by

$$Var(t_1) = \frac{\sigma_y^2}{gm} - \frac{1}{gm^2}\sum_{i=1}^{m}\Delta_{y(i)}^2, \tag{2}$$

where $\frac{\sigma_y^2}{gm}$ is equal to the variance of the mean estimator under SRS sample of the same size m, whereas the term $\Delta_{y(i)} = \mu_{y(i)} - \mu_y$ shows deviation of the ith order statistics mean $\mu_{y(i)}$ from the overall population mean μ_y. Eq. (2) shows that the conventional mean estimator t_1 is more precise than the conventional mean estimator under SRS given that $\mu_{y(i)} \neq \mu_y$.

2.1. The perfect and imperfect ranking

If the ranking is determined solely by the characteristics of the study variable itself, it is known as perfect ranking. This can be done using the visual assessments of the surveyor. Though this type of ranking is easy and inexpensive, one can reasonably assume that it is free from error. This type of ranking is usually used when there is a natural hierarchy or order among the sampling units. In situations where perfect ranking is not possible, [3] proposed an alternative way of ranking the units of the study variable. He ranked units of the study variable using the order of a closely related auxiliary variable, assuming that information is available or can be easily obtained on the auxiliary variable. He termed this imperfect ranking, and some researchers referred to this ranking with errors, since it is possible to place a larger unit before any smaller unit when units are observed. This is expected when the association between the study variable and the concomitant variable is low. This type of error generally reduces the efficiency of estimates. In our study, we consider both types of ranking for estimating the population mean while comparing the reproducibility of mean estimators under RSS.

3. Mean estimators involving the auxiliary variable

The auxiliary variables play an important role in the estimation of parameters for the study variables. Information on auxiliary variables can be used to design more efficient estimators of population parameters. In this section, we review three commonly discussed mean estimators in RSS that involve known information on the population mean of the auxiliary variable. The expressions for bias and MSE of these estimators are also provided, for which the following notation and symbols are used:

$$V_y = \left(\frac{N-n}{n}\right)C_y^2 - \frac{1}{gm^2\mu_y^2}\sum_{i=1}^{m}\Delta_{y(i)}^2,$$

$$V_x = \left(\frac{N-n}{n}\right)C_x^2 - \frac{1}{gm^2\mu_x^2}\sum_{i=1}^{m}\Delta_{x(i)}^2,$$

and

$$V_{yx} = \left(\frac{N-n}{n}\right)\rho C_y C_x - \frac{1}{gm^2\mu_y\mu_x}\sum_{i=1}^{m}\Delta_{y(i)}\Delta_{x(i)},$$

where N and n show the population and sample size, respectively. ρ is the correlation coefficient, while C_y and C_x show the coefficient of variation for the study variable and the auxiliary variable, respectively.

3.1. The regression estimator

The regression estimator under RSS was proposed by [4]. This estimator combines the information provided by the difference between the population mean and sample mean of the auxiliary variable along with the covariance between the study variable and the auxiliary variable, leading to a more precise and unbiased estimate of the population mean of the study variable. The regression estimator is given by

$$t_2 = \bar{y}_{rss} + \beta\left(\mu_x - \bar{x}_{rss}\right), \tag{3}$$

where $\beta = \frac{\sigma_{yx}}{\sigma_x^2}$ is population regression coefficient which can be estimated by $\hat{\beta} = \frac{s_{yx}}{s_x^2}$ using the sample information. μ_x shows known population mean of the study variable, whereas $\bar{x}_{rss}$ is the mean of the sample of the auxiliary variable under RSS computed as $\bar{x}_{rss} = \frac{1}{gm}\sum_{j=1}^{g}\sum_{i=1}^{m} X_{i(i)j}$. The MSE of regression estimator is given by

$$MSE\left(t_2\right) = \mu_y^2 V_y\left(1 - \frac{V_{yx}}{V_x^2}\right). \tag{4}$$

3.2. The ratio estimator

The ratio estimator under RSS was proposed by [5]. This estimator utilizes the information obtained from the ratio of the known population mean to the sample mean of the auxiliary variable. In order to develop a more precise estimator, this approach aims to capitalize on the proportionality between the auxiliary and study variables to increase the precision of mean estimates for the study variable. The ratio estimator is given by

$$t_3 = \bar{y}_{rss}\left(\frac{\mu_x}{\bar{x}_{rss}}\right) \tag{5}$$

The bias and MSE of the ratio estimator are given by

$$Bias\left(t_3\right) = \mu_y\left(V_x - V_{yx}\right) \tag{6}$$

and

$$MSE\left(t_3\right) = \mu_y^2\left(V_y + V_x - 2V_{yx}\right) \tag{7}$$

3.3. The exponential ratio estimator

The exponential ratio estimator under RSS was proposed by [6]. This estimator utilizes the information obtained from the exponential ratio of the known population mean to the sample mean of the auxiliary variable. To develop a more efficient estimate of the population mean for the study variable, this approach capitalizes on the exponential relationship. This method is preferred when the data under study have potentially extreme values. The exponential ratio estimator is given by

$$t_4 = \bar{y}_{rss}\exp\left(\frac{\mu_x - \bar{x}_{rss}}{\mu_x + \bar{x}_{rss}}\right) \tag{8}$$

the bias and MSE of the exponential ratio estimator are given by

$$Bias\left(t_4\right) = \mu_y\left(\frac{3}{8}V_x - \frac{1}{2}V_{yx}\right) \tag{9}$$

and

$$MSE\left(t_4\right) = \mu_y^2\left(V_y + \frac{1}{4}V_x + V_{yx}\right) \tag{10}$$

4. Statistical reproducibility

Reproducibility refers to the ability of any scientific study or methodology to produce findings and conclusions similar to those of previous studies or experiments. It is considered an important property of any research and has recently received a great deal of attention from researchers. An overview of recent studies on statistical reproducibility can be studied in [25], while a description of many aspects of reproducibility was provided by [26]. Interestingly, little attention has been paid to the reproducibility of the outcomes of any statistical technique or method, or, generally speaking, the reproducibility of statistical inferences, which are frequently an integral component of investigations. Statistical reproducibility can be defined straightforwardly as if an experiment were repeated under identical conditions, would it lead to the same findings of the statistical analysis as the findings based on the data from the original experiment? Initially, the reproducibility of statistical inferences was studied by [14], who highlighted a widespread misconception about the p-value in hypothesis testing, specifically that a smaller p-value would indicate strong reproducibility. Goodman referred to this as replicability. [15] endorses the concept of Goodman that the p-value and the reproducibility probability are different measures and that inconsistency can be expected between the test results of individual studies. However, he highlighted the significance of the p-value and its relationship with the reproducibility probability. [27] provide a summary of studies proposing a statistical reproducibility measure. [20] introduced an interesting idea, utilizing estimated power as a measure of reproducibility in the event that the null hypothesis is rejected. [28,29] used this method on a number of basic statistical tests. [27] proposed a novel approach to quantifying statistical reproducibility, viewing it as a predictive inference problem, with the aim of determining whether future experiments would yield the same results. [30] proposed a Bayesian predictive approach to address reproducibility as a problem in predictive inference.

5. The parametric predictive bootstrapping

[22] recently proposed a novel bootstrapping technique designed primarily for predictive inference based on parametric models. This method predicts future values based on the assumption that they come from data with specific parameters. The method begins by generating a future observation from an assumed distribution with estimated parameters based on data of sized m. The data is updated by adding this future observation, increasing its size to $m+1$. Another future observation is drawn from the assumed distribution with updated estimates of the parameters. In total, this process is repeated m times to draw m future observations, such that the drawing of every new observation is based on the estimated parameters of the updated data. The generated future observations makes a PPB sample of size m observations, which works well as a method for predictive inference. The steps involved in the formation of a PPB sample from any original data of size m are described below.

1. Consider a random sample $\left\{y_1,\ldots,y_m\right\}$, and estimated parameter θ.
2. Using the maximum likelihood estimation (MLE) or any other estimation method, estimate the parameter θ of the assumed distribution by $\hat{\theta}$.
3. Randomly draw a future observation y_1^* from the fitted distribution $F\left(y;\hat{\theta}\right)$.
4. Update the data by adding y_1^* so that $\left\{y_1,\ldots,y_m,y_1^*\right\}$, the sample size increases to $m+1$.
5. Considering updated data, repeat steps 2–4 and draw another future observation, y_2^*, update the data by adding this observation.
6. In total, steps 2–4 are repeated m times to obtain m future observations $\left\{y_1^*,\ldots,y_m^*\right\}$. This is called PPB sample of size m.

This method of drawing future observations leads to greater variation in the PPB sample compared to other bootstrapping methods, like those proposed by [31]. This method is similar to nonparametric predictive inference bootstrapping (NPI-B), except for the fact that the PPB method assumes that the data comes from a known distribution with known parameters, while the NPI-B method does not consider any parametric assumptions for generating future observations. PP-B bootstrapping samples are not restricted to observed values, unlike

Efron's bootstrapping, which restricts future observations to already observed values. [22] presented the PPB method for the reproducibility probability (RP) of some parametric tests. He argued that the test reproducibility is naturally predictive inference problem, which is consistent with the PP-B method. The explicitly predictive nature of PP-B provides an appropriate formulation for inferring RP, as the nature of RP is explicitly predictive as well. They compared the performance of PPB for RP with the NPI-B method, which also has a predictive nature but does not assume a parametric mode.

6. Reproducibility of mean estimators under RSS

In this section, a mathematical framework for comparing the reproducibility of mean estimators under RSS in the context of the PPB method is presented. Following [22], $RP(\epsilon)$ is the probability that mean estimates computed from reproduced samples will fall within the range of ϵ deviation from the mean estimates based on the original sample, given that the sampling procedure is repeated under the same conditions.

Consider any set of observed values in the setup of RSS method, where m original observations y_i; $i = 1, \dots, n$ are independently and identically distributed. We assume that the distribution $F(y;\theta)$ of y_i is known with parameter θ. Using the steps discussed in Section 5, obtain a PPB sample of m future values and replace it with the original set. This procedure is repeated for all m sets to finally obtain the setup of PPB-RSS. A reproduced sample of PPB is obtained from the setup of PPB-RSS, and the future sample mean is estimated using the discussed mean estimators. Using a similar procedure, we also estimate the future sample mean for the auxiliary variable. In total, we produce M future samples using PPB, and estimate the mean of the study variable and the auxiliary variable from them. We also compute the mean of the original RSS sample.

Let t_O be the mean estimate of the original sample, whereas t_B shows mean estimates based on reproduced samples using PPB. Then the Absolute Average Deviation (AAD) and Mean Square Deviation (MSD) between the estimates are computed as

$$AAD = \frac{1}{M}\sum_{i=1}^{M}\left|t_O - t_{B_i}\right| \tag{11}$$

and

$$MSD = \frac{1}{M}\sum_{i=1}^{M}\left(t_O - t_{B_i}\right)^2. \tag{12}$$

To obtain numerous values of AAD and MSD, this entire process is iterated a large number of times (say, D). Let ϵ be any real valued positive quantity, then $RP_1(\epsilon)$ is the reproducibility probability that AAD is equal to or less than ϵ. Similarly, $RP_2(\epsilon)$ is the reproducibility probability that MSD is equal to or less than ϵ. These probabilities are mathematically computed as

$$RP_1(\epsilon) = \Pr(AAD \leqslant \epsilon) \tag{13}$$

and

$$RP_2(\epsilon) = \Pr(MSD \leqslant \epsilon). \tag{14}$$

It is worth note that M represents the initial layer of simulation for generating future samples to compute values of AAD and MSD, while D shows the subsequent layer of simulation for repeating the entire process to calculate $RP_1(\epsilon)$ and $RP_2(\epsilon)$ values. When the $RP_1(\epsilon)$ and $RP_2(\epsilon)$ are plotted against a range of ϵ values, it provides a visual summary to compare the reproducibility of different mean estimators. We call this ϵ-reproducibility in terms of AAD and MSD for mean estimators, where the quantity ϵ should be in a closed interval $[0, +\infty]$.

The algorithm 1 given below summarizes the steps involved in computing the ϵ-reproducibility of mean estimators under RSS using the PPB approach.

Algorithm 1 Computing the ϵ-reproducibility of mean estimators using PPB approach

Data: The original RSS sample of size m.
Result: $RP(\epsilon_1)$, $RP(\epsilon_2)$
Draw a bivariate original sample using procedures of RSS. Calculate original mean using mean estimators
for *each method of RSS* **do**
  Apply PPB to generate bootstrapped sets in RSS samples
  Replace bootstrapped set with original sets in RSS sample
  Draw PPB-RSS samples from bootstrapped data
  Compute sample mean using mean estimators based on PPB-RSS samples.
end
for $i \leftarrow 1$ **to** M **do**
  Reiterate Steps 1-8 and compute AAD and MSD using Equation (11) and (12), respectively.
end
for $j \leftarrow 1$ **to** D **do**
  Reiterate Steps 1-11 and compute $RP_1(\epsilon)$ and $RP_2(\epsilon)$ using Equation (13) and (14), respectively.
end

7. Simulations

This section presents the simulation procedure and its results for comparing the ϵ-reproducibility of mean estimators under RSS using the PPB approach. Using R software, algorithm 1 is used to compute the ϵ-reproducibility, such that, we first generate the population values for the concomitant variable (say, X_i). We also generate values for the standardized normal population Z_i, which aids us in establishing the correlation between the study variable Y and the auxiliary variable X. The values of the study variables are generated using the relationship $Y_i = \rho X_i + Z_i\sqrt{1-\rho^2}$, where ρ is the correlation coefficient between Y and X and its value is fixed as 0.90 in order to establish a strong association between Y and X. We compare ϵ-reproducibility considering two populations, that is, (1) normal population and (2) exponential population. We also compare ϵ-reproducibility, assuming perfect ranking and imperfect ranking. The ϵ-reproducibility is also compared for various set sizes, i.e., $m = 3, 5, 7$. In Eqs. (11) and (12), we use $M = 1000$ to compute AAD and MSD. This whole process is repeated $D = 1000$ times to compute the respective reproducibility probabilities $RP_1(\epsilon)$ and $RP_2(\epsilon)$. We use different colors on plots to show the ϵ−reproducibility of different mean estimators. Lines in red, black, green, and blue indicate the respective ϵ−reproducibility for conventional, regression, ratio, and exponential ratio estimators based on RSS. The x−axis shows the magnitude of the ϵ value, while the y−axis shows the probability of observing it. Results are given below.

Simulation results given in Figs. 1–8 show that the ϵ−reproducibility of the regression estimator increases earlier than the ϵ−reproducibility of any other mean estimator, indicating that regression is the highest reproducible estimator. Similarly, the conventional mean estimator shows the lowest ϵ-reproducibility as compared to other mean estimators. The ϵ-reproducibility of the ratio and exponential ratio estimators is quite similar and intermediate between the conventional and regression estimators. The ϵ-reproducibility of the exponential ratio estimator is higher in the case of an exponentially distributed population as compared to a normally distributed population. The plots also show that the ϵ-reproducibility of mean estimators is higher for larger sample sizes as compared to small sample sizes. Additionally, the ϵ-reproducibility is higher in the case of perfect ranking as compared to the case of imperfect ranking. Generally, the ϵ-reproducibility of the regression estimator is higher than other mean estimators, while the conventional estimator exhibits the lowest ϵ−reproducibility in all cases.

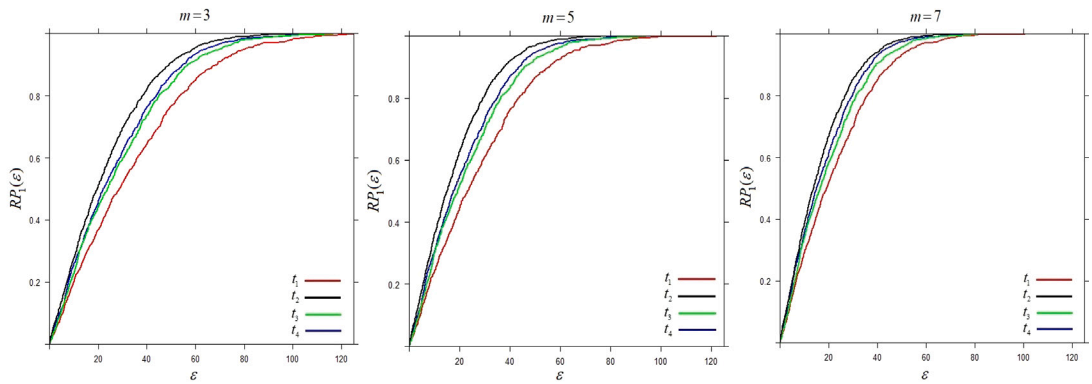

Fig. 1. $RP_1(\epsilon)$ of estimates for different sample size in case of normal distribution under perfect ranking. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

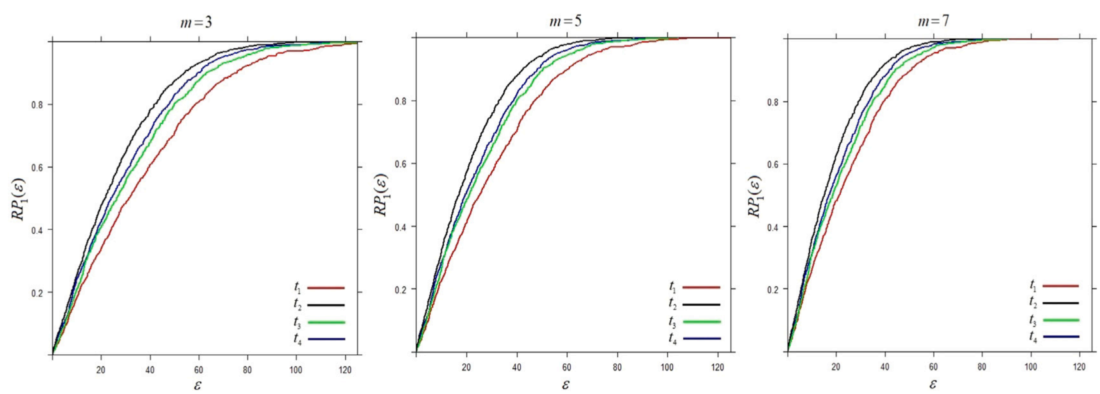

Fig. 2. $RP_1(\epsilon)$ of estimates for different sample size in case of normal distribution under imperfect ranking. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

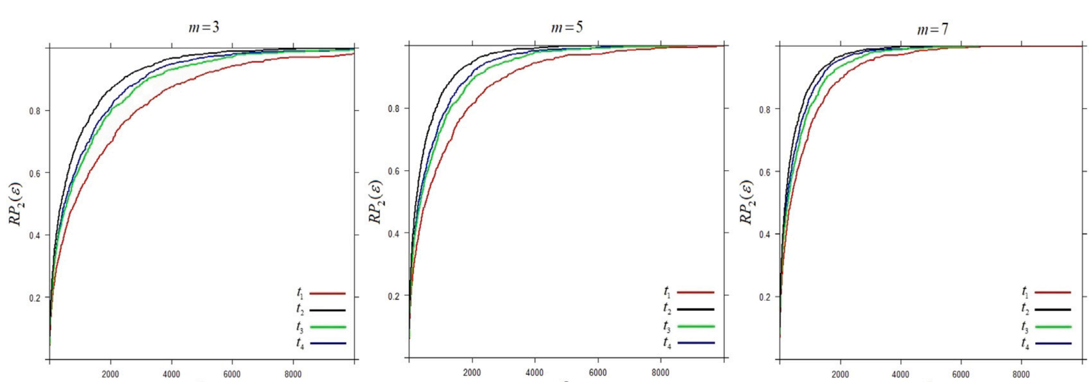

Fig. 3. $RP_2(\epsilon)$ of estimates for different sample size in case of normal distribution under perfect ranking. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

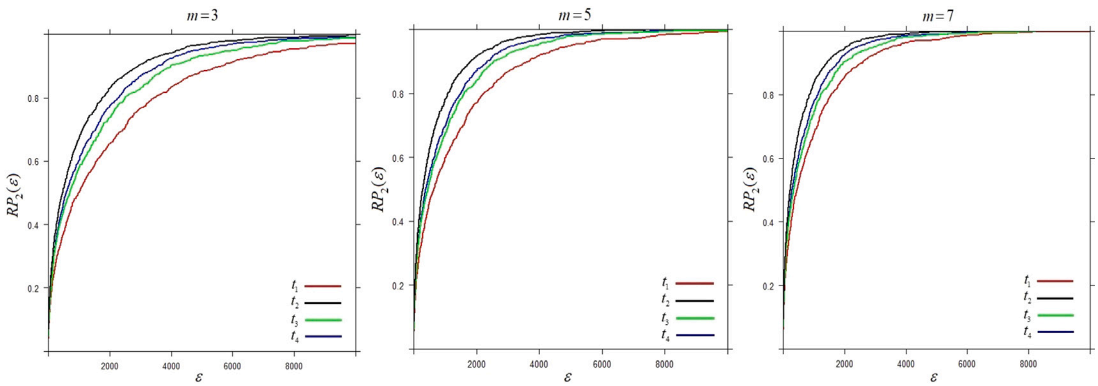

Fig. 4. $RP_2(\epsilon)$ of estimates for different sample size in case of normal distribution under imperfect ranking. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

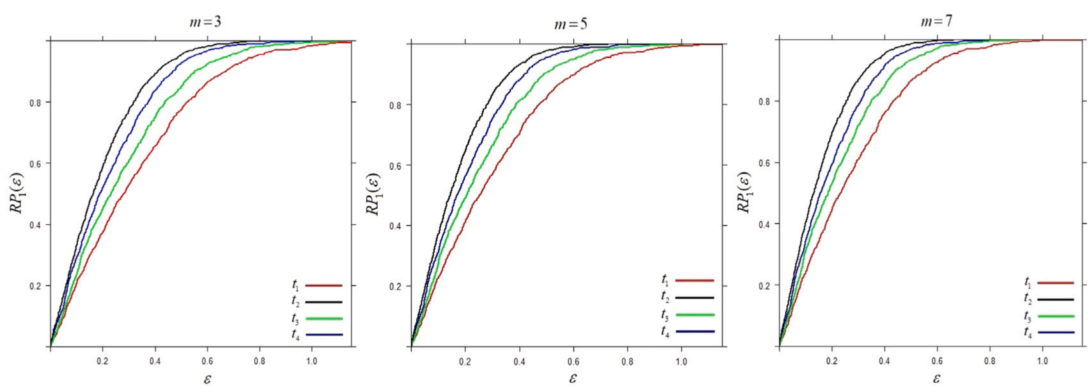

Fig. 5. $RP_1(\epsilon)$ of estimates for different sample size in case of exponential distribution under perfect ranking. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

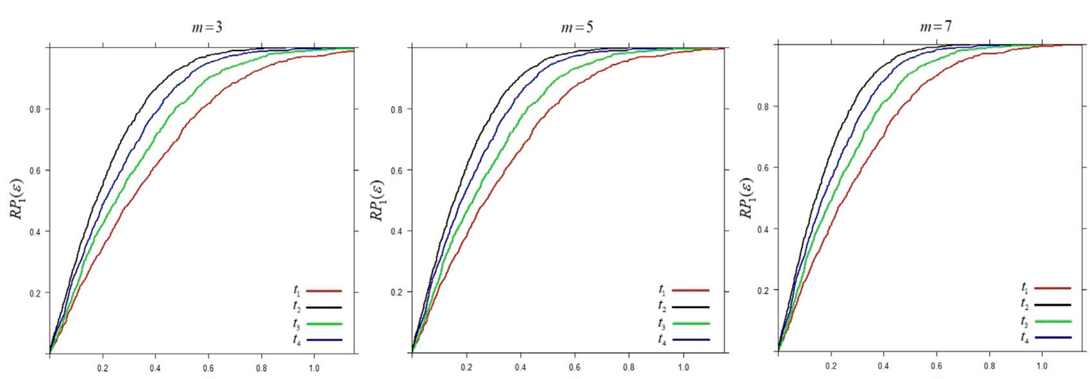

Fig. 6. $RP_1(\epsilon)$ of estimates for different sample size in case of exponential distribution under imperfect ranking. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

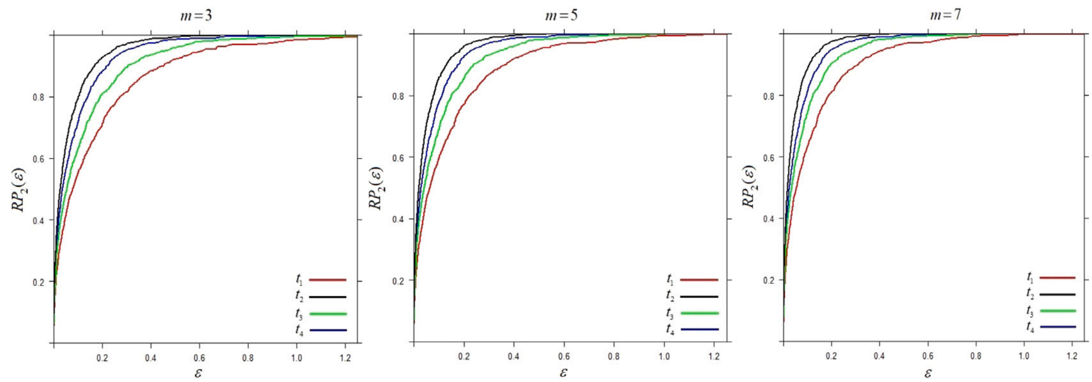

Fig. 7. $RP_2(\epsilon)$ of estimates for different sample size in case of exponential distribution under perfect ranking. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

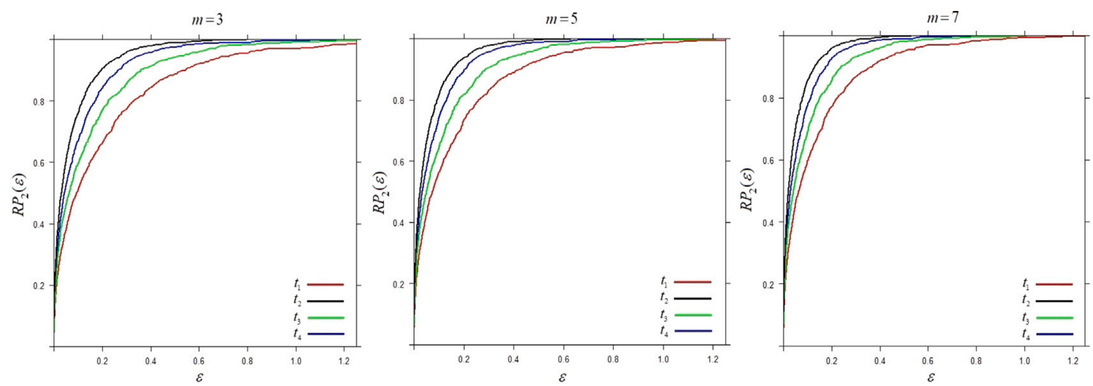

Fig. 8. $RP_2(\epsilon)$ of estimates for different sample size in case of exponential distribution under imperfect ranking. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

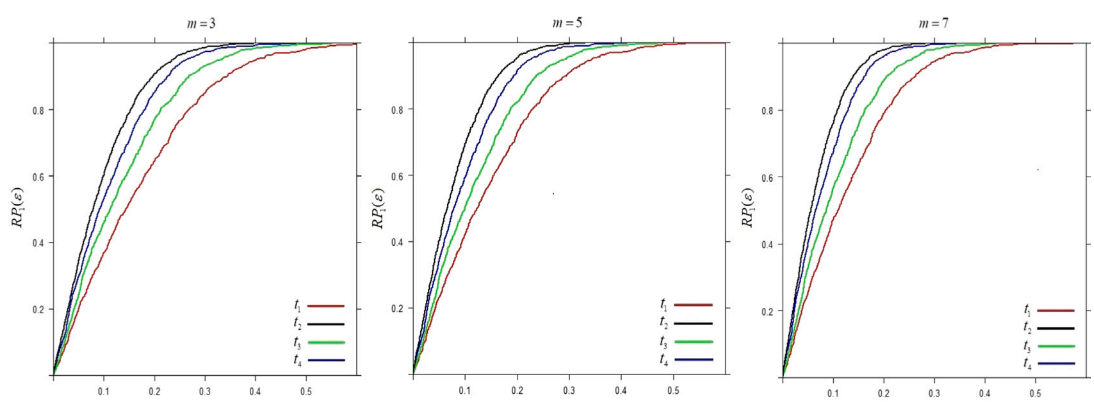

Fig. 9. $RP_1(\epsilon)$ of estimates under for different sample sizes under perfect ranking in case of real data.

8. Application to real-life data

In this section, we consider the dataset pertaining to the measurements of the physical characteristics of abalone, as initially collected by [32]. The study variable Y and the auxiliary variable X are taken as

Y = The whole weight of abalone (in grams)

X = The length of abalone (in mm), its the longest shell measurement.

We collect the original samples from this data and estimate the mean using the formulas of the four estimators discussed above. The remaining simulation procedure is the same as discussed in Algorithm 1, and the simulation results are presented in the graphs 9–12 given below.

9. Conclusions

Estimation of population parameters based on sample data is an important technique of statistical inference. Ranked set sampling (RSS)

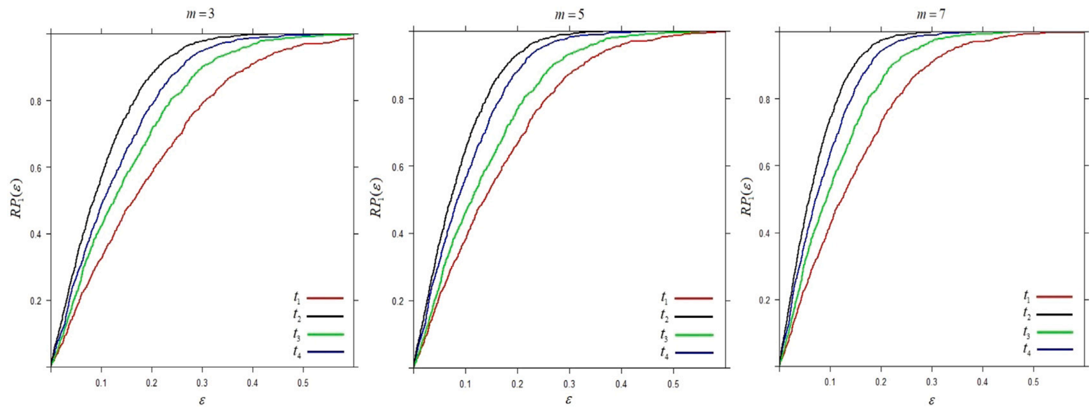

Fig. 10. $RP_1(\varepsilon)$ of estimates under for different sample sizes under imperfect ranking in case of real data.

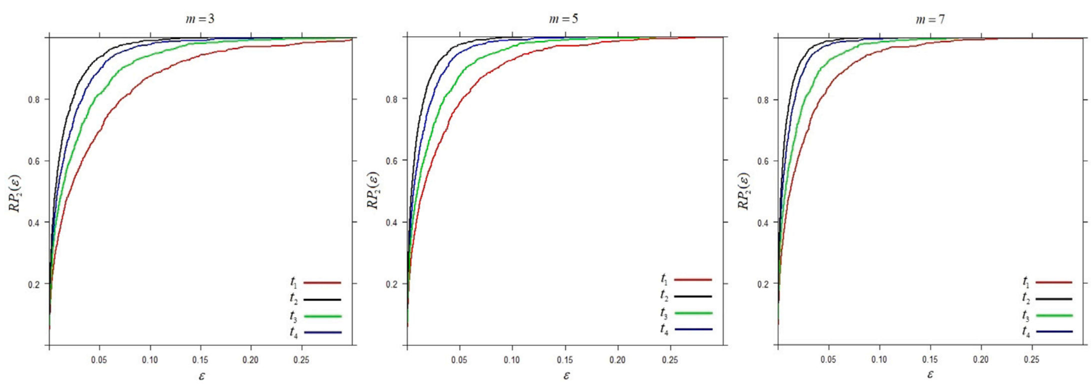

Fig. 11. $RP_2(\varepsilon)$ of estimates under for different sample sizes under perfect ranking in case of real data.

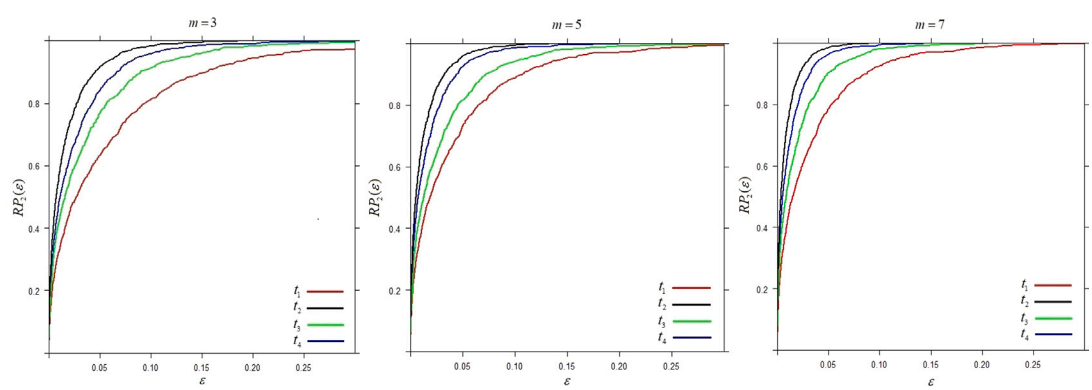

Fig. 12. $RP_2(\varepsilon)$ of estimates under for different sample sizes under imperfect ranking in case of real data.

is an efficient sampling method for collecting sample data. In this study, we investigated the reproducibility of four well-known mean estimators, called conventional, ratio, exponential ratio, and regression estimators under RSS, employing parametric predictive bootstrapping. Through simulation studies, we evaluated the reproducibility of these estimators for varying sample sizes in both perfect and imperfect ranking situations. Data on abalone was considered to provide real-life applications for this study. The findings indicated that the regression is the most reproducible estimator, whereas the conventional estimator is the least reproducible estimator.

Our methodology presented in this paper for computing and comparing the reproducibility of mean estimators can be extended to investigate and compare the reproducibility of other mean estimators in the existing literature. Additionally, our presented method can serve as a measure for evaluating newly proposed estimators and comparing them with other existing estimators. Moreover, this study can be expanded to examine and compare the reproducibility of estimators for variance and other parameters, providing valuable insights into statistical estimation.

Funding disclosure

This research did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

Animal welfare

No animals were involved in this study. Only data about animals was used in accordance with ethical guidelines.

Plagiarism and originality

The authors affirm that the manuscript is original and has not been published elsewhere, in whole or in part. Proper citation and attribution have been provided for all sources of information or ideas not their own.

CRediT authorship contribution statement

Syed Abdul Rehman: Writing – original draft, Visualization, Formal analysis. **Tahani Coolen-Maturi:** Methodology, Investigation. **Frank P.A. Coolen:** Supervision, Methodology, Conceptualization. **Javid Shabbir:** Writing – review & editing, Supervision, Methodology, Conceptualization.

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

The data used in this study is publicly available from the UCI Machine Learning Repository at the URL: https://archive.ics.uci.edu/dataset/1/abalone.

References

[1] G. McIntyre, A method for unbiased selective sampling, using ranked sets, Aust. J. Agric. Res. 3 (4) (1952) 385–390.
[2] K. Takahasi, K. Wakimoto, On unbiased estimates of the population mean based on the sample stratified by means of ordering, Ann. Inst. Statist. Math. 20 (1) (1968) 1–31.
[3] L. Stokes, Ranked set sampling with concomitant variables, Comm. Statist. Theory Methods 6 (12) (1977) 1207–1211.
[4] L. Philip, K. Lam, Regression estimator in ranked set sampling, Biometrics 53 (3) (1997) 1070–1080.
[5] C. Kadilar, Y. Unyazici, H. Cingi, Ratio estimator for the population mean using ranked set sampling, Statist. Papers 50 (2009) 301–309.
[6] G.K. Vishwakarma, S.M. Zeeshan, C.N. Bouza-Herrera, Ratio and product type exponential estimators for population mean using ranked set sampling, in: Ranked Set Sampling, Elsevier, 2019, pp. 251–257.
[7] E. Strzalkowska-Kominiak, M. Mahdizadeh, On the Kaplan–Meier estimator based on ranked set samples, J. Stat. Comput. Simul. 84 (12) (2014) 2577–2591.
[8] M. Mahdizadeh, E. Strzalkowska-Kominiak, Resampling based inference for a distribution function using censored ranked set samples, Comput. Statist. 32 (2017) 1285–1308.
[9] J. Frey, T.G. Feeman, Finding the maximum efficiency for multistage ranked-set sampling, Comm. Statist. Theory Methods 47 (17) (2018) 4131–4141.
[10] M. Mahdizadeh, On entropy-based test of exponentiality in ranked set sampling, Comm. Statist. Simulation Comput. 44 (4) (2015) 979–995.
[11] M. Mahdizadeh, N.R. Arghami, Quantile estimation using ranked set samples from a population with known mean, Comm. Statist. Simulation Comput. 41 (10) (2012) 1872–1881.
[12] J. Yin, Y. Hao, H. Samawi, H. Rochani, Rank-based kernel estimation of the area under the ROC curve, Stat. Methodol. 32 (2016) 91–106.
[13] S. Wang, W. Chen, R. Yang, Fisher information in ranked set sampling from the simple linear regression model, Comm. Statist. Simulation Comput. 53 (3) (2024) 1274–1284.
[14] S.N. Goodman, A comment on replication, p-values and evidence, Stat. Med. 11 (7) (1992) 875–879.
[15] S. Senn, Comment on 'a comment on replication, p-values and evidence', by SN Goodman (letter to the editor), Stat. Med. 21 (2002) 2437–2444.
[16] J. Miller, What is the probability of replicating a statistically significant effect? Psychon. Bull. Rev. 16 (2009) 617–640.
[17] E.J. Posavac, Using p values to estimate the probability of a statistically significant replication, in: Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences, Vol. 1, 2002, pp. 101–112, 2.
[18] J. Shao, S.-C. Chow, Reproducibility probability in clinical trials, Stat. Med. 21 (12) (2002) 1727–1742.
[19] P.R. Killeen, An alternative to null-hypothesis significance tests, Psychol. Sci. 16 (5) (2005) 345–353.
[20] D. De Martini, Reproducibility probability estimation for testing statistical hypotheses, Statist. Probab. Lett. 78 (9) (2008) 1056–1061.
[21] C. Begley, L. Ellis, Drug development: Raise standards for preclinical cancer research, Nature 483 (2012) 531–533.
[22] A. Aldawsari, Parametric Predictive Bootstrap and Test Reproducibility (Ph.D. thesis), Durham University, 2023.
[23] F.P.A. Coolen, S. BinHimd, Nonparametric predictive inference bootstrap with application to reproducibility of the two-sample Kolmogorov–Smirnov test, J. Stat. Theory Pract. 14 (2020) 1–13.
[24] F.P.A. Coolen, H.N. Alqifari, Nonparametric predictive inference for reproducibility of two basic tests based on order statistics, Revstat 16 (2) (2018) 167–185.
[25] F.P.A. Coolen, T. Coolen-Maturi, Statistical reproducibility, in: M. Lovric (Ed.), International Encyclopedia of Statistical Science, Springer, Heidelberg, 2024.
[26] H. Atmanspacher, S. Maasen, Reproducibility: Principles, Problems, Practices, and Prospects, John Wiley & Sons, 2016.
[27] F.P.A. Coolen, S. BinHimd, Nonparametric predictive inference for reproducibility of basic nonparametric tests, J. Stat. Theory Pract. 8 (2014) 591–618.
[28] L. De Capitani, D. Martini, On stochastic orderings of the Wilcoxon rank sum test statistic—with applications to reproducibility probability estimation testing, Statist. Probab. Lett. 81 (8) (2011) 937–946.
[29] L. De Capitani, D. Martini, Reproducibility probability estimation and testing for the Wilcoxon rank-sum test, J. Stat. Comput. Simul. 85 (3) (2015) 468–493.
[30] D. Billheimer, Predictive inference and scientific reproducibility, Amer. Statist. 73 (sup1) (2019) 291–295.
[31] B. Efron, Bootstrap methods: Another look at the Jackknife, Ann. Statist. 7 (1) (1979) 1–26.
[32] W. Nash, T. Sellers, S. Talbot, A. Cawthorn, W. Ford, Abalone, UCI Machine Learning Repository, 1995, http://dx.doi.org/10.24432/C55C7W.