Appl. Math. J. Chinese Univ.
2021, 36(2): 269-277

Parametric estimation for the simple linear regression model under moving extremes ranked set sampling design

YAO Dong-sen   CHEN Wang-xue*   LONG Chun-xian

Abstract. Cost effective sampling design is a major concern in some experiments especially when the measurement of the characteristic of interest is costly or painful or time consuming. Ranked set sampling (RSS) was first proposed by McIntyre [1952. A method for unbiased selective sampling, using ranked sets. Australian Journal of Agricultural Research 3, 385-390] as an effective way to estimate the pasture mean. In the current paper, a modification of ranked set sampling called moving extremes ranked set sampling (MERSS) is considered for the best linear unbiased estimators(BLUEs) for the simple linear regression model. The BLUEs for this model under MERSS are derived. The BLUEs under MERSS are shown to be markedly more efficient for normal data when compared with the BLUEs under simple random sampling.

§1 Introduction

The ranked set sampling(RSS) technique was introduced by McIntyre (1952). This technique is useful for cases when the variable of interest can be more easily ranked than quantified. This is the case in many environmental and ecological studies. For example, the ranking of hazardous waste sites with respect to their contamination levels can be made easier by a visual inspection of defoliation or soil discoloration, while making actual measurement of toxic chemicals and assessing their environmental impact may be very costly (Barabesi et al., 2001). For other applications of RSS in ecology see Halls et al. (1966) and Chen et al. (2004).

In RSS one first draws m^2 units at random from the population and partitions them into m sets of m units. The m units in each set are ranked without making actual measurements.

Received: 2019-12-17.   Revised: 2020-03-24.
MR Subject Classification: 62D05.
Keywords: simple linear regression model, best linear unbiased estimator, simple random sampling, ranked set sampling, moving extremes ranked set sampling.
Digital Object Identifier(DOI): https://doi.org/10.1007/s11766-021-3993-1.
Supported by the National Natural Science Foundation of China(11901236), the Scientific Research Fund of Hunan Provincial Science and Technology Department(2019JJ50479), the Scientific Research Fund of Hunan Provincial Education Department(18B322), the Winning Bid Project of Hunan Province for the 4th National Economic Census([2020]1), the Young Core Teacher Foundation of Hunan Province([2020]43) and the Fundamental Research Fund of Xiangxi Autonomous Prefecture(2018SF5026).
*Corresponding author.

From the first set of m units the unit ranked the lowest is chosen for actual measurements. From the second set of m units the unit ranked the second lowest is measured. The process is continued until the unit ranked the largest is measured from the m-th set of m units. We call this process a one cycle RSS of size m.

Takahasi et al. (1968) established a very important statistical foundation for the theory of RSS. They showed that the mean of the RSS is an unbiased estimator for the population mean and has a higher efficiency than the mean of simple random sampling (SRS). Barreto et al. (1999) studied best linear unbiased estimators (BLUEs) of parameters for the simple linear regression model under RSS and showed that the BLUEs are more efficient than that under simple random sampling for normal data. For more studies of RSS refer to Chen et al. (2017) and He et al. (2018). However, ranking accuracy affects the efficiency of the estimator. In order to reduce the error of ranking and keep optimality inherited in the original RSS procedure, Al-Odat et al. (2001) introduced the concept of varied set size RSS, which is coined here as Moving Extremes Ranked Set Sampling (MERSS).

The procedure of MERSS is described as follows:

(1) Select m simple random sample sets of size 1, 2,..., m, respectively.

(2) Order the elements of each set by visual inspection or other relatively inexpensive methods, without actual measurement of the characteristic of interest.

(3) Measure accurately the maximum ordered observation from the first set, then the second set,..., the last set.

(4) Step (3) is repeated on another m sets of size 1, 2,..., m respectively, however the minimum ordered observations are measured instead of the maximum ordered observations. We call this process a one cycle MERSS of size $2m$.

In the literature, there are numbers of studies focused on parametric inference of distributions under MERSS. Chen et al. (2019) studied BLUEs of the parameters for the Pareto distribution under MERSS. He et al. (2019) studied BLUEs of the parameters for the log-logistic distribution under MERSS.

In this paper, we are interested in studying the BLUEs for the simple linear regression under MERSS. The BLUEs for this model under MERSS are derived. The BLUEs under MERSS are shown to be markedly more efficient for normal data when compared with the BLUEs under SRS.

§2 Moving extremes ranked set sample

According to MERSS procedure which is described in Section 1. Let $\{y_{i1}^1, y_{i2}^1, ..., y_{ii}^1\}$ and $\{y_{i1}^2, y_{i2}^2, ..., y_{ii}^2\}$ be $2m$ sets of random samples, $i = 1, 2, ..., m$. Let $x_{ii} = \max\{y_{i1}^1, y_{i2}^1, ..., y_{ii}^1\}$ and $y_{1i} = \min\{y_{i1}^2, y_{i2}^2, ..., y_{ii}^2\}$. Thus the moving extremes ranked set sample with a one cycle is defined as $\{x_{11}, x_{22}, ..., x_{mm}, y_{11}, y_{12}, ..., y_{1m}\}$. Note that the elements of this sample are independent of each other. x_{ii} is distributed as the i-th order statistic from a simple random sample of size i. y_{1i} is distributed as the 1-th order statistic from a simple random sample of size i.

§3 BLUEs for the simple linear regression model under MERSS

Let $\{x_{11j}, x_{22j}, ..., x_{mmj}, y_{11j}, y_{12j}, ..., y_{1mj}\}$ be moving extremes ranked set sample from the population Y for each value of the predictor variable $X = x_j$, j=1,2,...,n. The conditional mean and variance of Y are respectively

$$E(Y\,|X = x\,) = \alpha + \beta x$$

and

$$var(Y\,|X\,) = \sigma^2.$$

The reduced order statistics at each value of X, take the form

$$u_{iij} = \frac{x_{iij} - \alpha - \beta x_j}{\sigma}$$

with means η_{ii} and variances τ_{ii} and

$$u_{1ij} = \frac{y_{1ij} - \alpha - \beta x_j}{\sigma}$$

with means η_{1i} and variances τ_{1i}, $i = 1, 2, ..., m$, when Y is a continuous variable with distribution function

$$F\left(\frac{y - \mu}{\sigma}\right).$$

Then we have

$$E\left(x_{iij}\right) = \alpha + \beta x_j + \sigma\eta_{ii} \tag{3.1}$$

with

$$var(x_{iij}) = \sigma^2\tau_{ii} \tag{3.2}$$

and

$$E\left(y_{1ij}\right) = \alpha + \beta x_j + \sigma\eta_{1i} \tag{3.3}$$

with

$$var(y_{1ij}) = \sigma^2\tau_{1i}. \tag{3.4}$$

A matrix form for the model is

$$\boldsymbol{Y} = \boldsymbol{X\theta} + \boldsymbol{\varepsilon}, \tag{3.5}$$

where

$$\mathbf{Y} = \begin{bmatrix} x_{111} \; ... \; x_{mm1} \; y_{111} \; ... \; y_{1m1} & ... & x_{11n} \; ... \; x_{mmn} \;\; y_{11n} \; ... \; y_{1mn} \end{bmatrix}',$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & ... & 1 & 1 & ... & 1... & 1 & ... & 1 & 1 & ... & 1 \\ x_1 & ... & x_1 & x_1 & ... & x_1... & x_n & ... & x_n & x_n & ... & x_n \\ \eta_{11} & ... & \eta_{mm} & \eta_{11} & ... & \eta_{1m}... & \eta_{11} & ... & \eta_{mm} & \eta_{11} & ... & \eta_{1m} \end{bmatrix}',$$

$$\boldsymbol{\theta} = \begin{bmatrix} \alpha & \beta & \sigma \end{bmatrix}'$$

and $\boldsymbol{\varepsilon}$ is the random error vector with $E\left(\boldsymbol{\varepsilon}\right) = \mathbf{0}$ and

$$var(\boldsymbol{\varepsilon}) = \sigma^2 diag\left\{\tau_{11}, ..., \tau_{mm}, \tau_{11}, ..., \tau_{1m}, ..., \tau_{11}, ..., \tau_{mm}, \tau_{11}, ..., \tau_{1m}\right\}.$$

Using the same argument of Lloyd (1952) shown above, the BLUE of $\boldsymbol{\theta} = \begin{bmatrix} \alpha & \beta & \sigma \end{bmatrix}'$ can be obtained

$$\widehat{\boldsymbol{\theta}}_{BLUE,\ MERSS} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \tag{3.6}$$

with

$$var(\widehat{\boldsymbol{\theta}}_{BLUE,\ MERSS}) = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\sigma^2, \tag{3.7}$$

where $\boldsymbol{V} = diag\left\{\tau_{11}, ..., \tau_{mm}, \tau_{11}, ..., \tau_{1m}, ..., \tau_{11}, ..., \tau_{mm}, \tau_{11}, ..., \tau_{1m}\right\}$,

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} = \begin{bmatrix} n\sum\limits_{i=1}^{m}\left(\tau_{ii}^{-1}+\tau_{1i}^{-1}\right) & \sum\limits_{j=1}^{n}x_j\sum\limits_{i=1}^{m}\left(\tau_{ii}^{-1}+\tau_{1i}^{-1}\right) & n\sum\limits_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}+\tau_{1i}^{-1}\eta_{1i}\right) \\ \sum\limits_{j=1}^{n}x_j\sum\limits_{i=1}^{m}\left(\tau_{ii}^{-1}+\tau_{1i}^{-1}\right) & \sum\limits_{j=1}^{n}x_j^2\sum\limits_{i=1}^{m}\left(\tau_{ii}^{-1}+\tau_{1i}^{-1}\right) & \sum\limits_{j=1}^{n}x_j\sum\limits_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}+\tau_{1i}^{-1}\eta_{1i}\right) \\ n\sum\limits_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}+\tau_{1i}^{-1}\eta_{1i}\right) & \sum\limits_{j=1}^{n}x_j\sum\limits_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}+\tau_{1i}^{-1}\eta_{1i}\right) & n\sum\limits_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}^2+\tau_{1i}^{-1}\eta_{1i}^2\right) \end{bmatrix}$$

and

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} = \begin{bmatrix} \sum\limits_{j=1}^{n}\sum\limits_{i=1}^{m}\left(\tau_{ii}^{-1}x_{iij}+\tau_{1i}^{-1}y_{1ij}\right) \\ \sum\limits_{j=1}^{n}x_j\sum\limits_{i=1}^{m}\left(\tau_{ii}^{-1}x_{iij}+\tau_{1i}^{-1}y_{1ij}\right) \\ \sum\limits_{j=1}^{n}\sum\limits_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}x_{iij}+\tau_{1i}^{-1}\eta_{1i}y_{1ij}\right) \end{bmatrix}.$$

After some calculations and simplifications, we can obtain

$$\begin{aligned}\widehat{\alpha}_{BLUE,\ MERSS} = \frac{1}{\Delta}\Bigg[&a_{11}\sum_{j=1}^{n}\sum_{i=1}^{m}\left(\tau_{ii}^{-1}x_{iij}+\tau_{1i}^{-1}y_{1ij}\right) + a_{12}\sum_{j=1}^{n}x_j\sum_{i=1}^{m}\left(\tau_{ii}^{-1}x_{iij}+\tau_{1i}^{-1}y_{1ij}\right) \\ &+a_{13}\sum_{j=1}^{n}\sum_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}x_{iij}+\tau_{1i}^{-1}\eta_{1i}y_{1ij}\right)\Bigg],\end{aligned} \tag{3.8}$$

$$\widehat{\beta}_{BLUE,\ MERSS} = \frac{1}{\Delta}\left[a_{12}\sum_{j=1}^{n}\sum_{i=1}^{m}\left(\tau_{ii}^{-1}x_{iij}+\tau_{1i}^{-1}y_{1ij}\right) + a_{22}\sum_{j=1}^{n}x_j\sum_{i=1}^{m}\left(\tau_{ii}^{-1}x_{iij}+\tau_{1i}^{-1}y_{1ij}\right)\right], \tag{3.9}$$

$$\widehat{\sigma}_{BLUE,\ MERSS} = \frac{1}{\Delta}\left[a_{13}\sum_{j=1}^{n}\sum_{i=1}^{m}\left(\tau_{ii}^{-1}x_{iij}+\tau_{1i}^{-1}y_{1ij}\right) + a_{33}\sum_{j=1}^{n}\sum_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}x_{iij}+\tau_{1i}^{-1}\eta_{1i}y_{1ij}\right)\right], \tag{3.10}$$

where

$$a_{11} = n\sum_{j=1}^{n}x_j^2\sum_{i=1}^{m}\left(\tau_{ii}^{-1}+\tau_{1i}^{-1}\right)\sum_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}^2+\tau_{1i}^{-1}\eta_{1i}^2\right) - \left(\sum_{j=1}^{n}x_j\right)^2\left[\sum_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}+\tau_{1i}^{-1}\eta_{1i}\right)\right]^2,$$

$$a_{12} = n\sum_{j=1}^{n}x_j\left[\sum_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}+\tau_{1i}^{-1}\eta_{1i}\right)\sum_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}+\tau_{1i}^{-1}\eta_{1i}\right) - \sum_{i=1}^{m}\left(\tau_{ii}^{-1}+\tau_{1i}^{-1}\right)\sum_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}^2+\tau_{1i}^{-1}\eta_{1i}^2\right)\right],$$

$$a_{13} = \sum_{i=1}^{m}\left(\tau_{ii}^{-1}+\tau_{1i}^{-1}\right)\sum_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}+\tau_{1i}^{-1}\eta_{1i}\right)\left[\left(\sum_{j=1}^{n}x_j\right)^2 - n\sum_{j=1}^{n}x_j^2\right],$$

$$a_{22} = n^2\left\{\sum_{i=1}^{m}\left(\tau_{ii}^{-1}+\tau_{1i}^{-1}\right)\sum_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}^2+\tau_{1i}^{-1}\eta_{1i}^2\right) - \left[\sum_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}+\tau_{1i}^{-1}\eta_{1i}\right)\right]^2\right\},$$

$$a_{33}=\left[\sum_{i=1}^{m}\left(\tau_{ii}^{-1}+\tau_{1i}^{-1}\right)\right]^{2}\left[n\sum_{j=1}^{n}x_{j}^{2}-\left(\sum_{j=1}^{n}x_{j}\right)^{2}\right]$$

and

$$\begin{aligned}\Delta=&\,n^{2}\left(\sum_{j=1}^{n}x_{j}^{2}\right)\left[\sum_{i=1}^{m}\left(\tau_{ii}^{-1}+\tau_{1i}^{-1}\right)\right]^{2}\sum_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}^{2}+\tau_{1i}^{-1}\eta_{1i}^{2}\right)+2n\left(\sum_{j=1}^{n}x_{j}\right)^{2}\left[\sum_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}+\tau_{1i}^{-1}\eta_{1i}\right)\right]^{2}\\&\sum_{i=1}^{m}\left(\tau_{ii}^{-1}+\tau_{1i}^{-1}\right)-n^{2}\left(\sum_{j=1}^{n}x_{j}^{2}\right)\left[\sum_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}+\tau_{1i}^{-1}\eta_{1i}\right)\right]^{2}\sum_{i=1}^{m}\left(\tau_{ii}^{-1}+\tau_{1i}^{-1}\right)\\&-n\left(\sum_{j=1}^{n}x_{j}\right)^{2}\left[\sum_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}+\tau_{1i}^{-1}\eta_{1i}\right)\right]^{2}\sum_{i=1}^{m}\left(\tau_{ii}^{-1}+\tau_{1i}^{-1}\right)\\&-n\left(\sum_{j=1}^{n}x_{j}\right)^{2}\sum_{i=1}^{m}\left(\tau_{ii}^{-1}\eta_{ii}^{2}+\tau_{1i}^{-1}\eta_{1i}^{2}\right)\left[\sum_{i=1}^{m}\left(\tau_{ii}^{-1}+\tau_{1i}^{-1}\right)\right]^{2}.\end{aligned}$$

The variances and covariances of the estimators are

$$var(\widehat{\alpha}_{BLUE,\ MERSS})=\sigma^{2}\frac{a_{11}}{\Delta},\tag{3.11}$$

$$var(\widehat{\beta}_{BLUE,\ MERSS})=\sigma^{2}\frac{a_{22}}{\Delta}\tag{3.12}$$

and

$$var(\widehat{\sigma}_{BLUE,\ MERSS})=\sigma^{2}\frac{a_{33}}{\Delta}\tag{3.13}$$

and $cov(\widehat{\alpha}_{BLUE,\ MERSS},\ \widehat{\beta}_{BLUE,\ MERSS})=\sigma^{2}\frac{a_{12}}{\Delta}$, $cov(\widehat{\alpha}_{BLUE,\ MERSS},\ \widehat{\sigma}_{BLUE,\ MERSS})=\sigma^{2}\frac{a_{13}}{\Delta}$ and $cov(\widehat{\beta}_{BLUE,\ MERSS},\ \widehat{\sigma}_{BLUE,\ MERSS})=0$.

In the special case where the continuous random variable is symmetric about zero, then from David (1981) and Balakrishnan et al.(1991) we have

$$\eta_{ii}=-\eta_{1i}\tag{3.14}$$

and

$$\tau_{ii}=\tau_{1i}.\tag{3.15}$$

Use the formulas (3.14) and (3.15) for (3.8), (3.9) and (3.10), then we can obtain

$$\widehat{\alpha}_{BLUE,\ MERSS}=\frac{1}{\Delta^{*}}\left[a_{11}^{*}\sum_{j=1}^{n}\sum_{i=1}^{m}\tau_{ii}^{-1}\left(x_{iij}+y_{1ij}\right)+a_{12}^{*}\sum_{j=1}^{n}x_{j}\sum_{i=1}^{m}\tau_{ii}^{-1}\left(x_{iij}+y_{1ij}\right)\right],\tag{3.16}$$

$$\widehat{\beta}_{BLUE,\ MERSS}=\frac{1}{\Delta^{*}}\left[a_{12}^{*}\sum_{j=1}^{n}\sum_{i=1}^{m}\tau_{ii}^{-1}\left(x_{iij}+y_{1ij}\right)+a_{22}^{*}\sum_{j=1}^{n}x_{j}\sum_{i=1}^{m}\tau_{ii}^{-1}\left(x_{iij}+y_{1ij}\right)\right],\tag{3.17}$$

$$\widehat{\sigma}_{BLUE,\ MERSS}=\frac{a_{33}^{*}}{\Delta^{*}}\sum_{j=1}^{n}\sum_{i=1}^{m}\tau_{ii}^{-1}\eta_{ii}\left(x_{iij}-y_{1ij}\right),\tag{3.18}$$

where

$a_{11}^{*}=4n\sum_{j=1}^{n}x_{j}^{2}\sum_{i=1}^{m}\tau_{ii}^{-1}\sum_{i=1}^{m}\tau_{ii}^{-1}\eta_{ii}^{2}$,

$a_{12}^{*}=-4n\sum\limits_{j=1}^{n}x_j\sum\limits_{i=1}^{m}\tau_{ii}^{-1}\sum\limits_{i=1}^{m}\tau_{ii}^{-1}\eta_{ii}^{2},$

$a_{22}^{*}=4n^2\sum\limits_{i=1}^{m}\tau_{ii}^{-1}\sum\limits_{i=1}^{m}\tau_{ii}^{-1}\eta_{ii}^{2},$

$a_{33}^{*}=4\left(\sum\limits_{i=1}^{m}\tau_{ii}^{-1}\right)^2\left[n\sum\limits_{j=1}^{n}x_j^2-\left(\sum\limits_{j=1}^{n}x_j\right)^2\right]$

and

$\Delta^{*}=8n\sum\limits_{i=1}^{m}\tau_{ii}^{-1}\eta_{ii}^{2}\left(\sum\limits_{i=1}^{m}\tau_{ii}^{-1}\right)^2\left[n\sum\limits_{j=1}^{n}x_j^2-\left(\sum\limits_{j=1}^{n}x_j\right)^2\right].$

The variances and covariances of the estimators are

$$var(\widehat{\alpha}_{BLUE,\ MERSS})=\sigma^2\frac{a_{11}^{*}}{\Delta^{*}}, \tag{3.19}$$

$$var(\widehat{\beta}_{BLUE,\ MERSS})=\sigma^2\frac{a_{22}^{*}}{\Delta^{*}} \tag{3.20}$$

and

$$var(\widehat{\sigma}_{BLUE,\ MERSS})=\sigma^2\frac{a_{33}^{*}}{\Delta^{*}} \tag{3.21}$$

and $cov(\widehat{\alpha}_{BLUE,\ MERSS},\ \widehat{\beta}_{BLUE,\ MERSS})=\sigma^2\frac{a_{12}^{*}}{\Delta^{*}}$ and

$cov(\widehat{\alpha}_{BLUE,\ MERSS},\ \widehat{\sigma}_{BLUE,\ MERSS})=cov(\widehat{\beta}_{BLUE,\ MERSS},\ \widehat{\sigma}_{BLUE,\ MERSS})=0.$

§4 The relative efficiency of the BLUEs

Let $\{y_{1j},y_{2j},...,y_{mj},...,y_{2mj}\}$ $(j=1,2,...,n)$ be the simple random sample from Y for each value of the predictor variable $X=x_j$. Then BLUEs of parameters (α,β,σ) are respectively obtained as below

$$\widehat{\alpha}_{BLUE,\ SRS}=\bar{y}-\widehat{\beta}_{BLUE,\ SRS}\bar{x}, \tag{4.1}$$

$$\widehat{\beta}_{BLUE,\ SRS}=\frac{\sum\limits_{i=1}^{2m}\sum\limits_{j=1}^{n}\left(y_{ij}-\bar{y}\right)\left(x_j-\bar{x}\right)}{\frac{2m}{n}\left[n\sum\limits_{j=1}^{n}x_j^2-\left(\sum\limits_{j=1}^{n}x_j\right)^2\right]} \tag{4.2}$$

and

$$\widehat{\sigma^2}_{BLUE,\ SRS}=\frac{1}{2mn-2}\sum_{i=1}^{2m}\sum_{j=1}^{n}\left(y_{ij}-\widehat{\alpha}_{BLUE,\ SRS}-\widehat{\beta}_{BLUE,\ SRS}x_j\right)^2. \tag{4.3}$$

We have for normal error structure

$$var(\widehat{\alpha}_{BLUE,\ SRS})=\frac{\sigma^2\sum\limits_{j=1}^{n}x_j^2}{2m\left[n\sum\limits_{j=1}^{n}x_j^2-\left(\sum\limits_{j=1}^{n}x_j\right)^2\right]}, \tag{4.4}$$

$$var(\widehat{\beta}_{BLUE,\ SRS}) = \frac{\sigma^2}{\frac{2m}{n}\left[n\sum_{j=1}^{n} x_j^2 - \left(\sum_{j=1}^{n} x_j\right)^2\right]}, \tag{4.5}$$

$$var(\widehat{\sigma^2}_{BLUE,\ SRS}) = \frac{\sigma^4}{mn-1} \tag{4.6}$$

and $cov(\widehat{\alpha}_{BLUE,\ SRS},\ \widehat{\beta}_{BLUE,\ SRS}) = -\frac{\sigma^2\sum_{j=1}^{n} x_j}{2m\left[n\sum_{j=1}^{n} x_j^2 - \left(\sum_{j=1}^{n} x_j\right)^2\right]}$ and

$cov(\widehat{\alpha}_{BLUE,\ SRS},\ \widehat{\sigma^2}_{BLUE,\ SRS}) = cov(\widehat{\beta}_{BLUE,\ SRS},\ \widehat{\sigma^2}_{BLUE,\ SRS}) = 0.$

Combining (3.19), (3.20), (4.4) with (4.5), we can respectively obtain the efficiencies $\widehat{\alpha}_{BLUE,\ MERSS}$ with respect to (w.r.t.) $\widehat{\alpha}_{BLUE,\ SRS}$ and $\widehat{\beta}_{BLUE,\ MERSS}$ w.r.t. $\widehat{\beta}_{BLUE,\ SRS}$

$$eff(\widehat{\alpha}_{BLUE,\ SRS},\ \widehat{\alpha}_{BLUE,\ MERSS}) = \frac{var(\widehat{\alpha}_{BLUE,\ SRS})}{var(\widehat{\alpha}_{BLUE,\ MERSS})} = \frac{\sum_{i=1}^{m}\tau_{ii}^{-1}}{m} \tag{4.7}$$

and

$$eff(\widehat{\beta}_{BLUE,\ SRS},\ \widehat{\beta}_{BLUE,\ MERSS}) = \frac{var(\widehat{\beta}_{BLUE,\ SRS})}{var(\widehat{\beta}_{BLUE,\ MERSS})} = \frac{\sum_{i=1}^{m}\tau_{ii}^{-1}}{m}. \tag{4.8}$$

To calculate the efficiency of $\widehat{\sigma^2}_{BLUE,\ MERSS}$ w.r.t. $\widehat{\sigma^2}_{BLUE,\ SRS}$, we use the well known result on the approximate form of the sampling variance of a random sample standard deviation (see Kendall et al., 1979, p. 46). This implies

$$var(\widehat{\sigma}_{BLUE,\ SRS}) = \frac{\sigma^2}{4\,(mn-1)}. \tag{4.9}$$

Combining (3.21) with (4.9), we can obtain the efficiency $\widehat{\sigma}_{BLUE,\ MERSS}$ w.r.t. $\widehat{\sigma}_{BLUE,\ SRS}$

$$eff(\widehat{\sigma}_{BLUE,\ SRS},\ \widehat{\sigma}_{BLUE,\ MERSS}) = \frac{n}{2(mn-1)}\sum_{i=1}^{m}\eta_{ii}^2\tau_{ii}^{-1}. \tag{4.10}$$

The following simulation result is calculated after 10000 times sampling.

Table 1. Efficiency of $\widehat{\alpha}_{BLUE,\ MERSS}$ w.r.t. $\widehat{\alpha}_{BLUE,\ SRS}$ and $\widehat{\beta}_{BLUE,\ MERSS}$ w.r.t. $\widehat{\beta}_{BLUE,\ SRS}$.

m	efficiency	m	efficiency
2	1.2448	11	2.2557
3	1.4513	12	2.3144
4	1.5614	13	2.4004
5	1.7108	14	2.4657
6	1.8194	15	2.5135
7	1.9265	16	2.5657
8	2.0049	17	2.6211
9	2.0957	18	2.6693
10	2.1927	19	2.7287

Table 2. Efficiency of $\widehat{\sigma}_{BLUE,\ MERSS}$ w.r.t. $\widehat{\sigma}_{BLUE,\ SRS}$.

m	n=5	n=7	n=10
2	0.1298	0.1303	0.1255
3	0.3140	0.3110	0.3036
4	0.5174	0.5037	0.4989
5	0.7207	0.7153	0.7024
6	0.9279	0.9245	0.9206
7	1.1364	1.1306	1.1174
8	1.3258	1.3281	1.3236
9	1.5533	1.5168	1.5317
10	1.7263	1.7097	1.7113
11	1.9148	1.9143	1.8990
12	2.1152	2.0884	2.0888
13	2.2790	2.2800	2.2624
14	2.4515	2.4338	2.4359
15	2.6381	2.6104	2.6241
16	2.8107	2.7727	2.7609
17	2.9504	2.9583	2.9447
18	3.1322	3.1238	3.1173
19	3.2974	3.2524	3.2414

From Table 1 and Table 2, we may be conclude the following:

(1) These efficiencies increase with the increase of m.

(2) $\widehat{\alpha}_{BLUE,\ MERSS}$ is more efficient than $\widehat{\alpha}_{BLUE,\ SRS}$ for $m > 2$.

(3) $\widehat{\beta}_{BLUE,\ MERSS}$ is more efficient than $\widehat{\beta}_{BLUE,\ SRS}$ for $m > 2$.

(4) $\widehat{\sigma}_{BLUE,\ MERSS}$ is more efficient than $\widehat{\sigma}_{BLUE,\ SRS}$ for $m > 7$.

(5) In conclusion, the BLUEs of α and β for the simple linear regression model under MERSS are more efficient than that under SRS for $m > 2$.

(6) In conclusion, the BLUE of σ for the simple linear regression model under MERSS is more efficient than that under SRS for $m > 7$.

(7) In conclusion, the BLUEs for the simple linear regression model under MERSS are more efficient that under SRS for moderate m.

§5 Conclusions

RSS has previously proven advantageous for estimating population parameters compared with nonparametric, maximum likelihood, and least squares estimation. In this paper, we have obtained the BLUEs in the classes of linear combinations of the moving extremes ranked set sample values for the simple linear regression models with replicated observations. The BLUEs under MERSS are shown to be markedly more efficient for normal data when compared with the BLUEs under simple random sampling. The estimation of parameters is the first step in the use of MERSS for the simple linear regression models. A further stage would be to extend the use of MERSS methods to multiple and multivariate regression models.

References

[1] M T Al-Odat, M F Al-Saleh. *A variation of ranked set sampling*, Journal of Applied Statistical Science, 2001, 10(2): 137-146.

[2] N Balakrishnan, C Cohen. *Order Statistics and Inference: Estimation Methods*, Academic Press, San Diego, 1991.

[3] L Barabesi, A El-Sharaawi. *The efficiency of ranked set sampling for parameter estimation*, Statistics and Probability Letters, 2001, 53(2): 189-199.

[4] M C M Barreto, V Barnett. *Best linear unbiased estimators for the simple linear regression model using ranked set sampling*, Environmental and Ecological Statistics, 1999, 6(2): 119-133.

[5] W X Chen, M Y Xie, M Wu. *Maximum likelihood estimator of the parameter for a continuous one parameter exponential family under the optimal ranked set sampling*, Journal of Systems Science and Complexity, 2017, 30(6): 1350-1363.

[6] W X Chen, R Yang, D S Yao, C X Long. *Pareto parameters estimation using moving extremes ranked set sampling*, Statistical Papers, 2019, https://doi.org/10.1007/s00362-019-01132-9.

[7] Z H Chen, Z D Bai, B K Sinha. *Ranked set sampling, theory and applications*, Lecture Notes in Statistics, Springer, New York, 2004.

[8] H A David. *Order Statistics*, 2nd ed., John Wiley, New York, 1981.

[9] L K Halls, T R Dell. *Trial of ranked set sampling for forage yields*, Forest Science, 1966, 12(1): 22-32.

[10] X F He, W X Chen, W S Qian. *Maximum likelihood estimators of the parameters of the log-logistic distribution*, Statistical Papers, 2020, 61(5): 1875-1892.

[11] X F He, W X Chen, R Yang. *Log-logistic parameters estimation using moving extremes ranked set sampling design*, Applied Mathematics-A Journal of Chinese Universities (Series B), 2021, 36(1): 99-113.

[12] M G Kendall, A Stuart. *The Advanced Theory of Statistics, Volume 1: Inference and Relationship*, 4th ed, Griffin, London, 1979.

[13] E H Lloyd. *Generalized least-squares theorem, In Contributions to order statistics*, A E Sarhan, B G Greenberg (eds), (1992), John Wiley, New York, pp: 20-7, 1952.

[14] G A McIntyre. *A method of unbiased selective sampling, using ranked sets*, Australian Journal of Agricultural Research, 1952, 3(4): 385-390.

[15] K Takahasi, K Wakimoto. *On unbiased estimates of the population mean based on the sample stratified by means of ordering*, Annals of the Institute of Statistical Mathematics, 1968, 20(1): 1-31.

Department of Mathematics and Statistics , Jishou University, Jishou 416000, China.

Email: chenwangxue2017@163.com