Journal of Statistical Research of Iran
Vol. 1 No. 2 pp. 117-119

مجله‌ی پژوهش‌های آماری ایران
دوره‌ی ۱، شماره‌ی ۲، پاییز و زمستان ۱۳۸۳، صص ۱۹۰-۱۹۲

Row and Column Elimination Sampling Design +1 and its Efficiencies

M. Salehi M and M. A. Jamalzadeh*

Extended Abatract. It is a traditional way in biological, sociological, agricultural and geological studies to partition a geographical area into quadrats and then take a sample of them by a particular sampling design. We study the relevant characteristic of quadrats to estimate a parameter of the population. We suppose that the variable of interest has a positive spatial autocorrelation. Sampling designs which produce an appropriate coverage of the population will increase the precision of the parameter estimator, (Schreuder et al, 1993).
Hájek (1959), under a model with a positive spatial autocorrelation, illustrated that the systematic sampling is an optimum sampling design for one dimensional population. However, systematic and stratified samplings with only one sample in each stratum, are two traditional sampling designs that cover the population region well (McKenzie et al, 1991). Unfortunately, there is no unbiased estimator for these two sampling designs. Simple Latin Square Sampling (SLSS) design is another design which provides a good coverage for population. Also, this design has no variance estimator and it is considered as a weak point in practice. Munholland and Borkowski, (1995) introduced Simple Latin Square Sampling +1 (SLSS+1). They suggest that taking one additional sampling unit helps to provide an unbiased variance estimator. However, two other problems still exist concerning SLSS and SLSS+1. The population has to be a square and also the sample size be restricted to square root of the population size.

Salehi (2002) introduced Systematic Simple Latin Square Sampling (SSLSS) for which the population shape is a polyhedron being embedded by squares. Borkowski (1998) also introduced Simple Latin Square Sampling +k (SLSS+k) design. In these sampling designs the sample size can be more than the square root of population size. Lawry and Bellhouse (1992) illustrated optimal prop-

* Corresponding author.

erties of a SLSS design under a spatial correlated model. Salehi (2002, 2004) illustrated the optimal properties of SSLSS design under the same spatial correlation model.

There remains one question to be answered. What sampling design could be applied if we want to take a sample with a sample size less than the number of sections of one side of the square population? Borkowski (2003) introduced the Simple Latin Square Sampling -k (SLSS-k) but presented no variance estimator. Hence, this design may not be applicable in practice, since the precision of the estimator can not be determined. To solve this problem and extend the SLSS-k design to a rectangular population we introduce Row and Column Elimination Sampling +1 (RCES+1). We then present its population mean estimator, the variance and variance estimator using Horvitz-Thompson estimator.

Suppose that the population has been partitioned into a $R \times C$ rectangular array of units and these units are indexed by $1, 2, 3, \ldots, RC$, where R is the number of rows and C is the number of columns. Suppose we want to take a sample of size $n+1$ where $n \leqslant \min\{R, C\}$. In practice, we randomly choose a unit and then eliminate the row and column which the selected unit lie in. The second unit will be selected from among the remaining units. This process continues to select n sample. We call this set of units a Row and Column Elimination Sample (RCES). In this design some of the joint inclusion probabilities are zero. Therefore, there is no unbiased variance estimator. So we select another sample from among units which have not been selected before. We name this sampling design a Row and Column Elimination Sample +1 (RCES+1). It can be shown that the inclusion probability for unit j is $\pi_j = \frac{n+1}{RC}$. Using Horvitz-Thompson (1952) estimator an unbiased estimator for mean of the population (μ) is as follows:

$$\hat{\mu} = \frac{1}{RC}\sum_{i=1}^{n+1}\frac{y_i}{\pi_i} = \frac{1}{n+1}\sum_{i=1}^{n+1} y_i = \bar{y}_i,$$

where y_i is the value under consideration for ith unit.

The variance of the mean and its unbiased estimate is simply derived by determining the joint inclusion probabilities, π_{ij}, based on Horvitz-Thompson estimator and its unbiased estimator for variance.

To evaluate the RCES design we compare the variance of RCES design with that of Simple Random Sampling (SRS) without replacement. So we define efficiency as $\text{eff}(\hat{\mu}) = \frac{\text{var}(\bar{y}_s)}{\text{var}(\hat{\mu})}$ where $\text{var}(\bar{y}_s)$ is variance of the sample mean for the SRS design with equivalent sample size to the sampling design under examination. We will illustrate in a theorem that the RCES design is more efficient than the SRS design if and only if the following inequality holds:

$$\sum_{i=1}^{RC} y_i \bar{\omega}_i^c < \sum_{i=1}^{RC} y_i \bar{\omega}_i$$

where $\hat{\omega}_i$ is the mean of units which lie in the row and column of the ith unit and $\hat{\omega}_i^c$ is the mean of units which do not lie in the row and column of the ith unit.

This inequality holds if and only if the covariance of $\hat{\omega}_i^c$s and y_is is greater than the corresponding value for $\hat{\omega}_i^c$s and y_is. If there is a positive autocorrelation value or a horizontal or vertical trend exists in the population we expect that the linear relationship between each unit value with other units in the same row or column is greater than the corresponding value for the units that are not in the same row or column. Consequently, RCES+1 design is an efficient design for populations with horizontal or vertical trend.

We calculate the efficiency of data from a rare population of the freshwater mussels in Cacapon West Virginia River with different sample sizes to evaluate the RCES+1 design. This population has a slight vertical trend. A part of the Cacapon River is partitioned into 800 sections. The population is of 40 rows and 20 columns. The variance of SRS, RCES+1 and the efficiency for sample size of 2 to 21 is given. The calculated efficiencies shows that the RCES+1 is more efficient than corresponding SRS.

In brief, the RCES+1 is such a design that can be applied for any population of rectangular arrays. This design has an unbiased variance estimator and has been planed for situations in which sample size is less than the minimum of the number of rows and columns plus one. The design do not permit repeated sample in rows or columns except for one. Thus it gives a good coverage of the population region. This article illustrate that if there is a positive spatial autocorrelation between units, which means that neighbor units have more similar values, the RCES+1 is an efficient design.

Keywords. Latin Square Sampling Design; Horvitz-Thompson Estimator; Autocorrelation Coeficient; Spatial Data Analysis.

M. Salehi. M
School of Mathematical Sciences,
Isfahan University of Technology,
Isfahan, Iran.
e-mail: *salehi_m@cc.iut.ac.ir*

M. A. Jamalzadeh
Statistical Research Center,
Shahid Fakouri St., Babataher St., Dr. Fatemi Ave.,
Tehran, Iran.
e-mail: *amin_jamalzadeh@srtc.ac.ir*

The full version of the paper, in Persian, appears on pages 187-200.