Available online at http://www.journalcra.com

International Journal of Current Research
Vol.11, pp.222-227, December, 2010

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

REVIEW ARTICLE

DETERMINANT SAMPLING SCHEME FOR RATIO AND REGRESSION METHODS OF ESTIMATION

Subramani, J

Department of Statistics, Pondicherry University, R V Nagar, Kalapet,
Puducherry – 605 014, India

ARTICLE INFO

Article History:
Received 12th September, 2010
Received in revised form
15th October, 2010
Accepted 13th November, 2010
Published online 28th December, 2010

Key words:
Determinant Sampling,
Simple Random sampling,
Systematic Sampling,
Natural population,
Linear trend, Trend free sampling.

ABSTRACT

In this paper, determinant sampling scheme for estimation of a finite population mean is extended for the ratio and regression methods of estimation. The relative performance of determinant sampling along with those of the simple random and systematic sampling are assessed for certain natural populations for estimation of the finite population means through the methods of ratio and regression estimation.

INTRODUCTION

As stated by Subramani (2009), it is of interests for many research workers in their statistical investigations to estimate the finite population mean $\bar{Y}=\frac{1}{N}\sum_{i=1}^{N}Y_i$ on the basis of a random sample selected from the population U , where U is a finite population $\mathrm{S}=\{\mathrm{u}_1,\mathrm{u}_2,...,\mathrm{u}_3\}$ of N distinct and identifiable units. Let Y be a real variable with value Y_i measured on $\mathrm{U}_\mathrm{i}, \mathrm{i}=1, 2, 3, ..., \mathrm{N}$ giving a vector $\mathrm{Y}=(\mathrm{Y}_1, \mathrm{Y}_2 ,..., \mathrm{Y}_\mathrm{N})\cdot$ Any ordered sequence $\mathrm{S}=\{\mathrm{u}_1,\mathrm{u}_2,...,\mathrm{u}_3\}=\{\mathrm{U}_{\mathrm{i}1} ,\mathrm{U}_{\mathrm{i}2} ,...,\mathrm{U}_{\mathrm{in}}\}, 1\le i_l \le N \textit{ and } 1\le l\le n$ is called a random sample of size n to be used to get an estimate for the population mean $\bar{Y}$.

In several sampling schemes including, simple random sampling, systematic sampling, trend free Sampling, diagonal systematic sampling, generalized diagonal systematic sampling, determinant sampling schemes the resulting sample mean $\bar{y}$ turns out to be an unbiased estimate of the population mean $\bar{Y}$, where as the other sampling schemes such as balanced systematic sampling, centered systematic sampling, truncated systematic sampling, etc, the resulting estimators are not unbiased unless there is a perfect liner trend among the population values. For a detailed discussion about the above sampling schemes and their advantageous, the readers are referred to Cochran (1997), Murthy (1967), Fountain and Pathak (1989), Mukerjee and Sengupta (1990), Rao (1969), Subramani (2000; 2009 and 2010). Subramani and Tracy (1999) and the references

***Corresponding author:** *drjsubramani@yahoo.co.in*

cited there in. It is often used the auxiliary information in estimation techniques of sample surveys to increase the precision of the sample mean. The two important and widely used methods are the ratio and regression methods of estimation, which provide better estimators than the usual estimators based on simple random sampling. These methods involve an auxiliary variable X correlated with the study variable Y, and the pair of values (x_i, y_i) is measured for each and every unit in the sample. Further it is assumed that the population mean $\overline{X}$ of the auxiliary variable X is known in advance. This has motivated the present study and the determinant sampling scheme has been extended for estimating finite population means through the methods of ratio and regression estimators. The explicit expressions for the variance of ratio and regression estimators based on determinant sampling scheme are derived. Further the relative performance of determinant sampling, simple random sampling and systematic sampling schemes are assessed for certain natural populations considered by Rao (1969). As a result it has been observed that the determinant sampling scheme performs better than the simple random sampling and the linear systematic sampling means for estimating the finite population mean using ratio and regression methods of estimation.

2. Determinant Sampling Scheme

For the sake of simplicity and for the benefit of the readers, the steps involved in selecting a determinant sample of size n from a population of size $N = kn$ are reproduced here. Let $N = kn$ where $n \le k$, be the population size. The population units $\mathrm{U}_1, \mathrm{U}_2, ..., \mathrm{U}_\mathrm{N}$ are arranged in a $n \times k$ matrix M (say) and the j-th row of M is denoted by $\mathrm{Rj}, \mathrm{j} = 1, 2, ..., \mathrm{n}$. The elements of R_j are $\{\mathrm{U(j-1)k+i, i = 1, 2, ..., k}\}$. The determinant sampling scheme consists of drawing n units from the matrix M randomly such that the selected n units are from different rows and from different columns of the matrix M. Hence no two selected units are not from the same row or from the same column. The steps involved in the determinant sampling scheme for selecting a random sample of size n are given below:

1. Arrange the N population units $\mathrm{U}_1, \mathrm{U}_2, ..., \mathrm{U}_\mathrm{N}$ in an $n \times k$ matrix M (say).
2. Select n random numbers $\mathrm{t}_1, \mathrm{t}_2, ..., \mathrm{t}_\mathrm{n}$ one by one such that $\mathrm{t}_\mathrm{i} \ne \mathrm{t}_\mathrm{j}$ whenever $i \ne \mathrm{j}$, t_i being the random number obtained at the i^{th} draw, with $1 \le t_i \le k$ and $1 \le i \le n$.
3. The selected sampling units are $\mathrm{U}_{\mathrm{t}_1}, \mathrm{U}_{\mathrm{k+t}_2}, ..., \mathrm{U}_{\mathrm{(n-1)k+t}_\mathrm{n}}$, located respectively in the positions $(1, \mathrm{t}_1), (2, \mathrm{t}_2)...(\mathrm{n}, \mathrm{t}_\mathrm{n})$ of the matrix M.

It is to be noted that the first order and second order inclusion probabilities are obtained as given below:

$\pi_i = \frac{1}{k}, \ \ for\ i = 1,2,3...N$ and

$$\pi_{ij} = \begin{bmatrix} \frac{1}{k(k-1)} & \textit{if i and j are from different rows and from different column} \\ 0 & \textit{Otherwise} \end{bmatrix}$$

The first order inclusion probabilities are the same for both the systematic sampling and the determinant sampling schemes but the difference is on the second order inclusion probabilities. The two units in the same column will get the same probability $\frac{1}{k}$ in the case of systematic sampling where as the two units from different rows and from different columns will get the same probability $\frac{1}{k(k-1)}$ in the case of determinant sampling and zero for other pair of units.

Let Y_{ij} be the observation corresponding to the unit in the i^{th} row and j^{th} column, which is corresponding to the unit $U_{(i-1)k+j}$, then the sample observations are denoted by $\mathrm{S} = \{\mathrm{Y}_{\mathrm{t}_1}, \mathrm{Y}_{\mathrm{k+t}_2}, ..., \mathrm{Y}_{\mathrm{(n-1)k+t}_\mathrm{n}}\}$. The mean of a determinant sample is denoted by $\overline{y}_d$ and its variance is given by

$V(\bar{y}_d) = \frac{1}{Nn(k-1)}\left[k(N-1)S^2 - NnS_c^2 - NkS_r^2\right]$ where

$S^2 = \frac{1}{(N-1)}\sum_{i=1}^{k}\sum_{j=1}^{n}(Y_{ij} - \bar{\bar{Y}})^2$, $S_c^2 = \frac{1}{k}\sum_{i=1}^{k}(\bar{Y}_i - \bar{\bar{Y}})^2$,

$S_r^2 = \frac{1}{n}\sum_{i=1}^{n}(\bar{Y}_j - \bar{\bar{Y}})^2$

$\bar{Y}_i = \frac{1}{n}\sum_{j=1}^{n} Y_{ij}$, $\bar{Y}_j = \frac{1}{k}\sum_{i=1}^{k} Y_{ij}$ and $\bar{\bar{Y}} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n} Y_{ij}$.

It has also been proved by Subramani and Tracy (1999) that the determinant sampling scheme performs better than the simple random sampling, systematic sampling and stratified random sampling with unit per stratum schemes for the populations with a perfect linear trend among the population values. Consider the hypothetical population with a perfect linear trend among the population values. Then the values of N population units are in arithmetic progression. That is, $\mathrm{Y_i = a + ib}, \quad \mathrm{i = 1, 2, ..., N}$, where a and b are constants. For the above population with a linear trend, the variances of the simple random sample mean $V(\bar{y}_r)$, systematic sample mean $V(\bar{y}_{sy})$, determinant sample man $V(\bar{y}_d)$ and stratified random sample mean with unit per stratum $V(\bar{y}_{st})$ are obtained as given below:

$V(\bar{y}_r) = \frac{(k-1)(N+1)b^2}{12}$, $V(\bar{y}_{sy}) = \frac{(k-1)(k+1)b^2}{12}$,

$V(\bar{y}_d) = \frac{(k-n)[k+1]b^2}{12n}$ and $V(\bar{y}_{st}) = \frac{(k-n)(k+1)b^2}{12n}$

By comparing the various variance expressions given above, one can easily find that determinant sampling is more efficient than the other sampling schemes. In fact $V(\bar{y}_d) \le V(\bar{y}_{st}) \le V(\bar{y}_{sy}) \le V(\bar{y}_r)$ .The equality sign attains only when $n = 1$. The following theorem gives the expression for covariance between $\bar{x}_d$ and $\bar{y}_d$, which will be useful for deriving the variances of the Ratio estimator $\bar{y}_{Rd}$ and linear regression estimator $\bar{y}_{ld}$ of the population mean $\bar{Y}$ based on determinant sampling. Here suffices Rd and ld are used to represent ratio and linear regression estimators based on determinant sampling.

Theorem 2.1: If x_i, y_i are a pair of values observed on the i^{th} unit of the population of size N and $\bar{x}_d, \bar{y}_d$ are the corresponding means of a determinant sample of size n, then the covariance is given by

$$Cov(\bar{x}_d, \bar{y}_d) = \frac{1}{Nn(k-1)}\left[k(N-1)S_{xy} - NnS_{cxy} - NkS_{rxy}\right]$$

$S_{xy} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \bar{\bar{X}})(Y_{ij} - \bar{\bar{Y}})$,

$S_{cxy} = \frac{1}{k}\sum_{i=1}^{k}(\bar{X}_i - \bar{\bar{X}})(\bar{Y}_i - \bar{\bar{Y}})$,

$S_{rxy} = \frac{1}{n}\sum_{i=1}^{n}(\bar{X}_j - \bar{\bar{X}})(\bar{Y}_j - \bar{\bar{Y}})$, $\bar{Y}_i = \frac{1}{n}\sum_{j=1}^{n} Y_{ij}$,

$\bar{Y}_j = \frac{1}{k}\sum_{i=1}^{k} Y_{ij}$ and $\bar{\bar{Y}} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n} Y_{ij}$.

The above theorem can be easily proved by following the steps similar to the theorem 2.3 of Cochran (1977, page 25). When each $X_{ij} = Y_{ij}$ then the expression given above turns out to be the expression given for $V(\bar{y}_d)$.

3. Ratio Methods of Estimation based on Determinant Sampling Scheme

In general the ratio estimator $\hat{\bar{\bar{Y}}}_R$ of the population mean $\bar{\bar{Y}}$ and its variance $V(\hat{\bar{\bar{Y}}}_R)$ are respectively given by

$$\hat{\bar{\bar{Y}}}_R = \frac{\bar{y}}{\bar{x}}\bar{X} \qquad (3.1)$$

and

$$V(\hat{\bar{\bar{Y}}}_R) = V(\bar{y}) + R^2V(\bar{x}) - 2RCov(\bar{x}, \bar{y}) \qquad (3.2)$$

where $\bar{x}$ and $\bar{y}$ are the sample means of the auxiliary variable X and the study variable Y. By substituting the sample means of the variables X and Y obtained through various sampling schemes in the above equation (3.1), one may get the corresponding ratio estimators for the population mean $\bar{Y}$. Similarly by substituting the corresponding expressions of $V(\bar{y})$, $V(\bar{x})$ and

$Cov(\bar{x},\bar{y})$ obtained through various sampling schemes in the equation (3.2) given above, one may get the appropriate expressions for the variance of the ratio estimators. The derivations of the above expressions are straightforward and hence they are omitted here and only the final results are given. For the sake of convenience of the readers, the variances of the ratio estimators of the population mean based on simple random sampling, systematic sampling and determinant sampling schemes are given below:

$$V(\hat{\bar{Y}}_{Rr}) = \frac{(N-n)}{Nn}\left[S_y^2 + R^2 S_x^2 - 2RS_{xy}\right] \qquad (3.3)$$

$$V(\hat{\bar{Y}}_{Rsy}) = S_{cy}^2 + R^2 S_{cx}^2 - 2RS_{cxy} \qquad (3.4)$$

$$V(\hat{\bar{Y}}_{Rd}) = \frac{1}{Nn(k-1)}\begin{bmatrix} k(N-1)(S_y^2 + R^2 S_x^2 - 2RS_{xy}) \\ -Nn(S_{cy}^2 + R^2 S_{cx}^2 - 2RS_{cxy}) - Nk(S_{ry}^2 + R^2 S_{rx}^2 - 2RS_{rxy}) \end{bmatrix} \qquad (3.5)$$

4. Regression Methods of Estimation based on Determinant Sampling Scheme

In general the linear regression estimator $\hat{\bar{Y}}_{lr}$ of the population mean $\bar{Y}$ and its variance $V(\hat{\bar{Y}}_{lr})$ are respectively given by

$$\hat{\bar{Y}}_{lr} = \bar{y} + b(\bar{X} - \bar{x}) \qquad (4.1)$$

and $V(\hat{\bar{Y}}_{lr}) = V(\bar{y}) + b^2 V(\bar{x}) - 2bCov(\bar{x},\bar{y})$ (4.2)

where b is the known constants; $\bar{x}$ and $\bar{y}$ are the sample means of the auxiliary variable X and the study variable Y. By substituting the sample means of the variables X and Y obtained through various sampling schemes in the above equation (4.1), one may get the corresponding linear regression estimators for the population mean $\bar{Y}$. Similarly by substituting the corresponding expressions of $V(\bar{y})$, $V(\bar{x})$ and $Cov(\bar{x},\bar{y})$ obtained through various sampling schemes in the equation (4.2) given above, one may get the appropriate expressions for the variance of the linear regression estimators. If the value of b is unknown then it can be replaced by the estimate of the regression coefficient. However it has been shown in literature (Cochran, 1977) that $V(\bar{y}_{lr})$ attains minimum when $b = \frac{Cov(x,y)}{V(x)}$. For assessing the relative performance of determinant sampling with that of simple random sampling and systematic sampling schemes, we have considered the minimum value of the respective variances. The derivations of the above expressions are straightforward and hence they are omitted here and only the final results are given. For the sake of convenience of the readers, the variances of the linear regression estimators of the population mean based on simple random sampling, systematic sampling and determinant sampling schemes are given below:

$$V(\hat{\bar{Y}}_{lr}) = \frac{(N-n)}{Nn}\left[S_y^2 + b_r^2 S_x^2 - 2b_r S_{xy}\right] \qquad (4.3)$$

$$V(\hat{\bar{Y}}_{lrsy}) = S_{cy}^2 + b_{sy}^2 S_{cx}^2 - 2b_{sy} S_{cxy} \qquad (4.4)$$

$$V(\hat{\bar{Y}}_{lrd}) = \frac{1}{Nn(k-1)}\begin{bmatrix} k(N-1)(S_y^2 + b_d^2 S_x^2 - 2b_d S_{xy}) \\ -Nn(S_{cy}^2 + b_d^2 S_{cx}^2 - 2b_d S_{cxy}) - Nk(S_{ry}^2 + b_d^2 S_{rx}^2 - 2b_d S_{rxy}) \end{bmatrix} \qquad (4.5)$$

Where $b_r = \frac{Cov(x_r, y_r)}{V(x_r)}$, $b_{sy} = \frac{Cov(x_{sy}, y_{sy})}{V(x_{sy})}$ and $b_d = \frac{Cov(x_d, y_d)}{V(x_d)}$. That is, b_r, b_{sy} and b_d are respectively the regression coefficients obtained through simple random sampling, systematic sampling and determinant sampling schemes.

5. Relative Performance of Determinant Sampling for a Certain Natural Populations

To evaluate the relative performance of ratio and linear regression estimators based on determinant sampling over simple random and systematic sampling schemes, we have selected 9 natural populations. The populations numbered from 1 to 7 are taken from Cochran (1977).These populations are also considered by Rao (1969) to assess the relative performance of several ratio and regression estimators. The populations numbered 8 and 9 are taken from Murthy (1967). Table 5.1 provides the source, nature of the study variable Y and the

Table 5.1: Description of the Populations considered for the Empirical Study

Population Number	*Source*	*Study Variable Y*	*Auxiliary Variable X*	Population Size N
1	Cochran (1977) Page 152 Cities 1-49	Size of the City in USA 1930	Size of the City in USA 1920	49
2	Cochran (1977) Page 152 Cities 25-49	Size of the City in USA 1930	Size of the City in USA 1920	25
3	Cochran (1977) Page 152 Cities 1-24	Size of the City in USA 1930	Size of the City in USA 1920	24
4	Cochran (1977) Page 182	No. of Polio cases in the non inoculated group	No. of not inoculated group	34
5	Cochran (1977) Page 182	No. of Polio cases in the Placebo group	No. of Placebo children	34
6	Cochran (1977) Page 203	Actual weight of peaches	Eye estimated weight of peaches	10
7	Cochran (1977) Page 325	No. of Persons in a block	No. of rooms in a block	10
8	Murthy (1967) Page 399	Area under wheat in 1964	Cultivated area in 1961	34
9	Murthy (1967) Page 399	Area under wheat in 1964	Cultivated area in 1963	34

Table 5.2: Comparison of Ratio Estimators based on simple random, systematic and determinant sampling for the population

Population Number	n	$V(\bar{y}_{Rr})$	$V(\bar{y}_{Rsy})$	$V(\bar{y}_{Rd})$
1	7	75.9807	45.8327	80.2353
2	5	139.4397	104.6067	135.2278
3	2	200.6973	322.8438	177.7578
	3	127.7119	84.1308	135.1133
	4	91.2265	136.7700	77.5781
4	2	2.6903	2.8459	2.7224
5	2	2.2649	2.4256	2.3298
6	2	2.9484	2.9913	2.4026
7	2	60.0763	73.4718	55.9298
8	2	2078.2930	1819.6810	2160.1780
9	2	424.5352	510.3750	411.7852

Table 5.3: Comparison of Linear Regression Estimators based on simple random, systematic and determinant sampling for the population

Population Number	n	$V(\bar{y}_{lr})$	$V(\bar{y}_{lrsy})$	$V(\bar{y}_{lrd})$
1	7	67.1616	43.1821	69.4816
2	5	89.1116	13.6676	107.3137
3	2	200.2793	293.3779	177.3828
	3	127.4473	78.9668	134.9287
	4	91.0371	133.7642	77.2539
4	2	2.5292	2.8441	2.5769
5	2	2.2198	2.3759	2.2531
6	2	1.9652	2.1747	1.4815
7	2	54.9156	56.7016	48.5703
8	2	1935.8830	1553.2380	2021.5720
9	2	418.7012	501.1426	408.8789

auxiliary variable X and the population size N. We have considered all possible cases of k and n and hence we have totally 11 cases. The variances of ratio estimators and the linear regression estimators obtained through determinant sampling, simple random sampling and systematic sampling schemes are respectively presented in Table 5.2 and Table 5.3.

The following strategies are used to evaluate the relative performance of the ratio and linear regression estimators based on these three different sampling schemes.

1. An estimator is said to be the best it is has the minimum variance among the estimators considered here.
2. An estimator is said to be the worst it is has the maximum variance among the estimators considered here.
3. An estimator is said to be a favourable estimator if it has minimum variance more frequently and maximum variance rarely.

The following are the tentative conclusions we have arrived at the following conclusions based on the results of the empirical study conducted and the results presented in Tables 5.2 and 5.3 respectively for the cases of ratio estimators and the linear regression estimators. The observations made on the relative performance of these different sampling schemes are as follows:

- Among the 9 populations with 11 possible cases considered, the variances of the ratio estimators presented in Table 5.2, the simple random sampling, systematic sampling and determinant sampling have minimum variances respectively in 2, 4 and 5 cases. On the other hand these estimators have maximum variances respectively in 1, 7, 3 cases.
- Similar results have been obtained from Table 5.3 for the case of linear regression estimators. That is, the regression estimators based on simple random sampling, systematic sampling and determinant sampling schemes have minimum (maximum) variances respectively in 2(0), 4(7) and 5(4) cases.

Hence we conclude that for practical purposes one may use the determinant sampling scheme for estimating the population mean through the methods of ratio and linear regression estimation compared to simple random sampling and systematic sampling schemes.

REFERENCES

Cochran, W.G. 1977. Sampling Techniques, 3rd Edition, John Wiley and Sons, New York

Fountain, R.L and Pathak, P.L.1989. Systematic and non-random sampling in the presence of Linear Trends Communications in statistics. *Theory and Methods*, 18: 2511-2526.

Mukerjee, R and Sengupta, S 1990. Optimal Estimation of a Finite Population Means in the Presence of Linear Trend, *Biometrika.,* 77: 625-630.

Murthy, M.N. 1967. Sampling theory and Methods, Statistical Publishing House, Calcutta, India.

Rao, J N K. 1969. Ratio and Regression Estimators – in New Developments in Survey Sampling, Ed. By N L *Johnson and H.Smith, Jr.,* 213-234.

Subramani, J. 2000. Diagonal Systematic Sampling Scheme for Finite Populations, Jour. Ind. Soc. Ag. Statistics 53(2): 187-195.

Subramani, J. 2009 Further Results on Diagonal Systematic Sampling Scheme for Finite Populations, Journal of Indian Society of Agricultural Statistics.63 (3): 277-282**.**

Subramani, J. 2010 Generalization of Diagonal Systematic Sampling Scheme for Finite Populations. *Model Assisted Statistics and Applications,* 5: 117-128.

Subramani, J. and Tracy, D. S. 1999. Determinant Sampling Scheme for Finite Populations. *Internl. J. Math. and Statist. Sci.,* 8(1): 27-41.