THE EGYPTIAN STATISTICAL JOURNAL
ISSR, CAIRO UNIV, VOL.39 NO.2,1995

An Estimator in Cluster Sample Combined with Stratification

by: Mounira A. Hussein*

Introduction:

Cluster sampling is a simple random sample in which each sampling unit is a group of elements. It may provide maximum information at minimum cost when a frame listing population elements is not available. It is also preferable when the cost of obtaining observations increases with increasing distance between elements (Scheaffer, Mendenhall and ott,1983). Cluster sampling can be combined with stratified sampling. Sample means and ratio estimators are illustrated for cluster sampling combined with stratification in most sampling literature (Cochran , 1977). An alternative product estimator was proposed independently by Serivenkataramana (1980) and Bandyopadhyay (1980) for a positive correlated auxiliary variable in simple random sample. This estimator was proposed for stratified random sample (Hussein ,1992). In this paper we are proposing it for cluster sampling combined with stratification. We are also proposing strategy to get the best estimator in cluster sample combined with stratification.

Study Objectives:

This paper is organized to accomplish the following objectives :

1- Introducing the alternative product type of estimator for first stage cluster sample combined with stratification.

2-Making use of the criteria for choosing between estimators in simple random sample , we are choosing between estimators in each stratum in order to get the best estimator in each stratum which in turn is used to get the best estimator in cluster sample combined with stratification.

3-Numerical example is shown for the purpose of comparison among estimators. The best estimator with its efficiency compared with other estimators are illustrated.

Notations and definitions:

The following notations are used in this paper :

1- N_h and $n_h < N_h$ are the number of clusters in the population and in the sample respectively in stratum h .

2- Y_h is the population total of the characteristic under study in stratum h.

3-M_h is the total of elements in the stratum h.

---

* : Mounira A. Hussein is an Assistant Professor , College of Commerce, Menoufia University.

4- m_{ih} is the number of the elements in the cluster i and the stratum h.
5- $\hat{y}_h = N_h \bar{y}_h$ and $\hat{M}_h = N_h \bar{m}_h$ are unbiased estimators for Y_h and M_h based on a sample of size n_h .
6- $\bar{m}_h$ and $\bar{y}_h$ are the sample means for the cluster size and the characteristic under study respectively.
7- C_{yyh}, C_{mmh} ,and C_{ymh} are the relative variance for the characteristic under study , relative variance for cluster size and their relative covariance respectively.

An alternative product estimator for first stage cluster sample:

In this section we are proposing the alternative product estimator for first stage cluster sample. An estimator for the total population using the ratio estimator for the first stage cluster sample was proposed by Scheaffer, Mendenhall, and Ott (1983) as follows :

$$\hat{y}_{rh} = \frac{\bar{y}_h}{\bar{m}_h} M_h \qquad (1)$$

In the same way we can propose the alternative product estimator for first stage cluster sample as follows:

$$\hat{y}_{ah} = \frac{\bar{y}_h}{\bar{m}_h} M_h^* \qquad (2)$$

$$where\ M_h^* = \frac{(N_h M_h - n_h \hat{M}_h)}{N_h - n_h}$$

$$\hat{M}_h = N_h \bar{m}_h$$

Best Estimator in Cluster Sample Combined with Stratification:

In this section we are presenting a strategy to get a best estimator in cluster sample combined with stratification as follows:

First : We consider each stratum as a separate population and choose a simple random sample of clusters from each stratum. The best estimator in each stratum can be chosen according to a criteria for simple random sample introduced by Serivenkataramana (1980) as follows::

1- The sample mean is the best estimator when the correlation between the characteristic under study and the cluster size in the stratum is low within the following limits:

$$0.0 < \rho_{ymh} < \frac{1}{2} g_h \frac{C_{mh}}{C_{yh}} \qquad (3)$$

$$where\ g_h = \frac{n_h}{N_h - n_h}$$

and C_{yh} and C_{mh} are the square roots of the relative variances.

2- The ratio estimator is the best estimator when the correlation between the characteristic under study and the cluster size in the stratum is high within the following limits:

$$\frac{1}{2}(1+g_h)\frac{C_{mh}}{C_{yh}} < \rho_{ymh} < 1 \qquad (4)$$

3- The alternative product type estimator is the best estimator when the correlation between the characteristic under study and the cluster size in the stratum is not low and not high within the following limits :

$$\frac{1}{2} g_h \frac{C_{mh}}{C_{yh}} < \rho_{ymh} < \frac{1}{2}(1+g_h)\frac{C_{mh}}{C_{yh}} \qquad (5)$$

Second: We will express the final estimator as a separate estimator for stratified sample . For example , if we have two strata we have six alternatives that can be considered as follows:

1- The ratio estimator in the first stratum and the product estimator in the second stratum.
2- The sample mean estimator in the first stratum and the ratio estimator in the second.
3- The sample mean estimator in the first stratum and the alternative product estimator in the second .
4- The ratio estimator for both the first and the second stratum .
5- The alternative product estimator for both the first and the second stratum .
6- The sample mean estimator for both the first and the second stratum.

Table (1) illustrates the formula for each estimator and its mean square error. Table (2) illustrates the separate estimator for the six alternative cases.

Table (1) : Estimators of the Total Population for Specific Stratum h.

Estimator	*Mean Square Error*
$\hat{Y}_h$	$\frac{1-f_h}{n_h}\hat{Y}_h^2 C_{yyh}$
$\hat{Y}_{rh}$	$\frac{1-f_h}{n_h}\hat{Y}_h^2 \{C_{yyh}-2C_{ymh}+C_{mmh}\}$
$\hat{Y}_{ah}$	$\frac{1-f_h}{n_h}\hat{Y}_h^2 \{C_{yyh}-2g_h C_{ymh}+g_h^2 C_{mmh}\}$

Table (2) : Final Estimators for the Total Population .

Estimator	*Mean Square Error*
$\sum_{h=1}^{2}\hat{Y}_h$	$\sum_{h=1}^{2}\frac{1-f_h}{n_h}\hat{Y}_h^2 C_{yyh}$
$\sum_{h=1}^{2}\hat{Y}_{rh}$	$\sum_{h=1}^{2}\frac{1-f_h}{n_h}\hat{Y}^2_h(C_{yyh}-2C_{ymh}+C_{mmh})$
$\sum_{h=1}^{2}Y_{ah}$	$\sum_{h=1}^{2}\frac{1-f_h}{n_h}\hat{Y}^2_h(C_{yyh}-2g_hC_{ymh}+g_h^2C_{mmh})$
$\hat{Y}_1+\hat{Y}_{r2}$	$\frac{1-f_1}{n_1}\hat{Y}^2_1C_{yy1}+\frac{1-f_2}{n_2}\hat{Y}_2^2(C_{yy2}-2C_{ym2}+C_{mm2})$
$\hat{Y}_1+\hat{Y}_{a2}$	$\frac{1-f_1}{n_1}\hat{Y}^2_1C_{yy1}+\frac{1-f_2}{n_2}\hat{Y}_2^2(C_{yy2}-2g_2C_{ym2}+g_2^2C_{mm2})$
$\hat{Y}_{r1}+\hat{Y}_{a2}$	$\frac{1-f_1}{n_1}\hat{Y}^2_1[C_{yy1}-2C_{ym1}+C_{mm1}]+\frac{1-f_2}{n_2}\hat{Y}_2^2\{C_{yy2}-2g_2C_{ym2}+g_2^2C_{mm2}\}$

Numerical Example:

The following example is used by Scheaffer , Mendenhall and Ott (1983) for ratio estimator in cluster sampling combined with stratification. "A sociologist wants to estimate the average income per adult male as well as the total income in a certain small city. No List of resident adults is available. So cluster sample is recommended. Since the city has two districts that are different in the economic levels, cluster sample combined with stratification is suggested . In the first stratum 10 blocks are sampled from 168 blocks while 25 blocks are sampled from 415 blocks from the second stratum".The data on total income and number of adult males are utilized using ratio estimator. We applied the proposed estimator in order to get more efficient estimate according to the following steps.

1- Applying the criterion for choosing among estimators for stratum 1, We found that:

$$0.5229 = \frac{1}{2}(1+g_1)\frac{C_{m1}}{C_{y1}} < \rho_{m1} = 0.975 \qquad (6)$$

This means that the ratio estimator is the best estimator for stratum 1.

2- Applying the criterion for choosing among estimators for stratum 2, We found that:

$$.031 = \tfrac{1}{2} g_2 \frac{C_{m2}}{C_{y2}} < \rho_{ym2} = 0.303 < \tfrac{1}{2}(1+g_2)\frac{C_{m2}}{C_{y2}} = .509 \qquad (7)$$

This means that the alternative product estimator is the best estimator for stratum 2. Table (3) and Table (4) illustrate the estimate , and the relative efficiency for the three different estimators for stratum 1 and stratum 2 respectively.

Table (3) : Relative Efficiency of Different Estimators of Stratum 1.

	Estimate	$\sqrt{MSE}$	Relative Efficiency
$\hat{Y}$	9189600	1666305	1
$\hat{Y}_r$	9153878	370621	4.50
$\hat{Y}_a$	9151645	1565327	1.06

Table (4) : Relative Efficiency of Estimators of Stratum 2.

Estimator	Estimate	$\sqrt{MSE}$	Relative Efficiency
$\hat{Y}$	22061400	1752792.5	1
$\hat{Y}_r$	22003311	2026217.5	.87
$\hat{Y}_a$	22057666	1723232.1	1.02

Therefore , the proposed final estimator is $\hat{Y}_{r1} + \hat{Y}_{a2}$.

Comparison between the proposed estimator and ratio estimator:

In the following we prove that the mean square error of the proposed estimator is less than the mean square error of the ratio estimator in some situations.

Since

$$M(\hat{Y}_{r1} + \hat{Y}_{a2}) < M(\sum_{h=1}^{2} Y_{rh}) \qquad (8)$$

$$\text{iff} \quad -2g_2 C_{ym2} + g_2^2 C_{mm2} < -2C_{ym2} + C_{mm2} \qquad (9)$$

$$\text{iff} \quad C_{ym2} < \frac{1}{2}(1+g_2) C_{mm2} \qquad (10)$$

Similarly, this condition can be satisfied for some strata in other situations.

Results and conclusion:

The ratio estimator is a best estimator for stratum 1. This result agrees with the criterion stated before. The product type estimator is a best estimator for stratum 2. This result also agrees with the criterion stated before Therefore, the best estimator for one stage cluster sample combined with stratification is the last estimator in table(2).This estimator is the most efficient estimator. It is worthwhile to attempt to find the best estimator when we have (L) strata and expressing the final estimator and it's mean square error in matrix form. The author is working on this and she will publish it later if she is successful

References:

1- ***Bandyopadhyay S.*** (1980), Improved ratio and product estimators, Sankhya, 2, series c, pt.1 and 2, pp.45-49.

2-***Cochran, W. G.***(1977), Sampling techniques, 3rd edition, John Wiley New York.

3-***Chang , D.*** S.(1986), The asymptotic distribution of multivariate product estimator, Chinese Journal of mathematics, Vol.14, No. 3.

4- ***Hussein*, *M. A*.** (1992), Alternative Estimators for Stratified Random Sampling.The Egyptian Statistical Journal, ISSR ,Cairo University ,Vol. 35 , No. 2 1992.

5-***Ouyang* , *Z.* ,*Srivastava* , *J.N.* and *Schreuder* , *H .T.***(1992), A general ratio estimator and its application in model based inference, Ann. Inst. Statist. Math. Vol.45, no. 1, 113-127.

6-***Sekkappan* , *R.*** M.(1986) , Estimation in Sampling from finite populations under the general linear regression model, Journal of Indian statistical association ,Vol. 24, 91-98.

7- ***Scheaffer* , *R. L.* , *Mendenhall*, *W.* and *Ott*, *L*.** (1983) . Elementary Survey Sampling . Wadsworth , Inc.

8-***Srivenkataramana*** ,*T.*(1980) , Dual to the ratio estimator in sample surveys , Biometrika, 67,1, pp. 193-204.