# COMPARATIVE ANALYSIS OF VARIANCE UNDER SAMPLES ALLOCATION FOR DEVELOPMENTAL PROGRAM EVALUATION

Rajiv PANDEY[1]

- *ABSTRACT: Evaluation of program, implemented at large scale in different phases can be ascertained by proper allocation of sampling units from each phase of program for further analysis. Method has been proposed for sample allocation in different strata (phase) with consideration that impacts in successive phases (implemented at different time) follow arithmetic progression (Pandey and Verma, 2008). Present paper proposes variance for the design under the allocation method. Comparison has also been made and discussed with the variance of proportional and optimum allocation methods.*

- *KEYWORDS: Impact evaluation; sample allocation; arithmetic progression; weight; efficiency.*

## 1 Introduction

The estimation of any parameter of interest for a large population depends primarily on sampling theory, which deals with the properties of the estimates from a statistical sample (Cochran, 1977). However, for the heterogeneous population the allocation of sample size in different homogeneous strata is crucial for estimation of parameters (Aoyama, 1954).

For a heterogeneous population, the samples are allocated into various strata depending on the nature of the population. Ideally, the sample allocation should be optimized so that the precision is maximized within the cost constraint. The simplest form of optimal allocation is to make the sampling fraction in stratum proportional to the standard deviation in the stratum, and inversely proportional to the square root of the cost of including a unit from the stratum in the sample. That is, more heterogeneous and cheaper strata are sampled at higher rates (Dalenius and Hodges, 1959; Cochran, 1977; Sukhatme *et al.*, 1984).

In stratified sampling, for a given sample size, approximate minimization of variation depends on number of strata, sample allocation within strata, population variance within strata, population size within strata, and strata boundary break points (Cochran, 1977). However, Pandey and Verma (2008) and Pandey (2010) have discussed the case of sample allocation by assigning weights for those designs, where all information, except sample allocation within strata is known apriori. These design dealt with the case of implementation of developmental program in phased manner, under the assumptions that the units of different strata received different impacts and follows some

[1] Indian Council of Forestry Research & Education, Dehradun, Utterakhand, 248006, India. E-mail: *rajivfri@yahoo.com*

trend such as arithmetic and geometric. Such program includes various governments and non-governments sponsored programs for large region and implemented at time intervals such as educational mission, health (child care, maternal care, AIDS eradication), poverty eradication. The impact of these programs differs in successive phases and may follow known and unknown trend such as arithmetic, geometric and others.

This paper addresses the properties of the estimator and compare with the conventional proportional and optimum allocation method for sample allocation within strata through integration of weight of response under different phases with assumption that the response in successive phases has additive effect. The efficiency of estimator for phased evaluation under combined arithmetic and geometric impacts has been dealt elsewhere (Pandey, 2011). Actual practical examples are not being readily discussed, however, chances of availability of such situation is very high.

## 2 Sample allocation and variance as per Pandey and Verma (2008)

Pandey and Verma (2008) consider the case of sample allocation for population of size N containing the population units $N_1, N_2, ..., N_h$ of a development programme implemented into "h" different phases under the consideration that the impact of the programme is uniformly distributed within each phases and response follows arithmetic trend with respect to different phases. Therefore, the sample weight is as follows:

$$\varpi_i = \frac{(h-i+1)\,N_i}{N_{Hy}} \quad , \tag{1}$$

where

h - number of development programme phase
i - stratum number, i = 1, 2, 3, …, h
$N_i$ - Actual numbers of units (beneficiaries) in the $i$ th stratum
$N_{Hy}$ - Total numbers of units in the population adjusted by the impact or phase factor with product of actual population size in each stratum i.e. $\sum[\{h-i+1\}N_i]$.

Under this case, the unbiased estimator of population mean is as follows

$$\overline{y}_{stm} = \sum_{i=1}^{h} \omega_i \overline{y}_i \quad , \tag{2}$$

where $\overline{y}_{stm}$ is the mean of the character and $\omega_i$ is weight.

And, the variance of $\overline{y}_{stm}$ is as

$$V(\overline{y}_{stm}) = \sum_{i=1}^{h} \frac{\omega_i^2 S_i^2}{n_i} - \sum_{i=1}^{h} \frac{\omega_i S_i^2}{N_i} . \tag{3}$$

## 3 Properties of estimator

The estimated variance of $\overline{y}_{stm}$ may be rewritten as follows after the substitution of value of $\omega_i$ as per Pandey and Verma (2008)

$$\omega_i = \frac{n_i}{n}.$$

With substitution of equation (1), the expression may be written as,

$$\Rightarrow n_i = \omega_i n = \frac{n[h-i+1]N_i}{\sum_{i=1}^{h}[h-i+1]N_i}. \tag{4}$$

With substitution of equation (4) in equation (3), the estimated variance may be written as

$$V(\overline{y}_{stm}) = \sum_{i=1}^{h}\frac{n_i}{n^2}S_i^2 - \sum_{i=1}^{h}\frac{n_i}{n}.\frac{S_i^2}{N_i}$$

$$= \sum_{i=1}^{h}\frac{\{h-i+1\}N_i}{\sum_{i=1}^{h}[\{(h-i+1)\}N_i]}.\frac{S_i^2}{n} - \sum_{i=1}^{h}\frac{\{(h-i+1)\}}{\sum_{i=1}^{h}[\{(h-i+1)\}N_i]}.S_i^2. \tag{5}$$

To be more simple, substitute $\sum[\{h-i+1\}N_i] = N_{Hy}$ (Say).

Therefore, the above equation of estimated variance may be written as follows

$$V(\overline{y}_{stm}) = \sum_{i=1}^{h}\frac{\{h-i+1\}N_i}{N_{Hy}}.\frac{S_i^2}{n} - \sum_{i=1}^{h}\frac{\{(h-i+1)\}}{N_{Hy}}.S_i^2. \tag{6}$$

On the other hand, the variance for simple random sampling (SRS) will be as follows:

$$V(\overline{y}_{sr}) = (1-\frac{n}{N})\frac{S^2}{n}. \tag{7}$$

Where, $S^2$ is sample variance under SRS.

And variance under conventional proportional allocation method

$$V(\overline{y}_{\text{Prop}}) = (\frac{1}{n}-\frac{1}{N})\sum_{i=1}^{h}\omega_i S_i^2 = \frac{1}{nN}(1-\frac{n}{N})\sum_{i=1}^{h}N_i S_i^2. \tag{8}$$

And variance under Optimum allocation method

$$V(\overline{y}_{Ney}) = \frac{1}{n}(\sum_{i=1}^{h}\omega_i S_i)^2 - \frac{1}{N}\sum_{i=1}^{h}\omega_i S_i^2 . \tag{9}$$

Then this method is efficient to SRS and stratified sampling with conventional proportional and Neyman allocation, if and only if, the $V(\overline{y}_{sr}) - V(\overline{y}_{stm})$; $V(\overline{y}_{\Pr op}) - V(\overline{y}_{stm})$; and $V(\overline{y}_{Ney}) - V(\overline{y}_{stm})$ is positive.

Under comparison, the difference of variance between conventional proportional and proposed proportional allocation method may be as follows

$$V(\overline{y}_{\Pr op}) - V(\overline{y}_{stm}) = \frac{1}{nN}(1-\frac{n}{N})\sum_{i=1}^{h} N_i S_i^2 - \sum_{i=1}^{h}\frac{\{h-i+1\}N_i}{N_{Hy}}.\frac{S_i^2}{n} + \sum_{i=1}^{h}\frac{\{(h-i+1)\}}{N_{Hy}}.S_i^2 . \tag{10}$$

And, if finite population correction (fpc) terms ignored then, the equation (10) may be written as follows:

$$V(\overline{y}_{\Pr op}) - V(\overline{y}_{stm}) = \frac{1}{nN}\sum_{i=1}^{h} N_i S_i^2 - \frac{1}{n}\sum_{i=1}^{h}\frac{\{(h-i+1)\}}{N_{Hy}}N_i S_i^2 . \tag{11}$$

Equation 11 may be written as follows

$$V(\overline{y}_{\Pr op}) - V(\overline{y}_{stm}) = \frac{1}{nh}\sum_{i=1}^{h} S_i^2 - \frac{1}{nN_{Hy}}\sum_{i=1}^{h}(h-i+1)N_i S_i^2 . \tag{12}$$

This will be a positive quantity, if, the first term is more than second term for the left hand side (LHS). This may be also true, if the individual coefficient of $S_i^2$ for first term of LHS is greater than the corresponding coefficient of $S_i^2$ for second term of LHS for each coefficient.

$$\begin{aligned}
\frac{1}{h} > \frac{(h-i+1)N_i}{N_{Hy}} \quad &\Rightarrow \quad \sum[(h-i+1)N_i] > h\{(h-i+1)\}N_i \\
&\Rightarrow Nh + N - h^2 N_i - hN_i > \sum iN_i - ihN_i \\
&\Rightarrow N(h+1) - hN_i(h+1) > \sum iN_i - ihN_i \\
&\Rightarrow (h+1)(N-hN_i) > \sum iN_i - iN + i(N-hN_i) \\
&\Rightarrow (h-i+1)(N-hN_i) > \sum iN_i - iN .
\end{aligned} \tag{13}$$

Thus, the proposed estimator will be efficient than the conventional proportional method $V(\overline{y}_{\Pr op}) \geq V(\overline{y}_{stm})$, if and only if, equation 13 holds. The all phase (strata) will have equal number of units.

Based on the similar analogy, for $V(\overline{y}_{Ney}) - V(\overline{y}_{stm})$ should also be a positive quantity and may be written as follows, without inclusion of the finite population correction (fpc) term

$$V(\overline{y}_{Ney}) - V(\overline{y}_{stm}) = \frac{1}{nN^2}(\sum_{i=1}^{h} N_i S_i)^2 - \frac{1}{nN_{Hy}}\sum_{i=1}^{h}(h-i+1)N_i S_i^2 . \qquad (14)$$

That is, Eq. (14) can be written as follows,

$$V(\overline{y}_{Ney}) - V(\overline{y}_{stm}) = \frac{1}{nh^2}(\sum_{i=1}^{h} S_i)^2 - \frac{1}{nN_{Hy}}\sum_{i=1}^{h}(h-i+1)N_i S_i^2 . \qquad (15)$$

This equation is sensitive in respect to total number of strata h, and the difference will be positive, if and only if, for individual *i*, following will be true.

$$\frac{1}{h^2} > \frac{(h-i+1)N_i}{N_{Hy}} \Rightarrow \sum[(h-i+1)N_i] > h^2\{(h-i+1)\}N_i . \qquad (16)$$

This condition is essentially required for the proposed estimator to be more efficient than the Optimum allocation.

## Conclusion

It can be concluded that the proposed estimator is case sensitive as far as efficiency is concerned with respect to conventional proportional and Optimum allocation method and logically, with the simple random sample too. However, theoretically sounds well than these all. The proposed estimator will be equal to the simple random sample, if each stratum contains same number of units, however, less efficient than proportional and optimum allocation method.

PANDEY, R. Análise de variância comparativa sobre alocação de amostras na avaliação de programas de desenvolvimento. *Rev. Bras. Biom.*, São Paulo, v.29, n.2, p.198-203, 2011.

- RESUMO: Avaliação de programa implementado em larga escala em diferentes fases pode ser determinada por uma distribuição adequada de unidades de amostragem de cada fase do programa para análise posterior. Métodos tem sido propostos para a alocação da amostra em diferentes estratos (fases), considerando que os impactos nas diversas fases (implementado em horários diferentes) crescem em progressão aritmética (Pandey e Verma, 2008). O presente trabalho propõe para a verificação do projeto o método de repartição. Comparação também foi feita e discutida com a variação dos métodos de alocação proporcional e ótima.

- PALAVRAS-CHAVE: Avaliação de impacto; alocação de amostra; progressão aritmética; peso; eficiência.

## References

AOYAMA, H. A study of stratified random sampling. *Ann. Inst. Stat. Math.*, Beachwood, v.6, n.1, p.1-36, 1954.

COCHRAN, W. G. *Sampling techniques*. 3.ed. New Delhi: John Wiley & Sons, 1977. 428p.

DALENIUS, T.; HODGES, J. L. Minimum variance stratification. *J. Am. Stat. Assoc.*, New York, v.54, n.285, p.88-101, 1959.

PANDEY, R. Samples allocation for evaluation of developmental program: A case of combined additive and multiplicative impact on units of different phases. *Rev. Bras. Biom.*, São Paulo, v.29, n.1, p.14-24, 2011.

PANDEY, R. Samples allocation for program evaluation: a case of multiplicative effect on units due to developmental initiatives in phased manner. *Folia Forestalia Polonica, Series A*, Raszyn, v.52, n.2, p.83-88, 2010.

PANDEY, R.; VERMA, M. R. Samples allocation in different strata for impact evaluation of developmental programme. *Rev. Mat. Estat.*, São Paulo, v.26, n.4, p.103-112, 2008.

SUKHATME, P. V.; SUKHATME, B. V.; SUKHATME, S.; ASOK, C. *Sampling theory of survey with application*. 3.ed. Ames: Iowa State University Press, 1984. 526p.

Received in 14.04.2011

Approved after revised in 17.06.2011