

SAMPLES ALLOCATION FOR EVALUATION OF DEVELOPMENTAL PROGRAM: A CASE OF COMBINED ADDITIVE AND MULTIPLICATIVE IMPACT ON UNITS OF DIFFERENT PHASES

Rajiv PANDEY¹

- **ABSTRACT:** Evaluation of any program, implemented in various phases for determining existing level requires systematic study of the units selected from the target population. This helps to appreciate existing status of the initiatives and possible intervention for further improvements, if needed. Therefore, inferences based on the selected units should be precise and appropriate. This can be achieved by proper allocation of sampling units from each phase of program for further analysis. This paper discusses the sample allocation problem for the population in which interventions has been made on phased manner and realized impacts are governed by additive and multiplicative effect in respective phases and follows arithmetic-geometric progression. A method of proportional allocation has been proposed by assigning weight based on the impact factor. The mean and variances are also worked out. A comparison is also made based on hypothetical case with standard techniques.
- **KEY WORDS:** Stratified sampling; impact evaluation; sample allocation; geometric; arithmetic progression; weight.

1 Introduction

Developmental programs under the aegis of government or public agency are being in vogue in the developing country. These development initiatives need to be evaluated for effective output and for exploring suitable options for further improvements against ineffectiveness to achieve the intended objectives. The evaluation process explores the existing situations by studying either the total or some representative units/sections of these initiatives. The study of total units is complex and resource intensive, therefore sample evaluation is a better option. Therefore, selection of units for sample study should be efficient to appreciate the actual existing situations. However, keeping in view of large spatial focus of development initiatives implemented in various phases in different time scale at different locations such as poverty alleviation schemes in various districts, literacy drive in various regions, the selection of units for evaluation should considered all sort of possible variation. These variations can be creep due to the nature of population units, the spatial location, phases of implementation and time of implementation.

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

The population under consideration is highly skewed due to lagged and distinct impact arise due to phase implementation at different time period. This signifies the importance of time factor i.e. phases. Therefore, implementation phase may be used as criterion of stratification for sampling. This advocates the suitability of stratified sampling approach for evaluation of program as it naturally provide the number of strata formed; and the stratum boundaries. On another hand, sample allocation within the strata is complex keeping in view of the time dependent impact on the respective population units. That is the initial units may have different impacts and thus the realised response in comparison to the units of the latter stages. This problem may be addressed by attaching suitable weight to the units based on perceived impacts on the units. This makes all units comparable by nullifying the impacts of implementation phase and therefore, minimise the bias on resulting estimates of impact evaluation.

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) . The case of number of strata determination, sample allocation within the strata, strata boundary has been discussed for various designs by many authors in statistical literature (Aoyama, 1954; Dalenius and Hodges, 1959; Ekman, 1959). However, the present design, all information is known apriori, except sample allocation within strata.

On the another hand, due to the implementation of developmental program in phased manner, the units of different strata received different impacts. Therefore, the conventional allocation of sample size in different strata depending on the characteristics of strata will not account the actual imposed variability. Rather, the heterogeneity in the population based on the nature of impact may be utilised for sample allocation within strata by assigning the weights in terms of response factor to the respective strata. The case of additive impacts (Pandey and Verma, 2008) and multiplicative impact (Pandey, 2010) in respective units has been already been discussed.

This paper addresses the problem of sample allocation within strata through integration of weight of response under different phases with assumption that the response in successive phases has combined additive and multiplicative effect. Actual practical examples are not being readily discussed, however, chances of availability of such situation is very high.

2 Sample sizes determination in different strata

The principal of optimum allocation is basis for assigning samples in different strata for apriori scientifically fixed sample size (Sukhatme *et. al.*, 1984). For the present study, let program has been implemented in h (say) phases in N units/villages/sectors. The phase is criterion of stratification. Then

$$N = \sum_{i=1}^h N_i . \quad (1)$$

A total of 'n' sample will be selected randomly for impact evaluation. Then,

$$n = \sum_{i=1}^h n_i . \quad (2)$$

Under, proportional allocation, sample size in different phase/strata will be as follows:

$$n_i = n \frac{N_i}{N} . \quad (3)$$

This, n_i will not serve the purpose keeping in view that the realized impact under different phases will be different due to the time lag in implementation (Pandey and Verma, 2008). This can be addressed by selecting units, in such a way that the impact on all units should be treated identical through assigning weight to each stratum depending on the impact of the programme (Pandey, 2010).

Consider that the impact of the programme is uniformly distributed within each phases and response follows arithmetico-geometric sequence i.e. combined additive and multiplicative in nature with respect to different phases. In other words, if there are h phase, and the impact of last phase (the most recent one) is 'M', and the impact of immediate preceding phase is 'Q' then $Q = (M+D)U$, where U is constant multiplicative factor i.e. improvement factor due to time lag (realized impact) and D is additional additive factor i.e. improvement factor due to time lag (realized impact) between the preceded and succeeded ones. In other words, we can say that the units, which were received the programmes in preceding to last phase, will have $\{(M+D)U - M\}$ times more impact than the units, which receive the last phase of programme with the above assumptions.

Mathematically, consider population of size N containing the N_1, N_2, \dots, N_h units under 'h' phases of developmental program. Therefore based on the analogy of previous para, the total actual impact will be $\{M+(h-1)D\}U^{(h-1)}$, times more than the last phase, for the first phase of N_1 units. Hence, the hypothetical stratum size for first phase may be $\{M+(h-1)D\}U^{(h-1)}N_1$ for comparative purposes with last phase. Similarly, for the second phase with N_2 units, total hypothetical population will be $\{M+(h-2)D\}U^{(h-2)}N_2$. Therefore, analogically, there will be $\{M+(h-i)D\}U^{(h-i)}N_i$ units for the i^{th} phase (i^{th} stratum) with N_i units. There will be $(M)N_h$ units (beneficiaries) in the last phase of implementation of development programme (Table 1).

Table 1 - Formula for allocation of sample size under arithmetico geometric impact evaluation

Stratum number	Stratum population size	Stratum sample size, Say	Impact of programme (impact factor)	Stratum population size due to calibration of impact (hypothetical size)	Estimated stratum weight with calibration of impact	Stratum sample size
1	N_1	n_1	$\{M + (h-1)D\}U^{(h-1)}$	$\{M + (h-1)D\}U^{(h-1)}N_1$	$\frac{\{M + (h-1)D\}U^{(h-1)}N_1}{N_{Hy}} = \omega_1$, Say	$n\omega_1$
2	N_2	n_2	$\{M + (h-2)D\}U^{(h-2)}$	$\{M + (h-2)D\}U^{(h-2)}N_2$	$\frac{\{M + (h-2)D\}U^{(h-2)}N_2}{N_{Hy}} = \omega_2$, Say	$n\omega_2$
3	N_3	n_3	$\{M + (h-3)D\}U^{(h-3)}$	$\{M + (h-3)D\}U^{(h-3)}N_3$	$\frac{\{M + (h-3)D\}U^{(h-3)}N_3}{N_{Hy}} = \omega_3$, Say	$n\omega_3$
4	N_4	n_4	$\{M + (h-4)D\}U^{(h-4)}$	$\{M + (h-4)D\}U^{(h-4)}N_4$	$\frac{\{M + (h-4)D\}U^{(h-4)}N_4}{N_{Hy}} = \omega_4$, Say	$n\omega_4$
...
...
...
$h-1$	N_{h-1}	n_{h-1}	$\{M + D\}U$	$\{M + D\}UN_{h-1}$	$\frac{(M + D)UN_{h-1}}{N_{Hy}} = \omega_{h-1}$, Say	$n\omega_{h-1}$
h	N_h	n_h	M	MN_h	$\frac{MN_h}{N_{Hy}} = \omega_h$, Say	$n\omega_h$
Total	N	n	$\frac{M(1-U^h)}{(1-U)} + UD \frac{(1-hU^{(h-1)} + (n-1)U^h)}{(1-U)^2}$	$N_{Hy} (\text{Say}) = \sum \{M + (h-i)D\}U^{(h-i)}N_i$	$\omega_1 + \omega_2 + \dots + \omega_h = 1$	n

Based on the above logic, the general formula for weight based on the impact and implementation phase is

$$\omega_i = \frac{\{M + (h-i)D\}U^{(h-i)}N_i}{N_{Hy}}, \quad (4)$$

where:

h - number of development programme phase;

i - stratum number, $i = 1, 2, 3, \dots, h$;

N_i - Actual numbers of units (beneficiaries) in the i th stratum;

N_{Hy} - Total numbers of hypothetical units in the population adjusted by the impact or phase factor with product of actual population;

U - Constant multiplicative factor of preceding phase with succeeding one;

D - Constant additive factor of preceding phase with succeeding one;

M - Realised impact of the last phase.

3 Estimation of mean and variance for the proposed method

The unbiased estimator of the population mean $\bar{Y} = \sum_{i=1}^h W_i \bar{Y}_i$ of study variable Y, will be

$$\bar{y}_{st} = \sum_{i=1}^h W_i \bar{y}_i, \quad (5)$$

and, the variance of \bar{y}_{st} is given by

$$V(\bar{y}_{st}) = \sum_{i=1}^h \frac{W_i^2 S_i^2}{n_i} - \sum_{i=1}^h \frac{W_i S_i^2}{N_i}. \quad (6)$$

Analogically, unbiased estimator of population mean under the proposed weight will be

$$\bar{y}_{stm} = \sum_{i=1}^h \omega_i \bar{y}_i, \quad (7)$$

where \bar{y}_{stm} is the mean of the character and ω_i has been defined in equation (4).

Similarly, the variance of \bar{y}_{stm} will be estimated as

$$V(\bar{y}_{stm}) = \sum_{i=1}^h \frac{\omega_i^2 S_i^2}{n_i} - \sum_{i=1}^h \frac{\omega_i S_i^2}{N_i}. \quad (8)$$

4 Proposed proportional sample allocation method

The samples allocation to different strata depends on the size of the strata as well as temporal variation between various strata i.e. phase, which follows arithmetico-geometric sequence. Analogically, it can be derived as follows:

From the i^{th} strata of stratified population

$$n_i \propto N_i . \quad (9)$$

Weight assigning due to program effect i.e. factor due to arithmetico-geometric impact

$$n_i \propto \{M + (h - i)D\}U^{(h-i)} . \quad (10)$$

Based on (9) and (10) we get

$$n_i \propto \{M + (h - i)D\}U^{(h-i)}N_i . \quad (11)$$

This can be rewritten as

$$n_i = k\{M + (h - i)D\}U^{(h-i)}N_i , \quad (12)$$

where k is constant.

Summation on both sides in (12) will result into total sample size for left hand side,

$$\sum_{i=1}^L n_i = n = k \sum_{i=1}^L \{M + (h - i)D\}U^{(h-i)}N_i . \quad (13)$$

However, for proportional allocation with hypothetical population

$$n = kN_{Hy} . \quad (14)$$

Now putting the value of k in Equation 12, the sample size from i^{th} strata will be

$$n_i = \left[\frac{\{M + (h - i)D\}U^{(h-i)}N_i}{N_{Hy}} \right] n . \quad (15)$$

Let, $w_i = \left[\frac{\{M + (h - i)D\}U^{(h-i)}N_i}{N_{Hy}} \right]$, hence

$$n_i = w_i . n . \quad (16)$$

This Equation 16 can be used to allocate sample sizes in different strata for impact evaluation of a development programme, which has been implemented in phases and have arithmetico-geometric impact in successive units.

Properties of proposed estimator

The estimated variance of \bar{y}_{stm} may be rewritten as follows after the substitution of value of ω_i

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

where,

$$\Rightarrow n_i = \omega_i n = \frac{n \{ [M + (h-i)D] U^{(h-i)} N_i \}}{\sum_{i=1}^h \{ [M + (h-i)D] U^{(h-i)} N_i \}}. \quad (17)$$

With substitution of Equation 17 in Equation 8, the estimated variance may be written as

$$V(\bar{y}_{stm}) = \sum_{i=1}^h \frac{n_i}{n^2} S_i^2 - \sum_{i=1}^h \frac{n_i}{n} \cdot \frac{S_i^2}{N_i} = \sum_{i=1}^h \frac{\{ [M + (h-i)D] U^{(h-i)} N_i \}}{\sum_{i=1}^h \{ [M + (h-i)D] U^{(h-i)} N_i \}} \cdot \frac{S_i^2}{n} - \sum_{i=1}^h \frac{\{ [M + (h-i)D] U^{(h-i)} \}}{\sum_{i=1}^h \{ [M + (h-i)D] U^{(h-i)} N_i \}} \cdot S_i^2. \quad (18)$$

To be more simple, substitute $\sum_{i=1}^h \{ [M + (h-i)D] U^{(h-i)} N_i \} = N_{Hy}$ (Say), therefore, the above equation of estimated variance may be written as follows

$$V(\bar{y}_{stm}) = \frac{1}{n} \sum_{i=1}^h \frac{\{ [M + (h-i)D] U^{(h-i)} \}}{N_{Hy}} N_i S_i^2 - \sum_{i=1}^h \frac{\{ [M + (h-i)D] U^{(h-i)} \}}{N_{Hy}} \cdot S_i^2. \quad (19)$$

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

$$V(\bar{y}_{sr}) = \left(1 - \frac{n}{N}\right) \frac{S^2}{n}, \quad (20)$$

where, S^2 is sample variance under SRS.

And variance under conventional proportional allocation method

$$V(\bar{y}_{Prop}) = \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{i=1}^h \omega_i S_i^2 = \frac{1}{nN} \left(1 - \frac{n}{N}\right) \sum_{i=1}^h N_i S_i^2. \quad (21)$$

And variance under Neyman allocation method

$$V(\bar{y}_{Ney}) = \frac{1}{n} \left(\sum_{i=1}^h \omega_i S_i\right)^2 - \frac{1}{N} \sum_{i=1}^h \omega_i S_i^2. \quad (22)$$

Then the proposed method will be efficient in comparison to SRS and stratified sampling with conventional proportional and Neyman allocation, if and only if, the $V(\bar{y}_{sr}) - V(\bar{y}_{stm})$; $V(\bar{y}_{Prop}) - V(\bar{y}_{stm})$; and $V(\bar{y}_{Ney}) - V(\bar{y}_{stm})$ is a positive quantity.

Hence, for comparative purposes, the variance between conventional proportional and proposed proportional allocation method may be written as follows

$$V(\bar{y}_{Prop}) - V(\bar{y}_{stm}) = \frac{1}{nN} \left(1 - \frac{n}{N}\right) \sum_{i=1}^h N_i S_i^2 - \frac{1}{n} \sum_{i=1}^h \frac{\{M + (h-i)D\}U^{(h-i)}}{N_{Hy}} N_i S_i^2 - \sum_{i=1}^h \frac{\{M + (h-i)D\}U^{(h-i)}}{N_{Hy}} S_i^2. \quad (23)$$

And, if finite population correction (fpc) terms ignored then, analogically, the Equation 23 may be written as follows:

$$V(\bar{y}_{Prop}) - V(\bar{y}_{stm}) = \frac{1}{nN} \sum_{i=1}^h N_i S_i^2 - \frac{1}{n} \sum_{i=1}^h \frac{\{M + (h-i)D\}U^{(h-i)}}{N_{Hy}} N_i S_i^2. \quad (24)$$

Equation 24 may be written as follows

$$V(\bar{y}_{Prop}) - V(\bar{y}_{stm}) = \frac{1}{nh} \sum_{i=1}^h S_i^2 - \frac{1}{nN_{Hy}} \sum_{i=1}^h \{M + (h-i)D\}U^{(h-i)} N_i S_i^2. \quad (25)$$

Equation 25 will be a positive quantity, if and only if, the individual coefficients of S_i^2 for first term of left hand side (LHS) is greater than the coefficient of S_i^2 for second term of LHS, i.e.

$$\frac{1}{h} > \frac{\{M + (h-i)D\}U^{(h-i)} N_i}{N_{Hy}} \Rightarrow \sum [\{M + (h-i)D\}U^{(h-i)} N_i] > h \{M + (h-i)D\}U^{(h-i)} N_i.$$

Based on mathematical logic, it can be claim that this will held for each i $\{i = 1(1)h\}$ with the fact that the RHS is the sum of all i 's of LHS, particularly for small number of phases (h). That is

$$V(\bar{y}_{Prop}) \geq V(\bar{y}_{stm}).$$

Based on the similar analogy, it can be concluded that $V(\bar{y}_{Ney}) - V(\bar{y}_{stm})$ is also a positive quantity. Consider, finite population correction (fpc) terms ignored, the difference can be written as follows

$$V(\bar{y}_{Ney}) - V(\bar{y}_{stm}) = \frac{1}{nN^2} \left(\sum_{i=1}^h N_i S_i\right)^2 - \frac{1}{n} \sum_{i=1}^h \frac{\{M + (h-i)D\}U^{(h-i)}}{N_{Hy}} N_i S_i^2. \quad (26)$$

That is, Equation 26 can be written as follows,

$$V(\bar{y}_{Ney}) - V(\bar{y}_{stm}) = \frac{1}{nh^2} (\sum_{i=1}^h S_i)^2 - \frac{1}{nN_{Hy}} \sum_{i=1}^h \{M + (h-i)D\} U^{(h-i)} N_i S_i^2. \quad (27)$$

The value of Equation 27 is sensitive in respect to total number of strata h . However, the difference will be positive, if and only if,

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

With this condition the proposed estimator will be efficient than the Neyman allocation.

Therefore, it can be concluded that the proposed estimator is case sensitive as far as efficiency is concerned with respect to traditional proportional and Neyman allocation method and logically, with the simple random sample too. However, theoretically sounds well than these all.

5 Empirical study

Consider a hypothetical development programme, which is being implemented in five phases in five different clusters of villages. The five clusters were having 20, 30, 40, 50 and 60 number of villages, where programme were launched in five consecutive phases. It was assumed that the development programme follows arithmetic-geometric sequence i.e. having combined additive and multiplicative effects in succeeding villages. In this case, conventional allocation may lead to improper selection keeping in view of high impact in initial villages. The allocation of samples in different phases using conventional proportional allocation and new proportional allocation is reported in Table 2 for comparative analysis.

Conclusion

Mid term impact evaluation of any program is essential for achieving the actual intended purposes. This may achieved by evaluating a part of the program i.e. some of the units. The intrinsic and extrinsic variability in population units must be considered for sample allocation. However, allocation of samples in different strata based on phased implementations needs special attention due to its temporal impact on the units. This may be attained by assigning the appropriate weight based on the characteristics of temporal impacts on the units. This study considers the issue of implementation of developmental program in phased manner for the arithmatico-geometric sequence nature of impacts with a hypothetical example.

Overall, this approach considers, sample size determination based on classical sampling approach based on pre specified margin of error and population variability, however, allocation of sample within strata is based on external influence on the units. This may provide better and theoretical efficient estimate keeping in view of real coverage incorporating the significance of impacts.

Table 2 - Numerical example for five year implemented programme

Stratum number	Population size N	Stratum sample size (proportional allocation, n = 20)	Impact factor (M=1, D=1.5, U=1.5)	Hypothetically stratum population N_{Hy}	Weight ω_i	Sample size for evaluation of programme n_i
1	20	2	34.4375	688.75	$688.75/1853.125 \approx 0.37$	$7.43 \approx 7$
2	30	3	18.5625	556.875	$556.875/1853.125 \approx 0.30$	$6.01 \approx 6$
3	40	4	9	360	$360/1853.125 \approx 0.19$	$3.88 \approx 4$
4	50	5	3.75	187.5	$187.5/1853.125 \approx 0.10$	$2.02 \approx 2$
5	60	6	1	60	$60/1853.125 \approx 0.03$	$0.65 \approx 1$
Total	200	20	67.75	1853.125	1	20

PANDEY, R. Alocação de amostra para avaliação de programa de desenvolvimento: um caso de impactos aditivo e multiplicativo combinados em unidades estratificadas. *Rev. Bras. Biom.*, São Paulo, v.29, n.1, p.14-24, 2011.

- *RESUMO: A avaliação de qualquer programa, implementado em várias etapas para determinar o nível existente, requer estudo sistemático das unidades selecionadas da população alvo. Isso ajuda a apreciar a situação atual das iniciativas e intervenções possíveis para melhorar ainda mais, se necessário. Portanto, as inferências com base nas unidades selecionadas devem ser precisas e adequadas. Isto pode ser conseguido por uma distribuição adequada de unidades amostrais de cada fase do programa, para posterior análise. Este artigo discute o problema de alocação de amostra para a população na qual as intervenções são feitas de forma estratificada para perceber impactos que são regidas pelo efeito aditivo e multiplicativo nas respectivas fases e segue uma progressão aritmético-geométrica. Um método da partição proporcional foi proposto através da atribuição de pesos com base no fator de impacto. As médias e variâncias também são estudadas. A comparação também é feita com base em caso hipotético, com as técnicas convencionais.*
- *PALAVRAS CHAVE: Amostragem estratificada; avaliação de impacto; alocação da amostra; progressão aritmética; progressão geométrica.*

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.
- EKMAN, G. An approximation useful in univariate stratification. *Ann. Math. Stat.*, Beachwood, v.30, p.219-229, 1959.
- 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*. New Delhi: Indian Society of Agriculture Statistics, 1984. 542p.

Received in 19.08.2010

Approved after revised in 05.04.2011