SAMPLES ALLOCATION IN DIFFERENT STRATA FOR IMPACT EVALUATION OF DEVELOPMENTAL PROGRAMME

Rajiv PANDEY[1]
Med Ram VERMA[2]

- *ABSTRACT: Mid term impact evaluation of any program/projects is presently in vogue in developing countries. The allocation of samples in different strata based on phased implementation needs special attention due to its temporal impact on the units. We have considered the problem of allocation of sample size in different strata for impact evaluation of the development programme. We have proposed a method of sample allocation, which can be used if the development programme is implemented in different phases.*
- *KEYWORDS: Stratified sampling; impact evaluation; allocation of samples; weight for sample size.*

1 Introduction

For understanding a phenomenon and locating variables that pattern the events in a certain fashion required a systematic study of relevant information obtained through efficient collection and logical analysis of proper sample size. In this context, the decision about the size of a sample is crucial and important as too large sample implies wastage of resources, and too small sample diminishes the utility of results. Sampling theory, which deals with the properties of the estimates from a statistical sample, provides the solution of it (Cochran 1977). However, for the heterogeneous population the allocation of sample size in different homogeneous strata is crucial for estimation of parameters.

The heterogeneity in the population under study may be intrinsic in nature or may arisen due to some artificial or extrinsic causes. The implementation of developmental program viz., poverty alleviation program in any region in phased manner may be one of the examples. It is due to the realization of impact since start of the project. The units, which are treated in first phase, will have more time to realize the actual effect either linearly or nonlinearly, therefore have more impact logically. Therefore to get the precise estimate of the parameter for impact evaluation, the impact has to be accounted for through assigning temporal weights to these units as per the actual facilitation of the program. If it is not being accounted then the impact evaluation will not be actual and proper. This assigning weight is possible, if the influence pattern of the impact is known

[1] Statistical Branch, Forest Research Institute, Dehradun, Utterakhand, India. E-mail: *rajivfri@yahoo.com*
2 Division of Agricultural Economics & Statistics, ICAR Research Complex for NEH Region, Umroi Road, Umiam (Barapani) Meghalaya Pin 793 103, India. E-mail: *mrverma19@yahoo.co.in*

and generally linear. The lack of this information affects the wisdom of practical decision that will be the made from the estimation.

The above logic will be utilized for allocation of samples for parameter estimation. As the main objective of stratification is to give a better cross-section of the population so as to gain a higher degree of relative precision. To achieve this, formation of strata, number of strata to be constructed, allocation of sample size within each stratum, and analysis of data from a stratified design are critical. Out of these four, the pertinent point for impact evaluation is allocation of sample size in different strata. Conventionally, it can be decided by considering three factors, viz; total number of units in the stratum i.e. stratum size, variability within the stratum, and cost in taking observations per sampling unit in the stratum. However, it becomes more critical, once the required information is time dependent. For example, one may wish to evaluate the impact of a developmental programme during the project period, which has been implemented in phased manner in various villages of a federal state. The number of villages considered for implementation of the programme may also be varied in different phase as it depends on administrative, geographical and many other resources. This makes the situation more complex. The conventional allocation of sample size through assigning weight in different strata depending on the characteristics of strata will not account the variability, which is aroused due to phased wise implementation of programme. In fact, phased wise operation influences the study variable as per operation. Therefore, it is logical to consider it, for weight before the allocation of sample size in the stratum. One way to address this issue is to assign equal importance to all the units of population under study by accounting the phase wise impact based on its time of implementation. This paper attempts to address it through integration of this information for weight estimation.

2 Choice of sample sizes in different strata

The desired information about the population and its precision is based on a stratified sample in stratified sampling, which depends on the size of samples in different strata. The sample size in different strata is being fixed in advance scientifically by the experimenter. The guiding principal in the determination of the n_i, the stratum sample size is to choose them in such a way as to estimate the mean on the desired information with the desired precision for a minimum cost or with maximum precision for a given cost, thus making the most effective use of the resourses available. The allocation of the sample to different strata made in accordance with this principle is called the principle of optimum allocation (Sukhatme et. al. 1984).

The intuitively obvious points for optimum allocation are:

1. The large size of stratum, the large should be the size of the sample to be selected from it.
2. The larger the variability within a stratum, the large should be the size of the sample from it, and
3. The cheaper the cost per unit in a stratum, the larger should be the size of the sample from that stratum.

Various methods are available for allocation of sample in different stratum depending on the characteristics of resources. They are equal allocation, proportional allocation due to Bowley (1926) (provides a self-weighting sample; estimates can be made with greater speed and a higher degree of precision) and Neyman allocation due to Neyman (1934) for fixed cost and fixed variance. However these allocations consider the available information about the population and condition of the experimenter. This sort of allocation is true for the population where the variable under consideration is not receiving/ attracting any external influence and if it so, then it is equal for all units of the population.

3 Proposed method - Effect of development programme on the population units

Symbolically, let there are N villages/units, where operation has been performed in phase manner in h (say) phases. Then

$$N = N_1 + N_2 + N_3 + ...N_h = \sum_{i=1}^{h} N_i \qquad (1)$$

A total of 'n' sample has to be studied for impact evaluation, which has been decided through standard technique. The size of sample in different strata considering the phased operation as stratification criteria will be:

$$n = n_1 + n_2 + n_3 + ...n_h = \sum_{i=1}^{h} n_i \qquad (2)$$

In stratified sampling the sample size for proportional allocation is given by

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

This n_i is not estimated based on the equal impact criterion as the village/units of the first phase will have more impact than the succeeding units. Therefore, theoretically the impact evaluation will not be proper. This may also happen due to the fact that the realized impact under different phases will be different due to the time lag in implementation. Generally the impact of program for the former will be more than the latter phase due to the realized impact of developmental programme for longer duration. To address it, one has to make the units in such a way that the impact on all units should be same. This can be achieved through assigning weight. This weight may be assigned depending on the impact of the programme being implemented. Suppose the impact of the programme is uniformly and equally distributed with respect to different phases i.e. if there are h phase, then the impact of each phase is 1/h (say). In other words, we can say that the units/village, which were received the programmes 3 years ago, if implementation has occurred in the yearly basis is having 3 times more impact than the units/villages, which receive the programme after three years. Under the consideration/assumptions that the impact is linearly additive in phased manner. Mathematically, let us consider

population of size N containing the population units $N_1, N_2, N_3.....N_L$. Suppose a development programme has to be implemented in different phases for different population units. Let $N_1, N_2, N_3.....N_L$ are the beneficiaries of the development programme implemented in first, second, third and L-th year. Suppose that development programme is implemented for h years. Therefore, on the first phase i.e. the first year, if there are N_1 units, then the actual impact will be 'h' times more than the last beneficiaries. Hence, for comparison with the last units, the hypothetical stratum size will be hN_1. In the second phase development programme will provide benefit to N_2 units in a year. Hence total number of beneficiaries will be $(h-1)N_2$. Similarly the number of beneficiaries for the third phase will be $(h-2)N_3$. In general there will be $(h-i+1)N_i$ beneficiaries for the i-th stratum. There will be N_h units beneficiaries in last year of implementation of development programme. This has been represented in Table 1 for more clarity.

4 Estimation of calibrated weights in proposed method

The weight of the sample allocation will not be true for the cases when the variable under study is influenced by some external causes or actions. For example: the impact of poverty reduction programmes on the village of a state. Since the number of beneficiaries of the development progamme depends on the phase of the implementation of the development programme and number of beneficiaries in each stratum per year. Due to the dynamic nature of the population from phase to phase general weight $W_i = N_i / N$ cannot be applied for estimation of sample mean, sample variance and allocation of the sample size. Hence general formula for weight in such problem is given by

$$\omega_i = \frac{(h-i+1)N_i}{N_{Hy}} \tag{4}$$

where

h = number of years for which development programme is implemented

i = stratum number , i=1,2,3,.....h

N_i = numbers of beneficiaries in the i th stratum

N_{Hy} = Sum of total numbers of beneficiaries adjusted by the impact or phase factor with product of stratum population.

5 Estimation of mean and variance for the proposed method

To estimate the population mean $\overline{Y} = \sum_{i=1}^{L} W_i \overline{Y}_i$ of the study variable Y, the estimator used is given by:

$$\overline{y}_{st} = \sum_{i=1}^{L} W_i \overline{y}_i \tag{5}$$

which is unbiased estimator of the population mean.

The variance of $\overline{y}_{st}$ under proportional allocation is given by

$$V(\overline{y}_{st}) = \sum_{i=1}^{L} \frac{W_i^2 S_i^2}{n_i} - \sum_{i=1}^{L} \frac{W_i S_i^2}{N_i} \tag{6}$$

In stratified sampling, before drawing a sample, the sampler has to decide about the allocation of the sample sizes to the strata. Equal, proportional and optimum allocations are well known in sampling literature. In practice any one type of allocation is selected according to the nature of the population and applied to all the strata (Ahsan et al. 2005).

For estimating the mean of the particular character under development programme when the development programme is implemented in different phases i.e. years. We cannot use stratum weight W_i as it does not take into account the temporal variation in development programme. So we have to use modified weight ω_i, which is defined in equation (4). Hence unbiased estimator of population mean is given by:

$$\overline{y}_{stm} = \sum_{i=1}^{L} \omega_i \overline{y}_i \tag{7}$$

where $\overline{y}_{stm}$ is the mean of the character under stratified sampling when modified weights are used.

The variance of $\overline{y}_{stm}$ is given by:

$$V(\overline{y}_{stm}) = \sum_{i=1}^{L} \frac{\omega_i^2 S_i^2}{n_i} - \sum_{i=1}^{L} \frac{\omega_i S_i^2}{N_i} \tag{8}$$

6 Proposed allocation method for impact evaluation of developmental program

Now the question arises how to allocate samples to different strata. The proportional allocation cannot be applied for allocating the samples to strata for the impact evaluation of development programme, because it does not take into account the temporal variation.

So we modified proportional allocation for the impact evaluation of development programme. After the strata are formed, the overall sample size n is allocated to strata using a mixed strategy between two sampling allocation: proportional allocation and optimum allocation. In general proportional allocation is appropriate when different parts of population are proportionally represented in the sample (Buddha Kuisomsiri and Parthanadee, 2008). In usual proportional allocation method number of units selected from each stratum directly depends on the number of units in the stratum. But in modified proportional allocation number of units selected from each stratum depends directly on the number of units in the stratum and the temporal effect. Mathematically we can say that

$$n_i \propto N_i \tag{9}$$

$$n_i \propto (h-i+1) \tag{10}$$

From (9) and (10) we get

$$n_i \propto (h-i+1)\, N_i \tag{11}$$

$$n_i = k\ (h-i+1)\, N_i \tag{12}$$

Taking the summation on both sides in (12) we get

$$\sum_{i=1}^{L} n_i = k \sum_{i=1}^{L} (h-i+1)\, N_i \tag{13}$$

$$n = kN_{Hy} \tag{14}$$

Now putting the value of k in equation (12) we get

$$n_i = \left[\frac{(h-i+1)N_i}{N_{Hy}} \right].n \tag{15}$$

Since from the equation (4) $\omega_i = \dfrac{(h-i+1)N_i}{N_{Hy}}$, hence we can write that

$$n_i = \omega_i .n \tag{16}$$

Equation (16) can be used to allocate sample sizes in different strata for impact evaluation of a development programme.

Now putting the value of n_i in equation (8) we get estimator of variance under modified proportional allocation method.

$$V(\bar{y}_{stm})_P = \sum_{i=1}^{L} \frac{\omega_i S_i^2}{n} - \sum_{i=1}^{L} \frac{\omega_i S_i^2}{N_i} \qquad (17)$$

7 Numerical study

Let us consider that a development programme is being implemented in five years (five phases) in different villages. In the first year development programme is started in 20 villages. In the second year development programme is implemented in another 30 villages. Similarly development programme is being implemented in different 40, 50 and 60 villages in third, fourth and fifth year respectively. If the development programme is being implemented with the same pace and number of beneficiaries increase proportional with the number of years for which the development programme is implemented. The allocation of samples in different phases using proportional allocation and proposed allocation is given in Table 2.

Conclusion

Mid term impact evaluation of any program/projects is presently in vogue in developing countries. The allocation of samples in different strata based on phased implementations needs special attention due to its temporal impact on the units. The accounting of temporal impact for allocation of samples will provide better mid-term evaluation of program.

Acknowledgement

The authors are highly thankful to the Editor and the learned referee for the valuable suggestions, which helped to bring the paper in the present form.

PANDEY, R.; VERMA, M. R. Alocação de amostras em diferentes estratos para avaliar o impacto em programas de desenvolvimento. *Rev. Bras. Biom.*, São Paulo, v.26, n.4, p.103-112, 2008.

- RESUMO: Avaliação do impacto médio de qualquer programa em desenvolvimento esta em uso nos país em desenvolvimento. A alocação de amostras em diferentes estratos, levando-se em conta a fase de implementação, precisa especial atenção devido ao impacto temporal. Foi considerado o problema da alocação do tamanho de amostras em diferentes estratos para avaliar seu impacto no desenvolvimento do projeto. Foi proposto um método para alocação de amostras, que pode ser usado se o desenvolvimento do projeto é implementado em diferentes fases.
- PALAVRAS-CHAVE: Amostragem estratificada, avaliação do impacto, alocação de amostras, ponderação para o tamanho da amostra.

References

AHSAN, M. J. , NAJMUSSEH, A.R.; KHAN, M. G. M. Mixed allocation in stratified sampling. *Aligarh J. Stat.*, Aligarh, v.25, p.87-97, 2005.

BUDDHA KULSOMSIRI, J.; PARTHANADEE, P. Stratified random sampling for estimating billing accuracy in health care systems. *Health Care Manage. Sci.*, New York, v.11, p 41-54, 2008 .

BOWLEY, A. L. Measurements of precision attained in sampling. *Bull. Int. Stat. Inst.*, Amsterdam, v.22, p.1-62,1926.

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

NEYMAN, J. On the two different aspects of the representative methods. *J. R. Stat. Soc.,* Cambridge, , v.97, p.558-606, 1934.

SUKHATME, P. V., SUKHATME, B. V., SUKHATME, S.; ASOKE, C. Sampling theory of survey with application. *Ind. Soc. Agric. Stat.*, New Delhi, India, 1984.

Received in 23.09.2008.

Approved after revised in 03.02.2009.

Table 1 - Formula for allocation of sample size for 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	h	$h\ N_1$	$\frac{hN_1}{N_{Hy}} = \omega_1$, Say	$n\omega_1$
2	N_2	n_2	$h-1$	$(h-1)N_2$	$\frac{(h-1)N_2}{N_{Hy}} = \omega_2$, Say	$n\omega_2$
3	N_3	n_3	$h-2$	$(h-2)N_3$	$\frac{(h-2)N_3}{N_{Hy}} = \omega_3$, Say	$n\omega_3$
4	N_4	n_4	$h-3$	$(h-3)N_4$	$\frac{(h-3)N_4}{N_{Hy}} = \omega_4$, Say	$n\omega_4$
….	….	….	…..	….	….	….
…..	….	….	…..	….	…	….
…..	….	….	…..	…..	…..	…..
$h-1$	N_{h-1}	n_{h-1}	2	$2N_{h-1}$	$\frac{2N_{h-1}}{N_{Hy}} = \omega_{h-1}$, Say	$n\omega_{h-1}$
h	N_h	n_h	1	N_h	$\frac{N_h}{N_{Hy}} = \omega_h$, Say	$n\omega_h$
h	N	n	$\frac{h(h+1)}{2}$	N_{Hy} **(Say)**	$\omega_1 + \omega_2 + ... + \omega_n = 1$	n

Table 2 - Numerical Example for five year implemented programme

Stratum number	Population size N	Stratum sample size (proportional allocation	Impact factor	Hypothetically stratum population N_{Hy}	Weight ω_i	Sample size for evaluation of programme n_i
1	20	2	5	100	1/5	4
2	30	3	4	120	6/25	$120/25 \approx 5$
3	40	4	3	120	6/25	$120/25 \approx 5$
4	50	5	2	100	1/5	4
5	60	6	1	60	3/25	$2.4 \approx 2$
Total	200	20	-	500	1	20