ISSN: 2454-1907
DOI: https://doi.org/10.29121/ijetmr.v5.i1.2018.54

IJETMR

International Journal of Engineering Technologies and Management Research
A Knowledge Repository

STOCHASTIC OPTIMIZATION IN MULTIVARIATE STRATIFIED DOUBLE SAMPLING DESIGN

Ziaul Hassan Bakhshi [*1]
[*1] Associate Professor, Department of Management Studies, JETGI, Jahangirabad, India

Abstract:
This paper deals with optimum allocation of sample size in stratified double sampling when costs are considered as random in the objective function. When costs function are random, by applying modified E. model, objective function is converted into an equivalent deterministic form. A Numerical example is presented to illustrate the computational procedure and the problem is solved by using LINGO Software.

Keywords: *Optimum Allocation; Stratified Double Sampling; Modified E-Model; Chance Constrained Programming.*

Cite This Article: Ziaul Hassan Bakhshi. (2018). "STOCHASTIC OPTIMIZATION IN MULTIVARIATE STRATIFIED DOUBLE SAMPLING DESIGN." *International Journal of Engineering Technologies and Management Research*, 5(1), 115-122. DOI: https://doi.org/10.29121/ijetmr.v5.i1.2018.54.

1. Introduction

When strata weights are unknown in stratified sampling, double sampling technique may be used to estimate them. At first a large simple random sample from the population without considering the stratification is drawn and sampled units belonging to each stratum are recorded to estimate the unknown strata weights. A stratified random sample is then obtained comprising of simple random subsamples out of the previously selected units of the strata. If the problem of non-response is there, then these subsamples may be divided into classes of respondents and non-respondents. A second subsample is then drawn out of non-respondents and an attempt is made to obtain the information. This procedure is called Double Sampling for Stratification (DSS). Okafor (1994) derived DSS estimators based on the subsampling of non-respondents. Najmussehar and Bari (2002) discussed an optimum double sampling design by formulating the problem as a mathematical programming problem and dynamic programming technique is used to solve it.

The double sampling design was first introduced by Neyman (1938) in survey research. Kokan (1963) proposed a nonlinear programming solution in multivariate surveys but did not discuss its applicability to double sampling. Kokan and Khan (1967) described an analytical solution of an allocation problem in multivariate surveys and also discuss its application to double sampling. One of the major problems in double sampling is determining the number of samples required in each phase to give the desired accuracy for the maximum economy. The efficiency of double

ISSN: 2454-1907
DOI: 10.5281/zenodo.1171883

sampling depends on two things: (1) the precision of the mathematical relationship, and (2) the cost of direct measurements compared to indirect estimates.

Stratified sampling design is the most widely used sampling design for estimating the population parameter of a heterogeneous population. it deals with the proportion of estimates constructed from a stratifies random sample and with the optimum choice of the sizes of the samples to be selected from various strata either to maximize the precision of the constructed estimate for a fixed cost or to minimize the cost of the survey for a fixed precision of the estimate. The sample sizes allocated according to either of the above criteria is called an optimum allocation in the sample. The optimum allocation in stratified double sampling is well known for a university population.

The case when sampling variances are random in the constraint has been dealt with Diaz Garcia (2007). Bakhshi and Javed (2009) applied modified E-model for solving the multivariate allocation problem when costs are considered as random in the objective function. Many other authors like, Rao, (1973), Rao, S. S. (1979), Melaku, A. (1986), Kall, P., Wallace, S.W. (1994), Prekopa, A. (1995), Kozak, M. (2005),Varshneyet. al. (2011), Haseen et.al. (2012) Iftekhar S., Ali Q. M., Ahsan M. J., (2014) , Saini M., Kumar A., (2015), other authors discussed many others discussed about double sampling technique for stratification under various situations.

In this paper, a method of optimum allocation for multivariate stratified double sampling is developed. The problems of determining the optimum allocations are formulated as Nonlinear Programming problems (NLPP) in which each NLPP has a convex objective function and a single linear cost constraint. Several techniques are available for solving these NLPPs, better known as Convex Programming Problems (CPP). The problem of optimum allocation is formulated as stochastic nonlinear programming problem (SNLPP) in which costs are treated as random variables , by applying modified E. model , problem is converted into an equivalent deterministic objective function. A Numerical problem is solved by using LINGO Software.

2. Formulation of the Problem

Consider a finite population of size N be stratified into k strata, with N_i (i=1, 2,…, k) units belonging to the i^{th} strata such that $\sum_{i=1}^{k} N_i = N$, which are homogeneous within themselves and whose means are widely different. The strata weights are used in estimating unbiasedly the mean or the total of the character under study. If these weights are not known, the technique of double sampling can be used, which consists of selecting a preliminary sample of n' by SRSWOR to estimate the strata weights and then further selecting a subsample of n units with that n_i (i=1, 2,…, k) units from thei^{th} stratum, to collect information on the character under study such that $\sum_{i=1}^{k} n_i = n$.

Let W_i : $\frac{N_i}{N}$ be proportion of units falling in thei^{th} stratum

ISSN: 2454-1907
DOI: 10.5281/zenodo.1171883

And w_i : $\frac{n_{i'}}{n'}$ be proportion of 1st sample units falling in thei^{th} stratum.

As an estimator of the population mean $\bar{Y}$ we can take
$\bar{y}_{std} = \sum_{i=1}^{k} w_i \bar{y}_i$is an unbiased estimator of the population mean $\bar{Y}$.

Where $\bar{y}_i$ is the sample mean for the study variable in the i^{th}stratum.

The variance of $\bar{y}_{std}$ is

$$V(\bar{y}_{std}) = \sum_{i=1}^{k} \left[w_i^2 + \frac{w_i(1-w_i)}{n'} \right] \frac{S_i^2}{n_i} + \sum_{i=1}^{k} \frac{w_i \left(\bar{Y}_{hj} - \bar{Y}\right)^2}{n'}$$

If we neglect the finite population corrections on the population as well as in the individual strata and further the term $\frac{w_i(1-w_i)}{n'}$ is negligibly small in comparison to w_i^2 [Cochran, (1970)], we obtain

$$V(\bar{y}_{std}) = \frac{\sum_{i=1}^{k} w_i^2 S_i^2}{n} + \frac{\sum_{i=1}^{k} w_i \left(\bar{Y}_{hj} - \bar{Y}\right)^2}{n'}$$

Denoting by $\sum_{i=1}^{k} w_i^2 S_i^2$ by V_n and $\sum_{i=1}^{k} w_i \left(\bar{Y}_{hj} - \bar{Y}\right)^2$by $V_{n'}$,
We get $V(\bar{y}_{std}) = \frac{V_n}{n} + \frac{V_{n'}}{n'}$

In allocating the sample size n to different strata we use Neyman allocation where $\sum_{i=1}^{k} n_i = n$

$n_i$being the sample size in the i^{th} stratum. The cost function is of the form
$C = C_1 n' + C_2 n + C_0$,
Where C_1 is smaller than C_2 and C_0 is the fixed cost.

The problem is to choose n and n' in such a way that the total cost is minimum subject to the tolerance limits on the variance of the various characters.

Thus the optimum allocation problem stated as:

Min. C = $nC_n + n'C_{n'}$
Subject to $\frac{V_n}{n} + \frac{V_{n'}}{n'} = V_j$

Where V_j is the specified variance.

The problem at hand addresses the issue of minimization the total cost subject to a predetermined specified variance requirement in addition to other constraints. In this paper we have discussed unknown cost.

ISSN: 2454-1907
DOI: 10.5281/zenodo.1171883

$$\left.\begin{array}{l} \underbrace{Min.}_{n,n'} \; C = \sum_{i=1}^{m} C_i n_i + C'n' + C_0 \\ Subject\ to \quad \frac{V}{n} + \frac{V'}{n'} = V_j \\ where\ V = \sum_{i=1}^{m} W_i^2 S_i^2 \quad and \quad V' = \sum_{i=1}^{m} W_i \left(\bar{Y}_j - \bar{Y}\right)^2 \\ n_i \geq 2, \quad n' \geq 5 \end{array}\right\} \tag{2.1}$$

For this, we use Modified-E Model approach which is described as given below:

3. Modified E-Model Approach

We have the following non- linear programming problem:

$$\left.\begin{array}{l} \underbrace{Min.}_{n,n'} \; C = \sum_{i=1}^{m} C_i n_i + C'n' + C_0 \\ Subject\ to \quad \frac{V}{n} + \frac{V'}{n'} = V_0 \\ where\ V = \sum_{i=1}^{m} W_i^2 S_i^2 \ and \ V' = \sum_{i=1}^{m} W_i \left(\bar{Y}_j - \bar{Y}\right)^2 \\ n_i \geq 2\,, \ n' \geq 5 \end{array}\right\} \tag{3.1}$$

The costs c_i are assumed to be independently and normally distributed random variables. Then the objective function will also be normally distributed random variables with mean $E\left(\sum_{i=1}^{m} c_i n_i + c'n' + C_0\right)$ and variance $V\left(\sum_{i=1}^{m} c_i n_i + c'n' + C_0\right)$.

If $c_i \sim N(\mu_i, \sigma_i^2)$, then its p.d.f. will be

$$f(c_i) = \frac{1}{\sigma_i \sqrt{2\pi}}\, e^{-\frac{1}{2\sigma_i^2}(c_i - \mu_i)^2}, \quad i = 1, \ldots, m.$$

The Joint distribution of $(c_1, \ldots, c_m)$, will be given by

$$f(\underline{c}') = \frac{1}{\prod_{i=1}^{m} \sigma_i (2\pi)^{m/2}}\, e^{-\frac{1}{2\sigma_i^2}\sum_{i=1}^{m}(c_i - \mu_i)^2}$$

$$E\left(\sum_{i=1}^{m} c_i n_i + C_0\right) = E\left(\sum_{i=1}^{m} c_i n_i\right) + C_0$$

ISSN: 2454-1907
DOI: 10.5281/zenodo.1171883

$$= \sum_{i=1}^{m} n_i\, E(c_i) + n' c_{\mu}^{'} + C_0$$

$$= \sum_{i=1}^{m} n_i \mu_i + n' c_{\mu}^{'} + C_0, \text{ (say)}$$

$$V\left(\sum_{i=1}^{m} c_i n_i + C_0\right) = V\left(\sum_{i=1}^{m} c_i n_i\right) + V(c'n')$$

$$= \sum_{i=1}^{m} n_i^2\, Var(c_i) + n'^2\, \sigma_{c'}^2$$

$$= \sum_{i=1}^{m} n_i^2 \sigma_{c_i}^2 + n'^2\, \sigma_{c'}^2. \quad \text{(Say)} \tag{3.2}$$

By applying modified E-model technique, our objective function will be

$$\left.\begin{aligned} &\min_{n} .C = K_1\, E\left(\sum_{i=1}^{m}(c_i n_i) + c'n' + C_0\right) + K_2 \sqrt{V\left(\sum_{i=1}^{m}(c_i n_i) + c'n' + C_0\right)} \\ &\textit{Subject to} \\ &\qquad \frac{\mathrm{V}}{\mathrm{n}} + \frac{V'}{n'} = V_0 \qquad\qquad i = 1,2,..., m. \\ &\textit{and} \qquad V = \sum_{i=1}^{m} W_i \left(\bar{y}_i - \bar{y}_{std}\right)^2 \quad \textit{and} \quad V' = \sum_{i=1}^{m} W_i^2\, S_i^2 \\ &\qquad n_i \geq 2\ ,\ n' \geq 5 \end{aligned}\right\} \tag{3.3}$$

This is a deterministic objective function. Here K_1 and K_2 are non-negative constants and their values show the relative importance of the expectation and variance of $\left(\sum_{i=1}^{m} c_i n_i + c'n' + C_0\right)$, (see Rao 1979, pp. 599). Thus, the equivalent deterministic problem to the stochastic problem (3.3) can be expressed as follows:

$$\left.\begin{aligned} &\min_{n}\ C = K_1 \left(\sum_{i=1}^{m} n_i \mu_{c_i} + n' c_{\mu}^{'} + C_0\right) + K_2 \left(\sum_{i=1}^{m} n_i^2 \sigma_{c_i}^2 + n'^2\, \sigma_{c'}^2\right)^{1/2} \\ &\textit{Subject to} \\ &\qquad \frac{\mathrm{V}}{\mathrm{n}} + \frac{V'}{n'} = V_0 \qquad\qquad i = 1,2,.., m. \\ &\textit{and} \qquad V = \sum_{i=1}^{m} W_i \left(\bar{y}_i - \bar{y}_{std}\right)^2 \quad \textit{and} \quad V' = \sum_{i=1}^{m} W_i^2\, S_i^2 \qquad i = 1,2,..., m. \\ &\qquad n_i \geq 2\ ,\ n' \geq 5 \qquad\qquad i = 1,2,..., m \end{aligned}\right\} \tag{3.4}$$

ISSN: 2454-1907
DOI: 10.5281/zenodo.1171883

Since the objective function (3.4) is given in terms of the cost expected values and variance of $\left(\sum_{i=1}^{m} c_i n_i + c'n' + C_0\right)$ which are unknown (by hypothesis), then we will use estimators of $E\left(\sum_{i=1}^{m} c_i n_i + c'n' + C_0\right)$ and variance $V\left(\sum_{i=1}^{m} c_i n_i + c'n' + C_0\right)$.

Thus, the optimal equivalent deterministic problem to the stochastic problem is given by

$$\left.\begin{aligned} &\min_{n} C = K_1 \hat{E}\left(\sum_{i=1}^{m} n_i \mu_{c_i} + c'n' + C_0\right) + K_2 \sqrt{\hat{V}\left(\sum_{i=1}^{m} n_i^2 \sigma_{c_i}^2 + n'^2 \sigma_{c'}^2 + C_0\right)} \\ &\textit{Subject to} \\ &\qquad \frac{\mathrm{V}}{\mathrm{n}} + \frac{V'}{n'} = V_0 \qquad\qquad i = 1,2,..,m. \\ &\textit{and} \qquad V = \sum_{i=1}^{m} W_i \left(\bar{y}i - \bar{y}_{std}\right)^2 \ \textit{and} \quad V' = \sum_{i=1}^{m} W_i^2 S_i^2 \qquad i = 1,2,...,m. \\ &\qquad n_i \geq 2 \ , \ n' \geq 5 \qquad\qquad i = 1,2,...,m. \end{aligned}\right\} \tag{3.5}$$

Estimator of $E\left(\sum_{i=1}^{m} c_i n_i + c'n' + C_0\right)` = E\left(\sum_{i=1}^{m} c_i n_i\right) + E\left(c'n'\right) + C_0$

$$= \sum_{i=1}^{m} n_i E(c_i) + n' c_{\mu}^{'} + C_0$$

$$= \sum_{i=1}^{m} n_i \bar{c}_i + n' c_{\mu}^{'} + C_0 . \tag{3.6}$$

Estimator of $V\left(\sum_{i=1}^{m} c_i n_i + c'n' + C_0\right) = \hat{V}\left(\sum_{i=1}^{m} c_i n_i + c'n' + C_0\right)$

$$= \hat{V}\left(\sum_{i=1}^{m} c_i n_i\right) + V\left(c'n'\right)$$

$$= \sum_{i=1}^{m} n_i^2 \hat{V}(c_i) + n'^2 \sigma_{c'}^2$$

$$= \sum_{i=1}^{m} n_i^2 \sigma_{c_i}^2 + n'^2 \sigma_{c'}^2 \tag{3.7}$$

Finally, our equivalent deterministic objective function subject to the variance with given precision will be

ISSN: 2454-1907
DOI: 10.5281/zenodo.1171883

$$\left.\begin{aligned} \min_{n} C &= K_1 \left(\sum_{i=1}^{m} n_i \bar{c}_i + n' c'_{\mu} + C_0\right) + K_2 \sqrt{\left(\sum_{i=1}^{m} n_i^2 \sigma_{c_i}^2\right) + n'^2 \sigma_{c'}^2 + C_0} \\ &\textit{Subject to} \\ &\frac{V}{n} + \frac{V'}{n'} = V_0 \qquad i = 1,2,..,m. \\ &\textit{and} \qquad V = \sum_{i=1}^{m} W_i \left(\bar{y}_i - \bar{y}_{std}\right)^2 \;\; \textit{and} \;\; V' = \sum_{i=1}^{m} W_i^2 S_i^2 \qquad i = 1,2,...,m. \\ &n_i \geq 2 \;,\; n' \geq 5 \qquad i - 1,2,...,m \end{aligned}\right\} \quad (3.8)$$

4. Numerical Illustration

Consider an allocation problem, in which population is stratified into two strata, with the following information:

Table 1: Data for two strata

Stratum *i*	n_i	n'	W_i	S_i^2	c_i
1	40	154	22	5.55	15
2	46	189	21.45	3.4	22

Using the above information, the non-linear programming problem (NLPP) takes the following form:

$$\left.\begin{aligned} &\min_{n\,,\,n'} C = 0.65\left(19n_1 + 24n_2 + 33n' + 25\right) + 0.35\left(25n_1^2 + 16n_2^2 + 36n'^2\right)^{1/2} \\ &\textit{Subject to} \\ &\frac{1.11}{n_1} + \frac{1.032}{n_2} + \frac{1.28}{n'} = 0.211 \\ &n_1, n_2 \geq 2 \;\; \textit{and} \;\; n' \geq 5 \end{aligned}\right\}$$

The above non-linear programming problem (NLPP) is solved by LINGO computer program. Since are sample sizes respectively, it must be an integer. The optimal solution so obtained after 75213 iterations as follows:

$$C = 7241.85, \;\; n_1 = 14 \,,\, n_2 = 12 \;\; n' = 28$$

5. Conclusion

This paper has provided an in-depth study of costs function in multivariate stratified double sampling in which costs are considered as random in the objective function. We have considered the costs parameters as normal random variables and converted the allocation problem as a problem of stochastic non-linear programming problem (SNLPP). An equivalent deterministic

ISSN: 2454-1907
DOI: 10.5281/zenodo.1171883

formulation of the non-linear programming problem (NLPP) was established by using modified E. model. There have been many successful attempts to determine optimum allocation of NLPP in which costs are assumed to be an exact value. But to solve the above problem with random costs by NLPP will be much more complicated. We have used the LINGO software to obtain the optimum solution of random costs constraint within no time. If a large number of subsystems are involved in a deterministic model, suitable nonlinear programming software can be used to solve the problem.

References

[1] Diaz Garcia, J. A., and Garay Tapia, M.M. Optimum allocation in Stratified surveys: Stochastic Programming, *Computational Statistics and Data Analysis*, 51, 2007, 3016-3026.
[2] Haseen, S., Iftekhar, S., Ahsan, M.J. and Bari, A. A Fuzzy Approach for Solving Double Sampling Design in Presence of Non-Response, International Journal of Engineering Science and Technology, 4, 2012, 2542-2551.
[3] Iftekhar S., Ali Q. M., Ahsan M. J. Compromise Allocation for Combined Ratio Estimates of Population Means of a Multivariate Stratified Population Using Double Sampling in Presence of Non-Response" Open Journal of Optimization, 3, 2014, 68-78.
[4] Javaid, S., Bakhshi, Z. H. and Khalid, M. M. Optimum allocation in Stratified Sampling with random costs, *International Review of Pure and Applied Mathematics,* 5(2), 2009, 363-370.
[5] Kall, P., Wallace, S.W. Stochastic programming, Wiley, New York. 1994.
[6] Kokan, A. R. Optimum allocation in multivariate surveys. *Journal of Royal Statistical Society*, A, 126: 1963, 557-565.
[7] Kokan A.R., Khan S.U. *Optimum allocation in multivariate surveys: an analytical solution*. *J R Stat Soc B.*, 29(1), 1967, 115–125.
[8] Kozak, M. On Stratified two stage Sampling: Optimum stratification and sample allocation between strata and sampling stages, *Model Assisted Statistics and Applications*, 1(1), 2005, 23-29.
[9] Melaku, A. Asymptotic Normality of the optimum allocation in multivariate stratified random sampling, *Indian J. Statist*. 48 (Series B), 1986, 224-232.
[10] Najmussehar, and Bari, A.: Double sampling for stratification with sub-sampling the non-respondents: a dynamic programming approach,. The Aligarh Journal of Statistics, 22, 2002, 27-41.
[11] Neyman, J. Contribution to the theory of sampling human populations. Journal of the American Statistical Association, 33, 1938, 101-116.
[12] Okafor, F. C. On double sampling for stratification with sub- sampling the non- respondents. Aligarh J Statist 14, 1994 13–23.
[13] Prekopa, A. Stochastic Programming, Kluwer Academic Publishers, Series Mathematics and its Applications. 1995.
[14] Rao, J. N. K. on double sampling for stratification and analytical surveys, Biometrika, 60, 1973, 125-133.
[15] Rao, S. S. Optimization Theory and Applications, Wiley Eastern Limited. 1979.
[16] Saini M., Kumar A. Optimum Allocation in Stratified Two Stage Design by Using Double Sampling for Multivariate Surveys, Prob. Stat. Forum, 08, 2015, 19-23.
[17] Varshney R., Najmussehar and Ahsan, M.J. An Optimum Multivariate Stratified Double Sampling Design in Non-Response" Optimization Letters, 6, 2011, 993-1008.

*Corresponding author.
E-mail address: bakhshistat@ gmail.com