Allocation of Stratified Random Sample Using Meta Goal Programming

Ebrahim Moussa[1], Ramadan Hamed[2], Asmaa Helmy[3].

[1]*Department of Statistics, Faculty of Commerce, Zagazig University.*

[2]*Department of Statistics, Faculty of Economic and Political Science, Cairo University.*

[٣]*Department of Statistics, Faculty of Commerce, Al-Azhar University.*

Abstract

In this paper, a meta goal programming model (MGP) is introduced to determine the optimum allocation of stratified random sample. The main objectives in this research are using the suggested model which is flexible enough to interactive with the decision maker and find practical solutions, in addition to minimize the estimated variance in the sampling, beside minimize the variance of estimated variance, and also minimize the cost and time of users in the study.

Keywords: Meta Goal Programming, Optimum Allocation, Stratified Random Sample.

1- Introduction

Sampling is the process by which inference is made to the whole by examining only a part. Sampling surveys are conducted on different cultural and scientific aspects. The use of sampling surveys arose from the need to minimize the time and effort that is greatly consumed when using complete enumeration. Moreover, although the cost per observation in sample surveys is higher than in complete enumeration, the overall cost of the sample survey will be much less. Furthermore, sometimes obtaining data by complete enumeration is not possible as in destructive tests such as testing the life of electric bulbs and haematological testing (Som, 1996).

In addition, more comprehensive data can be obtained using sample surveys as it is possible to make use of the highly trained and competent personal or the specialized equipment that are limited in availability. Hence, sample surveys offer more scope and flexibility regarding the types of information that can be collected which are impractical to obtain using complete enumeration. Furthermore, sample surveys can produce more accurate results as opposed to complete enumeration. And this is because the volume of work in surveys that rely on sampling is much less. So, it is possible to employ staff of higher quality and more careful supervision of the processing of results can be provided (Cochran, 1977).

Stratified random sample is one of the most often used sampling schemes. The main problem in stratified random sample is to allocate the total sample size in to different strata. Sample allocation between strata has a big influence on a precision of studied estimators. There are many techniques which are being used for allocation of sample size, such as equal share allocation, proportional allocation and optimum allocation, it is called the classical methods.

In this methods, the allocation to different strata is determined by minimizing the variance of the estimator for a given total cost or minimizing the cost for a given level of precision which measured by the variance of the estimator. These are classical methods sometimes suffer from limitations such as that is inability to optimize several objectives simultaneously, producing non-integer values for the sample sizes in some cases, producing a sample size lager than corresponding stratum size.

Survey sample design in general and sample allocation problems in particular can benefit from a mathematical programming formulation, especially when operational or sample size constraints lead away from straightforward or closed from solutions. The problem of allocating the sample to different strata arises, when applying stratified sampling and has been tackled from different perspectives. Mathematical programming (MP) has been used to solve the multi-variate sample design optimization problem and it also appropriate for a wide range of other problems encountered in sample design (Elsybaey, 2013).

Nonetheless, mathematical programming (MP) has many tools that can overcome these limitations faced by classical methods. Thus, many researchers tried to tackle this problem using mathematical programming approaches.

Many authors used mathematical programming to solve allocation problem in stratified sampling. Some of them are Khan et al., (2003, 2014), Khowaja (2013), Swan (2013), Kozak (2014), Raghav et al., (2014), Mohammad et al., (2016) and many more.

This paper is concerned with using the suggested meta goal programming to represent the problem of optimizing the size sample in stratified random sample.

2- Stratified Random Sample

In stratified random sampling the most important problem faced by the sample survey practitioner is to allocate the total sample size in to different strata. Suppose there is a finite population of size N units divided into L strata of size $N_h, h = 1,2, \dots, L,$ (Swan, 2013)

$$\sum_{h=1}^{L} N_h = N \qquad (1)$$

A sample of size n_h is selected from the h_{th} stratum following any sampling design to observe character y. The total sample size $\sum_{h=1}^{L} n_h = n$ is fixed in advance. Let Y_{hi} be the value of the character y for the j_{th} unit in the h_{th} stratum, $j = 1,2, \ldots, s$.

Define $\bar{Y} = \sum_{h=1}^{L} W_h \bar{Y}_h$, where $W_h = \frac{N_h}{N}$ and

$$\bar{Y}_h = \frac{1}{N_h} \sum_{h=1}^{N_h} y_{hj} \tag{2}$$

Under simple random sampling without replacement an unbiased estimate of population mean $\bar{Y}$ is given by

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h \ , \tag{3}$$

where the sample mean of the h_{th} stratum $\bar{y} = \frac{1}{n_h} \sum_{h=1}^{n_h} y_{hj}$ is an unbiased estimate of $\bar{Y}_h$.

The sampling variance of $\bar{y}_{st}$ is given by

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} \left[\frac{1}{n_h} - \frac{1}{N_h}\right] {W_h}^2 {S_h}^2$$

$$= \sum_{h=1}^{L} \frac{{W_h}^2 {S_h}^2}{n_h} \tag{4}$$

where ${S_h}^2$ is the sample variance computed from the h_{th} stratum.

(2-1) **Methods of Stratified Random Sample Allocation**

This part will introduce the most commonly used classical methods of allocation which are;

(2-1-1) **Equal Share Allocation**

This type of allocation divides the total sample n into equal shares among the L strata in the population,

$$n_h = \frac{n}{L} \qquad h = 1,2, \dots, L. \tag{5}$$

Where n_h is the sample size drawn from stratum h.

Given that the total cost C is fixed and takes the following linear form;

$$c = c_0 + \sum_{h=1}^{L} c_h n_h \tag{6}$$

where c_n the cost per sampling unit in the h^{th} stratum, c is the total budget available and c_0 is the overhead (fixed) cost.

The total sample size will be as follows:

$$n = L\frac{c-c_0}{\sum_h c_h} \tag{7}$$

(2-1-2) Proportional Allocation

Here, the total sample is allocated to the different strata in proportion to the total number of units in strata N_h (i.e.n_h is proportional to N_h),

$$\frac{n_h}{N_h} = \frac{n}{N} \quad or \quad \frac{n_h}{n} = \frac{N_h}{N} \rightarrow \quad n_h = n\frac{N_h}{N} \tag{8}$$

$h = 1,2, \dots, L$. Where N is the total population size.

For a fixed cost, the linear cost function (6) gives the total sample size in proportional allocation as follows (som, 1996):

$$n = \frac{c-c_0}{\sum_h w_h c_h} \tag{9}$$

where;

$$w_h = \frac{N_h}{N} = \frac{n_h}{n}. \tag{10}$$

(2-1-3) Optimum Allocation under Multi-objectives

Neman (1934) proposed a method to determine optimum sample sizes the strata by minimizing $V(\bar{y}_{st})$, subject to $\sum_{h=1}^{L} n_h = n$ which gives,

$$n^1{}_{h\,opt} = n\frac{w_h s_h}{\sum w_h s_h} \tag{11}$$

Another optimum allocation (Ross, 1961) may be derived by maximizing the stability of the estimated variance of the stratified estimate or otherwise by minimizing the variance of the estimated variance of $\bar{y}_{st}$, given by

$$V\left(V\left(\bar{y}_{j(st)}\right)\right) = V\left[\sum_{h=1}^{L} W_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) S_h^2\right] \tag{12}$$

where S_h^2 is the sample variance computed from the h_{th} stratum. For large N_h, $h=1, 2, \ldots, L$,

$$V\left(V\left(\bar{y}_{j(st)}\right)\right) = \sum_{h=1}^{L} \frac{W_h^4 S_h^4}{n_h^3} (\beta_{2h} - 1) \tag{13}$$

where β_{2h} is the coefficient of kurtosis of the character y under study in the stratum. The optimum sample size which minimizes

$V(V(\bar{y}_{st}))$ using usual technique for fixed $\sum_{h=1}^{L} n_h = n$ is given by

$$n_{h\ opt}^{2} = n\frac{w_h s_h (\beta_{2h}-1)^{1/4}}{\sum_h w_h s_h (\beta_{2h}-1)^{1/4}} \tag{14}$$

Now we have two sets of optimum sample sizes $n_{h\ opt}^{1}$ and $n_{h\ opt}^{2}$, that is, one by minimizing the variance of the estimate and another by minimizing the variance of the estimated variance of the estimate for fixed sample size n (Swan,2013). In this paper, we are using mathematical programming to find compromise allocations.

(2-1-3) Using Mathematical Programming in Allocation Sampling

Mathematical Programming has several advantages over classical methods. First, it offers the ability to optimize several objectives simultaneously and it has the benefit of assigning priorities to different objectives. Also, several constraints could be suggested. Second, mathematical programming can guarantee that the optimal allocation has integer sample sizes for the different strata. Third, it can ensure that oversampling does not occur (Sabry, 2012). Oversampling happens when the sample size in one or more strata is larger than the stratum size (Elsybaey, 2013).

Mathematical programming could be used to determine the sample size. The problem of deriving statistical information on population characteristics, based on sample data, can be formulated as an optimization problem in which we wish to minimize the cost of the survey, which is a function of the sample size, size of sampling unit, the sampling scheme, and the scope of the survey, subject to the restriction that the loss in precision is within a certain prescribed limit. Or alternatively, we may minimize the loss in precision, subject to the restriction that the cost of the survey is within the given budget (Arthanari and Yadolah, 1981).

3- Goal Programming

The goal programming (GP) model is one of the well-known multi-objective mathematical programming (MOP) models. This model allows to take into account simultaneously several objectives in a problem for choosing the most satisfactory solution within a set of feasible solutions (Aouni et al., 2005).

The GP technique was first used by Charnes and Cooper in 1960. This solution approach has been extended by Ljiri (1965), Lee (1972), and others. This model allows taking into account simultaneously many objectives while the decision making is seeking the best solution from among a set of feasible solutions. The goal programming is a special type of technique. This technique uses the Simplex method for finding optimum solution of a single or multi-dimensional objective function with a given set of constraints which are expressed in linear form (Sen and Nandi, 2012).

This paper examines the problem of determining an optimum allocation in multivariate stratified random sampling, when the population means of several characteristics are to be estimated. Survey sample design in general and sample allocation problems in particular can benefit from a mathematical programming formulation, especially when operation or sample size constraints lead away from straightforward or closed-form solutions.

The main objective of this paper is to using goal programming model to deal with allocation problem in stratified random sample to obtain the optimization allocation of the sample size. The other objectives in this study are minimize the variance of the estimator of the variable, minimize the variance of the estimated variance of this variable, and minimize the total cost and time that are using in sampling for survey.

In this model, some goals are represented including that the sample cost should not exceed a fixed limit; the time needed for the sampling process in kept within a specific range, and the variances of the estimates do not exceed specific values. Assuming that we have the following goals:

- The total cost of sampling for surveys should not exceed the value C.

- The total time of conducting surveys should not exceed the value T.

- The variance of the estimated variable for the characteristic j under the study, should not exceed the value V_{j1}, $j = 1, \dots, s$. (Where s denotes the total number of the decision variables).

- The variance of the estimated variance for the characteristic j under the study, should not exceed the value V_{j2}, $j = 1, \dots, s$. (Where s denotes the total number of the decision variables).

Using the previous definitions, notation, and the objectives of the model are presented, the goal programming approach can be formulated as,

Find n_h that minimize Z

$$Z = \sum_{i=1}^{k} (dp_i + dn_i), \qquad i = 1,2, \dots, k \qquad (15)$$

Subject to:

$$V_i\left(\bar{y}_{j(st)}\right) + dn_i - dp_i = V_{j1}, \qquad (16)$$

$$V_i\left(V\left(\bar{y}_{j(st)}\right)\right) + dn_i - dp_i = V_{j2}, \qquad (17)$$

$$\sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} R_h n_h + dn_i - dp_i = C, \qquad (18)$$

$$\sum_{h=1}^{L} t_h n_h + dn_i - dp_i = T, \qquad dn_i, dp_i \geq 0, \qquad (19)$$

$$\text{Min}\,(100, N_h) \leq n_h \leq N_h,\ n_h \text{ iteger}, \qquad h = 1,2, \dots, L, \qquad (20)$$

$$\sum_{h=1}^{L} n_h = n, \qquad j = 1, \dots, s. \qquad (21)$$

where, k Total number of goal functions,

s Total number of decision variables,

i Goal function index, $i = 1, 2, \dots, k$,

j Decision variables index, $j = 1, 2, \dots, s$,

dn_i, dp_i Negative and positive deviation variables of the ith goal,

R_h The travel cost per unit in the hth stratum, $h = 1, 2, \dots L$,

$R_h n_h$ The cost of visiting the n_h selected units in the hth stratum approximately.

V_{j1} The prefixed variance of the estimator of the population mean , and V_{j2} the variance of the estimated variance for the population mean for the study, which will be obtained through the solution using single objective model.

As previously mentioned that $V_j(\bar{y}_{i(st)})$ denotes the variance of the estimator of the variable under study, the estimator is $\bar{y}_{st}$ with the restriction on the fixed total sample size n.

4- Meta Goal Programming

In this section, a meta goal programming (MGP) is formulated, which allows the decision maker to establish requirements on different achievement functions, rather than limiting his opinions to the requirements of a single variant. In this sense, this approach could be used as a second stage after a traditional goal programming problem has been solved, and once the decision maker is shown the deviations from his original goals. The proposed approach can be considered like a "sensitivity analysis" of the solutions obtained as well as of the own model structure. This scheme requires a certain interaction with the decision maker in order to adjust the target value of the meta goals, but it is much more flexible than usual goal programming formulations and besides, it lets the decision maker to clarify his knowledge about his actual structure of preferences (Uria. et al., 2002).

Meta goal is considered as a simultaneous cognitive evaluation on the degree of achievements for original decision goals considered in a GP model. The meta goal expressed as the utility function of the model, evaluates undesired deviation of each of the goal function b_i in order to communicate concisely with decision makers the overall status of decision outcomes (Bhargava et al., 2015).

Depending on decision objective that the goal function models, the undesired deviation can be dp_i goal that needs to be equal or less than the target, dn_i goal that needs to be equal or larger than the target, or both dn_i and dp_i goal that needs to be attained exactly. To mathematically model meta goal programming based decision problems, the following notation are introduced to represent the parameters and variables involved in the model.

Let, b_i an appropriate normalizing factor, according to type of problem functions (Caballero et al., 2006), let i be goal function index for every existing priority level, $i = 1,2, \dots, k,$ and let w_i be relative weighting factors assigned to be undesired goal deviations of i^{th} goal function in the priority level, let α_i , β_i be negative and positive deviations of meta goal variant in the i^{th} priority level, let Q_i be target value of meta goal variant in the i^{th} priority level. Every undesired deviation considered dp_i, is normalized by dividing its value by the corresponding target. The normalization is necessary

to ensure all deviations are measured under the same scale.

Using the previous definitions and notation, and using goal programming model in the last section, the suggested meta goal programming model for allocation the stratified random sample can be formulated as follows:

Find n_h that minimize $\sum(\beta_i + \alpha_i), \qquad i \in k.$ (22)

Subject to:

$$V_i\left(\bar{y}_{j(st)}\right) + dn_i - dp_i = V_{j1}, \qquad \text{(goals)} \qquad (23)$$

$$V_i\left(V\left(\bar{y}_{j(st)}\right)\right) + dn_i - dp_i = V_{j2}, \qquad (24)$$

$$\sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} R_h n_h + dn_i - dp_i = C, \qquad (25)$$

$$\sum_{h=1}^{L} t_h n_h + dn_i - dp_i = T, \qquad (26)$$

$$\sum_{i}^{k} w_i \frac{dp_i}{b_i} + \alpha_i - \beta_i = 0\ , \qquad \text{(meta goal)} \qquad (27)$$

$$i \in k,\ \alpha_i \geq 0,\ \beta_i \geq 0, \qquad dn_i \geq 0,\ dp_i \geq 0, \quad (28)$$

$$\text{Min}\,(100, N_h) \leq n_h \leq N_h,\ n_h \text{ iteger}, \quad h = 1,2,\ldots,L, \qquad (29)$$

$$\sum_{h=1}^{L} n_h = n, \qquad j = 1,\ldots,s. \qquad (30)$$

We want to reduce the deviations in meta goal programming.

This paper is concerned with using the suggested meta goal programming (MGP) to represent the problem of optimizing the size sample in stratified random sample. This approach can be more flexible than the usual goal programming models allowing to the decision makers to establish target values not only for the goals but also for another uses the additive property to aggregate the deviational variables of the membership functions to minimize them.

5- Numerical Illustration

Using the data from Egypt Enterprise Survey 2016, which are obtained on Investment Climate Assessment in Egypt, the population were used in this survey are 7186 firms, where representative private sector in Egypt, while the sample are 1827firms, where this sample reflecting 7 different subsectors, 27 governorates. For using this data and making the total sales variable used in this proposed model.

Using $n_h = 824, 288, 121, 282, 103, 127, 82,$ which obtained through the previous study. Some of important calculations were extracted from this surveys on which this proposed method is based, Will be presented show. In **table 1** below,

Table 1

H	Nh	Wh	Sh	B2h	Rhnh	Chnh	thnh
1	3374	0.4695241	329506304	186.9853	11,520.00	57,680.00	960.00
2	996	0.1386028	196840143	135.0614	17,070.00	20,160.00	960.00
3	124	0.0172558	490520032	85.75437	15,984.00	8,470.00	960.00
4	1943	0.2703869	1554189164	235.2463	17,316.00	19,740.00	1,120.00
5	491	0.0683273	103507093	81.6382	20,730.00	7,210.00	1,120.00
6	168	0.0233788	41051767.7	33.8133	27,930.00	8,890.00	1,200.00
7	90	0.0125244	94325067.8	81.9015	31,770.00	5,740.00	1,200.00

The meta goal programming model for this problem using Egypt Enterprise Survey 2016 can be formulated as:

Find n_h minimize $\sum_{i=1}^{4}(\beta_i + \alpha_i) \quad i = 1,2, \dots ,4$ (31)

S.t:

$$\sum_{h=1}^{7} W_h^2 S_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) + dn_1 - dp_1 = 5.59529\text{E+}14 \quad (32)$$

$$\sum_{h=1}^{7} \frac{W_h^4 S_h^4}{n_h^3} (\beta_{2h} - 1) + dn_2 - dp_2 = 3.25943\text{E}+29 \quad (33)$$

$$\sum_{h=1}^{7} c_h n_h + \sum_{h=1}^{7} R_h n_h + dn_3 - dp_3 = 270{,}210 \quad (34)$$

$$\sum_{h=1}^{7} t_h n_h + dn_4 - dp_4 = 7{,}520 \quad (35)$$

$$\frac{dp_1}{5.59529\text{E}+14} + \frac{dp_2}{3.25943\text{E}+29} + \frac{dp_3}{270{,}210} + \frac{dp_4}{7{,}520} + \alpha_i - \beta_i = 0 \quad (36)$$

$$Min\ (100, N_h) \le n_h \le N_h,\ n_h \text{ iteger}, \qquad j=1 \quad (37)$$

$$\sum_{h=1}^{L} n_h = 1827 \qquad h = 1,2,\dots,7, \quad (38)$$

$$\alpha_i, \beta_i \ge 0, \qquad dn_i, dp_i \ge 0 \quad (39)$$

In this paper used the software GAMS (General Algebraic Modeling System) models, calls GAMS to solve these models, and then shows the final solutions for these models. we obtained the following solution $n_1 = 808, n_2 = 100,\ n_3 = 100,\ n_4 = 529,\ n_5 = 100,$ $n_6 = 100, n_7 = 90$. Table 2 compares the result which obtained from proposed method with the observed value which obtained from previous study using others method which different of the proposed method.

Table 2

Objectives	The observer value	The optimum value	Reduced %
$V(\bar{y}_{j(st)})$	5.59529E+14	2.72706E+14	51%
$V\left(V(\bar{y}_{j(st)})\right)$	3.25943E+29	4.96242E+28	84%
Total cost	270210.00	267794.0142	0.8%
Total time	7520.00	7518.408	0.02%

6- Conclusions

The suggested meta goal programming method is used to obtained to the optimum allocation of stratified random sample, with the achievement of the others goals sought the researcher in this study of reducing the estimated variance, variance of the estimated variance, and also reducing the cost and time in the sample. By applying the proposed model to real data for Egypt Enterprise Survey 2016 we have come up with the same time and cost we can have a more accuracy estimation. Hence, this creates a mathematical programming problem with contradicting goals. Thus, the use of the suggested meta goal programming approach would be essential in this case.

References

1. Arthanari, T.S., and Dodge, Y. (1981): Mathematical Programming in Statistics. John Wiley and Sons, New York.

2. Aouni, B., Abdelaziz, F., and Martel, J. (2005): Decision Makers Preferences Modeling in the Stochastic Goal Programming, European Journal of Operation research 162-610-618.

3. Bhargava, A.K., Singh, S.R., and Bansal, D. (2015): Production Planning Using Fuzzy Meta Goal Programming Model, Indian Journal of Science and Technology, Vol8 (34) /70706.

4. Caballero, R., Ruiz, F., Uria, M., Romero, C. (2006): Interactive Meta Goal Programming, European Journal of Operation Research, 175,135-154.

5. Cochran, W. G. (1977): Sampling Techniques, John Wiley and Sons, New York.

6. Elsybaey, R. (2013): Use of Mathematical Programming in Sample Surveys, NTTS, FULL PAPER-77.PDF.

7. Khan, M.G.M., Khan, E.A., Ahsan, M.T. (2003): An Optimal Multivariate Stratified Sampling Design Using Dynamic Programming, The university of the South Pacific, Hamdard University and Ali Garth Muslim University.

8. Khan, S.U., Muhammad, Y.S., and Afgan, N. (2014): Multi-Objective Compromise Allocation Stratified Sampling in the Presence of Non-Response Using Quadratic Cost Function, Quaid-1-Azam University, Islamabad, Vol. 5.No. 13.

9. Khowaja, S., Ghufran, S., Ahsan, M.J. (2013): Multi Objective Optimization for Optimum Allocation in Multivariate Stratified Sampling with Quadratic cost, University Laval.

10. Kozak, M. (2014): On Sample Allocation in Multivariate Surveys, University of Otago.

11. Muhammad, Y.S., Hussain, I., Shoukrt, A.H. (2016): Multivariate Multi Objective Allocation in Stratified Random Sampling a Game theoretic Approach, Doi, 10.1371/Journal Phone.167705.

12. Raghava, Y.S., Ali, I., Bari, A. (2014): Multi Objective Nonlinear Programming Problem Approach in Multivariate Stratified Sample Surveys in the Case of Case of non-response, the University of Melbourne of Libraries.

13. Ross, A. (1961): Variance Estimates in Optimum Sample Designs, J. Amer. Stat. Assoc, 56,135-142.

14. Sabry, D. M. (2012): A Mathematical Programming Approach Stratified Random Sampling, Unpublished M. Sc. Thesis, Department of Statistics, Cairo University, Cairo, Egypt.

15. Sen, N., Nandi, M. (2012): Goal Programming, its Application in Management Sectors-Special Attention in to Plantation Management A review, International Journal of Scientific and Research Publications, Issn,2250-3153.

16. Som, R.K. (1996): Practical Sampling Techniques, 2nd Edition,Marcel Dekker, Inc., New York.

17. Swain, A.K.P.C. (2013): A Note on Optimum Allocation in Stratified Random Sampling, Utkal University, Bhubanewar -751004.

18. Urıa, M.V.R., Caballero, R., Ruiz, F., Romero, C. (2002): Decision Aiding Meta Goal Programming, European Journal of Operational Research, 136, 422–429.