Behavior of Mathematical Goal Programming For Determining Optimum Stratum Boundaries

Fatma S. Abo- El.Hassan[1], Ramadan Hamid[2], Elham A.Ismail[3] and Safia M. Ezzat[4]

[1,3&4]Faculty of Commerce, Al-Azhar University (Girls' Branch Cairo) ,Egypt ;fatmasayed85@azhar.edu.eg,[2]Faculty of Economics and Political Science, Cairo University, Egypt .

Abstract

Stratified random sampling is an effective sampling technique for estimating the population characteristics. The determination of strata boundaries and the allocation of sample size to the strata is one of the most critical factors in maximizing the precision of the estimates. Most surveys are conducted in an environment of severe budget constraints and a specific time is required to finish the survey. So cost and time are especially very important objectives of most surveys thus they are necessitating to be under consideration. The study suggested Mathematical goal programming model for determining optimum stratum boundaries for an exponential study variable under multiple objectives model when cost and time are under consideration. To evaluate the performance for the suggested model for the exponential distribution a numerical example is presented. The results of the suggested mathematical goal programming are satisfying.

Key words: Stratified Random sampling, Optimum stratum boundary, Exponential distribution, Mathematical Goal programming, Time , Cost.

1. Introduction and literature Review:

In stratified random sampling. The basic idea is that the internally strata units should be as homogeneous as possible, that is, stratum variances should be as small as possible.

The equations for determining the optimum stratum boundaries was first provided by Dalenius 1950. Khan et al (From 2002 to 2009) studied the optimum strata width as a Mathematical Programming Problem that was solved using the dynamic programming technique. The study concerned with variables which follows triangular, uniform, exponential, normal, right triangular, Cauchy and power distribution. When the study variable has a pareto frequency distribution, Rao et al.(2014) suggested a procedure for determining optimum stratum boundary and optimum strata size of each stratum. Fonolahi and Khan(2014) presented a solution to evaluate the optimum strata boundaries When the measurement unit cost varies throughout the strata, when the variable distribution is exponentially distributed. Reddy et. al. (2016) solved the same problem when multiple survey variables are under consideration. Danish et al. (2017) presented optimum strata boundaries as a non-linear programming problem when the cost per unit varies throughout the strata. Reddy et. al. (2018) formulated the stated problem under Neyman allocation. Where the auxiliary variables follow Weibull distributions. Danish and Rizvi (2019) suggested a non-linear programming model to determine optimum strata boundaries for two auxiliary variables. Reddy and Khan (2020) implemented the problem of optimum stratum boundary for various distributions using R package.

The aim of this study is to determine optimum stratum boundary (OSB), Optimum sample size, Optimum cost and Optimum time when the study used one variable as basis for stratification.

2. Optimum Stratum Boundaries Model:

Let the study variable u from population stratified into J strata and $\overline{U}$ is the estimate of population mean. Let u_0 be the smallest and u_J largest values of the stratification variable u. To determine the intermediate stratum boundaries from the smallest to the largest then the variance of stratified sample mean $equal\ \sum_{h=1}^{J} w_h \bar{u}_h$ under proportional allocation,

$$V(\bar{u}_{st})_{pro} = \frac{1}{n}\sum_{h=1}^{J} W_h S_h^2 - \frac{1}{N}\sum_{h=1}^{J} W_h S_h^2 \qquad (1)$$

is small as possible , where $\bar{u}_h$ is the sample mean in stratum h $, h = 1,2, \dots, J$, W_h is the proportion of population units in stratum h and S_h^2 is the variance of stratum for variable u in the h^{th} stratum and n is the sample size chosen from population N. If the finite population correction (fpc) is ignored, then the minimization of variance given by (1) reduces to minimization of

$$\sum_{h=1}^{J} W_h S_h^2 \qquad (2)$$

The problem of determining OSB is to find $J-1$ intermediate points in the interval $|u_0, u_J|$, let the distance between the smallest and largest values of the stratification variable u to be equal

$$u_J - u_0 = q \qquad (3)$$

If the study variable u has a defined frequency function $f(u)$, and the boundaries of the h^{th} stratum are (u_{h-1}, u_h), then

$$W_h = \int_{u_{h-1}}^{u_h} f(u)\,du \qquad (4)$$

$$S_h^2 = \frac{1}{W_h}\int_{u_{h-1}}^{u_h} u^2 f(u)\,du - \mu_h^2 \qquad (5)$$

Where,

$$\mu_h = \frac{1}{W_h}\int_{u_{h-1}}^{u_h} uf(u)\,du \qquad (6)$$

using (4) , (5) and (6) , $W_h S_h$ in (2) can be represented as a function of u_h and u_{h-1}. i.e $f_h(u_h, u_{h-1}) = W_h S_h$. Thus the objective function is to obtain $u_1 \le u_2 \le \cdots \le u_{J-2} \le u_{J-1}$. That is adequate to the MPP:

Minimize $\sum_{h=1}^{J} f_h(u_h, u_{h-1})$

Subject to $u_0 \le u_1 \le u_2 \le \cdots \le u_{J-2} \le u_{J-1} \le u_J$

Let $q_h = u_h - u_{h-1} \ge 0$ denotes the width of the h^{th} $(h = 1, 2, \ldots, J)$ stratum. With above definition (3) is expressed as

$$\sum_{h=1}^{J} q_h = \sum_{h=1}^{J} (u_h - u_{h-1}) = u_J - u_0 = q$$

For k^{th} point

$$y_k = y_0 + q_1 + q_2 + \cdots + q_k = y_{k-1} + q_k \qquad (7)$$

As a result, determining OSB is the same as determining OSW (optimal stratum width) as MPP:

Minimize $\sum_{h=1}^{J} f_h(q_h, u_{h-1})$

Subjec to

$$\sum_{h=1}^{J} q_h = q \qquad (8)$$

And $q_h \ge 0 , h = 1,2, \ldots, J$

When $h = 1$ the function $f_1(q_1, u_0)$ transforms into a function in q_1 only where u_0 is known. Moreover When $h = 2$ the function $f_2(q_2, u_1) = f_2(q_2, u_0 + q_1)$ transforms into a function in q_2 only where u_1 is known.

As a result, the MPP can be written as a function in q_h as:

Minimize $\sum_{h=1}^{J} f_h(q_h)$

Subject to

$$\sum_{h=1}^{J} q_h = q \qquad (9)$$

And $q_h \geq 0, h = 1,2,\ldots,J$

3. The Suggested Mathematical Goal Programming Model:

The suggested Mathematical goal programming model for evaluating OSB and optimum sample size allocation to the strata was presented in this section.

The suggested mathematical goal programming constraints are as follows:

1- The aggregate of the optimum stratum width be equal to the distribution's range.
2- The cost (not exceed a fixed limit according to budget of survey) was added to the model as objective constrain need to minimize.
3- The time is another important constraint which needed for the sampling process within a specific range.

To optimally determine stratum boundary, allocate the sample to the different strata for variable u defined in [a,b] the problem is to partion u into J strata such that $a = u_0 \leq u_1 \leq \cdots \leq u_{J-1} \leq u_J = b$, let

$u_J - u_0 = q$, define $q_h = u_h - u_{h-1}$ thus the required stratification points is $u_h = u_{h-1} + q_h$.

The suggested Goal Programming (GP) approach can be formulated as, find q_h, u_h, n_h , c_h and t_h which:

Minimize

$$\sum_{i=1}^{k}(dp_i + dn_i) \qquad i = 1,2,3 \qquad (10)$$

Subject to

$$\sum_{h=1}^{J} \frac{W_h^2 S_h^2}{n_h} + dn_1 - dp_1 = v \qquad (11)$$

$$\sum_{h=1}^{J} c_h n_h + dn_2 - dp_2 = C \qquad (12)$$

$$\sum_{h=1}^{J} t_h n_h + dn_3 - dp_3 = T \qquad (13)$$

$$\sum_{h=1}^{J} q_h = q \qquad (14)$$

$$u_h = u_{h-1} + q_h \qquad (15)$$

$$\sum_{h=1}^{J} n_h = n \qquad ,h = 1,2,\dots,J \qquad (16)$$

$$, q_h \geq 0 \ , 1 \leq n_h \leq N_h \ , \ dp_i, dn_i \geq 0$$

Where , k Total number of goal functions,

n_h: Sample size of the hth stratum

$n = \sum_{h=1}^{J} n_h$: Total sample size

c_h: per unit cost of the h^{th} stratum

C total cost

t_h: time per unit of the h^{th} stratum

T total time

v prefixed variance of the estimator of the population mean

dp_i, dn_i positive and negative deviation variables of the i^{th} goal,

$(i = 1,2,3)$ is goal function index where first goal is to minimize $V(\bar{u}_{st})$, second and third goals are to minimize cost and time of collecting data per unit in each stratum respectively,

$\sum_{h=1}^{J} \frac{W_h^2 S_h^2}{n_h} = V(\bar{u}_{st})$ if the finite population correction is ignored, N_h: Stratum size of the h^{th} stratum.

4. Numerical example:

This section concerned with the numerical example for the suggested Mathematical goal programming model, the numerical example take the following step:-

1. The study variable which used followed exponential distribution because of its simple mathematical form as an application of the idea of a multi-objective model for obtaining optimum stratum boundary and allocation the sample into different strata.

   Let the variable under study (u) follows an exponential distribution with parameter > 0 , that is

$$f(u) = \begin{cases} \theta e^{-\theta u} , u \geq 0 \\ 0 \quad otherwise \end{cases} \qquad (17)$$

By using (4), (5),(6) and (17) the term W_h and S_h^2 can be expressed as

$$W_h = e^{-\theta u_{h-1}}\left(1 - e^{-\theta q_h}\right) \qquad (18)$$

$$, \quad S_h^2 = \frac{\left(1 - e^{-\theta q_h}\right)^2 - (\theta q_h)^2 e^{-\theta q_h}}{\theta^2 (1 - e^{-\theta q_h})^2} \qquad (19)$$

2. To determine the OSB and optimum allocation into sample strata which result in minimum possible variance of the estimator so the study suggested mathematical programming to calculate the variance and applied the value as initial value where $v = .002$, $v = .001, v = .0006, v = .0004$ when $J = 2, J = 3, J = 4$ and $J = 5$ respectively.
3. The suggested Mathematical goal programming which using for determining the OSB substituting values of $\theta = 1.057$ which chosen arbitrarily when sample size $n = 100$ followed exponential distribution with $u_0 = 0$, $u_J = 6.459$ and $q = 6.459$. Where θ is the chosen exponential distribution parameter and (u_0 ,u_J) are the chosen observation of smallest and largest values of stratification variable u, q is the different between largest and smallest . The study applied the suggested model when the fixed value of cost =12000 , the specific rang of time =150 are chosen arbitrary.
4. T

he suggested Mathematical goal programming is applied several times when $J = 2, J = 3, J = 4 \; and \; J = 5$ to be sure that the performance for the suggested approach led to satisfied results when cost and time are under consideration.

Using (18) and (19) the suggested goal programming model (10-16) when the study variable u is given by (17), can be formulated as:

Minimize

$$\sum_{i=1}^{k}(dp_i + dn_i) \qquad i = 1,2,3 \qquad (20)$$

Subject to

$$\left\{\sum_{h=1}^{J}\frac{1}{n_h}\left\{e^{-\theta u_{h-1}}\left(1-e^{-\theta q_h}\right)*\frac{\left(1-e^{-\theta q_h}\right)^2-(\theta q_h)^2e^{-\theta q_h}}{\theta^2(1-e^{-\theta q_h})^2}\right\}\right\}$$
$$+dn_1-dp_1 = v \qquad (21)$$

$$\sum_{h=1}^{J} c_h n_h + dn_2 - dp_2 = C \qquad (22)$$

$$\sum_{h=1}^{J} t_h n_h + dn_3 - dp_3 = T \qquad (23)$$

$$\sum_{h=1}^{J} q_h = q \qquad (24)$$

$$u_h = u_{h-1} + q_h \qquad (25)$$

$$\sum_{h=1}^{J} n_h = n \qquad ,h = 1,2,\dots,J \qquad (26)$$

$$, q_h \geq 0 \ , \ 1 \leq n_h \leq N_h \ , \ dp_i, dn_i \geq 0$$

Solving the suggested goal programming model (20-26) by using a GAMS program.

Table (1): Results for OSB, Optimum sample size of the variance function for exponential distribution when $J = 2$,$J = 3$,$J = 4$,$J = 5$ respectively.

No. of strata (J)	Optimum strata width OSW (q_h)	Optimum strata boundary OSB (u_h)	Sample size (n_h)	C_h	T_h	Optimum value of variance
2	1.175	1.175	48	5474.75	8.2	.00038
	5.284	6.459	52	6520.8	140.4	
3	0.711	0.711	32	3705.6	12.8	.000099
	1.161	1.871	33	3851.1	13.2	
	4.588	6.459	35	4438	122.5	
4	0.512	0.512	24	2805.6	9.6	.000029
	0.705	1.216	24	2812.8	9.6	
	1.144	2.361	25	2950	10	

No. of strata (J)	Optimum strata width OSW (q_h)	Optimum strata boundary OSB (u_h)	Sample size (n_h)	C_h	T_h	Optimum value of variance
	4.098	6.459	27	3429	121.5	
5	0.400 0.508 0.698 1.127 3.726	0.400 0.908 1.607 2.733 6.459	19 20 20 20 21	2236 2356 2360 2376 2664.9	27.9 29.4 29.6 29.8 33	.0000072

Table (1) showed the results when suggested goal programming model applied when the study variable followed exponential distribution. The suggested Mathematical goal programming model for determining OSB and optimum allocation of sample size to the strata are repeated when $J = 2, J = 3, J = 4 and\ J = 5$ to be sure the results are consistent. According to each number of strata $J = 2, J = 3, J = 4 and\ J = 5$, the suggested model each time calculate the optimum staratum width and optimum stratum boundary in satisfactory way. The new minmum value if variance is 0.00038,0.000099,0.000029 and 0.0000072 respectively which are less than the initial values which calculated before. Sample size is divided in satisfactory way according to the number of strata. The suggested model divided the time and cost and reducing them as much as possible.

5. Conclusion:

The study suggested Mathematical goal programming model to determine the optimum strata boundary by univariate variable in multiobjective problem.

Where,

- The study variable (u) which used follows exponential distribution.
- Applying the values v as initial value of variance of the estimator (which calculated using Khan and Sharama (2015)) where $v = .002, v = .001, v = .0006, v = .0004$ $when\ J = 2, J = 3, J = 4\ and\ J = 5$ respectively.
- Suggested Mathematical goal programming model for determining OSB and optimum allocation of sample size to the strata.
- The cost (not exceed a fixed limit according to budget of survey) was added to the model as objective constrain need to minimize.
- The time is another important constraint which needed for the sampling process within a specific range.

The results for the suggested Mathematical goal programming model are:

1. The suggested Mathematical goal programming model for determining OSB and optimum allocation of sample size to the strata are repeated when $J = 2, J = 3, J = 4 and\ J = 5$ to be sure the results are consistent.
2. According to each number of strata $J = 2, J = 3, J = 4 and\ J = 5$, the suggested model each time calculate the optimum staratum width and optimum stratum boundary in satisfactory way.

3. The new minimum value if variance is 0.00038, 0.000099, 0.000029 and 0.0000072 respectively which are less than the initial values which calculated before.
4. Sample size is divided in satisfactory way according to the number of strata.
5. The suggested model divided the time and cost and reducing them as much as possible.

References

1. Danish, F., Rizvi, S.E.H., Jeelani, M.I. and Reashi, J.A. (2017)"Obtaining strata boundaries under proportional allocation with varying cost of every unit". Pakistan Journal of Statistics and Operation Research, 13(3), 567-574.
2. Danish, F., Rizvi, S.E.H (2019)" Optimum Stratification by Two Stratifiction Variables Using Mathematical Programming" Pakistan Journal of Statistics Vol. 35(1), 11-24.
3. Fanolahi, A.V. and Khan, M.G.M. (2014)" Determing the optimum strata boundaries with constant cost factor". Conference: IEEE Asia-Pacific World Congress on Computer Science and Engineering (APWC), At Plantation Island, Fij.
4. Khan, M. G. M., Nand, N., & Ahmad, N. (2008). "Determining the optimum strata boundary points using dynamic programming". Survey Methodology, 34(2), 205-214.
5. Khan, M. G. M., Ahmad, N. and Khan, Sabiha (2009)."Determining the optimum stratum boundaries using mathematical programming". J. Math. Model. Algorithms, Springer, Netherland, 8(4): 409- 423.
6. Khan, E.A., Khan, M.G.M. and Ahsan, M.J. (2002). "Optimum stratification: A mathematical programming approach". Culcutta Statistical Association Bulletin, 52 (special), 205-208.
7. Khan, M.G.M., Sehar, N. , and Ahsan, M.J. (2005). "Optimum stratification for exponential study variable under Neyman allocation". Journal of Indian Society of Agricultural Statistics, 59(2), 146-150.

8. Rao, D.K. , Khan, M.G.M. and Reddy, K.G. (2014)."Optimum stratification of a skewed population". International journal of mathematical, computational, natural and physical engineering, 8(3):497-500.
9. Reddy, K. G., M. G. Khan, and D. Rao (2016). "A procedure for computing optimal stratum boundaries and sample sizes for multivariate surveys". JSW 11(8), 816–832.
10. Reddy K.G., Khan M.G.M., Khan S. (2018) "Optimum strata boundaries and sample sizes in health surveys using auxiliary variables". PLoS ONE 13(4): e0194787.
11. Reddy KG and Khan MGM (2020)" stratifyR: An R Package for optimal stratification and sample allocation for univariate populations " Australian & New Zealand Journal of Statistics 62 (3), 383-405