مجلة التجارة والتمويل

/https://caf.journals.ekb.eg

كلية التجارة – جامعة طنطا

العدد : الرابع

ديسمبر ٢٠٢١

Multi-objective Mathematical Programming Model for Optimum Stratification in Multivariate Stratified Sampling

Ramadan Hamid
Professor of Statistics and Operation Research
Faculty of Economics and Political Science,
Cairo University, Egypt.
Research Professor Social
Research Center, AUC
Email: ramadanh@aucegypt.edu

Elham A. Ismail
Professor of Statistics and Operation Research
Statistics DEPT
Deen of Faculty of commerce - AL Azhar University Girl's Branch
Nasr City, Cairo, Egypt
Email: comm.g.dean@azhar.edu.eg

Safia M. Ezzat
Teaching in Statistics DEPT
Faculty of commerce - AL Azhar University Girl's Branch
Nasr City, Cairo, Egypt
Email: safiahamed@azhar.edu.eg

Fatma S. Abo- El.Hassan
Teaching Assistant in Statistics DEPT
Faculty of commerce - AL Azhar University Girl's Branch
Nasr City, Cairo, Egypt
Mobile: +20-10-08816160
Email: fatmasayed85@azhar.edu.eg

Multi-objective Mathematical Programming Model for Optimum Stratification in Multivariate Stratified Sampling

Abstract

Obtaining accurate and reliable data when conducting surveys requires a long time, budget and large workforce so cost and time are especially very important objectives of most surveys thus they are necessitating to be under consideration. Most surveys are conducted in an environment of severe budget constraints and a specific time is required to finish the survey. The data is used to estimate the parameters, determine the characteristics of the population under study, and the possibility of prediction and decision-making.

The study suggested mathematical goal programming model does not depend on a lot of data or surveys, but depends on the distribution, parameters of any population through previous data for the community. A multi-objective model depend on time and cost used to evaluate the performance of the suggested mathematical goal programming model for exponential distribution.

The suggested Mathematical goal programming model used for getting Optimum Stratum Boundary and allocate sample size into different strata using two auxiliary variables as stratification factors. A numerical example is presented and the results of the suggested mathematical goal programming are satisfying.

Key words: Multivariate Stratified Random sampling, Optimum Stratum Boundary, Exponential distribution, Mathematical Goal programming, Time , Cost.

1. Introduction

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 (2002,2005,2008 and 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.

In this study auxiliary variable(s), which can be historical data, have also been utilized to improve the precision of study variable estimations. When the auxiliary variable's frequency distribution is known.

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

2. Optimum Stratum Boundaries Model

Let the target population consisting of "N" units be stratified into I strata based on p auxiliary variables $x_1, x_2, \dots., x_p$ and estimation of the mean study variable is of interest.

Consider that the study variable has the regression model of the form

$$z = \lambda(x_1, x_2, \dots., x_p) + \varepsilon \tag{1}$$

Where, $\lambda(x_1, x_2, \dots., x_p)$ is a linear or (non linear function) of x_r $(r = 1,2, \dots. p)$ and ε is an error term such that $E(\varepsilon \backslash x_1, x_2, \dots., x_p) = 0$ and $var(\varepsilon \backslash x_1, x_2, \dots., x_p) = \Psi(x_1, x_2, \dots., x_p) > 0$ for all x_r .

For simplicity, two auxiliary variables are taken as the basis of stratification with one study variable according to Danish and Rizvi (2019). They divided the whole population into $I * J$ strata on the basis of two auxiliary variables say y and q such that the number of units in the $(i, j)^{th}$ stratum is N_{ij} . a sample of size "n" is to be drawn from the whole population and suppose that the allocation of sample size to the $(i,j)^{th}$ stratum is n_{ij} $(i = 1,2,\dots,I; j = 1,2,\dots,J)$.

The value of population unit in the $(i,j)^{th}$ stratum be denoted by z_{ijl} $(l = 1,2,\dots,N_{ij})$. since the study variable is denoted by z . the unbiased estimate of population $\bar{z}$ is

$$\bar{z}_{st} = \sum_{i=1}^{I} \sum_{j=1}^{J} w_{ij} \bar{z}_{ij} \qquad (2)$$

Where $w_{ij} = \frac{N_{ij}}{N}$ is the stratum weight for the $(i,j)^{th}$ and $\bar{z}_{ij} = \frac{1}{n_{ij}} \sum_{l=1}^{n_{ij}} z_{ijl}$, with variance is given by

$$var(\bar{z}_{st}) = \sum_{i} \sum_{j} \left(\frac{1}{n_{ij}} - \frac{1}{N_{ij}}\right) w_{ij}^2 \sigma_{ijz}^2 \qquad (3)$$

Where, $\sigma_{ijz}^2 = \frac{1}{N_{ij}} \sum_{l=1}^{N_{ij}} (z_{ijl} - \bar{z}_{ij})^2$ (4)

If the finite population correction f.p.c is ignored $var(\bar{z}_{st})$ can be expressed as $var(\bar{z}_{st}) = \sum_i \sum_j \frac{w_{ij}^2 \sigma_{ijz}^2}{n_{ij}}$ (5)

Let the regression model of z on y and q be given as

$$z = \lambda(y,q) + \varepsilon \qquad (6)$$

Where, $\lambda(y,q)$ is linear or nonlinear function of y and q, ε is error term such that $E(\varepsilon \backslash y,q) = 0$ and $var(\varepsilon \backslash y,q) = \Psi(y,q) > 0 \ \forall\, y \in (a,b)$ and $q \in (c,d)$.

Under model (6) the stratum mean μ_{ijz} and the stratum variance σ_{ijz}^2 can be written as

$\mu_{ijz} = \mu_{ij\lambda}$ and $\sigma_{ijz}^2 = \sigma_{ij\lambda}^2 + \sigma_{ij\varepsilon}^2$ (7)

Where , $\mu_{ij\lambda}$ is the expected values of $\lambda(y,q)$ and $\sigma^2_{ij\lambda}$ is the variance of $\lambda(y,q)$ in the $(i,j)^{th}$ stratum and $\sigma^2_{ij\varepsilon}$ is the variance of error term in the $(i,j)^{th}$ stratum if λ and ε are uncorrelated .

Take into account uncorrelated between y and q. let , $f(y)$and $f(q)$be the frequency function of the auxiliary variables y and q respectively , defined in the interval [a,b] and [c,d] .

If the population mean of the study variable z is estimated under the variance (3) , then the problem of determining the strata boundaries is to cutup the ranges $h = b - a$ and $k = d - c$ at $(I - 1)$ and $(J - 1)$ intermediate points as $a = y_0 \leq y_1 \leq \cdots \leq y_{I-1} \leq y_L = b$ and

$c = q_0 \leq q_1 \leq \cdots \leq q_{J-1} \leq q_J = d$.

If the finite population correction f.p.c is ignored, then the minimization of variance $var(\bar{z}_{st})$ in (3) can be expressed as

Minimizing $\sum_i \sum_j \frac{w^2_{ij}\sigma^2_{ijz}}{n_{ij}}$ (8)

While using (7) equation (8) can be written as

Minimizing $\sum_i \sum_j \frac{w^2_{ij}(\sigma^2_{ij\lambda}+\sigma^2_{ij\varepsilon})}{n_{ij}}$ (9)

If (y,q) , $\lambda(y,q)$, $\Psi(y,q)$ are known and also integrable then w_{ij},$\sigma^2_{ij\lambda}$ and $\sigma^2_{ij\varepsilon}$ can be obtained as a function of boundary points $(y_{i-1}, y_i, q_{j-1}, q_j)$ by using the following expression

$$w_{ij} = \int_{y_{i-1}}^{y_i} \int_{q_{j-1}}^{q_j} f(y,q)\, dy\, dq \qquad (10)$$

$$\sigma^2_{ij\lambda} = \frac{1}{w_{ij}} \int_{y_{i-1}}^{y_i} \int_{q_{j-1}}^{q_j} \lambda^2(y,q) f(y,q)\, dy\, dq - \mu^2_{ij\lambda} \qquad (11)$$

$$\mu_{ij\lambda} = \frac{1}{w_{ij}} \int_{y_{i-1}}^{y_i} \int_{q_{j-1}}^{q_j} \lambda(y,q) f(y,q)\, dy\, dq \tag{12}$$

And $(y_{i-1}, y_i, q_{j-1}, q_j)$ are the boundary points of the $(i,j)^{th}$ stratum. Thus the objective function (9) could be expressed as the function of boundary points $(y_{i-1}, y_i, q_{j-1}, q_j)$ only . let

$$\phi_{ij}(y_{i-1}, y_i, q_{j-1}, q_j) = \frac{w_{ij}^2(\sigma_{ij\lambda}^2 + \sigma_{ij\varepsilon}^2)}{n_{ij}} \tag{13}$$

And the ranges as

$$h = b - a = y_I - y_0 \tag{14}$$

$$k = d - c = q_J - q_0 \tag{15}$$

Then in the bivariate stratification, a problem of determining stratum boundary (y_i, q_j) is to break up the ranges (14),(15) at intermediate points to estimate $y_1 \le y_2 \le \cdots \le y_{I-2} \le y_{I-1}$ and $q_1 \le q_2 \le \cdots \le q_{J-2} \le q_{J-1}$ .then the problem of obtaining OSB (y_i, q_j) is to

$$minimize \sum_i \sum_j \phi_{ij}(y_{i-1}, y_i, q_{j-1}, q_j) \tag{16}$$

$$subject\ to \quad a = y_0 \le y_1 \le \cdots \le y_{I-1} \le y_I = b$$
$$c = q_0 \le q_1 \le \cdots \le q_{J-1} \le q_J = d$$

Let $h_i = y_i - y_{i-1}$ and $k_j = q_j - q_{j-1}$ denote the total length or width of $(i,j)^{th}$ stratum. With the above definition of h_i and k_j the equation (14) and (15) can be expressed as

$$\sum_i h_i = \sum_i (y_i - y_{i-1}) = b - a = h \tag{17}$$

$$\sum_j k_j = \sum_j (q_j - q_{j-1}) = d - c = k \tag{18}$$

Then the $(i,j)^{th}$ stratification point $y_i: i = 1,2, \ldots\ldots, I-1,\ \ q_j: j = 1,2, \ldots\ldots, J-1$ is expressed as $y_i = y_0 + h_1 + \cdots + h_i = y_{i-1} + h_i$, $q_j = q_0 + k_1 + \cdots + k_j = q_{j-1} + k_j$

Restating the problem of determining OSB as the problem of determining optimum points $(\sum_i h_i, \sum_j k_j)$ adding equation (17), (18) as a constraint , the problem (16) can be treated as an equation problem of determining optimum strata width (O.S.W) $h_1, h_2, \ldots\ldots, h_I$ and $k_1, k_2, \ldots\ldots, k_J$ and can be expressed as Mathematical Programming Problem(M.P.P)

$$minimize \sum_i \sum_j \phi_{ij}(y_{i-1}, y_i, q_{j-1}, q_j) \qquad (19)$$

$$subject\ to \sum_i h_i = h$$

$$\sum_j k_j = k$$

$, i = 1,2, \ldots., I\ and\ j = 1,2, \ldots., J$ and $h_i \geq 0, k_j \geq 0$

For $i = 1$, $j = 1$ the term $\phi_{11}(h_1, y_0, k_1, q_0)$ in the objective function (19) is a function of h_1, k_1 alone as y_0, q_0 are known , similar the second term $\phi_{22}(h_2, y_1, k_2, q_1) = \phi_{22}(h_2, y_0 + h_1, k_2, q_0 + k_1)$ will become a function of h_2, k_2 alone once h_1, k_1 is known , and so on then stating the objective function as a function of h_i, k_j alone , a M.P.P can be written as

$$minimize \sum_i \sum_j \phi_{ij}(h_i, k_j) \qquad (20)$$

$$subject\ to \sum_i h_i = h$$

$$\sum_{j} k_j = k$$

$, i = 1,2, \ldots, I \ and \ j = 1,2, \ldots, J$ and $h_i \geq 0, k_j \geq 0$

3. The Suggested Mathematical Goal Programming Model

The suggested Mathematical goal programming model for evaluating OSB and optimum sample size allocation to the strata when the number of strata $(I * J)$ and the total sample size (n) are predetermined ,was presented in this section.

Assume that the regression model defined in equation (6) is a linear as:

$$z = A + By + Eq + e$$

When the error term and two auxiliary variables are independent, we get

$$\sigma_{ijz}^2 = B^2 \sigma_{ijy}^2 + E^2 \sigma_{ijq}^2 \tag{21}$$

Where, B and E are the estimates of regression coefficients.

$$\sigma_{ijy}^2 = \frac{1}{w_{ij}} \int_{q_{j-1}}^{q_j} \int_{y_{i-1}}^{y_i} y^2 f(y, q)\, dy\, dq - \mu_{ijy}^2 \tag{22}$$

$$\sigma_{ijq}^2 = \frac{1}{w_{ij}} \int_{y_{i-1}}^{y_i} \int_{q_{j-1}}^{q_j} q^2 f(y, q)\, dy\, dq - \mu_{ijq}^2 \tag{23}$$

$$w_{ij} = \int_{y_{i-1}}^{y_i} \int_{q_{j-1}}^{q_j} f(y, q)\, dy\, dq \tag{24}$$

$$\mu_{ijy} = \frac{1}{w_{ij}} \int_{q_{j-1}}^{q_j} \int_{y_{i-1}}^{y_i} y f(y, q)\, dy\, dq \tag{25}$$

$$\mu_{ijq} = \frac{1}{w_{ij}} \int_{y_{i-1}}^{y_i} \int_{q_{j-1}}^{q_j} q f(y, q)\, dy\, dq \tag{26}$$

To optimally determine stratum boundaries and allocate the sample to the different strata under multi-objective model.

Applying the variance formula in (8) and substituting in (21).Thus the model can be formulated as:

$$minimize \sum_{i}\sum_{j}\frac{w_{ij}^{2}(B^{2}\sigma_{ijy}^{2}+E^{2}\sigma_{ijq}^{2})}{n_{ij}} \qquad (27)$$

$$subject\ to \sum_{i} h_i = h$$

$$,\sum_{j} k_j = k$$

$$\sum_{i}\sum_{j} n_{ij} = n$$

$, i = 1,2, \ldots, I\ and\ j = 1,2, \ldots, J$, $1 \leq\ n_{ij} \leq N_{ij}$

and $h_i \geq 0, k_j \geq 0$

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.

Then the suggested Goal programming approach can be formulated as:

Minimize $dp_1 + dn_1 + dp_2 + dn_2 + dp_3 + dn_3$ (28)

Subject to

$$\sum_i \sum_j \frac{w_{ij}^2 (B^2 \sigma_{ijy}^2 + E^2 \sigma_{ijq}^2)}{n_{ij}} + dn_1 - dp_1 = v \qquad (29)$$

$$\sum_i \sum_j c_{ij}\, n_{ij} + dn_2 - dp_2 = C \qquad (30)$$

$$\sum_i \sum_j t_{ij}\, n_{ij} + dn_3 - dp_3 = T \qquad (31)$$

$$\sum\nolimits_{i=1}^{I} h_i = h \qquad (32)$$

$$y_i = y_{i-1} + h_i \qquad (33)$$

$$\sum\nolimits_{j=1}^{J} k_j = k \qquad (34)$$

$$q_j = q_{j-1} + k_j \qquad (35)$$

$$\sum_i \sum_j n_{ij} = n \qquad i = 1,2,\dots,I \quad , j = 1,2,\dots,J \qquad (36)$$

$$h_i \geq 0, k_j \geq 0\ ,\ 1 \leq \ n_{ij} \leq N_{ij}\ ,\quad dp_1, dp_2, dp_3, dn_1, dn_2, dn_3 \ \geq 0$$

Where ,

n_{ij}: Sample size of the ijth stratum

$n = \sum_i \sum_j n_{ij}$: Total sample size

c_{ij}: per unit cost of the ijth stratum

C: total cost

t_{ij}: time per unit of the ijth stratum

T total time

v prefixed variance of the estimator of the population mean

$dp_1, dp_2, dp_3, dn_1, dn_2, dn_3$ are positive and negative deviation variables of goals where first goal is to minimize $V(\bar{z}_{st})$, second and third goals are to minimize cost and time of collecting data per unit in each stratum respectively, $\frac{w_{ij}^2(B^2\sigma_{ijy}^2+E^2\sigma_{ijq}^2)}{n_{ij}} = V(\bar{z}_{st})$ if the finite population correction is ignored, N_{ij}: Stratum size of the ij^{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. Because of its simple mathematical form the study chosen the two auxiliary variables which chosen followed exponential distribution as an application of the idea of a multi-objective model for obtaining optimum stratum boundary and allocation the sample into different strata with pdf as:

$$f(y) = \begin{cases} \theta e^{-\theta y}, y \geq 0 \\ 0 \ \ otherwise \end{cases} \tag{37}$$

$$, f(q) = \begin{cases} \lambda e^{-\lambda q}, q \geq 0 \\ 0 \ \ otherwise \end{cases} \tag{38}$$

By using (22), (23),(24) and (37),(38) the term W_{ij} , $\sigma_{ijy}^2$and σ_{ijq}^2 can be expressed as

$$W_{ij} = e^{-\theta y_i} e^{-\lambda q_j} \left(e^{\theta h_i} - 1\right)\left(e^{\lambda k_j} - 1\right) \tag{39}$$

$$\sigma_{ijy}^2 = \frac{\frac{1}{\theta^2}\left(e^{\theta h_i} - 1\right)^2 - h_i^2 e^{\theta h_i}}{(e^{\theta h_i} - 1)^2} \tag{40}$$

$$\sigma_{ijq}^2 = \frac{\frac{1}{\lambda^2}\left(e^{\lambda k_j} - 1\right)^2 - k_j^2 e^{\lambda k_j}}{\left(e^{\lambda k_j} - 1\right)^2} \tag{41}$$

2. To determine the OSB and optimum allocation into sample strata which result in minimum possible variance of the estimator let y and q followed exponential distribution.
3. Where θ and λ are the chosen exponential distribution parameters and $(y_0, y_I), (q_0, q_J)$ are the chosen observation of smallest and largest values of stratification variables y and q respectively. h and k are the different between largest and smallest .
4. To evaluate the performance for the suggested model , some parameters were randomly selected for two auxiliary variables y and q chosen from some published researches in the field of study have $\theta = .08$ and $\lambda = .05$ respectively as distribution parameters when sample size $n = 100$ and $y_0 = 1.5$, $y_I = 21.5$ and $h = 20$, $q_0 = 1$, $q_J = 16$ and $k = 15$.
5. The study applied the suggested model to calculate the variance when the initial value of variance $v = 13.6$ (which calculated using Khan and Sharama (2015)), the fixed value of cost =12000, the specific rang of time =1500 were chosen arbitrary and the coefficients of regression are $B = .3$ and $E = .7$ the values of regression coefficients were chosen for the variation in the variables.

The suggested goal programming model (28-36) when the auxiliary variables y and q is given by (38) by Using (39), (40) and (41), can be formulated as:

Minimize $dp_1 + dn_1 + dp_2 + dn_2 + dp_3 + dn_3$ (42)

Subject to

$$\sum_i \sum_j \frac{\left(e^{-\theta y_i} e^{-\lambda q_j} \left(e^{\theta h_i}-1\right)\left(e^{\lambda q_j}-1\right)\right)^2 \left(.3^2 \left(\frac{\frac{1}{\theta^2}\left(e^{\theta h_i}-1\right)^2 - h_i^2 e^{\theta h_i}}{\left(e^{\theta h_i}-1\right)^2} \right) + \left(.7^2 \left(\frac{\frac{1}{\lambda^2}\left(e^{\lambda k_j}-1\right)^2 - k_j^2 e^{\lambda k_j}}{\left(e^{\lambda k_j}-1\right)^2} \right)\right)\right)}{n_{ij}} + dn_1 - dp_1 = 13.6 \qquad (43)$$

$$\sum_i \sum_j c_{ij}\, n_{ij} + dn_2 - dp_2 = 12000 \qquad (44)$$

$$\sum_i \sum_j t_{ij}\, n_{ij} + dn_3 - dp_3 = 1500 \qquad (45)$$

$$\sum_{i=1}^{I} h_i = 20 \qquad (46)$$

$$y_i = y_{i-1} + h_i \qquad (47)$$

$$\sum_{j=1}^{J} k_j = 15 \qquad (48)$$

$$q_j = q_{j-1} + k_j \qquad (49)$$

$$\sum_i \sum_j n_{ij} = 100 \qquad i = 1,2,\dots,I \quad ,j = 1,2,\dots,J \qquad (50)$$

$$, h_i \geq 0, k_j \geq 0\,,\; 1 \leq \; n_{ij} \leq N_{ij}\,,\quad dp_1, dp_2, dp_3, dn_1, dn_2, dn_3 \;\geq 0$$

5. Results

The study solved the suggested goal programming model (42-50) by using a GAMS program and the results as follows:

Figure 1: showed the OSB for six expected number of strata for two independent auxiliary variables y and q for the total of six strata, three along y variable and two along q variable with exponential distribution.

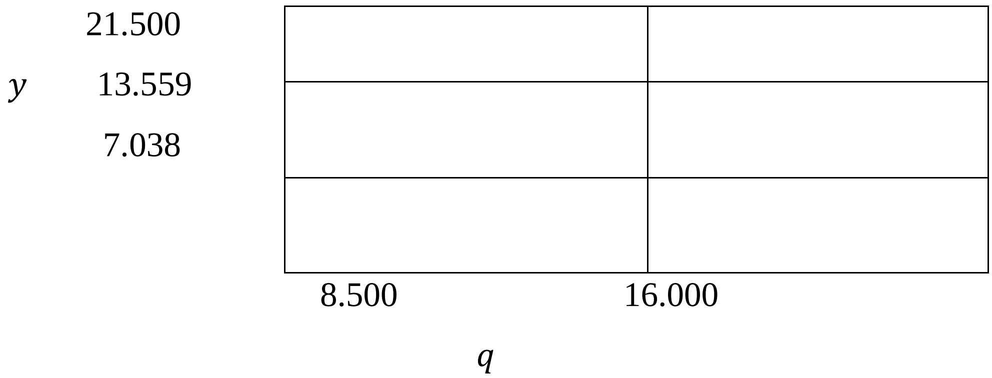

Table 1: The OSB Results for six expected number of strata for two independent auxiliary variables y and q with exponential distribution.

OSW	**OSB**	**Sample Size (n_{ij})**	C_{ij}	T_{ij}	**Objective function**
(05.538,7.5)	**(07.038,8.5)**	**18.14≈18**	**2410**	**296**	**.081**
(6.521,7.5)	**(13.559,8.5)**	**18.2≈18**	**2428**	**296**	
(7.941,7.5)	**(21.500,8.5)**	**18.3≈19**	**2459**	**296**	
(05.538,7.5)	**(07.038,16)**	**15.04≈15**	**900**	**150.75**	
(6.521,7.5)	**(13.559,16)**	**15.09≈15**	**1866**	**150.75**	
(7.941,7.5)	**(21.500,16)**	**15.18≈ 15**	**1875**	**277.5**	

The suggested Mathematical goal programming model for determining OSB and optimum allocation of sample size to the strata results showed in table (1)are:

1. The suggested Mathematical goal programming model calculate the optimum stratum width and optimum stratum boundary in satisfactory way where the new minimum value of variance is .081 which are less than the initial value $v = 13.6$ which calculated before
2. Sample size is divided in satisfactory way according to the number of strata where the suggested model determine the size

of strata ($n_1 = 18, \dots n_4 = 15$) at the range of the total sample size ($n = 100$).

3. The suggested model divided the time and cost and reducing them as much as possible where the model was not outside the permitted range of the proposed cost-time values.
4. As a result, we can infer that employing a single study variable with two auxiliary variables while taking cost and time into account And that's an extension the exciting technique with Khan and Sharama (2015).

6. Conclusion

The study suggested Mathematical goal programming model to determine the optimum strata boundary by bi-variate variables in multi-objective problem with minimum variance.

1. The new minimum value of variance .081 which is less than the initial value ($v = 13.6$) which chosen before
2. The cost and the time (not exceed a fixed limit according to budget and time of survey) was added to the model as objective constrain need to minimize.
3. The suggested mathematical goal programming help researcher or any statistician to predict Optimum Stratum Boundary through the available information represented by the parameters of the appropriate distribution that fit the nature of the data.

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., and 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, 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.
5. 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.
6. Khan, M. G. M., Nand, N., and Ahmad, N. (2008). "Determining the optimum strata boundary points using dynamic programming". Survey Methodology, 34(2), 205-214.
7. Khan, M. G. M., Ahmad, N. and Khan, S. (2009)."Determining the optimum stratum boundaries using mathematical

programming". J. Math. Model. Algorithms, Springer, Netherland, 8(4): 409- 423.

8. Khan, M.G.M. and Sharma, S. (2015). “Determining Optimum Strata Boundaries and Optimum Allocation in Stratified Sampling”, Aligarh Journal of Statistics, Vol. 35, 23-40.
9. 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.
10. Reddy, K. G., Khan, M. G. and D. Rao (2016). "A procedure for computing optimal stratum boundaries and sample sizes for multivariate surveys". JSW 11(8), 816–832.
11. 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.
12. 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.