Countries Classification According to Covid-19 Data, Based on Aged-65 and Older, Using Multi-Objective Mathematical Goal Programming Model

By

Dr. Fatma S. Abo El-Hassan

Lecturer of Statistics, Faculty of Commerce, Al-Azhar University, Girls' Branch, Cairo, Egypt

Dr. Ramadan Hamed

Professor of Statistics, Faculty of Economics and Political Science, Cairo University, Egypt.

Research Professor Social Research Center, AUC

Dr. Elham A. Ismail

Professor of Statistics, Faculty of Commerce, Al-Azhar University, Girls' Branch, Cairo, Egypt

Dr. Safia M. Ezzat

Lecturer of Statistics, Faculty of Commerce, Al-Azhar University, Girls' Branch, Cairo, Egypt

Scientific Journal for Financial and Commercial Studies and Research (SJFCSR)

Faculty of Commerce – Damietta University

Vol.4, No.2, Part 1., July 2023

APA Citation:

Abo El-Hassan, F. S.; **Hamed**, R.; **Ismail,** E. A. and **Ezzat**, S. M. (2023). Countries Classification According to Covid-19 Data, Based on Aged-65 and Older, Using Multi-Objective Mathematical Goal Programming Model, ***Scientific Journal for Financial and Commercial Studies and Research***, Faculty of Commerce, Damietta University, 4(2)1, 863-885.

Website: https://cfdj.journals.ekb.eg/

Countries Classification According to Covid-19 Data, Based on Aged-65 and Older, Using Multi-Objective Mathematical Goal Programming Model

Dr. Fatma S. Abo El-Hassan; Dr. Ramadan Hamed; Dr. Elham A. Ismail and Dr. Safia M. Ezzat

Abstract

Stratified random sampling is an efficient sampling technique for estimating the characteristics of a population. Identifying stratum borders and allocating sample size to strata are two of the most critical components for increasing estimate accuracy. Most surveys are conducted under severe budget constraints, and the survey must be completed within a set time frame. Hence, the majority of surveys include cost and time considerations as highly important goals. Thus, they are necessitating to be under consideration. In many cases, auxiliary variables are utilized in the absence of the main study variable or missing values, indicating that this hypothesis was examined in the study.

The paper proposed a mathematical goal programming model for determining optimum stratum boundaries and allocating sample size to different strata utilizing one auxiliary variable as stratification factor when cost and time are taken into consideration. It is not necessary to have all the data to get the optimum stratum boundary; rather, it is sufficient to know information about the parameters of the distribution based on the researcher's experience or other previous studies. Therefore, here the mathematical goal programming is proposed to make it easier for any researcher or statistician to make a prediction the optimum stratum boundary using the information represented by the parameters of the appropriate distribution that fit the nature of the data. In addition, the paper employed Covid-19 data to evaluate the performance of the proposed model using the auxiliary variable (population at old age), where it distributed exponential distribution and the results of the suggested model are satisfying.

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

1. Introduction

Stratified random sampling is one of the most important types of statistical 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.

Dalenius 1950 presented the first formulae for determining the optimum stratum boundaries. The optimum stratum width has been discussed by Khan et al. (2002, 2005, 2008, and 2009) as a mathematical programming problem could be solve by the dynamic programming method. They focused on variables that followed distributions: uniform, normal, exponential, triangular, right triangular, power and Cauchy. Rao et al. (2014) suggested a procedure for determining optimum stratum boundary and optimum strata size of each stratum When the study variable has a pareto frequency distribution.

In their paper, Fonolahi and Khan (2014) proposed a method to evaluate the optimum strata boundaries when the variable distribution is exponentially distributed and also when the measurement unit cost will vary throughout of the strata. When multiple survey variables are under consideration, Reddy et al. (2016) presented a computational method for obtaining optimum stratum boundary that using the dynamic programming methodology. Also provided a numerical example for calculating the optimum stratum boundary for the two primary study variables. Danish et al. (2017) discussed optimum strata boundaries as a non-linear programming problem that's when the cost per unit different throughout of the strata. The same problem was presented by Reddy et al. (2018) but under the Neyman allocation. Danish and Rizvi (2019) suggested a non-linear programming model to determine optimum strata boundaries use two auxiliary variables are follow Weibull distributions. Using R package, Reddy and Khan (2020) solved the problem of optimum stratum boundary for various distributions.

Hamid et. al (2021) suggested Mathematical goal programming model for getting Optimum Stratum Boundary and allocate sample size into different strata with two auxiliary variables as stratification factors using covid-19 data. Abo El.Hassan et al. (2022) suggested a mathematical goal programming model for determining the optimum stratum boundaries for an exponential variable and also cost and time are taken into consideration. Utilizing Covid-19 data to evaluate the efficiency of the proposed model.

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

This study's objective is to determine optimum stratum boundary (OSB), optimal sample size, optimal cost and optimal time when an auxiliary variable is utilized as the basis for stratification.

2. Methods and Methodology

Let the target population consisting of "N" units be stratified into I strata based on auxiliary variable y and estimation of the mean study variable is of interest.

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

$$z = \lambda(u) + \varepsilon \qquad (1)$$

Where, $\lambda(u)$ is linear or nonlinear function of u, ε is error term with conditional expectation $E(\varepsilon|u) = 0$ and variance $V(\varepsilon|u) = \varphi(u) > 0 \; \forall \; u \in [a, b]$.

The auxiliary variable is taken as the basis of stratification with one study variable according to Danish et. al (2018). They divided the whole population into I strata on the basis auxiliary variable u such that the number of units in the $(i)^{th}$ stratum is N_i . a sample of size "n" is to be drawn from the whole population and suppose that the allocation of sample size to the $(i)^{th}$ stratum is n_i $(i = 1,2, \dots, I)$.

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

$$\bar{z}_{st} = \sum_{i=1}^{I} W_i \bar{z}_i \qquad (2)$$

Where $W_i = \frac{N_i}{N}$ is the stratum weight for the i^{th} and $\bar{z}_i = \frac{1}{n_i}\sum_{l=1}^{n_i} z_{il}$, with variance is given by

$$var(\bar{z}_{st}) = \sum_{i} (\frac{1}{n_i} - \frac{1}{N_i}) W_i^2 \sigma_{iz}^2 \qquad (3)$$

Where,

$$\sigma_{iz}^2 = \frac{1}{N_i}\sum_{l=1}^{N_i}(z_{il} - \bar{z}_i)^2 \qquad (4)$$

If the finite population correction is ignored $var(\bar{z}_{st})$ can be expressed as

$$var(\bar{z}_{st}) = \sum_i \sum_j \frac{w_{ij}^2 \sigma_{ijz}^2}{n_{ij}} \qquad (5)$$

Under model (1) the stratum mean μ_{iz} and the stratum variance σ_{iz}^2 can be written as

$$\mu_{iz} = \mu_{i\lambda} \text{ and } \sigma_{iz}^2 = \sigma_{i\lambda}^2 + \sigma_{i\varepsilon}^2 \qquad (6)$$

Where, $\mu_{i\lambda}$ is the expected values of $\lambda(u)$ and $\sigma_{i\lambda}^2$ is the variance of $\lambda(u)$ in the $(i)^{th}$ stratum and $\sigma_{i\varepsilon}^2$ is the variance of error term in the $(i)^{th}$ stratum if λ and ε are uncorrelated . Let, $f(u)$ be the frequency function of the auxiliary variable u, defined in the interval [a,b].

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 $h = b - a$ at $(I - 1)$ intermediate points as $a = u_0 \leq u_1 \leq \cdots \leq u_{I-1} \leq u_L = b$.

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

$$\text{Minimizing } \sum_i \frac{W_i^2 \sigma_{iz}^2}{n_i} \tag{7}$$

While using (6) equation (7) can be written as

$$\text{Minimizing } \sum_i \frac{w_i^2(\sigma_{i\lambda}^2+\sigma_{i\varepsilon}^2)}{n_i} \tag{8}$$

If $f(u)$, $\lambda(u)$, $\varphi(u)$ are known and also integrable then w_i,$\sigma_{i\lambda}^2$ and $\sigma_{i\varepsilon}^2$ can be obtained as a function of boundary points (u_{i-1}, u_i) by using the following expression

$$w_i = \int_{u_{i-1}}^{u_i} f(u)\, du \tag{9}$$

$$\sigma_{i\lambda}^2 = \frac{1}{w_i}\int_{u_{i-1}}^{u_i} \lambda^2(u) f(u)\, du \; - \mu_{i\lambda}^2 \tag{10}$$

$$\mu_{i\lambda} = \frac{1}{w_i}\int_{u_{i-1}}^{u_i} \lambda(u) f(u)\, du \tag{11}$$

and (u_{i-1}, u_i) are the boundary points of the $(i)^{th}$ stratum.

Thus, the objective function (8) could be expressed as the function of boundary points (u_{i-1}, u_i) only. Let

$$\phi_i(u_{i-1}, u_i) = \frac{W_i^2(\sigma_{i\lambda}^2 + \sigma_{i\varepsilon}^2)}{n_i} \tag{12}$$

and the ranges as

$$h = b - a = u_I - u_0 \tag{13}$$

Then in the bivariate stratification, a problem of determining stratum boundary (u_i) is to break up the ranges (13) at intermediate points to estimate $u_1 \leq u_2 \leq \cdots \leq u_{I-2} \leq u_{I-1}$. Then the problem of obtaining OSB (u_i) is to

$$minimize \sum_i \phi_i(u_{i-1}, u_i) \tag{14}$$

$$subject\ to \quad a = u_0 \leq u_1 \leq \cdots \leq u_{I-1} \leq u_I = b$$

Let $h_i = u_i - u_{i-1}$ denotes the total length or width of $(i)^{th}$ stratum. With the above definition of h_i the equation (13) can be expressed as

$$\sum_i h_i = \sum_i (u_i - u_{i-1}) = b - a = h \tag{15}$$

Then the $(i)^{th}$ stratification point u_i: $i = 1,2, \ldots \ldots, I-1$, is expressed as $u_i = u_0 + h_1 + \cdots + h_i = u_{i-1} + h_i$.

Restating the problem of determining OSB as the problem of determining optimum points ($\sum_i h_i$) adding equation (15) as a constraint, the problem (14) can be treated as an equation problem of determining optimum strata width $h_1, h_2, \ldots \ldots, h_I$ and can be expressed as mathematical programming problem

$$minimize \sum_i \phi_i(u_{i-1}, u_i) \tag{16}$$

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

$$, i = 1,2, \ldots., I \text{ and } h_i \geq 0$$

For $i = 1$ the term $\phi_1(h_1, u_0)$ in the objective function (16) is a function of h_1 alone as u_0 are known , similar the second term $\phi_2(h_2, u_1) = \phi_2(h_2, u_0 + h_1)$ will become a function of h_2 alone once h_1 is known , and so on then stating the objective function as a function of h_i alone , a Mathematical programming problem can be written as

$$minimize \sum_i \phi_i(h_i) \tag{17}$$

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

$$, i = 1,2, \ldots., I \text{ and } h_i \geq 0$$

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) and the total sample size (n) are predetermined, was presented in this section.

Assume that the regression model defined in equation (1) is a linear as:

$$z = A + Bu + \varepsilon \tag{18}$$

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

$$\sigma_{iz}^2 = B^2\sigma_{iu}^2 + \sigma_{i\varepsilon}^2 \tag{19}$$

Where, B is the estimate of regression coefficient and where $\sigma_{i\varepsilon}^2$ is the variance of e in the i^{th} stratum. By applying the variance formula in (7) and substituting in (19). Thus, the mathematical programming problem model can be formulated as:

$$minimize \sum_i \frac{W_i^2(B^2\sigma_{iu}^2 + \sigma_{i\varepsilon}^2)}{n_i} \tag{20}$$

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

$$, i = 1,2, \dots, I \ , 1 \le \ n_i \le N_i \text{and } h_i \ge 0$$

If the total sample size n for a stratified survey is predetermined, a reasonable criterion for obtaining the optimum allocation n_i is to minimize the variance of the stratified sample mean given in (20). So, the allocation of sample sizes into different strata can be added in the suggested model. On the other hand, the cost of survey is a major factor of sample allocation to various strata in addition to that the time needed to collect data for the sample may be another objective need to minimize. Suggested mathematical goal programming model will be adopted to multiple objectives to optimally determine stratum boundary, allocate the sample to the different strata take in consideration the sample cost should not exceed a fixed limit and the time needed for the sampling process in kept within a specific range in the form of Goal Programming Model.

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 aggregate of sample sizes in each stratum be equal to predetermined sample size.

3- The cost (not exceed a fixed limit according to budget of survey) was added to the model as objective constrain need to minimize.

4- 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:

$$minimize\ \ dp_1 + dn_1 + dp_2 + dn_2 + dp_3 + dn_3 \tag{21}$$

Subject to

$$\sum_i \frac{W_i^2(B^2 s_{iu}^2 + s_{i\varepsilon}^2)}{n_i} + dn_1 - dp_1 = v \tag{22}$$

$$\textstyle\sum_i c_i\, n_i + dn_2 - dp_2 = C \tag{23}$$

$$\sum_i t_i\, n_i + dn_3 - dp_3 = T \tag{24}$$

$$\sum_{i=1}^{I} h_i = h \tag{25}$$

$$u_i = u_{i-1} + h_i \tag{26}$$

$$\sum_i n_i = n \qquad i = 1,2,\dots,I \tag{27}$$

$, h_i \geq 0\ ,\ 1 \leq\ n_i \leq N_i\ ,\ \ dp_1, dp_2, dp_3, dn_1, dn_2, dn_3\ \geq 0$

Where,

n_i: Sample size of the ith stratum

$n = \sum_i n_i$: Total sample size

c_i: per unit cost of the ith stratum

C: total cost

t_i: time per unit of the ith 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, N_i: Stratum size of the ith stratum.

The Goal Programming Approach for Exponential distribution:

If we take here, a variable of interest that follows the exponential distribution as follows.

Let the variable under study (u) follows an exponential distribution with parameter $\theta > 0$, that is probability density function

$$f(u) = \begin{cases} \theta e^{-\theta u}, u \geq 0 \\ 0 \;\; otherwise \end{cases} \tag{28}$$

By using the following equations

$$W_i = \int_{u_{i-1}}^{u_i} f(u)\, du \tag{29}$$

$$S_i^2 = \frac{1}{W_i} \int_{u_{i-1}}^{u_i} u^2 f(u)\, du - \mu_i^2 \tag{30}$$

Where,

$$\mu_i = \frac{1}{W_i} \int_{u_{i-1}}^{u_i} u f(u)\, du$$

the term W_i and S_i^2 can be expressed as the boundaries of the ith stratum are (u_{i-1}, u_i), then

$$W_i = \int_{u_{i-1}}^{u_i} \theta e^{-\theta u}\, du$$

The result of integrals gives W_i as,

$$W_i = e^{-\theta u_i}\left[e^{\theta h_i} - 1\right] \tag{31}$$

Where $h_i = u_i - u_{i-1}$

Similar that, $\mu_i$is as follows:

$$\mu_i = \frac{1}{W_i} \int_{u_{i-1}}^{u_i} (u)\, \theta e^{-\theta u} \;\; du \tag{32}$$

$$\mu_i = \frac{(e^{\theta h_i} - 1)(u_i + \frac{1}{\theta}) - h_i e^{\theta h_i}}{(e^{\theta h_i} - 1)} \tag{33}$$

Similarly, from (30) the stratum variance is

$$S_i^2 = \underbrace{\frac{1}{W_i} \int_{u_{i-1}}^{u_i} (u^2)\, \theta e^{-\theta u}\, du}_{Q} - \mu_i^2 \tag{34}$$

The integral of part B is

$$Q = \left[u_{i-1}^2 e^{-\theta u_{i-1}} - u_i^2 e^{-\theta u_i}\right] + \frac{2}{\theta}\left[-\left(u_i e^{-\theta u_i} - u_{i-1} e^{-\theta u_{i-1}}\right) - \frac{1}{\theta}\left(e^{-\theta u_i} - e^{-\theta u_{i-1}}\right)\right]$$

Using the relation $h_i = u_i - u_{i-1}$, thus doing some mathematical steps we get

$$Q = e^{-\theta u_i}\left[e^{\theta h_i}\left[(u_i - h_i) + \frac{1}{\theta}\right]^2 + \frac{1}{\theta^2}e^{\theta h_i} - \left(u_i + \frac{1}{\theta}\right)^2 - \frac{1}{\theta^2}\right]$$

Finally, by substituting W_i and $\mu_i$form (31) and (33) in (34), we get

$$S_i^2 = \frac{\frac{1}{\theta^2}\left(e^{\theta h_i - 1}\right)^2 - h_i^2 e^{\theta h_i}}{\left(e^{\theta h_i - 1}\right)^2} \qquad (35)$$

Using (29) and (35) the suggested goal programming model (21-27) when the study variable u is given by (28), can be formulated as:

$$minimize \sum_{g=1}^{G}(dp_g + dn_g) \qquad g = 1,2, \dots, G \qquad (36)$$

Subject to

$$\left\{\sum_{i=1}^{I}\frac{1}{n_i}\left\{\left(e^{-\theta u_i}\left[e^{\theta h_i} - 1\right]\right)^2\left\{B^2 * \frac{\left(e^{\theta h_i} - 1\right)^2 - (\theta h_i)^2 e^{\theta h_i}}{\theta^2(e^{\theta h_i} - 1)^2} + \mu_{\varphi(u)}\right\}\right\}\right\} + dn_1 - dp_1 = v \qquad (37)$$

$$\sum_{i=1}^{I} c_i n_i + dn_2 - dp_2 = C \qquad (38)$$

$$\sum_{i=1}^{I} t_i n_i + dn_3 - dp_3 = T \qquad (39)$$

$$\sum_{i=1}^{I} h_i = h \qquad (40)$$

$$u_i = u_{i-1} + h_i \qquad (41)$$

$$\sum_{i=1}^{I} n_i = n \qquad ,i = 1,2, \dots, I \qquad (42)$$

$, h_i \geq 0 \ , \ 1 \leq n_i \leq N_i \ , \ dp_g, dn_g \geq 0$

u is a variable defined in [a,b] interval

I is number of strata, $i = 1,2, \dots, I$

$a = u_0 \leq u_1 \leq \cdots \leq u_{I-1} \leq u_I = b$

$h = u_I - u_0 \ , h_i = u_i - u_{i-1}$

dp_g, dn_g positive and negative deviation variables of the gth goal

G Total number of goal functions,

N_i: Stratum size of the ith stratum.

N : total population size

n_i: Sample size of the ith stratum

$n = \sum_{i=1}^{I} n_i$: Total sample size

v prefixed variance of the estimator of the population mean

c_i: per unit cost of the ith stratum

C : total cost

t_i: time per unit of the ith stratum

T : total time

4. Application of Mathematical Goal Programming in Covid-19 Data Analysis

This section explains the application of the suggested approach in Covid-19 data analysis using auxiliary variable with lognormal distribution. The suggested model is formulated, solved and the results are discussed.

Although all age groups are at risk of contracting COVID-19, older persons are at a significantly higher risk of mortality and severe disease following infection UN COVID (2020).

This section uses auxiliary variable (population at old) which distributed log normal distribution as the following steps:

The study used Covid-19 data. The data are obtained from World Health Organization https://libguides.umn.edu/HealthStatistics for whole the world to compute infection ratio variable for each state where the infection ratio is equal to total number of cases divided by the number of populations in 2020. The data has the infection ratio as the study variable. Using aged-65 older is the auxiliary variable which defined as share of the population that is 65 years and older.

Using the available part of the complete data for the variable under study and the auxiliary variable, the relationship between them was studied using regression methods and the regression coefficient was estimated.

The linear regression model defined in (18) was fitted significantly for this data with B=.562. Table (1) shows the ANOVA results and table (2) shows the model parameters estimates.

Table (1): ANOVA results: ANOVA[b]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	753.506	1	753.506	79.209	.000[a]
	Residual	1636.216	172	9.513		
	Total	2389.722	173			

a. Predictors: (Constant), Aged-65 older

b. Dependent Variable: infection ratio

Table (2): Model parameters: Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.428	.401		1.070	.286
	Aged-65 older	.334	.037	.562	8.900	.000

a. Dependent Variable: infection rate

The R^2 value for this model is 0.31.

1- The estimated parameters are significant. The frequency histogram for (u= the variable under consideration) in fig (1). Based on the statistical test results:

- Chi-squared statistical test was done with (Deg. of freedom: (number of categories with frequency greater than 5)-1= 6, statistic= 9.5014, P-value= 0.14728).

The distribution of aged-65 older u reads as an exponential distribution as (28) with the estimated value of the parameter is $\theta = 0.1153$

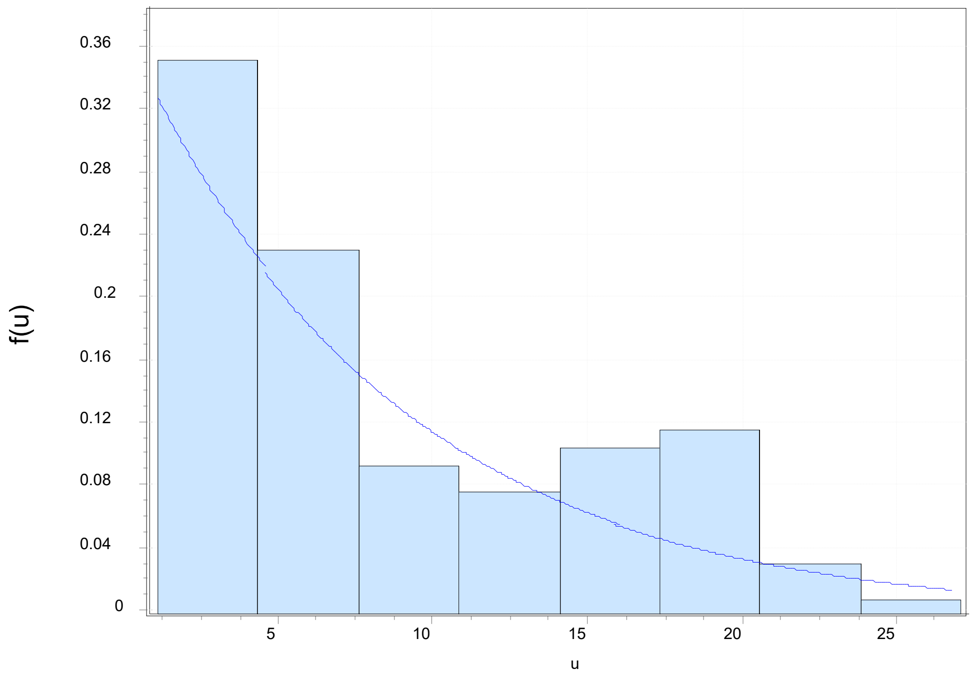

Fig (1): The histogram and the fitted Exponential distribution of the auxiliary variable u

2- Substituting W_i (31) and s_{iu}^2 (35) in suggested multi objective goal programming approach (37-43) and put B=.562, the expected stratum variance of the error is obtained as $\mu_{\varphi(u)} = MSE = 9.513$ we get:

$$minimize\ dp_1 + dn_1 + dp_2 + dn_2 + dp_3 + dn_3 \tag{43}$$

Subject to

$$\sum_i \left[\frac{\left(e^{-\theta u_i}\left[e^{\theta h_i} - 1\right]\right)^2 \left(.562^2 \left(\frac{\frac{1}{\theta^2}\left(e^{\theta h_i - 1}\right)^2 - h_i^2 e^{\theta h_i}}{\left(e^{\theta h_i - 1}\right)^2} \right) + 9.513 \right)}{n_i} \right] + dn_1 - dp_1 = v \tag{44}$$

$$\sum_i c_i n_i + dn_2 - dp_2 = C \tag{45}$$

$$\sum_i t_i n_i + dn_3 - dp_3 = T \tag{46}$$

$$\sum_{i=1}^{I} h_i = h \tag{47}$$

$u_i = u_{i-1} + h_i$ (48)

$\sum_i^I n_i = n \qquad i = 1,2,\dots,I$ (49)

$, h_i \geq 0 \ , \ 1 \leq \ n_i \leq N_i \ , \ \ dp_1, dp_2, dp_3, dn_1, dn_2, dn_3 \ \geq 0$

3- The suggested multi objective goal programming approach used the value of v as initial value where $v = 0$ when $I = 2, I = 3, I = 4$ and $I = 5$ respectively to be sure that the performance for the suggested approach led to satisfied results when cost and time are under consideration.

4- The suggested multi objective goal programming approach which using for determining the OSB substituting values of $B = 0.562, \theta = 0.1153$ for the population $N = 185$ when sample size $n = 50$ followed exponential distribution with $u_0 = 1.144$, $u_I = 27.049$ and $h = 25.905$. Where B is the regression coefficient, θ is the exponential distribution parameter.

5- (u_0 ,u_I) are the chosen observation of smallest and largest values of stratification variables u,h is the different between largest and smallest. The study applied the suggested model when the fixed value of cost =12000, the specific range of time =150 are chosen arbitrary.

Solving the suggested goal programming model (43-49) by using a GAMS software.

5. Results

This section explains the results for the second phase when auxiliary variable follows exponential distribution. Table (3) showed the results for OSB, Optimum sample size of the variance function for exponential distribution when $I = 2$, $I = 3$, $I = 4$, $I = 5$ respectively.

Table (3): Results for aged-65 older of COVID-19 variable which distributed exponential distribution

No. of strata (I)	Optimum strata width OSW (h_i)	Optimum strata boundary OSB (u_i)	Sample size (n_i)	Total cost per stratum (C_i) L.E	Total cost per stratum (T_i) hours	Optimum value of variance
2	h_1 =7.938	$u_1 = u_0 + h_1$ =9.82	27.966≈28	7413	92.5	0.008
	h_2 =17.967		22.034≈22	4588	57.5	
3	h_1 =6.197	$u_1 = u_0 + h_1$ =7.341	22.676≈22	6595	88	0.002
	h_2 =8.655	$u_2 = u_1 + h_2$ =15.996	16.814≈17	3774	50	
	h_3 =11.053		10.510≈11	1518	5	
4	h_1 =5.130	$u_1 = u_0 + h_1$ =6.274	19.967≈20	6450	95	0.002
	h_2 =6.612	$u_2 = u_1 + h_2$ =12.886	14.992≈15	3630	55	
	h_3 =6.864	$u_3 = u_2 + h_3$ =19.750	9.288≈9	1350	1	
	h_4 =7.299		5.753≈6	557	1	
5	h_1 =4.630	$u_1 = u_0 + h_1$ =5.774	18.476≈18	6082	90	0.002
	h_2 =5.820	$u_2 = u_1 + h_2$ =11.594	14.090≈14	3598	58	
	h_3 =5.584	$u_3 = u_2 + h_3$ =17.178	8.725≈8	1276	1	
	h_4 =5.218	$u_4 = u_3 + h_4$ =22.396	5.395≈6	592	1	
	h_5 =4.653		3.315≈4	242	1	

Table (3) presents OSB and variance when $I = 2, I = 3, I = 4\ and\ I = 5$ strata for the suggested model. fig (2) shows the results graphically for number of strata corresponding to variance.

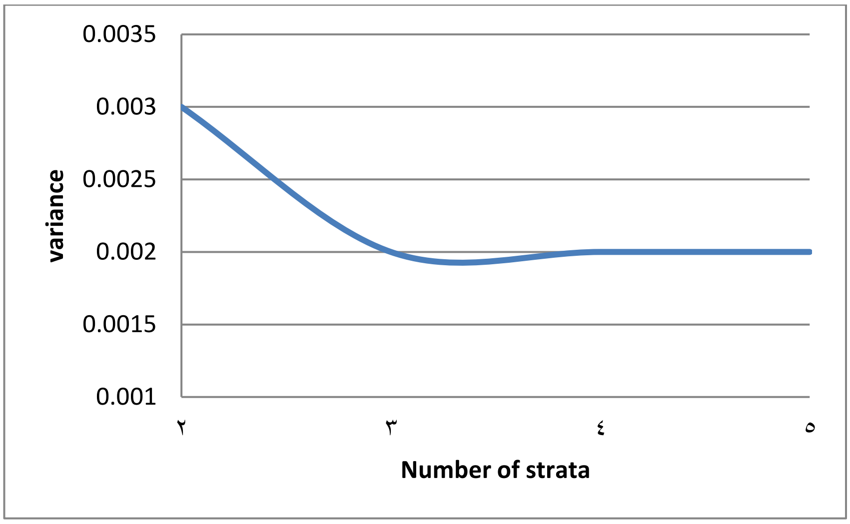

Fig (2): Strata variance corresponding to number of strata

It's clear from the fig (2) that there's no increase in efficiency at $I = 3, I = 4\ and\ I = 5$, because the variance values are the same when $I = 3, I = 4\ and\ I = 5$ that mean there is no reason to increase the number of strata from 3 or 4 strata.

The optimum stratum boundary was (7.34, 15.99) and it means that:

- The first strata (from first point in distribution of aged-65 and older variable to 7.34). The first stratum represents countries with less prevalence of infection. It has 102 countries. Of these, 50 are on Africa, most notably (Egypt, Libya, Sudan, Cameroon, Ghana, and Tanzania). 33countries in Asia (Qatar, Kuwait, Oman, Bahrain, Afghanistan, Iraq, Saudi Arabia (. 9 countries in North America and 7 countries in South America.
- The second strata (from 7.34 to 15.99 in distribution of aged-65 and older variable). The second stratum represents countries with medium prevalence of infection. It includes 39 countries. 3 countries on the continent of Africa (Tunisia-Seychelles-Rodcius) 10 countries in Asia, some of them Turkey Lebanon Israel. The continent of Europe appears in this stratum and includes 10 countries and North America 11 countries and South America 5 countries.
- The countries with the highest infection prevalence ratio were 41 countries, represented third stratum from 15.99 to end of the stated distribution. All in Asia, Europe, and North America. Europe's continent

tops the list with 35 countries, for example (Italy, France, Germany, Denmark, Poland, Australia, Portugal) North America has 3 countries (Canada - Dominica - Saint Kitts), 3 countries in Asia (Japan - China - Taiwan).

6. Evaluation of The Suggested Approach

This section explains another way to evaluate the performance of the suggested multi-objectives approach using three comparisons between the multi-objectives suggested approach and two classical approaches. The two classical methods which used in compare are $Cum\sqrt{f}$ method and MPP according to Danish 2018 method or khan (2009) and others where they all use the same model. The previous three phases used to compare between the multi-objectives suggested approach and two classical approaches. The comparison between the three approaches takes the following steps:

1- The study used Covid-19 data. The data are obtained from World Health Organization https://libguides.umn.edu/HealthStatistics for whole the world to compute infection rate variable for each state where the infection rate is equal to total number of cases divided by the number of populations in 2020. The data has the infection rate as the study variable while aged-65 older is the auxiliary variable.

2- To apply $Cum\sqrt{f}$ method, the Covid-19 available data have been grouped into 20 equal classes. The class frequencies and their cumulative square roots are shown in table (4).

Table (4): Frequency distribution of u and $Cum\sqrt{f(u)}$

No. of classes	Classes	Frequency of aged-65 older $f(u)$	$Cum\sqrt{f(u)}$
1	(1.144,2.496)	11	3.32
2	2.496,3.848)	45	10.02
3	(3.848,5.200)	22	14.72
4	(5.200,6.552)	12	18.18
5	(6.552,7.904)	14	21.92
6	(7.904,9.256)	8	24.75
7	(9.256,10.608)	4	26.75
8	(10.608,11.960)	7	29.39
9	(11.960,13.312)	4	31.39
10	(13.312,14.664)	5	33.63
11	(14.664,16.016)	10	36.79
12	(16.016,17.368)	6	39.24
13	(17.368,18.720)	6	41.69
14	(18.720,20.072)	13	45.30
15	(20.072,21.424)	3	47.03
16	(21.424,22.776)	1	48.03
17	(22.776,24.128)	1	49.03
18	(24.128,25.480)	0	49.03
19	(25.480,26.480)	1	50.03
20	(26.480,27.049)	1	51.03

The boundaries according to $cum\sqrt{f}$ and Khan (2009) methods when $I=2, I=3, I=4\ and\ I=5$ strata are illustrated in table (4).

The boundaries according to $cum\sqrt{f}$ method when I=2 strata are corresponding to nearest point to $(\frac{51.03}{2})$ = 10.608

The boundaries according to $cum\sqrt{f}$ method when I=3 strata are corresponding to nearest point to ($\frac{51.03}{3}$, $\frac{2(51.03)}{3}$) = 6.552, 16.016

The boundaries according to $cum\sqrt{f}$ method when I=4 strata are corresponding to nearest point to ($\frac{51.03}{4}$, $\frac{2(51.03)}{4}$, $\frac{3(51.03)}{4}$) = 5.2, 10.608, 17.368

The boundaries according to $cum\sqrt{f}$ method when I=5 strata are corresponding to nearest point to ($\frac{51.03}{5}$, $\frac{2(51.03)}{5}$, $\frac{3(51.03)}{5}$, $\frac{4(51.03)}{5}$)= 3.848, 7.904, 13.312, 18.72

3- To apply Danish et. Al (2018) method and others the equation as following is being used

$$minimize\ \sum_{i=1}^{I}\frac{W_i^2(S_{iz}^2)}{n_i}$$

Subject to

$$\sum_{i=1}^{I} h_i = h \qquad (50)$$

and $h_i \geq 0\ , i = 1,2,\dots,I$

4- substituting equations (19), (31) and (35) in (50) equation as:

$minimize$

$$\sum_{i}^{I}\left[\frac{\left(e^{-\theta u_i}\left[e^{\theta h_i}-1\right]\right)^2\left(.562^2\left(\frac{\frac{1}{\theta^2}\left(e^{\theta h_i}-1\right)^2-h_i^2e^{\theta h_i}}{\left(e^{\theta h_i}-1\right)^2}\right)+9.513\right)}{n_i}\right]$$

Subject to

$$\sum_{i=1}^{I} h_i = h \qquad (51)$$

and $h_i \geq 0\ , i = 1,2,\dots,I$

5- To apply the suggested approach the equations (43-49) are used.

6- GAMS software is used to calculate the three approaches and the results presented in table (5).

Table (5): Comparison results for the three methods.

I	$cum\sqrt{f}$		MPP Khan (2009)		Suggested multi objective Goal Programming approach		Relative efficiency (over)	
	OSB	variance	OSB	variance	OSB	variance	$cum\sqrt{f}$	Khan (2009)
2	10.608	0.137	9.861	0.018	9.680	0.008	17.125	2.25
3	6.552 16.016	0.0636	7.341 15.996	0.004	7.140 16.28	0.002	31.8	2
4	5.200 10.608 17.368	0.0283	6.274 12.886 19.750	0.004	6.885 13.917 20.377	0.002	14.15	2
5	3.848 7.904 13.312 18.72	0.0241	5.774 11.595 17.178 22.396	0.004	6.249 12.328 18.567 24.097	0.002	12.05	2

The results of comparison with two classical methods are presented in table (5), which shows the OSB and variance values for $cum\sqrt{f}$, Danish et. al (2018) methods and suggested multi objective Goal Programming approach when $I = 2, I = 3, I = 4\ and\ I = 5$ strata. The variance values for the three approaches are calculated and the value of the lowest variance was for the suggested approach. The third column in table (5) showed the efficiencies of the suggested approach over the other two methods. Thus, the suggested approach increases the precision of the estimate compared to the two other methods. The suggested approach can be applied for determining OSB in addition to optimum sample sizes, cost, and time over different strata with minimum variance.

7. Conclusion

This study presented the suggested mathematical goal programming model with minimum variance for determining OSB and sample allocation when the estimate population mean of the study variable is of the frequency distribution of its auxiliary variable was known under multiple objectives when cost and time are under consideration. The study applied the original covid-19 data from WHO to evaluate the performance of the suggested mathematical goal programming model. The data has the infection rate as the study variable, computed it for each country where the infection rate is equal to total number of cases divided by the number of populations in 2020, while aged-65 older is the auxiliary variable. The results were as: The three strata represent countries with less, medium, and high prevalence of infection. The group of countries in each group according to the division of the suggested approach is like what happened in reality.

References

1. Abo_El. Hassan, F. S., Hamed, R., Ismail, E. A., & Ezzat, S. M. (2022). Optimum stratum boundaries and sample sizes for Covid-19 data in Egypt. *Plos one*, *17*(7), e0271220.
2. Dalenius, T. (1950). The problem of optimum stratification. *Scandinavian Actuarial Journal*, *1950*(3-4), 203-213.
3. Danish, F. (2018). *Optimum Stratification with Auxiliary Information using Mathematical Programming* (Doctoral dissertation, Sher-E-Kashmir University of Agricultural Sciences and Technology of Jammu).
4. Danish, F., Rizvi, S. E. H., Jeelani, M. I., & Reashi, J. A. (2017). Obtaining strata boundaries under proportional allocation with varying cost of every unit. *Pakistan Journal of Statistics and Operation Research*, 567-574.
5. Danish, F., & Rizvi, S. E. H. (2019). Optimum stratification by two stratification variables using mathematical programming. *Pak. J. Statist*, *35*(1), 11-24.
6. Hamid, R., Ismail, E. A., Ezzat, S. M. & Abo_El. Hassan, F. S. (2021). Multi-objective Mathematical Programming Model for Optimum Stratification in Multivariate Stratified Sampling. *Trade and finance Journal, 41(4),* 189-208.
7. Fonolahi, A. V., & Khan, M. G. (2014). Determining the optimum strata boundaries with constant cost factor. In *Asia-Pacific World Congress on Computer Science and Engineering* (pp. 1-6). IEEE.
8. Khan, E. A., Khan, M. G., & Ahsan, M. J. (2002). Optimum stratification: a mathematical programming approach. *Calcutta Statistical Association Bulletin*, *52*(1-4), 323-334.
9. Khan, M. G., Sehar, N., & Ahsan, M. J. (2005). Optimum stratification for exponential study variable under Neyman allocation. *Journal of the Indian Society of Agricultural Statistics*, *59*(2), 146-150.

10. Khan, M. G. M., Nand, N., & Ahmad, N. (2008). Determining the optimum strata boundary points using dynamic programming. *Survey methodology*, *34*(2), 205-214.
11. Khan, M. G., Ahmad, N., & Khan, S. (2009). Determining the optimum stratum boundaries using mathematical programming. *Journal of Mathematical Modelling and Algorithms*, *8*, 409-423.
12. Rao, D. K., Khan, M. G., & Reddy, K. G. (2014). Optimum stratification of a skewed population. *World Academy of Science, Engineering and Technology-International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering*, *8*(3), 492-495.
13. Reddy, K. G., Khan, M. G., & Rao, D. (2016). A Procedure for Computing Optimal Stratum Boundaries and Sample Sizes for Multivariate Surveys. *J. Softw.*, *11*(8), 816-832.
14. Reddy, K. G., Khan, M. G., & Khan, S. (2018). Optimum strata boundaries and sample sizes in health surveys using auxiliary variables. *Plos one*, *13*(4), e0194787.
15. Reddy, K. G., & Khan, M. G. (2020). stratifyR: An R Package for optimal stratification and sample allocation for univariate populations. *Australian & New Zealand Journal of Statistics*, *62*(3), 383-405.
16. United Nations COVID. (2020). Secretary-general’s policy brief: The impact of COVID-19 on older persons.

تصنيف الدول وفقاً لبيانات Covid-19 بناءً على السن ٦٥ عام فأكثر باستخدام نموذج برمجة الأهداف الرياضية متعدد الأهداف

د. فاطمة سيد أبو الحسن؛ د. رمضان حامد؛ د. إلهام عبد الرازق إسماعيل؛ د. صفيه محمود عزت

الملخص العربي

العينات العشوائية الطبقية هي أسلوب للحصول على عينات تستخدم لتقدير خصائص السكان، ويعد تحديد حدود الطبقة وتخصيص حجم العينة للطبقات من أهم المكونات لزيادة دقة التقدير. وغالبا ما يتم إجراء الاستطلاعات او المسوح في ظل قيود شديدة على الميزانية، وفي حدود مدة زمنية محددة. ايضا في كثير من الحالات، يتم استخدام المتغيرات المساعدة في حالة عدم وجود متغير الدراسة الرئيسي. استخدمت الدراسة المتغيرات المساعدة كما اخذت في اعتبرها التكلفة والوقت كأهداف مهمة للغاية. تقترح الدراسة نموذجًا لبرمجة الهدف الرياضي لتحديد حدود الطبقة المثلى وتخصيص حجم العينة لطبقات مختلفة باستخدام متغير إضافي كعامل تقسيم مع أخذ التكلفة والوقت في الاعتبار. وتتميز الدراسة بانه ليس من الضروري الحصول على جميع البيانات للحصول على حدود الطبقة المثلى؛ بل يكفي معرفة معلومات عن معلمات توزيع البيانات بناءً على خبرة الباحث أو دراسات سابقة أخرى. لذلك اقترحت الدراسة برمجة الهدف الرياضي لتسهيل قيام أي باحث أو إحصائي بالتنبؤ بحدود الطبقة المثلى باستخدام المعلومات التي تمثلها معلمات التوزيع المناسب الذي يتناسب مع طبيعة البيانات. بالإضافة إلى ذلك استخدمت الدراسة بيانات Covid-19 لتقييم أداء النموذج المقترح باستخدام المتغير المساعد (السكان في سن الشيخوخة-٦٥ فأكثر)، وكانت نتائج النموذج المقترح مرضية.

الكلمات المفتاحية:

العينات العشوائية الطبقية، الحدود المثلى للطبقة، التوزيع الأسي، برمجة الهدف الرياضية، الوقت، التكلفة.