Advances and Applications in Statistics
© 2024 Pushpa Publishing House, Prayagraj, India
https://pphmjopenaccess.com
https://doi.org/10.17654/0972361724060
Volume 91, Number 9, 2024, Pages 1137-1151 **P-ISSN: 0972-3617**

A COMPARISON OF APPROXIMATE OPTIMAL STRATIFICATION WITH OTHER METHODS OF STRATIFICATION USING PROPORTIONAL ALLOCATION

Ghadah A. Alsakkal, Talal Abdul Razak Al Hasso and Mowafaq Muhammed Al-kassab

Department of Computer Engineering
College of Engineering
Tishk International University
Erbil, Iraq
e-mail: ghada.alsakkal@tiu.edu.iq

Department of Statistics and Informatics
College of Computers and Mathematics
University of Mosul
Iraq
e-mail: talal744740@uomosul.edu.iq

Received: May 30, 2024; Accepted: July 11, 2024

2020 Mathematics Subject Classification: 62D05.

Keywords and phrases: proportional allocation, stratification, optimum stratum boundaries, cumulative frequency, efficiency.

How to cite this article: Ghadah A. Alsakkal, Talal Abdul Razak Al Hasso and Mowafaq Muhammed Al-kassab, A comparison of approximate optimal stratification with other methods of stratification using proportional allocation, Advances and Applications in Statistics 91(9) (2024), 1137-1151. https://doi.org/10.17654/0972361724060

Published Online: July 17, 2024

Department of Mathematics Education
College of Education
Tishk International University
Erbil, Iraq
e-mail: mowafaq.muhammed@tiu.edu.iq

Abstract

One of the main reasons in stratification is to get more accurate estimates by producing gain in the precision of these estimates. To achieve this, we can determine the optimum stratum boundaries. Many procedures were developed to obtain this optimum boundary, and several approximate rules proposed as a result of the complicated calculation involved in solving the theoretical equations to obtain the optimum points of stratification. In this article, we present a comparison between the cumulative $f^{6/7}$ suggested by the authors with other given approximate methods suggested previously. Uniform, right triangular, exponential, normal and chi-square distributions are compared. For certain values of the parameters of these distributions, the cumulative $f^{6/7}$ method is favorable compared with these approximate optimal stratification methods.

1. Introduction

Stratified random sampling is an important sampling technique utilized in estimating the unknown parameters of the studied population [1]. In stratified sampling, the sampling-frame is divided into several non-overlapping groups or strata L, in such a way that the strata constructed are internally homogeneous with respect to the studied variable [2]. An advantage of stratified sampling design is that when a stratum is homogeneous, the measurements of the study variable (y) vary little from each other and the precise estimate of y can be obtained from a small sample in that stratum. Thus, combining these estimates from all L strata, the design produces a gain in the precision of estimate of the variable in the whole population [1, 2]. The stratification by convenience manner is not always a reasonable criterion as the strata so obtained may not be internally homogeneous with respect to a variable of interest. Thus, we must look for the optimum stratum boundaries (OSB) that maximize the precision of the estimators [3, 4]. The primary consideration involved in determining OSB is that the strata should be as internally homogeneous as possible to achieve maximum accuracy. The stratum variance should be as small as possible [2, 3]. When a single variable is of interest and the stratification is made based on this study variable, an ideal situation is that the distribution of the study variable is known and the OSB can be determined by dividing the range of its distribution at suitable points. This problem of determining the OSB, when both the study and stratification variables are identical, was first discussed by Dalenius and Hodges [4]. They presented a set of minimal equations which are usually difficult to solve for OSB because of their implicit nature. Iterative method is used to the OSB [4]. Many authors have also attempted to determine the global OSB [8, 9]. Unnithan [6] proposed an iterative method that requires a suitable initial solution. Hence, subsequently the attempts for determining approximate optimum stratum boundaries have been made by several authors [3, 5, 10, 11]. For a skewed population where a certainty stratum is necessary, Lavallée and Hidiroglou [12] proposed an algorithm to obtain stratum boundaries for a power allocated stratified sample. After reviewing Lavallée and Hidiroglou's algorithm, a modified algorithm that incorporates the different relationships between the stratification and study variables was proposed [19]. There are several other algorithms available in the literature [16, 18, 19]. Section 2 presents the formulation of the problem of optimum stratum boundaries. Section 3 presents four approximate optimal stratification methods proposed by [7, 9, 11, 13]. A comparison between these approximate methods using some theoretical distributions is given in Section 4. Finally, conclusions are given in Section 5.

2. Formulation of the Problem

Let the population under consideration be divided into 'L' strata and a stratified random sample of size n be drawn from it. Let the hth stratum contain N_h units with Y-values Y_{hi} $(i = 1, 2, ..., N_h)$ so that $N = \sum_{h=1}^{L} N_h$, and the population mean and the population variance in the hth stratum are, respectively,

$$\mu_h = \frac{1}{N_h} \sum_{i=1}^{N_h} Y_{hi}$$

and

$$\sigma_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (Y_{hi} - \mu_h)^2.$$

The population mean is $\mu = \sum_{h=1}^{L} W_h \mu_h$, where $W_h = N_h / N$. We denote the sample size in stratum h by n_h and the 'ith' observed Y-value in stratum h by y_{hi}. The 'hth' stratum mean and the stratified sample mean are $\bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$ and $\bar{y}_{st} = \frac{1}{n} \sum_{h=1}^{L} w_h \bar{y}_h$, respectively, where $n = \sum_{h=1}^{L} n_h$, $\bar{y}_{st}$ is an unbiased estimator of μ. The variance of $\bar{y}_{st}$ depends on how the strata are constructed, how the sample is allocated and whether stratification is done by Y or by some auxiliary variable X.

The population Y-values are the values of independent and identically distributed variables generated from a distribution with probability density $f(y)$. The optimal construction of strata using proportional allocation has been determined in [4, 14].

For the construction of L strata by choosing $L - 1$ stratum boundaries $y_1 < y_2 < \cdots < y_{l-1}$ on the Y-scale, $\mathrm{Var}(\bar{y}_{st})$ is minimum when the following equations are satisfied:

$$y_h = \frac{\mu_h + \mu_{h+1}}{2}; \quad h = 1, 2, ..., L - 1,$$

where

$$\mu_h = \frac{1}{w_h}\int_{Yh-1}^{Yh} yf(y)dy$$

with $w_h = \int_{Yh-1}^{Yh} f(y)dy$. Exact solution for obtaining the stratum boundaries y_h is not easy, approximate methods are given in Section 3.

3. Approximate Optimal Stratification Methods

In this section, some approximate methods are presented.

3.1. Cumulative $f^{1/3}$

Stratification method is studied in [7]. The cumulative $f^{1/3}$ (abbreviated cum $f^{1/3}$) is constructed first, and then the cum $f^{1/3}$ scale is partitioned into equal intervals. The variance of the stratified mean $\bar{y}_{1st}$ using this stratification and allocation method is given by

$$\mathrm{Var}(\bar{y}_{1st}) = H^3(Y)/12nL^2, \text{ where } H(Y) = \int_{-\infty}^{\infty} f^{1/3}(y)dy.$$

3.2. Cumulative $f^{1/2}$

It is studied and recommended in several books and articles [1, 5, 7, 9, 10]. The cumulative $f^{1/2}$ is formed, and then the cum $f^{1/2}$ scale is divided into equal intervals. The allocation consists of taking equally as many observations from each stratum. An approximation to the variance of the stratified mean $\bar{y}_{2st}$, using this stratification and allocation method, is given by

$$\mathrm{Var}(\bar{y}_{2st}) = K^2(Y)/12nL^2, \text{ where } K(Y) = \int_{-\infty}^{\infty} f^{1/2}(y)dy.$$

3.3. Cumulative $f^{5/6}$

The cumulative $f^{5/6}$ method of stratification is studied in [14]. The cumulative $f^{5/6}$ is formed and cum $f^{5/6}$ scale is divided into equal intervals. The variance of the stratified mean $\bar{y}_{3st}$ using this stratification and allocation method is given by

$$\operatorname{Var}(\bar{y}_{3st}) = C^{6/5}(Y)/12nL^2, \text{ where } C(Y) = \int_{-\infty}^{\infty} f^{5/6}(y)dy.$$

3.4. Cumulative $f^{6/7}$

This method is studied and used in [19]. First the cumulative $f^{6/7}$ is formed, and then the cum $f^{6/7}$ scale is divided into equal intervals. The variance of the stratified mean $\bar{y}_{4st}$ using this stratification and allocation method is given by

$$\operatorname{Var}(\bar{y}_{4st}) = A^{7/6}(Y)/12nL^2, \text{ where } A(Y) = \int_{-\infty}^{\infty} f^{6/7}(y)dy.$$

4. Theoretical Applications

In this section, we compare the approximate optimal stratification methods given in Section 3 using some theoretical distributions.

4.1. Using the uniform distribution $U(c, c+d)$, with the cum $f^{1/3}$, the variance of the stratified mean $\bar{y}_{1st}$ is

$$v_1 = V(\bar{y}_{1st}) = \frac{H^3(y)}{12nL^2} = \frac{d^2}{12nL^2}, \text{ where } H(y) = d^{\frac{2}{3}},$$

with cum $f^{1/2}$,

$$v_2 = V(\bar{y}_{2st}) = \frac{K^2(y)}{12nL^2} = \frac{d}{12nL^2}, \text{ where } K(y) = d^{\frac{1}{2}},$$

with cum $f^{5/6}$,

$$v_3 = V(\bar{y}_{3st}) = \frac{c^{6/5}(y)}{12nL^2} = \frac{d^{\frac{1}{5}}}{12nL^2}, \text{ where } C(y) = d^{\frac{1}{6}},$$

and for cum $f^{6/7}$,

$$v_4 = V(\bar{y}_{4st}) = \frac{A^{7/6}(y)}{12nL^2} = \frac{d^{\frac{1}{6}}}{12nL^2}, \text{ where } A(y) = d^{\frac{1}{7}}.$$

Comparisons between the four methods in terms of the efficiency of cum $f^{6/7}(v_4)$ relative to the other three approximate methods are given in Table 1.

Table 1. Comparing cum $f^{6/7}$ with the other three methods

Uniform distribution			
d	$R_{1u} = \frac{v_4}{v_1}$	$R_{2u} = \frac{v_4}{v_2}$	$R_{3u} = \frac{v_4}{v_3}$
1.0	1	1	1
1.5	0.4755	0.7133	0.9866
2.0	0.2806	0.5612	0.9772
3.0	0.1334	0.4003	0.9640
4.0	0.0787	0.3150	0.9548
5.0	0.0523	0.2615	0.9478

We see from Table 1, that for $d > 1.0$, cum $f^{6/7}$ is more efficient than cum $f^{1/3}$, cum $f^{1/2}$ and cum $f^{5/6}$, respectively.

4.2. Using right triangular distribution whose p.d.f. is given by

$$f(y) = \begin{cases} \frac{2(b-y)}{(b-a)^2}; & a \le y \le b, \\ 0; & \text{otherwise,} \end{cases}$$

with the cum $f^{1/3}$, the variance of the stratified mean $\overline{y}_{1st}$ is

$$v_1 = V_{Prop}(\overline{y}_{1st}) = \frac{H^3(y)}{12nL^2} = \frac{0.84(b-a)^2}{12nL^2}, \text{ where } H(y) = 0.945(b-a)^{\frac{2}{3}},$$

with cum $f^{1/2}$,

$$v_2 = V_{Prop}(\overline{y}_{2st}) = \frac{K^2(y)}{12nL^2} = \frac{\frac{1}{2}(b-a)^2}{12nL^2}, \text{ where } K(y) = \frac{b-a}{\sqrt{2}},$$

with cum $f^{5/6}$,

$$v_3 = V_{Prop}(\overline{y}_{3st}) = \frac{c^{6/5}(y)}{12nL^2} = \frac{0.96(b-a)^{\frac{1}{5}}}{12nL^2}, \text{ where } C(y) = 0.97(b-a)^{\frac{1}{6}},$$

and with cum $f^{6/7}$,

$$v_4 = V_{Prop}(\overline{y}_{4st}) = \frac{A^{7/6}(y)}{12nL^2} = \frac{0.1(b-a)^{\frac{1}{6}}}{12nL^2}, \text{ where } A(y) = 0.14(b-a)^{\frac{1}{7}}.$$

Comparisons between the four methods in terms of the efficiency of cum $f^{6/7}(v_4)$ relative to the other three approximate methods are given in Table 2.

Table 2. Comparing cum $f^{6/7}$ with the other three methods

Right triangular			
$b-a$	$R_{1r} = \frac{v_4}{v_1}$	$R_{2r} = \frac{v_4}{v_2}$	$R_{3r} = \frac{v_4}{v_3}$
0.5	0.4241	0.7127	0.1064
1.0	0.1190	0.2000	0.1040
1.5	0.0566	0.0951	0.1026
2.0	0.0334	0.0561	0.1016
4.0	0.0094	0.0157	0.0993
5.0	0.0062	0.0105	0.0986

We see from Table 2 that for $b - a \geq 0.5$, cum $f^{6/7}$ is more efficient than cum $f^{1/3}$, cum $f^{1/2}$, and cum $f^{5/6}$, respectively.

4.3. Using exponential distribution $\exp(\lambda)$, with the cum $f^{1/3}$, the variance of the stratified mean $\bar{y}_{1st}$ is

$$v_1 = V_{Prop}(\bar{y}_{st}) = \frac{H^3(y)}{12nL^2} = \frac{27/\lambda^2}{12nL^2}, \text{ where } H(y) = \frac{3}{\lambda^{2/3}},$$

with cum $f^{1/2}$,

$$v_2 = V_{Prop}(\bar{y}_{st}) = \frac{K^2(y)}{12nL^2} = \frac{1/\lambda}{12nL^2}, \text{ where } K(y) = \lambda^{\frac{-1}{2}},$$

with cum $f^{5/6}$,

$$v_3 = V_{Prop}(\bar{y}_{st}) = \frac{C^{6/5}(y)}{12nL^2} = \frac{1.24\lambda^{\frac{-1}{5}}}{12nL^2}, \text{ where } C(y) = \frac{6}{5\lambda^{\frac{1}{6}}}$$

and with cum $f^{6/7}$, $v_4 = V_{Prop}(\bar{y}_{st}) = \frac{A^{7/6}(y)}{12nL^2} = \frac{1.197\lambda^{\frac{-1}{6}}}{12nL^2}$, where

$$A(y) = \frac{7}{6\lambda^{\frac{1}{7}}}.$$

Comparisons between the four methods in terms of the efficiency of cum $f^{6/7}(v_4)$ relative to the other three approximate methods are given in Table 3.

Table 3. Comparing cum $f^{6/7}$ with the other three methods

Exponential distribution			
λ	$R_{1\exp} = \frac{v_4}{v_1}$	$R_{2\exp} = \frac{v_4}{v_2}$	$R_{3\exp} = \frac{v_4}{v_3}$
0.1	0.000650	0.04393	0.89081
1.0	0.044300	0.29930	0.96180
2.0	0.157864	0.53329	0.98426
3.0	0.331979	0.74766	0.99764
3.2	0.373676	0.78898	0.99978
4.0	0.562549	0.95021	1.00724
4.2	0.615186	0.98964	1.00888
5.0	0.846887	1.14440	1.01475
5.4	0.975217	1.22020	1.01736

We see from Table 3 that for $0.1 \le \lambda \le 3.2$, cum $f^{6/7}$ is more efficient than cum $f^{5/6}$, more efficient than cum $f^{1/2}$ for $0.1 \le \lambda \le 4.2$ and it is more efficient than cum $f^{1/3}$ for $0.1 \le \lambda \le 5.4$.

4.4. Using normal distribution $N(\mu, \sigma^2)$ with the cum $f^{1/3}$, the variance of the stratified mean $\bar{y}_{1st}$ is

$$v_1 = V_{Prop}(\bar{y}_{st}) = \frac{H^3(y)}{12nL^2} = \frac{2.22}{\sigma 12nL^2}, \text{ where } H(y) = 1.305/\sigma^{\frac{1}{2}},$$

with cum $f^{1/2}$,

$$v_2 = V_{Prop}(\bar{y}_{st}) = \frac{K^2(y)}{12nL^2} = \frac{5.013\sigma}{12nL^2}, \text{ where } K(y) = 2.24\sigma^{\frac{1}{2}},$$

with cum $f^{5/6}$,

$$v_3 = V_{Prop}(\bar{y}_{st}) = \frac{C^{\frac{6}{5}}(y)}{12nL^2} = \frac{1.342\sigma^{\frac{1}{5}}}{12nL^2}, \text{ where } C(y) = 1.278\sigma^{\frac{1}{6}}$$

and with cum $f^{6/7}$,

$$v_4 = V_{Prop}(\bar{y}_{st}) = \frac{A^{\frac{7}{6}}(y)}{12nL^2} = \frac{1.273\sigma^{\frac{1}{6}}}{12nL^2}, \text{ where } A(y) = 1.23\sigma^{\frac{1}{7}}.$$

Comparisons between the four methods in terms of the efficiency of cum $f^{6/7}(v_4)$ relative to the other three approximate methods are given in Table 4.

Table 4. Comparing cum $f^{6/7}$ with the other three methods

Normal distribution			
σ	$R_{1n} = \frac{v_4}{v_1}$	$R_{2n} = \frac{v_4}{v_2}$	$R_{3n} = \frac{v_4}{v_3}$
0.223	0.6117	0.8884	0.9999
1	0.0391	0.2544	0.9511
4	0.0031	0.0801	0.9081
8	0.0009	0.0450	0.8874
16	0.0002	0.0252	0.8671

We see from Table 4 that for $\sigma \geq 0.223$, cum $f^{6/7}$ is more efficient than cum $f^{1/3}$, cum $f^{1/2}$ and cum $f^{5/6}$, respectively.

4.5. Using chi-square distribution $\chi^2(k)$ and k degrees of freedom, with the cum $f^{1/3}$, the variance of the stratified mean $\bar{y}_{1st}$ is

$$v_1 = V_{Prop}(\bar{y}_{st}) = \frac{\Gamma^3\left(\frac{k+4}{6}\right)6^{(k/2)+2}}{12nL^2 * 2^{k/2}\Gamma\left(\frac{k}{2}\right)},$$

where

$$H(y) = \frac{\Gamma\left(\frac{k+4}{6}\right)6^{(k/6)+(2/3)}}{2^{k/6}\left[\Gamma\left(\frac{k}{2}\right)\right]^{1/3}}$$

with cum $f^{1/2}$,

$$v_2 = V_{Prop}(\bar{y}_{st}) = \frac{K^2(y)}{12nL^2} = \frac{\Gamma^2\left(\frac{k+2}{4}\right)2^{(k/2)+2}}{12nL^2 * \Gamma\left(\frac{k}{2}\right)},$$

where

$$K(y) = \frac{\Gamma\left(\frac{k+2}{4}\right)2^{(k/4)+1}}{\left[\Gamma\left(\frac{k}{2}\right)\right]^{1/2}},$$

with cum $f^{5/6}$,

$$v_3 = V_{Prop}(\bar{y}_{st}) = \frac{C^{\frac{6}{5}}(y)}{12nL^2} = \left(\frac{6}{5}\right)^{\left(\frac{5k+2}{10}\right)} \frac{\Gamma^{\frac{6}{5}}\left(\frac{5k+2}{12}\right)2^{(1/5)}}{12nL^2 * \Gamma\left(\frac{k}{2}\right)},$$

where

$$C(y) = \left(\frac{6}{5}\right)^{\left(\frac{5k+2}{12}\right)} \frac{\Gamma\left(\frac{5k+2}{12}\right)2^{(1/6)}}{\left[\Gamma\left(\frac{k}{2}\right)\right]^{5/6}}$$

and with cum $f^{6/7}$,

$$v_4 = V_{Prop}(\bar{y}_{st}) = \frac{A^{\frac{7}{6}}(y)}{12nL^2} = \frac{\left\{\frac{\left(\frac{7}{3}\right)^{\left(\frac{3k+1}{6}\right)}\Gamma^{\frac{7}{6}}\left(\frac{3k+1}{7}\right)}{2^{\left(\frac{k}{2}\right)}\Gamma\left(\frac{k}{2}\right)}\right\}}{12nL^2},$$

where

$$A(y) = \frac{\left(\frac{7}{3}\right)^{\left(\frac{3k+1}{7}\right)} \Gamma\left(\frac{3k+1}{7}\right)}{2^{(3k/7)}\left[\Gamma\left(\frac{k}{2}\right)\right]^{6/7}}.$$

Comparisons between the four methods in terms of the efficiency of cum $f^{6/7}(v_4)$ relative to the other three approximate methods are given in Table 5.

Table 5. Comparing cum $f^{6/7}$ with the other three methods

Chi-square distribution			
k	$R_{1\,Chi} = \frac{v_4}{v_1}$	$R_{2\,Chi} = \frac{v_4}{v_2}$	$R_{3\,Chi} = \frac{v_4}{v_3}$
1	0.0233	0.2458	0.9612
2	0.0124	0.1680	0.9398
3	0.0084	0.1356	0.9299
12	0.0022	3.255E-05	4.0208E-5
24	0.0011	2.041E-12	0.35E-10

We see from Table 5 that for $k \geq 1$, cum $f^{6/7}$ is more efficient than cum $f^{1/3}$, cum $f^{1/2}$ and cum $f^{5/6}$, respectively.

5. Conclusions

The cum $f^{6/7}$ has proved to be the most efficient out of the presented approximate methods [17]. This method is applied using five theoretical distributions namely uniform, right triangular, exponential, normal, and chi-square. It is concluded that:

- The cum $f^{6/7}$ is more efficient than the cum $f^{1/3}$, cum $f^{1/2}$ and cum $f^{5/6}$ for the uniform distribution when $d > 1.0$.

- The cum $f^{6/7}$ is more efficient than cum $f^{1/3}$, cum $f^{1/2}$ and cum $f^{5/6}$ for the right triangular distribution when $(b-a) \geq 0.5$.

- It is also noted that for the exponential distribution, the cum $f^{6/7}$ is more efficient than cum $f^{5/6}$ when $0.1 \leq \lambda \leq 3.2$, more efficient than cum $f^{1/2}$ when $0.1 \leq \lambda \leq 4.2$ and more efficient than the cum $f^{1/3}$ when $0.1 \leq \lambda \leq 5.4$.

- For the normal distribution, it is concluded that the cum $f^{6/7}$ is more efficient than the cum $f^{1/3}$, cum $f^{1/2}$ and cum $f^{5/6}$ when $\sigma \geq 0.223$.

- Finally, it is concluded that the cum $f^{6/7}$ is more efficient than the cum $f^{1/3}$, cum $f^{1/2}$ and cum $f^{5/6}$ for the chi-square distribution when $k \geq 1$.

References

[1] T. Dalenius, Sampling in Sweden, Almqvist and Wiksell, Stockholm, 1957.

[2] W. Cochran, Sampling Techniques, Wiley, N.Y., 1963.

[3] L. Kish, Survey Sampling, Wiley, New York, 1965.

[4] T. Dalenius and J. L. Hodges, Minimum variance stratification, J. Amer. Statist. Assoc. 54 (1959), 88-101.

[5] R. J. Serfling, Approximately optimal stratification, J. Amer. Statist. Assoc. 63 (1968), 1298-1309.

[6] V. K. G. Unnithan, The minimum variance boundary points of stratification, Sankhya 40(C) (1978), 60-72.

[7] G. Ekman, An approximation useful in univariate stratification, J. Amer. Statist. Assoc. 30 (1959), 219-229.

[8] W. Cochran, Comparison of methods for determining stratum boundaries, Bull. Int. Statist. Inst. 3(2) (1961), 345-358.

[9] I. Hess, V. K. Sethi and T. R. Balakrishnan, Stratification: a practical investigation, J. Amer. Statist. Assoc. 61 (1966), 74-90.

[10] I. Thomson, Comparison of approximately optimal stratification given proportional allocation with other methods of stratification and allocation, Metrika 23 (1976), 15-25.

[11] R. Singh, Approximately optimal stratification on the auxiliary variable, J. Amer. Statist. Assoc. 66 (1971), 829-833.

[12] P. Lavallée and M. Hidiroglou, On the stratification of skewed populations, Survey Methodology 14(1) (1988), 33-43.

[13] M. M. T. Al-kassab, Approximately optimal stratification using proportional allocation, Tanmeat Alrafideen Journal 41 (1993) pages.

[14] M. M. T. Al-kassab and T. H. Aldaghestani, Using the linear model to find stratum boundaries based on the proportional allocation: a comparative study, Mutah Journal for Research and Studies 1294 (1997), 195-209.

[15] R. Valliant, A. H. Dorfman and R. M. Royall, Finite Population Sampling and Inference: A Prediction Approach, Wiley, New York, 2000.

[16] Heidar Pirzade, S. Sara and Abdelwahab Hamou-Lhadj, The concept of stratified sampling of execution traces, The 19th IEEE International Conference on Program Comprehension, ICPC 2011, Kingston, ON, Canada, June 22-24, 2011.

[17] G. A. Alsakkal and M. M. Al-kassab, Approximate methods to obtain the optimum stratum boundaries: a comparative study, Eurasian Journal of Science and Engineering 7(1) (2021), 197-204.

[18] M. Kozak, Optimal stratification using random search method in agricultural surveys, Stat. Transit. 6 (2004), 797-806.

[19] L.-P. Rivest, A generalization of the Lavallée and Hidiroglou algorithm for stratification in business surveys, Surv. Methodol. 28(2) (2002), 191-198.