International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2020): 7.803

On Comparative Analysis of Optimum Allocation Procedures in a Multivariate Stratified Sampling

Mohammad Idris Umar[1], Rabo Yahaya Saidu[2], Ibrahim Y Adamu[3], Audi Najib[4]

[1]Lecturer Department of Statistics, Nasarawa State University Keffi Nigeria
mohammedidrisu.nsuk.edu.ng

[2]Lecturer Department of Statistics, Nasarawa State University Keffi Nigeria
saiduraboy.nsuk.edu.ng

[3]Lecturer Department of Mathematics and Statistics, Federal Polytechnic, Nasarawa
ibrosta111[at]gmail.com

[4]Lecturer Department of Statistics Nasarawa State University, Keffi
aisyaku[at]nsuk.edu.ng

Abstract: ***In multivariate sampling, the major interest is on the problem of estimation of several population characteristics which often make conflicting demands on the sampling procedure. In this type of survey, the best allocation for one item may not in general be the best for another. There is the need to come up with compromise solution in a survey with many characteristics under study. This paper focuses on comparing some techniques of optimum sample allocation which are Yates/Chatterjee, Booth and Sedransk and Vector Maximum Criterion (VMC) on five sets of real life data stratified into six strata and two variates with desired variances using: (i.) Method of maximum variances with fixed n and (ii.) arbitrary fixing of variances. The stratum sample size nh among the classes are obtained to ascertain the criterion that will produce the smallest n. Based on the set of data collected and used for the empirical study it was discovered that Vector Maximum Criterion (VMC), Booth and Sedransk are superior to Yates/Chatterjee.***

Keywords: Multivariate stratified sampling, optimum allocation, strata, stratum - weight, true variance, vector maximum criterion

1. Introduction

In multivariate sampling, the major interest is on the problem of estimation of several population characteristics such as age, family, expenditure and mean income. These characteristics often make conflicting demands on the sampling procedure (Sukhature, 1970). A procedure that is likely to decrease the variance of the estimate of one characteristic may very well increase the estimate of another.

Khan, Ali, Raghav and Bari (2013) submitted that in multivariate stratified sampling where more than one characteristic are to be estimated, an allocation which is optimum for one characteristic may not be optimum for other characteristics. In such situations a compromise criterion is needed to work out a usable allocation which is optimum for all characteristics in some sense. Such an allocation may be called a ‘Compromise Allocation’.

Attila, (1997) said the problem of optimum allocation of sample sizes in a sample survey when a single characteristics is being studied under a given sampling procedure is that which minimizes the cost of the survey for a desired precision or the variance of the sample estimate for a given budget of the survey.

In light of this, several optimality criteria have been developed over the years by different authors in a survey where many variables are under study. However, the problem is the choice of best allocation.

Yates (1953) suggested a criterion in which the sample specifies the variances that he wants for the estimates of each variate. A more reasonable criterion suggested by Dalenius (1957) is to minimize the total cost subject to the condition that the variances of the estimates for different variables do not exceed certain pre - assigned quantities.

However, this study focuses on comparing some techniques to identify which method is superior in producing the best optimal allocation for a given desire variance.

2. Optimum Allocations Solutions and Techniques

2.1 Optimum Allocations Solutions

Compromise solution

Several optimality criteria are found in the literatures. The first was suggested by Neyman (1953). It was observed in this approach that unless the stratum variances are distributed in the same way Neyman allocation is of limited value, because on allocation which is optimum for one variable may be quite unstable for another. When this occurred, he suggested that the sample be distributed among different strata in proportion to their sizes. Cochran, (1963) and others suggested the criteria of minimizing the sum of relative variances namely:

$$\sum_{h} \left(\frac{\sum_{i=1}^{l} \frac{P_i^{2} S_i^{2}}{n_i}}{\frac{1}{h}\left(\sum_{i=1}^{l} P_i S_i\right)^{2}} \right) \qquad (2.1)$$

Subject to the condition that

$$\sum_{i=1}^{l} n_i = n \; .$$

$p_i = \frac{N_i}{N}$

Is the ratio of the number of units in the i^{th} stratum to the number of units in the population and S_i is the ratio of the number of units in the j^{th} stratum to the number of units in the population.

Loss Function Solution

In some survey, the optimum allocations for individual variates differ so much that there is no obvious compromise. Some principles are needed to determine the allocation to be used. Two useful ones suggested by Yates (1960) are presented.

The first applies to surveys with a specialized objectives, in which the loss due to an error of given size in an estimate can be measured in terms of money or utility. Let *L (Z)* denotes the loss that will be incurred in a decision through an error of amount *Z* in the estimate. Although the actual value of *Z* is not predictable in advance, the sampling theory enables us to find the frequency distribution *f (Z, n)* of *Z* which, for a specified sampling method will depend on the sample size *n*. Hence the expected loss for a given size of sample size *n* is:

$$L\,(n) = \int L\,(Z) \int (z, n) dz \qquad (2.2)$$

The purpose in taking the sample is to diminish this loss. If *C (n)* is the cost of a sample of size n, a reasonable procedure is to choose *n*, to minimize

$$C\,(n) + L\,(n) \qquad (2.3)$$

Since this is the total cost involved in taking the sample and in making decisions from its result. The choice of *n* determines both the optimum size of sample and the most advantageous degree of precision.

Iterative Solution

In this approach, the cost C= C_0+ $\Sigma C_h nh$ is minimized subject to the tolerances v_j and the conditions 0≤ *nh* ≤ *Nh*. The problem is one in nonlinear programming.

According to Chatterjee (1967), the first steps is to work out the optimum allocation for each variate separately and to find the cost satisfying its tolerance. Take the variate separately and to find the cost satisfying its tolerance. Take the variate, say Y_1 for which the cost C_1 is highest and examine whether the optimum *nh* values for Y_1 satisfy all the other *(K-1)* tolerances. If so, we use this allocation and problem is solved; because no other allocation will satisfy the tolerances V_1 for y_1 at a cost as low as C_1. If some of the tolerances are not met, the problem is more difficult.

2.2 Techniques of Optimum Sample Allocation Yates/ Chatterjee procedure: -

1. Set the desire variances to be used.
2. Obtain the resulting variances for a sample of size *n* i.e.

$$\lambda v\left(\bar{y}_{ist}\right)_{=} \sum \frac{W_h^{2} S_{ih}^{2}}{nh} = \frac{1}{n} \sum \frac{W_h^{2} S_{ih}^{2}}{\frac{nh}{n}} \qquad (2.4)$$

3. Obtain

$$\frac{{}_1nh}{n} = \frac{W_h S_{1h}}{\sum W_h S_{1h}} \quad \text{and} \quad \frac{{}_2nh}{n} = \frac{W_h S_{2h}}{\sum W_h S_{2h}}$$

4. Obtain the value of λ

5. Obtain

$$nh = \frac{n\sqrt{\lambda\left({}_1nh\right)^2 + (1-\lambda)\left({}_2nh\right)^2}}{\sum \sqrt{\lambda\left({}_1nh\right)^2 + (1-\lambda)\left({}_2nh\right)^2}} \qquad (2.5)$$

Variables used

nh is the optimum sample size in stratum *h* from variable *j*
λ is the Langrage multiplier
W_h is the stratum weight
S_h is the true variance
n is the sample size

Booth and Sedransk Procedure: -

1. Set the desire variances to be used
2. Obtain $a_{1=} \frac{V_2}{V_1+V_2}$ and $a_{2=} \frac{V_1}{V_1+V_2}$, v_1 and v_2 are the variances for variate 1 and 2 respectively.
3. Obtain $V^* = \frac{2V_1V_2}{V_1+V_2}$
4. Equate $L = a_1 v\left(\bar{y}_{1st}\right) + a_2 v\left(\bar{y}_{2st}\right) = V^*$. L is the quadratic less function.
5. Obtain $\left(\sum W_h A_h\right)^2$; $Ah = \sqrt{\sum_{i=1}^{2} a_i S_{in}^{2}}$
6. Obtain $n = \frac{\left(\sum W_h A_h\right)^2}{l}$
7. Hence, obtain $nh = n\left(\frac{W_h A_h}{\sum W_h A_h}\right)$

Vector maximum criterion (VMC) procedures:

1. Obtain the value of the efficient feasible point for a total sample size *n*,

$$nh = \frac{n\left(\sum_{i=1}^{p}\alpha_i S_{ih}{}^2\right)^{\frac{1}{2}} N_h}{\sum_{h=1}^{L} N_h\left(\sum_{i=1}^{P}\alpha_i S_{ih}{}^2\right)^{\frac{1}{2}}} = \frac{n\left(\sum a_i S^2{}_{ih}\right)^{\frac{1}{2}} W_h}{\sum W_h\left(\sum a_i S^2{}_{ih}\right)} \tag{2.6}$$

Where α_i is the weight for the variate *i* such that $\Sigma x_i = 1$

2. For several values of α_i obtain corresponding values of n
$v(\bar{y}_{1st})$, $nv(\bar{y}_{2st})$ and $\frac{v(\bar{y}_{1st})}{v(\bar{y}_{2st})}$
3. Present the value in a table called efficient point tables
4. Set the desired variances
5. Obtain the actual values of $\frac{V_1}{V_2}$
6. Obtain the value of α_i corresponding to the values of $\frac{V_1}{V_2}$
7. Draw the graphs of nv_1, nv_2 and $\frac{V_1}{V_2}$ against α_i on the same axis
8. On the graphs, trace the values of the relative variances set to $\frac{V_1}{V_2}$
9. Obtain the value of α_i and trace it to the other two curves of nv_1 and nv_2
10. Through the value of α_i obtain the corresponding values of nv_1 and nv_2 respectively, then substitute into

$$nh = \left(\frac{n\left(\sum \alpha_i S_{ih}{}^2\right)^{\frac{1}{2}} Nh}{\sum Nh\left(\sum \alpha_i S_{ih}{}^2\right)^{\frac{1}{2}}}\right) \tag{2.7}$$

4. Results and Discussion

Table 3.1: Set of data used for Analysis

Stratum	*Nh*	W_h 1	S_{1h} 2	S_{2h} 3	W_hS_{1h} 4	W_hS_{2h} 5
1	6	0.094	41.9082	18.9032	3.9394	1.7773
2	23	0.438	12.5502	17.4111	5.4970	7.6261
3	10	0.156	26.2175	29.0423	4.0899	4.5306
4	8	0.125	17.2009	21.2556	2.1501	2.6570
5	6	0.094	16.5918	26.3106	1.5596	2.4732
6	6	0.094	12.1168	18.7296	1.1390	1.7606

Table 3.2 below shows the results of the analysis obtained on the distribution of the sample sizes with relative variances based on given *n* and also the results obtained on the distribution of the sample sizes on setting arbitrary variances.

Table 3.2: Sample sizes generated for different data classified by techniques and types of relative variance.

	Based on given *n*					Based on arbitrary variance				
Techniques	D_1	D_2	D_3	D_4	D_5	D_{1A}	D_{2B}	D_{3C}	D_{4D}	D_{5E}
Yates/Chatterjee	10	11	22	12	38	22	16	26	15	62
Booth and Sedransk	10	10	18	12	37	22	16	21	15	61
VMC	10	10	18	12	37	22	15	21	15	61

The results and findings of the table 1.2 shows that VMC produced the smallest sample size *nh* i.e. VMC dominates in D_{2B} while none of the techniques dominates in (D_1, D_4, D_{1A} and D_{4D}). However, both VMC and Booth and Sendransk dominate in (D_2, D_3, D_5, D_{3C}, and D_{5E} respectively).

Hence the results show that VMC and Booth and Sedransk procedures are superior to that of Yates/Chatterjee in the sense that the procedures dominate Yates/Chatterjee in majority of the results.

5. Conclusion

In this study, it was discovered that both Booth and Sedransk is less cumbersome to compute, followed by VMC while Yates/ Chatterjee is most cumbersome to compute. Hence Booth and Sedransk and VMC are superior to Yates/ Chatterjee. The result clearly brings out the fact that the best allocation is not always obvious and that sufficient care is necessary in the choice of allocation of the sample sizes to different strata with several items.

References

[1] Ali, I, Raghav, S. Y. and Bari, A. (2013): Compromise allocation in multivariate stratified Surveys with stochastic quadratic cost function, *Journal of Statistical Computation and Simulation*, Vol.83, issue 5, Pp 962 - 976

[2] Attila, C. (1997): Optimum allocations in stratified random sampling via holder's inequality, *the Statistician*, Vol.46, No.3, Pp 439 – 441.

[3] Bethel, F. (1989): Bayes and Minimax prediction in finite population. *Journal of Statistical Planning* 60, 127 - 135.

[4] Cochran, W. G. (1963): Sampling techniques. NewYork: John Wiley and Sons2nd edition. Diaz - Garcia, J. A and Cortez, L. U. (2008): Multi - objective optimization for optimum allocation in multivariate stratified sampling survey.

[5] Dalenius, T. (1957): On the application of cluster analysis to growing season precipitation in North America, Journal *of American Statistical Association*, 8, 897 - 931.

[6] Hunt, N. And Tyrell, S. (2004): Stratified sampling Coventry university press. http: //www.conventry. ac.

uk/ec/ - nhunt/meths/strati. html (accessed February 28, 2011).

[7] Khan, M., Ali I,. Raghav, Y and Bari, A. (2012): Allocation in multivariate stratified surveys with Non - linear random cost function, *American Journal of Operations Research*, Vol.2 No 1, Pp 100 - 105

[8] Neyman, J. (1953): Some experiments in the numerical analysis of archaeological data. *Biometrika,* 53, 311 - 324

[9] Sukhature (1970): Multivariate Economic. New York: John Willy and Sons, 2nd edition Yates, F. (1960): Optimum allocation in stratified surveys. *The Statistician*, 2nd edition.