Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.5, No.9, 2015

On The Efficiency of Some Techniques For Optimum Allocation In Multivariate Stratified Survey.

AMAHIA, G.N

Department of Statistics, University of Ibadan, Nigeria,

FAWEYA, O.

Department Of Mathematical Sciences, Ekiti State University, Ado-Ekiti, Nigeria.

AYOOLA, F.J

Department of Statistics, University of Ibadan, Nigeria

Abstract

In multivariate stratified sampling, the major concern is on the problem of estimation of more than one population characteristics which often make conflicting demands on sampling technique. In this type of survey, 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. This study is focuses on the efficiency of some techniques for optimum sample allocation which are Yates/Chatterjee, Booth and Sedransk and Vector maximum criterion (VMC) on the set of real life data stratified into six strata and two variates with desired variances using: (i) method of minimum variance with fixed sample size and (ii) an arbitrary fixing of variances. The stratum sample sizes n_h among the classes were obtained to examine the criterion that will produce the smallest n. In this paper, it was discovered that VMC and Booth and Sedransk are superior to Yates/Chatterjee. Even though, no universal conclusion can be drawn, the work 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.

Keywords:- Stratified Survey; Optimum Allocation; Vector Maximum Criterion (VMC); Yates/Chatterjee;Booth and Sedransk

Introduction

In multivariate sampling, more than one population characteristics are estimated. These characteristics may be of conflicting nature (Sukhatme, 1970). When stratified sampling is used, a procedure that is likely to decrease the variance of the estimate of one characteristic may very well increase the estimate of another.

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 well defined; It is that which minimize the cost of the survey for a desired precision or the variance of the sample estimate for a given budget of the survey. Meanwhile, typical univariate optimum sample allocation strategy failed when a number of characteristics are simultaneously under study as in the most survey situation where the possibility of irreconcilable individual allocation between characters (variables) become real.

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, these includes: Neyman (1934), Dalenius (1957) Yates (1960); Kokan and Khan (1967), Chatterjee (1968), Adelakun (2001) etc.

However, this study is anchored on the efficiency of some techniques of optimum sample allocation when desired variances are set to see which method is superior in producing the best (optimal) allocation for a given desired variance.

In stratified sampling, the values of the sample size n_h in the respective strata are chosen by the sampler. They may be selected to minimize variance of $\left(V(\bar{y}_{st})\right)$for a specified cost of taking the sample or to minimize the cost for a specified value of $V(\bar{y}_{st})$. Thus, it is generally considered in these two forms:-

(i) Fixed precision-Least cost formulation:- Given the cost function $C = C_o + \sum C_h n_h$ and some prescribed value (V) for the variance of the stratified sample means, we determine the stratum sample sizes such that the cost function $C = C_o + \sum C_h n_h$ is minimized.

(ii) Fixed cost-Best precision formulation:- Given the cost function $C = C_o + \sum C_h n_h$ and a fixed budget c, we determine the stratum sample size such that the variance of the stratified sample mean is at the minimum.

Several optimality criteria are found in the literature which were contributed by various authors to allocation problems through compromise solution, loss function and iterative solution.

The use of linear programming in sample survey to determine allocations when several characters are under study was first suggested by Dalenius (1953), and Nordbotten (1956) illustrated this approach by a numerical example. Non linear programming techniques can also be used to solve allocation problems in sample surveys, and this possibility was briefly mentioned by Dalenius (1957) who suggested minimizing a weighted average of precision.

Kokan (1963) define an optimum allocation for a multivariate survey as one that minimizes the cost of obtaining estimates with error smaller than previously specified numbers at a previously specified confidence level. He then showed how information on the various stratum variances could be used to obtain near optimum allocation. The multivariate sampling problem was proposed as a non-linear multi-objective programming problem by Kokan and Khan (1967).

A compromise allocation was suggested by Cochran (1977) for various characters, whereas Omule (1985) used dynamic programming to obtain a compromise allocation. Khan et al; (1997) used integer programming to obtain a compromise solution in multivariate stratified sampling. Daiz etal (2006) proposed stochastic programming approach to the allocation problem.

Yates (1960) in his approach suggested that the sampler specifies the variances that he wants for the estimates of each variate while Chatterjee (1968) following Yates (1960), et al illustrates a method of allocation in multivariate surveys that minimize the cost of obtaining estimates with variances not bigger than previously specified numbers. The various stratum variances are assumed known.

In a related problem, Booth and Sedransk (1969) pointed out that in default of a computer program a good approximation to the solution of Yates can often be obtained.

Adelakun (2001) proposed an alternative solution seen as a compromise solution designed as vector maximum criterion (VMC) which is a modification of the criterion advanced by Yates

(1961, 1967) and Chatterjee (1972, 1987). In the VMC, normalized vectors are to be used as weight rather than arbitrary numbers used by Yates and assumed by Chatterjee VMC differs from Yates criterion because it provides a base for deciding the choice of allocation when the sample size is specified in advance. It also provides a basis for judging the reasonableness of specified precision.

2.0 THE DATA AND METHODS USED FOR THE EMPIRICAL STUDY (MATERIAL AND METHODS)

2.1 Data Source

In this empirical study, five sets of real life data were used. Each set of data was divided into two variates and was stratified into six strata. The data were drawn from a survey on education and incidence.

The first set of data shown in table 1 deals with performance of student in Mathematics by sex in Abuja Secondary School JSCE for the year 2001/02. The percentage of male that passed (X_1) was taken as the first variate while that of female was taken as the second variate (X_2) and the data were stratified into six strata as shown below:-

Table 1: Student performance in Mathematics by sex in Abuja secondary school JSCE for year 2001/2002

Stratum No	N_h	W_h	S_{1h}	S_{2h}
Abaji (1)	3	0.1000	19.9440	15.1652
Municipal (2)	12	0.4000	25.1702	23.8943
Gwagwalada (3)	5	0.1667	21.8147	38.8280
Kuje (4)	2	0.0667	53.5351	15.2947
Kwali (5)	4	0.1333	9.3707	30.1106
Bwari (6)	4	0.1333	26.6815	12.9823

The second set of data shown in Table 2 deals with percentage passed in English (X_1) and Mathematics (X_2) mock result in Abuja Secondary Schools for years 2002 and 2004. The proportion passed in English (X_1) was taken as the first variate while that of Mathematics (X_2) was taken as second variate and the data were stratified into six strata as shown below:-

Table 2: Percentage passed in English and Mathematics Mock Result in Abuja secondary school for year 2002 and 2004.

Stratum No	N_h	W_h	S_{1h}	S_{2h}
Abaji (1)	6	0.094	41.9082	18.7951
Municipal (2)	28	0.438	12.5502	17.4111
Gwagwalada (3)	10	0.156	26.2175	29.0423
Kuje (4)	8	0.125	17.2009	21.2556
Kwali (5)	6	0.094	16.5918	26.3106
Bwari (6)	6	0.094	12.1168	18.7296

The third set of data shown in Table 3 deals with poverty incidence by state for the year 1996 (X_1) and 2004 (X_2). The data on year 1996 (X_1) was taken as the first variate and that of the year 2004 (X_2) was taken as the second variate. The data were stratified into six strata as shown below:-

Table 3: Poverty Incidence by state for the year 1996 and 2004.

Stratum No	N_h	W_h	S_{1h}	S_{2h}
NW (1)	7	0.19	7.0446	15.6653
NE (2)	6	0.16	8.9556	12.5675
NC (3)	7	0.19	9.3761	16.1231
SW (4)	6	0.16	8.0928	13.7378
SE (5)	5	0.14	2.8482	9.1739
SS (6)	6	0.16	10.1103	9.0433

The fourth set of data shown in Table 4 deals with primary enrolment ratio by State and Sex. The male ratio (X_1) was taken as the first variate and female ratio (X_2) as the second variate while the data were stratified into six strata as follows:-

Table 4: Primary Enrolment ratio by state and sex.

Stratum No	N_h	W_h	S_{1h}	S_{2h}
NW (1)	7	0.19	10.7281	12.5254
NE (2)	6	0.16	11.2073	11.9057
NC (3)	7	0.19	11.3669	14.1095
SW (4)	6	0.16	3.5408	4.7972
SE (5)	5	0.14	2.3391	5.2223
SS (6)	6	0.16	2.2602	2.2548

The last set of data shown in Table 5 deals with sets of scores of 180 students in the promotion examination which were randomly selected from six schools in Municipal Area Council of Abuja.

The scores in English (X_1) was taken as the first variate while that of Physics (X_2) was used as second variate and the data were stratified into six strata as shown below:-

Table 5: Scores of 180 students in the promotion Examination from six schools in Municipal Area Council, Abuja.

Stratum No	N_h	W_h	S_{1h}	S_{2h}
1	30	0.1667	9.4034	9.9547
2	28	0.1667	7.3928	8.4547
3	30	0.1667	9.3207	8.7057
4	30	0.1667	8.1439	9.1955
5	30	0.1667	11.7671	13.3782
6	30	0.1667	13.0227	9.3105

2.2 Statistical Analysis

For each of the five sets of data used, desired variances were set using:-

(i) Method of minimum variance calculation with a given sample of size n i.e.

$$V_{min}(\bar{y}_{st}) = \frac{(\sum W_h S_{ih})^2}{n} - \frac{(\sum W_h S^2{}_{ih})}{N} \qquad (i = 1,2)$$

Where W_h is the stratum weight

$S^2{}_h$ is the true variance.

N is the total number of units.

(ii) Arbitrary variances fixed.

The three techniques namely:-

(a) Yates/Chatterjee techniques

(b) Booth and Sedransk techniques

(c) Vector maximum criterion (VMC) were applied to each set of the data with their set desired variances.

2.2.1 Yates/Chatterjee procedure:-

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

$$\lambda V(\bar{y}_{st}) = \sum \frac{W^2{}_h S^2{}_{ih}}{n_h} = \frac{1}{n} \sum \frac{W^2{}_h S^2{}_{ih}}{\frac{n_h}{n}} - - - - - - - - - \qquad (2)$$

3. Obtain $\frac{1^{n_h}}{n} = \frac{W_h S_{1h}}{\sum W_h S_{ih}}$ and $\frac{2^{n_h}}{n} = \frac{W_h S_{2h}}{\sum W_h S_{ih}}$ $- - - - - - - - - - - - - -$ (3)

4. Obtain the values of λ

5. Obtain $n_h = \frac{n\sqrt{\lambda(1^{n_h})^2 + (1-\lambda)(2^{n_h})^2}}{\sum \sqrt{\lambda(1^{n_h})^2 + (1-\lambda)(2^{n_h})^2}}$ …………………………………… (4)

2.2.2 Booth and Sedransk Procedure:-

1. Set the desired variances to be used.
2. Obtain $a_1 = V_2/V_1 + V_2$ and $a_2 = V_1/V_1 + V_2$.where $V_1$and V_2 are the variances for variate 1 and 2 respectively.
3. Obtain $V^* = 2\,V_1 V_2/V_1 + V_2$ …………………………………………… ..(5)
4. Equate $L = a_1 V(\bar{y}_{1st}) + a_2 V(\bar{y}_{2st}) = V^*$………………………………… ..(6)
   Where L is the quadratic loss function.
5. Obtain $(\sum W_h A_h)^2$ where $A_h = \sqrt{\sum_{i=1}^{2} a_i S^2{}_{ih}}$ …………………………… (7)
6. Obtain $n = (\sum W_h A_h)^2/L$……………………………………………………(8)
7. Hence, obtain n_h=$n\left(\frac{W_h A_h}{\sum W_h A_h}\right)$ ……………………………………………… (9)

2.2.3 Vector Maximum Criterion (VMC) procedures.

1. Obtain the value of the efficient feasible point for a total sample size of n. i.e.

$$n_h = \frac{n(\sum_{i=1}^{2}\alpha_i S^2{}_{ih})^{1/2} N_h}{\sum_{h=1}^{L}(\sum_{i=1}^{2}\alpha_i S^2{}_{ih})^{1/2}} = \frac{n(\sum\alpha_i S^2{}_{ih})^{1/2} W_h}{\sum W_h(\sum_{i=1}^{p}\alpha_i S^2{}_{ih})^{1/2}} \text{ .... } \quad .(10)$$

Where α_i is the weight for the variate i such that $\sum \alpha_i$ =1

$$V(\bar{y}_{st}) = \frac{\sum W^2{}_h S^2{}_{ih}}{n_h} = \frac{(\sum W^2{}_h S^2{}_h)\sum W_h(\sum\alpha_i S^2{}_{ih})^{1/2}}{W_h(\sum\alpha_i S^2{}_{ih})^{1/2}} \text{..........................}(11)$$

2. For several values of α_i obtain corresponding values of

   n$V(\bar{y}_{1st})$, n$V(\bar{y}_{2st})$ and $nV(\bar{y}_{1st})/nV(\bar{y}_{2st})$.

3. Present the values in a table called efficient point tables.
4. Set the desired variances.
5. Obtain the actual values of V_1/V_2.
6. Obtain the value of α_i corresponding to the values of V_1/V_2.
7. Draw the graphs of nV_1, nV_2 and V_1/V_2 against α_i on the same axis.
8. On the graphs, trace the values of the relative variances set to 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

$$n_h = \frac{n(\sum\alpha_i S^2{}_{ih})^2 N_h}{\sum N_h(\sum\alpha_i S^2{}_{ih})^{1/2}} \text{..}(12)$$

3.0 DATA ANALYSIS

3.1 Analysis of data set 1 by fixing n:-

3.1.1 Using Yates/Chatterjee procedures based on setting relative variances with n=7

Table 4.8: Calculation of $W_h S^2{}_{1h}$ and $W_h S^2{}_{2h}$

Stratum	N_h	W_h	S_{1h}	S_{2h}	$W_h S_{1h}$	$W_h S_{2h}$	$(W_h S_{2h})^2 / W_h S_{1h}$	$(W_h S_{1h})^2 / W_h S_{2h}$	$W_h S^2{}_{1h}$	$W_h S^2{}_{2h}$
1	3	0.100	19.9440	15.1652	1.9944	1.5165	1.1531	2.6229	39.7763	22.9983
2	12	0.400	25.1702	23.8943	10.0681	9.5577	9.0732	10.6058	253.4156	228.3750
3	5	0.167	21.8147	38.8280	3.6431	6.4848	11.5431	2.0467	79.4721	251.7715
4	2	0.067	53.5351	15.2947	3.5869	1.0247	0.2927	12.5557	192.0225	15.6732
5	4	0.133	9.3707	30.1106	1.2463	4.0047	12.8682	0.3879	11.6787	120.5842
6	4	0.133	26.6815	12.9823	3.5486	1.7268	0.8403	7.2924	94.6830	22.4158
				TOTAL	24.0874	24.3152	35.7706	35.5114	671.0482	661.8189

$$V_1 \le 60.5178$$

$$V_2 \le 62.4007$$

$$V_1(\bar{y}_{1st}) = \frac{(24.0874)^2}{n} \Longrightarrow n = 9.5873$$

$$V_2(\bar{y}_{2st}) = \frac{(24.3152)^2}{n} \Longrightarrow n = 9.4747$$

If allocation 2 is used with n=9.4747, the variance obtained for y_1 is

$V_2(\bar{y}_{2st}) = \frac{(24.3152)(35.5114)}{9.4747}$

=91.1339

This value is larger than 60.5178 specified for V_1 , hence we seek a compromise allocation that satisfies both tolerances exactly.

Using langrange multipliers $\lambda_1 \, and \, \lambda_2$, we find the values of n_h.

For any value of λ, we have $V(\bar{y}_{1st}) = \frac{\phi_1(\lambda)}{n}$, $V(\bar{y}_{2st}) = \frac{\phi_2(\lambda)}{n}$

Hence, we shall find λ and n such that

$$\frac{\phi_1(\lambda)}{n} = V_1 = 60.5178$$

$$\frac{\phi_2(\lambda)}{n} = V_2 = 62.4007$$

For $\phi_1(\lambda)$, when $\lambda = 1$, $\phi_1(\lambda) = 580.2028$, when $\lambda = 0$, $\phi_2(\lambda) = 863.4668$.

This gives a parabolic approximation as:

$\frac{\phi_1(\lambda)}{n}$ =580.2028+283.2640$(1-\lambda)^2$=60.5178 ---(1)

=863.4668-566.528λ+283.2640λ^2=60.5178n --(2)

For $\phi_2(\lambda)$ when λ=1, $\phi_2(\lambda)$=591.2290, when $\lambda = 0$, $\phi_1(\lambda)$=861.6208

Also this gives a parabolic approximation as

$\frac{\phi_2(\lambda)}{n} = \frac{591.2290+270.3968\lambda^2}{n} = 62.4007$ ---(3)

$\Longrightarrow$ 591.2290+270.3918λ^2=62.4006n

$\Longrightarrow$ 9.4747+4.33λ^2=n --(4)

Substituting from (4) in (2) and solving for λ and n we obtain

λ=0.5225≅ 0.52, $1-\lambda$=0.48

Let $r_n = \sqrt{\lambda\left(\frac{1n_h}{n}\right)^2 + (1-\lambda)\left(\frac{2n_h}{n}\right)^2}$

Where $\frac{1n_h}{n} = \frac{W_h S_{1h}}{\sum W_h S_{1h}}$ and $\frac{2n_h}{n} = \frac{W_h S_{2h}}{\sum W_h S_{2h}}$

We have table 4.9 below :-

Table 4.9: Calculation of n_h

Stratum	$\left(\frac{1n_h}{n}\right)^2$	$\left(\frac{2n_h}{n}\right)^2$	r_h	$\frac{r_h}{\sum r_h} = \frac{n_h}{n}$	$\frac{W^2{}_h S^2{}_{1h}}{n_n/n}$	$\frac{W^2{}_h S^2{}_{2h}}{n_n/n}$	n_h
1	0.0069	0.0039	0.0739	0.649	53.8245	35.4356	0.7139≅ 1
2	0.1747	0.1545	0.4062	0.3566	284.2587	256.1683	3.9226≅ 4
3	0.0229	0.0711	0.2146	0.1884	70.4468	233.2093	2.0724≅ 2
4	0.0222	0.0018	0.1114	0.0978	131.5527	10.7363	1.0758≅ 1
5	0.0027	0.0271	0.1200	0.1054	14.7368	152.1596	1.1594≅ 1
6	0.0217	0.0710	0.2130	0.1869	67.3759	15.9542	2.0559≅ 2
		TOTAL	**1.1391**	**1.0000**	**622.1954**	**693.6633**	**11**

To obtain the resulting variance from the sample, we use

$$\lambda V(\bar{y}_{ist}) = \sum \frac{W^2{}_h S^2{}_{ih}}{n_h} = \frac{1}{n}\sum \frac{W^2{}_h S^2{}_{ih}}{n_h/n}$$

Thus $\lambda V(\bar{y}_{1st}) = \frac{622.1954}{n} = 60.5178$

$$\Rightarrow n = 10.2812 \cong 10$$

$$\lambda V(\bar{y}_{2st}) = \frac{693.6633}{n} = 62.4007$$

$\Rightarrow n = 11.1163.$

The two values of n are so close that we accept this allocation and take

$n = 11.1163 \cong 11.$

3.1.2 Using Booth and Sedransk procedure based on setting relative variance with n=7

$V_1 \leq 60.5178$ and $V_2 \leq 62.4007$

$$a_1 = \frac{V_2}{V_1 + V_2} = \frac{62.4007}{122.9185} = 0.5077$$

$$a_2 = \frac{V_1}{V_1 + V_2} = \frac{60.5178}{122.9185} = 0.4923$$

$$V^* = \frac{2V_1V_2}{V_1 + V_2} = 61.4448$$

$$L = a_1V(\bar{y}_{1st}) + a_2V(\bar{y}_{2st}) = V^*$$

$$= 0.5077V(\bar{y}_{1st}) + 0.4923V(\bar{y}_{2st}) = 61.4448$$

Using $A_h = \sqrt{a_1S^2{}_{1h} + a^2{}_2S^2{}_{2h}}$ and $n_h = \frac{n(W_hA_h)}{\sum W_hA_h}$

Where $n = \frac{(\sum W_hA_h)^2}{L}$

We have the table 4.10:-

Table 4.10: Calculation of n_h

Stratum	W_h	${S_{1h}}^2$	${S_{2h}}^2$	A_h	W_hA_h	n_h
1	0.100	397.7631	315.1651	17.7529	1.7753	0.7033≅ 1
2	0.400	633.5390	602.7221	24.5504	9.8201	3.8908≅ 4
3	0.167	475.8811	983.8009	31.3656	5.2381	2.0754≅ 2
4	0.067	2866.0069	1570.2357	39.6262	2.6550	1.0519≅ 1
5	0.133	87.8100	490.9238	22.1568	2.9469	1.1676≅ 1
6	0.133	168.5401	444.4043	21.0809	2.8038	1.1109≅ 1
				TOTAL	25.2392	**10**

$$n = \frac{(25.2392)^2}{61.4448} \cong 10$$

3.1.3 Using Vector Maximum Criterion Procedure (V.C.M)

$$n_h = \frac{(\sum \alpha_i S^2{}_{ih})^{1/2} N_h n}{\sum N_{h(\sum \alpha_i S^2{}_{ih})}{}^{1/2}} = \frac{(\sum \alpha_i S^2{}_{ih})^{1/2} W_h n}{\sum W_{h(\sum \alpha_i S^2{}_{ih})}{}^{1/2}}$$

Let $A_h = (\sum \alpha_i S^2{}_{ih})^{1/2} W_h$

$$\Rightarrow n_h = \frac{A_h}{\sum A_h}$$

Then we have table 4.12:-

Table 4.12: Calculation of A_h values for various α_i

Stratum	W_h	$S^2{}_{1h}$	$S^2{}_{2h}$	W_hS_{1h}		$\alpha_1 = 0$ $\alpha_2 = 1$	$\alpha_1 = 0.1$ $\alpha_2 = 0.9$	$\alpha_1 = 0.2$ $\alpha_2 = 0.8$	$\alpha_1 = 0.3$ $\alpha_2 = 0.7$
1	0.100	397.9833	229.9833	3.9776	1.6743	1.5165	1.5709	1.6234	1.6743
2	0.400	633.5390	570.9376	101.3662	9.7136	9.5577	9.6100	9.6619	9.7136
3	0.167	475.8811	1507.6136	13.2718	5.7805	6.2585	6.2585	6.0242	5.7805
4	0.067	2866.0069	233.4278	12.8655	2.1435	1.0247	1.4939	1.8475	2.1435
5	0.136	87.8100	87.8100	1.5533	3.4194	3.8196	3.8196	3.6250	3.4194
6	0.133	711.9024	168.5401	12.5928	2.4217	1.7266	1.9856	2.2144	2.4217
			TOTAL	**145.6272**	**25.1530**	**24.3145**	**24.7385**	**24.9964**	**25.1530**

Stratum	$\alpha_1 = 0.4$ $\alpha_2 = 0.6$	$\alpha_1 = 0.5$ $\alpha_2 = 0.5$	$\alpha_1 = 0.6$ $\alpha_2 = 0.4$	$\alpha_1 = 0.7$ $\alpha_2 = 0.3$	$\alpha_1 = 0.8$ $\alpha_2 = 0.2$	$\alpha_1 = 0.9$ $\alpha_2 = 0.1$	$\alpha_1 = 1$ $\alpha_2 = 0$	$\alpha_1 = 0.52$ $\alpha_2 = 0.48$	$\alpha_1 = 0.78$ $\alpha_2 = 0.22$
1	1.7236	1.7716	1.8184	1.8639	1.9084	1.9519	1.9944	1.7811	1.8996
2	9.7651	9.9162	9.8671	9.9177	9.9681	10.0182	10.0681	9.8264	9.9580
3	5.5260	5.2592	4.9781	4.6802	4.3620	4.0186	3.6481	5.2042	4.4274
4	2.4034	2.6378	2.8530	3.0530	3.2407	3.4182	3.4182	2.6822	3.2041
5	3.2006	2.9657	2.7105	2.4287	2.1095	1.2463	1.2463	2.9165	2.1771
6	2.6126	2.7905	2.9577	3.1160	3.2666	3.5486	3.5486	2.8248	3.2370
TOTAL	**25.2313**	**25.2410**	**25.1848**	**25.0595**	**24.8553**	**24.5499**	**24.0874**	**25.2352**	**24.9032**

Using $V(\bar{y}_{1st}) = \frac{\sum_{h=1}^{L}(W^2{}_hS^2{}_{ih})}{n_h} = \frac{\sum(W^2{}_hS^2{}_{ih})(\sum A_h)}{A_hn}$

$$\Rightarrow nV(\bar{y}_{1st}) = \frac{\sum(W^2{}_hS^2{}_{ih})(\sum A_h)}{A_h}$$

We obtained the table 4.13:-

Table 4.13

VMC TABLE OF EFFICIENT POINTS

α_i	$nV(\bar{y}_{1st})$	$nV(\bar{y}_{2st})$	$V(\bar{y}_{1st})/V\bar{\bar{y}}_{2st}$
0	863.4566	591.9111	1.4588
0.1	756.0416	595.9775	1.2686
0.2	705.4881	604.6556	1.1668
0.3	673.1817	615.3139	1.0940
0.4	649.6622	627.9226	1.0346
0.5	631.2525	642.9668	0.9818
0.6	616.1906	661.3997	0.9316
0.7	603.5691	708.9544	0.8571
0.8	592.9788	716.9594	0.8271
0.9	584.5664	765.9021	0.7632
1.0	580.1988	861.5750	0.6734

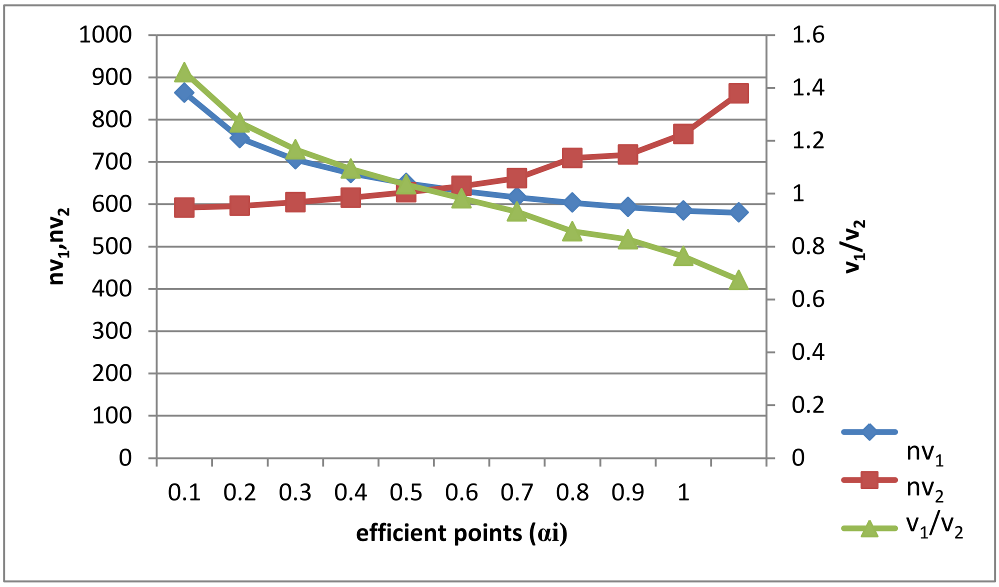
Using VMC based on setting relative variances with n=7

$$V(\bar{y}_{1st}) \leq 60.5178$$

$$V(\bar{y}_{2st}) \leq 62.4007$$

Thus, $V_1/V_2 = 0.9698$

Graphs of nV_1 ,nV_2 and V_1/V_2 against efficient point (α_i)



From the graph, we obtain the values of α_1 ,nV_1 , nV_2 corresponding with $V_1/\mathrm{V}_2 = 0.9698$ as 0.52, 627.80 and 646.00 respectively.

For nV_1 ,

$$n = \frac{627.8}{60.5178} = 10.3738 \cong 10$$

For nV_2 ,

$$n = \frac{646}{62.4007} = 10.3524 \cong 10$$

For $\alpha_1 = 0.52$, $\alpha_2 = 0.48$

$$n_1 = \frac{1.7811}{25.2352} \times 10.3524 = 0.7306 \cong 1$$

$$n_2 = \frac{9.8264}{25.2352} \times 10.3524 = 4.0311 \cong 4$$

$$n_3 = \frac{5.2042}{25.2352} \times 10.3524 = 2.1350 \cong 2$$

$$n_4 = \frac{2.6822}{25.2352} \times 10.3524 = 1.1003 \cong 1$$

$$n_5 = \frac{2.9165}{25.2352} \times 10.3524 = 1.1965 \cong 1$$

$$n_6 = \frac{2.8248}{25.2352} \times 10.3524 = 1.1588 \cong \frac{1}{10}$$

With the use of Yates/Chatterjee, Booth and Sedransk, and Vector Maximum Criterion (VMC) procedures for optimum allocation problems in multivariate survey on five sets of numerical real life data, the summary of the tabulated results are shown below

TABULATED RESULTS OF THE DATA ANALYSIS

Table 1 shows the results obtained on the distribution of the sample sizes with relative variances based on given n from the five data sets.

Data set	**Techniques**	**Strata**						
		1	2	3	4	5	6	**Total**
1	Yates/Chatterjee	1	4	2	1	1	2	11
	Booth & Sedransk	1	4	2	1	1	1	10
	VMC	1	4	2	1	1	1	10
2	Yates/Chatterjee	2	3	2	1	1	2	10
	Booth & Sedransk	2	3	2	1	1	1	10
	VMC	2	3	2	1	1	1	10
3	Yates/Chatterjee	4	3	4	3	5	3	22
	Booth & Sedransk	4	3	4	3	1	3	18
	VMC	4	3	4	3	1	3	18
4	Yates/Chatterjee	3	3	3	1	1	1	12
	Booth & Sedransk	3	3	3	1	1	1	12
	VMC	3	3	3	1	1	1	12
5	Yates/Chatterjee	5	5	6	6	8	8	38
	Booth & Sedransk	6	5	6	5	8	7	37
	VMC	6	5	6	5	8	7	37

TABLE 2 shows the results obtained on the distribution of the sample size on setting arbitrary variances.

Data set	**Techniques**	**Strata**						
		1	2	3	4	5	6	**Total**
1	Yates/Chatterjee	1	6	3	2	1	3	16
	Booth & Sedransk	1	6	3	2	2	2	16
	VMC	1	6	3	2	1	2	15
2	Yates/Chatterjee	4	7	5	3	2	1	22
	Booth & Sedransk	3	7	5	3	2	2	22
	VMC	3	7	5	3	2	1	22
3	Yates/Chatterjee	5	3	5	4	6	3	26
	Booth & Sedransk	4	4	5	3	2	3	21
	VMC	4	4	5	3	1	4	21
4	Yates/Chatterjee	4	3	5	1	1	1	15
	Booth & Sedransk	4	3	5	1	1	1	15
	VMC	4	4	4	1	1	1	15
5	Yates/Chatterjee	10	8	10	9	13	12	62
	Booth & Sedransk	10	8	9	9	13	12	61
	VMC	10	8	9	9	13	12	61

TABLE 3:-

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_1	D_2	D_3	D_4	D_5
Yates/Chatterjee	11	10	22	12	38	16	22	26	15	62
Booth and Sedransk	10	10	18	12	37	16	22	21	15	61
VMC	10	10	18	12	37	16	22	21	15	61

SUMMARY AND CONCLUSION

Summary, the results in the tables show that VMC and Booth and Sedransk procedures are superior to Yates/Chatterjee in the sense that the procedures dominate Yates/Chatterjee in majority of the results in terms of sample sizes with relative variances based on given n and on setting arbitrary variances.

CONCLUSION

In this research work, we discovered based on the set of data collected and used for the empirical studies, that VMC and Booth and Sedransk are superior to Yates/Chatterjee.

It was also discovered that some strata has one observation in some tables, hence there will be no need to estimate. Then, we collapse the affected strata to form a stratum.

Even though, no general conclusion can be drawn, the study 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

Adelakun, A.A. (2001) “Optimum Allocation in Multi-item sample Design: Some Contributions”. Unpublished Ph.D. Thesis, University of Ibadan, Ibadan.

Attila, C. (1997) “Optimum Allocation in Stratified Random Sampling Via Holder’s inequality”. The Statistician (1997) Vol. 46. No 3 pp. 439-441.

Benne, F.S and David O.Y. (1972) “Allocation in Stratified Sampling as a Game”.Journal of the American Statistical Association September (1972), Vol. 67. No 339 pp. 684-686.

Booth, G. and Sedransk, J.,(1967) “Designing several factor analytical surveys”, Proceeding of the Social Statistical Association,

Chatterjee, .S. (1968) “Multivariate Stratified Surveys”. Journal of the American statistical Association Vol 63, No.322, PP. 530-534.

Cochran, W.G. (1963) “Sampling Techniques” John Wiley and Sons New York. Second Edition.

Dalenius, T. (1950). The problem of optimum stratification. Skandinavisk Aktuarietidskrift **33,** 203-213.

Dalenius, T. (1953). The multivariate sampling problem. Skandinavisk Aktuarietidskrift **36,** 92-102.

Khan, M.F. et al (2012). "Allocation in Multivariate Stratified Surveys with Non-linear Random Cost Function". American Journal of Operations Research, Vol. 2 pp.100-105.

Kokan, A.R. and Khan, S. (1967). Optimum Allocation in Multivariate surveys: An analytical solution. Journal of the Royal Statistical Society, Series B, **29**, 115-125.

Kozak, M. and Wang H. (2010). "On stochastic optimization in sample allocation among strata". METRON-International Journal of Statistics, Vol.68 No 1 pp.95-103

Omule, .S. A. Y. (1985). Optimum design in Multivariate sampling. Biometrical Journal **27**, 907-912.

Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and Asok, C. (1984). Sampling Theory of Surveys with Applications. Third edition. Ames, Iowa: Iowa State University Press.

Winkler, W. E. (2009). "Sample Allocation and Stratification". U.S. Bureau of Census, Statistical Research Division Report RRS 2009-08.

Yates, F. (1966). Sampling Methods for Censuses and Surveys. Third Edition. London Griffin.