

A NOTE ON OPTIMUM ALLOCATION IN STRATIFIED RANDOM SAMPLING

A K P C Swain¹

Former Professor of Statistics, Utkal University, Bhubaneswar -751004, India

ABSTRACT

In this paper a goal programming technique of finding a compromise optimum allocation in stratified random sampling is suggested, which near about minimizes both the variance of the estimate and variance of the estimated variance.

KEY WORDS: Stratified random sampling, optimum allocation, goal programming.

MSC:62D05

RESUMEN

En este trabajo se sugiere el uso de la técnica de programación por metas para hallar una fijación óptima de compromiso en el muestreo estratificado aleatorio, el cual minimiza aproximadamente la varianza del estimado y la varianza estimada.

1. INTRODUCTION

In stratified random sampling the most important problem faced by the sample survey practitioner is to allocate the total sample size into different strata. Suppose there is a finite population of size N units divided into L

strata of sizes N_h $h = 1, 2, \dots, L$, $\sum_{h=1}^L N_h = N$. A sample of size n_h is selected from the h th stratum

following any sampling design to observe character y . The total sample size $\sum_{h=1}^L n_h = n$ is fixed in advance.

Let Y_{hi} be the value of the character y for the i th unit in the h th stratum, $i = 1, 2, \dots, N_h$, $h = 1, 2, \dots, L$

Define $\bar{Y} = \sum_{h=1}^L W_h \bar{Y}_h$, where $W_h = N_h / N$ and $\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} Y_{hi}$.

Under simple random sampling without replacement an unbiased estimate of population mean $\bar{Y}$ is given by

$$\bar{y}_{st} = \sum_{h=1}^L W_h \bar{y}_h,$$

where the sample mean of the h th stratum $\bar{y} = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$ is an unbiased estimate of $\bar{Y}_h$. The sampling

variance of $\bar{y}_{st}$ is given by

$$\begin{aligned} V \bar{y}_{st} &= \sum_{h=1}^L \left(\frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 S_h^2 \\ &= \sum_{h=1}^L W_h^2 S_h^2 / n_h \end{aligned}$$

¹ e-mail: akpcs@rediffmail.com

for large N_h $h = 1, \dots, L$

2. OPTIMUM ALLOCATION UNDER MULTI-OBJECTIVES

Neyman(1934) proposed a method to determine optimum sample sizes for the strata by minimizing $V \bar{y}_{st}$,

subject to $\sum_{h=1}^L n_h = n$ which gives

$$n_{h \text{ opt}}^1 = n \frac{W_h S_h}{\sum W_h S_h}, \quad h = 1, \dots, L.$$

Another optimum allocation(Ross,1961) may be derived by maximizing the stability of the estimated variance of the stratified estimate or otherwise by minimizing the variance of the estimated variance of $\bar{y}_{st}$, given by

$$V \text{ v } \bar{y}_{st} = V \left[\sum W_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) s_h^2 \right]$$

where s_h^2 is the sample variance computed from the h th stratum. For large N_h $h = 1, \dots, L$,

$$V \text{ v } \bar{y}_{st} \cong \sum_{h=1}^L \frac{W_h^4 S_h^4}{n_h^3} \beta_{2h} - 1$$

where β_{2h} is the coefficient of kurtosis of the character y under study in the h th stratum. The optimum sample sizes which minimize $V \text{ v } \bar{y}_{st}$ using usual technique for fixed $\sum_h n_h = n$ is given by

$$n_{h \text{ opt}}^2 = n \frac{W_h S_h \beta_{2h} - 1^{1/4}}{\sum_h W_h S_h \beta_{2h} - 1^{1/4}}$$

Now we have two sets of optimum sample sizes $n_{h \text{ opt}}^1$ and $n_{h \text{ opt}}^2$, that is, one by minimizing the variance of the estimate and another by minimizing the variance of the estimated variance of the estimate for fixed sample size n . In this paper goal programming approach of finding compromise allocations is discussed and is compared with Chatterjee's(1967) technique of finding compromise allocations.

3.COMPROMISE OPTIMUM ALLOCATIONS

The simplest one is due to Cochran(1963)to take average of two optimum allocations, if the allocations do not vary widely, that is,

$$n_{h \text{ opt}}^c = \frac{n_{h \text{ opt}}^1 + n_{h \text{ opt}}^2}{2}$$

Chatterjee (1967) suggested a compromise allocation for the multi-objectives as

$$n_{h \text{ opt}}^{c*} = n \frac{\sqrt{n_{h \text{ opt}}^{12} + n_{h \text{ opt}}^{22}}}{\sum \sqrt{n_{h \text{ opt}}^{12} + n_{h \text{ opt}}^{22}}}$$

This choice minimizes the proportional increase due to use of actual allocation n_h instead of optimum allocation,

$$\frac{1}{Lk} \sum_{j=1}^k \sum_{h=1}^L \frac{n_h - n_{h \text{ opt } j}^2}{n_h}$$

where k is the number of objectives and L is the number of strata.

Alternatively, one may obtain compromise solutions using **goal programming technique** setting the goal for each objective as the minimum variance under the individual optimum allocation. Goal programming was first formulated by Charnes, Cooper and Ferguson (1955) who considered its application in single objective linear programming problem and subsequently found many engineering and industrial applications (Sciederjans, 1995, Deb, 2001 and Jones and Tamiz, 2010) to find a compromise solution which simultaneously satisfy a number of goals or objectives to the extent possible. The main purpose behind the goal programming methodology is to find solutions to multi-objective problem when there does not exist solution which meets the targets in all objectives and the task therefore boils down to find solutions which minimize deviations from targets.

4. NUMERICAL ILLUSTRATION

Consider the following summary data (Table-1) pertaining to complete enumeration of 1000 villages in a certain district. The villages were stratified according to their agricultural area into four strata. The population values of the strata standard deviations and strata coefficients of kurtosis of the area under wheat are given below along with the values of strata sizes. Number of villages in the sample = $n = 340$

Table 1

Strata	Size N_h	W_h	S_h	(1) β_{2h}	(2) β_{2h}	(3) β_{2h}	(4) β_{2h}
1	285	0.285	56.3	1.5	5.5	1.5	3.5
2	355	0.355	116.4	2.5	3.5	2.0	5.5
3	226	0.226	186.0	3.5	2.5	2.5	7.5
4	134	0.134	363.1	5.5	1.5	3.0	9.5

Define $V_1 = \sum W_h^2 S_h^2 / n_h$ and $V_2 = \sum W_h^4 S_h^4 \beta_{2h} - 1 / n_h^3$

Case 1: $\beta_2 = 1.5, 2.5, 3.5, 5.5$

The minimizing V_1 and V_2 separately the individual Neyman optimum allocations are respectively

$$n^1 = 37, 95, 96, 112 \text{ and } n^2 = 25, 85, 98, 132$$

$$L_1 = V_1(\min) = \frac{1}{n} \sum W_h S_h^2 = 64.47564$$

$$L_2 = V_2(\min) = \frac{1}{n^3} \sum W_h S_h^4 \beta_{2h} - 1^{1/4} = 28.50126$$

For the application of goal programming the optimum values of n_1, n_2, n_3 and n_4 are, so as to satisfy the following goals.

$$\sum W_h^2 S_h^2 / n_h \geq 64.47564 \text{ (i.e. } V_1(\min))$$

$$\sum W_h^4 S_h^4 \beta_{2h} - 1 / n_h^3 \geq 28.50126 \text{ (i.e. } V_2(\min))$$

$$n_1 + n_2 + n_3 + n_4 = 340, 2 \leq n_1 \leq 285, 2 \leq n_2 \leq 355, 2 \leq n_3 \leq 226, 2 \leq n_4 \leq 134,$$

where n_1, n_2, n_3, n_4 are integers.

Using positive deviational variables we formulate the goal programming model as

Min: $dp1 + dp2$

Subject to $\sum W_h^2 S_h^2 / n_h - dp1 = 64.47564$ and

$$\sum W_h^4 S_h^4 \beta_{2h} - 1 / n_h^3 - dp2 = 28.50126$$

with other constraints as above.

Using Lingo Software the compromise optimum values of n_h 's using goal programming model are :

$$n_1 = 29, n_2 = 88, n_3 = 97 \text{ and } n_4 = 126.$$

The resulting variances V_1 and V_2 under compromise allocations using goal programming model are

$$V_{1c} = 65.28659 \text{ and } V_{2c} = 28.93667$$

The compromise solutions using Chatterjee's technique are $n_1 = 31$, $n_2 = 90$, $n_3 = 97$, $n_4 = 122$.

As such, the resulting variances are $V'_{1c} = 64.89865$ and $V'_{2c} = 29.55304$.

Table 2. Variances under different allocations

Cases	Individual variances under Neyman optimum allocation		Variances under compromise allocation using goal programming		Variances under Chatterjee's compromise allocation	
-----	$V_{1\min}$	$V_{2\min}$	V_{1c}	V_{2c}	V'_{1c}	V'_{2c}
Case 1	64.47564	28.50126	65.28659	28.93667	64.89865	29.55304
Case 2	64.47564	17.85832	65.36542	18.40473	65.04679	18.80187
Case 3	64.47564	16.34300	64.73394	16.50806	64.64065	16.62011
Case 4	64.47564	71.89904	64.90418	71.97986	64.58266	73.04560

Case 2: $\beta_2 = 5.5, 3.5, 2.5, 1.5$

Individual optimum values: $V_1(\min) = 64.47564$ with $\underline{n} = 37, 95, 96, 112$ and

$V_2(\min) = 17.85832$ with $\underline{n} = 49, 109, 97, 85$

Compromise solutions using goal programming model are $\underline{n} = 45, 103, 96, 96$

with resulting variances $V_{1c} = 65.36542$ and $V_{2c} = 18.40473$

As regards Chatterjee's compromise allocation, $\underline{n} = 43, 102, 96, 99$ with variances $V'_{1c} = 65.04679$ and $V'_{2c} = 18.80187$.

Table 3. Loss of Relative Efficiency(%) of the compromise allocations compared to individual optimum allocations

Cases	Allocation using goal programming			Allocation using Chatterjee's Technique		
	Goal 1	Goal 2	Total	Goal 1	Goal 2	Total
Case 1	1.2421	1.5047	2.7468	0.6518	3.5590	4.2108
Case 2	1.3612	2.9689	4.3301	0.8781	5.0183	5.8965
Case 3	0.3990	0.9999	1.3989	0.2553	1.6673	1.9226
Case 4	0.6603	0.1123	0.7726	0.1653	1.5696	1.7353

Case 3: $\beta_2 = 1.5, 2.0, 2.5, 3.0$

Individual optimum values: $V_1(\min) = 64.47564$ with $\underline{n} = 37, 95, 96, 112$ and $V_2(\min) = 16.34300$

with $\underline{n} = 28, 88, 99, 124$

Compromise solutions using goal programming model are $\underline{n} = 32, 91, 98, 119$

with variance $V_{1c} = 64.73394$ and $V_{2c} = 16.50806$

Chatterjee's compromise solutions are computed as $\underline{n} = 33, 92, 97, 118$ with variance $V'_{1c} = 64.64065$ and $V'_{2c} = 16.62011$.

Case 4: $\beta_2 = 3.5, 5.5, 7.5, 9.5$

Individual optimum values: $V_1(\min) = 64.47564$ with $\underline{n} = 37, 95, 96, 112$ and $V_2(\min) = 71.89904$ with $\underline{n} = 30, 89, 99, 122$

Compromise solutions using goal programming model are $\underline{n} = 31, 89, 99, 121$ with variance

$V_{1c} = 64.90418$ and $V_{2c} = 71.97986$

Further, Chatterjee's compromise allocations are computed as $\underline{n} = 34, 92, 97, 117$ with $V_{1c}' = 64.58266$ and $V_{2c}' = 73.04560$.

5. CONCLUSION

As the individual optimum allocations under two different objectives vary widely Cochran's (1963) rule is not taken into account to compute compromise allocations and as such Chatterjee's technique and goal programming technique for finding compromise allocations in multi-objective problem are compared through the given numerical illustration. The numerical illustration shows that the total loss of efficiency of the compromise allocations using goal programming model happens to be less than the total loss of efficiency using compromise allocations as proposed by Chatterjee (1967). As theoretical comparison between two alternative approaches is difficult to be carried out, further numerical investigations are necessary to arrive at some stable conclusions. The present paper is intended to suggest an alternative method of obtaining compromise solution following goal programming approach for multi-objective optimization.

RECEIVED JULY 2012
REVISED OCTOBER, 2012

REFERENCES

- [1] CHATTERJEE, S. (1967): A note on optimum stratification. **Skand. Akt.**, 50, 40-44.
- [2] CHARNES, A., COOPER, W.W. and FERGUSON, R (1955): Optimal estimation of executive compensation by linear programming, **Management Sciences**, 1, 138-151.
- [3] COCHRAN, W. G. (1963): **Sampling Techniques**. John Wiley and Sons, New York.
- [4] DEB, K. (2001): Solving Goal Programming Problems using Multi-objective Genetic Algorithms, **Journal of the Operational Research Society**, 52, 291-302.
- [5] JONES, D.F. and TAMIZ, M. (2010): **Practical Goal Programming**. Springer, Berlin.
- [6] NEYMAN, J. (1934): On the two different aspects of the representative methods: the method of stratified sampling and the method of purposive selection. **J. Roy. Statist. Soc.**, 97, 558-606.
- [7] ROSS, A. (1961): Variance estimates in optimum sample designs. **J. Amer. Stat. Assoc.**, 56, 135-142.
- [8] SCNIEDERJANS, M.J. (1995): **Goal Programming Methods and Applications**. Kluwer Publishers, Boston.