Research Article

Minimal equations to find OBP in the estimation of pooled variance under equal allocation and fixed allocation

P B Bharate[1*], P C Gupta[2], J C Agrawal[3]

[1]Associate Professor and Head, [3]Assistant Professor, Department of Statistics, Pratap College Amalner Dist. Jalgaon, Maharashtra, INDIA.
[2]Ex. Professor and Head, Department of Statistics, Veer Narmada, South Gujrat University, Surat, Gujarat, INDIA.
Email: pbbharate@gmail.com

Abstract

Bharate and Gupta (2013) suggested an estimator of the pooled variance and its variance is obtained when the parameter of interest is pooled variance. They have obtained minimal equations to obtain the OBP under the proportional allocation. In the present paper we have obtained the expression for variance of estimator under the equal allocation and fixed allocation. Minimal equations to get optimum boundary points(OBP) are derived by minimizing the variance of the estimator. The stratification is done on the study variable. Further, by assuming rectangular distribution within each stratum, optimum boundary points are obtained.

Key words: optimum stratification, study variable, fixed allocation, equal allocation, minimal equations, and optimum boundary points.

***Address for Correspondence:**
Dr. P B Bharate, Associate Professor and Head, Department of Statistics, Pratap College Amalner Dist. Jalgaon, Maharashtra, INDIA.
Email: pbbharate@gmail.com

Received Date: 30/08/2014 Accepted Date: 16/09/2014

Access this article online	
Quick Response Code:	Website: www.statperson.com
	DOI: 30 September 2014

INTRODUCTION

One of the basic design operations of stratified random sampling, construction of strata is very important with respect to the efficiency of design. If the knowledge of number of strata and the type of allocation is assumed the variance of the estimator of population mean (total) remains only a function of W_h and σ_h, the weight and the variance of h^{th} stratum; h=1,2,…..L. Whereas variance of estimator of pooled variance remains a function of of W_h , σ_h and μ_{4hy} .For a given population, these sets of parameters depend upon the choice of boundary points. The change in stratification changes the values of boundary points, which affects the variance of the estimator. Therefore, problem of optimum stratification is to choose one set of boundary points from all such sets, which give the minimum variance. A faulty choice of stratification may lead to considerable increase in the variance of the estimator. An appropriate choice of stratum boundary points is, therefore, of utmost importance. Dalenius (1950) was the first to discuss the problem of optimum stratification. He assumed the knowledge of frequency distribution of study variable and stratified on same variable for proportional and Neyman allocation. Dalenius and Gurney (1951) suggested the use of auxiliary variable for determination of optimum stratification boundary points (OSBP). Mahalanobis (1952) suggested a rule known as “Equal Aggregate Method.

How to site this article: P B Bharate, P C Gupta, J C Agrawal. Minimal equations to find OBP in the estimation of pooled variance under equal allocation and fixed allocation. *International Journal of Statistika and Mathemtika* Aug-Oct 2014; 11(3): 177-182. http://www.statperson.com (accessed 01 October 2014)

Ayoma(1954) derived equidistance stratification by applying the mean value theorem to the equations obtained by Dalenius (1950). Kitagawa (1956) gave name to Mahalonobis's (1952) suggestion as "Principle of equipartition". Dalenius and Hodges J.L. (1959) suggested "cumulative square root method" as an approximation to OSBP under optimum allocation. In this method, the cum $\sqrt{f}$ is divided in as many equal parts as the number of strata to be formed. i. e. cum $\sqrt{f}$ = constant. Ekman (1959) has given a method of finding AOSB for proportional allocation when the stratification is done on the study variable. Cochran (1961) observed that the cumulative square root rule used to obtain AOSB works extremely well with both theoretical as well as actual distributions. Sethi (1963) studied the problem from different angle and gave ready-made tables giving strata boundaries for some standard distributions. Des Raj (1964) obtained the equations giving the OBP for equal allocation by minimizing the variance of the stratified mean per unit estimator under equal distribution of sample to each stratum. Gupta (1970) also studied optimum stratification in case of ratio and product method of estimation, which minimizes the generalized variance of the estimates of mean of more than one character based on auxiliary character X under the proportional allocation. Singh (1971) gave the method of finding approximate OBP. Rajyaguru (1999) and Rajyaguru and Gupta (2002) have suggested an alternative aspect of optimum stratification. Instead of minimizing the variance of the stratification estimator, they minimized the weighted square co-efficient of variation under proportional allocation both when stratification variable is study variable and when it is different**.** Singh and Sukhatme (1969) proposed cube root of the probability function rule for determining strata boundaries. Lavallée and Hidiroglou (1988) derived an iterative procedure for stratifying skewed populations into a take-all stratum and a number of take-some strata such that the sample size is minimized for a given level of reliability. Other recent contributions include Hedlin (2000) who revisited Ekman's rule, Dorfman and Valliant (2000) who compared model based stratified sampling with balanced sampling, and Rivest (2002) who constructed a generalisation of the Lavallée and Hidiroglou algorithm by providing models accounting for the discrepancy between the stratification variable and the survey variable. All the workers have considered the parameter of interest as mean (total).Bharate and Gupta (2013) studied the problem of optimum stratification when parameter of interest is pooled variance. They have obtained minimal equations to obtain the OBP under the proportional allocation.Also an approximation is derived.In this paper, we have obtained an estimator of pooled variance and minimal equations for OSB are obtained by minimizing the variance of the estimator for equal and fixed allocation. This aspect of stratification is studied when stratification variable is study variable.

ESTIMATOR OF POOLED VARIANCE σ^2

Pooled sample variance s^2 as an unbiased estimator of pooled variance σ^2.
where

$$s^2 = \sum_{h=1}^{L} W_h s_h^2$$ = pooled sample variance and

$$\sigma^2 = \sum_{h=1}^{L} W_h \sigma_h^2$$ = Pooled population Variance. (1)

Variance Of The Estimator (s^2**)** is given by

$$Var\left(s^2\right) = \sum_{h=1}^{L} \frac{W_h^2 \mu_{4h}}{n_h} - \sum_{h=1}^{L} \frac{W_h^2 \sigma_h^4}{n_h} \tag{2}$$

MATHEMATICAL FORMULATION OF THE PROBLEM

We shall make following assumptions.
The variable Y has a continuous probability density function *f(y)* and the first four moments of Y exist
Population is infinite.
Though, these assumptions will not, in general, be satisfied, yet in practice they will be approximately satisfied.
Let $-\infty, y_1, y_2, \ldots\ldots y_{h-1}, y_h, y_{h+1}, \ldots\ldots y_{L-1}, \infty$ denote the boundaries of L strata, where L is fixed in advance.
In h^{th} stratum, we define

$W_h = \int_{y_{h-1}}^{y_h} f(y)\,dy$ =proportion of population units in the h[th] stratum (3)

$\mu_{hy} = \frac{1}{W_h}\int_{y_{h-1}}^{y_h} yf(y)\,dy$ =mean of the character y

for h[th] stratum (4)

$\sigma_{hy}^2 = \frac{1}{W_h}\int_{y_{h-1}}^{y_h} (y-\mu_{hy})^2 f(\mathrm{y})\,\mathrm{dy}$ = variance of the

character y for the h[th] stratum (5)

$\mu_{4hy} = \frac{1}{W_h}\int_{y_{h-1}}^{y_h} (y-\mu_{hy})^4 f(y)\,dy$

=Fourth central moment of character y (6)

Our problem is to find strata boundary points, $y_1, y_2, \ldots\ldots y_{h-1}, y_h, y_{h+1}, \ldots\ldots y_{L-1}$ such that Var(s^2) is minimum. We note that as y_h changes then W_h, W_{h+1}, μ_h, μ_{h+1}, σ_h, σ_{h+1}, μ_h, μ_{h+1} change. Also n_h, n_{h+1} may change**.**

Further, from (5) we have

$$\frac{\partial}{\partial y_h}\left(W_h \sigma^2_{hy}\right) = \left(y_h - \mu_{hy}\right)^2 f\left(y_h\right) \tag{7}$$

STRATIFICATION UNDER EQUAL ALLOCATION

For equal allocation $n_h = \frac{n}{L}$. Therefore, for the equal allocation $Var(s^2)$ becomes

$$Var\left(s^2\right) = \frac{L}{n}\left[\sum_{h=1}^{L} W_h^2 \mu_{4hy} - \sum_{h=1}^{L} W_h^2 \sigma_{hy}^4\right] \text{..........(8)}$$

Differentiating (8) partially with respect to y_h and equating it to zero, we get

$$\frac{\partial}{\partial y_h}\left(W_h^2 \mu_{4hy}\right) + \frac{\partial}{\partial y_h}\left(W_{h+1}^2 \mu_{4h+1y}\right) = \frac{\partial}{\partial y_h}\left(W_h^2 \sigma_{hy}^4\right) + \frac{\partial}{\partial y_h}\left(W_{h+1}^2 \sigma_{h+1y}^4\right) \text{......................(9)}$$ Now

$$\frac{\partial}{\partial y_h}\left(W_h^2 \mu_{4hy}\right) = 2W_h \mu_{4hy}\frac{\partial W_h}{\partial y_h} + \frac{W_h^2 \partial \mu_{4hy}}{\partial y_h} \quad \text{.........(10)}$$ But, we have

$$\frac{\partial}{\partial y_h}\left(W_h \mu_{4hy}\right) = \left(y_h - \mu_{hy}\right)^4 f\left(y_h\right)$$

i.e $\mu_{4hy}\frac{\partial W_h}{\partial y_h} + W_h \frac{\partial \mu_{4hy}}{\partial y_h} = \left(y_h - \mu_{hy}\right)^4 f\left(y_h\right)$ $\therefore W_h \frac{\partial \mu_{4hy}}{\partial y_h} = \left[\left(y_h - \mu_{hy}\right)^4 - \mu_{4hy}\right] f\left(y_h\right)$

Therefore, we get $W_h^2 \frac{\partial \mu_{4hy}}{\partial y_h} = W_h\left[\left(y_h - \mu_{hy}\right)^4 - \mu_{4hy}\right] f\left(y_h\right)$.....(11)

Putting value from equation (11) into (10), we get.

$$\frac{\partial}{\partial y_h}\left(W_h^2 \mu_{4hy}\right) = W_h\left[\mu_{4hy} + (y_h - \mu_{hy})^4\right] f(y_h) \quad ....(12)$$

Similarly, $\frac{\partial}{\partial y_h}\left(W_{h+1}^2 \mu_{4h+1y}\right) = -W_{h+1}\left[\mu_{4h+1y} + (y_h - \mu_{h+1y})^4\right] f(y_h)$ .....(13)

Again $\frac{\partial}{\partial y_h}\left(W_h^2 \sigma_{hy}^4\right) = 2W_h\sigma_{hy}^4 \frac{\partial W_h}{\partial y_h} + 4W_h^2\sigma_{hy}^3 \frac{\partial \sigma_{hy}}{\partial y_h}$ ....(14)

But we have $\frac{\partial \sigma_{hy}}{\partial y_h} = \frac{\left[\left(y_h - \mu_{hy}\right)^2 - \sigma_{hy}^2\right]}{2W_h\sigma_{hy}} f\left(y_h\right)$ ………..(15)

Putting the value of $\frac{\partial \sigma_{hy}}{\partial y_h}$ from equation (15) in (14), we get $\frac{\partial}{\partial y_h}\left(W_h^2 \sigma_{hy}^4\right) = 2W_h\sigma_{hy}^2\left(y_h - \mu_{hy}\right)^2 f(y_h)$ ....(16)

Similarly $\frac{\partial}{\partial y_h}\left(W_{h+1}^2 \sigma_{h+1y}^4\right) = -2W_{h+1}\sigma_{h+1y}^2\left(y_h - \mu_{h+1y}\right)^2 f(y_h)$ ....(17) Putting values from (12), (13), (16) and (17) in equation (9),we get

$$W_h\left[\mu_{4hy} + (y_h - \mu_{hy})^4\right] - W_{h+1}\left[\mu_{4h+1y} + (y_h - \mu_{h+1y})^4\right] = = 2W_h\sigma_{hy}^2\left(y_h - \mu_{hy}\right)^2 - 2W_{h+1}\sigma_{h+1y}^2\left(y_h - \mu_{h+1y}\right)^2$$

$$\Rightarrow W_h\left\{(\mu_{4hy} - \sigma_{hy}^4) + \left[(y_h - \mu_{hy})^2 - \sigma_{hy}^2\right]^2\right\} =$$

$$= W_{h+1}\left\{(\mu_{4h+1y} - \sigma_{h+1y}^4) + \left[(y_h - \mu_{h+1y})^2 - \sigma_{h+1y}^2\right]^2\right\}$$

$$h = 1, 2, ...L-1 \quad ....(18)$$

Equation (18) gives minimal equations for obtaining optimum boundary points. If assumption of rectangular distribution in each stratum is made, then

$$\mu_{hy} = \frac{y_h + y_{h-1}}{2}, \sigma_{hy}^2 = \frac{(y_h - y_{h-1})^2}{12} \text{ and}$$

$$\mu_{4hy} = \frac{1}{5}\frac{(y_h - y_{h-1})^4}{2^4}$$

By putting the values in the L.H.S. of equation (18), the L.H.S. of equation (18) becomes,

$$\frac{W_h}{30}\left[(y_h - y_{h-1})^4\right]$$

Similarly, R.H.S. of equation (18) becomes

$$\frac{W_{h+1}}{30}\left[(y_{h+1} - y_h)^4\right]$$

Therefore, equation (18) leads to

$$W_h(y_h - y_{h-1})^4 = W_{h+1}(y_{h+1} - y_h)^4$$

$$\Rightarrow y_h = \frac{W_h^{1/4} y_{h-1} + W_{h+1}^{1/4} y_{h+1}}{W_h^{1/4} + W_{h+1}^{1/4}} \quad ....(19)$$

If $W_h^{1/4} = W_{h+1}^{1/4}$ for all h, equation (19) reduces to same equation as obtained for estimation of mean of the character.

STRATIFICATION UNDER FIXED ALLOCATION

For fixed allocation n_h are decided in advance. Therefore, for the fixed allocation $Var(s^2)$ becomes

$$Var\left(s^2\right) = \left[\sum_{h=1}^{L} \frac{W_h^2 \mu_{4hy}}{n_h} - \sum_{h=1}^{L} \frac{W_h^2 \sigma_{hy}^4}{n_h}\right] ....(20)$$

Differentiating (20) partially with respect to y_h and equating it to zero, we get

$$\frac{\partial}{\partial y_h}\left(\frac{W_h^2 \mu_{4hy}}{n_h}\right) + \frac{\partial}{\partial y_h}\left(\frac{W_{h+1}^2 \mu_{4h+1y}}{n_{h+1}}\right) =$$

$$\frac{\partial}{\partial y_h}\left(\frac{W_h^2 \sigma_{hy}^4}{n_h}\right) + \frac{\partial}{\partial y_h}\left(\frac{W_{h+1}^2 \sigma_{h+1y}^4}{n_{h+1}}\right)$$

$$.............(21)$$

But

$$\frac{\partial}{\partial y_h}\left(\frac{W_h^2 \mu_{4hy}}{n_h}\right) = \frac{n_h \frac{\partial}{\partial y_h}\left(W_h^2 \mu_{4hy}\right) - W_h^2 \mu_{4hy} \frac{\partial n_h}{\partial y_h}}{n_h^2}$$

$$= \frac{1}{n_h}\left[\frac{\partial}{\partial y_h}\left(W_h^2 \mu_{4hy}\right)\right] \quad \text{since} \quad \frac{\partial n_h}{\partial y_h} = 0$$

Therefore, using (14), we get

$$\frac{\partial}{\partial y_h}\left(\frac{W_h^2 \mu_{4hy}}{n_h}\right) =$$

$$= \frac{W_h}{n_h}\left[\mu_{4hy} + (y_h - \mu_{hy})^4\right] f(y_h) \quad ....(22)$$

Similarly, we have

$$\frac{\partial}{\partial y_h}\left(\frac{W_{h+1}^2 \mu_{4h+1y}}{n_{h+1}}\right) =$$

$$-\frac{W_{h+1}}{n_{h+1}}\left[\mu_{4h+1y} + (y_h - \mu_{h+1y})^4\right] f(y_h) ....(23)$$

From (16), we get

$$\frac{\partial}{\partial y_h}\left(\frac{W_h^2 \sigma_{hy}^4}{n_h}\right) =$$

$$\frac{2W_h \sigma_{hy}^2}{n_h}\left(y_h - \mu_{hy}\right)^2 f(y_h).. \quad ....(24)$$

Similarly, from (17), we have

$$\frac{\partial}{\partial y_h}\left(\frac{W_{h+1}^2 \sigma_{h+1y}^4}{n_{h+1}}\right) =$$

$$= -\frac{2W_{h+1}\sigma_{h+1y}^2}{n_{h+1}} (y_h - \mu_{h+1y})^2 f(y_h) ....(25)$$

Putting values from (22), (23), (24) and (25) in equation (21),we get

$$\frac{W_h\left\{(\mu_{4hy}-\sigma_{hy}^4)+\left[(y_h-\mu_{hy})^2-\sigma_{hy}^2\right]^2\right\}}{n_h}=$$

$$=\frac{W_{h+1}\left\{(\mu_{4h+1y}-\sigma_{h+1y}^4)+\left[(y_h-\mu_{h+1y})^2-\sigma_{h+1y}^2\right]^2\right\}}{n_{h+1}}$$

$$h=1,2,........L-1...........(26)$$

Equations (26) give minimal equations for obtaining optimum boundary points.

If we assume rectangular distribution in each stratum, then we get optimal equations as below,

$$\frac{W_h(y_h-y_{h-1})^4}{n_h}=\frac{W_{h+1}(y_{h+1}-y_h)^4}{n_{h+1}}$$

$$\Rightarrow y_h=\frac{\left(\frac{W_h}{n_h}\right)^{1/4}y_{h-1}+\left(\frac{W_{h+1}}{n_{h+1}}\right)^{1/4}y_{h+1}}{\left(\frac{W_h}{n_h}\right)^{1/4}+\left(\frac{W_{h+1}}{n_{h+1}}\right)^{1/4}}\ ....(27)$$

Equations (27) give OBP in case of fixed allocation assuming rectangular distribution in each stratum.

CONCLUSION

The Minimal equations obtained here are implicit in nature. To solve these equations either rigorous computer programming is required or some approximate equations can be obtained.

REFERENCES

1. AYOMA H. (1954) : “A study of stratified random sampling” ; *Ann Inst. Stat. Math.*, 6, 1-136.
2. BHARATE AND GUPTA(2013): “Optimum stratification in the estimation of pooled variance under proportional allocation”; *Intr.Jounal of Statistika and Mathematika*, Volume 7, Issue 2, 24-27
3. DALENIUS T. (1950) : “The problem of optimum stratification I”, *Skand. Akt.* ,33, 203-213.
4. DALENIUS T. and GURNEY. M. (1951) : “The Problem of optimum stratification II” *Skand. Akt*, 34, 139 -148.
5. DALENIUS T. and HODGES. J. L. (1957) : “ The choice of stratification points” *Skand. Akt*., 40, 198-203.
6. DALENIUS T. and HODGES. J.L. (1959) : “Minimum variance stratification” *JASA,* 54, 88-101.
7. DES RAJ (1964) : ‘”On formulation of strata of equal aggregate size”,*JASA*, 59, 481-486
8. DORFMAN A.H., VALLIANT R.(2000):"Stratification by size revised", *Journal of Official Statistics*, 16, 2, 139–154.
9. EKMAN G. (1959): “An approximation useful in univariate stratification”; *Ann. Math. Stat*., 30, 219-229.
10. GUPTA P. C. (1970): “On some estimation problems in sampling using auxiliary in formation”, *Ph.D. thesis. IASRI* , New Delhi.
11. HEDLIN (2000). “A Procedure for Stratification by an Extended Ekman Rule”. *Journal of Official Statistics,* 16, 15-29.
12. KITAGAWA.T. (1956) “Some contribution to the design of sample surveys”. *Sankhya*, 17, 1-36.
13. LAVELLEE, P. and HIDIROGLOU, M (1988),”On the Stratification of Skewed Populations”,*Survey Methodology*, 14, 33-43.
14. MAHALANOBIS P.C. (1952) : “Some aspect of the design of sample surveys” *Sankhya,* 12, 1-17
15. RAJYAGURU ARTI (1999)”Some aspects of theory of optimum stratification using coefficient of variation as norm and estimation of coefficient of variation from finite population”; *Ph.D. thesis ,Surat India*
16. RAJYAGURU and GUPTA (2002)”On alternative aspect of optimum stratification”, *Guj. Stat. Review*, 29 ,101-112
17. RIVEST, L.P. (2002). “A Generalization of the Lavellee-Hidiroglou Algorithm for Stratification In Business Surveys”, *Survey Methodology*, 28, 191-198

Source of Support: None Declared
Conflict of Interest: None Declared