Original Research Article

Stratification in ratio estimators: a practical approach

Abstract: Whenever the population is stratified, there are two methods of obtaining ratio estimator i.e. separate and combined ratio estimators. Stratifying the population geographically may yield inefficient estimates as within strata homogeneity in such case cannot be maintained. The objective of this paper was to illustrate the best stratification rule under two methods of allocation in case of separate and combines ratio estimators. So this paper presents the estimates of mean and variance under proportional and Neyman allocation of mango production of 325 orchardists of Himachal Pradesh, collected through well designed survey and then stratified into L (number of strata) = 5 and 6 demarcated using four stratification rules i) Equalization of Strata Total ii) Equalization of cumulative $\sqrt{f(y)}$ iii) Equalization of cumulative $\sqrt[3]{f(y)}$ and iv) Equalization of cumulative $\frac{1}{2}[r(y) + f(y)]$ for varying number of sample sizes (n = 60, 90, 120).
Key words: separate ratio estimators, combined ratio estimators, stratification, Mango production

Introduction: There are many practical situations when the variable under study is from the population that is grouped in strata. Then to obtain the ratio estimate from such population, there are two types of estimators that can be used viz. separate and combined ratio estimators. Usually the population is stratified on the basis of geographical locations but this may cause the within strata variances to be severe which in turn will not yield efficient estimates. So the four methods i) Equalization of Strata Total ii) Equalization of cumulative $\sqrt{f(y)}$ iii) Equalization of cumulative $\sqrt[3]{f(y)}$ and iv) Equalization of cumulative $\frac{1}{2}[r(y) + f(y)]$ were used for stratification rather grouping on geographical basis and then the estimates of mean and variance for L (number of strata) = 5 and 6 and for varying sample sizes of n = 60, 90 and 120 under proportional and Neyman allocation, respectively were estimated.

Material methods:

Selection of sample: The primary data of 325 mango orchardists from five major mango growing districts of Himachal Pradesh, India were collected through well designed survey. The sample was selected with the help of multi-stage sampling in which 30 % of the blocks were selected randomly in first stage and from chosen blocks the orchardists were selected randomly. The primary data were collected on mango production and area (auxiliary variable) through survey of these selected orchardists. Mango production y being the study variable was then estimated by first stratifying the area under mango x (auxiliary variable) by using four stratification rules

Stratification rules: The four stratification rules that were used to stratify were *i) Equalization of strata total*: Mahalanobis (1952) proposed the equalization of strata total $(N_h\mu_h)$ with equal allocation.

ii) Equalization of cumulative $\sqrt{f(y)}$: Dalenius and Hodges (1957) proposed formation of strata by equalizing the cumulative $\sqrt{f(y)}$, where f(y) is the frequency function.

iii) Equalization of cumulative $\frac{1}{2}\{r(y) + f(y)\}$: Durbin (1959) proposed the equalization of the cumulative frequencies of a distribution, g(y), which is in between the original distribution f(y) and a rectangular distribution r(y) over the range (y_o, y_L) of y.

iv) Equalization of cumulative $\sqrt[3]{f(y)}$: Singh and Sukhatme (1969) suggested another method of construction of strata, which is called equal intervals on cumulative $\sqrt[3]{f(y)}$, where f(y) is the frequency function of the character under study.

Sample allocation: The sample size was allocated by proportional and Neyman allocation

i) Proportional allocation: In this method, allocation of a given sample size 'n' to different strata is done in proportion to stratum weight i.e. in the h^{th} stratum $n_h = nW_h$ where, $W_h = \frac{N_h}{N}$.

ii) Neyman allocation: Most of the times, a survey statistician has to work within a fixed budget and therefore, the sampling variance has to be minimized for a given cost. In this case, the sample size in the h^{th} stratum is given by $n_h = n\frac{W_hS_h}{\sum_{h=1}^{L} W_hS_h}$.

Estimate of mean and variance: The estimate of average for **separate ratio estimator** is given by $\hat{\bar{y}}_{RS} = \sum_{h=1}^{L} W_h\hat{R}_h\bar{X}_h$, where $\hat{R}_h = \bar{y}_h/\bar{x}_h$ and $\bar{x}_h \neq 0$ and the variance for the separate ratio estimator is given by

$$v(\hat{\bar{y}}_{RS}) = \sum_h W_h^2 \frac{(1-f_h)}{n_h}\left(s_{yh}^2 + R_h^2 s_{xh}^2 - 2R_h\rho_h s_{yh}s_{xh}\right)$$

Often in practical situation we may encounter small n_m's and to overcome this difficulty Hansen, Hurwitz and Gurney (1946) suggested a combined ratio estimator. The estimate of the average for **combined ratio estimator** is given by $\hat{\bar{y}}_{RC} = \sum_{h=1}^{L} \hat{R}_h\bar{X}_h$, where $\hat{R} = \bar{y}_{st}/\bar{x}_{st}$ where $\bar{y}_{st} = \sum_h \frac{N_h}{N}\bar{y}_h$ and $\bar{x}_{st} = \sum_h \frac{N_h}{N}\bar{x}_h$ and the variance for the combined ratio estimator is given by

$$v(\hat{\bar{y}}_{RC}) = \sum_h W_h^2 \frac{(1-f_h)}{n_h}\left(s_{yh}^2 + R^2 s_{xh}^2 - 2R\rho_h s_{yh}s_{xh}\right)$$

where s_{yh}^2 and s_{xh}^2 are the sample variance of the h^{th} stratum.

Results and Discussion: The estimates of mean and variance of mango production under proportional and Neyman allocation are presented in Table 1 and 2. For illustration the area under mango was used as auxiliary information and subject to stratification by these four stratification rules.

The results from these tables revealed minimum estimate of the variance was found to be 0.039 by using combined ratio estimate under Neyman allocation by using equalization of cumulative $\sqrt[3]{f(y)}$ for L=6 and n=120. Usually combined ratio estimates are less efficient than separate ratio estimator. But Table 1 reveals combined ratio estimate is superior to separate ratio estimate in few cases mostly for L=6. This is because as number of strata increases, the allocation of sample for each stratum decreases. Separate ratio estimate is always superior when the sample collected from each stratum is large.

However as results revealed in Table 1, minimum estimate of variance of estimated mean for mango production under Neyman allocation by using separate ratio estimator is found to be 0.40 by using equalization of cumulative $\sqrt[3]{f(y)}$ for L = 6 and n=120.

Moreover, Table 1 and 2 reveals that there is not much difference in the estimates between both the methods. It was seen that as the number of strata and sample size was increased the estimate of the variance decreased uniformly. Under Neyman allocation the relative precision of estimate of variance of estimated mean by using combined ratio estimate to that of separate ratio estimate was found to be 102.56%.

On similar notes estimate of the variance of estimated mean was calculated when the population was stratified using proportional allocation with above mentioned stratification methods for varying number of strata (L=5 & 6) and for n = 60, 90 and 120, respectively. The estimates of variances are shown in Table 2.

Minimum estimate of variance was found to be 0.048 by using combined ratio estimator using equalization of cumulative $\sqrt[3]{f(y)}$ for L=6 and n =120. It was found to be smaller than separate ratio estimate because of small sample sizes allocated due to large number of strata. Under proportional allocation the relative precision of estimate of variance of estimated mean of mango production by using combined ratio estimator with respect to separate ratio estimator was found to be 102.08%.

So estimate of average value can be obtained with the help of these estimators. Under Neyman and proportional allocation using equalization of cumulative $\sqrt[3]{f(y)}$ for L=6 and n=120, the estimate of mean of mango production of orchardists of the state by combined ratio estimator is found to be 8.36 and 8.03 tons, respectively.

Also Table 3 represents the gain in precision of separate and combined ration estimators over simple random sample without replacement of the respective sample sizes.

Maximum gain in efficiency in case when population was stratified by using Neyman allocation when compared to SRSWOR was observed to be 637.14% when population was stratified using equalization of cumulative $\sqrt[3]{f(y)}$ rule for L=6 and n = 120.

This reveals that for large sample and large number of strata, the estimate of variance of mean with separate and combined ratio estimate gains large efficiency and this method can be used to estimate mango area and production in the state.

References:

Dalenius T and Hodges JL. 1957. The Choice of Stratification Points. *Skandinavisk Aktuarietidskrift* 40:198–203.

Durbin J. 1959. Review of sampling in Swede. *Journal of Royal Statistical Society* 122:246-248.

Hansen MH, Hurwitz WN and Gurney M. 1946. Problems and methods of the sample survey of busisiness. *Journal of American Statistical Association* 41(234):173-189.

Mahalanobis PC. 1952. Some aspects of the design of sample survey. *Sankhya* 12:1-7.

Singh R and Sukhatme PV. 1969. Optimum stratification. *Annals of the Institute of Statistical Mathematics* 21:515-528.

Singh R. 1971. Approximately optimum stratification on auxiliary variable. *Journal of American Statistical Association* 66: 829-833.

Singh R and Prakash D. 1975. Optimum stratification for equal allocation. *Annals of the Institute of Statistical Mathematics* 27:273 -280.

Upadhyaya LN, Singh HP and Singh S. 2004. A Class of Estimators for Estimating the Variance of the Ratio Estimator. *Journal of the Japan Statistical Society* 34(1):47-63.

Vishwakarma GK and Singh HP. 2011. Separate ratio-product estimator for estimating population mean using auxiliary information. *Journal of Statistical Theory and Applications* 10(4):653-664.

Table 1 Estimates of $v(\hat{\bar{y}})$ of Mango Production using separate and combined ratio estimators when stratified through Neyman allocation (production in tons)

Separate Ratio Estimates						Combined Ratio Estimates					
Equalization of Strata Total			Equalization of cumulative $\sqrt{f(y)}$			Equalization of Strata Total			Equalization of cumulative $\sqrt{f(y)}$		
sample	Strata		sample	Strata		sample	Strata		sample	Strata	
	5	6		5	6		5	6		5	6
60	0.117	0.104	**60**	0.175	0.105	**60**	0.113	0.103	**60**	0.175	0.108
90	0.084	0.067	**90**	0.081	0.048	**90**	0.084	0.067	**90**	0.080	0.048
120	0.069	0.060	**120**	0.063	0.042	**120**	0.069	0.060	**120**	0.063	0.042
Equalization of cumulative $\sqrt[3]{f(y)}$			Equalization of cumulative $\frac{1}{2}[r(y) + f(y)]$			Equalization of cumulative $\sqrt[3]{f(y)}$			Equalization of cumulative $\frac{1}{2}[r(y) + f(y)]$		
sample	Strata		sample	Strata		sample	Strata		sample	Strata	
	5	6		5	6		5	6		5	6
60	0.130	0.101	**60**	0.161	0.108	**60**	0.128	0.102	**60**	0.161	0.108
90	0.093	0.051	**90**	0.115	0.073	**90**	0.093	0.050	**90**	0.115	0.073
120	0.069	0.040	**120**	0.055	0.056	**120**	0.069	0.039	**120**	0.056	0.056

Table 2 Estimates of $v(\hat{\bar{y}})$ mean of Mango Production using separate and combined ratio estimators when stratified through proportional allocation (production in tons)

Separate Ratio Estimates						Combined Ratio Estimates					
Equalization of Strata Total			Equalization of cumulative $\sqrt{f(y)}$			Equalization of Strata Total			Equalization of cumulative $\sqrt{f(y)}$		
sample	Strata		sample	Strata		sample	Strata		sample	Strata	
	5	6		5	6		5	6		5	6
60	0.141	0.093	**60**	0.144	0.111	**60**	0.141	0.093	**60**	0.144	0.111
90	0.095	0.069	**90**	0.099	0.081	**90**	0.095	0.066	**90**	0.099	0.075
120	0.090	0.058	**120**	0.079	0.052	**120**	0.089	0.060	**120**	0.079	0.063
Equalization of cumulative $\sqrt[3]{f(y)}$			Equalization of cumulative $\frac{1}{2}[r(y) + f(y)]$			Equalization of cumulative $\sqrt[3]{f(y)}$			Equalization of cumulative $\frac{1}{2}[r(y) + f(y)]$		
sample	Strata		sample	Strata		sample	Strata		sample	Strata	
	5	6		5	6		5	6		5	6
60	0.098	0.087	**60**	0.085	0.093	**60**	0.098	0.088	**60**	0.085	0.094
90	0.084	0.065	**90**	0.071	0.076	**90**	0.084	0.065	**90**	0.071	0.075
120	0.069	0.049	**120**	0.061	0.061	**120**	0.069	0.048	**120**	0.062	0.062

Table 3 Gain in efficiency of separate (a) and combined (b) ratio estimates with respect to simple random sample without replacement (Neyman allocation)

Equalization of Strata Total				
sample	Strata			
	5		6	
	a	b	a	b
60	144.87	153.28	172.24	178.35
90	166.77	166.03	235.15	235.12
120	198.33	197.20	242.44	241.67

Equalization of cumulative $\sqrt{f(y)}$				
sample	Strata			
	5		6	
	a	b	a	b
60	63.20	63.31	172.24	165.69
90	177.04	179.66	367.70	369.35
120	227.75	225.67	428.10	432.44

Equalization of cumulative $\sqrt[3]{f(y)}$				
sample	Strata			
	5		6	
	a	b	a	b
60	119.89	123.59	182.97	178.94
90	140.68	139.81	458.56	468.82
120	199.55	199.82	620.12	637.14

Equalization of cumulative $\frac{1}{2}[r(y) + f(y)]$				
sample	Strata			
	5		6	
	a	b	a	b
60	77.11	77.40	164.18	163.76
90	94.30	93.87	206.68	206.59
120	271.03	269.77	264.92	265.07