Original Research Article

Estimation of Guava Production in Himachal Pradesh by using Optimum Stratification

Abstract : The present study focuses on standardization of sampling technique and comparison of sample allocation methods. The goal of stratification is to provide a better cross-section of the population in order to increase relative accuracy. For this purpose,Primary data on area and production of guava were obtained from 275 respondents of Himachal Pradesh through a well-designed and pre-tested survey approach. The optimum stratification points were found by using the auxiliary variable "area under guava" as the stratification variable. Four methods, namely, Equalization of strata totals, Equalization of cumulative $\sqrt{f(y)}$, Equalization of cumulative 1/2 [r(y)+f(y)] and Equalization of cumulative$\sqrt[3]{f(y)}$, were used for the construction of approximate optimum strata boundaries for varying numbers of strata (L= 2,3,4,5) and sample sizesn_i= 60, 90, 120. The sample was allocated to different strata according to proportional and Neyman allocation methods.The minimum estimate of the variance of $\hat{\bar{y}}$ of guava production and maximum gain in efficiency was found to be 0.008 and 418.11 percent respectivelyin Equalization of cumulative $\sqrt[3]{f(y)}$ rule for n =120 and L=5 under Neymann allocation.

Keywords : Guava, Stratification, Stratified random sampling, Optimum Strata Boundaries, Neyman allocation, proportional allocation.

Introduction

The primary goal of stratification in sample survey design is to lower the sample variation of the estimates and to provide a better cross-section of the population in order to increase relative accuracy. The best characteristic to find these optimum strata boundaries is with the study variable itself. The next best presumably is the frequency distribution of some other variable highly correlated to the study variable. In the present study “area under guava plantation” was used as the auxiliary variable which is highly correlated with the Estimation variable “Production of Guava”. It has been seen that it is always profitable in terms of precision that the variance of the estimate decreases as there is increase in number of strata. The stratified random sampling yields unbiased estimate of the population mean and its standard error provide confidence interval in which the possible value of the population mean lies. The primary data on 275 guava orchardists were collected from five districts of Himachal Pradesh viz. Bilaspur, Hamirpur, Kangra, Una and Sirmour. Data were collected through well planned survey from these locations randomly. Data were collected through well designed questionnaire on socio-economic status, area and production of Guava in the mentioned districts of Himachal Pradesh. Guava production as the study variable, number of trees and area under Guava cultivation as the auxiliary information were used in the estimation of guava production and area under guava plantation. The auxiliary variable considered in the problem is a size variable that holds a common model for a whole population (Khan et al. 2009). The pioneering work was done by Dalenius (1950) for optimum stratification regarding stratified random sampling estimates. He considered the problem for study variable itself as the stratification variable. Dalenius and Gurney (1951) considered the problem of optimum stratification with respect to an auxiliary variable so as to minimize the variance of stratified random sampling estimate

The commonly used standard stratification methods of construction of strata (Sukhatme, et al., 1983), viz., equalization of strata total, equalization of cumulative of $\sqrt{f(y)}$, equalization of cumulative of ½ {r(y) + f(y)} and equalization of cumulative $\sqrt[3]{f(y)}$ have been tried to find out the optimum points of stratification for varying number of strata 2 to 5. The relative efficiencies of different methods of strata were examined when the number of strata was 2, 3, 4 and 5 under Proportional and Neyman sample allocation methods. Further, relative efficiencies for different methods of estimation were also examined to estimate the total production of guava in Himachal Pradesh.

Material and method

Multistage random sampling technique was employed for the selection of units.The primary data of 275 guava orchardists were collected from purposively selected five major guava growing districts of Himachal Pradesh, through well designed survey.In the first stage, more than 30% blocks of each districts were selected randomly. In second stage more than 30 % of farmers from each selected block were taken. In present study, total production was considered as study variable and area under guava was considered as auxiliary variable, as it was highly correlated with the study variable. The procedures of constructing approximately optimum strata boundaries (AOSB) for different allocation methods are as given below:

i) Equalization of strata total

Mahalonobis (1952) proposed the equalization of strata total $(N_h \mu_h)$ with equal allocation. He suggested that when the number of strata is predominated, say L, a practical method of stratification is to stratify the whole population into a set of L strata such that the total value of the character remains the same for each stratum. The main advantage of this rule is its simplicity. Hansen, Hurwitz and Madow (1953) demonstrated that this method lead to efficient stratification, if the strata coefficient of variation is same.

ii)Equalization of cumulative $\sqrt{\mathbf{f(y)}}$

Dalenius and Hodges (1957) proposed formation of strata by equalizing the cumulative $\sqrt{f(y)}$, where f(y) is the frequency function. In deriving the rule, it is assumed that the distribution is bounded and that the number of strata is large. Since f(y) is generally unknown, f(x) is used in place of f(y), where x is an auxiliary variable highly and positively correlated with y. If the cumulative total is H, the approximate strata boundaries (x_i) are given by $\frac{iH}{L}, i = 1, 2, \ldots, L-1$.

iii) Equalization of cumulative $\frac{1}{2}\{\mathbf{r(y) + f(y)}\}$

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. that is r(y) is taken as $r(y) = \frac{F(y_L)}{y_L - y_0}$ and the optimum points of stratification were obtained by equalizing the cumulative of stratification to cumulative of the function g(y) = $\frac{1}{2}\{r(y) + f(y)\}$.

iv) Equalization of cumulative $\sqrt[3]{\mathbf{f(y)}}$

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. In this method, the value of $\sqrt[3]{f}$ are cumulative where y is the character under study. Since f(y) is unknown, f(x) is used in place of f(y), where x is an auxiliary variable highly and positively correlated with y. If the cumulative cube root total is H, the approximately optimum strata boundaries are given by $\frac{iH}{L}, i = 1, 2, \dots, L - 1$.

Allocation of sample size: 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 hth stratum $n_h = nW_h$ where, $W_h = \frac{N_h}{N}$. Using this method of allocation, the estimator of variance of the estimate $\bar{y}_{st}$ reduces to

$$\hat{V}(\bar{y}_{st})_P = \left(\frac{1}{n_h} - \frac{1}{N_h}\right) \sum_{h=1}^{L} W_h s_h^2$$

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 hthstratum is given by $n_h = n\frac{W_h S_h}{\sum_{h=1}^{L} W_h S_h}$. Then, using this method of allocation, the estimator of the variance of the estimate $\bar{y}_{st}$ becomes:

$$\hat{V}(\bar{y}_{st})_N = \frac{1}{n}\left(\sum_{h=1}^{L} W_h s_h\right)^2 - \frac{1}{N}\sum_{h=1}^{L} W_h s_h^2$$

Results and Discussion

The information on area and production of guava was collected from the selected respondents. Table 1 gives the frequency distribution of the respondents according to the area under guava. The data revealed that distribution of holdings was highly skewed and most of the units (128) were located in the 0-0.25 class interval followed by 52 units in 0.50-0.75 class interval. For the present study the optimum points of stratification along with percentage of the orchardists falling in respective strata, as shown in Table 2, were determined by using four standard stratification methods namely equalization of strata total, equalization of cumulative of $\sqrt{f(y)}$, , equalization of 1/2 {r(y)+ f(y)} and equalization of cumulative of $\sqrt[3]{f(y)}$ and their relative efficiencies for estimating total production of guava were analyzed. Allocation of the sample to different strata was made in accordance with commonly used methods viz., Equal allocation, Proportional allocation, and Neyman allocation. Singh and Parkash (1975) considered the problem of optimum stratification on the auxiliary variable x for equal allocation.

Table1: Frequency Distribution of Area (ha) and Cumulative total of number of respondents by using different stratification method

Area (ha)	Frequency N_h	Mid values	Equalization of Strata Total		Equalization of cumulative $\sqrt{f(y)}$		Equalization of cumulative $\sqrt[3]{f(y)}$		Equalization of cumulative $\frac{1}{2}[r(y)+f(y)]$			
			$N_h\mu_h$	Cum. $N_h\mu_h$	$\sqrt{f(y)}$	Cum.$\sqrt{f(y)}$	$\sqrt[3]{f(y)}$	Cum.$\sqrt[3]{f(y)}$	r(y)	r(y)+f(y)	$\frac{1}{2}[r(y)+f(y)]$	Cum.$\frac{1}{2}[r(y)+f(y)]$
0 - 0.25	128.00	0.13	16.00	16.00	11.31	11.31	5.04	5.04	0.01	128.01	64.01	64.01
0.25 - 0.50	35.00	0.38	13.13	29.13	5.92	17.23	3.27	8.31	0.06	35.06	17.53	81.54
0.50 - 0.75	52.00	0.63	32.50	61.63	7.21	24.44	3.73	12.04	0.10	52.10	26.05	107.58
0.75 - 1.00	22.00	0.88	19.25	80.88	4.69	29.13	2.80	14.85	0.14	22.14	11.07	118.65
1.00 - 1.25	25.00	1.13	28.13	109.00	5.00	34.13	2.92	17.77	0.18	25.18	12.59	131.24
1.25 - 1.50	5.00	1.38	6.88	115.88	2.24	36.37	1.71	19.48	0.22	5.22	2.61	133.85
1.50 - 1.75	1.00	1.63	1.63	117.50	1.00	37.37	1.00	20.48	0.26	1.26	0.63	134.48
1.75 - 2.00	5.00	1.88	9.38	126.88	2.24	39.60	1.71	22.19	0.30	5.30	2.65	137.14
2.00 - 2.25	1.00	2.13	2.13	129.00	1.00	40.60	1.00	23.19	0.34	1.34	0.67	137.81
2.25 - 2.50	1.00	2.38	2.38	131.38	1.00	41.60	1.00	24.19	0.39	1.39	0.69	138.50
Total	**275.00**		**131.375**		**41.603**		**24.189**				**138.502**	

Optimum strata boundaries: Table 2 represents the demarcation points under various stratification rules along with percentage of respondents that falling in respective stratum. Under stratification by Equalization of strata totals, for L= 2, two points of demarcation point was0.75 ha. The percentage of number of orchardists that fall in 1^{st} and 2nd stratum was found to be 78 and 22 respectively. For L=3, twoAOSB were found to be 0.48 and 1.11 ha with 59, 34 and 7 percent of orchardists that fall in first, second ,and third stratum, respectively. Similarly we can check for all stratification rules. The area under guava (ha) which is correlated with the study variable guava production (tons) was subjected to stratification. The proportional and Neyman estimates of the variances of $\hat{\bar{y}}$were worked out with varying number of strata (L=2, 3, 4 and 5) under four methods of stratification and are presented in the Table 3 and Table 4. (The smaller values of variances are due to conversion of study variable in metric tons.)

Comment [P1]: Please provide the table before reference

Estimate of Variance of Guava production of under guava plantation under optimum stratification

Table 3 and 4reveals that the variance estimate is decreasing as the sample size (n) and the number of strata (L) are increasing under all stratification methods that prove their optimality. Under proportional allocation, the minimum estimated variance of $\hat{\bar{y}}$ of guava production of all the four stratification methods for varying strata an sample sizes was found to be 0.01 in Equalization of cumulative $\sqrt[3]{f(y)}$ rule for n =120 and L=5.The minimum estimate of the variance of $\hat{\bar{y}}$ of guavaproduction was found to be 0.008in Equalization of cumulative $\sqrt[3]{f(y)}$ rule for n =120 and L=5 under Neyman allocation.Similarly, Mathew et al. (2013) investigated the efficiency of Neyman allocation procedure over equal and proportional allocation procedures and found that Neyman allocation procedure was the best and most efficient for estimating the average and the variance of the prices of Peak Milk (Nigeria Made) in the markets in Abeokuta.

Comment [P2]: Please provide the table before Reference

Gain in efficiency due to stratification:

Variances owing to Proportional allocation and Neyman allocation were compared with simple random sampling variances to determine the increase in efficiency of stratification (L >1) over no stratification (L = l), and the findings are presented in Table 5 and Table 6. The Table indicated that there is a considerable gain in efficiency due to stratification, but the maximum gain in efficiency is observed when the strata are constructed through the Equalization of cum

$\sqrt[3]{f}$method. It is also observed that this gain in efficiency increases with the increase in the number of strata and sample size. For estimation of guava production by using proportional allocation, the maximum gain in efficiency was observed to be 365.60 percent using the Equalization of cumulative $\sqrt[3]{f(y)}$ rule. Under Neyman allocation, maximum gain in efficiency was observed for n=120 and L=5, which was 418.11 percent using Equalization of of cumulative $\sqrt[3]{f(y)}$.

Conclusion

It may be concluded that Equalization of cumulative $\sqrt[3]{f}$methodmay be used for greater efficiencies to estimate the production of guava in the study area.Anju Sharma *et al.* (2017) compared the different allocation procedures viz., Equal, Proportional and Optimum/Neyman in a stratified random sampling of skewed populations under different distributions and samples sizes and concluded that with the increase in number of strata from 2 to 4 and sample size from 10 to 40, equalization of cumulative of$\sqrt[3]{f(y)}$ method along with Neyman allocation resulted in least variance (0.89) and maximum percentage gain in efficiency (20418.16).

References

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

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

Khan, Ahmad MGM, and Khan S. 2009. Determining the Optimum Stratum Boundaries Using Mathematical Programming. *J Math Model Algor***8**:409.

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

Mathew OO, Sola AF, Oladiran BH and AA. 2013. Efficiency of Neyman Allocation Procedure over other Allocation Procedures in Stratified Random Sampling. American *Journal of Theoretical and Applied Statistics***2**:122-127.

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

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

Sharma A, Mahajan PK and Belwal OK. 2017. Statistical investigation through stratified random sampling for apple production in Himachal Pradesh. *Journal of Applied and Natural Science* **9**:1718–1723.

Table 2: Optimum strata boundaries and percentage of orchardists that fall in respective stratum

Strata	Equalization of Strata Total					Strata	Equalization of cumulative $\sqrt{f(y)}$				
	I	II	III	IV	V		I	II	III	IV	V
2	0.80					**2**	0.62				
Percentage	**79**	**21**				**Percentage**	**72**	**28**			
3	0.61	1.06				**3**	0.36	0.93			
Percentage	**72**	**15**	**14**			**Percentage**	**55**	**31**	**14**		
4	0.53	0.80	1.16			**4**	0.23	0.62	1.10		
Percentage	**59**	**19**	**15**	**7**		**Percentage**	**46**	**26**	**21**	**7**	
5	0.45	0.68	0.97	1.22		**5**	0.18	0.48	0.78	1.21	
Percentage	**58**	**13**	**15**	**7**	**7**	**Percentage**	**42**	**17**	**19**	**15**	**7**
Strata	**Equalization of cumulative $\sqrt[3]{f(y)}$**					**Strata**	**Equalization of cumulative $\frac{1}{2}[r(y) + f(y)]$**				
	I	**II**	**III**	**IV**	**V**		**I**	**II**	**III**	**IV**	**V**
2	0.75					**2**	0.32				
Percentage	**78**	**22**				**Percentage**	**48**	**52**			
3	0.48	1.11				**3**	0.18	0.60			
Percentage	**59**	**34**	**7**			**Percentage**	**59**	**34**	**7**		
4	0.36	0.76	1.30			**4**	0.14	0.32	0.71		
Percentage	**48**	**31**	**17**	**5**		**Percentage**	**48**	**31**	**17**	**5**	
5	0.24	0.59	0.97	1.48		**5**	0.11	0.22	0.52	0.82	
Percentage	**46**	**25**	**15**	**11**	**3**	**Percentage**	**31**	**15**	**14**	**20**	**20**

Table 3 :Estimate of variance of different stratification methods for different sample allocation methods (Proportional allocation)

Sample	**Equalization of strata totals**				**Sample**	**Equalization of cum $\frac{1}{2}[r(y)+f(y)]$**			
	2	**3**	**4**	**5**		**2**	**3**	**4**	**5**
60	0.075	0.066	0.057	0.050	**60**	0.069	0.047	0.045	0.038
90	0.045	0.037	0.029	0.024	**90**	0.044	0.036	0.032	0.030
120	0.042	0.035	0.025	0.019	**120**	0.028	0.025	0.020	0.015
Sample	**Equalization of cum $\sqrt{f}$**				**Sample**	**Equalization of cum $\sqrt[3]{f}$**			
60	0.072	0.052	0.044	0.039	**60**	0.056	0.048	0.040	0.038
90	0.035	0.03	0.026	0.025	**90**	0.029	0.024	0.022	0.020
120	0.022	0.020	0.019	0.016	**120**	0.019	0.016	0.012	0.01

Table 4: Estimate of variance of different stratification methods for different sample allocation methods (Neyman allocation)

Sample	Equalization of strata totals				Sample	Equalization of cum $\frac{1}{2}[\mathbf{r}(\mathbf{y}) + \mathbf{f}(\mathbf{y})]$			
	2	**3**	**4**	**5**		**2**	**3**	**4**	**5**
60	0.077	0.069	0.052	0.050	**60**	0.059	0.045	0.037	0.024
90	0.051	0.033	0.025	0.024	**90**	0.034	0.020	0.016	0.012
120	0.018	0.012	0.007	0.005	**120**	0.015	0.009	0.006	0.005
Equalization of cum $\sqrt{\mathbf{f}}$					**Equalization of cum $\sqrt[3]{\mathbf{f}}$**				
Sample	**Equalization of cum $\sqrt{\mathbf{f}}$**				**Sample**	**Equalization of cum $\sqrt[3]{\mathbf{f}}$**			
60	0.076	0.042	0.031	0.023	**60**	0.067	0.050	0.024	0.022
90	0.046	0.028	0.018	0.011	**90**	0.033	0.024	0.015	0.012
120	0.018	0.012	0.007	0.006	**120**	0.013	0.011	0.009	0.004

Table 5: Percentage gain in efficiency due to stratification (Proportional allocation)

Sample Sizes	**Number of strata**				**Number of strata**			
	Equalization of strata totals				**Equalization of cum $\frac{1}{2}[\mathbf{r}(\mathbf{y}) + \mathbf{f}(\mathbf{y})]$**			
	2	**3**	**4**	**5**	**2**	**3**	**4**	**5**
60	9.99	43.76	108.23	138.91	8.64	96.50	161.93	257.88
90	24.20	50.89	192.79	225.48	75.07	140.40	189.24	269.36
120	28.09	57.91	196.49	230.36	80.04	160.53	211.47	324.00
	Equalization of cum $\sqrt{\mathbf{f}}$				**Equalization of cum $\sqrt[3]{\mathbf{f}}$**			
60	14.53	57.32	86.95	292.64	9.98	30.40	63.14	135.94
90	38.14	66.85	154.68	358.53	89.40	128.49	195.42	232.83
120	61.64	154.17	190.63	333.65	166.91	241.04	233.46	365.60

Table 6: Percentage gain in efficiency due to stratification (Neyman allocation)

Sample Sizes	Number of strata				Number of strata			
	Equalization of strata totals				Equalization of cum $\frac{1}{2}[\mathbf{r}(\mathbf{y})+\mathbf{f}(\mathbf{y})]$			
	2	**3**	**4**	**5**	**2**	**3**	**4**	**5**
60	6.34	18.62	57.94	65.33	22.55	64.37	250.15	265.58
90	10.00	68.26	119.15	131.63	70.14	132.60	271.60	364.50
120	48.34	111.86	285.11	330.22	97.19	133.84	272.17	398.43
	Equalization of cum $\sqrt{\mathbf{f}}$				**Equalization of cum $\sqrt[3]{\mathbf{f}}$**			
60	8.64	96.50	161.93	257.88	38.30	81.92	121.83	243.78
90	20.53	97.86	207.26	388.11	63.94	177.08	250.54	355.04
120	46.87	112.67	260.98	390.56	75.67	193.60	333.18	418.11