



## Statistical investigation through stratified random sampling for apple production in Himachal Pradesh

Anju Sharma<sup>1\*</sup>, P. K. Mahajan<sup>1</sup> and O. K. Belwal<sup>2</sup>

<sup>1</sup>Dr Y. S. Parmar University of Horticulture and Forestry, Nauni, Solan-173230 (H.P.), INDIA

<sup>2</sup>H. N. B. Garhwal University, Srinagar-Garhwal-246174 (Uttarakhand), INDIA

\*Corresponding Author. Email: anjusharma\_uhf@yahoo.com

Received: August 14, 2016; Revised received: February 7, 2017; Accepted: August 14, 2017

**Abstract:** The present study focuses on standardization of sampling technique and comparison of different types of sample allocation methods in combination with various stratification tools (optimum strata boundaries, number of strata and optimum sample size etc.) for obtaining efficient estimators of area and production of apple in Himachal Pradesh. For this purpose, various aspects involved in optimum stratification with reference to data collected from the selected orchardists in district Shimla, during the year 2011-12 have been analyzed. The variable "Area under Apple" was chosen as the stratification variable as it had high correlation ( $r=0.96$ ) with the estimation variable "Production of Apple". Four methods of construction of strata viz., equalization of strata total, equalization of cumulative of  $\sqrt{f(y)}$ , equalization of cumulative of  $\frac{1}{2}\{r(y)+f(y)\}$  and equalization of cumulative of  $\sqrt[3]{f(y)}$  were used and their relative efficiencies for estimating total production of apple in the study district of the state have been analyzed. The critical examination of the result revealed 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). Thus, equalization of cumulative of  $\sqrt[3]{f(y)}$  method with  $L>2$  can profitably be applied for the estimation of apple production in the study district of the state Himachal Pradesh, India.

**Keywords:** Gain in efficiency, Multi-stage sampling, Neyman allocation, Relative efficiency

### INTRODUCTION

Among horticultural crops of the state, apple, grown in temperate and dry temperate zones, is so far the most dominating leading cash fruit crop and accounts for about 49 per cent of total area under fruit crops (2,24,352 hectares) and more than 83 per cent of the total fruits production (7,51,938 MT). The state is well recognized as the 'Apple State of India'. The area under apple and its production has increased from 88,673 hectares and 49,129 MT in 1999-00 to 1,09,553 hectares and 6,25,199 MT in 2014-15 (Anonymous, 2016), respectively. A suitable technique for estimating area and production of apple with a desired degree of precision will help in planning for achieving sustainable production plans. The accuracy and adequacy of basic statistics regarding resource availability and their types determines the success of agricultural development planning. Sample surveys are carried out to collect information from different agro climatic zones using stratification methods based on geographical contiguity. However, this is not possible when primary purpose is to obtain estimate with maximum precision. Therefore, the stratified sampling consists in classifying the population units into certain number of groups having minimum variance called strata and then select-

ing samples independently from each group or stratum. An appropriate estimate for the population as a whole is obtained by suitably combining the stratum wise estimates of the characteristic under study. The method of  $\sqrt[3]{f(y)}$  choosing the best boundaries that makes strata internally homogeneous, given some sample allocation, is known as Optimum Stratification (Singh and Sukhatme, 1969). In order to make the strata internally homogenous, the strata should be constructed in such a way that the strata variances for the characteristic under study be as small as possible. The factors that influence the reduction of variance are choice of stratification variable, number of strata, and determination of strata boundaries and allocation of samples. Proper statistical based estimation technique for the estimation of area and production of fruit crops is lacking and presently followed convenience based stratification method leads to over or under estimates, thereby affects the horticultural planning process. The present study aims at constructing the Optimum Strata Boundaries (OSB) and other related aspects of optimum stratification with a view to improve upon efficacy over current methods in use.

### MATERIALS AND METHODS

Shimla is the leading apple producing district in Hima-

chal Pradesh with a production of 3,36,753 MT which accounts for more than fifty percent of total apple production of the State (Anonymous, 2009), was selected purposively for the study at the first stage. However, during 2014-15, a production of 4,07,751 MT was obtained from 38,781 hectares under apple cultivation in Shimla district (Anonymous, 2016), accounting for 65.22 and 35.40 percent of total apple production and area under apple respectively of the State. In the first instance, a complete listing of development blocks in the apple growing areas of the district was done from the records of the concerned revenue office. The selection of sample was made through the process of multi-stage random sampling (Cochran, 1963). Chopal, Jubbal- Kotkhai and Theog blocks (30 per cent of total blocks) were selected out of nine blocks of district Shimla, viz., Shimla, Jubbal-Kotkhai, Rohru, Narkanda, Theog, Rampur, Chopal, Chirgaon, Mashobra-Basantpur at the second stage. These blocks were divided into two groups viz., (i) those panchayats having more than thirty per cent of the area under cultivation and (ii) those panchayats having less than thirty per cent of the area under cultivation, as per the practice in vogue. Selection of panchayats constituted the third stage. At the third stage, a 20 per cent random sample from each group was chosen. List of villages in selected panchayats was prepared. Again villages were divided into two categories viz., (i) those villages having more than thirty per cent of the area under cultivation and (ii) those villages having less than thirty per cent of the area under cultivation. At the fourth stage, 20 percent of villages from each category were selected. Thus, three clusters (Chopal, Kotkhai and Theog) were formed. A complete enumeration of all the selected villages was done and the list of the households was prepared from panchayat offices and with the help of Patwari of the village. Only those orchardists having fifty or more apple plants bearing on their farm were included in the list. From the list of commercial apple growers so prepared, fifty households from each cluster were selected based on probability proportion method. Thus a sample of 50 growers from each block i.e. a random sample of 150 respondents (orchardists) was selected through a two-step approach as suggested by Stein (1945) and Cox (1952) at the fifth stage. Similar methodology has been followed by Mahajan *et al.* (2004) and Pathania *et al.* (2014). The variable "Area under Apple" was chosen as the stratification variable as it had high correlation ( $r=0.96$ ) with the estimation variable "Production of Apple" (Chandel, 1984). 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  $\frac{1}{2} \{r(y) + f(y)\}$  and equalization of cumulative of  $\sqrt[3]{f(y)}$  have been tried to find out the optimum points of stratification for varying number of strata 2 to 4. The relative efficiencies of different methods of strata were examined when the number of strata was 2, 3 and 4 under different sample allocation methods. Further, relative efficiencies for different methods of estimation were also examined to estimate the total production of apple in Himachal Pradesh.

## RESULTS AND DISCUSSION

The information on area and production of apple was collected from the selected respondents. Table 1 gives the frequency distribution of the respondents according to the area under apple. The data revealed that distribution of holdings was highly skewed and most of the units (67) were located in the 0-2 class interval followed by 30 units in 2-4 class interval.

For the present study the optimum points of stratification, 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  $\frac{1}{2} \{r(y) + f(y)\}$  and equalization of cumulative of  $\sqrt[3]{f(y)}$  and their relative efficiencies for estimating total production of apple in the study district of the state 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. They proposed cum  $f(y)$  rule for obtaining AOSB. For allocation proportional to strata totals, with SRS, the AOSB were obtained by Yadava

**Table 1.** Construction of strata by equalization of strata totals.

Class Interval (Area in ha.)	N <sub>h</sub> (Number of respondents)
0-2	67
2-4	30
4-6	17
6-8	13
8-10	9
10-12	6
12-14	7
14-16	0
16-18	1
Total	150

**Table 2.** Strata boundaries for varying number of strata by different methods.

Stratum	Number of strata			
	1	2	3	4
	<b>Equalization of strata totals method</b>			
L = 2	7.14 (81.33)	<sup>3</sup> 7.15 (18.67)		
L = 3	4.92 (70.67)	9.46 (17.33)	<sup>3</sup> 9.47 (12.00)	
L = 4	3.78 (64.00)	7.14 (17.33)	10.82 (11.33)	<sup>3</sup> 10.83 (7.33)
	<b>Equalization of cumulative of <math>\sqrt{f(y)}</math> method</b>			
L = 2	4.77 (68.00)	<sup>3</sup> 4.78 (32.)		
L = 3	2.72 (56.67)	7.41 (26.00)	<sup>3</sup> 7.42 (17.33)	
L = 4	1.86 (41.33)	4.77 (26.67)	8.98 (19.33)	<sup>3</sup> 8.99 (12.67)
	<b>Equalization of cumulative of <math>\frac{1}{2}\{r(y) + f(y)\}</math> method</b>			
L = 2	4.11 (66.00)	<sup>3</sup> 4.12 (34.00)		
L = 3	2.09 (49.33)	7.26 (33.33)	<sup>3</sup> 7.27 (17.33)	
L = 4	1.53 (36.67)	4.11 (29.33)	9.23 (22.00)	<sup>3</sup> 9.24 (12.00)
	<b>Equalization of cumulative of <math>\sqrt[3]{f(y)}</math> method</b>			
L = 2	5.78 (76.00)	<sup>3</sup> 5.79 (24.00)		
L = 3	3.44 (60.00)	8.49 (26.67)	<sup>3</sup> 8.50 (13.33)	
L = 4	2.43 (56.00)	5.78 (20.00)	10.01 (14.67)	<sup>3</sup> 10.02 (9.33)

Note: The figure in parenthesis indicates the percentage of holdings falling in different stratum.

and Singh (1984). Wang and Aggarwal (1984) obtained OSB when stratification variable is positively skewed. Later on, Aggarwal and Singh (1984) discussed this problem under proportional allocation method. Rizvi (1998) considered the case of optimum stratification for the study of two variables using auxiliary information for various sampling schemes viz. proportional method of allocation, compromise method of allocation, linear regression, product, mixture and varying probability of selection, etc. Gunning and Horgan (2004) gave a simple and practicable algorithm for constructing stratum boundaries in such a way that the coefficients of variation are equal in each stratum is derived for positively skewed populations. The new algorithm is shown to compare favourably with the cumulative root frequency method (Dalenius and Hodges 1957) and the Lavallée and Hidiroglou (1988) approximation method for estimating the optimum stratum boundaries. The problem of determining optimum strata boundaries (OSB), which minimizes the variance of the estimated population mean under a weighted stratified balanced sampling, has been discussed as a mathematical programming problem (using dynamic programming technique) when strata are formed based on a single auxiliary variable with a varying measurement cost per units across strata by Khan *et al.* (2009).

**Comparison of efficiencies of different methods of stratification:** The relative efficiencies of above different methods of construction of strata were examined for varying number of strata (i.e. L=2,3 and 4) when the total sample size i.e. number of respondents under study was allocated to different strata by three allocation methods (Cochran, 1961); namely, equal, proportional and Neyman allocation methods. The variances under these allocations have been presented in Table 3.

The Table 3 reveals that the variance term goes on decreasing with the increase in number of strata in general for a particular sample size. Also the variance term decreases when the sample size is increased. It is interesting to note that the decrease in variance is least in case of Neyman allocation, which is in agreement with the theory. Barnabas and Sunday (2014) 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 the Optimum allocation procedure to be the most efficient. This encouraging result formed the basis of selecting Neyman allocation for further investigations. 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.

The critical examination of Variance of the apple production for varying number of strata (L = 2,3, and 4) for different sample sizes under different sample allocation methods (Table 3) reveals that for L= 2, and for varying sample size, equalization of cumulative of  $\frac{1}{2}\{r(y) + f(y)\}$  method provided the least variance (4.91) followed by equalization of cumulative of  $\sqrt[3]{f(y)}$  method (5.04) in case of Neyman allocation. The similar results were also obtained in case of equal allocation. Minimum variance was obtained as 6.46 for a sample of 40 units by equalization of cumulative of  $\frac{1}{2}\{r(y) + f(y)\}$  method followed by 6.89, which was obtained by equalization of cumulative of  $\sqrt[3]{f(y)}$  method. This suggests that the method of equalization of

**Table 3.** Variance of the apple production for varying number of strata (L = 2,3, and 4) for different sample sizes under different sample allocation methods.

Equal allocation						
Sample size	Equalization of strata total			Equalization of cumulative of $\sqrt{f(y)}$		
	2	3	4	2	3	4
10	46.19	42.84	12.63	16.10	14.07	11.42
20	14.42	14.18	11.73	8.42	7.44	5.93
30	11.16	11.14	9.34	9.47	6.73	3.30
40	10.63	9.95	8.72	6.97	3.37	2.75
Sample size	Equalization of cumulative of $\frac{1}{2}\{r(y) + f(y)\}$			Equalization of cumulative of $\sqrt[3]{f(y)}$		
	2	3	4	2	3	4
10	42.99	14.65	13.64	28.45	14.31	13.50
20	24.65	14.47	11.36	7.67	4.05	2.19
30	9.53	6.45	3.12	6.90	3.42	1.02
40	6.46	3.21	2.86	6.89	1.49	<b>0.93</b>
Proportional allocation						
Sample size	Equalization of strata total			Equalization of cumulative of $\sqrt{f(y)}$		
	2	3	4	2	3	4
10	66783.55	9689.27	42064.95	3388.33	2571.97	1068.42
20	48194.53	43698.48	33808.12	11138.63	5162.53	3363.59
30	39443.77	69243.85	17198.53	8365.87	5629.26	4088.58
40	57743.01	39604.74	16066.63	34875.54	9486.27	11216.22
Sample size	Equalization of cumulative of $\frac{1}{2}\{r(y) + f(y)\}$			Equalization of cumulative of $\sqrt[3]{f(y)}$		
	2	3	4	2	3	4
10	1393.82	1175.05	2607.39	4428.51	2385.26	2478.66
20	17983.72	2283.59	3937.22	3767.58	1952.41	2710.06
30	19223.13	5869.71	3452.93	8887.54	2477.20	<b>1693.54</b>
40	18658.95	9326.81	10572.18	11713.72	9648.39	6407.67
Neyman allocation						
Sample size	Equalization of strata total			Equalization of cumulative of $\sqrt{f(y)}$		
	2	3	4	2	3	4
10	41.97	39.76	28.40	14.93	8.88	7.29
20	14.27	7.39	5.78	8.40	3.87	3.35
30	9.69	5.65	5.11	8.24	3.51	3.13
40	9.70	5.00	4.46	6.83	3.08	2.38
Sample size	Equalization of cumulative of $\frac{1}{2}\{r(y)+f(y)\}$			Equalization of cumulative of $\sqrt[3]{f(y)}$		
	2	3	4	2	3	4
10	30.40	13.71	13.14	12.34	5.66	4.91
20	24.20	5.84	3.75	11.84	5.27	2.02
30	9.44	2.97	2.87	7.41	1.88	1.37
40	4.91	2.82	1.63	5.04	1.29	<b>0.89</b>

**Table 4.** Percentage gain in efficiency due to stratification – Neyman allocation.

Sample size	Equalization of strata total			Equalization of cumulative of $\sqrt{f(y)}$		
	2	3	4	2	3	4
10	1585.112	1672.869	2342.128	4456.397	7495.136	9125.448
20	2316.400	4475.514	5720.218	3934.115	8544.264	9880.055
30	2731.382	4681.413	5174.884	3210.141	7531.940	8456.238
40	1879.368	3645.601	4085.474	2667.833	5915.052	7662.258
Sample size	Equalization of cumulative of $\frac{1}{2}\{r(y) + f(y)\}$			Equalization of cumulative of $\sqrt[3]{f(y)}$		
	2	3	4	2	3	4
10	2188.177	4850.348	5060.950	5391.201	11744.711	13552.632
20	1365.972	5656.100	8825.345	2791.016	6274.251	16368.953
30	2802.037	8917.899	9229.391	3568.944	14111.667	19260.955
40	3714.756	6454.398	11163.964	3613.809	14179.077	<b>20418.160</b>

cumulative of  $\frac{1}{2}\{r(y) + f(y)\}$  can profitably be used for estimating apple production in Himachal Pradesh. Although, the least variance was shown by the method of equalization of cumulative of  $\frac{1}{2}\{r(y) + f(y)\}$ , the results for proportional allocation showed an erratic trend and hence cannot be generalized.

However, for number of strata greater than two, the equalization of cumulative of  $\sqrt[3]{f(y)}$  method proved to be best giving the least variances both in Neyman and equal allocations. However, the variance is least in case of Neyman allocation (0.89) as is suggested by the results, shown in Table 3, when number of strata considered was 3 and 4.

**Gain in efficiency due to stratification:** During the present study, in order to gauge the gain in efficiency of stratification ( $L > 1$ ) over no stratification ( $L = 1$ ), variances due to Neyman allocation have been compared with simple random sample variances. The results have been presented in Table 4. The Table 4 indicates that there is considerable gain due to stratification; but the maximum gain in efficiency is observed when the strata are constructed through “Equalization of cumulative of  $\sqrt[3]{f(y)}$ ” method followed by “Equalization of cumulative of  $\frac{1}{2}\{r(y) + f(y)\}$ ” method. It is also evident that percentage gain in efficiency increases with the increase in number of strata from 2 to 4 and sample size from 10 to 40. Similar methods were tried by Mahajan *et al.* (2004) for estimating apple production. They suggested using the stratified random sampling method along with equalization of strata total method for estimating the production of apple in Kullu district of Himachal Pradesh.

Therefore, it may be concluded that Equalization of cumulative of  $\sqrt[3]{f(y)}$  method with  $L > 2$  be used for greater efficiency in estimating the production of apple in Himachal Pradesh. With this finding, an attempt has also been made to estimate the total production of apple in Himachal Pradesh during 2011-12. From the selected 150 holdings, an optimum sample of 50 was randomly selected. This random sample was allocated according to Neyman allocation to four strata. Subsequently, the estimates of apple production along with the standard errors were obtained using simple random sampling ( $L = 1$ ) and stratified random sampling for 50 randomly selected holdings. The result indicated that the stratified random sampling was more precise method of estimation, with least standard error (63.52) as compared to simple random sampling method with standard error 256.43, hence better method of estimation.

## Conclusion

The results amply suggested that the stratified random sampling method of estimation, together with Equalization

of cumulative of  $\sqrt[3]{f(y)}$  method with number of strata more than two, be used for estimating the production of apple in Shimla district in particular and Himachal Pradesh state in general.

## REFERENCES

- Anonymous (2009). Statistical outline of Himachal Pradesh 2009-10. Department of Economics and Statistics, Shimla, Himachal Pradesh, India.
- Anonymous (2016). Statistical outline of Himachal Pradesh 2015-16. Department of Economics and Statistics, Shimla, Himachal Pradesh, India.
- Aggarwal, V. and Singh, R. (1984). On optimum stratification with proportional allocation for a class of Pareto distribution. *Commun. Statist. Theor. Math.* 13: 3107-3116.
- Barnabas, A. F. and Sunday, A. O. (2014). Comparison of allocation procedures in a stratified random sampling of skewed populations under different distributions and sample sizes. *International Journal of Innovative Science, Engineering & Technology.* 1 (8), 218-225.
- Chandel, S.R.S. (1984). A Handbook of Agricultural Statistics. 256. Achal Prakashan.
- Cochran, W. G. (1961). Comparison of methods of determining stratum boundaries. *Bull. Inst. Stat.* 38: 345-358.
- Cox, D. R. (1952). Estimation by double sampling. *Biometrika.* 39: 217-227.
- Dalenius, T. (1950). The problem of optimum stratification. *Skand. Akt.* 33: 202-213.
- Dalenius, T. and Gurney, M. (1951). The problem of optimum stratification. *Skand. Aktuarietidskr.* 34: 133-148.
- Dalenius, T. and Hodges, J. L. (1957). The Choice of Stratification Points. *Skandinavisk Aktuarietidskrift.* 198-203.
- Gunning, P. and Horgan, J.M. (2004). A new algorithm for the construction of stratum boundaries in skewed populations. *Survey Methodology.* 30(2), 159-166.
- Khan, M.G.M., Ahmad, N. and Khan, S. (2009). Determining the Optimum Stratum Boundaries Using Mathematical Programming. *J Math Model Algor* (2009) 8: 409.
- Lavallée, P., Hidioglou, M. (1988). On the stratification of skewed populations. *Sur. Methodol.* 14: 3-43.
- Mahajan, P. K. Kumar, S. and Negi, Y. S. (2004). Optimum stratification for estimating apple production in Himachal Pradesh. *The Horticulture Journal.* 17 (1):25-28.
- Mathew, O.O., Sola, A. F., Oladiran, B. H. and A. A. (2013). Efficiency of Neyman Allocation Procedure over other Allocation Procedures in Stratified Random Sampling. *American Journal of Theoretical and Applied Statistics.* 2(5):122-127.
- Pathania H, Juneja B, and Mahajan, P.K. (2014). Estimation of apple production in Kullu district of Himachal Pradesh. *Mathematical Sciences International Research Journal.* 3(1) 462-464.
- Rizvi (1998). Optimum stratification for two study variables using auxiliary information. Unpublished Ph.D. Thesis, PAU, Ludhiana.
- Singh, R. and Sukhatme, P.V. (1969). Optimum stratification. *Annals of the Institute of Statistical Mathematics.* 21: 515-528.
- Singh, R. and Parkash, D. (1975). Optimum stratification for equal allocation. *Annals of the Institute of Statistical*

- Mathematics. 27: 273-280.
- Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok, C. (1983). Sampling Theory Surveys applications. IASRI Publication.
- Stein, C. (1945). A two sample test for a linear hypothesis whose power is independent of the variance. *Annals of the Institutes of Statistical Mathematics.* 16: 243-258.
- Wang, M. C. and Aggarwal V.(1984). Stratification under a particular Pareto distribution. *Commun. Statist. Theor. Math.* 13: 711-735.
- Yadava, S. S. and Singh, R.(1984). Optimum stratification for allocation proportional to strata trials for simple random sampling scheme. *Commun. Statist. Theor. Math.* 15: 2793-2806.