Original Research Article

Optimum stratification for estimation of mango production in Himachal Pradesh

Abstract: Optimum stratification brings gain in precision in estimation of a characteristic of the population with limited time, money and human power. The primary data of area and production of mango of 325 mango orchardists of Himachal Pradesh were collected through well designed survey. The area under mango, auxiliary variable, was then subject to stratification in order to stratify the mango production, study variable. Four stratification methods: Equalization of Strata Total, Equalization of cumulative $\sqrt{f(y)}$, Equalization of cumulative $\sqrt[3]{f(y)}$, and Equalization of cumulative $\frac{1}{2}[r(y)+f(y)]$, were used for stratification of area under mango production into varying number of strata L = 3, 4, 5, 6. From each strata a SRSWOR sample was drawn of size n_i which was allocated by using proportional and Neyman allocation. After that estimate of mean and variance were computed for varying number of strata and for varying number of sample sizes n = 60, 90 and 120 allocated by proportional and Neyman allocation under these four stratification rules. The gain in precisions was also computed and is presented. It was found that $\sqrt[3]{f(y)}$ for L = 6 and n = 120 yield minimum variance and maximum gain in precision.

Key word: Stratification, Stratified random sampling, Estimation of mean and variance, Mango production, Optimum Strata Boundaries, Neyman allocation, proportional allocation

Introduction

Construction of optimum strata boundaries is of the most importance as it demarcates the optimum points on the frequency distribution such that the variance is reduced. 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 mango plantation" was used as the auxiliary variable which off course is highly correlated with the study variable. 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 325 mango orchardists were collected from 5 major mango growing 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 Mango in the mentioned districts of Himachal Pradesh. Mango production as the study variable, number of trees and area under mango cultivation as the auxiliary information were used in the estimation of mango production and area under mango plantation. The area under mango was then subject to stratification rules for varying number of strata L = 3, 4, 5 and 6. Then for varying sample sizes n = 60, 90 and 120 were allocated for each stratum by using proportional and Neyman allocation. Estimates of mean and variance were then computed under the same and presented in subsequent heading.

Material and Method

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 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 hth stratum 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

Optimum strata boundaries: Stratification rules were used for construction of strata boundaries are presented in Table 1 and Table 2 represents the demarcation points under various stratification rules and percentage of respondents that fall in hth stratum. Under stratification by Equalization of strata totals, for L=3, two points of demarcation were 1.08 and 2.04 ha respectively. The percentage of number of orchardists that fall in 1st, 2nd and 3rd stratum was found to be 66.15%, 21.53% and 12.38%. For L=4, three approx. OSB were found to be 0.88%, 1.56% and 2.37% ha with 56%, 22.77%, 13.23% and 8% of orchardists that fall in 1st, 2nd, 3rd and 4th stratum, respectively. Similarly we can check for all stratification rules.

The area under mango (ha) which is correlated with the study variable mango 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=3, 4, 5 and 6) 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.)

Estimation of Mango production: Minimum estimated variance of $\hat{\bar{y}}$ of mango production of all the four stratification method for varying strata and sample sizes under proportional allocation was found to be 0.055 in equalization of cumulative $\sqrt[3]{f(y)}$ rule for n =120 and L=6. Neyman allocation is always precise as compared to proportional allocation. Minimum estimate of variance of $\hat{\bar{y}}$ of mango production was found to be 0.032 in equalization of cumulative $\sqrt[3]{f(y)}$ rule for n =120 and L=6. The Tables 3 and 4 revealed that as the number of strata (n) and sample size (n) increases the variance is uniformly decreasing. The results revealed that under Neyman Allocation by using equalization of cumulative $\sqrt[3]{f(y)}$ rule gave minimum estimate of variance of the $\hat{\bar{y}}$ and provided 7.636 tons as an unbiased estimate of an average production of mango orchardists in the state. And the estimate of mango production in the state is estimated to be 48043.76 MT for the year 2021.

Estimation of area under mango: Area under mango was used as auxiliary information and was subjected to stratification to estimate the study variable mango production. This can also be used to estimate the unbiased estimate of mean and variance of $\hat{\bar{y}}$ of the area under mango itself. The mean and its variance of area under mango are presented in Tables 5 and 6. Minimum estimated variance of $\hat{\bar{y}}$ of area under mango of all the four-stratification method for varying strata and sample sizes under proportional allocation was found to be 0.0105 in equalization of cumulative $\sqrt[3]{f(y)}$ rule for n =120 and L=6. Variances of area under mango were also worked out by using Neyman allocation. Minimum variance of $\hat{\bar{y}}$ of area under mango was found to be 0.0101 in equalization of cumulative $\sqrt{f(y)}$ rule for n =120 and L=6. The result revealed that minimum estimate of variance of the $\hat{\bar{y}}$ was found to be under Neyman Allocation by using equalization of cumulative $\sqrt{f(y)}$ rule which provided 11.98 bigha i.e.0.96 ha as an unbiased estimate for the mean area under mango. This means that on an average according to the sample survey, mango orchardists have 11.98 bigha under mango cultivation in the state

Gain in efficiency: To check the gain in efficiency due to stratification over no stratification, percentage gain of varying sample and strata sizes using all the four methods were calculated. The percentage gain in efficiency due to proportional and Neyman allocation over no stratification of mango production are presented in Tables 7 and 8. In order to assess the gain in efficiency the estimates were compared with estimates obtained through Simple random sample without replacement of the respective sizes. It can be seen from the mentioned tables that there is uniform percent gain in efficiency as the number of strata (L) and sample size (n) is increasing.

For estimation of mango production by using proportional allocation, the maximum gain in efficiency was observed to be 273.75 % using the equalization of cumulative $\sqrt[3]{f(y)}$ rule. Under Neyman allocation maximum gain in precision was observed for n=120 and L=6 which was 544.58%. The relative precision of estimate of variance of $\hat{\bar{y}}$ from Neyman allocation to that of proportional allocation comes out to be 171.87%.

For estimation of Area under mango plantation by using proportional allocation, maximum gain in efficiency was found to be 338.23% using equalization of cumulative $\sqrt{f(y)}$ rule. In case of estimation of variance of $\hat{\bar{y}}$ when sample is drawn through Neyman allocation, maximum gain in efficiency was observed to be 391.03% using equalization of cumulative $\sqrt{f(y)}$ rule. The relative precision of the estimate of variance of $\hat{\bar{y}}$ by Neyman allocation to that of proportional allocation is 111.88%.

The results of the study suggests that with precise sampling technique such as stratified random sampling with stratification methods like equalization of cumulative $\sqrt{f(y)}$ and equalization of cumulative $\sqrt[3]{f(y)}$ can provide efficient estimate for the area under mango plantation and mango production in the state.

References

Bharti, Mahajan P. and Gupta R. (2017). Standardization of sampling technique for the estimation of apple production in Himachal Pradesh. *The Bioscan*, *12*(2): 1119-1121.

Dalenius, T. and Hodges, J. (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.

Mahalanobis, P. (1952). Some aspects of the design of sample survey. *Sankhya*, *12*: 1-7.

Sharma, A., Mahajan P. and Belwal. (2017). Statitsical investigation through stratified random sampling for apple production in Himachal Pradesh. *Journal of Applied and Natural Science*, *9*(3): 1718-1723.

Singh, R. and Sukhatme, P. (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.

Verma, M. and Rizvi, S. (2007). Optimum stratification for PPS sampling using auxiliary information. *Journal of Indian Society of Agricultural Statistics*, *61*: 66-76.

Table 1: Frequency Distribution of Area and Cumulative total of number of respondents by using different stratification method

classes	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.$\frac{1}{2}[r(y)+f(y)]$
0.00-0.55	126.00	0.28	34.65	34.65	11.22	11.22	5.01	5.01	0.04	126.04	63.02	63.02
0.55-1.10	93.00	0.83	76.73	111.38	9.64	20.87	4.53	9.54	0.15	93.15	46.57	109.59
1.10-1.65	44.00	1.38	60.50	171.88	6.63	27.50	3.53	13.07	0.26	44.26	22.13	131.72
1.65-2.20	33.00	1.93	63.53	235.40	5.74	33.25	3.21	16.28	0.37	33.37	16.69	148.41
2.20-2.75	11.00	2.48	27.23	262.63	3.32	36.56	2.22	18.51	0.48	11.48	5.74	154.15
2.75-3.30	11.00	3.03	33.28	295.90	3.32	39.88	2.22	20.73	0.59	11.59	5.80	159.95
3.30-3.85	3.00	3.58	10.73	306.63	1.73	41.61	1.44	22.17	0.70	3.70	1.85	161.80
3.85-4.40	2.00	4.13	8.25	314.88	1.41	43.03	1.26	23.43	0.82	2.82	1.41	163.21
4.40-4.95	1.00	4.68	4.68	319.55	1.00	44.03	1.00	24.43	0.93	1.93	0.96	164.17
4.95-5.50	1.00	5.23	5.23	324.78	1.00	45.03	1.00	25.43	1.04	2.04	1.02	165.19
Total	**325**		**324.78**		**45.03**		**25.43**				**165.19**	

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

Strata	**Equalization of Strata Total**						**Strata**	**Equalization of cumulative $\sqrt{\mathbf{f(y)}}$**					
	I	**II**	**III**	**IV**	**V**	**VI**		**I**	**II**	**III**	**IV**	**V**	**VI**
3	1.08	2.04					**3**	0.77	1.89				
%	66.15	21.54	12.31				**%**	48.31	38.46	13.23			
4	0.88	1.56	2.37				**4**	0.55	1.24	2.29			
%	56.00	22.77	13.23	8.00			**%**	38.77	30.46	22.76	8.00		
5	0.77	1.27	1.85	2.69			**5**	0.44	0.94	1.61	2.66		
%	48.31	23.08	15.08	8.00	5.54		**%**	28.62	31.69	18.77	15.38	5.54	
6	0.69	1.08	1.56	2.04	2.88		**6**	0.37	0.77	1.24	1.89	2.91	
%	46.15	20.00	12.62	8.92	7.69	4.62	**%**	19.69	28.62	20.92	17.54	8.62	4.62
Strata	**Equalization of cumulative $\sqrt[3]{\mathbf{f(y)}}$**						**Strata**	**Equalization of cumulative $\frac{1}{2}[\mathbf{r(y)} + \mathbf{f(y)}]$**					
	I	**II**	**III**	**IV**	**V**	**VI**		**I**	**II**	**III**	**IV**	**V**	**VI**
3	0.97	2.37					**3**	0.48	1.11				
%	61.23	30.77	8.00				**%**	30.77	36.62	32.62			
4	0.71	1.59	2.89				**4**	0.36	0.78	1.46			
%	46.77	32.00	16.61	4.62			**%**	19.69	28.92	28.31	23.08		
5	0.56	1.20	2.02	3.20			**5**	0.29	0.59	0.98	1.66		
%	39.38	29.23	19.08	9.85	2.46		**%**	10.77	29.85	21.23	19.08	19.08	
6	0.47	0.97	1.59	2.37	3.48		**6**	0.24	0.48	0.78	1.11	1.85	
%	30.77	30.46	17.54	13.23	6.15	1.85	**%**	5.85	24.92	17.85	18.77	19.08	13.54

Table 3 Estimate of variance of $\hat{\bar{y}}$ of Mango Production using proportional allocation (production in metric tons)

Equalization of Strata Total					Equalization of cumulative $\sqrt{f(y)}$				
sample	**Strata**				**sample**	**Strata**			
	3	**4**	**5**	**6**		**3**	**4**	**5**	**6**
60	0.163	0.156	0.150	0.101	**60**	0.230	0.165	0.140	0.114
90	0.117	0.114	0.103	0.075	**90**	0.115	0.102	0.084	0.073
120	0.100	0.098	0.092	0.064	**120**	0.096	0.092	0.069	0.066
Equalization of cumulative $\sqrt[3]{f(y)}$					**Equalization of cumulative $\frac{1}{2}[r(y) + f(y)]$**				
sample	**Strata**				**sample**	**Strata**			
	3	**4**	**5**	**6**		**3**	**4**	**5**	**6**
60	0.176	0.126	0.104	0.103	**60**	0.166	0.143	0.126	0.120
90	0.119	0.093	0.087	0.070	**90**	0.113	0.108	0.084	0.088
120	0.096	0.081	0.066	0.055	**120**	0.104	0.093	0.079	0.077

Table 4 Estimate of variance of $\hat{\bar{y}}$ of Mango Production using Neyman allocation (production in metric tons)

Equalization of Strata Total					Equalization of cumulative $\sqrt{f(y)}$				
sample	**Strata**				**sample**	**Strata**			
	3	**4**	**5**	**6**		**3**	**4**	**5**	**6**
60	0.133	0.120	0.109	0.093	**60**	0.142	0.135	0.120	0.095
90	0.102	0.102	0.077	0.055	**90**	0.089	0.085	0.075	0.058
120	0.071	0.069	0.059	0.049	**120**	0.082	0.068	0.067	0.042
Equalization of cumulative $\sqrt[3]{f(y)}$					**Equalization of cumulative $\frac{1}{2}[r(y) + f(y)]$**				
sample	**Strata**				**sample**	**Strata**			
	3	**4**	**5**	**6**		**3**	**4**	**5**	**6**
60	0.178	0.110	0.086	0.070	**60**	0.175	0.137	0.113	0.092
90	0.089	0.085	0.055	0.054	**90**	0.097	0.101	0.088	0.062
120	0.068	0.047	0.041	0.032	**120**	0.089	0.050	0.046	0.043

Table 5 Estimate of variance of $\hat{\hat{y}}$ of Area under mango using proportional allocation **(Area in bigha)**

Equalization of Strata Total					Equalization of cumulative $\sqrt{f(y)}$				
sample	**Strata**				**sample**	**Strata**			
	3	**4**	**5**	**6**		**3**	**4**	**5**	**6**
60	0.0300	0.0242	0.0216	0.0216	**60**	0.0297	0.0219	0.0216	0.0201
90	0.0261	0.0235	0.0207	0.0201	**90**	0.0287	0.0230	0.0187	0.0159
120	0.0257	0.0204	0.0181	0.0115	**120**	0.0271	0.0181	0.0128	0.0113
Equalization of cumulative $\sqrt[3]{f(y)}$					**Equalization of cumulative $\frac{1}{2}[r(y) + f(y)]$**				
sample	**Strata**				**sample**	**Strata**			
	3	**4**	**5**	**6**		**3**	**4**	**5**	**6**
60	0.0272	0.0210	0.0176	0.0128	**60**	0.0297	0.0270	0.0250	0.0167
90	0.0234	0.0204	0.0164	0.0120	**90**	0.0237	0.0206	0.0155	0.0119
120	0.0204	0.0158	0.0151	0.0105	**120**	0.0201	0.0200	0.0116	0.0106

Table 6 Estimate of variance of $\hat{\hat{y}}$ of Area under mango using Neyman allocation **(Area in bigha)**

Equalization of Strata Total					Equalization of cumulative $\sqrt{f(y)}$				
sample	**Strata**				**sample**	**Strata**			
	3	**4**	**5**	**6**		**3**	**4**	**5**	**6**
60	0.0282	0.0213	0.0184	0.0170	**60**	0.0235	0.0223	0.0203	0.0202
90	0.0269	0.0207	0.0166	0.0127	**90**	0.0225	0.0216	0.0193	0.0104
120	0.0263	0.0192	0.0154	0.0114	**120**	0.0204	0.0190	0.0145	0.0101
Equalization of cumulative $\sqrt[3]{f(y)}$					**Equalization of cumulative $\frac{1}{2}[r(y) + f(y)]$**				
sample	**Strata**				**sample**	**Strata**			
	3	**4**	**5**	**6**		**3**	**4**	**5**	**6**
60	0.0272	0.0248	0.0235	0.0131	**60**	0.0266	0.0245	0.0219	0.0138
90	0.0271	0.0222	0.0201	0.0124	**90**	0.0203	0.0193	0.0168	0.0126
120	0.0223	0.0193	0.0169	0.0115	**120**	0.0198	0.0159	0.0136	0.0109

Table 7 Percentage gain in efficiency due to stratification under proportional allocation of mango production **(Mango Production in tons)**

Equalization of Strata Total					**Equalization of cumulative $\sqrt{f(y)}$**				
sample	**Strata**				**sample**	**Strata**			
	3	**4**	**5**	**6**		**3**	**4**	**5**	**6**
60	74.76	83.57	89.98	184.20	**60**	24.29	72.99	104.10	151.62
90	90.96	96.06	115.73	196.68	**90**	94.45	118.03	166.22	205.80
120	106.33	108.89	123.69	223.50	**120**	113.27	123.63	197.92	210.30
Equalization of cumulative $\sqrt[3]{f(y)}$					**Equalization of cumulative $\frac{1}{2}[r(y)+f(y)]$**				
sample	**Strata**				**sample**	**Strata**			
	3	**4**	**5**	**6**		**3**	**4**	**5**	**6**
60	62.38	127.28	174.76	176.62	**60**	71.98	99.66	127.17	138.78
90	88.17	138.81	155.48	217.82	**90**	97.75	106.56	165.89	154.25
120	113.27	153.85	211.95	273.75	**120**	97.87	121.67	161.26	167.36

Table 8 Percentage gain in efficiency due to stratification under Neyman allocation of mango production **(Mango Production in tons)**

Equalization of Strata Total					**Equalization of cumulative $\sqrt{f(y)}$**				
sample	**Strata**				**sample**	**Strata**			
	3	**4**	**5**	**6**		**3**	**4**	**5**	**6**
60	114.12	138.13	162.30	208.72	**60**	101.75	111.31	138.29	199.77
90	119.13	119.47	188.66	303.52	**90**	150.64	162.32	196.78	283.57
120	190.90	195.93	251.48	317.57	**120**	151.75	203.83	205.05	387.31
Equalization of cumulative $\sqrt[3]{f(y)}$					**Equalization of cumulative $\frac{1}{2}[r(y)+f(y)]$**				
sample	**Strata**				**sample**	**Strata**			
	3	**4**	**5**	**6**		**3**	**4**	**5**	**6**
60	60.86	158.95	231.35	310.90	**60**	63.42	108.91	153.64	210.05
90	150.64	162.32	303.36	317.06	**90**	130.06	121.34	154.95	260.99
120	201.80	338.76	398.47	544.58	**120**	130.21	308.16	351.70	381.11