Heliyon 10 (2024) e31693

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Unlocking the secrets of apple harvests: Advanced stratification techniques in the Himalayan region

Abdullah Ali H. Ahmadini [a], Faizan Danish [b,*], Rafia Jan [c], Aafaq A. Rather [d], Yashpal Singh Raghav [a], Irfan Ali [e,**]

[a] Department of Mathematics, College of Science, Jazan University, P.O. Box. 114, Jazan 45142, Kingdom of Saudi Arabia
[b] Department of Mathematics, School of Advanced Sciences, Vellore Institute of Technology-Andhra Pradesh (VIT- AP) University, Inavolu, Beside AP Secretariat, Amaravati AP-522237, India
[c] Department of Statistics, Govt. Degree College Bejbehara, Anantnag, J&K, India
[d] Symbiosis Statistical Institute, Symbiosis International (Deemed University), Pune, 411004, India
[e] Department of Statistics and Operations Research, Aligarh Muslim University, Aligarh, 202002, India

ARTICLE INFO

Keywords:
Apple production
Stratified random sampling
Optimum stratification
Neyman allocation
Relative efficiency

ABSTRACT

This study focuses on standardizing sampling techniques and comparing various methods of sample allocation to effectively estimate apple area and production in the Himalayan region of India. We investigate different stratification tools,in formulating a sampling plan using information gathered from select orchardists in the locale during the 2016-17 period, it becomes essential to explore diverse methodologies to define the most suitable stratum boundaries, ascertain the requisite number of strata, and identify the optimal sample size. The stratification process, underpinned by the "Area under Apple" variable, which demonstrates a pronounced association with apple production, assumes a central role in this endeavor Several methods are utilized to construct strata, such as equalizing strata totals, cumulative equalization, equalization of ½{r(x) + f(x)} and equalization of $\sqrt{f(x)}$. We assess their efficiencies in estimating total apple production in the study district. The combination of the "Cum $\sqrt{f(x)}$" of Neyman allocation demonstrates the lowest variance and the highest efficiency within a range of 2–4 strata, coupled with an increase in sample size from 10 to 40. Consequently, it can be inferred that the "Cum $\sqrt{f(x)}$" method, particularly with $L > 2$, is preferable for estimating apple production in the Himalayan region of India.

1. Introduction

The apple's historical roots extend to Eastern Europe and Western Asia, yet its cultivation has flourished worldwide. Today, China stands as the foremost producer of apples, trailed closely by the United States, Poland, India, and Turkey. Notably, India's Himalayan region boasts an impressive apple cultivation landscape, spanning approximately 277.27 thousand hectares and yielding a remarkable 2241.71 thousand metric tons [1]. India's stature as a substantial apple exporter further underscores its prominence in the global market. export volume of 21085.23 metric tons valued at 5776.36 lakhs. Several states in India cultivate apples, with Jammu and

* Corresponding author.
** Corresponding author.
E-mail addresses: faizan.danish@vitap.ac.in (F. Danish), irfii.st@amu.ac.in (I. Ali).

https://doi.org/10.1016/j.heliyon.2024.e31693
Received 14 March 2024; Received in revised form 20 May 2024; Accepted 20 May 2024
Available online 22 May 2024

Kashmir (J&K) being the highest producer, followed by Himachal Pradesh, Uttarakhand, and Andhra Pradesh. In Jammu and Kashmir, the annual area and production of apples (in thousands) are 136.54 ha and 1672.71 metric tons, respectively, accounting for 49.24 % and 74.61 % of the total [1] or [2]. The major districts in Jammu and Kashmir contributing the most to apple production are as follows: [district names and associated production and area data would be listed here in].

Major Apple Producing District's in J&K [2]

District	Area (000 ha)	Production (000 Mt)
Baramulla	25.20	529.27
Shopian	21.61	248.04
Kulgam	18.91	209.21
Kupwara	19.02	278

Figs. 1 and 2 display mapsrepresenting Apple's Production (000 Mt) and cultivated area (000 ha) in major districts of J&K, India. It's important to note that the Jammu and Kashmir district map is included solely for representational purposes.

Figs. 3 and 4 represent pie charts illustrating the Production (000 Mt) and Area (000 ha)of Apple cultivation in major districts of J&K, India.

Fig. 5 displays a histogram representing theProduction (000 Mt) and the Area (000 ha)allocated to Apple cultivation in major districts of J&K, India.

Various methods have been proposed for stratification in different situations. For instance, the authors [3–6], and [7] have put forth different methods. Additionally, the authors [8,9], and [10] have utilized two stratification variables in their studies. Below is a paraphrased version of the paragraph you provided:

In a study, the researchers [11] introduced a Mathematical Goal Programming model to optimize stratum boundaries for an exponential study variable, addressing various cost and time objectives. Another study [12] explored the linear approximation of the Multivariate Stratified Sampling problem and illustrated it with practical examples. The researchers [13] proposed a Goal Programming method to determine optimal strata boundaries using bivariate variables in a multi-objective framework focused on minimizing variance. Additionally, researchers [14] developed an accurate resource allocation method, which was implemented alongside the BRKGA (Biased Random Key Genetic Algorithm) and GRASP (Greedy Randomized Adaptive Search Procedure) algorithms. In another approach, researchers [15] used dynamic programming to find stratification points for two correlated variables, comparing these with points obtained from other methods across various frequency distributions, and assessed accuracy using percentage relative efficiency from variance estimates. Further efforts, such as the Geometric method by researchers [16], aimed at skewed populations but showed limited applicability. Finally, researchers [17] discussed a method for calculating optimal stratum boundaries (OSB) and optimal stratum size (OSS) based on known per-unit stratum measurement costs or the survey's probability density function, demonstrating this method with empirical data from Wave 18 of the HILDA Survey general release dataset.

Fig. 1. Production(000 Mt) of Apple in major districts of J&K, India.

Fig. 2. Area (000 ha) of Apple in major districts of J&K, India.

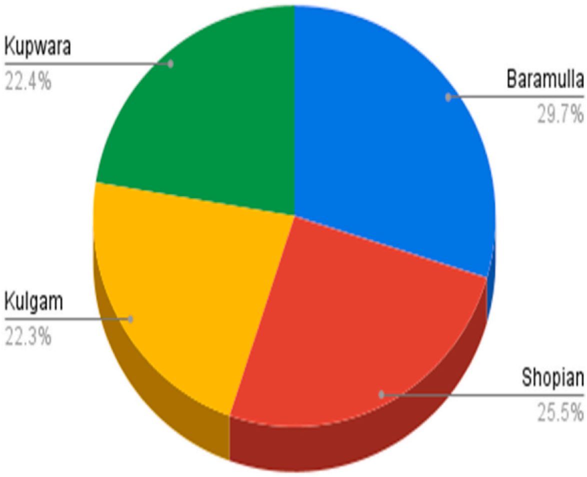


Fig. 3. Percentage of area (000 ha) of major apple producing districts of J&K, India.

2. Stratification and objectives of the study

Stratified sampling involves dividing a heterogeneous population into homogeneous sub-populations called strata to achieve optimum stratification, which minimizes variance. The main goal of stratification is to acquire a more representative sample from the population while improving relative precision. Typically, conducting a stratified sample survey involves five primary steps:

1. Selecting the stratification variable(s).
2. Determining the number of strata.
3. Defining boundaries for each stratum.
4. Allocating sample sizes to each stratum.
5. Choosing the sampling design within each stratum.

The objectives of this study include:

1. Developing a standardized sampling technique for estimating the area and production of apples in Jammu and Kashmir.

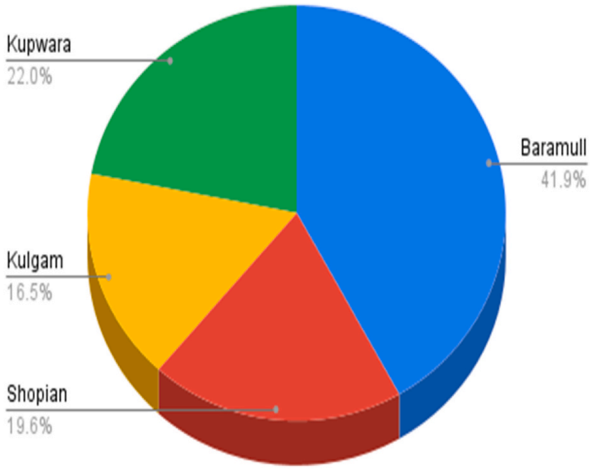


Fig. 4. Percentage of production (000 Mt) of major apple producing districts of J&K, India.

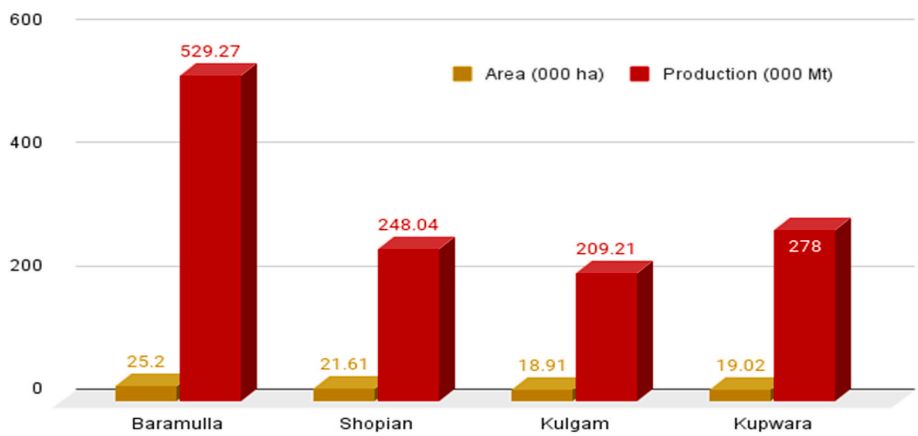


Fig. 5. Major apple producing districts in J&K, India.

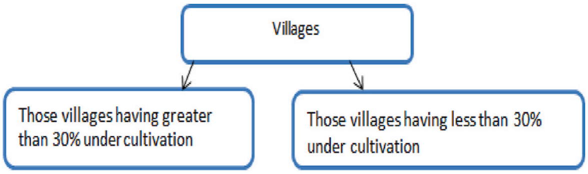
2. Comparing various methods of sample allocation to find the most efficient approach for estimating apple production.
3. Analyzing different stratification tools, including identifying OSB, determining the number of strata, and finding the ideal sample size.
4. Exploring the association of the "Area under Apple" variable and the estimation variable "Production of Apple."
5. Evaluating the performance of different methods for constructing strata.
6. Assessing the relative efficiencies of these methods in estimating aggregate response of apples in the study district of Shopian, Jammu and Kashmir, based on data collected from selected orchardists during the 2016-17 period.
7. Determining the most practical combination of the stratification method and Neyman allocation to minimize variance and maximize the percentage gain in efficiency.
8. Establishing the optimal number of strata and the range of sample sizes that yield the most accurate results for estimating apple production in the study district.
9. Conclude and recommend the preferred method for calculating aggregate response in the region, considering the combination of the equalization of $\sqrt{f(x)}$ method, based on the findings from the study.

3. Material and methods

Shopian, located in Jammu and Kashmir, is a prominent district known for apple production. It contributes 248.04 thousand metric tons (000 Mt) to the total apple production in Jammu and Kashmir, accounting for 14.82 % of the overall production by author [18].

To choose our sample, we utilized a multistage random sampling technique. Our first step involved creating a thorough inventoryof blocks having apple production. This compilation relied on records provided by the appropriate revenue officer. From this list, we randomly selected 30 % of the total blocks. During the selection process, we divided the blocks into two categories: those with more than 30 % cultivation and those with less than 30 % cultivation.

In the next stage, we proceeded to select the panchayats. Within each group, we randomly sampled 20 % of the panchayats. Subsequently, we prepared a list of villages within the selected panchayats. Once again, we classified the selected villages into two categories based on their cultivation levels: those with more than 30 % cultivation and those with less than 30 % cultivation. The diagram below illustrates this process:



In the fourth stage of our sampling process, we randomly selected 20 % of villages from each of the two categories identified earlier. This resulted in the formation of three clusters. We conducted a complete enumeration and prepared a list of households in the desired panchayats. We only included individuals with more apple plants in their orchards during this step.

In the fifth stage, we followed a two-step approach suggested by authors [19,20] to select respondents from each block within the three selected clusters. This methodology closely resembles the approach utilized by the authors [21,22]. With 1-β = 90 % and a type I of 5 % (α = 0.05), the expected number of individuals to be studies will be about 150 was clacluated as per the authors [23].The stratification variable chosen for this study was "Area under apple." With r = 0.96 with target variable, which is the resposne of apples, this factor played a pivotal role in determining the stratification pattern. The author [24], laid the groundwork for stratification in the context of stratified random sampling estimates, pioneering the concept of optimum stratification. Building upon his foundation, the authors [25], delved deeper into this realm tackling the challenge of optimum stratification by incorporating auxiliary variables to reduce variation in strataified design.

For constructing strata, as investigated by the authors [26], encompass Equalization of Strata Total, Equalization of ½ {r(y) + f(y)}, Cum $\sqrt{f(x)}$, and Cum $\sqrt[3]{f(x)}$. These approaches were explored to ascertain the optimal stratification points across varying numbers of strata, ranging from 2 to 4. Equalization of Strata Total involves creating strata of equal size while ensuring homogeneity within each stratum based on the total of the strata. Equalization of ½ {r(y) + f(y)} entails averaging the frequency of two distributions, where the frequency for each variable's class interval is generated, and then calculating the average of both frequencies to establish the stratification points.

Furthermore, the Equalization of Cumulative method, initially proposed by authors [27,28], equally stratify strata on the basis of frequency. This method, which remains widely utilized, offers an approximation for constructing strata. In this method, the frequencies for each class interval is first created by using any proposed method, and then the square root of each frequency is taken corresponding to each class interval. Finally, the stratification points are created based on the final column created by taking the square root of each class interval. In addition, the Equalization of Cumulative of $\sqrt[3]{f(x)}$ which was pioneered by author [3] in which the frequencies are made using $\sqrt[3]{f(x)}$ for each class interval, and then cube root is taken for all the frequencies generated by the method and proposed by the author [3]. Once the cube root frequencies are generated, the strata boundaries will be constructed using the final column of cube root frequencies.

The authors [29] introduced a mathematical programming approach to derive stratification points for one auxiliary variable and one study variable. By formulating an objective function and assuming distributions for the auxiliary variable, they generated strata boundaries, leveraging a dynamic programming approach to optimize the objective function. In parallel, the authors [30,31] explored alternative methods, utilizing frequency distributions of auxiliary variables to devise approximation techniques for determining Optimum Strata Boundaries (OSB). Additionally, the authors [32] framed the stratification challenge as a Mathematical Programming Problem (MPP), especially when total starta is predetermined. Meanwhile, The authors [29] proposed a method to derive optimal strata boundaries while minimizing the cost function under proportional allocation.

Table 1
Distribution of area and production.

Area in ha	N_h (Size of hth Stratum)
0–2	70
2–4	27
4–6	15
6–8	15
8–10	10
10–12	5
12–14	6
14–16	2
16–18	0
Total	**150**

4. Results and discussion

The frequency distribution of the responses recorded from all the respondents with regards to area and production are presented in Table 1 or Fig. 6, Figs. 7 and 8below:

Figs. 6–8, along with Table 1, reveal a pronounced skewness in the distribution of responses. Notably, a considerable majority of the units (70) were observed within the 0–2 Class Interval, followed by 27 units within the 2–4 Class Interval. Moreover, in this investigation, Table 2 presents the optimal stratification points determined through the utilization of four standard stratification methods:

In the tables provided earlier, the allocation of samples across different strata was determined using commonly adopted methods, including equal, proportional, and Neyman allocation techniques. The authors [33] focused specifically on addressing the optimal stratification challenge for equal allocation based on the auxiliary variable x. Similarly, the authors [34,35] focused on deriving optimal stratum boundaries, particularly when the stratification variable displayed positive skewness. The author [36] extended this investigation by examining optimal stratification for studying two variables alongside auxiliary information. This study encompassed a range of sampling schemes. Moreover, the authors [37] proposed a pragmatic algorithm for constructing stratum boundaries to ensure uniform coefficients of variation across each stratum, particularly for populations exhibiting positive skewness. Additionally, the author [38] introduced a method employing a dynamic programming approach to determine the stratification points.

5. Comparison of efficiencies of different methods of stratification

The efficiencies of various strata construction methods were analyzed for different numbers of strata (L = 2, 3, and 4). This analysis was conducted by distributing the total sample size, or the number of respondents, across the different strata using three allocation techniques: equal, proportional, and Neyman allocation methods, as described by Cochran (1961). The resulting variances for these strata boundaries are displayed in Table 3.

Table 3 demonstrates a general decrease in the variance term with an enhancement in sample size. Moreover, the variation tends to decrease as the sample size increases. Notably, the decrease in variance is minimal with Neyman allocation, consistent with theoretical expectations (the authors [39,40]). Further examination of the apple production variance across varying strata numbers (L = 2, 3, and 4) and diverse sample sizes under different sample allocation methods reveals compelling results. For instance, when L = 2 and with varying sample sizes, the Cum ½{r(y) + f(y)} method demonstrates the lowest variance (5.91), closely followed by the (6.04) under Neyman allocation. Similar trends are observed for equal allocation. Specifically, with a sample size of 40 units, the equalization of cumulative of ½{r(y) + f(y)} method achieves the minimum variance (7.46), succeeded by 7.89 with Cum $\sqrt[3]{f(x)}$. This suggests the potential utility of the Cum ½{r(y) + f(y)} method for estimating apple production in Jammu and Kashmir. However, while Cum $\sqrt[3]{f(x)}$ method displays the least variance, proportional allocation exhibits an inconsistent trend, limiting its generalizability. Nevertheless, L > 2 the Cum procedure perpformed better to be the most effective, yielding the lowest variances under both Neyman and equal allocations. Specifically, the variance is minimal with Neyman allocation (1.89), particularly evident when considering three or four strata, as indicated in Table 3.

6. Gain in efficiency due to stratification

In this investigation, the work aimed to assess the accuracy enhancement achieved through subdivision (L > 1) compared to no subdivision (L = 1), specifically focusing on variances attributed to Neyman allocation versus simple random sample variances. The

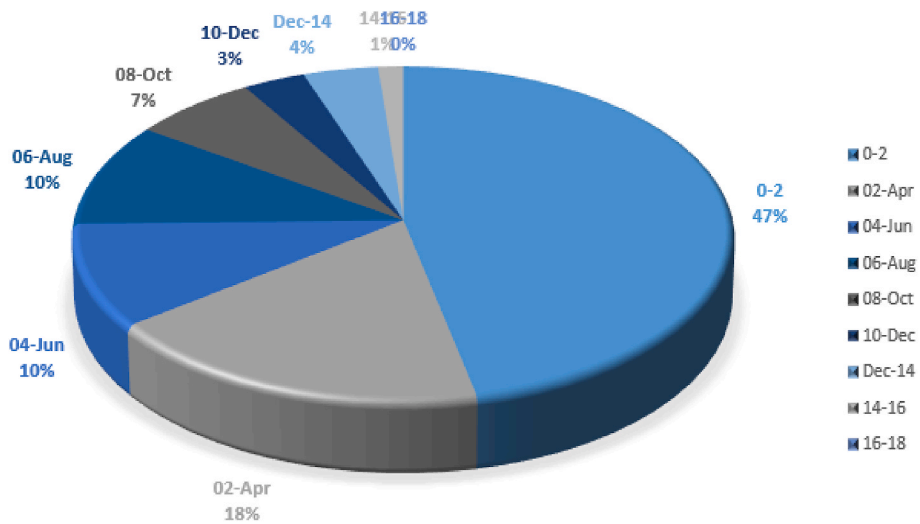


Fig. 6. Distribution of area and production.

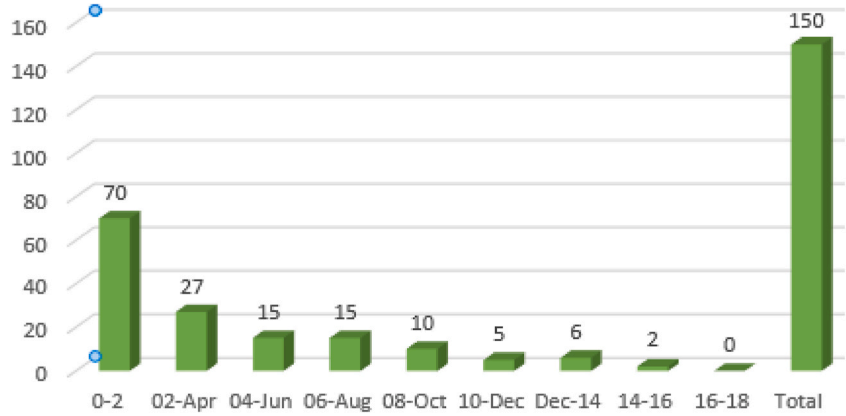


Fig. 7. Distribution of area and production.

Fig. 8. Distribution of area and production.

Table 2
Strataboundariesforvaryingnumbersofstratabydifferentmethods.

Stratum	Number of Strata			
	1	2	3	4
		Method of equality on starta total		
2	7.2374 (82.3343)	7.2456 (17.6785)		
3	4.8228 (71.6710)	9.4778 (16.3375)	9.4875 (12.0025)	
4	3.7991 (63.0002)	7.4963 (17.3345)	11.0545 (12.3375)	11.0623 (7.3356)
		Cum √f(y) Procedure		
2	4.7952 (69.0000)	4.8075 (31.0052)		
3	2.6276 (55.6701)	7.5145 (27.0063)	7.5252 (17.3342)	
4	1.4942 (40.3345)	5.7765 (27.6776)	9.1232 (19.6735)	9.1369 (12.3363)
Cum (½{r(y) + f(y)}) Procedure				
2	4.1572 (76.0012)	4.1654 (24.0025)		
3	2.5942 (50.3314)	7.7845 (32.3375)	7.7976 (17.3345)	
4	1.5412 (37.6777)	4.5274 (28.3343)	9.4235 (21.0025)	9.4378 (13.0078)
Cum $\sqrt[3]{f(x)}$ Procedure				
2	5.8978 (76.1275)	5.9075 (23.8885)		
3	3.4152 (59.0053)	8.4945 (25.6752)	8.5075 (15.3363)	
4	2.4217 (55.0035)	5.7963 (21.0045)	10.2252 (15.6775)	10.23 (8.3386)

Note: The percentage of holdings falling within different strata is denoted by the figure in parentheses.

outcomes are detailed in Table 4.

According to the data presented in Table 4, a substantial increase in efficiency is evident as a result of stratification. The most notable efficiency gain is observed when employing the "Equalization of cumulative of $\sqrt[3]{f(x)}$", followed closely by the "Equalization of

Table 3
Varaince for different sample size.

Equal Allocation						
Sample Size	Method of equality on starta total			Cum $\sqrt{f(x)}$		
	2	**3**	**4**	**2**	**3**	**4**
10	47.1778	43.9285	13.6589	17.1585	15.0578	12.4369
20	15.4375	15.1675	12.7263	9.4355	8.6552	6.9332
30	12.1845	12.1845	10.3652	10.3523	7.6545	4.3252
40	11.5663	10.9556	9.6545	7.9252	4.3525	3.7814
Sample Size	**Cum (½ {r(y) + f(y)})**			**Cum $\sqrt[3]{f(x)}$**		
	2	**3**	**4**	**2**	**3**	**4**
10	43.9978	15.6578	14.6478	29.4655	15.3178	14.5285
20	25.6552	15.4752	12.3652	8.6375	5.0552	3.2552
30	10.5345	7.4554	4.1656	7.9215	4.4276	2.0615
40	7.4641	4.2141	3.9663	7.8925	2.5658	1.9533
ProportionalAllocation						
Sample Size	Method of equality on starta total			Cum $\sqrt{f(x)}$		
	2	**3**	**4**	**2**	**3**	**4**
10	6531.2578	935.1378	4367.5963	369.5926	258.5358	138.2656
20	4635.3845	4638.7652	3648.2725	1386.5721	536.7263	376.4822
30	3678.1841	63879.5445	16934.3785	837.4658	537.6432	451.1723
40	5318.2945	3679.5142	1928.8323	3569.4956	953.1832	134.8521
Sample Size	**Cum (½{r(y) + f(y)})**			**Cum $\sqrt[3]{f(x)}$**		
	2	**3**	**4**	**2**	**3**	**4**
10	198.2745	195.8186	260.7378	423.5978	259.5652	281.7679
20	1563.8756	267.4663	364.1636	356.2822	167.6453	246.8582
30	1398.5756	536.8045	372.5963	825.4675	256.3423	**153.5223**
40	1593.2585	964.9169	1356.2845	1358.2269	964.2863	634.8223
NeymanAllocation						
Sample Size	Method of equality on starta total			Cum $\sqrt{f(x)}$		
	2	**3**	**4**	**2**	**3**	**4**
10	42.3156	40.7656	29.4130	15.9356	9.9985	8.2956
20	15.2863	8.2569	6.2840	9.4032	4.8232	4.6263
30	10.6933	6.6203	6.1156	9.2456	4.5153	4.1632
40	10.7016	6.0104	5.4608	7.8262	4.0825	3.3878
Sample Size	**Cum (½{r(y) + f(y)})**			**Cum $\sqrt[3]{f(x)}$**		
	2	**3**	**4**	**2**	**3**	**4**
10	31.4085	14.7156	14.1485	13.3475	6.6676	5.9193
20	25.2063	6.8265	4.7562	12.8453	6.2783	3.0537
30	10.4554	3.9752	3.8745	8.4163	2.8865	2.3876
40	**5.9150**	3.8263	2.6363	**6.0406**	2.2679	**1.8696**

Table 4
Percentage increase in efficiency.

Sample Size	Method of equality on starta total			Cum $\sqrt{f(x)}$		
	2	**3**	**4**	**2**	**3**	**4**
10	795.1148	836.7348	1174.2818	2223.5719	3747.2756	4564.4485
20	1158.0046	2236.5796	2860.7146	1965.2478	4271.2873	4940.0556
30	1382.1768	2440.8163	2580.7108	1605.2003	3713.5825	4223.2356
40	939.2863	1823.5718	2044.3013	1334.2717	2958.0525	3831.6313
Sample Size	**Cum(½{r(y) + f(y)})**			**Cum $\sqrt[3]{f(x)}$**		
	2	**3**	**4**	**2**	**3**	**4**
10	1094.1775	2425.3486	2530.9508	2695.2016	6822.7126	6726.6386
20	685.2815	2828.1005	4413.1006	1346.0166	6137.2818	8188.6196
30	1401.3752	4459.2813	4615.3913	1788.9445	7062.6186	9630.2795
40	1857.4128	3228.3989	5632.9648	1806.8093	7084.2765	**10209.1808**
	Danish et al.(2017)					
	2	**3**	**4**			
10	6738.0038	17056.7879	16816.676			
20	3365.0480	15343.2631	20471.5536			
30	4472.3696	17656.5586	24075.7862			
40	4517.02363	17710.6934	25522.9534			

cumulative of ½{r(y) + f(y)}" approach. It is clear that as the number of strata increases from 2 to 4 and the sample size grows from 10 to 40, there is a corresponding rise in efficiency percentage. Similar methodologies were utilized by Ref. [21] in their examination of estimating apple production. They advocated for employing the stratified random sampling method in conjunction with the "Equalization of strata total" approach to estimate Appple response in HP. Based on these findings, it can be inferred that the "Equalization of cumulative of" method with $L > 2$ should be employed for estimating apple production in Jammu and Kashmir with enhanced efficiency. Building upon these conclusions, an endeavor was made to calculate aggregate sum of response in J&K for the year 2017–18. A sample of 50 from 150 were selected and the sample was then allocated to four strata using Neyman allocation. The outcomes indicated that stratified random sampling offered a more precise estimation method, with a lower standard error of 58.74 compared to the standard error of 236.17 under Random sampling design. Consequently, Startified sampling design emerges as a superior method for estimating apple production in terms of precision.

7. Conclusion

The results highlight the effectiveness of employing a stratified random sampling approach in combination with the "Equalization of cumulative" method for estimating apple production in the Shopian district and across the broader landscape of the Jammu and Kashmir state. It is notable that Neyman allocation outperforms equal and proportional allocation methods, indicating its potential for enhancing precision. Furthermore, our findings highlight the significance of sample size in improving estimation accuracy. Increasing the sample size from 10 to 40 consistently enhances precision, with the "Equalization of cumulative" method demonstrating the highest accuracy, followed by the "Equalization of cumulative of ½{r(y) + f(y)}" method. By these insights, it is strongly recommended that stakeholders adopt the proposed approach of stratified random sampling with the "Equalization of cumulative" method and Neyman allocation for robust apple production estimates in both Shopian district and Jammu and Kashmir state. By implementing these methods and scaling up the sample size, stakeholders can significantly enhance the precision and reliability of production estimates, thereby facilitating informed decision-making and resource allocation in the apple industry.

8. Future scope

The future scope for the study's findings and recommendations are as follows:

8.1. Validation and replication

To strengthen the credibility of the study's findings, future research should focus on validating the results in other apple-producing regions within Jammu and Kashmir state. Replicating the study in different geographic locations with varying environmental and agricultural conditions would help confirm the robustness and applicability of the proposed method.

8.2. Scaling up to national level

Considering the significance of apple production in India, extending the research to a national level would be beneficial. Conducting a comprehensive study that includes multiple apple-producing states could provide valuable insights for policymakers and stakeholders in the Indian agricultural sector. Such research could also aid in formulating more effective and targeted policies for apple production and resource allocation.

8.3. Integration of advanced technologies

As technology advances, future studies should explore the integration of cutting-edge techniques, like as remote sensing, machine learning, and geospatial analysis, to improve the precision and efficiency of apple production estimation. Utilizing satellite imagery, drones, and advanced statistical algorithms could lead to more accurate and real-time estimations.

8.4. Addressing data limitations

The study's findings are based on data collected during a specific period. To enhance the reliability of the estimation method, researchers should consider collecting data over a more extended period or on a more frequent basis. Additionally, efforts should be made to obtain data from a larger sample size to ensure comprehensive coverage of apple orchards in the region.

8.5. Implications for resource management

Future research could explore the implications of the study's findings on resource management and planning in the apple industry. Understanding the accurate estimates of apple production can aid in optimizing resource allocation, such as water, fertilizers, and labour, leading to improved productivity and sustainability in apple cultivation.

8.6. Comparative studies

Conducting comparative studies between different agricultural products and crops using similar estimation methods could provide valuable insights into the strengths and limitations of the proposed approach. Such comparisons could contribute to a better understanding of the method's effectiveness in various agricultural contexts.

8.7. Economic analysis

In addition to precision and accuracy, future research could delve into the economic aspects of adopting the proposed estimation method. Analyzing the cost-benefit ratio and return on investment of implementing stratified random sampling with the "Equalization of cumulative" method and Neyman allocation would help stakeholders make informed decisions about its practicality and potential benefits.

Thus the study's recommendations lay a strong foundation for enhancing apple production estimation in Shopian district, Jammu, and Kashmir state. By addressing the future scope outlined above, researchers can contribute to advancing agricultural survey methodologies and supporting evidence-based decision-making in the apple industry at regional and national levels.

CRediT authorship contribution statement

Abdullah Ali H. Ahmadini: Writing – original draft, Visualization, Validation, Resources, Project administration, Funding acquisition, Formal analysis, Conceptualization. **Faizan Danish:** Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Rafia Jan:** Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Aafaq A. Rather:** Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yashpal Singh Raghav:** Writing – original draft, Visualization, Validation, Resources, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Irfan Ali:** Writing – review & editing, Visualization, Validation, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia, for funding this research work through project number ISP-2024. Furthermore, the authors express their sincere gratitude to the editor-in-chief, the associate editor, and the anonymous reviewers for their valuable feedback, which significantly enhanced the quality of the manuscript and aligned it with the high standards of the journal.

References

[1] Horticulture Statistics at a Glance, Govt.of India, ministry of agriculture and farmers welfare, department of agriculture, cooperation and farmers welfare, Horticulture Statistics Division (2017 (2017) 1–481.
[2] Digest of Statistics 2016-17,Director of Economics and Statistics, Govt. of Jammu and Kashmir, India.
[3] R. Singh, Approximately optimum stratification on the auxiliary variable, J. Am. Stat. Assoc. 66 (1971) 829–833.
[4] T. Dalenius, M. Gurney, The problem of optimum stratification-II, Scand. Actuar. J. 1951 (1–2) (1951) 133–148, https://doi.org/10.1080/03461238.1951.10432134.
[5] F. Danish, M.I. Jeelani, J.A. Reashi, Obtaining strata boundaries under proportional allocation with varying cost of every unit, Pak. J. Statistics Oper. Res. 13 (3) (2017) 567–574, https://doi.org/10.18187/pjsor.v13i3.1719.
[6] F. Danish, S.E.H. Rizvi, Optimum stratification in bivariate auxiliary variables under neyman allocation, J. Mod. Appl. Stat. Methods 17 (13) (2018), https://doi.org/10.22237/jmasm/1529418671.
[7] K.G. Bhuwaneshwar, M.I. Ahamed, Optimum stratification for a generalized auxiliary variable proportional allocation under a superpopulation model, Commun. Stat. Theor. Methods (2020) 1–16.
[8] F. Danish, S.E.H. Rizvi, C. Bouza, On approximately optimum strata boundaries using two auxiliary variables, RevistaInvestigacionOperacional 41 (3) (2020) 445–460.
[9] F. Danish, S.E.H. Rizvi, Approximately optimum strata boundaries for two concomitant stratification variables under proportional allocation, Statistics in Transition new series 22 (4) (2021) 19–40, https://doi.org/10.21307/stattrans-2021-036.
[10] F. Danish, S.E.H. Rizvi, Optimum stratification by two stratifying variables using mathematical programming, Pakistan Journal of Statistics 35 (1) (2019) 11–24.
[11] F. S Abo-El Hassan, R. Hamid, E. A Ismail, S. M Ezzat, Behavior of mathematical goal programming for determining optimum stratum boundaries, 130–143 (2021) (1) 26 المجلةالعلميةلقطاعكلياتالتجارة.
[12] ShokryaSaleh A. Alshqaq, Abdullah Ali H. Ahmadini, Irfan Ali, Nonlinear stochastic multi-objective optimization problem in multivariate stratified sampling design, Math. Probl Eng. 2022 (2022) 1–16.
[13] R. Hamid, E. A Ismail, S. M. Ezzat, Multi-objective mathematical programming model for optimum stratification in multivariate stratified sampling, 189–208 (2021) (4) 41 التجارةوالتمويل.
[14] J.A. Brito, L. de Lima, P.H. Gonzalez, B. Oliveira, N. Maculan, Heuristic approach applied to the optimum stratification problem, Oper. Res. 55 (2) (2021) 979–996.
[15] F. Danish, S.E.H. Rizvi, Approximately optimum strata boundaries under super population model, Int. J. Math. Oper. Res. (2022), https://doi.org/10.1504/IJMOR.2022.10052805.

[16] K.G. Reddy, M.G. Khan, S. Khan (2028). Optimum strata boundaries and sample sizes in health surveys using auxiliary variables, PLoS One 13 (4), https://doi.org/10.1371/journal.pone.0194787.
[17] K.G. Reddy, M.G.M. Khan, Constructing efficient strata boundaries in stratified sampling using survey cost, Heliyon 9 (11) (2023) e21407.
[18] W.G. Cochran, Comparison of methods of determin- ing stratum boundaries, Bull. Inst. Stat. 38 (1961) 345–358.
[19] C. Stein, A two sample test for a linear hypothesis whose power is independent of the variance, Annals of the Institutes of Statistical Mathematics 16 (1945) 243–258.
[20] D.R. Cox, Estimation by double sampling, Biometrika 39 (1952) 217–227.
[21] P.K. Mahajan, S. Kumar, Y.S. Negi, Optimum stratification for estimating apple production in Himachal Pradesh, The Horticulture Journal 17 (1) (2004) 25–28.
[22] H. Pathania, B. Juneja, P.K. Mahajan, Estimation of apple production in Kullu district of Himachal Pra- desh, Mathematical Sciences International Research Journal 3 (1) (2014) 462–464.
[23] F. Faul, E. Erdfelder, AG And Lang, A. Buchner, G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences, Behav. Res. Methods 39 (2) (2007) 175–191.
[24] T. Dalenius, The problem of optimum stratification, Skand.Akt. 33 (1950) 202–213.
[25] T. Dalenius, M. Gurney, The problem of optimum stratification, Skand.Aktuartidskr. 34 (1951) 133–148.
[26] P.V. Sukhatme, B.V. Sukhatme, S. Sukhatme, C. Asok, Sampling Theory Surveys Applications, IASRI Publication, 1983.
[27] T. Dalenius, J.L. Hodges, Minimum variance stratification, J. Am. Stat. Assoc. (1959) 88–101.
[28] T. Dalenius, J.L. Hodges, The choice of stratification points, SkandinaviskAktuarietidskrift (1957) 198–203.
[29] F. Danish, S.E.H. Rizvi, M.I. Jeelani, J.A. Reashi, Obtaining strata boundaries under proportional allocation with varying cost of every unit pak, j.stat.oper.res. 13 (3) (2017) 567–574.
[30] S.E.H. Rizvi, J.P. Gupta, R. Singh, Approximately optimum stratification for two study variable using auxiliary information, J. Indian Soc. Agric. Stat. 53 (3) (2002) 287–298.
[31] R.K. Gupta, R. Singh, P.K. Mahajan, Approximate optimum strata boundaries for ratio and regression estimators, Aligarh Journal of Statistics 25 (2005) 49–55.
[32] M.G.M. Khan, N. Ahmad, S. Khan, Deter- mining the optimum stratum boundaries using math- ematical programming, J. Math. Model. Algorithm. 8 (2009) 409.
[33] R. Singh, D. Parkash, Optimum stratification for equal allocation, Ann. Inst. Stat. Math. 27 (1975) 273–280.
[34] S.S. Yadava, R. Singh, Optimum stratification for allocation proportional to strata trials for simple random sampling scheme, Commun.Statist.Theor.Math. 15 (1984) 2793–2806.
[35] M.C. Wang, AggarwalV, Stratification under a particular Pareto distribution, Commun.Statist.Theor.Math. 13 (1984) 711–735.
[36] Rizvi, Optimum Stratification for Two Study Variables Using Auxiliary Information, Unpublished Ph.D. Thesis, PAU, Ludhiana, 1998.
[37] P. Gunning, J.M. Horgan, A new algorithm for the construction of stratum boundaries in skewed populations, Surv. Methodol. 30 (2) (2004) 159–166.
[38] F. Danish, Construction of stratification points under optimum allocation using dynamic programming, Pak. J. Statistics Oper. Res. 15 (2) (2019) 341–355, https://doi.org/10.18187/pjsor.v15i2.2635.
[39] A.F. Barnabas, A.O. Sunday, 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) (2014) 218–225.
[40] O.O. Mathew, A.F. Sola, B.H. Oladiran, A. A, Efficiency of neyman allocation procedure over other allocation procedures in stratified random sampling, Am. J. Theor. Appl. Stat. 2 (5) (2013) 122–127.