EVALUATION OF SAMPLE SIZE AND EFFICIENT FIELD SAMPLING PLAN IN HDP APPLE ORCHARDS

Tabasum Mushtaq[1], Mushtaq A. Lone[2], S. A. Mir[3], Sonali Kedar Powar[4], Aafaq A. Rather[5,*], Adil H. Khan[6], Faizan Danish[7]

•

[1,2,3]FOH, SKUAST –K, Jammu and Kashmir, India
[4]Department of Computer Science, Faculty of Science & Technology, Vishwakarma University, Pune
[5,*]Symbiosis Statistical Institute, Symbiosis International (Deemed University), Pune-411004, India
[6]Department of Statistics, University of Kashmir, Srinagar-190006, J&K, India
[7]Department of Mathematics, School of Advanced Sciences, VIT-AP University, Amravati, Andhra Pradesh-522237, India
[1]tabumushtaq2010@gmail.com, [2]mushtaqstat11@gmail.com, [3]mir_98@msn.com,
[4]sonali.powar@vupune.ac.in, [5,*]aafaq7741@gmail.com, [6]khanadil_192@yahoo.com,
[7]danishstat@gmail.com

Abstract

An essential stage in research is choosing an adequate sample size and sampling strategy. In order to obtain the most accurate estimates possible when surveying high density apple orchards, this paper provides the proper procedure for selecting the sample and an effective sampling strategy. For this study, primary information gathered during a two-year period from the SKUAST-Kashmir exotic apple block Plate I was employed. This investigation was conducted using the TCSA of exotic apple trees of the Gala and Fuji types. The sample was obtained using a variety of sampling techniques in order to find the parameters of population. Findings revealed that using proportional allocation of a stratified sample technique, in both the varieties, produces the most efficient population parameter estimates.

Keywords: Sample size, Stratification, Proportional Allocation, Gain in Efficiency.

1. Introduction

After data have been gathered, the most crucial goal of data analysis is to make generalizations about the population based on sample data. One of the most commonly requested questions by the investigators is "How big a sample is necessary." Inferences from the research cannot accurately reveal the reality of entire population when the sample size isn't calculated properly. The accuracy and caliber of research are impacted by inappropriate, insufficient, or excessive sample numbers. Choosing a sample size and addressing non-response bias are crucial in a quantitative survey design. The ability of quantitative methods to draw conclusions about vast populations that would be too expensive to research using smaller populations is one of their significant advantages [1]. Sample size is one of the four associated characteristics of a survey design that can impact the recognition of important disparity, correlations, or relations [2]. Precision, risk, and variability in the variables assessed are typically three criteria that need to be specified besides the research's objective and size of population to allot the right size of sample.

The "degree of precision" is used to find out the size of sample. The permitted discrepancy between the estimated value and the population value is known as the "degree of precision." Particularly, it is a measurement of how closely an estimate resembles the population's real distribution of a property [3]. You may refer to the degree of precision as sampling error. Cochran asserts that the level of mistakes that one is ready to accept in the sample estimates can be used to achieve the required precision. The SE is the discrepancy between the statistic & the relevant parameter of population [4]. It depends on how much hazard a researcher is eager to take when the data is used to inform their choices. Frequently it is stated as a percentage. If the degree of uncertainty for sampling is ±5 (%), and 70 (%) of the sample members ascribe a certain criterion, it can be inferred that 65% to 75% of members of the population also attribute that criterion. Higher expense and greater sample sizes are necessary for high levels of precision [5].

For the level of confidence or risk level the Central Limit Theorem's provides the basis [6]. Approximately 95% of the sample values in a normal distribution are within two standard deviations of the mean, or the real value of the population.

The degree of variability in the attributes being measured is referred to the distribution of the traits in the population. The sample size increases with population heterogeneity in order to achieve a particular degree of precision. Smaller sample sizes are ideal for populations that are less variable (more homogeneous). Remember that a proportion of 50% denotes a higher degree of variability than one of 20% or 80%. In order to calculate more cautious sample size, a proportion of 0.5 is frequently employed because it represents the population's maximum variability.

The choice of an appropriate survey plan, is a key component of the collection of information in any decision that is supported by science. A solid sampling strategy is essential to assure; the data are adequate to support the necessary inferences. It's critical to have reliable information while making science-based decisions. Creating a sample plan that appropriately reflects the issue under investigation is essential for obtaining valid and reliable data. Sampling is crucial in survey designs necessitating the human population and is receiving more and more heed from sociologists, pharmacologists, designers, accountants, physiologists, and medical professionals involved in business dealings, as well as from those in the fields of education, public administration, biostatistics, and even sociology, accounting, economics, anthropology, and political science [8]. Sampling plans are provided in order to end in fundamental investigation challenges, particularly in the social science disciplines and their uses [9]. In order to bring out interpretations about the population value from the sample value, the sampling plan also incorporates the assignments of estimate & selection [10]. With error-free measures, there are issues when extrapolating survey results from one group to another and often a larger population. The standard errors would differ depending on the sample designs. The main goal of sampling design is to choose the least error-prone option. An efficient sampling approach within a population display an adequate elicitation of relevant data, giving a significant understanding of the key elements of the population [7]. A sampling method that is cost-effective, effortless to use, bring in fair estimates, and lessens the consequences of sample-related volatility is therefore required. Multiple demographic parameters are frequently estimated in sample surveys; these parameters may be mutually exclusive. For the purpose of ensuring that samples represent all relevant points of view, stratified sampling has been established. The whole diversified population of interest is divided into homogenous classes, sometimes called as strata, each of which is similar within itself, using the stratification method of sample design. Depending on how important each stratum is to the population as a whole, the sample size changes for each stratum. Then, individual stratum sampling will be carried out [11]. There are several uses for stratified sample designs. These include obtaining more accurate estimates for fascinating domains. They also include increasing the true value of estimates for the entire population being gathered in the investigation [12]. The purpose of the experimentation is to establish sample size and choose the best sampling plan for surveys of apple orchards with a high population density. For this investigation, the high-density apple block (HDP, plate-1) of SKUAST-primary Kashmir's data on a variety of tree/fruit properties were employed. Gala and Fuji were two types of high-density apple trees selected in the study based on trunk cross-sectional area.

2. Methodology

According to Cochran (1977), the sample size n can be estimated by mentioning the margins of error for the survey items that are thought to be most important. For each of these significant factors, an estimation of the required sample size is first made. The data can be used to determine the sampling strategy. The largest can be used as the sample size if the values for the variables of interest are close together, and one can be sure that the sample size will produce the desired findings. Keeping in mind that a sample size that is too small will result in more estimation errors for demographic parameters and a sample size that is too large will result in higher survey costs to calculate the necessary sample size, a compromise between estimation accuracy and cost has to be made. Cochran's formula for sample size is:

$$n_0 = \frac{z^2\sigma^2}{e^2} \tag{1}$$

Where n_o is the sample size (without fpc factor), z = value for selected alpha level (1.96), σ = standard deviation = 0.17 and e= margin of error = 0.05.

Hence, for the population of 270, the calculated sample size is 47.704. But the sample size surpasses 5% of the population (270*0.05=13.5), Cochran's [4]. Correction formula must be incorporated to evaluate the optimum sample size. Put in the fpc factor out-turn in optimum sample size n, enumerated as in formula:

$$n = \frac{n_0 N}{n_0 + (N-1)} \tag{2}$$

$$n = \frac{47.704 * 270}{47.704 + (270-1)} = 41 \tag{3}$$

Where population size =270, n_o = required return sample size = 47.704, n_o = optimum return sample size.

Therefore, by using the Cochran's 1977 formula, the optimum sample size selected for study is 41 with the margin of error set up to 5% (0.05) and an alpha level of 1.96. In our study, sample selection methods included both stratified random sampling and plain random sampling. The TCA of each type of exotic apple plant was divided into three strata (TCA). The following guiding principles were followed in the stratification process: I The strata, or range of TCAs, do not overlap and make up the entire population as a whole. (ii) Within each stratum, or range of TCA, there is homogeneity [13].

The proportional allocation method was used to allocate the samples throughout the several TCA ranges. The sample fraction, n/N, in this approach is constant throughout all strata. With the help of this allocation, a sample that can estimate sample size more quickly and precisely was created. The allocation of a given sample of size n to different stratum was done in proportion to their sizes i.e. in the i^{th} stratum, $n_i = n\frac{N_i}{N}$ where n represents sample size, N_i refers to population size of the i^{th} strata and N is the size of population. In this investigation, population size (N) = 270; sample size (n) = 41

The variance of estimate of population mean in S_tRS (proportional allocation) and SRS (*wor*) is given by

$$Var(\overline{y_{st}})_{\text{Prop}} = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{i=1}^{k} W_i S_i^2 \tag{4}$$

$$Var(\overline{y}_n)_{Rand} = \left(\frac{1}{n} - \frac{1}{N}\right) S^2 \tag{5}$$

$$\text{where} \quad S^2 = \frac{1}{N-1}\sum_{i=1}^{k}\sum_{j=1}^{N_i}\left(Y_{ij} - \overline{Y}_N\right)^2 \tag{6}$$

$$S^2 = \sum_{i=1}^{k}(N_i - 1)S_i^2 + \sum_{i=1}^{k} N_i\left(\overline{Y}_{N_i} - \overline{Y}_N\right)^2 \tag{7}$$

The sample size in every stratum proportionate allocation differs relying on the total quantity of plants in each classified TCA, for duo exotic varieties. The quantitative values are SHOWN in Tables 1 and 2.

Table 1: *Grouping of TCA for Gala and Fuji: an empirical study*

Variety	Strata	N_i	n_i	$\overline{Y}_{N_i}$	$\overline{Y}_{N_i}^2$	$N_i\overline{Y}_{N_i}$	$N_i\overline{Y}_{N_i}^2$	S_i	S_i^2	N_iS_i	$N_iS_i^2$	$(N_i-1)S_i^2$
Gala	1	77	1	6.03	36.36	464.31	2799.79	0.86	0.72	66.22	56.95	56.21
	2	150	3	8.88	78.85	1332	11828.16	1.32	1.724	198.00	261.36	259.62
	3	43	07	12.78	163.33	549.54	7023.12	1.48	2.19	63.64	94.19	92.00
	Total	270	41	27.69	278.54	2345.85	21651.07	3.66	4.67	327.86	412.50	407.82
Fuji	1	48	07	3.16	9.99	151.68	479.31	0.98	0.96	47.04	46.10	72.99
	2	153	23	6.45	41.60	986.85	6365.18	1.81	3.28	276.93	501.24	488.14
	3	69	11	10.69	114.28	737.61	7885.05	2.27	5.15	156.63	355.55	216.42
	Total	270.00	41	20.30	165.86	1876.14	14729.54	5.06	9.39	480.60	902.89	777.55

Table 2: *Calculation of variances*

Sampling Strategy	HDP Variety	
	Gala	Fuji s
$Var(\overline{y}_n)_{Rand}$	0.128	0.189
$Var(\overline{y_{st}})_{Prop}$	0.032	0.069

Table 2 lists variations found in Gala and Fuji as a result of SRS &. S_tRS. Gala and Fuji hold variances of 0.118 and 0.189 under SRS, respectively, and .032 and .069 under S_tRS with proportionate allocation.

It is essential to examine the gain in efficiency (GE) brought on by separate kinds of allocations in order to determine how they affect sample size [16]. The efficiency improvement attributed to proportional allocation over SRS (wor) is given in equation below

$$(E-1) = \frac{Var(\overline{y}_n)_{Rand} - Var(\overline{y_{st}})_{Prop}}{Var(\overline{y_{st}})_{Prop}} \tag{8}$$

Table 3: *Gain in Efficiency and Percentage gain in Efficiency Due to Stratification*

Variety	Gain in Efficiency	% Gain in Efficiency
Gala	3.05	305
Fuji	1.34	134

According to Table 3, the efficiency gain (GE) for Gala and Fuji is 3.05 (305%) and 1.34 (134%) correspondingly.

3. Conclusion

The purpose of sample size formulas "is not merely to give a samples number," rather than to determine whether tens, hundreds, or thousands of plants are needed, but to scrutinize the study design, including a review of the validity and reliability of data collecting. Finally, it is concluded that proportional allocation in stratified sampling provides the best accurate estimates over simple random sampling for population parameters estimation in both of the HDP exotic apple types under investigation. This has been determined after using different sampling plans to obtain the sample statistic for population parameters.

References

[1] Singh, H., Caste and Premarket, (2022). Discrimination: Access to Civic Amenities and Healthcare Facilities in Rural Punjab, *Contemporary Voice of Dalit*, p. 2455328X2211069, doi: 10.1177/2455328X221106908.

[2] Gupta, S., Haq, A., and Varshney, R., (2022). Problem of Compromise Allocation in Multivariate Stratified Sampling Using Intuitionistic Fuzzy Programming, *Annals of Data Science 2022*, pp. 1–20, doi: 10.1007/S40745-022-00410-Y.

[3] Li, C., Wang, J., Wang, L. Hu, L. and Gong, P., (2014). Comparison of Classification Algorithms and Training Sample Sizes in Urban Land Classification with Landsat Thematic Mapper Imagery, *Remote Sensing 2014, Vol. 6, Pages 964-983*, vol. 6, no. 2, pp. 964–983, doi: 10.3390/RS6020964.

[4] Varshney, R., Khan, M. G. M., Fatima, U., and Ahsan, M. J., (2015). Integer compromise allocation in multivariate stratified surveys, *Ann Oper Res*, vol. 226, no. 1, pp. 659–668, doi: 10.1007/S10479-014-1734-Z/TABLES/2.

[5] Miranda, C., Santesteban, L. G., Urrestarazu, J., Loidi, M. and Royo, J. B., (2018). Sampling Stratification Using Aerial Imagery to Estimate Fruit Load in Peach Tree Orchards, *Agriculture 2018, Vol. 8, Page 78*, vol. 8, no. 6, p. 78, doi: 10.3390/AGRICULTURE8060078.

[6] McCravy, K. W., (2018). A Review of Sampling and Monitoring Methods for Beneficial Arthropods in Agroecosystems, *Insects 2018, Vol. 9, Page 170*, vol. 9, no. 4, p. 170, doi: 10.3390/INSECTS9040170.

[7] Varshney, R., Najmussehar, Ahsan, M. J., Varshney, R. and Ahsan, M. J., (2011). An optimum multivariate stratified double sampling design in presence of non-response, *Optimization Letters 2011 6:5*, vol. 6, no. 5, pp. 993–1008, doi: 10.1007/S11590-011-0329-8.

[8] Peers, I., (2006). Statistical Analysis for Education and Psychology Researchers: Tools for researchers in education and psychology, *Statistical Analysis for Education and Psychology Researchers*, doi: 10.4324/9780203985984.

[9] Ranjan, R., Sinha, R., Khot, L. R., Hoheisel, G. A., Grieshop, M., and Ledebuhr, M., (2021). Spatial Distribution of Spray from a Solid Set Canopy Delivery System in a High-Density Apple Orchard Retrofitted with Modified Emitters, *Applied Sciences 2021, Vol. 11, Page 709*, vol. 11, no. 2, p. 709, doi: 10.3390/APP11020709.

[10] Mukherjee, S., Baral, M. M., Pal, S. K., Chittipaka, V., Roy, R., and Alam, K., (2022). Humanoid robot in healthcare: A Systematic Review and Future Research Directions," *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, pp. 822–826, doi: 10.1109/COM-IT-CON54601.2022.9850577.

[11] Roy, R., Baral, M. M., Pal, S. K., Kumar, S., Mukherjee S., and Jana, B., (2022). Discussing the present, past, and future of Machine learning techniques in livestock farming: A systematic literature review, *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, pp. 179–183, doi: 10.1109/COM-IT-CON54601.2022.9850749.

[12] Vishwakarma, D. K., Kumar, R., Kumar, A., Kushwaha, N. L., Kushwaha, K. S., and Elbeltagi, A., (2022). Evaluation and development of empirical models for wetted soil fronts under drip irrigation in high-density apple crop from a point source," *Irrig Sci*, vol. 1, pp. 1–24, doi: 10.1007/S00271-022-00826-7/FIGURES/17.

[13] Ranjan, R., Sinha, R., Khot, L. R., Hoheisel, G. A., Grieshop, M. J. and Ledebuhr, M., (2021). Effect of Emitter Modifications on Spray Performance of a Solid Set Canopy Delivery System in a High-Density Apple Orchard, *Sustainability 2021, Vol. 13, Page 13248*, vol. 13, no. 23, p. 13248, doi: 10.3390/SU132313248.