



Research Paper

Cluster sampling of survey data and their application agriculture using R- software

■ M. IQBAL JEELANI, A.H. MIR, S. MAQBOOL, NAGEENA NAZIR AND
AJAZ A. SHAH

See end of the paper for authors' affiliations

Correspondence to :

NAGEENA NAZIR,
Division of Agricultural
Statistics, S.K. University
of Agricultural Sciences
and Technology (K),
Shalimar, SRINAGAR
(J&K)
INDIA

Paper History :

Received : 05.09.2011;

Revised : 06.11.2011;

Accepted : 15.12.2011

ABSTRACT : A common motivation for cluster sampling is to reduce the average cost per interview given a fixed budget this can allow an increased sample size. Assuming a fixed sample size the technique gives more accurate results when most of the variation in the population is within the groups, not between them. The present work is an attempt to show that cluster sampling is more efficient than simple random sampling provided the mean square within the clusters is maximum and there is a negative intra-class correlation coefficient between elements within clusters as relative efficiency of cluster sampling increases with increase in mean square within clusters. Different estimators of cluster sampling are applied and their results are compared with simple random sampling using the same sample size. Different computer programmes are developed using R-software. All these functions are run on real data set generated on Apple crop in year 2010-11 from district Ganderbal of Kashmir valley.

KEY WORDS : Cluster sampling, Simple random sampling, Intra-class correlation coefficient, R-software.

HOW TO CITE THIS PAPER : Iqbal Jeelani, M., Mir, A.H., Maqbool, S., Nazir, Nageena and Shah, Ajaz A. (2012). Cluster sampling of survey data and their application agriculture using R- software, *Internat. Res. J. agric. Eco. & Stat.*, 3 (1) : 35-39 .

INTRODUCTION

Cluster sampling is a sampling technique used when natural groupings are evident in a population. Cluster sampling is operationally more convenient, less time consuming and importantly cost wise efficient as compared to simple random sampling. In cluster sampling problems like imperfect sampling frames, improper stratification, hidden periodicity etc. do not arise. Cluster sampling is a technique where the entire population is divided into groups or clusters and a random sample of the selected clusters are included in the sample. The basic task here is to specify appropriate clusters or to divide the population into appropriate clusters. Clusters are generally made up of neighbouring units or of compact areas and therefore, the units within a cluster tend to have similar characteristics. As a simple rule, the number of units in a cluster should be small and the number of clusters should be large. Units within a cluster should be as heterogeneous as possible, but there should be homogeneity between cluster means. Each cluster should be a small scale representation of total population. A random sample technique is then used on any relevant cluster

to choose which clusters to include in the study.

The efficiency of cluster sampling has been studied by Smith (1938) and Hansen and Hurtwiz (1944), where it has been discussed that the relative efficiency of cluster sampling increases with the increase in mean square within clusters. On the basis of many agricultural surveys Jessen (1942), Mahalonobis (1944) and Hendricks (1944) developed a general law to predict how mean square within clusters changes with the size of cluster. Hansen and Hurtwiz (1944) discussed that in many practical situations, cluster size is positively correlated with the variable under study and in these cases, it is advisable to select the clusters with probability proportional to the number of elements in the cluster. A good discussion of numerical values of intra-class correlation coefficient for different elements within cluster in cluster sampling have been given by Hurtwiz and Madow (1953). They have shown the intra-class correlation coefficient as a "measure of homogeneity" of the clusters in cluster sampling.

In this paper a description of cluster sampling as compared to simple random sampling in reference to apple data has been given. The computations have been done by newly developed