



Research Paper

Stratified sampling allocation using R-software

■ S. MAQBOOL AND NAGEENA NAZIR

See end of the paper for authors' affiliations

Correspondence to :

S. MAQBOOL

Department of
Agricultural Statistics,
Sher-e-Kashmir,
University of Agricultural
Sciences and Technology
(K), SRINAGAR (J&K)
INDIA

ABSTRACT : Stratified sampling techniques have been widely used for surveys due to the benefits of efficiency. The purpose of stratification is to partition the population into disjoint sub-populations so that the power consumption characteristics within each sub-population are more homogenous than those in the original population. In this paper, we have developed two computer programs in R-software which can be implemented for stratified sampling allocation using equal and proportional allocation methods. The can also be used for drawing inferences of related parameters.

KEY WORDS : Stratified sampling, Even sample allocation, Proportional sample allocation, R-software

HOW TO CITE THIS PAPER : Maqbool, S. and Nazir, Nageena (2012). Stratified sampling allocation using R-software, *Internat. Res. J. agric. Eco. & Stat.*, **3** (2) : 326-328.

Paper History : Received : 19.04.2012; Revised : 06.07.2012; Accepted : 10.08.2012

INTRODUCTION

The precision of an estimate of the population mean or total besides sample size also depends on the variability among the units of the population. Therefore, apart from increasing the sample size, another possible way to increase the precision of the estimate could be to divide the population units into certain number of groups.. Smaller samples could then be selected from each of the groups so formed such that the total number of sampled units over all of the groups equals the required overall sample size. The groups thus formed are called strata and the process of forming strata is known as stratification.

The essence of the stratification is to classify the population into sub-populations or strata, based on some supplementary information, or it is even better to be able to select separate samples from each of the strata. The benefit of stratification derives from the fact that the sample sizes in the strata are controlled by the sampler, rather than being randomly determined by the sampling process. The strata sample sizes are usually proportional to the strata population size, being also known as proportionate stratification. R-Software is an advanced statistical computing system that is similar to S-Plus. It can be linked with C, C++ and FORTRAN codes, when necessary it can run computationally intensive tasks. R-software includes tools for data manipulation, matrix operations,

data analysis, graphical display, linear and non-linear regression, time series analysis, clustering, density estimation, Bayesian analysis, MCMC technique, neural networks and classical statistics. R-software developers can adopt these languages these language to develop a wide range of modules, many of which have been available in a public domain and are also available via the comprehensive R Archive Network (CRAN) [cran.us.r-project.org].

With its wide range of features, R-software has the potential of performing with the role to be the statistical and graphical engine.

MATERIALS AND METHODS

The following notation will be used to define the sample allocation problem. The decision variable of interest is the sample size for each stratum. The suffix h stands for h^{th} -stratum, $h=1, 2, 3, \dots, N_k$ is the population size and n_k is the sample size in the stratum h , where $N=\sum N_h$ and $n=\sum n_h$ represent the total population and sample sizes, respectively.

We define $f_k = n_k / N_k$ to be the sampling fraction in the stratum h , Moreover \bar{Y}_h and \bar{y}_h are the population mean and sample mean in the stratum h .

S_h^2 and s_h^2 are the population element variance and sample element variance in the stratum h .