BAYES AND MINIMAX PROCEDURES FOR ESTIMATING THE ARITHMETIC MEAN OF A POPULATION WITH TWO-STAGE SAMPLING

By Om P. Aggarwal

Iowa State University

1. Introduction. Two-stage sampling (or sub-sampling) in sample surveys is a procedure under which the population is considered as divided into a number of "clusters" or groups of units of the population and sampling is carried out in two stages. At first a sample of clusters is selected, and thus clusters become the first-stage sampling units or primary sampling units (p.s.u.'s); then a sample is taken from each of the selected p.s.u.'s. The ultimate units of the population are these second-stage sampling units. Stratification and cluster sampling are special cases of two-stage sampling when the rate of sampling is 100% at the first and the second stage, respectively. When the rate of sampling at each stage is less than 100%, one generally uses the term "two-stage sampling" or "subsampling."

In the current practice of choosing a survey design, the statisticians use one of the two principles: (i) to get an estimator of maximum precision for a given total cost of the survey, or (ii) to get an estimator of a given precision for a minimum total cost of the survey. The allocation of the resources for a given survey is usually carried out keeping one or the other of these two principles as the guide. The author [1] considered jointly the losses resulting from the errors in the estimators and from the cost of sampling, and obtained Bayes and minimax procedures for the estimation of mean in the case of an infinite population as well as a finite population. The loss function was taken as the sum of two components, one proportional to the square of the error of the estimator and the other proportional to the cost of obtaining and processing the sample. Both the case of a simple random sample and stratified random samples were discussed and a formula was obtained for the optimum allocation of the resources with a simple cost function.

In this paper we shall discuss two-stage sampling. We shall also use, for simplicity, the term "clusters" for first-stage units. The two cases, infinite and finite populations, shall be treated separately. For the sake of generality, we shall consider the case where the clusters are of unequal sizes and obtain the results for equal size clusters as a special case.

2. Infinite populations. Unequal-sized sampling at the second stage. Consider the situation where a statistician is required to estimate the mean μ of some random variable Y with a known upper bound σ_b^2 (> 0) for variance. He chooses a random sample of some predetermined size, say m $(\geqq 1)$, but is not allowed to or is unable to observe the values obtained, say $\mu_1, \mu_2, \cdots, \mu_m$. He is, however.

Received 8 December 1965; revised 20 April 1966.