Simple random sample equivalent survey designs reducing undesirable units from a finite population

Horng-Jinh Chang¹, Chih-Li Wang¹ and Kuo-Chung Huang²

¹Department of Management Sciences, Tamkang University, Tamsui, Taipei, Taiwan 25137, R.O.C., e-mail: chj@mail.tku.edu.tw
²Department of Business Administration, Chung-Yu Junior College of Business Administration, Keelung, Taiwan, 201, R.O.C.

Received: January 29, 2002; revised version: August 14, 2002

In this paper, we consider fixed size sampling plans for which the first order inclusion probabilities are identical for all units and the second order inclusion probabilities are constant for every pair-wise unit. The statistical conditions are identified under which these plans are equivalent to the usual simple random sampling plan. These sampling plans are constructed to reduce undesirable units.

Key words: Horvitz-Thompson estimator; inclusion probability; sampling without replacement.

1. Introduction

A typical survey sampling set up consists of a population \( U = (u_1, u_2, ..., u_N) \) on \( N \) labeled units with a value \( y_i \) attached to the unit \( u_i \) for \( i = 1, ..., N \). From the finite population \( U \) a sample
of size \( n \) is drawn without replacement. One of the problems of interest is to estimate \( \bar{Y} = \sum_{i=1}^{N} y_i / N \), the population mean, by observing the \( y \)-values on a subset of units in the population. Throughout this paper we will refer to the set \( S \) itself as a sample, and our choice for the estimator of population mean will be the well-known Horvitz-Thompson (1952) estimator

\[
\hat{Y} = \frac{1}{N} \sum_{i \in s} y_i, \tag{1.1}
\]

that has certain desirable statistical properties. Within this framework the only relevant features of a sampling plan are the first and second order inclusion probabilities. A popular choice is to select a sampling plan that has both its first and its second order inclusion probabilities equal. It is noted that a simple random sampling (SRS) plan is the most famous plan to achieve this. It is known that under simple random sampling, the first and the second order inclusion probabilities are respectively given by \( \pi_i = n/N \), for every individual unit, and \( \pi_{ij} = n(n-1)/N(N-1) \), for all pairs of units, \( i \neq j \). Having obtained these probabilities, the Horvitz-Thompson (1952) estimator reduces to the usual sample mean \( \bar{y} = \sum_{i \in s} y_i / n \). The sample mean is unbiased with variance given by

\[
V(\bar{y}) = \frac{N-n}{Nn} S^2, \tag{1.2}
\]

which can be unbiasedly estimated by

\[
\hat{V}(\bar{y}) = \frac{N-n}{Nn} s^2, \tag{1.3}
\]

where \( S^2 = \sum_{i=1}^{N} (y_i - \bar{Y})^2 / (N-1) \) and \( s^2 = \sum_{i \in s} (y_i - \bar{y})^2 / (n-1) \) have their usual meanings. It is clear that, for a fixed-size sampling design, the sampling error remains unchanged if the values \( \pi_i \) and \( \pi_{ij} \) are the same as that of simple random sampling.

In comparing different sampling designs, it always focuses on the accuracy levels of sample estimators. Except for the magnitude of sampling error, one may further exam the survey costs required since the budget is always restricted. Even though a sampling design is of high efficiency in estimation, it may still not be preferred if the survey costs required is quite high. And thus, another criterion for a good