



StatNews #11

What is a Complex Survey? October 1996

In a wide number of fields, data from surveys are increasingly being made available to researchers for secondary use. Often, these surveys are complex. For example, many national health and economic surveys are carried out by federal statistical agencies using complex designs. Also, many surveys of animal and human populations that academic researchers collect themselves involve complex sampling designs. This newsletter describes what features make a survey complex.

The purpose of sampling in surveys is to take measurements on a representative portion of the population so that the whole population does not have to be measured. Each observation in the sample can be thought of as representing a certain number of population members; this ratio is the sampling proportion. In a simple survey, each observation in the sample represents the same number of population members. Complex surveys differ from simple surveys in two fundamental ways.

First, a complex survey samples observations within strata, typically with disproportionate sampling probabilities. The population is divided into sub-populations, i.e., strata, and observations are sampled from within each stratum. For example, in many national surveys, strata are defined based upon geographic region. Observations sampled from one stratum may represent more or less population members than observations sampled from another stratum. That is, the sampling proportions across strata differ. Often adjustments to the sampling probabilities are made after the data are collected; this is called post-stratification adjustment. In order to have the sample represent the population, information on sampling proportions--usually called sample weights--is required.

Second, a complex survey has multiple stages of sampling. At every stage except the lowest stage, clusters of observations are sampled. At the lowest stage, individual observations are sampled. For example, a survey of school children may be done by sampling first schools, then classrooms within schools, and finally children within classrooms. This type of sampling is often required because it is logistically impossible, difficult, or expensive to sample observations directly. The use of multi-stage cluster sampling means that observations can not be assumed to be independent as is commonly done for a simple survey; observations that are from the same cluster will likely be more similar to each other than to observations from a different cluster.

In a subsequent newsletter, I will discuss the consequences of disproportionate, cluster sampling for the analysis of complex surveys, discuss software options, and provide references.

Author: Edward A. Frongillo, Jr.