

Sampling

by David A. Freedman
Department of Statistics
University of California
Berkeley, CA 94720

The basic idea in sampling is extrapolation from the part to the whole—from “the sample” to “the population.” (The population is sometimes rather mysteriously called “the universe.”) There is an immediate corollary: the sample must be chosen to fairly represent the population. Methods for choosing samples are called “designs.” Good designs involve the use of probability methods, minimizing subjective judgment in the choice of units to survey. Samples drawn using probability methods are called “probability samples.”

Bias is a serious problem in applied work; probability samples minimize bias. As it turns out, however, methods used to extrapolate from a probability sample to the population should take into account the method used to draw the sample; otherwise, bias may come in through the back door. The ideas will be illustrated for sampling people or business records, but apply more broadly. There are sample surveys of buildings, farms, law cases, schools, trees, trade union locals, and many other populations.

SAMPLE DESIGN

Probability samples should be distinguished from “samples of convenience” (also called “grab samples”). A typical sample of convenience comprises the investigator’s students in an introductory course. A “mall sample” consists of the people willing to be interviewed on certain days at certain shopping centers. This too is a convenience sample. The reason for the nomenclature is apparent, and so is the downside: the sample may not represent any definable population larger than itself.

To draw a probability sample, we begin by identifying the population of interest. The next step is to create the “sampling frame,” a list of units to be sampled. One easy design is “simple random sampling.” For instance, to draw a simple random sample of 100 units, choose one unit at random from the frame; put this unit into the sample; choose another unit at random from the remaining ones in the frame; and so forth. Keep going until 100 units have been chosen. At each step along the way, all units in the pool have the same chance of being chosen.