

Statistical Methods for Sampling Cross-classified Populations under Constraints

Louis-Paul Rivest

Université Laval

This presentation is concerned with populations that can be seen as a matrix with N rows and M columns. For instance, in a survey of tourists that visit a region, the rows of the matrix are N sites that can be visited by tourists while the columns are M days in the tourist season. The goal is to select a sample, among the MN possible units, that fulfills several constraints. The number of sites visited on a given day is often fixed; it depends on the manpower allocated to the survey. On the other hand, the number of visits to a site is set in advance and depends on the site's importance. The sample can be seen as a matrix Z with N rows and M columns with a 1 in cell (i,j) if site i is sampled on day j and 0 otherwise. Because of the constraints the row and the column totals of a sample matrix Z are fixed. The sample is selected according to a uniform distribution on all the matrices Z fulfilling these constraints. At each sampled point, the survey variable y is observed. The goal is to estimate the mean of y over the MN entries of the population matrix. This is done using the Horwitz-Thompson estimator. An unbiased estimator of its variance is proposed. Multilevel generalizations of the proposed sampling plan are also discussed; they are useful to address additional constraints. The analysis of data from a creel survey carried out with the proposed methodology is also presented.