Estimating the distance distribution of subpopulations for a large-scale complex survey

Jianqiang Wang

Iowa State University

E-Mail: TBP

Abstract: Many finite populations targeted by sample surveys consist of homogenous subpopulations with respect to the variables being collected. We propose a sample-based fully nonparametric estimator for the subpopulation distribution functions of the distances between elements and the subpopulation centers, and explore a general definition of distance metric and different ways to define the subpopulation centers. The usual mean vector and multivariate L_1 median are discussed as measures of center. We describe the theoretical properties of the estimator. The asymptotic variance of the estimator involves a gradient vector for which we propose a nonparametric estimator and this plug-in estimator is examined in simulation study. And a jackknife variance estimator is proposed as an alternative.