A contamination model for multivariate zero-inflated data

Davide Di Cecco, Ugo Guarnera, Danila Filipponi[1]

Abstract

We present a parametric model to estimate potential erroneous values in zero-inflated multivariate data. The model consists of a mixture of two (multivariate) Tobit models. One component would model units having an additional source of (erroneous) variability, and allows us to identify potentially erroneous data both among the zero values and in the positive part of the distribution.

Keywords: Selective Editing, Tobit model, finite mixture model

[1] Davide Di Cecco (mailto:dicecco@istat.itdicecco@istat.it), ISTAT, Italy; Ugo Guarnera (mailto:ugo.guarnera@agenziaentrate.itugo.guarnera@agenziaentrate.it), Agenzia delle entrate, Italy; Danila Filipponi (mailto:dafilipp@istat.itdafilipp@istat.it), ISTAT, Italy. The views and opinions expressed are those of the author and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.