Fitting models to complex survey data: a review

Pedro Luis do Nascimento Silva. Ph.D. en Estadística de la Universidad de Southampton, Inglaterra. Investigador Principal en la Metodología de Muestreo en el Instituto de Investigación de Ciencias de la Estadística de Southampton.Universidad de Southampton, Inglaterra.

Abstract: Data from complex sample surveys are frequently analysed with the aim of making inference about statistical models that are assumed to hold for a corresponding population. Such survey data typically differ from other data sets in three main aspects:

- the data is usually clustered due to the use of multi-stage cluster samples, implying that observations within the same cluster are correlated;
- sample units are often selected with unequal probabilities; when these probabilities are related to the outcome variable, the sampling becomes informative and the model holding for the sample is then different from the model holding in the population;
- survey data are almost inevitably subject to nonresponse, often of considerable magnitude, which again may distort the population model if the response propensity is correlated with the outcome (not missing at random nonresponse).

Several approaches have been proposed for modelling and analysing complex survey data with one or more of these features. A review of the various approaches available and a discussion of their relative merits is provided, considering a real modelling problem using data from a complex household sample survey.