

Robust estimation with survey data

Authors

- Tobias Schoch (University of Applied Sciences Northwestern Switzerland)

Abstract

Outlier detection and handling is a non-trivial task, even when the data are regarded as a random sample from an infinite population. In this context (i.e., classical statistics), outliers are typically considered to be generated by a model other than the one under study. Compared to classical statistics, outliers are a very different concept in finite population sampling. In the context of sampling (design-based inference), where no statistical model is assumed, outliers are extreme values that deviate from the bulk of the data. In addition, unlike in classical statistics, we also have to consider the sampling weights. Observations that are not considered outliers (i.e., that are in the bulk of the data) can still strongly influence an estimator due to their large sampling weight (influential values). An estimator or procedure is called (qualitatively) robust if it is resistant or insensitive to the presence of outliers and influential values. In principle, robust estimation can be implemented in two ways: i) detection and treatment of outliers, or ii) direct application of robust estimation techniques. We limit our attention to the latter approach. The `robsurvey` package implements: i) basic robust estimators of the mean and total (e.g., robust Horvitz-Thompson estimator), robust survey regression, and model-assisted estimation (e.g., robust generalized regression estimator, GREG). In the talk, we will take a look at some of the methods and illustrate them with examples from business surveys.

References

No References available