A Constrained Optimisation Approach for Adjusting Misclassified Proxy Variables

Fabrizio Solari, Romina Filippini, Valeria Quondamstefano, Gaia Rocchetti[1]

Abstract

When proxy variables replace unobserved targets, bias does not only affect point estimation but propagates to inference, prediction, and decisions in structurally consequential ways. This paper proposes a deterministic and distributionally coherent alternative to weighting, stochastic adjustment, or model-based replacement. The key idea is to treat proxy correction as a constrained reconstruction problem: the proxy is adjusted to restore consistency with the target distribution while preserving, as far as possible, the genuine individual-level signals it carries. The resulting correction is a controlled realignment of information that explicitly balances fidelity to observed structure with adherence to known constraints. An application to educational attainment data shows that the adjusted proxies are interpretable, reproducible, and statistically valid for both aggregate and unit-level analyses.

Keywords: misclassification, confusion matrix, constrained optimisation, transportation problem.

[1] Fabrizio Solari (solari@istat.it), Istat, Italy; Romina Filippini (filippini@stat.it), Istat, Italy; Valeria Quondamstefano (quondamstefano@istat.it), Istat, Italy; Gaia Rocchetti (grocchetti@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.