A Replacement Approach for the Analysis of Supposed Fake Data

Massimiliano Pastore
Department of Psychology
University of Cagliari
mpastore@unica.it

Luigi Lombardi
Department of Cognitive Science
University of Trento
lombardi@form.unitn.it

Keywords: Fake Data, Perturbation Analysis, Replacements.

Abstract

One of the major problems in psychological measurement is that in some circumstances one has no basis to assume that subjects always respond honestly. In fact some individuals tend to distort their responses in order to reach specific goals. For example, in personnel selection contexts it is possible that some subjects fake a personality questionnaire to match the ideal candidate's profile (positive impression management). Similarly, in administration of diagnostic tests individuals often attempt to malinger posttraumatic stress disorder (PTSD) in order to secure financial gain, to secure treatment, or to avoid charges from a crime. Supposed fake data confront the researcher with a crucial question: If my data contained possibly fake datapoints, would be the answer to my research question be different from what it is? Even in the clearest case – that is, randomly fake data – the answer is not necessarily obvious, because even the random perturbation of data constitutes a biased information that decreases the efficiency of parameter estimates and decreases the sensitivity of statistical techniques.

This paper reflects an attempt to contribute to the modeling of methods of treating possible fake data in simple statistical techniques. A new approach, called SGR (Sample Generation by Replacements) (Lombardi, Pastore & Nucci, 2003), is adopted in order to provide a perturbation model and an empirical procedure to generate a structured collection of perturbations. SGR is a combinatorial method that can be applied to discrete data with a restricted number of values (e.g. Likert scale) and consists of three different components:

1. a perturbation model,
2. a simple procedure to generate perturbed samples from a given real data set,
3. an index to evaluate fluctuations of statistical properties to perturbed data samples.

We present an illustrative application of SGR to the study of the sensitivity of some simple statistical indices. Finally we discuss the relation of the SGR method with standard Monte Carlo simulation studies.

References

Lombardi, L., Pastore, M. & Nucci, M. (2003). Sensitivity of fit indices to structured perturbations: a replacement approach. Manuscript submitted for publication.