

A New Method of Generating Synthetic Data for Public Use Files

Inês Ribeiro Viana

FEP, Faculdade de Economia do Porto, inessrviana@hotmail.com, Porto, Portugal

*Pedro Campos**

LIAAD/INESC TEC e FEP, Universidade do Porto, Porto, Portugal, pcampos@fep.up.pt

Keywords: Public use Files (PUF), Synthetic Data, Statistical Disclosure Control

Abstract

The availability of statistical information in PUF (Public Use Files), which can be accessed by every citizen, endangers the protection of individual respondents. To ensure compliance to confidentiality, there are a set of statistical disclosure control methods that can be applied to the data sets before they are published, aimed at reducing the risk of individual identification. In this paper, we present a new method (MEAC - Minimum Absolute Error of Variable Combinations) to generate synthetic data in microdata files.