



Contribution ID: 42

Type: **not specified**

## Clustering for Detecting Careless Respondents: A Case Study

The presence of careless respondents represents a well-known threat to the quality of survey data. Respondents who provide inattentive or random answers can distort statistical analyses, reduce measurement reliability, and bias substantive conclusions. A variety of indicators have been proposed in the literature to detect such respondents, including response pattern measures such as longstring indices, intra-individual response variability (IRV), and multivariate distance measures such as Mahalanobis distance. However, these indicators are often applied individually and require arbitrary thresholds, which may limit their effectiveness in complex datasets.

This work proposes a data-driven framework for identifying careless respondents that integrates several classical diagnostic indicators with clustering algorithms. First, multiple indicators capturing different aspects of response behavior are computed for each respondent, including measures of response consistency, variability, and multivariate outlyingness. These indicators are then jointly analyzed using unsupervised clustering techniques to identify groups of respondents with similar response patterns. In this way, respondents exhibiting atypical combinations of diagnostic indicators can be detected without relying on predefined cutoffs.

The proposed approach allows the simultaneous exploitation of complementary information contained in multiple quality indicators and provides a flexible and scalable tool for detecting potentially careless respondents. An empirical study based on survey data illustrates how the integration of traditional indicators with clustering methods can improve the identification of problematic response patterns and support more reliable data cleaning procedures in questionnaire-based research.

### Special/ Invited session

### Classification

Both methodology and application

### Keywords

careless responding, machine learning, smart data

**Primary authors:** BARZIZZA, Elena (University of Padova); Ms DISEGNA, Marta (University of Padova); Ms NAZIRI, Sara (Stat4Value); VILLANOVA, Silvia (University of Padova)

**Presenter:** VILLANOVA, Silvia (University of Padova)

**Track Classification:** Data Analytics and Data Science: Case Studies