

# Quantifying effects of rating-scale response bias: robustness properties of correlation measures

Aurore Archimbaud<sup>1</sup>, Max Welz<sup>1</sup>, and Andreas Alfons<sup>1</sup>

<sup>1</sup>Erasmus School of Economics, Erasmus University Rotterdam, Netherlands.

**Keywords:** Rating-scale data; bias curves; breakdown value; careless responding

Data obtained from questionnaires are common in many scientific fields, such as psychology, business, and social sciences. We focus on rating-scale data, which are measured by having respondents choose one answer category out of a fixed number of ordered answer categories for a given questionnaire item. Being ordered and discrete, rating-scale data are typically interpreted as numeric values. Moreover, they are often treated as interval data by assuming equal distances between adjacent values for performing classical statistical analyses. In psychometrics, rating-scale data are often a discrete measurement of a latent continuous variable (such as a personality trait), and multiple items measuring the same latent variable are called a *psychometric scale*.

Not much is known about the theoretical robustness properties (at the distributional level) of standard estimators when applied to rating scales. Due to the bounded and discrete nature of such data, existing results from robustness theory are often not applicable because of their implicit assumptions that outlying data points may be characterized by an arbitrarily large magnitude. In rating-scale data, one important example of outliers are responses of individuals with low or little motivation to comply with questionnaire instructions, correctly interpret item content, and provide accurate responses. Such individuals are called careless respondents, and there is substantial empirical evidence that even a low prevalence of careless respondents of about 5–10% can jeopardize the validity of research findings in questionnaire-based studies [1]. Another important example is *response faking* [2] being characterized by respondents who actively try to misrepresent themselves by, for instance, malingering or socially desirable responding.

Leveraging robustness theory, we study the statistical effects of rating-scale response biases at the distributional level, with a focus on correlational measures due to their key role in factor analysis, structural equation models and reliability measurements. In particular, we derive the bias curves for the Pearson correlation. These bias curves measure the estimation bias caused by having a fraction of participants responding carelessly or with the intention to misrepresent themselves in a certain way. In addition, we determine the maximum possible bias for the worst case of the most harmful type of contamination through maximum bias curves depending on the fraction of outliers. Finally, we investigate the breakdown value, which is the minimum contamination fraction required to flip the sign of the correlation estimator when the good data points are perfectly correlated. Specifically, we study how the adverse effects of response biases depend on the number of answer categories, the number of items in a psychometric scale, the construct reliability, and the properties of the psychometric scale itself. We find that already low prevalence of rating-scale response biases can render correlation measures fundamentally invalid. Furthermore, we provide freely available software in R for computation and visualization of bias curves.

- [1] V. B. Arias, L. Garrido, C. Jenaro, A. Martínez-Molina, and B. Arias, “A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data,” *Behavior Research Methods*, vol. 52, pp. 2489–2505, 2020.
- [2] D. S. Nichols, R. L. Greene, and P. Schmolck, “Criteria for assessing inconsistent patterns of item endorsement on the MMPI: Rationale, development, and empirical trials,” *Journal of Clinical Psychology*, vol. 45, no. 2, pp. 239–250, 1989.