RRTCS: An R Package for Randomized Response Techniques in Complex Surveys

Applied Psychological Measurement
2016, Vol. 40(1) 78–80
© The Author(s) 2015
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0146621615605090
apm.sagepub.com

María del Mar Rueda¹, Beatriz Cobo¹, and Antonio Arcos¹

Abstract

Randomized response (RR) techniques may be used to compile more reliable data, to protect the respondent's confidentiality, and to avoid an unacceptable rate of nonresponse when the information requested is sensitive (e.g., concerning racism, drug use, abortion, delinquency, AIDS, or academic cheating). Standard RR methods are used primarily in surveys that require a binary response to a sensitive question, and seek to estimate the proportion of people presenting a given (sensitive) characteristic. Nevertheless, some studies have addressed situations in which the response to a sensitive question results in a quantitative variable. RR methods are usually developed assuming that the sample is obtained using simple random sampling. However, in practice, most surveys are complex and involve stratification, clustering, and an unequal probability of selection of the sample. Data from complex survey designs require special consideration with regard to the estimation of finite population parameters and to the corresponding variance estimation procedures, due to the reality of significant departures from the simple random sampling assumption. In such a complex survey design, unbiased variance estimation is not easy to calculate, because of clustering and the involvement of (generally complex) second-order inclusion probabilities. In view of these considerations, a new computer program has been developed to provide a method for estimating the parameters of sensitive characteristics under a variety of complex sampling designs.

Keywords

surveys, statistics, randomized response technique

Program Description

RRTCS is a new (R Development Core Team, 2010) package to perform the point and interval estimation of linear parameters using data obtained from RR surveys under complex sampling designs. The package works with a wide range of sampling designs, including stratified sampling, cluster sampling, unequal probabilities sampling, and any combination of these. The package consists of 21 main functions, each of which implements one of the following RR procedures for complex surveys:

¹University of Granada, Spain

Corresponding Author:

María del Mar Rueda, Facultad de Ciencias, Avda, Fuentenueva, Granada, 18071, Spain.
Email: mrueda@ugr.es

- *RR procedures to estimate parameters of a qualitative sensitive characteristic:* Christofides model (Christofides, 2003); Devore model (Devore, 1977); forced response model (Boruch, 1972); Horvitz model (Greenberg, Abul-Ela, Simmons, & Horvitz, 1969); Horvitz model with unknown B (Chaudhuri, 2011, p. 42); Kuk model (Kuk, 1990); Mangat model (Mangat, 1992); Mangat model with unknown B (Chaudhuri, 2011, p. 53); Mangat and Singh model (Mangat & Singh, 1990); Mangat, Singh, and Singh model (Mangat, Singh, & Singh, 1992); Mangat, Singh, and Singh model with unknown B (Chaudhuri, 2011, p. 54); Singh and Joarder model (Singh & Joarder, 1997); Soberanis-Cruz model (Soberanis-Cruz, Ramírez Valverde, Pérez Elizalde, & González Cossio, 2008); and Warner model (Warner, 1965).
- *RR procedures to estimate parameters of a quantitative sensitive characteristic:* Bar-Lev model (Bar-Lev, Bobovitch, & Boukai, 2004), Chaudhuri and Christofides model (Chaudhuri & Christofides, 2013, p. 97), Diana and Perri 1 model (Diana & Perri, 2010, p. 1877), Diana and Perri 2 model (Diana & Perri, 2010, p. 1879), Eichhorn and Hayre model (Eichhorn & Hayre, 1983), Eriksson model (Eriksson, 1973), and Saha model (Saha, 2007).

The package includes an additional function that provides variance estimates of the RR estimators using resampling methods (Wolter, 2007) under stratified, cluster, and unequal probabilities sampling. These include the jackknife method (Quenouille, 1949), the Escobar-Berger method (Escobar & Berger, 2013), and the Campbell-Berger-Skinner method (Berger & Skinner, 2005). Finally, the package includes 20 data sets with observations from different surveys conducted in real and simulated populations using different RR techniques.

Availability, Documentation, and Distribution

The *RRTCS* package is available free of charge from the website <http://www.r-project.org> and works under Windows, Linux, and MacOS. A reference manual (in PDF format) is also available from the same website. Some examples to illustrate its use in real surveys related to sensitive issues are included in this manual on CRAN. The current version of *RRTCS* is 0.0.1. Version 3.1.3 (or later) of the R software should be installed for optimal functioning of *RRTCS*.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The development of the *RRTCS* package was partly funded by Ministerio de Educación, Cultura y Deporte (Grant MTM2012-35650 and FPU grant program, Spain) and by Consejería de Economía, Innovación, Ciencia y Empleo (Grant SEJ2954, Junta de Andalucía, Spain).

References

- Bar-Lev, S. K., Bobovitch, E., & Boukai, B. (2004). A note on randomized response models for quantitative data. *Metrika*, 60, 255-260.
- Berger, Y. G., & Skinner, C. J. (2005). A jackknife variance estimator for unequal probability sampling. *Journal of the Royal Statistical Society: Series B*, 67, 79-89.

- Boruch, R. F. (1972). Relations among statistical methods for assuring confidentiality of social research data. *Social Science Research*, 1, 403-414.
- Chaudhuri, A. (2011). *Randomized response and indirect questioning techniques in surveys*. London, England: Chapman & Hall.
- Chaudhuri, A., & Christofides, T. C. (2013). *Indirect questioning in sample surveys*. Berlin, Germany: Springer-Verlag.
- Christofides, T. C. (2003). A generalized randomized response technique. *Metrika*, 57, 195-200.
- Devore, J. L. (1977). A note on the randomized response technique. *Communications in Statistics-Theory and Methods*, 6, 1525-1529.
- Diana, G., & Perri, P. F. (2010). A new scrambled response models for estimating the mean of a sensitive quantitative character. *Journal of Applied Statistics*, 37, 1875-1890.
- Eichhorn, B., & Hayre, L. S. (1983). Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning and Inference*, 7, 307-316.
- Eriksson, S. A. (1973). A new model for randomized response. *International Statistical Review*, 41, 40-43.
- Escobar, E. L., & Berger, Y. G. (2013). A new replicate variance estimator for unequal probability sampling without replacement. *Canadian Journal of Statistics*, 41, 508-524.
- Greenberg, B. G., Abul-El, A. L., Simmons, W. R., & Horvitz, D. G. (1969). The unrelated question RR model: Theoretical framework. *Journal of the American Statistical Association*, 64, 520-539.
- Kuk, A. Y. C. (1990). Asking sensitive questions indirectly. *Biometrika*, 77, 436-438.
- Mangat, N.S. (1992) Two stage randomized response sampling procedure using unrelated question. *Journal of the Indian Society of Agricultural Statistics*, 44, 82-87.
- Mangat, N. S., & Singh, R. (1990). An alternative randomized response procedure. *Biometrika*, 77, 439-442.
- Mangat, N. S., Singh, R., & Singh, S. (1992). An improved unrelated question randomized response strategy. *Calcutta Statistical Association Bulletin*, 42, 277-281.
- Quenouille, M. H. (1949). Problems in plane sampling. *The Annals of Mathematical Statistics*, 20, 355-375.
- R Development Core Team. (2010). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from <http://www.R-project.org>
- Saha, A. (2007). A simple randomized response technique in complex surveys. *Metron*, LXV, 59-66.
- Singh, S., & Joarder, A. H. (1997). Unknown repeated trials in randomized response sampling. *Journal of the Indian Statistical Association*, 30, 109-122.
- Soberanis-Cruz, V., Ramírez Valverde, G., Pérez Elizalde, S., & González Cossio, F. (2008). Randomized response sampling in finite populations: A unifying approach. *Agrociencia*, 42, 537-549.
- Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60, 63-69.
- Wolter, K. M. (2007). *Introduction to variance estimation* (2nd ed.). New York, NY: Springer.