

## Applied Psychological Measurement

### **RRTCS: An R package for Randomized Response Techniques in Complex Surveys**

|                  |  |
|------------------|--|
| Journal:         | <i>Applied Psychological Measurement</i> |
| Manuscript ID:   | APM-15-06-083.R2                         |
| Manuscript Type: | Computer Program Exchange                |
| Keywords:        | Computer program, Surveys, Statistics    |
|                  |  |

SCHOLARONE™  
Manuscripts

## RRTCS: An R package for Randomized Response Techniques in Complex Surveys

María del Mar Rueda<sup>1</sup>, Beatriz Cobo<sup>1</sup>  
and Antonio Arcos<sup>1</sup>

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 (for example, concerning racism, drug use, abortion, delinquency, AIDS or academic cheating). Standard RR methods are used primarily in surveys which 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.

### Program Description

*RRTCS* is a new R 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:

- *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 et al. 1969), Horvitz model with unknown B (Chaudhuri 2011, page 42), Kuk model (Kuk 1990), Mangat model (Mangat 1992), Mangat model with unknown B (Chaudhuri 2011, page 53), Mangat and Singh model (Mangat and Singh 1990), Mangat, Singh and Singh model (Mangat, Singh and Singh 1992), Mangat, Singh and Singh model with unknown B (Chaudhuri 2011, page 54), Singh and Joarder model (Singh and Joarder 1997), Soberanis-Cruz model (Soberanis-Cruz et al. 2008) and Warner model (Warner 1965).

- *RR procedures to estimate parameters of a quantitative sensitive characteristic*: Bar-Lev model (Bar-Lev et al. 2004), Chaudhuri and Christofides model (Chaudhuri and Christofides 2013, page 97), Diana and Perri 1 model (Diana and Perri 2010, page 1877), Diana and Perri 2 model (Diana and Perri 2010, page 1879), Eichhorn and Hayre model (Eichhorn and Hayre 1983), Eriksson model (Eriksson 1973) and Saha model (Saha 2007).

The package includes an additional function which provides variance estimates of the RR estimators using resampling methods (Wolter 2007) under stratified, cluster and unequal

<sup>1</sup> University of Granada (Spain)

### Corresponding Author:

María del Mar Rueda, University of Granada (Spain).  
E-mail: [mrueda@ugr.es](mailto:mrueda@ugr.es)

probabilities sampling. These include the jackknife method (Quenouille 1949), the Escobar-Berger method (Escobar and Berger 2013) and the Campbell-Berger-Skinner method (Berger and Skinner 2005). Finally, the package includes 20 datasets 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) declare they have no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

### Funding

The author(s) disclose 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 programme, 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* 378 (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*. Chapman & Hall, London.
- Chaudhuri, A., & Christofides, T. C. (2013). *Indirect Questioning in Sample Surveys*. Springer-Verlag, Berlin, Heidelberg.
- 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(3), 508-524.
- Greenberg, B. G., Abul-Ela, 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.

- 1  
2  
3 Kuk, A. Y. C. (1990). Asking sensitive questions indirectly. *Biometrika* 77, 436-438.  
4 Mangat, N. S., & Singh, R. (1990). An alternative randomized response procedure. *Biometrika*  
5 77, 439-442.  
6 Mangat, N. S., Singh, R., & Singh, S. (1992). An improved unrelated question randomized  
7 response strategy. *Calcutta Statistical Association Bulletin* 42, 277-281.  
8 Quenouille, M. H. (1949). Problems in Plane Sampling. *The Annals of Mathematical Statistics*  
9 20, 355-375.  
10 R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*.  
11 Vienna, Austria: R Foundation for Statistical Computing, Available from [http://www.R-](http://www.R-project.org)  
12 [project.org](http://www.R-project.org)  
13 Saha, A. (2007). A simple randomized response technique in complex surveys. *Metron* LXV, 59-  
14 66.  
15 Singh, S., & Joarder, A. H. (1997). Unknown repeated trials in randomized response sampling.  
16 *Journal of the Indian Statistical Association* 30, 109-122.  
17 Soberanis-Cruz, V., Ramírez Valverde, G., Pérez Elizalde, S., & González Cossio, F. (2008).  
18 Muestreo de respuestas aleatorizadas en poblaciones finitas: Un enfoque unificador.  
19 *Agrociencia* 42(5), 537-549.  
20 Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer  
21 bias. *Journal of the American Statistical Association* 60, 63-69.  
22 Wolter, K. M. (2007). *Introduction to Variance Estimation*. 2nd Edition, Springer.  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## APM: Final Manuscript Submission Checklist

*Please note that all materials should, if possible, be submitted online, as a revision of your manuscript. This allows the manuscript's final files to proceed directly, via the Manuscript Central System, to Production after the Editor has completed his final check. It is not necessary any longer to send electronic files to us via e-mail.*

- ( ) Author Contact Information for all authors if possible, and for the corresponding author at the absolute minimum, containing up-to-date contact information for each author, including:

- Institutional affiliation, with desired affiliation to be printed with the manuscript if different from current affiliation
- Work address
- Work phone number
- Work fax number
- Work e-mail address
- Home address\*
- Home phone number\*
- Other e-mail address (if any)\*

Corresponding author: Maria del Mar Rueda Garcia, Faculty of Science. Department of Statistics and O. R. University of Granada. Spain. Work phone number: +34958240494. Work fax number: +34958243267. Work e-mail address: mrueda@ugr.es. Home address: Luis del Mármol Carvajal, 2 3ºB, 18001, Granada, Spain. Home phone number: +34629199659

Beatriz Cobo Rodriguez, Faculty of Science. Department of Statistics and O. R. University of Granada. Spain. Work fax number: +34958243267. Work e-mail address: beacr@ugr.es.

Antonio Arcos Cebrian, Faculty of Science. Department of Statistics and O. R. University of Granada. Spain. Work fax number: +34958243267. Work e-mail address: arcos@ugr.es.

- ( ) The final version of the manuscript (preferably in MS Word format; if submitting in LaTeX/TeX format, please follow the LaTeX/TeX submission instructions attached separately).

Submitted online

- ( ) Electronic files for all figures (camera-ready; see attachment for information on preparing camera-ready figures: "SAGE Artwork Submission Guidelines.")

No figures

- ( ) Abstract (should be part of the manuscript itself)

It is a part of the manuscript

- ( ) Keywords (should be part of the manuscript itself)

No keywords

- ( ) Tables (should be part of the manuscript itself). Also please make sure all the tables and equations are in their raw formats (word, excel, LaTeX, etc.), not as images. We cannot accept

- 1  
2  
3 tables or equations in image format because we are unable to edit them and they can become  
4 disfigured.  
5  
6 No tables  
7  
8 ( ) References (should be part of the manuscript itself)  
9 It is a part of the manuscript  
10  
11 ( ) Acknowledgements, if any (should be part of the manuscript itself)  
12 It is a part of the manuscript  
13  
14 ( ) Appendices, if any (If the appendices are accepted as a part of the main article, they must be  
15 submitted as a part of the Main Document; if they are accepted as online supplements, they  
16 must be submitted as a separate Supplementary File.)  
17  
18 No appendix  
19  
20 ( ) Permission to use any reproduced or copyrighted material (if needed)  
21 It is not needed  
22  
23 ( ) This checklist, completed.  
24  
25 Online as supplementary file  
26  
27

28 \* Needed to facilitate communications in case of a change in institutional affiliation. Will not be used  
29 unless contact cannot be made using work information. (Since Sage Publishers is working during  
30 vacations, weekends, etc., they like to have alternative contact information available in case something  
31 urgent arises and you are not at work.)  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60