The Journal of Open Source Software

splithalf: robust estimates of split half reliability

Sam Parsons[1]

1 Department of Experimental Psychology, University of Oxford

DOI: 10.21105/joss.03041

Software

- Review
- Repository
- Archive

Editor: Olivia Guest

Reviewers:

- @jpsnijder
- @Nathaniel-Haines
- @rMassimiliano

Submitted: 11 January 2021
Published: 23 April 2021

splithalf: robust estimates of split half reliability

The R package **splithalf** provides tools to estimate the internal consistency reliability of cognitive measures. In particular, the tools were developed for application to tasks that use difference scores as the main outcome measure, for instance the Stroop score or dot-probe attention bias index (average RT in incongruent trials minus average RT in congruent trials).

The methods in **splithalf** are built around split half reliability estimation. To increase the robustness of these estimates, the package implements a permutation approach that takes a large number of random (without replacement) split halves of the data. For each permutation the correlation between halves is calculated, with the Spearman-Brown correction applied (Spearman, 1904). This process generates a distribution of reliability estimates from which we can extract and plot summary statistics (e.g. average and 95% HDI).

Statement of need

While many cognitive tasks yield highly robust effects, the same task may not yield reliable individual differences (Hedge et al., 2018). As these measures are used in questions of individual differences researchers need to have some psychometric information for the outcome measures. Recently, it was proposed that psychological science should set a standard expectation for the reporting of reliability information for cognitive and behavioural measures (Parsons et al., 2019). **splithalf** was developed to support this proposition by providing a tool to easily extract internal consistency reliability estimates from behavioural measures.

Usage

The *splithalf* function estimates reliability for a variety of scoring types, including average scores, difference scores, and difference-of-difference scores. The function is also extendable to other scoring calculations, e.g. signal detection. Reliability can be estimated for response time and accuracy rate outcomes. A plotting option is also provided for users to examine the distribution of reliability estimates.

Users can perform a reliability multiverse analysis to examine the stability of reliability estimates, given a rage of data processing decisions (Parsons, 2020). By passing a list of data processing decisions (e.g. maximum and minimum RT cutoffs) and a *splithalf* object into the *splithalf.multiverse* function. *testretest.multiverse* can be used to explore a similar multiverse of data processing decisions on test retest reliability. Multiverse output can be plotted with *plot_multiverse*.

splithalf requires the **tidyr** (Wickham & Henry, 2019) and **dplyr** (Wickham et al., 2018) packages for data handling within the functions. The **robustbase** package is used to extract median scores when applicable. The computationally heavy tasks (extracting many random half samples of the data) are written in c++ via the R package **Rcpp** (Eddelbuettel et al.,

2018). Test retest multiverse analyses require the **psych** package (Revelle, 2019). Figures use the **ggplot** package (Wickham, 2016), raincloud plots use code adapted from Allen et al. (Allen et al., 2019), and the **patchwork** package (Pedersen, 2019) is used for plotting the multiverse analyses.

Comparison to other software

splithalf is the only package to implement all of these tools, in particular reliability multiverse analyses. Some other R packages offer a bootstrapped approach to split-half reliability: **multicon** (Sherman, 2015), **psych** (Revelle, 2019), and **splithalfr** (Pronk, 2020).

Acknowledgements

This work was supported by a grant from the Economic and Social Research Council (ESRC; Grant reference ES/R004285/1). I would also like to thank Anne-Wil Kruijt, James Bartlett, Andrew Jones, and Jean-Paul Snijder for their input on the package. I also thank all that have asked questions, posed suggestions, and shared bug reports - the **splithalf** package is more stable, and more powerful, because of them.

References

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*, 63. https://doi.org/10.12688/wellcomeopenres.15191.1

Eddelbuettel, D., Francois, R., Allaire, J., Ushey, K., Kou, Q., Russell, N., Bates, D., & Chambers, J. (2018). *Rcpp: Seamless r and c++ integration*. https://CRAN.R-project.org/package=Rcpp

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Parsons, S. (2020). *Exploring reliability heterogeneity with multiverse analyses: Data processing decisions unpredictably influence measurement reliability* [Preprint]. https://doi.org/10.31234/osf.io/y6tcz

Parsons, S., Kruijt, A.-W., & Fox, E. (2019). *Psychological science needs a standard practice of reporting the reliability of cognitive behavioural measurements*. Advances in Methods; Practices in Psychological Science. https://doi.org/doi.org/10.1177/2515245919879695

Pedersen, T. L. (2019). *Patchwork: The composer of plots*. https://CRAN.R-project.org/package=patchwork

Pronk, T. (2020). *Splithalfr: Extensible bootstrapped split-half reliabilities*. https://CRAN.R-project.org/package=splithalfr

Revelle, W. (2019). *Psych: Procedures for psychological, psychometric, and personality research*. Northwestern University. https://CRAN.R-project.org/package=psych

Sherman, R. A. (2015). *Multicon: Multivariate constructs*. https://CRAN.R-project.org/package=multicon

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72. https://doi.org/10.2307/1412159

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4

Wickham, H., François, R., Henry, L., & Müller, K. (2018). *Dplyr: A grammar of data manipulation*. https://CRAN.R-project.org/package=dplyr

Wickham, H., & Henry, L. (2019). *Tidyr: Tidy messy data*. https://CRAN.R-project.org/package=tidyr