The Journal of Open Source Software

rmcorrShiny: A web application for repeated measures correlation

Laura R. Marusich[1] **and Jonathan Z. Bakdash**[2, 3]

1 U.S. Army Combat Capabilities Development Command Army Research Laboratory South at the University of Texas at Arlington **2** U.S. Army Combat Capabilities Development Command Army Research Laboratory South at the University of Texas at Dallas **3** Department of Psychology and Special Education, Texas A&M–Commerce

DOI:

Software

- Review
- Repository
- Archive

Submitted:
Published:

Summary

The most common techniques for calculating the correlation (e.g., Pearson correlation) between two variables are based on the assumption that each data point of paired measures represents an independent observation. Take, for example, a study that calculates the correlation between a person's age and the volume of a region of the brain. In this example, each individual contributes a data point consisting of a brain volume and an age. However, it is not uncommon for studies to use repeated measures designs, such as a study that collected the brain region volume and age at two different time points (Raz et al., 2005). Each participant in this study contributed two (repeated) data points of paired measures. Repeated measures of the same individual are no longer independent observations and should not be analyzed as such. Erroneously modeling repeated measures data as independent observations is surprisingly prevalent in published research, even though such results will generally be misleading (Aarts, Verhage, Veenvliet, Dolan, & Van Der Sluis, 2014; Bakdash et al., 2020; Lazic, 2010). A common solution to this problem is to use aggregated data: first taking an average of the repeated measures data of each person so that each person again contributes a single data point, and then calculating the correlation from these averages.

Instead of aggregation, an alternative solution is to calculate the repeated measures correlation (Bakdash & Marusich, 2017; Bland & Altman, 1995a, 1995b), which assesses the common intra-individual (within-participants) association for paired repeated measures data. The repeated measures correlation technique is conceptually similar to a null multilevel model, with a common slope but varying intercept for each individual. Calculating the repeated measures correlation has multiple potential benefits. It is simpler and more straightforward to implement than a multilevel model, with the potential for far greater statistical power than aggregation. It also has the potential to provide insights into patterns among individuals that aggregation may obscure (Bakdash & Marusich, 2017).

We previously developed the `rmcorr` R package (Bakdash & Marusich, 2020) to make the repeated measures correlation technique widely available for researchers; it has since also been adapted as a function in the Pingouin statistics package (Vallat, 2018) for Python. However, the use of both of these packages requires some facility with programming languages and thus they are not universally accessible.

Here we introduce the `rmcorrShiny` application, which provides an intuitive graphical interface for computing and plotting the repeated measures correlation (see Figure 1 below demonstrating the application using brain area volume and age data from Raz (2005)).

The primary features of `rmcorrShiny` include:

- The ability to import data in different file formats or use one of four included sample datasets.
- The display of raw output from `rmcorr` as well as formatted output for reporting results.
- Multiple options to generate and customize `rmcorr` plots (making use of the `ggplot2` package (Wickham, 2016; Wickham et al., 2020) and palettes from the `RColorBrewer` (Neuwirth, 2014) and `pals` (Wright, 2019) packages).
- Customized R code using the data and options chosen by the user that can be directly pasted and executed in R to produce the same output as in `rmcorrShiny`.
- The ability to download plots (in a variety of file formats) or a .zip file of all output.

Note that many features in ''rmcorrShiny" were based on modifications of code in the `Raincloud-shiny` app ("Raincloud-shiny," 2021) (Laura: I like the new language, but what do you think about the citation? I found his name through Twitter, but maybe it's kind of creepy to do that? He doesn't have his name on the app or on the github)

Screenshot (placeholder until we have a near-final version?)

`rmcorrShiny` can be used in a web browser **here** or it can be installed from Github and run in R, using the following commands (BIG TODO):

```
devtools::install_github("lmarusich/rmcorrShiny")
library(rmcorrShiny)
rmcorrShiny::rmcorrShiny_run()
```

Acknowledgements

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Combat Capabilities Development Command Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation.

We acknowledge contributions from....

References

Aarts, E., Verhage, M., Veenvliet, J. V., Dolan, C. V., & Van Der Sluis, S. (2014). A solution to dependency: Using multilevel analysis to accommodate nested data. *Nature Neuroscience*, *17*(4), 491–496. doi:10.1038/nn.3648

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*, 456. doi:10.3389/fpsyg.2017.00456

Bakdash, J. Z., & Marusich, L. R. (2020). *Rmcorr: Repeated measures correlation*. Retrieved from https://CRAN.R-project.org/package=rmcorr

Bakdash, J. Z., Marusich, L. R., Kenworthy, J. B., Twedt, E., & Zaroukian, E. G. (2020). Statistical significance filtering overestimates effects and impedes falsification: A critique of endsley (2019). *Frontiers in Psychology*, *11*, 3669. doi:10.3389/fpsyg.2020.609647

Bland, J. M., & Altman, D. G. (1995a). Statistics notes: Calculating correlation coefficients with repeated observations: Part 1correlation within subjects. *BMJ*, *310*(6977), 446. doi:10.1136/bmj.310.6977.446

Bland, J. M., & Altman, D. G. (1995b). Calculating correlation coefficients with repeated observations: Part 2correlation between subjects. *BMJ*, *310*(6980), 633. doi:10.1136/bmj.310.6980.633

Lazic, S. E. (2010). The problem of pseudoreplication in neuroscientific studies: Is it affecting your analysis? *BMC neuroscience*, *11*(1), 1–17. doi:10.1186/1471-2202-11-5

Neuwirth, E. (2014). *RColorBrewer: ColorBrewer palettes*. Retrieved from https://CRAN.R-project.org/package=RColorBrewer

Raincloud-shiny. (2021). Retrieved from https://gabrifc.shinyapps.io/raincloudplots/

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., Dahle, C., et al. (2005). Regional brain changes in aging healthy adults: General trends, individual differences and modifiers. *Cerebral Cortex*, *15*(11), 1676–1689. doi:10.1093/cercor/bhi044

Vallat, R. (2018). Pingouin: Statistics in python. *Journal of Open Source Software*, *3*(31), 1026. doi:10.21105/joss.01026

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., et al. (2020). *Ggplot2: Create elegant data visualisations using the grammar of graphics*. Retrieved from https://CRAN.R-project.org/package=ggplot2

Wright, K. (2019). *Pals: Color palettes, colormaps, and tools to evaluate them*. Retrieved from http://kwstat.github.io/pals/