The Journal of Open Source Software

SimplyAgree: An R package and jamovi Module for Simplifying Agreement and Reliability Analyses

Aaron R. Caldwell[1,2]

1 United States Army Research Institute of Environmental Medicine, Natick, MA, the United States of America **2** Oak Ridge Institute of Science and Education, Oak Ridge, TN, the United States of America

DOI: 10.21105/joss.04148

Software

- Review
- Repository
- Archive

Editor: Frederick Boehm

Reviewers:

- @slwu89
- @itismeghasyam

Submitted: 01 February 2022
Published: 22 March 2022

Summary

Accurate and reliable measurements are critical to quantitative research efforts. Based on citation counts of Bland & Altman (1986) and Weir (2005) alone[1], researchers appear to highly value methods to quantify the accuracy and reliability of measurement tools used for research. This article introduces the `SimplyAgree` R package and jamovi module as user-friendly solutions for estimating agreement and reliability (R Core Team, 2020; The jamovi project, 2021) for continuous measurements. Updates and additional details on `SimplyAgree` can be found on the package's website. `SimplyAgree` was created for applied physiologists to use when evaluating different physiological measurements (e.g, comparing the measurement of oxygen consumption between two competing gas analyzers). However, the functions within `SimplyAgree` can be utilized by any researcher wanting to evaluate agreement between two continuous measurements, or for evaluating the reliability of a continuous measure[2].

Statement of Need

A number of new methods have been developed in the past three decades to improve the calculation of the limits of agreement (Lin, 1989; Shieh, 2019; Zou, 2011) and other measures of measurement reliability (Carrasco, Phillips, Puig-Martinez, King, & Chinchilli, 2013; Weir, 2005). However, to the author's best knowledge, statistical software — particularly open source software — to implement these statistical analyses are lacking. While some software may provide the limits of agreement analysis outlined by Bland & Altman (1986, 1999), few, if any, account for multiple observations within the same research subject (Zou, 2011) or include hypothesis tests of agreement (Shieh, 2019). `blandr` (Datta, 2017) and `BlandAltmanLeh` (Lehnert, 2015) exist to aid the creation of Bland-Altman type plots, but do not offer the tests detailed by Shieh (2019) or Zou (2011). `MethComp` (Carstensen, Gurrin, Ekstrøm, & Figurski, 2020) is the most comprehensive with regards to limits of agreement but also lacks the methods outlined by Shieh (2019). The `cccrm` package (Carrasco et al., 2013) does already provide concordance correlation coefficients as output, and is the basis of the concordance calculations within `SimplyAgree`. Lastly, `psych` (Revelle, 2020) and `ICC` (Wolak, Fairbairn, & Paulsen, 2012) provide some measures of reliability but lack the ability to calculate standard error of the measurement. Moreover, only `blandr` has a jamovi module but the capabilities are limited to just the traditional Bland-Altman limits.

`SimplyAgree` is also created to be easy to use for those with limited programming experience. The output from the functions is intentionally verbose and typically provides more than 1

---

[1] Bland & Altman (1986) and Weir (2005) have greater than 50000 and 4000 citations, respectively, at the time of writing this manuscript. Many of these publications are original research investigations.

[2] Discrete and categorical measurements are not supported at the time of writing this manuscript

measure of agreement or reliability. The hope is that researchers who use these functions will not limit their interpretations of the results to 1 numeric value, but instead look at the totality of the results (estimates and confidence intervals) for their interpretation.

Additionally, many researchers may not have the skills necessary to write the code, from scratch, in order to implement many of the newest techniques. The jamovi project (2021) is a open source statistical platform that provides a graphical user interface (GUI), and therefore is an accessible source for researchers, or even students, without coding experience. Therefore, a jamovi module of `SimplyAgree` was also created in order to reach those researchers who may not have the coding expertise required to effectively use the R package.

Current R Capabilities

The R package `SimplyAgree`, currently v0.0.2 on the comprehensive R archive network (CRAN), implements a number of useful agreement and reliability analyses.

The current release of the R package can be downloaded directly from CRAN in R:

```
install.packages(”SimplyAgree”)
```

Or, the developmental version, can be downloaded from GitHub:

```
devtools::install_github(”arcaldwell49/SimplyAgree”)
```

There are 2 vignettes that document the major functions within the package that can be found on the package's website (https://aaroncaldwell.us/SimplyAgree). Overall, there are 6 fundamental functions, all with generic `plot` and `print` methods, within the R package:

1. `agree_test`: Simple Test of Agreement. This is function performs agreement analyses on two vectors of the same length, and is designed for analyses that were described by Bland & Altman (1986, 1999). In addition to providing the traditional Bland-Altman limits of agreement, the function provides a hypothesis test (Shieh, 2019), and provides the concordance correlation coefficient (Lin, 1989).
2. `agree_reps`: Test of Agreement for Replicate Data. This function provides the limits of agreement described by Zou (2011) for data where the mean, per subject, does not vary. In addition, the concordance correlation coefficient, calculated by U-statistics, is also provided in the output (Carrasco et al., 2013).
3. `agree_nest`: Test of Agreement for Nested Data. This function provides the limits of agreement described by Zou (2011) for data where the mean, per subject, *does* vary. Similar to the replicate data function, the concordance correlation coefficient, calculated by U-statistics, is provided in the output (Carrasco et al., 2013).
4. `loa_mixed`: Bootstrapped Limits of Agreement for Nested Data. This function calculates limits of agreement using a non-parametric bootstrap method, and can allow the underlying mean to vary (replicate data) or not (nested data).
5. `blandPowerCurve`: Power Analysis for Bland-Altman Limits of Agreement. This function implements the formula outlined by Lu et al. (2016). This allows for power calculations for the Bland & Altman (1999) limits of agreement. The function `find_n` can then be used to find the sample size at which adequate power (defined by the user) is achieved.
6. `reli_stats`: Reliability Statistics. This function calculates and provides as output the statistics outlined by Weir (2005). This includes an array of intraclass correlation coefficients, the coefficient of variation, and the standard error of measurement.

Current jamovi Capabilities

The jamovi module can be added to the jamovi directly from the "add module" tab in the GUI.

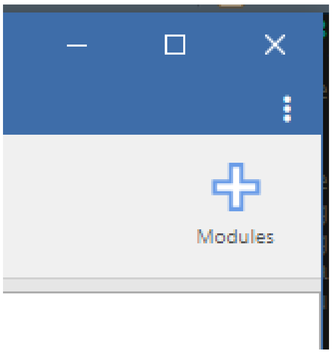


Figure 1: How to add a module in jamovi.

The `SimplyAgree` module is then available on the main menu, and within it there are three analysis options.

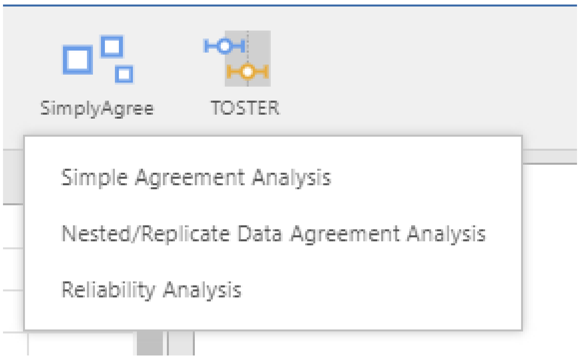
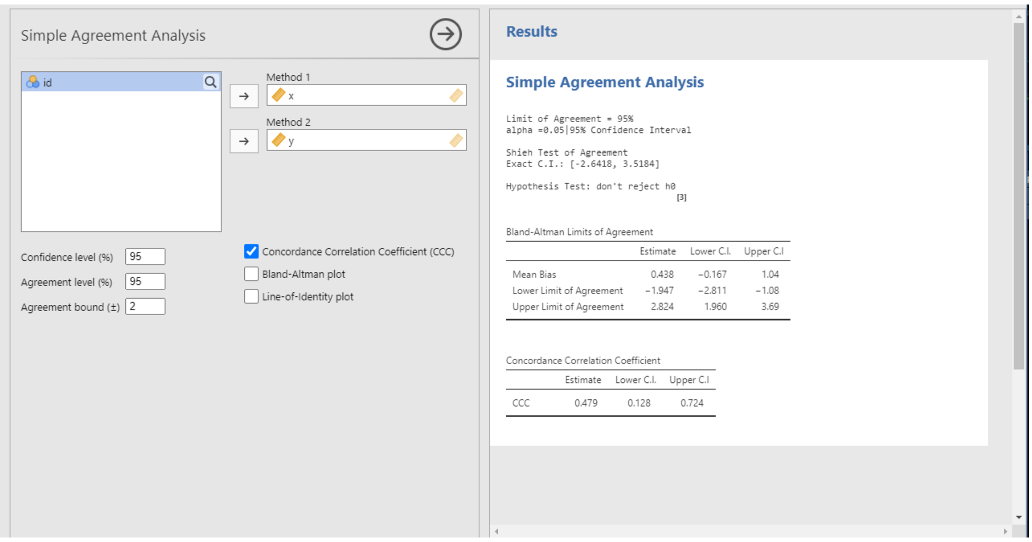


Figure 2: SimplyAgree in jamovi.

The three analysis options essentially enable jamovi users to complete some of the same analyses available in the R package.

1. The simple agreement analysis incorporates the `agree_test` function. Users have the option of including the concordance correlation coefficient, and plots of the data.



Confidence level (%) 95

Agreement level (%) 95

Agreement bound (±) 2

☑ Concordance Correlation Coefficient (CCC)
☐ Bland-Altman plot
☐ Line-of-Identity plot

Simple Agreement Analysis

```
Limit of Agreement = 95%
alpha =0.05|95% Confidence Interval

Shieh Test of Agreement
Exact C.I.: [-2.6418, 3.5184]

Hypothesis Test: don't reject h0
                                  [3]
```

Bland-Altman Limits of Agreement

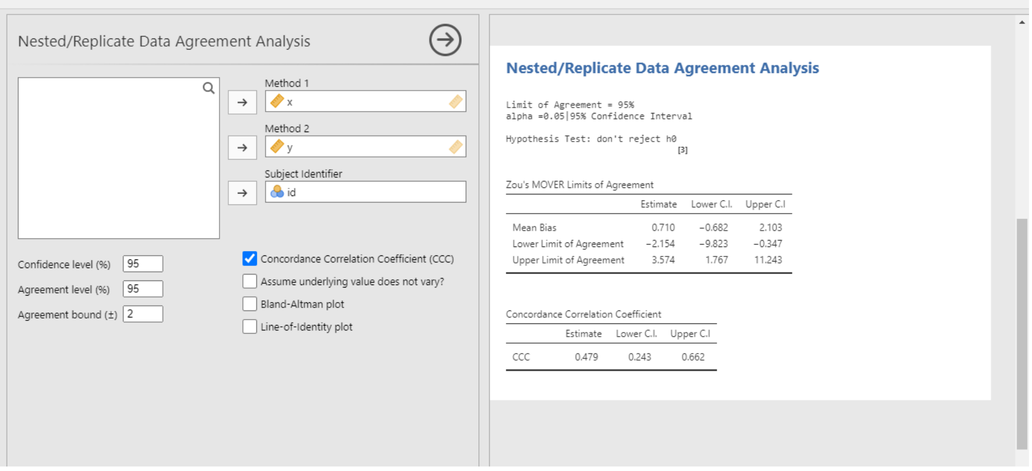
	Estimate	Lower C.I.	Upper C.I
Mean Bias	0.438	−0.167	1.04
Lower Limit of Agreement	−1.947	−2.811	−1.08
Upper Limit of Agreement	2.824	1.960	3.69

Concordance Correlation Coefficient

	Estimate	Lower C.I.	Upper C.I
CCC	0.479	0.128	0.724

Figure 3: Sample Output from the Simple Agreement Analysis.

2. The nested/replicate agreement analysis uses the `agree_nest` and `agree_reps` function to perform the analyses. The `agree_reps` function is used if "Assume underlying value does not vary?" is selected; otherwise `agree_nest` is used.



Confidence level (%) 95

Agreement level (%) 95

Agreement bound (±) 2

☑ Concordance Correlation Coefficient (CCC)
☐ Assume underlying value does not vary?
☐ Bland-Altman plot
☐ Line-of-Identity plot

Nested/Replicate Data Agreement Analysis

```
Limit of Agreement = 95%
alpha =0.05|95% Confidence Interval

Hypothesis Test: don't reject h0
                                  [3]
```

Zou's MOVER Limits of Agreement

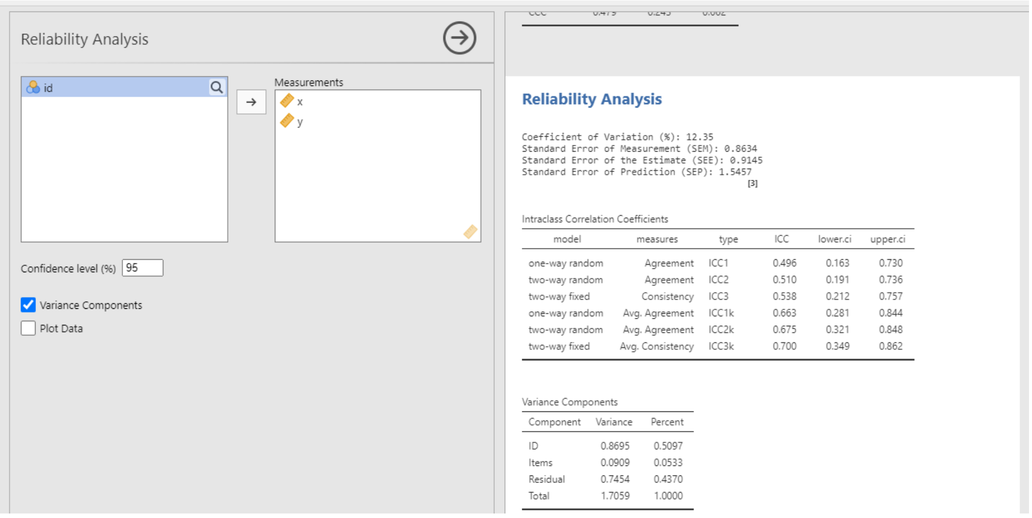
	Estimate	Lower C.I.	Upper C.I
Mean Bias	0.710	−0.682	2.103
Lower Limit of Agreement	−2.154	−9.823	−0.347
Upper Limit of Agreement	3.574	1.767	11.243

Concordance Correlation Coefficient

	Estimate	Lower C.I.	Upper C.I
CCC	0.479	0.243	0.662

Figure 4: Sample Output from the Nested/Replicate Agreement Analysis.

3. The reliability analysis utilizes `reli_stats` to calculate reliability statistics.



Confidence level (%) 95

☑ Variance Components

☐ Plot Data

Reliability Analysis

```
Coefficient of Variation (%): 12.35
Standard Error of Measurement (SEM): 0.8634
Standard Error of the Estimate (SEE): 0.9145
Standard Error of Prediction (SEP): 1.5457
                                        [3]
```

Intraclass Correlation Coefficients

model	measures	type	ICC	lower.ci	upper.ci
one-way random	Agreement	ICC1	0.496	0.163	0.730
two-way random	Agreement	ICC2	0.510	0.191	0.736
two-way fixed	Consistency	ICC3	0.538	0.212	0.757
one-way random	Avg. Agreement	ICC1k	0.663	0.281	0.844
two-way random	Avg. Agreement	ICC2k	0.675	0.321	0.848
two-way fixed	Avg. Consistency	ICC3k	0.700	0.349	0.862

Variance Components

Component	Variance	Percent
ID	0.8695	0.5097
Items	0.0909	0.0533
Residual	0.7454	0.4370
Total	1.7059	1.0000

Figure 5: Sample Output from the Reliability Analsyis.

Acknowledgements

I would like the thank Ashley Akerman for his kind feedback during the development of `SimplyAgree`.

The opinions or assertions contained herein are the private views of the author and are not to be construed as official or reflecting the views of the Army or the Department of Defense. Any citations of commercial organizations and trade names in this report do not constitute an official Department of the Army endorsement of approval of the products or services of these organizations. No authors have any conflicts of interest to disclose. Approved for public release; distribution is unlimited.

References

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, *327*(8476), 307–310. doi:10.1016/s0140-6736(86)90837-8

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*(2), 135–160. doi:10.1177/096228029900800204

Carrasco, J. L., Phillips, B. R., Puig-Martinez, J., King, T. S., & Chinchilli, V. M. (2013). Estimation of the concordance correlation coefficient for repeated measures using SAS and R. *Computer Methods and Programs in Biomedicine*, *109*(3), 293–304. doi:10.1016/j.cmpb.2012.09.002

Carstensen, B., Gurrin, L., Ekstrøm, C. T., & Figurski, M. (2020). *MethComp: Analysis of agreement in method comparison studies*. Retrieved from https://CRAN.R-project.org/package=MethComp

Datta, D. (2017). *Blandr: A Bland-Altman method comparison package for R*. doi:10.5281/zenodo.824514

Lehnert, B. (2015). *BlandAltmanLeh: Plots (slightly extended) Bland-Altman plots*. Retrieved from https://CRAN.R-project.org/package=BlandAltmanLeh

Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, *45*(1), 255. doi:10.2307/2532051

Lu, M.-J., Zhong, W.-H., Liu, Y.-X., Miao, H.-Z., Li, Y.-C., & Ji, M.-H. (2016). Sample size for assessing agreement between two methods of measurement by Bland-Altman method. *The International Journal of Biostatistics*, *12*(2). doi:10.1515/ijb-2015-0039

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Revelle, W. (2020). *Psych: Procedures for psychological, psychometric, and personality research*. Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Shieh, G. (2019). Assessing agreement between two methods of quantitative measurements: Exact test procedure and sample size calculation. *Statistics in Biopharmaceutical Research*, *12*(3), 352–359. doi:10.1080/19466315.2019.1677495

The jamovi project. (2021). *jamovi*. Retrieved from https://www.jamovi.org

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength and Conditioning Research*, *19*(1), 231. doi:10.1519/00124278-200502000-00038

Wolak, M. E., Fairbairn, D. J., & Paulsen, Y. R. (2012). Guidelines for estimating repeatability. *Methods in Ecology and Evolution*, *3*(1), 129–137. doi:10.1111/j.2041-210X.2011.00125.x

Zou, G. (2011). Confidence interval estimation for the Bland-Altman limits of agreement with multiple observations per individual. *Statistical Methods in Medical Research*, *22*(6), 630–642. doi:10.1177/0962280211402548