Pingouin: statistics in Python

Raphael Vallat[1]

1 Department of Psychology, University of California, Berkeley.

Summary

Python is currently the fastest growing programming language in the world, thanks to its ease-of-use, fast learning curve and its numerous high quality packages for data science and machine-learning. Surprisingly however, Python is far behind the R programming language when it comes to general statistics and for this reason many scientists still rely heavily on R to perform their statistical analyses.

In this paper, we present `Pingouin`, an open-source Python package aimed at partially filling this gap by providing easy-to-use functions for computing some of the main statistical tests that scientists use on an every day basis. This includes basics functions such as ANOVAs, ANCOVAs, post-hoc tests, non-parametric tests, effect sizes, as well as more advanced functions such as Bayesian T-tests (Rouder, Speckman, Sun, Morey, & Iverson, 2009), repeated measures correlations (Bakdash & Marusich, 2017), robust correlations (Pernet, Wilcox, & Rousselet, 2012) and circular statistics (Berens, 2009), to cite but a few. `Pingouin` is written in Python 3 and is mostly built on top of the Pandas (McKinney, 2010) library, therefore allowing a fluid integration within a data analysis pipeline. `Pingouin` comes with an extensive documentation and API as well as with several Jupyter notebook examples.

References

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Front. Psychol.*, *8*, 456.

Berens, P. (2009). CircStat: A MATLAB toolbox for circular statistics. *J. Stat. Softw.*, *31*(10), 1–21.

McKinney, W. (2010). Data structures for statistical computing in python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th python in science conference* (pp. 51–56).

Pernet, C. R., Wilcox, R., & Rousselet, G. A. (2012). Robust correlation analyses: False positive and power validation using a new open source matlab toolbox. *Front. Psychol.*, *3*, 606.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.*, *16*(2), 225–237.