scikit-posthocs: Pairwise multiple comparison tests in Python

Maksim A. Terpilowski

1 Institute of Evolutionary Physiology and Biochemistry, Saint Petersburg, Russia

Summary

scikit-posthocs is a Python package providing multiple pairwise comparison tests (post hoc). Statisticians, data scientists, and researchers will find it useful in a statistical analysis routine to assess the differences between group levels if a statistically significant result of a parametric or nonparametric analysis of variance (ANOVA) test has been obtained.

Most statistical software packages (such as SPSS and Stata) provide a number of post hoc tests. However, Python ecosystem still lacks implementation of these tests for factorial and block design data. Currently, Student, Wilcoxon, Mann-Whitney tests which are not adapted to multiple pairwise comparisons are offered by SciPy. Statsmodels package includes TukeyHSD test which needs extra actions to be fluently integrated into a data analysis pipeline.

scikit-posthocs package is aimed at filling this gap by providing a number of nonparametric and parametric pairwise comparisons tests as well as outliers detection algorithms implemented in Python.

Currently, the following tests are implemented in this package:

1. Parametric pairwise multiple comparisons tests: Scheffe, Student T, Tamhane T2, and TukeyHSD test.
2. Non-parametric tests for factorial design: Conover, Dunn, Dwass-Steel-Critchlow-Fligner, Mann-Whitney, Nashimoto-Wright (NPM), Nemenyi, van Waerden, and Wilcoxon test.
3. Non-parametric tests for block design: Conover, Durbin and Conover, Miller, Nemenyi, Quade, Siegel test.

scikit-posthocs provides tests for outliers detection: interquartile range (IQR) test, Grubbs test, Tietjen-Moore test, and generalized (extreme Studentized deviate) ESD test. It also has plotting functionality to present the results of pairwise comparisons as a heatmap (significance plot, see figure).

scikit-posthocs is compatible with Python 2 and 3 versions, relies heavily and extends the functionality of statsmodels, SciPy and PMCMRplus packages (Seabold & Perktold, 2010), (Jones, Oliphant, & Peterson, 2001), (Pohlert, 2018). It is also integrated with Pandas (McKinney, 2010) and Numpy (T. E. Oliphant, 2006) for efficient computations and data analysis. The package is fully documented and comes with a Jupyter notebook example.
Figure 1: Significance plot

References


