




# Statmanager-kr: A User-friendly Statistical Package for Python in Pandas

Changseok Lee <sup>1</sup>

<sup>1</sup> DYPHI Research Institute, DYPHI Inc., Daejeon, Republic of Korea

DOI: [10.21105/joss.06642](https://doi.org/10.21105/joss.06642)

## Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

---

Editor: [Teon Brooks](#)  

## Reviewers:

- [@catstats](#)
- [@behinger](#)

Submitted: 15 January 2024

Published: 10 October 2024

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

## Summary

Python is one of the most accessible and adaptable programming languages, utilized in various research domains, including statistics. However, few statistical packages inherit these characteristics of Python, leaving researchers unfamiliar with programming languages dependent on other expensive software. To address this gap, Statmanager-kr has been developed to provide non-programmers with convenient access to statistical functions. Statmanager-kr is designed to be compatible with `Pandas.DataFrame` and enables statistical analyses using a single method with a relatively small number of parameters. With `SciPy` and `statsmodels` ensuring the validity of analyses, Statmanager-kr offers functions for hypothesis testing, comparing between-group and within-group differences, regression, correlations, data visualization, and more.

## Statement of need

Statmanager-kr is a statistical package for Python in Pandas. This package provides methods commonly used for null hypothesis significance testing (NHST), which is of interest to researchers in various fields ([Moon, 2020](#)). It is also possible to test for normality or equivariance using the Shapiro-Wilk, Levene, or Fmax tests.

Most of the statistical software available today is difficult to use, as a previous study reported that one of the challenges students face in statistics courses was “using software” ([Murtonen & Lehtinen, 2003](#)). Although there are basic statistical libraries in Python, such as `SciPy` ([Virtanen et al., 2020](#)) and `statsmodels` ([Seabold & Perktold, 2010](#)), they are quite complex. While some studies require complex and detailed statistical modeling and analysis, there are also many studies that require only a few hypothesis tests. Therefore, the development of an easy-to-use statistical package would be of great benefit to these researchers.

To achieve this, Statmanager-kr has been designed to run analyses with only three lines of code: 1) read data as a `Pandas.DataFrame`, 2) create a `Stat_Manager` object, 3) execute the `progress()` method. Therefore, users can use Statmanager-kr as long as they know Pandas functions to read the data, such as `read_csv()` or `read_excel()`. It also includes functions to visualize the results depending on the analysis method.

## Related Work

Recent advances in the field of statistics have been achieved through the emergence of user-friendly packages, such as Pingouin ([Vallat, 2018](#)). Pingouin is an easy-to-use statistics package that offers a wide range of analytical functions. Like Pingouin, Statmanager-kr is similar in that it aims to be a user-friendly statistics package.

However, Statmanager-kr and Pingouin differ in their target users. Since Statmanager-kr is designed for researchers with limited programming experience, it focuses on keeping the workflow short and concise; therefore, Statmanager-kr was designed to allow users to apply analyses and obtain results by always running a single method, `progress()`, in a similar way. On the other hand, Pingouin was developed for users with a relatively high level of programming knowledge and experience; therefore, in terms of workflow, Pingouin offers more comprehensive and fine-tunable analysis methods and provides more detailed analysis results. Also, Statmanager-kr only works with `Pandas.DataFrame`, while Pingouin has the advantage of being compatible with a wider range of date types.

Another difference is related to post-hoc analysis and visualization. When running an analysis that allows post-hoc analysis, such as ANOVA, in Statmanager-kr, post-hoc analysis like Bonferroni correction can be performed by adding the `posthoc` parameter to the `progress` method. It is possible to visualize the results by using `figure()` as a method chaining in Statmanager-kr. Although Pingouin does not provide the functions to directly visualize the results of an analysis, it does support the generation of graphs that are very useful from a statistical perspective, such as paired plots, shift plots, and circular mean plots. Pingouin also has the advantage of supporting a wider range of post-hoc analyses, although it requires a separate method.

In conclusion, depending on the researcher's programming experience and the purpose of the study, Statmanager-kr and Pingouin can be used differently. Researchers who are familiar with programming may be better suited to use Pingouin as it supports more analysis methods and customization. On the other hand, Statmanager-kr is designed to be used by researchers who are not familiar with programming and coding, but want to get quick results.

## Features

Learn more about the features in the [official documentation](#). The instructions for running each analysis are also described in [the manual section of the official documentation](#). [Korean versions of the official documentation](#) are also available.

## Acknowledgements

Author declares no conflicts of interests.

## References

- Moon, S. M. (2020). *Statistics for the Social Sciences: Moving Toward an Integrated Approach*. Cognella Academic Publishing. ISBN: 978-1516519613
- Murtonen, M., & Lehtinen, E. (2003). Difficulties experienced by education and sociology students in quantitative methods courses. *Studies in Higher Education*, 28(2), 171–185. <https://doi.org/10.1080/0307507032000058064>
- Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. *Proceedings of the 9th Python in Science Conference*, 57(61), 10–25080. <https://doi.org/10.25080/Majora-92bf1922-011>
- Vallat, R. (2018). Pingouin: statistics in Python. *Journal of Open Source Software*, 3(31), 1026. <https://doi.org/10.21105/joss.01026>
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., & others. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>