How to Perform the Nemenyi Post-Hoc Test in Python with Statsmodels

Authored by

stats writer

December 6, 2025

ARABPSYCHOLOGY.COM

RECOMMENDED CITATION

stats writer (2025). *How to Perform the Nemenyi Post-Hoc Test in Python with Statsmodels*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=106162

The Nemenyi Post-Hoc Test is a crucial statistical procedure utilized when comparing the means of multiple related groups, typically following a significant result from an omnibus test like the Friedman Test. It serves as a necessary safeguard against the inflation of Type I error rates associated with performing multiple pairwise comparisons. While historical implementations often relied on statistical suites, this robust procedure is readily accessible in the Python programming environment through specialized libraries such as `scikit-posthocs`. Executing this test involves careful data preparation, running the preliminary omnibus test to confirm global differences, and finally applying the post-hoc procedure to obtain detailed pairwise comparison results.

Introduction to the Nemenyi Post-Hoc Test

The Nemenyi procedure is a powerful statistical tool categorized as a non-parametric multiple comparison test. It is specifically designed to analyze data where the strict assumptions required for traditional parametric tests, such as the assumption of normality or homogeneity of variances, are not met, or when dealing primarily with ordinal data. Crucially, the Nemenyi test functions as a dedicated follow-up test, or post-hoc analysis, performed only after an initial omnibus test has already provided evidence that differences exist among group medians. This sequential approach is vital for pinpointing exactly which specific pairs of groups contribute to the overall statistically significant difference identified by the initial test.

Unlike methods tailored for independent samples, the Nemenyi test is most often applied after performing the Friedman Test, which is the appropriate omnibus test for experimental designs involving related samples, such as repeated measures where the same subjects are measured across three or more different conditions. The fundamental statistical methodology of the Nemenyi method involves calculating a critical difference (CD) value based on the ranking of observations within each subject or block. This calculation allows for cautious, simultaneous comparisons across all possible pairs of groups while effectively maintaining the family-wise error rate ($text{FWER}$) at the desired significance level. This rigorous control over potential false positives is essential for ensuring that any claimed differences are robust, reliable, and scientifically defensible.

It is important to understand the precise statistical context for employing this specific test. It operates as the non-parametric analog to the post-hoc procedures that follow a repeated measures Analysis of Variance (ANOVA). If the study were instead designed with independent groups, the analysis would typically begin with the Kruskal-Wallis H test, followed by an appropriate post-hoc test like Dunn's test. Since our focus here is on repeated measures involving related samples, we will adhere strictly to the methodology that necessitates the use of the Friedman Test and subsequently, the Nemenyi Post-Hoc Test.

ARABPSYCHOLOGY.COM

Why We Need Post-Hoc Analysis: The Role of the Friedman Test

Before any Nemenyi procedure can be meaningfully executed, the preliminary omnibus test--the Friedman Test--must confirm the existence of a global effect across the experimental conditions. The Friedman test is designed to assess whether there is a statistically significant difference in the central tendencies of three or more related samples. If this test yields a sufficiently large test statistic and a correspondingly small P-value (typically compared against the chosen significance level, $alpha = 0.05$), we are justified in rejecting the null hypothesis, thereby confirming that the group distributions are not all identical.

However, the finding from the Friedman test is inherently limited. It only indicates an overall difference; it fails to specify the location of that difference. The outcome only tells us that "at least one pair of groups differs significantly." If the analysis were to conclude at this stage, the researcher's understanding would be fundamentally incomplete. This ambiguity is precisely why a dedicated follow-up procedure is necessitated. Conversely, if the initial P-value derived from the Friedman test is not statistically significant (i.e., $p ge 0.05$), the analysis must cease, as there is insufficient evidence to suggest any differences warranting further pairwise exploration.

The underlying statistical necessity for the Nemenyi test stems from the need to rigorously control the Family-Wise Error Rate ($text{FWER}$). When an experiment involves three groups (let's say A, B, and C), there are three distinct pairwise comparisons possible: (A vs. B), (A vs. C), and (B vs. C). If a researcher were to conduct three separate t-tests, each using an alpha level of 0.05, the cumulative probability of committing at least one Type I error (a false positive) across the entire family of tests increases dramatically, compromising the reliability of the findings. The Nemenyi test, through its built-in adjustment mechanism, ensures that the overall confidence level for the entire set of comparisons remains robust, confirming the reliability of identified differences.

Step 1: Setting Up Data for Paired Non-Parametric Analysis

We will illustrate the application of the Nemenyi test in Python using a practical research example. Imagine a scenario where a pharmacologist seeks to compare the efficacy of three distinct treatments--Drug A, Drug B, and Drug C--on patient reaction times. Crucially, the same group of patients is tested sequentially under all three conditions, creating a classic repeated measures or related samples design. The primary objective is to determine if the median reaction times are statistically equivalent across the three drugs or if specific treatments result in significantly different performance metrics.

In this simulated experiment, the reaction time (measured in seconds) is recorded for 10 distinct patients under the influence of each of the three drugs. The structure of the input data is critical: the observations must be organized into separate arrays where the index position within each array corresponds to the specific patient being measured. This specific arrangement allows the

ARABPSYCHOLOGY.COM

statistical procedure to correctly account for the dependency inherent in the paired nature of the observations, which is fundamental for both the Friedman and subsequent Nemenyi tests.

We define the three data arrays in Python to accurately reflect the collected response times for each of the pharmacological treatments. These structures ensure that the statistical analysis correctly interprets the relationships between the measurements:

group1 =
group2 =
group3 =

These arrays, representing Drug A, Drug B, and Drug C respectively, form the foundational input required for initiating our statistical analysis.

Step 2: Executing the Friedman Test in Python

As mandated by the statistical workflow, the initial step requires the execution of the Friedman Test. We leverage the robust statistical capabilities provided by the `scipy.stats` library in Python for this purpose. The function `friedmanchisquare()` is called, taking the separate arrays representing the related groups as its mandatory arguments. A successful computation here generates the test statistic and the corresponding omnibus P-value, establishing the justification for the subsequent Nemenyi analysis.

The necessary code execution is shown below, demonstrating the import and application of the Friedman test using our defined group data:

```
from scipy import stats

#perform Friedman Test
stats.friedmanchisquare(group1, group2, group3)

FriedmanchisquareResult(statistic=13.3513513, pvalue=0.00126122012)
```

The interpretation of these results relies on the established principles of statistical hypothesis testing. The Friedman test is designed to test the following competing hypotheses:

The null hypothesis (H0): The median reaction time is equal for all three drug populations.
The alternative hypothesis (Ha): At least one drug population's median reaction time differs significantly from the others.

In this specific outcome, the calculated test statistic is **13.35135**, and the resulting P-value is

ARABPSYCHOLOGY.COM

0.00126. Since this P-value (0.00126) is significantly smaller than the conventional significance threshold of $alpha = 0.05$, we possess compelling evidence to reject the null hypothesis. The conclusion is that the type of drug administered has a statistically significant effect on patient response times, thereby providing the necessary statistical justification to proceed to the Nemenyi post-hoc test to identify the specific pairwise differences.

Step 3: Installing Necessary Libraries for Nemenyi Analysis

Crucially, the complex calculation required for the Nemenyi post-hoc test is not natively integrated within standard Python libraries such as `scipy` or `statsmodels`. To perform this analysis efficiently, researchers rely on a specialized, third-party package known as `scikit-posthocs`. This library is specifically engineered to offer a comprehensive suite of post-hoc comparison tests suitable for both parametric and non-parametric data, positioning it as an indispensable tool for advanced comparative statistics in Python.

Before proceeding with the analytical code, this external library must be successfully installed into the current Python environment. This process is standardized and easily accomplished using Python's primary package management system, `pip`. Successful installation ensures that the specific functions required for the analysis, particularly `posthoc_nemenyi_friedman()`, are correctly indexed and available for immediate execution in the following step.

The necessary installation command is executed in the terminal or command line interface:

pip install scikit-posthocs

Once the package dependencies are resolved and the library is confirmed operational, we can confidently prepare the data in the required structure and proceed with the definitive pairwise comparison analysis.

Step 4: Performing the Nemenyi Post-Hoc Test

The implementation of the Nemenyi post-hoc test requires careful attention to the data format expected by the `scikit-posthocs` function. The `posthoc_nemenyi_friedman()` function requires the input data to be organized such that the rows represent the subjects (the repeated measure units or blocks), and the columns represent the different treatment conditions or groups. Since our initial data arrays (`group1`, `group2`, `group3`) were defined in a column-wise manner (each array being a column of data), we must first consolidate them and then perform a transposition.

We utilize the powerful numerical capabilities of the `numpy` library to combine the three Python lists into a single multi-dimensional `numpy` array. The critical step is the subsequent transposition operation, achieved using `data.T`. This reshaping transforms the array orientation from (Groups

ARABPSYCHOLOGY.COM

$times$ Patients) to the required (Patients $times$ Groups) format, which is essential for the Nemenyi test to correctly interpret the data as related samples and accurately conduct the within-subject ranking procedure.

The following code block demonstrates the necessary data consolidation, transformation, and the final execution of the Nemenyi post-hoc test:

```
import scikit_posthocs as sp
import numpy as np

#combine three groups into one array
data = np.array()

#perform Nemenyi post-hoc test
sp.posthoc_nemenyi_friedman(data.T)
```

```
0 1 2
0 1.000000 0.437407 0.065303
1 0.437407 1.000000 0.001533
2 0.065303 0.001533 1.000000
```

A crucial technical note: the transposition (`data.T`) is absolutely mandatory for related samples analysis. Failure to transpose the array would result in the function incorrectly treating the treatment groups as subjects and the subjects as multiple dependent variables, which violates the fundamental statistical assumption of the Friedman Test post-hoc procedure, yielding erroneous results.

Step 5: Interpreting the Pairwise Comparison Results

The output generated by the Nemenyi post-hoc test is a symmetrical matrix (or a similar data structure) where both the row and column indices represent the groups being compared (0, 1, and 2 corresponding to group1, group2, and group3, respectively). The values contained within this matrix are the adjusted P-values for every possible pairwise comparison, accounting for the multiple testing adjustments necessary to control the $text{FWER}$.

We analyze the off-diagonal entries to determine which specific group pairings demonstrate statistical significance. The matrix output provides the following key adjusted P-values:

P-value comparing Group 0 and Group 1 (Drug A vs. Drug B): **0.4374**
P-value comparing Group 0 and Group 2 (Drug A vs. Drug C): **0.0653**
P-value comparing Group 1 and Group 2 (Drug B vs. Drug C): **0.0015**

ARABPSYCHOLOGY.COM

The final statistical inference is made by comparing these adjusted P-values against our predefined significance level, $alpha = 0.05$. If the adjusted P-value is less than 0.05, we reject the null hypothesis of no difference for that specific pair.

Based on our findings, we draw the following conclusions:

For the comparison between Group 0 and Group 1 (P = 0.4374), since $0.4374 > 0.05$, we conclude there is no statistically significant difference in median reaction times.
For the comparison between Group 0 and Group 2 (P = 0.0653), since $0.0653 > 0.05$, we conclude there is no statistically significant difference, although the difference is marginally close to significance.
For the comparison between Group 1 and Group 2 (P = 0.0015), since $0.0015 < 0.05$, we confidently reject the null hypothesis, concluding that there is a highly significant difference in reaction times between Drug B and Drug C.

In summary, the Nemenyi post-hoc test successfully localized the overall significant effect found by the Friedman test. The analysis confirms that among the three drugs tested, only the difference between Drug B (Group 1) and Drug C (Group 2) achieves the threshold for statistical significance, providing the researcher with a clear and reliable conclusion regarding the comparative efficacy of the treatments.

ARABPSYCHOLOGY.COM