

Learning the Friedman Test: A Python Tutorial for Non-Parametric Analysis

Authored by
Mohammed loot

November 8, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Learning the Friedman Test: A Python Tutorial for Non-Parametric Analysis*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=12687>

The [Friedman Test](#) is an indispensable [non-parametric](#) statistical procedure, functioning as the robust alternative to the standard [Repeated Measures ANOVA](#). This test is meticulously engineered for analyzing complex experimental designs involving dependent samples, where the primary analytical goal is to definitively assess whether statistically significant differences exist among the central tendencies of three or more related groups. Its utility is paramount for researchers and data scientists who analyze data derived from longitudinal studies or within-subject experiments--scenarios where the same subjects or entities are measured repeatedly under varying conditions or across multiple sequential time points. By controlling for inherent individual variability, the Friedman Test provides a powerful and precise method for comparison.

A key advantage of employing the Friedman Test, in contrast to its parametric equivalent, lies in its freedom from stringent distributional assumptions. Specifically, it does not require the underlying population data to follow a normal distribution, making it exceptionally reliable when handling [ordinal data](#) or datasets that exhibit severe violations of normality. For modern data scientists and statisticians navigating the complexities of real-world data, mastering the implementation of this test within a powerful programming environment like Python is absolutely essential. This comprehensive tutorial delivers a detailed, step-by-step methodology on how to accurately execute the [Friedman Test](#) utilizing the specialized statistical tools available within the Python ecosystem, ensuring rigorous and defensible statistical inference.

The Friedman Test: Non-Parametric Power for Dependent Samples

The core utility of the Friedman Test is realized in situations characterized by a classic [randomized block design](#) where the measurements taken within each block--or, more commonly, within each subject--are inherently dependent upon one another. A typical example involves comparing the effects of multiple treatments, such as administering Drug A, Drug B, and Drug C to the same cohort of patients. Since the observations are intrinsically related (e.g., Patient 1's response to Drug A is related to their response to Drug B), traditional tests designed for independent samples are statistically inappropriate and would lead to inflated error rates. The Friedman Test elegantly resolves this dependency issue by transforming the raw scores into ranks within each block, calculating the test statistic based on these internal rankings.

This approach of ranking the data significantly minimizes the detrimental influence of extreme values (outliers) and underlying distributional irregularities. By working with ranks, the [non-parametric](#) method provides a remarkably reliable framework for [hypothesis testing](#), particularly when the rigid assumptions demanded by parametric tests--such as normality or sphericity--cannot be reasonably satisfied by the data. The test ultimately computes a [chi-square statistic](#), which assesses the consistency of the differences observed across the treatment groups. A deep conceptual understanding of the circumstances requiring this powerful non-parametric method is critical for ensuring that the statistical conclusions drawn from repeated measures experiments are

both valid and scientifically robust.

In summary, the statistical focus of the Friedman Test is to determine whether the observed differences in the central location (usually mediated by median ranks) among the various conditions are substantially greater than what could be plausibly attributed to random chance alone, operating under the formal assumption that the treatments have no differential effect. This procedure serves as a powerful and often essential preliminary step in the analysis of complex experimental designs that necessitate within-subject comparisons, setting the stage for more focused follow-up analyses if the initial omnibus test proves significant.

Crucial Assumptions and Data Prerequisites

While the Friedman Test is categorized as [non-parametric](#), granting it flexibility regarding data distribution, it is not entirely free of foundational assumptions. To ensure the validity and reliability of the analytical results, researchers must confirm that their data adheres to several specific prerequisites before proceeding with execution. Foremost among these is the requirement for a [randomized block design](#). This means that every single treatment condition must be administered to the exact same set of subjects, or that subjects must be organized into matched blocks where observations across conditions are related. Secondly, the dependent variable must be measured at least on an [ordinal scale](#). Although the test can certainly utilize interval or ratio data, it is crucial to remember that the procedure internally converts these scores into ranks, effectively treating them as ordinal data during the calculation phase.

A critical structural prerequisite is the need for three or more related groups or conditions ($k \geq 3$). The Friedman Test is specifically designed for scenarios involving multiple treatments. If the comparison involves only two related groups, the statistically correct and more powerful non-parametric alternative is the [Wilcoxon Signed-Rank Test](#). Furthermore, the test strictly assumes that the blocks (subjects) themselves are independent of one another. This independence means that the outcome or response observed in one patient must not influence, bias, or otherwise correlate with the response of any other patient within the study cohort.

Finally, consistency in measurement is paramount: each subject must have a recorded observation for every single treatment condition being tested. Missing data points can complicate or invalidate the analysis, as the ranking process relies on a complete set of scores across all conditions for each block. By diligently verifying these prerequisites, researchers can confidently select and apply the [Friedman Test](#), knowing that it represents the most appropriate statistical methodology for drawing robust and accurate inferences from their specific dependent dataset. Failure to meet these core assumptions risks producing unreliable statistical inference and potentially leading to flawed conclusions regarding treatment efficacy or differences between conditions.

Practical Application: A Clinical Research Case Study

To effectively illustrate the practical utility and application of the Friedman Test, let us examine a typical scenario encountered in clinical and pharmaceutical research. Imagine a pharmaceutical researcher dedicated to investigating whether three distinct experimental drugs--designated Drug A, Drug B, and Drug C--exhibit differential effects on human reaction time. Recognizing that individual physiological variability can severely confound results, the researcher consciously adopts a repeated measures design, administering all three drug conditions to the same standardized cohort of 10 volunteer patients over a period of time. The chosen primary outcome measure is the patient's reaction time, meticulously recorded in seconds, following the administration of each specific drug.

The fundamental scientific question driving this investigation is focused on the equality of central tendencies: Are the mean reaction times across the three drugs statistically equivalent, or does at least one drug induce a significantly different reaction time compared to the others? Given the impossibility of guaranteeing the strict normality of biological reaction time distributions across a small sample size, coupled with the inherent dependency of the samples (because the same subjects are measured repeatedly), the Friedman Test emerges as the optimal and most scientifically sound analytical method. The steps that follow detail the precise execution of this test within Python, providing the statistical mechanism necessary to address this pivotal research hypothesis.

Implementing the Friedman Test in Python using SciPy

The foundational practical step in Python involves the crucial task of organizing the collected experimental data into appropriate numerical arrays or lists. In statistical programming environments, it is essential that each condition or treatment group is represented by a separate data structure, with the individual measurements meticulously aligned by subject across all arrays. For our specific clinical case study, we must define three distinct arrays, which correspond directly to the reaction times observed under Drug A (group1), Drug B (group2), and Drug C (group3). This critical alignment--where the index position within each array corresponds to a specific patient--is what formally preserves the dependent nature of the repeated measurements, which is central to the test's validity.

We will utilize the standard Python list structure to store these numerical observations. This native structure is highly efficient and seamlessly compatible with the sophisticated statistical functions readily available within the widely accepted [SciPy](#) library, the cornerstone of scientific computing in Python. The raw data collected from the 10 patients across the three experimental conditions is entered precisely as follows, setting the stage for computation:

Step 1: Enter the data.

```
group1 =  
group2 =  
group3 =
```

It is imperative at this stage to verify that all defined groups contain an equal number of observations ($n=10$ in this instance), corresponding exactly to the total number of subjects who participated in the study. This meticulous setup confirms the data's structural readiness for the subsequent stage of statistical processing and computational analysis.

Executing the Friedman Test using SciPy

Python's extensive statistical analysis capabilities are principally driven by the [SciPy](#) library, which provides the dedicated function required for this specific non-parametric analysis: `friedmanchisquare`. This specialized function is conveniently housed within the `scipy.stats` module. To proceed with the analysis, we must first import the necessary modules, and subsequently pass our previously defined group arrays directly as arguments into the function call.

The `friedmanchisquare` function is expertly designed to manage multiple input arrays simultaneously, automatically treating each array as a separate condition within the repeated measures design. Behind the scenes, it internally calculates the required within-subject ranks, computes the [chi-square statistic](#) approximation of the [Friedman Test](#), and calculates the corresponding [p-value](#). This automated process streamlines the complex statistical steps into a single, straightforward operation:

Step 2: Perform the Friedman Test.

The execution process is remarkably straightforward, requiring only the importation of the SciPy library and the subsequent call to the specific test function with the data arrays included:

```
from scipy import stats
```

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

```
(statistic=13.3514, pvalue=0.00126)
```

The resulting output delivers two critically important numerical values: the calculated test statistic and the associated [p-value](#). These two numbers form the absolute basis for our inferential conclusion regarding whether the observed differences in reaction times associated with the three drugs are statistically meaningful. The test statistic quantifies the magnitude of the difference observed across the calculated ranks, while the p-value rigorously quantifies the probability of

observing such a large or larger difference if, in reality, the drugs truly had no differential effect whatsoever.

Interpreting Results and Essential Post-Hoc Analysis

Interpreting the numerical results derived from the Friedman Test mandates a clear and formal understanding of the underlying [null hypothesis](#) and its competing alternative hypothesis. These hypotheses formally define the relationship that the statistical test is designed to evaluate, establishing the framework for our decision-making process:

The test operates under the following formalized framework:

The Null Hypothesis (H₀): This posits that the distributions (or the median reaction times, in practical context) of the populations from which the three samples were drawn are statistically equal. In the context of our study, this means the average reaction time for Drug A, Drug B, and Drug C are all indistinguishable from one another.

The Alternative Hypothesis (H_a): This asserts that at least one population median is significantly different from the others. Substantively, this suggests that the specific type of drug administered does have a genuine and measurable influence on the patient's reaction time.

Step 3: Interpret the results.

In our current example, the calculated test statistic is **13.3514**, and the corresponding [p-value](#) is **0.00126**. To arrive at a formal statistical decision regarding the [null hypothesis](#), we must compare this calculated p-value against a predetermined significance level, conventionally denoted as α and typically set at 0.05. The core decision rule in frequentist statistics stipulates that if the p-value is less than the chosen significance level (α), we are statistically obligated to reject the null hypothesis (H_0). Since 0.00126 is substantially smaller than 0.05, we confidently reject H_0 . This rejection provides strong statistical evidence, signifying that the differences observed in the reaction times across the three drugs are highly unlikely to have occurred simply by random chance variation.

Consequently, we conclude that the type of drug administered leads to statistically significant differences in patient response time. However, while the [Friedman Test](#) successfully established that a significant difference exists among the three groups, it is an omnibus test--meaning it does not specify precisely which pairs of groups are the source of this difference. To move beyond the general conclusion of "difference exists" and identify the specific pair-wise differences (e.g., does Drug A differ from Drug B?), a necessary follow-up procedure known as [post-hoc analysis](#) is absolutely required.

For the Friedman Test, the appropriate non-parametric [post-hoc analysis](#) usually involves applying

the [Wilcoxon Signed-Rank Test](#) to all possible pairwise comparisons (A vs B, A vs C, B vs C). Crucially, this must be combined with a specialized correction mechanism, such as the Bonferroni correction or the preferred Nemenyi test, designed to control the family-wise error rate and maintain statistical rigor across multiple comparisons. Implementing this final layer of analysis ensures that the researcher can provide specific, actionable insights--such as definitively identifying the most effective or slowest-acting drug--thereby completing the statistical investigation and delivering a comprehensive answer to the initial research question.