

# Learning the Friedman Test: A Guide to Non-Parametric Comparison of Related Groups

Authored by  
**Mohammed loot**

November 8, 2025

## RECOMMENDED CITATION

Mohammed loot (2025). *Learning the Friedman Test: A Guide to Non-Parametric Comparison of Related Groups*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=13089>

The **Friedman Test** is a highly valued statistical procedure, serving as the **non-parametric alternative** to the one-way repeated measures **ANOVA** (Analysis of Variance). This powerful statistical tool is specifically designed to analyze data derived from matched samples or **block designs**, where the same group of subjects or units is measured across three or more different conditions or time points. Its fundamental purpose is to determine whether a **statistically significant difference** exists between the central tendencies of these groups when dealing with dependent samples.

Unlike traditional parametric tests, the Friedman Test is robust against violations of key assumptions, such as normality, because its calculation relies entirely on ranking the data within each subject or block. This makes it an essential choice when researchers are working with **ordinal data** or when the underlying population distributions are unknown or clearly non-normal, ensuring valid conclusions can still be drawn from complex experimental designs involving repeated measurements.

## Prerequisites for Applying the Friedman Test

The core requirement for appropriately employing the Friedman Test is the presence of three or more related samples. This specialized setup, often referred to as a **repeated measures design** or randomized block design, necessitates that the same individuals or blocks are measured under all conditions being tested. The test rigorously evaluates whether the central tendencies (specifically the median or mean ranks) of the measurements differ significantly across these experimental conditions.

Researchers typically turn to the Friedman Test in experimental settings where minimizing individual subject variability is paramount. By utilizing the same subjects for every treatment, the test effectively controls for inherent between-subject differences, focusing the analysis purely on the effect of the treatments themselves. This inherent dependency structure--where observations within a row are linked--is what mandates the use of non-parametric tests like Friedman, as standard independent sample tests would be statistically inappropriate and potentially lead to inaccurate findings.

## Common Applications and Practical Scenarios in Research

The Friedman Test is widely applied across diverse disciplines, ranging from clinical medicine to behavioral psychology, whenever data is collected under multiple, dependent conditions. The two most frequent types of applications involve measuring changes in a variable over time or comparing performance under various experimental treatments or stimuli.

A typical application involves measuring the mean scores of subjects during three or more distinct time points. For example, a sports researcher might track the resting heart rate of a cohort of

subjects: one month before beginning a new strenuous training program, one month after starting the program, and finally, two months into the program. By performing the Friedman Test on the ranked heart rate data, the researcher can definitively assess whether there is a significant systemic change in the mean resting heart rate across these three sequential time periods, providing crucial insight into the program's efficacy and longitudinal impact.

A second major application involves assessing the mean scores of subjects when exposed to three or more different conditions or stimuli. Consider a psychological study where participants are asked to watch three distinct movie trailers and then rate each one based on their enjoyment. Since every subject provides a rating for all three movies, their responses are inherently dependent. The Friedman Test allows us to determine if the mean enjoyment ratings differ significantly among the three movies, confirming if one stimulus was consistently preferred over the others within the same population sample.

## Executing the Friedman Test: A Detailed Statistical Example

To clearly illustrate the practical application of this test, suppose we aim to determine if the mean reaction time of patients is significantly different when they are administered three distinct pharmaceutical drugs (Drug 1, Drug 2, and Drug 3). For this study, we recruit 10 patients and measure each patient's reaction time (in seconds) after taking each of the three drugs on separate occasions. Since each patient serves as their own control and is measured in all three conditions, the data requires a repeated measures analysis, making the Friedman Test the appropriate statistical choice. The raw data collected from this hypothetical experiment is presented below:

Patient	Drug 1	Drug 2	Drug 3
Patient 1	4	5	2
Patient 2	6	6	4
Patient 3	3	8	4
Patient 4	4	7	3
Patient 5	3	7	2
Patient 6	2	8	2
Patient 7	2	4	1
Patient 8	7	6	4
Patient 9	6	4	3
Patient 10	5	5	2

The subsequent analysis proceeds through defined, sequential steps, starting with the formal hypothesis definition and concluding with result interpretation based on the calculated test statistic and the associated p-value.

## Formulating Hypotheses and Calculating the Test Statistic

The formal process of the Friedman Test begins by meticulously defining the statistical hypotheses. We must establish the **null hypothesis** ( $H_0$ ), which posits that there is no difference in the central tendency across the treatment groups, and the alternative hypothesis ( $H_a$ ), which suggests that at least one group's central tendency differs significantly from the rest.

The hypotheses for our drug reaction time study are formally stated as follows:

**The null hypothesis ( $H_0$ ):**  $\mu_1 = \mu_2 = \mu_3$  (The median reaction times across the three drug populations are all equal.)

**The alternative hypothesis ( $H_a$ ):** At least one population median is different from the rest (i.e., the drugs have a differential effect on reaction time).

The calculation phase involves ranking the data within each patient (row) and then summing these ranks across the treatments (columns) to derive the Friedman **test statistic** ( $Q$ ). Although understanding the manual process is crucial for conceptual grounding, in practical research settings, we rely on advanced statistical software or dedicated online calculators to manage these complex computations efficiently.

Using a statistical calculator with the input data provided above demonstrates the necessary setup for computation:

Group 1	Group 2	Group 3	Group 4	Group 5
4	5	2		
6	6	4		
3	8	4		
4	7	3		
3	7	2		
2	8	2		
2	4	1		
7	6	4		
6	4	3		
5	5	2		

After the calculation is executed by the software, it provides the critical output values necessary for interpretation: the test statistic  $Q$  and the corresponding **p-value**. This output confirms the necessary metrics needed to make a sound statistical decision.

CALCULATE

Test Statistic  $Q$ : 12.35000

p-value: 0.00208

## Interpreting the Results and Standardized Reporting

Based on the software analysis, the resulting **test statistic** is  $Q = 12.35$ , and the corresponding p-value is  $p = 0.00208$ . To interpret these findings accurately, we must compare the calculated p-value against a predetermined significance level (alpha, typically set at 0.05). Since the p-value of 0.00208 is significantly less than the critical alpha value of 0.05, we possess overwhelming evidence to reject the **null hypothesis** ( $H_0$ ).

The rejection of the null hypothesis leads us directly to the conclusion that the median reaction time is statistically unequal across the three drugs. In practical terms, this signifies that the type of drug administered leads to statistically significant differences in patient response time. This finding generally warrants further investigation through [post-hoc testing](#) (such as the Conover or Nemenyi tests) to determine precisely which specific pairs of drugs differ from one another, rather than just knowing that a difference exists.

The final and crucial step in any statistical analysis is clearly and concisely reporting the findings, ensuring all critical parameters of the test are included. This standardized reporting includes the statistical test used, the sample size, the calculated test statistic (Q), and the associated [p-value](#). Clarity and adherence to standard format are essential for the scientific community to accurately assess the results.

A **Friedman Test** was conducted on 10 patients to examine the effect that three different drugs had on response time, with each patient utilizing all three drugs once.

Results showed that the type of drug used led to statistically significant differences in response time ( $Q = 12.35$ ,  $p = 0.00208$ ).

## Implementing the Friedman Test Using Statistical Software

While mastering the manual calculation provides a necessary understanding of the underlying mathematical principles, modern researchers consistently utilize powerful statistical programming languages or specialized software packages for complex data analysis. These tools streamline the ranking and calculation process, ensuring accuracy and efficiency for large datasets.

The following resource provides practical guidance on implementing the Friedman Test using common statistical tools, enabling researchers to integrate this powerful non-parametric method into their standard workflow:

[How to Perform the Friedman Test in Python](#)