

# Learning Repeated Measures ANOVA with Python: A Step-by-Step Guide

Authored by  
**Mohammed Iooti**

November 8, 2025

## RECOMMENDED CITATION

Mohammed Iooti (2025). *Learning Repeated Measures ANOVA with Python: A Step-by-Step Guide*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=12689>

## The Power of Repeated Measures ANOVA: A Foundation

A [Repeated Measures ANOVA](#) (Analysis of Variance) represents a sophisticated [statistical technique](#) designed for comparing the means of three or more groups that are inherently related. Its defining characteristic, which sets it apart from a standard one-way ANOVA, is the requirement that the **same subjects** participate in, and are measured across, all levels or conditions of the independent variable. This experimental approach is frequently referred to as a within-subjects design.

This type of design is indispensable in areas such as developmental psychology, clinical trials, and longitudinal studies, where researchers must track changes within an individual over a period of time or in response to different stimuli. By repeatedly measuring the **same subjects**, the RM-ANOVA significantly minimizes the impact of inter-subject variability--the natural differences existing between individuals--which is a major source of error in between-subjects designs. This reduction in unexplained variance inherently leads to greater [statistical power](#), making the analysis more sensitive to detecting genuine treatment effects.

The primary objective of conducting an RM-ANOVA is to rigorously determine if there is a [statistically significant difference](#) among the means observed across these repeated measurements. This tutorial will provide a comprehensive, step-by-step methodology for executing a one-way repeated measures ANOVA. We will leverage the robust open-source capabilities of the [Python](#) ecosystem, specifically relying on the **pandas** library for efficient data manipulation and the specialized **statsmodels** library for performing the necessary statistical computations, ensuring a transparent and reproducible research pipeline.

## Setting the Stage: A Pharmacological Efficacy Study

To demonstrate the practical application of the Repeated Measures ANOVA, we will explore a typical scenario rooted in clinical research. Imagine a team of researchers who wish to evaluate whether four distinct pharmacological treatments (drugs) have differential effects on human reaction time. Since baseline reaction speed varies widely among individuals, employing a within-subjects design is strategically crucial to isolate the effect of the drug itself from pre-existing individual differences.

In this controlled experiment, a small cohort of five patients is recruited. The core element of the design is that the reaction time of **each patient** is meticulously recorded four separate times--once following the administration of each of the four different drug treatments. Because the dependent variable (reaction time) is measured repeatedly on the identical individuals across the various levels of the independent variable (the four drugs), the RM-ANOVA is the only statistically appropriate method for analysis. Attempting to use a standard ANOVA would violate the critical assumption of independence between observations, potentially yielding skewed results and

leading to erroneous conclusions regarding the drugs' comparative efficacy.

Our formal hypothesis posits that at least one of the administered drugs will produce a population mean reaction time that is distinct from the others. To test this hypothesis rigorously, we must first correctly format the raw data within the [Python](#) environment, a prerequisite that ensures the subsequent statistical analysis is valid and accurate.

## Step 1: Data Preparation and Structuring for Analysis

The initial and most vital step for any statistical analysis performed in [Python](#) involves preparing and structuring the raw experimental data correctly. For this task, we utilize the powerful array manipulation capabilities of **NumPy** and the data structuring efficiency of **Pandas** DataFrames. It is essential to note that the `AnovaRM` function within the `statsmodels` library requires the input data to be organized in the **long format**. This specific format dictates that every single observation--meaning a subject's measured response under one specific condition--must occupy its own row within the DataFrame.

Therefore, our resulting DataFrame must contain three mandatory columns: a unique identifier for the **subject** (the patient), a column defining the **within-subjects factor** (the specific drug level administered), and the column containing the **dependent variable** (the measured response time). The code below demonstrates how to construct this sample dataset efficiently, followed by a preview that illustrates the required long-format structure.

```
import numpy as np
import pandas as pd
```

```
#create data
df = pd.DataFrame({'patient': np.repeat(, 4),
                  'drug': np.tile(, 5),
                  'response': })
```

```
#view first ten rows of data
df.head
```

```
patient drug response
0 1 1 30
1 1 2 28
2 1 3 16
3 1 4 34
4 2 1 14
5 2 2 18
```

```
6 2 3 10
7 2 4 22
8 3 1 24
9 3 2 20
```

This careful organization confirms the structure: Patient 1 has four separate measurements recorded, corresponding precisely to the four levels of the drug factor (1 through 4) in the 'response' column. The function `np.repeat` systematically lists each patient ID four times, while `np.tile` cycles the drug levels, ensuring the conditions are correctly aligned with the corresponding reaction times. This prepared and validated [Pandas](#) DataFrame is now ready for the rigorous statistical modeling phase.

## Step 2: Executing the Repeated Measures ANOVA using Statsmodels

Once the data is correctly structured in the long format, the next crucial step is to execute the ANOVA calculation. We turn to the [statsmodels](#) library, which is essential for statistical modeling and econometrics within the [Python](#) environment. We specifically employ the `AnovaRM` function, which is expertly designed to manage the dependencies inherent in repeated measures designs. This function requires several precise parameters to define the experimental structure and the relationship between the variables.

The parameters passed to `AnovaRM` are intuitive yet critical: `data` (the prepared [Pandas](#) DataFrame), `depvar` (the dependent variable, 'response'), `subject` (the column identifying the unique individuals, 'patient'), and `within` (the list containing the name of the within-subject factor, which is 'drug'). Correct specification of these factors is paramount, as it enables the function to accurately partition the total variance observed into the variance explained by the drug effect and the residual error variance specific to the within-subjects design.

The following code block initiates the RM-ANOVA calculation and subsequently prints the resulting ANOVA table, providing the foundational metrics required for testing our core hypothesis.

```
from statsmodels.stats.anova import AnovaRM
```

```
#perform the repeated measures ANOVA
print(AnovaRM(data=df, depvar='response', subject='patient', within=).fit())
```

```
Anova
```

```
=====
```

```
F Value Num DF Den DF Pr > F
```

```
-----
```

```
drug 24.7589 3.0000 12.0000 0.0000
```

=====

The output is a concise summary table that delivers the core statistical information: the F Value, the degrees of freedom (Numerator and Denominator), and the critical P-value ( $Pr > F$ ). These metrics form the basis for interpreting the findings and determining whether the pharmacological treatments produced a statistically significant effect on the patients' reaction times.

### Step 3: Interpreting the Statistical Results and Drawing Conclusions

The entire structure of the RM-ANOVA is built upon the comparison of the [null hypothesis](#) ( $H_0$ ) against the alternative hypothesis ( $H_a$ ). These hypotheses formalize the statistical question being addressed by the experiment:

**The Null Hypothesis ( $H_0$ ):**  $\mu_1 = \mu_2 = \mu_3 = \mu_4$ . This states that the population mean reaction times are equivalent across all four drug treatments.

**The Alternative Hypothesis ( $H_a$ ):** At least one population mean is different from the rest. This suggests that at least one drug yields a mean reaction time that differs significantly from the others.

The interpretation of the results hinges on the magnitude of the [F-test statistic](#) and its corresponding p-value. The F-ratio is fundamentally a ratio of the variance explained by the factor (the drug effect) to the residual error variance (the unexplained variance within subjects). A substantial F-value indicates that the differences observed between the drug means are considerably larger than what would be expected purely due to random measurement error.

Reviewing the output provided by the [statsmodels](#) analysis, we observe that the [F-test statistic](#) for the drug effect is calculated as **24.7589**. Crucially, the associated p-value ( $Pr > F$ ) is **\*\*0.0000\*\***. Given that this calculated p-value (0.0000) is far below the widely accepted significance threshold (alpha level) of 0.05, we possess robust evidence to reject the null hypothesis. This rejection confirms that the specific type of drug administered has a statistically significant influence on reaction time. In essence, the average reaction times under the four different drug conditions are not all identical within the population.

While the RM-ANOVA confirms the existence of a difference among the groups, it does not pinpoint which specific drug pairs are responsible for this variance. To precisely determine the nature of the pairwise differences (e.g., comparing Drug 1 versus Drug 3), researchers must conduct follow-up analyses known as [post-hoc tests](#). These tests, such as Tukey's HSD or Bonferroni correction, are essential for maintaining statistical integrity by adjusting for the increased risk of Type I errors that arise from multiple comparisons.

## Step 4: Formal Reporting and Contextualizing the Findings

The final stage of a rigorous statistical analysis involves reporting the results accurately and transparently, often adhering to specific style guidelines like those set by the American Psychological Association (APA). A formal statistical report must include the type of test conducted, the degrees of freedom associated with the F-statistic, the calculated F-value itself, and the precise p-value (or a boundary if it is extremely small).

For our pharmacological case study, the essential components extracted directly from the ANOVA summary table are:

**F-value:** 24.7589

**Numerator Degrees of Freedom (Num DF):** 3 (derived from the number of factor levels minus one)

**Denominator Degrees of Freedom (Den DF):** 12 (calculated as the Numerator DF multiplied by the number of subjects minus one)

**P-value:** < 0.001 (reported conservatively due to the extremely small calculated value)

We can now synthesize these components into a clear, formal statement that provides the necessary context regarding the experimental design and the definitive conclusion drawn from the hypothesis test. This formal presentation is vital for ensuring the reproducibility and credibility of the research findings.

The following is an exemplary formal report of the findings:

A one-way repeated measures ANOVA was conducted on 5 participants to examine the effect of four distinct drug treatments on response time. Results indicated that the type of drug administered led to **statistically significant differences** in mean response time ( $F(3, 12) = 24.76, p < 0.001$ ).

This succinct conclusion confirms the rejection of the null hypothesis and firmly establishes that the four drugs do not produce equivalent effects on patient reaction times. As previously noted, further [post-hoc tests](#) would be the necessary next step to identify the specific nature of these differences and determine which drug regimens are statistically superior or inferior to others.

## Beyond the Basics: Advanced RM-ANOVA Considerations

Successfully executing and interpreting the one-way repeated measures ANOVA is a fundamental achievement in quantitative data analysis. However, advanced statistical practice requires awareness of potential complications, particularly the assumption of [sphericity assumption](#). Sphericity relates to the equality of variances of the differences between treatment levels, and its violation is common in RM-ANOVA. When sphericity is violated, adjustments (such as the

Greenhouse-Geisser or Huynh-Feldt corrections) must be applied to the degrees of freedom to ensure the F-test remains accurate.

For researchers committed to deepening their expertise in within-subjects designs, continued exploration of statistical theory and software capabilities is highly recommended. The [statsmodels](#) library offers comprehensive documentation covering these advanced corrections and a wide range of other statistical tests vital for robust empirical research. By consistently practicing data manipulation with [Pandas](#) and integrating these powerful statistical methods, analysts can strengthen their capacity to conduct rigorous, publishable research using the advanced, open-source tools available in [Python](#).