

How to Perform a Repeated Measures ANOVA by Hand: A Step-by-Step Guide

Authored by
stats writer

March 13, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Perform a Repeated Measures ANOVA by Hand: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=135563>

Understanding the Foundations of Repeated Measures ANOVA

The **repeated measures ANOVA**, often referred to as a within-subjects **analysis of variance**, serves as a sophisticated **statistical** instrument designed to evaluate differences across multiple related groups. Unlike a standard one-way ANOVA, which compares independent groups, this method is specifically tailored for scenarios where the same individuals are measured under different conditions or across multiple time points. By utilizing the same subjects, researchers can effectively control for individual differences that might otherwise obscure the **statistical significance** of the treatment effect. This reduces the amount of unexplained **variance**, thereby increasing the power of the test to detect real differences between the experimental conditions.

Performing these calculations manually provides an invaluable educational perspective on how variance is partitioned within a dataset. While modern **statistical software** can generate results in milliseconds, the manual process reveals the relationship between total variation, treatment effects, and individual differences. This approach requires a meticulous organization of data, typically in a **table** format where rows represent individual subjects and columns represent the various measurements or time intervals. Understanding the mathematical mechanics behind the **F-statistic** allows researchers to better interpret their findings and recognize the limitations of their **experimental design**.

The primary goal of a **repeated measures ANOVA** is to test the **null hypothesis**, which posits that the **means** of all treatment levels are equal. If the resulting **p-value** is less than the chosen **significance level**, usually denoted as **alpha**, the null hypothesis is rejected in favor of the **alternative hypothesis**. This indicates that at least one treatment level differs significantly from the others. In the following sections, we will walk through a comprehensive example involving medical research to demonstrate how to execute these calculations by hand, ensuring each component of the **ANOVA** table is accurately derived.

Experimental Setup: Investigating Drug Reaction Times

To illustrate the step-by-step process of a **repeated measures ANOVA**, consider a clinical scenario where researchers are investigating the efficacy of three distinct pharmacological agents. The objective is to determine whether these different drugs influence the reaction times of patients in a statistically meaningful way. In this study, five unique patients are selected as the test subjects. Each of these five patients is administered all three drugs at different times, and their reaction times are recorded in seconds. This design ensures that the variation between conditions is not caused by differences between the people in each group, as the same people appear in every group.

The data collected from this experiment is crucial for our calculation. Below is the visual

representation of the raw scores obtained during the clinical trial:

Patient	Drug 1	Drug 2	Drug 3
Patient 1	30	28	16
Patient 2	14	18	10
Patient 3	24	20	18
Patient 4	38	34	20
Patient 5	26	28	14

In this dataset, each row corresponds to a single patient, while each column corresponds to one of the three drugs. By observing the reaction times, we can see fluctuating values. Our mission is to determine if the observed fluctuations in the **mean** reaction times for Drug 1, Drug 2, and Drug 3 are due to the actual effects of the drugs themselves or simply due to random **sampling error**. Because the same subjects are tested repeatedly, we must account for the **variance** attributed to individual patient differences, which is the hallmark of the **repeated measures** approach.

This tutorial will guide you through the six essential steps required to perform the analysis. We will calculate the **Total Sum of Squares** (SST), the **Between-Groups Sum of Squares** (SSB), the **Subject Sum of Squares** (SSS), and the **Error Sum of Squares** (SSE). By the end of this exercise, we will have a complete **ANOVA** summary table and the necessary information to reach a definitive statistical conclusion regarding the drugs' impact on reaction speeds.

Step 1: Calculating the Total Sum of Squares (SST)

The initial phase of our analysis involves calculating the **Total Sum of Squares** (SST). This value represents the aggregate amount of variation present across the entire dataset, encompassing every observation across all patients and all drug conditions. Essentially, SST measures how much each individual **data point** deviates from the grand **mean** of all the observations. To find this, we use a formula that relates the **variance** of the entire data pool to the total number of observations minus one.

The mathematical formula for SST is expressed as **SST = s2total(ntotal - 1)**. In this equation, **s2total** refers to the **variance** calculated for the entire set of fifteen observations (5 patients x 3 drugs), while **ntotal** represents the total count of those observations. By calculating the **mean** of all values and then determining the squared deviations of each point from that **mean**, we arrive at the total variation that must eventually be partitioned into its constituent parts.

For our specific drug study example, the **variance** for the entire dataset is calculated to be 64.2667. With a total of 15 observations, our **degrees of freedom** for the total variation is 14 (15 -

1). Multiplying these figures ($64.2667 * 14$) results in an SST of **899.7**. This **899.7** serves as the "total pie" of variation that we will now begin to slice into portions attributed to the drugs, the subjects, and the residual error.

Step 2: Determining the Between-Groups Sum of Squares (SSB)

Once the total variation has been established, we move to **Step 2**, which is the calculation of the **Between-Groups Sum of Squares (SSB)**. This specific metric quantifies the amount of variation that can be attributed to the differences between the treatment conditions--in this case, the three drugs. It helps us understand how much the average reaction time for each drug deviates from the overall average of all drugs combined. If the drugs have very different effects, the SSB will be a large number; if the drugs have similar effects, the SSB will be small.

The formula for calculating SSB is **$SSB = \sum n_j(x_j - x_{total})^2$** . Here, the Greek symbol Σ indicates that we must sum the results for each group. The variable n_j represents the number of observations within a specific group (which is 5 in our example), x_j is the mean of that specific group, and x_{total} is the grand mean of the entire dataset. By squaring the difference between each group's average and the total average, then multiplying by the group size, we isolate the treatment-related variation.

In our experiment, we first determine the means for the three drugs and the grand mean (22.533). Applying the formula, we calculate: $(5)(26.4 - 22.533)^2 + (5)(25.6 - 22.533)^2 + (5)(15.6 - 22.533)^2$. The summation of these three values yields an SSB of **362.1**. This indicates that 362.1 units of our total 899.7 units of variation are directly associated with the differences between the three medications.

Step 3: Isolating the Subject Sum of Squares (SSS)

The unique feature of the repeated measures ANOVA is **Step 3**: calculating the **Subject Sum of Squares (SSS)**. In a standard ANOVA, this variation would be lumped into the error term, but here we can isolate it. SSS measures the consistency of differences between the participants themselves. For instance, some patients might naturally have faster reaction times across all drugs compared to others. By identifying this "subject-level" variation, we can remove it from the error term, making our statistical test more sensitive to the drug effects.

The mathematical formula for SSS is **$SSS = (\sum r^2k / c) - (N^2 / rc)$** . In this notation, $\sum r^2k$ represents the sum of the squared totals for each patient across all conditions. The variable c refers to the number of conditions (3 drugs), r represents the number of patients (5 subjects), and N is the grand total sum of all observations. This calculation effectively determines how much of the total variation is due to the fact that people are inherently different from one another.

For our patient data, we first sum the reaction times for each of the five patients across the three

drugs, square those sums, and add them together. Using the totals of 74, 42, 62, 92, and 68 for each patient, the calculation becomes: $((74^2 + 42^2 + 62^2 + 92^2 + 68^2) / 3) - (3382 / 15)$. Completing this **arithmetic** gives us an SSS of **441.1**. This high value suggests that a significant portion of the total variation in the study is simply due to the inherent differences in speed between the five patients.

Step 4: Quantifying the Error Sum of Squares (SSE)

After calculating the variation for the total dataset, the treatments, and the subjects, we must find the remaining "leftover" variation, known as the **Error Sum of Squares (SSE)**. This is often called the **residual** variation. It represents the noise in the data that cannot be explained by the drugs or the individual differences between patients. In a **repeated measures** context, SSE represents the interaction between the subjects and the treatments--essentially, how differently individuals react to the drugs in ways that aren't consistent across the board.

Finding the SSE is a straightforward process of subtraction. Since we know that the total variation (SST) is composed of treatment variation (SSB), subject variation (SSS), and error variation (SSE), we can use the formula: **SSE = SST - SSB - SSS**. This logical partitioning ensures that all components of the variance are accounted for and that the sum of the individual parts equals the total variation we calculated in the first step of our analysis.

Using the figures derived in the previous steps, we perform the subtraction: 899.7 (SST) - 362.1 (SSB) - 441.1 (SSS). This calculation results in an SSE of **96.5**. This relatively low error value is a good sign for our experiment; it suggests that once we account for the differences between drugs and the differences between people, there is very little unexplained "noise" remaining in our reaction time measurements. This high level of control is what makes the **repeated measures design** so effective in clinical research.

Step 5: Constructing the ANOVA Summary Table

With all the **Sum of Squares** values now calculated, we can proceed to construct the **ANOVA** summary table. This table is a standard format used to organize statistical data so that the **F-statistic** can be easily derived. The table consists of several columns: Source of Variation, Sum of Squares (SS), **Degrees of Freedom** (df), **Mean Squares** (MS), and finally, the F-value.

The **degrees of freedom** are calculated based on the number of groups and subjects. For the "Between" (drug) source, the df is the number of groups minus one ($3 - 1 = 2$). For the "Subject" source, the df is the number of participants minus one ($5 - 1 = 4$). The "Error" df is the product of the drug and subject degrees of freedom ($2 * 4 = 8$). **Mean Squares** are then calculated by dividing each SS value by its corresponding df. Finally, the **F-statistic** is found by dividing the MS Between by the MS Error.

Our completed **ANOVA** table for the drug reaction study appears as follows:

Source	Sum of Squares (SS)	df	Mean Squares (MS)	F
Between	362.1	2	181.1	15.006
Subject	441.1	4	110.3	
Error	96.5	8	12.1	

In this table, the **Mean Square Between** (181.1) represents the average variance explained by the drugs, while the **Mean Square Error** (12.1) represents the average unexplained variance. By dividing 181.1 by 12.1, we obtain an **F-statistic** of **15.006**. This number is the core result of our test, representing the ratio of treatment effect to error. The higher this number is, the more likely it is that the drugs actually caused a difference in reaction times.

Step 6: Interpreting the Final Results

The final step in our manual **repeated measures ANOVA** is to interpret the **F-statistic**. To determine if our calculated value of 15.006 is statistically significant, we must compare it to a "critical value" from the **F-distribution** table. This table provides the threshold that the **F-statistic** must exceed for us to reject the **null hypothesis** at a specific **alpha level**, which is typically set at 0.05 in most scientific research.

To find the correct critical value, we need three pieces of information: the **significance level** ($\alpha = 0.05$), the numerator **degrees of freedom** (df between = 2), and the denominator **degrees of freedom** (df error = 8). Looking at a standard **F-distribution** table with these parameters, we find that the critical value is **4.459**. This is the minimum value our F-statistic would need to reach to suggest that the drug effects were not just a result of random chance.

Because our calculated F-statistic of **15.006** is substantially greater than the critical value of **4.459**, we have sufficient evidence to **reject the null hypothesis**. We can conclude with high confidence that there is a statistically significant difference between the **mean** reaction times across the three drug conditions. This suggests that the pharmacological intervention indeed had an effect on the patients, providing the researchers with the quantitative proof needed to move forward with their study or clinical recommendations.