

Perform a Repeated Measures ANOVA in SAS

Authored by
Mohammed loot

November 1, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Perform a Repeated Measures ANOVA in SAS*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=7653>

The [repeated measures ANOVA](#) (Analysis of Variance) represents a cornerstone of statistical methodology, particularly valuable in experimental psychology, medicine, and social sciences. This technique is specifically engineered to determine whether a [statistically significant difference](#) exists among the means of three or more related groups. What fundamentally distinguishes this approach is its reliance on the **within-subjects design**, where the same individuals are measured multiple times under different conditions or across various time points.

The use of repeated measures designs offers substantial methodological and statistical advantages over independent-samples designs. By testing the same subjects repeatedly, the analysis effectively controls for **inter-individual variability**--the inherent, unique differences between participants that often inflate the error term in traditional ANOVAs. Removing this source of variance significantly enhances the statistical power of the analysis, making it easier to detect genuine treatment effects when they exist.

This comprehensive article serves as an expert guide, providing a precise, step-by-step demonstration of how to execute a rigorous repeated measures ANOVA within the powerful [SAS](#) statistical software environment. We will navigate the essential procedures within SAS to accurately model the crucial **within-subjects factor**. Furthermore, we will focus on the rigorous interpretation of the resulting statistical output, ensuring the conclusions drawn are both valid and meaningful in a research context.

Understanding the Repeated Measures Design and Data Structure

Prior to initiating any statistical computation, researchers must possess a clear conceptualization of the data structure demanded by a repeated measures analysis. Unlike a standard independent-samples ANOVA, where each participant contributes one single, unique data point, the repeated measures framework necessitates that each subject provides data across all levels of the independent variable. This structural difference is key because it allows the statistical model to isolate and remove the variability attributed solely to individual participant characteristics from the overall error term, thereby sharpening the focus on the treatment effect.

The defining feature of this experimental design is the inherent dependence of observations; since the same individuals are measured repeatedly, their scores across conditions are correlated. Failing to properly account for this **correlation among observations** would violate the assumption of independence necessary for standard statistical tests and lead to inflated Type I error rates. This necessity for managing correlation is precisely why specialized analytical tools, such as the **proc glm** (General Linear Model) procedure in SAS, are essential for accurate modeling.

When preparing data for repeated measures analysis in SAS programming, the data is optimally arranged in the **'long' format**. In this configuration, each row represents a distinct observation (e.g., a specific measurement taken after one drug application), rather than representing a single

subject across all conditions. The long format is ideally suited for the powerful [proc glm](#) procedure, as it enables the researcher to designate the Subject identifier as a formal factor within the statistical model. This crucial step effectively acts as a blocking mechanism, allowing the software to partition out and control for the variance that exists purely between subjects.

Step 1: Defining and Creating the Data Structure in SAS

To illustrate the practical application of this method, let us consider a common hypothetical research scenario. A clinical researcher intends to investigate the comparative efficacy of four specific pharmaceutical agents (labeled Drug 1, Drug 2, Drug 3, and Drug 4) on human reaction time. The primary research question is whether these four distinct drug conditions elicit significantly different average reaction times in the patient population. To establish a robust within-subjects design, the researcher enrolls five patients, and the reaction time for every patient is meticulously measured following the sequential administration of all four drugs.

The measured reaction times, recorded in milliseconds, are systematically organized in the provided table. It is paramount to observe the structure: Subject 1 contributes four separate measurements, corresponding exactly to the four levels of the independent variable (Drug condition). This arrangement unequivocally confirms the data adheres to the principles of a **within-subjects design**, where the treatments are repeatedly applied to the same experimental unit.

Patient	Drug 1	Drug 2	Drug 3	Drug 4
Patient 1	30	28	16	34
Patient 2	14	18	10	22
Patient 3	24	20	18	30
Patient 4	38	34	20	44
Patient 5	26	28	14	30

The next critical step involves translating this tabulated data into an executable programmatic structure suitable for the SAS environment. We must define three essential variables: **Subject** (the unique participant identifier), **Drug** (the categorical treatment level, ranging from 1 to 4), and **Value** (the measured reaction time, which functions as the dependent variable). The following SAS code snippet demonstrates the creation of the dataset, designated as `my_data`, utilizing the fundamental `data` and `datalines` statements, establishing the long format required for the analysis:

```
/*create dataset: defining Subject, Drug (treatment level), and Value (reaction time)*/  
data my_data;  
input Subject Drug Value;  
datalines;
```

```
1 1 30
1 2 28
1 3 16
1 4 34
2 1 14
2 2 18
2 3 10
2 4 22
3 1 24
3 2 20
3 3 18
3 4 30
4 1 38
4 2 34
4 3 20
4 4 44
5 1 26
5 2 28
5 3 14
5 4 30
;
run;
```

Step 2: Executing the Repeated Measures ANOVA using PROC GLM

In the SAS statistical suite, the highly versatile and robust procedure for managing analysis of variance designs, particularly those involving complex structures or correlated data like repeated measures, is the **proc glm** (General Linear Model). This procedure is selected because it provides the necessary flexibility to properly specify the subject-level structure, thereby enabling the critical separation of variance attributable to the subjects (which acts as a blocking factor) from the variance directly attributable to the specific treatment effects (the four drug conditions).

The operational syntax for initiating this analysis requires the careful construction of three fundamental statements contained within the **proc glm** block: the `class` statement, the `model` statement, and the concluding `run` statement. The `class` statement plays an essential role by identifying all nominal or categorical variables within the dataset; in this specific analysis, it is absolutely crucial that both `Subject` and `Drug` are explicitly designated as classification variables. Failure to classify `Subject` correctly prevents the procedure from utilizing it as a control factor.

The defining structure is established in the `model` statement, where the relationship between the dependent variable and the independent factors is mathematically specified. We designate `Value` (reaction time) as the variable being predicted, and both `Subject` and `Drug` are entered as independent factors predicting `Value`. By including `Subject` in the model statement alongside the primary treatment factor (`Drug`), we achieve the necessary **variance partitioning**. The total sum of squares is effectively divided into three orthogonal components: variability occurring between subjects, variability attributed to the within-subjects treatment (`Drug`), and the residual error. This sophisticated partitioning mechanism is precisely how [proc glm](#) executes the repeated measures calculation when the input data is configured in the efficient long format.

```
/*perform repeated measures ANOVA using PROC GLM*/
```

```
proc glm data=my_data;  
class Subject Drug;  
model Value = Subject Drug;  
run;
```

Step 3: Analyzing the General Linear Model Output

The successful execution of the [proc glm](#) statement triggers the generation of a comprehensive output report from SAS. For the specific objective of interpreting the repeated measures ANOVA, researchers must concentrate their analytical efforts on the standard [ANOVA table](#). This essential table serves as the summary of the variance decomposition process, providing critical statistics such as the sums of squares (SS), the corresponding degrees of freedom (df), the mean squares (MS), the calculated F statistic, and, most importantly, the associated probability (p-value).

The image presented below illustrates the core ANOVA output generated by the SAS procedure, meticulously detailing the various sources contributing to the variability observed in the reaction time measurements:

The GLM Procedure

Dependent Variable: Value

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	1379.000000	197.000000	20.96	<.0001
Error	12	112.800000	9.400000		
Corrected Total	19	1491.800000			

R-Square	Coeff Var	Root MSE	Value Mean
0.924387	12.31302	3.065942	24.90000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Subject	4	680.8000000	170.2000000	18.11	<.0001
Drug	3	698.2000000	232.7333333	24.76	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Subject	4	680.8000000	170.2000000	18.11	<.0001
Drug	3	698.2000000	232.7333333	24.76	<.0001

When assessing the primary research hypothesis--which concerns the existence of differences in reaction time across the four distinct drug conditions--our analytical focus must be placed exclusively on the row labeled **Drug**. In this specific model setup, the variance attributed to the `Subject` factor is deliberately partitioned out but generally disregarded for the main hypothesis test concerning the treatment effect. The inclusion of the Subject factor serves a highly technical purpose: it ensures that the error term used to test the within-subjects effect (Drug) is purified, having been stripped of the variance caused by stable individual differences. This purification leads to a more precise and statistically powerful F-test for the treatment.

Step 4: Drawing Statistical Conclusions from the P-Value

Statistical inference within the framework of ANOVA relies upon a formal procedure of hypothesis testing. The primary goal is to determine, based on empirical evidence, whether the observed dispersion between the group means is sufficiently large to be confidently attributed to the drug treatment itself, or whether it is merely the result of random sampling variability (chance). For this particular experiment, the hypotheses guiding our decision-making process are rigorously defined as follows:

H0: The Null Hypothesis. The [null hypothesis](#) posits that all population group means--specifically, the mean reaction times under Drug 1, Drug 2, Drug 3, and Drug 4--are statistically equal. This

hypothesis asserts that the type of drug administered has absolutely no systematic effect on reaction time.

HA: The Alternative Hypothesis. The alternative hypothesis asserts that at least one of the group means is statistically different from the others. This implies that the factor of drug type does indeed exert a significant, measurable influence on the patient's reaction time performance.

By carefully scrutinizing the critical components of the ANOVA output table provided by SAS, we extract the two primary statistics essential for making a decision regarding the effect of the **Drug** factor:

The F Statistic (the calculated [F Value](#)) for the Drug effect: **24.76**

The Probability Value (the calculated [p-value](#)) for the Drug effect: **<.0001**

The standard procedure requires comparing the calculated p-value against a predetermined threshold of significance, known as alpha (α), which is conventionally set at 0.05 (or 5%). The fundamental decision rule is unequivocal: if the p-value is found to be less than or equal to α , the researcher must reject the null hypothesis (H_0) in favor of the alternative hypothesis (H_A). In this specific analysis, the resulting p-value for the Drug factor ($<.0001$) is dramatically smaller than the established $\alpha = .05$ threshold. Consequently, we forcefully reject the null hypothesis. This decisive rejection furnishes robust statistical evidence to support the conclusion that the mean reaction time is not constant across the four administered drug treatments. In applied, practical terms, the specific type of drug utilized has a **statistically significant impact** on the patient's measured reaction time.

Advanced Considerations and Further Resources for Repeated Measures ANOVA

While the `proc glm` approach demonstrated here provides a powerful and standard method for balanced repeated measures designs, researchers should be aware of advanced statistical considerations necessary for comprehensive analysis. A crucial assumption underlying the traditional repeated measures ANOVA is the assumption of **sphericity**. Sphericity refers to the condition where the variances of the differences between all pairs of repeated measures levels are equal. When this assumption is violated, the F-ratio becomes positively biased, potentially leading to inflated Type I errors. Researchers often employ Mauchly's test to evaluate sphericity and apply correction factors (like Greenhouse-Geisser or Huynh-Feldt) if a violation is detected.

Furthermore, rejecting the overall null hypothesis only confirms that a difference exists somewhere among the means, but it does not specify which pairs of means differ. To pinpoint these specific differences, researchers must conduct **post-hoc comparisons** or planned contrasts. In SAS, these pair-wise comparisons are typically executed using the dedicated `LSMEANS` statement within

the [proc glm](#) procedure, often adjusted using methods like Bonferroni or Tukey's HSD to control the family-wise error rate.

For research involving longitudinal data, unbalanced designs, or complex covariance structures, the **proc mixed** procedure in SAS often offers superior analytical flexibility compared to the rigid requirements of [proc glm](#). The mixed modeling approach allows for direct modeling of the covariance structure (e.g., autoregressive, compound symmetry) among repeated observations, which can provide a more accurate representation of complex data relationships. Exploring these alternative procedures is highly beneficial for deepening your mastery of within-subjects data analysis.

To further enhance your understanding of the theoretical underpinnings and explore more complex applications of within-subjects designs in SAS, consider exploring the following advanced tutorials and documentation:

Detailed guidance on interpreting the F-ratio and degrees of freedom in [ANOVA](#), including the calculation of mean squares.

Tutorials focused on performing complex post-hoc comparisons using the `LSMEANS` statement in SAS to determine specific mean differences.

A comprehensive introduction to the **proc mixed** procedure, detailing why it is frequently the preferred method for longitudinal and highly complex repeated measures data structures.

By diligently adhering to these steps, you can confidently execute and accurately interpret a repeated measures ANOVA utilizing the powerful [SAS](#) software. This ensures that your research conclusions are statistically sound, methodologically rigorous, and properly contextualized within the field of study.