

# How to Perform a Two-Way ANOVA in Stata: A Step-by-Step Guide

Authored by  
**stats writer**

March 8, 2026

## RECOMMENDED CITATION

stats writer (2026). *How to Perform a Two-Way ANOVA in Stata: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=134499>

## Understanding the Theoretical Foundation of Two-Way ANOVA

The **Two-Way ANOVA** (Analysis of Variance) serves as a cornerstone of frequentist statistics, specifically designed to investigate the influence of two distinct **categorical variables** on a single continuous response variable. Unlike a one-way ANOVA, which only considers a single factor, the two-way approach allows researchers to observe how different independent variables act in isolation and, more importantly, how they interact with one another. This methodology is indispensable in experimental design where multiple conditions are tested simultaneously to understand complex relationships that a simpler model might overlook.

When performing this analysis in **Stata**, the primary objective is to evaluate whether the **mean** of the dependent variable significantly differs across the levels of the factors provided. Stata utilizes the **anova** command to partition the total **variance** in the data into components attributable to the main effects of each factor and the interaction effect between them. By calculating the **F-statistic** for each component, Stata provides a rigorous mathematical framework for rejecting or failing to reject the **null hypothesis** regarding group equality.

The versatility of the Two-Way ANOVA in Stata is further enhanced by its ability to handle both balanced and unbalanced designs. In a balanced design, each combination of factor levels contains the same number of observations, leading to straightforward calculations. However, in real-world clinical or social science research, datasets are often unbalanced. Stata's robust computational engine automatically adjusts for these discrepancies, ensuring that the estimated effects remain reliable. This makes it a preferred tool for professionals who require a comprehensive and efficient way to perform multi-factor analyses.

## Establishing Research Objectives and Hypotheses

Before executing any commands in Stata, it is critical to clearly define the research objectives. The purpose of a two-way ANOVA is twofold: first, to determine if there is a main effect for each independent variable, and second, to determine if there is an **interaction** effect between the two factors. An interaction occurs when the effect of one factor on the response variable depends on the level of the second factor. Identifying these relationships is vital for drawing nuanced conclusions from experimental data.

In the context of the example provided, we seek to understand how medical treatments and pre-existing conditions influence physiological changes. Specifically, we examine how different types of drugs and the category of a patient's disease impact the change in **systolic blood pressure**. By framing the study this way, we can ask whether a specific drug works better for one disease than another, or if the drug's efficacy is consistent regardless of the patient's condition.

The statistical hypotheses for a Two-Way ANOVA involve testing three distinct sets of null

hypotheses. These include: the null hypothesis that the means of all levels of Factor A are equal, the null hypothesis that the means of all levels of Factor B are equal, and the null hypothesis that there is no interaction between Factor A and Factor B. If the **p-value** associated with any of these tests falls below a predetermined threshold (commonly 0.05), the result is considered statistically significant, indicating that the factors do indeed exert a measurable influence on the outcome.

## The Case Study: Analyzing the Systolic Dataset

To illustrate the practical application of these concepts, we will utilize a built-in Stata dataset known as **systolic**. This dataset is a classic example used in biostatistics to demonstrate factorial ANOVA. It contains data for 58 different individuals, providing a sufficient sample size to observe meaningful statistical trends. The structure of the data allows us to model the change in systolic blood pressure as a function of two categorical predictors: the specific drug administered and the type of disease the patient is suffering from.

The variables included in this dataset are as follows:

**Drug:** A categorical variable representing the different medications tested.

**Disease:** A categorical variable representing the classification of the patient's illness.

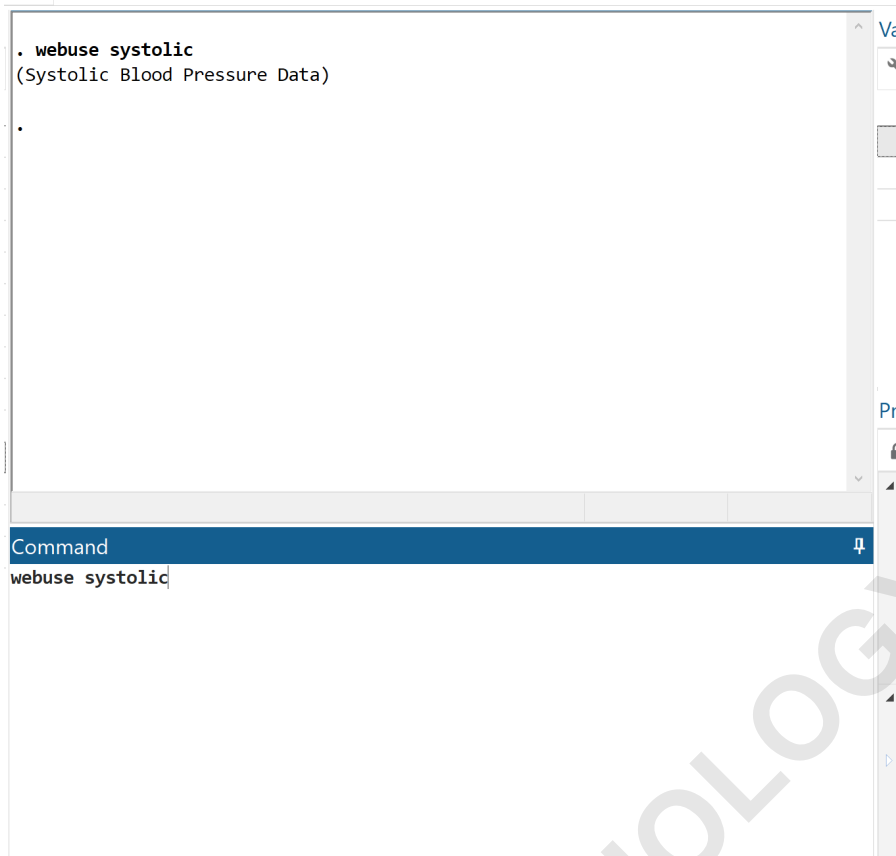
**Systolic:** A continuous **dependent variable** representing the change in blood pressure measurements.

By using this standardized dataset, we can follow a replicable workflow that mirrors professional data analysis practices. The goal is to determine whether the type of drug used and the patient's disease type have a significant impact on blood pressure changes, either independently or through a combined interaction. This step-by-step approach ensures that the researcher understands not just the "how" of the software, but the "why" of the statistical logic.

### Step 1: Data Acquisition and Environment Preparation

The first step in any Stata analysis is the proper loading of the data into the system's memory. Stata provides a convenient way to access example datasets through the **webuse** command. This ensures that the user is working with a clean, pre-formatted version of the data, which is ideal for learning the nuances of the software. To begin, navigate to the command console and enter the instruction to retrieve the file from the official repository.

Specifically, you should type **webuse systolic** into the command box and press Enter. This command prompts Stata to fetch the dataset over the internet and load it into the current session. Once the data is successfully loaded, the variables will appear in the "Variables" pane on the right side of the interface. This sets the stage for the exploratory and analytical steps that follow, ensuring that all necessary parameters are available for manipulation.



It is worth noting that proper **data management** is essential before running any complex model. While the "systolic" dataset is already optimized, in a real-world scenario, you would need to check for missing values, verify variable types, and ensure that the categorical variables are correctly labeled. Stata's command-line interface provides a high degree of control over these preparatory tasks, allowing for a seamless transition from raw data to statistical modeling.

## Step 2: Performing Exploratory Data Analysis

Before proceeding to the formal Two-Way ANOVA calculation, a thorough visual inspection of the raw data is highly recommended. **Exploratory Data Analysis** (EDA) allows the researcher to identify outliers, understand the distribution of values, and get a general sense of the differences between groups. In Stata, the Data Editor provides a spreadsheet-like view of the dataset, which is invaluable for confirming that the data has loaded correctly and reflects the intended experimental design.

To view the data, look at the top menu bar and navigate to **Data > Data Editor > Data Editor (Browse)**. This opens a separate window where you can scroll through the records of all 58 patients. By looking at the "systolic," "drug," and "disease" columns, you can observe the variations in blood pressure changes across different treatment groups. This manual check is a best practice

that helps prevent errors in model specification later in the process.

Data Editor (Browse) - [systolic]

File Edit View Data Tools

systolic[1] 42

	drug	disease	systolic
1	1	1	42
2	1	1	44
3	1	1	36
4	1	1	13
5	1	1	19
6	1	1	22
7	1	2	33
8	1	2	26
9	1	2	33
10	1	2	21
11	1	3	31
12	1	3	-3
13	1	3	25
14	1	3	25
15	1	3	24
16	2	1	28
17	2	1	23
18	2	1	34

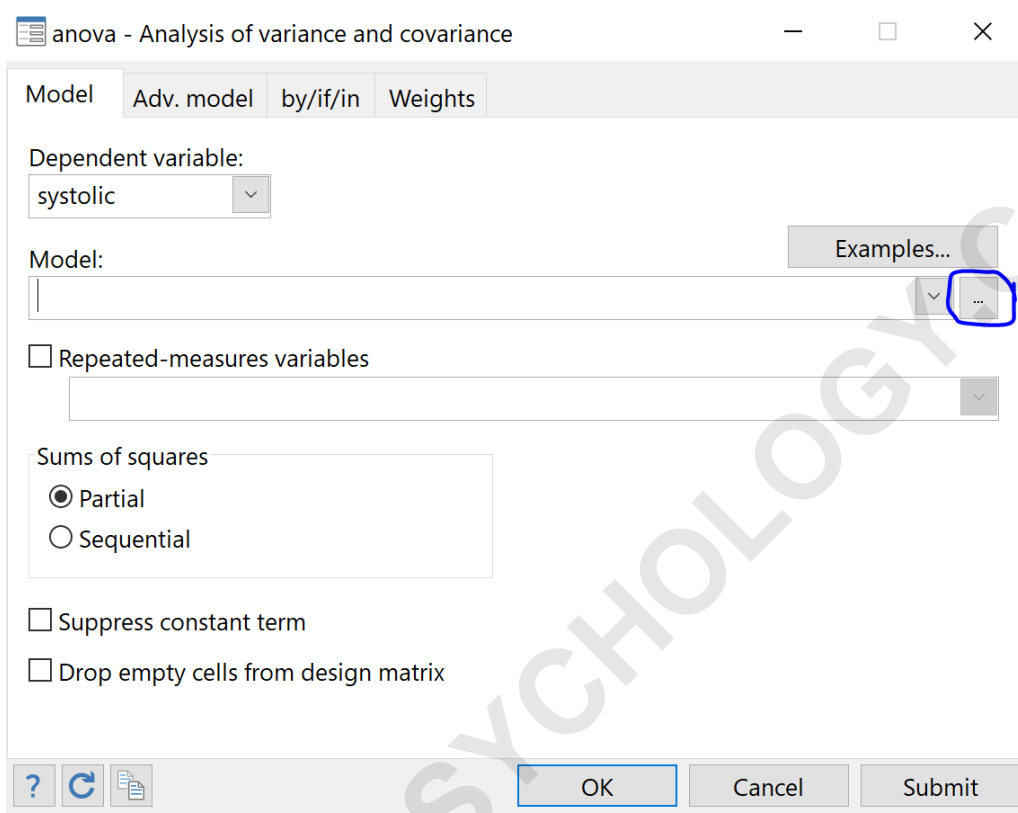
While the browser window is excellent for a quick glance, more advanced users might also use the **summarize** or **tabulate** commands to generate descriptive statistics. For example, tabulating the means of the systolic variable by drug and disease can provide early hints about which factors might turn out to be significant in the final ANOVA table. This combination of visual browsing and summary statistics forms a robust foundation for the formal analysis.

### Step 3: Configuring the Two-Way ANOVA Model

With the data verified, we can now move to the execution of the **Analysis of Variance**. Stata's graphical user interface makes it simple to specify complex models without needing to memorize syntax immediately. To access the ANOVA dialog box, go to the top menu and click on **Statistics > Linear models and related > ANOVA/MANOVA > Analysis of variance and covariance**. This will open a window where you can define the dependent variable and the independent factors.

In the "Dependent variable" dropdown menu, select **systolic**. This identifies the outcome we are

trying to predict. Next, we need to define the model, including the interaction term. Instead of just selecting the variables individually, click the three dots (ellipsis) next to the "Model" dropdown. This opens the "Variable addition" dialog, which allows you to specify the interaction between drug and disease. Using the notation **drug##disease** tells Stata to include the main effect of drug, the main effect of disease, and the interaction effect between the two.



Create varlist with factor variables

Type of variable

Factor variable

Polynomial variable

Continuous variable

Specification:

2-way full factorial

	Variables	c.	Base
Variable 1:	drug	<input type="checkbox"/>	Default
Variable 2:	disease	<input type="checkbox"/>	Default

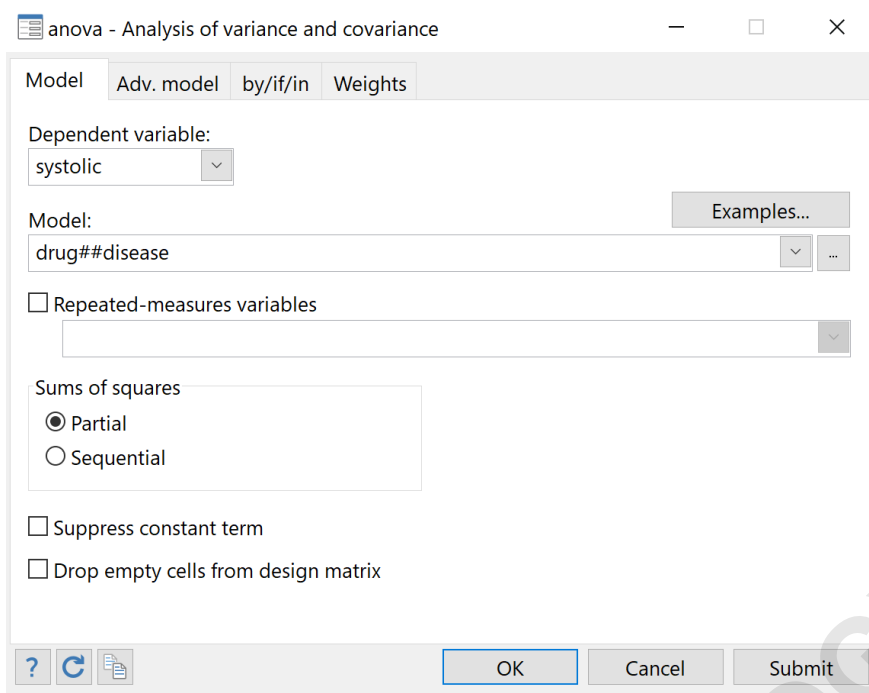
Add to varlist

Varlist:

drug##disease

OK Cancel

Once you have configured the interaction term, the main window will update to show the full model specification. It is important to ensure that the interaction term is present, as omitting it would change the analysis to an additive model, which assumes the factors do not influence each other. After verifying the settings, simply click **OK** to run the command. Stata will then process the data and display the results in the main output window.



#### Step 4: Interpreting the ANOVA Table and Statistical Metrics

The output generated by Stata is a standard ANOVA table, which contains several crucial metrics for **statistical significance**. The most important columns are the "Partial SS" (Sum of Squares), "df" (**Degrees of Freedom**), "MS" (Mean Square), and the "Prob > F" (the p-value). Each row in the table corresponds to a different part of the model, including the overall model, the individual factors, the interaction term, and the residual error.

To interpret the results, we focus primarily on the p-values in the "Prob > F" column. These values tell us the probability of observing our data if the null hypothesis (no effect) were true. In our specific case study, the results provide a clear picture of the influences at play. By examining the table, we can draw specific conclusions about the drugs being tested and the disease categories being studied.

```
. anova systolic drug##disease
```

	Number of obs =	58	R-squared =	0.4560	
	Root MSE =	10.5096	Adj R-squared =	0.3259	
Source	Partial SS	df	MS	F	Prob>F
Model	4259.3385	11	387.21259	3.51	0.0013
drug	2997.4719	3	999.15729	9.05	0.0001
disease	415.87305	2	207.93652	1.88	0.1637
drug#disease	707.26626	6	117.87771	1.07	0.3958
Residual	5080.8167	46	110.45254		
Total	9340.1552	57	163.86237		

Based on the output provided in the example, we can observe the following critical findings:

**Interaction Effect (drug#disease):** The p-value is 0.3958. Since this is much greater than 0.05, we conclude there is no significant interaction between drug and disease.

**Main Effect of Disease:** The p-value is 0.1637, indicating that the patient's disease type does not have a statistically significant effect on the change in systolic blood pressure.

**Main Effect of Drug:** The p-value is 0.0001. This is highly significant, suggesting that the type of drug administered has a profound impact on the outcome variable.

## Step 5: Post-Hoc Analysis and Further Diagnostics

While the initial Two-Way ANOVA tells us that a factor like "Drug" is significant, it does not specify which particular drugs differ from one another. To determine this, researchers often perform **post-hoc tests**, such as the Tukey HSD or Bonferroni adjustment. These tests compare all possible pairs of group means while controlling for the increased risk of Type I errors that comes with multiple comparisons. Stata makes these tests easy to perform immediately after running the main ANOVA command.

Additionally, it is a professional requirement to check the assumptions of the ANOVA model. These include the normality of residuals and the homogeneity of variances (homoscedasticity). Stata provides diagnostic tools like the **predict** command to generate residuals and **ovtest** or **hettest** for variance checks. If these assumptions are violated, the results of the ANOVA might be misleading, and the researcher may need to consider data transformations or non-parametric alternatives.

In cases where a significant interaction is found (unlike our current example), the researcher should not interpret the main effects in isolation. Instead, they should analyze "simple effects," which look at the effect of one factor at each specific level of the other factor. This deeper level of

analysis is what makes the Two-Way ANOVA such a powerful tool for understanding the complexities of experimental data, particularly in fields like medicine and psychology.

## Step 6: Formal Reporting of Statistical Results

The final stage of the analysis is the formal reporting of the findings. In academic and professional settings, it is essential to communicate the results clearly, including the F-statistics, degrees of freedom, and p-values. A well-structured report allows other researchers to verify your conclusions and understand the magnitude of the effects found in the study. Following standard reporting guidelines, such as those from the American Psychological Association (APA), ensures consistency and clarity.

For the analysis conducted in this tutorial, the report would be structured as follows:

A two-way ANOVA was conducted on a sample of 58 individuals to examine the effects of **drug type** and **disease classification** on the change in **systolic blood pressure**. The analysis revealed that there was no statistically significant interaction between the effects of drug and disease on systolic blood pressure ( $F(6, 46) = 1.06, p = 0.3958$ ). Furthermore, the main effect of disease was found to be non-significant ( $F(2, 46) = 1.89, p = 0.1637$ ).

However, the main effect of drug was statistically significant ( $F(3, 46) = 9.05, p = 0.0001$ ), indicating that different medications resulted in different average changes in blood pressure. These results suggest that while the choice of medication is a critical factor in managing systolic blood pressure, the patient's disease category and the interaction between drug and disease did not play a significant role in this specific experimental context. This type of clear, data-driven reporting is the ultimate goal of performing a Two-Way ANOVA in Stata.