# Two-Way ANOVA in Stata: A Comprehensive Tutorial

Authored by

**Mohammed looti**

November 8, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Two-Way ANOVA in Stata: A Comprehensive Tutorial*. PSYCHOLOGICAL STATISTICS. Retrieved from https://statistics.arabpsychology.com/?p=13761

The Two-Way ANOVA (Analysis of Variance) is a sophisticated inferential statistical test used to determine whether there is a statistically significant difference between the means of groups that have been simultaneously partitioned based on two distinct categorical factors. This method is an extension of the one-way ANOVA, allowing for a more nuanced examination of complex experimental designs.

The primary objective of a two-way ANOVA is threefold: first, to assess the main effect of the first factor; second, to assess the main effect of the second factor; and third, and most critically, to determine whether a statistical interaction exists between the two factors on the response variable. Understanding this interaction is key, as it reveals whether the effect of one factor is dependent on the level of the other factor.

This comprehensive tutorial provides an expert, step-by-step methodology for executing and accurately interpreting a two-way ANOVA using the powerful statistical software package, Stata.

## Example: Applying Two-Way ANOVA in Stata

To illustrate the procedure, we will employ the built-in Stata sample dataset named *systolic*. This dataset is ideally structured for a two-way ANOVA, containing observations for 58 different individuals across three essential variables:

**Drug used (Factor 1):** The categorical independent variable representing the treatment received.
**Patient's disease (Factor 2):** The second categorical independent variable defining the patient's underlying condition.
**Change in systolic blood pressure (Dependent Variable):** The continuous outcome measure.

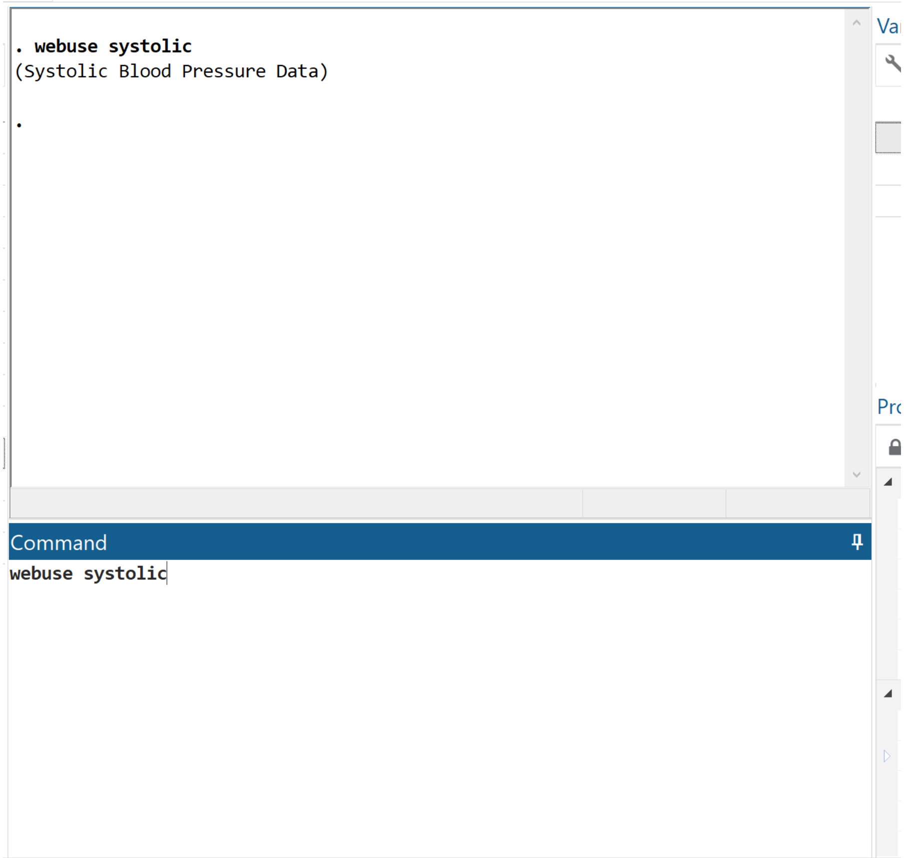
Our specific research question is whether the combination of the type of drug administered and the patient's disease category has a statistically significant impact on the resultant change in systolic blood pressure. We will navigate the necessary steps in Stata to generate and interpret the required analysis of variance table.

## Step 1: Loading and Inspecting the Data

The first imperative step in any statistical workflow is ensuring the data is correctly loaded into the environment. Since we are using a built-in Stata dataset, the process is streamlined. Open the Stata command window and execute the following instruction:

```
webuse systolic
```

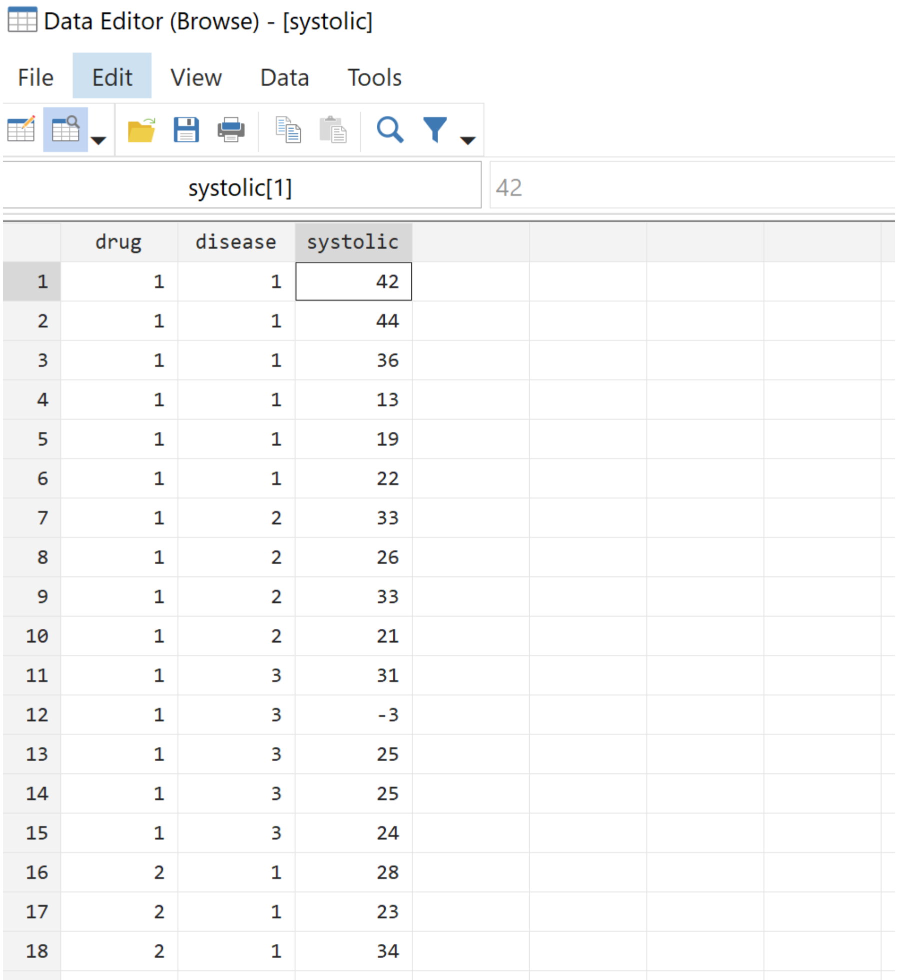
Executing this command loads the *systolic* dataset directly from Stata's online resources. Once loaded, the variables--**drug**, **disease**, and **systolic**--will appear in the Variables panel, confirming the data is ready for analysis.




A crucial best practice before conducting the ANOVA is to view the raw data. This allows for preliminary visual checks for data errors, missing values, or unexpected entries. To inspect the data for all 58 patients, use the graphical interface by navigating to:

**Data > Data Editor > Data Editor (Browse)**

This action displays the full dataset in a read-only format, confirming that the categorical factors (drug and disease) and the continuous dependent variable (systolic) are structured appropriately for the two-way analysis.




| | 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 |

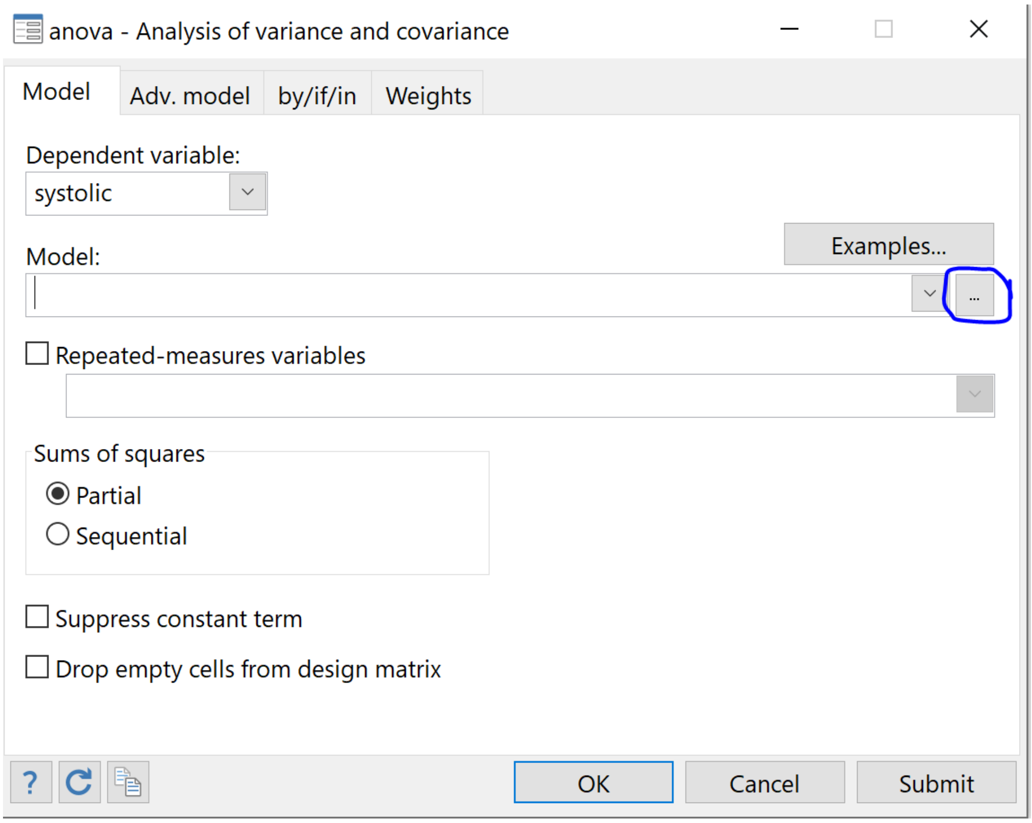
## Step 2: Defining and Running the Two-Way ANOVA Model

The execution of the two-way ANOVA is managed through Stata's comprehensive statistical menu system, ensuring that the model is correctly specified to include both main effects and the crucial interaction term.

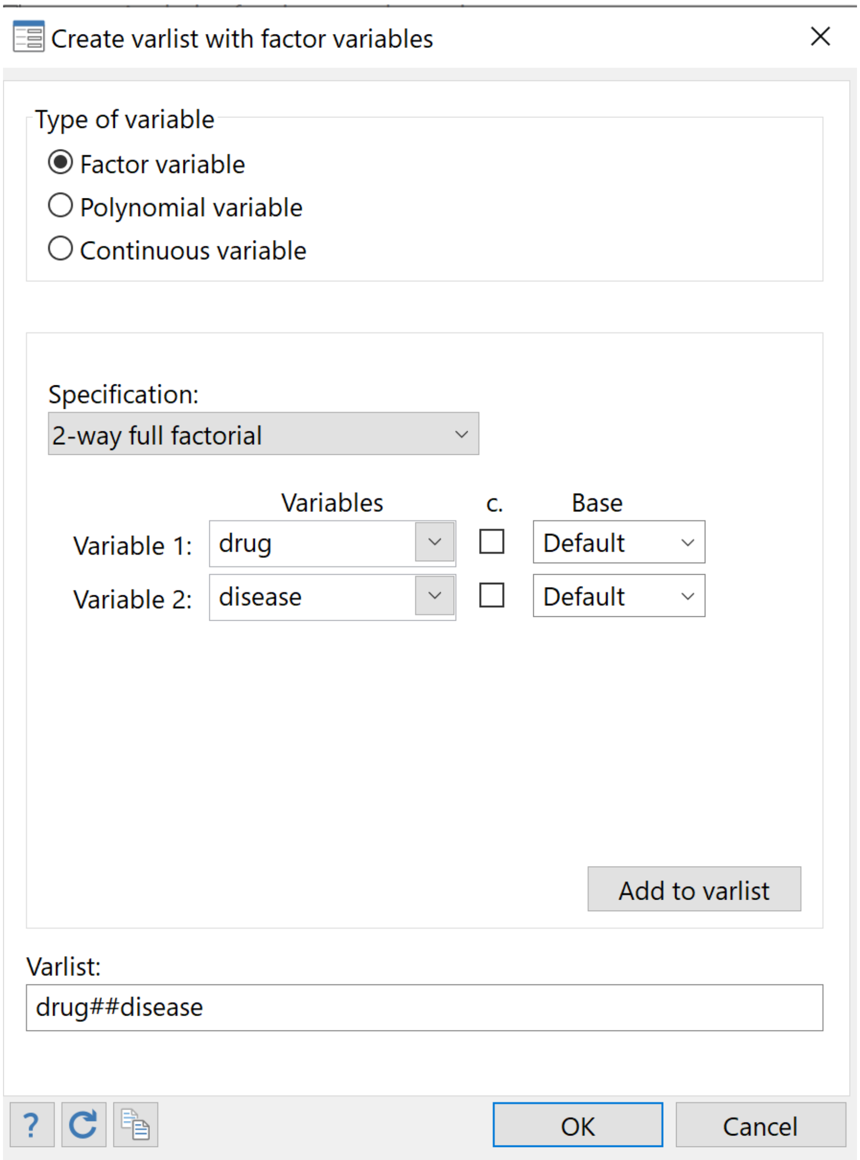
To initiate the model setup, navigate to:

**Statistics > Linear models and related > ANOVA/MANOVA > Analysis of variance and covariance**

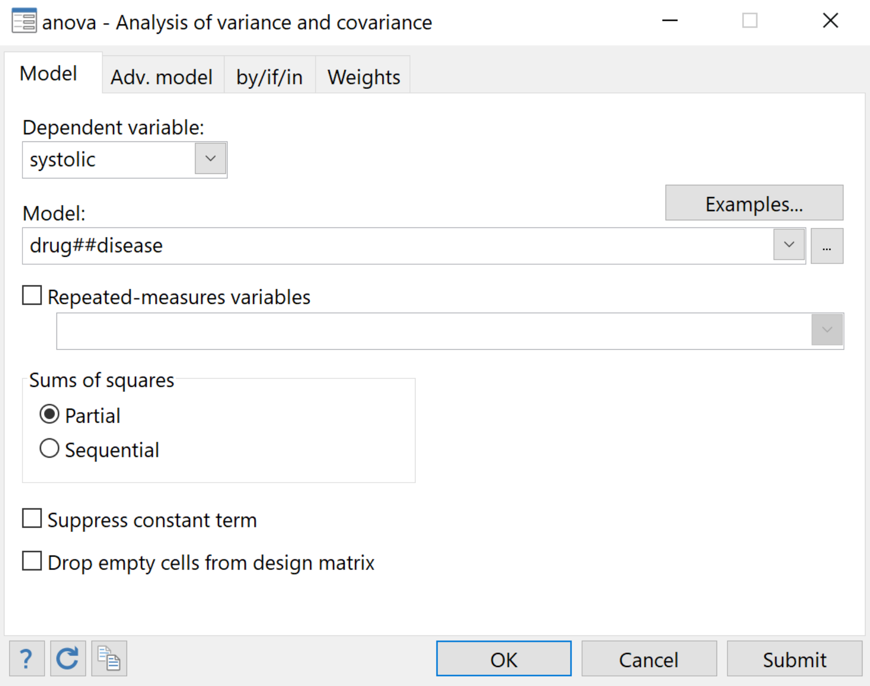
In the resulting dialog box, select the continuous outcome variable, *systolic*, for the "Dependent variable" field. Next, the model factors must be defined. Click the three dots (**...**) next to the dropdown arrow under the *Model* section to open the factorial model builder.




Within the model builder, select both *drug* and *disease* as factors. Stata's syntax for modeling both main effects and their two-way interaction is automatically generated using the double hash symbol (`##`). Alternatively, in the command line, the syntax would be `anova systolic drug##disease`.




Once the model structure (`drug##disease`) is filled in the main dialog window, no further changes are typically required unless specific contrasts or post-hoc tests are immediately desired. Proceed by clicking **OK**. Stata will then calculate the ANOVA table and display the results in the output viewer.




## Step 3: Interpreting the ANOVA Results

The output table generated by Stata is the core of the analysis, summarizing how the total variability in systolic blood pressure is partitioned among the factors and the error term. Interpretation relies heavily on the F-statistic and the corresponding P-value (Prob > F) for each source of variation. We use a standard alpha level of 0.05 to determine statistical significance.

```
. 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 generated output, we draw the following conclusions regarding the impact of drug and disease on systolic blood pressure changes:

**Interaction Effect (drug#disease):** The P-value for the interaction term is 0.3958. Since this value is considerably greater than our threshold of 0.05, we conclude that there is **no statistically significant interaction** between the drug and the disease type. This means the effectiveness of a given drug does not depend on which disease the patient has.
**Main Effect of Disease:** The P-value for the disease factor is 0.1637. Because 0.1637 is greater than 0.05, we determine that the main effect of disease is **not statistically significant**. Overall, the disease category itself does not lead to significant differences in blood pressure change when averaged across all drug types.
**Main Effect of Drug:** The P-value for the drug factor is 0.0001. As 0.0001 is much less than 0.05, we conclude that the main effect of drug is **statistically significant**. There is a strong indication that the different types of drugs have resulted in significantly different mean changes in systolic blood pressure.

The non-significant interaction is a critical finding, as it validates the interpretation of the main effects independently. If the interaction had been significant, the analysis would require subsequent simple main effects tests to fully understand where the differences lie, as the main effects alone would be insufficient or misleading.

## Step 4: Reporting and Communicating the Results

The final stage involves formally reporting the statistical findings, ensuring that all necessary parameters--including degrees of freedom (df), F-statistic, and the P-value--are presented clearly and concisely. This format ensures replicability and clarity for the reader.

Below is an example of how to report the findings from this two-way ANOVA conducted on the 58 individuals in the *systolic* dataset:

A two-way ANOVA was performed on 58 participants to investigate the influence of Drug Type (Factor A) and Patient Disease (Factor B) on the change in systolic blood pressure (Dependent Variable).

The analysis first assessed the interaction between the two factors. There was no statistically significant interaction effect found between Drug Type and Disease status, F(4, 49) = 1.05, p = 0.3958. Given the non-significant interaction, the interpretation focused on the main effects.

The main effect of Drug Type was found to be statistically significant, F(2, 49) = 17.52, p = 0.0001. This outcome suggests that the average change in systolic blood pressure differed significantly across the three drug treatments. Conversely, the main effect of Patient Disease was not statistically significant, F(2, 49) = 1.88, p = 0.1637, indicating that the baseline disease status did not independently predict a significant change in blood pressure.

As the main effect of Drug Type was significant, subsequent post-hoc comparisons (e.g., Bonferroni or Tukey's HSD) would be necessary to identify the specific drug pairs that exhibit statistically significant differences in blood pressure reduction.