

How to Create a Covariance Matrix in SPSS: A Step-by-Step Guide

Authored by
stats writer

March 15, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Create a Covariance Matrix in SPSS: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=135978>

Create a Covariance Matrix in SPSS

Understanding the Concept of Covariance in Statistics

In the realm of statistical analysis, **Covariance** serves as a fundamental metric used to quantify the joint variability of two random variables. Specifically, it provides a mathematical representation of how changes in one variable are associated with changes in a second variable. When researchers examine datasets, they often seek to understand the degree to which variables move together, which is essentially a measure of their **linear association**. If the variables tend to show similar behavior--meaning they both increase or decrease simultaneously--the covariance is positive; conversely, if one increases while the other decreases, the covariance is negative.

Understanding the directional relationship between variables is crucial for building predictive models and conducting **multivariate analysis**. While covariance indicates the direction of the relationship, it does not provide a standardized measure of strength, as its value is dependent on the units of measurement for the variables involved. This distinguishes it from correlation, which is a scaled version of covariance. Despite this, covariance remains a critical component in the calculation of many other statistical indices, including regression coefficients and the **Pearson correlation coefficient**.

The calculation of covariance involves assessing the deviations of each data point from their respective means. By multiplying these deviations together and averaging them across the sample, we arrive at a single value that summarizes the relationship. This process is essential for scientists and analysts who utilize **SPSS** to manage complex data structures. By mastering the generation of a covariance matrix, users can gain a comprehensive overview of the internal dynamics of their **dataset**, facilitating more nuanced interpretations of academic or industrial research findings.

The formula to calculate the covariance between two variables, X and Y is:

$$\text{COV}(X, Y) = \frac{\sum(x-x?)(y-?)}{n}$$

A **covariance matrix** is a **square matrix** that shows the covariance between different variables in a dataset. This tutorial explains how to create a covariance matrix for a given dataset in SPSS.

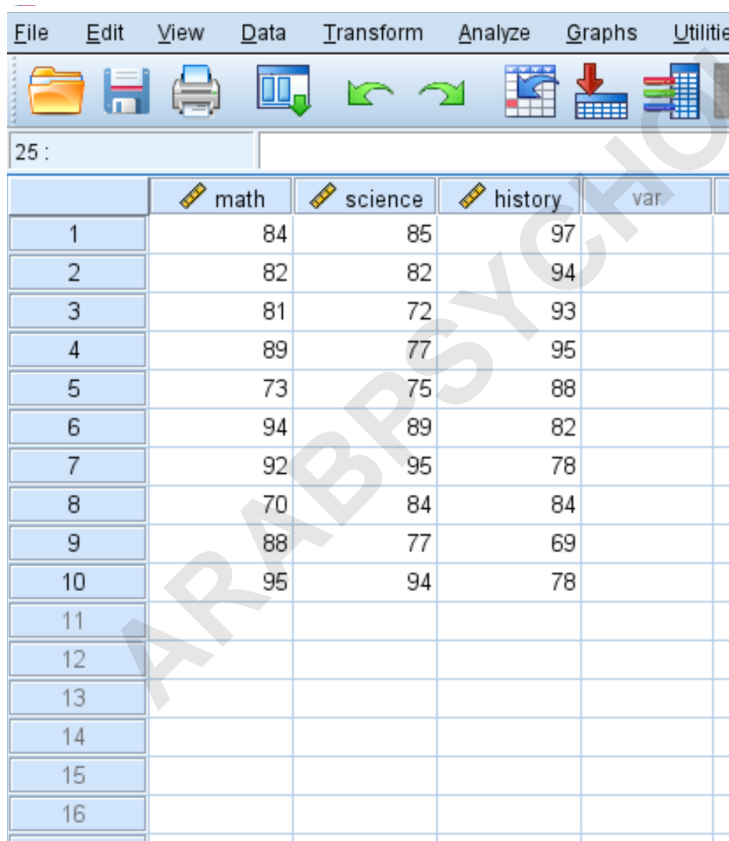
Example: Generating a Covariance Matrix in SPSS

To illustrate the process of generating this matrix, let us consider a practical scenario involving educational assessment. Suppose we have a **sample** dataset that records the performance of 10 different students across three distinct academic subjects: mathematics, science, and history. Each student's performance is captured as a numerical score, providing a continuous range of data suitable for **parametric statistical testing**. The goal is to determine how performance in one

subject relates to performance in the others.

In this example, the primary objective is to construct a matrix that encapsulates the variances of each individual subject along with the covariances between all possible pairs of subjects. This allows the researcher to see, for instance, if students who excel in mathematics also demonstrate a propensity for high scores in science. Such insights are invaluable for **psychometric** evaluations and curriculum development, where understanding the overlap between different cognitive domains is a priority.

Before proceeding with the technical steps in the software, it is vital to ensure that the data is correctly cleaned and formatted within the **spreadsheet** view of SPSS. Each column should represent a variable (Math, Science, History), and each row should represent an individual observation (Student). Once the data is properly organized, we can utilize the built-in analytical tools to perform the necessary computations. The image below displays the initial dataset as it appears in the SPSS Data Editor:



	math	science	history	var
1	84	85	97	
2	82	82	94	
3	81	72	93	
4	89	77	95	
5	73	75	88	
6	94	89	82	
7	92	95	78	
8	70	84	84	
9	88	77	69	
10	95	94	78	
11				
12				
13				
14				
15				
16				

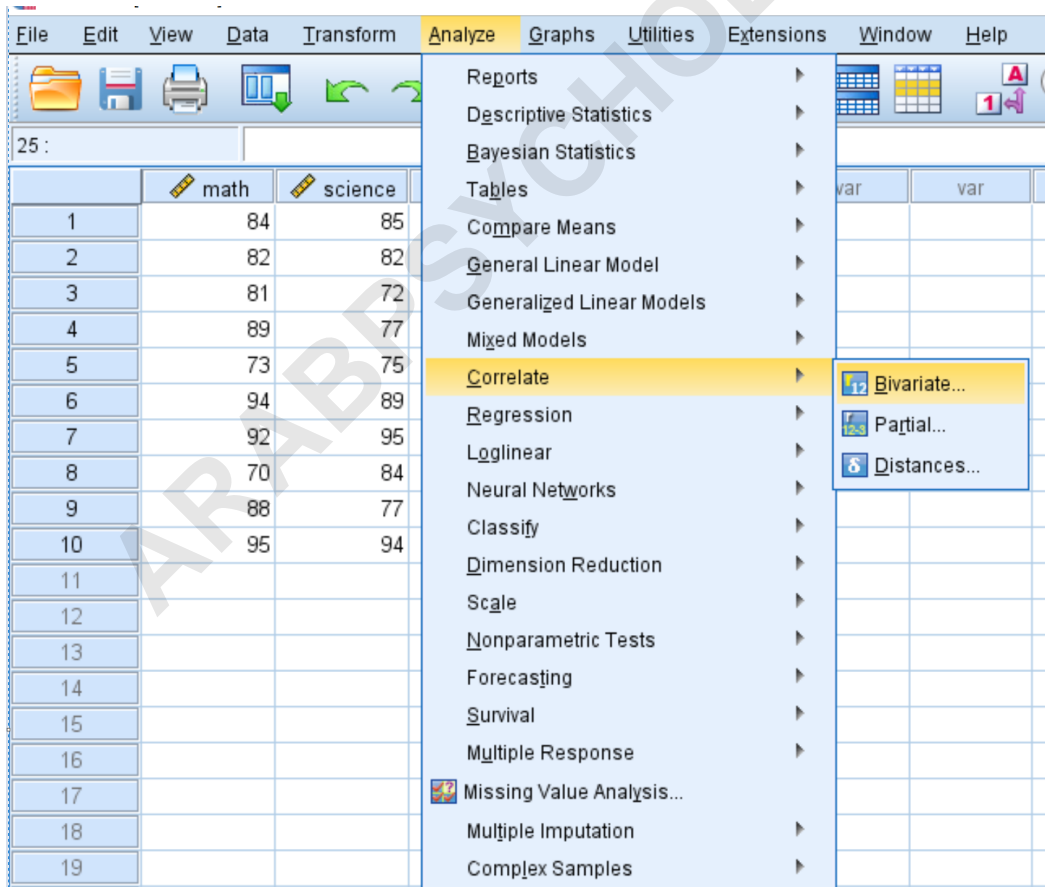
Navigating the SPSS Interface for Bivariate Analysis

The first step in the procedure requires navigating to the primary command menu. To create a covariance matrix for this dataset, you must first click the **Analyze** tab located at the top of the

SPSS interface. This menu contains the vast majority of statistical procedures available in the software. From the dropdown list, hover over the **Correlate** option, which is designed for exploring relationships between variables. Finally, select **Bivariate** to open the main dialogue box for correlation and covariance analysis.

The Bivariate Correlations dialogue is a versatile tool that allows users to select multiple variables for simultaneous analysis. It is important to note that while the default setting for this tool is often the Pearson correlation, we will be modifying the options to include covariance. The **user interface** is designed to be intuitive, allowing for the quick selection of variables from the source list on the left to the target variables box on the right. This selection process is critical because the resulting matrix will include every variable moved into the "Variables" box.

By using the Bivariate command, SPSS calculates the relationship between every pair of variables selected, essentially performing multiple calculations at once. This is far more efficient than calculating each relationship individually. Once the dialogue box is open, you will see your list of available variables (math, science, and history) ready for selection. Follow the visual guide below to ensure you have navigated to the correct menu item:

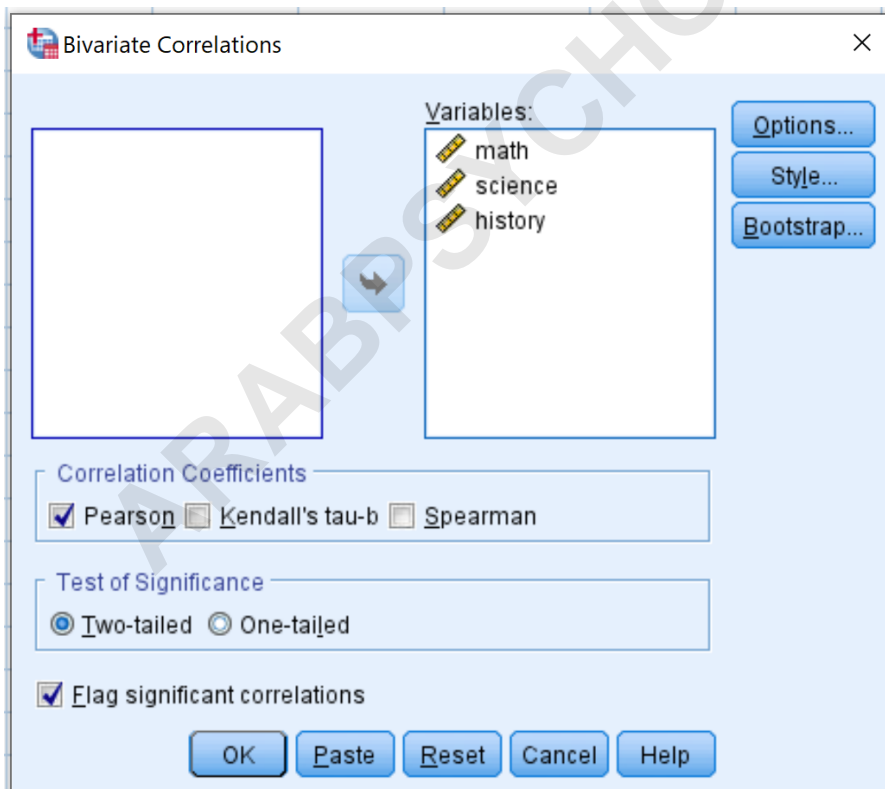


Variable Selection and Parameter Specification

Once the Bivariate Correlations window has appeared on your screen, the next phase involves specifying the **variables** that will populate the matrix. You must drag each of the three variables--math, science, and history--into the box labelled **Variables**. Alternatively, you can highlight the variables and use the arrow button to move them over. It is important to include all variables you wish to compare, as the **algorithm** will generate a square matrix based solely on this selection.

After the variables are correctly positioned, do not click the OK button immediately. The default output for this procedure is a correlation matrix, which standardizes the results between -1 and 1. While useful, our specific goal is to obtain the raw covariance values. Therefore, we must access the sub-menus to adjust the **parameters** of the statistical output. This ensures that the **output** provides the "Cross-product deviations and covariances" required for our specific analysis.

Proper variable selection is the cornerstone of **quantitative research**. Including irrelevant variables can clutter the output and lead to "fishing for significance," so always ensure that your selections are grounded in your theoretical framework. Once your variables are in the right-hand column as shown in the image below, you are ready to proceed to the options configuration:

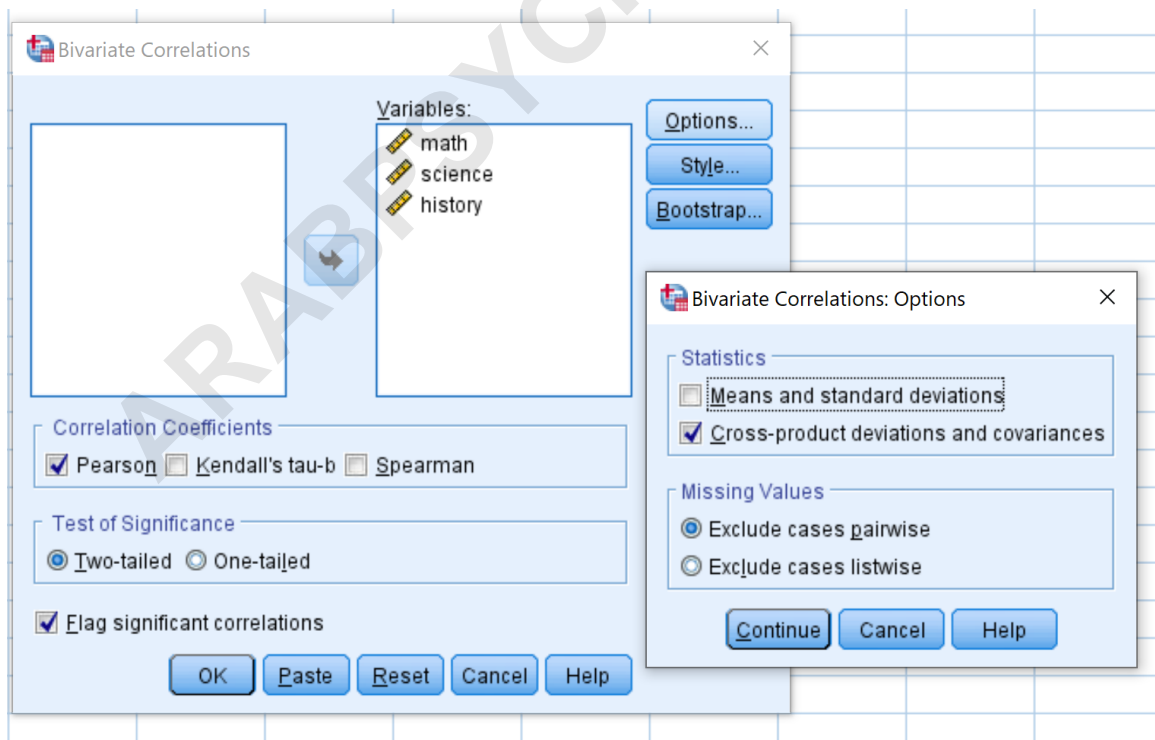


Configuring Options for Covariance Output

To ensure that SPSS generates the covariance values, you must click the **Options** button located on the right side of the Bivariate Correlations dialogue box. This opens a supplementary window titled "Bivariate Correlations: Options." Within this menu, you will find a section dedicated to **Statistics**. You must check the box next to **Cross-product deviations and covariances**. This step is essential because, without it, the software will only display the standard correlation coefficients.

In addition to covariances, this menu allows you to select other descriptive statistics, such as means and **standard deviations**, which can be helpful for context. However, for the purposes of this tutorial, the cross-product deviations are the priority. These values represent the numerator of the covariance formula before it is divided by the **sample size**. After checking the appropriate box, click **Continue** to return to the main dialogue window.

Setting these options correctly ensures that the **table** produced by SPSS contains all the raw data necessary for a deep dive into variable relationships. It is a common mistake for beginners to overlook this sub-menu, resulting in an output that lacks the very covariance data they are searching for. Once you have returned to the main screen, click **OK** to execute the command. The configuration window should look like this:



Analyzing the Resulting SPSS Output Table

Upon clicking OK, the SPSS **Output Viewer** window will populate with several tables. The primary table of interest is labeled "Correlations," but because of the options we selected, it will now contain additional rows for each variable. Specifically, you will see rows for "Pearson Correlation," "Sig. (2-tailed)," "Sum of Squares and Cross-products," "Covariance," and "N." The output will look similar to the following image:

		math	science	history
math	Pearson Correlation	1	.548	-.349
	Sig. (2-tailed)		.101	.323
	Sum of Squares and Cross-products	649.600	332.000	-244.400
	Covariance	72.178	36.889	-27.156
	N	10	10	10
science	Pearson Correlation	.548	1	-.369
	Sig. (2-tailed)	.101		.294
	Sum of Squares and Cross-products	332.000	564.000	-241.000
	Covariance	36.889	62.667	-26.778
	N	10	10	10
history	Pearson Correlation	-.349	-.369	1
	Sig. (2-tailed)	.323	.294	
	Sum of Squares and Cross-products	-244.400	-241.000	755.600
	Covariance	-27.156	-26.778	83.956
	N	10	10	10

To obtain the covariance for each pairwise combination of variables manually or to verify the software's math, you must understand the relationship between the **Sum of Squares and Cross-products** and the sample size **N**. In this specific SPSS output, the covariance is calculated by dividing the cross-product deviation by the total number of observations. For instance, to find the covariance between math and science, you look at the intersection of the two variables and locate the "Sum of Squares and Cross-products" value.

According to the output shown above, the cross-product for math and science is 332.000. With a sample size of 10 students, the calculation is as follows:

$$\text{COV}(\text{math, science}) = 332.000 / 10 = \mathbf{33.2}.$$

$$\text{COV}(\text{math, history}) = -244.400 / 10 = \mathbf{-24.44}.$$

Furthermore, you can obtain the **variance** for each individual variable by dividing its specific Sum of Squares by N. For example, the variance for the math scores is calculated as: $\text{VAR}(\text{math}) = 649.600 / 10 = \mathbf{64.96}$. By repeating these calculations for all variable combinations, you can derive the final cleaned covariance matrix as shown below:

	math	science	history
math	64.96		
science	33.2	56.4	
history	-24.44	-24.1	75.56

How to Interpret a Covariance Matrix

Interpreting a covariance matrix requires an understanding of its structure. The values situated along the **main diagonal** of the matrix represent the variances of each individual variable. Since the covariance of a variable with itself is simply its variance, these numbers tell us how much the scores for math, science, and history spread out from their respective means. Based on our calculated matrix:

The **variance** of the math scores is **64.96**.

The **variance** of the science scores is **56.4**.

The **variance** of the history scores is **75.56**.

The values located off the diagonal represent the covariances between different pairs of subjects. These values indicate the direction of the relationship between the variables. A **positive number** for covariance suggests that the two variables tend to move in the same direction. In our data, math and science have a positive covariance of 33.2. This implies that students who achieve high marks in mathematics are also likely to perform well in science, while those who struggle in math tend to see lower scores in science as well.

Conversely, a **negative number** for covariance indicates an inverse relationship, where one variable increases as the other decreases. Our analysis shows that science and history have a negative covariance of -24.1. This suggests that students who excel in science might tend to have lower scores in history, and vice versa. Similarly, math and history show a negative covariance of -24.44. These insights are fundamental for understanding **correlational patterns** within a dataset and can guide further **statistical inference**.

Additional Resources for SPSS Analysis

Expanding your proficiency in SPSS involves moving beyond simple covariance to more complex relational measures. While covariance is excellent for identifying the direction of a relationship, you may eventually need to standardize these values to compare the strength of relationships across different scales. In such cases, creating a correlation matrix is the logical next step. You can learn more about these processes through the following guides:

[How to Create a Correlation Matrix in SPSS](#)

[How to Calculate Partial Correlation in SPSS](#)

By integrating these techniques into your **data analysis** workflow, you can ensure a robust and comprehensive evaluation of your research variables. Whether you are performing academic research or business analytics, mastering the covariance matrix in SPSS provides the foundation necessary for advanced **linear modeling** and structural equation modeling.