

How to Create a Covariance Matrix in Google Sheets

Authored by
stats writer

December 20, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Create a Covariance Matrix in Google Sheets*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=108196>

The **Covariance Matrix** is an indispensable statistical tool for analyzing the linear relationships between multiple **variables** within a dataset. Utilizing **Google Sheets**, constructing this matrix is a highly efficient and straightforward process, accessible even to novice data analysts. The fundamental workflow involves meticulous data entry, followed by the application of the dedicated covariance functions available in Sheets (such as the COVARIANCE.P or COVAR functions) to calculate the pairwise associations. Once these **covariance** values are determined, they are organized into a square matrix structure. This organized format allows data professionals to swiftly interpret the direction and magnitude of the linear relationships between every pairing of **variables**, providing crucial insights into data behavior.

The concept of **Covariance** serves as the foundation for the matrix. It is a statistical measure quantifying how changes in one variable are associated with changes in a second variable. More precisely, **covariance** determines the degree to which two variables exhibit a linear association, meaning they move together either positively (both increase or both decrease) or negatively (one increases while the other decreases).

Understanding Covariance: The Core Statistical Concept

In practical applications across finance, engineering, and social sciences, **covariance** helps us understand dependency structures within complex systems. A high positive covariance suggests that larger values of one variable tend to consistently correspond with larger values of the other. For example, if advertising spend and sales revenue have a high positive covariance, it indicates that increasing the budget generally leads to higher sales figures.

Conversely, a large negative **covariance** indicates an inverse relationship. If a stock's price and the unemployment rate show negative covariance, it suggests that as unemployment rises, the stock price tends to fall. It is critical to note that while covariance measures the direction of the relationship, its raw magnitude is dependent on the measurement units of the variables involved. Therefore, covariance alone cannot be used as a standardized measure of relationship strength; for standardization, the calculation must be converted into correlation.

The underlying mathematical definition of population covariance between two variables, X and Y , is derived by summing the product of the deviations of each data point from its respective mean, divided by the number of observations (n). This formula is essential to understanding what the Sheets function calculates behind the scenes:

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

The Architecture of the Covariance Matrix

A **covariance matrix** (often represented as Σ in statistical literature) is a highly organized, square matrix designed to simultaneously display the covariances between all possible pairings of variables within a multivariate dataset. If a dataset contains k variables (e.g., three subjects: Math, Science, History), the resulting matrix will be of size $k \times k$ (a 3x3 matrix).

This structured approach is invaluable because it centralizes all relational data into one object, providing a comprehensive, holistic view of the dataset's internal dependencies. It also inherently manages the redundancy of pairwise calculations. For instance, the covariance between Math and Science is mathematically identical to the covariance between Science and Math. Consequently, the **covariance matrix** is always symmetric around its main diagonal (the element at row i , column j is equal to the element at row j , column i).

The principal diagonal of the matrix holds the **variance** of each individual variable (which is mathematically the covariance of a variable with itself). The off-diagonal elements are the values that truly reflect the relationship between different pairs of variables. This structure is a prerequisite for many multivariate statistical techniques, including portfolio optimization in finance and feature engineering in machine learning.

Preparing Your Data in Google Sheets

To create the matrix in **Google Sheets**, data organization is paramount. We will use the provided dataset, which tracks the test scores of 10 students across three subjects: Math, Science, and History. Each variable must occupy its own column, ensuring consistent arrangement for calculation.

The structure of the data requires that we define clear ranges for each variable. For instance, if Math scores are in Column B (B2:B11), Science scores are in Column C (C2:C11), and History scores are in Column D (D2:D11), these ranges will be the inputs for our covariance function. This organized input structure guarantees that the covariance calculation correctly pairs the corresponding scores for all 10 students when comparing one subject to another.

Below is the sample dataset utilized for this example, showing the required column structure:

fx |

| | A | B | C | D |
|----|-------------|----------------|----------------|---|
| 1 | Math | Science | History | |
| 2 | 84 | 85 | 97 | |
| 3 | 82 | 82 | 94 | |
| 4 | 81 | 72 | 93 | |
| 5 | 89 | 77 | 95 | |
| 6 | 73 | 75 | 88 | |
| 7 | 94 | 89 | 82 | |
| 8 | 92 | 95 | 78 | |
| 9 | 70 | 84 | 84 | |
| 10 | 88 | 77 | 69 | |
| 11 | 95 | 94 | 78 | |
| 12 | | | | |
| 13 | | | | |
| 14 | | | | |
| 15 | | | | |
| 16 | | | | |
| 17 | | | | |

Given that we have three variables, the resulting **covariance matrix** will be 3x3. It is best practice to label the rows and columns of your matrix output clearly (e.g., placing "Math," "Science," and "History" as row and column headers) to ensure accurate interpretation later.

Implementing the COVAR Function in Google Sheets

In **Google Sheets**, the function used for calculating the population covariance is **COVAR()**, which is functionally equivalent to COVARIANCE.P. Since Sheets does not have a dedicated multivariate statistics package to generate the entire matrix automatically, we must calculate each of the nine necessary elements individually by entering the appropriate formula into the corresponding cell.

The COVAR function requires two input ranges, representing the two variables whose **covariance** you wish to determine. The syntax remains simple and direct:

COVAR(data_range_y, data_range_x)

For efficiency, when building the matrix, using absolute cell references (e.g., `B2:B11`) for the data arrays is highly recommended. This allows the formulas to be dragged across the 3x3 grid without the data ranges shifting incorrectly. You will define the first array (`data_range_y`) based on the row header and the second array (`data_range_x`) based on the column header.

Step-by-Step Calculation of the Matrix Elements

We initiate the matrix structure starting in a convenient location, such as cell B15. For the purpose of clarity, we will detail how the formulas are placed across the matrix grid:

Diagonal Elements (Variance): For the Math variance (Cell B15), the formula is `=COVAR(B2:B11, B2:B11)`. This compares the Math column against itself. Similarly, for Science (C16) and History (D17), the formula compares those columns to themselves.

Off-Diagonal Elements (Covariance): For the Math and Science covariance (Cell C15), the formula is `=COVAR(B2:B11, C2:C11)`. This measures how Math scores relate to Science scores. For the Science and Math covariance (Cell B16), the formula is `=COVAR(C2:C11, B2:B11)`. Due to symmetry, these two cells must yield the exact same result.

Completing the Matrix: This process is repeated for all nine cells until the full 3x3 matrix is populated with the corresponding covariance values.

The resulting **covariance matrix** is displayed in cells B15:D17, and the corresponding formulas used to populate these cells are shown in cells B21:D23 below. Observing the formulas in the range B21:D23 confirms the proper use of the COVAR function for each pairwise comparison.

| | A | B | C | D |
|----|--------------------------|------------------------|------------------------|------------------------|
| 1 | Math | Science | History | |
| 2 | 84 | 85 | 97 | |
| 3 | 82 | 82 | 94 | |
| 4 | 81 | 72 | 93 | |
| 5 | 89 | 77 | 95 | |
| 6 | 73 | 75 | 88 | |
| 7 | 94 | 89 | 82 | |
| 8 | 92 | 95 | 78 | |
| 9 | 70 | 84 | 84 | |
| 10 | 88 | 77 | 69 | |
| 11 | 95 | 94 | 78 | |
| 12 | | | | |
| 13 | <i>Covariance Matrix</i> | | | |
| 14 | | Math | Science | History |
| 15 | Math | 64.96 | | |
| 16 | Science | 33.2 | 56.4 | |
| 17 | History | -24.44 | -24.1 | 75.56 |
| 18 | | | | |
| 19 | <i>Formulas</i> | | | |
| 20 | | Math | Science | History |
| 21 | Math | =COVAR(A2:A11, A2:A11) | | |
| 22 | Science | =COVAR(A2:A11, B2:B11) | =COVAR(B2:B11, B2:B11) | |
| 23 | History | =COVAR(A2:A11, C2:C11) | =COVAR(B2:B11, C2:C11) | =COVAR(C2:C11, C2:C11) |
| 24 | | | | |
| 25 | | | | |

Once these formulas are entered and calculated, the matrix is finalized. This 3x3 structure now contains all the necessary relational metrics for detailed statistical investigation.

Interpreting the Results: Diagonals and Variance

Interpretation begins with the main diagonal. These values are particularly important as they represent the inherent variability, or **variance**, of each subject's scores. A variable's **variance** provides insight into how spread out the scores are relative to the average score for that subject.

By examining the diagonal values from our calculated matrix, we find:

The **variance** of the Math scores (Cell B15) is **64.96**.

The **variance** of the Science scores (Cell C16) is **56.4**.

The **variance** of the History scores (Cell D17) is **75.56**.

The subject with the highest **variance**, History, indicates that student performance in this subject is the most diverse, with scores being highly dispersed around the mean. Conversely, Science

scores exhibit the least variability among the three subjects.

| | | | | |
|----|-------------------|-------------|----------------|----------------|
| 12 | | | | |
| 13 | Covariance Matrix | | | |
| 14 | | Math | Science | History |
| 15 | Math | 64.96 | | |
| 16 | Science | 33.2 | 56.4 | |
| 17 | History | -24.44 | -24.1 | 75.56 |
| 18 | | | | |

Interpreting Off-Diagonal Elements: Analyzing Relationships

The off-diagonal elements contain the actual **covariance** values between distinct pairs of subjects. These are the key measures for assessing how the scores relate to one another. We analyze the sign and magnitude of these values.

From the matrix, we extract the three unique covariance pairings:

The covariance between Math and Science scores is **33.2**.

The covariance between Math and History scores is **-24.44**.

The covariance between Science and History scores is **-24.1**.

| | | | | |
|----|-------------------|-------------|----------------|----------------|
| 13 | Covariance Matrix | | | |
| 14 | | Math | Science | History |
| 15 | Math | 64.96 | | |
| 16 | Science | 33.2 | 56.4 | |
| 17 | History | -24.44 | -24.1 | 75.56 |
| 18 | | | | |

The sign of the covariance is the most straightforward indicator of the relationship type:

Positive Covariance (Math and Science): The positive value (**33.2**) signifies a direct relationship. High scores in Math tend to be paired with high scores in Science, and low scores in Math correspond with low scores in Science. This suggests that the skills measured by these two subjects are complementary or draw upon similar student aptitudes.

Negative Covariance (Math/History and Science/History): The negative values (**-24.44** and **-24.1**) indicate an inverse relationship. For example, high scores in Math tend to be associated with lower scores in History. This counter-intuitive finding suggests that students who excel in quantitative disciplines may, within this specific dataset, struggle comparatively in humanities. Negative covariances are critical in finance for understanding portfolio diversification, where assets

that move inversely can reduce overall risk.

Advanced Applications of the Covariance Matrix

While manually calculating the elements in **Google Sheets** is educational, the ultimate utility of the **Covariance Matrix** lies in its application to sophisticated statistical modeling. In multivariate analysis, this matrix is the bedrock for techniques designed to manage complex, correlated data.

One key application is Principal Component Analysis (PCA), which uses the covariance matrix to identify the dimensions of greatest variability (principal components). This allows analysts to reduce the complexity of the data while preserving the maximum amount of information. Furthermore, in machine learning, understanding the covariance between features is crucial for developing accurate models and avoiding multicollinearity, where highly correlated input variables can destabilize the model coefficients.

Mastering the construction of this matrix, even using the basic functions available in **Google Sheets**, is an essential skill that bridges foundational statistics and advanced data science.

For further learning on creating covariance matrices in different platforms, explore these resources:

[How to Create a Covariance Matrix in Excel](#)

[How to Create a Covariance Matrix in R](#)

[How to Create a Covariance Matrix in Python](#)

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