How to Calculate Sxy: A Step-by-Step Guide with Examples

Authored by

stats writer

November 22, 2025

In statistics, **Sxy** is a critical intermediate calculation used to quantify the shared linear variability between two variables, X and Y. Sxy stands for the **Sum of Products of Deviations**. Although Sxy is conceptually linked to both covariance and the correlation coefficient, it is fundamentally the unscaled numerator used in determining the slope of a linear regression line. Unlike the correlation coefficient, Sxy's value is not constrained to the range of -1 to 1, as it is highly dependent on the scale of the original data.

For example, if we were calculating the correlation coefficient (r), we would use the covariance of X and Y, and divide it by the product of the standard deviation of X and the standard deviation of Y. However, **Sxy** itself is merely the first step toward calculating these metrics. If a dataset yields an Sxy of 59, this high positive value immediately suggests a strong, positive relationship, indicating that as X increases, Y tends to increase as well, which is the foundational information required for model fitting.

Understanding the Sxy Statistic

In the realm of descriptive statistics, **Sxy** represents the sum of the product of the differences between the X values and the mean of X, and the differences between the Y values and the mean of Y. This value measures the total co-variation between the two sets of data points, determining the direction (positive or negative) and magnitude of their relationship before any standardization or scaling takes place.

This fundamental calculation is most frequently employed when we are manually determining the parameters required to fit a simple linear regression line to a dataset. Though modern statistical software automates this process, calculating Sxy by hand provides deep insight into the underlying mechanics of least squares estimation, ensuring a thorough understanding of how the slope coefficient is derived from the raw data.

The sign of Sxy is particularly informative: a positive Sxy indicates that X and Y tend to increase or decrease together (positive relationship), whereas a negative Sxy suggests that as X increases, Y tends to decrease (negative relationship). If Sxy is close to zero, it indicates a very weak or non-existent linear relationship.

The Formula for Calculating Sxy

We use the following specific formula to calculate **Sxy**, which is the cornerstone for determining the slope of the regression line:

$Sxy = \Sigma(x_i - x)(y_i - y)$

This equation utilizes several standard statistical notations that require precise definition for accurate calculation. Understanding each component is essential for performing the subsequent steps in linear regression analysis:

Σ: This Greek symbol represents the operation of **summation**, indicating that we must add up all the calculated product terms for every data pair (xi, yi) in the dataset.
xi: This denotes the ith individual observation or value found in the X variable set.
x: This represents the arithmetic **mean** (average) value of all observations in the X variable set.
yi: This denotes the ith individual observation or value found in the Y variable set.
y: This represents the arithmetic **mean** (average) value of all observations in the Y variable set.

The following example shows how to utilize this formula in practice by systematically calculating each deviation and product term for a small sample dataset.

Example: Calculating Sxy by Hand

Suppose we are tasked with fitting a simple linear regression model to a given dataset consisting of paired observations for variables X and Y. This hypothetical dataset contains six pairs of observations, and our first goal is to determine the Sxy value for these six points.

The dataset provided for calculation is visualized below, showing the corresponding values for the independent variable X and the dependent variable Y:

x	y
1	8
2	12
2	14
3	19
5	22
8	21

To calculate Sxy accurately, we must first perform two preliminary steps: determining the mean of X and determining the mean of Y.

Step 1: Calculate the Mean of X (x?)

The calculation of the mean of X involves summing all the individual xi values and dividing this total by the number of observations (n), which is 6 in this case. This central tendency measure serves as the reference point for calculating the deviations of X.

x? = (1 + 2 + 2 + 3 + 5 + 8) / 6 = 21 / 6 = **3.5**

This result, 3.5, represents the sample mean for the X variable. Every subsequent calculation for the X deviation will subtract this value from the observed data point to find the centered deviation.

Step 2: Calculate the Mean of Y (?)

Similarly, the mean value for Y must be calculated to establish the central reference point for the Y deviations. We sum all yi values and divide by the total number of observations (n=6).

? = (8 + 12 + 14 + 19 + 22 + 21) / 6 = 96 / 6 = **16**

The sample mean for the Y variable is 16. These mean values (x?=3.5 and ?=16) are crucial for calculating the product of deviations, which forms the core of the **Sxy** statistic.

Step 3: Calculate the Sum of Products (Sxy)

With both means calculated, we can now proceed to the most critical step: calculating the sum of the products of the deviations, **Sxy**. This involves creating a table to systematically calculate (xi - x?), (yi - ?), and finally, the product of these two deviations for each row. The final **Sxy** value is the sum of this last column.

Using these values, the following screenshot shows how to calculate the value for Sxy:

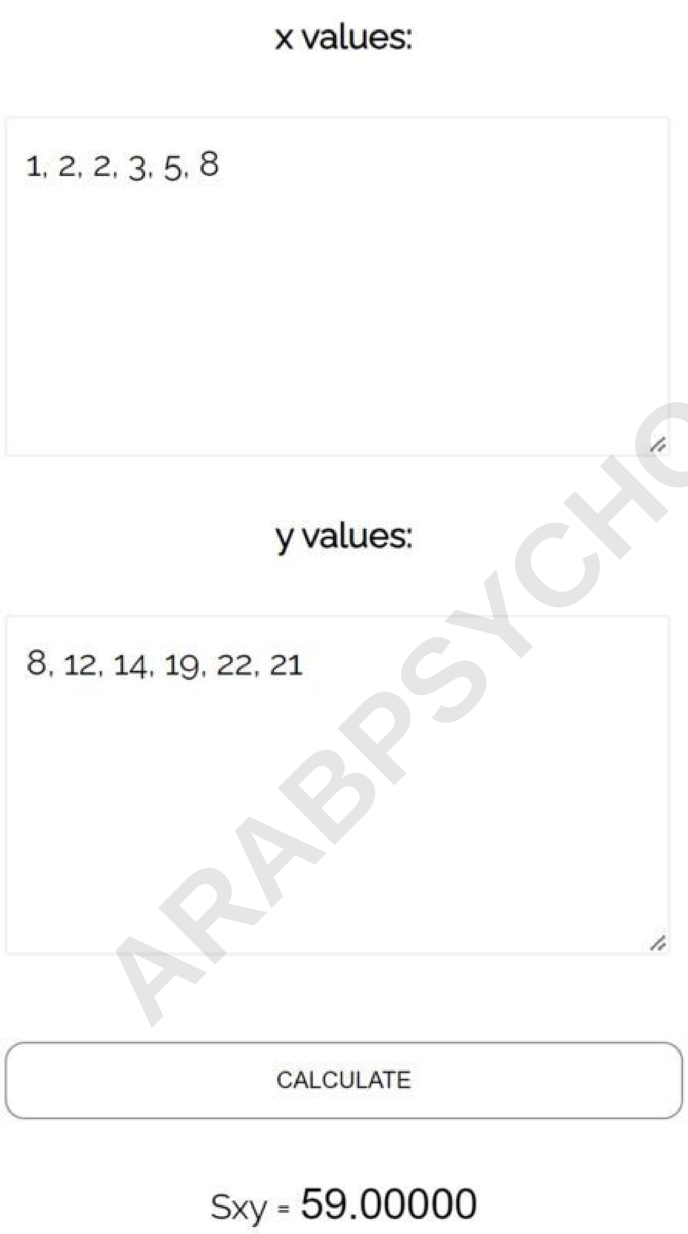
x	y	$x - x_{bar}$	$y - y_{bar}$	$(x - x_{bar})(y - y_{bar})$
1	8	-2.5	-8	20
2	12	-1.5	-4	6
2	14	-1.5	-2	3
3	19	-0.5	3	-1.5
5	22	1.5	6	9
8	21	4.5	5	22.5
			Σ	59

As shown in the calculated table, the sum of the products of the deviations, **Sxy**, for this specific dataset is **59**. A positive Sxy indicates a positive covariance and a positive linear association between X and Y.

Verification Using Statistical Software

While calculating **Sxy** by hand is vital for theoretical understanding, in professional settings, statistical software or dedicated online calculators are used for efficiency and accuracy. We can verify our manual calculation of 59 using a statistical tool designed for linear regression analysis.

Using a software package allows us to input the raw data and generate the necessary descriptive statistics automatically. The output below confirms the result:



The statistical output confirms that the calculator returns a value of **59**, which perfectly matches the value derived from our detailed, step-by-step manual calculation. This verifies the accuracy of both the means calculated and the final summation performed.

The Role of Sxy in Linear Regression Coefficients

The primary reason for calculating **Sxy** is its essential role in determining the slope (b) and intercept (a) of the best-fit line in a simple linear regression model. The general equation for a simple linear model is:

y = a + bx

Where 'a' is the intercept and 'b' is the slope. The slope coefficient (b) represents the estimated change in Y for every one-unit increase in X. It is calculated using the following relationship:

b = Sxy / Sxx

Here, Sxx is the sum of squares for X, calculated as Sxx = Σ(xi - x?)². Once the slope 'b' is calculated, the intercept 'a' is easily determined:

a = y - bx

The calculation for **Sxy** is just one calculation that we must perform in order to fit a simple linear regression model.

Distinguishing Sxy from Covariance and Correlation

While **Sxy** measures co-variation, it is important to understand how it relates to, and differs from, related statistical measures like covariance and the correlation coefficient (r). All three are fundamentally based on the product of deviations, but they differ in scaling and interpretation.

The sample covariance (Sxy) is calculated by taking **Sxy** and dividing it by (n-1), where n is the sample size. The formula is Cov(x, y) = Sxy / (n-1). Covariance provides an average measure of the linear relationship, but like Sxy, its magnitude is still dependent on the units of X and Y, making comparison across different datasets difficult.

In contrast, the correlation coefficient (r) standardizes this relationship by dividing the covariance by the product of the standard deviation of X and the standard deviation of Y. This normalization results in a unitless measure that ranges from -1 to 1, offering a clear interpretation of the strength and direction of the linear relationship, irrespective of the scale of the original variables. Therefore, Sxy is the raw, unscaled sum that feeds into these more standardized metrics.

Conclusion and Related Concepts

The calculation of **Sxy** is a fundamental skill for anyone engaging in quantitative analysis, particularly in the fields of economics, engineering, or social statistics. It represents the vital preliminary step toward quantifying the linear relationship between two variables, forming the numerator required for establishing the slope of the regression line. By calculating the difference of each observed data point from its respective mean and summing the resulting products, we gain insight into the magnitude and direction of the co-variation.

Understanding how Sxy relates to Sxx (the sum of squares for X) and Syy (the sum of squares for Y) is essential for full comprehension of bivariate analysis. While Sxy measures the relationship between X and Y, Sxx and Syy measure the variance within X and Y, respectively. These three components combined are necessary for a comprehensive analysis of the variability in the dataset.

The following tutorials explain how to perform other common tasks in statistics: