

# How to Easily Perform Bivariate Analysis in Excel

Authored by  
**stats writer**

December 2, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to Easily Perform Bivariate Analysis in Excel*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=103645>

Understanding the relationship between two distinct variables is fundamental to data analysis. Bivariate analysis in Excel offers a powerful, accessible way to explore how pairs of variables--such as sales figures and customer satisfaction scores--influence each other. This methodology moves beyond simple descriptive statistics, aiming to quantify the extent and nature of their connection.

Typically, this analysis begins with creating a scatter plot to visualize the data distribution. Following visualization, analysts often apply a linear regression model to mathematically define the relationship. The output of this model is critical, as it calculates the best-fit line through the data points and provides key metrics, notably the coefficient of determination (R-squared). This R-squared value is essential for measuring the strength and goodness-of-fit of the correlation between the two variables. Mastering bivariate techniques in Excel allows users to identify significant relationships and trends, enabling more informed, data-driven decision-making.

## Understanding the Core Concept: Bivariate Analysis

The term bivariate analysis is derived from the prefix "bi," meaning two, inherently signifying the study of two variables simultaneously. The primary goal of this statistical approach is to ascertain if a relationship exists between the pair of variables, and if so, to describe the nature, strength, and direction of that relationship. This initial exploratory step is crucial before attempting predictive modeling.

While many advanced statistical techniques exist, Excel provides three readily accessible and powerful methods for performing bivariate analysis. These methods offer increasing levels of detail, moving from visual inspection to quantified mathematical relationships.

These three common methods are:

Scatter plots (Visualizing the relationship)

Correlation Coefficients (Quantifying the strength and direction)

Simple Linear Regression (Modeling the predictive relationship)

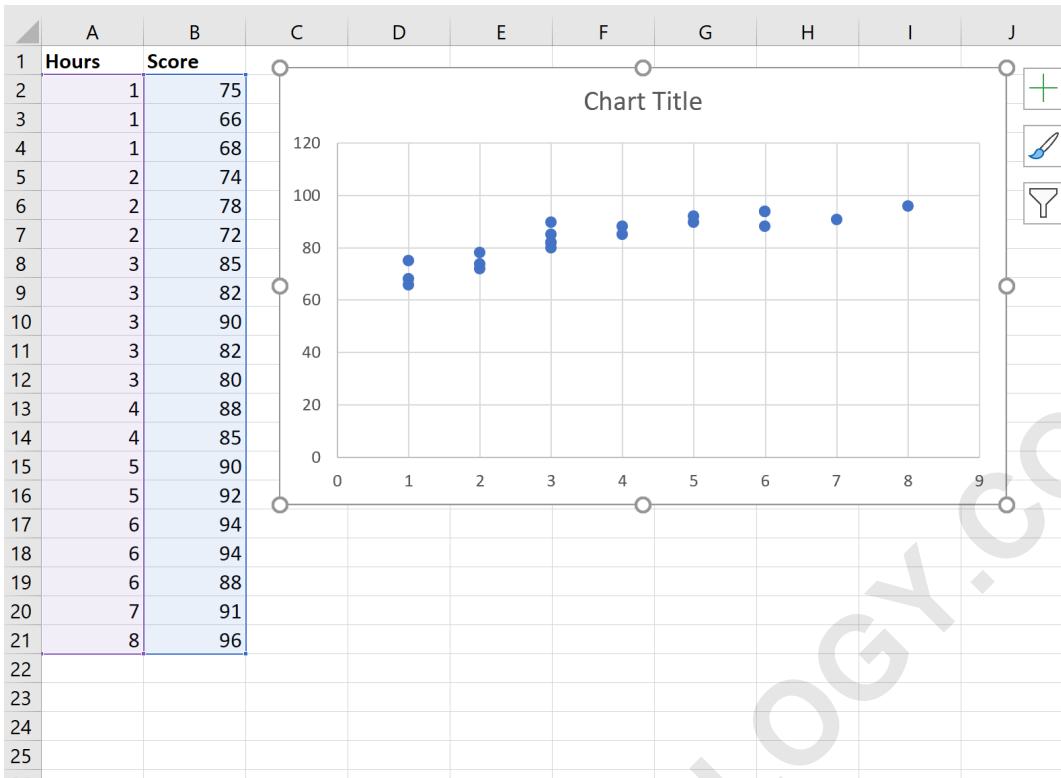
To demonstrate these techniques practically within the Microsoft Excel environment, we will utilize a sample dataset. This dataset tracks 20 different students and includes two primary variables: the total number of **Hours spent studying** and the corresponding **Exam score received**. This specific context provides a clear, intuitive example of how changes in an independent variable (Hours Studied) may predict changes in a dependent variable (Exam Score).

	A	B	C	D	E	F	G
1	<b>Hours</b>	<b>Score</b>					
2	1	75					
3	1	66					
4	1	68					
5	2	74					
6	2	78					
7	2	72					
8	3	85					
9	3	82					
10	3	90					
11	3	82					
12	3	80					
13	4	88					
14	4	85					
15	5	90					
16	5	92					
17	6	94					
18	6	94					
19	6	88					
20	7	91					
21	8	96					
22							
23							
24							

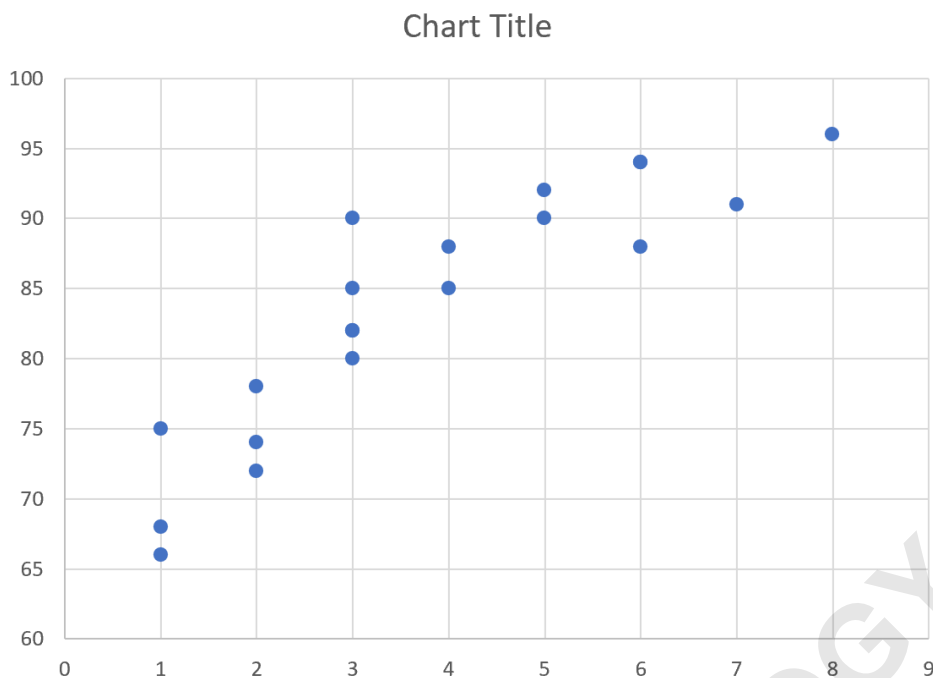
## 1. Creating and Interpreting Scatterplots

The scatter plot is the most immediate and intuitive method for assessing a bivariate relationship. By plotting the independent variable (Hours Studied) on the X-axis and the dependent variable (Exam Score) on the Y-axis, we can visually determine the direction (positive, negative, or none) and form (linear or curved) of the correlation.

To generate this visualization in Excel, follow these precise steps: First, highlight the data range containing both variables, specifically cells **A2:B21** in our sample data. Next, navigate to the **Insert** tab located on the top ribbon menu. Within the **Charts** group, select the option labeled **Insert Scatter Chart**. This action will immediately render the initial visual representation of the study hours versus the exam scores.



For enhanced clarity and to better focus on the cluster of data points, it is often necessary to refine the axis limits. Since exam scores generally range higher than zero, adjusting the Y-axis minimum can improve interpretability. To modify the Y-axis, double-click on the axis itself. A **Format Axis** panel will appear on the right side of your screen. Under the **Axis Options** heading, redefine the **Minimum** bound to 60 and the **Maximum** bound to 100. Excel will automatically update the chart to reflect these new boundaries, as illustrated below.



Upon reviewing the adjusted scatter plot, we can clearly observe a **positive relationship** between the two variables. The trend suggests that as the value on the X-axis (Hours Studied) increases, the corresponding value on the Y-axis (Exam Score) tends to increase as well. The tightness of the data points around an imaginary line also suggests a strong relationship, leading us naturally to the next quantitative step: calculating the correlation coefficient.

## 2. Calculating the Correlation Coefficient

While a scatter plot provides visual confirmation of a relationship, the Pearson Correlation Coefficient (often denoted as 'r') is the standard statistical measure used to numerically quantify the strength and direction of a linear relationship between two variables. This coefficient ranges from -1.0 to +1.0, where values closer to +1 indicate a strong positive relationship, values closer to -1 indicate a strong negative relationship, and values near 0 suggest a weak or non-existent linear relationship.

Excel simplifies this calculation significantly using the built-in `CORREL` function. To compute the correlation between Hours Studied (Array 1) and Exam Score (Array 2), we input the respective data ranges. Array 1 corresponds to cells A2 through A21, and Array 2 corresponds to B2 through B21. The formula is structured as follows:

**`=CORREL(A2:A21, B2:B21)`**

Executing this formula yields a correlation coefficient of **0.891**. This high value confirms the initial

visual assessment from the scatter plot, indicating a **strong positive correlation** between the number of hours a student spends studying and the score they ultimately receive on the exam. This strong positive value suggests that the variables move almost perfectly in lockstep.

	A	B	C	D	E	F	G
1	<b>Hours</b>	<b>Score</b>					
2	1	75		0.891306			
3	1	66					
4	1	68					
5	2	74					
6	2	78					
7	2	72					
8	3	85					
9	3	82					
10	3	90					
11	3	82					
12	3	80					
13	4	88					
14	4	85					
15	5	90					
16	5	92					
17	6	94					
18	6	94					
19	6	88					
20	7	91					
21	8	96					
22							
23							

### 3. Implementing Simple Linear Regression

Simple linear regression represents the pinnacle of bivariate analysis, as it not only quantifies the relationship but also establishes a predictive model. This statistical technique fits a straight line to the data that minimizes the sum of squared differences between the observed data points and the line itself (the method of least squares). The resulting equation allows us to predict the value of the dependent variable based on a known value of the independent variable.

To execute a linear regression model in Excel, we must utilize the Analysis ToolPak add-in. If you do not see the **Data Analysis** option under the **Data** tab on the ribbon, you must first enable this feature via Excel Options. Once enabled, navigate to the **Data** tab, locate the **Analyze** group, and click on **Data Analysis**. In the dialog box that appears, select **Regression** and then click **OK**.

The Regression input panel requires careful specification of the variables. For the **Input Y Range**,

enter the dependent variable (Exam Score: **\$B\$1:\$B\$21**). For the **Input X Range**, enter the independent variable (Hours Studied: **\$A\$1:\$A\$21**). It is important to check the **Labels** box, as our selection includes the header row, ensuring Excel correctly identifies the variables in the output summary. Keep the default settings for Confidence Level (95%) and select a desired **Output Range** to display the results on the current sheet, then click **OK**.

	A	B	C	D	E	F	G	H	I	J
1	<b>Hours</b>	<b>Score</b>								
2	1	75								
3	1	66								
4	1	68								
5	2	74								
6	2	78								
7	2	72								
8	3	85								
9	3	82								
10	3	90								
11	3	82								
12	3	80								
13	4	88								
14	4	85								
15	5	90								
16	5	92								
17	6	94								
18	6	94								
19	6	88								
20	7	91								
21	8	96								
22										
23										
24										
25										
26										

Regression

Input

Input Y Range:  ↑

Input X Range:  ↑

Labels  Constant is Zero

Confidence Level:  %

Output options

Output Range:  ↑

New Worksheet Ply:

New Workbook

Residuals

Residuals  Residual Plots

Standardized Residuals  Line Fit Plots

Normal Probability

Normal Probability Plots

OK Cancel Help

## Interpreting the Regression Output

Once the analysis is complete, Excel generates a detailed summary table. This table contains crucial metrics for evaluating the quality and coefficients of the fitted linear regression model, including the multiple R, the R Square, and the intercept and slope coefficients.

E	F	G	H	I	J
SUMMARY OUTPUT					
<i>Regression Statistics</i>					
Multiple R	0.891305974				
R Square	0.794426339				
Adjusted R Square	0.783005581				
Standard Error	4.170589387				
Observations	20				
<i>ANOVA</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	1209.911315	1209.911	69.55986	1.3474E-07
Residual	18	313.088685	17.39382		
Total	19	1523			
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	
Intercept	69.0733945	1.965132634	35.14948	4.85E-18	
Hours	3.847094801	0.46126824	8.340255	1.35E-07	

From the output, we can extract the coefficient values needed to construct our prediction equation. The intercept coefficient (or Y-intercept) is found under the "Intercept" row, and the slope coefficient for the independent variable is found under the "Hours Studied" row. The fitted equation takes the standard linear form ( $Y = \text{Intercept} + \text{Slope} * X$ ). Using the displayed coefficients, our predictive model is:

$$\text{Exam Score} = 69.0734 + 3.8471 * (\text{Hours Studied})$$

This equation provides strong insight. The intercept (69.0734) suggests that a student who studies zero hours is predicted to receive a base score of approximately 69.07. More importantly, the slope coefficient (3.8471) indicates that for every additional hour a student dedicates to studying, their exam score is expected to increase by an average of **3.8471 points**. Furthermore, the R-Squared value (Multiple R-Squared) found in the Summary Output table confirms the relationship's strength: 0.794, meaning approximately 79.4% of the variation in Exam Score can be explained by the variation in Hours Studied, reinforcing the strong findings of the correlation coefficient.

We can now leverage this equation for predictive purposes. Suppose we wish to estimate the score of a student who studies for exactly 3 hours. We substitute this value into our derived regression equation:

Exam Score =  $69.0734 + 3.8471 * (\text{Hours Studied})$

Exam Score =  $69.0734 + 3.8471 * (3)$

Exam Score = 81.6147

The prediction suggests that a student studying for 3 hours is estimated to receive a score of **81.6147**. This demonstrates the power of bivariate analysis in transforming observational data into actionable, predictive insights.

## Conclusion and Further Reading

By using scatter plots for visualization, correlation coefficients for quantification, and simple linear regression for prediction, Excel provides a comprehensive suite of tools for robust bivariate analysis. These techniques are essential for identifying statistically significant relationships that drive accurate forecasting and strategic business decisions.

For those interested in deepening their understanding of these statistical fundamentals, the following resources provide additional information about bivariate analysis and related statistical methods.