

How do you perform bivariate analysis in Excel? Can you provide examples?

Authored by
stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *How do you perform bivariate analysis in Excel? Can you provide examples?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=165417>

Bivariate analysis in Excel is a statistical method used to examine the relationship between two variables. This can be done by creating a scatter plot or by using the built-in correlation function. The first step is to organize the data into two columns, with each column representing one variable. Then, the data can be plotted on a scatter plot to visualize the relationship between the variables. Additionally, the correlation function can be used to calculate the correlation coefficient, which measures the strength and direction of the relationship. An example of bivariate analysis in Excel could be analyzing the relationship between a person's age and their income level. By plotting the data on a scatter plot and calculating the correlation coefficient, one can determine if there is a positive or negative relationship between age and income. This method can be used in various fields, such as finance, economics, and social sciences, to gain insights and make informed decisions.

Perform Bivariate Analysis in Excel (With Examples)

The term bivariate analysis refers to the analysis of two variables. You can remember this because the prefix "bi" means "two."

The purpose of bivariate analysis is to understand the relationship between two variables

There are three common ways to perform bivariate analysis:

1. Scatterplots

2. Correlation Coefficients

3. Simple Linear Regression

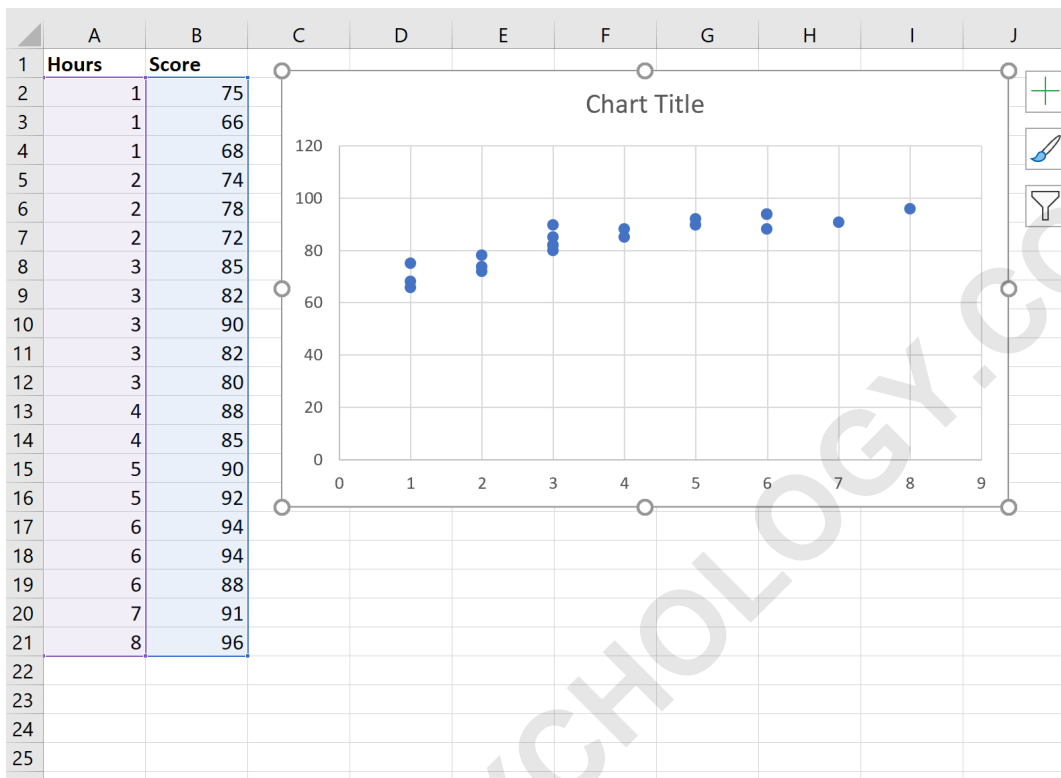
The following example shows how to perform each of these types of bivariate analysis in Excel using the following dataset that contains information about two variables: (1) Hours spent studying and (2) Exam score received by 20 different students:

| | A | B | C | D | E | F | G |
|----|-------|-------|----|---|---|---|---|
| 1 | Hours | Score | | | | | |
| 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. Scatterplots

To create a scatterplot of hours vs. score, we can highlight cells A2:B21, then click the Insert tab along

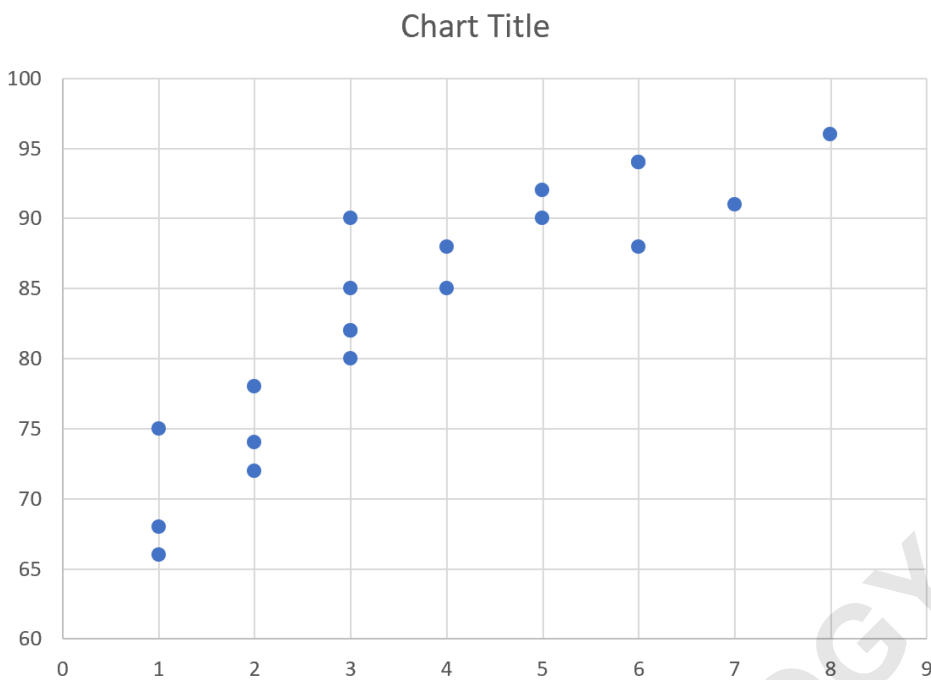
the top ribbon, then click **Insert Scatter Chart** within the **Charts** group:



We can also modify the y-axis limits to gain a better view of the data points.

To do so, double click the y-axis. In the **Format Axis** panel that appears on the right side of the screen, click **Axis Options** and then change the **Minimum** and **Maximum** bounds to 60 and 100, respectively.

The y-axis will automatically update:



The x-axis shows the hours studied and the y-axis shows the exam score received.

From the plot we can see that there is a positive relationship between the two variables. As hours studied increases, exam score tends to increase as well.

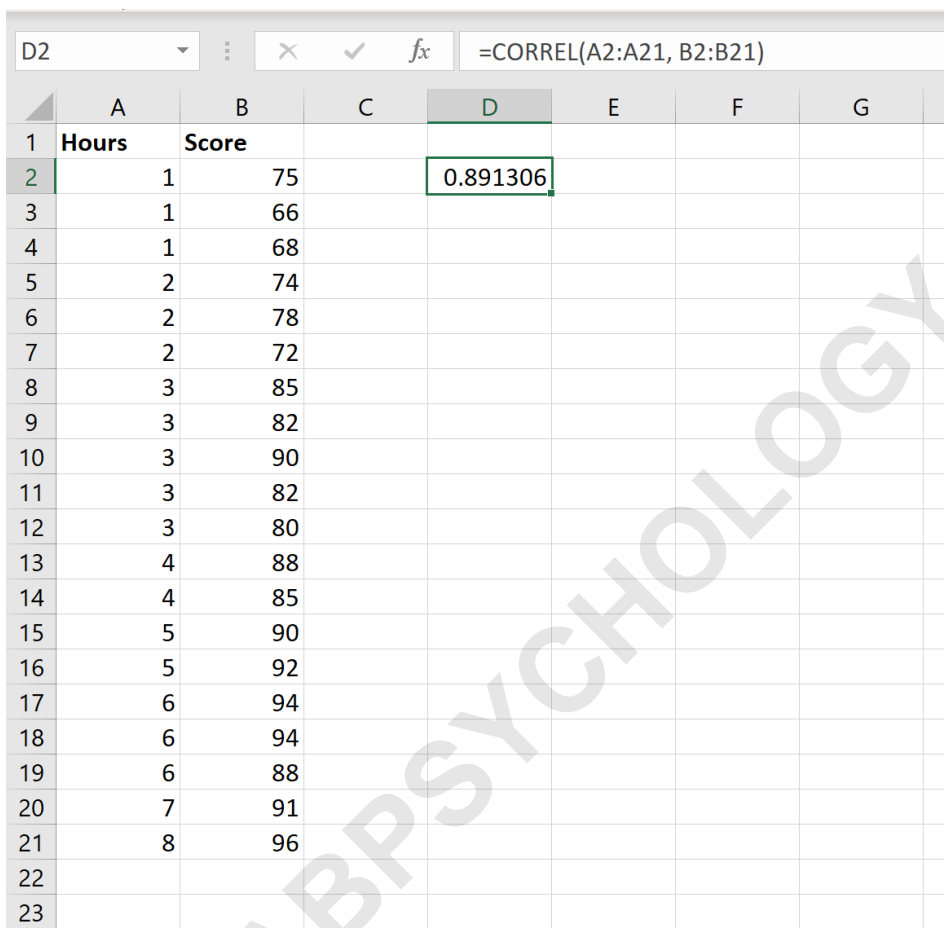
2. Correlation Coefficients

A Pearson Correlation Coefficient is a way to quantify the linear relationship between two variables.

We can use the following formula in Excel to calculate the correlation coefficient between hours studied and

exam score:

=CORREL(A2:A21, B2:B21)



| | A | B | C | D | E | F | G |
|----|-------|-------|---|----------|---|---|---|
| 1 | Hours | Score | | | | | |
| 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 | | | | | | | |

The correlation coefficient turns out to be 0.891.

This value is close to 1, which indicates a strong positive correlation between hours studied and exam score received.

3. Simple Linear Regression

Simple linear regression is a statistical method we can use to quantify the relationship between two variables.

To fit a simple linear regression model in Excel, click the Data tab along the top ribbon, then click the Data Analysis option in the Analyze group. In the new panel that appears, click Regression and then click OK.

Note: If you don't see the Data Analysis option, you need to first load the .

In the panel that appears, enter the following information and then click OK:

| | A | B | C | D | E | F | G | H | I | J |
|----|--------------|--------------|----|---|---|---|---|---|---|---|
| 1 | Hours | Score | | | | | | | | |
| 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 [?] [X]

Input

Input **Y** Range: [↑]

Input **X** Range: [↑]

Labels **C**onstant is **Z**ero

Confidence Level: %

Output options

Output Range: [↑]

New Worksheet **P**ly:

New **W**orkbook

Residuals

Residuals **R**esidual **P**lots

Standardized Residuals **L**ine **F**it **P**lots

Normal Probability

Normal Probability **P**lots

[OK] [Cancel] [Help]

Once you click OK, the results of the regression model will appear:

| 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 | |

The fitted regression equation turns out to be:

$$\text{Exam Score} = 69.0734 + 3.8471 * (\text{hours studied})$$

This tells us that each additional hour studied is associated with an average increase of 3.8471 in exam score.

We can also use the regression equation to estimate the score that a student will receive based on their total hours studied.

For example, a student who studies for 3 hours is estimated to receive a score of 81.6147:

Exam Score = 69.0734 + 3.8471*(hours studied)
Exam Score = 69.0734 + 3.8471*(3)
Exam Score = 81.6147

Additional Resources

The following tutorials provide additional information about bivariate analysis:

ARABPSYCHOLOGY.COM