

# How to calculate Partial Correlation in Excel?

Authored by  
**stats writer**

December 26, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to calculate Partial Correlation in Excel?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=108990>

The calculation of partial correlation in Excel requires a systematic approach that moves beyond simple correlation matrices. Initially, data must be meticulously organized within the spreadsheet. The process then involves calculating the required bivariate correlation coefficients--typically using the Pearson correlation coefficient--before applying a dedicated mathematical formula to isolate the relationship between two variables while holding the effect of a third variable constant. This advanced technique helps researchers achieve a more nuanced understanding of complex relationships in multivariate data sets, moving beyond the limitations of simple bivariate analysis. While Excel does not offer a direct built-in function for partial correlation, we can leverage its powerful statistical functions, such as the CORREL function, to perform the necessary computations efficiently and accurately, ensuring that the results provide meaningful insights into the underlying data structure.

## Introduction to Correlation and Confounding Variables

In the realm of statistics, the standard correlation coefficient is fundamentally employed to quantify the linear strength and direction of the relationship observed between two distinct variables. This measure, most commonly the **Pearson correlation coefficient** (often denoted as  $r$ ), provides a snapshot of how two variables move together. However, real-world data is rarely confined to just two variables; often, a third factor, known as a confounding variable, may influence both variables under observation, thereby distorting or spuriously inflating the perceived association between them. Understanding this limitation is crucial, as relying solely on bivariate correlation can lead to incorrect conclusions about the true underlying dynamics of the system being studied.

Consider a scenario where researchers are studying the relationship between ice cream sales and instances of drowning. A simple correlation calculation might reveal a strong positive association. However, this correlation is likely spurious, driven by a third variable: the ambient temperature. When temperatures rise (the confounding variable), both ice cream sales and swimming activities (and subsequently, drowning incidents) increase. To accurately assess the intrinsic link between ice cream sales and drowning risk, if any exists, we must mathematically remove the influence of temperature. This necessity highlights a significant gap in bivariate analysis and establishes the critical role of controlling for external factors when seeking robust statistical inference.

This is precisely where the technique of **partial correlation** becomes indispensable. The core objective of partial correlation is to measure the degree of association between two variables after controlling for or adjusting for the effect of one or more additional variables. By mathematically neutralizing the influence of these extraneous factors, we can isolate the unique covariance shared exclusively by the two variables of primary interest. This method yields a correlation coefficient that offers a purer, more focused measure of association, providing a far more reliable foundation for theory development and practical decision-making than simple correlation alone.

## Understanding the Need for Partial Correlation

Partial correlation is a sophisticated statistical tool designed to untangle complex relationships obscured by mediating or confounding variables. For instance, suppose we aim to measure the association between the number of hours a student spends studying and the final exam score they ultimately achieve. A straightforward correlation might be high, but this relationship is often intertwined with the student's pre-existing academic ability, perhaps quantified by their current grade in the class. A student with a high current grade might naturally study more effectively or already possess deep subject knowledge, making their high exam score potentially less attributable solely to the recent hours spent studying.

In this educational example, the "current grade" acts as the control variable (or confounding factor). By employing a **partial correlation analysis**, we can statistically hold the current grade constant for all students in the sample. This allows us to determine the correlation between study hours and exam scores as if every student started with the identical current grade. The resultant partial correlation coefficient reveals the true, isolated association between studying and performance, free from the biasing influence of prior academic standing. If the simple correlation was 0.70 but the partial correlation drops to 0.20, it strongly suggests that the original relationship was largely mediated by the student's initial capability.

Thus, the primary utility of partial correlation lies in its ability to enhance the precision of our statistical models. It transforms a potentially misleading bivariate correlation into a controlled estimate that better reflects the direct relationship between the variables of interest. This makes it an essential technique in fields like psychology, economics, and social sciences, where observed relationships are often interconnected through multiple simultaneous processes. The calculation requires the input of the original Pearson correlation coefficient values calculated for all possible pairs of variables involved, establishing the foundation for the subsequent mathematical adjustment.

### Case Study Setup: Data Preparation in Excel

To demonstrate the calculation of partial correlation effectively, we will utilize a practical example involving student performance data within the familiar environment of Excel. Our objective is to determine the correlation between study hours and final exam score while systematically controlling for the student's current grade. This requires a dataset structured with three variables (Current Grade, Hours Studied, and Final Exam Score) across multiple observations (students).

Suppose our dataset encompasses information for 10 hypothetical students, detailing their performance metrics:

Current grade in a class (A proxy for prior knowledge or ability)

Hours spent studying for the final exam (The independent variable)

Final exam score (The dependent variable)

It is crucial that this data is input cleanly into adjacent columns within the Excel spreadsheet. Assigning clear labels to each column is essential for clarity during the correlation calculation steps. For organizational purposes, we can designate Variable A as the Final Exam Score, Variable B as Hours Studied, and Variable C as the Current Grade (the control variable). This consistent nomenclature will simplify the application of the partial correlation formula later on.

	A	B	C	D	E	F
1	<b>Grade</b>	<b>Hours</b>	<b>Exam</b>			
2	82	4	88			
3	88	3	85			
4	75	6	76			
5	74	5	70			
6	93	4	92			
7	97	5	94			
8	83	8	89			
9	90	7	85			
10	90	4	90			
11	80	6	93			
12						
13						
14						
15						
16						
17						
18						
19						
20						

Once the data is correctly entered as shown above, we can proceed to the sequential steps required to find the partial correlation between hours studied and exam score while controlling for the current grade. The following methodology outlines a robust, manual approach that ensures precise control over the calculation process, which is necessary since Excel lacks an automated partial correlation function.

### Step 1: Calculating All Pairwise Correlations

The foundational requirement for computing the partial correlation ( $r_{AB.C}$ ) is the calculation of all three bivariate correlation coefficients among the variables A, B, and C. These coefficients measure the simple linear relationship between every possible pair of variables before any control adjustment is made. If A is Exam Score, B is Hours Studied, and C is Current Grade, we need the following three correlation values:

$r_{A,B}$ : The simple correlation between Exam Score (A) and Hours Studied (B).

$r_{A,C}$ : The simple correlation between Exam Score (A) and Current Grade (C).

$r_{B,C}$ : The simple correlation between Hours Studied (B) and Current Grade (C).

In Excel, these correlations are calculated using the built-in `CORREL` function. For example, to find  $r_{A,C}$ , one would use the formula `=CORREL(Column_A_Data, Column_C_Data)`. These values represent the raw, unadjusted relationships, reflecting any potential inflation or suppression caused by the third variable. It is advisable to dedicate a specific cell in Excel to store each of these calculated values, labeled clearly, as they form the inputs for the next crucial step.

After applying the `CORREL` function to the dataset, we obtain a correlation matrix containing these three essential pairwise coefficients. For our hypothetical data, these calculations might yield strong positive correlations, reflecting the natural tendency for these educational metrics to increase together. This matrix confirms the presence of linear associations that we must now scrutinize through the partial correlation formula to determine which associations are direct and which are confounded.

	A	B	C	D	E	F	G	H	I
1	<b>Grade</b>	<b>Hours</b>	<b>Exam</b>				<b>Formula Used</b>		
2	82	4	88		r(Hours, Exam)	-0.0595	=CORREL(B2:B11, C2:C11)		
3	88	3	85		r(Hours, Grade)	-0.25716	=CORREL(B2:B11, A2:A11)		
4	75	6	76		r(Exam, Grade)	0.724872	=CORREL(C2:C11, A2:A11)		
5	74	5	70						
6	93	4	92						
7	97	5	94						
8	83	8	89						
9	90	7	85						
10	90	4	90						
11	80	6	93						
12									
13									
14									
15									
16									
17									
18									
19									

As depicted in the screenshot, the resulting values-- $r_{A,B}$ ,  $r_{A,C}$ , and  $r_{B,C}$ --are now available. These numbers indicate the degree of shared variance between each pair. The high values here suggest that the three variables are significantly interrelated, which necessitates the use of partial correlation to disentangle these dependencies and ascertain the true, independent relationship between hours studied and the final score, removed from the effect of the current grade.

## Step 2: Applying the Partial Correlation Formula

With the three required pairwise correlation coefficients calculated, we can proceed to the core computation of the partial correlation. The formula for calculating the partial correlation between Variable A and Variable B, while controlling for Variable C (denoted as  $r_{AB.C}$ ), is mathematically defined as follows:

The formula is designed to remove the covariance that A and B share solely through their relationship with C. The numerator adjusts the raw correlation ( $r_{A,B}$ ) by subtracting the product of the correlations involving the control variable ( $r_{A,C} * r_{B,C}$ ). The denominator acts as a scaling factor, ensuring the resulting coefficient remains within the standardized range of -1 to +1, reflecting the reduction in variance due to the controlled factor.

$$\text{Partial correlation } (r_{AB.C}) = (r_{A,B} - r_{A,C} * r_{B,C}) / (\sqrt{((1-r_{A,C}^2)(1-r_{B,C}^2))})$$

Translating this statistical formula into a working Excel equation requires meticulous cell referencing to the results obtained in Step 1. If we assume  $r_{A,B}$  is in cell D2,  $r_{A,C}$  is in D3, and  $r_{B,C}$  is in D4, the Excel formula would combine these references using the appropriate algebraic operations. Specifically, the use of the `SQRT` function is necessary to calculate the square root in the denominator, and the exponent operator (`^2`) is used for squaring the r-values. The precision of the calculation is paramount, so ensuring correct parentheses placement according to the order of operations is vital for a valid result.

	A	B	C	D	E	F	G	H	I	J
1	<b>Grade</b>	<b>Hours</b>	<b>Exam</b>				<b>Formula Used</b>			
2	82	4	88		r(Hours, Exam)	-0.0595	=CORREL(B2:B11, C2:C11)			
3	88	3	85		r(Hours, Grade)	-0.25716	=CORREL(B2:B11, A2:A11)			
4	75	6	76		r(Exam, Grade)	0.724872	=CORREL(C2:C11, A2:A11)			
5	74	5	70							
6	93	4	92		Partial r(Hours, Exam)	0.190626	=(F2-F3*F4)/SQRT((1-F3^2)*(1-F4^2))			
7	97	5	94							
8	83	8	89							
9	90	7	85							
10	90	4	90							
11	80	6	93							
12										
13										
14										
15										
16										

As demonstrated in the screenshot, applying this formula yields a partial correlation coefficient. In this example, the partial correlation calculated is **0.190626**. This value is significantly lower than the original simple correlation ( $r_{A,B}$ , which was 0.81237), indicating that a substantial portion of the

original relationship between Hours Studied and Final Exam Score was indeed explained by or mediated through the student's Current Grade. The remaining correlation (0.190626) represents the relationship that persists even after the influence of the current grade has been statistically accounted for. However, a correlation value alone does not confirm whether this residual relationship is reliable or simply due to random sampling fluctuations; thus, we must proceed to assess its statistical significance.

### Step 3: Assessing Statistical Significance (t-test)

Once the partial correlation coefficient ( $r_p = 0.190626$ ) is calculated, the next critical step is to determine if this correlation is sufficiently strong to be considered statistically significant, meaning it is unlikely to have occurred purely by chance. This assessment is typically performed using a t-test. The null hypothesis ( $H_0$ ) for this test is that the true population partial correlation is zero, implying no relationship exists after controlling for the third variable.

The test statistic,  $t$ , for the partial correlation can be calculated using a slightly adjusted formula compared to a simple correlation t-test. Since we have controlled for one variable ( $C$ ), we lose one additional degree of freedom. The formula for the test statistic  $t$  is given by:

$$t = r_p \sqrt{(n-k-1) / \sqrt{(1-r_p^2)}}$$

Where:  $r_p$  is the partial correlation coefficient;  $n$  is the sample size (number of students,  $n=10$ ); and  $k$  is the number of control variables ( $k=1$ , Current Grade). In the simplified notation often used for partial correlation with one control variable, this is written as:  $t = r \sqrt{(n-3) / \sqrt{(1-r^2)}}$ .

To calculate this in Excel, we substitute the known values. The resulting  $t$  statistic provides a measure of how far our calculated partial correlation deviates from zero, relative to the sampling variability. We must also determine the degrees of freedom (df), which in this case is calculated as  $n - k - 1$ , or  $10 - 1 - 1 = 8$ , or simply  $n - 3 = 7$  based on the common notation for first-order partial correlation. Using  $df = 7$  is appropriate for this specific test.

	A	B	C	D	E	F	G	H	I	J
1	<b>Grade</b>	<b>Hours</b>	<b>Exam</b>				<b>Formula Used</b>			
2	82	4	88		r(Hours, Exam)	-0.0595	=CORREL(B2:B11, C2:C11)			
3	88	3	85		r(Hours, Grade)	-0.25716	=CORREL(B2:B11, A2:A11)			
4	75	6	76		r(Exam, Grade)	0.724872	=CORREL(C2:C11, A2:A11)			
5	74	5	70							
6	93	4	92		Partial r(Hours, Exam)	0.190626	=(F2-F3*F4)/SQRT((1-F3^2)*(1-F4^2))			
7	97	5	94							
8	83	8	89		t	0.51377	=F6/SQRT(1-F6^2)*SQRT(COUNT(A2:A11)-3)			
9	90	7	85		df	7	=COUNT(A2:A11)-3			
10	90	4	90		p	0.623228	=T.DIST.2T(F8, F9)			
11	80	6	93							
12										
13										
14										
15										
16										
17										
18										

The computation shown in the corresponding Excel screenshot illustrates the derivation of the  $t$  statistic. Our calculated test statistic,  $t$ , is found to be **0.51377**. The determination of the statistical significance relies on comparing this  $t$  value against a critical value from the  $t$ -distribution or, more commonly, deriving the corresponding P-value using the appropriate Excel function, such as `T.DIST.2T` for a two-tailed test, which is standard for testing correlation hypotheses. The P-value indicates the probability of observing a correlation this extreme if the null hypothesis were true.

## Interpreting the Final Results and Significance

The final step involves synthesizing the partial correlation coefficient, the test statistic, and the resulting P-value to draw definitive conclusions about the relationship between hours studied and exam score, conditional on the current grade. Our calculated test statistic  $t$  is **0.51377**, and the degrees of freedom are confirmed as 7 ( $n-3$ ). The corresponding P-value, calculated using the  $t$  statistic and the degrees of freedom, is **0.623228**.

To evaluate statistical significance, we compare the P-value to a predetermined significance level ( $\alpha$ , often set at 0.05). The decision rule dictates that if the P-value is less than  $\alpha$  ( $P < 0.05$ ), we reject the null hypothesis and conclude that the partial correlation is significant. Conversely, if the P-value is greater than  $\alpha$  ( $P > 0.05$ ), we fail to reject the null hypothesis, suggesting that the observed partial correlation is not robustly different from zero.

In our case, the P-value of 0.623228 is substantially greater than the conventional  $\alpha$  level of 0.05. Therefore, we must conclude that the partial correlation of 0.190626 between hours studied and exam score, after controlling for the current grade, is not **statistically significant**. This finding carries an important practical implication: while the two variables initially appeared highly

correlated ( $r = 0.81$ ), nearly all of that predictive power was attributable to the shared variance with the student's current grade. Once the effect of the current grade is removed, the remaining direct relationship between hours studied and the final score is too weak ( $r_p = 0.19$ ) to be reliably distinguished from random noise, suggesting that in this specific sample, hours studied had little unique linear impact on the final score beyond what was already explained by the initial academic standing.

ARABPSYCHOLOGY.COM