

How to Calculate Spearman Rank Correlation in Excel: A Step-by-Step Guide

Authored by
stats writer

March 9, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Calculate Spearman Rank Correlation in Excel: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=134884>

The Fundamentals of Correlation in Statistical Analysis

In the vast field of **statistics**, the concept of correlation serves as a fundamental pillar for understanding how different **variables** interact with one another. Correlation measures the degree to which two sets of data move in relation to each other, providing insights into potential dependencies or associations. When we analyze **correlation coefficients**, we typically look at a numerical value that ranges from -1 to 1. A value of 1 signifies a perfect positive relationship, meaning as one variable increases, the other increases in a perfectly predictable manner. Conversely, a value of -1 represents a perfect negative relationship, where one variable increases while the other decreases systematically. A value of 0 indicates that no discernible relationship exists between the datasets being studied.

While there are several methods to quantify these relationships, the most common is the **Pearson correlation**, which measures linear relationships between continuous data. However, in many real-world scenarios, data may not meet the strict assumptions of linearity or normality required for Pearson's method. This is where **non-parametric statistics** become essential. These methods allow researchers to draw conclusions without assuming a specific distribution for the data, making them highly versatile for various types of scientific and business inquiries. By focusing on the relative order of data points rather than their exact numerical values, we can uncover patterns that linear models might overlook.

The **Spearman Rank Correlation** is a specific non-parametric measure that assesses how well the relationship between two variables can be described using a monotonic function. Unlike linear correlation, Spearman's method is ideal for ranked data or ordinal variables, such as the ranking of students in a classroom or the preference order of consumer products. It is particularly robust against **outliers**, as the ranking process effectively limits the influence of extreme values that would otherwise skew a traditional linear analysis. Understanding this method is crucial for anyone looking to perform deep data analysis using **Microsoft Excel**.

Defining the Spearman Rank Correlation Coefficient

The **Spearman Rank Correlation** coefficient, often denoted by the Greek letter rho (ρ), is a statistical measure designed to evaluate the strength and direction of the association between two ranked variables. It is essentially the Pearson correlation coefficient calculated between the rank values of those variables. This means that instead of comparing the raw scores of two datasets, such as test scores or heights, we compare their positions within their respective groups. This approach is highly effective when dealing with data that is **ordinal** or when the relationship between variables is non-linear but still monotonic--meaning the variables tend to move in the same relative direction but not at a constant rate.

One of the primary advantages of utilizing the **Spearman Rank Correlation** is its flexibility regarding the distribution of data. Because it relies on ranks rather than raw values, it does not require the assumption that the data follows a **normal distribution**. This makes it a preferred choice for social sciences, psychology, and biology, where data often exhibits skewness or contains significant outliers. For example, if you were comparing the rank of a student's math exam performance against their rank in a science exam, the Spearman method would provide a clear indication of whether high performance in one subject consistently corresponds with high performance in the other, regardless of the actual point spread between the students.

In **Microsoft Excel**, calculating this coefficient is a streamlined process that involves two main phases: ranking the raw data and then applying a correlation formula to those ranks. While some manual methods involve calculating the squared differences between ranks, modern software allows us to use built-in functions to achieve the same result with greater accuracy and speed. This tutorial will guide you through a practical example involving ten students, demonstrating how to transform raw exam scores into a meaningful **correlation coefficient** that can inform educational or administrative decisions.

Step 1: Data Organization and Preparation

The first step in any successful statistical analysis is the meticulous organization of your dataset. In **Microsoft Excel**, you should begin by creating a clear structure where each **variable** is contained within its own column. For this specific example, we are investigating the relationship between math exam scores and science exam scores. You should enter the names of the subjects or students in the first column and their corresponding scores in the adjacent columns. Proper labeling ensures that the data remains interpretable as you proceed through the subsequent calculation steps.

Consistency is key when entering data for **Spearman Rank Correlation**. Each row should represent a single observational unit--in this case, an individual student. Ensure that there are no missing values in your columns, as empty cells can cause errors in the ranking functions. If a student is missing a score for one subject, it is generally standard practice to exclude that student from the correlation analysis entirely to maintain the integrity of the **sample size**. Accuracy at this stage prevents the propagation of errors throughout the analytical workflow.

Once your scores are entered, your spreadsheet should resemble the structure shown in the following image. This visual organization allows you to easily verify that each student has a pair of scores, which is a prerequisite for calculating any form of correlation. By having the data neatly laid out in **Microsoft Excel**, you set the foundation for the automated ranking formulas that will follow in the next step.

	A	B	C	D	E
1	Student	Math	Science		
2	Austin	70	90		
3	Bob	78	94		
4	Craig	90	79		
5	Devon	87	86		
6	Elizabeth	84	84		
7	Frank	86	83		
8	George	91	88		
9	Helen	74	92		
10	Isaiah	83	76		
11	Jessica	85	75		
12					
13					
14					
15					
16					

Step 2: Implementing the Ranking Methodology

After the raw data has been entered, the next phase of the **Spearman Rank Correlation** process is to convert these scores into ranks. To do this in **Microsoft Excel**, we utilize the **RANK.AVG** function. This specific function is superior to the standard RANK function because it handles "ties" (instances where two or more students have the same score) by assigning them the average of the ranks they would have otherwise occupied. This ensures that the sum of the ranks remains constant and that the resulting **correlation coefficient** is not biased by identical data points.

To begin ranking, you will create two new columns, labeled "Math Rank" and "Science Rank." In the first cell of the Math Rank column (D2), you will enter a formula that references the student's score and the entire range of scores for that subject. By using absolute cell references (denoted by the dollar signs, e.g., \$B\$2:\$B\$11), you ensure that the range remains fixed as you copy the formula down the column. This allows Excel to compare each individual score against the entire group to determine its relative position.

The specific formulas to be entered into cells D2 and E2 are as follows:

Cell D2: =RANK.AVG(B2, \$B\$2:\$B\$11, 0)

Cell E2: =RANK.AVG(C2, \$C\$2:\$C\$11, 0)

	A	B	C	D	E	F
1	Student	Math	Science	Math Rank	Science Rank	
2	Austin	70	90	10	3	
3	Bob	78	94			
4	Craig	90	79			
5	Devon	87	86			
6	Elizabeth	84	84			
7	Frank	86	83			
8	George	91	88			
9	Helen	74	92			
10	Isaiah	83	76			
11	Jessica	85	75			
12						
13						
14						
15						

Once the initial formulas are set for the first student, you must apply them to the rest of the class. This is easily achieved by highlighting the cells containing the formulas and the empty cells below them for all ten students. By using the fill-down command or clicking and dragging the fill handle, Excel will automatically calculate the ranks for every data point in the set.

	A	B	C	D	E	F
1	Student	Math	Science	Math Rank	Science Rank	
2	Austin	70	90	10	3	
3	Bob	78	94			
4	Craig	90	79			
5	Devon	87	86			
6	Elizabeth	84	84			
7	Frank	86	83			
8	George	91	88			
9	Helen	74	92			
10	Isaiah	83	76			
11	Jessica	85	75			
12						
13						
14						
15						
16						

After applying the formulas, you should see a list of ranks ranging from 1 to 10 for each subject. If there were ties in the scores, you would see fractional ranks (like 5.5). Using the **Ctrl+D** shortcut is a rapid way to ensure all students are ranked correctly based on their performance relative to their peers. This transformation from raw scores to ranks is the critical step that defines the **Spearman**

Rank Correlation method.

	A	B	C	D	E	F
1	Student	Math	Science	Math Rank	Science Rank	
2	Austin	70	90	10	3	
3	Bob	78	94	8	1	
4	Craig	90	79	2	8	
5	Devon	87	86	3	5	
6	Elizabeth	84	84	6	6	
7	Frank	86	83	4	7	
8	George	91	88	1	4	
9	Helen	74	92	9	2	
10	Isaiah	83	76	7	9	
11	Jessica	85	75	5	10	
12						
13						
14						
15						

Step 3: Calculating the Correlation Coefficient

With the data now successfully ranked, the final calculation of the **Spearman Rank Correlation** coefficient is straightforward. In **Microsoft Excel**, the **CORREL()** function is used to find the Pearson correlation between two arrays. Since the Spearman coefficient is mathematically equivalent to the Pearson correlation of the ranks, we simply apply this function to our newly created rank columns. This eliminates the need for complex manual calculations involving squared differences and the traditional Spearman formula, reducing the likelihood of human error.

Choose an empty cell in your spreadsheet to display the final result. In this cell, you will input the function referencing the two arrays of ranks. For example, if your Math ranks are in column D and your Science ranks are in column E, your formula would target those specific ranges. Excel will then process the relationship between these two sets of ranks and provide a single numerical value that encapsulates the entire relationship between the two exam scores across the student population.

	A	B	C	D	E
1	Student	Math	Science	Math Rank	Science Rank
2	Austin	70	90	10	3
3	Bob	78	94	8	1
4	Craig	90	79	2	8
5	Devon	87	86	3	5
6	Elizabeth	84	84	6	6
7	Frank	86	83	4	7
8	George	91	88	1	4
9	Helen	74	92	9	2
10	Isaiah	83	76	7	9
11	Jessica	85	75	5	10
12					
13	Spearman Correlation:	=CORREL(D2:D11, E2:E11)			
14					
15					

In our specific classroom example, performing this calculation yields a Spearman rank **correlation coefficient** of **-0.41818**. This number provides the first quantitative look at how student performance in Math relates to performance in Science. The negative sign is immediately informative, suggesting that there might be an inverse relationship between the two subjects in this particular group of students, though we must look deeper into the significance of this number before drawing final conclusions.

	A	B	C	D	E
1	Student	Math	Science	Math Rank	Science Rank
2	Austin	70	90	10	3
3	Bob	78	94	8	1
4	Craig	90	79	2	8
5	Devon	87	86	3	5
6	Elizabeth	84	84	6	6
7	Frank	86	83	4	7
8	George	91	88	1	4
9	Helen	74	92	9	2
10	Isaiah	83	76	7	9
11	Jessica	85	75	5	10
12					
13	Spearman Correlation:	-0.41818			
14					
15					

Step 4: Interpretation and Statistical Significance

Obtaining a numerical value for the **Spearman Rank Correlation** is only half of the analytical journey; the next step is interpretation. A coefficient of **-0.41818** indicates a moderate negative correlation. In practical terms, this suggests that students who ranked higher in Math tended to rank somewhat lower in Science, and vice versa. However, because this value is not close to -1, the relationship is far from perfect, indicating significant variability in student performance across the two subjects.

To determine if this observed relationship is meaningful or simply a result of random chance, we must evaluate its **statistical significance**. This involves testing the **null hypothesis**, which posits that there is actually no correlation between the variables in the broader population. To do this, we compare the absolute value of our calculated coefficient to a **critical value** found in a standard Spearman rank correlation table. This table provides thresholds based on the **sample size** (n) and the chosen **significance level** (often denoted as alpha, α).

n	α				
	0.10	0.05	0.025	0.01	0.005
5	0.800	0.900	1.000	1.000	-
6	0.657	0.829	0.886	0.943	1.000
7	0.571	0.714	0.786	0.893	0.929
8	0.524	0.643	0.738	0.833	0.881
9	0.483	0.600	0.700	0.783	0.833
10	0.455	0.564	0.648	0.745	0.794
11	0.427	0.536	0.618	0.709	0.755
12	0.406	0.503	0.587	0.678	0.727
13	0.385	0.484	0.560	0.648	0.703
14	0.367	0.464	0.538	0.626	0.679
15	0.354	0.446	0.521	0.604	0.654
16	0.341	0.429	0.503	0.582	0.635
17	0.328	0.414	0.488	0.566	0.618
18	0.317	0.401	0.472	0.550	0.600
19	0.309	0.391	0.460	0.535	0.584
20	0.299	0.380	0.447	0.522	0.570
21	0.292	0.370	0.436	0.509	0.556
22	0.284	0.361	0.425	0.497	0.544
23	0.278	0.353	0.416	0.486	0.532
24	0.271	0.344	0.407	0.476	0.521
25	0.265	0.337	0.398	0.466	0.511
26	0.259	0.331	0.390	0.457	0.501
27	0.255	0.324	0.383	0.449	0.492
28	0.250	0.318	0.375	0.441	0.483
29	0.245	0.321	0.368	0.433	0.475
30	0.240	0.306	0.362	0.425	0.467

For our example with a **sample size** of $n = 10$ and a standard **significance level** of 0.05, the **critical value** is **0.564**. Since our calculated absolute value (0.41818) is less than the critical value (0.564), we fail to reject the null hypothesis. This means that although we observed a negative correlation in our sample, it is not statistically significant. We cannot confidently claim that a relationship exists between Math and Science scores for all students based on this small group of ten individuals. This conclusion highlights the importance of sample size in **statistics**; larger samples are generally required to detect moderate correlations with high confidence.

Advanced Considerations: When to Use Spearman Over Pearson

Choosing between **Pearson correlation** and **Spearman Rank Correlation** is a critical decision in data analysis. Pearson's method is highly sensitive to the exact distance between values and assumes a linear relationship. If your data contains **outliers**--values that are significantly higher or lower than the rest of the group--Pearson may produce a misleading coefficient. Spearman, by contrast, focuses on the "order" of the data, which makes it far more resilient to these anomalies. If one student in our example had scored a 100 while everyone else scored in the 60s, the Spearman rank would only see them as "Rank 1," whereas Pearson would be heavily influenced by the 40-point gap.

Another factor is the type of relationship you expect to find. If you suspect that as one variable increases, the other increases, but not necessarily at a constant rate (a monotonic but non-linear relationship), Spearman is the superior tool. For instance, the relationship between "hours spent studying" and "exam score" might plateau after a certain point; this non-linear but consistent trend is captured more accurately by rank-based methods. Using **Microsoft Excel** to calculate both coefficients can sometimes provide a more holistic view of the data's behavior.

Finally, the level of measurement of your **variables** dictates the choice of test. If your data is **ordinal**--meaning it represents categories with a natural order, like "satisfied," "neutral," and "dissatisfied"--you cannot use Pearson because the "distance" between satisfaction levels is not numerically defined. Spearman is specifically designed for such cases. By understanding these nuances, you can ensure that your statistical analysis is both appropriate for your data type and scientifically rigorous.

Summary of Calculating Spearman Correlation in Excel

Mastering the **Spearman Rank Correlation** in **Microsoft Excel** is a valuable skill for any data analyst or researcher. The process, while grounded in complex **non-parametric statistics**, is made accessible through Excel's intuitive functions. By following the systematic approach of organizing data, applying the **RANK.AVG** function to handle ties fairly, and utilizing the **CORREL()** function to derive the final coefficient, you can efficiently analyze relationships within any ranked

dataset. This workflow transforms raw, potentially messy information into a clear, quantifiable measure of association.

Beyond the calculation itself, the ability to interpret the results and determine **statistical significance** is what separates a basic user from an expert. Understanding that a **correlation coefficient** must be compared against a **critical value** ensures that you do not over-interpret findings that could be due to random variation. Always remember to consider your **sample size** and the context of your data when drawing conclusions about the strength and direction of the relationships you discover.

As you continue to use **statistics** in your professional or academic work, the **Spearman Rank Correlation** will remain a robust tool in your analytical arsenal. Whether you are analyzing student performance, market trends, or biological data, this method provides a reliable way to uncover the underlying structure of your data without being misled by outliers or non-linear patterns. With the power of **Microsoft Excel**, these insights are only a few formulas away.