Learning to Calculate Spearman’s Rank Correlation Coefficient in Excel: A Step-by-Step Guide

Authored by

Mohammed looti

November 8, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Learning to Calculate Spearman’s Rank Correlation Coefficient in Excel: A Step-by-Step Guide*. PSYCHOLOGICAL STATISTICS. Retrieved from https://statistics.arabpsychology.com/?p=13511

In the expansive discipline of statistics, the concept of **correlation** stands as a foundational principle. It serves the crucial function of quantifying both the strength and the specific direction of the linear or monotonic relationship that exists between any two distinct numerical variables. A thorough understanding of correlation is invaluable for researchers, data scientists, and analysts, enabling them to make informed predictions about potential outcomes and accurately observe how fluctuations in one variable may exert an influence on the dynamics of the other. The intensity of this measured relationship is encapsulated by the **correlation coefficient**, a highly standardized metric that is universally confined to a precise numerical range, spanning from -1 to +1.

This universal scale provides immediate and robust insight into the fundamental nature of the relationship under observation, allowing for exceptionally clear and standardized interpretation across various fields of study:

-1 (Perfect Negative Correlation): This specific value signifies a perfect, inverse relationship between the two variables. It means that as the value of the first variable increases consistently, the value of the second variable decreases consistently, maintaining a precise and proportional relationship throughout the dataset.

0 (No Linear Correlation): A coefficient that registers exactly zero indicates that no detectable linear relationship exists between the two variables. They are understood to move independently of one another, with changes in one not systematically predicting changes in the other.

1 (Perfect Positive Correlation): This value represents a perfect, direct relationship. As the value of the first variable increases, the value of the second variable also increases consistently and proportionally.

While the most frequently employed measure is the parametric Pearson correlation coefficient, a highly specialized technique known as the Spearman Rank Correlation (often referred to as Spearman's rho) is frequently the preferred method in situations where the data is either non-normally distributed, or when the variables inherently represent ordered categories (ordinal data). This robust method specifically measures the correlation coefficient between the ranked positions of the observations, rather than relying on their raw, absolute numerical values. This makes it particularly suitable for assessing the monotonic relationship between, for example, a student's percentile rank in a mathematics examination versus their corresponding rank in a science examination within the same academic cohort. This comprehensive tutorial provides a step-by-step, practical guide detailing exactly how to calculate the **Spearman rank correlation coefficient** with precision and accuracy utilizing the powerful features of Microsoft Excel.

The Core Principles of Spearman Rank Correlation

The **Spearman Rank Correlation** (symbolized by the Greek letter ρ or *r*s) is fundamentally a non-parametric measure designed to rigorously assess how effectively the relationship between two variables can be described using a monotonic function. Crucially, unlike the Pearson correlation, which strictly searches for a linear pattern, Spearman's correlation only mandates that as one variable demonstrates an increasing trend, the other variable tends to either consistently increase or consistently decrease, regardless of the precise rate or form of that change. This foundational characteristic renders the Spearman method highly robust against the influence of extreme outliers and makes it exceptionally useful for analyzing ordinal data or continuous data where the underlying assumption of normal distribution cannot be reliably met or is demonstrably violated.

The methodological core of Spearman's technique revolves around the strategic conversion of all raw data points into their corresponding ordinal ranks. In this process, the observation with the highest numerical score is conventionally assigned rank 1, the second-highest score receives rank 2, and the process continues sequentially. A vital mathematical detail arises when ties occur (e.g., two or more students achieving the exact same score): in such instances, the average of the ranks that would have been assigned to those observations is instead uniformly assigned to all tied observations. This crucial step is usually handled automatically and accurately by sophisticated statistical software packages, including Excel. Once the original raw scores have been systematically replaced by their corresponding rank variables, the standard Pearson correlation formula is then applied directly to these newly generated rank variables. This process ensures that the statistical focus is maintained solely on the relative ordering and comparative structure of the data points, rather than on the potentially skewed or non-normal distribution of the raw values.

The applications of this powerful non-parametric technique are incredibly widespread, extending across numerous academic and professional disciplines, including psychology, ecology, finance, and social sciences. To illustrate the practical methodology, we will now apply this Spearman technique to a dataset to meticulously determine the relationship between the academic performance ranks of ten students across two distinct subjects: Mathematics and Science. This detailed, practical exercise will systematically solidify the required procedural steps necessary for accurate and reliable calculation within the Microsoft Excel environment.

Setting Up the Data in Microsoft Excel (Step 1: Data Entry)

To initiate the calculation process for the **Spearman rank correlation coefficient**, the first and most critical step is the precise organization of the raw dataset. For the purposes of this guiding example, we will utilize the raw examination scores in Math and Science for a modest sample size of 10 students. The procedure begins by meticulously entering the raw numerical data into two adjacent, dedicated columns within a new Excel worksheet. It is considered best practice, and highly recommended, to label the columns clearly and descriptively (e.g., "Student," "Math Score," and "Science Score") to diligently maintain data integrity, maximize clarity, and ensure readability throughout the entire subsequent calculation process.

For enhanced clarity, we will assume that the student names are systematically listed in Column A, the raw Math scores occupy Column B, and the raw Science scores are entered into Column C. This initial setup is fundamentally straightforward, yet it represents the absolute foundation upon which all subsequent ranking and correlation calculations will be meticulously built. It is imperative to ensure that the initial data entry is absolutely precise, as any inaccuracies introduced at this stage will inevitably propagate and distort the results through the ranking and final correlation stages. The visual representation below displays this initial, foundational data entry stage. Note the clear and logical organization of the scores, which prepares them optimally for the essential ranking process--the defining characteristic of the robust Spearman method.

	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					

The Critical Step: Calculating Ranks (Step 2: Transformation)

The decisive element that distinguishes the Spearman correlation is the necessary transformation of the raw numerical scores into corresponding ordinal ranks. This essential step is accomplished efficiently and accurately in Excel by employing a specific function that has been engineered to correctly handle instances of tied scores. We must now designate and create two new adjacent columns within the spreadsheet--one specifically dedicated to storing the Math Ranks (Column D) and a second for the Science Ranks (Column E)--to house the newly transformed data.

The function officially recommended by Excel for this precise ranking purpose is the **RANK.AVG()** function. This powerful function is mathematically designed to assign the statistically correct average rank to any scores that are found to be tied, which is the mathematically required procedure for accurately calculating Spearman's rho. The syntax structure for this function requires the input of three distinct arguments: first, the specific number (score) that you intend to rank;

second, the absolute range of all numbers within the dataset (the reference array); and third, the order parameter (conventionally 0 for descending order, which ensures that the highest numerical score is correctly assigned the rank of 1).

To calculate the ranks for the first student, Austin, located in row 2, the specific formulas utilized to calculate the Math rank (Cell D2) and the Science rank (Cell E2) are presented below. It is absolutely crucial to note the mandatory use of absolute references (signified by the inclusion of dollar signs, e.g., `$B$2:$B$11`) for the reference range argument. This precise application ensures that when the formula is subsequently copied or dragged down to calculate the remaining ranks, the ranking reference range remains fixed and constant, effectively preventing catastrophic calculation errors:

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

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

Immediately following the successful entry of these two formulas into cells D2 and E2, the preliminary rank calculation for the first observation will be clearly visible, as demonstrated in the image provided directly below:

	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						

To efficiently complete the ranking operation for all 10 students in the sample size, simply select both cells D2 and E2 simultaneously, and then utilize the fill handle to drag the selection down to perfectly encompass the remaining rows (D3 through E11). As an alternative, a more rapid method is to highlight the entire target range (D2:E11) and then use the keyboard shortcut **Ctrl+D** (Fill Down) to automatically and instantaneously populate the ranks for every single student within the dataset. This critical action systematically transforms all the raw scores into their accurate corresponding ordinal ranks, thereby preparing the data for the final, conclusive correlation coefficient calculation.

	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						

The resulting data table, now comprehensively populated with the calculated ranks for both academic subjects, definitively confirms that the essential transformation step has been executed successfully. We now possess two pristine columns of purely ranked data (Columns D and E) that accurately represent the relative performance ordering of the students, effectively divorced from the absolute numerical value and potential distribution issues of their original scores.

	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						

Determining the Correlation Coefficient (Step 3: Calculation)

With the dataset successfully and accurately converted into ranked variables, the final determination of the **Spearman rank correlation coefficient** becomes surprisingly straightforward within the Excel environment. This simplicity arises because Spearman's method is, by definition, mathematically equivalent to applying the traditional Pearson correlation formula specifically to the ranks. Consequently, we can conveniently and reliably utilize Excel's powerful, built-in statistical function, **CORREL()**.

The **CORREL()** function requires the input of precisely two arguments: Array1, which is the range encompassing the first set of values (the independent variable), and Array2, which is the range encompassing the second set of values (the dependent variable). In the context of our current analysis, Array1 must be defined as the range containing the Math Ranks (D2:D11), and Array2 must be defined as the range containing the Science Ranks (E2:E11). It is critically important to strictly select the columns containing the ranks and absolutely avoid selecting the columns containing the original raw scores for this final step.

We can strategically place the resulting formula in any vacant cell located outside the main data table (for instance, cell F14) and structure the calculation precisely as follows:

```
=CORREL(D2:D11, E2:E11)
```

The image provided immediately below visually illustrates the formula being accurately entered and subsequently executed, clearly demonstrating the final, decisive step in the correlation calculation procedure:

	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					

Upon successful execution, the calculated **correlation coefficient** for this specific dataset is revealed to be **-0.41818**. This obtained negative value definitively signals the presence of a moderate, inverse monotonic relationship between a student's rank in Mathematics and their rank in Science. The result suggests a general trend where students who achieved a higher rank in Math tended to achieve a slightly lower rank in Science, and vice versa. However, merely calculating this coefficient is often deemed insufficient for rigorous academic or professional analysis; the next essential step involves determining whether this observed relationship holds genuine statistical meaning.

	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					

Interpreting and Testing for Statistical Significance (Step 4: Hypothesis Testing)

The calculated correlation coefficient of **-0.41818** provides concrete information regarding the magnitude and the precise direction of the observed monotonic relationship, unequivocally confirming a negative correlation. Nevertheless, within the framework of rigorous statistical analysis, it is an absolute requirement to assess whether this result is indicative of a true underlying relationship present within the larger population, or if it is merely an artifact of random sampling chance inherent in the specific sample selected. This crucial assessment is precisely the objective of testing for statistical significance.

To accurately determine the significance of the Spearman rank correlation, researchers traditionally rely upon a pre-calculated table of critical values, which is based on the underlying distribution of the test statistic. These authoritative tables meticulously account for both the specific sample size (denoted as *n*) and the chosen significance level (alpha, conventionally set at α = 0.05). The critical value itself represents the minimum absolute value that the calculated coefficient must either meet or exceed in order to be confidently considered statistically significant at the pre-determined alpha level.

The established decision rule for hypothesis testing is transparent and highly straightforward: If the **absolute value** of our calculated correlation coefficient (in our case, |-0.41818| = 0.41818) is numerically greater than the critical value retrieved from the table for our specific sample size and alpha level, then we are justified in concluding that the observed correlation is statistically significant. Conversely, if the absolute value is numerically less than the critical value, the relationship is deemed statistically insignificant, meaning we do not possess sufficient evidence to confidently reject the null hypothesis, which posits that no correlation genuinely exists in the population.

	α				
n	**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

Referring to the provided table of critical values, we must now precisely identify our relevant parameters. Our sample size is n = 10 students. Using the industry-standard significance level of 0.05 (representing a two-tailed test), we trace along the row corresponding to n = 10 to the specific column marked α = 0.05, which yields a critical value of **0.564**. Since our calculated absolute correlation coefficient (**0.41818**) is not numerically larger than this critical value (0.564), we are compelled to conclude that the observed negative correlation between the Math and Science performance ranks is **not statistically significant**. While a negative relationship was indeed found within the confines of this specific sample, the statistical evidence is ultimately not strong enough to confidently suggest that this relationship holds true for the larger student population from which this limited sample was originally drawn.