

# How to Easily Calculate Intraclass Correlation Coefficient (ICC) in Excel

Authored by  
**stats writer**

December 6, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to Easily Calculate Intraclass Correlation Coefficient (ICC) in Excel*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=106458>

The Intraclass Correlation Coefficient (ICC) is a powerful statistical measure fundamental to assessing the consistency or absolute agreement among multiple observations or ratings of the same subjects. Unlike Pearson's correlation, which measures the linear association between two variables, the ICC specifically quantifies the degree of similarity or homogeneity within groups relative to the total variation. This makes it the preferred metric for determining inter-rater reliability, especially when dealing with quantitative or continuous data derived from two or more raters or measurement methods.

When researchers seek to validate a measurement instrument or ensure standardization across multiple assessors, calculating the ICC becomes essential. Its value inherently reflects how much of the total variability in the measurements is attributable to differences between the subjects being rated, as opposed to variability caused by inconsistencies between the raters themselves. A high ICC value suggests strong agreement, indicating that the observed differences in scores primarily reflect true differences among the subjects, rather than measurement error introduced by the raters.

While specialized statistical software like R or SPSS often handle ICC calculations, Microsoft Excel offers a robust method using the built-in Analysis Toolpak add-on, leveraging the results of an Analysis of Variance (ANOVA). This guide provides a comprehensive, step-by-step approach to calculating and interpreting the ICC for scenarios involving multiple raters and multiple subjects directly within the Excel environment, ensuring clean and reliable results.

## Understanding the Intraclass Correlation Coefficient (ICC)

The concept of the Intraclass Correlation Coefficient is rooted in partitioning the variance observed in a dataset. In simple terms, it compares the variability of scores between the subjects (the signal) against the variability of scores within the subjects (the noise, often measurement error or rater disagreement). The resulting coefficient ranges strictly from 0 to 1. A value of 0 signifies absolutely no reliability or agreement among the raters, meaning the differences between ratings are as large as the differences between the subjects. Conversely, a value of 1 indicates perfect agreement, where all raters assign identical scores to the same subject, suggesting flawless reliability.

It is crucial to distinguish ICC from inter-rater correlation measures like Pearson's  $r$ . Pearson's  $r$  only captures the linear relationship and does not detect systematic bias or differences in means between raters. For instance, two raters could have a high Pearson's correlation if Rater A consistently scores 10 points higher than Rater B, even though their absolute agreement is low. The ICC, particularly in certain models, specifically accounts for these systematic differences, making it a superior measure for evaluating true agreement and reliability in measurement studies. This emphasis on absolute agreement ensures that the measurement standard is truly consistent across all observers.

The chosen statistical framework for calculating the ICC involves ANOVA, as ANOVA provides the necessary Mean Squares (MS) components--Mean Squares Between Subjects, Mean Squares Within Subjects, and Mean Squares Error--which are the building blocks of the ICC formula. By utilizing the variance components derived from the ANOVA table, we can construct a ratio that precisely reflects the proportion of variance attributable to the differences between subjects that we are interested in measuring.

## The Role of ICC in Reliability Testing

The primary application of the ICC lies in studies requiring robust evidence of measurement consistency. This is particularly relevant in fields such as psychology, education, medicine, and engineering, where assessments often rely on subjective judgment or non-standardized measurement equipment. If a clinical diagnosis or educational assessment score heavily depends on the specific rater involved, the measurement system lacks statistical reliability, undermining the validity of any conclusions drawn.

Consider the example we will explore: multiple judges rating college entrance exams. If Judge 1 rates a specific exam as excellent (score 90) while Judge 4 rates the same exam as poor (score 50), the system exhibits low reliability. The ICC computation formally quantifies this discrepancy. By aiming for a high ICC (typically above 0.75 or 0.80), researchers establish confidence that their measurement process is stable and objective, minimizing the influence of idiosyncratic rater variability. This confirmation is often a required component of psychometric validation studies.

Before proceeding with the calculation, it is necessary to confirm that the Excel environment is properly set up. The technique relies entirely on accessing the specialized statistical functions provided by the Analysis Toolpak, which is not activated by default in most Excel installations. Ensuring this crucial add-on is available saves time and guarantees access to the sophisticated ANOVA procedure required for this methodology.

## Prerequisites: Setting Up the Data and Analysis Toolpak

To begin the calculation process, two main prerequisites must be fulfilled: the data must be correctly structured, and the statistical add-on must be enabled. For the ICC calculation using the ANOVA approach in Excel, the data must be organized in a rectangular format, where subjects (or items) occupy the rows and the raters (or observations) occupy the columns. This specific matrix structure is mandatory for the Excel ANOVA function to correctly identify the factors and levels necessary for partitioning variance.

The second prerequisite involves activating the Data Analysis Toolpak. If you click the **Data** tab and do not see the **Data Analysis** option in the Analysis group on the far right of the ribbon, you must enable it. This typically involves navigating to the Excel Options menu, selecting Add-ins,

choosing Excel Add-ins from the Manage dropdown, and checking the box next to "Analysis Toolpak." Once activated, this tool provides access to specialized statistical procedures, including the critical **Anova: Two-Factor Without Replication** test, which we will employ.

Failure to correctly structure the data or activate the Analysis Toolpak will prevent the successful completion of the subsequent steps. This method is highly sensitive to the input format, as the ANOVA procedure treats rows as one factor (subjects) and columns as the other factor (raters), which are essential for deriving the necessary Mean Squares values for the ICC formula.

## Step 1: Structuring the Dataset in Excel

For illustrative purposes, assume a scenario where 10 distinct college entrance exams (our subjects) were independently rated by four different judges (our raters). Each cell in the matrix represents the score given by a specific judge to a specific exam. The proper structure is critical for defining the factors in the subsequent statistical analysis.

We label the rows (A2:A11) as the different exams and the columns (B1:E1) as the different judges. The dataset must be contiguous, meaning there should be no empty rows or columns within the input range selected for the ANOVA analysis. This arrangement defines our experiment as a measurement consistency study where all subjects are rated by all raters--a classic balanced design required for the Two-Factor Without Replication model.

The data might look like this, representing the scores assigned by the four judges (J1 through J4) to the ten exams (E1 through E10):

	A	B	C	D	E	F	G
1	Exam	Judge A	Judge B	Judge C	Judge D		
2	1	1	2	0	1		
3	2	1	3	4	2		
4	3	3	8	1	3		
5	4	6	4	5	3		
6	5	6	5	5	6		
7	6	7	5	6	4		
8	7	8	7	6	6		
9	8	9	9	9	8		
10	9	8	8	8	8		
11	10	7	8	8	9		
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							

Note that the inclusion of header labels (Exam, Judge 1, Judge 2, etc.) is highly recommended for clarity and must be accounted for when selecting the input range in the next step, ensuring the "Labels" box is checked in the ANOVA setup window.

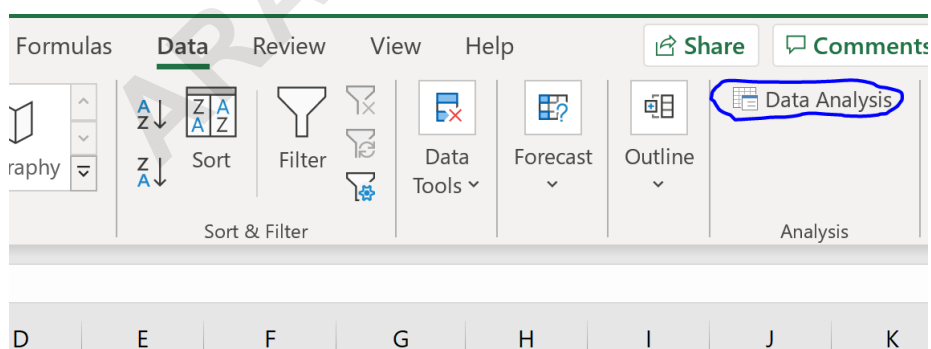
## Step 2: Executing the Two-Factor ANOVA Without Replication

The core statistical engine for calculating the ICC in this specific scenario is the **Anova: Two-Factor Without Replication** test. This model is chosen because we have two factors--Subjects (Exams) and Raters (Judges)--and each subject provides only one score per rater (i.e., there are no repeated measures within the cells of the matrix). The use of this particular ANOVA partition allows us to isolate the variance due to subjects, the variance due to raters, and the residual error variance.

To execute the analysis, navigate to the **Data** tab and click the **Data Analysis** option. From the list of statistical tools, select **Anova: Two-Factor Without Replication** and click **OK**. The Input Range should encompass the entire dataset, including the row and column headers, as illustrated below. In our example, this corresponds to cells A1:E11:

	A	B	C	D	E	F	G
1	Exam	Judge A	Judge B	Judge C	Judge D		
2	1	1	2	0	1		
3	2	1	3	4	2		
4	3	3	8	1	3		
5	4	6	4	5	3		
6	5	6	5	5	6		
7	6	7	5	6	4		
8	7	8	7	6	6		
9	8	9	9	9	8		
10	9	8	8	8	8		
11	10	7	8	8	9		
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							

Upon opening the ANOVA dialogue box, ensure the following parameters are correctly configured: the Input Range (A1:E11), Grouped By Rows (indicating subjects are grouped by rows), and the checkbox for Labels in First Row and First Column must be selected. Finally, specify an Output Range where the ANOVA results will be displayed, typically on the same sheet or a new worksheet. Clicking **OK** generates the extensive ANOVA output tables.



If the **Data Analysis** option is missing, it confirms that the Analysis Toolpak has not been successfully loaded. Review the prerequisite steps to ensure the add-in is enabled before proceeding, as the manual calculation of these variance components is complex and error-prone.

Below is the configuration window for the ANOVA analysis, showing the input range and selection settings:

	A	B	C	D	E	F	G
1	Exam	Judge A	Judge B	Judge C	Judge D		
2	1	1	2	0	1		
3	2	1	3	4	2		
4	3	3	8	1	3		
5	4	6	4	5	3		
6	5	6	5	5	6		
7	6	7	5	6	4		
8	7	8	7	6	6		
9	8	9	9	9	8		
10	9	8	8	8	8		
11	10	7	8	8	9		

Anova: Two-Factor Without Replication

Input Range:

Labels

Alpha:

Output options

Output Range:

New Worksheet Ply:

New Workbook

The resulting output contains descriptive statistics for each subject and rater, but the critical information for the ICC lies within the ANOVA table itself.

## Interpreting the ANOVA Output for ICC Calculation

Once the ANOVA is executed, a detailed set of results appears. We are primarily concerned with the second table, labeled "ANOVA," which decomposes the total variance into three key components: Between Samples (Rows, or Exams/Subjects), Between Columns (Judges/Raters), and Error (Residual). Each component is associated with its respective Sum of Squares (SS), Degrees of Freedom (df), and, most importantly, the Mean Squares (MS).

The Mean Squares (MS) values are the heart of the ICC calculation. MS is calculated by dividing SS by df, and it represents an estimate of the variance associated with that specific source of variation. Specifically, we need three values from the ANOVA output for the calculation of the ICC (assuming the common Two-Way Random Effects, Absolute Agreement model):

**MSRows (MSBetween Subjects):** Reflects variability between the exams (the subjects we are measuring). This is the desired signal.

**MSColumns (MSBetween Raters):** Reflects systematic differences in the mean scores assigned by the judges. This component accounts for systematic bias.

**MSError (MSResidual):** Represents the unexplained variance or interaction effect, often treated as random error or measurement noise.

The ANOVA table provides these necessary values, typically presented as follows:

G	H	I	J	K	L	M	N
Anova: Two-Factor Without Replication							
	<i>SUMMARY</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>		
	1	4	4	1	0.666667		
	2	4	10	2.5	1.666667		
	3	4	15	3.75	8.916667		
	4	4	18	4.5	1.666667		
	5	4	22	5.5	0.333333		
	6	4	22	5.5	1.666667		
	7	4	27	6.75	0.916667		
	8	4	35	8.75	0.25		
	9	4	32	8	0		
	10	4	32	8	0.666667		
	Judge A	10	56	5.6	8.488889		
	Judge B	10	59	5.9	5.877778		
	Judge C	10	52	5.2	8.622222		
	Judge D	10	50	5	7.777778		
	ANOVA						
	<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
	Rows	231.525	9	25.725	15.30744	1.85E-08	2.250131
	Columns	4.875	3	1.625	0.966942	0.422653	2.960351
	Error	45.375	27	1.680556			
	Total	281.775	39				

From this table, we extract MSRows, MSColumns, and MSError. We also need to know k, which is the number of raters (in this case, k=4), and n, the number of subjects (n=10). These components will be substituted directly into the Intraclass Correlation Coefficient formula in the next step to derive the final reliability score.

### Step 3: Calculate the Intraclass Correlation Coefficient

We can use the following formula (ICC(2,1) for Single Raters, Absolute Agreement) to calculate the ICC among the raters, where k is the number of raters (k=4). This formula is based on the components derived from the ANOVA variance partitioning:

G	H	I	J	K	L	M	N
Anova: Two-Factor Without Replication							
	<i>SUMMARY</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>		
	1	4	4	1	0.666667		
	2	4	10	2.5	1.666667		
	3	4	15	3.75	8.916667		
	4	4	18	4.5	1.666667		
	5	4	22	5.5	0.333333		
	6	4	22	5.5	1.666667		
	7	4	27	6.75	0.916667		
	8	4	35	8.75	0.25		
	9	4	32	8	0		
	10	4	32	8	0.666667		
	Judge A	10	56	5.6	8.488889		
	Judge B	10	59	5.9	5.877778		
	Judge C	10	52	5.2	8.622222		
	Judge D	10	50	5	7.777778		
	ANOVA						
	<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
	Rows	231.525	9	25.725	15.30744	1.85E-08	2.250131
	Columns	4.875	3	1.625	0.966942	0.422653	2.960351
	Error	45.375	27	1.680556			
	Total	281.775	39				
ICC	0.782	$=(K23-K25)/((K23+J24*K25+(J24+1)*(K24-K25))/(J23+1))$					

By substituting the MS values from the ANOVA table into the appropriate cells in Excel, the intraclass correlation coefficient (ICC) turns out to be **0.782**.

Here is how to interpret the value of an intraclass correlation coefficient, according to conventional guidelines:

**Less than 0.50:** Poor reliability

**Between 0.5 and 0.75:** Moderate reliability

**Between 0.75 and 0.9:** Good reliability

**Greater than 0.9:** Excellent reliability

Thus, we would conclude that an ICC of **0.782** indicates that the exams can be rated with "good" reliability by different raters, confirming strong inter-rater agreement.

## A Note on Calculating ICC Models

There are several different versions of an ICC that can be calculated, depending on the following three factors, which dictate the specific formula derived from the ANOVA variance components:

**Model:** One-Way Random Effects, Two-Way Random Effects, or Two-Way Mixed Effects

**Type of Relationship:** Consistency or Absolute Agreement

**Unit:** Single rater or the mean of raters

In the previous example, the ICC that we calculated used the standard assumptions for assessing absolute agreement among a randomly selected pool of raters:

**Model:** Two-Way Random Effects

**Type of Relationship:** Absolute Agreement

**Unit:** Single rater

For a detailed explanation of these assumptions, particularly the distinction between consistency and absolute agreement and the implications of fixed vs. random effects models, researchers should consult foundational psychometric literature to ensure the correct statistic is reported.

The Intraclass Correlation Coefficient (ICC) is used to determine if items (or subjects) can be rated reliably by different raters.

The value of an ICC can range from 0 to 1, with 0 indicating no agreement among raters and 1 indicating perfect reliability.

This tutorial provides a step-by-step example of how to calculate ICC in Excel, leveraging the Analysis Toolpak.

## Step 1: Create the Data

Suppose four different judges were asked to rate the quality of 10 different college entrance exams. The results are shown below:

	A	B	C	D	E	F	G
1	Exam	Judge A	Judge B	Judge C	Judge D		
2	1	1	2	0	1		
3	2	1	3	4	2		
4	3	3	8	1	3		
5	4	6	4	5	3		
6	5	6	5	5	6		
7	6	7	5	6	4		
8	7	8	7	6	6		
9	8	9	9	9	8		
10	9	8	8	8	8		
11	10	7	8	8	9		
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							

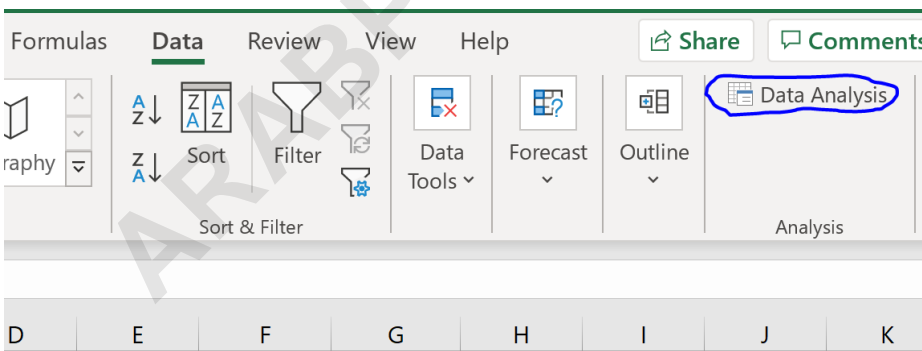
## Step 2: Fit an ANOVA

In order to calculate the ICC for these ratings, we first need to fit an **Anova: Two-Factor Without Replication**. This partitioning of variance is essential for the calculation.

To do so, highlight cells A1:E11, encompassing all data and labels, as follows:

	A	B	C	D	E	F	G
1	Exam	Judge A	Judge B	Judge C	Judge D		
2	1	1	2	0	1		
3	2	1	3	4	2		
4	3	3	8	1	3		
5	4	6	4	5	3		
6	5	6	5	5	6		
7	6	7	5	6	4		
8	7	8	7	6	6		
9	8	9	9	9	8		
10	9	8	8	8	8		
11	10	7	8	8	9		
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							

Next, click the **Data** tab along the top ribbon and then click the **Data Analysis** option under the **Analysis** group:



If you don't see this option available, you need to first install the [Analysis Toolpak](#).

In the dropdown menu that appears, click **Anova: Two-Factor Without Replication** and then click **OK**. In the new window that appears, fill in the following information, ensuring the Input Range is correct and Labels is checked, and then click **OK**:

	A	B	C	D	E	F	G
1	Exam	Judge A	Judge B	Judge C	Judge D		
2	1	1	2	0	1		
3	2	1	3	4	2		
4	3	3	8	1	3		
5	4	6	4	5	3		
6	5	6	5	5	6		
7	6	7	5	6	4		
8	7	8	7	6	6		
9	8	9	9	9	8		
10	9	8	8	8	8		
11	10	7	8	8	9		

Anova: Two-Factor Without Replication

Input  
 Input Range:

Labels  
 Alpha:

Output options  
 Output Range:    
 New Worksheet Ply:   
 New Workbook

The following ANOVA results will appear, providing the necessary Mean Squares (MS) values:

G	H	I	J	K	L	M	N
Anova: Two-Factor Without Replication							
	<i>SUMMARY</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>		
	1	4	4	1	0.666667		
	2	4	10	2.5	1.666667		
	3	4	15	3.75	8.916667		
	4	4	18	4.5	1.666667		
	5	4	22	5.5	0.333333		
	6	4	22	5.5	1.666667		
	7	4	27	6.75	0.916667		
	8	4	35	8.75	0.25		
	9	4	32	8	0		
	10	4	32	8	0.666667		
	Judge A	10	56	5.6	8.488889		
	Judge B	10	59	5.9	5.877778		
	Judge C	10	52	5.2	8.622222		
	Judge D	10	50	5	7.777778		
	ANOVA						
	<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
	Rows	231.525	9	25.725	15.30744	1.85E-08	2.250131
	Columns	4.875	3	1.625	0.966942	0.422653	2.960351
	Error	45.375	27	1.680556			
	Total	281.775	39				

### Step 3: Calculate the Intraclass Correlation Coefficient

We can use the following formula (ICC(2,1)) to calculate the ICC among the raters, utilizing the MS components derived from the ANOVA:

G	H	I	J	K	L	M	N
Anova: Two-Factor Without Replication							
	<i>SUMMARY</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>		
	1	4	4	1	0.666667		
	2	4	10	2.5	1.666667		
	3	4	15	3.75	8.916667		
	4	4	18	4.5	1.666667		
	5	4	22	5.5	0.333333		
	6	4	22	5.5	1.666667		
	7	4	27	6.75	0.916667		
	8	4	35	8.75	0.25		
	9	4	32	8	0		
	10	4	32	8	0.666667		
	Judge A	10	56	5.6	8.488889		
	Judge B	10	59	5.9	5.877778		
	Judge C	10	52	5.2	8.622222		
	Judge D	10	50	5	7.777778		
	ANOVA						
	<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
	Rows	231.525	9	25.725	15.30744	1.85E-08	2.250131
	Columns	4.875	3	1.625	0.966942	0.422653	2.960351
	Error	45.375	27	1.680556			
	Total	281.775	39				
ICC	0.782	=(K23-K25)/(K23+J24*K25+(J24+1)*(K24-K25)/(J23+1))					

The intraclass correlation coefficient (ICC) turns out to be **0.782**.

Here is how to interpret the value of an intraclass correlation coefficient, according to standard reliability scales:

**Less than 0.50:** Poor reliability

**Between 0.5 and 0.75:** Moderate reliability

**Between 0.75 and 0.9:** Good reliability

**Greater than 0.9:** Excellent reliability

Thus, we would conclude that an ICC of **0.782** indicates that the exams can be rated with "good" reliability by different raters.

## A Note on Calculating ICC Models

There are several different versions of an ICC that can be calculated, depending on the following three factors, which relate to the experimental design:

**Model:** One-Way Random Effects, Two-Way Random Effects, or Two-Way Mixed Effects

**Type of Relationship:** Consistency or Absolute Agreement

**Unit:** Single rater or the mean of raters

In the previous example, the ICC that we calculated used the following assumptions:

**Model:** Two-Way Random Effects

**Type of Relationship:** Absolute Agreement

**Unit:** Single rater

For a detailed explanation of the differences between these models and when to apply them, consult established statistical resources.