Understanding and Calculating the Intraclass Correlation Coefficient (ICC) in Excel

Authored by

Mohammed looti

November 5, 2025

The Intraclass Correlation Coefficient (ICC) stands as a cornerstone in research methodology, serving as a vital reliability statistic. It is specifically designed to quantify the degree of agreement or consistency between multiple quantitative measurements taken by different observers, instruments, or raters on the same set of subjects or items. Understanding the ICC is essential for validating measurement instruments and ensuring data quality in quantitative studies.

The calculated ICC value is constrained to fall strictly between 0 and 1. A score near 0 indicates negligible consistency or **poor reliability** among the measurements, suggesting high variability or systematic disagreement. Conversely, an ICC approaching 1 signifies **near-perfect agreement** and robust reliability, confirming that the raters are applying the measurement criteria consistently. This detailed guide provides a precise, step-by-step methodology for calculating this essential coefficient using **Microsoft Excel**, leveraging its built-in analytical capabilities.

Step 1: Preparing and Structuring the Data Set for Analysis

To execute the ICC calculation successfully in Excel, the data must be meticulously organized in a specific format suitable for Analysis of Variance (ANOVA). We will use a typical scenario involving inter-rater reliability: four independent judges (labeled Raters 1 through 4) assessing the quality of ten distinct items (Items 1 through 10), such as college entrance essays or clinical observations.

The critical structural requirement is that the items (the subjects being rated) must be arranged down the rows, while the raters (the sources of measurement) must be arranged across the columns. This organization ensures Excel correctly partitions the variance attributable to differences among items versus differences among raters. The raw scores for our hypothetical example are presented in the spreadsheet snippet 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							

Adhering to this row-by-item and column-by-rater structure is the foundational prerequisite for proceeding to the statistical analysis in the next step.

Step 2: Executing the Two-Factor Without Replication ANOVA

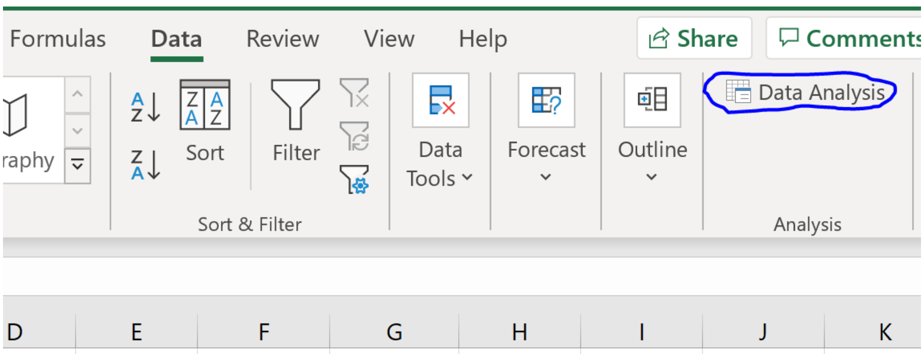
The Intraclass Correlation Coefficient is fundamentally calculated by comparing different sources of variability within the data. This requires applying an ANOVA model. For reliability studies where every rater assesses every item once, the appropriate model is the **Anova: Two-Factor Without Replication**. This model is ideal because it accounts for variance caused by the items themselves (rows) and variance caused by the raters (columns), while assuming no interaction effect.

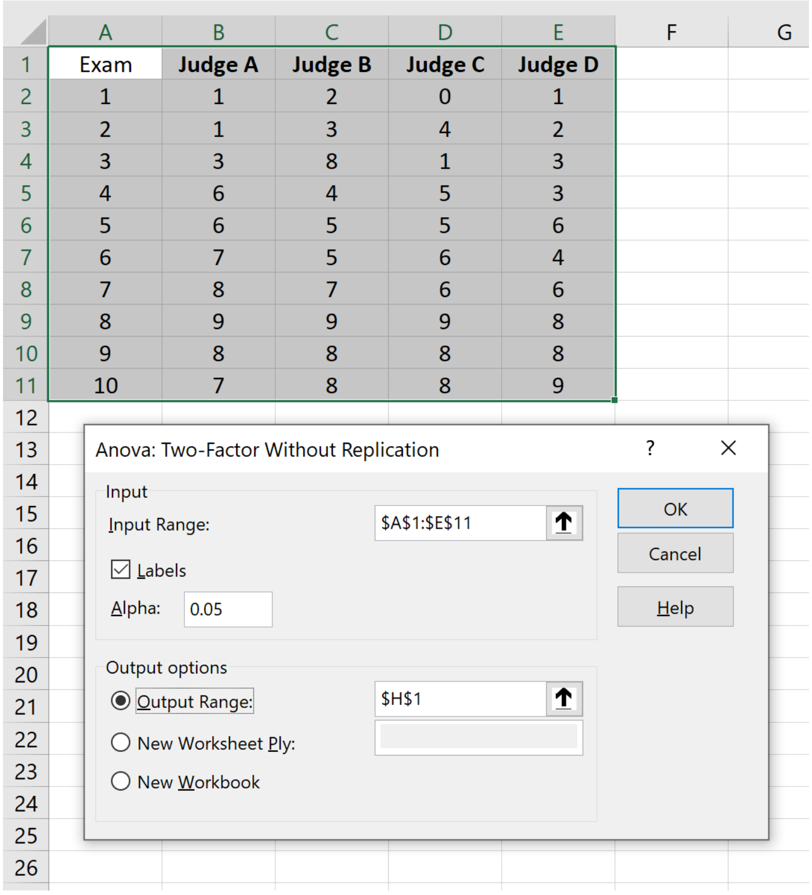
To initiate the analysis, first select the entire data range, including the row and column headers (cells A1:E11) as illustrated below. Headers are crucial for proper data labeling:

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

Next, access the analytical tools via the **Data** tab on the Excel ribbon. Locate the **Analysis** group on the far right and click the **Data Analysis** button. If this option is missing, you must first enable the [Data Analysis ToolPak](#) add-in through Excel Options before proceeding.

In the Data Analysis dialog box, select **Anova: Two-Factor Without Replication**. Ensure the input range matches your selection (A1:E11), confirm that the **Labels** box is checked to include headers, and specify an output location, typically a new worksheet. Proceed by clicking **OK**:





Step 3: Extracting Key Mean Square Values from the ANOVA Summary

Once the ANOVA is executed, Excel produces a comprehensive summary table, usually placed on a new spreadsheet tab. This table serves as the foundation for the ICC calculation, as the coefficient itself is a ratio of variance components derived directly from this output.

We must carefully isolate three specific values from the "MS" (Mean Square) column, which represent the variance estimates necessary for the ICC formula:

MS for Rows (MSR): Represents the variance attributed to differences among the items (the subjects being rated).
MS for Columns (MSC): Represents the variance attributed to differences among the raters (systematic bias).
MS for Error (MSE): Represents the residual or unexplained variance (random error).

The image below displays the typical structure of the ANOVA output, clearly indicating where these critical Mean Square values are located in the summary table:

Anova: Two-Factor Without Replication

SUMMARY	*Count*	*Sum*	*Average*	*Variance*
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

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
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 4: Applying the Intraclass Correlation Coefficient Formula

The calculation we perform here corresponds to the ICC formula for the **two-way random effects model** focusing on **absolute agreement** among single measures. This particular model is the most stringent test of reliability, requiring raters to assign scores that are numerically identical, not just rank-consistent. The formula integrates the Mean Square values derived in the previous step:

Anova: Two-Factor Without Replication

SUMMARY	*Count*	*Sum*	*Average*	*Variance*
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

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
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))

In this context, the variables are defined as:

MSR: Mean Square for Rows (Variance due to Items)
MSC: Mean Square for Columns (Variance due to Raters)
MSE: Mean Square for Error (Residual Variance)
k: The number of raters (which is 4 in our current example)

By substituting the numerical results from the Excel ANOVA output into this equation, we arrive at an Intraclass Correlation Coefficient (ICC) result of approximately **0.782**. This numerical result now requires interpretation within established guidelines to assess the actual level of agreement.

Step 5: Interpreting the Reliability Magnitude

The final step in the process is giving practical meaning to the calculated ICC value. While a higher value always indicates better reliability, researchers often rely on established conventions to classify the strength of agreement. One widely cited framework for interpreting inter-rater reliability coefficients was provided by Cicchetti (1994), which categorizes reliability into four distinct levels based on the coefficient magnitude:

Less than 0.50: Indicates **Poor reliability**, suggesting the measurement system is inconsistent and unreliable.
Between 0.50 and 0.75: Suggests **Moderate reliability**, acceptable in exploratory research contexts.
Between 0.75 and 0.90: Demonstrates **Good reliability**, suitable for most applied settings where measurement precision is important.
Greater than 0.90: Signifies **Excellent reliability** or near-perfect consistency among raters.

Applying these standards to our result, the ICC of **0.782** falls squarely within the "Good reliability" range. We can confidently conclude that the panel of judges consistently applied the scoring criteria, validating the quality of the measurement process for the college entrance exams.

Step 6: Critical Considerations for Choosing the Correct ICC Model

It is crucial for accurate statistical reporting to recognize that the Intraclass Correlation Coefficient is not a monolithic statistic. A variety of ICC formulas exist, and selecting the correct version is entirely dependent on the specific design of the research study. Misapplication of the formula can lead to incorrect conclusions regarding consistency. The choice hinges on three defining factors:

Statistical Model: Determining whether the data fits a One-Way Random Effects, Two-Way Random Effects, or Two-Way Mixed Effects model. This dictates how variance components (e.g., rater bias) are treated in the calculation.
Type of Relationship Measured: The focus must be either on **Consistency** (do raters maintain the same relative ranking of items?) or **Absolute Agreement** (do raters provide nearly identical numerical scores?).
Unit of Reliability: Whether the researcher needs the reliability estimate for a **Single Rater** (how reliable is any one rater from the group?) or for the **Mean of all Raters** (how reliable is the average score of the entire panel?).

The standard calculation derived from the **Excel Two-Factor Without Replication ANOVA** inherently assumes a specific configuration, often referred to as ICC(A,1) in statistical literature: Two-Way Random Effects, Absolute Agreement, and Single Rater reliability. Researchers must meticulously match their research design to the appropriate ICC formula. Consulting specialized statistical textbooks or software documentation is recommended when dealing with complex data structures that deviate from this standard assumption.