

What is the definition and example of the Intraclass Correlation Coefficient?

Authored by
stats writer

April 27, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the definition and example of the Intraclass Correlation Coefficient?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=139954>

The Intraclass Correlation Coefficient (ICC) is a statistical measure used to assess the degree of similarity or agreement between multiple measurements or observations within a group or cluster. It is commonly used in research studies to evaluate the reliability or consistency of measurements taken by different individuals or methods.

An example of ICC would be a study where the blood pressure of a group of participants is measured by three different healthcare professionals. The ICC would then be used to determine the level of agreement between the three measurements, indicating how consistent the results were among the different professionals. A higher ICC value indicates a stronger agreement between the measurements, while a lower value suggests a lack of consistency. Overall, the ICC is a valuable tool for evaluating the reliability and reproducibility of data within a group or cluster.

Intraclass Correlation Coefficient: Definition + Example

An intraclass correlation coefficient (ICC) is used to measure the of ratings in studies where there are two or more raters.

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

In simple terms, an ICC is used to determine if items (or subjects) can be rated reliably by different raters.

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

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

Here's a brief description of the three different models:

1. One-way random effects model: This model assumes that each subject is rated by a different group of randomly chosen raters. Using this model, the raters are considered the source of random effects. This model is rarely used in practice because the same group of raters is usually used to rate each subject.

2. Two-way random effects model: This model assumes that a group of k raters is randomly selected from a population and then used to rate subjects. Using this model, both the raters and the subjects are considered sources of random effects. This model is often used when we'd like to generalize our findings to any raters who are similar to the raters used in the study.

3. Two-way mixed effects model: This model also assumes that a group of k raters is randomly selected from a population and then used to rate subjects. However, this model assumes that the group of raters we chose are the *only* raters of interest, which means

we aren't interested in generalizing our findings to any other raters who might also share similar characteristics as the raters used in the study.

Here's a brief description of the two different types of relationships we might be interested in measuring:

1. Consistency: We are interested in the systematic differences between the ratings of judges (e.g. did the judges rate similar subjects low and high?)

2. Absolute Agreement: We are interested in the absolute differences between the ratings of judges (e.g. what is the absolute difference in ratings between judge A and judge B?)

Here's a brief description of the two different units we might be interested in measuring:

1. Single rater: We are only interested in using the ratings from a single rater as the basis for measurement.

2. Mean of raters: We are interested in using the mean of ratings from all judges as the basis for measurement.

Note: If you want to measure the level of agreement between two raters who each rate items on a , you should instead use .

How to Interpret Intraclass Correlation Coefficient

**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**

The following example shows how to calculate an intraclass correlation coefficient in practice.

Example: Calculating Intraclass Correlation Coefficient

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

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

Suppose the four judges were randomly selected from a population of qualified entrance exam judges and that we'd like to measure the absolute agreement among judges and that we're interested in using the ratings from a single rater perspective as the basis for our measurement.

We can use the following code in R to fit a two-way random effects model, using absolute agreement as the relationship among raters, and using single as our unit of interest:

```
#load the interrater reliability packagelibrary(irr)
```

```
#define data
```

```
data <- data.frame(A=c(1, 1, 3, 6, 6, 7, 8, 9, 8, 7),
```

B=c(2, 3, 8, 4, 5, 5, 7, 9, 8, 8),

C=c(0, 4, 1, 5, 5, 6, 6, 9, 8, 8),

D=c(1, 2, 3, 3, 6, 4, 6, 8, 8, 9))

#calculate ICC

icc(data, model = "twoway", type = "agreement", unit = "single")

Model: twoway

Type : agreement

Subjects = 10

Raters = 4

ICC(A,1) = 0.782

F-Test, H0: $r_0 = 0$; H1: $r_0 > 0$

F(9,30) = 15.3 , p = 5.93e-09

95%-Confidence Interval for ICC Population Values:

0.554 < ICC < 0.931

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

Based on the rules of thumb for interpreting ICC, we would conclude that an ICC of 0.782 indicates that the

exams can be rated with "good" reliability by different raters.

The following tutorials provide in-depth explanations of how to calculate ICC in different statistical software:

ARABPSYCHOLOGY.COM