

Intraclass Correlation Coefficients (ICC) with the irrICC Package

Kilem L. Gwet, Ph.D.

2021-08-25

```
library(irrICC)
```

Abstract

irrICC is an R package that provides several functions for calculating various Intraclass Correlation Coefficients (ICC). This package follows closely the general framework of inter-rater and intra-rater reliability presented by Gwet (2014).

All input datasets to be used with this package must contain a mandatory “Target” column of all subjects that were rated, and 2 or more columns “Rater1”, “Rater2”, . . . showing the ratings assigned to the subjects. The Target variable must represent the first column of the data frame, and every other column is assumed to contain ratings from a rater. Note that all ratings must be numeric values for the ICC to be calculated. For example, here is a dataset “iccdatal” that is included in this package:

```
iccdatal
#>   Target   J1   J2   J3   J4
#> 1      1  6.0  1.0  3.0  2
#> 2      1  6.5  NA  3.0  4
#> 3      1  4.0  3.0  5.5  4
#> 4      5 10.0  5.0  6.0  9
#> 5      5  9.5  4.0  NA  8
#> 6      4  6.0  2.0  4.0  NA
#> 7      4  NA  1.0  3.0  6
#> 8      4  8.0  2.5  NA  5
#> 9      2  9.0  2.0  5.0  8
#> 10     2  7.0  NA  2.0  6
#> 11     2  8.0  NA  2.0  7
#> 12     3 10.0  5.0  6.0  NA
```

The first column “Taget” (the name Target can be replaced with any other name you like) contains subject identifiers, while J1, J2, J3, J4 are the 4 raters (referred to here as Judges) and the ratings they assigned to the subjects. You will notice that the Target column contains duplicates, indicating that some subjects were rated multiple times. Moreover, none of these judges rated all subjects as seen by the presence of missing ratings identified with the symbol NA.

Two other datasets, iccdatal2, and iccdatal3 come with the package for you to experiment with. Even if your data frame contains several variables, note that only the Target and the Rater columns must be supplied as parameters to the functions. For example the iccdatal2 data frame contains a variable named Group, which indicates the specific group each Target is categorized. It must be excluded from the input dataset as follows: iccdatal2[,2:6].

Computing various ICC values

To determine what function you need, you must first have a statistical description of experimental data. There are essentially 3 statistical models recommended in the literature for describing quantitative inter-rater reliability data. These are commonly refer to as model 1, model 2 and model 3.

- **Model 1**

Model 1 is uses a single factor (hence the number 1) to explain the variation in the ratings. When the factor used is the subject then the model is referred to as Model 1A and when it is the rater the model is named Model 1B. You will want to use Model 1A if not all subjects are rated by the same roster of raters. That raters may change from subject to subject. Model 1B is more indicated if different raters may rate different rosters of subjects. Note that while Model 1A only allows for the calculation of inter-rater reliability, Model 1B on the other hand only allows for the calculation of intra-rater reliability.

Calculating the ICC under Model 1A is done as follows:

```
icc1a.fn(iccdatal)
#>      sig2s      sig2e      icc1a n r max.rep min.rep Mtot ov.mean
#> 1 1.761312 5.225529 0.2520899 5 4      3      1 40      5.2
```

It follows that the inter-rater reliability is given by 0.252, the first 2 output statistics being the subject variance component 1.761 and error variance component 5.226 respectively. You may see a description of the other statistics from the function's documentation.

The ICC under Model 1B is calculated as follows:

```
icc1b.fn(iccdatal)
#>      sig2r      sig2e      icc1b n r max.rep min.rep Mtot ov.mean
#> 1 4.32087 3.365846 0.5621217 5 4      3      1 40      5.2
```

It follows that the intra-rater reliability is given by 0.562, the first 2 output statistics being the rater variance component 4.321 and error variance component 3.366 respectively. A description of the other statistics can be found in the function's documentation.

- **Model 2**

Model 2 includes a subject and a rater factors, both of which are considered random. That is, Model 2 is a pure random factorial ANOVA model. You may have Model 2 with a subject-rater interaction and Model 2 without subject-rater interaction. Model 2 with subject-rater interaction is made up of 3 factors: the rater, subject and interaction factors, and is implemented in the function *icc2.inter.fn*.

For information, the mathematical formulation of the full Model 2 is $y_{ijk} = \mu + s_i + r_j + (sr)_{ij} + e_{ijk}$, where y_{ijk} is the rating associated with subject i , rater j and replicate (or measurement) k . Moreover, μ is the average rating, s_i subject i 's effect, r_j rater j 's effect, $(sr)_{ij}$ subject-rater interaction effect associated with subject i and rater j , and e_{ijk} is the error effect. The other statistical models are similar to this one. Some may be based on fewer factors or the assumptions applicable to these factors may vary from model to model. Please read Gwet (2014) for a technical discussion of these models.

Calculating the ICC from the iccdatal dataset (included in this package) and under the assumption of Model 2 with interaction is done as follows:

```
icc2.inter.fn(iccdatal)
#>      sig2s      sig2r      sig2e      sig2sr      icc2r      icc2a n r max.rep min.rep
#> 1 2.018593 4.281361 1.315476 0.4067361 0.251627 0.8360198 5 4      3      1
#>      Mtot ov.mean
#> 1 40      5.2
```

This function produces 2 intraclass correlation coefficients **icc2r** and **icc2a**. While **iccr** represents the inter-rater reliability estimated to be 0.252 , **icc2a** represents the intra-rater reliability

estimated at 0.836. The first 3 output statistics are respectively the the subject, rater, and interaction variance components.

The ICC calculation with the iccdata1 dataset and under the assumption of Model 2 without interaction is done as follows:

```
icc2.nointer.fn(iccdata1)
#>      sig2s      sig2r      sig2e      icc2r      icc2a n r max.rep min.rep Mtot ov.mean
#> 1 2.090769 4.34898 1.598313 0.2601086 0.801157 5 4      3      1 40      5.2
```

The 2 intraclass correlation coefficients have now become $icc2r=0.26$ and $icc2a=0.801$. That is the estimated inter-rater reliability slightly went up while the intra-rater reliability coefficient slightly went down.

- **Model 3**

To calculate the ICC using the iccdata1 dataset and under the assumption of Model 3 with interaction, you should proceed as follows:

```
icc3.inter.fn(iccdata1)
#>      sig2s      sig2e      sig2sr      icc2r      icc2a n r max.rep min.rep Mtot
#> 1 2.274732 1.315476 0.2007963 0.5823786 0.6530007 5 4      3      1 40
#>      ov.mean
#> 1      5.2
```

Here, the 2 intraclass correlation coefficients are given by $icc2r = 0.582$ and $icc2a = 0.653$. The estimated inter-rater reliability went up substantially while the intra-rater reliability coefficient went down substantially compared to Model 2 with interaction.

Assuming Model 3 without interaction, the same coefficients are calculated as follows:

```
icc3.nointer.fn(iccdata1)
#>      sig2s      sig2e      icc2r      icc2a n r max.rep min.rep Mtot ov.mean
#> 1 2.241792 1.470638 0.6038611 0.6038611 5 4      3      1 40      5.2
```

It follows that the 2 ICCs are given by $icc2r = 0.604$ and $icc2a = 0.604$. As usual, the omission of an interaction factor leads to a slight increase in inter-rater reliability and a slight decrease in intra-rater reliability. In this case, both become identical.

References:

1. Gwet, K.L. (2014) *Handbook of Inter-Rater Reliability*, 4th Edition. Advanced Analytics, LLC.