Different types of ICC in R

Brice Ozenne

July 1, 2017

Contents

In this document, we will consider measurements done at several sessions over several individuals. We assume no missing data, i.e., all individuals have measurements for all sessions. The question we would like to answer is what is the reliability of the measurements?

Recommended references:

- Kenneth O. McGraw and S. P. Wong **Forming Inferences About Some Intraclass Correlation Coefficients**, Psychological Methods, 1996, Vol. l,No. 1,30-46

- Lawrence Lin, A.S. Hedayat, Wenting Wu, **Statistical Tools for Measuring Agreement**, 2012, Springer.

1 Preparation

1.1 Load R packages

```r
library(data.table, quietly = TRUE, verbose = FALSE)
library(lava, quietly = TRUE, verbose = FALSE)
library(lme4, quietly = TRUE, verbose = FALSE)
library(nlme, quietly = TRUE, verbose = FALSE)
library(irr, quietly = TRUE, verbose = FALSE)
library(psych, quietly = TRUE, verbose = FALSE)
```

1.2 Data generation

```r
n.id <- 10
n.scan <- 2

set.seed(10)
m <- lvm(c(Ytest,Yretest) ~ 2*eta)
latent(m) <- ~eta
distribution(m, ~Ytest) <- lava::gaussian.lvm(mean = 1, sd = 1)
distribution(m, ~Yretest) <- lava::gaussian.lvm(mean = 1, sd = 1)

dtW.icc1 <- as.data.table(lava::sim(m, n.id))
dtW.icc1[, eta := NULL]
dtW.icc1[, Id := as.character(1:.N)]
dtL.icc1 <- melt(dtW.icc1, id.vars = "Id", value.name = "Y", variable.name = "session")

dfW.icc1 <- as.data.frame(dtW.icc1)
dfL.icc1 <- as.data.frame(dtL.icc1)

n.rater <- 5

mm <- lvm(c(Y1,Y2,Y3,Y4,Y5) ~ 2*eta)
latent(mm) <- ~eta
distribution(mm, ~Y1) <- lava::gaussian.lvm(mean = 1, sd = 1)
distribution(mm, ~Y2) <- lava::gaussian.lvm(mean = 1, sd = 1)
distribution(mm, ~Y3) <- lava::gaussian.lvm(mean = 1, sd = 1)
distribution(mm, ~Y4) <- lava::gaussian.lvm(mean = 1, sd = 1)
distribution(mm, ~Y5) <- lava::gaussian.lvm(mean = 1, sd = 1)

dtW.icc2 <- as.data.table(lava::sim(mm, n.id))
dtW.icc2[, eta := NULL]
dtW.icc2[, Id := as.character(1:.N)]
dtL.icc2 <- melt(dtW.icc2, id.vars = "Id", value.name = "Y", variable.name = "session")

dfW.icc2 <- as.data.frame(dtW.icc2)
dfL.icc2 <- as.data.frame(dtL.icc2)
```

2 ICC1: fixed number of sessions (e.g. test, re-test) assuming no session effect

Using ICC1, we assess the absolute agreement among measurements repeated over one random factor (here patient identity) assuming no other effect. The number of sessions is fixed meaning that it is precisely those sessions that are interesting.

Example: we want to plan a study to assess treatment efficacy with two sessions: baseline and post-treatment. EEG is used to measure treatment efficacy. Prior to this study, another study is performed to assessed the stability of the EEG measurement between two sessions. We don't expect systematic difference in EEG signal between the sessions.

ICC1 is equivalent to the following mixed model:

$$Y_{i,t} = \alpha + u_i + \varepsilon_{i,t} \tag{1}$$

$$\text{where} \quad \begin{aligned} u_i &\sim \mathcal{N}\left(0, \tau^2\right) \\ \varepsilon_{i,t} &\sim \mathcal{N}\left(0, \sigma^2\right) \end{aligned}$$

$$ICC_{1,agreement}^{single} = \frac{\tau^2}{\tau^2 + \sigma^2} \tag{2}$$

$$ICC_{1,agreement}^{average} = \frac{\tau^2}{\tau^2 + \frac{\sigma^2}{n}} \tag{3}$$

2.1 using mixed models

```
lme_ICC1 <- lmer(Y ~ 1 + (1|Id), data = dtL.icc1)

sigma_id <- attr(VarCorr(lme_ICC1)[[1]],"stddev")
sigma_error <- attr(VarCorr(lme_ICC1),"sc")

ICC1.lme <- c(
single = as.double(sigma_id^2/(sigma_error^2+sigma_id^2)),
average = as.double(sigma_id^2/(sigma_error^2/n.scan+sigma_id^2))
)

ICC1.lme
```

```
   single    average
0.5191614 0.6834842
```

2.2 using generalized least squares

```
gls_ICC1 <- gls(Y ~ 1, correlation = corCompSymm(form = ~ 1|Id), data = dtL.icc1)

# correlation matrix
unclass(cov2cor(getVarCov(gls_ICC1)))
```

```
         [,1]      [,2]
[1,] 1.0000000 0.5191607
[2,] 0.5191607 1.0000000
```

2.3   using anova

```r
aov_ICC1 <- aov(Y ~ Error(Id), data = dtL.icc1)

MSB <- summary(aov_ICC1)[["Error: Id"]][[1]]["Mean Sq"]
MSW <- summary(aov_ICC1)[["Error: Within"]][[1]]["Mean Sq"]

ICC1.aov <-  c(
single = as.double((MSB - MSW)/(MSB + (n.scan-1)*MSW)),
average = as.double((MSB - MSW)/MSB)
)

ICC1.aov
```

```
   single    average
0.5191614 0.6834842
```

2.4   using existing packages

```r
# single
cat("single \n")
irr::icc(dfW.icc1[,c("Ytest", "Yretest")],
     model="oneway", type="consistency")$value
psych::ICC(dfW.data[,c("Ytest", "Yretest")])[[1]]$ICC[1]

# average
cat("average \n")
irr::icc(dfW.icc1[,c("Ytest", "Yretest")],
     model="oneway", type="consistency", unit = "average")$value
psych::ICC(dfW.data[,c("Ytest", "Yretest")])[[1]]$ICC[4]
```

```
single
[1] 0.5191614
[1] 0.5191614
average
[1] 0.6834842
[1] 0.6834842
```

3   ICC3: fixed number of sessions (e.g. test, re-test) accounting for a session effect

Using ICC3, we assess the absolute agreement among measurements repeated over one random factor (here patient identity) accounting for a possible session effect. The number of sessions is fixed meaning that it is precisely those sessions that are interesting.

Example: we want to plan a study (called main study) to assess treatment efficacy with two sessions: baseline and post-treatment. EEG is used to measure treatment efficacy. Prior to this study, another study is performed to assessed the stability of the EEG measurement between two sessions. Unfortunately the two measurements were not performed using the same EEG machine so we suspect a systematic difference in EEG signal between the sessions. The investigator will make sure that this won't happen in the main study. Therefore the systematic difference is not inherant to the EEG technic (i.e. won't be observed in the main study).

Note: In test re-test analysis this should not be the case since we replicate a measurement under the same conditions.

Using ICC3, we assess the absolute agreement among measurements repeated over one random factor account for a possible session effect. Compared to ICC1, this will lead to:

- a lower σ^2 (some of the residual variance is explained)

- a higher τ (individual measurements are better correlated since there is less arbitrary variation between them)

- a higher ICC

ICC3 is equivalent to the following mixed model:

$$Y_{i,t} = \alpha + \beta_t t + u_i + \varepsilon_{i,t} \tag{4}$$

where $u_i \sim \mathcal{N}\left(0, \tau^2\right)$
t is treated as a categorical variable
$\varepsilon_{i,t} \sim \mathcal{N}\left(0, \sigma^2\right)$

$$ICC_{1,agreement}^{single} = \frac{\tau^2}{\tau^2 + \sigma^2} \tag{5}$$

$$ICC_{1,agreement}^{average} = \frac{\tau^2}{\tau^2 + \frac{\sigma^2}{n}} \tag{6}$$

3.1 using mixed models

```r
lme_ICC2 <- lmer(Y ~ session + (1|Id), data = dtL.icc1)

sigma_id <- attr(VarCorr(lme_ICC2)[[1]],"stddev")
sigma_error <- attr(VarCorr(lme_ICC2),"sc")

ICC2.lme <- c(
    single = as.double(sigma_id^2/(sigma_error^2+sigma_id^2)),
    average = as.double(sigma_id^2/(sigma_error^2/n.scan+sigma_id^2))
)

ICC2.lme
```

```
   single    average
0.7495312 0.8568366
```

3.2 using generalized least squares

```r
gls_ICC2 <- gls(Y ~ session, correlation = corCompSymm(form = ~ 1|Id), data = dtL.icc1)

# correlation matrix
unclass(cov2cor(getVarCov(gls_ICC2)))
```

```
          [,1]      [,2]
[1,] 1.0000000 0.7495308
[2,] 0.7495308 1.0000000
```

3.3 using anova

```r
aov_ICC2 <- aov(Y ~ session + Error(Id), data = dtL.icc1)

MSB <- summary(aov_ICC2)[["Error: Id"]][[1]]["Mean Sq"]
MSW <- summary(aov_ICC2)[["Error: Within"]][[1]]["Residuals","Mean Sq"]

ICC2.aov <-  c(
    single = as.double((MSB - MSW)/(MSB + (n.scan-1)*MSW)),
    average = as.double((MSB - MSW)/MSB)
)

ICC2.aov
```

```
   single    average
0.7495312 0.8568366
```

3.4 using existing packages

```r
# single
cat("single \n")
psych::ICC(dfW.data[,c("Ytest", "Yretest")])[[1]]$ICC[3]

# average
cat("average \n")
psych::ICC(dfW.data[,c("Ytest", "Yretest")])[[1]]$ICC[6]
```

```
single
[1] 0.7495312
average
[1] 0.8568366
```

4 Heterogenous variance between sessions

The previous ICC are not corrected for a possible difference in variance between sessions:

```r
gls_heteroV <- gls(Y ~ session,
            weights = varIdent(form = ~ 1|session),
            correlation = corCompSymm(form = ~ 1|Id),
            data = dtL.icc1)

# correlation matrix
unclass(cov2cor(getVarCov(gls_heteroV)))
```

```
        [,1]     [,2]
[1,] 1.000000 0.763223
[2,] 0.763223 1.000000
```

In test re-test analysis this should not be the case since we replicate a measurement under the same conditions.

5 ICC2: random number of sessions

Using ICC2, we assess the absolute agreement among measurements repeated over two random factors, e.g. patient identity and sessions. We don't account for a possible session effect. The number of sessions is random meaning that we are not precisely interested in those sessions. A more realistic example would be to consider raters instead of sessions.

Example: we want to study the reliability of the grades given by teachers. We include 10 teachers that will grade 20 students. We are not interested in those specific teachers since further studies may be performed in other schools with other teachers.

ICC2 is equivalent to the following mixed model:

$$Y_{i,j,t} = \alpha + u_i + v_j + \varepsilon_{i,j,t} \tag{7}$$

where
$$u_i \sim \mathcal{N}\left(0, \tau^2\right)$$
$$v_j \sim \mathcal{N}\left(0, \delta^2\right)$$
$$\varepsilon_{i,t} \sim \mathcal{N}\left(0, \sigma^2\right)$$

$$ICC_{2,agreement} = \frac{\tau^2}{\tau^2 + \sigma^2} \text{ or } ICC_{2,consistency} = \frac{\tau^2}{\tau^2 + \delta^2 + \sigma^2} \tag{8}$$

$$ICC_{2,agreement}^{average} = \frac{\tau^2}{\tau^2 + \frac{\sigma^2}{n}} \text{ or } ICC_{2,consistency}^{average} = \frac{\tau^2}{\tau^2 + \frac{1}{n}(\delta^2 + \sigma^2)} \tag{9}$$

5.1 using mixed models

```
lme_ICC2r <- lmer(Y ~ 1 + (1|session)+ (1|Id), data = dtL.icc2)

sigma_id <- attr(VarCorr(lme_ICC2r)[[1]],"stddev")
sigma_scan <- attr(VarCorr(lme_ICC2r)[[2]],"stddev")
sigma_error <- attr(VarCorr(lme_ICC2r),"sc")

ICC2r.lme <- c(
    consistency.single = as.double(sigma_id^2/(sigma_error^2+sigma_id^2)),
    agreement.single = as.double(sigma_id^2/(sigma_error^2+sigma_scan^2+sigma_id^2)),
    consistency.average = as.double(sigma_id^2/(sigma_error^2/n.rater+sigma_id^2)),
    agreement.average = as.double(sigma_id^2/(sigma_error^2/n.rater+sigma_scan^2/n.rater+
    sigma_id^2))
)

ICC2r.lme
```

consistency.single	agreement.single	consistency.average	agreement.average
0.8057232	0.8001849	0.9539944	0.9524334

5.2   using generalized least squares

Don't know how to specify non-nested random effects

5.3   using anova

```r
aov_ICC2r <- aov(Y ~ Error(session + Id), data = dtL.icc2)

MSB <- summary(aov_ICC2r)[["Error: Id"]][[1]]["Mean Sq"]
MSscan <- summary(aov_ICC2r)[["Error: session"]][[1]]["Residuals","Mean Sq"]
MSW <- summary(aov_ICC2r)[["Error: Within"]][[1]]["Residuals","Mean Sq"]

ICC2r.aov <-  c(
    consistency.single = as.double((MSB - MSW)/(MSB + (n.rater-1)*MSW)),
    agreement.single = as.double((MSB - MSW)/(MSB + (n.rater-1)*MSW + n.rater * (MSscan-
    MSW)/n.id)),
    consistency.average = as.double((MSB - MSW)/MSB),
    agreement.average = as.double((MSB - MSW)/(MSB + (MSscan-MSW)/n.id))
)

ICC2r.aov
```

```
consistency.single    agreement.single consistency.average    agreement.average
        0.8057232          0.8001849          0.9539944          0.9524334
```

5.4   using existing packages

```r
# single
cat("single \n")
irr::icc(dfW.icc2[,c("Y1", "Y2", "Y3", "Y4", "Y5")],
     model="twoway", type="consistency")$value
irr::icc(dfW.icc2[,c("Y1", "Y2", "Y3", "Y4", "Y5")],
     model="twoway", type="agreement")$value
psych::ICC(dfW.icc2[,c("Y1", "Y2", "Y3", "Y4", "Y5")])[[1]]$ICC[3:2]

# average
cat("average \n")
irr::icc(dfW.icc2[,c("Y1", "Y2", "Y3", "Y4", "Y5")],
     model="twoway", type="consistency", unit = "average")$value
irr::icc(dfW.icc2[,c("Y1", "Y2", "Y3", "Y4", "Y5")],
     model="twoway", type="agreement", unit = "average")$value
psych::ICC(dfW.icc2[,c("Y1", "Y2", "Y3", "Y4", "Y5")])[[1]]$ICC[6:5]
```

```
single
[1] 0.8057232
[1] 0.8001849
[1] 0.8057232 0.8001849
average
[1] 0.9539944
[1] 0.9524334
[1] 0.9539944 0.9524334
```

6 Different correlation between sessions

When considering more than 2 sessions, the previous ICC assume the same correlation between sessions (or raters). This can be relaxed using an unstructured covariance matrix:

```
gls_heteroC <- gls(Y ~ session,
            correlation = corSymm(form = ~ 1|Id),
            data = dtL.icc3)

# correlation matrix
M.icc <- unclass(cov2cor(getVarCov(gls_heteroC)))
print(M.icc)
```

```
          [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 1.0000000 0.8257162 0.6932362 0.8742870 0.7906416
[2,] 0.8257162 1.0000000 0.7640218 0.7507234 0.8421196
[3,] 0.6932362 0.7640218 1.0000000 0.5166739 0.6906413
[4,] 0.8742870 0.7507234 0.5166739 1.0000000 0.5647855
[5,] 0.7906416 0.8421196 0.6906413 0.5647855 1.0000000
```

Note that the mean ICC:

```
mean(setdiff(unique(M.icc),1))
apply(M.icc, 1, function(x){mean(setdiff(x,1))})
```

```
[1] 0.7312846
[1] 0.7959702 0.7956452 0.6661433 0.6766174 0.7220470
```

does not match the "normal" ICC:

```
gls_homoC <- gls(Y ~ session,
            correlation = corCompSymm(form = ~ 1|Id),
            data = dtL.icc3)

# correlation matrix
unclass(cov2cor(getVarCov(gls_homoC)))
```

```
          [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 1.0000000 0.8057232 0.8057232 0.8057232 0.8057232
[2,] 0.8057232 1.0000000 0.8057232 0.8057232 0.8057232
[3,] 0.8057232 0.8057232 1.0000000 0.8057232 0.8057232
[4,] 0.8057232 0.8057232 0.8057232 1.0000000 0.8057232
[5,] 0.8057232 0.8057232 0.8057232 0.8057232 1.0000000
```

```
irr::icc(dfW.icc2[,c("Y1", "Y2", "Y3", "Y4", "Y5")],
      model="oneway", type="consistency")$value
```

```
[1] 0.7999099
```

Note: small difference between the function `icc` and `gls`. Due to non optimal optimization in `gls`?