

## estat icc — Estimate intraclass correlations

[Description](#)
[Remarks and examples](#)
[Menu for estat](#)
[Stored results](#)
[Syntax](#)
[Methods and formulas](#)
[Options](#)
[Also see](#)

## Description

`estat icc` is for use after estimation with `mixed`, `meintreg`, `metobit`, `melogit`, `meqrllogit`, and `meprobit`. `estat icc` is also for use after estimation with `meglm` in cases when the fitted model is a linear, logit, or probit mixed-effects model.

`estat icc` displays the intraclass correlation for pairs of responses at each nested level of the model. Intraclass correlations are available for random-intercept models or for random-coefficient models conditional on random-effects covariates being equal to 0. They are not available for crossed-effects models or with residual error structures other than independent structures.

## Menu for estat

Statistics > Postestimation

## Syntax

```
estat icc [ , llevel(#) ]
```

## Options

`level(#)` specifies the confidence level, as a percentage, for confidence intervals. The default is `level(95)` or as set by `set level`; see [\[U\] 20.8 Specifying the width of confidence intervals](#).

## Remarks and examples

[stata.com](#)

See, for instance, example 2 in [\[ME\] mixed postestimation](#) and examples 1 and 4 in [\[ME\] melogit postestimation](#).

## Stored results

`estat icc` stores the following in `r()`:

### Scalars

<code>r(icc#)</code>	level-# intraclass correlation
<code>r(se#)</code>	standard errors of level-# intraclass correlation
<code>r(level)</code>	confidence level of confidence intervals

### Macros

<code>r(label#)</code>	label for level #
------------------------	-------------------

### Matrices

<code>r(ci#)</code>	vector of confidence intervals (lower and upper) for level-# intraclass correlation
---------------------	---

For a  $G$ -level nested model, # can be any integer between 2 and  $G$ .

## Methods and formulas

### Intraclass correlations

Consider a simple, two-level random-intercept model, stated in terms of a latent linear response, where only  $y_{ij} = I(y_{ij}^* > 0)$  is observed for the latent variable,

$$y_{ij}^* = \beta + u_j^{(2)} + \epsilon_{ij}^{(1)}$$

with  $i = 1, \dots, n_j$  and level-2 groups  $j = 1, \dots, M$ . Here  $\beta$  is an unknown fixed intercept,  $u_j^{(2)}$  is a level-2 random intercept, and  $\epsilon_{ij}^{(1)}$  is a level-1 error term. In a mixed-effects linear and probit regression, errors are assumed to be normally distributed with mean 0 and variance  $\gamma$ . In a mixed-effects logistic regression, errors are assumed to be logistic with mean 0 and variance  $\gamma$ . Random intercepts are assumed to be normally distributed with mean 0 and variance  $\sigma_2^2$  and to be independent of error terms.

The intraclass correlation for this model is

$$\rho = \text{Corr}(y_{ij}^*, y_{i'j}^*) = \frac{\sigma_2^2}{\gamma + \sigma_2^2}$$

where  $\gamma = \sigma_1^2$  for a mixed-effects linear regression,  $\gamma = 1$  for a mixed-effects probit regression, and  $\gamma = \pi^2/3$  for a mixed-effects logistic regression. The intraclass correlation corresponds to the correlation between the latent responses  $i$  and  $i'$  from the same group  $j$ .

Now consider a three-level nested random-intercept model,

$$y_{ijk}^* = \beta + u_{jk}^{(2)} + u_k^{(3)} + \epsilon_{ijk}^{(1)}$$

for measurements  $i = 1, \dots, n_{jk}$  and level-2 groups  $j = 1, \dots, M_{1k}$  nested within level-3 groups  $k = 1, \dots, M_2$ . Here  $u_{jk}^{(2)}$  is a level-2 random intercept,  $u_k^{(3)}$  is a level-3 random intercept, and  $\epsilon_{ijk}^{(1)}$  is a level-1 error term. The random intercepts are assumed to be normally distributed with mean 0 and variances  $\sigma_2^2$  and  $\sigma_3^2$ , respectively, and to be mutually independent. The error terms are also independent of the random intercepts.

We can consider two types of intraclass correlations for this model. We will refer to them as level-2 and level-3 intraclass correlations. The level-3 intraclass correlation is

$$\rho^{(3)} = \text{Corr}(y_{ijk}^*, y_{i'j'k}^*) = \frac{\sigma_3^2}{\gamma + \sigma_2^2 + \sigma_3^2}$$

This is the correlation between latent responses  $i$  and  $i'$  from the same level-3 group  $k$  and from different level-2 groups  $j$  and  $j'$ .

The level-2 intraclass correlation is

$$\rho^{(2)} = \text{Corr}(y_{ijk}^*, y_{i'jk}^*) = \frac{\sigma_2^2 + \sigma_3^2}{\gamma + \sigma_2^2 + \sigma_3^2}$$

This is the correlation between latent responses  $i$  and  $i'$  from the same level-3 group  $k$  and level-2 group  $j$ . (Note that level-1 intraclass correlation is undefined.)

More generally, for a  $G$ -level nested random-intercept model, the  $g$ -level intraclass correlation is defined as

$$\rho^{(g)} = \frac{\sum_{l=g}^G \sigma_l^2}{\gamma + \sum_{l=2}^G \sigma_l^2}$$

The above formulas also apply in the presence of fixed-effects covariates  $\mathbf{X}$  in a random-effects model. In this case, intraclass correlations are conditional on fixed-effects covariates and are referred to as residual intraclass correlations. `estat icc` also uses the same formulas to compute intraclass correlations for random-coefficient models, assuming 0 baseline values for the random-effects covariates, and labels them as conditional intraclass correlations.

Intraclass correlations will always fall in  $[0,1]$  because variance components are nonnegative. To accommodate the range of an intraclass correlation, we use the logit transformation to obtain confidence intervals. We use the delta method to estimate the standard errors of the intraclass correlations.

Let  $\hat{\rho}^{(g)}$  be a point estimate of the intraclass correlation and  $\widehat{\text{SE}}(\hat{\rho}^{(g)})$  be its standard error. The  $(1 - \alpha) \times 100\%$  confidence interval for  $\text{logit}(\rho^{(g)})$  is

$$\text{logit}(\hat{\rho}^{(g)}) \pm z_{\alpha/2} \frac{\widehat{\text{SE}}(\hat{\rho}^{(g)})}{\hat{\rho}^{(g)}(1 - \hat{\rho}^{(g)})}$$

where  $z_{\alpha/2}$  is the  $1 - \alpha/2$  quantile of the standard normal distribution and  $\text{logit}(x) = \ln\{x/(1-x)\}$ . Let  $k_u$  be the upper endpoint of this interval, and let  $k_l$  be the lower. The  $(1 - \alpha) \times 100\%$  confidence interval for  $\rho^{(g)}$  is then given by

$$\left( \frac{1}{1 + e^{-k_l}}, \frac{1}{1 + e^{-k_u}} \right)$$

## Also see

[ME] [meglm](#) — Multilevel mixed-effects generalized linear model

[ME] [meintreg](#) — Multilevel mixed-effects interval regression

[ME] [melogit](#) — Multilevel mixed-effects logistic regression

[ME] [meprobit](#) — Multilevel mixed-effects probit regression

[ME] [meqrlogit](#) — Multilevel mixed-effects logistic regression (QR decomposition)

[ME] [metobit](#) — Multilevel mixed-effects tobit regression

[ME] [mixed](#) — Multilevel mixed-effects linear regression

[U] [20 Estimation and postestimation commands](#)