

[The Analysis Factor](#)

- [Home](#)
- [Consulting](#)
 - [Quick Question Consultations](#)
 - [Hourly Statistical Consulting](#)
 - [Results Section Review](#)
 - [Statistical Project Services](#)
- [Workshops](#)
 - [Live Online Workshops](#)
 - [On Demand Workshops](#)
 - [In Person Workshops](#)
- [Membership](#)
- [Training](#)
- [Free Webinars](#)
 - [Sign up for the Next Webinar](#)
 - [Webinar Recordings](#)
- [About](#)
 - [Our Privacy Policy](#)
 - [Our Team](#)
 - [About Karen Grace-Martin](#)
- [Contact](#)

The Intraclass Correlation Coefficient in Mixed Models

by Karen Grace-Martin

The ICC, or Intraclass Correlation Coefficient, can be very useful in many statistical situations, but especially so in Linear Mixed Models.

Linear Mixed Models are used when there is some sort of clustering in the data.

Two common examples of clustered data include:

- individuals were sampled within sites (hospitals, companies, community centers, schools, etc.). The site is the cluster.
- repeated measures or longitudinal data where multiple observations are collected from the same individual. The individual is the cluster in which multiple observations are grouped.

$$Y_{ij} = \beta_0 + \beta_1 X + u_i + \varepsilon_{ij}$$

$$u_i \sim \text{iid } N(0, \sigma_u^2)$$

for subject i

$$\varepsilon_{ij} \sim \text{iid } N(0, \sigma^2)$$

for outcome j

Observations from the same cluster are *usually* more similar to each other than observations from different clusters. If they are, you can't use statistical methods on these data to that assume independence, because estimates of variance, and therefore p-values, will be incorrect.

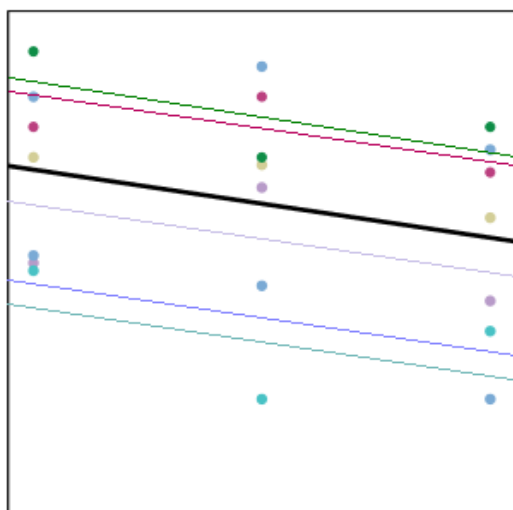
Mixed models not only account for the correlations among observations in the same cluster, they give you an estimate of that correlation.

At the right is the equation of a very simple linear mixed model. This has a single fixed independent variable, X , and a single random effect u . For simplicity, I'm going to assume that X is centered on it's mean. This is also known as a random intercept model.

The subscripts i and j on the Y indicate that each observation j is nested within cluster i .

The u represents the random intercept for each cluster. It's really a residual term that measures the distance from each subject's intercept around the overall intercept β_0 . Rather than calculate an estimate for every one of those distances, the model is able to just estimate a single variance σ_u .

That variance parameter estimate is the between-cluster variance. The variance of the residuals is the within-cluster variance. Their sum is the total variance in Y that is not explained by X .



If there is no real correlation among observations within a cluster, the

cluster means won't differ. It's only when some clusters have generally high values and others have relatively low values that the values within a cluster are correlated.

In the graph on the right, each cluster has its own trajectory of a different color. The thick black line represents the overall trajectory, averaged across all clusters.

Some clusters, like the magenta one, have all three values above the overall (black) mean. Those values will be correlated, because they're all relatively high. Simultaneously, those three points have a high mean.

Likewise, the turquoise cluster has all three values below the overall (black) mean. Again, those values will be correlated, because they're all relatively low. And the turquoise mean is quite low.

And so it goes. When some clusters have generally high values and others have generally low, (in other words, where there is consistency among a cluster's responses), there is variation among the clusters' means. This is the between-cluster variance.

The within-cluster variance represents how far each point is to the cluster specific mean. In other words, what the variation of the magenta points around the magenta trajectory?

In this graph, it's pretty small. Because those magenta points are all pretty high, they are quite close to their trajectory, and there is not a lot of within-cluster variation.

$$\text{Cov}(Y_{ij}, Y_{ij'}) = u_i = \sigma_0^2$$

$$\text{Corr}(Y_{ij}, Y_{ij'}) = \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2}$$

The ratio of the between-cluster variance to the total variance is called the **Intraclass Correlation**. It tells you the proportion of the total variance in Y that is accounted for by the clustering.

It can also be interpreted as the correlation among observations within the same cluster.

Why ICC is useful

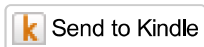
1. It can help you determine whether or not a linear mixed model is even necessary. If you find that the correlation is zero, that means the observations within clusters are no more similar than observations from different clusters. Go ahead and use a simpler analysis technique.
2. It can be theoretically meaningful to understand how much of the overall variation in the response is explained simply by clustering. For example, in a repeated measures psychological study you can tell to what extent mood

is a trait (varies among people, but not within a person on different occasions) or state (varies little on average among people, but varies a lot across occasions).

3. It can also be meaningful to see how the ICC (as well as the between and within cluster variances) changes as variable are added to the model.



Could you use some affordable ongoing statistical training, along with the opportunity to ask questions about statistical topics that have been stumping you? Consider joining our [Data Analysis Brown Bag program](#) with monthly online seminars and question & answer sessions.



Related Posts

- [Some Good References for Multilevel Regression Models](#)
- [Random Intercept and Random Slope Models Webinar](#)
- [Mixed Up Mixed Models](#)
- [Free Webinar: Fixed and Random Factors in Mixed Models: What IS the difference?](#)
- [R code for the Random Intercept and Random Slope Models Webinar](#)

Tagged as: [Intraclass Correlation Coefficient](#), [Mixed Models](#)

{ 2 comments... read them below or [add one](#) }

Eleanor Carson [September 4, 2013 at 7:31 am](#)

Karen, thank you for writing the article, The Intraclass Correlation Coefficient in Mixed Models. I'm enrolled in your upcoming "Analyzing Repeated Measures Data Workshop," and I fit the criterion of "not for you" because of my lack of knowledge of the basic principles in the workshop. I have about 3 weeks to learn as many of those basic principles as possible, before my own lectures start, because I must be able to analyze my PhD research data within a longitudinal repeated-measures design, whether I know the concepts or not. After reading your article, I was able to understand the introduction of West et al's book on Linear Mixed Models, and I feel that there is some hope! I couldn't understand, before. Your article was not only well-written and easy to understand, but explained the principles in a way that I can remember. Thank you!

[Reply](#)