級内相関係数についての覚書

村山航*

平成 20 年 9 月 14 日

級内相関係数 (intraclass correlation) は信頼性指標の 1 つとして頻繁に用いられる。基本的には，測定の対象 (object of measurements) の分散が全分散に占める割合と定義される[1] 。この級内相関には，いくつかの種類があり，それぞれで意味するところが違う。また，近年では階層線形モデル (hierarchical linear model; HLM) の文脈でも級内相関という言葉は頻繁に用いられる。この覚書では，特に一般化可能性理論 (generalizability theory; Brennan, 2001[2]) の観点から，級内相関係数のタイプの整理を行う。

データ構造

級内相関係数を算出する場合の，基本的なデータ構造を表 1 に示す。一般化可能性理論と同じく，測定の対象 (object of measurement) とは，こちらがその測定の対象としている（信頼性を知りたい対象となる）ものである。通常は，人 (participants) がここになる。測定の対象は，何らかの形で繰り返し測定 (measurement) される。測定の次元には，評定者 (judge) やテスト項目などが考えられる。データ構造をみると分かるように，級内相関係数を算出するときのデータ構造は，2 要因（もしくは 1 要因）であり，それ以上の要因は基本的に考えない。以下，測定値は x_{ij} の形で記す。i $(i = 1, 2, ..., n)$ は測定の対象を，j $(j = 1, 2, ..., k)$ は測定を示す添え字である。

表 1: 基本的なデータ構造

	測定 (measurement)				
測定の対象	1	2	3	$...j$	$...k$
1	x_{11}	x_{12}	x_{13}	$...x_{1j}$	$...x_{1k}$
2	x_{21}	x_{22}	x_{23}	$...x_{2j}$	$...x_{2k}$
.	.	.	.	.	.
.	.	.	.	.	.
i	x_{i1}	x_{i2}	x_{i3}	$...x_{ij}$	$...x_{ik}$
.	.	.	.	.	.
.	.	.	.	.	.
n	x_{n1}	x_{n2}	x_{n3}	$...x_{nj}$	$...x_{nk}$

*日本学術振興会・東京工業大学 mail: murakou@orion.ocn.ne.jp

[1]Shrout & Fleiss (1979) は，級内 “相関” 係数のもう 1 つの定義として “ICC（級内相関係数） is the correlation between one measurement (either a single rating or a mean of several ratings) on a target and another measurement obtained on that target.” と記している。この意味で，級内相関係数は相関であり，値が負になることもある。

[2]http://www4.ocn.ne.jp/ murakou/statistics.htm の ppt ファイル参照。

一要因モデル

一要因モデルは，測定が測定の対象それぞれにネストされた構造になっている場合に使用する。たとえば，ある筆記式テストの答えを，それぞれの回答者に対して，異なる 3 人の評定者が評定するような場合である。言い換えるならば，測定の対象ごとに j の順序が入れ替わっても問題がないようなケースである。

このとき，データは測定の対象だけを要因とした一要因分散分析モデルで分析される。得られる分散成分は σ_p^2（測定の対象間の分散成分）と σ_e^2（測定の対象内の分散成分＋誤差）であり，級内相関係数 (ICC) は

$$ICC(1) = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_e^2} \tag{1}$$

で推定される。これは，Shrout & Fleiss (1979) がケース 1 (case 1) と表現した級内相関係数である。ここでは，McGraw & Wong (1996) に基づき $ICC(1)$ と表記する。括弧内の 1 は，case 1 の 1 ではなく，この級内相関係数が，単一の測定値に対するものであることを示したものである。詳しくは後述する。

階層線形モデルの文脈でよく計算される級内相関係数も，$ICC(1)$ と基本的に同じである。$ICC(1)$ がネストされたデータ構造で計算されるものであり，階層線形モデルもネストされたデータに適用されることからも分かるであろう。階層線形モデルの場合，測定の対象が人そのものではない場合も多い。例えば，測定の対象が学校のクラスで，測定が生徒だった場合などが考えられる。また，階層線形モデルの文脈における級内相関係数は，測定の信頼性というよりも，各測定の対象（グループ）内での測定の非独立性を示す指標という意味合いが強い。

二要因変量モデル

測定の対象と測定がクロスしており，さらに測定が変量効果だと考えられる場合（たとえば，記述式テストの評価者が母集団からのランダムサンプルだと考えられる場合），二要因変量モデルの分散分析で分析され，級内相関係数が推定される。これは，Shrout & Fleiss (1979) がケース 2 (case 2) と表現した級内相関係数である。得られる分散成分は，σ_p^2（測定の対象間の分散成分）と σ_r^2（測定間の分散成分），$\sigma_{pr,e}^2$（両者の交互作用と誤差成分の混合）である。ここでは添え字として一般化可能性理論でよく使われる p（測定の対象，participant），r（評定者, rater）を用いているが，別に測定が評定者でなくてもこのモデルは適用可能である。

一致性と一貫性

一般化可能性理論では，ここで 2 つの信頼性係数が推定可能であった。dependability coefficient Φ と，generalizability coefficient $\boldsymbol{E}\rho^2$ である。それぞれ以下の式で算出される。

$$\Phi = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_r^2 + \sigma_{pr,e}^2} \tag{2}$$

$$\boldsymbol{E}\rho^2 = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{pr,e}^2} \tag{3}$$

Φ には，σ_r^2 が誤差成分に入っている。従って，たとえ測定の対象の順位をどの評定者も同じようにつけていたとしても，その絶対値が一致しなければ，この値は低くなる[3]。従って，級内相関係数の文脈では，この式で算出されたものを一致性 (agreement)（もしくは絶対一致）を示す級内相関係数と呼ぶ。以下では McGraw らの表現に従い，$ICC(A,1)$ と表記する。

一方，$\boldsymbol{E}\rho^2$ は，交互作用項のみが誤差成分として加えられている。従って，それぞれの評定者の絶対値が一致していなくても，測定の対象の順位を同じように評定していたならば（"A 君は B 君より点が高い" ということが，評定者の間で一貫していたならば），この値は基本的に高くなる。従って，級内相関係数の文脈では，この式で算出されたものを一貫性 (consistency) を示す級内相関係数と呼ぶ[4]。以下では McGraw らの表現に従い，$ICC(C,1)$ と表記する。

Shrout & Fleiss (1979) はケース 2 の級内相関係数として，一致性の級内相関係数のみを示しているが，実際は二要因変量モデルで推定できる級内相関係数には，一致性の指標と一貫性の指標の両方があるので混乱しないように注意が必要である。

相関係数による信頼性との違い

一致性の指標と一貫性の指標の違いをもう少しだけ具体的にみてみる。表 2 は Bartko (1976) が用いた例を少し改変したものである。測定の対象（回答者）が 5 人，評定者が 2 人いたとき，以下の 3 つの事例を考える。

1 つめの事例は，測定の対象の順位だけでなく，絶対値も評定者間で一致している。従って，$ICC(C,1)$ だけでなく，$ICC(A,1)$ も 1 になる。一方， 2 つめの事例は，回答者の順位は一致しているが，絶対値は一致していない。従って，$ICC(C,1)$ は 1 だが，$ICC(A,1)$ はもっと低い値になる。具体的には，0.238 となる[5]。

このように考えると，$ICC(C,1)$ は，評定者間の相関係数と同じだと思う人も出てくるかもしれない[6]。しかし，実際には大きな違いがある。3 つめの事例は，rater A の評定と rater B の評定の間の相関係数は 1 である。しかし，このとき $ICC(C,1)$ は，0.80 にしかならない。

一般に，rate A の評定値 y が rater B の評定値 x によって，回帰式 $y = ax + b$ で完全に予測できる場合，両者の相関係数は 1 になる。一方で，$ICC(C,1)$ が 1 になるのは，実は $y = x + b$ で完全に予測される場合のみである。回帰係数が 1 という制約が入っている。すなわち，事例 2 のように，片方の評定がもう片方の評定に一定の値を加えているような場合にのみ $ICC(C,1)$ は 1 になるのである。事例 3 のように，回帰係数が 1 ではないときには，たとえ評定者間で順位が完全に一致していても，$ICC(C,1)$ は 1 にはならない[7]。

二要因混合モデル

測定の対象と測定がクロスしており，さらに測定が固定効果だと考えられる場合（たとえば，記述式テストの評定者が最初から固定されている場合），二要因混合モデルの分散分析で分析され，級内相関係数が推定される。これは，Shrout & Fleiss (1979) がケース 3 (case 3) と表現した級内相関係数である。

[3] 詳しくは，冒頭の脚注で記した一般化可能性理論を説明したファイルを参照のこと。
[4] SPSS では agreement を "一致性"，consistency を "絶対一致" と呼んでいるようなので注意。
[5] R の psy パッケージの icc 関数を使うと，$ICC(C,1)$ と $ICC(A,1)$ はすぐに算出できる。
[6] 評定者間の相関係数は，評定者が 2 人のときのみに算出可能であるが，今回はそのような場合での比較をしている。
[7] このような場合というのは，得てして評定者（測定）間で評定値の分散が異なる場合によく起こる。級内相関係数というのは，測定間で分散が等しいという前提を持っていることから生じる特徴である

表 2: 3 つの信頼性係数の違いを示す事例

測定の対象	事例 1		事例 2		事例 3	
	rater A	rater B	rater A	rater B	rater A	rater B
1	1	1	1	5	1	3
2	2	2	2	6	2	5
3	3	3	3	7	3	7
4	4	4	4	8	4	9
5	5	5	5	9	5	11

二要因混合モデルを用いた場合にも，級内相関係数には，一致性と一貫性の指標の 2 種類が存在する。ここで重要なのは，二要因混合モデル (case 3) を用いた場合も，最終的に算出される一致性と一貫性の指標が $ICC(C,1)$ と $ICC(A,1)$ と一致することである（ここでは記さないが，検定などの結果も一致する）。もちろん，背後にあるモデルが違うので，case 2 と case 3 の場合の解釈の違い（測定が固定効果か変量効果か）には気をつける必要がある。

Shrout & Fleiss (1979) は case 2 として一致性の指標しか，case 3 として一貫性の指標しか紹介しなかった。そのため，case 2 = 一致性の指標，case 3 = 一貫性の指標，という誤解を時おり見かける。しかし，変量モデルか混合モデルかという次元は，一致性か一貫性かという次元とは独立であり，両者を分けて考える必要がある。本稿では，case 2 と case 3 では解釈上の違いがあるものの，値は一致することを踏まえ，ここでの一致性の指標と一貫性の指標をそれぞれ $ICC(C,1)$, $ICC(A,1)$ のように，変量効果モデル (case 2) の場合と同じ形で表記することとする (McGraw & Wong, 1996)。繰り返しになるが，解釈のときには，これらがどちらのモデル (case 2 or case 3) から来たものなのかを明確にしておかなければならない。

複数の測定の平均値の信頼性

これまでの級内相関係数は，単一の観測値に対する信頼性であった。一方，現実的には，たとえば 3 人の評定者の評定平均を使用する場合など，複数の測定 (k とする) の平均値の信頼性を知りたい場合も多い。このときの信頼性は，一要因モデルの場合，$ICC(k)$，一貫性の指標では $ICC(C,k)$，一致性の指標では $ICC(A,k)$ と表記する。

これらの推定方法は，一般化可能性理論の決定研究 (D study) における信頼性の推定方法と同じである。すなわち，測定に関する分散を k で割ったものを誤差分散とする。以下がそれぞれの級内相関係数の算出式である。

$$ICC(k) = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_e^2/k} \tag{4}$$

$$ICC(A,k) = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_r^2/k + \sigma_{pr,e}^2/k} \tag{5}$$

$$ICC(C,k) = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{pr,e}^2/k} \tag{6}$$

なお，よく使われる Cronbach の α 係数は，$ICC(C,k)$ と同じである。また，階層線形モデルにおける信頼性（Raudenbush & Bryk, 2002）も，概念的に $ICC(k)$ と等しい。

参考文献

Bartko, J. (1976). On various intraclass correlation reliability coefficients. *Psychological Bulletin*, *83*, 762-765.

McGraw, K. O. & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*, 30-46.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing reliability. *Psychological Bulletin*, *86*, 420-428.