

複数または層別された一致性研究における一致係数 AC₁ の均一性検定と共通 AC₁ の推定

久留米大学大学院医学研究科・小野薬品工業（株） 本田 主税
久留米大学バイオ統計センター 大山 哲司

背景・目的：測定には必ず誤差が伴う．誤差が小さく，一致性，再現性がよいという性質は臨床データに望ましい性質であり，信頼性と呼ばれる．2 名の評価者が対象を＋か－のいずれかに分類する場合の信頼性評価には，Cohen の κ 係数（以下 κ ）がしばしば用いられてきた． κ は，観察された一貫性に対して偶然の一貫性を調整することによって得られるが，観察対象者集団の有病（陽性）割合によって大きく変動し，評価が一貫しない割合（バイアス）の変化と逆の動きをする問題が指摘されている．近年，複数の代替手法が提案されているが，なかでも Gwet（2008）により提案された一致係数 AC₁ は，有病割合の偏りに対して頑健で，様々な分野で活用されている．

2 名の評価者が対象を＋か－のいずれかに分類する評価系が複数（ここでは K 回と表す）ある場合，これらの一貫性は同じとみなせるのかを検討したい場合がある．また，同じとみなせた場合，共通の一貫性係数を算出したい場合もある．本研究では，層のある場合の AC₁ の均一性検定を導出し，その性能を評価する．さらに，均一性が示された場合の共通 AC₁ の推定についても検討する．

モデル：第 k 層 ($k = 1, \dots, K$) の被験者 ($i = 1, \dots, n$) を評価者 r ($r = A, B$) が＋と判定するとき， $X_{kir} = 1$ ，それ以外の場合を $X_{kir} = 0$ とする．層毎に，評価者に対して周辺均一性を仮定し，確率 $P(X_{kir} = 1|i) = P_{ki}$ 期待値 $E(P_{ki}) = \pi_k$ ，分散 $V(P_{ir}) = \sigma_k^2$ とする．このとき第 k 層の AC₁（以下 γ_k ）は以下に表せる．

$$\gamma_k = \frac{1 + 2[\sigma_k^2 - 2\pi_k(1 - \pi_k)]}{1 - 2\pi_k(1 - \pi_k)}$$

均一性の検定を導くために，尤度理論を用いる． 2×2 表のうち，2 者の評価が一致しない項は 1 つの Category にまとめ，それぞれの頻度と確率を下表のようにまとめる．

Category	Ratings	Frequency	Probability
1	+, +	x_{1k}	$P_{1k}(\gamma_k)$
2	+, + or -, +	x_{2k}	$P_{2k}(\gamma_k)$
3	-, -	x_{3k}	$P_{3k}(\gamma_k)$
Total		n_k	1

第 k 層の対数尤度は $l_k(\gamma_k, \pi_k) = x_{1k} \log P_{1k}(\gamma_k) + x_{2k} \log P_{2k}(\gamma_k) + x_{3k} \log P_{3k}(\gamma_k)$ と表され，均一性の帰無仮説 $H_0: \gamma_k = \gamma$ に対するスコア統計量 T は

$$T(\tilde{\gamma}, \tilde{\pi}) = \sum_k^K \frac{\tilde{R}_k^2 \tilde{D}_k}{n_k(\tilde{B}_k \tilde{D}_k - \tilde{C}_k^2)}$$

と導ける． $\tilde{\gamma}, \tilde{\pi} = (\tilde{\pi}_1, \dots, \tilde{\pi}_K)'$ はそれぞれ， H_0 のもとでの最尤推定量， $\tilde{B}_k, \tilde{D}_k, \tilde{C}_k, \tilde{R}_k$ はそれぞれ計算をまとめた項である． T が自由度 $K-1$ のカイ二乗分布に従うことを利用して検定を行う．検定の結果， γ が各層で共通とみなせる場合，共通 γ 推定量は H_0 のもとでの最尤推定量，信頼区間は漸近近似法，プロファイル尤度法にて導出する．

評価：均一性のスコア検定及び共通 γ 推定量の性能評価をモンテカルロシミュレーションにより実施し，臨床試験データを用いて実例の解析を実施する．結果は当日示す．

提案手法は，複数回実施された一貫度評価の要約，複数のグループからの論文報告のメタ解析，層別解析による層の違いの検討に活用できると考えている．また，提案手法は医学研究のみならず，計量生物，心理測定，行動科学などの特性の一貫度評価にも応用可能である．

参考文献：

- Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol. 2008; 61: 29-48.
- Ohyama T. Statistical inference of agreement coefficient between two raters with binary outcomes. Commun Stat Theory Methods. 2019; Feb. DOI:10.1080/03610926.2019.1576894.