様式Ｃ－１９

科研費 KAKENHI

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年　5 月 17 日現在

機関番号：13101
研究種目：若手研究（B）
研究期間：　2010　～　2012
課題番号：22730506
研究課題名（和文）　テストデータへの階層的線形モデルの適用

研究課題名（英文）　Application of hierarchical linear modeling to test data

研究代表者
杉澤　武俊（SUGISAWA TAKETOSHI）
新潟大学・人文社会・教育科学系・准教授
研究者番号：30361603

研究成果の概要（和文）：階層的なテストデータの統計的分析において，より適切な手法を用いるための方法論的検討を行った。実際のテストデータを想定した人工データによるシミュレーションから，あるテストの信頼性に関する複数の研究結果を統合する際に，より正確な結果を得るための変換方法を明らかにした。また，縦断的データを用いた研究において 2 群の変化の様子の違いを調べることのできる 2 つの統計モデルの検出力が本質的には同等であるということを示した。

研究成果の概要（英文）：Methodologies for analyzing hierarchically structured test data were investigated. Two simulation studies showed that either log or cubic root transformation performs better than other transformations in reliability generalization studies which adopt hierarchical linear modeling, and that hierarchical linear modeling is as powerful as latent growth modeling in the case of comparison between growth rates of two groups.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2010 年度	800,000	240,000	1,040,000
2011 年度	600,000	180,000	780,000
2012 年度	500,000	150,000	650,000
年度			
年度			
総　計	1,900,000	570,000	2,470,000

研究分野：社会科学
科研費の分科・細目：心理学・教育心理学
キーワード：テストデータ・階層線形モデル・信頼性・検定力

１．研究開始当初の背景

テストデータは，データ収集のデザインによって様々な階層構造を持つのが普通である。たとえば，複数の学校を対象に一斉にテストを実施すれば，各受験者の得点データはそれぞれが属している学級やさらに上位の階層である学校にネストされることになる。あるいは，同一の受験者を対象に縦断的にデータを収集して成長の様子を測定すれば，複数の測定時点のデータが各受験者にネストされるという構造を持っている。このような階層性を無視した従来の分析法では，誤った結論を導いたり，データの持つ重要な情報を捨ててしまったりすることになるという指摘がしばしばなされている。このようなデータの階層性を適切に扱うことのできる統計モデルに階層的線形モデル（Hierarchical Linear Model；HLM）がある。欧米と比較

して我が国では，このような階層性を考慮した適切な分析が行われることはきわめて少ないのが現状である。

階層構造をモデルに組み込むことによって，通常の分析結果よりも妥当な結論が導けるという保守的かつ消極的な理由だけではなく，通常の分析では捨てられてしまう有益な情報をもっと積極的に活用していくこともできる。

HLM の利点が特に活かされるケースとして(1)メタ分析および(2)縦断データの分析が挙げられる。

(1)テストデータに関するメタ分析として近年提案された概念として，テストの信頼性係数をメタ分析する Reliability Generalization（Vacha-Haase, 1998）がある。テストの信頼性は，ツールとしてのテストそのものの性質ではなく，ツールと測定対象となった集団との相互作用等によって変動しうるものである。従って，同一のテストツールを用いた複数の研究で報告されている信頼性係数をメタ分析することによって，同一のテストツールを用いたときの信頼性係数の変動の大きさや，測定対象の属性などの研究デザイン等に関わる要因のうち何が信頼性係数の高低に影響するのかを知ることができる。これらの情報から，あるテストツールを用いたときの信頼性係数の値を予測したり，信頼性をより高めていくための示唆を得ることができる。

しかし，この信頼性係数のメタ分析というアイデアは比較的新しいものであるため，理論的・方法論的に検討するべき課題が残されている。その大きなものの１つとして，分布の歪みへの対処が挙げられる。信頼性係数の分布は負にゆがんだ分布となりやすいと考えられるが，メタ分析の理論的な枠組みに HLM を採用した場合，変数の正規性を前提とするため，この分布の歪みが結果に何らかの影響を及ぼす可能性がある。分布の歪みに対しては，一般的に変数変換による対処法がとられるが，信頼性係数の場合は，直感的に Fisher の Z 変換が用いられたり（Beretvas et al, 2002），立方根に基づいた変換（Rodriguez & Maeda, 2006），対数変換（Miyazaki, 2007）などが用いられたりしている。Miyazaki(2007)はこれらの方法を比較して，得られる結果が異なることを指摘しているが，どのような変換を施せばより妥当な結論が得られるのかということは明らかにされていない。

(2)縦断的なテストデータの分析は，個人の成長の様子を把握したり，能力等の変化を見ることで教育効果の検討を行ったり，追跡調査による入学試験等の妥当性の検討を行ったりする場合など，さまざまな応用場面が考えられる。

縦断データの分析モデルの１つとして，成長曲線モデルを用いる場合がしばしば見られる。成長曲線モデルの理論的枠組みとして，HLM を用いるものと，構造方程式モデル（SEM）の下位モデルである潜在曲線モデルを用いるものがあるが，各測定時点における従属変数を１つの観測変数（複数項目の合成得点を含む）とすれば，この両者は本質的に同等であることが知られている。しかし，縦断データの分析において，HLM を用いた場合と，潜在曲線モデルを用いた場合で，得られる結果にどのような相違があるのかについては必ずしも明らかになっていない。

２．研究の目的

本研究では主として以下の２点について明らかにすることを目的とする。

(1)信頼性のメタ分析における変換方法の検討

複数の信頼性研究の結果を統合する信頼性係数のメタ分析を HLM の枠組みで行う際に，各研究で得られた信頼性係数の推定値に対してどのような変換を適用するのが妥当であるのか，複数の変換方法の候補を比較することで検証する。

(2)縦断データの群間比較における検定力

複数時点で測定を行う縦断データの変化の様子を群間で比較する状況において，分析モデルとして HLM を採用した場合と潜在曲線モデルを採用した場合で得られる結論にどのような相違が生じるのかについて，検定力の観点から検証する。

３．研究の方法

(1) 信頼性のメタ分析

以下の手順により，人工データを発生させたシミュレーションを行う。

【データの発生】

信頼性研究 k の個人 j に対して，反復測定がなされる仮想的な状況での測定値を Yijk と表し，以下のような３水準の階層線形モデル(HLM)を考える。

・レベル１：測定値レベル

$Y_{ijk} = \pi_{0jk} + e_{ijk}, \quad e_{ijk} \sim N(0, \sigma_e^2)$

ここで，π_{0jk} は個人 j の真の得点，e_{ijk} は測定誤差で，σ_e は測定の標準誤差を表す。

・レベル２：個人レベル

$\pi_{0jk} = \beta_{00k} + r_{0jk}, \quad r_{0jk} \sim N(0, \tau_k)$

ここで，β_{00k} は研究 k における特性値の母平均，r_{0jk} は個人差を反映した残差を表す。

・レベル３：研究レベル
$\beta_{00k} = \gamma_{000} + u_{00k}, \quad u_{00k} \sim N(0, \tau_\beta)$
ここで, γ_{000} は特性値の全体的な平均, u_{00k} は研究間の差異を反映した残差を表す。

このとき，尺度の項目数を a とすると，研究 k における信頼性係数の真値 α_k は以下の式で与えられる(Miyazaki & Skaggs, 2008)。

$\alpha_k = \tau_k / (\tau_k + \sigma_e^2 / a)$

実際の研究では，この真値に標本誤差が加わる。そこで，研究 k における標本の大きさを n_k ，信頼性係数の推定値を R_k とすると，次の関係が成り立つことを利用し(Van Zyl, Neudecker, & Nel, 2000)，標本誤差を加える。

$(1 - R_k)(1 - \alpha_k) \sim F_{[(nk-1)(a-1),(nk-1)]}$

この方法に基づいて，テストの項目数，サンプルサイズ, 各研究内の分散(τ_k)の発生のさせ方などの条件を変化させ，信頼性係数の観測地を発生させる。

【信頼性係数の変換方法】
各信頼性研究で得られる信頼性係数の推定値を R と表すと，変換方法の候補として，以下の 6 つを考える。

1. 変換なし (T_1), $T_1 = R$

2. Fisher の Z 変換 (T_2), $T_2 = \frac{1}{2}\ln\frac{1+R}{1-R}$

3. $\sqrt{R}$ の Z 変換 (T_3), $T_3 = \frac{1}{2}\ln\frac{1+\sqrt{R}}{1-\sqrt{R}}$

4. Miyazaki (2007)の対数による変換 (T_4), $T_4 = -\frac{1}{2}\ln(1-R)$

5. Rodriguez & Maeda (2006)の立方根による変換 (T_5), $T_5 = (1-R)^{1/3}$

6. 立方根による変換の修正版 (T_6), $T_6 = 1-(1-R)^{1/3}$

【分析モデル】
変換後の信頼性係数の値について，以下のような 2 水準のモデルを仮定し，母数の推定を行う。

・レベル１：信頼性係数のレベル
$T_k = \delta_k + e_k, \quad e_k \sim N(0, V_k)$
研究 k において得られる信頼性係数は，その真の値に標本誤差が加わったものとする。標本誤差の分散 Vk はサンプルサイズ等から理論的に導出される既知の値とする。

・レベル２：研究のレベル
$\delta_k = \gamma_0 + u_k, \quad u_k \sim N(0, \tau)$

各研究における信頼性係数の真の値は，そのツールによる信頼性係数の平均値に，各研究独自の要因による誤差が加わったものとする。
【結果の評価】
上記分析モデルに基づいて推定された母数の推定値が，真の値とどの程度近いかを評価するために，母数の真値と推定値との差（バイアス）を，その統計量の標準誤差で割って標準化した指標を用いる。

(2) 縦断データの検定力
縦断データとして，3 時点の反復測定データを扱う。成長曲線として直線を当てはめ，2 群間で平均的な直線の傾きに違いがあるかどうかを検定する状況を想定した人工データを発生させ，そのデータに対して階層線形モデルと潜在曲線モデルの 2 通りの方法を適用し，群間の効果の検定を行う。データの発生は以下のような 2 水準の階層線形モデルを採用した。

・レベル１：測定値レベル
個人 i の時点 $t(t=0,1,2)$における観測値を x_{it} と表す。
$x_{it} = \beta_{0i} + \beta_{1i}\, t + e_{it}, \quad e_{it} \sim N(0, \sigma^2)$

・レベル２：個人レベル
$\beta_{0i} = \gamma_{00} + u_{0i}, \quad u_{0i} \sim N(0, \tau_0)$
$\beta_{1i} = \gamma_{10} + \gamma_{11} W_i + u_{1i}, \quad u_{1i} \sim N(0, \tau_1)$
ただし，W_i は個人 i が所属する群を表し，0 または 1 の値をとる。

上記のモデルに基づいて発生させた人工データに対して，階層線形モデルと潜在曲線モデルの両方を適用し，群間の効果が有意であるかどうかを判定する。階層線形モデルは，上記のデータ発生モデルと同じモデルを指定し，成長曲線の傾きの群間差を表す γ_{11} の統計的に有意と判定された回数の割合を調べ，経験的な検定力の推定値とする。潜在曲線モデルについては，図１のパス図で示されたモデルをそれぞれの群に適用する多母集団同時分析を行う。その際，すべての母数が 2 群間で等値であるという制約を置いたモデルと，傾きの平均値のみ等値制約を置かないモデルの 2 通りのモデルを適用し，それぞれのモデルの χ^2 適合度指標の差を用いた尤度

比検定により，群間の傾きの差について検定を行う。潜在曲線モデルについても，階層線形モデルと同様に，2 つのモデル間で適合度の差が有意（群間の傾きの平均が有意）であると判定された回数の割合を調べる。

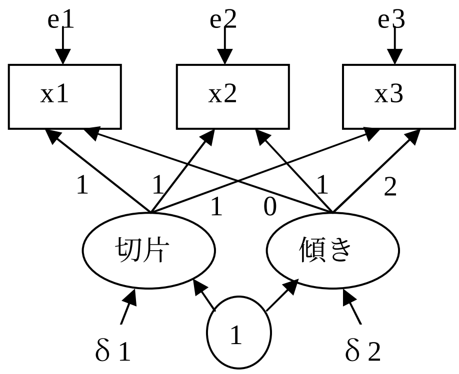

図１．潜在曲線モデル

４．研究成果

(1)信頼性のメタ分析

シミュレーションの結果，各変換法のもとでの固定効果(γ_0)の再現性については図２，変量効果(τ)の再現性については図３のようになった。(τ_k を一様分布から発生させた場合。)

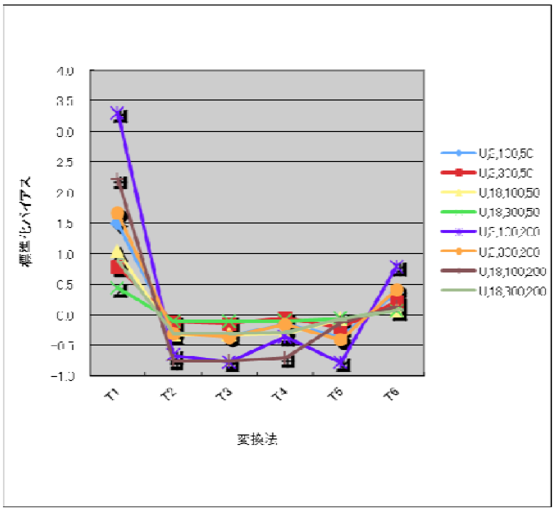

図２．固定効果母数の再現性

標準化されたバイアスが 0 に近いほど母数の再現性が高いと解釈されることから，対数による変換(T_4)と立方根による変換(T_5)がより望ましい方法であることが示唆された。この結果により，階層線形モデルの枠組みでテストの信頼性係数のメタ分析をより正確に行うための方法論的な指針が示唆されたといえる。また，それと同時に，本研究において考案された，測定モデルに基づいた信頼性係数の人工データ発生法は，従来，ベータ分布等に基づいて恣意的に行われていた方法に比べて，より理論的根拠の強いものであり，人工データを用いた信頼性係数のメタ分析の方法論的研究そのものの向上にも寄与するものと考えられる。

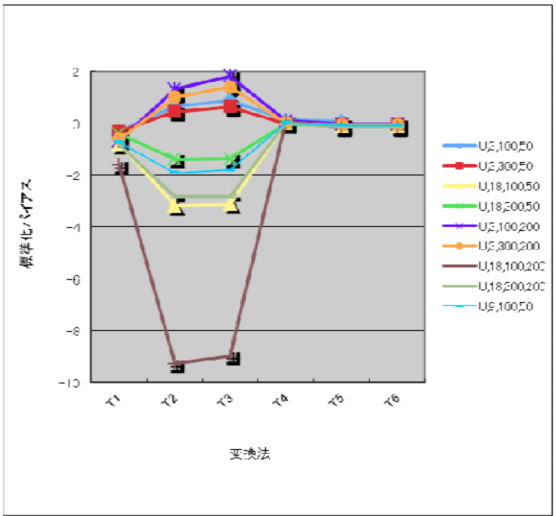

図３．変量効果母数の再現性

(2)縦断データの検定力

発生させる人工データのサンプルサイズや 2 群間の傾きの平均差（効果）の条件を変えてシミュレーションを行った結果，いずれの条件においても全体的に階層線形モデルの方がやや p 値が小さめになる傾向があるものの，有意水準(5%)近辺の値を取らない限りは 2 つのモデル間で有意性の判断に不一致が出ることはなかった。すなわち，2 群の比較における検定力の観点からは，階層線形モデルと潜在曲線モデルは同等の手法であるということができる。この結果は，縦断データにおける群間差を調べる研究において，分析者が抱きがちなモデル選択に対する疑問を解消するものであるということができる。ただし，本研究では説明変数を含まないシンプルな 3 時点データのみを想定していたため，より一般的なモデルでも同様の結果になるのかについて，包括的な検討を行っていく必要がある。

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 1 件）

(1) 杉澤武俊，測定・評価に関する動向と方法論研究のススメ，教育心理学年報，査読無，50 巻，2011，126-135

〔学会発表〕（計 2 件）

(1) 杉澤武俊，信頼性係数のメタ分析の枠組みにおける測定モデルに基づいた人工データ生成法, 日本教育心理学会第 54 回総会，2012 年 11 月 24 日，琉球大学

(2) Sugisawa，T.， Effects of two-stage sampling on the test for correlation coefficients.， The 75th Annual Meeting of Psychometric Society, July 7, 2010, The University of Georgia, USA.

６．研究組織

(1)研究代表者

杉澤　武俊（SUGISAWA TAKETOSHI）

新潟大学・人文社会・教育科学系・准教授

研究者番号：30361603