構造方程式モデリングにおける適合度指標とモデル改善について： 展望とシミュレーション研究による新たな知見

星 野 崇 宏*, 岡 田 謙 介**, 前 田 忠 彦***

Fit Indices and Model Modification in Structural Equation Modeling : A Review and new findings

Takahiro HOSHINO, Kensuke OKADA and Tadahiko MAEDA

In this paper we surveyed several fit indices, model modification methods, and other related previous studies in structural equation modeling. Inconsistency in results was detected from the review of the literature. To obtain more reliable knowledge, we performed simulation studies some ten times larger in scale than previous studies to examine the soundness of *Golden Rules* and the relationship between indices. Also, bias in the distribution of likelihood ratio statistics which depends both sample size and model complexity was indicated as a result of another large simulation study. This could be one of the causes of inconsistency in former studies of fit indices. We also examined fit indices for structural equation modeling with mean structure, multiple groups and categorical indicators. Stepwise estimation methods, model modification methods, and *Item Parceling Issues* were also reviewed.

Key words: covariance structure analysis, fit index, cut-off criteria, bias in the distribution of likelihood ratio statistics, item parceling issues

1. 問 題 意 識

構造方程式モデリング, 特にその観測変数が連続変数である場合の下位モデルである共分散構造分析 (Jöreskog, 1970; 狩野・三浦, 2003; 豊田, 1993) は, 学問分野で利用されるだけでなく, 近年マーケティングなど産業界においても意思決定を促す強力なツールとして認識されつつある。学問分野や産業界で構造方程式モデリングが非常に頻繁に利用されている理由のひとつは, モデルが複数考えられる状況で「提案した

モデルを確証する」こと, または「いくつかの競合的なモデルを比較し, その中からより良い一つのモデルを選択する」ことが可能である (と一般には思われている) ことにある。

また, 最尤推定法など一般的に利用されている推定法は「そのモデル全体が正しい」時に正しい推定値を与えてくれる方法である。もしもモデルが誤っていた場合には, 推定値は信頼できず, 正しい部分のパス係数の大きさなども解釈を誤る可能性がある。そこで, モデルの適合が良くないときにはなんらかのモデルの修正を行うことが必要である。

したがって, 構造方程式モデリングを実際に利用する際には, 候補となるモデルの適合を何らかの指標で調べたり, さらに競合するモデル間の優劣を比較するという「モデルの検証」の段階がもっとも重要なプロセスであると言って良い。

しかし適合度指標はこれまでに多数提案されており, 先行研究からどの指標がどのような性質を有しているかなどについての総合的なレビューは現在皆無で

*東京大学教養学部 大学院総合文化研究科
(Faculty of Arts and Sciences/Graduate School of Arts and Sciences, The University of Tokyo)

**東京大学大学院総合文化研究科
(Graduate School of Arts and Sciences, The University of Tokyo)

***統計数理研究所

(The Institute of Statistical Mathematics)

連絡先: 〒153-8902 東京都目黒区駒場3-8-1 星野
崇宏 宛

E-mail: bayesian@jasmine.ocn.ne.jp

ある。また、どの指標が、どのような状況において望ましいのかについての経験的規則 (=“Golden rule” としばしば呼ばれる) は多数存在するが、それらの知見は一貫していない。さらに、モデルの適合度が高くない場合にどのようなことをするとよいか、ということに関してもコンセンサスが得られているとは言い難い。

本稿では、これまでに提案されてきたモデル全体の適合度指標、モデル修正に関する方法とそれらに関する様々な先行研究のレビューを行い、先行研究の知見が一貫していない理由についての考察を行う。さらに、これまでの適合度指標についての議論は主に共分散構造分析に対するものであったが、平均構造を含むモデルや多群モデル、離散変数がある場合の適合度指標についても言及する。

また、近年注目を集めている段階的なモデル構築・検証方法、項目の尺度化の影響に関する議論 (いわゆる Item Parceling Issue) などについても触れる。具体的には、第2節では尤度比検定統計量によるモデル評価の問題について述べる。第3節では構造方程式モデリングにおいてこれまで提案されている適合度指標を様々な観点から分類し、紹介する。第4節では各指標の比較やいわゆる “Golden Rule” についての先行研究を展望する。第5節ではこれまでの先行研究よりも大規模なシミュレーション研究により、各適合度指標の関係と Golden rule による感度を調べる。適合度指標のほとんどは尤度比検定統計量の関数であったり、適合度指標の導出において漸近理論を用いているため、この問題は重要である。第6節では適合度指標に関する先行研究でこれまで一貫した知見が得られなかったひとつの理由として「尤度比検定統計量の分布が漸近理論による分布と大きく偏りやすい」ことを示す。さらに、平均構造や多群・離散観測変数がある場合の適合度指標については第7節で、モデルの修正について、そして段階的なモデル構築と推定についての議論については第8節で展望する。最後に、全体のまとめといくつかの議論を示す。

2. 構造方程式モデリングにおける尤度比検定とその問題点

複雑なモデルにおいて、データとの適合度を考える際に一般的に最もよく利用されるのが尤度比検定である。

尤度比検定は、あるモデル M_0 (このモデルに対応する、パラメータに対する仮説を H_0 とする) が、それよりも制約のない (より上位の) モデル M_1 (同じくこの仮説を H_1 とする) に比べて当てはまりが悪いかどうかを尤度比の観点から確かめる検定であり、(1) 推定法として最尤法を用いる (2) サンプルサイズが大きい、という条件が満たされる場合にさまざまな統計モデルにおいて利用することができる (ただし最尤法以外の推定法でも同様の漸近的な検定手法が利用できる)。

ここでモデル M_0 の下での尤度を $L(\theta_0 | Data, M_0)$ と置く。ただし θ_0 はモデル M_0 の下でのパラメータである。同様にモデル M_1 の下での尤度を $L(\theta_1 | Data, M_1)$ と置く。ただし θ_1 はモデル M_1 の下でのパラメータとする。

このとき、尤度比検定統計量

$$T = -2 \log \frac{L(\hat{\theta}_0 | Data, M_0)}{L(\hat{\theta}_1 | Data, M_1)} \quad (1)$$

は、モデル M_0 が正しい時には自由度 df の χ^2 分布に漸近的に従う (Serfling, 1980)。ただし自由度 df はモデル M_1 をモデル M_0 に制約したときの制約されたパラメータ数である。さらに $\hat{\theta}_0$, $\hat{\theta}_1$ はそれぞれの推定量である。また T は χ^2 分布に従う確率変数の性質として、その期待値は自由度に、分散は自由度の2倍になる。

ここで、この尤度比検定を構造方程式モデリングに適用した場合の欠点は大きく分けて以下の四つである。

1: 二つのモデルが入れ子構造になっている必要がある。

尤度比検定を含む統計的仮説検定では一般に、帰無仮説のモデル M_0 は対立仮説のモデル M_1 の部分モデルとなっている必要がある (このことを、入れ子になっている (nested) という)。したがって、一方がもう一方の部分モデルになっていない二つのモデル間の比較はできない (ただし、入れ子になっていないモデル間での検定は、例えば Vuong (1989) など複数提案されている)。

2: 示したいモデルを反証はできても確証はできない。

尤度比検定を行う場合に注意するべきは、 t 検定な

ど他の検定方法と同様に、この検定では「モデル M_0 が正しい」という帰無仮説 H_0 をデータから棄却して「モデル M_1 が正しい」と積極的に言うことはできるが、「モデル M_1 が正しい」という対立仮説を棄却することはできないことである。つまり「モデル M_0 が正しい」と積極的に言うことはできないのである。したがって、尤度比検定ではより制約のない（関心のない）モデルの方しか積極的に正しいと言うことができないことになる。一般の統計モデル、例えばロジスティック回帰分析などでは、「回帰係数がゼロである」という帰無仮説を棄却して「回帰係数がゼロでない」と言うことにより、回帰に意味があることを示すために尤度比検定が利用される。これは、「回帰係数がゼロである」というモデルよりも「回帰係数がゼロでない」というモデルの方がより制約のないモデルであるからである。この場合、研究者が示したいのは一般的には「回帰係数がゼロでない」という対立仮説であり、統計的仮説検定は自分の示したいモデルを確認することができる。しかし構造方程式モデリングにおいて尤度比検定を利用する場合は、自分の示したいモデルを反証することしかできない。

ここで、議論を特に変数がすべて連続である場合である共分散構造分析に限定する。この場合、考えられる中で最も制約のないモデル（一般に「飽和モデル」(saturated model) と呼ばれる）は「母共分散行列に構造が入っていない」というものになる。しかし、飽和モデルが成立するということは実質科学的には何の知見にもならない。あくまで、観測変数間の関係が（共通の潜在変数によって説明されるなどの）ある構造を示すという「制約のあるモデル」が成立することを示すことが実質科学的には重要である。

研究者は、このモデルを対立仮説側のモデル M_1 として、自分のモデルを M_0 に制約を加えた帰無仮説のモデル M_0 として検定を行うことになる。つまり、共分散構造分析の文脈では自分の仮説するモデルの適合を飽和モデルと比較する際には、自分のモデルを「棄却」することはできても確認することはできないということになる。

3: サンプルサイズに対する依存

尤度比検定は大標本では飽和モデル以外のどんなモデルも棄却し、小標本では検出力が低くなりどんなモデルも保持されやすいという点で構造方程式モデリングへの適用には非常に問題がある（Bentler & Bonett,

1980）。上記2の問題点を逆手に取ると、研究者が恣意的にあえてサンプルサイズを少なく取ることによって検定の検出力を下げ、「自分のモデルを少なくとも棄却はさせない」ということも可能になってしまう。

また、尤度比検定の p 値を適合度の基準として考えると、たとえ帰無仮説のモデルが同じ説明力を持っていたとしても、サンプルサイズが大きいと「当てはまりが悪い」と判断され、サンプルサイズが小さいと「当てはまりが良い」と判断されることになる。

4: 常に複雑なモデルに対して有利

構造方程式モデリングの文脈のみならず実質科学におけるモデリングの原則として、得られたモデルは同様の説明力を有する時には簡潔な方が有用とされる（Occam's razor: オッカムのカミソリ）。これに対して尤度比検定は常に対立仮説により複雑なモデルを置くため、必ず複雑なモデルが否定されない構造になっている。また上記3にもあるように、サンプルサイズが大きいと検出力が増し、必ずといっていいほど複雑なモデルが採用されてしまう。これは、帰無仮説と対立仮説との差が実質科学的にあまり意味がなくても常に対立仮説に有利になる、という統計的仮説検定そのものが有する問題でもある。

3. さまざまな適合度指標とその分類

構造方程式モデリングにおいて尤度比検定は上記のような問題点を持つということを明示的に取り上げた Bentler & Bonett (1980) の論文以来、検定統計量以外の形でデータへのモデルのフィットを測るための適合度指標が多数提案されている。

この節では共分散構造分析における様々な適合度指標を、まずは Anderson & Gerbing (1992) 及び Fan, Thompson & Wang (1999) の分類を一部改変し、適合度の数理的な意味及び導出の類似性の観点から四つのタイプに分けて紹介する。

3.1. 数理的な意味及び導出の類似性の観点からの分類

絶対的指標 (Absolute Fit Indices)

モデルがデータに当てはまっているかどうかを説明力や予測力とモデルの複雑さの観点から表す指標であり、結果として飽和モデルとの比較の指標になっている場合が多い。

この分類に属する指標はさらに三つに分類することができる。まず一つめには情報量基準であり、 AIC , $CAIC$, BIC などが含まれる。

これらの情報量基準は観測変数が離散である場合の、より一般的な構造方程式モデル全般にも利用できる指標である。ただし、一般に共分散構造分析の文脈では各指標をそのまま利用せず、例えば AIC では「関心のあるモデルでの AIC 」-「飽和モデルでの AIC 」のように、飽和モデルとの差を利用する場合が多い。すなわち、

$$\begin{aligned} AIC &= T - 2df, \\ CAIC &= T - (\log(N) + 1)df, \\ BIC &= T - \log(N)df \end{aligned} \quad (2)$$

となる。ここで例えばすべての観測変数が連続変数である場合は

$$T = (N-1)\{tr(\Sigma^{-1}(\hat{\theta})S) - \log|\Sigma^{-1}(\hat{\theta})S| - p\}, \quad (3)$$

つまり T は対立仮説を「母共分散行列に構造が入っていない」としたとき (= 飽和モデル) の尤度比検定統計量である。ただし θ は関心のあるモデルでのパラメータ、 $\hat{\theta}$ はその推定量、 S は観測変数の不偏共分散行列、 p は観測変数の数である (ここで S として標本共分散行列を利用するならば、式 (3) は $N-1$ でなく N 倍となる)。

また式 (2) から明らかなように、複雑なモデルに対するペナルティの強さは $AIC < BIC < CAIC$ である。

以降 T を、それぞれ利用する推定法によって、最尤法では T_{ML} 、一般化最小二乗法では T_{GLS} 、最小二乗法では T_{OLS} と置く。つまりそれぞれ、

$$T_{ML} = (N-1)\{tr(\Sigma^{-1}(\theta)S) - \log|\Sigma^{-1}(\theta)S| - p\}, \quad (4)$$

$$T_{GLS} = (N-1)\frac{1}{2}tr\{([S - \Sigma(\theta)]S^{-1})^2\}, \quad (5)$$

$$T_{OLS} = (N-1)\frac{1}{2}tr\{[S - \Sigma(\theta)]^2\}, \quad (6)$$

となる。

二つめとしてはクロスバリデーションによる真のモデルとの差についての指標である。特に ECVI (クロスバリデーションによる真のモデルとの差の指標の期待値) の推定値として CAK と CK (Browne & Cudeck, 1989) という二つの指標がある。

$$CAK = \frac{T}{N-1} + \frac{2q}{N-1}, \quad CK = \frac{T}{N-1} + \frac{2q}{N-p-2} \quad (7)$$

ただし q は推定すべきパラメータ数である。ここで、

$$AIC = CAK \times (N-1) - p(p+1) \quad (8)$$

の関係がある。推定法として最尤法を用いる場合は、 CAK でなく CK を利用することが推奨されている (Browne & Cudeck, 1993)。

これらの指標は、モデルが正しい場合の共分散行列の真値 $\Sigma(\theta^0)$ とその推定値 $\Sigma(\hat{\theta})$ の距離を利用した指標であるが、これに対して非心度を用いた指標 (後述) では、 $\Sigma(\theta^0)$ と $\Sigma(\theta^*)$ の距離を計算しているという点で大きな違いがある (ただし θ^* は $\Sigma(\theta)$ と飽和モデルでの共分散行列の真値 Σ^0 の距離を最小化する θ である: Bandalos, 1997; Cudeck & Henly, 1991)

最後に、データの共分散行列とモデルによって説明される共分散行列との乖離を示す指標として、 GFI , $AGFI$, RMR , $SRMR$ がある。

$$GFI = 1 - \frac{tr[W(S - \hat{\Sigma})^2]}{tr[(WS)^2]} \quad (9)$$

ただし $\hat{\Sigma}$ は関心のあるモデルでの共分散行列の推定値であり、最尤法の場合は $W = \hat{\Sigma}^{-1}$ 、最小二乗法の場合は $W = I$ (Jöreskog & Sörbom, 1986), GLS では推定時に利用する重み行列で、多くの場合 S^{-1} (Tanaka & Huba, 1985) がそれぞれ利用される。

またそれぞれの推定法での GFI に対して $AGFI$ は

$$AGFI = 1 - \frac{p(p+1)}{2df}(1 - GFI) \quad (10)$$

で定義される。ここで $GFI \geq AGFI$ の関係があることが容易にわかる。

RMR (Root Mean-Square Residual: Jöreskog & Sörbom, 1986) は、

$$RMR = \sqrt{\frac{2}{p(p+1)} \sum_{i=1}^p \sum_{j=1}^p (s_{ij} - \hat{\sigma}_{ij})^2} \quad (11)$$

と定義される。ただし s_{ij} は標本共分散行列の第 ij 要素であり、 $\hat{\sigma}_{ij}$ は母共分散行列の第 ij 要素 σ_{ij} のモデルによる推定値である。 RMR は観測変数の分散の大きさに依存するという性質があるため、現在では相関係数の形で置き換える Standardized Root Mean-Square Residual ($SRMR$)

$$SRMR = \sqrt{\frac{2}{p(p+1)} \sum_{i=1}^p \sum_{j=1}^p ((s_{ij} - \hat{\sigma}_{ij}) / (s_{ii}s_{jj}))^2} \quad (12)$$

が一般的に利用されている。

相対的指標 (Incremental/Relative Fit Indices)

相対的指標は、評価したいモデルよりもさらに制約の強いベースラインモデルを考え、そのモデルに比べてどれくらいフィットがよくなったかを表す指標であり、Bentler & Bonett (1980) によって提案された。

一般にベースラインモデルとしては、各観測変数に関連が無い（独立である）というモデルを立て、それとの乖離を見る指標がほとんどである。しかし、独立モデルはあまりに非現実的な仮定であり、全ての相関が等しいモデルをベースラインモデルにするべき (Marsh, 1998; Rigdon, 1998) といった意見や、段階的に適用するべき (Widaman & Thompson, 2003) といった意見もある。

相対的指標の代表的なものは *NFI*, *IFI*, *TLI* (*NNFI*), *RFI* などである。

NFI (Normed Fit Index: Bentler & Bonett, 1980), *IFI* (Incremental Fit Index: Bollen, 1989) はそれぞれ

$$NFI = \frac{T_b - T_h}{T_b}, \quad IFI = \frac{T_b - T_h}{T_b - df_h} \quad (13)$$

である。

TLI (Tucker-Lewis Index: Tucker & Lewis, 1973), または *NNFI* (NonNormed Fit Index: Bentler & Bonett, 1980) は

$$TLI = \frac{(T_b/df_b) - (T_h/df_h)}{(T_b/df_b) - 1} \quad (14)$$

RFI (Relative Fit Index: Bollen, 1986) は

$$RFI = \frac{(T_b/df_b) - (T_h/df_h)}{(T_b/df_b)} \quad (15)$$

である。

ただしここで T_b はベースラインモデルでの、 T_h は関心のあるモデルでの T の値であり、 df_h は関心のあるモデルと飽和モデルの母数の数の差（つまり関心のあるモデルでの自由度）、 df_b はベースラインモデルと飽和モデルの母数の数の差（つまりベースラインモデルでの自由度）である。また、*NFI* と *RFI* は

$$RFI = 1 - \frac{df_b}{df_h}(1 - NFI) \quad (16)$$

という線形の関係が成立する。

相対的指標は、シミュレーション研究から推定法の影響を受けやすいことが指摘されている (Sugawara & MacCallum, 1993)。特に最尤法ではなく一般化最小二乗法で推定をした場合、間違ったモデルでも大きくなりやすい (Fan, Thompson & Wang, 1999) とされる。

尤度比検定統計量の非心度を用いた指標 (Non-centrality-based Indices)

自分の考えているモデル M_0 と、飽和モデル M_1 の比較を尤度比検定によって行う場合を考える。

尤度比検定統計量 T_{ML} は帰無仮説 (H_0) が正しいときには、自由度 df の χ^2 分布に漸近的に従い、その期待値は df になる。これに対して、対立仮説 (H_1) が正しい場合は、自由度 df で非心度 λ の非心 χ^2 分布に漸近的に従うことが知られている。また、非心 χ^2 分布の性質から T_{ML} の期待値は $E(T_{ML}) = \lambda + df$ となる。そこで、非心度は

$$\lambda = T_{ML} - df \quad (17)$$

と近似することができる。

ここでもし自分のモデル M_0 が正しければ、 λ は 0 となるはずである。そこで、Bentler (1990) は非心度 $T_{ML} - df$ がゼロに近いモデルを良いモデルであるとみなす指標をいくつか提案した。その中で現在もっともよく利用されているのが *CFI* (Comparative Fit Index: Bentler, 1990) である。*CFI* は

$$CFI = 1 - \frac{\max[(T_h - df_h), 0]}{\max[(T_h - df_h), (T_b - df_b), 0]} \quad (18)$$

と定義されるが、これが相対的指標というより非心度を用いた指標である理由は、その導出過程にある。*CFI* は非心度 $T_h - df_h$ を基準化するためにベースラインモデルの検定統計量及び自由度を用いているのである。また式から、*CFI* は T_h が 0 になる（つまり関心のあるモデルが飽和モデルと当てはまりが変わらない）時に 1 に、ベースラインモデルと当てはまりが変わらないときに 0 になるという性質がある。

同様に非心度を用いた指標としては *RMSEA*, *RNI*, *ghat*, *Mc* などがある。

RMSEA (Root Mean Square Error of Approximation: Steiger & Lind, 1980)

$$RMSEA = \sqrt{\frac{\max[(T_h - df)/(N-1), 0]}{df}} \quad (19)$$

RNI (Relative Noncentrality Index: McDonald

& Marsh, 1990)

$$RNI = \frac{(T_b - df_b) - (T_h - df_h)}{(T_b - df_b)} = 1 - \frac{(T_h - df_h)}{(T_b - df_b)} \quad (20)$$

gamma hat (*ghat*) (Steiger, 1989)

$$ghat = \frac{p}{p + 2[(T_h - df)/(N - 1)]} \quad (21)$$

McDonald's centrality index (*Mc*: McDonald & Marsh, 1990)

$$Mc = \exp\left\{-\frac{1}{2}[(T_h - df)/N]\right\} \quad (22)$$

儉約的指標 (Parsimonious Fit Indices)

ベースラインモデルと評価したいモデルの自由度の比を掛けることで、さらに複雑なモデルにペナルティーを課す指標を儉約的指標といい、*PGFI* (Mulaik *et al.*, 1989), *PNFI* (Parsimonious Normed Fit Index: James, Mulaik & Brett 1982), *PNFI2*, *PCFI* がある。これらは上記の分類で紹介された適合度指標に $\frac{df_h}{df_b}$ を掛けたものとなる。例えば

$$\begin{aligned} PGFI &= \frac{df_h}{df_b} GFI, \\ PNFI &= \frac{df_h}{df_b} NFI, \\ PNFI2 &= \frac{df_h}{df_b} IFI \end{aligned} \quad (29)$$

であり、もとの指標より必ず値は低下する。低下の度合いは、評価したいモデルが複雑であるほど大きくなる。つまり単純なモデルほど高い適合度を示すという点で複雑なモデルにペナルティーを課しており、パラメータを「儉約」する指標である。

3.2. Tanaka (1993) の分類

上記の 4 分類は各適合度指標の数理的な意味及び導出の類似性からの分類である。一方以下の Tanaka (1993) の分類も有名である。これは、各指標を六つの性質から分類するものである。

1: 記述的か推測的か

記述的指標はモデルとデータの当てはまりを、得たデータから記述するものであり、推測的指標は母数としての適合度をデータから推測するという考え方に基づくものである。

2: 単純なモデルを採用するか

複雑なモデルにより強くペナルティーを科すかどうかについてであり、*GFI* に対する *AGFI* や、情報量基準、儉約的指標が単純なモデルを採用する指標に分類される。

3: 基準化されているか

(0, 1) の範囲に (近似的にでも) 収まるように基準化されているかどうかということであり、情報量基準などを除き、多くの指標は基準化されている。ただし、*IFI* や *TLI* は 1 を大きく超えるなどといったことがありえることが知られている。

4: 絶対的か相対的か

なにかしらのベースラインモデルとの比較を行う指標が相対的指標(前節でも述べた)、とくにそうした比較を行わない指標が絶対的指標である。非心度を用いた指標のいくつかは、相対的指標としての側面も有する。

5: 推定法に依存するか

推定法によって指標の値が変わるかどうかという問題以前に、尤度比検定や情報量基準は最尤法でのみ使える指標である。また、*GFI* などは推定法によって指標の式が変わるという点で、推定法に依存する。

6: サンプルサイズに依存するか

サンプルサイズに指標が依存するかどうかということには二つの意味がある。まずは「計算式にサンプルサイズが入っている」という意味である。尤度比検定統計量や情報量基準、*ghat* や *RMSEA* が計算そのものにサンプルサイズが入っている指標であり、*GFI* などはこの観点からはサンプルサイズに依存しない。

一方、「指標の分布がサンプルサイズに依存する」という意味もある。Marsh, Balla & McDonald (1988) のシミュレーション研究などからは、サンプルサイズに依存しないのは *TLI*, *IFI* などごく限られた指標のみであるとされている。

前者の意味ではサンプルサイズに依存しないが、後者の意味では依存するという指標として *GFI* がある。*GFI* は定義上ではサンプルサイズに依存しないように見えるが、その分布は既知であり(後述)、サンプルサイズに依存することは知られている。例えば同じ程度のフィットでもサンプルサイズが大きいと、値が

大きくなる傾向があることはシミュレーション研究によっても知られている (Bollen, 1989; Gerbing & Anderson, 1993)。

4. 各指標の比較や“Golden Rule”についての先行研究

4.1. 先行研究での各指標の比較

ここでは (1) サンプルサイズの影響, (2) モデルの複雑さ (観測変数の数の影響), (3) 指標の分布についての情報, の三つの観点での各指標の比較を行った研究を概説する。

サンプルサイズの影響

Anderson & Gerbing (1984) は, サンプルサイズ, 1 潜在変数あたりの観測変数数, 因子負荷, 因子数, 因子間相関を操作したモンテカルロ実験を行い, 分散分析によって各種適合度指標のばらつきを説明する要因を調べた。その結果, 適合度指標のばらつきを群を抜いてよく説明するのはサンプルサイズ N であることがわかった。つまり, サンプルサイズが小さいと適合度指標の値がばらついてしまうということである。また彼らは, GFI はサンプルサイズとは独立だという Jöreskog & Sörbom (1981) の言明に対して, 明示的に式の中に N が入っていないだけで結局サンプルサイズの影響は受けていると反論した。

また, 彼らは N が小さいとき, 標本尤度比検定統計量 T が χ^2_{df} 分布とは異なった分布をすることをも指摘した。同じ指摘は Bentler (1990) にも見られる。

Hu, Bentler & Kano (1992) は, 尤度比検定統計量に与えるサンプルサイズ, 非正規性, および母数推定法の影響をモンテカルロ実験により調べた。真のモデル下でサンプルサイズが検定統計量に与える影響については, $N \leq 250$ の条件で T の標本値が大きく推定されてしまうバイアスが確認された。この論文は構造方程式モデリング一般に利用できる T の補正の必要性にも言及しているが, 実際に補正式を導いてはいない (この問題については 5 節を参照)。

最近の研究としては, Curran, Bollen, Paxton, Kirby & Chen (2002) が, 真のモデルと様々な誤設定されたモデルの下での T の挙動を調べるモンテカルロ実験を行っている。真のモデルの下での T の挙動を見ると, 小標本条件下で大きく χ^2 分布から逸脱するという一貫した結果が得られた。

適合度指標のサンプルサイズ依存性は, 他にも多く

のシミュレーション研究によって繰り返し指摘されている (Marsh *et al.*, 1988; La Du & Tanaka, 1989; Fan *et al.*, 1999)。

感度と特異度

適合度指標の性能を評価する基準として, 正しいモデルにおいて適合度を高く評価できる (=感度が高い) かどうか, また誤ったモデルにおいて適合度を低く評価できる (=特異度が高い) かどうかということは自然な指標となる。そこでこれまで様々なシミュレーション研究が行われてきた。例えば Fan *et al.* (1999) のシミュレーション研究では, 誤ったモデルにおいて適合度が低く算出される (=特異度が高い) という点に関しては GFI , $AGFI$, Mc , $RMSEA$ が優れているという結論が得られている。ただし彼らの用いたモデル自体はシンプルなものであり, またここでのモデルの誤設定の度合いもあまり大きくない。

Hu & Bentler (1998) のシミュレーションでは, 確認的因子分析モデルを利用している。適合度指標同士の相関からは GFI , $AGFI$, NFI , RFI , CAK , CK と TLI , IFI , RNI , CFI , Mc , $RMSEA$ のクラスターに分かれることがわかった。また $SRMR$ はどれとも相関が低かった。彼らは誤ったモデルを識別できない (=特異度が低い) という観点から, 第一のクラスターに属する指標は利用しないほうがよいとしている。一方, 後者のクラスターと $SRMR$ は識別力がある (Hu & Bentler (1998) では $ML/GLS/ADF$ の三つの推定法についてそれぞれシミュレーション研究の結果が得られているので詳しくは参照)。特に, 誤った因子構造を見つけるのには $SRMR$ が一番よく, 因子負荷については TLI , IFI , RNI , CFI , $RMSEA$ が同様によいことがわかった。推定法として ML か GLS を用いる場合は加えて TLI , IFI , RNI , CFI , $ghat$, Mc , $RMSEA$ を, ADF 法を用いる場合は加えて TLI , IFI , RNI , CFI を用いるとよいとされている。ただし, TLI , Mc , $RMSEA$ はサンプルサイズが少ない場合 (250 以下) ではそれぞれ範囲が広がりすぎたり, バイアスが生じたり, 正しいモデルを棄却する確率が高くなるという観点から利用しないほうがよいとしている。

観測変数の数の影響

構造方程式モデリングにおけるモデルの複雑さは, 自由母数の多さや潜在変数, 観測変数の数などさまざまな観点から定義することができるが, 多くの先行研究では「1 潜在変数あたりの観測変数の数」として定義され, 比較されてきた。

例えば Ding, Velicer & Harlow (1995) は確認的因子分析モデルに従うデータを発生させ、そのモデルを用いて推定を行った(つまり感度の研究である)。結果として、(1) 一般的に因子に対する観測変数の数が増えると適合度指標 (ここで調べたのは *NFI*, *TLI*, *RNI*, *CFI*) は下がる(例: *TLI* は2変数/因子では1.005, 6変数/因子では0.984)。 (2) その傾向はサンプルサイズが小さいほど顕著であり、サンプルサイズが多いと *TLI* 以外は1に近づく(これらの現象の原因については5節を参照)。

また, Kenny & McCoach (2003) のシミュレーション研究では、正しいモデルで推定を行った場合は、先行研究と同様の結果が出たが、*RMSEA* は他の指標と異なり変数が多いと値が低くなる(フィットが良いという示唆を与える)ことがわかった。

さらに、正しいモデルとは異なった(誤った)モデルで推定を行う際には、*RMSEA* では変数が増えると値が下がるが、*CFI* と *TLI* では設定したモデルの形によって変数が増えると上がる場合と下がる場合があることがわかった。このことは、変数が多い時は *RMSEA* を利用すると誤ったモデルでも「フィットが良い」とみなしてしまう可能性があることを示している。

指標の分布について

GFI と *AGFI* については、*GFI* の値を g_0 , *AGFI* の値を a_0 とすると、 $\text{tr}(\hat{\Sigma}^{-1}S)=p$ の時、つまり推定法に最尤法を用いてかつ構造が定数変換に関して不変の場合 (Bock & Bargmann, 1966; Browne, 1974) に、

$$\chi_0^2 = \frac{Np}{2} \left(\frac{1}{g_0} - 1 \right), \quad \chi_0^{*2} = \frac{Np}{2} \left[\frac{p(p+1)}{2df(1-a_0)} - 1 \right] \quad (24)$$

が自由度 df の χ^2 分布に従うことが分かっている (Maiti & Mukherjee, 1990)。したがって χ^2 分布の性質から、同じ *GFI*, *AGFI* の値でもその分散はサンプルサイズに反比例することがわかる。

また、*RMSEA* は大標本では非心カイ二乗分布を使って近似でき、その95%信頼区間は

$$\left(\sqrt{\frac{\lambda_L}{(N-1)df}}, \sqrt{\frac{\lambda_U}{(N-1)df}} \right) \quad (25)$$

である。ただし自由度 df , 非心度 λ の非心 χ^2 分布の累積密度関数を $\chi^2(x | \lambda, df)$ とすると、 λ_L, λ_U は

$$\begin{aligned} \chi^2(x = T_{ML} | \lambda_L, df) &= 0.95 \text{ となる } \lambda_L \\ &\text{if } \chi^2(0 | \lambda_L, df) \geq 0.95 \\ \lambda_L &= 0 \quad \text{otherwise} \end{aligned} \quad (26)$$

$$\begin{aligned} \chi^2(x = T_{ML} | \lambda_U, df) &= 0.95 \text{ となる } \lambda_U \\ &\text{if } \chi^2(0 | \lambda_U, df) \geq 0.05 \\ \lambda_U &= 0 \quad \text{otherwise.} \end{aligned} \quad (27)$$

となる (Browne & Cudeck, 1993)。

また *RMR*, *SRMR* の漸近分散も計算できることがわかっていて (Ogasawara, 2001)。

4.2. “Golden Rule”, “Rule of Thumb” 又は “Cut-off Criteria” について

これまで、適合度指標は一般的に0.90以上が望ましい (Bentler & Bonett, 1980) とされ、これが長年基準とされてきた。

また *RMSEA* は0.05以下だと “close fit”, 0.08 までなら “fair fit”, 0.1 までなら “moderate fit”, 0.1 以上だとよくない (Browne & Cudeck, 1993; Browne & Mels, 1990; Stieger, 1989) とされてきた。

しかし近年ではさまざまなシミュレーション研究から基準としては0.90は低すぎるとされ、一般に0.95以上が推奨されるようになってきた。例えば Shevlin & Miles (1998) はシミュレーション研究から、非常に小さな(真のモデルが1因子6変数の)モデルでも、誤った2因子モデルを用いても *GFI* の下位25%点でさえ0.90以上になることから、「*GFI* が0.9以上ならモデルの当てはまりがよい」というルールは特異度が低いルールであるとした(因子負荷各0.5, 200データ, サンプルサイズ50~800)。

また同様のシミュレーション研究から Hu & Bentler (1998, 1999) は、最尤法を用いた場合、*CFI*, *IFI*, *TLI*, *RNI*, *ghat* については0.95以上、*Mc* については0.90以上、*SRMR* については0.08以下、*RMSEA* については0.06以下という基準を提案している。さらに彼らは、異なる適合度指標を併用すると効果的であるとしている。また、小標本(250以下)では *TLI*, *Mc* は低くなりすぎ、また *RMSEA* は高くなりすぎるので好ましくないとしている。

現在上記の Hu & Bentler (1998, 1999) の基準が非常に影響力を持っていることに対して、Marsh, Hau & Wen (2004) は追試実験を行い、彼らの研究で対象にした「(真のデータ発生モデルと異なる) 誤ったモデル」の設定を問題視し、基準を過剰に一般化して適用することの問題点を指摘している。また彼らは、「適合度がある値以上なら当てはまりがよい」という絶対基

準としての“Cut-off Criteria”を用いると、一般にサンプルサイズが大きくなると特異度が下がることを指摘した。

4.3. どの指標がよく利用されるか？

McDonald & Ho (2002) は 1995 年から 1997 年にかけて、心理学の有名学術誌 13 誌に掲載された論文の中で基準を満たしたものの 41 を調べた。結果として、全ての論文で尤度比検定統計量の値と自由度が報告されていた。また尤度比検定の結果をどれくらいそのまま受け取っているかという点に関しては、提案されたモデルと飽和モデルを比較した尤度比検定で有意でなかった（つまり尤度比検定からは自分のモデルが棄却されなかった）ものは五つ（標本サイズ 70～461）であった。基本的に「尤度比検定での棄却・保持をモデル選択やモデル確認に直接結びつけない」というコンセンサスが得られていることが示されたと言える。そのほかの指標で取り上げられたものは、それぞれ CFI (21 回), RMSEA (20 回), GFI (15 回), TLI (13 回), NFI (9 回) の順であった。

4.4. 同値モデルに関する注意

これらの適合度指標を利用する際に注意すべき点として、同値モデルの存在がある。適合度指標によってあるモデルの当てはまりが高いと判断されたとしても、そのモデルが唯一のモデルであるという保証はない。なぜなら、同じ適合度を示す別のモデルが存在する可能性は極めて高いからである。

例えば MacCallum, Wegener, Uchino & Fabrigar (1993) のレビューによると、心理学の代表的な 3 雑誌で 1988 年から 1991 年に掛けて掲載された 99 論文のうち、同値モデルが存在するモデルは 53 もあった。このように代表的な学術論文においても同値モデルの問題は無視されてきたが、因子間の関係などが逆転しても適合度が変わらない同値モデルも存在することがしばしばあり、結果の解釈そのものが大きく変更される可能性があるという点で非常に重要である。

この問題に関して、Lee & Hershberger (1990), Mayekawa (1994), Stelzl (1986) が一つのモデルから同値モデルを導出する方法を示しており、大津 (1996) は同値性の自動判定を行うプログラムの作成法も示している。また、Bekker, Merckens & Wansbeek (1994) が同値モデルの問題について詳しいので参照されたい。

5. 各適合度指標の関係と感度についてのシミュレーション研究

Hu & Bentler (1998) はある確認的因子分析モデルにおいて、代表的な 15 の適合度指標の様々なサンプルサイズでの相関係数を算出し、前節に記載したような結果を得ている。しかし、この研究からはモデルの複雑さによって適合度指標間の相関関係にどのような変化が生じるかがわからない。またシミュレーションの繰り返し数が少ない（この問題については次節でより詳しく論じる）。

そこで本節では、Hu & Bentler (1998) より非常に多い繰り返し数（水準あたり 10,000）分シミュレーションデータを発生させ、1 潜在変数あたりの観測変数の数を変化させることでモデルの複雑さによる適合度指標間の関係の変化を調べた。また同時に、いわゆる Golden Rule による各適合度指標の感度を調べた。シミュレーションは以下のように実施した。

(1) モデル

Curran *et al.* (2002) と同じ単純な（確認的因子分析モデルではない）モデルを用いた。因子数は 3 とした。また観測変数については、1 因子あたりの観測変数の数を 3 変数、5 変数、7 変数、10 変数の 4 条件を用意した。この設定は、1 因子あたりの観測変数数が適合度指標に影響を及ぼすという報告（Anderson & Gerbing, 1984; Ding, Velicer & Harlow, 1995）をふまえ、これを操作することを意図したものである。母数の真値は図 1 のように設定した。

(2) データ生成と推定

観測変数のシミュレーションデータセットは、次のように生成された。まず、モデル母数の真値を用いて真の分散共分散行列を計算した。次に、平均が零ベクトル・分散共分散行列がいま計算した真値である多変量正規分布から、乱数発生により N 標本の観測データ

図 1. 本研究で用いられたモデルと母数の真値

を発生させた。乱数発生法には、さまざまな良い性質を持つとされる Mersenne Twister を用いた (Matsumoto & Nishimura, 1998)。また、ここでは Hu & Bentler (1998) のシミュレーション結果との比較も考え、これまで紹介してきた指標に加えて CN (Hoelter, 1983) を加えた 15 の指標 (NFI , RFI , TLI , IFI , RNI , CFI , GFI , $AGFI$, $ghat$, CAK , CK , Mc , CN , $SRMR$, $RMSEA$) を算出した。

(3) 繰り返し

上記のデータセット生成を、1 水準につき 10,000 回繰り返した。

(4) サンプルサイズ

生成する 1 データセットあたりのサンプルサイズ N は、 $N=50, 200, 400, 10,000$ の 4 条件であった。

ここで、いわゆる Golden rule として適合度指標が 0.9 以上、及び 0.95 以上 (ただし小さい方が適合が良い指標である $RMSEA$, $SRMR$, CAK , CK については 0.1 以下、及び 0.05 以下。また CN は 200 以上; Hoelter, 1983) を採択するルールを用いた場合、10,000 データセットのうち真のモデルを正しく採択する回数 (つまり感度) を表 1・表 2・表 3 に示した。

また各適合度指標を標準化し、さらに上述の小さい方が良い 4 指標は他の指標と方向性が逆であるので -1 倍して、多次元尺度法を用いて 2 次元で表示したものを図 2, 図 3 に示した。ここで、距離はユークリッド距離を用いた。ストレスは最大のもので 0.053 であり、また RSQ はすべて 0.995 以上であり、2 次元でデータの変動を十分説明している。このとき各条件ごとに標準化しているため、 NFI と RFI , CAK と CK , GFI と $AGFI$ は同値になる。したがって、 RFI と CK , $AGFI$ を除外した 12 指標の関係を考察する。

これらの図から得られる知見を以下に記載する。

- 1 TLI , IFI , RNI , $ghat$, CAK , Mc が主クラスターを構成する。
- 2 $RMSEA$, CFI はゆるやかな第二クラスターを構成し、サンプルサイズが大きいほど主クラスターから分離する。
- 3 $SRMR$ は他の指標とは分離して布置する。
- 4 CN も他の指標と分離して布置するが、 $SRMR$ ほどではない。また、分離の度合いは観測変数数が小さくサンプルサイズが大きい時に大きい。
- 5 GFI はサンプルサイズが小さく観測変数数が大きいときには他の指標とは分離するが、サンプルサイズが大きくなるにつれて主クラスターに吸

表 1. Golden Rule の感度 ($AGFI$, CAK , CFI , CK , CN)

			潜在変数あたりの観測変数数			
指標	基準	N	3	5	7	10
$AGFI$	0.9	50	266	0	0	0
		200	9,986	9,804	4,346	0
		400	10,000	10,000	10,000	10,000
		10,000	10,000	10,000	10,000	10,000
	0.95	50	2	0	0	0
		200	5,742	29	0	0
		400	9,965	9,653	2,919	0
		10,000	10,000	10,000	10,000	10,000
CAK	0.1	50	0	0	0	0
		200	0	0	0	0
		400	0	0	0	0
		10,000	10,000	10,000	10,000	10,000
	0.05	50	0	0	0	0
		200	0	0	0	0
		400	0	0	0	0
		10,000	10,000	10,000	10,000	1,328
CFI	0.9	50	9,823	9,181	7,010	662
		200	10,000	10,000	10,000	10,000
		400	10,000	10,000	10,000	10,000
		10,000	10,000	10,000	10,000	10,000
	0.95	50	8,651	6,137	2,519	27
		200	10,000	10,000	10,000	10,000
		400	10,000	10,000	10,000	10,000
		10,000	10,000	10,000	10,000	10,000
CK	0.1	50	0	0	0	0
		200	0	0	0	0
		400	0	0	0	0
		10,000	10,000	10,000	10,000	10,000
	0.05	50	0	0	0	0
		200	0	0	0	0
		400	0	0	0	0
		10,000	10,000	10,000	10,000	1,282
CN	200	50	48	0	0	0
		200	9,668	9,553	9,337	8,518
		400	10,000	10,000	10,000	10,000
		10,000	10,000	10,000	10,000	10,000

表 2. Golden Rule の感度 (*GFI*, *ghat*, *IFI*, *Mc*, *NFI*)

			潜在変数あたりの観測変数数			
指標	基準	N	3	5	7	10
GFI	0.9	50	7,513	0	0	0
		200	10,000	10,000	9,910	35
		400	10,000	10,000	10,000	10,000
		10,000	10,000	10,000	10,000	10,000
	0.95	50	419	0	0	0
		200	9,994	3,392	0	0
		400	10,000	10,000	9,672	9
		10,000	10,000	10,000	10,000	10,000
ghat	0.9	50	9,972	9,563	7,431	505
		200	10,000	10,000	10,000	10,000
		400	10,000	10,000	10,000	10,000
		10,000	10,000	10,000	10,000	10,000
	0.95	50	9,174	6,664	2,682	24
		200	10,000	10,000	10,000	10,000
		400	10,000	10,000	10,000	10,000
		10,000	10,000	10,000	10,000	10,000
IFI	0.9	50	9,891	9,358	7,392	813
		200	10,000	10,000	10,000	10,000
		400	10,000	10,000	10,000	10,000
		10,000	10,000	10,000	10,000	10,000
	0.95	50	8,833	6,420	2,725	34
		200	10,000	10,000	10,000	10,000
		400	10,000	10,000	10,000	10,000
		10,000	10,000	10,000	10,000	10,000
Mc	0.9	50	8,912	4,331	822	0
		200	10,000	9956	9487	7121
		400	10,000	10,000	9999	9915
		10,000	10,000	10,000	10,000	10,000
	0.95	50	7,200	3,027	488	0
		200	9,936	8,965	7,331	4,461
		400	10,000	9,971	9,636	8,420
		10,000	10,000	10,000	10,000	10,000
NFI	0.9	50	4,563	0	0	0
		200	10,000	9,995	8,999	146
		400	10,000	10,000	10,000	10,000
		10,000	10,000	10,000	10,000	10,000
	0.95	50	203	0	0	0
		200	9,794	1,406	0	0
		400	10,000	9,994	8,581	32
		10,000	10,000	10,000	10,000	10,000

表 3. Golden Rule の感度 (*RFI*, *RMSEA*, *RNI*, *SRMR*, *TLI*)

			潜在変数あたりの観測変数数			
指標	基準	N	3	5	7	10
RFI	0.9	50	499	0	0	0
		200	9,950	9,649	5,334	10
		400	10,000	10,000	10,000	9,999
		10,000	10,000	10,000	10,000	10,000
	0.95	50	8	0	0	0
		200	5,563	80	0	0
		400	9,951	9,595	4,486	3
		10,000	10,000	10,000	10,000	10,000
RMSEA	0.1	50	8,904	9,583	9,730	9,038
		200	10,000	10,000	10,000	10,000
		400	10,000	10,000	10,000	10,000
		10,000	10,000	10,000	10,000	10,000
	0.05	50	5,969	4,323	2,017	59
		200	9,241	9,954	9,998	10,000
		400	9951	10,000	10,000	10,000
		10,000	10,000	10,000	10,000	10,000
RNI	0.9	50	9,823	9,181	7,010	662
		200	10,000	10,000	10,000	10,000
		400	10,000	10,000	10,000	10,000
		10,000	10,000	10,000	10,000	10,000
	0.95	50	8,651	6,137	2,519	27
		200	10,000	10,000	10,000	10,000
		400	10,000	10,000	10,000	10,000
		10,000	10,000	10,000	10,000	10,000
SRMR	0.1	50	9,636	8,670	7,520	6,486
		200	10,000	10,000	10,000	10,000
		400	10,000	10,000	10,000	10,000
		10,000	10,000	10,000	10,000	10,000
	0.05	50	2,006	63	3	0
		200	9,952	9,745	9,404	8,773
		400	10,000	10,000	10,000	10,000
		10,000	10,000	10,000	10,000	10,000
TLI	0.9	50	8,989	8,301	5,923	433
		200	10,000	10,000	10,000	10,000
		400	10,000	10,000	10,000	10,000
		10,000	10,000	10,000	10,000	10,000
	0.95	50	7,275	5,233	2,041	24
		200	9,940	9,995	9,997	9,996
		400	10,000	10,000	10,000	10,000
		10,000	10,000	10,000	10,000	10,000

図2. MDSの結果のプロット (観測変数3及び5の場合)

図3. MDS の結果のプロット (観測変数 7 及び 10 の場合)

収される。

- 6 *NFI* は $N=400$ までは主クラスターから (*SRMR* ほどではないが) 分離して布置するが, $N=10,000$ では主クラスターに吸収される。
- 7 $N=10,000$ では, どの条件でも *SRMR* 以外は指標が一直線上に並ぶ, N が小さいと *NFI* も直線から離れる。
- 8 サンプルサイズが 10,000 の時は, 観測変数数が違っていても, 布置はトポロジカルに保存されている。

また, *Mc* と *ghat* などのように式の形がまったく違うものであっても, すべての条件でほぼ同じ布置をしていたものもある。“7”の *SRMR* 以外が一直線に並ぶことに必然性があるのか, そしてその事実は何を表すかなどは, 今後の研究課題としたい。

なお, Hu & Bentler (1998) は分布の非正規性を 7 水準に, 標本数を 6 水準に操作して, 今回と同じ 15 個の適合度指標のふるまいを調べるシミュレーション研究を行った。しかし繰り返し数は 200 であり, われわれの 10,000 回に比べ大幅に少なかった。彼らは, これらすべての条件をつぶして作った相関行列を元に, 指標間の類似性について議論している (p. 438)。彼らの結果 (最尤推定による) は,

- 1 *NFI*, *RFI*, *GFI*, *AGFI*, *CAK*, *CK* が第一クラスターを主構成する。
- 2 *TLI*, *IFI*, *RNI*, *CFI*, *Mc*, *RMSEA* が第二クラスターを構成する。
- 3 *CN* と *SRMR* は, 他の指標との類似度は低い, というものであった。*SRMR* が他の指標と異なった性質を持つことについては, 我々と Hu & Bentler (1998) の結果とは一致している。また, *CN* についても, 我々の結果でも大標本数・小観測変数数条件では他の指標と分離したことから, 彼らの結果と整合的だといえる。しかし, 我々の結果では繰り返し数を 10,000 回と大きくとり, 標本数や観測変数数を細かに操作したことによって, 各指標の類似性と相違点をよりきめ細かく調べることが可能になった。

6. 「尤度比検定統計量の偏り」による先行研究の統合的理解と新しい Golden rule に向けて

これまで見てきたように, 大部分の適合度指標は尤度比検定統計量の関数とみなすことができる。尤度比検定統計量の漸近分布は χ^2 分布であることが知られ

ており, 3 節での相対的指標や非心度を利用した指標の多くではこの性質から指標の導出がなされている。

しかし, そもそもサンプルサイズが小さい時には尤度比検定統計量の分布は理論上の χ^2 分布から乖離することが知られており (Curran *et al.*, 2002; 星野・橋本・繁樹, 2001; Hu, Bentler & Kano, 1992), 一般的には理論上の χ^2 分布よりも大きめに偏る。Marsh, Hau, Balla & Grayson (1998) が行ったシミュレーション研究では, サンプルサイズが 50 の時は χ^2/df の比が 2 変数/因子で 0.788, 12 変数/因子では 1.476 であったのに対し, サンプルサイズが 1,000 の時は 2 変数/因子で 0.969, 12 変数/因子では 1.015 であった。ここから Marsh らは, 尤度比検定統計量は小標本で大きな偏りがあり, 観測変数が少ないときは理論上の分布より小さめに, 観測変数が多いときは大きめに偏るとしている。ただし, 2 観測変数/因子というモデルは非常に極端なものである。一般的には小標本では尤度比検定統計量は理論分布よりも大きめになると思われる。

尤度比検定統計量 T だけでなく様々な適合度指標がサンプルサイズに依存することは, 多くのシミュレーション研究により指摘されている。これまでに発表されてきた代表的なシミュレーション研究について, 研究の概要と結論とを表 4 にまとめた。しかし, これらの先行研究には 1 実験水準あたりの繰り返し数が十分大きいとは言えない, という問題があった。先行研究での繰り返し数は, 多くても 500, 通常は 200 前後である。これだけの繰り返し数では, 誤差変動の影響が無視できないほど大きくなる可能性が高い。実際, 先行研究におけるサンプルサイズと T の平均値の大きさととの関係は, 必ずしも単調減少せず, 変動を含むものが多い (e.g. Marsh *et al.*, 1988: p. 400 Table 3; Hu *et al.*, 1992: p. 356 Table 3; Curran, West & Finch, 1996: p. 22 Table 1)。

そこでここでは, このような尤度比検定統計量の理論分布との乖離の度合いを調べるための実験を, これまでの先行研究よりも非常に多い繰り返し数で行った。

(1) モデル

前節と同じ図 1 のモデルを利用した。また観測変数については, 1 因子あたりの観測変数の数を 3 変数から 10 変数まで 1 変数づつ変更した 8 条件を用意した。

(2) データ生成と推定

前節と同様である。但しここでは, 真のモデルに対

表 4. 過去の主なシミュレーション研究

著者	Anderson & Gerbing (1984)	Marsh, Balla & McDonald (1988)	La Du & Tanaka (1989)
論文誌	<i>Psychometrika</i>	<i>Psychological Bulletin</i>	<i>Journal of Applied Psychology</i>
主要な目的	標本誤差などが、収束・不適解・適合度指標に与える影響	標本数が適合度指標に与える影響	標本数, 推定法, モデル誤設定が適合度指標に与える影響
モデル	CFA	CFA	CFA
標本数	* (50, 75, 100, 150, 300)	* (25, 50, 100, 200, 400, 800, 1,600)	* (35, 100, 300)
因子負荷量	* (0.6, 0.9, 混合条件)	* (大, 小)	固定 (0.55–0.75)
観測変数数	* (4–16)	9	6
因子数	* (2, 3, 4)	3	2
観測変数数: 因子数 (I/F) 比	* (2, 3, 4)	3	3
因子相関 or 潜在回帰係数	* (0.3, 0.5)	有り	0
推定法	ML	ML	* (ML, GLS)
非正規性	正規	正規	正規
モデル誤設定	なし	* (真のモデル, 誤設定モデル, 実データの再サンプリング×2)	* (真のモデル, 誤設定モデル, 実データの再サンプリング)
繰り返し数	100	10	50
実装	LISREL	LISREL	SAS
適合度指標	T , GFI , $AGFI$, RMR , TLI	EF , LHR , T , T/df , RMR , GFI , $AGFI$, CAK , CSK , CN . および各々を2種の incremental index にしたもの	GFI (ML), GFI (GLS), NFI (ML), NFI (GLS)
結論	<ul style="list-style-type: none"> ・非収束と不適解は, 小標本数・小 I/F 比で多く生じた ・GFI, $AGFI$, RMR には, 標本数・I/F 比, 因子数の有意な影響が見られた ・T の p 値には, どの要因の影響も見られなかった 	<ul style="list-style-type: none"> ・NFI, GFI, RFI は標本数の影響を強く受けた ・TLI, および同種の新しい4つの指標は相対的に標本数の影響を受けなかった 	<ul style="list-style-type: none"> ・標本数はどの適合度指標にも影響した ・推定法に合った適合度指標がより精度よい結果を返した ・誤ったモデルは各適合度指標に異なった影響を与えていた
著者	Hu, Bentler & Kano (1992)	Ding, Velicer & Harlow (1995)	Curran, West & Finch (1996)
論文誌	<i>Psychological Bulletin</i>	<i>Structural Equation Modeling</i>	<i>Psychological Methods</i>
主要な目的	標本数と分布が適合度指標に与える影響	推定法, 観測変数/因子数, モデル誤設定, 不適解が適合度指標に与える影響	非正規性とモデル誤設定が適合度指標に与える影響
モデル	CFA	潜在回帰モデル	CFA
標本数	* (150, 250, 500, 1,000, 2,500, 5,000)	* (50, 100, 200, 500)	* (100, 200, 500, 1,000)
因子負荷量	固定 (0.70–0.80)	* (0.50, 0.70, 0.90)	固定 (0.35, 0.7)
観測変数数	15	* (6, 9, 12, 15, 18)	9
因子数	3	3	3
観測変数数: 因子数 (I/F) 比	5	* (2, 3, 4, 5, 6)	3
因子相関 or 潜在回帰係数	固定 (9.39–0.50)	* (0.50, 0.70, 0.90)	0.3
推定法	* (ML, GLS, ERLS, HK, ADF, SB)	* (GLS, ML)	* (ML, GLS, SB)
非正規性	* (正規, 6種の非正規)	正規	* (正規, 少し非正規, 大きく非正規)
モデル誤設定	なし	なし	* (真のモデル, 3種の誤設定)
繰り返し数	200	100	200
実装	EQS	EQS	EQS
適合度指標	T	T/df , NFI , CFI , RNI , Mc	T
結論	<ul style="list-style-type: none"> ・ADF は標本数が大きいときはいくらかはたらかなかった ・Satorra-Bentler の補正を適用した T が最もよくはたらいた 	<ul style="list-style-type: none"> ・I/F 比と標本数が小さいとき, 因子負荷が小さいときに不適解がよく生じた ・GLS は ML よりも多く不適解を生じた ・NFI を除く適合度指標には, 不適解の影響は見られなかった ・NFI はさまざまな条件に影響されやすい ・$NNFI$ を除く適合度指標は, 推定法の影響を受けた ・推定法, I/F 比, 標本数, 負荷量の交互作用も見られた 	<ul style="list-style-type: none"> ・歪度 2.0, 尖度 7.0 以上の非正規性では問題が生じた ・ADF は標本数が小さいときには問題が生じた ・ほとんどの条件でも, SB が ML, ADF よりも上手くはたらいた

著者	Hu & Bentler (1998)	Fan, Thompson & Wang (1999)	Jackson (2001)
論文誌	<i>Psychological Methods</i>	<i>Structural Equation Modeling</i>	<i>Structural Equation Modeling</i>
主要な目的	モデル誤設定が適合度指標に与える影響	標本数, 推定法, モデル誤設定が適合度指標に与える影響	標本数, 母数数が適合度指標に与える影響
モデル	CFA	潜在回帰モデル	CFA
標本数	* (150, 250, 500, 1,000, 2,500, 5,000)	* (50, 100, 200, 500, 1,000)	* (50, 100, 200, 400, 800)
因子負荷量	固定 (0.7-0.8)	固定 (0.6-1.2)	* (0.6, 0.8)
観測変数数	15	12	20
因子数	3	4	* (1, 2, 3, 4, 5)
観測変数数: 因子数 (I/F)比	* (固定 (5), 変動 (5-7))	4	* (4, 5, 6.67, 10, 20)
因子相関 or 潜在回帰係数	固定 (0.3-0.5)	固定 (1)	0.5
推定法	* (ML, GLS, ADF)	* (ML, GLS)	ML
非正規性	* (正規, 6 種の非正規)	正規	正規
モデル誤設定	* (単純構造モデル/複雑なモデル, 真のモデル/誤設定モデル)	* (真のモデル, 2 種の誤設定)	なし
繰り返し数	200	200	200
実装	SAS	SAS	SAS
適合度指標	NFI, BL86, TLI, BL89, RNI, CHI, GFI, AGFI, gamma hat, CAK, CK, Mc, CN, SRMR, RMSEA	GFI, AGFI, NFI, TLI, CFI, MC	T, GFI, RMSR, RMSEA, NFI, NNFI, CFI
結論	<ul style="list-style-type: none"> ML と GLS を用いるときは, 第一に SRMR を, 第二に TLI, BL89, RNI, CFI, gamma hat, Mc, RMSEA がよい ADF を用いるときは, 第一に SRMR を, 第二に TLI, BL89, RNI, CFI がよい ML が, GLS や ADF よりもよい 		
著者	Curran, Bollen, Paxton, Kirby & Chen (2002)	Kenny & McCoach (2003)	Jackson (2003)
論文誌	<i>Multivariate Behavioral Research</i>	<i>Structural Equation Modeling</i>	<i>Structural Equation Modeling</i>
主要な目的	標本数, モデル誤設定が非心度を与える影響	観測変数数が適合度指標に与える影響	標本数, 母数数が適合度指標に与える影響
モデル	潜在回帰・MIMIC 型モデル	CFA	CFA
標本数	* (50, 75, 100, 200, 400, 800, 1,000)	200	* (50, 100, 200, 400, 800)
因子負荷量	固定 (様々)	固定	* (0.6, 0.8)
観測変数数	* (9, 13, 15)	* (6, 12, 20)	20
因子数	3	* (1, 2)	* (1, 2, 3, 4, 5)
観測変数数: 因子数 (I/F)比	* (4, 5, 6)	* (3, 6, 10, 12, 20)	* (4, 5, 6.67, 10, 20)
因子相関 or 潜在回帰係数	固定	固定	0.5
推定法	ML	ML	ML
非正規性	正規	正規	正規
モデル誤設定	* (真のモデル, 4 種の誤設定)	* (3 種の誤設定)	* 制約の入れ方を操作 (congeneric, tau 等価, parallel)
繰り返し数	500	100	200
実装	EQS	EQS	SAS
適合度指標	T	TLI, CFI, RMSEA	T, GFI, RMSEA, Mc
結論	<ul style="list-style-type: none"> 誤設定が小〜中程度で, 標本数が大いときは, 非心カイ二乗分布の近似はよかった 変数間に相関のないベースラインモデルでは, どの条件でも T は非心カイ二乗分布に従わなかった RMSEA は, 誤ったモデルで観測変数数の増加に伴って一般によい値を示すようになった CFI と TLI は, 誤ったモデルで観測変数数の増加に伴って一般に悪い値を示すようになった 標本数/母数数比が大きいときには, 適合度指標も大きくなった ただしその相対的效果は小さかった 		

「*」印が操作された属性であることを表す。

する尤度比検定統計量 T を計算した(ここで対立仮説は飽和モデルである)。

(3) 繰り返し

上記のデータセット生成を、1水準につき 10,000 回繰り返した。

(4) サンプルサイズ

生成する 1 データセットあたりのサンプルサイズ N は、 $N=50, 75, 100, 150, 200, 300, 400, 1,000, 10,000$ の 9 条件であった。

比較のための指標として、自由度で割った尤度比検定統計量の平均、有意確率を 0.05 に設定した時の棄却確率をそれぞれ表 5, 6 に記載した。また、図示したものをそれぞれ図 4, 5 に記載した。ここで理論上、それぞれは 1 と 0.05 になるはずである。

実際、漸近理論が正当化できる状況、たとえば N が 10,000 ではこれらの指標は理論値にほぼ一致している。しかし、行動科学の多くの研究で一般的な $N=1,000$ 以下では、(1) 1 因子あたりの観測変数の数が多いほど、(2) サンプルサイズが小さいほど、平均として尤度比検定統計量は理論上の分布よりも正の方向に

移動することがわかる。その結果、正しいモデルの棄却確率は名義水準に比べて非常に大きくなる。例えば $N=50 \cdot 1$ 因子あたりの観測変数数 10 の条件において、99.52% ものデータが棄却される。この傾向はこのような極端な場合だけでなく、例えば行動科学分野で十分大きいとされる $N=400$ においても、10.98% が棄却される。いま有意確率は 0.05 に設定してあったから、正しいモデルの期待棄却率は 5% であるはずであるが、これだけのずれが生じてしまっている。

さらにこの乖離を視覚的に理解するため、1 因子あたりの観測変数数が 10 のときに、もっとも近似がよいと考えられる $N=10,000$ の条件と、もっとも近似が悪いと考えられる $N=50$ の条件とにおける標本尤度比検定統計量のヒストグラムと漸近理論に基づくカイ二乗分布とを図 6・図 7 に図示した。 N が 10,000 と大きい前者ではたしかにより近似が得られているが、後者の図 7 からはこの問題の深刻さが明確に理解できる。理論上の尤度比検定統計量の平均は (χ^2 分布の性質より) 自由度と等しくなるはずであるが、やはり実際は大きめになってしまう。したがって非心度も常に正の

表 5. 尤度比検定統計量を自由度で割ったものの平均

観測変数数	サンプルサイズ								
	50	75	100	150	200	300	400	1,000	10,000
3	1.0733	1.0497	1.0397	1.0281	1.0218	1.0140	1.0129	1.0007	1.0021
	0.0033	0.0033	0.0032	0.0032	0.0031	0.0031	0.0031	0.0031	0.0031
4	1.1301	1.0811	1.0602	1.0410	1.0271	1.0215	1.0179	1.0049	1.0013
	0.0023	0.0022	0.0021	0.0021	0.0021	0.0021	0.0021	0.0021	0.0021
5	1.1607	1.1021	1.0710	1.0441	1.0334	1.0220	1.0173	1.0056	1.0000
	0.0018	0.0017	0.0016	0.0016	0.0016	0.0016	0.0016	0.0015	0.0016
6	1.1971	1.1198	1.0882	1.0525	1.0394	1.0258	1.0181	1.0078	0.9996
	0.0015	0.0014	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013	0.0013
7	1.2308	1.1370	1.0972	1.0644	1.0466	1.0326	1.0200	1.0096	1.0012
	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0011	0.0010	0.0010
8	1.2698	1.1573	1.1121	1.0691	1.0512	1.0335	1.0242	1.0094	1.0017
	0.0012	0.0011	0.0010	0.0010	0.0009	0.0009	0.0009	0.0009	0.0009
9	1.3153	1.1791	1.1255	1.0777	1.0574	1.0370	1.0293	1.0103	1.0008
	0.0011	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0008
10	1.3611	1.2007	1.1397	1.0884	1.0637	1.0406	1.0299	1.0115	1.0021
	0.0010	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007	0.0007	0.0007

下段は平均値の標準誤差をあらわす

表 6. 有意水準 0.05 のときの棄却率

観測変数数	サンプルサイズ								
	50	75	100	150	200	300	400	1,000	10,000
3	0.0816	0.0712	0.0626	0.0626	0.0587	0.0581	0.0548	0.0491	0.0510
4	0.1624	0.1102	0.0866	0.0748	0.0669	0.0648	0.0591	0.0523	0.0512
5	0.2682	0.1610	0.1199	0.0914	0.0790	0.0665	0.0637	0.0558	0.0500
6	0.4460	0.2478	0.1716	0.1174	0.0932	0.0764	0.0674	0.0562	0.0483
7	0.6491	0.3521	0.2354	0.1601	0.1189	0.0886	0.0744	0.0598	0.0514
8	0.8428	0.5048	0.3356	0.1875	0.1396	0.1018	0.0851	0.0640	0.0517
9	0.9660	0.6748	0.4559	0.2477	0.1847	0.1253	0.1009	0.0635	0.0522
10	0.9952	0.8330	0.5929	0.3298	0.2301	0.1428	0.1098	0.0695	0.0522

図 4. 尤度比検定統計量を自由度で割ったものの平均

図 5. 有意水準 0.05 での棄却率

バイアスを有することになり、非心度を用いた諸指標及び相対的指標は小標本では小さめの値になることがわかる。その結果、例えば Hu & Bentler (1998, 1999) で示された「*RMSEA* が小標本では大きめになるバイアスがある」ということになる（これまでの先行研究では因子負荷の大きさが議論されてきたことがある。

しかしここでは詳しく取り上げないが、本研究でのセッティングでは因子負荷を変えてもほとんど結果は変わらないという結果を我々は得ている。

この結果から考えられる既存の尤度比検定統計量 T を利用した検定及び適合度指標によるモデルの選択の最大の問題点は、因子あたりの観測変数を増やす

図6. 標本尤度比検定統計量のヒストグラム ($N=50$)と漸近理論に基づく χ^2 分布

図7. 標本尤度比検定統計量のヒストグラム ($N=10,000$)と漸近理論に基づく χ^2 分布

ことで、自分の立てたモデルがより否定される方向に誤って判断されやすいという点にある。これは、潜在変数の測定の妥当性を向上させようと努力するほど、その努力が仇となって報われなくなることを示している。特に情報量を一定にするために、サンプルサイズが少ない場合は観測変数を増やそうとする場合が多い。このことは例えば村上 (2002) が「項目全体としてカバーする個人差の範囲を拡大し、構成概念の測定

の偏りを小さくする」と述べているが、このような努力が、より不当に正しいモデルを否定する方向に向かせることになる。

この問題に対して、岡田・星野・繁樹 (2004; 投稿中) では、Wakaki, Eguchi & Fujikoshi (1990) の Bartlett 補正の議論を利用して、共分散構造分析での尤度比検定統計量の Bartlett 補正を行い、小標本で χ^2 分布への近似が非常によくあることを確認している。たとえばサンプルサイズが $N=50$ で因子数2、因子あたりの観測変数数5の確認的因子分析モデルで、名義の有意水準を5%にした時の補正前の尤度比検定の棄却率は12.1%と非常に高かった。しかし補正を行うことで5.4%に押さえることができた。また、Bartlett 補正後の T を用いた各適合度指標はサンプルサイズに依存しない方向に補正されることがわかった。例えばRMSEAは上記のモデルで補正前は $N=50$ で平均0.02656、 $N=300$ で0.01890と1.41倍ほどの違いがあったが、補正後はそれぞれ平均が0.01909、0.01776と1.07倍に収まった。

これまでの先行研究では、サンプルサイズが小さい時のシミュレーション研究による適合度指標の性質の検討に関して、使用したモデルによって知見が異なったものになるという問題があり、議論が収束しなかった。これは、これまでの先行研究におけるシミュレーションでの繰り返し数が少なかったことによる変動の大きさの問題だけではない。今回示されたように、尤度比検定統計量の理論分布からの乖離が「因子あたりの観測変数の数」に代表されるような「モデルの自由度・複雑さ」に依存する、という問題がありながら、それを考慮しなかったことに起因する可能性が大いにあると考えられる。したがって、Bartlett 補正などを用いた検討が重要であろう。

また、このような補正を通じて、潜在変数の測定の妥当性を向上させようとする努力が、少なくともモデルの正当化にマイナスにははたらかないようにすることができるという利点がある。

多くの適合度指標は、非心度やベースラインモデルとの差などの漸近理論に基づいて導出されている。したがって、自由度の大きいときやサンプルサイズの小さいときには Bartlett 補正などを用いて、漸近理論による分布からの乖離を補正する必要がある。しばしばこのような問題に対して既存の T を使ったブートストラップ法を用いることが良いのではないかという議論があるが、適合度指標は漸近理論により導出されて

いる場合が多く、 T の補正を通じて適合度指標の式そのものを補正しなくては、この問題は解決しない。

これまで、Hu & Bentler (1998, 1999) などのような適合度指標に対する Golden rule の探索の研究は盛んに行われており、一方で Marsh *et al.* (2004) のようにこれに反論する研究も行われている。上記の T の理論分布からの乖離、及び補正後のサンプルサイズに対する頑健性から、Bartlett 補正後の尤度比検定統計量 T を用いた適合度指標についての Golden rule の探索は有望であると考えられる。

7. 平均構造のあるモデル、多群モデル、離散観測変数のある場合への応用について

7.1. 平均構造のある場合と多群モデルの場合

平均構造のあるモデルや、多群モデルにおいては、これまで紹介してきた適合度指標の内、 GFI 、 $AGFI$ 、 $(S)RMR$ などの「データの共分散行列とモデルによる共分散行列の推定値の距離」を用いた指標を利用することはできない。また、他の指標においても尤度比検定統計量を適切なものに置き換える必要がある。平均構造のある場合にこれを行わないと、まったく異なったモデルが採用されることが十分にありえる (Wicherts & Dolan, 2004)。ここで平均構造のある場合のモデルでは尤度比検定統計量は以下になる。

$$T = (N-1) \{ (\bar{y} - \mu(\hat{\theta}))' \Sigma^{-1}(\hat{\theta}) (\bar{y} - \mu(\hat{\theta})) + \text{tr}(\Sigma^{-1}(\hat{\theta})S) - \log |\Sigma^{-1}(\hat{\theta})S| - p \}. \quad (28)$$

ただし $\bar{y}$, S は観測変数の平均ベクトルと不偏共分散行列である。

また、因子平均の群間比較などを行いたい場合は、最低限切片項と因子負荷を群間で共通にする「測定不変 (Measurement invariance)」の仮定が成立することを確かめる必要がある (Meredith, 1993)。

7.2. 離散観測変数のある場合

さらに難しいのが、観測変数の一部に離散変数が存在する場合である。

そもそも離散観測変数が存在する時は、離散変数を連続変数とみなす場合と、離散変数用のモデリングをする場合に分かれる (狩野・三浦, 2003)。

前者の方略を採用した場合に関しては、順序尺度でカテゴリー数の少ない場合でもうまく行くことが多い (萩生田・繁樹, 1996) ことが知られているが、データ

の性質によって適合度指標が正しく機能する場合もしない場合もある (Hutchinson & Olmos, 1998) ことに注意する必要がある。

後者の観測変数に離散変数がある場合の構造方程式モデリングは、項目反応理論 (Item Response Theory) モデルを下位モデルとして含むものなど様々なものが提案されている (Moustaki & Knott, 2000; Muthén, 1978; Sammel, Ryan & Legler, 1997)。

その中で、最も古くから実際に利用されているのは、ポリコリック相関係数やポリシリアル相関係数を用いる方法である。 $LISREL$ や $Mplus$ など順序尺度の観測変数をモデリングすることができる構造方程式モデリングのソフトの場合、以下の二段階推定法を行っている

1: ポリコリック相関係数・ポリシリアル相関係数の推定

各カテゴリカル観測変数の背後に連続潜在変数を仮定し、それらの間、またはそれらと連続観測変数の間の相関係数をポリコリック相関係数 (Polychoric Correlation Coefficient) やポリシリアル相関係数 (Polyserial Correlation Coefficient) で推定する。

2: 構造方程式モデリングの母数推定

推定された相関係数、共分散行列の推定値とその推定値の共分散行列による一般化最小二乗法を用いて構造方程式モデリングの母数を推定する。

この解析法を確立した研究者の一人である Muthén (1993) は、離散観測変数のある場合は (1) ポリコリック・ポリシリアル相関係数のフィット、(2) 構造方程式間のフィット、の二つのレベルのフィットを確認する必要性を強調している。

具体的には、Muthén (1993) はレベル 1 のフィットについては、例えばポリコリック相関係数については各相関係数ごとに (2 変数のペアごとに) モデルの期待度数と観測度数の差をピアソンの χ^2 検定を用いて調べる方法を、またポリシリアル相関係数については各相関係数ごとに、尤度比検定を利用することを提案している。

また、レベル 2 のフィットについては、一般化最小二乗法の理論から χ^2 検定を行うか、共分散行列の推定値とモデルで表現される共分散行列を用いて GFI などの指標を計算することができる。

このレベル 2 での適合度指標の挙動については複数の研究がある (Muthén & Kaplan, 1992; Potthast,

1993; Hutchinson & Olmos, 1998) が、一般に一般化最小二乗法の χ^2 検定は正しいモデルも棄却する傾向にバイアスを持つとされる。また、GFI や AGFI は分布が歪む場合には正しいモデルに対しても低くなりがちとなる。例えば Hutchinson & Olmos (1998) による単純なモデル (1 因子に 4 指標で 4 因子) でのシミュレーション研究では (χ^2 , *RMSEA*, *TLI*, *CFI*, *IFI*, *RFI*, *Mc*) の中でポリシリアル相関係数を用いた GLS による *RMSEA* が一番安定しているとして推奨している。*TLI*, *CFI*, *IFI* もあまり悪くない (特に *CFI* と *IFI* は全ての条件で平均 0.95 を超えている) と報告している。

ここで、Muthén の提案した二つのレベルでの適合度の確認法の問題点としては、(1) この場合の全体としての飽和モデルを考えることが出来ないために、モデル全体の (レベル 1 もレベル 2 も含めた) 適合度指標が (情報量基準以外では) 提案されていないということ、(2) レベル 1 のフィットについては、全変数を一括してフィットをチェックすることができないということである。

近年ソフトウェアの普及により、離散観測変数を含む構造方程式モデリングの利用が多くなってきた。したがって、全体としての飽和モデルを提案し、モデル全体での適合度をチェックする方法、または次善の策として、レベル 1 のフィットを全変数一括してチェックする方法の開発が今後望まれている。

8. モデル修正と段階的なモデル構築について

8.1. モデル修正について

ここでは修正指標を用いたモデルの修正、モデルに適合しない変数の除去、Item Parceling の話題を紹介する。

モデルの固定母数を自由母数にする

構造方程式モデリングにおいて、適合度があまり高くない、かつ実質科学的な示唆が無いときは一般に修正指標 (Modification Index: MI; Jöreskog & Sörbom, 1986) が利用されることが多い。これは、モデルの中のある一つの固定母数を自由母数にした時の *LM* 検定 (Lagrange multiplier test) 統計量の値のことであり、自由度 1 の χ^2 分布に従うことから、3.84 以上であれば H_0 が棄却される。

ここで *LM* 検定を利用する理由は、尤度比検定では「修正前の推定値」も「修正後の推定値」も必要である

のに対して、*LM* 検定統計量の値は「修正前の推定値」だけで計算ができるからである。

また、修正指標と *AIC* の関係として、漸近的に「修正による *AIC* の減少分 = 修正指標 - 2」であることが容易に示せる (Kaplan, 2000)。したがって、修正指標が 2 より大きい固定母数を自由母数にすると *AIC* が改善することになり、これを利用したモデル修正を行うこともできる。

なお、*LM* 検定や尤度比検定では修正指標を 3.84 と比較するが、*AIC* を用いた方法では修正指標を 2 と比較する。したがって、*AIC* を用いた方法のほうがより対立モデルを支持しやすくなっている。

また、モデルの段階的修正によって誤った結果が導かれるということが Mac-Callum, Roznowski & Necowitz (1992) によって示されている。彼らは大規模データ (3694 人) からのリサンプリングを行ってデータを複数生成し、生成されたデータのモデルフィットを行い段階的に 4 回モデルを修正したところ、1 データあたりのサンプルが少ないとき ($N=100 \sim 800$) は、全データで修正を行った結果の最終モデルと一致するデータセットは全体の半分以下だった。

このように有限母集団から一部抽出した場合でも母集団での結果と大きく異なることから、一般的に仮定される無限母集団に対するモデル修正の一般化はより危険であると彼らは結論している。

モデルに適合しない変数を除去する

しばしばモデルに適合しない観測変数を除去することによって、適合度指標は劇的に改善することがあるため、多くの研究で変数の除去が行われる。

しかし、一般に潜在変数の指標変数を適合度だけに注目して観測変数を減らすと、内容的妥当性に問題が生じることが多い (南風原, 2002)。この問題は Bandwidth-fidelity dilemma (Cronbach & Gleser, 1965)、つまり尺度構成の際にその因子の内容を豊富にするようにさまざまな項目を含めるほど、一貫性 (信頼性) が低減するという問題とちょうど逆であり、モデル適合度を上げるために、本来の解析の対象である潜在変数とは異なった潜在変数間の関係を調べることになってしまわないように十分この問題には注意するべきである。

また、適合度を改善する目的でモデルに含まれる因子による説明率が低い変数をターゲットとして除去することがしばしば行われるが、これは適合度を改善させない場合が多い。なぜなら、一般に適合度が低い原

因はモデリングされていない部分にある場合、たとえば誤差相関が大きい場合が多いからである。この場合は、誤差相関があると思われる一連の変数群を除去することで適合度はしばしば劇的に改善する。

誤差共分散の導入について

誤差共分散があると考えられる場合に、これをモデリングすることで、測定方程式部分を改善することも多くの研究者によって提案されている(狩野 2002b; Rubio & Gillespie, 1995)。しかし誤差共分散の導入には注意すべき点がいくつかある。そもそも誤差すべてに共分散を導入することはモデルの識別性から不可能であり、一部の誤差間に共分散を導入するには、そこだけ共分散が存在することに実質科学的意味があることが要求される。また、誤差共分散が多くなると、その共分散を利用して共通因子を構成できる可能性があるなど、同値モデルの存在の可能性が増してくる。特に、潜在変数の多くは、それを測定する観測変数との関係によって命名され、そのモデルによる解析結果の解釈が行われる。したがって誤差共分散の存在によって潜在変数と観測変数の関係が単純でなくなったときに、当該潜在変数の表現する内容が変化することが十分考えられるので注意すべきである。

8.2. Item Parceling Issues: 尺度化について

Item Parceling (Bandalos & Finny, 2001; Hagtvet & Nasser, 2004; 狩野, 2002a; Little, Cunningham, Shahar & Widaman, 2003) とは、同一因子を測定していると考えられる多数の指標の合計点(尺度得点)をその因子の「観測値」とみなし、尺度得点間の回帰分析などにより、因子間の関係(つまり構造方程式のパラメータ)を推定するという方法であり、心理学や医学を中心に広く行われている。その理由としては(Bandalos & Finny, 2001),

- 1 個々の指標を投入するよりも信頼性があるはずである (Cattell & Burdsal, 1975; Kishton & Widaman, 1994)。
- 2 各項目が3件法などの場合でも、尺度得点にする分布が正規分布に近づくため、正規性を仮定している解析においては有利である。
- 3 モデルを小さくできる。観測変数の数に対するサンプルサイズの比が高くなり、推定が安定する。これに対して個々の指標を投入する場合は非常にモデルが大きくなり、推定が不安定になるはずである。

4 モデルフィットがよくなるはずである。

といった点が挙げられることが多い。これに対して、Item Parceling を批判する側からは

- I 上記の3は本当であるかどうかかわからない。逆に推定が安定しないという研究もある (MacCallum, Widaman, Zhang & Hong, 1999; Marsh, Hau, Balla & Grayson, 1998)。
- II 上記の4も本当であるかどうかかわからない。
- III 結果として得られる因子間の回帰(パス)係数などにバイアスが生じることがある (Bandalos, 2002)。

という批判もある。いずれにせよ、Item Parceling を行うのなら、最低限、合計した項目に一次元性があるという仮定を因子分析などでテストする必要があるであろう。

また、上記 III については、Item Parceling を行うことによって、因子間のパス係数の絶対値が小さくなる(正の係数なら過小評価) ことが Bandalos (2002) のシミュレーション研究によって報告されているが、この原因は星野、光永、繁樹、前川 (2003)、及び光永・星野・繁樹・前川 (2005) での「因子得点による回帰分析」の研究で示されている。すなわち、尺度得点ないし因子得点は推定の際の誤差分を含むため、希薄化が生じているのである。彼らの研究では因子得点を用いた構造方程式モデリングの母数推定におけるバイアスの修正法も示している。またこの問題については狩野 (2002a) も参照してほしい。

8.3. 段階的なモデル構築と推定

SEMの母数推定で一般に利用されている最尤法や最小二乗法などでは、構造方程式と測定方程式との母数推定を同時に行っている。したがって、構造方程式のモデル設定を変えることによって測定方程式の母数推定値が変動する (Burt, 1976; Kumar & Dillon, 1987; 椿, 2002; 星野, 2003) ことが指摘されている。具体的には、あるモデルの推定結果と、そのモデルに新しい潜在変数とそれを測定する観測変数を加えたモデルの推定結果を比較すると、元のモデル部分についても推定値が変化することによって、潜在変数の意味内容まで変化してしまうという問題である。この問題は一般に、測定方程式以外の部分のモデルフィットが悪いこと、例えば誤差変数間に共分散があるために起こる。またこのような場合、測定方程式だけが誤った推定がされるのではなく、モデリングされていない誤

差相関に引っ張られて構造方程式でのパスも過大に推定されることが多い。さらに、6変数でGFIが0.953という一見極めてモデル全体の適合度指標が高い場合にも2つのパスの大小を反対に推定してしまうような場合すら存在することが指摘されている(星野, 2003)。そこで星野(2003)は、多重指標モデルにおいてまず測定方程式部分のパラメータを推定し、それを固定して構造方程式部分を推定する方法を提案し、この方法が一致性などの性質を持つことを示している。また、Anderson & Gerbing (1988) は同様の関心事から、

- 1 まず全ての潜在変数が相関を持つように(構造方程式が飽和モデルであるように)モデルを構成して推定する。
- 2 関心のあるモデルを推定する。
- 3 上記の二つのモデルの差が(χ^2 検定などで)ないと判断されれば、関心のあるモデルを採用し、あると判断すれば、関心のあるモデルよりも構造方程式の推定の自由度のある(制約の少ない)モデルと飽和モデルを比較する

などといった段階的なモデルフィットを行う方法を提案している(この議論についてはAnderson & Gerbing, 1992; Fornell & Yi, 1992abも参照)。さらにFour step debateという議論が近年行われている(Bentler, 2000; Hayduk & Glaser, 2000; Markus, 2000; Mulaik & Millsap, 2000)。これは評価したいモデルのフィットを考えると、測定方程式のフィットと構造方程式のフィットを分離するという観点から考えられた方法という点でAnderson & Gerbing (1988)と同じである。具体的には

- 1 潜在変数はあるが、それ以外まったく制約の無いモデル＝探索的因子分析モデルを実行する。
- 2 測定方程式を評価したいモデルと同じにする。
- 3 構造方程式を評価したいモデルと同じにする。
- 4 さらに特定のパラメータを固定していく。

という四つの段階それぞれの当てはまりが良いかを、 χ^2 検定ではなく適合度指標で調べるという考え方である。

9. まとめと議論

本稿ではBentler & Bonnet (1980)以来問題とされてきた(特に行動科学分野への応用を考えた時の)構造方程式モデリングにおける尤度比検定の問題を整理した。そしてこのような問題意識からこれまで開発さ

れてきた様々な適合度指標の分類とそれらの比較についての先行研究のレビューを行った。また“Golden Rule”づくりの研究、つまりどの適合度指標がどれくらいのときにそのモデルを採用すべきかについての統一的な指標を探索している複数の先行研究をレビューし、これらの結論が一貫していないことを指摘した。さらに、大規模なシミュレーション研究から、適合度指標のGolden Ruleの感度と、適合度指標間の関係を調べた。

これまでの先行研究ではシミュレーションにおける1条件あたりの繰り返し数が100から400程度と少なく、標本変動の大きさが結果の非一貫性の一因である可能性を否定できない。そこで本研究では、繰り返し数を1条件あたり10,000回と大規模に設定した。その結果として(1)サンプルサイズが少ないと(2)1潜在変数あたりの観測変数数が多いと、尤度比検定統計量の分布が χ^2 分布から非常に大きく乖離することを再確認した。

適合度指標の大部分は尤度比検定統計量の関数であり、かつその漸近的性質を利用して開発されている場合が多いため、行動科学で一般に利用される状況設定では、適合度指標が有するはずのサンプルサイズに対する独立性などの性質が崩れる可能性が高い。本稿ではこれまでの先行研究の(一致していない)結果がこの問題によるものである可能性を指摘した。

この問題を回避するためには、Bartlett補正などの補正法の利用が可能である。

補正前の尤度比検定統計量の分布は、本研究でも示したようにサンプルサイズとモデルの大きさによって大きく偏ると考えられる。しかし補正後の尤度比検定統計量を用いれば、Golden Ruleを構成できる可能性がある。

さらに、平均構造があるモデルや多群モデル、離散観測変数が存在する場合の適合度指標についてのレビューを行った。また、モデル修正に関する議論と、尺度化の問題、段階的なモデル構築の重要性を述べた。

付記と謝辞

本研究は日本行動計量学会第7回春の合宿セミナーにて星野が講演した内容を大幅に再構成したものです。このような研究の機会を与えてくださいました実行委員会委員長である長岡技術科学大学の植野真臣先生はじめ実行委員の皆様、日本行動計量学会理事会の

皆様に感謝いたします。また、論文担当委員の岡本安晴先生 (日本女子大学) および匿名の査読者の方には、本論文の改稿に際して非常に参考となるご意見をいただきましたことを深謝いたします。

参 考 文 献

- Anderson, J.C. & Gerbing, D.W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, **49**, 155-173.
- Anderson, J.C. & Gerbing, D.W. (1988). Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach. *Psychological Bulletin*, **103**, 411-23.
- Anderson, J.C. & Gerbing, D.W. (1992). Assumptions and Comparative Strengths of the Two-Step Approach: Comment on Fornell and Yi. *Sociological Methods & Research*, **20**, 321-33.
- Bandalos, D.L. (1997). Assessing sources of error in structural equation models: The effects of sample size, reliability, and model misspecification. *Structural Equation Modeling*, **4**, 177-192.
- Bandalos, D.L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling*, **9**, 78-102.
- Bandalos, D.L. & Finney, S.J. (2001). Item parceling issues in structural equation modeling. In G.A. Marcoulides, & R.E. Schumacker (eds.), *New Developments and Techniques in Structural Equation Modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bekker, P.A., Merckens, A. & Wansbeek, T.J. (1994). *Identification, Equivalent Models, and Computer Algebra*. Boston: Academic Press.
- Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, **107**, 238-246.
- Bentler, P.M. (2000). Rites, wrongs, and gold in model testing. *Structural Equation Modeling*, **7**, 82-91.
- Bentler, P.M. & Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, **88**, 588-606.
- Bock, R.D. & Bargmann, R.E. (1966). Analysis of covariance structures. *Psychometrika*, **31**, 507-534.
- Bollen, K.A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley & Sons.
- Browne, M.W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*, **8**, 1-24.
- Browne, M.W. & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research*, **24**, 445-455.
- Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen, & J.S. Long (eds.), *Testing structural equation models*. Newbury Park, CA: Sage.
- Browne, M.W. & Mels, G. (1990). *RAMONA user's guide*. Unpublished report, Columbus, Ohio State University.
- Burt, R.S. (1976). Interpretational confounding of unobserved variables in structural equation models. *Sociological Methods and Research*, **5**, 3-52.
- Cattell, R.B. & Burdall, C.A. (1975). The radial parcel double factor design: A solution to the item-versus-parcel controversy. *Multivariate Behavioral Research*, **10**, 165-179.
- Cronbach, L.J. & Gleser, G.C. (1965). *Psychological tests and personnel decisions*. Urbana, IL: University of Illinois Press.
- Cudeck, R. & Henly, S.J. (1991). Model selection in covariance structures analysis and the "problem" of sample size: A clarification. *Psychological Bulletin*, **109**, 512-519.
- Curran, P.J., West, S.G. & Finch, J.F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, **1**, 16-29.
- Curran, P.J., Bollen, K.A., Paxton, P., Kirby, J. & Chen, F. (2002). The noncentral chi-square distribution in misspecified structural equation models: Finite sample results from a monte carlo simulation. *Multivariate Behavioral Research*, **37**, 1-36.
- Ding, L., Velicer, W.F. & Harlow, L.L. (1995). effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling*, **2**, 119-144.
- Fan, Z., Thompson, B. & Wang, L. (1999). Effects of Sample Size, Estimation Methods, and Model Specification on Structural Equation Modeling Fit Indexes. *Structural Equation Modeling*, **6**, 56-83.
- Fornell, C. & Yi, Y. (1992a). Assumptions of the Two-Step Approach to Latent Variable Modeling. *Sociological Methods & Research*, **20**, 291-320.
- Fornell, C. & Yi, Y. (1992b). Assumptions of the Two-Step Approach: Reply to Anderson and Gerbing. *Sociological Methods & Research*, **20**,

- 334-339.
- Gerbing, D.W. & Anderson, J.C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K.A. Bollen, & J.S. Long (eds.), *Testing structural equation models*. 40-65. Newbury Park, CA: Sage.
- 南風原朝和 (2002). モデル適合度の目標適合度. 行動計量学, **29**, 160-166.
- Hagtvet, K.A. & Nasser, F.M. (2004). How well do item parcels represent conceptually defined latent constructs? A two-facet approach. *Structural Equation Modeling*, **11**, 168-193.
- 萩生田伸子・繁樹算男 (1996). 順序付きカテゴリカルデータへの因子分析の適用に関するいくつかの注意点. 心理学研究, **67**, 1-8.
- Hayduk, L.A. & Glaser, D.N. (2000). Jiving the four-step, waltzing around factor analysis, and other serious fun. *Structural Equation Modeling*, **7**, 1-35.
- Hoelter, J.W. (1983). The analysis of covariance structures: Goodness-of-fit indices. *Sociological Methods and Research*, **11**, 325-344.
- 星野崇宏・橋本貴充・繁樹算男 (2001). ベイズ的アプローチによる因子平均の群間差・線形対比の有意性検定. 教育心理学研究, **49**, 31-40.
- 星野崇宏 (2003). 潜在変数への観測変数の回帰に関する問題と段階推定による解決. 心理学研究, **74**, 218-226.
- 星野崇宏・光永悠彦・繁樹算男・前川眞一 (2003). 潜在変数得点の推定値を用いた構造方程式の母数の推定について. 第31回行動計量学会大会発表抄録集.
- Hu, L. & Bentler, P.M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, **3**, 424-453.
- Hu, L. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, **6**, 1-55.
- Hutchinson, S.R. & Olmos, A. (1998). Behavior of description fit indexes in confirmatory factor analysis using ordered categorical data. *Structural Equation Modeling*, **5**, 344-364.
- Jöreskog, K.G. (1970). A general method for the analysis of covariance structure. *Biometrika*, **57**, 239-251.
- Jöreskog, K.G. & Sörbom (1986). *LISREL VI*, Mooresville, IN: Scientific Software Inc.
- Kenny, D.A. & McCoach, D.B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, **10**, 333-351.
- 狩野 裕 (2002a). 構造方程式モデリングは、因子分析、分散分析、パス解析のすべてにとって代わるのか? 行動計量学, **29**, 138-159.
- 狩野 裕 (2002b). 再討論: 誤差共分散の利用と特殊因子の役割. 行動計量学, **29**, 182-197.
- 狩野 裕・三浦麻子 (2003). グラフィカル多変量解析. 目で見える共分散構造分析. 現代数学社.
- Kaplan, D. (2000). *Structural Equation Modeling*. Thousand Oaks, CA: Sage.
- Kishton, J.M. & Widaman, K.F. (1994). Unidimensional versus domain representative parceling of questionnaire items: An empirical example. *Educational & Psychological Measurement*, **54**, 757-765.
- Kumar, A. & Dillon, W.R. (1987). The Interaction of Measurement and Structure in Simultaneous Equation Models with Unobservable Variables. *Journal of Marketing Research*, **24**, 98-105.
- La Du, T.J. & Tanaka, J.S. (1989). Influence of sample size, estimation method, and model specification on goodness-of-fit assessment in structural equation modeling. *Journal of Applied Psychology*, **74**, 625-635.
- Lee, S. & Hershberger, S. (1990). A simple rule for generating equivalent path models in covariance structure modeling. *Multivariate Behavioral Research*, **25**, 313-334.
- Little, T.D., Cunningham, W.A., Shahar, G. & Widaman, K.F. (2003). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, **9**, 151-173.
- MacCallum, R.C., Roznowski, M. & Necowitz, L.B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, **111**, 490-504.
- MacCallum, R.C., Wegener, D.T., Uchino, B.N. & Fabrigar, L.R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, **114**, 185-199.
- MacCallum, R.C., Widaman, K.F., Zhang, S. & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, **4**, 84-99.
- Maiti, S.S. & Mukherjee, B.N. (1990). A note on distributional properties of the Jöreskog-Sörbom fit indices. *Psychometrika*, **55**, 721-726.
- Markus, K.A. (2000). Conceptual shell games in the four-step debate. *Structural Equation Modeling*, **7**, 163-173.
- Marsh, H.W. (1998). The equal correlation baseline model: Comment and constructive alternatives. *Structural Equation Modeling*, **5**, 78-86.
- Marsh, H.W., Balla, J.R. & McDonald, R.P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: the effect of sample size. *Psychological Bulletin*, **103**, 391-410.

- Marsh, H.W., Hau, K.T., Balla, J.R. & Grayson, D. (1998). Is more ever too much? : The number of indicators per factors in confirmatory factor analysis. *Multivariate Behavioral Research*, **33**, 181-222.
- Marsh, H.W., Hau, K.T. & Wen, Z. (2004). In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, **11**, 320-341.
- Matsumoto, M. & Nishimura, T. (1998). Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computer Simulation*, **8**, 3-30.
- Mayekawa, S-I. (1994). Equivalent path models in linear structural equation models. *Behaviormetrika*, **21**, 79-96.
- McDonald, R.P. & Marsh, H.W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin*, **107**, 247-255.
- McDonald, R.P. & Ho. (2002). Principles and practice in reporting structural equation analysis. *Psychological Method*, **7**, 64-82.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, **58**, 525-543.
- 光永悠彦・星野崇宏・繁桝算男・前川真一 (2005). 因子スコアや潜在変数得点を用いた構造方程式モデルの母数推定の偏りの解決. 行動計量学, **32**, 21-33.
- Moustaki & Knott. (2000). Generalized latent trait models. *Psychometrika*, **65**, 391-411.
- Mulaik, S.A., James, L.R. Van Alstine, J., Bennett, N., Lind, S. & Stilwell, C.D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, **105**, 430-445.
- Mulaik, S.A. & Millsap, R.E. (2000). Doing the four-step right. *Structural Equation Modeling*, **7**, 36-73.
- 村上 隆 (2002). 心理測定の立場から見た因子分析と主成分分析. 日本行動計量学会第30回大会発表論文抄録集.
- Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, **43**, 551-560.
- Muthén, B. (1993). Goodness of fit with categorical and other nonnormal variables. In K.A. Bollen, & J.S. Long (eds.), *Testing structural equation models*. 205-234. Newbury Park, CA: Sage.
- Muthén, B. & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, **45**, 19-30.
- 岡田謙介・星野崇宏・繁桝算男 (2004). 構造方程式モデリングにおける適合度指標の改善について. 日本行動計量学会第32回大会発表抄録集.
- 岡田謙介・星野崇宏・繁桝算男 (投稿中). 構造方程式モデリングにおける Bartlett 補正を用いた尤度比検定統計量の改善.
- Ogasawara, H. (2001). Standard Errors of Fit Indices using Residuals in Structural Equation Modeling. *Psychometrika*, **66**, 421-436.
- 大津起夫 (1996). パスモデルにおける識別可能性の自動判定に向けて. 統計数理共同研究レポート, **100**, 統計数理研究所.
- Potthast, R. (1993). Confirmatory factor analysis of ordered categorical variables with large models. *British Journal of Mathematical and Statistical Psychology*, **46**, 273-286.
- Rigdon, E.E. (1998). The equal correlation baseline model for comparative fit assesment in structural equation modeling. *Structural Equation Modeling*, **5**, 63-77.
- Rubio, D.M. & Gillespie, D.F. (1995). Problems with error in structural equation models. *Structural Equation Modeling*, **2**, 367-378.
- Sammel, M.D., Ryan, L.M. & Legler, J.M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society, Series B*, **59**, 667-678.
- Serfling, R. (1980). *Approximation theorems of mathematical statistics*. Wiley and sons, New York.
- Shevlin, M. & Miles, J.N.V. (1998). Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Personality and Individual Differences*, **25**, 85-90.
- Steiger, J.H. (1989). *EzPATH: A supplementary module for SYSTAT and SYGRAPH*. Evanston, IL: SYSTAT.
- Steiger, J.H. (1998). A note on multiple sample extensions of the RMSEA fit index. *Structural Equation Modeling*, **5**, 173-180.
- Steiger, J.H. & Lind, J. (1980). Statistically based tests of the number of common factors. *Paper presented at the annual meeting of the Psychometric Society*. Iowa City, IA.
- Stelzl, I. (1986). Changing a causal hypothesis without changing the fit: Some rules for generating equivalent path models. *Multivariate Behavioral Research*, **21**, 309-331.
- Sugawara, H.M. & MacCallum, R.C. (1993). Effect of estimation method on incremental fit indexes for covariance structure models. *Applied Psychological Measurement*, **17**, 365-

- 377.
- Tanaka, J.S. (1993). Multifaceted conceptions of fit in structural equation models. In K.A. Bollen, & J.S. Long (eds.), *Testing structural equation models*. 10-39. Newbury Park, CA: Sage.
- Tanaka, J.S. & Huba, G.J. (1985). A general coefficient of determination for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, **42**, 233-239.
- 豊田秀樹 (1993). SAS による共分散構造分析. 東京大学出版会.
- 椿 広計 (2002). 「尺度化+回帰分析」の問題点に関する注意. *行動計量学*, **29**, 167-173.
- Vuong, Q.H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, **57**, 307-333.
- Wakaki, H., Eguchi, S. & Fujikoshi, Y. (1990). A class of tests for a general covariance structure. *Journal of Multivariate Analysis*, **32**, 313-325.
- Wicherts, J.M. & Dolan, C.V. (2004). A cautionary note on the use of information fit indexes in covariance structure modeling with means. *Structural Equation Modeling*, **11**, 45-50.
- Widaman, K.F. & Thompson, J.S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods*, **8**, 16-37.
- (2005 年 3 月 14 日受付, 2005 年 6 月 3 日最終修正)