

研究ノート Rのsemによる潜在成長モデルの解析

著者	清水 和秋, 紺田 広明
雑誌名	関西大学社会学部紀要
巻	42
号	1
ページ	129-146
発行年	2010-11-05
その他のタイトル	Latent Growth Modeling by sem in R
URL	http://hdl.handle.net/10112/4836

研究ノート

Rのsemによる潜在成長モデルの解析

清水 和秋・紺田 広明

Latent Growth Modeling by sem in R

Kazuaki SHIMIZU and Hiroaki KONDA

Abstract

The sem package in R was introduced in latent growth modeling analysis. The sample date in this analysis were selected from the published official batting average records of professional Japanese baseball players over a period of ten years. The result of latent growth modeling demonstrated that the quadratic form trajectory model fit the data well. The R command lists of this analysis were also presented.

Keywords: R, structure equation modeling, moment matrix, latent growth model

抄 録

Rのsemパッケージを、潜在成長モデルの分析のために紹介した。この分析の例データのために、10年にわたる日本のプロ野球選手の打撃記録を刊行されている公認記録から選んだ。潜在成長モデルの結果は、二次形式の軌跡モデルがよくデータに合うことを明らかにした。この分析のRコマンドリストも提示した。

キーワード: R, 構造方程式モデリング, モーメント行列, 潜在成長モデル

はじめに

心理統計解析の世界で大きな変化が起きている。理論的な面では、因子分析モデルを測定モデルとする構造方程式モデリング (Structural Equation Modeling: 以下SEM) が心理学研究方法の中心的な役割を担ってきているように、観測変数を対象としたある種の目的関数を最小化あるいは最大化するような重みづけの推定をベースとする方法論から、潜在変数を対象とした観測変数への重みづけや潜在変数間のパス係数を推定する方法論へと展開してきている (例えば, Cudeck & MacCallum (2007) や Mulaik (2009) など)。この流れは、解の推定の仕組みとしてのベイズ法 (例えば, Lee (2007) など)、推定値の信頼区間を得るための標準誤差の推定とその応用による推定値の評価 (例えば, Ogasawara (2005) など)、混合モデルによる下位集団のクラスタ化 (Hancock & Samuelsen, 2008) などの分野で急速な広がりを見せている。このような動きを取り入れる方向でSEMの代表的なソフトである LISREL, EQS, Mplus, Amosなどは、版の改訂を毎年のように重ねている。ソフト・ユーザーの立場に徹して、このような改訂に期待するということもできる。その一方で、心理データ解析の教育では、本質的な理解よりは、改訂を重ねるソフトの使い方の習得に時間を割かざるをえない状況となってきた。

統計ソフトの世界で起きている新しい動きは、R (Ihaka & Gentleman, 1996) によるものである。CRAN (Complete R Archive Network: <http://cran.r-project.org/> あるいは <http://www.r-project.org/>) には、R本体のパッケージの他に、ボランティアの貢献による膨大な量の数値解析・数理統計・統計解析やグラフィックスのパッケージが集積されている。フリーの統計ソフトと呼ばれるRの特徴は、インターネット上でCRANあるいはミラーサイトからフリーにこれらのパッケージをダウンロードすることができるということだけでなく、ソースコードも公開されているという点にある。Rでは、コンピュータ言語である Fortran や C あるいは C++ とのインターフェースを備えているので、過去にこれらの言語で作成されてきた知的遺産を組み込むことができる。そして、利用者が自分で解析用のプログラムを組むこともできる (リゲス, 2006)。

SASのIMLやSPSSのスク립トのように、利用者が自分の分析したい問題をプログラムとして書き下すこともできるが、その場合には、SASあるいはSPSSをプラットフォームとしてPC内に確保することが必要となる。Rでは同じように、入出力の命令形だけではなく、紺田 (2010) や清水 (2010) で使用例を示したように、統計解析で必要となる行列計算などもオブジェクトとして簡便な記述体系の下で使用することができる (間瀬,

2007)。Rはインタプリタの対話型を採用しているため、実行にはコンパイル型に比べれば時間はかかるが、記述したスクリプトを実行させながら結果を得るまでの過程を利用者がPCのコンソール画面上で体験することができるという特徴がある。

本稿では、Rで提供されているライブラリsem (Fox, 2006, 2009) を使って、潜在構造方程式モデリングの代表的な下位モデルである潜在成長モデル (Latent Growth Model: 以下LGM) の解析を行ってみることにする。

SEMは、よく知られているように、共分散構造モデルと平均構造モデルの2つの下位モデルからなる。因子分析にたとえると、前者では、因子と観測変数の関係を因子パターンと観測での独自性（誤差と特殊性）、そして、因子の分散と因子間の共分散（因子間の相関）を推定することになる。後者は、因子の平均を取り扱うモデルであり、構造平均 (structured mean) モデルとも呼ばれる (Sörbom, 1974)。

SEMでの解析では、パラメータ推定において、自由・拘束・固定の3種類の方法を組み合わせることによってモデルを構築し、標本から得られた平均・分散・共分散とこのモデルとの適合度を評価することになる。

LGMは、詳細は後ほど紹介するが、一見すると因子分析モデルのようである。異なる点は、因子と観測変数の関係に相当する因子パターンを自由推定ではなく固定パラメータとしたことである。そして、素点データを対象として、観測変数の切片をゼロ固定し、測定の開始時の得点の平均の位置と時間経過の中で変化の大きさ（線形であれば傾き）を固定パラメータとすることによって、これらの因子の平均に、変化の統計量を集約的にして表現する方法を確立した。

このようにLGMは、SEMの共分散構造と平均構造を組み合わせることによって実現したものである。ある意味では、解析対象のモデルの設定にSEMの特殊な方法を採用しているともいえる。ここでは、このモデル記述をRで行ってみることにする。そして、解析結果のモデルの適合度に関する指標にも検討を加えてみることにする。ここで取り上げる例は、結果の比較のために、清水 (2003, 2008) でAmosで解析したプロ野球の縦断的データである。

1. ライブラリsemが採用するSEMモデル

SEMのデータ解析の世界で使用されるソフトは、それぞれ共分散構造についての独自の理論モデルに基づいて構築されている。清水 (1994) で紹介したように内生変数と外生変数の2つの変数群を想定するLISRELモデルを、McArdle & McDonald (1984) は、観測

変数と潜在変数とを1つにまとめたより包括的なモデルであるRAM (Rreticular Action Model) へと発展させた。このモデルについては、清水 (1989) で紹介しているので、ここでは、最小限度の説明にとどめる。

RAMのモデル式は、観測変数と潜在変数とからなる確率変数ベクトル v と潜在外生変数と観測変数と潜在変数の攪乱変数からなるベクトル u 、そして、因子パターンやパス係数などの係数行列 A から構成される。すなわち、 $v = Av + u$ である。

この共分散構造のモデル式では、平均を取り扱うことはできない。McArdle & Epstein (1987) は、分散1の固定数を導入することにより、素点の回帰モデルとしてRAMを扱うことで、平均を推定することができることを明らかにしている。McArdle (2005) は、因子分析モデルで、この固定数を導入した平均の扱いについて紹介している。平均構造を推定するために、固定数1をモデルに導入する方式は、SEMの解析ソフトでは、Mxで導入されているだけである (Mxについては紺田 (2009) も参照)。

共分散構造モデル (RAM) と固定変数による平均構造モデルが内包されている sem というライブラリをRにFox (2009) が提供している。以下ではこの sem を取り上げることにする。

2. semでのモデル記述と解析

Rの sem は、他の統計解析と同じくコンソール上でコマンドを実行する形式であり、次の部から構成されている。

- 1) データ入力：素点入力から共分散行列の計算 (cov コマンド)、あるいは素点入力からモーメント行列の計算 (raw.moments コマンド)
- 2) モデル構築：specify.modelでの解析モデルの記述
- 3) パラメータ推定：sem コマンドを実行し、結果を推定結果に収容する。

共分散行列を対象とした場合

推定結果 <- sem (2で記述したモデル, 共分散行列, N = 標本数)

モーメント行列を対象とする場合には、

推定結果 <- sem (2で記述したモデル, モーメント行列,

N = 標本数, raw = TRUE, fixed.x = "固定変数の名前")

- 4) 結果の出力：summary (推定結果)

このコマンドを実行すると推定値 (標準誤差, 有意水準) そして、適合度指標が出力される。適合度指標としては、 χ^2 統計量に加えて、共分散構造を対象とした

場合にはGFI, AGFI, RMSEA, NFI, NNFI, CFI, SRMR, BICなどが出力される。

- 4) 修正指標の出力: mod.indices (推定結果)
- 5) 標準形式での出力: std.coef (推定結果)
- 6) 標準化残差の出力: standardized.residuals (推定結果)

SEMの解析では、以上のコマンドを目的に合わせ、組み合わせて使用することになる。

2.1 分析モデルの記述

パス図でのモデル構築がSEMの世界では一般的なものとなっている。現状のRのsemでは、specify.modelのセクションで、Rの一般的なコマンドとは別に、解析するモデルを記述しなければならない。ここではsemの解析モデルを記述するためのspecify.modelの関数を簡単に説明してみることにする。

まず、変数名は、Rでの入力した観測変数の名前がそのままモデルに組み込まれる。潜在変数については、モデルを特定するために新しく記述することになる。RAMでの記述体系では、観測変数の独自性分散と潜在変数の分散は、変数を定義することにより、自動的に付与されることになる。Amosなどでは、独自性を潜在変数として独立させて定義している。semでは、観測変数の自己分散の形式で独自性が取り扱われる。

次に、変数間の関係と推定するパラメータを記述するコマンドは、次の2つのカンマで区切られた記述項からなる。第1の記述項では、変数間の関係を定義する。パス関係は、 \rightarrow という記号で変数から変数への影響を定義する。共分散は \leftarrow という記号で2つの変数の共分散関係を記述することになる。第2の記述項は、これらの関係で推定するパラメータの名前を記述する欄である。同じスペルは同値拘束の対象となる。特定のスペルではなくNA (not available) と記述した場合には、次の第3の欄で記述する特定の値に固定することを指示することになる。第3の記述項では、固定パラメータの数値や自由推定の初期値を与えることができる。

```
解析モデル <- specify.model (モデルのファイル名)
```

あるいは、コンソールにて次のように記述する。

```
解析モデル <- specify.model ()
```

```
以下にモデルを記述
```

基本的な記述の原則を以下に整理してみる。

目的	第1の記述項	、	第2の記述項	、	第3の記述項
パスを推定	因子 f1	—>	観測変数 x1, パラメータ名,		初期値
パスを固定	因子 f1	—>	観測変数 x2, NA,		値
共分散を推定	因子 f1	<—>	因子 f2, パラメータ名,		初期値
独自分散を推定	変数 x1	<—>	変数 x1, パラメータ名,		初期値

以上を踏まえて、観測変数 (x1, x2, x3, x4) で2因子 (f1, f2) の specify.model を例示してみることにする。テキスト形式のファイルにモデルを記述し、これを読み込むこともできる。

#以降の記述はRではコメント扱いとなる。ここでは、簡単な解説を加えることにする。

```
cfa_model <- specify.model () # cfa_model という名前でのモデル
f1 —> x1, NA, 1 # x1のf1の因子パターンを1に固定
f1 —> x2, pa_x, 0.9 # pa_xを推定パラメータの名前とし、初期値を0.9
f2 —> x3, NA, 1 # x3のf2の因子パターンを1に固定
f2 —> x4, pa_x, 0.9 # x2のf1の因子パターンと同値
f1 <—> f2, cov_f1f2, 0.4 # f1とf2の共分散、初期値を0.4
f1 <—> f1, v_f1, 0.2 # 因子f1の分散、初期値を0.2
f2 <—> f2, v_f2, 0.2 # 因子f2の分散、初期値を0.2
x1 <—> x1, e_x1, 0.1 # 観測変数x1の独自性の分散、初期値を0.1
x2 <—> x2, e_x2, 0.1 # 観測変数x2の独自性の分散、初期値を0.1
x3 <—> x3, e_x3, 0.1 # 観測変数x3の独自性の分散、初期値を0.1
x4 <—> x4, e_x4, 0.1 # 観測変数x4の独自性の分散、初期値を0.1
```

ここで、変数 x2 と x4 の因子パターンに pa_x という名前を与えたのは、同値拘束を例示するためである。SEM では標本の分散・共分散行列を解推定の対象とするので、モデルを識別するためもある x1 と x2 の因子パターンを1に固定している。

因子や観測変数の平均値を推定するには、分散を1に固定した変数をモデルに挿入することになる。この変数を k とすると k から因子へのパスの推定値が因子の平均に相当する。そして、観測変数の平均も k からのパスによって、推定することが可能となる。この操作については、3以下でより詳しく説明する。

2.2 解の推定方法

semの解の推定には最尤法が提供されている。SEMの解析ソフトでは、最尤法での推定の前に、あらかじめ初期値を与え、ここから関数の最小値を求める繰り返し計算を始めるという仕組みを組み込んでいる。適切な初期値から始める理由は2つある。1つは、本来求める最小値ではない局小値に落ちることにならないようにするためである。もう1つは、計算時間を短くすることである。適切な初期値を与えないと収束せずに繰り返し長時間にわたるということも起きかねない。適切な初期の値をいくりにするかは、解析対象のモデルに関して蓄積されてきた結果が参考になるであろう。

3. 潜在成長モデル / 潜在曲線モデル

1つの変数を対象として、時間経過の中での起きる変化の様相を解析するために、複数の観測対象について、複数機会において測定したデータは、T技法データと呼ばれた。このデータ行列に通常の因子分析法を適用しようとしたのがCattellである。時間に関する因子を得ることができるという楽観的な見通しは、Cronbachからの強い批判を浴びることになった。批判のポイントは、清水（1999a, 2003）で紹介しているので詳細はここでは省略するが、探索的因子分析の手法の性質として、時間の因子を特定することができないというものである。確かに、因子数の決め方、解の推定方法、共通性の推定方法、因子軸の回転方法について、いずれの方法をデータ解析で使用するかによって結果が異なることになる。

因子パターンの値を固定するという発想から、この問題への突破口がMcArdle（1986, 1988）、McArdle & Epstein（1987）とMeredith & Tisak（1990）によって開かれた。因子分析では、潜在変数から観測変数へのパス係数すなわち因子パターンを、共通因子という枠内で推定する。J. J. McArdleは、上で述べた1の固定変数を導入するというアイデアを、素点の観測変数を対象として、潜在変数の平均の推定に導入した。

急速な応用への広がりを見せるこの分野では、この理論は潜在成長モデル（Latent Growth Model）あるいは潜在曲線モデル（Latent Curve Model）と呼ばれる。この2つの用語が併存している理由は、J. J. McArdleとW. Meredithの理論の発表の先陣争いという面もあるが、研究対象との関係にあると考えている。Meredith & Tisak（1990）は、以上の方法を、Tuckerlingとよばれることもある心理学での古典的な曲線を分解する方法論の流れの上で、潜在曲線モデル（Latent Curve Model）としている。これに対して、J. J. McArdleは、知能の生涯発達研究で収集されたデータを対象としてきたことから潜在成長モデル

(Latent Growth Model) としているのではないだろうか。

3.1 素点からの1次モーメント

伝統的な因子分析が解の推定の対象とするのは相関行列である。観測得点の形式としては標準得点とも考えることもできる。個々の変数の散らばりの大きさに違いがある場合には、標準化という操作によってその情報は失われることになる。共分散構造分析では、標本(分散・)共分散行列から母集団パラメータを推定し、モデルとの適合度を評価することになる。

変化の量は個々の得点の変動に現れる。分散では、観測時点ごとでの散らばりの違いをとらえることで、レベルの変化をとらえることはできない。集合統計量としての平均を扱うには、素点データが必要となる。データ解析の現場では、素点データに直接アクセスできない場合がある。モーメント行列から、共分散構造分析の解析ソフトで潜在変数である因子の平均を推定する方法を提案したのがSörbom(1974)であった。観測変数の分散・共分散行列を S 、平均ベクトルを m とすると標本のモーメント行列は次のように表すことができる。

$$M = \begin{bmatrix} S + mm' & m \\ m' & 1 \end{bmatrix} \quad (1)$$

なお、この行列の行と列の大きさは、観測変数の数に1を加えた数である。

3.2 素点の因子分析モデル

ここでは、McArdle(1988)、McArdle & Epstein(1987)そしてMeredith & Tisak(1990)を参考にしながら、因子の平均を推定するためのモデルの記述方法を検討してみることにする。

素点からの因子分析モデルを3つの観測変数、それぞれ y_{1i} 、 y_{2i} 、 y_{3i} について、因子を f_i 、各観測変数の独自性を u_{1i} 、 u_{2i} 、 u_{3i} として表してみることにする。

$$\begin{aligned} y_{1i} &= \tau_1 + \lambda_1 f_i + u_{1i} \\ y_{2i} &= \tau_2 + \lambda_2 f_i + u_{2i} \\ y_{3i} &= \tau_3 + \lambda_3 f_i + u_{3i} \end{aligned} \quad (2)$$

ここでは、観測変数の平均をそれぞれ τ_1 、 τ_2 、 τ_3 とし、因子パターンを λ_1 、 λ_2 、 λ_3 としている。そして、個人の得点には添え字 i を与えている。

相関行列からの探索的因子分析は、観測変数を標準得点とするので、平均をパラメータ推定の対象とすることはない。共分散構造でも同様である。観測変数の平均をモデルに組み込むのは集団比較や縦断的变化を検討する場合である。

観測変数の平均 τ_1, τ_2, τ_3 をゼロに固定したモデルへと上の式を変形してみる。

$$\begin{aligned} y_{1i} &= 0 + \lambda_1 \tilde{f}_i + u_{1i} \\ y_{2i} &= 0 + \lambda_2 \tilde{f}_i + u_{2i} \\ y_{3i} &= 0 + \lambda_3 \tilde{f}_i + u_{3i} \end{aligned} \quad (3)$$

さらに、因子パターンをすべて1に固定してみることにする。

$$\begin{aligned} y_{1i} &= 0 + 1\tilde{f}_i + u_{1i} \\ y_{2i} &= 0 + 1\tilde{f}_i + u_{2i} \\ y_{3i} &= 0 + 1\tilde{f}_i + u_{3i} \end{aligned} \quad (4)$$

この結果、因子得点を素点の形式で表し、同じ重みで観測変数を表現するができたことになる。

3.3 因子の平均の推定：R の sem でのモデルの記述

素点データが入手できない状況で \tilde{f}_i の平均を推定する方法については、Meredith & Tisak (1990) が(1)式のモーメント行列を共分散構造分析での共分散行列に相当するものとした最尤法によるパラメータの推定方法から検討を加えている。そして、適合度評価についても言及している。

(1)式の観測変数の平均を取り扱うために、McArdle & Epstein (1987) に従って、固定変数 k をここでは導入してみることにする。固定変数 k から変数へのパスにおいて、平均の推定値を得ることができるので、(4)式の観測変数の平均をゼロとする記述は、 k から観測変数へのパスを0に固定すればよいことになる。因子への k からのパスの値を推定することで、全変数に等しく1と置いた重みの因子の平均を推定することができることになる。

ここでは、ある変数を3回の観測機会測定したとして、これらの変数を (y_1, y_2, y_3) と表し、平均の因子を Level として、固定変数を k と表記する。なお、この k の分散は1に固定する。

```

# モーメント行列を求める方法① (sem の raw.moments 関数を使用)
LGM_mom <- raw.moments (cbind (LGM_data,1))
rownames (LGM_mom) <- list ("y1", "y2", "y3", "k")
colnames (LGM_mom) <- list ("y1", "y2", "y3", "k")
# モデルの記述
raw_score <- specify.model ()
Level -> y1, NA, 1 # 因子パターンを1に固定
Level -> y2, NA, 1 # 因子パターンを1に固定
Level -> y3, NA, 1 # 因子パターンを1に固定
Level <-> Level, v_L, 5.0 # Levelの分散, 初期値を5.0
y1 <-> y1, d1, 10.0 # 観測変数y1の独自性の分散, 初期値を10.0
y2 <-> y2, d2, 10.0 # 観測変数y2の独自性の分散, 初期値を10.0
y3 <-> y3, d3, 10.0 # 観測変数y3の独自性の分散, 初期値を10.0
k -> Level, m_L, 25.0 # Levelの平均, 初期値を25.0
k -> y1, NA, 0 # 観測変数y1の平均を0に固定
k -> y2, NA, 0 # 観測変数y2の平均を0に固定
k -> y3, NA, 0 # 観測変数y3の平均を0に固定
k <-> k, NA, 1 # 固定変数kの分散を1に固定
# パラメータの推定と結果の出力
LGM_result <- sem (LGM_model, LGM_mom, N = 121, raw = T, fixed.x
= "k")
summary (LGM_result)

```

3.4 傾き因子の記述

測定機会間の間隔がほぼ同じであれば、そして、線形で変化をモデル化できるなら、傾きの因子を単調関数として記述することができる。上の平均の因子をここではレベル因子として、これに傾きの因子を加えることによって、線形の変化を捉えることができる。この場合、SEMでは、Levelの因子 (\tilde{f}_{li}) に加えて、Slopeの因子 (\tilde{f}_{si}) を時間関数として定義し、それぞれの因子の平均を推定することで、発達過程での変化の関数を手にすることができる。観測機会の間隔が等しいならば、(4)式を次のように書き換えることができる。

$$\begin{aligned}
 y_{1i} &= 0 + 1\tilde{f}_{Li} + 0\tilde{f}_{Si} + u_{1i} \\
 y_{2i} &= 0 + 1\tilde{f}_{Li} + 1\tilde{f}_{Si} + u_{2i} \\
 y_{3i} &= 0 + 1\tilde{f}_{Li} + 2\tilde{f}_{Si} + u_{3i}
 \end{aligned}
 \tag{5}$$

解推定のための識別性を確保できる範囲内において、0 から 2 とした観測機会を推定パラメータとすることも可能であることも指摘しておきたい。

現象が非線形の場合には、適切な関数を定義することが求められる。清水 (1999a, b) でも一部を紹介したように、発達に関する心理学的研究は、ゴンベルツ関数など多数の関数についての知見を蓄積している。ここでは、2 次関数を例として取り上げてみることにする。二次の項 (Quadratic) の因子 (\tilde{f}_{Qi}) を(5)式に加えてみることにする。

$$\begin{aligned}
 y_{1i} &= 0 + 1\tilde{f}_{Li} + 0\tilde{f}_{Si} + 0^2\tilde{f}_{Qi} + u_{1i} \\
 y_{2i} &= 0 + 1\tilde{f}_{Li} + 1\tilde{f}_{Si} + 1^2\tilde{f}_{Qi} + u_{2i} \\
 y_{3i} &= 0 + 1\tilde{f}_{Li} + 2\tilde{f}_{Si} + 2^2\tilde{f}_{Qi} + u_{3i}
 \end{aligned}
 \tag{6}$$

4. 2 次関数モデルの推定

ここでは清水 (2003) で報告したプロ野球選手121名の10年間の打撃成績を R の sem での対象としてみることにする。なお、Amos によるこのデータの解析結果については、清水 (2008) でも紹介している。

以下は R でのメイン・プログラムである。

```

library (sem)      # ライブラリの設定
LGM_data <- read.csv ("kanzen_10year_obgeneki_genpon100.csv", header = T)
# モーメント行列を求める方法① (sem の raw.moments 関数を使用)
LGM_mom <- raw.moments (cbind (LGM_data.1)) # 固定変数を 1 で指定
rownames (LGM_mom) <- list ("y1", "y2", "y3", "y4", "y5", "y6", "y7", "y8",
                           "y9", "y10", "k")
colnames (LGM_mom) <- list ("y1", "y2", "y3", "y4", "y5", "y6", "y7", "y8",
                           "y9", "y10", "k")

```

```
# モデルの指定と推定
LGM_model <- specify.model("LGM_model_for_baseball.txt")
LGM_result <- sem(LGM_model, LGM_mom, N = 121, raw = T, fixed.x
                  = "k")
summary(LGM_result)
# 修正指標
mod.indices(LGM_result)
```

モデルを記述したファイル (LGM_model_for_baseball.txt) は次である。

```
Level -> y1, NA, 1 # レベルの因子：全観測変数を1に固定
Level -> y2, NA, 1
一部省略
Level -> y10, NA, 1
Slope -> y1, NA, 0 # 傾きの因子：初年を0に固定
Slope -> y2, NA, 1 # : 1年を1の増分とした単調増加
一部省略
Slope -> y10, NA, 9 # : 10年目を9に固定
Quadratic -> y1, NA, 0 # 2次の項：増加分を2乗して固定
Quadratic -> y2, NA, 1 # : 1の2乗
一部省略
Quadratic -> y10, NA, 81 # : 9の2乗
Level <-> Level, varL, 0.5 # レベル因子の分散を推定
Slope <-> Slope, varS, 0.1 # 傾き因子の分散を推定
Quadratic <-> Quadratic, varQ, 0.1 # 2次項の因子の分散を推定
Level <-> Slope, covLS, 0.1 # レベルと傾きの共分散を推定
Level <-> Quadratic, covLQ, 0.1 # レベルと2次項の共分散を推定
Slope <-> Quadratic, covSQ, 0.1 # 傾きと2次項の共分散を推定
y1 <-> y1, var_y1, 0.1 # 観測変数の独自性の推定
y2 <-> y2, var_y2, 0.1
一部省略
```

```

y10<-> y10, var_y10, 0.1
k  -> Level, meanL, 1.0          # レベル因子の平均を推定
k  -> Slope, meanS, 0.5         # 傾き因子の平均を推定
k  -> Quadratic, meanQ, 0.5    # 2次項の平均を推定
k  -> y1, NA, 0                # 観測変数の平均推定値を0に固定
k  -> y2, NA, 0
一部省略
k  -> y10, NA, 0
k  <-> k, NA, 1                # 固定変数kの分散を1に固定

```

注：前後の記述から明らかな記述箇所は、ここでは、省略している。

Rでの解析結果の前に、(1)式のモーメント行列をこの10変数に関して、示しておくことにする。ここでは、Rのprintコマンドを使用している。

```

> print ( LGM_mom )
Raw Moments
y1      y2      y3      y4      y5      y6      y7
y1  581.58058 609.52570 629.98653 637.65413 644.39612 647.43793 640.23132
y2  609.52570 657.76339 669.92504 679.11248 687.07554 690.18967 681.34769
y3  629.98653 669.92504 701.13215 703.78570 711.34231 716.18074 706.52744
y4  637.65413 679.11248 703.78570 721.98116 722.17116 725.87355 717.44107
y5  644.39612 687.07554 711.34231 722.17116 736.97488 734.67488 726.14992
y6  647.43793 690.18967 716.18074 725.87355 734.67488 747.66223 731.87835
y7  640.23132 681.34769 706.52744 717.44107 726.14992 731.87835 731.97603
y8  641.95987 681.91592 710.89617 720.52913 729.06134 736.60272 727.39001
y9  639.24711 680.54835 706.27099 715.70959 723.91091 730.01926 721.87017
y10 628.12562 670.36025 692.94364 703.05736 709.50058 717.28298 710.00190
k   23.78760 25.30331 26.26860 26.62645 26.94050 27.08512 26.79339
      y8      y9      y10      k
y1  641.95987 639.24711 628.12562 23.78760

```

```

y2  681.91592 680.54835 670.36025 25.30331
y3  710.89617 706.27099 692.94364 26.26860
y4  720.52913 715.70959 703.05736 26.62645
y5  729.06134 723.91091 709.50058 26.94050
y6  736.60272 730.01926 717.28298 27.08512
y7  727.39001 721.87017 710.00190 26.79339
y8  739.66204 726.94164 711.01175 26.91167
y9  726.94164 726.87934 707.84066 26.72727
y10 711.01175 707.84066 705.63983 26.22975
k   26.91167  26.72727  26.22975  1.00000

```

N = 121

次に、このモーメント行列を対象とした解析結果を示すことにする。Rのsemでは、共分散行列を対象とした場合とは異なり、適合度の指標としては、 χ^2 統計量とモデル比較の指標であるBIC (Bayesian Information Criterion) だけである。各変数の推定結果では、推定値、標準誤差、推定値の検定統計量 (Z値とP)、そして、推定値の定義が出力される。

```

> LGM_result <- sem (LGM_model, LGM_mom, N = 121, raw = T, fixed.
                      x = "k")
> summary (LGM_result)
Model fit to raw moment matrix.
Model Chisquare = 50.859 Df = 46 Pr (> Chisq) = 0.28817
BIC = -169.75
Normalized Residuals
  Min. 1st Qu. Median Mean 3rd Qu. Max.
-0.22200 -0.05250 0.00493 0.00134 0.06250 0.20500
Parameter Estimates
      Estimate Std Error z value Pr (> |z|)
covLS  -2.1758021 0.6988149 -3.1136 1.8484e-03 Slope <--> Level
covLQ   0.1766489 0.0656851  2.6893 7.1596e-03 Quadratic <--> Level

```

```

covSQ   -0.0843091 0.0312621 -2.6968 7.0000e-03 Quadratic <--> Slope
varL     9.2967230 1.9245189  4.8307 1.3607e-06 Level <--> Level
varS     1.0056592 0.3214117  3.1289 1.7547e-03 Slope <--> Slope
varQ     0.0077573 0.0032212  2.4082 1.6030e-02 Quadratic <--> Quadratic
var_y1   6.9280735 1.6532733  4.1905 2.7832e-05 y1 <--> y1
var_y2  10.4406756 1.5792173  6.6113 3.8097e-11 y2 <--> y2
var_y3   6.9536369 1.0321657  6.7369 1.6176e-11 y3 <--> y3
var_y4   8.6199964 1.2344365  6.9829 2.8906e-12 y4 <--> y4
var_y5   7.2077368 1.0848314  6.6441 3.0506e-11 y5 <--> y5
var_y6   7.3012136 1.0949831  6.6679 2.5953e-11 y6 <--> y6
var_y7   8.6438423 1.2657284  6.8291 8.5423e-12 y7 <--> y7
var_y8   8.7185126 1.3009970  6.7014 2.0642e-11 y8 <--> y8
var_y9   5.7035640 0.9843569  5.7942 6.8646e-09 y9 <--> y9
var_y10 10.0380151 1.7091282  5.8732 4.2752e-09 y10 <--> y10
meanL    24.0926756 0.3448185 69.8706 0.0000e+00 Level <--- k
meanS    1.1059577 0.1394401  7.9314 2.2204e-15 Slope <--- k
meanQ   -0.0983744 0.0138825 -7.0862 1.3780e-12 Quadratic <--- k
Iterations = 351

```

ここで得ることができた推定値は、清水（2003）で報告したAmosの結果とほぼ同じであり、違いは小数桁で第3位にだけみられた。 χ^2 統計量では、Amosの解では50.426（df = 46, P = 0.303）で、小数第1位でも違いがみられた。そして、mod.indicesコマンドによる修正指標の出力も、Amosのものとは大きな違いはみられなかった。

以上の結果から、Rのsemは、LGMというSEMでも特殊なモデルを対象とした最尤法によるパラメータ推定でも、市販のソフトと比べても遜色のない、十分に信頼できる結果を提供してくれるといえる。

5. おわりに

Rでは、プルダウンメニュー方式のRCommanderが提供されるようになってきているが、基本的にはコンソールからコマンドを一行ずつ打ち込むか、テキストファイルからコマンドを読み込む方式である。Rはインターネット接続の利用環境にあれば無料で自由に

ダウンロードできるが、PCの入門段階のユーザーインターフェースとしては、敷居が高いと言わざるを得ない。

SEMの世界では、観測変数を四角で潜在変数を円あるいは楕円で描き、変数間の関係をパス（一方向矢印）あるいは共分散（両方向矢印）で線を引くグラフィカルなインターフェース方式は、モデルの探索を試行錯誤的に展開するためには、非常に優れたツールとして高く評価されている。その一方で、よりよい適合度を求めるために修正指標の数値に振り回され、SEMの基本的な考え方からは外れたモデルを追求することになるというマイナス面もみられる。

SEMの世界を、Amosのようなソフトが、数式に馴染みの薄い応用の立場からの研究者にも広げることに貢献したことは確かである。その一方で、SEMの本質から外れた解の報告やパラメータ推定値を標準形式へと変換した値としてしか活用していない報告に接することもある。

ここでみてきたようにsemのspecify.modelは、モデルの数式やその統計的意味を記号の体系に置き換えたものである。このため、モデルを記述するためには、分散そして共分散さらには平均の概念を押さえることを求められる。ここで紹介してきたようなLGMのような応用段階へと研究を進めるには、分散・共分散そしてモーメントの概念は必須となっている。その意味で、SEMの活用を教育するという観点からみると、Rは魅力ある統計解析ツールであるといえる。

semに関してはFox(2006)が総括しているように、グラフィックスに対応していないだけでなく、多集団同時分析をこのままでは行うことはできない。標本数を同じとする複数の集団であれば、超行列にこれらを独立させて配置することにより、集団間でのパラメータ拘束も合わせて、解の推定が可能である。この同時分析の方法に関しては、清水(1997)がCOSANを使って報告している。

Fox(2006)は、さらに、パラメータの推定方法として、最尤法に加えて、漸近的分布自由理論の導入にも言及している。心理学では、多変量正規分布から乖離した分布を示すことが多い。最尤法が、このような乖離に対しては、頑健性を示すことは知られているが、今回のLGMの推定では、収束には351の繰り返しを要しているようである。Amosなどと比べても回数が多い。semでは、モデル構築時に初期値を設定しなければならない。パラメータの推定についての改良を望みたいことを指摘しておきたい。

引用文献

- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: Wiley.
- Cudeck, R., & MacCallum, R. C. (Eds.). (2007). *Factor analysis at 100: Historical development and future directions*. Mahwah, NJ: Lawrence Erlbaum.
- Fox, J. (2006). Structural Equation Modeling with the sem Package in R. *Structural Equation Modeling*, 13, 465-486.
- Fox, J. (2009). *Package 'sem': Structural equation models*. <<http://www.r-project.org>>, <<http://socserv.socsci.mcmaster.ca/~jfox/>> (2010年1月19日).
- Hancock, G. R., & Samuelsen, K. M. (Eds.). (2008). *Latent variable mixture models*. Charlotte, NC: Information Age Publishing.
- Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5, 299-314.
- 紺田広明 (2009). Mxを使用した構造方程式モデリング——Mx GUIによるモデル構築と解の推定—— 関西大学大学院心理学研究科心理学叢誌, 2, 1-12.
- 紺田広明 (2010). Amosによる動的因子分析——Rによるブロック・トリープツ行列の構成とDAFSモデルの推定—— 関西大学大学院心理学研究科心理学叢誌, 3, 9-22.
- Lee, S-Y. (2007). *Structural equation modeling: A Bayesian approach*. England: John Wiley.
- リゲス, U. (石田基広 訳) (2006). Rの基礎とプログラミング技法 シュプリンガー・ジャパン.
- 間瀬 茂 (2007). Rプログラミングマニュアル 数理工学社.
- McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics*, 16, 163-200.
- McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselrode & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed.) (pp. 561-614). New York, NY: Plenum Press.
- McArdle, J. J. (2005). The development of the RAM rules for latent variable structural equation modeling. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 225-273). Mahwah, NJ: Lawrence Erlbaum.
- McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, 58, 110-133.
- McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the Reticular Action Model. *British Journal of Mathematical and Statistical Psychology*, 37, 234-251.
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107-122.
- Mulaik, S. A. (2009). *Linear causal modeling with structural equations*. Boca Raton, FL: Chapman & Hall/CRC.
- 清水和秋 (1989). 検証的因子分析, LISRELそしてRAMの概要. 関西大学社会学部紀要, 20(2), 61-86.
- 清水和秋 (1994). JöreskogとSörbomによるコンピュータ・プログラムと構造方程式モデル. 関西大学社会学部紀要, 25(3), 1-41.
- 清水和秋 (1997). SASのCALISプロシジャによる多集団同時因子分析 関西大学情報処理センターフォーラム, 11, 26-38.
- 清水和秋 (1999a). 潜在成長モデルによる進路成熟の解析——不完全コーホート・データへの適用—— 関西大学社会学部紀要, 30(3), 1-47.

- 清水和秋 (1999b). キャリア発達 of 構造的解析モデルに関する比較研究 進路指導研究, 19(2), 1-12.
- 清水和秋 (2003). プロ野球選手の熟達の軌跡——潜在成長モデルの打撃成績への適用—— 悠峰職業科学研究所紀要, 11, 63-72.
- 清水和秋 (2008). 混合モデルによる熟達パターンの探索——プロ野球選手の熟達の軌跡を例として—— 関西大学社会学部紀要, 40(1), 17-37.
- 清水和秋 (2010). 項目因子分析で構成した尺度の因子パターン, 共通性, 信頼性そして因子的真实性 関西大学心理学研究, 1, 9-24.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, 27, 229-239.
- Ogasawara, H. (2005). Bias reduction of estimated standard error in factor analysis. *Behaviormetrika*, 32(1), 9-28.

—2010. 6. 30 受稿—