

## 共分散構造分析における数値解析上の課題と信頼性に関わる実証的分析

Reliability and Stability on the Numerical Analysis in Structural Equation Modeling

北海道大学大学院工学院 ○学生員 目黒達也 (Tatsuya Meguro)  
北海道大学大学院工学院 学生員 Borith Long  
北海道大学大学院工学研究院 正員 中辻隆 (Takashi Nakatsuji)

## 1. はじめに

土木計画学研究に関わる市民の意識構造を分析する際に、共分散構造分析 (Structural Equation Modeling、以下 SEM) が広く用いられるようになってきている。SEM は、パス図の自動描写などの高い利便性から、その分析には Amos<sup>1)</sup>をはじめとする専用ソフトウェアが広く用いられており、その計算過程はブラックボックスとなっている。汎用 SEM ソフトを用いて意識構造の分析を行った際に、解が存在しないこともしばしば経験することである。また、SEM では、解の存在や一意性についても、初期値により解の有無が起ること<sup>2)</sup>、残差分散の値が負値になること<sup>3)</sup>が指摘されてきた。さらに、著者らによる経験でも、以下のような現象が観測されている。

①残差分散が負値になったり、制約条件によってパス係数の推定値が変動すること。

②初期値の正負を変えると、パス係数の推定値の符号が反転すること。

本研究では、SEM における数値解析上の課題を分析するため、3 種類の SEM プログラムと 2 種類のサンプルデータを用い、様々な条件下における数値解析を行うことで信頼性の高い分析を行うための提言を行う。さらに、バンコクで実施した飲酒運転に関するアンケート調査による実測データを用い、その実証的分析を示す。

## 2. 分析方法

## 2.1 使用プログラム

本研究では、分析の信頼性を高めるために 3 種類の SEM プログラムを用いて分析を行った。以下にその 3 種類を示す。

- ① Amos18.0<sup>1)</sup> (P1)
- ② 自作した SEM プログラム (P2、自作ソルバー)
- ③ GA<sup>4)</sup> による最適化プログラム (P3)

なお P2 は、Excel のアドイン機能の 1 つである「ソルバー」を使って未知数を求める数値計算を実行する。また、P1 に対し、未知数に与える初期値に任意の値を与える機能および推定値の範囲に制約条件を設ける機能を付加している。それぞれのプログラムで条件を適切に設定した上で分析を行った場合、ほぼ同値の解がでることを確認している。

## 2.2 使用データ

本論文で使用するサンプルデータは、参考文献<sup>3)</sup>より引用した①と②の 2 種類である。

① 参考文献<sup>3)</sup> 図 6.1 の例題

「自分や家族の自分の趣味に対するイメージ」を表すモデル (3 章 3.1 図-1、3.2 図-2 で使用)。

② 参考文献<sup>3)</sup> 図 6.3 の例題

「収入や学歴が社会的地位に及ぼす影響の程度、社会的地位が人脈や知名度に及ぼす影響の程度」を表すモデル (3 章 3.3 図-3、図-4、図-5 で使用)。

## 2.3 分析手法

3 種類すべての分析プログラムにおいて、最尤法と最小 2 乗法により数値計算を実行している。なお、本論文での解は、すべて最尤法によるものである。

## 3. 共分散構造分析における数値解析上の問題点

1 章に示したように、SEM における問題点の検証分析を行った結果、大きく 2 つの数値解析上の問題点が明らかになった。以下にその 2 つの問題点を示す。さらに、解の多様性や一意性についてより深く分析するため、SEM に最適化計算である遺伝的アルゴリズム (GA) を適用した。

## 3.1 残差分散の制約条件による問題

1 つ目の問題は、残差分散が負値になったり、制約条件によってパス係数の推定値が変動することである。2 章 2.2、①のデータを用いて、P1 (Amos18.0、上段太字) により分析したパス係数および残差分散の値と P2 (自作ソルバー、下段斜字) により残差分散が負値にならない制約を加えて分析した値の比較を図-1 に示す。P1 の e2 および e6 に注目すると、残差分散の値が負値になっていることがわかる。これら負値は残差分散の値としてふさわしくなく、不適解であるといえる。そこで、P2 を用いてこの残差分散の値を 0~1 の間で導くよう、制約条件を加えて分析を行った結果が下段である。制約条件の有無により、残差分散だけでなくパス係数の値にも差が生じていることがわかる。また、各適合度についても両者ともほぼ同じ値がでていることを確認している。このような場合、制約を加えた解のほうがより信頼できる解であると考えられる。なお、ここでソルバー実行前に与えた初期値は、パス係数、残差分散とも 0.5 であり、パス係数は標準化解、P1 の残差分散は半標準化解、P2 の残差分散は非標準化解である。P2 を用いて制約をかけずに分析した場合、P1 と同値が得られている。

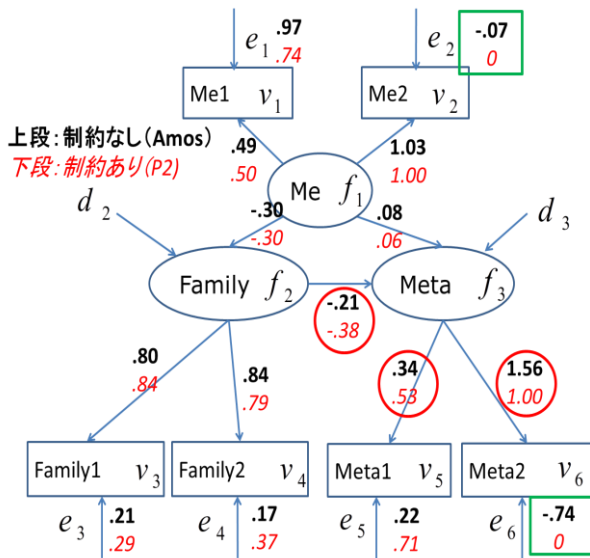


図-1 制約条件の有無による分析結果の比較 (P1,P2)

### 3.2 初期値による問題

2 つ目の問題は、初期値の正負を変えると、パス係数の推定値の符号が反転することである。2 章 2.2、①のデータを使用し、P2 を用いて様々な値を初期値に与えて分析を行った。図-2 に、パス係数、残差分散とも 0.5 を与えた初期値 A (上段太字) およびパス係数は-1~1 間の乱数を与え、残差分散には 0~1 間の乱数を与えた初期値 B (下段斜字) の 2 パターンにおけるパス係数の推定値 (標準化解) を示す。これらの値は、解として導かれる可能性のある範囲の値であり、初期値として妥当な値であるといえる。表-1 に、それぞれに与えた初期値を示す。図-2 に示すように、初期値 B は初期値 A に比べて絶対値は一致しているが、一部符号が反転した結果となっている。なお、この時の目的関数の値は同値であった。他のモデルでも同様の検証を行った結果、負値の初期値が入る位置に応じて、一部に符号の反転した解が生じることがわかった。

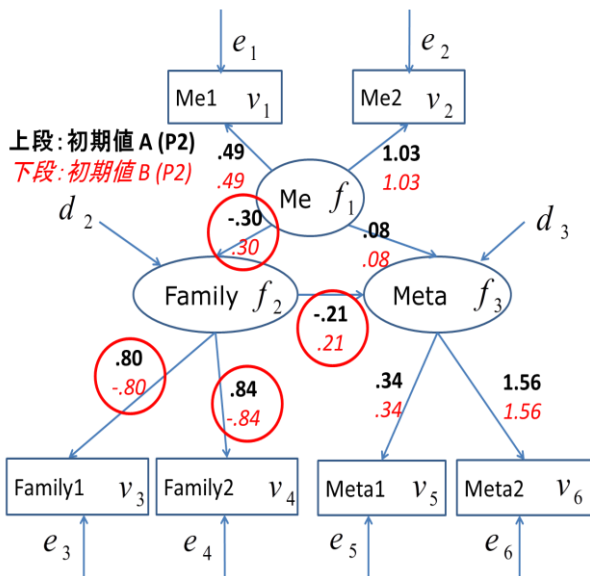


図-2 初期値Aと初期値Bの分析結果の比較 (P2)

表-1 初期値Aと初期値Bの値

|      | パス係数      | 残差分散    |
|------|-----------|---------|
| 初期値A | 0.5       | 0.5     |
| 初期値B | - 1~1間の乱数 | 0~1間の乱数 |

### 3.3 遺伝的アルゴリズムによる最適化

解の多様性や一意性についてより深く考察するため、遺伝的アルゴリズムを適用してSEMを行った。単純交叉数20、単一算術交叉数20、全体算術交叉数を20とし、一様突然変異数20、境界突然変異数20、非一様突然変異数は40である。その他のパラメータは、GAのデフォルト値を使用している。

SEMにGAを適用したプログラムP3を使用し、いくつかのモデルを用いて、様々な設定条件の下に解の多様性や一意性について更なる分析を行った。どのモデルにも共通して、繰り返し回数に応じてP2でのソルバー実行後の目的関数の値よりも観測変数の数に限りなく近づくが、推定値の制約条件をかけない、もしくはあまりに範囲が広ければ、激しく解が変動することがわかった。

また、2章2.1、②のデータ (図-3のモデル参照) を用いて、表-2に示す制約条件および初期値の下でパス係数と残差分散を求めた。このとき、母数を100、繰り返し回数を15000回とした。図-4は、図-3に示すパス係数 $\alpha_{d11}$ と目的関数の値が、繰り返し回数によってどのように変化するかを表すグラフである。横軸は繰り返し計算回数を対数目盛でとり、左縦軸はパス係数の推定値、右縦軸は目的関数の値を示している。なお、最尤法により計算を実行した。図-4を見ると、繰り返し回数が100回程度で目的関数は収束に近づくが、収束に近づいてからもなお、めまぐるしく解が変動していることがわかる。これは、他のパス係数や残差分散についても同様の傾向が見られた。

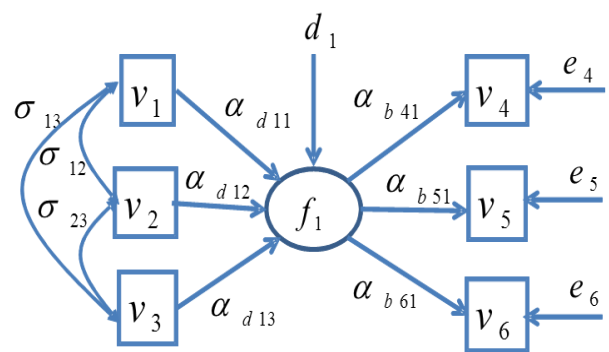


図-3 データ②の分析モデル

表-2 制約条件および初期値

|      | パス係数  | 残差分散 (対角) | 残差分散 (非対角) |
|------|-------|-----------|------------|
| 制約条件 | - 1~1 | 0~1       | - 1~1      |
| 初期値  | 0.5   | 0.5       | 0.5        |

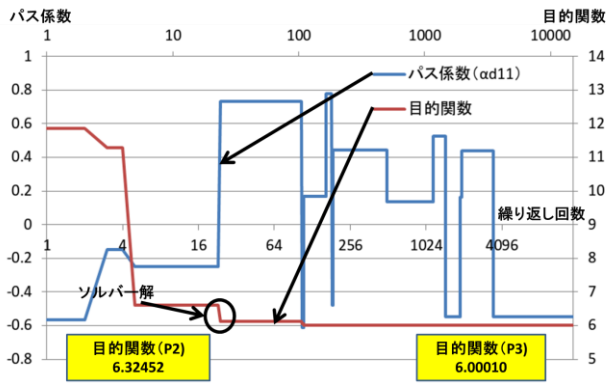


図-4 繰り返し回数に対するパス係数 $\alpha_{d11}$ と目的関数の値 (P3)

次に、P2（上段太字）、およびP3（下段斜字）で分析したパス係数の値の比較を図-5に示す。P2の初期値は、パス係数と残差分散の非対角要素が0.3、残差分散の対角要素が0.5であり、推定値の制約条件は設けていない。P3の初期条件は、表-2に示したものと同条件である。P2のソルバー実行後の目的関数値は6.32452、P3の計算終了時の目的関数値は6.00010であり、収束に非常に近い値である。図-5より、両者を比較すると、その結果も大きく異なっている。これらから、①ソルバー解がまだローカルミニマムに留まっていること、②パス係数や残差分散がわずかな目的関数値の変化で大きく変動すること、を表しているといえる。

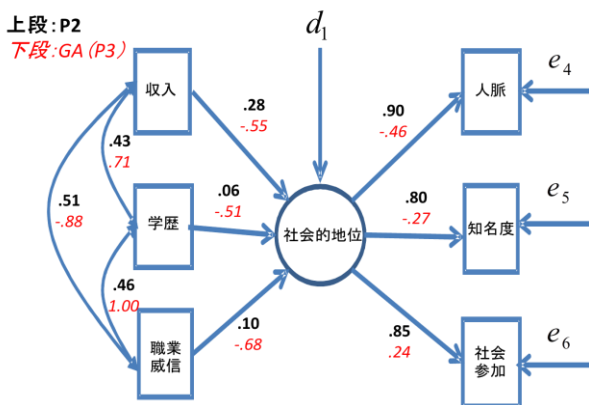


図-5 P2の解とGA解 (P3) の比較

### 3.4 数値解析的結論

3種類のSEMプログラムを用いて、SEMにおける数値解析上の課題を分析し、さらに解の多様性について言及した。初期値や制約条件、使用するソフトによって解が変動したりすることがSEMの弱点であり、そのような条件下で導いた解がどの程度信頼できる結果であるかはわからないのが現状であると考えられる。信頼性の高い分析を行うため、数値解析的研究からいえることを列挙する。

#### ①初期値

正の解を仮定する場合は正の初期値、負を仮定する場合は負の初期値を与える。正負の混在する乱数はいない。

#### ②制約条件

残差分散の対角成分には負値とならない制約を与える。

### ③GA最適化

分析の適合度が不安な場合や、仮定と大きく異なる解が出力された場合などはGAプログラムを利用する。推定値が正負に関わらず一定の値に収束し、多様な解が存在しないことを確認する。多様な解が存在する場合は、モデルの変更を行う。

## 4. 実証分析

前章まででは、SEMにおける数値解析的課題について言及し、さらに信頼性の高い分析を行うための提言を示した。本章では、そこで得られた数値解析的成果を実データを用いた分析に活かし、どのような結果を得られるか、その実証分析を行う。

### 4.1 アンケート調査

発展途上国における死亡事故の大きな一因として、飲酒運転が問題視されている。そこで、タイのバンコクを例に、飲酒運転に至る要因を定量的に分析するため、SEMでの分析を前提として飲酒運転挙動のアンケート調査を実施した。

#### (1)調査概要

バンコク市内の大学、企業、商業施設、空港において、直接配布、直接回収にて実施した。有効回答数N=96である。

#### (2)調査内容

学歴や年齢など個人属性に加え、表-3に示す質問に対し、1～5で評価する形式で実施した。なお、集計に際しては、各質問に飲酒運転リスクの高低を仮定し、数字が大きくなるほど飲酒運転リスクが高くなるよう設定した。例えば、飲酒量は多いほど飲酒運転リスクは高くなると仮定している。

表-3 アンケート調査質問票

|                                   |
|-----------------------------------|
| 1.あなたはお酒をどのくらいの頻度で飲みますか？          |
| 2.飲酒するときには通常どのくらいの量を飲みますか？        |
| 3.一度に6単位以上の飲酒をする頻度はどれくらいですか？      |
| 4.少しの飲酒なら運転しても大丈夫だと思う。            |
| 5.近距離であれば飲酒運転してもよいと思う。            |
| 6.飲酒運転をするととっさの判断力が鈍くなると思う。        |
| 7.飲酒運転をするととっさの反応が鈍くなると思う。         |
| 8.新聞やテレビで飲酒運転による事故の報道を見たことがありますか？ |
| 9.飲酒運転の危険さについての教育を受けたことがありますか？    |
| 10.飲酒運転に対する法律を厳しくすべきだと思う。         |
| 11.飲酒運転の取り締まりを増やすべきだと思う。          |

## 4.2 分析概要

P1を用い、最尤法にて分析を実行したところ、残差分散が負値となった。モデルの変更を避けるため、P2を用いて残差分散が負値をとらない制約を加えた上で、再分析を行った。本分析のモデルでは、その解は基本的に正の値をとると仮定している。そのため、初期値にはパス係数、残差分散ともすべて0.5の正值を与えた。なお、最終的な目的関数の値は13.907である。

## 4.3 分析結果

図-6 に、本分析結果のパス図を示す。また、表-4 に示すのが本分析結果の適合度である。

まず、図-6 のパス図に注目する。本モデルは、飲酒運転を先頭に、飲酒レベル、安全意識、技能意識、関心、法律の6つの潜在変数を仮定した。飲酒運転を起こす影響因子として、安全意識に次いで飲酒量や頻度などの飲酒レベルが高い結果となった。本結果より、学歴や年齢といった属性的要因よりも、安全意識や飲酒レベルといった意識的な部分が大きく影響していることが明らかになった。

次に、表-4 の適合度を参照する。アンケート調査のサンプル数  $N=96$  であることを考慮すると、1 以下になるのが望ましいといえる  $\chi^2/df$  が 1.4354 である。また、0.9 以上で当てはまりがよいとされる GFI が 0.884、0.05 以下で良好、0.1 以上で NG とされる RMSEA が 0.068 と、適合度には不安が残る結果となった。

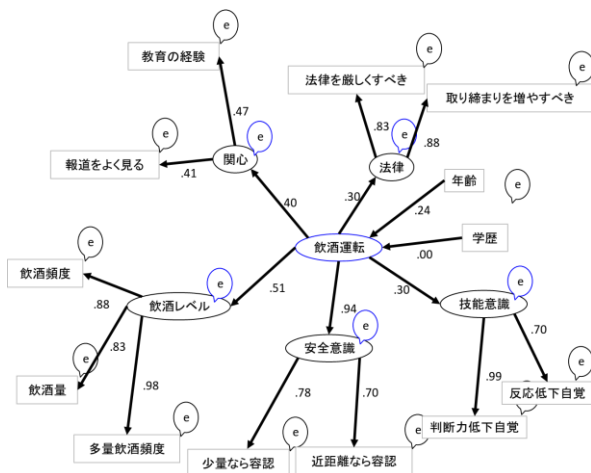


図-6 アンケートの分析結果

表-4 アンケート分析の適合度

| 適合度                |        |      |       |       |      |
|--------------------|--------|------|-------|-------|------|
| χ <sup>2</sup>     | 86.126 | GFI  | .884  | RMSEA | .068 |
| df                 | 60     | AGFI | .824  | NFI   | .834 |
| p                  | 1.5%   | SRMR | .087  | CFI   | .941 |
| χ <sup>2</sup> /df | 1.4354 | AIC  | -33.9 |       |      |

## 4.4 GA 最適化

前項より、適合度には不安が残る結果となったため、P3を用いて、GAによる非線形解析を行い、解の安定性を確認した。GAは多点的に最適解を探索するため、グローバルミニマムに入りやすいことが知られている。こ

の解析で、各推定値がソルバー解に向かって収束していくことが確認できた。以上より、P2で求められた解はある程度信頼性の高い結果であると考えられる。

## 4.5 考察

本分析前のモデル作成の段階では、学歴や関心の指標が高くなるのではと仮説を立てていた。しかしながら、本分析結果より、高度飲酒者への安全意識の改革が、飲酒運転根絶への近道であると考えられることがわかった。具体的には、飲酒運転が少量の飲酒や近距離でも危険だと自覚させる教育を行う必要があると考えられる。特にタイ人はアルコールに強い人が多く、少量なら安全と考える人の割合も高い。現に、現地では少量の飲酒ではあたりまえのように飲酒運転を行う人々が多々みられた。本分析結果にも、その実情が反映されているといえる。

## 5. 結論

本研究では、3種類のSEMプログラムを用いてSEMにおける数値解析上の課題を分析し、信頼性向上のための提言を行った。さらにアンケート調査から得られた実データを用いて実証分析を行い、タイの飲酒運転対策について考察した。初期値や制約条件によって解が変動したり、有無があることがSEMの弱点であり、そのような条件下で導いた解がどの程度信頼できる結果であるかはわからないのが現状であると考えられる。信頼性の高い解を得るには、与える初期値と残差分散の制約条件に加え、多様な解が存在する可能性がある結果の解釈をすることが必要であるといえる。

## 6. あとがき

今後の展望として、非線形SEMの開発を考えている。非線形の構造を説明することができないことがSEMのシステム上の限界である。人間の判断に影響を与える心理的構造は非線形であると考え、潜在変数にシグモイド関数を組み込み、人間の判断に影響を及ぼす心理的境界を明らかにしたいと考えている。人間の心理構造は、個人により感度が異なり、感度の高い人と低い人ではその感じ方は大きく異なるといえる。本研究では、シグモイド関数によりその感度の差を再現し、分析能力の高度化を図りたいと考えている。

## 参考文献

- 1) James L. Arbuckle : IBM SPSS Amos19 ユーザーズガイド、IBM、2011.
- 2) 豊田秀樹：共分散構造分析－構造方程式モデリング－[疑問編]、朝倉書店、2003.
- 3) 豊田秀樹：共分散構造分析－構造方程式モデリング－[入門編]、朝倉書店、1998.
- 4) Zbigniew Michalewicz : Genetic Algorithms+Data Structures = Evolution Programs、Springer、1996.
- 5) 小島隆矢：Excelで学ぶ共分散構造分析とグラフィカルモデリング、オーム社、2003.