SEMのソフトウェアMxによるグラフィカルモデリングの実行

豊田 秀樹*, 福中 公輔**, 川端 一光**

The Graphical Modeling by Software of SEM

Hideki TOYODA*, Kosuke FUKUNAKA**, Ikko KAWAHASHI**

In this research, it is shown that the graphical modeling is able to be run by the software of structural equation modeling. Its software is Mx.

The data for analysis were three correlation matrices used in the Japanese Society for Quality Control (1999), and using Mx, these data was analyzed by graphical modeling. In case 1, reproduction of a result was possible, but furthermore it was possible to carry out covariance selection from perspective of fit index. In case 2 and case 3, reproduction of a result was almost completely possible. In this result, the graphical modeling was able to be analyzed by framework for structural equation modeling, and it can be included for lower model of SEM.

A program of Mx and its manual was shown as an appendix in the last of this article.

Key words: graphical modeling, Mx, structural equation modeling

キーワード: グラフィカルモデリング, Mx, 構造方程式モデリング

1. 問題・目的

構造方程式モデリング (SEM, structural equation modeling, 豊田, 1998) は, 数理的に一般的な形式で表現されているために, そのソフトウェアを使用すると, 本来の目的ではなかった筈の統計モデルの推定・計算が可能となることがある. たとえば潜在成長モデルは, そもそも SEM とは異なった経緯で提案された統計モデルであるが, Browne (1993) によって理論的に SEM の枠組みに位置づけられ, その後 SEM のツールを利用して様々にモデルが発展している. また自己回帰や移動平均等の時系列モデルも Molenaar (1985) によって SEM の枠組みに位置づけられ, 因子分析との融合が成功して発展を続けている. 他にも, 一般化可能性係数や2母数正規累積の項目反応モデルや, 2段抽出モデルや, 意思決定法における AHP や,

一対比較モデルなどは, 後になって SEM の枠組みでも表現できることが判明したモデルである (豊田, 2000, 2006 など参照).

近年, グラフィカルモデリング (graphical modeling, 以下 GM と略記. 宮川, 1997, 宮川, 2004) の共分散選択 (Dempster, 1972) を利用した無向独立グラフが, 変数間の仮説が少ない場合に有効な知見を与える手法として注目されている. GM の目的は, 変数間の関係において事前には明確でない場合に, データに基づいて探索的にモデル化を行い, その構造の妥当性を検証することである. 一般的には, 複数の変数間の条件付独立の関係を, 変数を頂点とし, 関係を辺として, グラフ理論におけるグラフで表現する. 条件付独立であるか否かの判断は, 2変数間の偏相関係数が0に十分近いかどうかで判断する. 日本品質管理学会テクノメトリックス研究会は, これら一連の計算を行うために CGGM という GM に特化したソフトウェアを独自に開発した. 日本品質管理学会テクノメトリックス研究会 (1999) では, このソフトを用いて GM を実施している.

本資料論文の目的は, (1) SEM における共分散構造の枠組みで GM の数理的アイデアが表現できること

*早稲田大学

(Waseda University)

連絡先: 〒162-8644 新宿区戸山 1-24-1

**早稲田大学大学院

(Graduate School of Letters, Arts and Sciences,
Waseda University)

連絡先: 〒162-8644 新宿区戸山 1-24-1

を理論的に示すこと, (2)SEM による分析結果の数値が, 日本品質管理学会テクノメトリックス研究会(1999)の数値例に一致することを確認すること, (3)当該ソフトウェアのマニュアルには載っていない実行方法を解説し, 実践の便宜に供すること, の3点である。

2. 方法

母数ベクトル θ の関数で表現された共分散行列の構造 $\Sigma(\theta)$ (狩野・三浦, 2002) を

$$\Sigma(\theta) = G^{-1} \quad (1)$$

と表現する。ここで G は観測変数の共分散行列の逆行列である。したがって θ の要素は共分散行列の逆行列 G の要素を重複なく (たとえば対角成分を含む下三角要素) を取り出したものである。

GM の分析過程の一部ではその他の変数の影響を取り除いたときの変数 i と変数 j の偏相関が 0 であるか否かを考察する。偏相関 $\rho_{ij \cdot rest}$ は以下のように表される。

$$\rho_{ij \cdot rest} = \frac{-\sigma^{ij}}{\sqrt{\sigma^{ii}\sigma^{jj}}} \quad (2)$$

ここで σ^{ij} は考察の対象となっている相関行列の逆行列の第 i 行第 j 列の要素である。変数 i と変数 j の偏共分散の符号を反転し, それを第 i 行第 j 列に配した行列を P とすると,

$$P = \text{diag}(G)^{-1/2} G \text{diag}(G)^{-1/2} \quad (3)$$

と表現される。この関係を利用して (1) 式を書き直すと

$$\Sigma(\theta) = (\text{diag}(G)^{1/2} P \text{diag}(G)^{1/2})^{-1} \quad (4)$$

となる。偏相関が 0 であるか否かの考察は, 偏共分散が 0 であるか否かの考察と同等である。また 0 であるか否かの考察に符号は関係しない。一般に G の対角成分は非ゼロであり, それが成り立つ場合には, G の非対角成分 $g_{ij}(i \neq j)$ に関して

$$g_{ij} = 0 \Leftrightarrow \rho_{ij \cdot rest} = 0 \quad (5)$$

である。ゆえに, その他の変数の影響を取り除いたときの変数 i と変数 j の偏相関が 0 であるか否かの考察は, g_{ij} が 0 であるか否かの考察に帰着する。構造方程式モデリングの多くのソフトではこのような固定母数

を有するモデルの解を求めることは難しいが, Mx (Neale & Cardon, 1992)¹ のように変数間共分散行列の逆行列に固定母数の制約をいれることのできる構造方程式モデリングのソフトウェアを用いれば GM を実行できる。

またモデルの評価には, 本研究では AIC と RMSEA を用いている。これは, AIC は多くの尤度モデルとの比較が可能な適合度指標の中で代表的なものであり, RMSEA は SEM においてよく利用され, 最も信頼のおける指標の 1 つだからである。これらの客観的指標を利用することで, GM と他の多くの統計モデルとの直接的な比較が可能となる。

3. 結果

3.1. 適用例 1

データ概要 宮川 (1997) の第 1 章には 1995 年に行われた陸上競技 10 種目における日本人選手 50 人のデータが掲載されていた。日本品質管理学会テクノメトリックス研究会 (1999) の第 4 章では, この「陸上十種競技データ」の中から, “100 m 走 (秒)”, “走り幅跳び (m)”, “400 m 走 (秒)”, “110 m ハードル (秒)”, “1,500 m 走 (秒)” の 5 種目を選択し, GM が実施された。このとき計算された 5 種目間の標本相関行列を表 1 に示す。ただし, 各変数は “100 m: 100 m 走”, “Jump: 走り幅跳び”, “400 m: 400 m 走”, “110 H: 110 m ハードル”, “1,500 m: 1,500 m 走” を意味している。本適用例ではこの 5 種目の相関データを用いて分析を行った。

共分散選択 まず始めに, 表 1 の相関係数行列を基にして, Mx を用いて相関行列の逆行列を求め, そこから偏相関行列を算出した。このフルモデルにおける偏相関行列を表 2 に示す。このとき, 5 行 4 列目の要素である 1,500 m 走と 110 m ハードルの偏相関係数の絶対値が最小となった。これは言い換えると, 1,500 m 走と 110 m ハードルの偏共分散の値が 0 であるという帰無仮説の z 値が最小であるということを意味する。そこで次に, この要素が 0 となるように逆行列を推定し, その値を用いて偏相関係数を求めた。なお, このときのモデルをモデル 1 とした。

同様にして, 偏共分散の値が 0 であるという帰無仮説の z 値が最小である要素を, 順に 0 となるように推

¹ 詳しくは <http://www.vcu.edu/mx/> を参照。

表 1. 陸上十種競技から選択された 5 種目の相関データ

100 m	Jump	400 m	110 H	1,500 m
1.000				
-.221	1.000			
.467	-.119	1.000		
.353	-.247	.465	1.000	
-.178	-.052	.205	.031	1.000

表 2. 陸上競技 5 種目のフルモデルにおける偏相関行列

100 m	Jump	400 m	110 H	1,500 m
—				
-.180	—			
.434	.083	—		
.118	-.195	.359	—	
-.320	-.100	.323	.040	—

表 3. 適用例 1 のモデルの変遷に伴う適合度の推移

	FULL	モデル 1	モデル 2	モデル 3	モデル 4	モデル 5	モデル 6
0 制約		χ^2	χ^2	χ^2	χ^2	χ^2	χ^2
χ^2 値	0	.077	.388	.656	1.735	3.234	6.318
df	0	1	2	3	4	5	6
p 値	1	.78	.82	.88	.78	.66	.39
AIC	0	-1.923	-3.612	-5.344	-6.265	-6.766	-5.682
RMSEA	0	0	0	0	0	0	.033

表 4. 陸上競技 5 種目の最終モデルにおける偏相関行列（左下：Mx による結果，右上：CGGM による結果）

100 m	Jump	400 m	110 H	1,500 m
—	-.13	.42	.14	-.31
0	—	0	-.17	0
.479	0	—	.34	.31
0	-.220	.392	—	0
-.316	0	.303	0	—

定を繰り返し，モデルを更新していった。このときのモデルの変遷を適合度の推移とともに表 3 にモデル 6 まで示す。表 3 には χ^2 値とそのモデルにおける自由度および p 値，そして AIC と RMSEA を掲載した。また，0 制約には当該モデルで 0 に制約して推定する逆行列の要素を示した。

表 3 より χ^2 値はすべてのモデルにおいて 5% 水準で棄却されておらず，モデルのデータへの適合は比較的良好であるといえる。また，AIC および RMSEA の観点から，モデル 5 までは適合が良くなっているが，モデル 6 で適合は悪くなっていた。したがって，モデル 5 を本適用例における最終的なモデルとして採用する

図 1. 適用例 1 における最終モデルの偏相関係数付き無向独立グラフ

ことにした。このモデル 5 における最終的な偏相関行列を表 4 の左下三角行列に，CGGM による結果（日本品質管理学科テクノメトリックス研究会，1999）を表 4 の右上三角行列に示す。また，偏相関係数付き無向独立グラフを図 1 に示す。図 1 より $c_1 = \{100 \text{ m 走}, 1,500 \text{ m 走}, 400 \text{ m 走}\}$ というクリークが 1 つ形成されていることがわかった。

3.2. 適用例 2

データ概要 日本品質管理学会テクノメトリックス研究会（1999）の 5 章 1 節では数学系 5 教科の試験の成績データを用いて，GM が実施された。数学系 5 教科の変数は“力学 M”，“ベクトル解析 V”，“代数学 AI”，“解析学 An”，“統計学 S”の 5 つであった。この

表 5. 数学系 5 教科の相関データ

M	V	Al	An	S
1.000				
.553	1.000			
.547	.610	1.000		
.409	.485	.711	1.000	
.389	.437	.665	.607	1.000

表 6. 数学系 5 教科のフルモデルにおける偏相関行列

M	V	Al	An	S
—				
.328	—			
.231	.281	—		
-.002	.078	.432	—	
.024	.021	.357	.252	—

表 7. 適用例 2 のモデルの変遷に伴う適合度の推移

	FULL	モデル 1	モデル 2	モデル 3	モデル 4	モデル 5
0 制約		$\chi^2_{.41}$	$\chi^2_{.52}$	$\chi^2_{.51}$	$\chi^2_{.42}$	$\chi^2_{.31}$
χ^2 値	0	0	.040	.133	.873	10.123
df	0	1	2	3	4	5
p 値	1	.98	.98	.99	.93	.07
AIC	0	-2.000	-3.960	-5.867	-7.127	.123
RMSEA	0	0	0	0	0	.109

表 8. 数学系 5 教科の最終モデルにおける偏相関行列 (左下: M_x による結果, 右上: CGGM による結果)

M	V	Al	An	S
—	.33	.23	0	0
.331	—	.33	0	0
.236	.327	—	.45	.36
0	0	.452	—	.26
0	0	.364	.256	—

とき使用された学生 88 名の成績より算出された標本相関係数行列を表 5 に示す。本適用例ではこの数学系 5 教科の相関データを用いて分析を行った。

共分散選択 適用例 1 と同様にして、表 5 の相関行列からフルモデルの偏相関行列を計算した。この偏相関行列を表 6 に示す。表 6 より 4 行 1 列目の要素である力学と解析学の偏相関係数の絶対値が最小となっていることが分かる。そこで、この要素に対応する相関行列の逆行列の要素が 0 となるように推定し、これをモデル 1 とした。

その後、適用例 1 と同様に、偏共分散の値が 0 であるという帰無仮説の z 値が最小である要素を、順に 0 となるようにモデルの推定を繰り返し、適合度が悪く

なるまでモデルを更新した。このときの適合度の推移を表 7 に示す。表 7 より χ^2 値はすべてのモデルにおいて 5% 水準で棄却されておらず、モデルのデータへの適合は比較的良好であるといえる。また、AIC および RMSEA の観点から、モデル 4 までは適合が良くなっているが、モデル 5 で適合は悪くなっていた。したがって、モデル 4 を本適用例における最終的なモデルとして採用することにした。このモデル 4 における最終的な偏相関行列を表 8 の左下三角行列に、CGGM の結果 (日本品質管理学会テクノメトリックス研究会, 1999) を右上三角行列に示す。

3.3. 適用例 3

データ概要 日本品質管理学会テクノメトリックス研究会 (1999) の 5 章 1 節では、適用例 2 の数学系 5 教科の成績データ以外に、乗用車用内装樹脂の見栄えに関する“立体感”、“柔軟感”、“織細感”、“光沢感”の 4 変数のもとで測定された SD データに関して、GM を適用している。この分析で用いられた標本相関行列を表 9 に示す。またオブザベーション数は 70 (被験者 7 人 $\times$ 試料の個数 10) であった。本適用例ではこの標本相関行列を用いて分析を行った。

共分散選択 表 9 の相関行列からフルモデルの偏相関行列を計算したところ表 10 の結果が得られた。前 2

表 9. SD データの標本相関行列

立体感	柔軟感	繊細感	光沢感
1.000			
.720	1.000		
.487	.631	1.000	
-.257	-.389	-.559	1.000

表 10. SD データのフルモデルにおける偏相関行列

立体感	柔軟感	繊細感	光沢感
—			
.611	—		
.085	.375	—	
.070	-.087	-.442	—

表 11. 適用例 3 のモデルの変遷に伴う適合度の推移

	FULL	モデル 1	モデル 2	モデル 3	モデル 4
0 制約		χ^2	χ^2	χ^2	χ^2
χ^2 値	0	.338	.593	.813	26.665
df	0	1	2	3	4
p 値	1	.561	.743	.846	0
AIC	0	-1.662	-3.407	-5.187	18.665
RMSEA	0	0	0	0	.287

表 12. SD データの最終モデルにおける偏相関行列 (左下: Mx による結果, 右上: CGGM による結果)

柔軟感 A	立体感 B	光沢感 C	繊細感 D
—	.63	0	0
.627	—	.43	0
0	.436	—	-.46
0	0	-.463	—

例と同様に偏相関行列の要素に対応する相関行列の逆行列の要素が 0 となるように推定し、これをモデル 1 とした。

その後、適用例 1 と同様に、偏共分散の値が 0 であるという帰無仮説の z 値が最小である要素を、順に 0 となるようにモデルの推定を繰り返し、適合度が悪くなるまでモデルを更新した。このときの適合度の推移を表 11 に示す。表 11 よりモデル 3 の下で最良の解が得られていることが分かる。このモデル 3 における偏相関行列を表 12 の左下三角行列に、CGGM の結果(日本品質管理学会テクノメトリックス研究会, 1999) を右上三角行列に示す。

4. 考 察

適用例 1 において、日本品質管理学会テクノメトリックス研究会 (1999) では、逸脱度における p 値の観点からモデル 4 で共分散選択を終了し、最終的な解としてモデル 3 の偏相関行列を採用していた。このとき、Mx におけるモデル 4 までの各モデルの偏相関行列の変遷は、小数点第 2 位の精度で、日本品質管理学会テクノメトリックス研究会 (1999) での CGGM を使用した結果と一致していた。

しかし、CGGM を利用した研究において、共分散選択を終了する基準は多少恣意的であり、著者自らが共分散選択の続行可能性を示唆している。本研究では、AIC および RMSEA など各種適合度の観点から、モデル 4 以降も共分散選択を行うことが可能であることがわかった。その結果、本研究では最終的にはモデル 6 を棄却し、明確な客観的基準を根拠としてモデル 5 を採用することができた。

モデル 5 に関して図 1 より、クリーク c_1 は距離走、110 m ハードルと走り幅跳びは跳躍に関する種目であり、この 2 種類の種目の中心に位置するのは中距離走である 400 m 走であることが分かる。しかし、同じ距離走でも 100 m 走と 1,500 m 走に、同じ跳躍種目でも 110 m ハードルと走り幅跳びに負の関係が表れているのは興味深い。これは 400 m 走の能力が同程度の選手

を集めてきた場合、短距離走が得意な選手は長距離走を苦手とし、ハードルが得意な選手は走り幅跳びを苦手であるということの意味する。その逆もまた同様である。短距離走は瞬発力が、長距離走は持久力が重要な種目でありこのトレードオフの関係にも納得がいく。また、ハードルはいかに短く跳躍するかが、走り幅跳びはいかに長く跳躍するかが重要になる種目であり、瞬発力と持久力とのある種の類似したトレードオフ関係が存在すると考えられる。

一方、適用例2と適用例3に関しては、本研究における最終的なモデルがCGGMの結果により最終的に採用されたモデルと小数点第2位の精度でほぼ完全に一致していた。このように、GMは構造方程式モデリングのソフトウェアで完全に解析可能であるといえる。

5. 結 論

GMにおいて「モデル比較」というと、観測変数間にパスが引けるか引けないかという、共分散選択に基づくモデルの比較を意味している。しかし統計学には、因子分析モデルやパス解析モデルを始めとした、様々な統計モデルが存在するわけであり、これらの各種モデルとGMによるモデルとを比較することはきわめて重要であると考えられる。つまり、あるデータにGMを適用するのなら、そのモデルが他の統計モデルよりも最適であると判断できる根拠が必要であるということだ。

しかしこれまでの多くの研究では、これら各種モデルとGMとを客観的指標に基づいて、正しく比較されることはなかった。これは、そもそもGMという分析手法が多くの変数モデルに基づく分析手法と根本的に異なるものであると考えられていたため、仕方のないことでもある。

しかし本研究により、GMの無向独立グラフも構造方程式モデリングの枠組みで解くことができ、SEMの下位モデルに含めることが可能となった。これはすなわち、当該データに対してGMを適用するのが良いのか、あるいは因子分析モデルやパス解析モデルなどを適用するのが良いのかを、適合度という客観的基準を用いて、同じレベルで比較可能になったということである。

また、GMのもう一つの特徴として、これまでの研究ではほとんどの場合において観測変数のみに対して

しか適用されてこなかったというものがある。しかし、本研究によって構造方程式モデリングと融合されたことにより、潜在変数に関してもより柔軟にGMを適用できる可能性が開けた。このように観測変数を主な対象としていたGMを、それと同レベルの水準で潜在変数への適用の可能性を広げたことは、応用可能性という面で非常に有用だと考えられる。

参 考 文 献

- Browne, M.W. (1993). Structured latent curve models. In Cuadras, C.M. & Rao, C.R. (Eds.) *Multivariate Analysis : Future Directions 2* (pp. 171-197). Elsevier Science Publishers.
- Dempster, A.P. (1972). Covariance selection. *Biometrics*, 28, 157-175.
- 狩野 裕・三浦麻子 (2002). グラフィカル多変量解析—AMOS, EQS, CALISによる目で見える共分散構造分析—. 現代数学社.
- 宮川雅巳 (1997). グラフィカルモデリング. 朝倉書店.
- 宮川雅巳 (2004). 統計的因果推論—回帰分析の新しい枠組み. 朝倉書店.
- Molenaar, P.C.M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, 50, 181-202.
- Neale, M.C. & Cardon, L.R. (1992). *Methodology for Genetic Studies of Twins and Families*. Dordrecht, NL: Kluwer Academic Publishers.
- 日本品質管理学会テクノメトリックス研究会 (1999). グラフィカルモデリングの実際. 日科技連.
- 豊田秀樹 (1998). 共分散構造分析 [入門編]—構造方程式モデリング—. 朝倉書店.
- 豊田秀樹 (2000). 共分散構造分析 [応用編]—構造方程式モデリング—. 朝倉書店.
- 豊田秀樹 (2006). 購買心理を読み解く統計学. 東京図書.

付 録

プログラム

適用例1の陸上競技5種目で使用したMxのプログラムを以下に掲載する。なお、下記プログラムをそのまま走らせるとフルモデルに関する結果が得られる。

```

01 #NGroups 3
02 G1:
03 Calculation
04 Begin Matrices;
05 S Symm 5 5
06 End Matrices;
07 Specify S
08 01
09 02 03
10 04 05 06
11 07 08 09 10
12 11 12 13 14 15
13 End
14 G2:
15 Data NInput_vars= 5 Nobservations=50
16 Labels x1 x2 x3 x4 x5
17 KMatrix
18 1.000
19 -0.221 1.000
20 0.467 -0.119 1.000
21 0.353 -0.247 0.465 1.000
22 -0.178 -0.052 0.205 0.031 1.000
23 Matrices= Group 1
24 Covariances S~;
25 va 1 S~ 1 1
26 va 1 S~ 2 2
27 va 1 S~ 3 3
28 va 1 S~ 4 4
29 va 1 S~ 5 5
30 !va 0 S 5 4
31 !va 0 S 3 2
32 !va 0 S 5 2
33 !va 0 S 2 1
34 !va 0 S 4 1
35 !va 0 S 4 2
36 End
37 G3:
38 Data Calculation
39 Begin Matrices= Group 1;
40 I Iden 5 5
41 End Matrices;
42 Begin Algebra;
43 P= -(\sqrt(I.S)~*S*\sqrt(I.S)~);
44 End Algebra;
45 End

```

仕様

プログラムは全部で3つのグループから構成されており、その数は1行目で指定する。第1のグループは2行目から始まり、ここでは主に推定する行列の宣言

を行う。第2グループは14行目からであり、実際に母数推定を行うパートである。ここでは使用する相関行列や制約に関する記述も行う。第3のグループは37行目からであり、偏相関係数行列の計算を行う。

まず第1グループに関して、4行目から6行目で行列名の宣言をする。本例では行列名を“S”とし、また5行5列の対称行列であるので“Symm 5 5”とする。GM では0に制約していくのは相関行列の逆行列であるので、“S”を相関行列の逆行列とみなして後の計算式を構成していかなければならないことに注意する。7行目から12行目では推定する行列の要素を指定する。

次に第2グループでは最初に、使用するデータを記述する。15行目では相関行列の変数とオブザベーションの数を、16行目ではラベル名を、17行目から22行目では実際の相関係数行列を指定する。23行目でグループ1の行列を呼び出し、24行目で実際の推定式を記述する。ここで“~”は逆行列を表す記号であり、したがって“S~”は相関係数行列を意味することになる。しかし実際には、そのままでは推定されるのは共分散行列になってしまうので、25行目から29行目で対角要素を1に固定して推定を行うことで相関行列の推定を行えるようにする。30行目から35行目では、グループ3で計算された偏相関係数の絶対値が最小であった逆行列の要素を0に固定していく過程が記述されているが、今は“!”でコメントアウトしてある。

最後にグループ3であるが、グループ1で宣言された行列S以外に単位行列Iを新たに宣言し、42行目から44行目で偏相関行列Pを計算する。

使用方法

上記のフルモデルのプログラムを走らせ、計算された結果である偏相関行列Pの絶対値が最小である要素を探索する。本例では5行4列目になる。

次にこの要素を0に固定して推定できるように、7行目から12行目を

```

07 Specify S
08 01
09 02 03
10 04 05 06
11 07 08 09 10
12 11 12 13 0 15

```

のように、要素14を0に変更する。同時に30行目の

“!”を削除し、コメントアウトを解放する。そして再度プログラムを走らせる。この過程を適合度が悪くなるまで繰り返し、モデルを更新していくことによって、

適用例 1 と同様の結果を得ることができる。

(2006 年 7 月 1 日受付, 2007 年 7 月 31 日最終修正)