

まえがき

本書では共分散構造分析 (CSA, Covariance Structure Analysis), あるいは構造方程式モデリング (SEM, Structural Equation Modeling) と呼ばれる数理統計モデルの応用可能性を論じる。

共分散構造分析では、他の多くの多変量解析モデルと異なり、固定化した数理モデルをデータに当てはめるのではなく、データの性質・実質科学的知見・先行研究・常識・事前情報等を利用して分析者固有の数理モデルを構成することができる。この性質を利用すると、旧来から利用されている多くの数理モデルをも構造方程式モデルの下位モデルに帰着させることができる。これは、多くの数理モデルを統一的に説明するという理論的な興味深さばかりでなく、プログラムを 1 から組むことなしに独自のモデルを構成し、分析することが可能であるという応用的に優れた性質である。

応用編では、共分散構造モデルによって様々な数理モデルがどのように表現されるかを具体的に詳述する。モデルの含意とデータの特徴をパスダイアグラム・方程式・共分散構造によって表現し、統一的な視点から数理モデル群の展望を行う。応用編に登場するモデルをそのまま実行するだけでも相当な高度な分析が可能になるし、掲載されたモデルを学んだ後には、読者自身のデータに合わせて独自のモデルを構成できるようになっているに違いない。

モデルの表現方法は SEM の理論体系のなかでもっとも実用的な領域である。本書は入門編の続編として執筆されているので、入門編あるいは同程度の統計学的予備知識を必要とする。しかし入門編と比較すると、新出する数学的内容は、幸いなことによても少ない。入門編で険しい山に登ったとの印象を持った読者も、応用編の数理的負担は少ないと感じるはずである。わずかに入門編の線形代数の内容を越える部分があるけれども、それらに関しては中級の入門を付録につけたので、適宜参照していただきたい。

1998 年 11 月 17 日

豊田秀樹

共分散構造分析<応用編>

—構造方程式モデリング—

立教大学社会学部 豊田秀樹

目次

1 方程式モデルの表現	2	4.5 ベクトル自己回帰モデル	44
1.1 RAM (Reticular Action Model)	2	4.6 動的因子分析	46
1.2 EQS(EQuationS model)	3	4.7 時系列因子分析	48
1.3 LISREL(Linear Structural RElations model)	4		
1.4 モデル表現の相違	5		
1.5 モデル間の関係	6		
1.6 非確率ベクトル	8		
1.7 モデルの指定例	9		
1.8 問題	11		
2 因子分析法	12		
2.1 探索的因子分析	12		
2.2 確認的因子分析	16		
2.3 2次(高次)因子分析	16		
2.4 独自因子と誤差因子の分離	17		
2.5 イプサティブモデル	19		
3 実験データの解析	25		
3.1 1要因実験	25		
3.2 2要因実験	27		
3.3 3要因実験	30		
3.4 分割実験	33		
3.5 共分散分析	34		
3.6 因子の分散分析	35		
3.7 因子の共分散分析	37		
3.8 多変量分散分析	37		
4 時系列解析	39		
4.1 定常性とトーブリツツ行列	39	A. データ・プログラム	67
4.2 自己回帰モデル	41	A.1 第2章 因子分析	67
4.3 移動平均モデル	42	A.2 第3章 実験データの解析	69
4.4 自己回帰移動平均モデル	44	A.3 第4章 時系列解析	73
		A.4 第5章	75
		A.5 第6章	77
		A.6 第7章	77
		A.7 第8章	78
		A.8 第9章	78
		A.9 第10章	78
		A.10 第11章	78

図目次

1.1	RAM による表現	6
1.2	LISREL による表現	6
1.3	逐次モデル	10
1.4	誤差相関モデル	10
1.5	2 次因子分析	10
1.6	縦断モデル 1	11
1.7	縦断モデル 2	11
2.1	2 次因子分析	17
2.2	独自因子と誤差因子の分離	19
4.1	景気動向指数(一致指数)	42
4.2	景気動向指数(先行指数)	45
4.3	景気動向指数(遅行指数)	47
4.4	先行・一致・遅行指標の同時グラフ	49
5.1	遺伝因子分析モデル	56
5.2	遺伝因子分析モデル(一般児)	58
5.3	遺伝因子分析モデル(双生児)	58
6.1	変数 1 から変数 3 の多変量散布図	62
6.2	変数 1 のヒストグラム	63
6.3	変数 2 のヒストグラム	63

1 方程式モデルの表現

共分散構造モデル (covariance structure model) は、因子分析・テスト理論・行動遺伝学モデル・多相データ解析・実験計画・質的データ解析・多変量解析・パス解析・時系列解析などを統一的に表現した数理統計モデルであり、定量的な実証研究にしばしば利用される計量的分析手法である。相関行列や共分散行列を母数の関数として表現するというアイデアの起源を单一の研究に求めることは困難であるが、現在の相関・共分散構造の原形といってよいモデルは、まず Bock & Bargmann(1966)¹ によって提案された。ただし提案されたモデルは、理論的にも応用的にも必ずしも扱いやすいものではなかったので、その後 Jöreskog(1970a)² によって理論の整理がなされ、続いて Jöreskog & Sörbom(1976)³ の計算機プログラム LISREL (リズレル, LInear Structural RElations) によって応用的な重要性が広く認識されるようになった。

共分散構造分析には様々な呼び名があるが、Sörbom(1974)⁴ が、共分散構造ばかりでなく、平均値をも構造化して示したことを契機に種々の拡張がなされ、共分散構造モデルという名称が必ずしもふさわしい呼び名とはいえなくなった。近年では、構造方程式モデル (Structural Equation Model with latent variables, 潜在変数を伴う構造方程式モデル、SEM と略されることもある) という呼び名のほうが一般的になりつつある。分析の本来の目的が方程式中の母数の推定であり、共分散構造の分析はその手段であることを考えると、構造方程式モデルという呼び名の方が、より直接的であるかもしれない。

社会科学・人文科学・行動科学の研究領域では「知能」や「社会的地位」や「性格」や「購買力」などの構成概念を扱う必要がある。構成概念とは、とりあえずその存在を仮定することによって複雑に込み入った現象を比較的単純に理解することを目的として構成した概念である。潜在変数は因子、潜在特性などと呼ばれることもあり、厳密な測定が難しい構成概念を扱うために統計モデル

¹Bock, R.D., & Bargmann, R.E. (1966). Analysis of Covariance Structures. *Psychometrika*, **31**, 507-534.

²Jöreskog, K.G. (1970). A General Method for Analysis of Covariance structures. *Biometrika*, **57**, 239-251.

³Jöreskog, K.G. & Sörbom, D. (1976). *LISREL III: Estimation of Linear Structural Equation Systems by Maximum Likelihood Methods*. Chicago:National Educational Resources, Inc.

⁴Sörbom, 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.

に導入される仮想的な変数である。

SEM の研究領域は多岐にわたるが、モデルを応用する際に最も重要なのは、方程式モデルの表現方法である。Bock & Bargmann(1966) を改良した LISREL は、構成概念間の回帰モデル (同時方程式モデル) という非常に分かりやすい具体的表現を採用したので、様々な研究領域での実際的な応用を促した。しかしモデル表現が構成概念間の回帰モデルという観点からのみ強調されすぎ、SEM には構成概念間の回帰モデル以外にも様々な可能性があることが、応用現場に広まらないという副作用も生じさせてしまった。

1970 年代後半以降、そのような状況を開拓するために、LISREL よりも柔軟な 2 つのモデル表現が提案された。Bentler & Weeks(1980)⁵ による EQS (イーキューエス) モデル (EQuationS model) と McArdle(1980) と McArdle & McDonald(1984)⁶ による RAM (ラム, Reticular Action Model) である。本章では、複数のモデル表現の方法を比較し、平均・共分散構造の性質を調べていく。

1.1 RAM (Reticular Action Model)

RAM は、構造方程式モデルを、最も簡潔に効率よく記述することが可能な表現方法である。このため入門編では、混乱を避ける意味もあり、効率の良い RAM だけを用いてモデルを記述してきた。本節では、まず入門編で導入した構造方程式モデル RAM を復習する。

1.1.1 方程式の表現

RAM の構造方程式は

$$\mathbf{t} = \boldsymbol{\alpha}_0 + \mathbf{A}\mathbf{t} + \mathbf{u} \quad (1.1)$$

である。ここで \mathbf{t} は構造変数ベクトルである。構造変数には、構成概念 \mathbf{f} と、観測変数 \mathbf{x} とがあり、それを縦に並べて

$$\mathbf{t} = (\mathbf{f}' \ \mathbf{x}')' \quad (1.2)$$

⁵Bentler, P.M., & Weeks, D.G. (1980). Linear Structural Equations with Latent Variables. *Psychometrika*, **45**, 289-308.

⁶McArdle, J.J. (1980). Causal modeling applied to psychonomic systems simulation. *Behavior Research Methods and Instrumentation*, **12**, 193-209.

McArdle, J.J., & McDonald, R.P. (1984). Some algebraic properties of the reticular action model for moment structures. *British Journal of Mathematical and Statistical Psychology*, **37**, 234-251.

と表現する⁷. α_0 は構造変数が他の変数から影響を受ける前の平均値ベクトルである. \mathbf{u} は外生変数ベクトルである. 外生変数には, \mathbf{f} に関する残差変数 \mathbf{d} と, \mathbf{x} に関する残差変数 \mathbf{e} があり, それを縦に並べて

$$\mathbf{u} = (\mathbf{d}' \ \mathbf{e}')' \quad (1.3)$$

と表現する. (1.3) 式は外生変数(残差変数)なのだから, \mathbf{f} の i 番目の要素 f_i が外生変数であれば \mathbf{d} の i 番目の要素は f_i 自身となり, 内生変数であれば \mathbf{d} の i 番目の要素は誤差変数 d_i となる. 同様に, \mathbf{x} の i 番目の要素 x_i が外生変数であれば \mathbf{e} の i 番目の要素は x_i 自身となり, 内生変数であれば \mathbf{e} の i 番目の要素は誤差変数 e_i となる.

係数行列 \mathbf{A} は, 4 つの行列

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_a & \mathbf{A}_d \\ \mathbf{A}_b & \mathbf{A}_c \end{bmatrix} \quad (1.4)$$

から構成される. ここで

- \mathbf{A}_a : \mathbf{f} から \mathbf{f} への規定力を表現する係数行列
- \mathbf{A}_b : \mathbf{f} から \mathbf{x} への規定力を表現する係数行列
- \mathbf{A}_c : \mathbf{x} から \mathbf{x} への規定力を表現する係数行列
- \mathbf{A}_d : \mathbf{x} から \mathbf{f} への規定力を表現する係数行列

である. 以上の表記を考慮すると (1.1) 式は

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{x} \end{bmatrix} = \alpha_0 + \begin{bmatrix} \mathbf{A}_a & \mathbf{A}_d \\ \mathbf{A}_b & \mathbf{A}_c \end{bmatrix} \begin{bmatrix} \mathbf{f} \\ \mathbf{x} \end{bmatrix} + \begin{bmatrix} \mathbf{d} \\ \mathbf{e} \end{bmatrix} \quad (1.5)$$

と書き下すことができる.

1.1.2 平均・共分散構造

観測変数の共分散行列をモデルの母数によって構造化する. まず構造変数ベクトルから観測変数のみを

$$\mathbf{x} = \mathbf{G}\mathbf{t} \quad (1.6)$$

⁷縦ベクトル $\begin{bmatrix} a \\ b \end{bmatrix}$ は常に $(a' \ b')'$ と表現できる. 紙面の節約のために後者の表現が用いられることが多い.

を使って取り出す. (1.6) 式を選択方程式という. 式中の \mathbf{G} は

$$\mathbf{G} = [\mathbf{O} \ \mathbf{I}] \quad (1.7)$$

のようゼロ行列と単位行列を横にならべた矩形の定数行列である. 共分散構造を導くためには, まず $\mathbf{T} = (\mathbf{I} - \mathbf{A})^{-1}$ の存在を仮定し, 構造方程式を

$$\mathbf{t} = \mathbf{T}(\alpha_0 + \mathbf{u}) \quad (1.8)$$

と表現する. \mathbf{t} を (1.6) 式に代入して

$$\mathbf{x} = \mathbf{G}\mathbf{T}(\alpha_0 + \mathbf{u}) \quad (1.9)$$

を得る. $E[\mathbf{u}] = \mathbf{o}$ は自然に仮定できるので, 観測変数の期待値の構造は

$$\mu(\theta) = E[\mathbf{x}] = \mathbf{G}\mathbf{T}\alpha_0 \quad (1.10)$$

と導かれる. 共分散構造は

$$\begin{aligned} \Sigma(\theta) &= E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])'] \\ &\quad [\quad \mathbf{x} - E[\mathbf{x}] = \mathbf{G}\mathbf{T}(\alpha_0 + \mathbf{u}) - \mathbf{G}\mathbf{T}\alpha_0 = \mathbf{G}\mathbf{T}\mathbf{u} \text{ なので }] \\ &= E[\mathbf{G}\mathbf{T}\mathbf{u}\mathbf{u}'\mathbf{T}'\mathbf{G}'] = \mathbf{G}\mathbf{T}\Sigma_u\mathbf{T}'\mathbf{G}' \end{aligned} \quad (1.11)$$

と導かれる. ここで Σ_u は

$$\Sigma_u = E[\mathbf{u}\mathbf{u}'] = \begin{bmatrix} \Sigma_d & \Sigma_{de} \\ \Sigma_{ed} & \Sigma_e \end{bmatrix} \quad (1.12)$$

と表現された残差ベクトルの共分散行列である.

1.2 EQS(EQuationS model)

RAM はパス図に登場する変数を外生変数ベクトルと構造変数ベクトルに分けて配置することによってモデルを表現した((1.1) 式参照). 一方, パス図に登場する変数を, 主として外生変数ベクトルと内生変数ベクトルに分けて配置することによってモデルを表現する方法があり, これは EQS モデルと呼ばれている. 本節では, EQS モデルの特徴を紹介する.

1.2.1 方程式の表現

EQS の構造方程式は

$$\mathbf{t}_0 = \boldsymbol{\alpha}_0 + \mathbf{A}_0 \mathbf{t}_0 + \boldsymbol{\Gamma}_0 \mathbf{u} \quad (1.13)$$

である。ここで \mathbf{u} が外生変数ベクトルであることは (1.1) 式と変わりない。 \mathbf{t}_0 は、内生変数と観測変数を並べたベクトルである(ただし後述する (1.14) 式の選択方程式で観測変数を取り出す関係上、外生的観測変数も \mathbf{t}_0 に含まれている必要がある)。 \mathbf{A}_0 は内生変数から内生変数への規定力を表現する係数行列であり、 $\boldsymbol{\Gamma}_0$ は外生変数から内生変数への規定力を表現する係数行列である。

1.2.2 平均・共分散構造

(1.6) 式にならって、選択方程式

$$\mathbf{x} = \mathbf{G}\mathbf{t}_0 \quad (1.14)$$

を構成し、 $\mathbf{T}_0 = (\mathbf{I} - \mathbf{A})^{-1}$ の存在を仮定すると、構造方程式の誘導形は

$$\mathbf{t}_0 = \mathbf{T}_0(\boldsymbol{\alpha}_0 + \boldsymbol{\Gamma}_0 \mathbf{u}) \quad (1.15)$$

となる。 \mathbf{t}_0 を (1.14) 式に代入して

$$\mathbf{x} = \mathbf{G}\mathbf{T}_0(\boldsymbol{\alpha}_0 + \boldsymbol{\Gamma}_0 \mathbf{u}) \quad (1.16)$$

を得る。因子などの平均を含む部分に $E[\mathbf{u}] = \boldsymbol{\alpha}_u$ を仮定し、観測変数の平均構造は

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = E[\mathbf{x}] = \mathbf{G}\mathbf{T}_0(\boldsymbol{\alpha}_0 + \boldsymbol{\Gamma}_0 \boldsymbol{\alpha}_u) \quad (1.17)$$

である。共分散構造は

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])'] = \mathbf{G}\mathbf{T}_0\boldsymbol{\Gamma}_0\boldsymbol{\Sigma}_u\boldsymbol{\Gamma}_0'\mathbf{T}_0'\mathbf{G}' \quad (1.18)$$

と導かれる。

1.3 LISREL(Linear Structural Relations model)

1.3.1 方程式の表現

本節では、入門編で慣れ親しんだ RAM との関係から LISREL を導入する。LISREL では (1.2) 式中の構成概念ベクトル \mathbf{f} を

$$\mathbf{f} = (\xi' \eta')' \quad (1.19)$$

のように 2 つに分ける。 ξ は外生的な構成概念を表し、 η は内生的な構成概念を表す。 \mathbf{f} の表記に合わせて、(1.3) 式の残差ベクトル \mathbf{d} は

$$\mathbf{d} = (\xi' d_{\$}')' \quad (1.20)$$

と表現する。構成概念間の関係を記述する構造方程式には

$$\boldsymbol{\eta} = \boldsymbol{\alpha}_{\eta} + \mathbf{B}_{+}\boldsymbol{\eta} + \boldsymbol{\Gamma}\xi + d_{\$} \quad (1.21)$$

を用いる。 ξ は外生変数なので左辺に置かないこと、観測変数が無いので選択方程式を必要としないことが、RAM との相違点である。

選択方程式がない代わりに、構成概念と観測変数の関係を表現するための方程式を用いる。残差ベクトル \mathbf{d} や構成概念ベクトル \mathbf{f} の表記に合わせて、観測変数を

$$\mathbf{x} = (x_{\$}' y_{\$}')' \quad (1.22)$$

のように分ける。 $x_{\$}$ は ξ から影響を受ける観測変数とし、 $y_{\$}$ は η から影響を受ける観測変数とする。(1.22) 式の表記に合わせて、(測定) 誤差変数ベクトルも

$$\mathbf{e} = (e_x' e_y')' \quad (1.23)$$

のように分け、 $x_{\$}$ に関する測定状況を

$$x_{\$} = \boldsymbol{\alpha}_{x\$} + \mathbf{\Lambda}_x\xi + e_x \quad (1.24)$$

で、 $y_{\$}$ に関する測定状況を

$$y_{\$} = \boldsymbol{\alpha}_{y\$} + \mathbf{\Lambda}_y\eta + e_y \quad (1.25)$$

で表現する. (1.24) 式, (1.25) 式を測定方程式という. LISREL の外生変数ベクトルは

$$\mathbf{u} = (\mathbf{d}' \ \mathbf{e}')' = (\boldsymbol{\xi}' \ \mathbf{d}_\$' \ \mathbf{e}_x' \ \mathbf{e}_y')' \quad (1.26)$$

と分割され, その共分散行列は,

$$\boldsymbol{\Sigma}_u = \begin{bmatrix} \boldsymbol{\Sigma}_{\xi} & & & sym \\ \mathbf{O} & \boldsymbol{\Sigma}_{d\$} & & \\ \mathbf{O} & \mathbf{O} & \boldsymbol{\Sigma}_{ex} & \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \boldsymbol{\Sigma}_{ey} \end{bmatrix} \quad (1.27)$$

のように制約する.

1.3.2 平均・共分散構造

逆行列 $\mathbf{B}_\$ = (\mathbf{I} - \mathbf{B}_+)^{-1}$ の存在を仮定すると構造方程式の誘導形は

$$\boldsymbol{\eta} = \mathbf{B}_\$ (\boldsymbol{\alpha}_\eta + \boldsymbol{\Gamma} \boldsymbol{\xi} + \mathbf{d}_\$) \quad (1.28)$$

だから, $\mathbf{y}_\$$ は, 外生変数の重み付き和で

$$\mathbf{y}_\$ = \boldsymbol{\alpha}_{y\$} + \boldsymbol{\Lambda}_y (\mathbf{B}_\$ (\boldsymbol{\alpha}_\eta + \boldsymbol{\Gamma} \boldsymbol{\xi} + \mathbf{d}_\$)) + \mathbf{e}_y \quad (1.29)$$

と表現される. 誤差変数には $E[\mathbf{d}_\$] = \mathbf{o}$, $E[\mathbf{e}_y] = \mathbf{o}$, $E[\mathbf{e}_x] = \mathbf{o}$ が自然に仮定できる. 一方, 構成概念である $\boldsymbol{\xi}$ の期待値は, 平均構造を考察する際に重要であり, 0 に固定することは得策ではないので $E[\boldsymbol{\xi}] = \boldsymbol{\alpha}_\xi$ と表記する. 以上の仮定より

$$E[\mathbf{y}_\$] = \boldsymbol{\alpha}_{y\$} + \boldsymbol{\Lambda}_y (\mathbf{B}_\$ (\boldsymbol{\alpha}_\eta + \boldsymbol{\Gamma} \boldsymbol{\alpha}_\xi)) \quad (1.30)$$

$$E[\mathbf{x}_\$] = \boldsymbol{\alpha}_{x\$} + \boldsymbol{\Lambda}_x \boldsymbol{\alpha}_\xi \quad (1.31)$$

が導かれ, 観測変数の平均構造は

$$\begin{aligned} \boldsymbol{\mu}(\boldsymbol{\theta}) &= E[\mathbf{x}] = E[(\mathbf{y}_\$' \mathbf{x}_\$')] = (E[\mathbf{y}_\$]' \ E[\mathbf{x}_\$]')' \\ &= ((\boldsymbol{\alpha}_{y\$} + \boldsymbol{\Lambda}_y (\mathbf{B}_\$ (\boldsymbol{\alpha}_\eta + \boldsymbol{\Gamma} \boldsymbol{\alpha}_\xi)))' \ (\boldsymbol{\alpha}_{x\$} + \boldsymbol{\Lambda}_x \boldsymbol{\alpha}_\xi)')' \end{aligned} \quad (1.32)$$

と表現される. (1.27) 式の制約を考慮すると, $\mathbf{y}_\$$ の共分散構造は

$$\begin{aligned} \boldsymbol{\Sigma}_{y\$} &= E[(\mathbf{y}_\$ - E[\mathbf{y}_\$])(\mathbf{y}_\$ - E[\mathbf{y}_\$])'] \\ &= \boldsymbol{\Lambda}_y \mathbf{B}_\$ (\boldsymbol{\Gamma} \boldsymbol{\Sigma}_\xi \boldsymbol{\Gamma}' + \boldsymbol{\Sigma}_{d\$}) \mathbf{B}_\$' \boldsymbol{\Lambda}_y' + \boldsymbol{\Sigma}_{e_y} \end{aligned} \quad (1.33)$$

と表現され, $\mathbf{x}_\$$ の共分散構造は

$$\boldsymbol{\Sigma}_{x\$} = E[(\mathbf{x}_\$ - E[\mathbf{x}_\$])(\mathbf{x}_\$ - E[\mathbf{x}_\$])'] = \boldsymbol{\Lambda}_x \boldsymbol{\Sigma}_\xi \boldsymbol{\Lambda}_x' + \boldsymbol{\Sigma}_{e_x} \quad (1.34)$$

と表現される. $\mathbf{x}_\$$ と $\mathbf{y}_\$$ の共分散構造は

$$\boldsymbol{\Sigma}_{xy\$} = E[(\mathbf{x}_\$ - E[\mathbf{x}_\$])(\mathbf{y}_\$ - E[\mathbf{y}_\$])'] = \boldsymbol{\Lambda}_x \boldsymbol{\Sigma}_\xi \boldsymbol{\Gamma}' \mathbf{B}_\$' \boldsymbol{\Lambda}_y' \quad (1.35)$$

となる. したがって観測変数の共分散構造は分割行列の形式で

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Lambda}_y \mathbf{B}_\$ (\boldsymbol{\Gamma} \boldsymbol{\Sigma}_\xi \boldsymbol{\Gamma}' + \boldsymbol{\Sigma}_{d\$}) \mathbf{B}_\$' \boldsymbol{\Lambda}_y' + \boldsymbol{\Sigma}_{e_y} & \boldsymbol{\Lambda}_y \mathbf{B}_\$ \boldsymbol{\Gamma} \boldsymbol{\Sigma}_\xi \boldsymbol{\Lambda}_x' \\ \boldsymbol{\Lambda}_x \boldsymbol{\Sigma}_\xi \boldsymbol{\Gamma}' \mathbf{B}_\$' \boldsymbol{\Lambda}_y' & \boldsymbol{\Lambda}_x \boldsymbol{\Sigma}_\xi \boldsymbol{\Lambda}_x' + \boldsymbol{\Sigma}_{e_x} \end{bmatrix} \quad (1.36)$$

と導かれる. RAM や EQS と比較すると複雑である.

1.4 モデル表現の相違

RAM, EQS, LISREL の平均・共分散構造が導かれたので, ここで図 1.1 のモデルを 3 つの表現方法で記述してみよう. ただし, このモデルでは複雑な平均構造は考えないものとする. まず, 入門編でなじみ深い(1.1)式の RAM の構造方程式は以下の表現となる.

$$\begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix} = \boldsymbol{\alpha}_0 + \begin{bmatrix} 0 & 0 & 0 \\ \alpha_{a21} & 0 & 0 \\ \alpha_{a31} & \alpha_{a32} & 0 \\ \alpha_{b11} & 0 & 0 \\ \alpha_{b21} & 0 & 0 \\ 0 & \alpha_{b32} & 0 \\ 0 & \alpha_{b42} & 0 \\ 0 & 0 & \alpha_{b53} \\ 0 & 0 & \alpha_{b63} \end{bmatrix} \mathbf{O}_{9 \times 6} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix} + \begin{bmatrix} f_1 \\ d_2 \\ d_3 \\ e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{bmatrix} \quad (1.37)$$

一方 (1.13) 式の EQS の構造方程式は以下の表現となる.

$$\begin{bmatrix} f_2 \\ f_3 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix} = \boldsymbol{\alpha}_0 + \begin{bmatrix} 0 & 0 \\ \alpha_{a32} & 0 \\ 0 & 0 \\ 0 & 0 \\ \alpha_{b32} & 0 \\ \alpha_{b42} & 0 \\ 0 & \alpha_{b53} \\ 0 & \alpha_{b63} \end{bmatrix} \mathbf{O}_{8 \times 6} \begin{bmatrix} f_2 \\ f_3 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix} + \begin{bmatrix} \alpha_{a21} \\ \alpha_{a31} \\ \alpha_{b11} \\ \alpha_{b21} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \mathbf{I}_{8 \times 8} \begin{bmatrix} f_1 \\ d_2 \\ d_3 \\ e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{bmatrix} \quad (1.38)$$

RAM の表現では t に入っていた外生的構成概念 f_1 は、EQS の表現では t_0 に入っていない。また f_1 からの係数が Γ_0 に入っていることが RAM との相違である。 (1.12) 式のモデルの残差ベクトルの共分散行列は、RAM と EQS では共通しており、以下の表現となる。

$$\Sigma_u = \begin{bmatrix} \sigma_{f1}^2 & & & & & & \\ 0 & \sigma_{d2}^2 & & & & & \\ 0 & 0 & \sigma_{d3}^2 & & & & \\ 0 & 0 & 0 & \sigma_{e1}^2 & & & \\ 0 & 0 & 0 & 0 & \sigma_{e2}^2 & & \\ 0 & 0 & 0 & 0 & 0 & \sigma_{e3}^2 & \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{e4}^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{e5}^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{e6}^2 \end{bmatrix} \quad (1.39)$$

RAM と EQS の表現を比較すると、EQS のほうが行列が 1 つ多く、表現の簡潔さという観点からは冗長である。SEM の数理的な性質を調べる場合に RAM は便利である。ただし内生変数から内生変数への係数のないモデルを扱う場合に EQS を用いると、逆行列がなくなる（単位行列になる）という長所がある。

同じモデルを LISREL で表現すると、(1.21) 式の構造方程式は、

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \alpha_\eta + \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \end{bmatrix} [\xi_1] + \begin{bmatrix} d_{\$1} \\ d_{\$2} \end{bmatrix} \quad (1.40)$$

となる。RAM や EQS の表現と比較すると、小さな母数行列がたくさん登場する。(1.24) 式と(1.25)式の測定方程式は、それぞれ

$$\begin{bmatrix} x_{\$1} \\ x_{\$2} \end{bmatrix} = \alpha_{x\$} + \begin{bmatrix} \lambda_{x11} \\ \lambda_{x21} \end{bmatrix} [\xi_1] + \begin{bmatrix} e_{x1} \\ e_{x2} \end{bmatrix} \quad (1.41)$$

$$\begin{bmatrix} y_{\$1} \\ y_{\$2} \\ y_{\$3} \\ y_{\$4} \end{bmatrix} = \alpha_{y\$} + \begin{bmatrix} \lambda_{y11} & 0 \\ \lambda_{y21} & 0 \\ 0 & \lambda_{y32} \\ 0 & \lambda_{y42} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} e_{y1} \\ e_{y2} \\ e_{y3} \\ e_{y4} \end{bmatrix} \quad (1.42)$$

となる。外生変数の共分散行列は

$$\Sigma_\xi = \sigma_{\xi_1}^2, \quad \Sigma_{d\$} = \begin{bmatrix} \sigma_{d\$1}^2 & 0 \\ 0 & \sigma_{d\$2}^2 \end{bmatrix}, \quad \Sigma_{ex} = \begin{bmatrix} \sigma_{e_{x1}}^2 & 0 \\ 0 & \sigma_{e_{x2}}^2 \end{bmatrix},$$

$$\Sigma_{ey} = \begin{bmatrix} \sigma_{ey1}^2 & 0 & 0 & 0 \\ 0 & \sigma_{ey2}^2 & 0 & 0 \\ 0 & 0 & \sigma_{ey3}^2 & 0 \\ 0 & 0 & 0 & \sigma_{ey4}^2 \end{bmatrix}$$

と表現される。RAM と EQS では同じ母数名を与えたが、LISREL で同様のことを行うと、モデルの表現が分かりにくくなるので、図 1.1 と同じモデルを表現した図 1.2 に LISREL の母数名を書き込んだので比較していただきたい。

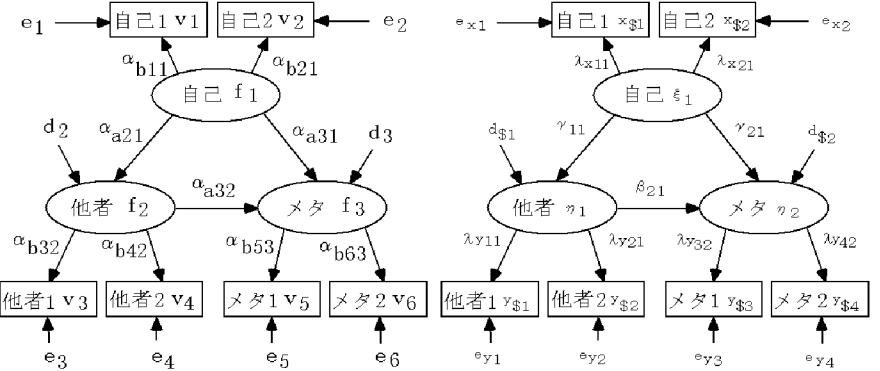


図 1.1: RAM による表現

図 1.2: LISREL による表現

1.5 モデル間の関係

LISREL と RAM の関係を調べて見よう。まず RAM における構造ベクトルは、LISREL の表記では

$$t = (f' \ v')' = (\xi' \ \eta' \ x_{\$}' \ y_{\$}')' \quad (1.43)$$

となる。同様に(1.26)式の残差(外生)変数ベクトルを利用すると LISREL は RAM の表現を用いて

$$\begin{bmatrix} \xi \\ \eta \\ x_{\$} \\ y_{\$} \end{bmatrix} = \alpha_0 + \begin{bmatrix} O & O & O & O \\ \Gamma & B_+ & O & O \\ \Lambda_x & O & O & O \\ O & \Lambda_y & O & O \end{bmatrix} \begin{bmatrix} \xi \\ \eta \\ x_{\$} \\ y_{\$} \end{bmatrix} + \begin{bmatrix} \xi \\ d_{\$} \\ e_x \\ e_y \end{bmatrix} \quad (1.44)$$

と表現することが可能である。(1.1)式(あるいは(1.5)式)と(1.44)式とを見比べると RAM の 1 部に制約を入れたものが LISREL であることがわかる。こ

れは LISREL が RAM の下位モデルであることを意味する。言い換えるならば、LISREL で表現できるモデルは必ず RAM で表現できるが、RAM で表現できるモデルは LISREL で表現できるとは限らないということを (1.44) 式は示している。

具体的には、以下のモデルは LISREL の本来の使用法では表現できない⁸。

限界 1: $\mathbf{A}_c = \mathbf{O}$ なので、構成概念の登場しない観測変数のパス解析を表現できない。また $\mathbf{A}_d = \mathbf{O}$ なので、観測変数が構成概念を形成するモデルを表現できない。

限界 2: (1.44) 式では \mathbf{A}_b に相当する部分がブロックダイアゴナルの形式

$$\mathbf{A}_b = \begin{bmatrix} \mathbf{\Lambda}_x & \mathbf{O} \\ \mathbf{O} & \mathbf{\Lambda}_y \end{bmatrix} \quad (1.45)$$

に制約されているので、 $\boldsymbol{\xi}$ と $\boldsymbol{\eta}$ の両方から直接影響をうける観測変数を表現できない。

限界 3: (1.27) 式の Σ_u がブロックダイアゴナルの形式で表現されているので 4 つの異なる種類の確率変数相互間の共分散を表現することができない。

(1.44) 式と同様に LISREL は EQS の表現 (1.13) 式を用いて

$$\begin{bmatrix} \boldsymbol{\eta} \\ \mathbf{x}_{\$} \\ \mathbf{y}_{\$} \end{bmatrix} = \boldsymbol{\alpha}_0 + \begin{bmatrix} \mathbf{B}_+ & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \mathbf{\Lambda}_y & \mathbf{O} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta} \\ \mathbf{x}_{\$} \\ \mathbf{y}_{\$} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{I} & \mathbf{O} & \mathbf{O} \\ \mathbf{\Lambda}_x & \mathbf{O} & \mathbf{I} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{O} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi} \\ \mathbf{d}_{\$} \\ \mathbf{e}_x \\ \mathbf{e}_y \end{bmatrix} \quad (1.46)$$

と表現することが可能である。表現 A によって記述できるモデルの集合を $M(A)$ と表記すると、(1.44) 式と (1.46) 式は

$$M(\text{LISREL}) \subset M(\text{RAM}) \quad (1.47)$$

$$M(\text{LISREL}) \subset M(\text{EQS}) \quad (1.48)$$

⁸ \mathbf{A}_a に相当する部分が

$$\mathbf{A}_a = \begin{bmatrix} \mathbf{O} & \mathbf{O} \\ \boldsymbol{\Gamma} & \mathbf{B}_+ \end{bmatrix}$$

に制約されているが、 $\boldsymbol{\xi}$ は外生変数なので、それに対する係数がゼロに制約されていることはモデルの表現を制約することにはならない。

のように、LISREL で記述できるモデルは必ず RAM や EQS で表現できることを意味している。しかし、LISREL の共分散構造 (1.36) 式は、標本共分散行列の近似に利用されるのであるから、仮に LISREL を

$$\mathbf{x}_{\$} = \mathbf{o}, \quad \boldsymbol{\Gamma} = \mathbf{I}, \quad \Sigma_{d\$} = \mathbf{O}, \quad \Sigma_{e_y} = \mathbf{O}, \quad \Sigma_{e_x} = \mathbf{O} \quad (1.49)$$

と制約し、

$$\mathbf{A}_y = \mathbf{G}, \quad \mathbf{B}_{\$} = \mathbf{T}, \quad \Sigma_{\xi} = \Sigma_u \quad (1.50)$$

と見なせば、LISREL を制約したものが RAM であると見なすことができる。また LISREL を

$$\mathbf{x}_{\$} = \mathbf{o}, \quad \Sigma_{d\$} = \mathbf{O}, \quad \Sigma_{e_y} = \mathbf{O}, \quad \Sigma_{e_x} = \mathbf{O} \quad (1.51)$$

と制約し、

$$\mathbf{A}_y = \mathbf{G}, \quad \mathbf{B}_{\$} = \mathbf{T}_0, \quad \boldsymbol{\Gamma} = \boldsymbol{\Gamma}_0, \quad \Sigma_{\xi} = \Sigma_u \quad (1.52)$$

と見なせば、LISREL を制約したものが EQS であると見なすことができる。つまり本来のモデルの使用法を離れれば、(1.50) 式と (1.52) 式は

$$M(\text{RAM}) \subset M(\text{LISREL}) \quad (1.53)$$

$$M(\text{EQS}) \subset M(\text{LISREL}) \quad (1.54)$$

であることを示している。言い換えるならば、RAM や EQS で記述できるモデルは必ず LISREL で表現できることが示されたことになる。

$M(A) \subset M(B)$ かつ $M(B) \subset M(A)$ が同時に成立つということは、2 つの集合が等しいということを意味している。したがって RAM, EQS, LISREL は、下位モデルの表現力という観点からは、見かけは違ってもみな同一のモデル⁹である。ならば、通常は、どれか 1 つの表現を用いればよい¹⁰ことになる。そこで入門編では

⁹ 豊田 (1991) 共分散構造分析の下位モデルとその適用例、教育心理学研究 38(4) 438-444.

¹⁰ モデルの表現方法には LISREL, EQS, RAM 以外にも

McDonald, R.P. (1978). A Simple Comprehensive Model for the Analysis of Covariance Structures. *British Journal of Mathematical and Statistical Psychology*, 31, 59-72.
の COSAN (コサン) モデル (A simple comprehensive model for the analysis of covariance structures model) がある。表現できるモデルの集合に関しては、COSAN も他の 3 つと同等であることが知られている。応用場面で利用されることが比較的少ないので詳細は割愛した。

1. 本来的な使用法から外れることなく、全てのモデルを表現でき
2. 平均・共分散構造の数理的な形が最も単純な

RAM に限定してモデルを学習した。ただし、見かけは違ってもみな同一のモデルだから、2つ以上知る必要はないという考え方は正しくない。数理的には等しくとも、特定の表現は特定の意図を強調するから見た目は重要である。たとえば LISREL は構成概念間のパス解析モデル（同時方程式モデル）という側面を強調した表現である。我々はモデルを数理として扱うばかりでなく、具体的な分析対象や実質科学的な理論との対応関係を重視するから、SEM がどのように表現されているかが、応用的には大切である。

本書・応用編の目的は SEM によって表現できる有用なモデルを、できるだけたくさん紹介することであるから、RAM, EQS, LISREL の関係を把握し、モデルの応用的な目的に合わせて適切な表現を用意する必要がある。場合によっては RAM, EQS, LISREL と本質的に同じでも、異なった表現を新たに構成する必要が生じる場合もある。

1.6 非確率ベクトル

SEM の最も単純なモデルに、一方の変数からもう一方の変数を予測する单回帰モデル

$$x_2 = \alpha_{20} + \alpha_{21}x_1 + e_2 \quad (1.55)$$

があった。この場合、 x_2 と x_1 は、それぞれ基準変数と予測変数と呼ばれている。单回帰モデルの応用場面は、

- 原料の重さから製品の重さを予測する場合、加熱時間から試料の硬さを予測する場合、音源までの距離から音圧を予測する場合、
- 無作為標本に対して模擬試験の成績から本試験の成績を予測する場合、最高血圧から最低血圧を予測する場合、身長から体重を予測する場合

など2種類に大別される。前者は予測変数が非確率変数のモデルである。重さを10グラムからはじめて、10グラムおきに100グラムまでとか、0秒から10秒間隔で3分までのように、予め定められた予測変数から基準変数を予測する場合である。これは「線形モデル」と呼ばれている。実験計画における分散分

析モデルは、予測変数が非確率変数（0か1の値だけをとる計画行列の変数）である線形モデルの特別な場合である。

後者は予測変数が確率変数のモデルである。まず標本を抽出し、抽出した標本の予測変数と基準変数を測定するから、基準変数ばかりでなく予測変数も確率的に変動する。こちらは、前者と区別するために「線形回帰モデル」と呼ばれることがある。

前者は実験データの分析に使われることが多く、後者は調査データの分析に使われることが多い。

これまで、構造方程式モデルを論じる際に、外生的観測変数（予測変数）も内生的観測変数（基準変数）も確率ベクトルとして扱ってきた。しかし線形モデルと同様に、外生的観測変数を非確率ベクトル（モデル中の所与の定数）として扱いたい場合は少なくない。たとえば

- 外生的観測変数が、0 m から 10 m 間隔で 100 m までのように予め値が固定されている場合
- 無作為標本であっても、外生的観測変数の分布が相当歪んでおり、分布を仮定せずに定数として扱いたい場合
- 実験データを SEM で分析するために、0 か 1 の値だけをとる実験計画を表現したい場合

などである。非確率変数である外生的観測変数を扱うことができれば、SEM の応用範囲は更に広がる。

1.6.1 方程式の表現

1.5 節では、多くの SEM の表現が、本質的に同じであることを導いたけれども、例外もある。下位モデルの表現力の拡大、あるいは応用可能性の観点から特筆できる例外は LISCOMP (リスコンプ analysis of LInear Structural equations with a COMprehensive measurement model, Muthen, 1984, 1987)¹¹ である¹²。

¹¹Muthen, B.O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49, 115-132.

Muthen, B.O. (1987). *LISCOMP: Analysis of linear Structural Equation with a Comprehensive Measurement Model*. Mooresville, IN: Scientific Software, Inc.

¹²LISCOMP は SEM システムの中でカテゴリカルな順序変数、上限下限のある変数を(1部で切断データをも)扱うことを最初に可能にしたソフトウェアである。順序変数、上限下限のある変

LISCOMP では従来の LISREL と同様に、測定方程式と構造方程式によってモデルを表現する。ただし測定方程式は 1 本である。それらは、それぞれ

$$\mathbf{y}_\$ = \alpha_{y\$} + \Lambda \boldsymbol{\eta} + \mathbf{e}_y \quad (1.56)$$

$$\boldsymbol{\eta} = \alpha_\eta + \mathbf{B}_+ \boldsymbol{\eta} + \Gamma \mathbf{x}_\$ + \mathbf{d}_\$ \quad (1.57)$$

である¹³。モデルに登場する変数や母数は LISREL のそれとほとんど同じである。ただし外生的観測変数を表現する $\mathbf{x}_\$$ は、見かけは同じでも、これまで登場した構造方程式モデルとは異なり、非確率変数である。簡単にいうと LISCOMP では $\mathbf{x}_\$$ を単なる数字(定数)として扱う。

1.6.2 平均・共分散構造

これまでのモデルと同様に LISCOMP にも 3 種類の仮定を入れて平均・共分散構造を導く。まず $\mathbf{B}_\$ = (\mathbf{I} - \mathbf{B}_+)^{-1}$ の存在を仮定する。次に、誤差変数の期待値はゼロである ($E[\mathbf{e}_y] = \mathbf{o}$, $E[\mathbf{d}_\$] = \mathbf{o}$) ことを仮定する。そして外生変数の確率ベクトル $\mathbf{u} = (\mathbf{e}'_y \mathbf{d}'_\$)'$ の共分散行列は

$$\Sigma_u = \begin{bmatrix} \Sigma_{ey} & \mathbf{0} \\ \mathbf{0} & \Sigma_{d\$} \end{bmatrix} \quad (1.58)$$

であると仮定する。外生的観測変数 $\mathbf{x}_\$$ は確率変数ではないので (1.58) 式には登場しない。

母数の推定の基本となるのは、 $\mathbf{x}_\$$ を定数として扱った場合の $\mathbf{y}_\$$ の期待値と分散¹⁴である。まず (1.57) 式から $\boldsymbol{\eta}$ に関する誘導形を求め、(1.56) 式に代入すると

$$\mathbf{y}_\$ = \alpha_{y\$} + \Lambda \mathbf{B}_\$ (\alpha_\eta + \Gamma \mathbf{x}_\$ + \mathbf{d}_\$) + \mathbf{e}_y \quad (1.59)$$

数は質問紙調査のデータには必ずといってよいほど含まれている。下位モデルの表現力という観点から、LISCOMP の提案は重要な発展である。

¹³ソフトウェアとしての LISCOMP には Δ という $\mathbf{y}_\$$ の尺度を調節する母数があるが、本章の内容とは直接は関係ないので省略する。

¹⁴ $\mathbf{x}_\$$ によって条件づけられた $\mathbf{y}_\$$ の期待値と分散といういい方もできる

を得る。この期待値

$$\begin{aligned} \mu_{y\$}(\boldsymbol{\theta}) &= E[\mathbf{y}_\$] = E[\alpha_{y\$} + \Lambda \mathbf{B}_\$ (\alpha_\eta + \Gamma \mathbf{x}_\$ + \mathbf{d}_\$) + \mathbf{e}_y] \\ &= E[\alpha_{y\$}] + \Lambda \mathbf{B}_\$ (E[\alpha_\eta] + \Gamma E[\mathbf{x}_\$] + E[\mathbf{d}_\$]) + E[\mathbf{e}_y] \\ &\quad [\alpha_{y\$}, \alpha_\eta, \mathbf{x}_\$ \text{ は定数であり, } E[\mathbf{d}_\$] = \mathbf{o}, E[\mathbf{e}_y] = \mathbf{o} \text{ だから}] \\ &= \alpha_{y\$} + \Lambda \mathbf{B}_\$ (\alpha_\eta + \Gamma \mathbf{x}_\$) \end{aligned} \quad (1.60)$$

が観測変数の期待値構造である。共分散構造は

$$\Sigma(\boldsymbol{\theta}) = V[\mathbf{y}_\$] = E[(\mathbf{y}_\$ - E[\mathbf{y}_\$])(\mathbf{y}_\$ - E[\mathbf{y}_\$])'] \quad (1.61)$$

を計算すればよい。(1.61) 式の期待値の中の平均からの偏差は

$$\begin{aligned} \mathbf{y}_\$ - E[\mathbf{y}_\$] &= (\alpha_{y\$} + \Lambda \mathbf{B}_\$ (\alpha_\eta + \Gamma \mathbf{x}_\$ + \mathbf{d}_\$) + \mathbf{e}_y) - (\alpha_{y\$} + \Lambda \mathbf{B}_\$ (\alpha_\eta + \Gamma \mathbf{x}_\$)) \\ &= \Lambda \mathbf{B}_\$ \mathbf{d}_\$ + \mathbf{e}_y \end{aligned} \quad (1.62)$$

であるから、共分散構造 (1.61) 式は

$$\begin{aligned} \Sigma(\boldsymbol{\theta}) &= E[(\Lambda \mathbf{B}_\$ \mathbf{d}_\$ + \mathbf{e}_y)(\Lambda \mathbf{B}_\$ \mathbf{d}_\$ + \mathbf{e}_y)'] \\ &\quad [(1.58) \text{ 式の仮定から } \Sigma_{ey} = \mathbf{O} \text{ だから}] \\ &= \Lambda \mathbf{B}_\$ E[\mathbf{d}_\$ \mathbf{d}'_\$] \mathbf{B}'_\$ \Lambda' + E[\mathbf{e}_y \mathbf{e}'_y] \\ &= \Lambda \mathbf{B}_\$ \Sigma_{d\$} \mathbf{B}'_\$ \Lambda' + \Sigma_{ey} \end{aligned} \quad (1.63)$$

のように導かれる。

$\mathbf{x}_\$$ を定数と見なし、 $\mathbf{y}_\$$ の期待値と分散によって母数を推定するという考え方には、従来の SEM にはなかった。定数として扱うことによって $\mathbf{x}_\$$ は単なる計画行列とみなせる。このため $\mathbf{x}_\$$ の分布型は何であってもかまわず、後の章に登場する適用例で示すように、明らかに歪みが観察される変数もモデルに組込むことができる。また、このアイデアによって SEM は、線形モデルを明示的に下位モデルに組み入れたことになる。

1.7 モデルの指定例

入門編では、全てのモデルを RAM によって表現してきた。本節ではそこで登場した 6 つのモデルを選び、EQS と LISREL で表現し、方程式モデルの相

違を読者に実感してもらう。LISRELには、通常の変数の取り扱いをしている限りにおいては、3つの限界があった。その中で、限界1は観測変数と構成概念を1対1対応させることによって、簡単に克服できる。(1.53)式と(1.54)式で明らかのように、限界2と限界3は、理論的には克服可能であることが分かっているが、煩雑なので(工夫している間にRAMやEQSを使ったほうが早いので)省略する。また平均構造は、ここでは考えないものとする。

1.7.1 逐次モデル

図1.3の逐次モデルは、EQSを用いて以下のように表現される。

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = \begin{bmatrix} O_{2 \times 2} & O_{2 \times 2} & 0 \\ O_{2 \times 2} & 0 & 0 \\ O_{3 \times 2} & \alpha_{43} & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} + \begin{bmatrix} I_{2 \times 2} & & O_{2 \times 2} \\ \alpha_{31} & \alpha_{32} & O_{2 \times 2} \\ 0 & \alpha_{42} & I_{2 \times 2} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ e_3 \\ e_4 \end{bmatrix}$$

$$\Sigma_u = \begin{bmatrix} \sigma_1^2 & & & \\ \sigma_{21} & \sigma_2^2 & & \\ 0 & 0 & \sigma_{e3}^2 & \\ 0 & 0 & 0 & \sigma_{e4}^2 \end{bmatrix}$$

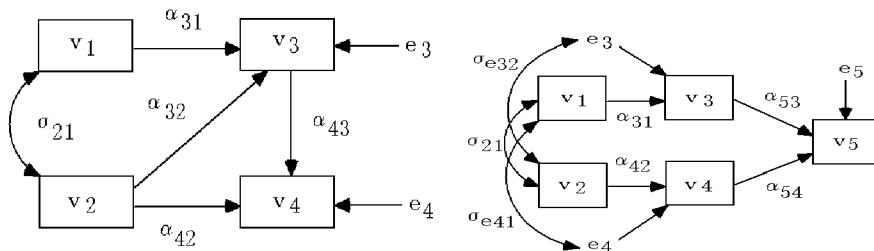


図1.3: 逐次モデル

限界1にもかかわらずLISRELを用いると逐次モデルは以下のように表現できる。

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ 0 & \gamma_{22} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} d_{\$1} \\ d_{\$2} \end{bmatrix}$$

$$\begin{bmatrix} x_{\$1} \\ x_{\$2} \end{bmatrix} = I_{2 \times 2} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} e_{x1} \\ e_{x2} \end{bmatrix}, \quad \begin{bmatrix} y_{\$1} \\ y_{\$2} \end{bmatrix} = I_{2 \times 2} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} e_{y1} \\ e_{y2} \end{bmatrix}$$

$$\Sigma_\xi = \begin{bmatrix} \sigma_{\xi_1}^2 & & \\ \sigma_{\xi_21} & \sigma_{\xi_2}^2 & \end{bmatrix}, \quad \Sigma_{d\$} = \begin{bmatrix} \sigma_{d\$1}^2 & & \\ 0 & \sigma_{d\$2}^2 & \end{bmatrix}, \quad \Sigma_{ex} = O, \quad \Sigma_{ey} = O$$

1.7.2 誤差相関モデル

図1.4の誤差相関モデルは、EQSを用いて以下のように表現される。限界3のためにLISRELの表現は省略する。

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} = \begin{bmatrix} O_{2 \times 2} & & O_{2 \times 3} & & 0 \\ O_{3 \times 2} & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & \\ \alpha_{53} & \alpha_{54} & 0 & 0 & \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} + \begin{bmatrix} I_{2 \times 2} & & O_{2 \times 3} & & \\ \alpha_{31} & 0 & & & \\ 0 & \alpha_{42} & I_{3 \times 3} & & \\ 0 & 0 & 0 & & \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix}$$

$$\Sigma_u = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & 0 & \sigma_{1e4} & 0 \\ \sigma_{21} & \sigma_2^2 & \sigma_{2e3} & 0 & 0 \\ 0 & \sigma_{e32} & \sigma_{e3}^2 & 0 & 0 \\ \sigma_{e41} & 0 & 0 & \sigma_{e4}^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{e5}^2 \end{bmatrix} \quad (1.64)$$

1.7.3 2次因子分析

図1.5の2次因子分析モデルは、EQSを用いて以下のように表現される。

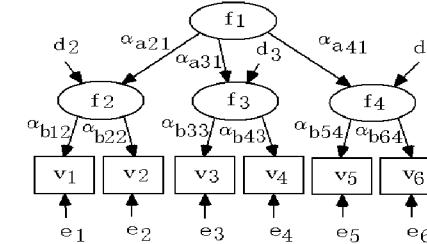


図1.5: 2次因子分析

$$\begin{bmatrix} f_2 \\ f_3 \\ f_4 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \alpha_{b12} & 0 & 0 \\ \alpha_{b22} & 0 & 0 \\ 0 & \alpha_{b33} & 0 \\ 0 & \alpha_{b43} & 0 \\ 0 & 0 & \alpha_{b54} \\ 0 & 0 & \alpha_{b64} \end{bmatrix} \begin{bmatrix} f_2 \\ f_3 \\ f_4 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix} + \begin{bmatrix} \alpha_{a21} \\ \alpha_{a31} \\ \alpha_{a41} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{bmatrix} \quad I_{9 \times 9}$$

$$\Sigma_u = \begin{bmatrix} \sigma_{f1}^2 & & & & & & & \\ 0 & \sigma_{d2}^2 & & & & & & \\ 0 & 0 & \sigma_{d3}^2 & & & & & \\ 0 & 0 & 0 & \sigma_{d4}^2 & & & & \\ 0 & 0 & 0 & 0 & \sigma_{e1}^2 & & & \\ 0 & 0 & 0 & 0 & 0 & \sigma_{e2}^2 & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{e3}^2 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{e4}^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{e5}^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{e6}^2 \end{bmatrix}$$

LISREL を用いると、以下のように表現される。

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} = O \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} + \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \\ \gamma_{31} \end{bmatrix} [\xi_1] + \begin{bmatrix} d_{\$1} \\ d_{\$2} \\ d_{\$3} \end{bmatrix}$$

$$\begin{bmatrix} y_{\$1} \\ y_{\$2} \\ y_{\$3} \\ y_{\$4} \\ y_{\$5} \\ y_{\$6} \end{bmatrix} = \begin{bmatrix} \lambda_{y11} & 0 & 0 \\ \lambda_{y21} & 0 & 0 \\ 0 & \lambda_{y32} & 0 \\ 0 & \lambda_{y42} & 0 \\ 0 & 0 & \lambda_{y53} \\ 0 & 0 & \lambda_{y63} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} + \begin{bmatrix} e_{y1} \\ e_{y2} \\ e_{y3} \\ e_{y4} \\ e_{y5} \\ e_{y6} \end{bmatrix}$$

$$\Sigma_{d\$} = \begin{bmatrix} \sigma_{d\$1}^2 & 0 & 0 \\ 0 & \sigma_{d\$2}^2 & 0 \\ 0 & 0 & \sigma_{d\$3}^2 \end{bmatrix}, \quad \Sigma_{ey} = \begin{bmatrix} \sigma_{ey1}^2 & \dots & O \\ O & & \sigma_{ey6}^2 \end{bmatrix}, \quad \Sigma_\xi = \sigma_{\xi1}^2$$

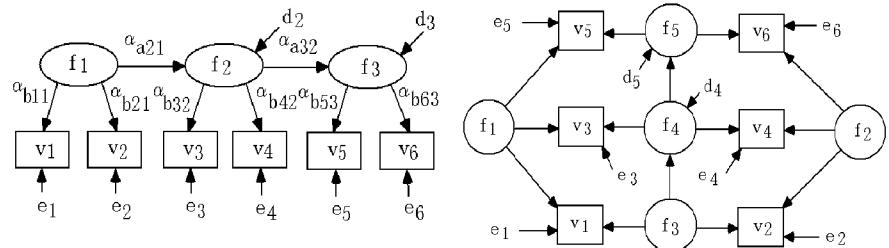


図 1.6: 縦断モデル 1

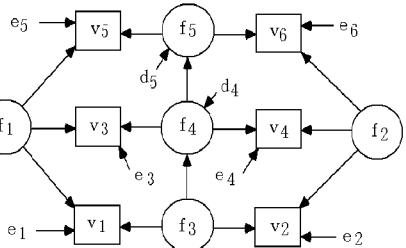


図 1.7: 縦断モデル 2

1.8 問題

上の 2 つのモデルを EQS で表現せよ。また縦断モデル 1 を LISREL で表現せよ。縦断モデル 2 は限界 2 の理由から表現しにくいことを確認せよ。

2 因子分析法

Spearman(1904)¹⁵を嚆矢とする因子分析法 (factor analysis) は、すでに 1 世紀程の歴史を有する数理モデルである。本章では、構造方程式モデルの観点から因子分析モデルを再考し、近年発達したバリエーションの中で、応用的に利用価値の高い方法を紹介する。

2.1 探索的因子分析

因子分析モデルでは、入門編第 4 章で導入したように観測変数 \mathbf{x} を

$$\mathbf{x} = \boldsymbol{\mu}_x + \mathbf{A}\mathbf{f} + \mathbf{e} \quad (2.1)$$

と表現する。 $\boldsymbol{\mu}_x$ は期待値ベクトル、 \mathbf{A} は因子負荷行列、 \mathbf{f} は因子ベクトル、 \mathbf{e} は誤差変数ベクトルである。

$$E[\mathbf{f}] = \mathbf{o}, \quad E[\mathbf{e}] = \mathbf{o}, \quad E[\mathbf{f}\mathbf{e}'] = \mathbf{O} \quad (2.2)$$

を仮定すると、観測変数の共分散構造は

$$\boldsymbol{\Sigma}(\theta) = \mathbf{A}\boldsymbol{\Sigma}_f\mathbf{A}' + \boldsymbol{\Sigma}_e \quad (2.3)$$

となる。因子の分散は、データに対する適合を損なうことなしに 1 に定めることができたから、 $\boldsymbol{\Sigma}_f$ の対角成分は 1 に固定することが可能である。このとき $\boldsymbol{\Sigma}_f$ は因子間相関行列と呼ばれる。

因子間の相関が全て無相関である ($\boldsymbol{\Sigma}_f = \mathbf{I}$) と仮定したモデル

$$\boldsymbol{\Sigma}(\theta) = \mathbf{A}\mathbf{A}' + \boldsymbol{\Sigma}_e \quad (2.4)$$

が用いられることが多く、これを「直交モデル」という。直交モデルに対して (2.3) 式を「斜交モデル」という。また (2.3) 式と (2.4) 式のモデルを合わせて、探索的因子分析モデルという。

¹⁵Spearman, C. (1904). General Intelligence, Objectively Determined and Measured. *American Journal of Psychology*, 15, 201-293.

探索的因子分析の直交モデルは $\boldsymbol{\Sigma}_e$ (対角行列) と因子負荷行列が全て自由母数であるようなモデルである。ただし任意の正規直交行列 $\mathbf{T}(\mathbf{T}'\mathbf{T} = \mathbf{I})$ を用いて

$$\begin{aligned} \boldsymbol{\Sigma}(\theta) &= \mathbf{A}\mathbf{A}' + \boldsymbol{\Sigma}_e = \mathbf{A}\mathbf{T}'\mathbf{T}\mathbf{A}' + \boldsymbol{\Sigma}_e \\ &= \mathbf{B}\mathbf{B}' + \boldsymbol{\Sigma}_e \end{aligned} \quad] \quad (2.5)$$

と変形されてしまうから¹⁶そのままでは識別されない。斜交解は、 $\boldsymbol{\Sigma}_f$ に自由母数が含まれているから、方程式の不定の程度がさらに大きくなり、識別されない。そこで従来は、探索的因子分析を実行する場合に、

$$\mathbf{A}'\mathbf{A} = \mathbf{D} \quad (\text{対角行列}) \quad (2.6)$$

という制約を(ソフトウェアが自動的に)入れて解¹⁷を求めることが多かった。したがって \mathbf{A} が自由母数の集まりであるといつても、個々別々に全く値が自由に定まるものではなく、全体的に制約が入っている。この制約は研究仮説を表現するための積極的な制約ではなく、モデルを識別させるための制約である。

2.1.1 SEM の枠組みでの実行

当初 SEM のソフトウェアで (2.6) 式タイプの制約を容易に入れられるものがなかったために、探索的因子分析は SEM の枠組みで行わないことが多かった。しかし一般的な観点から理論が整備されている SEM の知識を利用しないのはもったいない。そこでここでは探索的因子分析を SEM の枠組みで行う方法と、その意味を考察する。まず (2.6) 式が、方程式何本分の制約に相当するかを考えよう。(2.6) 式は、因子負荷行列を構成する任意の 2 本の列の積和が 0 であることを意味している。したがって制約は変数の組み合せ $(n_f \times (n_f - 1))/2$ 本分

¹⁶ サイズ 1×1 の正規直交行列はスカラーの 1 であり、 $\mathbf{A} = \mathbf{B}$ となるから 1 因子モデルは識別される。

¹⁷ 探索的因子分析を行う場合には、従来は「相関行列の対角要素に共通性の推定値を入れて相関行列の固有値問題を解く」方法が用いられることが多かった。しかし、共通性を推定して固有値問題を解く方法は、計算機が発達していなかった時代の便法であり、かつてのセントロイド法と同様に、すでに時代的使命を終えている。現在では選択する理由が全くない。共通性の推定を行なう方法は統計モデルの推定法として非常に特殊で人為的であり、この方法を習得しても他の統計モデルへの応用が利かない。また初心者への教育という点からは、この方法を理解するためには固有値問題という高度な知識が必要である(最小 2 乗法の習得のためにには固有値の知識が必要ないので因子分析の習得の障壁が非常に低くなる)。そして何よりも統計モデルの推定量の構成法として不正確である。現在では最小 2 乗法・一般化最小 2 乗法・最尤推定法の中から推定法を選ぶのが定石である。

である。因子間相関行列は対角成分がすでに 1 に固定されているし、対称行列であるから、因子負荷行列と同様に $(n_f \times (n_f - 1))/2$ 本分だけ余分に母数が必要である。以上の考察から

直交解 は $(n_f \times (n_f - 1))/2$ 個の制約が不足し、

斜交解 は、因子の分散を(多くの場合 1 に)固定しても、更に $n_f \times (n_f - 1)$ 個の制約が不足し

ていることが分かる。探索的因子分析において最初から斜交解を求めるることは少ないので、ここでは直交解の求め方を紹介する。SEM の枠組みで探索的因子分析の直交解を求めるためには、因子負荷を

$$a_{ij} = 0, \quad \text{ただし } i < j \text{ のとき} \quad (2.7)$$

と固定する方法が、分かり易く、しばしば用いられる。具体的には以下のように $n_f \times (n_f - 1)/2$ 個の母数を固定する。こうすれば(2.6)式と同じ数だけの制約が入り、出力された適合度指標や推測統計的指標を主因子解のものとして解釈することができる、

1 因子解 制約なし

2 因子解 $a_{12} = 0$

3 因子解 $a_{12} = a_{13} = a_{23} = 0$

4 因子解 $a_{12} = a_{13} = a_{23} = a_{14} = a_{24} = a_{34} = 0$

5 因子解 $a_{12} = a_{13} = a_{23} = a_{14} = a_{24} = a_{34} = a_{15} = a_{25} = a_{35} = a_{45} = 0$

6 因子解 $a_{12} = a_{13} = a_{23} = a_{14} = a_{24} = a_{34} = a_{15} = a_{25} = a_{35} = a_{45} = 0,$
 $a_{16} = a_{26} = a_{36} = a_{46} = a_{56} = 0$

2.1.2 集団行動データの分析

SEM の枠組みで行った探索的因子分析の具体例を紹介する。表 2.1 は、表 2.2 で示された 5 件法の項目¹⁸から計算された共分散行列である。「他者評価重視傾

¹⁸調査対象：立教大学社会学部産業関係学科学生、調査時期：1997 年 7 月、調査方法：集合回収調査、有効回収数：345 人

向」「付和雷同傾向」「集団形成傾向」と名づけられた構成概念を測るために 3 つずつ用意された変数である。

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1	1.245								
V2	0.651	1.134							
V3	0.666	0.698	1.117						
V4	0.176	0.074	0.178	1.565					
V5	0.226	0.119	0.113	0.469	1.291				
V6	0.253	0.122	0.182	0.360	0.268	1.574			
V7	0.348	0.201	0.242	0.192	0.196	0.144	0.727		
V8	0.175	0.151	0.129	0.115	0.095	0.013	0.283	0.658	
V9	-0.242	-0.195	-0.151	-0.188	-0.138	0.068	-0.319	-0.319	1.106

表 2.1: 集団行動データの共分散行列

変数	内容
	他者評価重視傾向
V1	人と違う行動を起こしたとき周りの目が気になりますか
V2	自分の噂話が広まることが気になりますか
V3	他人からどう思われているかが気になりますか 付和雷同傾向
V4	仲間につられて授業をサボってしまうことがありますか
V5	授業中に仲間につられて話してしまうことがありますか
V6	友達につられてトイレに付いていくことがありますか 集団形成傾向
V7	自分の意見よりも周囲の意見に合わせて行動するほうですか
V8	自分の意見よりも集団の意見を尊重するほうですか
V9	集団の意見と自分の意見が違っているときに、自分の意見を主張しますか(逆転項目)

表 2.2: 集団行動データの項目内容

表 2.1 のデータを用いて通常の 3 因子の因子分析を(最尤推定法で)行い、主因子解を求めるとき、因子寄与率は、それぞれ 5.119, 1.574, 0.897 であり、累積寄与率は 7.591 となる。因子寄与率は、変数の数 9 で割り、それぞれ 56.87%, 17.49%, 9.97% であり、累積寄与率は 84.33% である。この場合は 3 つの因子によって観測変数の分散の 84.33% が説明されたと解釈する。

因子寄与による説明率の考察は実用的に有効である。しかし「因子が観測変数の分散を説明している程度」という説明率の基準だけを用いると、因子数を適切に選択すれば、どのような相関・共分散行列の因子分析の結果を掲載してもかまわないという間違った認識を誘発する恐れがある。手元の相関行列は、もしかしたら因子分析モデルがあてはまらない相関行列なのではないだろうかという大切な視点を見逃してしまう。

寄与率が低くとも、適合度の良い因子分析の解も存在する。設定した因子が観測変数の分散をあまり説明していないという状態がデータによく当てはまっている場合である。逆にある程度寄与率が高くともモデル全体がデータに当てはまっていない場合もある。現実には因子分析モデルがあてはまらない相関・共分散行列はたくさんあるので、因子分析の結果を論文に掲載する場合には、GFI

等の適合度を計算し、因子分析を行ったことの妥当性を確認することが望ましい。共分散・相関行列には、因子数が 1 ~ 3 の、あるいはそれ以上の因子数の因子分析モデルがあてはまらない可能性もあるということを認識しなくてはいけない。どのような共分散・相関行列にも無条件に因子分析をする(してよい)という、しばしば無批判にうけいれられてしまう研究の定石は、改める必要がある。

n_f	χ^2 値	df	AIC	CAIC	CFI	GFI	AGFI	RMR	RMSEA
1	192.07	27	138.07	7.29	0.725	0.875	0.791	0.143	0.133
2	72.49	19	34.49	-57.54	0.911	0.951	0.885	0.085	0.091
3	16.87	12	-7.13	-65.25	0.992	0.989	0.959	0.028	0.034
4	4.75	6	-7.25	-36.31	1.000	0.997	0.977	0.014	0.000
3'	42.84	24	-5.16	-121.41	0.969	0.972	0.948	0.061	0.048

表 2.3: 集団行動データの探索的因子分析の SEM による適合度

因子寄与率は共通因子がデータの分散を説明している割合であり、適合度指標は因子分析モデルが手元の相関・共分散行列を説明している程度の指標である。このため適合度は、因子寄与率よりも優先して確認すべき指標である。たとえば 20 変数程度のデータに因子分析を実施し、ある因子数で累積寄与率が 0.5 であっても、GFI が 0.55 位では、その因子分析の結果はデータを説明するための候補から外さなくてはならない。

SEM の SW(ソフトウェア)用い、1 因子から 4 因子までのモデルで、表 2.1 の「集団行動データ」を探索的因子分析した結果を表 2.3 に示した(5 行目の $n_f = 3'$ のモデルに関しては後述する)。探索的因子分析の 1 因子から 3 因子までの具体的な解は、それぞれ表 2.4 から表 2.6 に示した((2.7) 式の規則に従ってモデルが指定されていることを確認されたい)。探索的因子分析の 4 因子解は、独自性の分散の推定値の 1 つが負になり、不適解になったので示していない。

3 因子モデル以上であれば、 χ^2 値の観点からは適合が悪くない。AIC は、不適解である 4 因子モデルを勧めているが、保守的な性質のある CAIC は 3 因子モデルを勧めている。また CFI, GFI, AGFI, RMR, RMSEA の値は、3 因子モデルであれば問題ないことを示している。4 因子モデルは不適解なので、ここでは 3 因子モデルを採用する¹⁹。

¹⁹標本数がもっと多く、数千もあるような場合には、 χ^2 値は大きくなりすぎてあてにできない。

	f_1
V1	0.755
V2	0.727
V3	0.751
V4	0.198
V5	0.215
V6	0.199
V7	0.445
V8	0.301
V9	-.292

表 2.4: 1 因子解

	f_1	f_2
V1	0.734	0.000
V2	0.743	0.185
V3	0.780	0.192
V4	0.190	0.221
V5	0.203	0.219
V6	0.188	0.039
V7	0.464	0.521
V8	0.312	0.490
V9	-.298	0.448

表 2.5: 直交 2 因子解

	f_1	f_2	f_3
V1	0.734	0.000	0.000
V2	0.747	0.219	0.000
V3	0.777	0.161	0.078
V4	0.221	-.495	0.193
V5	0.231	-.452	0.166
V6	0.214	-.261	0.274
V7	0.448	-.359	-.332
V8	0.298	-.278	-.497
V9	-.285	0.257	0.443

表 2.6: 直交 3 因子解

	f_1	f_2	f_3
V1	0.667	0.199	0.232
V2	0.766	0.028	0.137
V3	0.780	0.124	0.107
V4	0.030	0.559	0.132
V5	0.052	0.513	0.138
V6	0.126	0.414	-.043
V7	0.222	0.231	0.581
V8	0.096	0.037	0.635
V9	-.099	-.045	-.576

表 2.7: 直交バリマックス解

表 2.6 の 3 因子を初期解としてバリマックス回転した結果を表 2.7 に示した。表 2.6 は通常の因子分析の SW が output する初期解である主因子解とは見かけは異なっている。しかし、同じ因子空間を有しているので、表 2.6 を回転した結果は(バリマックス解に限らず、全ての回転解で) 主因子解を回転した結果と一致する。もちろん表 2.3 の指標は、通常の因子分析の SW が output する主因子解(最尤解)を評価している。表 2.7 を観察すると、当初のもくろみ通りに、変数が 3 つずつに分類されている様子が示されている。

2.1.3 斜交解のすすめ

回転解には、直交解と斜交解がある。従来は、直交解(バリマックス解)が、頻繁に用いられてきた。斜交解は計算時間がかかり、また、しばしば解釈不可能な解を与えることが多かったからである。しかし近年の計算機の進歩によって、計算の手間は問題でなくなった。更にプロマックス解²⁰という優れた斜交回転解が提案され、それが広まることにより、状況は一変しつつある。

表 2.8 には、表 2.7 のプロマックス斜交回転解を示した。斜交解は直交解と比較して因子パターンのコントラストが強く、因子負荷行列の解釈がしやすい。表

また探索的因子分析に関しては経験的に AIC が上手く働かないことが多いことを赤池氏自身が Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52, 317-332. で述べている。著者の経験でも CAIC のほうが適切な判断をしている場合が多い。

²⁰バリマックス解の要素を、それぞれ奇数乗(3 乗が多い)した行列に近づけるように、因子負荷行列を斜交回転する方法

	f_1	f_2	f_3
V1	0.654	0.094	0.097
V2	0.799	-.087	0.005
V3	0.809	0.018	-.048
V4	-.056	0.574	0.043
V5	-.028	0.522	0.053
V6	0.098	0.434	-.142
V7	0.102	0.131	0.556
V8	-.022	-.069	0.672
V9	0.008	0.052	-.605

表 2.8: 斜交プロマックス解

	f_1	f_2	f_3
V1	0.731	0.000	0.000
V2	0.764	0.000	0.000
V3	0.789	0.000	0.000
V4	0.000	0.599	0.000
V5	0.000	0.552	0.000
V6	0.000	0.363	0.000
V7	0.000	0.000	0.714
V8	0.000	0.000	0.586
V9	0.000	0.000	-.534

表 2.9: 確認的因子分析

	f_1	f_2	f_3
f_1	1.000	0.300	0.386
f_2	0.300	1.000	0.353
f_3	0.386	0.353	1.000

表 2.10: 因子間相関(プロマックス解)

	f_1	f_2	f_3
f_1	1.000	0.284	0.459
f_2	0.284	1.000	0.393
f_3	0.459	0.393	1.000

表 2.11: 因子間相関(確認的分析)

表 2.8 と表 2.7 を比較すると、表 2.8 のほうが、0 に近い要素とそうでない要素のコントラストが明らかに大きい。

また斜交解を用いると因子間相関の情報を利用して因子的妥当性の観点から因子数を決定したり、採用した解の考察を進めることができる。社会・人文・行動科学の研究に登場する因子は、互いに相関を持つのが自然である。表 2.10 を見ると、「他者評価重視傾向」「付和雷同傾向」「集団形成傾向」という 3 つの因子の相関はいずれも中程度の正の関係にあることが分かる。

斜交解は因子間相関が 0 に近い場合は直交解とほとんど同じ解になるが、その場合でも、因子間相関を 0 に固定しなくとも 0 に近く推定されているという証拠を積極的に示すために斜交解を用いた方が良い。研究ツールとしてのパリマックス解が与える情報は、プロマックス解が与える情報と比較して見るべき長所はない。プロマックス解の普及と計算機の進歩により、斜交回転解が探索的因子分析の標準的な出力になったといえよう。

2.2 確認的因子分析

確認的因子分析法（検証的因子分析ともいう）では、(2.3) 式の共分散構造に、実質科学的な知見から制約を入れてモデルを特定する。先の節において、斜交解は直交解よりも因子負荷のコントラストを強いと述べたが、確認的因子分析では、それを 1 歩進めて、実質科学的な観点から自由母数・制約母数・固定母数を指定する。表 2.2 の分類に基づいて各因子から 3 本ずつの影響指標を出し、残りは 0 に固定した解を表 2.9 に示す。プロマックス解のコントラストを、ある意味で完全にした解が得られている。また表 2.11 は、確認的因子分析による因子間相関行列である。表 2.10 と表 2.11 の因子間相関は、互いに同様な傾向を示しているが、モデルが異なっているので若干数値は異なっている。

表 2.3 の最下の 3' の行に、確認的因子分析の適合度指標を示した。CFI, GFI, AGFI, RMR, RMSEA はどれも十分に満足できる値である。3 因子の探索的因子分析の解と比較すると、AIC では探索的解が、保守的な CAIC では確認的解がよいとしている。観測変数の数が増えるにしたがって、探索的因子分析モデルよりも、確認的因子分析のほうが自由度が大きく（母数が少なく）なる傾向が顕著になり、モデル構成の観点からも有利になる。科学的研究の 1 つの大切なアプローチは、理論モデルによる予測であるから、データ収集に先立って実質科学的理論がある場合には、是非、確認的因子分析を使用したい。

2.3 2 次（高次）因子分析

2 次因子分析は、通常の因子分析における複数の因子が、更に小数の因子による因子分析で説明されるモデルである。このとき通常の因子分析における因子を 1 次因子といい、その背後に設定される因子を 2 次因子という。本節では 1 次因子を \mathbf{f}_1 と表記し、2 次因子を \mathbf{f}_2 と表記する。1 次因子が 2 次因子によって因子分析モデルで生成されるとすると 1 次因子の共分散（相関）行列は

$$\Sigma_{f1} = \mathbf{A}_2 \Sigma_{f2} \mathbf{A}'_2 + \Sigma_{e2} \quad (2.8)$$

と表現される。この式を (2.3) 式に代入すると、観測変数の共分散構造は

$$\Sigma(\theta) = \mathbf{A}_1 (\mathbf{A}_2 \Sigma_{f2} \mathbf{A}'_2 + \Sigma_{e2}) \mathbf{A}'_1 + \Sigma_{e1} \quad (2.9)$$

と表現される。ここで \mathbf{A}_i は i 次因子の因子負荷行列であり、 Σ_{ei} は i 次の誤差変数の共分散行列である。同様に 2 次因子が、更に小数の 3 次因子で説明されるモデル

$$\Sigma(\theta) = \mathbf{A}_1 (\mathbf{A}_2 (\mathbf{A}_3 \Sigma_{f3} \mathbf{A}'_3 + \Sigma_{e3}) \mathbf{A}'_2 + \Sigma_{e2}) \mathbf{A}'_1 + \Sigma_{e1} \quad (2.10)$$

あるいは、それ以上のモデルも考えることができ、それらをまとめて高次因子分析モデルという。(2.9) 式や (2.10) 式は因子分析モデルが入れ子になっている状態を明示的に示しており²¹、分かり易い。一方、高次因子分析が SEM の 1 つの下位モデルであるという観点からは、入門編で特定したように、統一された RAM の表現のほうが便利である。

2.3.1 1 因子モデルか多因子モデルか

1 次元尺度を構成しようとして項目群を作成したとき、データを集めて因子分析をした場合には 1 因子モデルが当てはまることが期待される。そして実際に当てはまりが良かったとする。しかし同じデータに多因子モデルを当てはめると、それもまた当てはまり、解釈可能である場合も多い。

²¹(2.9) 式の形状から、まず通常の因子分析で斜交解を求めておき、次に因子間相関行列をデータと見なして、再び因子分析を行い、2 次因子分析を行われることもあった。この方法は因子分析のソフトだけで 2 次因子分析が実行できるし、計算時間も少ないので多かったのでソフトやハードの進歩していかなかった時代には便利であった。しかし、因子分析を 2 回繰り返すことは、誤差関数を 2 回評価することであり、誤差が累積して正確な推定量にはならないので、計算機が発達した今日では、どちらかというとアドホックな方法である。

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
V1	1.00										
V2	0.42	1.00									
V3	0.47	0.56	1.00								
V4	0.48	0.56	0.53	1.00							
V5	0.33	0.23	0.24	0.18	1.00						
V6	0.13	0.12	0.14	0.08	0.64	1.00					
V7	0.30	0.30	0.40	0.29	0.46	0.45	1.00				
V8	0.20	0.26	0.33	0.25	0.57	0.55	0.66	1.00			
V9	0.25	0.21	0.18	0.36	0.27	0.24	0.38	0.33	1.00		
V10	0.23	0.23	0.24	0.29	0.34	0.29	0.36	0.42	0.48	1.00	
V11	0.31	0.35	0.46	0.34	0.33	0.28	0.52	0.51	0.44	0.53	1.00
SD	1.9	2.8	1.9	2.0	2.1	2.5	2.3	2.6	1.9	2.4	2.5

表 2.12: 仲間評価データの相関行列と標準偏差

たとえば单一の心理特性(たとえば「集団形成傾向」)を測定するために収集した項目を因子分析すると、1因子モデルの因子負荷は全体的に中程度の数値が得られることが多い。しかし、ためしに3因子までとて解を回転すると(たとえば「他者評価重視傾向」「付和雷同傾向」「集団形成傾向」のように)解釈可能な3因子構造が確認される、というような場合である。

あるいは企業ランキングのための多変量の評価データを因子分析すると(ランキングを計算するために集めたデータであるから)1因子モデルが当てはまるけれども、4因子解を回転したら「社会性」「成長力」「環境重視」「若さ」と名づけられる明快な4因子構造が確認されるなどのケースである。

上記の例の場合は、一般に1因子モデルと多因子モデルの解が、当然のように併記される。しかし1因子モデルは、2因子目以降が独自因子であることを仮定した解であるから、厳密にいうならば、互いに矛盾した提案をしていることになる。この様な状況を記述し、1つのモデルで統一的な説明をするためには、2次因子分析モデルが適している。

2.3.2 2次因子分析の例(仲間評価データの分析)

表 2.12 は Marsh & Hocevar(1985)²² で示された相関行列と標準偏差であり、クラスの中での評価の高さを測定するために集めた 11 の変数から計算されて

²²Marsh, H.W., and Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First-and higher-order factor models and their invariance across groups. *Psychological Bulletin*, 97, 562-582.

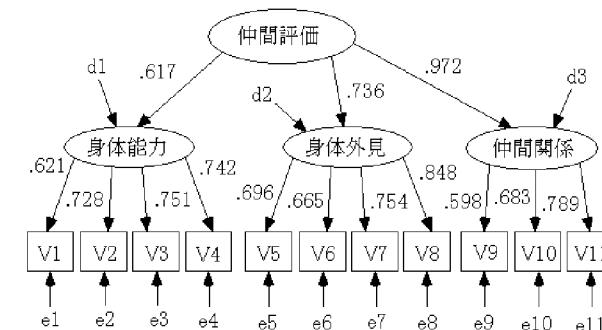


図 2.1: 2 次因子分析

いる。ただし評価の高さは、3つの下位領域を持っており、「身体能力」「身体外見」「仲間関係」であるとしている。このような場合に、従来は、まず1因子解を計算して、各変数の因子負荷が大きいことを確認し、全体として1つの特性を測っている変数であることを示し、続いて3因子解を求めて変数を分類した。しかし、図 2.1 のような2次因子分析モデルで分析することにより、全体として1つの特性を測定し、かつその特性が複数の下位概念から構成されていることを同時に示すことができる。図 2.1 には標準化解を示した。

2.4 独自因子と誤差因子の分離

スピアマンの知能の2因子説では、学力テストの相関行列が一般知能因子と特殊知能因子の2つで説明されることを示した。その後、サーストンは n_f 個の共通因子 f と n 個の独自因子²³ e とで、観測変数の相関行列を説明するモデル(2.1)式を提案し、探索的因子分析の原形を完成させた。ただし独自因子は、実質科学的な観点からは少なくとも2つの成分の和と考えたほうが有効である。2つの成分とは、本来の独自因子と測定誤差である。たとえば表 2.2 の V1, V2, V3 は「他者評価重視傾向」を測定するために用意されているが、その項目独自の安定した測定内容をも同時に有しているはずであり、これが本来の独自因子である。また測定変数の限界の結果としてのバラツキである測定誤差があり、本

²³構造方程式モデル一般では誤差変数と呼ぶが、探索的因子分析モデルの観点からは独自因子と呼ぶことが多い。

來の意味での独自因子と測定誤差は分離されないまま (2.1) 式の誤差項に含まれてしまっている。

この状況は、繰り返しのない 2 因子分散分析モデルの誤差項に、本来の誤差項と交互作用項が分離されないまま含まれているのと同じである。分散分析モデルではセル内で繰り返し測定を行い、誤差項と交互作用項を分離し、両者の分散を分解するのが一般的である。しかし従来、因子分析では独自因子と測定誤差を分離することは、あまりなされてこなかった。本節では、繰り返し測定を行うことによって独自因子と測定誤差を分離する方法を紹介する。

2.4.1 繰り返し測定による分離

同一の観測変数を同一の観測対象から複数回測定すると、その j 番目の測定状況は因子分析モデルを用いて

$$\mathbf{x}_j = \boldsymbol{\mu} + \mathbf{A}\mathbf{f} + \mathbf{e}_j \quad (2.11)$$

のように記述される。 $\boldsymbol{\mu}$ と \mathbf{A} は、変数固有の性質を表現する母数だから添え字 j には影響されない。 \mathbf{f} は観測対象の性質であるから、同様に添え字 j には影響されない。 \mathbf{e}_j は誤差項であるから測定機会 j に依存する。ここで \mathbf{e}_j を

$$\mathbf{e}_j = \mathbf{d}^* + \mathbf{e}_j^* \quad (2.12)$$

のように分解する。 \mathbf{d}^* は観測変数固有の本来の独自成分である。変数の性質であるから、 \mathbf{d}^* は j には依存させない。 \mathbf{e}_j^* は本来の測定誤差である。(2.12) 式を (2.11) 式に代入し

$$\mathbf{x}_j = \boldsymbol{\mu} + \mathbf{A}\mathbf{f} + \mathbf{d}^* + \mathbf{e}_j^* \quad (2.13)$$

を得る。右辺第 2 項と第 3 項を

$$\mathbf{v}^* = \mathbf{A}\mathbf{f} + \mathbf{d}^* \quad (2.14)$$

と置いて、(2.13) 式を

$$\mathbf{x}_j = \boldsymbol{\mu} + \mathbf{I}\mathbf{v}^* + \mathbf{e}_j^* \quad (2.15)$$

と書き換える。すると (2.14) 式と (2.15) 式は、1 次因子の因子負荷が 1 に固定された 2 次因子分析モデルに一致する。 $\boldsymbol{\mu}$ と \mathbf{e}_j^* の分散が j には影響されずに

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	0.910									
X2	0.739	0.945								
X3	0.356	0.318	0.983							
X4	0.316	0.299	0.778	0.999						
X5	0.561	0.556	0.404	0.394	1.262					
X6	0.557	0.549	0.403	0.391	1.061	1.211				
X7	0.195	0.183	0.185	0.172	0.221	0.202	0.977			
X8	0.120	0.129	0.168	0.166	0.208	0.191	0.346	1.007		
X9	0.177	0.217	0.148	0.141	0.398	0.380	0.145	0.110	1.070	
X10	0.202	0.262	0.117	0.126	0.454	0.419	0.140	0.093	0.470	1.075
平均	2.096	2.116	3.103	3.138	2.581	2.569	4.054	4.162	3.538	3.587

表 2.13: 独自因子と誤差因子の分離データの共分散行列

等しいという制約の下では、たった 2 回の繰り返し測定 ($j = 1, 2$) で、このモデルは識別される。

\mathbf{f} と \mathbf{d}^* と \mathbf{e}_j^* が互いに無相関と仮定すると

$$\sigma_{xi} = \text{共通因子の分散} + \sigma_{di*} + \sigma_{ej*} \quad (2.16)$$

が成立つ。両辺を σ_{xi} で割ると、各観測変数が 3 つの分散成分によって説明される割合を考察することができる。SEM の SW は、方程式ごとの決定係数を出力するものが多いから、観測変数 x_i の決定係数

$$R_{xi}^2 = \frac{\text{共通因子の分散} + \sigma_{di*}}{\sigma_{xi}} \quad (2.17)$$

と 1 次因子 v_i^* の決定係数

$$R_{vi*}^2 = \frac{\text{共通因子の分散}}{\text{共通因子の分散} + \sigma_{di*}} \quad (2.18)$$

を利用して、各成分の説明割合を

$$\text{誤差分散の説明割合} = 1 - R_{xi}^2 \quad (2.19)$$

$$\text{独自分散の説明割合} = (1 - R_{vi*}^2) \times R_{xi}^2 \quad (2.20)$$

$$\text{共通性の説明割合} = R_{vi*}^2 \times R_{xi}^2 \quad (2.21)$$

のように求めることができる。

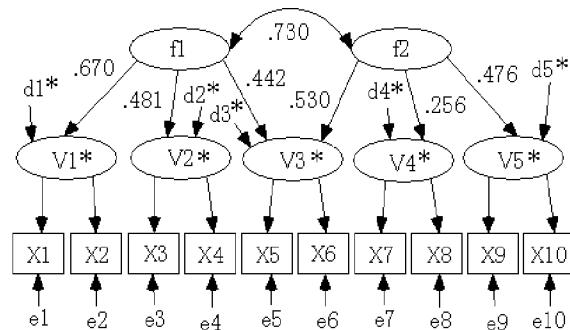


図 2.2: 独自因子と誤差因子の分離

2.4.2 数値例

表 2.13 は、5 つの変数を、それぞれ 2 回ずつ測定(奇数番号の変数が 1 回目の測定であり、それに 1 を足した偶数番号が、その変数の 2 回目の測定)したデータから計算した標本共分散行列と標本平均である。繰り返し測定による分離モデルで分析した結果が図 2.2 である。

制約のついた 5 種類の観測変数の平均の推定値は、最初の変数から順に、2.106, 3.120, 2.575, 4.108, 3.563 であり、決定係数は順に、0.797, 0.785, 0.858, 0.345, 0.437 であった。また 1 次因子の決定係数は順に、0.607, 0.298, 0.772, 0.191, 0.483 であった。以上の推定値から、観測変数の変動を分解すると表 2.14 のようになる。

2 番目の変数と 5 番目の変数は、共通性が 20% 程あることは共通している。通常の因子分析では、ここまでしか考察できない。しかし繰り返し測定による分離モデルで分析すると、前者は独自性の高い変数であり、後者は独自性の低い変数であることがわかる。

1 番目の変数と 2 番目の変数の共通性を比較すると 1 番目の変数の共通性が大きい。通常の因子分析では、ここまでしか考察できない。しかし分離モデルで分析すると、2 番目の変数は 1 番目の変数より独自性が大きく、再検査による測定の安定という観点からは、両者は同程度であることがわかる。

	共通性	独自性	誤差
V1	48.4	31.3	20.3
V2	23.4	55.1	21.5
V3	66.2	19.6	14.2
V4	6.6	27.9	65.5
V5	21.1	22.6	56.3

表 2.14: 3 つの要因による説明割合

2.5 イプサティブモデル

データ解析で一般的に用いられる変数は、集団に対して尺度を考え、測定対象(個人)を尺度上に位置づけることを目的とすることが多い。それに対して、測定対象(個人)毎に尺度を考え、変数を尺度上に位置づける場合があり、その様なデータをイプサティブデータ(ipsative data)という。イプサティブデータが重要となるのは、たとえば

1. 真理・実利・美・博愛・権力・宗教というシュプレンガーの 6 つの価値変数のどれを、どれくらい重要視しているか。
2. 戸外・機械・計算・科学・説得・美術・文芸・音楽・奉仕・書記の 10 の職業興味変数に、どれくらい興味を持っているか。
3. 予習・復習の個人目標を月別に自己評定した個人内評価。

などの場面である。「真理」という価値に対する重視度は、測定対象(個人)間で比較するよりは、むしろ他の 5 つの価値との相対的な重視度を個人内で評価したほうが解釈しやすい。同様に「戸外」に関連した職業に対する興味の強さは、別の生徒と比較するより、他の 9 つの興味との相対的な興味の強弱を比較したほうが、進路決定に役に立つことのほうが多いであろう。

たとえば、A 君が 10 の職業興味変数全てに 5 点の興味を持ち、B 君は、同様に全ての変数に 3 点の興味を持っていたとしよう。この場合、就業できる職業は通常は 1 人 1 つだから A 君のほうが B 君より全体的に興味が強いという解釈はあまり重要ではない。A 君も B 君も 10 の領域にまんべんなく興味を持っているという解釈が、進路指導には役に立つ。イプサティブデータは個人の特徴を記述するのにとても重要であるが、それにもかかわらず多变量の分析は、これまで、あまり行われなかった。それは

- 個人間で比較可能でないスコアであるにもかかわらず、相関係数を計算するためには標準化しなくてはならない
- 標本共分散・相関行列の行列式が 0 になって尤度が計算できなくなる
- 分析を強行しても解釈可能な結果が出ないことが多い

などの理由による。この中で、理由の 1 は共分散行列を分析することによって解決される。本節では、更に、Chan & Bentler (1993)²⁴ のアイデアを利用したイプサティブデータの分析方法を論じる。

2.5.1 イプサティブ変数の共分散行列

観測対象内の相対的な関係を記述しているイプサティブな変数 (\mathbf{x}_{ps} と表記する) は、変数の和が一定 c であるという状況が一般的である。たとえば 6 つの価値変数の合計は、どの人も 6 点ということにしておけば、 $c = 6$ である。ただし合計点が決まっている変数は全ての変数から c/n を引けば(先の例では各価値変数から 1 を引けば)合計点は常に 0 点となる。ここでは、合計点が 0 点に調整されているイプサティブな変数 \mathbf{x}_{ps} を議論の対象²⁵ とする。このとき \mathbf{x}_{ps} の背後に \mathbf{x} という直接観測することができない通常の(これまで扱ってきた)観測変数を仮定し、 \mathbf{x}_{ps} が

$$\mathbf{x}_{ps} = \mathbf{x} - \mathbf{1}\bar{x} \quad (2.22)$$

のように生成されたものとする。まず通常の変数を測定し、次にその n 個の変数の平均を引いて、イプサティブ変数を生成するという状況である。(2.22) 式に従って生成された変数は

$$\mathbf{1}'\mathbf{x}_{ps} = \mathbf{1}'(\mathbf{x} - \mathbf{1}\bar{x}) = \mathbf{1}'\mathbf{x} - \mathbf{1}'\mathbf{1}\bar{x} = \mathbf{1}'\mathbf{x} - n\bar{x} = 0 \quad (2.23)$$

のように和が 0 であることが容易に確認できる。このようなイプサティブ変数の共分散行列 (Σ_{ps} と表記する) の性質を調べてみよう。 Σ_{ps} の任意の行、たと

²⁴Chan, W., and Bentler, P.M. (1993). The covariance structure analysis of ipsative data. *Sociological Methods and Research*, 22, 2, 214-247.

²⁵したがって全体を比例配分することによってイプサティブな変数となった比率データは議論の対象から外れる。

えば i 行の共分散 σ_{psi} の和を計算すると

$$\begin{aligned} \sum_{j=1}^n \sigma_{psi} &= E[(x_{psi} - \mu_{psi})(x_{ps1} - \mu_{ps1})] + E[(x_{psi} - \mu_{psi})(x_{ps2} - \mu_{ps2})] \\ &\quad + \cdots + E[(x_{psi} - \mu_{psi})(x_{psn} - \mu_{psn})] \\ &= E[(x_{psi} - \mu_{psi}) \sum_{j=1}^n (x_{psj} - \mu_{psj})] = E[(x_{psi} - \mu_{psi})(0 - \sum_{j=1}^n E[x_{psj}])] \\ &= E[(x_{psi} - \mu_{psi})(0 - E[\sum_{j=1}^n x_{psj}])] = E[(x_{psi} - \mu_{psi})0] = 0 \end{aligned} \quad (2.24)$$

である。この関係は全ての行で成立し、対称行列では行で成立する性質は列でも成立するから

$$\Sigma_{ps}\mathbf{1} = \mathbf{o}, \quad \mathbf{1}'\Sigma_{ps} = \mathbf{o}' \quad (2.25)$$

である。この性質が標本共分散行列 \mathbf{S}_{ps} にも成立つ

$$\mathbf{S}_{ps}\mathbf{1} = \mathbf{o}, \quad \mathbf{1}'\mathbf{S}_{ps} = \mathbf{o}' \quad (2.26)$$

ことは同様に容易に証明できる。(2.26) 式の性質から、 \mathbf{S}_{ps} の行列式は 0 になってしまい、尤度が定義できなくなる。個人間で比較可能でないスコアであるにもかかわらず、相関係数を計算するためには標準化しなくてはならないという難点は、共分散行列を分析することによって回避できるが、標本共分散行列の行列式が 0 になって、尤度が計算できなくなるという問題が残る。

2.5.2 イプサティブ変数の平均・共分散構造

イプサティブ変数と通常の観測変数の関係を調べると

$$\begin{aligned} \mathbf{x}_{ps} &= \mathbf{x} - \mathbf{1}\bar{x} && [\bar{x} = n^{-1}\mathbf{1}'\mathbf{x} \text{ だから}] \\ &= \mathbf{x} - \mathbf{1}n^{-1}\mathbf{1}'\mathbf{x} = \mathbf{x} - n^{-1}\mathbf{1}\mathbf{1}'\mathbf{x} \\ &= (\mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}')\mathbf{x} && [(\mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}') = \mathbf{G}_{ps} \text{ とおいて}] \\ &= \mathbf{G}_{ps}\mathbf{x} \end{aligned} \quad (2.27)$$

であることがわかる。たとえば $n = 2, 3, 4$ である場合に \mathbf{G}_{ps} は、それぞれ

$$\begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix}, \begin{bmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{bmatrix}, \begin{bmatrix} 0.75 & -0.25 & -0.25 & -0.25 \\ -0.25 & 0.75 & -0.25 & -0.25 \\ -0.25 & -0.25 & 0.75 & -0.25 \\ -0.25 & -0.25 & -0.25 & 0.75 \end{bmatrix}$$

である。 \mathbf{G}_{ps} は定数行列であるから。 \mathbf{x}_{ps} は \mathbf{x} の重み付き和であることが示されたことになる。ここまで構造方程式モデルとは無関係なイプサティブ変数の一般的性質である。

ここでイプサティブ変数を構造方程式で表現するためには、(1.9) 式を (2.27) 式に代入し、

$$\mathbf{x}_{ps} = \mathbf{G}_{ps} \mathbf{GT}(\boldsymbol{\alpha}_0 + \mathbf{u}) \quad (2.28)$$

を得る。したがってイプサティブ変数の平均構造は

$$\boldsymbol{\mu}_{ps}(\boldsymbol{\theta}) = E[\mathbf{x}_{ps}] = \mathbf{G}_{ps} \mathbf{GT} \boldsymbol{\alpha}_0 = \mathbf{G}_{ps} \boldsymbol{\mu}(\boldsymbol{\theta}) \quad (2.29)$$

である。共分散構造は

$$\begin{aligned} \boldsymbol{\Sigma}_{ps}(\boldsymbol{\theta}) &= E[(\mathbf{x}_{ps} - E[\mathbf{x}_{ps}])(\mathbf{x}_{ps} - E[\mathbf{x}_{ps}])'] \\ &= E[\mathbf{G}_{ps} \mathbf{GT} \mathbf{uu}' \mathbf{T}' \mathbf{G}' \mathbf{G}_{ps}] = \mathbf{G}_{ps} \mathbf{GT} \boldsymbol{\Sigma}_u \mathbf{T}' \mathbf{G}' \mathbf{G}_{ps} \\ &= \mathbf{G}_{ps} \boldsymbol{\Sigma}(\boldsymbol{\theta}) \mathbf{G}_{ps} \end{aligned} \quad (2.30)$$

と導かれる。

以上の考察から、イプサティブ変数の平均構造は、通常の観測変数の平均構造に左から \mathbf{G}_{ps} かけた形式で表現され、イプサティブ変数の共分散構造は通常の観測変数の共分散構造の両側から \mathbf{G}_{ps} かけた形式で表現されることが示された。この性質はとても重要である。何故ならば、変換行列 \mathbf{G}_{ps} を使用する限りにおいて、イプサティブ変数の平均・共分散構造モデルは、原則的に通常の変数と同じ扱いができることが示されたからである。ただし、問題が 2 つある。

第 1 に、標本共分散行列 \mathbf{S}_{ps} の行列式の値が 0 になり、計算ができなくなる問題があった。この問題に対処するためには、イプサティブ変数の中から変数を 1 つ取り除いた（本書では最後の n 番目の変数を取り除くことにする）変数を、 \mathbf{S}_{ps}^* と定義し、通常の変数との関係を

$$\mathbf{x}_{ps}^* = \mathbf{G}_{ps}^* \mathbf{x} \quad (2.31)$$

と表現する。ここで \mathbf{G}_{ps}^* は \mathbf{x}_{ps} から削除した変数に対応する行を \mathbf{G}_{ps} から削除したサイズ $(n-1) \times n$ の行列である。右辺に \mathbf{x} がそのまま残っていることが重要である。平均・共分散構造は、 \mathbf{G}_{ps} を \mathbf{G}_{ps}^* に代えただけであるから、それぞれ

$$\boldsymbol{\mu}_{ps}^*(\boldsymbol{\theta}) = \mathbf{G}_{ps}^* \boldsymbol{\mu}(\boldsymbol{\theta}) \quad (2.32)$$

$$\boldsymbol{\Sigma}_{ps}^*(\boldsymbol{\theta}) = \mathbf{G}_{ps}^* \boldsymbol{\Sigma}(\boldsymbol{\theta}) \mathbf{G}_{ps}^{*\prime} \quad (2.33)$$

と表現される。(2.32) 式、(2.33) 式中の母数は、 $n-1$ 個のイプサティブ変数から計算された標本平均・標本共分散行列から推定することができ、この場合は標本共分散行列の行列式の値は一般的に 0 にはならないので、最尤推定法を実行することができる。興味深いことは、 $\boldsymbol{\theta}$ の中には、取り除いたイプサティブ変数に関する母数（たとえば因子負荷など）も含まれているということである。取り除いた変数に関する母数が推定できるというのは妙なものである。しかしイプサティブ変数は、そもそも $n-1$ の変数が手元にあれば残りの 1 つの変数は、言い当てることができるという性質があるのだから、1 つまでなら削除しても情報の損失はない（というよりも $n-1$ 個分の情報しか、もともとないといったほうが正確である）。

第 2 に識別問題がある。(2.33) 式左辺の $n-1$ 個のイプサティブ変数の共分散行列には、 $n(n-1)/2$ 個の分散・共分散があり、それが連立方程式の数になる。したがって通常の変数よりも識別条件が厳しくなり、 n 個のイプサティブ変数間の関係を記述するためには、 $n(n-1)/2$ 個以上の自由母数を使用しないことが必要条件となる。残念ながら広範囲なモデルの識別を簡単に判別する十分条件ではなく、識別されるモデルは経験と勘に頼って探さなければならない。しかし、通常の変数であれば当然識別されるであろうモデルが、イプサティブ変数の場合は識別されないことも少なくなく、通常のモデルに対する経験と勘が通じないこともある。

2.5.3 2 因子データの分析例

本節を含めて以下 3 つ節でイプサティブモデルの分析例を示す。ただしここでは因子分析を主体とし、平均構造は考えないものとする。共分散構造は、通

常 $\Sigma(\theta) = \mathbf{A}\Sigma_f\mathbf{A}' + \Sigma_e$ であるが、ここでは

$$\Sigma(\theta) = \mathbf{F}\Phi\mathbf{F}' = [\mathbf{AI}] \begin{bmatrix} \Sigma_f & \mathbf{O} \\ \mathbf{O} & \Sigma_f \end{bmatrix} \begin{bmatrix} \mathbf{A}' \\ \mathbf{I} \end{bmatrix} \quad (2.34)$$

と表現する。従って(2.33)式を考慮するとイプサティブ変数の因子分析モデルは

$$\Sigma_{ps}^*(\theta) = \mathbf{G}_{ps}^* \Sigma(\theta) \mathbf{G}_{ps}^{*\prime} = \mathbf{G}_{ps}^* \mathbf{F}\Phi\mathbf{F}' \mathbf{G}_{ps}^{*\prime} \quad (2.35)$$

と表現²⁶される。

表2.14は、2種類の特性を4種類の方法で測り、イプサティブ変数に変換した状態をシミュレートしたデータから計算した標本共分散行列である。2種類の特性はV1, V2, V3, V4とV5, V6, V7, V8とで測られており、V1, V5とV2, V6とV3, V7とV4, V8は、それぞれ同一の測定方法を用いた。

	V1	V2	V3	V4	V5	V6	V7	V8
V1	0.514							
V2	0.119	0.615						
V3	0.069	0.051	0.651					
V4	0.065	-0.007	-0.040	0.651				
V5	-0.207	-0.210	-0.186	-0.179	0.536			
V6	-0.156	-0.190	-0.190	-0.157	0.120	0.553		
V7	-0.222	-0.186	-0.176	-0.164	0.109	-0.013	0.679	
V8	-0.182	-0.192	-0.179	-0.169	0.017	0.033	-0.027	0.697

表2.15: イプサティブデータの共分散行列

以上の事前情報に照らして表2.14を従来の常識で観察すると、測定に成功してると、とてもいえない。何故ならばV1, V2, V3, V4の変数内の標本共分散が標本分散に比べて非常に小さい(相関が低い)し、V5, V6, V7, V8の変数内の

²⁶このように $\mathbf{F}_1 \mathbf{F}_2 \cdots \mathbf{F}_p \Phi \mathbf{F}_p \mathbf{F}_{p-1} \cdots \mathbf{F}_1$ の形式で、共分散構造を表現するのが McDonald (1978) の COSAN モデルである。

標本共分散も標本分散に比べて非常に小さい(相関が低い)からである。しかし

$$\mathbf{G}_{ps}^* = \begin{bmatrix} 0.875 & -0.125 & -0.125 & -0.125 & -0.125 & -0.125 & -0.125 & -0.125 \\ -0.125 & 0.875 & -0.125 & -0.125 & -0.125 & -0.125 & -0.125 & -0.125 \\ -0.125 & -0.125 & 0.875 & -0.125 & -0.125 & -0.125 & -0.125 & -0.125 \\ -0.125 & -0.125 & -0.125 & 0.875 & -0.125 & -0.125 & -0.125 & -0.125 \\ -0.125 & -0.125 & -0.125 & -0.125 & 0.875 & -0.125 & -0.125 & -0.125 \\ -0.125 & -0.125 & -0.125 & -0.125 & -0.125 & 0.875 & -0.125 & -0.125 \\ -0.125 & -0.125 & -0.125 & -0.125 & -0.125 & -0.125 & 0.875 & -0.125 \end{bmatrix}$$

を用い、同じ測定法の変数の因子負荷と独自分散は等しいという制約の下で確認的因子分析の解を求める

$$\hat{\mathbf{A}}' = \begin{bmatrix} 0.624 & 0.468 & 0.421 & 0.338 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.624 & 0.468 & 0.421 & 0.338 \end{bmatrix}$$

となる。因子間相関の推定値は-.070であり、ほぼ無相関であった。因子負荷は通常の心理検査で観察される程度の値であり、測定に失敗しているわけではない。また適合度は GFI=0.991, AGFI=0.987, CFI=1.000, RMSEA=0.000と極めて良く、モデルはデータの振舞いを説明している。イプサティブ変数の標本共分散行列の観察は、従来の常識が通用しないことの、1つの具体例である。

2.5.4 1因子データの分析例

イプサティブ変数の解析の難しさは、標本共分散行列の観察からの知見が役に立たないばかりではない。通常では、容易に識別されるモデルが識別されなくなることがある、モデル構成のカンが通用しない場合も少なくない。

たとえば表2.16は、1因子モデルに基づいて発生させた人工データからイプサティブスコアを計算し、それから計算した標本共分散行列である。計画行列として

$$\mathbf{G}_{ps}^* = \begin{bmatrix} 0.8333 & -0.1667 & -0.1667 & -0.1667 & -0.1667 & -0.1667 \\ -0.1667 & 0.8333 & -0.1667 & -0.1667 & -0.1667 & -0.1667 \\ -0.1667 & -0.1667 & 0.8333 & -0.1667 & -0.1667 & -0.1667 \\ -0.1667 & -0.1667 & -0.1667 & 0.8333 & -0.1667 & -0.1667 \\ -0.1667 & -0.1667 & -0.1667 & -0.1667 & 0.8333 & -0.1667 \end{bmatrix}$$

	V1	V2	V3	V4	V5	V6
V1	0.302					<i>sym</i>
V2	-0.083	0.423				
V3	-0.069	-0.097	0.483			
V4	-0.075	-0.110	-0.146	0.554		
V5	-0.028	-0.053	-0.077	-0.097	0.327	
V6	-0.047	-0.079	-0.093	-0.124	-0.072	0.415

表 2.16: 1 因子性の高いイプサティブデータの共分散行列

を用いて、共分散構造モデルを構成することができるはずである。しかしここで 1 因子モデルを用いて解析しても、まともな解は得られない。識別されず、初期値に依存して解釈不能な解がいくらでも得られたらし、初期値によっては発散して非常識な解に到達する場合もあった。制約母数を幾つか導入しても、あまり効果が見られなかった。 5×5 の標本共分散行列を 1 因子モデルで分析して(不適解はデータによって生じることはあっても)発散が観察されるというのは、常識的にも考えにくいことであり、イプサティブ変数の扱いの難しさを物語っている。

ただし、この現象は以下のように説明できる。元々の変数が 1 因子構造であるとすると、各被験者の 6 つの変数の平均値²⁷はその被験者の共通因子スコアの良い推定値である。このため 1 因子構造が明確なデータの場合は、各変数から 6 つの変数の平均値を引いてイプサティブ変数を作成した段階で、最も重要な共通因子の影響がデータ行列から取り除かれてしまう。表 2.16 の標本共分散行列には、そもそも誤差変数の情報しか残っていない。このため \mathbf{G}_{ps}^* で補正しても 1 因子モデルの分析に失敗したのである。共通因子が 2 以上あるばあいには、イプサティブ変数を作成した段階でそれらの主要な変動が除去されてしまうことはない。

2.5.5 通常の変数と混ざったモデルの提案

イプサティブ変数と通常の変数の関係を考察し、簡略な共分散構造を導いた Chan & Bentler (1993) のアイデアは卓抜している。しかし構造方程式モデル

²⁷通常の平均値ではなく、各被験者に関して 6 つの変数の値をたして 6 で割った値

を構築できるほどの変量数のイプサティブ変数は、実際にはあまり見かけないので使用できる場面は少ない。そもそも標本共分散行列の目視による考察ができるなかつたり、識別モデルに関する従来の常識が通用しにくくては、実用的とはいえない。

応用場面では、通常の変数の中に少しだけ数の少ないイプサティブ変数が含まれることが多いであろう。さらに変量数の少ない数種のイプサティブ変数が同時に登場するモデル(たとえば年齢と価値態度変数から職業興味を説明するモデル)を扱えれば実用的であろう。本節では、通常の変数と数種のイプサティブ変数を同時に扱う母数配置を導く。

まず観測変数 \mathbf{x} が $m + 1$ 個の部分ベクトル \mathbf{x}_l

$$\mathbf{x} = (\mathbf{x}'_0 \mathbf{x}'_1 \cdots \mathbf{x}'_l \cdots \mathbf{x}'_m)' \quad (2.36)$$

に分割されているものとする。そして \mathbf{x}_0 以外の m 個の観測変数ベクトルは、それぞれのベクトル内でイプサティブ変数に変換され、その結果

$$\mathbf{x}_{ps} = (\mathbf{x}'_0 \mathbf{x}'_{ps1} \cdots \mathbf{x}'_{psl} \cdots \mathbf{x}'_{psm})' \quad (2.37)$$

だけが観察されているものとする。このとき \mathbf{x}_0 以外の m 個のイプサティブ変数それぞれから、任意の 1 つの変数を取り除いたイプサティブ変数を \mathbf{x}_{psl}^* と表記し

$$\mathbf{x}_{ps}^* = (\mathbf{x}'_0 \mathbf{x}_{ps1}^{*'} \cdots \mathbf{x}_{psl}^{*'} \cdots \mathbf{x}_{psm}^{*'})' \quad (2.38)$$

を分析対象とする。サイズ n のベクトル \mathbf{x} とサイズ $n - m$ のベクトル \mathbf{x}_{ps}^* との関係は、(2.31) 式を考慮して

$$\mathbf{x}_{ps}^* = \mathbf{G}_{ps}^{**} \mathbf{x} \quad (2.39)$$

である。ただし \mathbf{G}_{ps}^{**} は

$$\mathbf{G}_{ps}^{**} = \begin{bmatrix} \mathbf{I} & & & & \\ & \mathbf{G}_{ps1}^* & & & \mathbf{O} \\ & & \ddots & & \\ & & & \mathbf{G}_{psl}^* & \\ & & & & \mathbf{O} \\ & & & & & \ddots \\ & & & & & & \mathbf{G}_{psm}^* \end{bmatrix} \quad (2.40)$$

である。 \mathbf{G}_{psl}^* は(2.31)式に準じて \mathbf{x}_{psl}^* と \mathbf{x}_l の関係を示した定数行列である。
平均構造と共に分散構造は(2.32)式、(2.33)式に準じて

$$\boldsymbol{\mu}_{ps}^*(\theta) = \mathbf{G}_{ps}^{**} \boldsymbol{\mu}(\theta) \quad (2.41)$$

$$\boldsymbol{\Sigma}_{ps}^*(\theta) = \mathbf{G}_{ps}^{**} \boldsymbol{\Sigma}(\theta) \mathbf{G}_{ps}^{**'} \quad (2.42)$$

と表現される。通常の変数とイプサティブ変数を同時に分析することによって、モデル構成時の識別問題に対して、経験と勘を活かせることが期待できる。

	V1	V2	V3	V4	V5	V6	V7	V8
V1	0.980							
V2	0.343	1.024						sym
V3	0.575	0.414	1.035					
V4	0.270	0.146	0.326	1.062				
V5	0.148	0.095	0.181	0.434	0.956			
V6	-0.024	-0.043	-0.022	0.173	0.126	0.378		
V7	-0.152	-0.076	-0.133	0.015	0.054	-0.140	0.476	
V8	0.176	0.119	0.155	-0.189	-0.180	-0.238	-0.336	0.574

表 2.17: 一部がイプサティブな共分散行列

表 2.17 は、通常の変数 V1, V2, V3, V4, V5 とイプサティブ変数 V6, V7, V8 とが混ざったデータ行列(イプサティブ変数は一組だから $m = 1$)から計算した共分散行列である。V1, V2, V3, V8 の変数で 1 つの特性を、V4, V5, V6, V7 の変数で別の 1 つの特性を測定している。通常の変数内では、変数が V1, V2, V3 と V4, V5 とに分類される状態が観察されるが、それ以外の部分は目視での考察は無理である。またイプサティブ変数 V6, V7, V8 内の標本共分散行列は、列和と行和がすべて 0 である²⁸ ことが確認される。そこで、上記の仮説を表現した 2 因子の確認的因子分析モデルで分析を行う。変数 V8 を削除し、計画行列としては、

$$\mathbf{G}_{ps}^{**} = \begin{bmatrix} \mathbf{I}_{5 \times 5} & \mathbf{O}_{5 \times 3} \\ & 0.6667 & -0.3333 & -0.3333 \\ \mathbf{O}_{2 \times 5} & -0.3333 & 0.6667 & -0.3333 \end{bmatrix} \quad (2.43)$$

²⁸ $0.378 - .140 - .238 = 0, -0.238 - .336 + 0.574 = 0, -0.140 + 0.476 - .336 = 0$

を用いる。特に付加的な制約を設けることもなく、因子負荷行列の推定値は

$$\hat{\mathbf{A}}' = \begin{bmatrix} 0.715 & 0.494 & 0.809 & 0.000 & 0.000 & 0.000 & 0.000 & 0.638 \\ 0.000 & 0.000 & 0.000 & 0.781 & 0.552 & 0.787 & 0.591 & 0.000 \end{bmatrix}$$

のように求まる。因子間相関の推定値は 0.470 であった。因子負荷は通常の心理検査で観察される程度の値であり、測定に失敗しているわけではない。またイプサティブ変数の標本共分散は解釈できなくとも、適合度は GFI=0.992, AGFI=0.986, CFI=1.000, RMSEA=0.000 と極めて良く、モデルはデータの振舞いを説明している。和が一定という制約の入っている変数が、分析したい変数の中に入っていることは珍しくない。その場合、分析を避けるのではなく、また、そのまま組み込むのではなく、本節で紹介した処理を行うと有効な分析ができる場合が少なくないはずである。

SEM の SW は、LISREL, COSAN, Mx など、従来は共分散構造を指定する形式のものが多かったので、使いこなすためにはユーザーの側に線形代数の初步的な知識が必要であった。一方、近年ではユーザーが扱いやすいように、EQS, AMOS など、スカラーの方程式を用いてモデルを指定できる SW が主流になってきた。実際、ほとんどの実用場面でスカラーの方程式によるモデルの指定は便利である。しかし本節で紹介したイプサティブモデルのような例外もある。イプサティブモデルは(2.42)式の形式から明らかのように、共分散構造を行列で指定するのが容易である。一方、同じイプサティブモデルをスカラーの方程式を用いてモデルを指定するのは非常に難しい。以降の章でも、行動遺伝学モデル・多相モデルその他の場面で、共分散構造を行列で指定するのは容易でも、スカラーの方程式を用いてモデルを指定するのは難しいモデルが時々登場する。

3 実験データの解析

構造方程式モデルは、しばしば非実験データの解析手法という観点から論じられる。ひとつの理由は、要因を統制しないデータを分析することが実際に多いからである。しかし要因を統制してデータを収集するか否かは、実験計画あるいは調査計画で扱われる問題であり、数理モデルの性質とは直接関係していない。要因を統制したデータの解析を SEM の枠組みで実行することは可能であるし、重要である。

SEM が非実験データ専用の解析手法と誤解されてしまうもう一つの理由は、分散分析と呼ばれる一群の解析手法を実行する際に SEM の SW を用いることが少なかったことにある。多くの分散分析モデルは、構造方程式モデルの下位モデルとして実行できることは知られているものの、分散分析の実行に特化された扱いやすい SW がたくさん提供されており、わざわざ SEM の SW を用いる必要がなかった。

しかし SEM の枠組みで実験データの解析手法を論じることは、2つの理由から重要である。1つは、SEM の柔軟な表現力を利用して、分析者自らが、既存の実験計画モデルで解析できない新しい分析モデルを構成できることである。この点に関しては、十分に研究が進んでいるとは言い難いが、本章でも幾つかの例を示す。

2つ目の理由は、既存のモデル(1・2元配置、繰り返し測定、分割実験、混合モデル等)を構造方程式で表現することによって、分散分析モデルに対する理解が深まるという利点である。分散分析に限らず、統計モデルには母数の配置とその推定という2つのプロセスがある。オーソドックスな分散分析の教程では、それらが必ずしも明確に分離して解説されていないことが多い。一方 SEM では、モデルの本質部分である母数の配置を的確に表現してやれば、推定部分は共通の原理で実行されるから、モデルの成り立ちを理解しやすい。仮に SEM の SW で分散分析を日常的に実践せずとも、SEM の下位モデルとしての表現を一度確認しておくだけで、統一的な視点からの分散分析の理解が促される。

3.1 1 要因実験

分散分析モデルは、観測可能な連続変数(これを特性値といふ)の平均値が、名義変数から受ける影響を記述したモデルである。特性値に影響を及ぼすと考

えられる多くの原因のうち、その実験で取り上げて調べられるものを「要因」という。要因は「因子」と呼ばれることもある。要因の数が1つのモデルを1要因実験モデル、あるいは一元配置モデル・完全無作為実験モデルといふ。また要因のとるさまざまな状態、あるいは条件を「水準」といふ。水準の数を「水準数」といふ。水準は「処理」と呼ばれることもある。1要因実験モデルは

$$y_{ij} = \mu + a_j + e_{ij} \quad (3.1)$$

と表現される。実験計画の分野では、このようなモデル式を「構造模型」と呼ぶ。 y_{ij} は要因 A の j 番目の水準の i 番目の繰返しにおける特性値である。また、 e_{ij} は y_{ij} に対応する測定誤差である。

要因を表わすときは、通常、ローマ字の大文字 A, B, C, D, E... が使われる。1要因実験では要因を A と表わし、添え字 j は要因 A の水準を表わすこととする。また繰返しを添え字 i で表現する²⁹。 μ は「一般平均」といわれ、すべての水準を込みにしたときの特性値の平均である。 a_j は水準 j の効果と呼ばれる。水準の効果は対応する水準の小文字を用いて表わす。水準 j の効果は、 μ_j を j 番目の水準の平均としたときに

$$a_j = \mu_j - \mu \quad (3.2)$$

と定義される。

ここで分析例を示そう。表 3.1 は亜硫酸ガス³⁰の空気中濃度を特性値とし、東京・池袋で、季節ごとに6日選び、測定した結果である。このデータでは「季節」が実験の要因であり、春、夏、秋、冬が「季節」の水準である。このデータを「亜硫酸ガスデータ」と呼ぶ。たとえば $j = 1$ のとき、春の亜硫酸ガスの濃度の平均 μ_1 が1年を通じた亜硫酸ガスの濃度の平均 μ よりも小さい場合は、春の効果 $a_1 (= \mu_1 - \mu)$ はマイナスの値となる。このとき、春には亜硫酸ガスの濃度が低くなる傾向があると解釈する。

²⁹ 入門編を通じ、ここまででは、添え字 i は変数を、添え字 j は実現値(分散分析の用語では「繰り返し」)を表現してきた。ところが実験データの解析モデルを表現する場合には、変数を表現する添え字が次節から複数必要になる。そこで添え字 i で実現値を表現し、添え字 j で変数を表現し、必要に応じて j, k, l, \dots を変数に割りあてていく。

³⁰ 亜硫酸ガスは、白いスモッグの主役といわれる公害物質である。燃料に含まれる硫黄分が、燃焼によって酸化した際に生じる硫黄酸化物の総称であり、2酸化硫黄がその主成分である。無色であるが、刺激臭をもち、慢性気管支炎をはじめとする呼吸器系疾患の有力な原因となる。このデータは、豊田(1994)違いを見ぬく統計学講談社ブルーバックス、より引用。

季節	春	夏	秋	冬
1	10	8	8	14
2	10	10	8	12
3	9	8	11	11
4	11	10	11	16
5	12	12	14	13
6	11	9	15	12
平均	10.5	9.5	11.2	13.0

表 3.1: 亜硫酸ガスデータ

1要因実験の構造模型は行列とベクトルを使って

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{a}\mathbf{X} + \mathbf{e} \quad (3.3)$$

と表現する。ただし \mathbf{y} は確率変数ではなく、特性値を表現する確率変数 y の実現値を横に並べた観測変数のベクトルである。同様に \mathbf{e} は確率変数ではなく、単一の誤差変数 e の実現値を横に並べたベクトルである。ここでは \mathbf{y} も \mathbf{e} もサイズ 24 の横ベクトルである。 \mathbf{a} は水準の効果を並べた母数行列(ベクトル)である。 $\boldsymbol{\mu}$ は、 μ を横に 24 個並べた横ベクトルである。 \mathbf{X} は 0 と 1 によって構成された計画行列である。

たとえば「亜硫酸ガスデータ」の構造模型は、

$$\mathbf{y} = \begin{bmatrix} y_{11} \cdots y_{61} & y_{12} \cdots y_{62} & y_{13} \cdots y_{63} & y_{14} \cdots y_{64} \end{bmatrix} \quad (3.4)$$

$$\mathbf{a} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \end{bmatrix} \quad (3.5)$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \end{bmatrix} = \begin{bmatrix} 111111 & 000000 & 000000 & 000000 \\ 000000 & 111111 & 000000 & 000000 \\ 000000 & 000000 & 111111 & 000000 \\ 000000 & 000000 & 000000 & 111111 \end{bmatrix} \quad (3.6)$$

$$\mathbf{e} = \begin{bmatrix} e_{11} \cdots e_{61} & e_{12} \cdots e_{62} & e_{13} \cdots e_{63} & e_{14} \cdots e_{64} \end{bmatrix} \quad (3.7)$$

と表現される。計画行列 \mathbf{X} は、4 本の横ベクトル x_1, x_2, x_3, x_4 から構成されている。(3.3) 式から (3.7) 式の表現は (3.1) 式と同等であることを確認していただきたい。ただし、定義の式からも明らかなように水準の効果は

$$0 = \sum_{j=1}^a a_j \quad (3.8)$$

のように、総和が 0 になるという制約の下で母数を推定する。ここで a は、要因 A の水準数である

3.1.1 SEM の枠組みでの実行

一般に (3.3) 式の方程式モデルを扱うためには、以下の 3 つの方法がある。

1. \mathbf{X} と \mathbf{e} に分布を仮定せず、入門編で導入した最小 2 乗の観点から解を求める方法。仮定が最も少ない方法。
2. \mathbf{X} に分布を仮定せず、 \mathbf{e} に分布を仮定し、第 1 章 (1.56) 式、(1.57) 式の観点から解を求める方法。正規分布が仮定されることが多い。
3. \mathbf{X} と \mathbf{e} に分布を仮定し、入門編で導入した最尤推定の観点から解を求める方法。多変量正規分布が仮定されることが多い。

一般線形モデルの観点からは、分散分析モデルは 2 番目の立場で記述される。第 1 章でも論じたように、2 番目の立場での構造方程式モデルの SW も存在しているが、広く流布するにいたってはおらず、1 番目と特に 3 番目が現時点での主流である。ただし \mathbf{X} と \mathbf{e} に相関がなければ、この 3 つの方法は、母数の推定値が一致することが知られている。 \mathbf{X} と \mathbf{e} に相関のある分散分析モデルはないので、解を求めるだけなら 3 つの方法のうち、どれを使ってもよいことになる。また \mathbf{X} と \mathbf{e} に相関がなければ、正規分布を仮定した 2 番目の方法と 3 番目の方法では標準誤差・母数の検定結果が一致することが知られている(1 番目の方法では、標準誤差・検定統計量は求まらない)。0 と 1 だけから構成される計画行列 \mathbf{X} に正規分布を仮定して 3 番目の方法を適用するのは、仮定を明らかに逸脱しており、釈然としない。しかし推定値と標準誤差は 2 番目の方法に一致するから、実用的には問題ない。ここでは SEM の SW の現在の主流である 3 番目の方法で解を求める³¹。

³¹ 2 番目の方法が、容易に実行できる SW が数多く発表されることが望まれる。

ただし分散分析の計画行列 \mathbf{X} のランクは行の数より小さい(行フルランクではない)から、そのままでは尤度が計算できない。そこで計画行列 \mathbf{X} の一部の変数だけを使用し、残りの行は構造方程式内部で生成する。たとえば(3.6)式の「亜硫酸ガスデータ」の計画行列のランクは 3 である。そこで \mathbf{X} は 2 行目から 4 行目まで³²をデータとして使用し(それぞれ $\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ と呼ぶ), 1 行目は

$$\mathbf{f}_1 = \mathbf{1}_{1 \times N} - \mathbf{x}_2 - \mathbf{x}_3 - \mathbf{x}_4 \quad (3.9)$$

という(サイズが 24 で、値がすべて分かっている横ベクトルの)潜在変数を利用する。便宜的な構成概念を用い、(3.3)式を書き換えて、

$$\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \bar{\mathbf{x}}_2 \\ \bar{\mathbf{x}}_3 \\ \bar{\mathbf{x}}_4 \\ \boldsymbol{\mu} \end{bmatrix} + \begin{bmatrix} 0 & -1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ a_1 & a_2 & a_3 & a_4 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \\ \mathbf{y} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{x}_2 - \bar{\mathbf{x}}_2 \\ \mathbf{x}_3 - \bar{\mathbf{x}}_3 \\ \mathbf{x}_4 - \bar{\mathbf{x}}_4 \\ \mathbf{e} \end{bmatrix} \quad (3.10)$$

という式を用いる。この式は、構造方程式モデルの一般式(1.1)式を、実現値で表現したものである。ここで $\bar{\mathbf{x}}_j$ は、 \mathbf{x}_j の平均が $1/4$ であるので、 $1/4$ を横に 24 個並べた横ベクトルとなる。

推定値・母数の検定統計量は表 3.2 に示した通りである。 $\hat{\mu} = 11.042$ は 1 年を通じての亜硫酸ガスの平均である。 $\hat{a}_1 = -0.542$ とは、春の亜硫酸ガスの平均が、年間の平均より 0.542 だけ少ないことを意味する。検定統計量は、その絶対値が 1.98 以上であれば 5% 水準で有意であると判断できるから、夏の亜硫酸ガスが少ない傾向、冬の亜硫酸ガスが多い傾向が有意であると解釈する。通常の 1 要因分散分析モデルでは、「 a_j のうち少なくとも 1 つが 0 でない」が検定仮説であるから、「季節」の主効果は有意であることが分かる。

3.2 2 要因実験

1 つの特性値に対して、影響を与える要因は 1 つとは限らない。同時に 2 つの要因が影響を及ぼすようなデータの解析に利用されるのが「2 要因実験」モ

³²1 行目から 3 行目まででも、何でも、どれでも 3 行ならよい

	a_1	a_2	a_3	a_4	μ	σ_e^2
推定値	-0.542	-1.542	0.125	1.958	11.042	3.160
検定統計量	-0.862	-2.453	0.199	3.116	30.431	3.464

表 3.2: 亜硫酸ガスデータ分析結果

ルである。2 要因実験は「2 因子実験」とか「2 元配置実験」と呼ばれることがある。本節では 2 要因実験のモデルを SEM で表現する。

2 要因実験モデルには、2 つの要因とも固定要因である母数モデルと、固定要因と変量要因が 1 つずつの混合モデルと、2 つの要因とも変量要因である変量モデルの 3 種類のモデルがある。ここでは母数モデルと混合モデルを構造方程式モデルで記述する。

3.2.1 母数モデル

2 要因実験のモデル式は

$$y_{ijk} = \mu + a_j + b_k + (ab)_{jk} + e_{ijk} \quad (3.11)$$

と表現される。添え字 i は実現値(実験計画の言葉では「繰り返し」という)を、添え字 j は要因 A の水準を、添え字 k は要因 B の水準を表す。この構造模型にしたがった実験を「繰り返しのある 2 要因実験」という。 a_j は要因 A の水準 j の効果、 b_k は要因 B の水準 k の効果、 $(ab)_{jk}$ は要因 A と要因 B の交互作用効果である。交互作用の効果には、

$$\sum_{j=1}^a (ab)_{jk} = 0, \quad \sum_{k=1}^b (ab)_{jk} = 0 \quad (3.12)$$

という制約がある。ここで b は要因 B の水準数である。

表 3.3 のデータを用いて分析例を示す。この表は、日本人観光客に人気の高いサンフランシスコとロサンゼルスにあるホテルをいくつか選んで料金を調べたものである。ダウンタウンにあるホテルと、ダウンタウンから 10 キロくらい離れた郊外にあるホテルを 9 軒ずつ無作為に選んで、ツインルーム 1 泊の料金を示した。

平均値を見ると、サンフランシスコではダウンタウンのほうが郊外よりも料

金が高い。しかしロサンゼルスでは、逆に、郊外のホテルのほうがダウンタウンよりも料金が高い³³。

単位：ドル	サンフランシスコ	平均
ダウンタウン	079 107 103 092 180 165 240 265 300	170
郊外	075 060 060 094 119 100 102 125 165	100
	ロサンゼルス	平均
ダウンタウン	095 099 070 116 170 145 205 200 210	145
郊外	153 078 075 092 115 155 250 340 380	182

表 3.3: サンフランシスコとロサンゼルスのホテルの料金

交互作用 $(ab)_{jk}$ は j 番目の都市の k 番目の場所のホテルの料金の全体の平均からのずれであり、 $j \times k$ 個の値を代表してあらわしたものである。添え字 i は亜硫酸ガスの実験のときと同じように繰返しをあらわしている。つまり $n = 9$ である。 (3.3) 式を用いてこのデータの母数を推定するためには

$$\mathbf{y} = \begin{bmatrix} y_{111} \cdots y_{911} & y_{112} \cdots y_{912} & y_{121} \cdots y_{921} & y_{122} \cdots y_{922} \end{bmatrix} \quad (3.13)$$

$$\mathbf{a} = \begin{bmatrix} a_1 & a_2 & b_1 & b_2 & (ab)_{11} & (ab)_{12} & (ab)_{21} & (ab)_{22} \end{bmatrix} \quad (3.14)$$

³³サンフランシスコはチャイナタウン、フィッシャーマンズワーフ、ユニオンスクエア、日本人街、どの見どころに行くにしても、ダウンタウンから 2 キロ以内にある。ゴールデンゲートブリッジは少し離れているけれども 5 キロくらいである。夜景のきれいな場所もダウンタウンに集まっている。サンフランシスコに行った観光客は、必ずしも郊外に出かける必要がない。逆に、郊外のホテルに泊った人はダウンタウンに行かねば面白くない。郊外のホテルが安いのはそのためである。

一方ロサンゼルスの観光の見どころは、ハリウッド、サンタモニカ、ビバリーヒルズ、ディズニーランドなど、どれをとってもみな郊外に位置している。この街のダウンタウンは、観光客にとって、街に入るときと出るときの通過場所という役割が強い。郊外のホテルのほうが料金が高いのにはそのためである。

$$\mathbf{X} = \begin{bmatrix} 111111111 & 111111111 & 000000000 & 000000000 \\ 000000000 & 000000000 & 111111111 & 111111111 \\ 111111111 & 000000000 & 111111111 & 000000000 \\ 000000000 & 111111111 & 000000000 & 111111111 \\ 111111111 & 000000000 & 000000000 & 000000000 \\ 000000000 & 111111111 & 000000000 & 000000000 \\ 000000000 & 000000000 & 111111111 & 000000000 \\ 000000000 & 000000000 & 000000000 & 111111111 \end{bmatrix} \quad (3.15)$$

$$\mathbf{e} = \begin{bmatrix} e_{111} \cdots e_{911} & e_{112} \cdots e_{912} & e_{121} \cdots e_{921} & e_{122} \cdots e_{922} \end{bmatrix} \quad (3.16)$$

と指定する。 (3.12) 式の制約に

$$\sum_{j=1}^a a_j = 0, \quad \sum_{k=1}^b b_k = 0 \quad (3.17)$$

の制約を加えて母数を推定する。

0 と 1 だけから構成される計画行列 \mathbf{X} に正規分布を仮定しても、推定値と標準誤差は \mathbf{X} を非確率変数として扱ったのと同じなので実用的には問題ない。そこで \mathbf{X} と \mathbf{e} に分布を仮定し、最尤推定法で解を求める。

\mathbf{X} のランクは 3 だから 4 行以上使用すると尤度が計算できなくなる。そこで \mathbf{X} には互いに線形独立な 1,3,6 行だけ使い(それぞれ $\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_6$ と呼ぶ), その他の行は

$$\mathbf{f}_2 = \mathbf{1}_{1 \times N} - \mathbf{x}_1 \quad (3.18)$$

$$\mathbf{f}_4 = \mathbf{1}_{1 \times N} - \mathbf{x}_3 \quad (3.19)$$

$$\mathbf{f}_5 = \mathbf{x}_1 - \mathbf{x}_6 \quad (3.20)$$

$$\mathbf{f}_7 = \mathbf{x}_3 - \mathbf{f}_5 \quad (3.21)$$

$$\mathbf{f}_8 = \mathbf{1}_{1 \times N} - \mathbf{f}_5 - \mathbf{x}_6 - \mathbf{f}_7 \quad (3.22)$$

という(値がすべて分かっている)潜在変数を利用する。構造方程式による表現は以下の通りである³⁴。

³⁴式中の \mathbf{a} は (3.14) 式のサイズ 8 の横ベクトルである

$$\begin{bmatrix} x_1 \\ f_2 \\ x_3 \\ f_4 \\ f_5 \\ o \\ x_6 \\ f_7 \\ f_8 \\ y \end{bmatrix} = \begin{bmatrix} \bar{x}_1 \\ 1 \\ \bar{x}_3 \\ 1 \\ f_5 \\ o \\ \bar{x}_6 \\ o \\ 1 \\ \mu \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 \end{bmatrix} + \begin{bmatrix} x_1 \\ f_2 \\ x_3 \\ f_4 \\ f_5 \\ o \\ x_6 \\ f_7 \\ f_8 \\ y \end{bmatrix} + \begin{bmatrix} x_1 - \bar{x}_1 \\ o \\ x_3 - \bar{x}_3 \\ o \\ o \\ x_6 - \bar{x}_6 \\ o \\ o \\ e \end{bmatrix}$$

	a_1	a_2	b_1	b_2	μ	σ_e^2
推定値	-14.36	14.36	8.42	-8.42	149.42	5334.54
検定統計量	-1.18	1.18	0.69	-0.69	12.27	1257.36
	$(ab)_{11}$	$(ab)_{12}$	$(ab)_{21}$	$(ab)_{22}$		
推定値	26.64	-26.64	26.64	-26.64		
検定統計量	2.19	-2.19	2.19	-2.19		

表 3.4: 2 要因分散分析母数の推定値

推定値・母数の検定統計量は表 3.4 に示した通りである。 $\hat{\mu} = 149.42$ は 36 のホテルの料金の平均である。 $\hat{a}_1 = -14.36$ とは、サンフランシスコの平均料金が 14 ドルほど安いことを意味する。 $\hat{b}_1 = 8.42$ とは、ダウンタウンの平均料金が 8 ドルほど高いことを意味する。検定統計量は、その絶対値が 1.98 以上であれば 5% 水準で有意であると判断できるから、交互作用だけが有意であることが確認される。

3.2.2 混合モデル

要因 A を変量要因、要因 B を母数要因とした場合の混合モデルの構造模型は

$$y_{ijk} = \mu + \ddot{a}_j + b_k + (\ddot{ab})_{jk} + e_{ijk} \quad (3.23)$$

である。変量要因の効果はダブルドットの付いた文字 \ddot{a}_j で表現する。変量要因と母数要因の交互作用は一般的に変量効果となるから、 $(\ddot{ab})_{jk}$ は変量効果(確率変数)である。また繰り返しがない場合の混合モデルの構造模型は

$$y_{jk} = \mu + \ddot{a}_j + b_k + e_{jk} \quad (3.24)$$

と表現できる。本節では (3.24) 式の繰り返しのないモデルを SEM の枠組みで表現する。混合モデルの要因 A の添え字 j は母集団からの j 番目の標本(実現値)を表現している。(3.24) 式では、 μ , b_k が母数であり、 \ddot{a}_j , e_{jk} が確率変数である。SEM 流の表記では、実現値の添え字は書かないので、とりあえず

$$y_k = \mu + \ddot{a} + b_k + e_k \quad V[e_k] = \sigma_e^2 \quad (3.25)$$

と書き換えておく。

(3.25) 式を SEM で表現するためには、まず因子負荷が 1 に固定された 1 因子の因子分析モデル

$$y_k = \mu_k + f_a + e_k \quad (3.26)$$

を導入し、そして因子を、さらに

$$f_a = \mu + d_a \quad (3.27)$$

と表現する。その上で

SEM による表記	2 要因混合モデルの表記
変数の期待値 μ_k	— 母数要因の水準 k の効果 b_k
因子 f_a の平均 μ_a	— 全平均 μ
因子 f_a の分散 σ_a^2	— 変量要因の分散成分の推定値 $V[\ddot{a}]$

と読み替え、誤差変数 e_k の添え字 k によらず誤差分散 σ_e^2 は等しいという制約を加えると、(3.25) 式の繰り返しがない場合の 2 要因実験混合モデルは、(3.26) 式と (3.27) 式に一致することが分かる。しかも 0-1 要素から構成される計画行列をデータとして扱う必要がないので、検定統計量と解だけでなく、SEM の通常の最尤モデルと完全に一致する。これが母数モデルとの大きな相違である。

表 3.5 は、1988 年にカナダのカルガリーで開催された冬季オリンピックのフィギュアスケート、女子シングルの技術点の採点結果である。縦に 11 人の選手、横に 9 人の審査委員が並んでおり、採点は 6 点満点の減点法で、少数第 1 位まで出している。このデータをスケーターを変量要因 A、審査委員を母数要因 B

	J1	J2	J3	J4	J5	J6	J7	J8	J9	平均
S1	4.8	4.7	4.8	5.1	5.0	4.6	4.9	4.3	4.8	4.78
S2	5.6	5.4	5.2	5.5	5.5	5.6	5.4	5.4	5.5	5.46
S3	5.4	5.5	5.5	5.7	5.6	5.7	5.7	5.1	5.3	5.50
S4	5.0	4.9	4.6	5.0	5.0	4.9	4.8	4.7	4.9	4.87
S5	5.0	5.1	5.0	5.4	5.4	5.3	5.3	5.4	5.2	5.23
S6	5.6	5.5	5.6	5.7	5.6	5.7	5.6	5.6	5.7	5.62
S7	5.7	5.7	5.8	5.6	5.7	5.8	5.6	5.6	5.7	5.69
S8	5.9	5.8	5.9	5.8	5.9	5.9	5.9	5.9	5.9	5.88
S9	5.8	5.8	5.9	5.8	5.7	5.8	5.8	5.9	5.8	5.81
S10	5.5	5.0	5.0	5.6	5.4	5.4	5.4	5.0	5.1	5.27
S11	5.5	5.8	5.6	5.7	5.6	5.7	5.7	5.7	5.7	5.67
平均	5.44	5.38	5.35	5.54	5.49	5.49	5.46	5.33	5.42	5.43

表 3.5: オリンピック女子フィギュアスケートの技術点

とし、繰り返しのない混合モデルを構造方程式モデルで表現すると

$$\begin{bmatrix} f_a \\ y_1 \\ y_2 \\ \vdots \\ y_9 \end{bmatrix} = \begin{bmatrix} \mu \\ \mu_1 \\ \mu_2 \\ \vdots \\ \mu_9 \end{bmatrix} + \begin{bmatrix} 0 & & & & f_a \\ 1 & O_{10 \times 9} & & & d_a \\ & & y_1 & & \\ & & 1 & & e_1 \\ & & \vdots & & e_2 \\ & & 1 & & \vdots \\ & & & & e_9 \end{bmatrix} \quad (3.28)$$

となる。外生変数の共分散行列は

$$\Sigma_u = \begin{bmatrix} \sigma_a^2 & & O & & \\ & \sigma_e^2 & & & \\ & & \sigma_e^2 & & \\ & & & \ddots & \\ O & & & & \sigma_e^2 \end{bmatrix} \quad (3.29)$$

であり、 $\mu_1 + \mu_2 + \dots + \mu_9 = 0$ という制約が入る。

母数要因Bの推定・検定結果を表 3.6 に示す。5%の危険率で、4番目の審査員は点数が甘く、3番目と8番目の審査委員は厳しいと判断できる。他の母

	μ_1	μ_2	μ_3	μ_4	μ_5
推定値	0.0030	-0.0515	-0.0788	0.1030	0.0576
検定統計量	0.0778	-1.3221	-2.0221	2.6442	1.4777
	μ_6	μ_7	μ_8	μ_9	
推定値	0.0576	0.0303	-0.1061	-0.0152	
検定統計量	1.4777	0.7777	-2.7220	-0.3889	

表 3.6: 2 要因分散分析母数の推定値 1

	μ	σ_a^2	σ_e^2
推定値	5.4333	0.1298	0.018788
検定統計量	49.6132	2.201	6.325

表 3.7: 2 要因分散分析母数の推定値 2

数の推定・検定結果を表 3.7 に示した。変量要因Aの分散は有意であり、5%の危険率でスケーターの間には実力差があると判断できる。

3.3 3 要因実験

3 要因実験モデルは、3つの要因が特性値に与える影響を調べる際に用いられる。2要因実験の場合、混合モデルは1種類しかないけれども3要因実験には混合モデルが2種類ある。母数要因が2つで変量要因が1つの場合と、逆に、母数要因が1つで変量要因が2つの場合である。たとえば要因Aが変量要因で要因BとCが母数要因である場合の構造模型は

$$y_{ijkl} = \mu + \ddot{a}_j + b_k + c_l + (\ddot{ab})_{jk} + (bc)_{kl} + (\ddot{ac})_{lj} + (\ddot{abc})_{jkl} + e_{ijkl} \quad (3.30)$$

のようになる。また繰り返しがない場合の混合モデルの構造模型は2次の交互作用と誤差項が区別付かなくなるから

$$y_{jkl} = \mu + \ddot{a}_j + b_k + c_l + (\ddot{ab})_{jk} + (bc)_{kl} + (\ddot{ac})_{lj} + e_{jkl} \quad (3.31)$$

である。変量要因と母数要因の交互作用は変量効果となるから、要因Aと要因Bの交互作用 $(\ddot{ab})_{jk}$ 、要因Aと要因Cの交互作用 $(\ddot{ac})_{lj}$ は変量効果(確率変数)

となり、要因Bと要因Cの交互作用 $(bc)_{kl}$ は固定効果(母数)となる。

本節では、繰り返しのない混合モデル(3.31)式を SEM の枠組みで表現する。混合モデルの要因Aの添え字 j は母集団からの j 番目の標本(実現値)を表現している。SEM 流の表記では、実現値の添え字は書かないことが多いので

$$y_{kl} = \mu + \ddot{a} + b_k + c_l + (\ddot{ab})_k + (bc)_{kl} + (\ddot{ac})_l + \epsilon_{kl} \quad (3.32)$$

と書き換える。ここで因子負荷が 1 に固定され、1 つの観測変数が 3 つの因子から影響を受ける因子分析モデル

$$y_{kl} = \mu_{kl} + f_a + f_{bk} + f_{cl} + \epsilon_{kl} \quad (3.33)$$

を考える、そして因子を

$$f_a = \mu_a + d_a \quad (3.34)$$

$$f_{bk} = \mu_k + d_{bk} \quad (3.35)$$

$$f_{cl} = \mu_l + d_{cl} \quad (3.36)$$

と表現する。

このとき

SEM による表記

変数の期待値 μ_{kl}

因子 f_a の平均 μ_a

因子 f_a の分散 σ_a^2

因子 f_{bk} の平均 μ_{bk}

因子 f_{bk} の分散 σ_b^2

因子 f_{cl} の平均 μ_{cl}

因子 f_{cl} の分散 σ_c^2

混合モデルの表記

– 交互作用 $(bc)_{kl}$

– 全平均 μ

– 変量要因Aの分散成分の推定値 $V[\ddot{a}]$

– 要因Bの主効果 b_k

– 要因AとBとの交互作用の分散成分 $V[(\ddot{ab})_k]$

– 要因Cの主効果 c_l

– 要因AとCとの交互作用の分散成分 $V[(\ddot{ac})_l]$

と読み替え、誤差変数 ϵ_{kl} の添え字 kl によらず誤差分散 σ_e^2 は等しいという制約を加えると、(3.33) 式から (3.36) 式までの方程式モデルは、(3.32) 式の繰り返しがない場合の 2 要因実験の混合モデルに完全に一致することが分かる。

前節では、フィギュアスケートの技術点の分析を行った。しかしフィギュアスケートの公式試合では技術的な観点からばかりでなく、芸術的な観点からも

	J1	J2	J3	J4
S1	4.8	4.7	4.8	5.1
S2	5.6	5.6	5.4	5.4
S3	5.2	5.2	4.8	5.5
S4	4.8	4.5	4.5	5.1
S5	5.0	5.1	4.9	5.3
S6	5.9	5.8	5.9	5.9
S7	5.6	5.0	5.4	5.7
S8	5.8	5.6	5.6	5.7
S9	5.9	5.7	5.8	5.9
S10	5.2	4.5	4.7	5.4
S11	5.7	5.5	5.5	5.7
平均	5.41	5.20	5.21	5.52

表 3.8: オリンピック女子フィギュアスケート芸術点

演技が採点され、その合計点で順位が決まる。表 3.8 のデータは、前出のカルガリーオリンピックのフィギュアスケートの芸術点である。ただし表 3.5 の審査委員のうち最初の 4 人分の採点結果だけを示している。スケーターは表 3.5 に示したとおり 11 人である。

表 3.5 と表 3.8 を合わせてみると、1 つのデータを特定するためには、「スケーター」と「審査委員」を指定するだけではなくて、その点が技術点なのか芸術点なのかを指定しなければならない。そこで要因Aとも要因Bともクロスする 3 つ目の要因Cとして「観点」という新たな要因を考える。この場合、要因Cの水準は「技術」と「芸術」の 2 つである。また技術点も最初の 4 人の審査委員の結果だけを用いることとし、要因Bの水準は 4 とする。

3 要因実験でも、3 つの要因の水準の組み合わせで指定される条件を「セル」という。そして各セルについて 2 つ以上の標本がある場合は「繰返しのある実験」という。ここでは「スケーター」と「審査委員」と「観点」の組み合わせでは 1 つの得点しかないので「繰返しのない 3 要因実験」である。

この場合の構造方程式モデルは

$$\begin{bmatrix} f_a \\ f_{b1} \\ f_{b2} \\ f_{b3} \\ f_{b4} \\ f_{c1} \\ f_{c2} \\ y_{11} \\ y_{21} \\ y_{31} \\ y_{41} \\ y_{12} \\ y_{22} \\ y_{32} \\ y_{42} \end{bmatrix} = \begin{bmatrix} \mu_a \\ \mu_{b1} \\ \mu_{b2} \\ \mu_{b3} \\ \mu_{b4} \\ \mu_{c1} \\ \mu_{c2} \\ \mu_{11} \\ \mu_{21} \\ \mu_{31} \\ \mu_{41} \\ \mu_{12} \\ \mu_{22} \\ \mu_{32} \\ \mu_{42} \end{bmatrix} + \begin{bmatrix} 1 & 1 & & 1 \\ & 1 & 1 & 1 \\ 1 & & 1 & 1 \\ & 1 & 1 & 1 \\ 1 & & & 1 \\ 1 & 1 & 1 & 1 \\ 1 & & 1 & 1 \\ 1 & & & 1 \end{bmatrix} O_{7 \times 15} + \begin{bmatrix} d_a \\ d_{b1} \\ d_{b2} \\ d_{b3} \\ d_{b4} \\ d_{c1} \\ d_{c2} \\ y_{11} \\ y_{21} \\ y_{31} \\ y_{41} \\ y_{12} \\ y_{22} \\ y_{32} \\ y_{42} \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{21} \\ e_{31} \\ e_{41} \\ e_{12} \\ e_{22} \\ e_{32} \\ e_{42} \end{bmatrix}$$

と表現される。外生変数の共分散行列は

$$\Sigma_u = \begin{bmatrix} \sigma_a^2 & & & & & & \\ & \sigma_b^2 & & & & & \\ & & \sigma_b^2 & & & & \\ & & & \sigma_b^2 & & & \\ & & & & \sigma_c^2 & & \\ & & & & & \sigma_c^2 & \\ & & & & & & \ddots \\ & & & & & & \\ & O & & & & & \\ & & & & & & \\ & & & & & & \sigma_e^2 \end{bmatrix} \quad (3.37)$$

であり、要因Bと要因Cの主効果に関しては、それぞれ $\mu_{b1} + \mu_{b2} + \mu_{b3} + \mu_{b4} = 0$, $\mu_{c1} + \mu_{c2} = 0$ という制約が入る。また要因Bと要因Cの交互作用に関しては、 $\sum_k \mu_{kl} = 0$ に関する制約 2つと $\sum_l \mu_{kl} = 0$ に関する制約が 4つ導入される。

	μ_{b1}	μ_{b2}	μ_{b3}	μ_{b4}	μ_{c1}	μ_{c2}
推定値	0.0420	-0.0898	-0.0989	0.1466	0.0466	-0.0466
検定統計量	1.1975	-2.5569	-2.8158	4.1752	1.6576	-1.6576

表 3.9: 3 要因分散分析母数の推定値 1

表 3.9 は母数要因の主効果に関する推定・検定結果である。 μ_{b1} から μ_{b4} までは要因B(審査委員)の主効果に関する母数であり、5%の危険率で審査委員2, 3が厳しく、審査委員4が甘いと判断できる。

要因Cの主効果 μ_{c1} と μ_{c2} は、技術点と芸術点の難易度の差を表わしている。帰無仮説は「技術点と芸術点の難易度は等しい」というものであり、もし棄却

されれば「技術点(あるいは芸術点)のほうが点を取りにくい」という解釈になるが、5%水準では有意な差とはいえない。

表 3.10 は、全平均と母数要因間の交互作用に関する推定・検定結果である。まず全平均 μ_a が高度に有意である。採点の平均値が0でないというだけのことであり、特に解釈上の意味はない。

	μ_a	μ_{11}	μ_{21}	μ_{31}	μ_{41}
推定値	5.3807	-0.0330	0.0443	0.0261	-0.0375
検定統計量	48.7430	-1.5741	2.1169	1.2485	-1.7913

表 3.10: 3 要因分散分析母数の推定値 2

このモデルの特徴は主効果ではなく、交互作用にある。繰返しのない2要因実験(3.24)式には交互作用の項はなかったけれども、3要因実験では繰返しがなくとも交互作用が3つも登場する。解釈の方法は2要因の場合と同じであり、一方の要因の水準ごとに他方の要因の水準の平均のパターンが異なると解釈する。

たとえば要因Bと要因Cの交互作用 μ_{ij} は、観点ごとの審査委員の点数の傾向の相違を意味する。 $\sum_{l=1}^2 \mu_{kl} = 0$ という制約があるので、表 3.10 には、 $l=1$ の効果の推定・検定結果だけを示している。 μ_{21} が5%水準で有意なので、この場合は、技術性を重視する審査委員と芸術性を重視する審査委員がいるという解釈をする。

表 3.11 は変量効果の推定・検定結果である。まず、変量要因Aの分散は有意であり、5%の危険率でスケーターの間には実力差があると判断できる。また競技会は選手の実力で勝負する場であるから、スケーターの実力の分散 σ_a^2 の推定値が1番大きいことは納得がいく。

次に要因Aと要因Bの交互作用 σ_b が5%水準で有意である。この場合はスケーターごとの審査委員の採点の傾向が異なっている。審査委員のスケーターに対する個人的好みによって、悪くいえば、えこひいきがあると解釈する。この交互作用の効果と、「審査委員の間に甘さに差がある」という要因Bの主効果とは区別しなくてはならない。そして要因Aと要因Cの交互作用 σ_c は、観点ごとのスケーターの点数の傾向が異なる程度を表現している。もし値が大きければ、スケーターには技術演技が得意な選手と芸術演技が得意な選手がいて、それぞれ得手不得手があると解釈する。ただしこのデータに関しては、 σ_c は5%水準で有意でない。

	σ_a^2	σ_b^2	σ_c^2	σ_e^2
推定値	0.122	0.012	0.014	0.013
検定統計量	2.038	2.352	1.812	3.873

表 3.11: 3 要因分散分析・分散成分の推定値

3.4 分割実験

本節では、分割実験と呼ばれるモデルを構造方程式で表現する。分割実験は、標本内実験、被験者内配置実験、経時測定の実験と呼ばれることもある。これまでのモデルと比較して多少複雑なので、このモデルに関しては分析対象となるデータを先に示す。

表 3.12 はアメリカのシカゴとオーランドにあるホテルの料金である。それぞれの街から無作為に選んだ 4 つのホテルのツインの料金を、夏季と冬季に分けて示している。 x_1 と x_2 は値が 1 であれば、それぞれ、そのホテルがシカゴあるいはオーランドにあることを意味する計画行列である。データの傾向としてシカゴは冬よりも夏のほうが料金が高く、オーランドは夏よりも冬のほうが料金が高い傾向が観察される³⁵。

要因 A を「都市」、要因 B を「ホテル」、要因 C を「季節」とすると、要因 A 「都市」は要因 B 「ホテル」を入れ子として含んでいる(たとえばオムニオリエンタルは、シカゴにあり、オーランドのデータはない)。要因 A 「都市」は要因 C 「季節」とはクロスしている(どちらの都市にも夏料金と冬料金がある)。このため、要因 A と要因 C のあいだに交互作用が仮定される。要因 B と要因 C はクロスしているが、繰返しがないから交互作用は考えないということである。ここで例のように 1 つの標本から複数の測定をすることを「反復測定」という。反復は、セルのなかで n 個の標本をとって測定する繰返しとは意味が違うことに注意する必要がある。また因子 A 「都市」を「標本間因子」といい、因子 C 「季節」を「標本内因子」という。また分割実験モデルでは、要因 A と要因 C を母数要因として、要因 B を変量要因として扱う。

³⁵ シカゴはアメリカ大陸の中央よりもやや北東、5 大湖のひとつであるミシガン湖の南西に位置している。冬の寒さは並大抵ではなく、摂氏マイナス 15 度になることも珍しくない。ビジネスマンはともかく、冬に訪れる観光客は、夏に比べてぐっと少ない。

オーランドはフロリダ半島の北部に位置し、ディズニーワールドやケネディー宇宙センターなどのアミューズメントパークで有名な街である。1 年中観光客が訪れててもよさそうだが、夏は暑い。摂氏 35 度を超える日もある。そのうえ湿度は高いし、蚊が多いしで魅力は半減する。

料金(US ドル) ホテル名	夏料金	冬料金	シカゴ	オーラ ンド x_2
	y_1	y_2	x_1	
ハイアット	135	089	1	0
オムニオリエンタル	089	079	1	0
ベスト ウエスタン	056	043	1	0
コンフォート イン	075	068	1	0
クオリティーホテル	069	089	0	1
ハンプトン イン	065	079	0	1
ホリデー イン	098	117	0	1
ハワードジョンソン	080	100	0	1

表 3.12: シカゴとオーランドの季節料金

構造方程式モデルでは

$$\begin{bmatrix} f_b \\ x_1 \\ f_2 \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \mu_b \\ 1/2 \\ 1 \\ a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} 0 & c_1 & c_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 1 & (ac)_{11} & (ac)_{12} & 0 & 0 \\ 1 & (ac)_{21} & (ac)_{22} & 0 & 0 \end{bmatrix} \begin{bmatrix} f_b \\ x_1 \\ f_2 \\ y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} d_b \\ x_1 - 1/2 \\ 0 \\ e_1 \\ e_2 \end{bmatrix}$$

と表現する。外生変数の共分散行列は

$$\Sigma_u = \begin{bmatrix} \sigma_b^2 & & & \\ & 1/4 & & \\ & & 0 & \\ & & & \sigma_e^2 \\ & & & & \sigma_e^2 \end{bmatrix} \quad (3.38)$$

である。このとき

SEM による表記	分割実験のモデルの表記
因子 f_b の平均 μ_b	— 全平均
因子 f_b の分散 σ_b^2	— 変量要因 B の分散成分
観測変数 y_j の平均 a_j	— 母数要因 A の主効果
因子 f_b への係数 c_l	— 母数要因 C の主効果
観測変数 y_j への係数 ac_{jl}	— 母数要因 A と C の交互作用

と読み替え、

- ・誤差変数 e_j の添え字 j よらず誤差分散 σ_e^2 は等しい
 - ・ $a_1 + a_2 = 0, c_1 + c_2 = 0, ac_{11} + ac_{12} = 0, ac_{21} + ac_{22} = 0, ac_{11} + ac_{21} = 0$
- という制約を加えると、分割実験のモデルが表現される。

	a_1	σ_b^2	c_1	ac_{11}	σ_e^2
推定値	0.1875	313.09	-3.9375	-9.3125	31.71
検定統計量	0.0942	1.808	-0.5998	-4.6774	2.00

表 3.13: 分割実験の母数の推定値

表 3.13 は、分割実験の母数の推定値と検定統計量である。母数要因の主効果と交互作用に関しては、上記の制約が入っているので、それぞれ 1 つずつ結果を示せばよい。主効果は有意でなく、交互作用のみ有意である。従って都市によって、季節料金は逆転していると判断する。

3.5 共分散分析

共分散分析は、回帰分析と分散分析とを同時に使う手法である。予測変数による基準変数の条件付き分布に分散分析を行うと考えてもよいし、水準ごとに回帰分析を行うと考えてもよい。前者の場合にはこれまでと同様に水準を 0-1 のダミー変数で表現して実行できる。ここでは後者の表現で分析例を示す。

水準の相違を母集団の相違として表現する場合は、

$$y^{(g)} = b^{(g)} + a^{(g)}x^{(g)} + e^{(g)}, \quad (g = 1, \dots, G) \quad (3.39)$$

$y^{(1)}$	$x^{(1)}$	$y^{(2)}$	$x^{(2)}$	$y^{(3)}$	$x^{(3)}$
15	29	20	22	14	33
19	49	34	24	20	45
21	48	28	49	30	35
27	35	35	46	32	39
35	53	42	52	34	36
39	47	44	43	42	48
23	46	46	64	40	63
38	74	47	61	38	57
33	72	40	55	54	56
50	67	54	54	56	78

表 3.14: 適性検査と生物の成績

となる。 g は水準を指示する添字であり、通常、水準によらず誤差分散は等しい

$$V[e^{(g)}] = \sigma_e^2, \quad (g = 1, \dots, G) \quad (3.40)$$

という仮定を設ける。線形モデルにおける共分散分析では、予測変数は非確率変数であるが、ここでは正規分布を仮定し、

$$M_1 : a^{(g)} \text{ も } b^{(g)} \text{ も水準 } g \text{ によって異なる}$$

$$M_2 : b^{(g)} \text{ は水準 } g \text{ よらず } b \text{ で一定}, a^{(g)} \text{ は異なる}$$

$$M_3 : a^{(g)} \text{ は水準 } g \text{ よらず } a \text{ で一定}, b^{(g)} \text{ は異なる}$$

$$M_4 : a^{(g)} \text{ と } b^{(g)} \text{ は水準 } g \text{ よらず一定で}, \text{ それぞれ } a, b$$

という 4 つのモデルを候補に挙げてみよう。 M_3 と M_4 を比較して、 M_3 が採用されれば、水準の効果があったと判断するのが、通常の共分散分析である。また、その分析に先立って、そもそも傾きが水準間で等しいか否かを検討するために M_1 と M_2 が比較の対象になる場合もある。

表 3.14 は Huitema (1980)³⁶ で示されたデータである。基準変数は「生物の試験」の成績であり、予測変数は「適性検査」の成績である。表 3.15 には、 M_1 から M_4 までのモデルの解の適合度を示した。多母集団の表現を利用すると、

³⁶Huitema, B. H. (1980) *The Analysis of Covariance and Alternatives*. New York: Wiley.

0-1 のダミー変数を使わないために、推測統計的な指標を参照することが可能になる。

	χ^2 値	df	AIC	CAIC	CFI	GFI	AGFI	RMR	RMSEA
M_1	0.374	2	-3.626	-7.042	1.000	0.981	0.826	0.162	0.000
M_2	0.940	4	-7.060	-13.892	1.000	0.955	0.795	0.201	0.000
M_3	0.635	4	-7.365	-14.198	1.000	0.963	0.832	0.162	0.000
M_4	3.641	6	-8.359	-18.607	1.000	0.934	0.803	0.192	0.000

表 3.15: 共分散分析の 4 つのモデルの適合度

自由度が小さいために適合度の値は総じて良く、 CFI, GFI, AGFI, RMR, RMSEA で比較することは難しい。 AIC, CAIC は M_4 が最適であることを示しており、共分散分析流には水準の効果は認められないという解釈となる。

通常の共分散分析では、水準によらず誤差分散は等しいという仮定を導入するが、多母集団モデルの場合は、それらの制約をはずすこともできる。表 3.15 で成績のよかつた M_3 と M_4 に関して、等分散の仮定をはずし

$M_5 : a^{(g)}$ は水準 g によらず a で一定, $b^{(g)}, \sigma_e^{(g)}$ は異なる

$M_6 : a^{(g)}$ と $b^{(g)}$ は水準 g によらず、それぞれ a, b で一定, $\sigma_e^{(g)}$ は異なる

というモデルの解を求め、適合度に関する指標を示したのが表 3.16 である。 SEM の枠組みでは、通常の分析の枠組みでは容易に計算できないモデルの解を比較的容易に求められる。 M_5 も M_6 も情報量規準の観点で M_4 には及ばず、水準の効果が認められない M_4 を採用することになる。

	χ^2 値	df	AIC	CAIC	CFI	GFI	AGFI	RMR	RMSEA
M_5	0.200	2	-3.800	-7.216	1.000	0.999	0.987	0.121	0.000
M_6	3.619	4	-4.381	-11.213	1.000	0.947	0.760	0.184	0.093

表 3.16: 誤差分散の等しくないモデルの適合度

3.6 因子の分散分析

本章ではここまで、実験データの解析に利用される分散分析モデルを SEM の枠組みで表現してきた。本節と次節では、測定方程式の機能を利用した分散分析モデルを考察する。たとえば入門編第 13 章第 5 節および第 6 節の問題 3 では、自己・他者・メタの因子の平均と分散の異同の分析を行っている。このモデルは標本(被験者)内 1 要因 3 水準の分散分析を因子に対して行っていることになるから、因子分析モデルと分散分析モデルを統合した分析モデルである。本節では、標本(被験者)間 1 要因 2 水準の分散分析を因子に対して行うという意味での因子分析と分散分析の統合モデルを紹介する。

平均	4.990	4.419	4.403	4.017	平均
S D	1.871	2.147	2.250	2.294	S D
変数	x_1	x_2	x_3	x_4	変数
相関	1.000	0.440	0.377	0.385	x_1
		1.000	0.515	0.483	x_2
x_1	1.000		1.000	0.503	x_3
x_2	0.364	1.000		1.000	x_4
x_3	0.376	0.498	1.000		
x_4	0.336	0.458	0.488	1.000	
変数	x_1	x_2	x_3	x_4	変数
平均	4.841	4.000	4.229	3.488	平均
S D	1.931	2.128	2.309	2.297	S D

表 3.17: 外国人労働者に対する要望

表 3.17 は Kuhnel(1988)³⁷ で示されたデータである。ドイツで行われた ALL-BUS と呼ばれる調査データの一部であり、外国人労働者に対する要望が測定されている。項目内容は

x_1 ：外国人労働者はライフスタイルを当地のそれに合わせるべきである。

x_2 ：仕事が短縮されたら外国人労働者は本国に戻るべきである。

x_3 ：外国人労働者の政治参加は禁止されるべきである。

³⁷Kuhnel, S.M. (1988). Testing MANOVA designs in LISREL. *Sociological Methods and Research*, 16, 504-523.

x_4 : 外国人労働者は内々で結婚するべきである。

である。7件尺度法であり、値が大きいほど外国人労働者に対して強い要望(ネガティブな印象)を持っていることを意味する。表3.17の上半分は1980年に、下半分は1984年に調査したデータの平均・SD・相関である。被験者は独立に抽出されている。

4年の「経時」を要因、1980年と1984年を水準とし、4つの変数の背後に「外国人労働者に対する要望」という因子を仮定し、分布の差を検討する。構造方程式モデルは

$$\begin{bmatrix} f^{(g)} \\ x_1^{(g)} \\ x_2^{(g)} \\ x_3^{(g)} \\ x_4^{(g)} \end{bmatrix} = \begin{bmatrix} \alpha_f^{(g)} \\ \alpha_{10}^{(g)} \\ \alpha_{20}^{(g)} \\ \alpha_{30}^{(g)} \\ \alpha_{40}^{(g)} \end{bmatrix} + \begin{bmatrix} 0 & & & & \\ \alpha_{11}^{(g)} & \mathbf{O}_{5 \times 4} & & & \\ \alpha_{21}^{(g)} & & \mathbf{O}_{5 \times 4} & & \\ \alpha_{31}^{(g)} & & & \mathbf{O}_{5 \times 4} & \\ \alpha_{41}^{(g)} & & & & \mathbf{O}_{5 \times 4} \end{bmatrix} \begin{bmatrix} f^{(g)} \\ x_1^{(g)} \\ x_2^{(g)} \\ x_3^{(g)} \\ x_4^{(g)} \end{bmatrix} + \begin{bmatrix} f^{(g)} \\ e_1^{(g)} \\ e_2^{(g)} \\ e_3^{(g)} \\ e_4^{(g)} \end{bmatrix} \quad (3.41)$$

と表現する。外生変数の共分散行列は対角行列である。

このモデルに、入門編第13章、第14章で導入した

制約1: 1つの因子の平均と分散は固定

$$\alpha_f^{(1)} = 0.0, \sigma_f^{2(1)} = 1.0$$

制約2: 観測変数の切片をそろえる

$$\alpha_{10}^{(1)} = \alpha_{10}^{(2)}, \alpha_{20}^{(1)} = \alpha_{20}^{(2)}, \alpha_{30}^{(1)} = \alpha_{30}^{(2)}, \alpha_{40}^{(1)} = \alpha_{40}^{(2)}$$

制約3: 測定の不变性

$$\alpha_{11}^{(1)} = \alpha_{11}^{(2)}, \alpha_{21}^{(1)} = \alpha_{21}^{(2)}, \alpha_{31}^{(1)} = \alpha_{31}^{(2)}, \alpha_{41}^{(1)} = \alpha_{41}^{(2)}$$

制約4: 測定誤差の等分散性

$$\sigma_{e1}^{2(1)} = \sigma_{e1}^{2(2)}, \sigma_{e2}^{2(1)} = \sigma_{e2}^{2(2)}, \sigma_{e3}^{2(1)} = \sigma_{e3}^{2(2)}, \sigma_{e4}^{2(1)} = \sigma_{e4}^{2(2)}$$

という4種類の制約を入れる。

SEMの枠組みで実験データの解析を行う場合には、大別して2種類の解釈方法がある。1つは、本章の1要因実験の節で紹介したように、ある係数の検定・区間推定の結果を利用する方法である。この方法には計算が1回で済むというメリットがある。もう一つは、共分散分析の節で紹介したように、複数の仮説に対応するモデルの中から最適なものを選ぶ方法である。この方法には、最適なモデル下での母数の推定値が得られるというメリットがある。

ここでは両者の折衷的な方法を紹介する。分析の主たる目的は「外国人労働者に対する要望」という因子の平均値に差があるか否かというものである。通常の分散分析では、誤差分散は水準間で等しいという制約が入るが、ここでは

それも考察の対象としてみよう。まず1つのモデルで全体的な傾向を把握するために、

$$M_1 : \alpha_f^{(2)} \text{ と } \sigma_f^{2(2)} \text{ を自由母数とする}$$

というモデルの解をもとめる。平均の推定値は $\hat{\alpha}_f^{(2)} = -0.223$ であり、その標準誤差と検定統計量は、それぞれ 0.030, -7.461 である。 $\alpha_f^{(1)}$ はデータに対する適合を損なうことなく制約 1 で 0.0 に固定されているから、明らかに有意な平均値の差である。分散の推定値は $\hat{\sigma}_f^{2(2)} = 0.982$ であり、標準誤差は 0.047 だったので、 $\sigma_f^{2(1)} = 1.0$ と比較して有意差はない。したがって

$$M_2 : \alpha_f^{(2)} \text{ のみ自由母数}, \quad \sigma_f^{2(2)} = 1.0$$

$$M_3 : \alpha_f^{(2)} = 0.0, \quad \sigma_f^{2(2)} \text{ のみ自由母数}$$

$$M_4 : \alpha_f^{(2)} = 0.0, \quad \sigma_f^{2(2)} = 1.0$$

というモデルを比較すると M_2 が最適であることが予想される。実際計算してみると適合度は表3.18のようになり、予想通りであることが確認できる。

	χ^2 値	df	AIC	CAIC	CFI	GFI	AGFI	RMR	RMSEA
M_1	86.711	14	58.711	-48.828	0.987	0.998	0.995	0.045	0.030
M_2	86.862	15	56.862	-58.358	0.987	0.998	0.995	0.044	0.029
M_3	142.799	15	112.799	-2.422	0.976	0.998	0.995	0.045	0.038
M_4	142.945	16	110.945	-11.957	0.977	0.998	0.995	0.045	0.037

表3.18: 因子の分散分析モデルの適合度

M_2 の平均の推定値は $\hat{\alpha}_f^{(2)} = -0.224$ であり、その標準誤差と検定統計量は、それぞれ 0.030, -7.491 であり、 M_1 とほとんど変わらない。因子負荷の推定値(標準化解)は、

$$v_1 = 0.539f_1 + 0.842e_1 \quad (3.42)$$

$$v_2 = 0.711f_1 + 0.703e_2 \quad (3.43)$$

$$v_3 = 0.711f_1 + 0.704e_3 \quad (3.44)$$

$$v_4 = 0.679f_1 + 0.734e_4 \quad (3.45)$$

である。 x_1 はその他の項目のように具体的な行動に関する質問ではない点で、少し独自成分が大きいようである。

3.7 因子の共分散分析

水準ごとの因子に分散分析が可能であるならば同じ考え方を共分散分析にも適用することが可能である。表 3.19 は、AMOS3.6 のマニュアルの分析例 16 に登場するデータである。 x_1 と x_2 はプリテスト、 x_3 と x_4 はポストテストの観測指標であり、上半分が対照群、下半分が実験群のデータである。

平均	18.381	20.229	20.400	21.343	平均
変数	x_1	x_2	x_3	x_4	変数
共分散	37.626	24.933	26.639	23.649	x_1
		34.680	24.236	27.760	x_2
x_1	50.084		32.013	23.565	x_3
x_2	42.373	49.872		33.443	x_4
x_3	40.760	36.094	51.237		
x_4	37.343	40.396	39.890	53.641	
変数	x_1	x_2	x_3	x_4	変数
平均	20.556	21.241	25.667	25.870	平均

表 3.19: プリテストとポストテストの成績

2つの指標の背後に、それぞれ「プリテスト」「ポストテスト」という因子を設定し、構造方程式モデルを

$$\begin{bmatrix} f_1^{(g)} \\ f_2^{(g)} \\ x_1^{(g)} \\ x_2^{(g)} \\ x_3^{(g)} \\ x_4^{(g)} \end{bmatrix} = \begin{bmatrix} \alpha_{f1}^{(g)} \\ \alpha_{f2}^{(g)} \\ \alpha_{10}^{(g)} \\ \alpha_{20}^{(g)} \\ \alpha_{30}^{(g)} \\ \alpha_{40}^{(g)} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \alpha_{f21}^{(g)} & 0 \\ \alpha_{11}^{(g)} & 0 \\ \alpha_{21}^{(g)} & 0 \\ 0 & \alpha_{32}^{(g)} \\ 0 & \alpha_{42}^{(g)} \end{bmatrix} \mathbf{O}_{6 \times 4} \begin{bmatrix} f_1^{(g)} \\ f_2^{(g)} \\ x_1^{(g)} \\ x_2^{(g)} \\ x_3^{(g)} \\ x_4^{(g)} \end{bmatrix} + \begin{bmatrix} f_1^{(g)} \\ d_2^{(g)} \\ x_1^{(g)} \\ x_2^{(g)} \\ x_3^{(g)} \\ x_4^{(g)} \end{bmatrix}$$

と構成する。外生変数の共分散行列は対角行列である。

このモデルに、

制約 1 : 1 つの群の因子の平均と分散は固定 (分散の代りに係数を 1 つ固定)

$$\alpha_{f1}^{(1)} = 0.0, \alpha_{f2}^{(1)} = 0.0, \alpha_{11}^{(1)} = 1.0, \alpha_{32}^{(1)} = 1.0$$

制約 2 : 観測変数の切片をそろえる

$$\alpha_{10}^{(1)} = \alpha_{10}^{(2)}, \alpha_{20}^{(1)} = \alpha_{20}^{(2)}, \alpha_{30}^{(1)} = \alpha_{30}^{(2)}, \alpha_{40}^{(1)} = \alpha_{40}^{(2)}$$

制約 3 : 測定の不变性 (そのうち 2 つは制約 1 との絡みで固定される)

$$\alpha_{11}^{(1)} = \alpha_{11}^{(2)} = 1.0, \alpha_{21}^{(1)} = \alpha_{21}^{(2)}, \alpha_{32}^{(1)} = \alpha_{32}^{(2)} = 1.0, \alpha_{42}^{(1)} = \alpha_{42}^{(2)}$$

制約 4 : 測定誤差の等分散性

$$\sigma_{e1}^{2(1)} = \sigma_{e1}^{2(2)}, \sigma_{e2}^{2(1)} = \sigma_{e2}^{2(2)}, \sigma_{e3}^{2(1)} = \sigma_{e3}^{2(2)}, \sigma_{e4}^{2(1)} = \sigma_{e4}^{2(2)}$$

という 4 種類の制約を入れる。

分析の主たる興味は因子「プリテスト」が所与である場合の「ポストテスト」の条件付き分布の平均の異同である。実験群の因子「ポストテスト」の平均の推定値は $\alpha_{f2}^{(2)} = 3.483$ であり、その標準誤差と検定統計量は、それぞれ 0.481, 7.240 であった。対照群の因子「ポストテスト」の平均は制約 1 において 0.0 に固定されていたから、有意差が見出されたと結論できる。

3.8 多変量分散分析

SEM の一般的なモデル化の道筋は、観測変数の平均や共分散を $\mu(\theta), \Sigma(\theta)$ のように母数の関数で表現することから始まるが、本節では μ, Σ そのものに制約を入れる方法を紹介する。

(1.1) 式に関して、 $\mathbf{t}^{(g)} = \mathbf{x}^{(g)}, \alpha_0^{(g)} = \mu^{(g)}, \mathbf{A}^{(g)} = \mathbf{O}, \mathbf{u}^{(g)} = \mathbf{v}^{(g)}$ とおくと、構造方程式モデルは

$$\mathbf{x}^{(g)} = \mu^{(g)} + \mathbf{v}^{(g)} \quad (3.46)$$

となる。このとき $\alpha_0^{(g)}$ に制約を入れることは、観測変数の平均に制約を入れることと同じになる。同様に外生変数の共分散行列 $\Sigma_u^{(g)}$ に制約を入れることは、観測変数の共分散行列 $\Sigma^{(g)}$ に直接制約を入れることと同じになる。

多変量分散分析モデルは、各水準の共分散行列が等しいという制約の下で、水準間の平均ベクトルの異同を分析するモデルである。たとえば前節の ALLBUS の調査データの場合には、水準数は 2 であるから、 $\Sigma^{(1)} = \Sigma^{(2)}$ という条件の下で

$$M_1 : \mu^{(1)} = \mu^{(2)}$$

$$M_2 : \mu^{(1)} \neq \mu^{(2)}$$

を比較すれば良い。 M_1 と M_2 の適合度 (表 3.20) を比較すると M_2 のほうが明らかによいことが分かる。

	χ^2 値	df	AIC	CAIC	CFI	GFI	AGFI	RMR	RMSEA
M_1	132.195	14	104.195	-3.344	0.978	0.999	0.997	0.043	0.038
M_2	29.234	10	9.234	-67.579	0.996	0.999	0.996	0.042	0.018

表 3.20: 多変量分散分析モデルの適合度

3.8.1 複合対称性の検定

平均や分散・共分散以外の母数を使用しないモデルは、多母集団のモデルばかりでなく 1 母集団のモデルを考えることもできる。繰り返し測定の分散分析では、各変数の分散が等しく、任意の組み合わせの共分散が等しいという仮定が導入される。

たとえば表 3.21($N = 9$, sepath の EX.17 参照) のデータに関しては Σ に

$$\Sigma = \begin{bmatrix} \sigma^2 & & \\ \sigma_r & \sigma^2 & \\ \sigma_r & \sigma_r & \sigma^2 \end{bmatrix} \quad (3.47)$$

という制約(帰無仮説)を入れて解を求めるとき、検定結果は、 $\chi^2 = 0.694$, $df = 4$, $p = 0.95$ であり、帰無仮説は棄却されない。ただし標本数が小さいので帰無仮説が棄却されていないということであり、標本数が大きくなったら適合度指標などを参照したほうがよい。

	x_1	x_2	x_3
x_1	3.10		
x_2	1.92	2.80	
x_3	1.82	2.00	3.80

表 3.21: Winer の分散分析データ

3.8.2 時間的変化の検討

多変量の縦断的データを分析する際には、時間の変化に伴って共分散(相関)行列が変化するのか否かを検討したくなる場合がある。たとえば表 3.22($N = 120$, sepath の EX.9 参照) のデータは、3 つの変数を間隔をおいて 2 回(それぞれ

x_1, x_2, x_3 と x_4, x_5, x_6) 測定したデータである。2 時点で共分散(相関)行列が変化していないという仮説は、 Σ に

$$\Sigma = \begin{bmatrix} \sigma_a^2 & & & & & \\ \sigma_{ab} & \sigma_b^2 & & & & \\ \sigma_{ac} & \sigma_{bc} & \sigma_c^2 & & & \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_a^2 & & \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{ab} & \sigma_b^2 & \\ \sigma_{61} & \sigma_{62} & \sigma_{63} & \sigma_{ac} & \sigma_{bc} & \sigma_c^2 \end{bmatrix} \quad (3.48)$$

という制約を入れて解を求めるべきだ。標本数が大きいので適合度指標を参照すると CFI=0.954, GFI=0.955, AGFI=0.843, RMR=0.169, RMSEA=0.129 となる。観測変数の数が少ないと GFI の値を重視することはできないし、RMSEA の値はよくないので、2 時点で共分散(相関)行列は異なったものであると判断される。

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1.00					
x_2	0.65	1.00				
x_3	0.54	0.68	1.00			
x_4	0.27	0.30	0.21	1.00		
x_5	0.32	0.21	0.27	0.59	1.00	
x_6	0.18	0.26	0.22	0.48	0.55	1.00
S D	2.10	3.00	2.40	1.60	3.30	2.20

表 3.22:

4 時系列解析

時系列解析は、しばしばファイナンス分野で収益やリスクの管理に使用される統計モデルである。大きなお金の動く場面で使用されることが多く、応用的関心や現実的要請が強い手法である。従来は SEM と時系列解析は全く関係の無いモデルであると考えられてきたが、Hershberger, Corneal & Molenaar(1994)³⁸ で動的因子分析モデルが SEM の枠組みで分析できることが示され、Toyoda(1997)³⁹ や Van Buuren(1997)⁴⁰ で SEM と時系列解析の関係が論じられ、時系列データを SEM で分析することの有効性が示されてきた。本章では基本的な事項を紹介する。

4.1 定常性とトープリツツ行列

4.1.1 時系列が 1 つの場合

時系列が 1 つの場合には、たとえば現在 (T 期) から、第 1 期までを

$$[X_T \ X_{T-1} \ X_{T-2} \cdots \ X_t \cdots \ X_1]' \quad (4.1)$$

と表現する。時系列解析の確率変数の扱いに関しては、これまでのデータ解析とは異なる発想を必要とする。 (4.1) 式では個々の X_t が確率変数である。しかも通常は T 個の確率変数から、それぞれ 1 個ずつの実現値 x_t しか観測しない。データ(実現値)も T 個である。1 つの確率変数に 1 つの実現値という状態では、モデルの識別はおろか、標本共分散行列すら計算できない。そこで標本共分散行列を計算できるように、相当地強い仮定を 3 つ

$$E[X_t] = \mu \quad (4.2)$$

$$V[X_t] = \sigma_0^2 \quad (4.3)$$

$$Co[X_t, X_{t-s}] = \sigma_s \quad (4.4)$$

³⁸Hershberger, S.L., Corneal, S.E., and Molenaar, P.C.M. (1994). Dynamic factor analysis: An application to emotional response patterns underlying daughter/father and stepdaughter/stepfather relationships. *Structural Equation Modeling*, 2, 31-52.

³⁹Toyoda, H. (1997). Time series factor analysis model: factors generated by autoregression and moving average process. *Sociological Theory and Methods*, 12, 1-14.

⁴⁰Van Buuren, A. (1997). Fitting ARMA time series by structural equation models. *Psychometrika*, 62, 215-236.

導入する。この仮定を「(弱) 定常性の仮定」という。それぞれ期待値・分散・共分散に関する仮定である。第 1, 第 2 の仮定は、添字 t によらず、確率変数の期待値と分散は、それぞれ一定値 μ, σ_0^2 であることを示している。第 3 の仮定は、共分散は添字 t ではなく、添字の差 $s (s = 0, \dots, S)$ によって規定されることを示している。定常性の仮定をいれても X_T から X_1 までの共分散行列をそのまま論じるのは大変なので、

$$\mathbf{x}_S = [X_t \ X_{t-1} \ X_{t-2} \ \cdots \ X_{t-s} \ \cdots \ X_{t-S}]' \quad (4.5)$$

のように時点 t から S 期前までの共分散行列 \mathbf{T}_S を議論の対象とする。 \mathbf{T}_S は

$$\mathbf{T}_S = \begin{bmatrix} \sigma_0^2 & & & & & & \text{sym} \\ \sigma_1 & \sigma_0^2 & & & & & \\ \sigma_2 & \sigma_1 & \sigma_0^2 & & & & \\ \vdots & \ddots & \ddots & \ddots & & & \\ \sigma_S & \sigma_{S-1} & \cdots & \sigma_1 & \sigma_0^2 & & \end{bmatrix} \quad (4.6)$$

という特殊な構造を持つことになり、これをトーパリツツ (Toeplitz) 行列という。トーパリツツ行列 \mathbf{T}_S (サイズは $(S+1) \times (S+1)$) の推定値である標本トーパリツツ行列 $\hat{\mathbf{T}}_S$ は、 (4.1) 式の実現値である時系列データ

$$[x_T \ x_{T-1} \ x_{T-2} \ \cdots \ x_t \ \cdots \ x_1]' \quad (4.7)$$

を用いて、以下のように計算する ($S < T$ である。 T は数十から数万以上、 S は 1 行であることが多い)。

まず時系列データを縦に $S + 1$ 本並べる。

$$\begin{bmatrix} x_T & x_{T-1} & x_{T-2} & \cdots & x_t & \cdots & x_1 \\ x_T & x_{T-1} & x_{T-2} & \cdots & x_t & \cdots & x_1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ x_T & x_{T-1} & x_{T-2} & \cdots & x_t & \cdots & x_1 \end{bmatrix} \quad (4.8)$$

次に、 s 行目を左に s 列ずらす。

$$\begin{bmatrix} & x_T & x_{T-1} & \cdots & x_2 & x_1 \\ & x_T & x_{T-1} & x_{T-2} & \cdots & x_1 \\ & \vdots & \vdots & \vdots & & \vdots \\ x_T & x_{T-1} & x_{T-2} & \cdots & x_1 \end{bmatrix} \quad (4.9)$$

そして、両端のはみ出した部分(空白が1つ以上ある列)を削除し、

$$\mathbf{X}_S = \begin{bmatrix} x_T & x_{T-1} & x_{T-2} & \cdots & x_t & \cdots & x_{S+1} \\ x_{T-1} & x_{T-2} & x_{T-3} & \cdots & x_{t-1} & \cdots & x_S \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{T-S} & x_{T-S-1} & x_{T-S-2} & \cdots & x_{t-S} & \cdots & x_1 \end{bmatrix} \quad (4.10)$$

のようにサイズ $((S+1) \times (T-S))$ のデータ行列 \mathbf{X}_S を作る。 \mathbf{X}_S の任意の列は(第1列に限らず)時系列変数 \mathbf{x}_S の実現値である。 \mathbf{X}_S は \mathbf{x}_S の実現値を $T-S$ 個含んだデータ行列である。したがって \mathbf{X}_S から計算した標本共分散行列 $\hat{\mathbf{T}}_S$ は、トーブリツ行列 \mathbf{T}_S の推定値として利用できる。 $\hat{\mathbf{T}}_S$ を標本トーブリツ行列といい、推定値の配列も(4.6)式と同じである。各共分散を計算する際の標本数($T-S$ 個)は等しくなる。

ただしこの計算方法だと、1行1列の値と2行2列の値、一般に*i*行*i*列の値は、全て σ_0^2 の推定値なのに値が互いに一致しない。さらに本来 T 個の標本が使えるのに S 個無駄にしていることにもなる。同様に、 \mathbf{X}_S の共分散行列では、2行1列の値と3行2列の値、一般に*i*+1行*i*列の値は、全て σ_1 の推定値なのに値が一致しない。本来 $T-1$ 個の標本が使えるのに $S-1$ 個無駄にしていることにもなる。一般的に、 \mathbf{X}_S の共分散行列では、*i*+*s*行*i*列の値は、全て σ_s の推定値なのに値が一致しない。本来 $T-s$ 個の標本が使えるのに $S-s$ 個無駄にしていることになる。

そこで σ_s の推定値は $T-s$ 個の標本を使って

$$\hat{\sigma}_s = \frac{1}{T-s} \sum_{k=1}^{T-s} (x_k - \bar{x})(x_{k+s} - \bar{x}) \quad (4.11)$$

で計算しておき、(4.6)式に従って推定値を配置する($T-s-1$ で割る場合もある)。本章では、こちらの計算方法で標本共分散行列 $\hat{\mathbf{T}}_S$ を計算する。

4.1.2 時系列が*n*個の場合

時系列が*n*個の場合には、たとえば現在(*T*期)から、第1期までを

$$[x_T \ x_{T-1} \ x_{T-2} \cdots \ x_t \cdots \ x_1] \quad (4.12)$$

と表現する(多変量のイタリック表示では、大文字は行列を意味して使えないので、小文字で確率変数とその実現値を表現する)。サイズは $n \times T$ である。上式

の一般項 \mathbf{x}_t は、*t*期の*n*個の確率変数

$$\mathbf{x}_t = [X_{t1} \ X_{t2} \ X_{t3} \cdots \ X_{ti} \cdots \ X_{tn}]' \quad (4.13)$$

から構成される。(4.12)式(4.13)式は、個々の X_{ti} が確率変数であり、しかも通常は $T \times n$ 個の確率変数から、それぞれ1個ずつの実現値 x_{ti} しか観測しない。1つの確率変数に1つの実現値という状態では、モデルの識別ばかりでなく、標本共分散行列すら計算できないのは、時系列が1つの場合と同じである。そこで、同様な仮定を3つ

$$E[\mathbf{x}_t] = \boldsymbol{\mu} \quad (4.14)$$

$$V[\mathbf{x}_t] = \boldsymbol{\Sigma}_0 \quad (4.15)$$

$$Co[\mathbf{x}_t, \mathbf{x}_{t-s}] = \boldsymbol{\Sigma}_s \quad (4.16)$$

導入する。この仮定を「多変量の(弱)定常性の仮定」という。第1、第2の仮定は、添字*t*によらず、確率変数の期待値ベクトルと共分散行列は、それぞれ一定値 $\boldsymbol{\mu}, \boldsymbol{\Sigma}_0$ であることを示している。第3の仮定は、異なる時期の確率変数間の共分散行列は添字*t*ではなく、添字の差*s*によって規定されることを示している。1変量の場合と同じように*t*期から S 期前までを考察の対象とし、時系列変数

$$\mathbf{x}_{Sn} = [x'_t \ x'_{t-1} \ x'_{t-2} \cdots \ x'_{t-s} \cdots \ x'_{t-S}]' \quad (4.17)$$

を導入する。この確率変数の共分散行列 \mathbf{T}_{Sn} は

$$\mathbf{T}_{Sn} = \begin{bmatrix} \boldsymbol{\Sigma}_0 & & & & sym \\ \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_0 & & & \\ \boldsymbol{\Sigma}_2 & \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_0 & & \\ \vdots & \ddots & \ddots & \ddots & \\ \boldsymbol{\Sigma}_S & \boldsymbol{\Sigma}_{S-1} & \cdots & \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_0 \end{bmatrix} \quad (4.18)$$

という構造を持つ共分散行列となる。これをブロック・トーブリツ行列という。

トーブリツ行列 \mathbf{T}_{Sn} の推定値である標本トーブリツ行列 $\hat{\mathbf{T}}_{Sn}$ は、まず(4.12)式の実現値である時系列データ行列を縦に $S+1$ 個並べる。次に、上から*s*個目の行列を左に*s*列ずらす。そして、両端のはみ出した部分(空白が1つ

以上ある列) を削除し,

$$\mathbf{X}_{Sn} = \begin{bmatrix} \mathbf{x}_{T-n} & \mathbf{x}_{T-1-n} & \mathbf{x}_{T-2-n} & \cdots & \mathbf{x}_{t-n} & \cdots & \mathbf{x}_{S+1-n} \\ \mathbf{x}_{T-1-n} & \mathbf{x}_{T-2-n} & \mathbf{x}_{T-3-n} & \cdots & \mathbf{x}_{t-1-n} & \cdots & \mathbf{x}_{S-n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{x}_{T-S-n} & \mathbf{x}_{T-S-1-n} & \mathbf{x}_{T-S-2-n} & \cdots & \mathbf{x}_{t-S-n} & \cdots & \mathbf{x}_{1-n} \end{bmatrix} \quad (4.19)$$

サイズ $((S+1) \times n) \times (T-S)$ のデータ行列 \mathbf{X}_{Sn} を作る。 \mathbf{X}_{Sn} の任意の列は時系列変数 \mathbf{x}_{Sn} の実現値である。したがって \mathbf{X}_{Sn} から計算した標本共分散行列 $\hat{\mathbf{T}}_{Sn}$ は、ブロック・トープリツツ行列 \mathbf{T}_{Sn} の推定量である。 $\hat{\mathbf{T}}_{Sn}$ を標本ブロック・トープリツツ行列という。多変量の場合も 1 変量の場合と同様に \mathbf{X}_{Sn} から直接に標本共分散行列を計算すると、1 つの母数行列に対して異なった推定値が求まってしまうし、無駄になる標本が生じる。そこで本章では $\hat{\Sigma}_0$ から $\hat{\Sigma}_S$ を個別に計算しておき、(4.18) 式に従って並べるという方法を用いる。

4.1.3 母数の推定

時系列変数を構造方程式で表現し、(ブロック・) トープリツツ行列を $\mathbf{T}_{Sn}(\theta)$ のように母数で構造化し、標本(ブロック・) トープリツツ行列 $\hat{\mathbf{T}}_{Sn}$ を使って、基礎編で論じた推定法で母数を推定する。最小 2 乗法は、推定方法として自然である。また自由度が 0 のモデル(飽和モデル) は適合度関数の値が 0 になり、最尤推定法の解は最小 2 乗法に一致する。

一方、非飽和モデル(自由度が 0 でないモデル)の解を最尤推定する場合には、推定法が仮定している条件をデータが満たしていないことを認識する必要がある。標本(ブロック・) トープリツツ行列を計算する際のデータ行列の各列は、互いに独立ではない。言い換えると 1 列目の実現値と 2 列目の実現値は、データが 1 つしか入れ替わっていないから、1 列目と 2 列目の同時出現確率は、1 列目の出現確率と 2 列目の出現確率の積にはならない。したがって χ^2 値はあてにならないし、それを元に計算される AIC, CFI, RMSEA は参照しないほうがよく、代わりに GFI, AGFI, RMR 等をデータの説明の程度として利用するとよい。

ただし推定値は実用的なレベルで使用して構わないことが、Van Buuren(1997) のシミュレーションで報告されている。また Gourieroux, Monfort,

& Trognon(1984)⁴¹では、標本数が多くなると推定量の統計的な性質もよくなることを、疑似尤度の観点から理論的に示している。

4.2 自己回帰モデル

自己回帰(autoregression) モデルは、観測対象が等間隔の経時測定の 1 変数の回帰モデルであり、過去と現在のデータから、将来を予測するときに最も頻繁に利用される統計モデルである。自己回帰モデルのモデル式は

$$x_t = \mu' + \sum_{s=1}^S \alpha_{t-s} x_{t-s} + e_t \quad (4.20)$$

あるいは、母平均が 0 の観測変数を使って

$$v_t = \sum_{s=1}^S \alpha_{t-s} v_{t-s} + e_t \quad (4.21)$$

と表現され、 $AR(S)$ と表記する。どちらも SEM の枠組みで表現できるが、本章では式展開が容易な後者を議論の対象とする。(4.21) 式には

$$E[e_t e_{t'}] = \begin{cases} \sigma_e^2 & t = t' \\ 0 & t \neq t' \end{cases} \quad (4.22)$$

$$E[v_t e_{t'}] = 0 \quad t < t' \quad (4.23)$$

という制約が入る。すなわち、誤差変数は時期によらず分散が一定であること、同時期以外の誤差変数とは無相関であること、以前の観測変数とは無相関であることが仮定されている。

4.2.1 AR(1) モデル

1 期前の変数から現在の変数を予測するモデルが $AR(1)$ である。モデル式は

$$v_t = \alpha_{t-1} v_{t-1} + e_t \quad (4.24)$$

⁴¹Gourieroux, C., Monfort, A., & Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. *Econometrica*, 17, 287-304.

である。推定すべき母数は α_{t-1} と σ_e^2 であるから、標本トープリツツ行列 $\hat{\mathbf{T}}_1$ を標本共分散行列として扱い単回帰分析を行うことにより SEM の SW で推定することができる。定常性の仮定を満足させる母数の範囲は

$$-1 \leq \alpha_{t-1} \leq 1 \quad (4.25)$$

であることが知られている（紙面の都合で証明は省略）。

4.2.2 AR(2) モデル

1期前と2期前の変数から現在の変数を予測するモデルが $AR(2)$ である。モデル式は

$$v_t = \alpha_{t-1} v_{t-1} + \alpha_{t-2} v_{t-2} + e_t \quad (4.26)$$

である。推定すべき母数は α_{t-1} と α_{t-2} と σ_e^2 であるから、標本トープリツツ行列 $\hat{\mathbf{T}}_2$ を用いて、予測変数 2 つの重回帰分析によって母数を推定する。定常性の仮定を満足させるには母数の範囲が

$$-1 \leq \alpha_{t-2} \leq 1 \quad (4.27)$$

$$\alpha_{t-1} + \alpha_{t-2} \leq 1 \quad (4.28)$$

$$-\alpha_{t-1} + \alpha_{t-2} \leq 1 \quad (4.29)$$

であることが知られている。SW で制約をいれてもよいし、制約を入れずに解を求め、条件外であれば「モデルがデータに適合していない」と判断する。

図 4.1 は、1988 年 1 月より 1994 年 1 月までの景気動向指数（一致指数）の DI (diffusion indexes)⁴² を示したものである。表 4.1 は標本トープリツツ行列 $\hat{\mathbf{T}}_2$ であり、表 4.2 は表 4.1 から計算した標本自己相関行列である。

図 4.1 のデータでモデル $AR(1)$ は

$$v_t = 0.667 v_{t-1} + e_t \quad (4.30)$$

⁴²出典は、経済企画庁編、季刊日本経済指標である。DI (diffusion indexes) は、景気に敏感な指標（系列）を選定し、その変化方向を合成した景気指標であり、景気局面の判断、予測と景気転換点（景気の山、谷）の判定に用いる。指標には 3 種類あり、それは先行指標（図 4.2）、一致指標、遅行指標（図 4.3）である。3ヶ月前と比較して好転している指標（系列）の割合を景気動向の指標とする。データを巻末の付録に掲載したので参照されたい。

のように推定された。標準誤差は 0.088 であり、検定統計量は 7.597 であった。1 期前（3ヶ月前）からの正の影響は確実にあると解釈される。モデル $AR(2)$ は

$$v_t = 0.611 v_{t-1} + 0.085 v_{t-2} + e_t \quad (4.31)$$

のように推定された。1 期前からの係数の標準誤差は 0.117 であり、検定統計量は 5.200 であり、2 期前からの係数の標準誤差も 0.117 であり、検定統計量は 0.720 であった。2 期前（6ヶ月前）からの影響はほとんどないと解釈される。

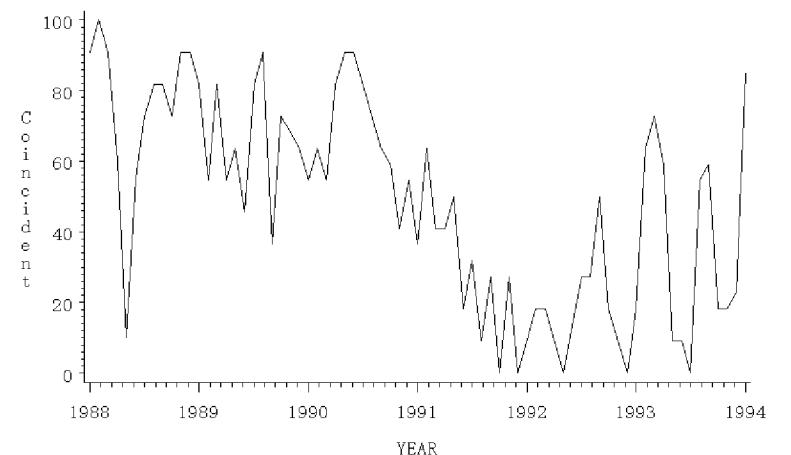


図 4.1: 景気動向指数（一致指数）

4.3 移動平均モデル

観測対象が等間隔の経時測定の 1 変数の回帰モデルであり、過去と現在のデータから、将来を予測するときに利用されるもう一つのモデルに移動平均（moving average）モデルがある。移動平均モデルのモデル式は、母平均が 0 の観測変数を使って

$$v_t = f_t - \sum_{s=1}^S \beta_{t-s} f_{t-s} \quad (4.32)$$

表 4.1: 標本トーブリツツ行列 $\hat{\mathbf{T}}_2$

一致 t	865.3		
一致 $t - 1$	577.2	865.3	
一致 $t - 2$	425.6	577.2	865.3

表 4.2: 標本自己相関行列

一致 t	1.0000		
一致 $t - 1$	0.6671	1.0000	
一致 $t - 2$	0.4919	0.6671	1.0000

と表現され、 $MA(S)$ と表記する。 (4.32) 式には

$$E[f_t f_{t'}] = \begin{cases} \sigma_f^2 & t = t' \\ 0 & t \neq t' \end{cases} \quad (4.33)$$

という制約が入る。ここでは誤差変数は、時期によらず分散が一定であること、同時期以外の誤差変数とは無相関であることが仮定されている。

4.3.1 $MA(1)$ モデル

1 期前の潜在変数の値が現在の変数の値に影響するモデルが $MA(1)$ である。モデル式は

$$v_t = f_t - \beta_{t-1} f_{t-1} \quad (4.34)$$

である。母数の範囲は（ここでは論じないが、反転可能性という性質を満たすために）

$$-1 \leq \beta_{t-1} \leq 1 \quad (4.35)$$

とする必要があることが知られている。推定すべき母数は β_{t-1} と σ_f^2 であるから、標本トーブリツツ行列 $\hat{\mathbf{T}}_1$ を標本共分散行列として扱い、直交した因子で

$$v_t = 1.00f_t + (-\beta_{t-1})f_{t-1} + e_t \quad (4.36)$$

$$v_{t-1} = 1.00f_{t-1} + (-\beta_{t-1})f_{t-2} + e_{t-1} \quad (4.37)$$

という係数に等価の制約 ($\beta_{t-1} = \beta_{t-2}$) のある確認的因子分析の方程式を指定する。ここで

$$\sigma_{et}^2 = \sigma_{et-1}^2 = 0 \quad (4.38)$$

$$\sigma_f^2 = \sigma_{ft}^2 = \sigma_{ft-1}^2 = \sigma_{ft-2}^2 \quad (4.39)$$

の制約を加えることによって (4.33) 式を表現し、 $MA(1)$ の母数を推定する。

4.3.2 $MA(2)$ モデル

1 期前と 2 期前の潜在変数の値が現在の変数の値に影響するモデルが $MA(2)$ である。モデル式は

$$v_t = f_t - \beta_{t-1} f_{t-1} - \beta_{t-2} f_{t-2} \quad (4.40)$$

である。反転可能性を満たせるための母数の範囲は

$$-1 \leq \beta_{t-2} \leq 1 \quad (4.41)$$

$$\beta_{t-1} + \beta_{t-2} \leq 1 \quad (4.42)$$

$$-\beta_{t-1} + \beta_{t-2} \leq 1 \quad (4.43)$$

であることが知られている。推定すべき母数は β_{t-1} と β_{t-2} と σ_f^2 であるから、標本トーブリツツ行列 $\hat{\mathbf{T}}_2$ を標本共分散行列として扱い、直交した因子で

$$v_t = 1.00f_t + (-\beta_{t-1})f_{t-1} + (-\beta_{t-2})f_{t-2} + e_t \quad (4.44)$$

$$v_{t-1} = 1.00f_{t-1} + (-\beta_{t-1})f_{t-2} + (-\beta_{t-2})f_{t-3} + e_{t-1} \quad (4.45)$$

$$v_{t-2} = 1.00f_{t-2} + (-\beta_{t-1})f_{t-3} + (-\beta_{t-2})f_{t-4} + e_{t-2} \quad (4.46)$$

という係数に等価の制約のある確認的因子分析の方程式を指定する。ここで

$$\sigma_{et}^2 = \sigma_{et-1}^2 = \sigma_{et-2}^2 = 0 \quad (4.47)$$

$$\sigma_f^2 = \sigma_{ft}^2 = \sigma_{ft-1}^2 = \sigma_{ft-2}^2 = \sigma_{ft-3}^2 = \sigma_{ft-4}^2 \quad (4.48)$$

の制約を加えることによって (4.33) 式を表現し、 $MA(2)$ の母数を推定する。

図 4.1 のデータでモデル $MA(1)$ は

$$v_t = 1.00f_t + 1.00f_{t-1} \quad (4.49)$$

のように推定された。モデルの(反転可能性のため)バウンズの制約を受けてしまい、モデルとデータが適合していないことが示されてしまった。モデル $MA(2)$ は

$$v_t = 1.00f_t + 2.000f_{t-1} + 0.432f_{t-2} \quad (4.50)$$

のように推定された。 $MA(1)$ の場合と同様にモデルのバウンズの制約を受けてしまい、こちらもモデルとデータが適合していない。

4.4 自己回帰移動平均モデル

AR モデルと MA モデルを合体させた予測モデルが自己回帰移動平均 $ARMA$ (autoregression and moving average) モデルである。モデル式は

$$v_t = \sum_{r=1}^R \alpha_{t-r} v_{t-r} + f_t - \sum_{r'=1}^{R'} \beta_{t-r'} f_{t-r'} \quad (4.51)$$

であり、 $ARMA(R, R')$ と表記する。モデルの仮定は(4.22)式、(4.23)式、(4.33)式である。たとえば $ARMA(1, 1)$ は

$$v_t = \alpha_{t-1} v_{t-1} + f_t - \beta_{t-1} f_{t-1} \quad (4.52)$$

である。母数の範囲は

$$-1 \leq \beta_{t-1} \leq 1 \quad (4.53)$$

$$-1 \leq \alpha_{t-1} \leq 1 \quad (4.54)$$

とする必要があることが知られている。推定すべき母数は α_{t-1} と β_{t-1} と σ_f^2 の 3 つであるから、標本トーブリツツ行列 $\hat{\mathbf{T}}_2$ を標本共分散行列として扱い

$$v_t = \alpha_{t-1} v_{t-1} + 1.00f_t + (-\beta_{t-1})f_{t-1} + e_t \quad (4.55)$$

$$v_{t-1} = \alpha_{t-1} v_{t-2} + 1.00f_{t-1} + (-\beta_{t-1})f_{t-2} + e_{t-1} \quad (4.56)$$

という係数に等価の制約のある方程式を指定し、さらに

$$\sigma_{et}^2 = \sigma_{et-1}^2 = 0 \quad (4.57)$$

$$\sigma_f^2 = \sigma_{ft}^2 = \sigma_{ft-1}^2 = \sigma_{ft-2}^2 \quad (4.58)$$

の制約の下で $ARMA(1, 1)$ の母数を推定する。

図 4.1 のデータでモデル $ARMA(1, 1)$ は

$$v_t = 0.690v_{t-1} + f_t - 0.081f_{t-1} \quad (4.59)$$

のように推定された。自己回帰部分の標準誤差は 0.069 であり、検定統計量は 9.942 であった。移動平均部分の標準誤差は 0.139 であり、検定統計量は -0.586 であった。1 期前(3 ケ月前)からの正の影響は確実にあるが、移動平均はこのデータを説明するのに適切ではないと解釈される。

4.5 ベクトル自己回帰モデル

ベクトル自己回帰(vector autoregression) モデルは、多変量の自己回帰モデルである。モデル式は、サイズ $n \times 1$ の母平均が 0 の観測変数ベクトルを使って

$$v_t = \mathbf{A}_1 v_{t-1} + \mathbf{A}_2 v_{t-2} + \cdots + \mathbf{A}_s v_{t-s} + \cdots + \mathbf{A}_S v_{t-S} + e_t \quad (4.60)$$

と表現され、 $VAR(S, n)$ と表記する。 \mathbf{A}_s は s 期前からの影響を表現するサイズ $n \times n$ の係数行列であり、(4.60) 式には

$$E[e_t] = \mathbf{o} \quad (4.61)$$

$$E[e_t e_t'] = \Sigma_e \quad (4.62)$$

$$E[e_t e_s'] = 0 \quad (t \neq s) \quad (4.63)$$

という制約が入る。ここでは、誤差変数は平均が 0 であること、時期によらず共分散行列は一定であること、同時期以外の誤差変数同士は無相関であることが仮定されている。母数は標本ブロック・トーブリツツ行列 \mathbf{T}_{Sn} を用いて推定する。たとえば $S = 2, n = 2$ の場合のベクトル自己回帰モデル $VAR(2, 2)$ は

$$v_{t1} = \alpha_{t-1, 11} v_{t-1, 1} + \alpha_{t-2, 11} v_{t-2, 1} + \alpha_{t-1, 12} v_{t-1, 2} + \alpha_{t-2, 12} v_{t-2, 2} + e_{t1} \quad (4.64)$$

$$v_{t2} = \alpha_{t-1, 21} v_{t-1, 1} + \alpha_{t-2, 21} v_{t-2, 1} + \alpha_{t-1, 22} v_{t-1, 2} + \alpha_{t-2, 22} v_{t-2, 2} + e_{t2} \quad (4.65)$$

である。係数についての 3 つの添字は「時期」「影響を受ける変数番号」「影響を与える変数番号」である。

図 4.2 は、1988 年 1 月より 1994 年 1 月までの先行指数の DI を示したものである。表 4.3 は、図 4.1 と図 4.2 による先行指数と一致指数の標本ブロック・

トープリツツ行列 $\hat{\mathbf{T}}_{22}$ であり、表 4.4 は表 4.3 から計算した標本自己相関行列である。このデータを用いた $VAR(2, 2)$ は

$$v_{t1} = 0.387v_{t-11} - 0.110v_{t-21} + 0.154v_{t-12} + 0.112v_{t-22} + e_{t1} \quad (4.66)$$

$$v_{t2} = 0.069v_{t-11} - 0.260v_{t-21} + 0.587v_{t-12} + 0.188v_{t-22} + e_{t2} \quad (4.67)$$

のように推定された。5%水準で有意なのは、 α_{t-111} と α_{t-122} だけである。どちらも 1 期前の自己から影響を受けているが、お互いには影響は与えていない。

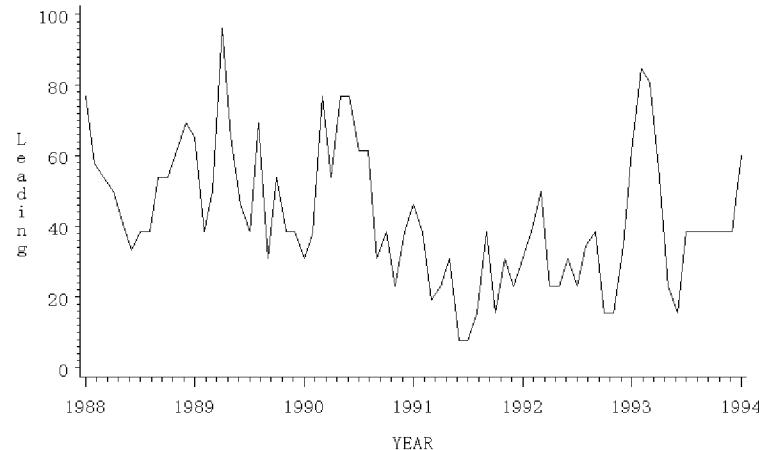


図 4.2: 景気動向指数 (先行指標)

表 4.3: 標本ブロック・トープリツツ行列 $\hat{\mathbf{T}}_{22}$

先行 t	370.0					
一致 t	373.7	865.3				
先行 $t-1$	212.7	248.6	370.0			
一致 $t-1$	315.1	577.2	373.7	865.3		
先行 $t-2$	121.8	134.4	212.7	248.6	370.0	
一致 $t-2$	266.6	425.6	315.1	577.2	373.7	865.3

表 4.4: 標本自己相関行列

先行 t	1.0000					
一致 t	0.6605	1.0000				
先行 $t-1$	0.5748	0.4393	1.0000			
一致 $t-1$	0.5569	0.6671	0.6605	1.0000		
先行 $t-2$	0.3292	0.2375	0.5748	0.4393	1.0000	
一致 $t-2$	0.4711	0.4919	0.5569	0.6671	0.6605	1.0000

4.5.1 グレンジャーの因果関係

経済時系列解析では、因果関係の方向性が議論されることが多い。その際、しばしば登場するのがグレンジャーの因果関係である。本来、グレンジャーの因果関係は $VAR(S, n)$ モデル特有の概念ではないが、ベクトル自己回帰モデルの分析で問題にされることが多いので、ここではその文脈で解説する。

ベクトル自己回帰モデルにおいて、変数 v_1 から変数 v_2 へグレンジャー因果的影響があるとは「過去の v_1 から t 期の v_2 への係数が 0 でない」ということである（0 でないというだけではなく、実感として影響するだけの絶対値が、本来は必要である）。先に分析した $VAR(2, 2)$ で例をあげるならば、

$$\textcircled{1} \quad \alpha_{t-112} \neq 0 \quad \text{あるいは} \quad \alpha_{t-212} \neq 0$$

$$\textcircled{2} \quad \alpha_{t-121} \neq 0 \quad \text{あるいは} \quad \alpha_{t-221} \neq 0$$

という命題を用いて、

(a) ①と②の両方が成り立つ：双方向の因果関係がある

(b) ①だけが成り立つ： v_2 が v_1 に因果的影響を与えている

(c) ②だけが成り立つ： v_1 が v_2 に因果的影響を与えている

(d) ①も②も成り立たない：互いに因果的な影響はない

のよう因果関係を判定し、このときグレンジャーの意味での因果関係が確認された（されなかった）という。

「グレンジャーの意味での」 というように限定的に表現するのには理由がある。それは入門編、第 9 章で論じたように第 3 の変数の影響が無いことを、こ

の分析だけでは示せないからである(そもそもデータ解析の結果のみから因果関係を導くことはできない)。

たとえば、 v_1 が「最高気温」で v_2 が前日の天気予報の「最高気温の予測値」とすると、グレンジャーの意味での因果関係が確認されても v_2 は v_1 の原因とはいえない。前日の大気や雲の状態という v_2 と v_1 に共通の、真の原因があることは明白である。

それではグレンジャーの意味での因果関係は無意味であるかといえば、そうではない。因果関係の方向性について、道具体的変数なしに論じることができる。これは大変有用で強力な性質である。たとえば(b)であれば、 v_2 が v_1 に因果的影響を与えてるのであってその逆ではないことを、かなりはっきりと導ける。これとて、原因と認定した変数は「稻妻から雷鳴が遅れて観測されただけで、両者に因果関係はない」と反論できる場合もあるが、依然として方向性の確認には有効である。

ただし第3の変数が有るかもしれない。測定の遅延が有るかもしれない、という可能性だけで思考・考察を止めてはいけない。実質科学的な知見を利用して、それらの可能性を吟味し、仮説を鍛えようとする限り、グレンジャーの意味での因果関係は有効であるし、手軽に利用できるモデルであるから計量経済学以外の分野でも、もっと使用されるべきツールである。

4.5.2 縦断・繰り返し測定との相違

多変量の時系列分析は、因果関係の吟味に比較的明確な知見を提供する。ただし入門編で登場したモデルの中で、シンプレックス構造解析や、ラグ付き変数による分析などの縦断データの解析は、広義の意味で時系列の解析モデルであり、基本的に時間の流れに沿ってだけパスが引けるから、因果の前後関係の吟味に関して明確な知見が得られる。その意味で本章で論じている時系列解析と同等の利点を有している。

狭義の時系列モデルと縦断データ解析モデルの相違は、データの測定状況の相違である。本章で扱った時系列モデルでは、 T 期の各時点で1回(多変量を含む)の測定しかしない。1時点で1回の標本抽出である。逆に T はある程度大きいし、確率変数の性質は T に関する添字 t には依存させられない。時期の差である s にだけ確率変数の特徴が現れるという強い仮定(定常性の仮定)を必要とする。その代わりに、多時点の時系列の性質を効率よく分析することがで

きる。

それに対して縦断データ解析モデルに共通していることは、各時点で多数の測定を行うということである。1時点で N 回の標本抽出を行う。逆に測定回数である T が小さくてよい(2回以上であれば縦断データと呼ばれる)ということである。確率変数の性質は T に関する添字 t に依存させて記述できる。たとえば、小学校1年時の学力の平均と分散は2年次より小さいなどという t に依存した記述は、縦断データ解析モデルでは簡単に表現できても、時系列モデルでは相当に難しい。

4.6 動的因子分析

因子分析法には「時期」を観測対象とする多変量データに直接因子分析を適用する p -技法があるが、この方法はラグ付き共分散を考慮できないという理論的な欠点を持っている(Anderson(1963)⁴³)。この問題を解決するために Molenaar(1985)⁴⁴ は、ラグ付き共分散を考慮した動的因子分析モデル(Dynamic factor analysis model)を提案した。動的因子分析法の概略は以下のとおりである。

分析対象となっている多変量時系列データの背後に Q 個の因子

$$\mathbf{f}_t = [f_{t1} \cdots f_{tq} \cdots f_{tQ}]' \quad (4.68)$$

を仮定する。また、時期 t における n 個の誤差変数を

$$\mathbf{e}_t = [e_{t1} \cdots e_{tj} \cdots e_{tn}]' \quad (4.69)$$

と表記する($Q < n$)。このとき動的因子分析モデルは

$$\mathbf{v}_t = \sum_{s=0}^S \mathbf{\Lambda}_s \mathbf{f}_{t-s} + \mathbf{e}_t \quad (4.70)$$

と表現される。 $\mathbf{\Lambda}_s$ は \mathbf{v}_t に対する \mathbf{f}_{t-s} からの影響力を示した因子負荷行列で

⁴³ Anderson, T.W. (1963). The use of factor analysis in the statistical analysis of multiple time series. *Psychometrika*, 28, 1-25.

⁴⁴ Molenaar, P.C.M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, 50, 181-202.

ある. 因子 \mathbf{f}_t には

$$E[\mathbf{f}_t] = \mathbf{o} \quad (4.71)$$

$$V[\mathbf{f}_t] = \mathbf{I} \quad (4.72)$$

$$Co[\mathbf{f}_t, \mathbf{f}_{t-s}] = \mathbf{O} \quad (4.73)$$

のような仮定が置かれている. また誤差変数 \mathbf{e}_t に関しては

$$E[\mathbf{e}_t] = \mathbf{o} \quad (4.74)$$

$$V[\mathbf{e}_t] = \mathbf{D}_0 \quad (4.75)$$

$$Co[\mathbf{e}_t, \mathbf{e}_{t-s}] = \mathbf{D}_s \quad (4.76)$$

という仮定が置かれる (\mathbf{D}_s ($s = 0, \dots, S$) は対角行列). 場合によっては $\mathbf{D}_s = \mathbf{O}$ ($s \neq 0$) が仮定されることもある.

Hershberger, Corneal & Molenaar (1994)⁴⁵は, 動的因子分析を共分散構造分析の SW で実行できることを示した. その方法は, まず見かけ上, 通常の因子分析モデルを用意し

$$\mathbf{v} = \mathbf{\Lambda} \mathbf{f} + \mathbf{e} \quad (4.77)$$

確率変数の各要素は

$$\mathbf{v} = [\mathbf{v}'_t \mathbf{v}'_{t-1} \cdots \mathbf{v}'_{t-s} \cdots \mathbf{v}'_{t-S}]' \quad (4.78)$$

$$\mathbf{f} = [\mathbf{f}'_t \mathbf{f}'_{t-1} \cdots \mathbf{f}'_{t-s} \cdots \mathbf{f}'_{t-2S}]' \quad (4.79)$$

$$\mathbf{e} = [\mathbf{e}'_t \mathbf{e}'_{t-1} \cdots \mathbf{e}'_{t-s} \cdots \mathbf{e}'_{t-S}]' \quad (4.80)$$

であり, 因子負荷行列は

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_0 & \mathbf{\Lambda}_1 & \cdots & \mathbf{\Lambda}_S & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_0 & \mathbf{\Lambda}_1 & \cdots & \mathbf{\Lambda}_S & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{\Lambda}_0 & \mathbf{\Lambda}_1 & \cdots & \cdots & \mathbf{\Lambda}_S \end{bmatrix} \quad (4.81)$$

⁴⁵Hershberger,S.L., Corneal,S.E., & Molenaar,P.C.M. (1994). Dynamic factor analysis: An application to Emotional Response patterns underlying daughter/father and stepdaughter/stepfather relationships. *Structural Equation Modeling*, 2, (1), 31-57. この論文では父娘関係の多変量時系列の分析が行われている. 本章で紹介した方法は経済分析ばかりでなく, 様々な領域での応用が期待される.

と制約する. (4.71) 式から (4.76) 式の仮定の下でプロック・トーブリツツ行列 \mathbf{T}_{Sn} は

$$\mathbf{T}_{Sn} = \mathbf{\Lambda} \mathbf{\Lambda}' + \mathbf{\Delta} \quad (4.82)$$

と分解される. ただし

$$\mathbf{\Delta} = \begin{bmatrix} \mathbf{D}_0 & & & & & & & sym \\ \mathbf{D}_1 & \mathbf{D}_0 & & & & & & \\ \mathbf{D}_2 & \mathbf{D}_1 & \mathbf{D}_0 & & & & & \\ \vdots & \ddots & \ddots & \ddots & & & & \\ \mathbf{D}_S & \mathbf{D}_{S-1} & \cdots & \mathbf{D}_1 & \mathbf{D}_0 & & & \end{bmatrix} \quad (4.83)$$

である. \mathbf{D}_s は対角行列とする.

図 4.3 は, 1988 年 1 月より 1994 年 1 月までの逓行指数の DI を示したものである. 表 4.5 は, 図 4.1 と図 4.2 と図 4.3 を合わせた先行指数と一致指数と逓行指標の標本プロック・トーブリツツ行列 \mathbf{T}_{23} であり, 表 4.6 は表 4.3 から計算した標本自己相関行列である. 3 つの時系列に同時に動的因子分析法を適用してみよう.

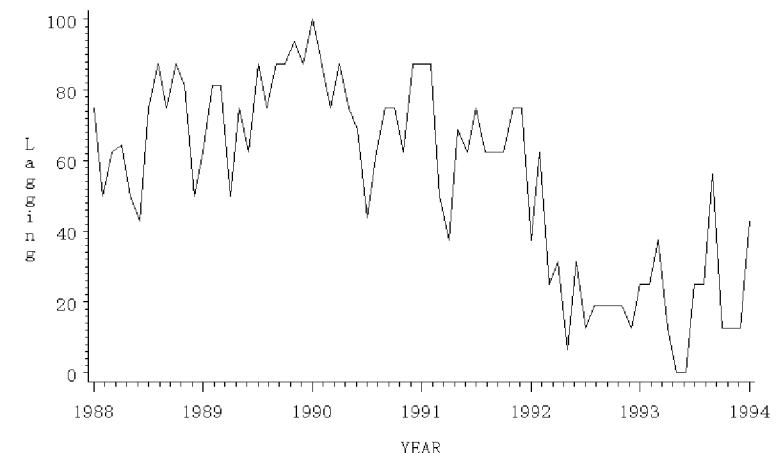


図 4.3: 景気動向指数 (逓行指数)

3 つの時系列の背後には, 指標の性質から「日本経済の景気」という 1 つの潜在変数を仮定するのが適当であろう ($Q = 1$). $S = 2$ としてモデルを特定す

表 4.5: 標本ブロック・トープリツツ行列 $\hat{\mathbf{T}}_{23}$

先行 t	370.0							
一致 t	373.7	865.3						
遅行 t	91.5	433.7	748.5					
先行 $t-1$	212.7	248.6	103.6	370.0				
一致 $t-1$	315.1	577.2	359.9	373.7	865.3			
遅行 $t-1$	83.0	377.3	585.0	91.5	433.7	748.5		
先行 $t-2$	121.8	134.4	98.7	212.7	248.6	103.6	370.0	
一致 $t-2$	266.6	425.6	307.6	315.1	577.2	359.9	373.7	865.3
遅行 $t-2$	48.9	332.4	537.0	83.0	377.3	585.0	91.5	433.7
								748.5

表 4.6: $\hat{\mathbf{T}}_{23}$ から計算した標本自己相関行列

先行 t	1.0000							
一致 t	0.6605	1.0000						
遅行 t	0.1739	0.5389	1.0000					
先行 $t-1$	0.5748	0.4393	0.1970	1.0000				
一致 $t-1$	0.5569	0.6671	0.4472	0.6605	1.0000			
遅行 $t-1$	0.1577	0.4688	0.7815	0.1739	0.5389	1.0000		
先行 $t-2$	0.3292	0.2375	0.1875	0.5748	0.4393	0.1970	1.0000	
一致 $t-2$	0.4711	0.4919	0.3822	0.5569	0.6671	0.4472	0.6605	1.0000
遅行 $t-2$	0.0929	0.4130	0.7174	0.1577	0.4688	0.7815	0.1739	0.5389
								1.0000

ると確率変数の各要素は

$$\mathbf{v} = [\mathbf{v}'_t \mathbf{v}'_{t-1} \mathbf{v}'_{t-2}]' \quad (9 \times 1) \quad (4.84)$$

$$\mathbf{f} = [f'_t f'_{t-1} f'_{t-2} f'_{t-3} f'_{t-4}]' \quad (5 \times 1) \quad (4.85)$$

$$\mathbf{e} = [\mathbf{e}'_t \mathbf{e}'_{t-1} \mathbf{e}'_{t-2}]' \quad (9 \times 1) \quad (4.86)$$

である。 \mathbf{v}_t は「先行」「一致」「遅行」の順に並んでいる。因子負荷行列は

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix} \quad (9 \times 5) \quad (4.87)$$

と制約する。 \mathbf{A}_0 も \mathbf{A}_1 も \mathbf{A}_2 もサイズは 3×1 である⁴⁶。独自性行列は

$$\Delta = \begin{bmatrix} \mathbf{D}_0 & & \\ \mathbf{D}_1 & \mathbf{D}_0 & \\ \mathbf{D}_2 & \mathbf{D}_1 & \mathbf{D}_0 \end{bmatrix} \quad (9 \times 9) \quad (4.88)$$

である。

最初に (4.87)(4.88) に従った分析をしたところ、不適解になってしまった。そこで「一致指標」は因子のみから影響を受ける（誤差なしで観測される）と仮定して母数を減らし、解を求めた。その結果、因子負荷の標準化解は

$$\hat{\mathbf{A}}_0 = \begin{bmatrix} 0.158 \\ 0.598 \\ 0.302 \end{bmatrix} \quad \hat{\mathbf{A}}_1 = \begin{bmatrix} 0.396 \\ 0.487 \\ 0.230 \end{bmatrix} \quad \hat{\mathbf{A}}_2 = \begin{bmatrix} 0.541 \\ 0.637 \\ 0.145 \end{bmatrix} \quad (4.89)$$

となった。「一致指標」の誤差変数は 0 と仮定しているので、因子の内容は「一致指標」の性質を強く受ける。「先行指標」の係数は古いものほうが大きいこと ($0.158 < 0.396 < 0.541$)、逆に「遅行指標」の係数は新しいものほうが大きい ($0.302 > 0.230 > 0.145$) のはこのためであろうと考えられる（「先行指標」は景気の変動の影響を早く受ける業種によって、「遅行指標」は遅く受ける業種によって計算されている）。GFI=0.912, AGFI=0.867 であった。

誤差変数の 1 次と 2 次の自己相関（標準化された自己共分散）は

$$\hat{\mathbf{D}}_1 = \begin{bmatrix} 0.444 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.741 \end{bmatrix} \quad \hat{\mathbf{D}}_2 = \begin{bmatrix} 0.190 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.679 \end{bmatrix} \quad (4.90)$$

であった。当然のことながら「先行指標」や「遅行指標」にも独自の自己相関要素があるということである。

4.7 時系列因子分析

動的因子分析法は、時系列的な因子の重み付き和で観測変数のラグ付き共分散を説明するモデルである。これに対して、本節では互いに関連のある多数の

⁴⁶ 因子が複数ある探索的因子分析モデルでは、モデルを識別させるために因子負荷行列 \mathbf{A}_s は広義の下三角行列とする。この場合は推定値を求めた後に、目的に応じて回転をする。また後述する時系列因子分析でも、この事情は共通している。本章では単因子の場合のみを数値例としてあげる。

時系列の背後に、観測できない小数の時系列が存在するという観点から提案された時系列因子分析モデル(前掲 Toyoda,1997)を紹介する。

観測変数 v_t は通常の因子分析モデルで

$$v_t = \Lambda_0 f_t + e_t \quad (4.91)$$

のように表現されているものとする。因子が複数ある探索的因子分析モデルでは、モデルを識別させるために因子負荷行列 Λ_0 は広義の下三角行列とする。そして因子 f_t が ARMA 過程に従って

$$f_t = B_1 f_{t-1} + B_2 f_{t-2} + \cdots + B_R f_{t-R} + u_t + \Gamma_1 u_{t-1} + \cdots + \Gamma_{R'} u_{t-R'} \quad (4.92)$$

のように生成されるものとする。ここで R は AR の次数であり、 R' は MA の次数である。(4.91) 式と (4.92) 式を RAM の表現を用いて同時に表現する方法に関しては、多少複雑なので Toyoda(1997) を参照されたい。

図 4.4 は、図 4.1、図 4.2、図 4.3 を重ねて描いたグラフである。3 つの指標はそれぞれ日本経済の景気動向を示している。大まかに見るとバブル経済がはじけた様子が明確に示されている。そこでこの 3 本の時系列の背後に潜在した「景気動向」と最初のモデルでは「景気動向」が ARMA(1,1) 過程 ($R = 1, R' = 1$) に従っているものと仮定する。(4.77) 式に従った表現をするならば確率変数の各要素は

$$v = [v'_t v'_{t-1} v'_{t-2}]' \quad (9 \times 1) \quad (4.93)$$

$$f = [f_t f_{t-1} f_{t-2}]' \quad (3 \times 1) \quad (4.94)$$

$$e = [e'_t e'_{t-1} e'_{t-2}]' \quad (9 \times 1) \quad (4.95)$$

である。 v_t は「先行」「一致」「遅行」の順に並んでいる。因子負荷行列は

$$\Lambda = \begin{bmatrix} \Lambda_0 & 0 & 0 \\ 0 & \Lambda_0 & 0 \\ 0 & 0 & \Lambda_0 \end{bmatrix} \quad (9 \times 3) \quad (4.96)$$

⁴⁷ 例えば銀行株、繊維株といふことがあるように、決まった業種で似たような株価の変動が見られる場合がある。同じ業界で似たような値動きをする複数の株の背後に「銀行株」「繊維株」のように業種名を命名した潜在変数を設定することは有効かもしれない。

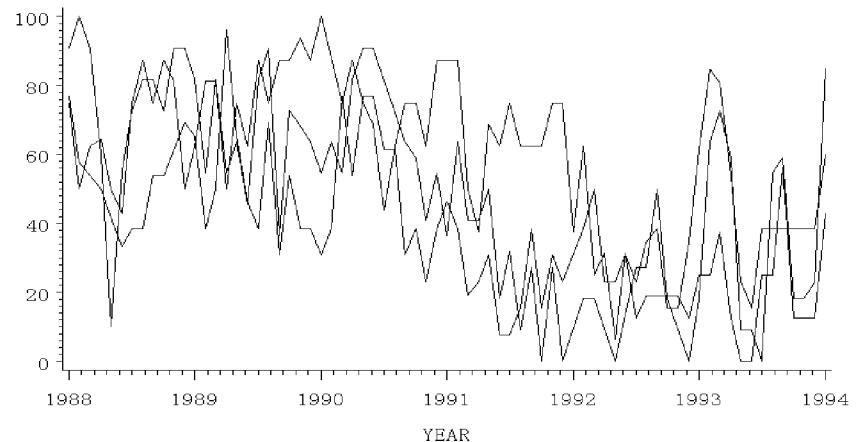


図 4.4: 先行・一致・遅行指標の同時グラフ

と指定する。因子の ARMA 過程は

$$f_t = \beta_{t-1} f_{t-1} + u_t + \gamma_{t-1} u_{t-1} \quad (4.97)$$

$$f_{t-1} = \beta_{t-2} f_{t-2} + u_{t-1} + \gamma_{t-2} u_{t-2} \quad (4.98)$$

のようく設定する。ただし因子の分散は t によらず一定なので、

$$\psi = \sigma_{ut}^2 = \sigma_{ut-1}^2 = \sigma_{ut-2}^2 = (\sigma_{ft-2}^2 - \beta_{t-1}^2 \sigma_{ft-2}^2) / (1 + \gamma_{t-1}^2) \quad (4.99)$$

という制約をいれる。 \hat{T}_{23} で推定した解を表 4.7 に示した。固定母数にはアスタリスクを付した。ただし $\phi = \sigma_{ft-2}^2$ である。GFI は 0.63 であり、AGFI は 0.55 であり、値が低いので改良を要する。

Model 2: 表 4.7 を観察すると、 $\hat{\gamma}_1 = 0.01$ であり、移動平均の効果はほとんどないことがわかる。そこで時系列因子が ARMA(1,0) 過程 ($R = 1, R' = 0$) に従っているものとする。(4.77) 式に従った表現をするならば、確率変数は

$$v = [v'_t v'_{t-1}]' \quad (6 \times 1) \quad (4.100)$$

$$f = [f_t f_{t-1}]' \quad (2 \times 1) \quad (4.101)$$

$$e = [e'_t e'_{t-1}]' \quad (6 \times 1) \quad (4.102)$$

であり、因子負荷行列は

$$\Lambda = \begin{bmatrix} \Lambda_0 & 0 \\ 0 & \Lambda_0 \end{bmatrix} \quad (6 \times 2) \quad (4.103)$$

と指定する。因子の ARMA 過程は

$$f_t = \beta_t f_{t-1} + u_t \quad (4.104)$$

のように設定する。ただし因子の分散は t によらず一定なので、

$$\sigma_{ut}^2 = \sigma_{ft-2}^2 - \beta_{t-1}^2 \sigma_{ft-2}^2 \quad (4.105)$$

という制約をいれる。 \hat{T}_{23} で推定した解を表 4.7 に示した。GFI は 0.75 であり、AGFI は 0.62 であり、依然として値が低いので改良を要する。

Model 3 : モデル 2 の残差行列 ($S - \hat{\Sigma}$) を観察すると

$$S - \hat{\Sigma} = \begin{bmatrix} 0.0 & & & & & \\ 0.6 & 0.0 & & & & \\ & & & & & \\ -97.7 & 1.2 & 0.0 & & & \\ & & & & & \\ 101.6 & -5.4 & -25.1 & 0.0 & & \\ & & & & & \\ 61.1 & -3.4 & 65.5 & 0.6 & 0.0 & \\ & & & & & \\ -45.8 & 82.9 & 435.7 & -97.7 & 1.2 & 0.0 \end{bmatrix} \quad sym. \quad (4.106)$$

である。大きな残差として $s_{63} - \hat{s}_{63} = 435.7$ と $s_{41} - \hat{s}_{41} = 101.6$ の 2 つがある。それらは「一致指標」と「遅行指標」の誤差の 1 次の自己相関の存在を示唆する。これは動的因子分析でも同様の結果が得られていた。そこで独自性行列を

$$\Delta = \begin{bmatrix} \delta_1 & & & & & \\ 0 & \delta_2 & & & & \\ & & & & & \\ 0 & 0 & \delta_3 & & & \\ \delta_{41} & 0 & 0 & \delta_1 & & \\ 0 & 0 & 0 & 0 & \delta_2 & \\ 0 & 0 & \delta_{63} & 0 & 0 & \delta_3 \end{bmatrix} \quad (4.107)$$

表 4.7: 時系列因子分析の分析結果

Goodness of fit indices	Model 1	Model 2	Model 3	
GFI	0.63	0.75	0.91	
AGFI	0.55	0.62	0.85	
Parameter Estimates				Standardized Value
$\hat{\lambda}_1$	1.0*	1.0*	1.0*	(0.67)
$\hat{\lambda}_2$	2.17	2.29	2.19	(0.96)
$\hat{\lambda}_3$	1.16	1.16	0.97	(0.47)
$\hat{\beta}_{t-1}$	0.74	0.68	0.74	
$\hat{\gamma}_{t-1}$	0.01	0.0*	0.0*	
$\hat{\phi}$	170.6	163.2	166.1	
$\hat{\psi}$	77.4	87.6	74.1	
$\hat{\delta}_1$	199.6	206.8	199.5	
$\hat{\delta}_2$	65.8	12.3	67.4	
$\hat{\delta}_3$	519.4	529.2	532.6	
$\hat{\delta}_{41}$	0.0*	0.0*	77.7	(0.38)
$\hat{\delta}_{52}$	0.0*	0.0*	0.0*	(0.0*)
$\hat{\delta}_{63}$	0.0*	0.0*	385.5	(0.72)

と指定して、再分析を行った。表 4.7 に示した解の GFI は 0.91 であり、AGFI は 0.85 であり、データとモデルの適合は悪くない。

自己回帰係数は $\hat{\beta}_1 = 0.74$ と推定され、かなり大きな影響を過去から受けている。誤差の自己回帰も「一致指標」と「遅行指標」で、それぞれ 0.38 と 0.72 であり、相当に大きい。因子負荷の標準化解は、 $\hat{\lambda}_1 = 0.67$, $\hat{\lambda}_2 = 0.96$, $\hat{\lambda}_3 = 0.47$ と十分に大きく、3 つの時系列の背後に 1 つの潜在した時系列を仮定したことの適切さが示されている。

4.7.1 最近の発展

本章では、入門的話題として定常性を仮定したモデルの紹介のみを行った。ただし定常を仮定したモデルは経時にともなう平均の変化(トレンド)を記述することが難しい。このため回帰分析を行ってから、その残差に定常性を仮定して分

析を行うなどの工夫が必要となる場合も多い(本章で扱った DI という指標は、3ヶ月前との比較の指標なので、定常性の仮定が無理ではない)。しかし最初から定常性を仮定しない非定常モデルもあり、Molenaar, De Gooijer & Schmitz(1992)⁴⁸ などが SEM のソフトで実行可能な記述を紹介している。

また非定常モデルを更に発展させた状態空間モデルと呼ばれる汎用時系列モデルもあり、状態空間モデルに登場する観測方程式と遷移方程式を SEM の測定方程式と構造方程式で表現することも可能である。更に Hamerle, Nagl & Singer(1991)⁴⁹ や Arminger(1986)⁵⁰ では、微分方程式モデル (differential equation models) と SEM との関係について論じている。

⁴⁸Molenaar, P.C.M., De Gooijer, J.G., & Schmitz, B. (1992). Dynamic factor analysis of nonstationary multivariate time series. *Psychometrika*, 57, 333-349.

⁴⁹Hamerle, A., Nagl, W. & Singer, H. (1991). Problems with the estimation of stochastic differential equations using structural equations models. *Journal of Mathematical Sociology*, 16, 201-220.

⁵⁰Arminger, G. (1986). linear stochastic differential equation models for panel data with unobserved variables. In Tuma, N.B. (ed.) *Sociological methodology*. Washington, DC, 187-212.

5 行動遺伝学

本章では、研究領域固有の優れたモデル構成の1つの良き目標として、心理学と量的遺伝学の学際領域における行動遺伝学 (behavior genetics) と呼ばれる研究分野に注目し、この領域で提案され、改良されていった共分散構造モデルを紹介する (Neale and Cardon, 1992, 豊田, 1997⁵¹). 共分散構造モデルが利用される他の多くの研究分野と比較して、行動遺伝学の研究分野では完成度の高い洗練された応用モデルが多数提案されているためである。その理由は、行動遺伝学がその発展の経緯において以下のような特徴的な歴史を持っていることに由来している。

量的遺伝学の研究分野では、共分散構造モデルが提案される相当以前から、観測変数の分散や共分散を遺伝子の関数として表現していた。優れた品種の家畜を作り出すためには、形質の分散・共分散を親の素質や生育環境の関数で構造化する必要があったためである (網羅的文献として Falconer, 1990⁵² がある)。育種が許されない人間行動遺伝学の研究分野でも双生児研究法を利用し、表現形の分散や共分散を構造化する伝統的な研究方法を踏襲していた (Vandenberg, 1965⁵³)。Jöreskog によって LISREL(linear structural relation) の考え方方が導入された 60 年代後半から 70 年代前半にかけて、最尤法の考え方を取り入れられて、ほとんど独自に共分散構造モデルの本質的なアイデアに到達していた (たとえば Vandenberg, 1968⁵⁴)。ただしこの時点では、両者は別々のモデルと考えられていた。

LISREL はその後、短い期間に急速に発展し、行動遺伝学における同様のアプローチの数理的研究水準をすぐに追い抜いてしまった。遺伝現象の解明に主たる関心のある行動遺伝学の研究領域よりも、モデルの性質それ自身に主たる関心が注がれた LISREL ほうが数理的な洗練のスピードが速かったためである。その後 EQS や RAM が提案されると、数学的なモデルの表現力という観点か

⁵¹Neale, M.C., & Cardon, L.R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht, NL: Kluwer Academic Publishers.

豊田秀樹 (1997) 共分散構造分析による行動遺伝学モデルの新展開. 心理学研究, 67, 464-473.

⁵²Falconer, D.S. (1990). *Introduction to quantitative genetics* (3rd ed.). New York: Longman Group Ltd.

⁵³Vandenberg, S. G. (ed.) (1965). *Method and goals in human behavior genetics*. Academic Press: New York.

⁵⁴Vandenberg, S. G. (ed.) (1968). *Progress in human behavior genetics*. Johns Hopkins University Press: Baltimore.

らは、それ等は皆同じモデルであることが解り (豊田, 1990⁵⁵), 共分散構造モデルは当初の予想より相当に一般的な数理モデルであることが明らかになった。その結果、観測変数の分散・共分散を親の素質や生育環境の関数で構造化する人間行動遺伝学のモデルは、共分散構造モデルの下位モデルの1つであることが示されてしまった。

このため行動遺伝学者たちには、共分散構造モデルをほぼ独力で開発しながら、理論的な貢献に関する評価がほとんど与えられていない。現在では他の研究領域と同様に、1つの応用分野として共分散構造モデルをデータ解析の主要なツールの一つとして利用している。しかし LISREL が提案される以前から分散・共分散を母数で構造化し、遺伝率を計算してきた行動遺伝学の研究分野では、完成された LISREL をそのまま導入して道具として利用したその他多くの研究分野と比較して、モデル構成に一日の長があり、構成されたモデルには様々な工夫が凝らされている。行動遺伝学における共分散構造モデルを展望することは、その他の研究分野におけるモデル構成の1つの良き到達目標を与えることになる。

5.1 多変量ACEモデル

行動遺伝学モデルは、すでに1度、入門編第14章において論じられており、そこでは行動遺伝学的考え方を端的にあらわしている遺伝ACEモデルが紹介されている。遺伝ACEモデルは、統計モデルとしては、2つの変数を2つの群で測定しているので4変数のモデルである。しかし実質科学的観点からは1つの特性の性質を調べるモデルであった。本章では、まず遺伝ACEモデルを多変量モデルに拡張する。遺伝ACEモデルを、実質的に多数の特性の性質を同時に分析できるように拡張したモデルが多変量ACEモデルである。

遺伝ACEモデルでは、双生児の特徴を測定した観測変数を表現型(phenotype)と呼ぶ。測定対象は一卵性双生児 (MZ: monozygotic twins) と二卵性双生児 (DZ: dizygotic twins) であり、多くの場合には双子が「双生児1」「双生児2」と呼ばれ、測定対象になる。ただしここでは、どちらかの性の(どちらでも良いが、どちらか一方の)一卵性双生児と二卵性双生児をモデル化の対象とする。

⁵⁵豊田秀樹 (1990). 共分散構造の表現. 教育心理学研究, Vol.38, No.4, p.438-444

5.1.1 モデル構成

b 個の特性

$$\mathbf{p} = [p_1 \ p_2 \ \cdots \ p_b]' \quad (5.1)$$

を考察する場合には、双生児 1 と双生児 2 でそれぞれ 1 回ずつ、一卵性双生児と二卵性双生児の 2 つの群で、計 4 回測定する（平均は 0 に基準化されているものとする）。この変数を群ごとに

$$\mathbf{p}_{MZ} = [\mathbf{p}'_{MZ1} \ \mathbf{p}'_{MZ2}]' \quad (5.2)$$

$$\mathbf{p}_{DZ} = [\mathbf{p}'_{DZ1} \ \mathbf{p}'_{DZ2}]' \quad (5.3)$$

のように表現する。ここで観測変数を加算的遺伝 (\mathbf{a} : additive genetic)・共有環境 (\mathbf{c} : common environment)・非共有環境 (\mathbf{e} : random environment) という 3 つの要因ベクトル（サイズ b の潜在変数）の和

$$\mathbf{p}_{MZ1} = \mathbf{a}_{MZ} + \mathbf{c} + \mathbf{e}_{MZ1} \quad (5.4)$$

$$\mathbf{p}_{MZ2} = \mathbf{a}_{MZ} + \mathbf{c} + \mathbf{e}_{MZ2} \quad (5.5)$$

$$\mathbf{p}_{DZ1} = \mathbf{a}_{DZ1} + \mathbf{c} + \mathbf{e}_{DZ1} \quad (5.6)$$

$$\mathbf{p}_{DZ2} = \mathbf{a}_{DZ2} + \mathbf{c} + \mathbf{e}_{DZ2} \quad (5.7)$$

と考える。これが多変量 A C E モデルのモデル式である。

遺伝要因は直接的に親から受け継いだ遺伝的素質を表す潜在変数である。一卵性双生児は同一の素質 (\mathbf{a}_{MZ}) を受け継ぎ、二卵性双生児は相関はあるけれども異なる素質 ($\mathbf{a}_{DZ1}, \mathbf{a}_{DZ2}$) を受け継ぐ。共有環境要因 (\mathbf{c}) は、いっしょに住む者の表現型を、いっしょに住むが故に類似させる潜在変数である。双生児の概念と直接的には関係ないので、添え字のつかない \mathbf{c} という単一の潜在変数が用いられる。非共有環境要因 (\mathbf{e}) は、各個体がそれぞれに遭遇する独自の環境が、観測変数に与える影響を表現した潜在変数である。

一卵性双生児の個人内では

$$\begin{aligned} E[\mathbf{a}_{MZ} \ \mathbf{c}'] &= E[\mathbf{a}_{MZ} \ \mathbf{e}'_{MZ1}] = E[\mathbf{c} \ \mathbf{e}'_{MZ1}] = E[\mathbf{a}_{MZ} \ \mathbf{e}'_{MZ2}] = E[\mathbf{c} \ \mathbf{e}'_{MZ2}] \\ &= \mathbf{O} \end{aligned} \quad (5.8)$$

の無相関が仮定される。二卵性双生児の個人内では

$$\begin{aligned} E[\mathbf{a}_{DZ1} \ \mathbf{c}'] &= E[\mathbf{a}_{DZ1} \ \mathbf{e}'_{DZ1}] = E[\mathbf{c} \ \mathbf{e}'_{DZ1}] = E[\mathbf{a}_{DZ2} \ \mathbf{c}'] = E[\mathbf{a}_{DZ2} \ \mathbf{e}'_{DZ2}] \\ &= E[\mathbf{c} \ \mathbf{e}'_{DZ2}] = \mathbf{O} \end{aligned} \quad (5.9)$$

が仮定される。また個人間でも (5.8) 式、(5.9) 式の一部の制約に加えて

$$E[\mathbf{a}_{DZ1} \ \mathbf{e}'_{DZ2}] = E[\mathbf{a}_{DZ2} \ \mathbf{e}'_{DZ1}] = E[\mathbf{e}_{MZ1} \ \mathbf{e}'_{MZ2}] = E[\mathbf{e}_{DZ1} \ \mathbf{e}'_{DZ2}] = \mathbf{O}$$

が仮定される。

一方、各要因内では、共分散行列

$$E[\mathbf{a}_{MZ} \ \mathbf{a}'_{MZ}] = E[\mathbf{a}_{DZ1} \ \mathbf{a}'_{DZ1}] = E[\mathbf{a}_{DZ2} \ \mathbf{a}'_{DZ2}] = \Sigma_a \quad (5.10)$$

$$E[\mathbf{c} \ \mathbf{c}'] = \Sigma_c \quad (5.11)$$

$$E[\mathbf{e}_{MZ1} \ \mathbf{e}'_{MZ1}] = E[\mathbf{e}_{MZ2} \ \mathbf{e}'_{MZ2}] = E[\mathbf{a}_{DZ1} \ \mathbf{e}'_{DZ1}] = E[\mathbf{a}_{DZ2} \ \mathbf{e}'_{DZ2}] = \Sigma_e \quad (5.12)$$

が仮定される。二卵性双生児の遺伝子は互いに半分が共有されているので、ポリジーンの考え方を導入し、

$$E[\mathbf{a}_{DZ1} \ \mathbf{a}'_{DZ2}] = 0.5 \times \Sigma_a \quad (5.13)$$

を仮定する。

以上の仮定より、観測変数の共分散行列は

$$\Sigma_{MZ} = E[\mathbf{p}_{MZ} \ \mathbf{p}'_{MZ}] = \begin{bmatrix} \Sigma_{MZ1} & \\ \Sigma_{MZ21} & \Sigma_{MZ2} \end{bmatrix} \quad (5.14)$$

$$= \begin{bmatrix} \Sigma_a + \Sigma_c + \Sigma_e & \\ 1.0 \times \Sigma_a + \Sigma_c & \Sigma_a + \Sigma_c + \Sigma_e \end{bmatrix} \quad (5.15)$$

$$\Sigma_{DZ} = E[\mathbf{p}_{DZ} \ \mathbf{p}'_{DZ}] = \begin{bmatrix} \Sigma_{DZ1} & \\ \Sigma_{DZ21} & \Sigma_{DZ2} \end{bmatrix} \quad (5.16)$$

$$= \begin{bmatrix} \Sigma_a + \Sigma_c + \Sigma_e & \\ 0.5 \times \Sigma_a + \Sigma_c & \Sigma_a + \Sigma_c + \Sigma_e \end{bmatrix} \quad (5.17)$$

のように、3 つの要因に対応する共分散行列で構造化される。もちろん一卵性双生児と二卵性双生児の間の共分散行列は計算できない。以上が、多変量遺伝 A C E モデルの特定と、その共分散構造である。

観測変数としては、知能・性格・血圧…等、遺伝と環境の影響を受ける全ての間隔尺度に適用可能な汎用モデルである。 Σ_a , Σ_c , Σ_e は、各変数の背後に別々に仮定した遺伝要因、共有環境要因、非共有環境要因の共分散行列である。共分散構造は、1変数の遺伝ACEモデルのスカラーの場所を行列で置き換えたものとなっている。各変数を個別に分析するのではなく多変数を同時に分析することによって、遺伝、共有環境、非共有環境の影響の程度を考察するばかりでなく、各要因の要因内における関係を考察することが可能になる。

このモデルのパラメータである3つの要因の共分散行列を推定するためには

$$\Sigma_a = \mathbf{C}_a \mathbf{C}'_a \quad (5.18)$$

$$\Sigma_c = \mathbf{C}_c \mathbf{C}'_c \quad (5.19)$$

$$\Sigma_e = \mathbf{C}_e \mathbf{C}'_e \quad (5.20)$$

のように各行列を下三角行列とその転置行列の積で表現するのが一般的である⁵⁶。ただし多変量ACEモデルを汎用分析モデルとして利用する場合には \mathbf{C}_a , \mathbf{C}_c , \mathbf{C}_e の係数を直接解釈してはならない。それらの表現は飽くまでもモデルを識別させるために行なっているのであり、解釈は Σ_a , Σ_c , Σ_e によって行なう。まず共分散構造モデルのソフトウェアで \mathbf{C}_a , \mathbf{C}_c , \mathbf{C}_e を推定し、その推定値を右辺に代入して、要因内の共分散行列の推定値を得る。

5.1.2 適用例

表5.1は、Neale and Cardon (1992, p254) で示された四分相関係数行列である。オーストラリア人の双生児から4種類の疾患(喘息、花粉症、埃アレルギー、湿疹)の経験の有無を調査し、各双生児をランダムに双生児1と双生児2に振り分け計算している。下三角部分が一卵性双生児、上三角部分が二卵性双生児の四分相関係数である。

⁵⁶左辺の右辺による表現をコレスキーディクションという

表 5.1: 各疾患の四分相関行列

双生児 1 喘息 1	1.000	.524	.588	.291	.262	.129	.079	.217
花粉症 1	.556	1.000	.749	.314	.170	.318	.171	.114
埃アレルギー 1	.573	.758	1.000	.279	.041	.262	.214	.087
湿疹 1	.273	.264	.309	1.000	.139	.093	.019	.313
双生児 2 喘息 2	.592	.366	.398	.232	1.000	.395	.684	.254
花粉症 2	.411	.593	.451	.145	.549	1.000	.723	.218
埃アレルギー 2	.434	.421	.518	.192	.640	.770	1.000	.276
湿疹 2	.087	.196	.193	.589	.145	.122	.218	1.000

下三角部分：一卵性双生児 (1232組), 上三角部分：二卵性双生児 (751組)

モデル中の母数の推定値は以下のようになる⁵⁷.

$$\hat{\Sigma}_a = \begin{bmatrix} 0.4359 & 0.3028 & 0.4336 & 0.0873 \\ 0.3028 & 0.5464 & 0.4119 & 0.1876 \\ 0.4336 & 0.4119 & 0.4925 & 0.2341 \\ 0.0873 & 0.1876 & 0.2341 & 0.5136 \end{bmatrix} \quad (5.21)$$

$$\hat{\Sigma}_c = \begin{bmatrix} 0.1557 & 0.0461 & -0.0355 & 0.1000 \\ 0.0461 & 0.0499 & 0.0212 & 0.0072 \\ -0.0355 & 0.0212 & 0.0357 & -0.0424 \\ 0.1000 & 0.0072 & -0.0424 & 0.0781 \end{bmatrix} \quad (5.22)$$

$$\hat{\Sigma}_e = \begin{bmatrix} 0.4153 & 0.1651 & 0.2251 & 0.0481 \\ 0.1651 & 0.4047 & 0.3200 & 0.0308 \\ 0.2251 & 0.3200 & 0.4767 & 0.0793 \\ 0.0481 & 0.0308 & 0.0793 & 0.4107 \end{bmatrix} \quad (5.23)$$

多変量ACEモデルは、1変量のACEモデルの分析結果を含んでいる。たとえば喘息は、遺伝と共有環境と非共有環境から、それぞれ43.6%, 15.6%, 41.5%説明される。花粉症は、それぞれ54.6%, 5.0%, 40.5%説明される。埃アレルギーは、それぞれ49.3%, 3.6%, 47.7%説明される。湿疹は、それぞれ51.4%, 7.8%, 41.1%説明される。しかし、そればかりでなく、各要因内の共通変動要因を解釈

⁵⁷四分相関係数を通常のピアソンの相関係数と見なした場合の最尤推定値を示した。入門編第12章で論じたように、0-1データからは四分相関係数とその漸近分散を求め、それを利用して漸近有効な推定量を求める方法が利用されている。しかし、そのためには元データか、あるいは四分相関の漸近分散が必要である。ここでは読者が追計算が可能になるように通常の最尤推定法を利用した。

することができる。たとえば Σ_a 中の喘息、花粉症、埃アレルギーに関連した非対角要素には、比較的大きな共分散が観察され、3つの疾患の背後には共通した遺伝的要因が存在することが示唆される。花粉症と埃アレルギーに対する非共有環境の共分散も 0.32 と小さくはなく、共通した原因が示唆される。

5.2 遺伝因子分析

多変量 ACE モデルは、要因ごとの共分散行列を推定するモデルであるから、特性の数が多くなると相互の関係を解釈するのが困難になる。そこで (5.18) 式から (5.20) 式の共分散構造中の要因毎の共分散行列に

$$\Sigma_a = \Lambda_a \Lambda'_a + \Psi_a \Psi_a \quad (5.24)$$

$$\Sigma_c = \Lambda_c \Lambda'_c + \Psi_c \Psi_c \quad (5.25)$$

$$\Sigma_e = \Lambda_e \Lambda'_e + \Psi_e \Psi_e \quad (5.26)$$

のような因子分析モデルを導入する。このように要因内の変数の関係を縮約的に表現したモデルが遺伝因子分析モデル (genetic factor analysis model) である。ここで右辺の第 1 項は因子負荷行列による共通因子空間であり、第 2 項は対角成分に独自性の係数を配した対角行列の 2 乗である。遺伝因子分析モデルは 1 つの観測変数に 3 つの独自成分が影響を与えるので、どれかが 0 以下になってしまふ場合が少なくない。通常の因子分析では独自因子の分散を推定するが、ここでは分散を 1 に固定して、その代わりに係数を推定する。

各要因の因子数が 1 である場合の、もっとも単純な遺伝因子分析モデルを図 5.1 に示す。ただし表現されているのはどちらか一方の双生児の 1 名分である(実際にはこのモデルの場合には 16 個の変数がある)。多くの行動遺伝学のモデルは類似した 4 つ(2 名 \times 2 組)のパートに分かれしており、ほとんどのモデルは、1 人分の変数あるいは 1 組の双生児分の変数に関係したパス図に全ての母数が登場する。そこで行動遺伝学の論文では、紙面の節約のために 1 人分だけ、あるいは 1 組分だけをパス図で表現することが多い。

行動遺伝学モデルは理解しにくいといわれることがある。それは通常のパス図の表現を見慣れた者にとって行動遺伝学モデルがおよそ識別されないモデルに見えてしまうためである。たとえば図 5.1 は、4 つの観測変数から 3 つの共通因子を抽出し、なおかつ 12 個の独自成分があるので、パス図の原則的なルール

表 5.2: 遺伝因子分析モデルの母数の推定値

	Λ_a	Ψ_a	Λ_c	Ψ_c	Λ_e	Ψ_e
喘息	0.66	0.00	0.37	0.00	0.32	0.57
花粉症	0.59	0.48	-0.07	0.00	0.47	0.44
埃アレルギー	0.70	0.00	-0.18	0.00	0.69	0.00
湿疹	0.27	0.69	0.06	0.19	0.12	0.63

表 5.3: 6 要因の説明割合

	Λ_a	Ψ_a	Λ_c	Ψ_c	Λ_e	Ψ_e
喘息	44%	0%	14%	0%	10%	32%
花粉症	35%	23%	0%	0%	22%	19%
埃アレルギー	50%	0%	3%	0%	47%	0%
湿疹	7%	47%	0%	4%	1%	40%

まるめ誤差の影響で和は正確に 100% にはならない

からは全く識別されないように見える。しかし実際にはこの 4 倍の観測変数が分析の対象になっている。

ここでは遺伝要因、共有環境要因、非共有環境要因それぞれに、共通因子を 1 つ仮定してモデルの母数を推定し、表 5.2 に推定値を示した。図 5.1 から明らかなように、各観測変数は 6 つの互いに無相関な分散 1 の潜在変数の重み付き和で表現されている。このため推定された係数の 2 乗和を計算すれば 1.0 になり、2 乗値は各因子からの説明割合となる。表 5.3 には百分率で表現した説明率を示した。

ACE の観点から考察すると、遺伝要因・共有環境要因・非共有環境要因の説明比率は、それぞれ喘息で 44%・14%・42%，花粉症で 58%・0%・41%，埃アレルギーで 50%・3%・47%，湿疹で 54%・4%・41% である(丸め誤差の影響で和は正確に 100% にはならない)。4 つの疾患に共通していることは、共有環境要因の説明比率が小さいことである。いっしょに住んでいるが故に共変して発現する疾患とはいえないということである。

遺伝要因内の説明割合を考察すると喘息・花粉症・埃アレルギーは共通因子で説明される割合が比較的高く、湿疹は低い。これは喘息、花粉症、埃アレル

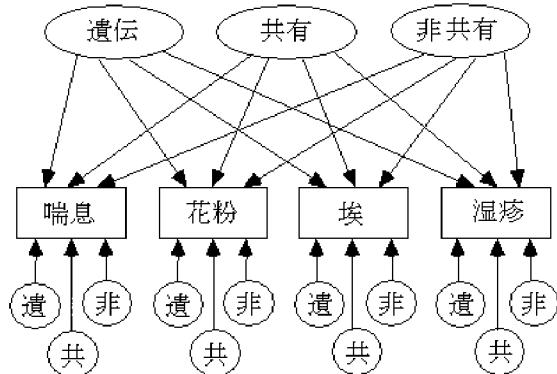


図 5.1: 遺伝因子分析モデル

ギーが、遺伝要因の中でも共通した原因によって引き起こされ、湿疹は独自の原因によって引き起こされるという解釈となる。喘息、花粉症、埃アレルギーは気管を通じて進入する物質に影響される疾患であり、共通因子は「気管因子」と命名できるかもしれない。

本節で分析された4つの疾患に皮膚病を幾つか加えれば、湿疹といっしょに第2因子を形成する可能性がある。このように(5.24)式から(5.26)式で分解された因子分析モデルは、1因子モデルに限定されるものではない。もちろん第2因子以上の因子を抽出する場合には、一般的な因子分析モデルと同様に、回転の不定性の問題が生じる。通常は、第2章で論じたように因子負荷行列の第m列に値が0のm-1個の固定母数を(多くの場合に右上に寄せて)設定して、識別問題を回避し、推定値を得てから目的に応じて回転する。

5.3 双生児と一般児の統合的モデル

双生児法による遺伝ACEモデルは、新しい知見を提供する可能性を持った手法であるが、現在、行動遺伝学の研究分野で一般的に用いられている双生児法には方法論的に2つの問題点がある。

5.3.1 双生児法の問題点

第1の問題点は、結果の安定性を確保し難いことである。行動遺伝学の研究分野では、主として双生児から収集した観測変数の共分散を、遺伝や環境の分散で記述し、それらを推定するというパラダイムを用いる。ACEモデルは母集団の確認的因子分析モデルであるから、分散分析を使ったモデルよりも標本数がたくさん必要である。ところが双生児研究では多くのデータを集めることが他の研究分野より難しい。

1つの解決策として、前節で紹介した遺伝因子分析を使用する方法が挙げられる。遺伝因子分析法はACEモデルと因子分析法を統合した統計モデルであり、共分散構造モデルの下位モデルとして記述されている。この方法を用いることにより、多くの観測変数から少数の分散成分を推定することが可能になり、遺伝率の計算に用いる母数の数は標本共分散の数に比べて相当少なくなる。しかし、貴重な双生児データを用いて遺伝率と同時に因子負荷を推定しなくてはならないので、遺伝率に関する母数を減らした効果は直接的には享受できない。

第2の問題点は、「双生児ゆえの特異性を認めない」という仮定を暗黙のうちに導入していることにある。双生児研究から得られた心理的・教育的知見を一般児(non-twin)まで一般化するためには、1人の双生児から計算した複数の観測変数の共分散行列は、一般児のそれと同じであるという理論的な仮定を認めなくてはならない。もし双生児の共分散行列と一般児の共分散行列の乖離が激しい場合には、双生児であるという属性が共分散に強く反映する観測変数ということになる。このため、その分析結果は、文字どおり双生児を分析した結果ということになり、一般児に関する知見を導くことは控えなくてはならない。

現実的には、双生児と一般児の共分散行列が正確に一致することはないので、乖離の悪影響は程度の問題となる。乖離が小さければ双生児統制法は遺伝と環境の影響を調べるのに有効であるし、乖離が大きければ有効でなくなる。しかし現状ではこの乖離の程度は、確認されずに分析され、考察されることが多い。

双生児統制法に関するこれら2つの欠点を同時に補うために、豊田・村石(1998)⁵⁸で提案された双生児と一般児のデータを統合的に扱う方法がある。具体的には遺伝因子分析モデルに注目し、因子構造に関しては双生児と一般児の因子負荷および因子間共分散に等価の制約を置いて母数を推定する。解を安定させることができ程の一般児の標本を確保することは、双生児の標本を同程

⁵⁸ 豊田秀樹・村石幸正(1998) 双生児と一般児による遺伝因子分析－Y-G性格検査への適用－教育心理学研究, 46, 255-261.

度確保することと比較して、ずっと容易である。遺伝因子分析モデルによって遺伝率に関する母数を減らし、しかも因子負荷の推定に関しては一般児のデータを利用して、分析結果を安定させることにより、第1の問題点を克服する。

同じ制約が、第2の問題点を同時に解決する理由は以下の通りである。双生児と一般児の因子負荷・因子間共分散が等しいという制約は、両者の個人内共分散行列が等しいという制約と同じであり、この制約の下で求められた解の適合度や標準化残差を確認することによって制約が妥当であるが否かを確認できる。標準化残差が小さい場合は、両者の共分散行列の乖離が少ないことを意味し、双生児データから推定された遺伝率による心理・教育的知見を一般児にまで広げて解釈できることの直接的な1つの証拠となる。逆に標準化残差が大きい場合には、両者の共分散行列の乖離が大きいことを意味し、その観測変数に関しては、分析結果を一般児にまで広げて解釈することはできないと判断する。このように第2の問題点に対処する。

5.3.2 Y-G 性格検査への適用

Y-G 性格特性の因子分析的研究は、続・織田・鈴木(1971)⁵⁹、玉井・田中・柏木(1985)⁶⁰、国生・柳井・柏木(1990)⁶¹など、尺度得点ではなく項目得点を基礎とした論文が多い。近年では福田・谷嶋・斎藤(1995)⁶²によって尺度得点を用いた因子分析結果が発表されており、ここでは2因子モデルが採用され、T尺度の分類が通常の類型判定基準に合わないことが報告されている。

本分析では、中学・高等学校の教育現場で実施されている類型判定基準の根拠となる2因子モデル、つまり D, C, I, N, O, Co の6尺度の背後に「情緒性の適応性」因子を仮定し、Ag, G, R, T, A, S の6尺度⁶³の背後に「広義の向性」因子を仮定するモデル(谷田部, 1993⁶⁴)を採用し、「広義の向性」に関して分析

⁵⁹ 続有恒・織田輝準・鈴木真雄(1971) 質問紙法による性格診断の方法論的吟味 II 教育心理学研究, 19, 21-33.

⁶⁰ 玉井寛・田中芳美・柏木繁男(1985) 項目単位の因子分析によるY-G テストの次元性の確証 心理学研究, 56, 292-295.

⁶¹ 国生理枝子・柳井晴夫・柏木繁男(1990) 新性格検査における併存的妥当性の検証 心理学研究, 61, 31-39.

⁶² 福田将史・谷嶋喜代志・斎藤朗(1995) YG 性格検査の再検討(2)-項目と尺度の相関と因子について- 日本心理学会第59回大会発表論文集, 43.

⁶³ Ag尺度-愛想の悪さ(Lack of Agreeableness)尺度、G尺度-一般的活動性(General Activity)尺度、R尺度-のんきさ(Rhathymia)尺度、T尺度-思考的外向(Thinking Extraversion)尺度、A尺度-支配性(Ascendance)尺度、S尺度-社会的外向(Social Extraversion)尺度

⁶⁴ 谷田部順吉(1993) 矢田部ギルフォード性格検査 岡堂哲雄(編) 心理テスト入門 こころの

を進める⁶⁵。

ACE モデルと因子分析モデルを統合する方法には大別して2種類ある。1つは前節で紹介した ACE の各要因を因子分析モデルで記述する方法である。もうひとつは、各因子を ACE モデルで記述する方法である。ここでは「広義の向性」という因子の遺伝規定性を調べることが目的であるから後者の方法を用いる。

表5.4は一卵性双生児、二卵性双生児、一般児の「広義の向性」を測る6尺度の相関行列⁶⁶である。

表 5.4: 広義の向性を測る 6 尺度の相関行列 (双生児)

Ag1	1.00	0.59	0.61	-.32	0.50	0.54	0.02	-.04	0.12	-.03	0.12	0.04
G1	0.25	1.00	0.40	-.09	0.66	0.64	0.01	0.19	0.17	0.17	0.33	0.30
R1	0.56	0.32	1.00	0.13	0.41	0.36	-.05	-.08	-.07	0.17	-.03	0.10
T1	-.06	0.08	0.26	1.00	-.08	0.09	-.34	-.10	-.32	0.25	-.34	-.17
A1	0.46	0.47	0.49	0.14	1.00	0.67	0.10	0.27	0.11	0.06	0.37	0.34
S1	0.39	0.48	0.52	0.26	0.79	1.00	0.04	0.28	0.11	0.24	0.27	0.30
Ag2	0.28	0.11	0.08	-.09	0.13	0.13	1.00	0.53	0.45	-.27	0.55	0.47
G2	0.15	0.30	0.18	-.01	0.29	0.25	0.23	1.00	0.37	-.04	0.71	0.71
R2	0.27	0.15	0.35	0.04	0.21	0.16	0.43	0.26	1.00	0.05	0.44	0.52
T2	0.00	0.00	0.15	0.32	0.10	0.12	-.13	-.02	0.18	1.00	-.08	0.15
A2	0.24	0.25	0.22	0.03	0.50	0.50	0.26	0.52	0.24	0.05	1.00	0.75
S2	0.18	0.25	0.26	0.11	0.44	0.50	0.21	0.55	0.35	0.20	0.73	1.00

下三角部分: 一卵性双生児

上三角部分: 二卵性双生児

図5.2には、まず一般児のパス図を示した。右側には広義の向性を測る6尺度から測定した観測変数が配置されている。影響指標には特に名前を付けていない。単純構造であるために、たとえば、「Ag尺度への影響指標」は1つしかないからである。左側には、因子ごとにACEの各要因が配置され、性格因子を規定している。4つの構成概念の分散は全て1に固定されている。母数は全て自由母数であり係数の意味は以下の通りである。性格因子への規定力は

科学増刊 日本評論社, 61-65.

⁶⁵ 2因子を同時に分析した結果に関しては、前掲 豊田・村石(1998)を参照されたい。

⁶⁶ 「測定対象」東京大学教育学部附属中・高等学校に1977年から1994年まで入学した生徒。ただし1979年、1990年、1992年に入学した生徒のデータは存在しない。「測定方法」矢田部・ギルフォード(Y-G)性格検査の実施要領に従い、当該性格検査を集合回収形式で実施「抽出方法」全数調査「データ形式」12の性格尺度の尺度得点「有効回収数」一卵性双生児-男性 87組・女性 100組、二卵性双生児-男性 11組・女性 13組・異性 20組、一般児-男性 670名・女性 639名

表 5.5: 広義の向性を測る 6 尺度の相関行列 (一般児)

Ag	1.00
G	0.28 1.00
R	0.26 0.24 1.00
T	-.07 0.21 0.44 1.00
A	0.37 0.59 0.19 0.11 1.00
S	0.35 0.51 0.35 0.32 0.83 1.00

a 加算的遺伝要因が「向性」を規定する程度

c 共有環境要因が「向性」を規定する程度

e 非共有環境要因が「向性」を規定する程度

である。要因間の相関は、通常の ACE モデルに準じて設定しない。

ただしこれら 3 個の母数は、図 5.2 のパス図だけでは識別されない。その代わり因子負荷である 6 個の影響指標は、一般児と双生児を含め 1771 名のデータによって安定期に推定される。

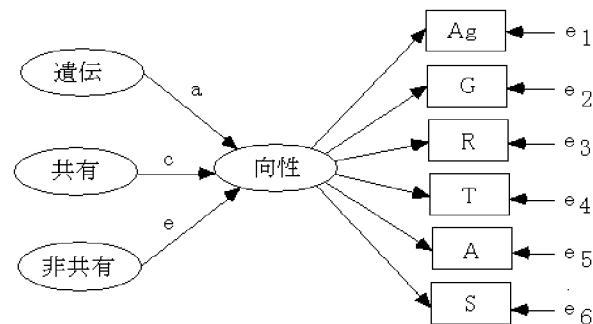


図 5.2: 遺伝因子分析モデル (一般児)

図 5.3 には、双生児のパス図を示した。両脇にそれぞれの双生児の尺度を 6 個、合計 12 変数が配置されている。観測変数・構成概念のラベルの後の数字は、ランダムに割り当てた双生児の番号である。共有環境要因には双生児の番号は付かない。性格因子への規定力、要因内の相関、および各尺度の影響指標は、双生児ばかりでなく一般児とも共通している。一般児のパス図にない母数 X は、遺伝要因内の双生児間の相関である。一卵性双生児のデータでは値が 1.0、二卵性双生児のデータでは値が 0.5 に固定される。

児ばかりでなく一般児とも共通している。一般児のパス図にない母数 X は、遺伝要因内の双生児間の相関である。一卵性双生児のデータでは値が 1.0、二卵性双生児のデータでは値が 0.5 に固定される。

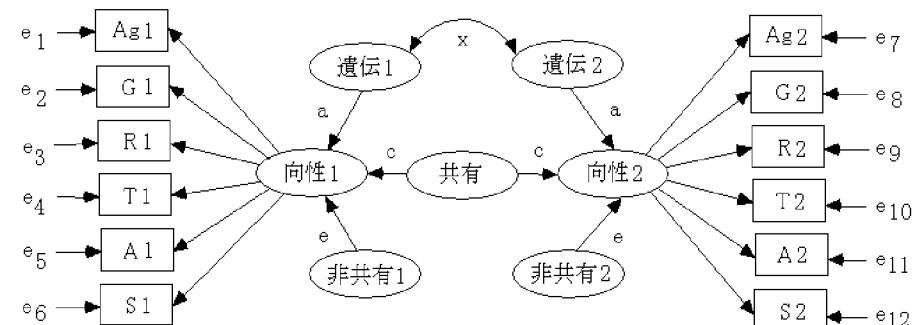


図 5.3: 遺伝因子分析モデル (双生児)

表 5.6: 因子負荷の推定値

	Ag	G	R	T	A	S
	0.49	0.68	0.47	0.12	0.86	0.87

表 5.7: 各要因からの説明率

加算的遺伝	共有環境	非共有環境
0.70	0.36	0.62

表 5.6 に、各尺度に対する各群共通の因子負荷の推定値を示す。福田・谷嶋・斎藤 (1995) の表 2 で示されたように T 尺度は、類型判定基準に合わないことが再確認されている。表 5.7 に、各性格因子への各要因からの説明率の推定値を示す。遺伝的影響・共有環境・非共有環境は、「広義の向性」因子の分散を、それぞれ 49%, 13%, 38% 説明している。広い意味での気質である「広義の向性」

は、遺伝的規定性が半分ほどある。逆に共有環境の影響は約1割であり、とても小さい。

5.4 優れたモデルの構成のために

前節までに、行動遺伝学の研究の中で提案してきた幾つかの特徴的な共分散構造モデルを概観した。それらのモデルは「多母集団」「等値の制約」「固定母数」等が組み合わされ、行動遺伝学特有の測定状況・研究目的を表現していた。本節では前述の行動遺伝学モデルを例に取り、他の多くの研究分野で共通して利用できる「モデル構成のノウハウ」を考察する。

5.4.1 反復測定・多群の測定と等値母数

遺伝 ACE モデルでは、1つの表現型(特性)の性質を調べるために少なくとも4回(2人×2群)、あるいは8回(4回×2世代)もの測定を行っていた。このため、たとえば遺伝ACEモデルでは、4つの変数の分散構造が一致するし、遺伝因子分析モデルでは各個人内の4つの共分散行列の構造が一致している。モデル構成に利用できる母数の数は、高々、構造が同じでない標本分散・共分散の数である。このため母数の推定の観点から考えると、行動遺伝学モデルは相当に無駄を含んだモデルといえる。ただし通常は、丁度識別のモデルは統計的安定性の考察ができないことが多いけれども、行動遺伝学モデルではそれが可能である。たとえば遺伝ACEモデルでは自由度が3であり統計的な安定性を考察できる。同じ構造の分散・共分散があるといって、それらは自由度として数えられるから、制限の範囲でモデルを表現できれば欠点とはならない。むしろ現象を的確に記述するためのモデル構築という観点からは、次の様な理由で、行動遺伝学モデルは望ましい母数配置を採用している。

社会－人文－行動科学の研究分野では、モデルの適合度が高く、母数の標準誤差が小さくても、必ずしも安定した知見を得られるとは限らない。同じ内容のデータを取り直して同じモデルで分析すると、結果が大きく変わって安定しないことをしばしば経験させられる。理学や工学分野の測定と比較して、データを取り直す際に微妙に測定状況が変化するためである。標準誤差や適合度指標は同一母集団からデータを抽出したときの指標であるから、測定状況が微妙に変化したときの結果の安定性は、それらの指標から直接読み取ることはでき

ないのである。

行動遺伝学モデルでは測定状況を微妙に変化させながら同一の表現型を複数回測定し、しかも等価の制約の付いた母数で表現する。本章の多変量 ACE モデルでは2群の分析のみを紹介したが、実際の双生児には5つのパターンがあるし、地域や年齢や人種などを層別要因とすると群の数はどんどん増えてゆく。それらのデータに共通の遺伝率を仮定し、その分析結果の適合度が高かったならば得られた知見は微妙な測定状況の変化に頑健なものとして利用することができる(もちろん、たとえば性別によって遺伝率が明らかに違う場合は異なる母数を設定する必要がある)。得られる知見が微妙な測定の変化に頑健でなければ、同じ母数で表現することに無理が生じて適合度は必ず下がる。行動遺伝学モデルは、1回の分析で、交差妥当性や頑健性を検討しているのである。

この長所を行動遺伝学以外の研究に敷延するならば、たとえば n 個の項目から構成される心理検査の因子構造を調べる場合に、検査を m の地域でそれぞれ l 回実施するということが考えられる。サイズ $l \times n$ の m 個の標本共分散行列に同一の項目に関する $l \times m$ 個の因子負荷と独自性は互いに等しいという制約の下で因子分析を行う。このとき適合度指標の値が高ければ、そこからの解釈は相当に安定したものとなるはずである。

一般化された共分散構造分析を行う場合も同じ考え方を利用できる。通常は、1つの構成概念を測定していると考えられる複数の観測変数を用意する。これは結果の頑健性の向上に寄与している。しかし、変数毎に異なった影響指標や形成指標を推定するのではなく、さらに条件を厳しくして繰り返し測定を行い、等値の母数を配置したらどうだろうか。制約が強くなるのでモデル探索をしても適合度を上げることは難しくなるだろう。しかし、そのような制約の下で適合度の高いモデルを構成できたならば、そこからの知見は頑健なものとなるし、母数が少ないので知見を利用し易くなる。

5.4.2 固定母数の利用

統計モデルはデータの性質を調べるための道具であるが、データに未知な部分が多く過ぎると安定的な解析を行えない。しかし、このとき実質科学的立場からの補助情報である固定母数を導入することができれば、データの未知な部分が相対的に少なくなり、より複雑なモデルの構成が可能になる。遺伝因子分析モデルは、1組の変数群に対して通常の因子分析モデル3つを同時に適用したの

と同じ情報量を提供している。条件を変えて、母集団の数を増やすだけではモデルは識別されないが、一卵性と二卵性を区別するたった2つの固定母数(1.0と0.5)を導入したために識別が可能になっている。

この事実は「母集団および変数群の特徴を記述するための少數の固定母数を導入することによって、想像以上に複雑なモデルを構築することが可能になる」という共分散構造モデルの性質を示している。因子分析を行う際にも、統計モデルのみにたよって因子負荷を全て推定させるのではなく、興味の対象となっている特性の意味から因子負荷の一部を0に固定したり、手引書に掲載された信頼性係数の値を利用して影響指標と独自性を固定する等の工夫は有効である。もちろん、思い込みに過ぎない不適切な固定母数を設けることは好ましくないが、それとて適合度指標やLM検定等によって間違いを修正することが可能であるから、固定母数はもっと積極的に使用するべきである。

5.4.3 多母集団モデルの利用法

構造方程式の集まりとして表現される統計モデルの母数は、推定のし易さという観点から、互いに対等ではない。たとえば、遺伝因子分析モデルでは、遺伝要因からの係数が双生児のデータからしか推定できないのに対して、因子負荷は一般児のデータからも双生児のデータからも推定できる。因子負荷のほうが明らかに推定しやすい。双生児と一般児の統合的モデルでは、双生児のデータだけでモデル中の全ての母数を推定するのではなく、双生児のデータと大量の一般児のデータを用いて因子負荷を推定することによって、モデル全体の母数の推定を安定させている。

また、たとえば鈴木・長田(1998)⁶⁷の企業評価モデルPRISMでは、まず標本を集めやすい企業の状態に関する変数によって推定される母数と、標本を集めにくい企業の評価に関する変数によって推定される母数とを区別する。そして評価変数を含まない大量の標本と評価変数を含む小数の標本を、多母集団モデルで同時に分析し、効率よく評価構造を導いている。多母集団モデルは、母集団の異同を比較する目的だけで使用されるものではない。このようにデータの集め易さの事情を反映した分析法は、他の分野でも極めて有益であり、積極的に導入を図る価値がある。

⁶⁷ 鈴木督久・長田公平(1998)企業モデルPRISMの開発 豊田秀樹編 共分散構造分析<事例編>—構造方程式モデリング— 第3章 北大路書房

6 上限と下限のあるデータの分析

提案された当初の 70 年代前半から SEM の母数は、最尤法で推定するのが標準であった。最尤法は推測統計的に優れた多くの性質を持っているが、確率変数に特定の分布を仮定しなくてはならない。現在のところ SEM の最尤法は、観測変数に多変量正規分布を仮定する方法が主流である。しかし SEM が多用される研究分野を考慮すると、必ずしも多変量正規分布の仮定が厳密に成立とは言い難い。標本が多くなった場合には、大数の法則により標本共分散行列は共分散行列に近づく。したがって推定値に関しては、最尤法における多変量正規分布の仮定は頑健であり、その使用に際して過度に神経質になる必要がないことが知られている。しかしその一方で、多変量正規分布からの逸脱が著しいデータに対しては、以下のような 3 つの方向の対処がある。

1 つは分布に強く依存しない母数の推定法による対処である。楕円分布を仮定した楕円分布法や 4 次のモーメントを利用した ADF 法によって多変量正規分布しないデータの分析を行う。2 つ目は、第 1 章で紹介したように外生的な観測変数を非確率変数として扱う対処である。外生的な観測変数で説明されない残差が正規分布するという仮定は、観測変数そのものが正規分布するという仮定より相当に緩やかである。

3 つ目は、正規分布しない理由を統計モデルとして記述し、それをモデルの 1 部に組み込む対処である。入門編第 12 章で紹介した順序データの扱いがその 1 つの例といえる。そこでは閾値という考え方を導入し、正規分布しない理由を統計モデルとして記述してから解を求めていた。同様なモデルとして有効なものに上限と下限のあるデータ⁶⁸の扱いがある。本章では、SEM の SW で既に実用の域に達している上限と下限のあるデータの扱いについて論じていく。

6.1 トービット変数

表 6.1 は観測変数の数が 5、標本数 100 の多変量データを 3 つに分割して示したものである。人工データであるが、データの背後には次のようなシナリオが設定されている。ある講義は同一のクラスを 2 人の先生が担当している。 u_1 は A 先生の授業内試験の結果である。A 先生の試験は易しく、多くの学生が満点である 100 点をとってしまった。 u_2 は B 先生の授業内試験の結果である。

⁶⁸ 正規分布しない理由を記述したモデルとして、近年、発展が著しいものに SEM における切断データの解析もある。しかし現時点ではまだ実用的に扱いやすいとはいえないで紹介は割愛する。

易しくなりすぎてはいけないと、問題を難しくしたところ最低点である 0 点を取る学生がたくさん出てしまった。そこで 2 人で相談して問題を作り、期末試験（変数では u_3 ）を実施したところ、満点もゼロ点もいなくなった。

u_4 と u_5 は、それぞれ A 先生と B 先生の最終評価であり、3 点 – 「優」、2 点 – 「良」、1 点 – 「可」、0 点 – 「不可」を意味している。素点を眺めると A 先生は甘く、B 先生は厳しいようである。ただし u_1 から u_5 は単一の授業の評価得点であるから、背後に「教科学力」という単一の構成概念を仮定した 1 因子モデルが期待される。

表 6.1: 5 つの評価変数

u_1	u_2	u_3	u_4	u_5	u_1	u_2	u_3	u_4	u_5	u_1	u_2	u_3	u_4	u_5
87	7	46	1	1	100	10	64	2	2	95	0	37	2	0
81	7	18	1	1	72	5	54	1	0	100	25	56	0	1
76	0	29	0	0	100	22	66	2	1	89	7	38	2	0
70	16	52	3	0	95	3	47	0	2	86	11	34	0	2
100	24	56	2	2	78	0	16	0	0	85	9	48	2	1
85	28	44	2	1	75	24	43	3	1	85	2	44	2	0
73	2	60	2	0	74	0	47	0	0	100	22	49	2	3
100	2	58	0	1	100	19	84	3	3	100	7	75	2	1
100	0	63	1	0	93	10	32	3	1	96	2	40	3	0
87	11	47	2	2	100	17	46	1	1	93	11	65	2	3
100	10	83	3	1	91	13	56	3	1	78	11	64	1	0
81	0	33	0	0	81	0	68	3	2	100	13	48	3	2
97	0	67	2	1	82	0	50	1	2	100	12	66	3	0
85	0	25	2	0	100	13	37	1	1	83	15	48	2	0
85	5	53	0	0	64	4	18	1	0	100	4	56	0	1
87	11	40	1	0	100	4	38	2	0	100	6	48	3	1
100	0	40	2	0	97	27	71	1	2	100	0	35	0	2
77	7	40	2	0	86	6	30	1	0	92	0	31	0	1
94	16	45	2	0	100	20	60	3	2	98	24	56	3	1
71	7	32	2	1	87	1	48	0	0	94	0	59	2	1
100	21	53	2	2	64	0	47	1	1	76	5	41	2	2
100	18	56	3	1	83	8	55	1	0	83	0	38	0	1
95	10	42	0	0	100	11	47	2	1	100	36	57	3	2
100	16	50	2	2	88	2	39	0	0	83	6	33	2	2
77	0	22	0	0	83	12	63	1	1	100	18	49	2	1
100	26	40	3	2	100	11	75	3	2	100	21	67	3	1
100	15	47	1	1	97	10	36	0	1	100	10	60	2	0
100	9	53	1	1	93	0	35	0	0	100	0	40	2	0
100	8	68	3	1	94	21	68	3	3	68	9	36	2	1
100	13	80	1	1	100	27	97	2	3	84	0	27	2	0
89	20	56	2	3	92	4	31	2	0	97	0	45	1	2
100	34	55	3	2	100	12	68	2	1	100	14	51	3	0
100	30	33	2	2	74	0	41	1	0					
83	18	31	0	2	73	5	34	2	0					

図 6.1 は連続変数である u_1 から u_3 までの多変量散布図⁶⁹を示したものである。1 行 1 列に示された u_1 のヒストグラムでは、100 点満点の学生が多数いることが観察される。易しいテストは全問正解する学生が少なくなく、100 点となって、それ以上の点はない。したがって 110 点の実力の学生の能力は、そのテストでは正確には測れない。逆に 2 行 2 列に示された u_2 のヒストグラムでは、0 点の学生が多数いることが観察される。難しいテストは全問誤答する学生が少なくなく、0 点となって、それ以下の点はない。したがって -20 点の実力の学生の能力は正確には測れない。このような尺度の上限や下限に多くの度数が観察される変数を「トービット変数 (tobit variable)」とか「センソード変数 (censored variable)」といい、そのようなデータを「トービットデータ」とか「センソードデータ」という。

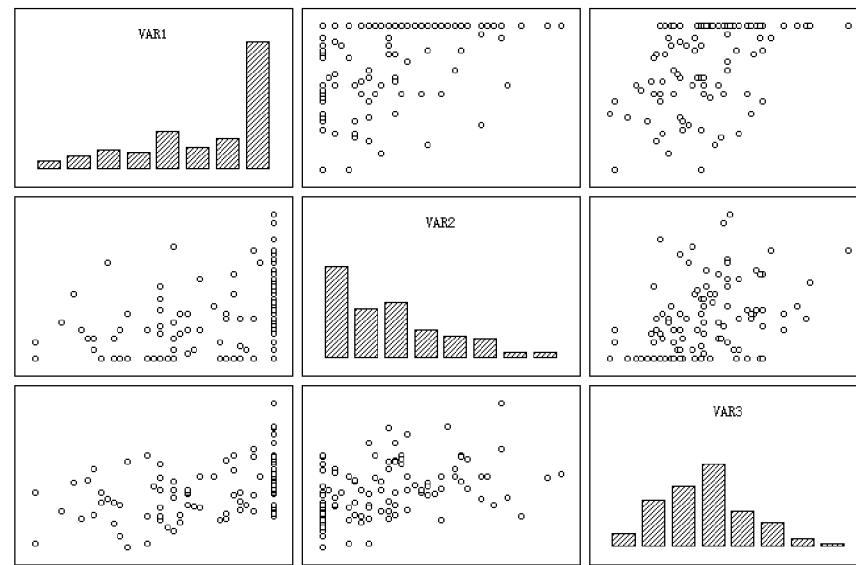


図 6.1: 変数 1 から変数 3 の多変量散布図

3 行 1 列や 3 行 2 列に示された散布図には、センソード変数と通常の変数間の（特定の方向から押し固められたような）典型的な形状が示されている。上

⁶⁹多変量散布図は、共分散行列における分散が配される位置にヒストグラムを、共分散が配される位置に通常の散布図を描いた行列形式の図である。

限や下限が存在せず、本来の測定がなされた場合と比較して相関係数も共分散も変化してしまうことが多い。前者、および後者におけるそのような変化を、それぞれ「天井効果」「床効果」という。2 行 1 列に示された散布図には天井効果と床効果が同時に観察されている。天井効果や床効果のある変数の標本平均や標本共分散行列は、そのまま分析するのではなく、それらの効果を統計モデルに組み込むことが有効である。

6.2 平均と分散の補正

尺度に上限や下限があり、上限値・下限値が比較的多数観測される変数 u のデータの発生機構をモデル化してみよう。まず、本来、測定したかった変数を x とする (x は観測変数として利用してきた文字であるが、ここでは必ずしも直接的には観測できないものとする)。つぎに変数 u は、変数 x の値に応じて

$$u = \begin{cases} \tau_b, & x \leq \tau_b \\ x, & \tau_b < x \leq \tau_a \\ \tau_a, & \tau_a < x \end{cases} \quad (6.1)$$

のように観測されるものとする。ここで τ_b は尺度の下限値であり、 τ_a は上限値である。 x の値が τ_b 以下の場合は、実際に観測される u の値は τ_b になる（たとえば 0 点以下の実力の人も 0 点が観測される）。逆に x の値が τ_a 以上の場合には、 u の値は τ_a になる（100 点以上の実力の人も 100 点が与えられる）。そして x の値が上限と下限の間である場合には、 x の値そのものが観測される。

表 6.1 の u_1 のように上限だけ考慮すればよい場合には $\tau_b = -\infty$ とする。以下の式展開では、 τ_b を含んだ項は無視する。逆に u_2 のように下限だけ考慮すればよい場合には $\tau_a = \infty$ とする。 τ_a を含んだ項は無視する。

以上のモデルに従って、本来、測定したかった変数 x の平均・分散に関する尤度関数を構成する。まず変数 u （の 1 個の実現値）に下限値が観測される確率は

$$p(u = \tau_b) = \int_{-\infty}^{\tau_b} f(x|\mu, \sigma^2) dx \quad (6.2)$$

である。ここで $f(x|\mu, \sigma^2)$ は正規分布の確率密度関数である。次に変数 $u = x$ が観測されたとすると、その確率密度は

$$p(u = x) = f(x|\mu, \sigma^2) \quad (6.3)$$

である. (6.2) 式と同様に, 上限値が観測される確率は

$$p(u = \tau_a) = \int_{\tau_a}^{+\infty} f(x|\mu, \sigma^2) dx \quad (6.4)$$

である.

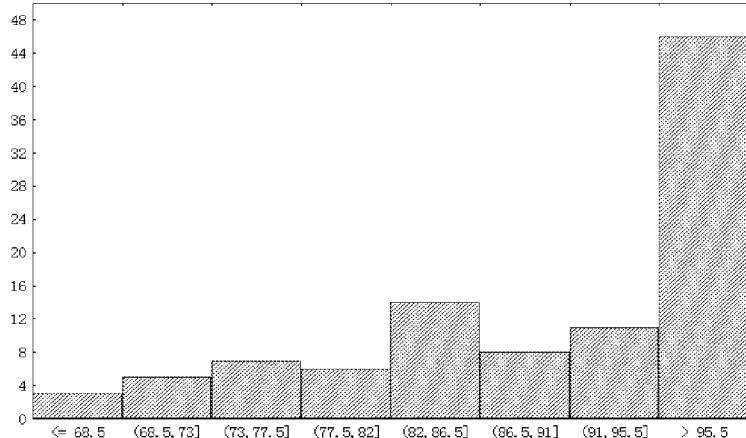


図 6.2: 変数 1 のヒストグラム

変数 u の N 個の実現値 \mathbf{u} が観測され, そのうち下限と上限が, それぞれ N_b 個, N_a 個含まれていた場合の μ, σ^2 の尤度関数は, (6.2) 式から (6.4) 式の表現を用いることにより

$$p(\mathbf{u}|\mu, \sigma^2) = p(u = \tau_b)^{N_b} \times \prod_{i=1}^{N-N_b-N_a} p(u = x_i) \times p(u = \tau_a)^{N_a} \quad (6.5)$$

のように構成される. したがって直接観測された u の平均と分散ではなく, 本来, 観測したかった x の平均と分散は, 対数尤度関数

$$\log p(\mathbf{u}|\mu, \sigma^2) = N_b \log p(u = \tau_b) + \sum_{i=1}^{N-N_b-N_a} \log p(u = x_i) + N_a \log p(u = \tau_a) \quad (6.6)$$

を最大化することによって, その最尤推定値を得る.

表 6.2 には, u_1, u_2 からそのまま計算した平均・分散・標準偏差と, (6.6) 式を最大化して推定した x_1, x_2 の平均・分散・標準偏差を示した. 変数 1 は上限

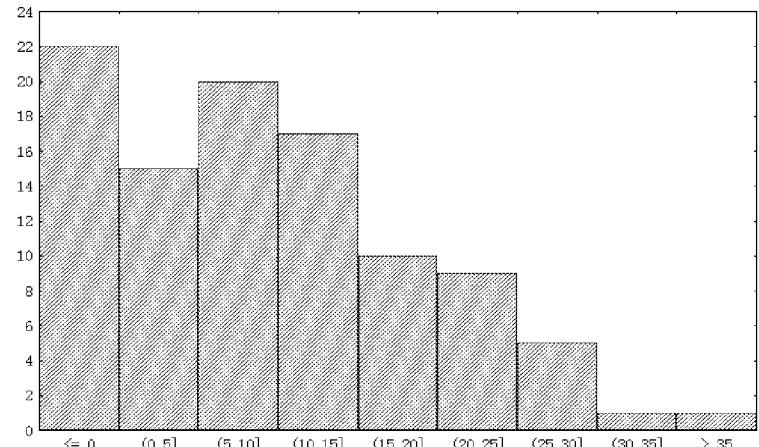


図 6.3: 変数 2 のヒストグラム

表 6.2: 平均・分散・標準偏差

変数名	平均	分散	標準偏差
u_1	90.61	106.14	10.30
u_2	10.12	81.12	9.01
x_1	94.91	242.37	15.56
x_2	8.75	120.78	10.99

の影響で, 本来, 平均が 94.91 であったところ 90.61 に下がっていたことが示されている. 標準偏差は, 本来, 15.56 であったところ 10.30 に狭まって観測されていたことが分かる. 同様に変数 2 は, 下限の影響で, 平均は 8.75 が 10.12 に上がっていたこと, 標準偏差は 10.99 が 9.01 に狭まって観測されていたことが分かる. 図 6.2 と図 6.3 に, それぞれ u_1, u_2 のヒストグラムを示したので, その違いを想像していただきたい.

6.3 平均と分散の補正

表 6.3 は, 表 6.1 の生データから直接計算した標本相関行列である. この相関行列は u_i の相関行列の推定値である. 一方, 上限・下限のある変数や, カテゴ

リカルな順序変数の背後に仮定した連続変数の相関行列の推定値を表 6.4 に示した。表 6.4 中の 5 行 4 列の 0.283 は、入門編第 12 章で紹介したポリコリック相関係数である。4 行 3 列と 5 行 3 列の 0.399 と 0.421 は、ポリシリアル相関係数である。

表 6.5 にはポリコリック相関係数とポリシリアル相関係数を推定する際に、同時に推定されたカテゴリカル変数の閾値である。おおざっぱに解釈すると B 先生の「優・良」は A 先生の「優」に相当し、B 先生の「可」は A 先生の「良」に相当しているようである。連続変数以外の変数を共分散構造分析に組み込む場合には、それが上限・下限のある変数であるときには補正された平均と分散（標準偏差）を、カテゴリカルな順序変数であるときには閾値を解釈することが重要であり、有用である。

表 6.4 には以下の 3 種類の新出の相関係数が記載されている。

- (1) トービット変数間の相関係数 (2 行 1 列の 0.429)
 - (2) トービット変数と正規（連続）変数との相関 (3 行 1,2 列の 0.509, 0.308)
 - (3) トービット変数とカテゴリカル変数との相関 (4 行 1,2 列, 5 行 1,2 列)
- 次節以降では、順にこれらの相関係数の導出を紹介する。

表 6.3: 祖点の相関行列

	テスト 1	テスト 2	テスト 3	評価 1	評価 2
テスト 1	1.000				
テスト 2	0.375	1.000			
テスト 3	0.440	0.365	1.000		
評価 1	0.250	0.426	0.371	1.000	
評価 2	0.339	0.507	0.385	0.242	1.000

表 6.4: 推定された相関行列

	テスト 1	テスト 2	テスト 3	評価 1	評価 2
テスト 1	1.000				
テスト 2	0.429	1.000			
テスト 3	0.509	0.380	1.000		
評価 1	0.322	0.479	0.399	1.000	
評価 2	0.385	0.530	0.421	0.283	1.000

表 6.5: u_4, u_5 の閾値

	τ_1	τ_2	τ_3
u_1	-0.842	-0.253	0.772
u_1	-0.332	0.583	1.555

6.4 トービット変数間の相関・共分散

2 つの観測変数 u_x と u_y とがあり、双方ともトービット変数であるとする。 u_x の上限と下限を、それぞれ τ_{ax}, τ_{bx} と表記し、 u_y の上限と下限を、それぞれ τ_{ay}, τ_{by} と表記する。

このとき u_x に下限が観測され、かつ u_y にも下限が観察される確率 p_{bb} は

$$p_{bb} = p(u_x = \tau_{bx}, u_y = \tau_{by}) = \int_{-\infty}^{\tau_{bx}} \int_{-\infty}^{\tau_{by}} f(x, y) dy dx \quad (6.7)$$

と表現される。ここで変数 x, y は、それぞれ u_x と u_y の背後に仮定した、上限・下限のない変数であり、 $f(x, y)$ は 2 変量正規分布の密度関数である。 $f(x, y)$ には、共分散と、2 つの変数の平均と分散の合計 5 つの母数が含まれており、正確には $f(x, y | \mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy})$ と表記される確率密度関数である。

同様の表記を利用して、 u_x に下限が観測され、かつ u_y には尺度内の値が観測される確率密度 p_{by} は

$$p_{by} = p(u_x = \tau_{bx}, u_y = y) = \int_{-\infty}^{\tau_{bx}} f(x, y) dx \quad (6.8)$$

であるし、 u_x に下限が観測され、かつ u_y には上限が観測される確率密度 p_{ba} は

$$p_{ba} = p(u_x = \tau_{bx}, u_y = \tau_{ay}) = \int_{-\infty}^{\tau_{bx}} \int_{\tau_{by}}^{+\infty} f(x, y) dy dx \quad (6.9)$$

である。

u_x に尺度内の値が観測され、かつ u_y に下限、尺度内の値、上限が観測され

る確率密度 p_{xb}, p_{xy}, p_{xa} は、それぞれ

$$p_{xb} = p(u_x = x, u_y = \tau_{by}) = \int_{-\infty}^{\tau_{by}} f(x, y) dy \quad (6.10)$$

$$p_{xy} = p(u_x = x, u_y = y) = f(x, y) \quad (6.11)$$

$$p_{xa} = p(u_x = x, u_y = \tau_{ay}) = \int_{\tau_{by}}^{+\infty} f(x, y) dy \quad (6.12)$$

と表現される。

同様に u_x に上限が観測され、かつ u_y に下限、尺度内の値、上限が観測される確率密度 p_{ab}, p_{ay}, p_{aa} は、それぞれ

$$p_{ab} = p(u_x = \tau_{ax}, u_y = \tau_{by}) = \int_{\tau_{bx}}^{+\infty} \int_{-\infty}^{\tau_{by}} f(x, y) dy dx \quad (6.13)$$

$$p_{ay} = p(u_x = \tau_{ax}, u_y = y) = \int_{\tau_{bx}}^{+\infty} f(x, y) dx \quad (6.14)$$

$$p_{aa} = p(u_x = \tau_{ax}, u_y = \tau_{ay}) = \int_{\tau_{bx}}^{+\infty} \int_{\tau_{by}}^{+\infty} f(x, y) dy dx \quad (6.15)$$

である。

(6.7) 式から (6.15) 式は確率と確率密度に 2 分される。確率は $p_{bb}, p_{ba}, p_{ab}, p_{aa}$ であり、実現値によって値は変わらない。一方、確率密度である $p_{by}, p_{xb}, p_{xy}, p_{xa}, p_{ay}$ は、実現値によって値は変わる。そこで添字 i を付して $p_{byi}, p_{xbi}, p_{xyi}, p_{xai}, p_{ayi}$ と表記する。

このとき、変数 u_x, u_y の実現値 $\mathbf{u}_x, \mathbf{u}_y$ が与えられたときの $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy}$ の尤度は

$$\begin{aligned} p(\mathbf{u}_x, \mathbf{u}_y | \mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy}) &= p_{bb}^{N_{bb}} \times \prod_{i=1}^{N_{by}} p_{byi} \times p_{ba}^{N_{ba}} \times \prod_{i=1}^{N_{xb}} p_{xbi} \\ &\times \prod_{i=1}^{N_{xy}} p_{xyi} \times \prod_{i=1}^{N_{xa}} p_{xai} \times p_{ab}^{N_{ab}} \times \prod_{i=1}^{N_{aa}} p_{ayi} \times p_{aa}^{N_{aa}} \end{aligned} \quad (6.16)$$

と表現される。ただし実現値は、 N に同じ添字を付けた数だけ観測されたものとする(たとえば p_{bb} は N_{bb} 個観察されてたものとする)。この式の対数をとることによって対数尤度関数が構成され、その値を最大化する平均・分散・共分散がそれぞれの最尤推定値となり、分散・共分散から相関係数を求める。

6.5 トービット変数と連続変数の相関

観測変数 u_x がトービット変数であり、 u_y が連続変数であるとき、 u_x に下限が観測され、かつ $u_y = y$ が観察される確率密度 p_{by} は

$$p_{by} = p(u_x = \tau_{bx}, u_y = y) = \int_{-\infty}^{\tau_{bx}} f(x, y) dx \quad (6.17)$$

である。同様に $u_y = y$ が観察され、 u_x が尺度内の値である場合と、 u_x が上限である場合の確率密度 p_{xy}, p_{ay} は、それぞれ

$$p_{xy} = p(u_x = x, u_y = y) = f(x, y) \quad (6.18)$$

$$p_{ay} = p(u_x = \tau_{ax}, u_y = y) = \int_{\tau_{ax}}^{+\infty} f(x, y) dx \quad (6.19)$$

である。確率密度は実現値によって値は変わるから、添字 i を付して $p_{byi}, p_{xyi}, p_{ayi}$ と表記すると、 $\mathbf{u}_x, \mathbf{u}_y$ が与えられたときの $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy}$ の尤度は

$$p(\mathbf{u}_x, \mathbf{u}_y | \mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy}) = \prod_{i=1}^{N_{by}} p_{byi} \times \prod_{i=1}^{N_{xy}} p_{xyi} \times \prod_{i=1}^{N_{ay}} p_{ayi} \quad (6.20)$$

と表現される。この式から対数尤度関数を構成し、母数の最尤推定値を求め、相関係数を計算する。

6.6 トービット変数とカテゴリカル変数の相関

本節では u_x がトービット変数であり、 u_y がカテゴリカルな順序変数である場合の相関係数について論じる。入門編第 12 章では順序のあるカテゴリカル変数の背後に平均 0, 分散 1 に標準化された変数 z 仮定した。トービット変数とは異なり、背後に仮定する変数は直接は観測されないので平均や分散を仮定しても意味が無いからである。ところが、入門編第 13,14 章で導入された期待値の構造化や多母集団のモデルを導入すると、カテゴリカル変数の背後に仮定する変数の平均や分散を推定することが意味を持つてくる(適用例は後述する)。ここでは入門編第 12 章とは異なり、 u_y が C 個のカテゴリカルな順序尺度による離散変数である場合は、

$$u_y = c \quad (\tau_c < y \leq \tau_{c+1}) \quad (6.21)$$

とする. ここで $c = 0, 1, 2, \dots, C - 1$ であり, $\tau_{i0} = -\infty$, $\tau_{iC} = +\infty$ である. y の平均と分散は μ_y, σ_y^2 とする. たとえば表 6.1 のように 4 段階尺度である場合には

$$u = \begin{cases} 0 & (-\infty < y \leq \tau_1) \\ 1 & (\tau_1 < y \leq \tau_2) \\ 2 & (\tau_2 < y \leq \tau_3) \\ 3 & (\tau_3 < y \leq +\infty) \end{cases} \quad (6.22)$$

となる. もちろん平均構造も多母集団も扱わない場合には $\mu_y = 0, \sigma_y^2 = 1$ と固定するのが一般的である. このとき, トービット変数とカテゴリカル変数の相関は以下のように計算する.

u_x に下限が観測され, かつ $u_y = c$ が観測される確率 p_{bc} は

$$p_{bc} = p(u_x = \tau_{bx}, u_y = c) = \int_{-\infty}^{\tau_{bx}} \int_{\tau_c}^{\tau_{c+1}} f(x, y) dy dx \quad (6.23)$$

である. ここで注意すべきことは, 添字 c は a や b とは異なり, 0 から C まで動く添字なので (6.23) 式は, $C + 1$ 個の確率を表現しているということである. 次に u_x に尺度内の値が観測され, かつ $u_y = c$ が観測される確率密度 p_{xc} は

$$p_{xc} = p(u_x = x, u_y = c) = \int_{\tau_c}^{\tau_{c+1}} f(x, y) dx \quad (6.24)$$

と表現される. 確率密度は実現値ごとに値が違うから (6.24) 式には添字 i を付けて p_{xci} と表記する. そして u_x に上限が観測され, かつ $u_y = c$ が観測される確率 p_{ac} は

$$p_{ac} = p(u_x = \tau_{ax}, u_y = c) = \int_{\tau_{ax}}^{+\infty} \int_{\tau_c}^{\tau_{c+1}} f(x, y) dy dx \quad (6.25)$$

と表現される.

以上のことから, $\mathbf{u}_x, \mathbf{u}_y$ が与えられたときの $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy}$ の尤度は

$$p(\mathbf{u}_x, \mathbf{u}_y | \mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy}) = \prod_{c=0}^C p_{bc}^{N_{bc}} \times \prod_{c=0}^C \prod_{i=1}^{N_{xc}} p_{xc} \times \prod_{c=0}^C p_{ac}^{N_{ac}} \quad (6.26)$$

と表現される. この式から対数尤度関数を構成し, 相関係数を求める. (6.16) 式, (6.20) 式, (6.26) 式の最適化には, 5 つの母数 $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy}$ が存在するが,

このうち平均と分散の推定値は 1 変数ごとに推定可能であった (トービット変数の場合には, (6.6) 式による推定値をもちい, 連続変数の場合には標本平均・標本分散を用いる). したがって (6.16) 式, (6.20) 式, (6.26) 式に関しては, 平均と分散は既に推定された値に固定し, 1 つの母数 (共分散) に関してだけ最適化することもできる.

6.7 トービット因子分析

トービット変数の因子分析モデルをトービット因子分析 (tobit factor analysis) という. 表 6.1 を用いて, 因子数 1 でトービット因子分析を行ない, 因子負荷と独自分散の推定値を表 6.6 に示した. 自由度は 5, $\chi^2 = 8.068$ であり, 限界水準は $p = 0.1523$ であり, 当てはまりは悪くない.

表 6.6: 因子分析結果

	テスト 1	テスト 2	テスト 3	評価 1	評価 2
因子負荷	0.652	0.783	0.649	0.585	0.736
独自分散	0.575	0.386	0.579	0.658	0.458

表 6.7 には, 1 因子モデルの制約のはいった相関行列の推定値 $\Sigma(\theta)$ を示した. トービット・カテゴリカル変数の相関行列は, 表 6.3 と表 6.4 の相関行列があるから, 合計 3 種類の相関行列がある.

表 6.7: 推定された相関行列

	テスト 1	テスト 2	テスト 3	評価 1	評価 2
テスト 1	1.000				
テスト 2	0.510	1.000			
テスト 3	0.423	0.508	1.000		
評価 1	0.381	0.458	0.380	1.000	
評価 2	0.480	0.577	0.478	0.431	1.000

A データ・プログラム

A.1 第2章 因子分析

```
/TITLE  
第2章 探索的因子分析、直交同時1因子解  
/SPECIFICATIONS  
VAR=9;CAS=345;MAT=COV;FO='(9F4.3)';  
/EQUATIONS  
V1==*F1+E1;  
V2==*F1+E2;  
V3==*F1+E3;  
V4==*F1+E4;  
V5==*F1+E5;  
V6==*F1+E6;  
V7==*F1+E7;  
V8==*F1+E8;  
V9==*F1+E9;  
/VARIANCES  
F1 = 1.00; E1 TO E9=*;  
/PRINT  
FIT=ALL;  
/MATRIX  
1245  
06511134  
066606981117  
0176007401781565  
02260119011304691291  
025301220182036002681574  
0348020102420192019601440727  
01750151012901150095001302830658  
-242-195-151-188-1380068-319-3191106  
/END
```

```
/TITLE  
第2章 探索的因子分析、直交同時2因子解  
/SPECIFICATIONS  
VAR=9;CAS=345;MAT=COV;FO='(9F4.3)';  
/EQUATIONS  
V1==*F1 +E1;  
V2==*F1**F2+E2;  
V3==*F1**F2+E3;  
V4==*F1**F2+E4;  
V5==*F1**F2+E5;  
V6==*F1**F2+E6;  
V7==*F1**F2+E7;  
V8==*F1**F2+E8;  
V9==*F1**F2+E9;  
/VARIANCES  
F1 TO F2= 1.00; E1 TO E9=*;  
/PRINT  
FIT=ALL;  
/MATRIX  
-----データは1因子解と同じ-----  
/END
```

```
/TITLE  
第2章 探索的因子分析、直交同時3因子解  
/SPECIFICATIONS
```

```
VAR=9;CAS=345;MAT=COV;FO='(9F4.3)';  
/EQUATIONS  
V1==*F1 +E1;  
V2==*F1**F2 +E2;  
V3==*F1**F2**F3+E3;  
V4==*F1**F2**F3+E4;  
V5==*F1**F2**F3+E5;  
V6==*F1**F2**F3+E6;  
V7==*F1**F2**F3+E7;  
V8==*F1**F2**F3+E8;  
V9==*F1**F2**F3+E9;  
/VARIANCES  
F1 TO F3= 1.00; E1 TO E9=*;  
/PRINT  
FIT=ALL;  
/MATRIX  
-----データは1因子解と同じ-----  
/END
```

```
/TITLE  
第2章 探索的因子分析、直交同時4因子解  
/SPECIFICATIONS  
VAR=9;CAS=345;MAT=COV;FO='(9F4.3)';  
/EQUATIONS  
V1==*F1 +E1;  
V2==*F1**F2 +E2;  
V3==*F1**F2**F3 +E3;  
V4==*F1**F2**F3**F4+E4;  
V5==*F1**F2**F3**F4+E5;  
V6==*F1**F2**F3**F4+E6;  
V7==*F1**F2**F3**F4+E7;  
V8==*F1**F2**F3**F4+E8;  
V9==*F1**F2**F3**F4+E9;  
/VARIANCES  
F1 TO F4= 1.00; E1 TO E9=*;  
/PRINT  
FIT=ALL;  
/MATRIX  
-----データは1因子解と同じ-----  
/END
```

```
/TITLE  
第2章 確認的因子分析  
/SPECIFICATIONS  
VAR=9;CAS=345;MAT=COV;FO='(9F4.3)';  
/EQUATIONS  
V1==*F1+E1; V2==*F1+E2; V3==*F1+E3;  
V4==*F2+E4; V5==*F2+E5; V6==*F2+E6;  
V7==*F3+E7; V8==*F3+E8; V9==*F3+E9;  
/VARIANCES  
F1 TO F3= 1.00; E1 TO E9=*;  
/COVARIANCES  
F1 TO F3=*;  
/PRINT  
FIT=ALL;  
/MATRIX  
-----データは1因子解と同じ-----  
/END
```

```
TITLE1, 第2章 探索的因子分析, 3因子解';  
DATA DATA1(TYPE=COV);  
INPUT _TYPE_ $ _NAME_ $ V1-V9;
```

```

CARDS;
COV V1 1.245 .
COV V2 0.651 1.134 .
COV V3 0.666 0.698 1.117 .
COV V4 0.176 0.074 0.178 1.565 .
COV V5 0.226 0.119 0.113 0.469 1.291 .
COV V6 0.253 0.122 0.182 0.360 0.268 1.574 .
COV V7 0.348 0.201 0.242 0.192 0.196 0.144 0.727 .
COV V8 0.175 0.151 0.129 0.115 0.095 0.013 0.283 0.658 .
COV V9 -.242 -.195 -.151 -.188 -.138 0.068 -.319 -.319 1.106
;RUN;
PROC FACTOR NFACT=3 ROTATE=PROMAX METHOD=ML;
RUN;

/TITLE
第2章 高次因子分析
/SPECIFICATIONS
VAR=11;CAS=345;MAT=COR;FO='(11F2.2)';
/EQUATIONS
V1 = F1+E1; V2 =*F1+E2; V3 =*F1+E3;
V4 =*F1+E4; V5 = F2+E5; V6 =*F2+E6;
V7 =*F2+E7; V8 =*F2+E8; V9 =*F3+E9;
V10=*F3+E10;V11=*F3+E11;
F1==*F4+D1; F2==*F4+D2; F3==*F4+D3;
/VARIANCES
F4=1.0; D1 TO D3=*; E1 TO E11=*;
/PRINT
FIT=ALL;
/STA
1.9 2.8 1.9 2.0 2.1 2.5
2.3 2.6 1.9 2.4 2.5
/MATRIX
1.
421.
47561.
4856531.
332324181.
13121408641.
3030402946451.
202633255755661.
25211836272438331.
2323242934293642481.
313546343328525144531.
/END

/TITLE
第2章 独自因子と誤差因子の分離
/SPECIFICATIONS
VAR=10;CAS=500;MAT=COV;ANA=MOM;FO='(10F4.3)';
/EQUATIONS
V1=*V999+F1+E1; V2=*V999+F1+E2;
V3=*V999+F2+E3; V4=*V999+F2+E4;
V5=*V999+F3+E5; V6=*V999+F3+E6;
V7=*V999+F4+E7; V8=*V999+F4+E8;
V9=*V999+F5+E9; V10=*V999+F5+E10;
F1=F6 +D1;
F2=F6 +D2;
F3=F6+*F7+D3;
F4= *F7+D4;
F5= *F7+D5;
/VARIANCES
F6 TO F7= 1.00; D1 TO D5=*; E1 TO E10=*;
/COVARIANCES
F6,F7=*;
/PRINT
FIT=ALL;
/MEANS
2.096 2.116 3.103 3.138 2.581 2.569 4.054
4.162 3.538 3.587
/MATRIX
910
739 945
356 318 983
316 299 778 999
561 556 404 3941262
557 549 403 39110611211
195 183 185 17202210202 977
120 129 168 16602080191 3461007
177 217 148 14103980380 14501101070
202 262 117 12604540419 140009304701075
/CONSTRAINTS
(E1,E1)=(E2,E2);
(E3,E3)=(E4,E4);
(E5,E5)=(E6,E6);
(E7,E7)=(E8,E8);
(E9,E9)=(E10,E10);
(V1,V999)=(V2,V999);
(V3,V999)=(V4,V999);
(V5,V999)=(V6,V999);
(V7,V999)=(V8,V999);
(V9,V999)=(V10,V999);
/END

OPTIONS LS=72;
TITLE1 '第2章イプサティップ因子分析, 6変数';
DATA DATA1(TYPE=COV);
INPUT _TYPE_ $ _NAME_ $ V1-V5;
CARDS;
COV V1 0.302 .
COV V2 -0.083 0.423 .
COV V3 -0.069 -0.097 0.483 .
COV V4 -0.075 -0.110 -0.146 0.554 .
COV V5 -0.028 -0.053 -0.077 -0.097 0.327
;RUN;
PROC CALIS COV ALL NOMOD EDF=399 GTOL=0.00001;
COSAN G(6,GEN)*A(7,GEN)*P(7,DIA);
MATRIX G
[1,] = 0.8333 -.1667 -.1667 -.1667 -.1667 ,
[2,] = -.1667 0.8333 -.1667 -.1667 -.1667 ,
[3,] = -.1667 -.1667 0.8333 -.1667 -.1667 ,
[4,] = -.1667 -.1667 -.1667 0.8333 -.1667 -.1667 ,
[5,] = -.1667 -.1667 -.1667 -.1667 0.8333 -.1667 ;
MATRIX A
[1,1]=A1(0.8) A2(0.7) A3(0.6) A4(0.5) A1 A2, [1,2]=6*1;
MATRIX P
[1,1]=1.0 D1-D4 D1 D2;
*VNAMES G X1-X5, A Y1-Y6, P F1 E1-E6;
BOUNDS 0.0< D1-D4 < 1.0;
RUN;

TITLE1 '第2章イプサティップ因子分析, 8変数';
DATA DATA1(TYPE=COV);
INPUT _TYPE_ $ _NAME_ $ V1-V7;

```

```

CARDS;
COV V1 0.514 .
COV V2 0.119 0.615 .
COV V3 0.069 0.051 0.651 .
COV V4 0.065 -0.007 -0.040 0.651 .
COV V5 -0.207 -0.210 -0.186 -0.179 0.536 .
COV V6 -0.156 -0.190 -0.190 -0.157 0.120 0.553 .
COV V7 -0.222 -0.186 -0.176 -0.164 0.109 -0.013 0.679
;RUN;
PROC CALIS COV ALL NOMOD EDF=399 MAXITER=1000;
COSAN G(8,GEN)*A(10,GEN)*P(10,SYM);
MATRIX G
[1,]= 0.875 -.125 -.125 -.125 -.125 -.125 -.125 -.125,
[2,]= -.125 0.875 -.125 -.125 -.125 -.125 -.125 -.125,
[3,]= -.125 -.125 0.875 -.125 -.125 -.125 -.125 -.125,
[4,]= -.125 -.125 -.125 0.875 -.125 -.125 -.125 -.125,
[5,]= -.125 -.125 -.125 -.125 0.875 -.125 -.125 -.125,
[6,]= -.125 -.125 -.125 -.125 -.125 0.875 -.125 -.125,
[7,]= -.125 -.125 -.125 -.125 -.125 -.125 0.875 -.125;
MATRIX A
[,1]=A1(0.8) A2(0.7) A3(0.6) A4(0.5),
[,2]=4*A. A1(0.8) A2(0.7) A3(0.6) A4(0.5),
[,3]=8*1;
MATRIX P
[1,1]=1.0 1.0 D1-D4 D1-D4,
[2,1]=P21(0.5);
BOUNDS 0.0< D1-D4 <1.0;
RUN;

```

TITLE1 '一部がイプサティブなデータの分析';
DATA DATA1(TYPE=COV); _TYPE_= 'COV';
INPUT _NAME_ \$ V1-V7;
CARDS;
V1 0.980
V2 0.343 1.024
V3 0.575 0.414 1.035
V4 0.270 0.146 0.326 1.062 . . .
V5 0.148 0.095 0.181 0.434 0.956 . .
V6 -.024 -.043 -.022 0.173 0.126 0.378 . .
V7 -.152 -.076 -.133 0.015 0.054 -.140 0.476
;RUN;
PROC CALIS COV ALL NOMOD EDF=399;
COSAN G(8,GEN)*A(10,GEN)*P(10,SYM);
MATRIX G
[1,]= 1,
[2,]= 0 1,
[3,]= 0 0 1,
[4,]= 0 0 0 1,
[5,]= 0 0 0 0 1,
[6,]= 0 0 0 0 0 0.6667 -.3333 -.3333,
[7,]= 0 0 0 0 0 -.3333 0.6667 -.3333;
MATRIX A
[,1]=A1(0.8) A2(0.7) A3(0.6) 0 0 0 0 A8(0.5),
[,2]=3*A A4(0.8) A5(0.7) A6(0.6) A7(0.5) 0,
[,3]=8*1;
MATRIX P
[1,1]=1.0 1.0 D1-D8,
[2,1]=P21(0.5);
BOUNDS 0.0< D1-D8 <1.0;
RUN;

A.2 第3章 実験データの解析

```

TITLE1 '第3章 1要因実験母数モデル';
DATA DATA1;
INFILE 'H:\Po3o1.DAT';
INPUT X1 V1-V4;
RUN;
PROC CALIS ALL NOMOD UCOV AUG NPOINT EDF=24;
VAR X1 V2-V4;
LINEQS
F1=INTERCEP-V2-V3-V4+D1,
X1=AO INTERCEP + A1 F1 + A2 V2+ A3 V3 + A4 V4 + E1;
STD
E1=Del1,D1=0.0;
A1=0.0-A2-A3-A4;
RUN;

10 1 0 0 0
10 1 0 0 0
09 1 0 0 0
11 1 0 0 0
12 1 0 0 0
11 1 0 0 0
08 0 1 0 0
10 0 1 0 0
08 0 1 0 0
10 0 1 0 0
12 0 1 0 0
09 0 1 0 0
08 0 0 1 0
08 0 0 1 0
11 0 0 1 0
11 0 0 1 0
14 0 0 1 0
15 0 0 1 0
14 0 0 0 1
12 0 0 0 1
11 0 0 0 1
16 0 0 0 1
13 0 0 0 1
12 0 0 0 1

TITLE1 '第3章 2要因実験母数モデル';
DATA DATA1;
INFILE 'H:\Po3o2.dat';
INPUT X1 T1 T2 R1 R2 I11 I12 I21 I22;
RUN; PROC PRINT; RUN;
PROC CALIS ALL NOMOD UCOV AUG NPOINT EDF=36;
VAR X1 T1 R1 I12 ;
LINEQS
FT2=INTERCEP-T1+D1,
FR2=INTERCEP-R1+D2,
FI11=T1-I12+D3,
FI21=R1-FI11+D4,
FI22=INTERCEP-FI11-I12-FI21+D5,
X1=AO INTERCEP + A1 T1 + A2 FT2+ B1 R1 + B2 FR2 + C11 FI11 +
C12 I12 + C21 FI21 + C22 FI22 + E1;
STD
E1=Del1,D1-D5=5*0.0;
A1=0.0-A2; B1=0.0-B2; C21=0.0-C22; C12=0.0-C22; C11=0.0-C12;

```

```

RUN;

079 1 0 1 0 1 0 0 0
107 1 0 1 0 1 0 0 0
103 1 0 1 0 1 0 0 0
092 1 0 1 0 1 0 0 0
180 1 0 1 0 1 0 0 0
165 1 0 1 0 1 0 0 0
240 1 0 1 0 1 0 0 0
265 1 0 1 0 1 0 0 0
300 1 0 1 0 1 0 0 0
075 1 0 0 1 0 1 0 0
060 1 0 0 1 0 1 0 0
060 1 0 0 1 0 1 0 0
094 1 0 0 1 0 1 0 0
119 1 0 0 1 0 1 0 0
100 1 0 0 1 0 1 0 0
102 1 0 0 1 0 1 0 0
125 1 0 0 1 0 1 0 0
165 1 0 0 1 0 1 0 0
095 0 1 1 0 0 0 1 0
099 0 1 1 0 0 0 1 0
070 0 1 1 0 0 0 1 0
116 0 1 1 0 0 0 1 0
170 0 1 1 0 0 0 1 0
145 0 1 1 0 0 0 1 0
205 0 1 1 0 0 0 1 0
200 0 1 1 0 0 0 1 0
210 0 1 1 0 0 0 1 0
153 0 1 0 1 0 0 0 1
078 0 1 0 1 0 0 0 1
075 0 1 0 1 0 0 0 1
092 0 1 0 1 0 0 0 1
115 0 1 0 1 0 0 0 1
155 0 1 0 1 0 0 0 1
250 0 1 0 1 0 0 0 1
340 0 1 0 1 0 0 0 1
380 0 1 0 1 0 0 0 1

TITLE1 '第3章 フィギュアスケート';
DATA DATA1;
INPUT X1-X9/X10-X18;
CARDS;
4.8 4.7 4.8 5.1 5.0 4.6 4.9 4.3 4.8
4.8 4.7 4.8 5.1 5.0 4.6 4.9 4.3 4.8
5.6 5.4 5.2 5.5 5.5 5.6 5.4 5.4 5.5
5.6 5.6 5.4 5.4 5.6 5.7 5.4 5.4 5.6
5.4 5.5 5.5 5.7 5.6 5.7 5.7 5.1 5.3
5.2 5.2 4.8 5.5 5.5 5.6 5.5 5.1 5.0
5.0 4.9 4.6 5.0 5.0 4.9 4.8 4.7 4.9
4.8 4.5 4.5 5.1 4.7 4.5 4.7 4.8 4.8
5.0 5.1 5.0 5.4 5.4 5.3 5.3 5.4 5.2
5.0 5.1 4.9 5.3 5.3 5.2 5.3 5.3 5.1
5.6 5.5 5.6 5.7 5.6 5.7 5.6 5.6 5.7
5.9 5.8 5.9 5.9 5.9 5.9 5.8 5.9 5.9
5.7 5.7 5.8 5.6 5.7 5.8 5.6 5.6 5.7
5.6 5.0 5.4 5.7 5.7 5.6 5.6 5.6 5.7
5.9 5.8 5.9 5.8 5.9 5.9 5.9 5.9 5.9
5.8 5.6 5.6 5.7 5.8 5.7 5.7 5.7 5.8
5.8 5.8 5.9 5.8 5.7 5.8 5.8 5.9 5.8
5.9 5.7 5.8 5.9 5.9 5.9 5.7 5.8 5.9

5.5 5.0 5.0 5.6 5.4 5.4 5.4 5.0 5.1
5.2 4.5 4.7 5.4 5.4 5.2 5.4 5.0 5.1
5.5 5.8 5.6 5.7 5.6 5.7 5.7 5.7 5.7
5.7 5.5 5.5 5.7 5.7 5.8 5.6 5.6 5.6
;RUN;
/*
TITLE1 '第3章 2要因混合モデル';
PROC CALIS ALL NOMOD UCOV AUG NOINT;
VAR X1-X9;
LINEQS
X1 =A1 INTERCEP + F + E1,
X2 =A2 INTERCEP + F + E2,
X3 =A3 INTERCEP + F + E3,
X4 =A4 INTERCEP + F + E4,
X5 =A5 INTERCEP + F + E5,
X6 =A6 INTERCEP + F + E6,
X7 =A7 INTERCEP + F + E7,
X8 =A8 INTERCEP + F + E8,
X9 =A9 INTERCEP + F + E9;
STD
E1-E9=9*Del,F=DF;
RUN;
*/
TITLE1 '第3章 3要因混合モデル';
PROC CALIS ALL NOMOD UCOV AUG NOINT M=LS;
LINEQS
X1 =A1 INTERCEP + F + F1 + F10 + E1,
X2 =A2 INTERCEP + F + F2 + F10 + E2,
X3 =A3 INTERCEP + F + F3 + F10 + E3,
X4 =A4 INTERCEP + F + F4 + F10 + E4,
X5 =A5 INTERCEP + F + F5 + F10 + E5,
X6 =A6 INTERCEP + F + F6 + F10 + E6,
X7 =A7 INTERCEP + F + F7 + F10 + E7,
X8 =A8 INTERCEP + F + F8 + F10 + E8,
X9 =A9 INTERCEP + F + F9 + F10 + E9,
X10=A10 INTERCEP+ F + F1 + F11 + E10,
X11=A11 INTERCEP+ F + F2 + F11 + E11,
X12=A12 INTERCEP+ F + F3 + F11 + E12,
X13=A13 INTERCEP+ F + F4 + F11 + E13,
X14=A14 INTERCEP+ F + F5 + F11 + E14,
X15=A15 INTERCEP+ F + F6 + F11 + E15,
X16=A16 INTERCEP+ F + F7 + F11 + E16,
X17=A17 INTERCEP+ F + F8 + F11 + E17,
X18=A18 INTERCEP+ F + F9 + F11 + E18,
F =MI INTERCEP+D,
F1 =MJ1 INTERCEP+D1,
F2 =MJ2 INTERCEP+D2,
F3 =MJ3 INTERCEP+D3,
F4 =MJ4 INTERCEP+D4,
F5 =MJ5 INTERCEP+D5,
F6 =MJ6 INTERCEP+D6,
F7 =MJ7 INTERCEP+D7,
F8 =MJ8 INTERCEP+D8,
F9 =MJ9 INTERCEP+D9,
F10=MK1 INTERCEP+D10,
F11=MK2 INTERCEP+D11;
STD
D=DI,D1-D9=9*DIJ,D10-D11=2*DIK,
E1-E18=18*DIJK;

MJ1=0.0-MJ2-MJ3-MJ4-MJ5-MJ6-MJ7-MJ8-MJ9;
MK1=0.0-MK2;

```

```

A1=0.0-A2-A3-A4-A5-A6-A7-A8-A9;
A10=0.0-A11-A12-A13-A14-A15-A16-A17-A18;
A1=0.0-A10;A2=0.0-A11;A3=0.0-A12;
A4=0.0-A13;A5=0.0-A14;A6=0.0-A15;
A7=0.0-A16;A8=0.0-A17;A9=0.0-A18;
RUN;

```

```

TITLE1 '第 3 章 ホテルの料金';
DATA DATA1;
INPUT X1 X2 V1 V2;
CARDS;
135 089 1 0
089 079 1 0
056 043 1 0
075 068 1 0
069 089 0 1
065 079 0 1
098 117 0 1
080 100 0 1
:RUN;
TITLE1 '第 3 章 分割実験、混合モデル';
PROC CALIS ALL NOMOD UCOV AUG NOINT EDF=8;
VAR X1 X2 V1;
LINEQS
FV2=INTERCEP - V1 + D1,
X1 =A1 INTERCEP + F + AC11 V1 + AC12 FV2 + E1,
X2 =A2 INTERCEP + F + AC12 V1 + AC12 FV2 + E2,
F =MU INTERCEP+ C1 V1 + C2 FV2 +D;
STD
D1=0.0, E1 E2= 2*DE, D=DA;
*COV E1 E2=DE21;
A1=0.0-A2; C1=0.0-C2;
AC11=0.0-AC12;
RUN;

```

```

/TITLE
第 3 章 共分散分析 第 1 グループ
/SPECIFICATIONS
GR=3;CAS=5;VAR=6;MA=RAW;ANA=MOM;DATA='P0305.dat';
/PRINT
FIT=ALL;
/EQUATIONS
V1=*V999**V2+E1;
/VARIANCES
V2=**; E1=**;
/COVARIANCES
V999,V2=**;
/END
/TITLE
第 3 章 共分散分析 第 2 グループ
/SPECIFICATIONS
CAS=5;VAR=6;MA=RAW;ANA=MOM;DATA='P0305.dat';
/EQUATIONS
V3=*V999**V4+E3;
/VARIANCES
V4=**; E3=**;
/COVARIANCES
V999,V4=**;
/END
/TITLE

```

```

第 3 章 共分散分析 第 3 グループ
/SPECIFICATIONS
CAS=5;VAR=6;MA=RAW;ANA=MOM;DATA='P0305.dat';
/EQUATIONS
V5=*V999**V6+E5;
/VARIANCES
V6=**; E5=**;
/COVARIANCES
V999,V6=**;
/CONSTRAINTS
(1,V1,V2)=(2,V3,V4)=(3,V5,V6);
(1,V2,V2)=(2,V4,V4)=(3,V6,V6);
(1,V1,V999)=(2,V3,V999)=(3,V5,V999);
(1,E1,E1)=(2,E3,E3)=(3,E5,E5);
/END

```

```

15 29 20 22 14 33
19 49 34 24 20 45
21 48 28 49 30 35
27 35 35 46 32 39
35 53 42 52 34 36
39 47 44 43 42 48
23 46 46 64 40 63
38 74 47 61 38 57
33 72 40 55 54 56
50 67 54 54 56 78

```

```

/TITLE
第 3 章 因子の分散分析 1980 年
/SPECIFICATIONS
GR=2;CAS=2925;VAR=4;MA=COR;ANA=MOM;
/PRINT
FIT=ALL;
/EQUATIONS
V1=*V999**F1+E1;
V2=*V999**F1+E2;
V3=*V999**F1+E3;
V4=*V999**F1+E4;
F1=0.0 V999+D1;
/VARIANCES
E1 TO E4 =**;D1=1.00;
/MATRIX
1.000
0.440 1.000
0.377 0.515 1.000
0.385 0.483 0.503 1.000
/MEANS
4.990 4.419 4.403 4.017
/STA
1.871 2.147 2.250 2.294
/END
/TITLE
第 3 章 因子の分散分析 1984 年
/SPECIFICATIONS
CAS=2967;VAR=4;MA=COR;ANA=MOM;
/EQUATIONS
V1=*V999**F1+E1;
V2=*V999**F1+E2;
V3=*V999**F1+E3;
V4=*V999**F1+E4;

```

```

F1=*V999+D1;
/VARIANCES
E1 TO E4 *=; D1 TO D2 *=;
/MATRIX
1.000
0.364 1.000
0.376 0.498 1.000
0.336 0.458 0.488 1.000
/MEANS
4.841 4.000 4.229 3.488
/STA
1.931 2.128 2.309 2.297
/CONSTRAINTS
(1,V1,V999)=(2,V1,V999);
(1,V2,V999)=(2,V2,V999);
(1,V3,V999)=(2,V3,V999);
(1,V4,V999)=(2,V4,V999);
(1,V1,F1)=(2,V1,F1);
(1,V2,F1)=(2,V2,F1);
(1,V3,F1)=(2,V3,F1);
(1,V4,F1)=(2,V4,F1);
(1,E1,E1)=(2,E1,E1);
(1,E2,E2)=(2,E2,E2);
(1,E3,E3)=(2,E3,E3);
(1,E4,E4)=(2,E4,E4);
/END

/TITLE
第 3 章 潜在共分散分析 統制群
/SPECIFICATIONS
GR=2;CAS=105;VAR=4;MA=COV;ANA=MOM;
/PRINT
FIT=ALL;
/EQUATIONS
V1=*V999+ F1+E1;
V2=*V999+*F1+E2;
V3=*V999+ F2+E3;
V4=*V999+*F2+E4;
F1= 0.0 V999 + D1;
F2= 0.0 V999 + *F1+D2;
/VARIANCES
E1 TO E4 *=; D1 TO D2 *=;
/MATRIX
37.626
24.933 34.680
26.639 24.236 32.013
23.649 27.760 23.565 33.443
/MEANS
18.381 20.229 20.400 21.343
/END
/TITLE
第 3 章 潜在共分散分析 実験群
/SPECIFICATIONS
CAS=108;VAR=4;MA=COV;ANA=MOM;
/EQUATIONS
V1=*V999+ F1+E1;
V2=*V999+*F1+E2;
V3=*V999+ F2+E3;
V4=*V999+*F2+E4;
F1=*V999+D1;
F2=*V999+*F1+D2;

/VARIANCES
E1 TO E4 *=; D1 TO D2 *=;
/MATRIX
50.084
42.373 49.872
40.760 36.094 51.237
37.343 40.396 39.890 53.641
/MEANS
20.556 21.241 25.667 25.870
/CONSTRAINTS
(1,V1,V999)=(2,V1,V999);
(1,V2,V999)=(2,V2,V999);
(1,V3,V999)=(2,V3,V999);
(1,V4,V999)=(2,V4,V999);
(1,V2,F1)=(2,V2,F1);
(1,V4,F2)=(2,V4,F2);
(1,F2,F1)=(2,F2,F1);
/END

/TITLE
第 3 章 多変量分散分析 1980 年
/SPECIFICATIONS
GR=2;CAS=2925;VAR=4;MA=COR;ANA=MOM;
/PRINT
FIT=ALL;
/EQUATIONS
V1=*V999+E1;
V2=*V999+E2;
V3=*V999+E3;
V4=*V999+E4;
/VARIANCES
E1 TO E4 *=;
/COVARIANCES
E1 TO E4 *=;
/MATRIX
1.000
0.440 1.000
0.377 0.515 1.000
0.385 0.483 0.503 1.000
/MEANS
4.990 4.419 4.403 4.017
/STA
1.871 2.147 2.250 2.294
/END
/TITLE
第 3 章 多変量分散分析 1984 年
/SPECIFICATIONS
CAS=2967;VAR=4;MA=COR;ANA=MOM;
/EQUATIONS
V1=*V999+E1;
V2=*V999+E2;
V3=*V999+E3;
V4=*V999+E4;
/VARIANCES
E1 TO E4 *=;
/COVARIANCES
E1 TO E4 *=;
/MATRIX
1.000
0.364 1.000
0.376 0.498 1.000

```

```

0.336 0.458 0.488 1.000
/MEANS
4.841 4.000 4.229 3.488
/STA
1.931 2.128 2.309 2.297
/CONSTRAINTS
(1,E1,E1)=(2,E1,E1);
(1,E1,E2)=(2,E1,E2);
(1,E1,E3)=(2,E1,E3);
(1,E1,E4)=(2,E1,E4);
(1,E2,E2)=(2,E2,E2);
(1,E2,E3)=(2,E2,E3);
(1,E2,E4)=(2,E2,E4);
(1,E3,E3)=(2,E3,E3);
(1,E3,E4)=(2,E3,E4);
(1,E4,E4)=(2,E4,E4);
/END
(1,V1,V999)=(2,V1,V999);
(1,V2,V999)=(2,V2,V999);
(1,V3,V999)=(2,V3,V999);
(1,V4,V999)=(2,V4,V999);

/TITLE
第3章 複合対称性の分析
/SPECIFICATIONS
CAS=9;VAR=3;MA=COV;ANA=COV;
/PRINT
FIT=ALL;
/EQUATIONS
V1=E1;
V2=E2;
V3=E3;
/VARIANCES
E1 TO E3 =*;
/COVARIANCES
E1 TO E3 =*;
/MATRIX
3.10
1.92 2.80
1.82 2.00 3.80
/CONSTRAINTS
(E1,E1)=(E2,E2)=(E3,E3);
(E1,E2)=(E1,E3)=(E2,E3);
/END

/TITLE
第3章 時間の変化の分析
/SPECIFICATIONS
CAS=120;VAR=6;MA=COR;ANA=COV;
/PRINT
FIT=ALL;
/EQUATIONS
V1=E1;
V2=E2;
V3=E3;
V4=E4;
V5=E5;
V6=E6;
/VARIANCES
E1 TO E6 =*;

```

```

/COVARIANCES
E1 TO E6 =*;
/MATRIX
1.00
.65 1.00
.54 .68 1.00
.27 .30 .21 1.00
.32 .21 .27 .59 1.00
.18 .26 .22 .48 .55 1.00
/STA
2.10 3.00 2.40 1.60 3.30 2.20
/CONSTRAINTS
(E1,E1)=(E4,E4);
(E2,E2)=(E5,E5);
(E3,E3)=(E6,E6);
(E1,E2)=(E4,E5);
(E1,E3)=(E4,E6);
(E2,E3)=(E5,E6);
/END

```

A.3 第4章 時系列解析

```

TITLE1, 景気動向指数(1988年1月より1994年1月まで);RUN;
DATA JAPANEC; INPUT X1-X3 Y @@;
LABEL X1='先行指数' X2='一致指数' X3='逕行指数';
CARDS;
76.9 90.9 75.0 73 57.7 100 50.0 72 53.8 90.9 62.5 71
50.0 60.0 64.3 70 41.7 10.0 50.0 69 33.3 55.0 42.9 68
38.5 72.7 75.0 67 38.5 81.8 87.5 66 53.8 81.8 75.0 65
53.8 72.7 87.5 64 61.5 90.9 81.3 63 69.2 90.9 50.0 62
65.4 81.8 62.5 61 38.5 54.5 81.3 60 50.0 81.8 81.3 59
96.2 54.5 50.0 58 65.4 63.6 75.0 57 46.2 45.5 62.5 56
38.5 81.8 87.5 55 69.2 90.9 75.0 54 30.8 36.4 87.5 53
53.8 72.7 87.5 52 38.5 68.2 93.8 51 38.5 63.6 87.5 50
30.8 54.5 100 49 38.5 63.6 87.5 48 76.9 54.5 75.0 47
53.8 81.8 87.5 46 76.9 90.9 75.0 45 76.9 90.9 68.8 44
61.5 81.8 43.8 43 61.5 72.7 62.5 42 30.8 63.6 75.0 41
38.5 59.1 75.0 40 23.1 40.9 62.5 39 38.5 54.5 87.5 38
46.2 36.4 87.5 37 38.5 63.6 87.5 36 19.2 40.9 50.0 35
23.1 40.9 37.5 34 30.8 50.0 68.8 33 7.7 18.2 62.5 32
7.7 31.8 75.0 31 15.4 9.1 62.5 30 38.5 27.3 62.5 29
15.4 0.0 62.5 28 30.8 27.3 75.0 27 23.1 0.0 75.0 26
30.8 9.1 37.5 25 38.5 18.2 62.5 24 50.0 18.2 25.0 23
23.1 9.1 31.3 22 23.1 0.0 6.3 21 30.8 13.6 31.3 20
23.1 27.3 12.5 19 34.6 27.3 18.8 18 38.5 50.0 18.8 17
15.4 18.2 18.8 16 15.4 9.1 18.8 15 34.6 0.0 12.5 14
61.5 18.2 25.0 13 84.6 63.6 25.0 12 80.8 72.7 37.5 11
53.8 59.1 12.5 10 23.1 9.1 0.0 9 15.4 9.1 0.0 8
38.5 0.0 25.0 7 38.5 54.5 25.0 6 38.5 59.1 56.3 5
38.5 18.2 12.5 4 38.5 18.2 12.5 3 38.5 22.7 12.5 2
60.0 85.0 42.9 1
; RUN;
*データのソート;PROC SORT; BY Y;RUN;

TITLE2, トーブリッツ(Toeplitz)行列の生成;RUN;
PROC IML;RESET NOLOG;START MAIN;
USE JAPANEC; READ ALL VAR{X1 X2 X3} INTO X3;
READ ALL VAR{X1 X2} INTO X2;

```

```

READ ALL VAR{X2} INTO X1;
C3=COVLAG(X3,3); CC3=TOEPLITZ(C3); C2=COVLAG(X2,3);
CC2=TOEPLITZ(C2); C1=COVLAG(X1,10); CC1=TOEPLITZ(C1);
CREATE CCC3 FROM CC3; APPEND FROM CC3; CLOSE CCC3;
CREATE CCC2 FROM CC2; APPEND FROM CC2; CLOSE CCC2;
CREATE CCC1 FROM CC1; APPEND FROM CC1; CLOSE CCC1;
FINISH;RUN MAIN;QUIT;

TITLE2 '一致のラグ 9 の共分散行列';RUN;
DATA C1(TYPE=COV);_TYPE_='COV';INPUT _NAME_ $ @@;SET CCC1;
LABEL COL1='一致指数 00' COL2='一致指数-1' COL3='一致指数-2'
COL4='一致指数-3' COL5='一致指数-4' COL6='一致指数-5'
COL7='一致指数-6' COL8='一致指数-7' COL9='一致指数-8'
COL10='一致指数-9';
CARDS;
COL1 COL2 COL3 COL4 COL5 COL6 COL7 COL8 COL9 COL10
;RUN;

TITLE2 '先行と一致のラグ 2 の共分散行列';RUN;
DATA C2(TYPE=COV);_TYPE_='COV';INPUT _NAME_ $ @@;
SET CCC2;
LABEL COL1='先行指数 00' COL2='一致指数 00'
COL3='先行指数-1' COL4='一致指数-1'
COL5='先行指数-2' COL6='一致指数-2';
CARDS;
COL1 COL2 COL3 COL4 COL5 COL6
;RUN;

TITLE2 '先行と一致と遅行指標のラグ 2 の共分散行列';RUN;
DATA C3(TYPE=COV);_TYPE_='COV';INPUT _NAME_ $ @@;SET CCC3;
LABEL COL1='先行指数 00' COL2='一致指数 00' COL3='遅行指数 00'
COL4='先行指数-1' COL5='一致指数-1' COL6='遅行指数-1'
COL7='先行指数-2' COL8='一致指数-2' COL9='遅行指数-2';
CARDS;
COL1 COL2 COL3 COL4 COL5 COL6 COL7 COL8 COL9
;RUN;

TITLE2 'AR (1) による自己回帰分析';RUN;
PROC CALIS DATA=C1 COV EDF=72 ALL NOMOD; VAR COL1-COL2;
LINEQS
COL1=G1_2 COL2 + E1;
STD E1 = DEL1;
BOUNDS -1<=G1_2<=1;
RUN;

TITLE2 'AR (2) による自己回帰分析';RUN;
PROC CALIS DATA=C1 COV EDF=72 ALL NOMOD;
VAR COL1-COL3;
LINEQS
COL1=G1_2 COL2 + G1_3 COL3 + E1;
STD E1 = DEL1;
BOUNDS -2<=G1_2<=2, -1<=G1_3<=1;
RUN;

TITLE2 'MA (1) による移動平均分析';RUN;
PROC CALIS DATA=C1 COV CORR EDF=72 ALL NOMOD;
VAR COL1-COL2;
LINEQS
COL1=1.00 FX1 + L1_2 FX2 + E1,
COL2=1.00 FX2 + L1_2 FX3 + E2;
STD E1 E2 FX1-FX3= 0.00 0.00 3*PHI1(672);
BOUNDS -1<=L1_2<=1;

RUN;
TITLE2 'MA (2) による移動平均分析';RUN;
PROC CALIS DATA=C1 COV EDF=72 ALL NOMOD;
VAR COL1-COL3;
LINEQS
COL1=1.00 FX1 + L1_2 FX2 + L1_3 FX3 + E1,
COL2=1.00 FX2 + L1_2 FX3 + L1_3 FX4 + E2,
COL3=1.00 FX3 + L1_2 FX4 + L1_3 FX5 + E3;
STD E1-E3 FX1-FX5 = 3*0.00 5*PHI1(672);
BOUNDS -2<=L1_2<=2, -1<=L1_3<=1;
RUN;

TITLE2 'ARMA (1, 1)';RUN;
PROC CALIS DATA=C1 COV EDF=72 ALL NOMOD;
VAR COL1-COL3;
LINEQS
COL1=G1_2(0.66) COL2+1.00 FX1+L1_2(0.66) FX2+E1,
COL2=G1_2 COL3 + 1.00 FX2 + L1_2 FX3 + E2;
STD E1-E2 FX1-FX3 = 2*0.00 3*PHI1(672);
BOUNDS -1<= L1_2 G1_2 <= 1;
RUN;

TITLE2 '多変量自己回帰 (VAR)';RUN;
PROC CALIS DATA=C2 COV EDF=72 ALL NOMOD;
LINEQS
COL1=B1_3 COL3+B1_4 COL4+B1_5 COL5+B1_6 COL6+E1,
COL2=B2_3 COL3+B2_4 COL4+B2_5 COL5+B2_6 COL6+E2;
STD E1 E2= DEL1 DEL2;
COV E1 E2= DEL21;
RUN;

TITLE2 '動的因子分析-因子数 1, ラグ 2';RUN;
TITLE3 '誤差項にラグ 2 の自己相関を仮定';RUN;
PROC CALIS DATA=C3 COV EDF=70 ALL NOMOD;
LINEQS
COL1=L1_1 FX1 + L1_2 FX2 + L1_3 FX3 +E1,
COL2=L2_1 FX1 + L2_2 FX2 + L2_3 FX3 +E2,
COL3=L3_1 FX1 + L3_2 FX2 + L3_3 FX3 +E3,
COL4=L1_1 FX2 + L1_2 FX3 + L1_3 FX4 +E4,
COL5=L2_1 FX2 + L2_2 FX3 + L2_3 FX4 +E5,
COL6=L3_1 FX2 + L3_2 FX3 + L3_3 FX4 +E6,
COL7=L1_1 FX3 + L1_2 FX4 + L1_3 FX5 +E7,
COL8=L2_1 FX3 + L2_2 FX4 + L2_3 FX5 +E8,
COL9=L3_1 FX3 + L3_2 FX4 + L3_3 FX5 +E9;
STD FX1-FX5=5*1.00,
E1-E9=DEL1 0.0 DEL3 DEL1 0.0 DEL3 DEL1 0.0 DEL3;
COV E4 E1=DEL41, E5 E2=0.000, E6 E3=DEL63,
E7 E4=DEL41, E8 E5=0.000, E9 E6=DEL63,
E7 E1=DEL71, E8 E2=0.000, E9 E3=DEL93;
RUN;

TITLE2 'Time Series Factor Analysis';RUN;
TITLE3 'MODEL1 ARMA(1,1)';RUN;
PROC CALIS DATA=C3 COV NOMOD EDF=70 M=LSML ALL;
LINEQS
COL1=1.00 FO          +E1,
COL2= L2 FO           +E2,
COL3= L3 FO           +E3,
COL4=          1.00 F1 +E4,
COL5=          L2 F1   +E5,
COL6=          L3 F1   +E6,

```

```

COL7=          1.00 F2  +E7,
COL8=          L2  F2  +E8,
COL9=          L3  F2  +E9,
FO= B1(0.74) F1 + 1.00 FU0 + G1(0.0) FU1 +D1,
F1= B1(0.74) F2 + 1.00 FU1 + G1(0.0) FU2 +D2;
STD E1-E9=DEL1-DEL3 DEL1-DEL3 DEL1-DEL3,
D1 D2= 2*0.00,F2 FU0 FU1 FU2=PHI(170) 3*PSI(77);
PSI=(PHI-B1*B1*PHI)/(1+G1*G1);RUN;

TITLE3 'MODEL2 ARMA(1,0)';RUN;
PROC CALIS DATA=C3 COV NOMOD EDF=70 ;
VAR COL1-COL6;
LINEQS
COL1=1.00 FO      +E1,
COL2= L2  FO      +E2,
COL3= L3  FO      +E3,
COL4=   1.00 F1    +E4,
COL5=   L2  F1    +E5,
COL6=   L3  F1    +E6,
FO = B1  F1      +DU1;
STD E1-E6=DEL1-DEL3 DEL1-DEL3,
DU1 = PSI,      F1 = PHI;
PSI=PHI-B1*B1*PHI;RUN;

TITLE3 'MODEL3 ARMA(1,0) 誤差自己相関あり';RUN;
PROC CALIS DATA=C3 COV NOMOD EDF=70 ;
VAR COL1-COL6;
LINEQS
COL1=1.00 FO      +E1,
COL2= L2  FO      +E2,
COL3= L3  FO      +E3,
COL4=   1.00 F1    +E4,
COL5=   L2  F1    +E5,
COL6=   L3  F1    +E6,
FO = B1  F1      +DU1;
STD E1-E6=DEL1-DEL3 DEL1-DEL3,
DU1 = PSI,      F1 = PHI;
COV E1 E4=DEL41, E6 E3=DEL63;
PSI=PHI-B1*B1*PHI;RUN;

```

A.4 第5章

```

! 多変量ACEモデル
! Australian Arithmetic Computation Data
G1: genetic structure
Data Calc NGroups=5
Matrices
A Lower 4 4 Free ! 遺伝
Compute A*A' /
End
G2: common environmental structure
Data Calc
Matrices
B Lower 4 4 Free ! 共有環境
Compute B*B' /
End
G3: specific environmental structure
Data Calc

```

```

Matrices
C Lower 4 4 Free ! 非共有環境
Compute C*C' /
End
G4: female MZ twin pairs
Data NInput_vars=8 NObservations=1232
Labels
asth1 hayf1 dust1 ecze1 asth2 hayf2 dust2 ecze2
PMatrix
1.000
.556 1.000
.573 .758 1.000
.273 .264 .309 1.000
.592 .366 .398 .232 1.000
.411 .593 .451 .145 .549 1.000
.434 .421 .518 .192 .640 .770 1.000
.087 .196 .193 .589 .145 .122 .218 1.000
!ACov File=ahdemzff.acv
Matrices
A Symm 4 4 = %E1 ! expected matrix of group 1
C Symm 4 4 = %E2 ! expected matrix of group 2
E Symm 4 4 = %E3 ! ..
Covariances A+C+E | A+C -
A+C | A+C+E /
End
G5: female DZ twin pairs
Data NInput_vars=8 NObservations=751
Labels
asth1 hayf1 dust1 ecze1 asth2 hayf2 dust2 ecze2
PMatrix
1.000
.524 1.000
.588 .749 1.000
.291 .314 .279 1.000
.262 .170 .041 .139 1.000
.129 .318 .262 .093 .395 1.000
.079 .171 .214 .019 .684 .723 1.000
.217 .114 .087 .313 .254 .218 .276 1.000
!ACov File=ahdedzff.acv
Matrices
A Symm 4 4 = %E1
C Symm 4 4 = %E2
E Symm 4 4 = %E3
H Full 1 1
Covariances A+C+E | H@A+C -
H@A+C | A+C+E /
Matrix H .5
Start .5 All
Options Multiple NDecimals=4
Options Iterations=300
Options Rsduals check
End

! 遺伝因子分析モデル ( Genetic Factor Model)
! Australian Arithmetic Computation Data
G1: genetic structure
Data Calc NGroups=5
Matrices
A Full 4 1 Free ! 遺伝の共通因子
B Diag 4 4 Free ! 遺伝の特殊因子
Compute A*A' + B*B'/

```

```

End
G2: common environmental structure
Data Calc
Matrices
C Full 4 1 Free ! 共有環境共通因子
D Diag 4 4 Free ! 共有環境共通因子
Compute C*C' + D*D'/
End
G3: specific environmental structure
Data Calc
Matrices
E Full 4 1 Free ! 非共有環境共通因子
F Diag 4 4 Free ! 非共有環境共通因子
Compute E*E' + F*F'/
End
G4: female MZ twin pairs
Data NInput_vars=8 NObservations=1232
Labels
asth1 hayf1 dust1 ecze1 asth2 hayf2 dust2 ecze2
PMatrix
1.000
.556 1.000
.573 .758 1.000
.273 .264 .309 1.000
.592 .366 .398 .232 1.000
.411 .593 .451 .145 .549 1.000
.434 .421 .518 .192 .640 .770 1.000
.087 .196 .193 .589 .145 .122 .218 1.000
!ACov File=ahdemzf.acv
Matrices
A Symm 4 4 = %E1 ! expected matrix of group 1
C Symm 4 4 = %E2 ! expected matrix of group 2
E Symm 4 4 = %E3 !
Covariances A+C+E | A+C -
A+C | A+C+E /
End
G5: female DZ twin pairs
Data NInput_vars=8 NObservations=751
Labels
asth1 hayf1 dust1 ecze1 asth2 hayf2 dust2 ecze2
PMatrix
1.000
.524 1.000
.588 .749 1.000
.291 .314 .279 1.000
.262 .170 .041 .139 1.000
.129 .318 .262 .093 .395 1.000
.079 .171 .214 .019 .684 .723 1.000
.217 .114 .087 .313 .254 .218 .276 1.000
!ACov File=ahdedzf.acv
Matrices
A Symm 4 4 = %E1
C Symm 4 4 = %E2
E Symm 4 4 = %E3
H Full 1 1
Covariances A+C+E | H@A+C -
H@A+C | A+C+E /
Matrix H .5
Start .5 All
Options Multiple NDecimals=4
Options Iterations=300
Options Rsdiduals check

```

```

End
/TITLE
一般児を含めた遺伝因子分析 一卵性
/SPECIFICATIONS
VARIABLES=12; GROUP=3; MA=COR;
/LABEL
V1 =AG1; V2=G1; V3=R1; V4=T1; V5=A1; V6 =S1;
V7 =AG2; V8=G2; V9=R2; V10=T2; V11=A2; V12=S2;
F1=向性 1; F2=向性 2; F3=向遺 1; F4=向遺 2;
F5=向共; F6=向非 1; F7=向非 2;
/EQUATIONS
V1 = F1+E1 ; V2 =*F1+E2 ; V3 =*F1+E3 ;
V4 =*F1+E4 ; V5 =*F1+E5 ; V6 =*F1+E6 ;
V7 = F2+E7 ; V8 =*F2+E8 ; V9 =*F2+E9 ;
V10=*F2+E10; V11=*F2+E11; V12=*F2+E12;
F1=*F3 + *F5 + *F6 + D1;
F2=*F4 + *F5 + *F7 + D2;
/VARIANCES
F3 TO F7 =1; D1 TO D2 =0; E1 TO E12 =*;
/COVARIANCES
F3 ,F4 =1.0;
/MATRIX
1.00
0.25 1.00
0.56 0.32 1.00
-.06 0.08 0.26 1.00
0.46 0.47 0.49 0.14 1.00
0.39 0.48 0.52 0.26 0.79 1.00
0.28 0.11 0.08 -.09 0.13 0.13 1.00
0.15 0.30 0.18 -.01 0.29 0.25 0.23 1.00
0.27 0.15 0.35 0.04 0.21 0.16 0.43 0.26 1.00
0.00 0.00 0.15 0.32 0.10 0.12 -.13 -.02 0.18 1.00
0.24 0.25 0.22 0.03 0.50 0.50 0.26 0.52 0.24 0.05 1.00
0.18 0.25 0.26 0.11 0.44 0.50 0.21 0.55 0.35 0.20 0.73 1.00
/END
/TITLE
二卵性
/SPECIFICATIONS
VARIABLES=12; MA=COR;
/LABEL
V1 =AG1; V2=G1; V3=R1; V4=T1; V5=A1; V6 =S1;
V7 =AG2; V8=G2; V9=R2; V10=T2; V11=A2; V12=S2;
F1=向性 1; F2=向性 2; F3=向遺 1; F4=向遺 2;
F5=向共; F6=向非 1; F7=向非 2;
/EQUATIONS
V1 = F1+E1 ; V2 =*F1+E2 ; V3 =*F1+E3 ;
V4 =*F1+E4 ; V5 =*F1+E5 ; V6 =*F1+E6 ;
V7 = F2+E7 ; V8 =*F2+E8 ; V9 =*F2+E9 ;
V10=*F2+E10; V11=*F2+E11; V12=*F2+E12;
F1=*F3 + *F5 + *F6 + D1;
F2=*F4 + *F5 + *F7 + D2;
/VARIANCES
F3 TO F7 =1; D1 TO D2 =0; E1 TO E12 =*;
/COVARIANCES
F3 ,F4 =0.5;
/MATRIX
1.00
0.59 1.00
0.61 0.40 1.00
-.32 -.09 0.13 1.00

```

```

0.50 0.66 0.41 -.08 1.00
0.54 0.64 0.36 0.09 0.67 1.00
0.02 0.01 -.05 -.34 0.10 0.04 1.00
-.04 0.19 -.08 -.10 0.27 0.28 0.53 1.00
0.12 0.17 -.07 -.32 0.11 0.11 0.45 0.37 1.00
-.03 0.17 0.17 0.25 0.06 0.24 -.27 -.04 0.05 1.00
0.12 0.33 -.03 -.34 0.37 0.27 0.55 0.71 0.44 -.08 1.00
0.04 0.30 0.10 -.17 0.34 0.30 0.47 0.71 0.52 0.15 0.75 1.00
/END
/TITLE
一般児
/SPECIFICATIONS
VARIABLES=6;MA=COR;
/LABEL
V1 =AG;V2 =G; V3 =R; V4=T; V5=A; V6=S;
F1=向性; F3=向違; F5=向共; F6=向非;
/EQUATIONS
V1 = F1+E1 ; V2 =*F1+E2 ; V3 =*F1+E3 ;
V4 =*F1+E4 ; V5 =*F1+E5 ; V6 =*F1+E6 ;
F1= *F3 + *F5 + *F6 + D1;
/VARIANCES
F3=1.0; F5=1.0; F6=1.0; D1 =0; E1 TO E6= *;
/COVARIANCES
/MATRIX
1.00
0.28 1.00
0.26 0.24 1.00
-.07 0.21 0.44 1.00
0.37 0.59 0.19 0.11 1.00
0.35 0.51 0.35 0.32 0.83 1.00
/CONSTRAINTS
!因子負荷に関する制約 (向性)
(1,V2,F1)=(1,V8 ,F2)=(2,V2,F1)=(2,V8 ,F2)=(3,V2,F1); !G
(1,V3,F1)=(1,V9 ,F2)=(2,V3,F1)=(2,V9 ,F2)=(3,V3,F1); !R
(1,V4,F1)=(1,V10 ,F2)=(2,V4,F1)=(2,V10 ,F2)=(3,V4,F1); !T
(1,V5,F1)=(1,V11 ,F2)=(2,V5,F1)=(2,V11 ,F2)=(3,V5,F1); !A
(1,V6,F1)=(1,V12 ,F2)=(2,V6,F1)=(2,V12 ,F2)=(3,V6,F1); !S
!誤差分散に関する制約 (向性)
(1,E1,E1)=(1,E7 ,E7 )=(2,E1,E1)=(2,E7 ,E7 )=(3,E1,E1); !Ag_E
(1,E2,E2)=(1,E8 ,E8 )=(2,E2,E2)=(2,E8 ,E8 )=(3,E2,E2); !G_E
(1,E3,E3)=(1,E9 ,E9 )=(2,E3,E3)=(2,E9 ,E9 )=(3,E3,E3); !R_E
(1,E4,E4)=(1,E10 ,E10)=(2,E4,E4)=(2,E10 ,E10)=(3,E4,E4); !T_E
(1,E5,E5)=(1,E11 ,E11)=(2,E5,E5)=(2,E11 ,E11)=(3,E5,E5); !A_E
(1,E6,E6)=(1,E12 ,E12)=(2,E6,E6)=(2,E12 ,E12)=(3,E6,E6); !S_E
!遺伝率に関する制約 (向性)
(1,F1,F3)=(1,F2,F4)=(2,F1,F3)=(2,F2,F4)=(3,F1,F3); !a 向
(1,F1,F5)=(1,F2,F5)=(2,F1,F5)=(2,F2,F5)=(3,F1,F5); !c 向
(1,F1,F6)=(1,F2,F7)=(2,F1,F6)=(2,F2,F7)=(3,F1,F6); !e 向
/PRINT
FIT=ALL;DIG=2;
/TEC
ITR=500;
/END

```

A.5 第6章

```

TI トーピット因子分析
DA IY=5 NO=100 NG=1 TR=OT VT=OT
TT

```

```

0 0 0 0 0
TE
-2 -1 0 3 3
CE
100.0 0.0
LA
' テスト 1 , テスト 2 , テスト 3 , 評価 1 , 評価 2 ,
MO MO=EF LE=1 UE=1
OU WF SV ST SS ES ED RS VE ET SE TV EC EK TO WE
RA FO UN='H:\p0601.dat'
(*)
```

A.6 第7章

```

data test; infile 'j:\classic.dat';input (x1-x5)(2.0);run;

title ' (弱) 同族テスト';
proc calis ucov aug noint all nomod m=ml;
lineqs x1 = mu1 intercept + a1 f + e1,
        x2 = mu2 intercept + a2 f + e2, x3 = mu3 intercept + a3 f + e3,
        x4 = mu4 intercept + a4 f + e4, x5 = mu5 intercept + a5 f + e5;
std f = 1.00, e1-e5 = del1 - del5;
run;

title ' (弱) 同族テスト (平均値等しい)';
proc calis ucov aug noint all nomod m=ml;
lineqs x1 = mu intercept + a1 f + e1,
        x2 = mu intercept + a2 f + e2, x3 = mu intercept + a3 f + e3,
        x4 = mu intercept + a4 f + e4, x5 = mu intercept + a5 f + e5;
std f = 1.00, e1-e5 = del1 - del5;
run;

title ' 信頼性の等しいテスト (強同族テスト)';
proc calis ucov aug noint all nomod m=ml;
lineqs x1 = m1 intercept + a1 f + e1,
        x2 = m2 intercept + a2 f + e2, x3 = m3 intercept + a3 f + e3,
        x4 = m4 intercept + a4 f + e4, x5 = m5 intercept + a5 f + e5;
std f = 1.00, e1-e5 = del1 - del5;
parms ro = 0.8;
del1=((1-ro)/ro)*a1**2; del2=((1-ro)/ro)*a2**2; del3=((1-ro)/ro)*a3**2;
del4=((1-ro)/ro)*a4**2; del5=((1-ro)/ro)*a5**2;
run;

title ' 信頼性の等しいテスト (強同族テスト, 平均値等しい)';
proc calis ucov aug noint all nomod m=ml;
lineqs x1 = mu intercept + a1 f + e1,
        x2 = mu intercept + a2 f + e2, x3 = mu intercept + a3 f + e3,
        x4 = mu intercept + a4 f + e4, x5 = mu intercept + a5 f + e5;
std f = 1.00, e1-e5 = del1 - del5;
parms ro = 0.8;
del1=((1-ro)/ro)*a1**2; del2=((1-ro)/ro)*a2**2; del3=((1-ro)/ro)*a3**2;
del4=((1-ro)/ro)*a4**2; del5=((1-ro)/ro)*a5**2;
run;

title ' 弱平行 (本質的にタウ等価な) テスト';
proc calis ucov aug noint all nomod m=ml;
lineqs x1 = mu1 intercept + a f + e1,
        x2 = mu2 intercept + a f + e2, x3 = mu3 intercept + a f + e3,
```

```

x4 = mu4 intercept + a f + e4, x5 = mu5 intercept + a f + e5;
std      f = 1.00, e1-e5 = del1 - del5;
run;

title '弱平行 (タウ等価な) テスト (平均値等しい)';
proc calis ucov aug noint all nomod m=ml;
lineqs   x1 = mu1 intercept + a f + e1,
          x2 = mu2 intercept + a f + e2, x3 = mu3 intercept + a f + e3,
          x4 = mu4 intercept + a f + e4, x5 = mu5 intercept + a f + e5;
std      f = 1.00, e1-e5 = del1 - del5;
run;

title '(強) 平行テスト';
proc calis ucov aug noint all nomod m=ml;
lineqs   x1 = mu1 intercept + a f + e1,
          x2 = mu2 intercept + a f + e2, x3 = mu3 intercept + a f + e3,
          x4 = mu4 intercept + a f + e4, x5 = mu5 intercept + a f + e5;
std      f = 1.00, e1-e5 = del1 - del5;
run;

title '(強) 平行テスト (平均値等しい)';
proc calis ucov aug noint all nomod m=ml;
lineqs   x1 = mu1 intercept + a f + e1,
          x2 = mu2 intercept + a f + e2, x3 = mu3 intercept + a f + e3,
          x4 = mu4 intercept + a f + e4, x5 = mu5 intercept + a f + e5;
std      f = 1.00, e1-e5 = del1 - del5;
run;

```

A.7 第8章

A.8 第9章

A.9 第10章

A.10 第11章