共分散構造分析による行動遺伝学モデルの新展開

立教大学 豊田秀樹¹

A new wave of behavior genetic modeling using covariance structure analysis

Hideki Toyoda (*Department of Sociology, Rikkyo University, Nishi-Ikebukuro, Toshima-ku, Tokyo 171*)

A number of useful methods for analyzing covariance structure have been proposed in the studies of human behavior genetics, reflecting the fact that the behavior genetic studies are one of the main origins of covariance structure model. In this paper, I review recent progress on methodology for behavior genetic studies of twins and families from the standpoint of the structural equation modeling. Especially, genetic ACE (additive genetic, common environment and random environment) model, multivariate ACE model, genetic factor analysis model and twin-parent model are focused upon. This review also discusses how to construct applied structural equation models which are useful for psychological research.

Key words: human behavior genetics, covariance structure analysis, ACE (additive genetic, common environment and random environment) model, genetic factor analysis model, structural equation modeling.

共分散構造モデル (covariance structure model) は、生物学分野におけるパス解析 (Wright, 1921) と心理学分野における因子分析 (Spearman, 1904) を理論的に統合し、Bock & Bargmann (1966) を端緒とする統計モデルである。そして Jöreskog (1970) によってモデルの応用可能性が認識され、コンピュータプログラム LISREL (LInear Structural RELations, Jöreskog & Sörbom, 1976) の改良にともなって社会・人文・行動科学の多くの分野で利用されるようになった。80 年代に入ると Bentler & Weeks (1980) が EQS を、McDonald (1978) が COSAN を、McArdle & McDonald (1984) が RAM というモデルを提案し、共分散の構造は洗練されていった。近年では構造方程式モデリング (SEM: structural equation modeling) と呼ばれることのほうが多くなってきている。発展の経緯に関しては、Bentler (1986)、Bollen (1989)、狩野 (1992)、豊田 (1992) 等が詳しい。

共分散構造モデルは、一般的な数理表現を採用している統計モデルであるために、適用される研究領域の個別の事情と分析者の学問的仮説に基づいて、データ固有の分析モデルを創造することができるという利点がある。これは、分析者自身がほとんどモデル構成を

行う余地のない他の多くの多変量解析モデルと根本的に異なる優れた性質といえる。

しかしこの利点は容易には生かし得ない。共分散構造モデルはむしろ使用するのが難しいと感じられることすらある。学問的仮説をモデル構成というフォーマットで十分に表現できない場合、あるいは分析対象の知識が集積していない研究の初期段階でモデルが構成できない場合は、共分散構造分析を実行することが難しい。データの形式が整っていれば取りあえず分析を実行することだけは可能な他の多くの多変量解析モデルと比較して、共分散構造分析の実用的使用が難しいといわれる一つの理由である。

また、たとえ実行はできても共分散構造モデルの表現力を十分に生かしたモデルが構成されているとは限らない。構成概念の間に定型の識別条件の範囲内でパスを引くという、モデル構成の汎用的なストラテジーしか利用されないことも少なくない。

優れたモデルを構成するためには、各研究領域の実質科学的な知見を利用した個々独自の工夫が必要になり、多くの研究分野に共通した一般的なノウハウは少ない。特定の研究領域で、データのもつ重要な情報を抽出するモデルを構成することは、それ自体が創造的な仕事である。

¹ 審査の過程で、審査委員の方々より多くの有益なコメントをいただいたことを、記してここに感謝の意を表します。

目 的

本論文では、研究領域固有の優れたモデル構成の一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、改良されてきた共分散構造の応用モデルを展望する (Neale & Cardon, 1992; 安藤, 1996)。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異な歴史をも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, 1960, 1990 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, 1965)。Jöreskog によって LISREL の考え方が導入された 60 年代後半から 70 年代前半にかけて、最尤法の考え方も取り入れられて、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, 1968)。但しこの時点では、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値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 のほうが数値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、数値的なモデルの表現力という観点からは、それ等は皆同じモデルであることがわかり (豊田, 1990)、共分散構造モデルは当初の予想より相当に一般的な数値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モデルは、共分散構造モデルの下位モデルの一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、数値的・解析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、一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さまざまな工夫が凝らされている。

現在、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一つの良き到達目標を与えることになる。また本論文の終節では、心理学の研究分野でモデル構成を行う際に利用できる一般的なノウハウを行動遺伝学モデルの中から展望する。

遺伝 ACE モデル

本節では、もっとも頻繁に用いられ、行動遺伝学的考え方を端的にあらわしている遺伝 ACE モデルを特定し、当分野に対する導入を図る。

遺伝 ACE モデルでは、知能・学力・性格・身体的特徴・医学的特徴等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。三つ子以上の場合はその内の 2 名がランダムに測定対象に選ばれることが多い。また双生児 1 と 2 は無作為に割り当てられる。一卵性双生児は男性のペアと女性のペアだけであるが、二卵性双生児には異性のペアがいるので、表現型の数は $10 (=5 \text{ 群} \times 2 \text{ 人})$ 種類となる。

但しここでは、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。遺伝 ACE モデルに限らず、ペアを 2 種類に限定しておいたほうが、続く節で紹介されるモデルの本質を簡潔に記述することができる。このため本節に限らず、以後もこのような状況でモデルの性質を論じる。

表現型の数は $4 (=2 \text{ ペア} \times 2 \text{ 人})$ 種類となり、平均を 0 に基準化した観測変数を $p_{MZ1}, p_{MZ2}, p_{DZ1}, p_{DZ2}$ とする。ここで観測変数を加算的遺伝 (A: additive genetic)・共有環境 (C: common environment)・非共有環境 (E: random environment) という三つの要因 (潜在変数) の重み付き和と考える。重み係数をそれぞれ a, c, e とすると、観測変数は

$$p_{MZ1} = aA_{MZ} + cC + eE_{MZ1} \quad (1)$$

$$p_{MZ2} = aA_{MZ} + cC + eE_{MZ2} \quad (2)$$

$$p_{DZ1} = aA_{DZ1} + cC + eE_{DZ1} \quad (3)$$

$$p_{DZ2} = aA_{DZ2} + cC + eE_{DZ2} \quad (4)$$

のように表現される。

遺伝要因としては、加算的遺伝要因と非加算的遺伝要因 (D: non-additive dominance genetic effect) が、状況に応じて導入されるが、本論文では加算的遺伝のみを議論の対象とし、これを遺伝要因と呼ぶ。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 (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った素質 (A_{DZ1}, A_{DZ2}) を受け継ぐ。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、ポリジーンの方法を導入し、両者の相関には 0.5 (芝, 1975, p. 12, (1.7) 式において, $2n_c = n$ を代入した状態) が仮定される。

共有環境要因 (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 C という単一の潜在変数が用いられる。

非共有環境要因 (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。このため個人ごとに異なった変数が用意され、互いに無相関が仮定される。

ここで各要因 (潜在変数) は互いに無相関で分散が 1 であると仮定すると、一卵性双生児と二卵性双生児の観測変数の共分散行列は、それぞれ

$$\Sigma_{MZ} = \begin{bmatrix} \sigma_{MZ1}^2 & \\ \sigma_{MZ21} & \sigma_{MZ2}^2 \end{bmatrix} \quad (5)$$

$$= \begin{bmatrix} a^2 + c^2 + e^2 & \\ 1.0 \times a^2 + c^2 & a^2 + c^2 + e^2 \end{bmatrix} \quad (6)$$

$$\Sigma_{DZ} = \begin{bmatrix} \sigma_{DZ1}^2 & \\ \sigma_{DZ21} & \sigma_{DZ2}^2 \end{bmatrix} \quad (7)$$

$$= \begin{bmatrix} a^2 + c^2 + e^2 & \\ 0.5 \times a^2 + c^2 & a^2 + c^2 + e^2 \end{bmatrix} \quad (8)$$

のように三つの母数で構造化される。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、遺伝 ACE 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すべての間隔尺度に適用可能な汎用モデルである。

遺伝 ACE モデルは、多群の母集団を同時に扱う共

分散構造分析の下位モデルの一つとして位置付けられ、Figure 1 のようにパス図を用いて表現される。ここで X は一卵性双生児の場合には値が 1.0、二卵性双生児の場合には値が 0.5 の固定母数である。

たとえば、Heath, Neale, Hewitt, Eaves, & Fulker (1989) には、双子の女子の単位期間当りのアルコール飲量の標本共分散行列が示されている。一卵性双生児 (171 組) と二卵性双生児 (101 組) でそれぞれ、

$$S_{MZ} = \begin{bmatrix} 1.280 & \\ 0.766 & 1.194 \end{bmatrix} \quad (9)$$

$$S_{DZ} = \begin{bmatrix} 1.077 & \\ 0.463 & 0.962 \end{bmatrix} \quad (10)$$

であり、標準化された母数の最尤推定値は

$$\hat{a} = 0.462, \quad \hat{c} = 0.622, \quad \hat{e} = 0.633 \quad (11)$$

となる。それぞれ 2 乗すれば、各要因の説明率が得られる。その結果、遺伝要因、共有環境要因、非共有環境要因は、それぞれ 21%、39%、40% 観測変数を説明していることになる。

日常生活で、我々が感じている親子の類似性は、大まかにいうならば遺伝要因と共有環境要因の和である。つまり“A さんが飲んだくれなのは家系だ”という否定的かつ固定的印象の説明率は 60% (= 21 + 39) ということになる。しかし行動遺伝学的な視点を導入すると、“変化させようがないのは、わずかに 21% であり、酒飲みの家族と別に暮らさせたり、禁アルコール教育をする余地が 79% もある”という可塑性を示唆する教育的知見が得られる。現在、我が国の発達心理学の多くの研究は、遺伝をタブー視するあまり、親子間の相関の原因を暗黙のうちに共有環境に求めている論文が少なくないが、行動遺伝学的な視点を導入しなければ考察できない側面があることを認識しなくてはならない。

多変量 ACE モデル

遺伝 ACE モデルは、統計モデルとしては、二つの変数を二つの群で測定しているので 4 変数のモデルであるが、実質科学的観点からは一つの特性の性質を調べるモデルであった。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。

考察を進める b 個の特性

$$p = [p_1 \ p_2 \ \cdots \ p_b] \quad (12)$$

があり、双生児 1 と双生児 2 で群内で 2 回、一卵性双生児と二卵性双生児の 2 つの群で、計 4 回測定した変数を、群ごとに

$$p_{MZ} = [p'_{MZ1} \ p'_{MZ2}] \quad (13)$$

$$p_{DZ} = [p'_{DZ1} \ p'_{DZ2}] \quad (14)$$

のように並べる。

各変数が個別の遺伝 ACE モデルによって生成され、かつ要因間の潜在変数は無相関 (要因内の潜在変

Figure 1. 遺伝 ACE モデル。

数間の相関は無相関とは限らない)と仮定すると、観測変数の共分散行列は

$$\Sigma_{MZ} = \begin{bmatrix} \Sigma_{MZ1} & \\ \Sigma_{MZ21} & \Sigma_{MZ2} \end{bmatrix} \quad (15)$$

$$= \begin{bmatrix} \Sigma_a + \Sigma_c + \Sigma_e & \\ 1.0 \times \Sigma_a + \Sigma_c & \Sigma_a + \Sigma_c + \Sigma_e \end{bmatrix} \quad (16)$$

$$\Sigma_{DZ} = \begin{bmatrix} \Sigma_{DZ1} & \\ \Sigma_{DZ21} & \Sigma_{DZ2} \end{bmatrix} \quad (17)$$

$$= \begin{bmatrix} \Sigma_a + \Sigma_c + \Sigma_e & \\ 0.5 \times \Sigma_a + \Sigma_c & \Sigma_a + \Sigma_c + \Sigma_e \end{bmatrix} \quad (18)$$

のように、三つの要因に対応する共分散行列で構造化される。

ここで 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三つの要因の共分散行列を推定するためには、まず

$$\Sigma_a = C_a C_a' \quad (19)$$

$$\Sigma_c = C_c C_c' \quad (20)$$

$$\Sigma_e = C_e C_e' \quad (21)$$

のように各行列を正三角行列とその転置行列の積 (コレスキー分解) で表現する。共分散構造モデルのソフトウェアで推定する母数は C_a , C_c , C_e であり、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。

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。そこで式 (16), (18) の共分散構造中の要因ごとの共分散行列に

$$\Sigma_a = \Lambda_a \Lambda_a' + \Psi_a \Psi_a' \quad (22)$$

$$\Sigma_c = \Lambda_c \Lambda_c' + \Psi_c \Psi_c' \quad (23)$$

$$\Sigma_e = \Lambda_e \Lambda_e' + \Psi_e \Psi_e' \quad (24)$$

のような因子分析モデルを導入し、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の部分であり、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一つの観測変数に三つの独自成分が影響を与えるので、どれかが 0 以下になってしまう場合が少なくない。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。係数を母数に推定することによって解空間を制限することな

く、ヘイウッドケースを回避できる。

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 Figure 2 に示す。但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四つ (2 名 × 2 組) のパートに分かれており、ほとんどのモデルは、双子の一方から測定した変数あるいは 1 組の双生児から測定した変数に関係した方程式にすべ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、パス図全体ではなく、1 人から測定した変数、あるいは双子 1 組から測定した変数に関係した部分のパス図だけを掲載することが多い。

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 Figure 2 には、四つの観測変数から三つの共通因子を抽出し、なおかつ 12 個の独自性が存在しているので、パス図の原則的なルール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て、それが省略されている。このような省略の流儀に慣れると行動遺伝学モデルは理解しやすくなる。

Table 1 は、Neale & Cardon (1992, p.254) 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 に振り分けて計算している。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。

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、共通因子を一つ仮定してモデルの母数を推定した。Table 2 には、四分相関係数を通常のピアソンの相関係数とみなした場合の最尤推定値を示した。近年、Muthen (1978) を端緒として、0-1 デー

Figure 2. 遺伝因子分析モデル。

Table 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)

Table 2
遺伝因子分析モデルの母数の推定値

	遺伝要因		共有環境要因		非共有環境要因	
	共通因子	独自因子	共通因子	独自因子	共通因子	独自因子
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

Table 3
6 要因の説明割合

	遺伝要因		共有環境要因		非共有環境要因	
	共通因子	独自因子	共通因子	独自因子	共通因子	独自因子
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%

丸め誤差の影響で和は正確に 100%にはならない。

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 (たとえば具体例としては豊田, 1996)。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。ここでは読者の追計算が可能になるように通常の最尤推定法を利用する。

Figure 2 から明らかなように、各観測変数は六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 乗値は各因子の説明割合となる。Table 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)。四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。

本節で分析された四つの疾患に皮膚病をいくつか加えれば、湿疹といっしょに第2因子を形成する可能性がある。このように式(16)、(18)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、一般の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、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（豊田，1992，pp.156-160）。

双生児家族モデル

行動遺伝学で扱う遺伝要因は、ここまでの議論ですでに明かなように、生体内の遺伝子の機能をそのまま表現した概念ではなく、多くの心理学研究で導入される構成概念（因子分析における因子など）と本質的に同じであり、実体はない。しかし実体がないということは、有用でないということではない。

たとえば経済学における景気動向指数にも実体はない。実体は、企業が倒産したとか業績を伸ばした（ミクロレベルのDNAの膨大な相互作用）等の個々の現象であるが、膨大な現象を眺めていても巨大な社会（生命体の複雑な営み）の全体的な変化を捕えきれない。そこで、実在しないけれども有用で我々の素朴な直感に合致する景気動向指数（遺伝要因という構成概念）を利用するのである。行動遺伝学における構成概念に関しても、その分析から得られる教育学的な知見を、実験室レベルの分子生物学から得ることは困難である。

本節では上記のような行動遺伝学の特徴をふまえて、家族内の二つの世代間で表現型がどのように受け継がれるかを表現する因果モデルを紹介する（Cardon, Fulker, & Jöreskog, 1991）。これまで行動遺伝学で利用する構成概念は、解釈が容易になるように互いに無相関であることが仮定されていた。無相関を仮定すれば各要因の分散の和が観測変数の分散の推定値に一致するためである。この中で非共有環境と他の二つの要因との間には無相関を仮定しても比較的無理がないだろう。しかし遺伝と共有環境には Figure 3(a) のように相関 s を仮定したほうが常識的な直感に合致する。何故ならば、たとえばピアニストの家に生まれた才能ある子供は、生まれたときからピアノに触れることができるように、才能と家庭環境は連関している可能性が高いためである。

ところが遺伝 ACE モデルでは、四つの観測変数の分散構造は互いに完全に一致している。このため標本分散と標本共分散の合計は六つなのに、母数は三つまでしか識別できず、遺伝 ACE モデルはちょうど識別モデルとなって、遺伝と共有環境の相関は推定できない。

ここでは逆回帰（reverse regression）を利用した Figure 3(b) のような工夫を行い、遺伝と共有環境の相関を表現する。ここで逆回帰モデル中の誤差変数間の共分散は、

$$E[r_a \ r_c] = s - (a + sc)(c + sa) \quad (25)$$

$$E[r_c \ r_e] = -(c + sa)e \quad (26)$$

$$E[r_a \ r_e] = -(a + sc)e \quad (27)$$

のように表現される。係数および誤差変数間の共分散をこのように表現すると、Figure 3(a) と Figure 3(b) のパス図は“遺伝”、“共有環境”の間の共分散が一致する。逆回帰とはモデルに登場する構造変数の間の共分散を変化させずに、パスの向きを逆方向にする技術である。

この考え方を拡張して夫婦の表現型それぞれに ACE モデルが成立すると仮定し、夫婦の観測変数間の相関が μ であるとする、Figure 4の ACE モデル中の外生変数間の相関は

$$E[em \ ef] = e^2\mu \quad (28)$$

$$E[em \ cf] = E[cm \ ef] = e(c + sa)\mu \quad (29)$$

$$E[em \ af] = E[am \ ef] = e(a + sc)\mu \quad (30)$$

$$E[am \ cf] = E[cm \ af] = (a + sc)(c + sa)\mu \quad (31)$$

$$E[am \ af] = (a + sc)^2\mu \quad (32)$$

$$E[cm \ cf] = (c + sa)^2\mu \quad (33)$$

のように表現することができる。これらの相関は、広義の選択結婚（assortative mating）の効果である。特に $E[am \ af]$ は夫婦間の遺伝的素質の相関であり、狭義の選択結婚の効果として解釈できる。

式(28)から式(33)に登場した母数の中で a, c, e は、子供の世代の ACE モデルによって表現されているから、両親と2人の子供の表現型を同時に測定すれば、

Figure 3. 回帰と逆回帰モデル。

識別可能なモデルを書くことができる。

具体的には、まず一卵性双生児のいる家族内で表現型がどのように受け継がれるかを記述する因果モデルとして Figure 5 を導入する。両親の表現型から本人の遺伝的要因への形成的指標は、逆回帰の知見から得られた制約母数を利用する。続いてその影響を 1/2 ずつ受けて“子供世代の遺伝的素質”を構成する。子供世代の表現型は、相関のある遺伝的要因と共有環境、およびそれらとは無相関な非共有環境との重み付き和で生成される。また親の表現型から子供の共有環境への影響 (z) も推定する。

二卵性双生児のいる家族内の因果モデルでは、Figure 6 のように各子供ごとに別々の“遺伝的素質”を構成する。また二つのパス図に登場するすべての構成

Figure 4. 選択結婚の効果。

Figure 5. 一卵性双生児の双生児家族モデル。

Table 4
一卵性双生児と両親の共分散行列

	双生児 1	双生児 2	母親	父親
双生児 1	0.95			
双生児 2	0.54	0.89		
母親	0.08	0.21	1.11	
父親	0.28	0.18	0.11	0.90

Table 5
二卵性双生児と両親の共分散行列

	双生児 1	双生児 2	母親	父親
双生児 1	1.21			
双生児 2	0.29	1.03		
母親	0.23	0.16	0.91	
父親	0.13	0.09	0.21	1.08

概念の分散が 1 に基準化されるように誤差変数の分散に制約が設けられる（豊田, 1992, p.78）。

Cardon, Fulker, & Jöreskog (1991) では、“社会的批判の恐れ (fear of social criticism)” がどのように受け継がれていくかに関する考察が行われ、Table 4 と Table 5 で示された一卵性双生児 (144 組) と二卵性双生児 (106 組) の標本共分散行列が分析された。

標準化された母数の推定値は

$$\hat{a}=0.79, \quad \hat{c}=0.32, \quad \hat{e}=0.66, \\ \hat{z}=-.40, \quad \hat{s}=-.32, \quad \hat{\mu}=0.15$$

である。“遺伝的要因”と“共有環境要因”の相関は、知能などの能力を分析すると正の値を得ることも多いけれども、本分析では s に負の推定値が得られている。この例は、行動遺伝学モデルを用いることにより、遺伝的規定に抗する環境要因の影響を積極的に示

Figure 6. 二卵性双生児の双生児家族モデル。

すことも可能であることを示している。

モデルに登場する変数の平均は0、分散は1が仮定されているから、各要因間の相関係数は

$$E[\widehat{em} \widehat{ef}] = 0.07 \quad (34)$$

$$E[\widehat{em} \widehat{cf}] = E[\widehat{cm} \widehat{ef}] = 0.01 \quad (35)$$

$$E[\widehat{em} \widehat{af}] = E[\widehat{am} \widehat{ef}] = 0.07 \quad (36)$$

$$E[\widehat{am} \widehat{cf}] = E[\widehat{cm} \widehat{af}] = 0.01 \quad (37)$$

$$E[\widehat{am} \widehat{af}] = 0.07 \quad (38)$$

$$E[\widehat{cm} \widehat{cf}] = 0.00 \quad (39)$$

と推定される。この特性に関しては選択的結婚の影響はほとんど無視できる程度である。行動遺伝学モデルによる分析結果は、ここでも遺伝と共有環境の相関、選択的結婚の効果について、世間の偏見を打ち消し、悲観的なステレオタイプを払拭する解釈をもたらしている。

行動遺伝学の研究過程では、Jensen (1969) の例を引くまでもなく、差別やタブーに直結するデータを解析するので、分析結果の解釈と使用には細心の注意を要する。しかし上記の例、およびアルコールの消費の分析例で示した教育可能性を行動遺伝学的アプローチ以外の方法で示すことは、ほぼ不可能である。大切なのは、客観的な分析結果を教育可能性の発見・拡大につながる観点から利用するという信念をもつことである (豊田・前田・柳井, 1992, p.217)。

考 察

——優れたモデルの構成によせて——

前節までに、行動遺伝学の研究の中で提案されてきたいく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”、“関数関係の制約”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。

反復測定・多群の測定と等値母数

本論文で紹介した行動遺伝学モデルでは、一つの表現型 (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、四つの変数の分散構造が $a^2 + c^2 + e^2$ のように一致するし、遺伝因子分析モデルでは各個人内の四つの共分散行列の構造が $\Sigma_a + \Sigma_c + \Sigma_e$ のように一致していた。モデル構成に利用できる母数の数は、たかだか、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贅沢なモデル構成をしている。通常は、ちょうど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。しかし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も、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、次の理由から、行動遺伝学モデルは望ましい母数配置を採用している。

現在、心理学研究の過程では、たとえば本調査に利用した項目の因子構造を調べる場合など“因子数を適当に選択すれば、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”という誤った認識がある。しかし、現実には因子分析モデルに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。

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くらいでは、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。

さらに問題なのは、標準誤差が小さく分析結果の適合度が高くとも、それだけでは、必ずしも安定した知見を得られるとは限らないということである。行動科学の研究分野では、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が少なくない。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。特に、心理学の研究では、知り合いのつてを頼って収集したデータを分析することが多いから、各種の指標をそのまま利用することはできない。測定状況が微妙に変化したときにも、そこから得られる知見は安定しているか (頑健性があるか) 否か、という視点が重要になってくる。理想的なサンプリングが困難な状況で、データごとに、そのデータのためだけの母数を推定していたのでは、その分析から得られた知見が安定したものなのか、不安定なものなのかを実感できないのである。

この問題に関して、行動遺伝学の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、しかも同一の母数で表現していた。本論文では2群の分析のみを紹介したが、実際の双生児には五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はずである（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）。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。

このことを一般の心理学研究に敷衍するならば、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である。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もとで因子分析を行う。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、一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、変数ごと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は上げることは難しくなるだろう。しかし、制約のもと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やすくなる。

固定母数・制約母数の利用

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す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や制約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三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二つの固定母数（1.0と0.5）を導入したために識別が可能になっている。

制約母数は家族モデルの中で巧みに導入されていた。このモデルは、識別問題の制約から自由パラメータに設定できない遺伝と共有環境の関係、選択結婚の効果等、日常生活における素朴な現象の記述を行う母数を推定するために、制約母数を利用していた。

これらの事例は“母集団および変数群の特徴を記述するための少数の固定・制約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すべ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制約する（豊田・前田・柳井, 1992, p. 198）等の工夫は有効である。もちろん、思い込みにす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（指定した固定母数がデータとの適合を損ねていないかどうか調べる検定）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・制約母数はもっと積極的に使用するべきである。

文 献

- 安藤寿康 1996 遺伝要因が教授・学習過程に及ぼす諸効果——双生児統制法による英語教授法比較研究——教育心理学研究, **44**, 223-233.
- Bentler, P. M. 1986 Structural modeling and psychometrika: A historical perspective on growth and achievements. *Psychometrika*, **51**, 35-51.
- Bentler, P. M., & Weeks, D. G. 1980 Linear structural equations with latent variables. *Psychometrika*, **45**, 289-308.
- Bock, R. D., & Bargmann, R. E. 1966 Analysis of covariance structures. *Psychometrika*, **31**, 507-534.
- Bollen, K. A. 1989 *Structural equations with latent variables*. New York: Wiley.
- Cardon, L. R., Fulker, D. W., & Jöreskog, K. G. 1991 A LISREL 8 model with constrained parameters for twin and adoptive families. *Behavior Genetics*, **21**, 327-350.
- Falconer, D. S. 1960 *Quantitative genetics*. Edinburgh: Oliver and Boyd.
- Falconer, D. S. 1990 *Introduction to quantitative genetics*. 3rd ed. New York: Longman Group Ltd.
- Heath, A. C., Neale, M. C., Hewitt, J. K., Eaves, L. J., & Fulker, D. W. 1989 Testing structural equation models for twin data using LISREL. *Behavior Genetics*, **19**, 9-36.
- Jensen, A. R. 1969 How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, **39**, 1-123.
- Jöreskog, K. G. 1970 A general method for analysis of covariance structures. *Biometrika*, **57**, 239-251.
- Jöreskog, K. G., & Sörbom, D. 1976 *LISREL III: Estimation of linear structural equation systems by maximum likelihood methods*. Chicago: National Educational Resources, Inc.
- 狩野 裕 1990 因子分析における統計的推測：最近の発展 行動計量学, **18**, 3-12.
- McArdle, J. J., & McDonald, R. P. 1984 Some alge-

- braic properties of the reticular action model for moment structures. *British Journal of Mathematical and Statistical Psychology*, **37**, 234-251.
- McDonald, R. P. 1978 A simple comprehensive model for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, **31**, 59-72.
- Muthen, B. O. 1978 Contributions to factor analysis of dichotomous variables. *Psychometrika*, **43**, 551-560.
- Neale, M. C., & Cardon, L. R. 1992 *Methodology for genetic studies of twins and families* Dordrecht, NL: Kluwer Academic Publishers.
- Spearman, C. 1904 General intelligence, objectively determined and measured. *American Journal of Psychology*, **15**, 201-293.
- 芝 祐順 1975 行動科学における相関分析法 第2版 東京大学出版会
- 豊田秀樹 1990 共分散構造の表現 教育心理学研究, **38**, 438-444
- 豊田秀樹 1992 SASによる共分散構造分析 東京大学出版会
- 豊田秀樹・前田忠彦・柳井晴夫 1992 原因をさぐる統計学 講談社
- 豊田秀樹 1996 公共事業への態度形成に関する因果モデル構成の試み 応用社会学研究, **38**, 41-50.
- Vandenberg, S. G. (Ed.) 1965 *Method and goals in human behavior genetics*. New York: Academic Press.
- Vandenberg, S. G. (Ed.) 1968 *Progress in human behavior genetics*. Baltimore: Johns Hopkins University Press.
- Wright, S. 1921 Correlation and causation. *Journal of Agricultural Research*, **20**, 557-585.
- 1996.5.20 受稿, 1996.11.9 受理——