再討論：誤差共分散の利用と特殊因子の役割

狩野　裕[1]

Rejoinder: Use of error covariances and the role of specific factors

Yutaka Kano

In this rejoinder, special attentions are paid to error covariances and specific factors in the comparison between SEM and traditional methods. When a factor analysis model receives a poor fit, it does not make sense to simply remove important variables although inconsistent with the factor analysis model, as pointed out by the discussants. It is to be emphasized that a better way than removing the variables is to allow for error covariances, in order to overcome the inconsistency problem. The model with error covariances guarantees the invariance of estimation results over item selection.

The discussants pointed out that an important difference between a scale score (sum of items) and a measurement model by effect indicators in SEM is that a scale score includes specific factors whereas a measurement model excludes them. Practitioners could use scale scores when they are interested in effects of specific factors as well as a common factor. It is argued, however, that appropriate use of error terms and a common factor in SEM can make better inference than the use of unidimensional scale scores, because the error terms of effect indicators contain information on specific factors and they can individually evaluate the effects of the common factor and the specific factors in SEM.

Other related topics are also discussed.

1　はじめに

本特集号の編集の労をとって下さった本誌編集委員の豊田秀樹氏と，狩野 (2002) について真摯な議論を提供した３名の討論者に感謝を申し上げる．各討論者は，構造方程式モデリング (SEM) の長所・短所について，やや異なった観点からの議論を展開している．結果として，方法論的な観点からも応用的観点からもバランスのとれた特集号になったのではないかと考えている．改めて，本企画のオーガナイザーとそれに応えた討論者に感謝したい．

ここでは,「立場の違い」や狩野 (2002) において議論しなかった「重要な論点」について再整理を試み，ユーザサイドからも方法論的観点からも議論をより明確にしたい．

２節，３節と４節，７節は，それぞれ，南風原 (2002)，椿 (2002)，鈴木 (2002) に対する再討論である．５節と６は南風原・椿両氏に共通するコメントに関連する再討論である．ただしこれらは原則であって，３名の討論者の議論は互いに関連していることがあり，実際は, A 氏に対するコメントの中に B 氏に対するものが混じっていたりする．やはり，テーマが絞られると討論にも「共通因子」が現れるようである．

[1]大阪大学 大学院人間科学研究科 (Osaka University, Graduate School of Human Sciences) Suita, Osaka 565-0871, JAPAN

2　指標の数と一般化可能性

測定モデル（検証的因子分析モデル）において観測変数を指標 (indicator variable) という．直接観測できない潜在変数 (or 構成概念) を測定する観測項目という意味合いをもつ．潜在変数と観測変数の関係において,「観測変数の背後に潜む共通因子」という捉え方と「潜在変数を測定するための道具としての観測変数 (指標)」という考え方の違いが「指標」という呼称に表れている．探索的因子分析の流れでは指標という用語はあまり用いられなかったのである．

2.1　少数個の指標変数しか準備できない場合

指標の数に関して最初に強調しておきたいことは，信頼性が十分高くなるよう十分多くの指標変数を準備できないことがあるということである．その場合は，SEM では検証的因子分析である測定モデルを用いることによって誤差をコントロールすることが可能になる．SEM は，十分観測項目が用意できない状況でも正確な分析を提供するという意味で有為な拡張を行ったのである．

では，信頼性が高くない尺度しか準備できない場合はどのように分析してきたのか．希薄化を無視するという傍若な方法は論外として，一つの方法は希薄化の修正公式を用いることであろう[2]:

$$\rho_{xy} = \frac{\rho_{XY}}{\sqrt{\rho_X \rho_Y}} \tag{2.1}$$

すなわち，真の得点間の相関係数 ρ_{xy} は尺度間の相関係数 ρ_{XY} をそれぞれの尺度の信頼性（ρ_X と ρ_Y）の幾何平均で除することによって得られる．本節では，まず (2.1) による伝統的な方法と SEM による方法との差異を議論する．

具体例で考えよう．二つの構成概念間 (F1 と F2 とする) の相関を求めたいとし，それぞれの概念は，3 個の尺度項目からなる尺度で測定するとしよう．それぞれの概念において尺度項目の因子負荷は等しく 0.5 と 0.6 であるとしよう．6 個の尺度項目が因子分析モデルに従っているとすれば，図 1 の (a) のように表すことができる．

尺度 F1, F2 の信頼性係数はそれぞれ 0.50, 0.63 であるから十分に高いとは言えない．構成概念間の真の相関が 0.70 であったとしたら，尺度得点間の相関は 0.39 と推定されるはずであり，信頼性の低さによって引起こされる希薄化の影響をもろに受けてしまう．この値は (2.1) 式を用いることで

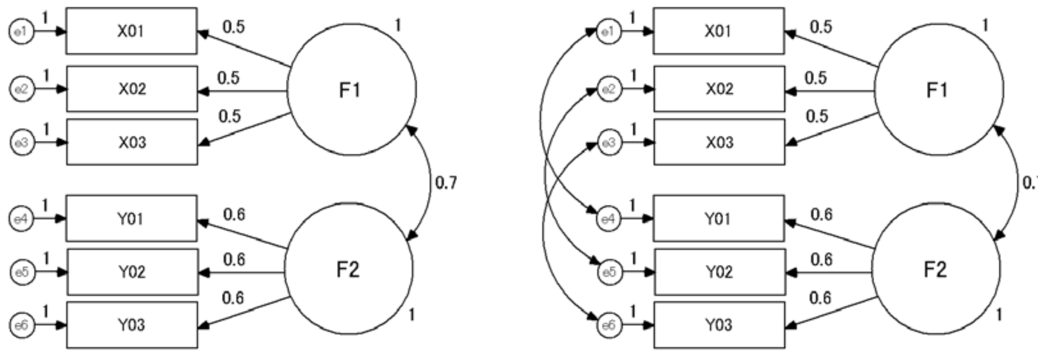
$$0.70 \times \sqrt{0.50 \times 0.63} = 0.39 \tag{2.2}$$

と計算することもできる．逆に，尺度間相関と信頼性の値から，構成概念間の真の相関を

$$\frac{0.39}{\sqrt{0.50 \times 0.63}} = 0.70 \tag{2.3}$$

と見積もることができ，この方法が上述したものである．

[2]南風原 (2002) の (1) 式と同一．狩野 (2002) の (3.1) 式にその導出方法がある．



(a) 因子分析モデルが成立している場合　　(b) 誤差に共分散がある場合

図 1: 信頼性が低い場合の分析

では，(2.3) による古典的方法と検証的因子分析による方法 (図 1(a) にもとづく分析) との違いはどこにあるのだろうか．相関の推定値は同一である．

それは，図 1(a) での分析では適合度を評価していることである．もし同モデルの適合度が悪く，たとえば図 1(b) のように誤差間に共分散が存在していたならば，尺度間相関が 0.39 とならず，誤差共分散によるバイアスが生じてしまう．したがって，公式 (2.1) による修正はうまく機能しない．

図 1(b) で導入された誤差共分散は，尺度「間」の誤差に設定されている．したがって，尺度ごとに因子分析しても誤差共分散の存在は判らない．両方の尺度を同時に (6 個の変数で) 分析すると，因子間相関が少し高めに推定され，因子負荷の単純構造が少し崩れるということが観測されるであろう．そして，もし適合度の吟味を行わないならば，そのまま，尺度間相関の算出へ進んでしまい，誤った相関係数を報告する可能性が高いと思われる．椿 (2002, 2.2 節) が指摘するように，希薄化された相関が誤差相関からの上乗せで緩和され，それが希薄化の解消と考えるとすれば，本末転倒である．

SEM では，真の状況が図 1(a) であれ (b) であれ，このようなモデルを正確に実現することができ，構成概念間の相関を正しく推定することができる[3]．

つぎに，図 1(b) の誤差共分散が生じるのはどのような場合かを考えてみたい．ひとつの典型例は，ある検査を 3 回実行し，１回目に (X01,Y01)，２回目に (X02,Y02)，3 回目に (X03,Y03) を測定したという状況である．同時に測定した項目の誤差間に共分散が生じることはしばしば見受けられる．南風原 (2002) の (2) 式のモデルで考えれば，e_x と e_y との相関であり，測定誤差の要因 (1)「体調や動機づけのレベルなど，被験者の内的状態の日々の変動」による相関である．もう一つの例は，X01 と Y01 の測定方法や測定者が同じであるが，X02 と Y02 や X03 と Y03 とは異なるという場合である．これらは方法因子 (method factor) による相関という．

南風原 (2002, 2.3 節) は「希薄化の修正公式は，心理学などの研究論文においてルーチン的に適用され，修正後の値が相関係数のよりよい推定値として報告されることはなかっ

[3]図 1(b) のモデルは識別可能でない．誤差共分散が等しいとか因子負荷が等しいなどの制約を一つ以上おくことで推定可能となる．

た」と述べ，再検査法によって信頼性を推定して公式 (2.1) を適用することが適切でない巧妙でかつ現実的な例を与えている．実は，図 1(b) は本質的に南風原の例と同一である．南風原の記号を用いて，彼の考察した状況を表すモデルをパス図で表せば，図 2(上) となる．ただし，一回目の測定を添え字 1 で二回目の測定を添え字 2 で表している．このパス図によ

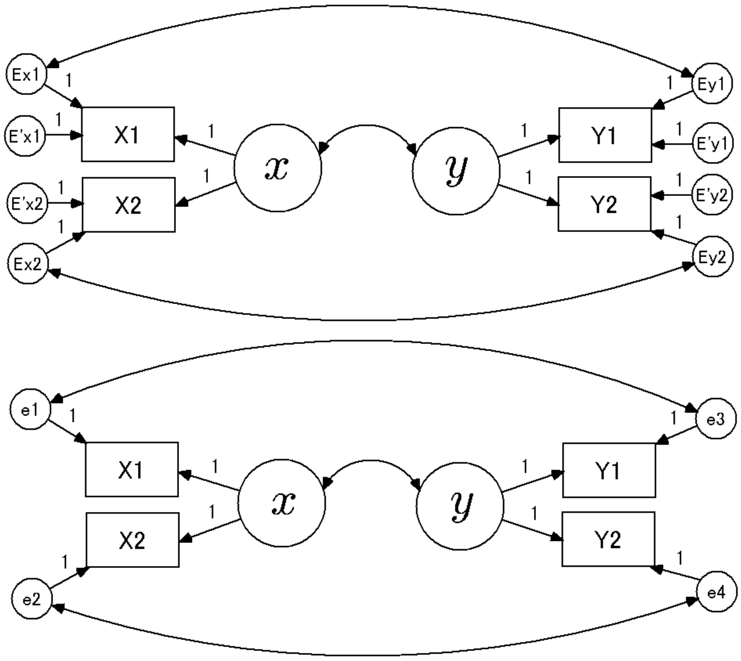


図 2: (上) 南風原のモデル　(下) SEM での分析モデル

るモデルでは $\mathrm{V}(E_{x1})$ と $\mathrm{V}(E'_{x1})$ などが区別できないので，これらの誤差分散を合併すると図 2(下) となる．誤差の合併は，$\mathrm{Cov}(x,y)$ や誤差共分散の推定にまったく影響を及ぼさないことに注意する．このモデルは図 1(b) で指標を 2 つにして因子負荷量を 1 とおいたものである．

南風原の指摘は，(2.3) による希薄化の修正は誤差共分散を無視することになり正しくないというものであり，それは正しい．SEM を用いて図 2(下) で分析するならば，$\mathrm{Cov}(x,y)$ と $\mathrm{Cor}(x,y)$ を正しく推定でき，そして，タイプ 1 の要因による誤差間の共分散 $\mathrm{Cov}(e_x,e_y)$ も正しく推定することができる．万一，図 2(下) のモデルで誤差共分散を導入しなかったならば，誤差共分散（の一部）は $\mathrm{Cov}(x,y)$ に反映されてしまう．南風原はこのことを「被験者の内的状態の日々の変動 (タイプ 1) による得点変動が誤差ではなく，潜在変数 f に含まれている $\cdots$」と表現している．しかし，SEM では誤差共分散を導入しない場合は「適合の悪化」を招くから，(2.3) による希薄化の修正のような間違いを犯すことはないのである．

以上のことから学ぶべきことは，適合度を評価することの重要性であり，そして，推定を段階に分けて行うのではなく，現象全体についてモデリングを行い一括して推定することの重要性である．

なお余談であるが，二回目の測定に欠測が生じた場合においても，欠測のメカニズムが

Missing at Random (MAR) であれば[4]，すべてのデータを活かしかつバイアスのない推定値を (市販ソフトウェアを用いて) 得ることができる．

南風原は希薄化公式が機能しない例を挙げ，ここではそれが SEM によって適切に分析されることを示したわけであるが，一方，尺度得点を計算する場合はどのようになるのであろうか．尺度 X と Y に尺度項目がそれぞれ p 個と q 個あるとし ($X = \sum_{i=1}^{p} X_i$，$Y = \sum_{j=1}^{q} Y_j$)，それらが以下の構造をもっているとする:

$$\begin{aligned} X_i &= x + e_x + e'_{xi} \quad (i = 1, \cdots, p) \\ Y_j &= y + e_y + e'_{yj} \quad (j = 1, \cdots, q) \end{aligned}$$

ここで，$e'_{x1}, \cdots, e'_{xp}$ と $e'_{y1}, \cdots, e'_{yq}$ は，それぞれ独立同一分布をもち，南風原 (2002) のように

$$\begin{aligned} \mathrm{Cov}(x, e_x) &= \mathrm{Cov}(x, e'_{xi}) = \mathrm{Cov}(e_x, e'_{xi}) = 0 \\ \mathrm{Cov}(y, e_x) &= \mathrm{Cov}(y, e'_{yj}) = \mathrm{Cov}(e_y, e'_{yj}) = 0 \\ \mathrm{Cov}(e'_{xi}, e'_{yj}) &= 0 \end{aligned}$$

を仮定する．このとき，尺度間相関は

$$\mathrm{Cor}(X, Y) = \frac{pq\mathrm{Cov}(x, y) + pq\mathrm{Cov}(e_x, e_y)}{\sqrt{\left(p^2\mathrm{V}(x) + p^2\mathrm{V}(e_x) + p\mathrm{V}(e'_{x1})\right)\left(q^2\mathrm{V}(y) + q^2\mathrm{V}(e_y) + q\mathrm{V}(e'_{y1})\right)}}$$

となる．もし $e_x = e_y = 0$ であるならば，誤差 $\mathrm{V}(e'_{x1})$ の影響は $\frac{1}{p}$ のオーダーで，$\mathrm{V}(e'_{y1})$ の影響は $\frac{1}{q}$ のオーダーで 0 に収束し，その結果，尺度間相関は真の相関に収束する．このことが，南風原 (2002, 2.1 節) による「尺度の信頼性が，各項目の信頼性を項目数に応じて高めたものとなること」に符号する．しかし，$e_x = e_y = 0$ でないならば，たとえ $p, q \to \infty$ としても，尺度間相関は

$$\mathrm{Cor}(X, Y) \to \frac{\mathrm{Cov}(x, y) + \mathrm{Cov}(e_x, e_y)}{\sqrt{\left(\mathrm{V}(x) + \mathrm{V}(e_x)\right)\left(\mathrm{V}(y) + \mathrm{V}(e_y)\right)}} \neq \frac{\mathrm{Cov}(x, y)}{\sqrt{\mathrm{V}(x)\mathrm{V}(y)}}$$

となって，真の得点間の相関をバイアスなく推定しているとは限らないのである．

2.2　多くの指標変数が準備できる場合

前節では，主に構成概念について少ない測定項目しか準備できない場合における SEM の特徴について述べた．本節では，測定項目が十分用意できる場合の SEM の役割について考えてみたい．

狩野 (2002, p.15) の言及「CFA の大きな特徴は，モデルが適合しているならば，変数の削除によって分析結果が大きな変化を受けないということである」に対して，南風原 (2002,

[4]MAR については岩崎 (2002) や Little & Rubin(1987) を参照されたい．

4.1 節) は「たまたま選ばれた項目群を用いて得られる結果から，測定したい内容を広くカバーした項目ユニバースを用いたときに得られるであろう結果への一般化可能性が低い」と述べている．狩野は「モデルが適合している」という状況での言及であるのに対して，南風原はモデルが適合していない，特に，多数の項目のある場合に 1 因子モデルが適合しないという状況での主張であるから，両者の主張に矛盾はない．重要なのは，項目ユニバースは厳密には因子分析モデルに従わないという指摘である．たとえば，一つの構成概念に 10 個の指標変数をもってくれば，まず（1 因子の）因子分析モデルは適合しない．そして，10 個の指標からたとえば任意の 3 つの変数を選んで指標とすれば，選ばれた変数の組によってその後の分析の結果が変化することは十分に考えられる．

南風原 (2002, 5.2 節) は,「モデルの適合度を犠牲にしても追求すべきこと（結果の一般化可能性および測定の妥当性）があるのではないか，モデル適合度を，研究の目的に照らして相対化してとらえる必要があるのではないか」と述べている．この問いに対する筆者の回答は，狭い意味での因子分析モデルに拘泥する必要はなく，例えば誤差共分散を導入するなどして最低限の適合度は確保したいというものである．

南風原氏が述べているように，典型的な因子分析モデルは，共通因子の影響を除いたときの偏相関がすべてゼロになるということである[5]．確かにこの仮定は制約的であり，変数が多くなれば成立しないことが多い．ということであるならば,「偏相関がゼロ」という仮定を緩めればよい．すなわち，偏相関が存在する変数の誤差間に共分散を認めたモデルを立てれば，モデルは適合することが多くなる．誤差共分散を導入したモデルにおいても，変数の選択に関する不変性が成立する．したがって，項目ユニバースに対する一般化可能性は高くなる．測定モデルとして，誤差が独立という仮定に縛られることはないのである．

観測変数の選択に関して不変性があるということは，原理的には，少数個の指標変数で構成概念が測定できるということである．そして，このことは回答者の負担減という意味で SEM のメリットと考えることもできよう．しかし，議論はそう単純ではなく，引き続き 6 節で議論される．

実は，合計得点である尺度得点と因子分析モデルによる測定 (結果指標) とは，希薄化の問題以外に，本質的に異なる側面がある．この問題は，少数個の指標変数しか準備できない状況でも同様であり，2.4 節で議論する．

2.3　適合の悪さの影響

南風原は「モデルの適合度を高めることが，研究の目標・目的にどれだけ適合したことなのか，どういう意味で重要なのか」という疑問を投げかけている．この疑問に関して，適切な例を用いてユーザに対して警鐘を鳴らしてこなかったのは，筆者を含めた方法論者の怠慢であると言ってよい．

一つの実例は，狩野 (2002, 2.2.2 節) で示した信頼性係数に生じるバイアスである．適合の悪い誤差共分散のないモデルでの分析結果を信用するならば，信頼性係数の推定値は $\alpha = 0.747$, $\hat{\rho} = 0.746$ となるが，バイアスのない本来の推定値は $\hat{\rho}' = 0.703$ である．

[5] この仮定は局所独立性とも言われる．

二つ目の例として，図 1(b) の状況のとき，誤差共分散を設定しない狭義の因子分析モデル (図 1(a)) で分析したとき，因子間相関にどのようなバイアスが生じるかを調べてみる．誤差共分散を 0.2(どの組合せに対しても一定)，$n = 200$ として，表 1 に，許した誤差共分散の数ごとに因子間相関の推定値と適合度（カイ 2 乗値）を示している．なお，母集団共分散行列を分析しているのでサンプリングエラーはない．したがって，誤差共分散の数=0 の場合は完全な適合となり，因子間相関においては真の値 0.70 が再現されている．

表 1: 適合度と推定値のバイアス ($n = 200$)

誤差共分散の数	因子間相関の推定値	適合度 カイ 2 乗値	df	p 値
0	0.70	0.000	8	1.000
1	0.80	9.416	8	0.308
2	0.87	21.517	8	0.006
3	0.92	34.631	8	0.000

表 1 から，適合が悪くなるにしたがって，因子間相関の推定値は真値である 0.70 から大きくずれていくこと ($0.70 \to 0.80 \to 0.87 \to 0.92$) が観察される．誤差共分散が存在するとそれを説明するために因子間相関の推定値がバイアスをもつのである．また，(同符号の) 誤差共分散がたくさん存在するとバイアスはより顕著になる．

誤差共分散の数が 1 の場合は，適合度検定は非有意を示している．しかし，誤差共分散の LM 検定を行えばその有意性を確認することができる．また，実際の状況ではサンプリングエラーが付加されるのでほとんどの場合でカイ 2 乗検定が有意になると考えられる．

ここまで 2 つの例を用いて，因子分析モデルの適合が悪い場合に，それを無視すると分析に生じる問題点を明らかにしてきたが，他にも別の問題が生じる可能性がある．モデルの仮定が崩れているとどんな問題が生じるのか，一般には予測することができないと思われる．

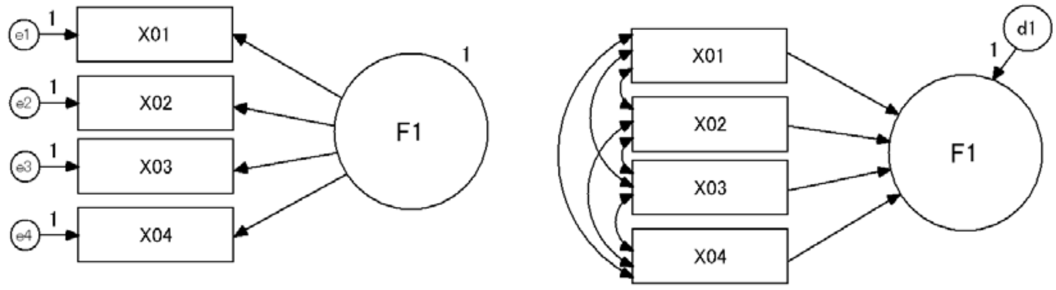
2.4 結果指標と原因指標

SEM による構成概念の測定方法は二つある．それらは結果指標と原因指標とよばれており，パス図で表すと図 3 となる．これらの指標は，指標変数が潜在変数から見て結果変数となっているか原因変数となっているかにおいて異なる[6]．尺度得点を用いることは原因指標と近い考え方を採っていることになる．因果の方向を正確に記述するという観点では結果指標のモデルが適切であったとしても，分析の目的によっては原因指標を採用することもある．

観測項目 X_i が 1 因子の因子分析モデルに従っているとしよう：

$$X_i = \lambda_i f + s_i + e_i \quad (i = 1, \cdots, p) \tag{2.4}$$

[6] 豊田 (1992, p.155) にも記述がある．



(a) 結果指標　　(b) 原因指標

図 3: 構成概念の測定方法

ここで s_i と e_i は特殊因子と誤差因子である[7]．一般に s_i と e_i とは区別できないため，それらの和を u_i と書いて，独自性 $\mathrm{V}(u_i)$ を推定対象とするのは狩野 (2002) で述べたとおりである．結果指標は，X_i から特殊因子と誤差因子を取り除いて共通因子 f を取り出すという機能がある．これは，南風原 (2002) の表現を借りると「タイプ 3 による要因を新たに誤差要因に加えて，それとタイプ 2 の要因による誤差を合わせて除去する」ということになる．

原因指標においてパス係数をすべて 1 とおくと

$$F_1 = \left(\sum_{i=1}^{p} \lambda_i\right) f + \sum_{i=1}^{p} s_i + \sum_{i=1}^{p} e_i + d_1 \tag{2.5}$$

となるから，撹乱項において $d_1 = 0$ とすれば尺度得点と一致する．このことからわかるように，尺度得点の利用は，実は SEM の一つの下位モデルと同等であって，狩野 (2002) で展開した「尺度化 versus SEM」という枠組みよりも，SEM のいくつかのモデルの中でどれを採用すべきかというのがより正確な記述といえよう．このことは椿 (2002) において指摘されている．

(2.5) において $\sum_{i=1}^{p} e_i$ は除去すべき要因であることは万人の認めるところであろうが，特殊因子 $\sum_{i=1}^{p} s_i$ については状況依存である[8]．特殊因子を排除するべき状況では結果指標を利用し，もし，特殊因子を無視しない場合は原因指標 (or 尺度得点) を用いるということも一案かもしれない．

(内容的) 妥当性は，構成概念のモノサシである尺度の重要な性質である．村上 (2002a) には，妥当性とは「尺度得点が研究者の想定している構成概念と対応している度合い」とある．妥当性をもたない尺度は意味がない．尺度においては，研究者が研究目的に合わせて定義した構成概念に関して，必要ないくつかの側面のすべてが測定されていなければならない．この意味で，複数個の質問項目が用意され，伝統的には，それらを加算することでモノサシを作ったわけである．特殊因子は構成概念のいくつかの側面を反映したものと考えるこ

[7] ここでは s_i を特殊因子とよんでいるが，実は，特殊因子間に共分散が設定されることがあり，この意味では，狭い意味での因子分析モデルで用いられている特殊因子という用語は適切でないかもしれない．もはや，その観測項目にしか関連しない “specific” な因子ではないからである．特殊因子間の共分散は誤差共分散の典型例である．

[8] d_1 の役割は別の機会に議論することとする．

とができる．狩野 (2002) でも取り上げた WISC-R の分析モデルでは，動作性因子を測定するテスト項目は「X07: 絵画完成」から「X12: 迷路」までの 6 つの側面を捉えているわけで，それらの個性 (特殊性) は置き換え可能ではない．すなわち，

$$s_i \neq s_j \quad (i \neq j)$$

である．古典的な尺度構成で観測項目の合計得点を用いてきたということは，複数個の側面をすべて合計した得点

$$\left(\sum_{i=1}^{p} \lambda_i\right) f + \sum_{i=1}^{p} s_i \tag{2.6}$$

をターゲットとしてきたと考えることができる．

結果指標によって定められるのは共通因子 f であるということは既に述べた．構成概念を数量化したものとして，(2.6) と f のどちらを用いるべきなのだろうか．筆者の意見は，(i) 特殊因子 s_i を加えるべきかどうかは研究の目的による，そして，(ii) 特殊因子 s_i を加えたいときでも観測項目の合計得点の利用は必ずしも万能でない，ということである．以下でそれを説明する．

項目の個性 (特殊性) を活かしたいとき，実は，伝統的な尺度得点 $\sum_{i=1}^{p} X_i$ の利用にはいくつか問題がある．一つ目は，尺度得点における個性 s_i の相対的な大きさが，加算する尺度項目の数に依存することである．簡単のため $\lambda_i = 1, \mathrm{V}(f) = \mathrm{V}(s_i) = \mathrm{V}(e_i) = 1$ としよう．このとき，尺度得点における s_i の変動の占める割合は

$$\frac{\mathrm{V}(s_i)}{\mathrm{V}(\sum_{i=1}^{p} X_i)} = \frac{1}{p(p+2)} \tag{2.7}$$

となる．したがって，X_i の個性 s_i のインパクトは尺度項目の数 p について $\frac{1}{p^2}$ のオーダーで小さくなっていく．また，$\sum_{i=1}^{p} X_i$ と $\sum_{i=1}^{p-1} X_i$ の違いは，X_p の個性である s_p の有り無しというだけではなく，両者では $f, s_1, \cdots, s_{p-1}$ の構成要素の割合も異なっている．この意味で，尺度得点は，尺度項目の選択にセンシティヴである[9]．他の問題の一つは，多次元性である．村上 (2002a) は心理測定の本質として「多次元性」を挙げている．1 因子の因子分析モデルで考えるならば，誤差を取り除いたとしても，1 次元である共通因子 f だけで多次元観測変数の変動を説明できないということであろう．すなわち，特殊性 s_i の存在，個性の存在を指摘しているわけである[10]．

筆者は，因子分析の価値の一つは，観測変数の変動を共通因子と特殊因子 (+誤差因子) へ分解することにあると考えている．因子分析は，1 因子の場合 p 次元の観測変数を $(p+1)$ 次元の潜在変数へ分解するわけであるが，これを単に加算することで 1 次元に落とすのではなく，多次元のまま次の段階の分析 (e.g., 回帰分析・相関分析) に活かすことはできないのだろうか．実は，SEM はそれをも可能にし，上述した尺度得点の問題を解決するのである．アイデアは単純で，誤差項からパスをひくのである．

[9] 原因指標の項目選択依存の問題は多くの論文で指摘されている．例えば，Diamantopoulos & Winklhofer (2001) をみよ．

[10] 加えて，s_i の間の独立性が崩れていることも指す．

2.5 誤差項からパスをひくモデル

構成概念を具現する潜在変数を f として，f を測定する指標変数を $X_1, \cdots, X_p$ とする．基準変数を Y とするとき，Y を f と e_i とから予測するモデルを考える：

$$
\begin{aligned}
X_i &= \lambda_i f + e_i \quad (i = 1, \cdots, p) \\
Y &= \gamma f + \gamma_1 e_1 + \cdots + \gamma_p e_p + e_{p+1}
\end{aligned}
$$

$p = 4$ のモデルをパス図で表せば図 4 となる．実はこのモデルは識別可能でないので推定に

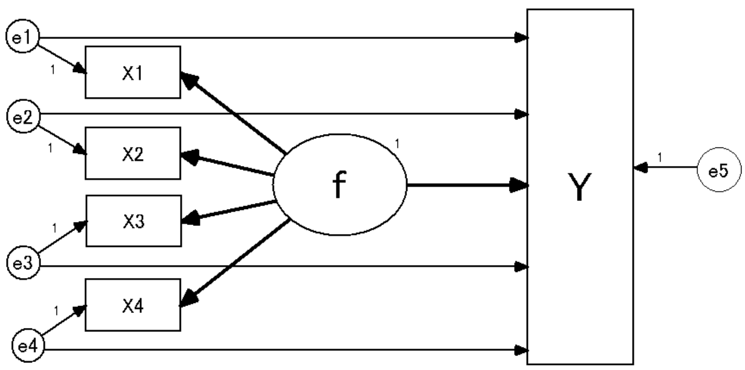


図 4: 誤差を説明変数とするモデル

は注意が必要である．まず，誤差からのパスを除いたモデルで分析する[11]．つぎに，図 4 のモデルにおいて f から Y へのパス γ を推定された値に固定したもとで二段階目の推定を行う[12]．もし，誤差からのパスが非有意であれば，個性の Y への効果は認められず，共通因子の効果のみを考えればよいということになる．すなわち，結果指標でよいということになる．

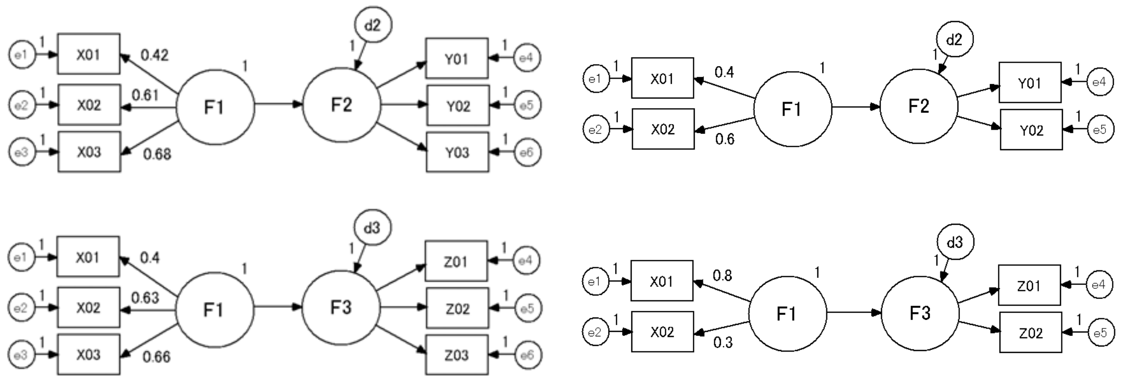
一方，次のような主張があるかもしれない．本モデルの誤差とは特殊因子と誤差因子を加えたものであるから，それを説明変数にすると希薄化が生じ，そのために特殊因子からの効果を正確に有意性検定できていない．尤もな批判である．もし，何らかの方法で特殊性 $\mathrm{V}(s_i)$ を見積もることができたならば，$s_i + e_i$ ではなく s_i の Y への影響 (希薄化されないもの) を推定することができる．Kano(2002) では，犯罪心理学における低セルフコントロールという文脈で同モデルを適用し，特殊因子の Y への効果の非有意性を報告している．

3 測定モデルの不変性

椿 (2002) の論点の一つは原因系測定構造の不変性である．ここで不変性という用語は，多母集団同時分析における因子不変性と関連するが直接それを意味しない．結果系の変数を取り替えたとき，原因系の測定モデルの推定値が変化するという問題である．椿の指摘どお

[11] 1 因子の因子分析モデルである．
[12] 詳細は Kano(2002) を参照されたい．

り，SEM にはその不変性がないが，たとえば，グラフィカルモデリングは不変性を有する．この問題は，南風原 (2002, 5.4 節) においても指摘されている．



(a) 3 指標の場合　　(b) 2 指標の場合

図 5: 指標の数と推定値

図 5(a) での 2 つのモデルをみてみよう．原因系の測定モデル (F1) を構成する変数は同一であるが，結果系の変数は異なっている．同じ変数で違った変数を予測するのである．F1 から F2 へのパス係数が F1 から F3 へのそれと異なるであろうことは容易に想像されるが，測定モデルの構造 (X01, X02, X03 の因子負荷量の推定値) が微妙に異なっており，これが，椿の指摘する不変性の崩れである．そして,「この種の要求[13]は，人文・社会科学研究では必要ないのであろうか」という疑問が投げかけられている．

図 5(a) の場合，母集団においては測定モデルは不変である，すなわち，(X01, X02, X03) の共分散行列から因子負荷量は一意的に定められる．しかし，推定においては，結果系変数との関連も情報として用いるので，不変性が成り立つような自然な制約はおかれていないということである．他方，グラフィカルモデリングは同時密度の分解を利用することで自然に不変性が成り立つ推定方法となっている．

教科書的な回答としては，不変性が必要な状況ならば多母集団の同時分析を行って等値制約をおいて推定すればよいということになる．また，測定モデルを共有するいくつかのモデルを合併して一つのモデルを構築するのも有効である．たとえば，図 5(a) の場合は図 6 のようになる．

しかし，より重要なのは，適合度が良ければ，測定モデルの違いは無視できる程度であること，裏返せば，測定モデルの推定値が大きく異なるような場合は，推定したモデルのうちいくつかのモデルの適合度が非常に悪いという認識をもつことである．適合度の良いモデルはこういった意味でも安心感を与えてくれる．

椿 (2002) はこの問題提起の後，モデルを飽和化することによって測定モデルの不変性を保つ推定法を提示している．

さて，実は，上記の議論は測定モデルが 2 指標の場合 (図 5(b)) にはまったく通用しないことに注意する．関連する問題が椿 (2002, 2.3 節) において指摘されている．図 5(b) の場

[13]不変性のこと

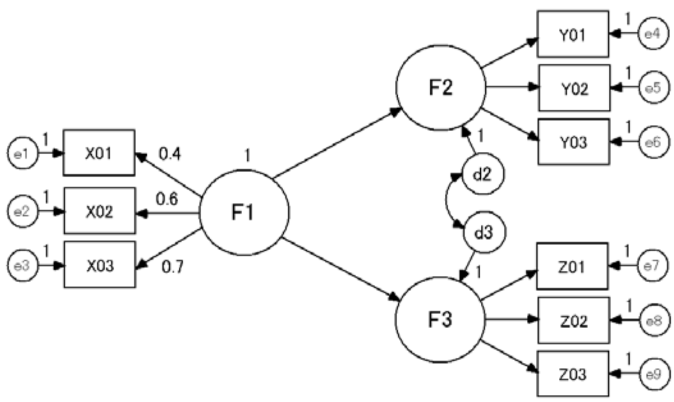


図 6: 測定モデルを共有する

合，X01 と X02 の分散共分散から因子負荷量を一意的に定めることができない．これは母集団レベルにおいてそうであり，測定モデル自身が識別可能でないことを示している．しかしながら，F2 を含めたモデル (4 変数) は一般に識別可能である．つまり，原因系の測定モデルは結果系の変数との関わりを利用して定められているのである[14]．結果として，結果系変数が異なれば，測定モデルの推定値も (もちろん母数も) 理論上異なることになる．実際，図 5(b) に提示した 2 つの測定モデルは大きく異なっている．

椿が指摘するように，測定モデルが 2 指標の場合は潜在変数の解釈に注意が必要である．図 5(b) に示された 2 つの F1 の解釈は異なるかもしれないのである．もう一つ注意しておかなければならないのは，2.4 節で議論した結果指標と原因指標についてである．結果指標の方は，3 指標以上でありかつモデルの適合がよければ，実質的に測定モデルの不変性が成立するのに対して，原因指標はそうではない，指標変数から潜在変数 F1 へのパス係数は他の変数との関連によってのみ決定される．したがって，同時に分析する他の変数が違えば F1 の意味が違ってくる可能性がある．

表 2 に（結果）指標の数と測定モデルの性質をまとめてある．指標の数が理論的には 4 以上が望ましいと言われる理由は，誤差と構成変数 (構成概念) の分離可能性，(母集団レベルの）モデルの不変性，適合度の吟味可能性，誤差共分散の導入可能性にある．

さて，椿の指摘する原因系測定構造の不変性の必要性はケースバイケースではないだろうか．不変性の性質は，議論や手続きを単純化するという観点において重要な役割を果たすのは事実である．心理学・行動科学で扱われる尺度という考え方は不変性を基盤においている．つまり，関連をみたい変数が何であろうと，尺度に変更を加えないのである．そう決めることで分析が単純化される．相手によって尺度項目や重みを変化させることはいたずらに分析を複雑にする可能性がある．

一方，正準相関分析のように相手の変数群とのかかわりで重みが決まるという分析方法もある[15]．主成分回帰では，説明変数の線形変換を決めるにあたって，従属変数への予測可

[14]結果系の測定モデルも同様．すなわち，原因系の変数との関係でもって定められている．
[15]椿 (2002, 2.1 節) で指摘済み．

表 2: （結果）指標の数と測定モデルの性質

指標の数	測定モデルの性質
1	識別不能．信頼性を既知とするか，誤差を 0(潜在変数=観測変数) とみなさざるを得ない
2	当該測定モデルだけでは識別不能．他の変数との関係で当該測定モデルが決定される (測定モデルは不変でない)
3	当該測定モデルだけで識別可能．飽和モデルのためモデル適合の吟味不可．誤差共分散の導入も不可
4 以上	当該測定モデルだけで識別可能．誤差共分散の吟味・導入が可能

能性が考慮されていないことがときに批判の対象になる．SEM では，既に指摘されたように 2 指標の測定モデルがそうである．測定モデルにこのような自由度を持たせることで説明力の向上が期待できる．一方，別の母集団との比較や時点間の比較を行いたいときは，測定モデルの不変性を要求することが多く，そのような場合には，SEM においては不変性を実現するモデリング (等値制約など) を行うことになる．

4　椿の分析モデル

椿 (2002) は 2.2 節において,「尺度化＋回帰分析」の代替として二つのタイプのモデリングを提案している．一つ目は因子得点を求めてから，単純な回帰分析ではなく，変量内誤差モデルで (偏) 回帰係数を求める方法である．狩野 (2002, 4.2 節) でも指摘したように因子得点経由の分析は勧めたくはないが，変量内誤差モデルへ持ち込むアイデアは面白い．一方，二つ目のモデルは，実質的に，本論文 2.5 節のモデル (図 4) と同等である．すべての誤差から基準変数へパスをひくと識別できないので，椿は一つのパスをゼロと固定しているのである．椿の議論では，飽和モデル化することによって，測定モデルの不変性を達成するところに主な目的があるようであるが，筆者は，説明変数の変動を共通変動要因と個々の変数の個性とに分解し，それぞれの影響を評価するところにその価値を求めたい[16]．この考え方は，重回帰分析における多重共線性の問題にも関係する．つまり，相関の高い複数個の説明変数の背後に潜在変数を設定することに意味があれば，図 4 のモデリングによって多重共線性の問題を排除できる可能性があるからである．誤差変数からのパスを「タブー」視することはない．

[16] 古典的科学観である「デカルトの分析と綜合」に符合する．

5　適合度の分解

モデルの評価は通常，全体的評価と局所評価によって行われる (豊田, 1992, 7.2 節, 7.3 節). 全体的評価は標本共分散行列 S とモデルによって推定された共分散行列 (implied covariance matrix) $\hat{\Sigma}$ との食違いの程度にもとづく．局所評価とは各パラメータの有意性検定 (Wald 検定) である．これらに加えて，南風原 (2002, 5.3 節)・椿 (2002, 2.2 節) は，別の観点からの評価基準の利用も勧めている．南風原は，測定モデルと構造モデルを個別に評価する．この方法は一部では既に実践されているが，テキストには明記されていないと思われる．

Mulaik らによって提唱されている SEM に関する方法論に “4-Step Procedure”がある[17]. それは，SEM を以下の 4 つのステップに分けて分析を進めるというものである．

Step 1. 探索的因子分析
Step 2. 検証的因子分析: 因子間相関に構造を設定しない（飽和モデル）
Step 3. 構造方程式モデル（2. において因子間相関に構造を設定したもの）
Step 4. 精緻化

ここで Step 4 の精緻化とは，Step 3 のモデルにおいて有意でないパスを落としたり，パラメータ間に制約をおくなどの最終段階のモデル化をさす．“4-Step Procedure” には，各段階において適合度を評価し，モデルが適合しない場合は次のステップには進まないというルールがある．南風原の指摘は，Step 2 と Step 3 での適合度の評価に対応する．

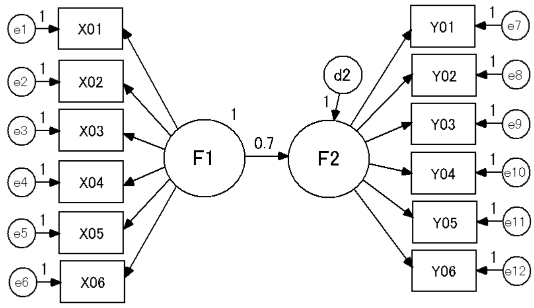
“4-Step Procedure” には，必ずしも探索的因子分析を実行できるとは限らないとか，構造のない因子相関をもつ検証的因子分析はいつも実行できるとは限らないなどの批判があり，状況によってはそれを適用できないこともある．しかし，適用できる状況もかなりあるわけで，そのようなときには分析手順の目安になる．すなわち，モデルの適合度を測定モデルの適合の良さと構造モデルの適合の良さに分解することは，より適切なモデルへ改変するための有益な情報になると考えられる．

一方，椿 (2002) は，1 因子の因子分析モデル (椿 (2002) の図 1) において，原因系変数の指標部分 $\boldsymbol{X}$ と基準変数 Y を区別したとき，Z (共通因子) を与えた下での $\boldsymbol{X}$ と Y の独立性を評価する適合度統計量の重要性を指摘している．また，その検定が，原因系変数の指標に付随する誤差から基準変数へのパスの有意性検定と同値であることも指摘している．すなわち，特殊因子の効果の吟味の必要性を訴えているわけである．

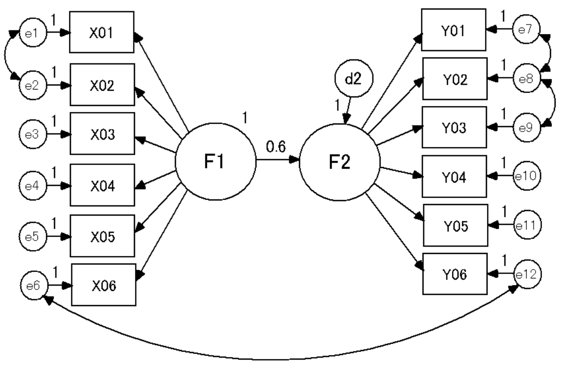
6　多くの指標変数がありモデルが適合しないときの処方箋

本節では，南風原の「適合度の相対的解釈」「少数個の下位尺度にまとめて新たに観測変数とすること」そして，南風原・椿両氏が指摘する「指標変数を間引くこと」について議論する．本節は，前節までの議論のまとめという意味合いももつ．

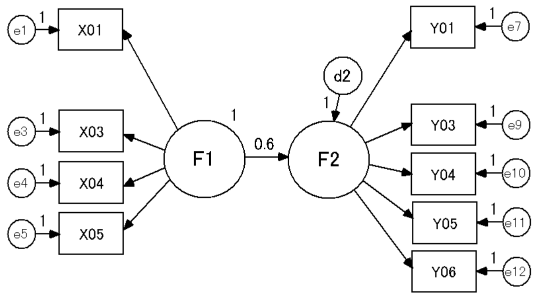
[17]詳しくは，James, Mulaik, & Brett(1982), Hayduk & Glaser(2000), Mulaik & Millsap(2000) を参照のこと．



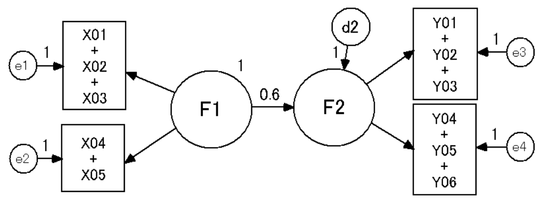
(a) 典型的な多重指標モデル：適合度は低い



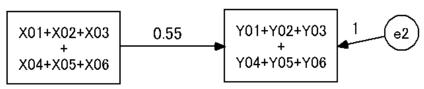
(b) 誤差共分散を導入：適合度は高い



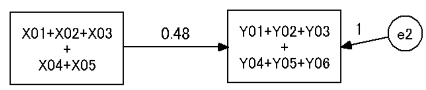
(c) 指標変数の除去：適合度は高い



(d) 下位尺度化：適合度は高い



(e) 尺度による回帰分析 (X06 を含む)



(f) 尺度による回帰分析 (X06 を含まない)

図 7: いくつかの分析方法

南風原は指標変数を間引くことについて，変数の選択に関する一般化可能性と指標として偏った項目だけが残ることによる妥当性の低下を問題にしている．村上 (2002b) は「項目全体としてカバーする個人差の範囲を拡大し，構成概念の測定の偏りを小さくするため」に十分な指標変数が必要なのだと述べている．いずれも尤もな意見である．

図 7 をみてみよう．(a) は測定モデルとして狭義[18]の因子分析モデルが採用された典型的な多重指標分析である．ただし，このモデルの適合は良くないと仮定する．このとき，(b)〜(f) の 5 つの処方箋を考える．もちろん，(a) で良いとする考えもある．(b) は (a) のモデルに誤差共分散を導入して適合度を高めたものである．(c) は変数を削除することによって適合度を高めたものである．(b) において誤差共分散が解消されるように変数が除去されていることに注目されたい．狭義の因子分析モデルに適合するように変数選択が行われた場合このような結果になる[19]．(d) では 6 つの尺度項目を二組に分割して合計得点を算出し，潜在

[18]誤差共分散を許さないという意味である．
[19]Kano (2001) 参照．SEFA や SCoFA によって変数選択するとこのような結果になる．これらのプログラム

変数の指標としたものである．ただし，X06 は除かれていることに注意する．それは e6 が e12 と関連があるからである．(e) と (f) は古典的な尺度得点による回帰分析である．なお，ここでの議論は結果指標を用い，誤差からのパスを考えていないので，各変数の個性 — 特殊因子 — の F2 への影響はないという状況を想定している．

(a) の適合度が悪いとき，筆者の結論は (b) または状況によっては (c) のモデルを採用することである．確かに適合度というのは一つの基準であって，研究を意味のあるものにするには妥当性など他にも重要な基準がある．それらも考慮する結果，モデルの適合を軽視したくなるという気持ちは理解できる．しかし，(a) において F1 から F2 へひいたパスの係数は 0.7 であるが，データに適合するモデルを採用した場合，パス係数 0.7 はかなり変化することが考えられる[20]．適合が十分でない場合，モデルの推定値の信憑性は高くないのである．それゆえ，たとえば信頼性係数を正確に見積もることができない．ただし，適合が悪いことを認めた上で (a) を報告することは，適合度の吟味を省略した (e) や (f) よりは良い．

さて，F1 と F2 の共通因子によって説明できない観測変数間の相関 — 偏相関 — を導入すれば，すなわち，適合の良いモデルに到達すれば，どのような変数を落としても (選択しても)，分析結果は大きくは変わらない[21]．この意味で，南風原のいう項目ユニバースに関する一般化可能性が保たれることになる．一方，モデル (a) ではそうはならない．未設定の誤差共分散の影響があちこちのパス係数へ反映されるからであり，南風原の指摘どおり，一般化可能性が乏しいということになる．

誤差共分散の導入については根強い抵抗感があるのは事実である．何の考慮もなく多くの誤差共分散を入れるのは勧められない．それは，データ収集前の項目の選定やワーディングの検討が不十分であったことを証明しているようなものである．また，多くの誤差共分散を入れなければいけない状況では共分散の入れ方に一義性がなく複数個のモデルが適合し得る．それを避けるためにも，事後分析 (モデルの修正) は最低限にしなければならない．誤差共分散を導入するためのもう一つの要件は，共分散を解釈できることである．なぜ，偏相関が残ったのかということが実質科学的にみて納得・了解[22]できなければならない．

さて，先に「どのような変数を落としても (選択しても)，分析結果は大きくは変わらない」と述べた．であるならば，一番小さなモデル (最少の指標変数) で分析してもよいと考えるかもしれない．しかし，それは正しくなく，いくつかの理由を挙げることができる．理論的な議論になるが，因子分析モデルには観測変数から因子得点が一意的に定められないという因子の不確定性という問題がある．そして，その問題は，観測変数の数が少ないほど顕著になる．したがって解釈という意味で，構成概念を規定するのに重要な指標変数はすべてモデルに含めて，潜在変数の意味の不定性を最小化しておかなければならない．二つ目は，一つ目と関連するのであるが，各被験者の因子得点が必要な場合，観測変数が少ないと推定 (予測) 精度が低くなるという問題である．この問題を避けるためには，信頼性が高くなるようそれなりの数の指標変数を用意しなければならない．

については，Kano & Harada(2000) を参照されたい．

[20] もちろん，因子負荷量の推定値も変化する．

[21] 母集団レベルでは完全に不変．各潜在変数に少なくとも 2 つの指標があり，そのうち 1 つは (b) において誤差共分散がないことが必要．

[22] 豊田 (1998) の表現を借りた．

より現実的な意味として，大は小を兼ねるが逆は成り立たないことを指摘しておく．小さなモデルでの分析結果のみを報告したとすれば「必要な変数をすべて利用した大きなモデルでの分析もうまくいくのであるが，何らかの理由で小さなモデルでの分析結果を報告している」とは理解されない．大きなモデルの分析に問題があったから小さなモデルでの分析結果を報告したのだと誤解されるだろう．

(c) のモデルの採否は，どこまで因子分析のモデルにこだわる必要があるかにかかっている．共通因子を与えた下で観測変数が独立に分布するという局所独立性が重要視される場合は，本モデルを採用することになろう．局所独立性がみたされていれば，観測変数を加算することで誤差が相殺しその影響を減少させることができる．一方，誤差間に共分散が存在すると加算しても誤差が相殺しない．局所独立性とそれなりの大きさの因子負荷量が保証されれば，観測変数の加算は信頼性の向上を意味するが，どちらかの仮定がくずれると，観測変数の加算の信頼性向上への効果は不透明になる．

狩野 (2002) において図 3 と表 5 で示された WISC-R の言語性尺度の分析をみてみよう．「X03: 算数」と「X06: 数唱」の誤差共分散が 0.34 と推定されている．この共分散は WISC-R の第 3 番目の因子 F3[23]の影響によるもので，「言語性因子」とは無縁のものと考えられる．表 5 によれば 6 項目からなる言語性尺度の信頼性は 0.703 であるが，もし，誤差共分散がなかったならば 0.735 となる[24]．これが局所独立性の効果である．また，この例においては，X03 と X06 の負荷量 0.30, 0.34 は「それなりの大きさの因子負荷量」という仮定をみたしておらず，これらを除いた 4 項目での信頼性は 0.766 となって 0.735 をも上回る．解釈という観点においては，X03 と X06 では，言語性因子の要素よりも F3 の要素の方が大きいことに注意しなければならない[25]．言語性尺度から X03 や X06 を除くべきかどうかは，これらの分析結果を考慮して，尺度の内容的妥当性に基づいて決定することになる．

モデルに合わない変数が同定されたとき，上記のように観測変数の再吟味が行われる．その結果として観測変数を落とすという場合がある．誤差共分散やモデルに合わない変数の同定は，指標変数の再吟味を促すという点で意義がある．

(d) のモデルは (i) 全変数を用いていること，(ii) 正しいパス係数 (0.6) が得られること，(iii) モデルがデータに適合しやすいこと，などの特徴を有した「賢い」モデルである．ここでは X06 が外されているが，もしこれを含めるならば e2 と e4 の間に共分散を入れることになる．X01+X02+X03 を指標とすることは，個々の変数の構造は問われず，合計得点のみの構造が問題になる．たとえば，e1 と e2 に相関があるということは，(a) では大問題であるが (d) では問題にならない．このように各変数に要求される仮定が弱められるがゆえに適合度が高められるのである．しかし，条件が緩いことは結論を弱めることになる．共通因子は，個々の観測変数の共通因子である保証はなく，合成変数の共通因子にしかすぎない．また，数理的には推定精度が少し落ちることがわかっている (Yuan, Bentler, & Kano, 1997; Kano, Bentler, & Mooijaart, 1993)．南風原の指摘するように分析結果が合成変数の作り方

[23]南風原 (1992) によると「転導性からの解放」「短期記憶」「数量的な能力」などと命名された因子であることは狩野 (2002) で紹介ずみである．

[24]Kano & Azuma (in press) は，誤差共分散が存在するときの尺度構成を支援するシステムを開発している．

[25]F3 から X03 と X06 への因子負荷が同じ大きさであるとするならば，因子負荷量は $0.58 (= \sqrt{0.34})$ となり，言語性因子からの負荷量である 0.30, 0.34 を大きく上回る．

に依存する可能性もある．(d) は非常に大きなモデルにおいて測定モデルの細部にこだわらないような状況に向いた「賢い」テクニックであると言えよう．誤差が調整できるという意味で，少なくとも (e) や (f) よりは良い．

(e) と (f) は伝統的な尺度得点による分析である．狩野 (2002) ではこの伝統的な方法の問題点を指摘してきた．本論文においても 2.1 節で指摘したように，2 つの尺度を構成する変数の誤差間に共分散がある場合は大きな問題になる．ここでもそうで，(b) でのように e6 と e12 との間に相関がある場合は，どちらかの変数を除去せざるを得ない．そうしたのが (f) である．ここでは両尺度の信頼性係数を 0.8 として回帰係数を計算してある．もし，変数 X06 を削除し忘れた場合は，相関が正の場合は 0.48 を過大推定 (0.55) するし，負の場合は過小推定することになる．したがって，このような相関の同定は，尺度得点を用いた分析の前にしておかなければならない．そのためには SEM が必須であろう．尺度を利用するにはその信頼性を見積もっておく必要がある．そのためには因子分析を行う．項目数が多くなると因子分析モデルは適合しない可能性が高くなるが，そのようなとき，誤差共分散を用いたモデリングが有効である．そのためにも SEM が必須である．

構成概念がたくさんあってその間の因果モデルを構築したいとする．測定モデル，そして構造モデルを作成して SEM を実行するのは理想であるが，モデルが大きくなりすぎて分析が難しいとしよう．たとえば，大きなモデルにはそれに見合ったサンプルサイズが必要になる．そのようなとき，すべての構成概念について尺度得点を求め，観測変数である尺度得点間のパス解析を行うということも考えられる．しかし，可能ならば，(d) のモデリングを行うほうが良い．そのためには，構成概念ごとに因子分析を行っておき，誤差相関の吟味を行う．この結果は指標変数を 2 つのグループに分けるときにも有用である．加えて，2 つか 3 つの潜在変数の測定モデルで分析し，(b) にあるような尺度間の誤差共分散が存在しないかどうかの吟味を行っておくのがよい．

本節の結果をまとめる．質問紙調査にもとづいて，ある構成概念に関する尺度構成を行いたいとする．因子分析の結果，適合が悪かったとしよう．適合を向上させるためただちに構成概念にとって重要な変数を落とすことは避けなければならないが，適合が悪い因子分析の結果をそのまま信用するのも良くない．因子負荷量などの推定値が不正確であり，信頼性係数も信用できないからである．そこで，適合度を勘案しながら，項目の内容を再吟味して構成概念の定義にそぐわない項目を落とすこと，そして，誤差共分散を導入することで適合の良いモデルを探索する．すなわち，図 7 で考えれば (b) と (c) を利用するのである．最終モデルにもとづいて信頼性係数などを計算しておく．

構成概念間の相関分析や回帰分析を行うときも同様である．尺度項目が確定しているときは，誤差共分散の設定を行う．特に尺度にまたがる相関に注意する必要がある．

適合の悪いモデルはいつも採用できないかというとそうでもない．十分に研究が積み重ねられた分野で，適合が悪くなる原因が明らかになっており，その原因の結果への影響が重大でないということがわかっているときは，モデル (a) を採用してもよいだろう．実務的な要請を重視して少ない観測項目で調査を行うこともあり得る．ただし，適合の悪いモデルを利用する場合は，常に，先に述べたような適合の悪さに起因する問題点を意識しておく必要

があろう.

7 多母集団同時分析の実践

鈴木 (2002) は，狩野 (2002) で指摘された「検証的因子分析が探索的因子分析を凌駕する点」を実践した論文である.

もしこの解析を伝統的な多変量解析法で分析するならば次のようになるだろう.

1. 年度ごとに探索的因子分析を計 10 回行い，3 因子の同定と目視によって 10 回の因子分析結果の同等性を確認する
2. 10 年分のデータを合算して探索的因子分析を 1 回行う
3. 因子構造が単純構造から非常に遠いので尺度化を断念し，2. の結果を用いて因子得点を計算する
4. 因子得点の予測値に基づいて，年度を横軸にとった推移図 (企業の層をつぶす) や企業ごとの推移図を描く

この伝統的な分析にはいくつか問題がある．因子構造の変化が統計的に確認できていないことはもちろんあるが，加えて，因子回転の不定性を超えてどのように因子不変性を確認するかということがある．2 つの母集団ならともかく，10 個の因子負荷行列の相等性の検討は易しくない[26]．また，2. での因子分析は，年度における平均の違いを調整していないので，データの共分散行列には平均の変化による変動が加算されているという問題がある．これは狩野 (2002, 4.3 節) で指摘されたものである．その結果，因子負荷量や因子間相関にバイアスが生じている可能性がある.

鈴木 (2002) によって分析されたデータは，経時データのように見えてそうではなく横断的なものである．それは，年度ごとの調査協力者が異なっているからである．一般には，同じようなデザインの経時データもあり得る．経時データに対しても，時点間の共分散に飽和モデルを適合させることで，多母集団の分析とほとんど同じ手続きで同時分析が可能である (狩野 他, 2000)．一方，時点間の因果の強さを評価したいときは，構造モデルとして対応する因子間の自己回帰モデルを考えればよい．なお，その場合には対応する観測変数の誤差間に共分散を設定することが多い.

8 まとめ

3 名の討論者への再討論を書き終えて感じることは，観測変数をむやみに減らすことに対する危機感と特殊因子の評価の重要性である．前者は，社会科学データの分析においても重要になりつつある適合度という新しい基準への不信感と言い換えてもよい．適合度は，因子分析で考えれば共通性 (因子寄与) が大きいとか単純構造に近いというような伝統的な基準に置き換わるものではない．従来の基準に加えて，適合度も分析時に吟味すべき重要な観

[26] 10 個の因子負荷行列が最も近くなるような回転を見つけるのは容易ではない.

点になったということである．

前節までの議論を以下にまとめる．

1. 仮定を吟味するというサイエンスの礼儀をまもるべきであろう．モデルを用いた分析を行うならばそのモデルの適合度を評価するのは当然のマナーではないか．適合が悪いと分析結果にどのような問題が生じているかよくわからない．したがってそこから引き出される結果の解釈に不安がある．
    - モデルは所詮現実を近似する「模型」であるから，現実との乖離があるのは仕方がない．しかし，モデルによる分析結果から何らかの解釈を行うためには，採用したモデルがデータにそれなりに適合するというのは必要なことであろう．解釈の雛型は，モデルの定義や性質から引き出されるものだからである．
    - モデルが適合していない場合のいくつかの問題点は 2.3 節で述べた．変数選択に関する不変性がなく，因子負荷量の推定値がバイアスをもち，信頼性係数が不正確に算出されるということであった．しかし，これら以外にも問題が生じるであろう．
    - 正確に言えば，モデルの適合が良いことは正確な解釈のための十分条件であって必要条件ではない．モデルがデータから乖離していても，分析結果やそこから引き出される解釈に誤りはないかもしれないのである．しかし，分析者は，モデルがデータから離れているとき，推定値のバイアスや分析結果の解釈に関して，どこに落とし穴があるかはよくわからないのであり，それゆえ，良い適合を要求しておけば安心感があるのである．
2. 観測変数が多い場合 1 因子の因子分析モデルが適合するのは稀である．多因子にして下位尺度を構成することが不適である場合は，因子を構成するほどではないが，観測変数間に偏相関が存在することは十分に考えられる．そして，それを説明するために誤差共分散を導入することができる．
    - その結果，変数選択に関する不変性，因子負荷量や信頼性を正しく見積もることができる．
3. 統計的にみて因子分析モデルに合わない観測変数を同定することは重要である．内容的に妥当だと考えられた変数であっても，モデル不適合のレッテルを貼ることで，項目内容の再吟味を行う機会が与えられるからである．誤差共分散を許したモデルでは，モデルに合わないと判断された変数の誤差に共分散が生じている．
4. 誤差共分散の大きさの吟味は重要である．因子で説明できる変動の方が，偏相関で説明される変動よりもずっと大きくあって欲しいからである．特に，尺度得点を用いる場合はこの吟味は重要であろう[27]．
5. 尺度得点を用いた伝統的な分析である「因子分析＋相関分析・回帰分析」を行う場合と SEM で分析する場合を比較すると，狩野 (2002) で強調した希薄化に関連する問題の他

[27] 誤差共分散を許した測定モデルを採用する場合は，因子の変動は偏相関から独立に取り出されるので，数理的には問題にならない．

に，観測変数の個性，すなわち特殊因子の扱いにおいて大きな違いがある．伝統的な分析では相関分析や回帰分析に特殊因子の影響が関わっているのに対して，結果指標を用いる SEM はそれらを排除している．

6. 特殊因子の影響にも興味があるときは，尺度得点の利用も一つの方法であるが，分析結果が項目選択に依存することや多次元性の問題がある．SEM を用いて誤差から基準変数への影響の大きさを評価することで，これらの問題を解決することができる．

永らく信じられてきた古典的テスト理論の基本モデルである $X_i = x_i + e_i$ において，e_i の独立性の崩れは，南風原の指摘するように理論的な考察はあったが，本質的な問題として顕在化してこなかったのではないか．すなわち,「探索的因子分析 (適合度の検討なし) ＋α係数＋尺度得点化」という伝統的なプロセスは，因子分析モデルからのズレとその分析結果への影響の評価をすり抜けてきたのである．ところが,「検証的因子分析＋適合度の評価」という方法論が登場したことによって，歴史をつんだこの伝統的方法論に疑問が投げかけられることになった．

心理学や行動学などの社会科学においてそのような細かい議論をしても仕方がないとか,因子分析モデルで記述できる現象は世の中に存在しないというような理由付けをして，新しい方法論を切って捨てることも可能ではある．しかし，因子分析モデルを基本に据えて，現象はモデルからどのように乖離しているのか，それは再現可能なズレなのか，そうだとすればどのように解釈やモデリングができるのか，再現不能なカオティックな現象であるならばその複雑さの程度はどのくらいか，それを除去するようなモデリングは可能か，など，一つでも現象の理解を深める方向に研究を進めたいと考える．適合が悪いというのはそのためのスタートラインである．

本稿では適合度の重要性を指摘してきた．しかし，残念ながら，どのように適合度を評価するかというより現実的な問題にはふれることができなかった．別の機会に議論したい．

参考文献

Diamantopoulos, A. & Winklhofer, H. (2001). Index construction with formative indicators: An alternative scale development. *Journal of Marketing Research*, **38**, 269-277.

南風原朝和 (2002). コメント：モデル適合度の目標適合度 — 観測変数の数を減らすことの是非を中心に —. 行動計量学，29, **-**.

Hayduk, L. A. & Glaser, D. N. (2000). Jiving the four-step, walzing around factor analysis, and other serious fun. *Structural Equation Modeling*, **7**, 1-35.

岩崎 学 (2002). 不完全データの統計解析．エコノミスト社.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data.* Sage Publications: CA.

Kano, Y. (2001). All about variable selection in factor analysis and structural equation modeling. A paper presented at the International Meeting of the Psychometric Society (IMPS2001). Osaka, Japan.

狩野 裕 (2002). 構造方程式モデリングは，因子分析，分散分析，パス解析のすべてにとって代わるのか？行動計量学，29, **-**.

Kano, Y. (2002). Multiple indicator model with explanatory specific factors. In preperation.

Kano, Y. & Azuma, Y. (2003). Use of SEM programs to precisely measure scale reliability. In Yanai, H. et al. (Eds), *New developments in psychometrics.* pp.141-148. Springer Verlag: Tokyo.

Kano, Y., Bentler, P. M., & Mooijaart, A. (1993). Additional information and precision of estimators in multivariate structural models. In *Statistical sciences and data analysis: Proceedings of the third pacific area statistical conference,* (K. Matusita, T. Hayakawa, et al., Eds.) pp. 187-196. VSP International Science Publisher: Zeist, The Netherlands.

Kano, Y. & Harada, A. (2000). Stepwise variable selection in factor analysis. *Psychometrika*, **65**, 7-22.

狩野裕，豊本満喜子，服部祥子，山田冨美雄, & 島井哲志 (2000). 対応のある共分散行列の同時分析．日本行動計量学会 第 28 回大会発表論文抄録集，pp.387-390.

Little, R. J. & Rubin, D. B. (1987). *Statistical analysis with missing data.* Wiley: New York.

Mulaik, S. A. & Millsap, R. E. (2000). Doing the four-step right. *Structural Equation Modeling*, **7**, 36-73.

村上 隆 (2002a). 質問紙法による心理尺度構成の概念と方法．日本行動計量学会春の合宿セミナー配布資料.

村上 隆 (2002b). 心理測定の立場から見た因子分析と主成分分析．日本行動計量学会 第 30 回大会発表論文抄録集.

鈴木督久 (2002). コメント：SEM による企業イメージのマネジメント．行動計量学，**29**, **-**.

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 (1998). 共分散構造分析 — 構造方程式モデリング — [入門編]. 朝倉書店.

椿 広計 (2002). コメント：「尺度化＋回帰分析」の問題点に関する注意．行動計量学，**29**, **-**.

Yuan, Ke-Hai, Bentler, P. M., & Kano, Y. (1997). On averaging variables in a confirmatory factor analysis model. *Behaviormetrika*, **24**, 71-83.