

心理測定の立場から見た因子分析と主成分分析

村上 隆（名古屋大学大学院・教育発達科学研究科）

1 何が問題なのか

筆者がここで語りたいのは、今日、社会科学の分野でも主流になろうとしているデータ分析の方法である共分散構造分析、あるいは、潜在方程式モデル（たとえば、狩野，2002）が、発達・教育・人格・社会的態度・対人関係等を扱う心理学の分野において、研究の発展を促進するというよりも、むしろ妨げているのではないかという懸念である。これに関して、著者は今まで、様々なところで現在の潮流に棹差す発言をしてきたが、かえって混乱を助長することが多かったようである。今になって考えてみると、その混乱の原因は、1）議論の対象となっているモデルの定式化が必ずしも明確でない。たとえば、今回取り上げられる因子分析と主成分分析の間の区別がはっきりしない。2）心理学者の目指しているものが統計家に理解されていない。3）心理学者自身も自分のしていることの意味、特に自分のデータの性質が本当のところはわかっていない、といったところにあったように思える。

因子分析と主成分分析というトピックは、より限定されたものであるが、これにも十分な意味があると感じる。そこには筆者が取り上げたいと思っていた問題の基本は含まれているし、これに関してなら、この時点で議論のための準備があるとも考えるからである。

2 因子分析と「因子分析」

因子分析が主成分分析とは異なるものであることは、狩野の議論によって既に十分明らかになったと思われるが、因子分析と主成分分析は、ともに、外的基準のない多変量解析の方法に属し、因子モデルにもとづいてデータの縮約的表現を目指す点は同じである。また、因子分析法として現在も使用頻度の高い主因子法を考えれば、そのアルゴリズムの中核部分は、 R 、 I 、 Ψ をそれぞれ変数間の相関行列、単位行列、共通性の推定値を要素とする対角行列とするとき、 $R - I + \Psi$ の固有値分解に帰着する。主成分分析の中核は、 R の固有値分解であり、この2つの解は多くの場合、極めてよく似たものとなる。換言すれば、因子分析と

主成分分析は、データとモデルの間の誤差に直交性という制約条件を課すか課さないかの違いであると言えるであろう。

さらに、単純構造を目指す（斜交）回転を含めて（探索的）因子分析と主成分分析は、相関関係の高低にもとづいて変数を分類するという機能において共通している。心理学者が従来必要としていたのはその機能であり、その意味で、因子分析と主成分分析をあわせて広義の因子分析と呼ぶことは許されるであろう。ここではそれを「因子分析」と表記する。

こう述べたからといって、筆者は2つの方法の差異を過小評価しているわけではない。そのことは追って明らかになるであろう。

3 心理学者にとって「因子分析」とは

最初に列挙したような分野の心理学の研究は、複数の質問項目に対する被験者の反応を合成して（多くの場合単純和、または、重み付きの和）尺度得点を作り、複数の尺度得点間の相関や、複数の明確に定義された群間の尺度得点の平均差にもとづいて進められることが多かった。尺度得点は、理論的に想定される何らかの構成概念（construct）の測定値であることが期待され、そのこと自体も検証を要する（妥当性検討， validation）。

構成概念のような抽象的なものをなぜ設定しなければいけないのかと言うことであるが、上のような尺度得点であらわされるものに適当な名前をつけなければ、それらを日常言語の中で語るができなくなるからであると考えられる。日常言語と離れてしまえば、心理学の成果を社会的に還元することが困難になる。

いずれにしても、複数の項目が何か共通した個人差を反映しているのでなければ、それらを合成して尺度得点を定義する意味がない。また、複数の尺度得点があるとすれば、それらは相互に異なるものを反映していなければ複数作る意味がない。負荷行列の単純構造への回転を含む「因子分析」を用いて、同じ「因子」に高く負荷する項目を拾っていけば、それらは相互に（相対的に）高い相関をもつという意味で、何らかの共通した個人差を反映しており、異なる「因子」に高く負荷する項

目間には、通常高い相関はない。したがって、各「因子」に高い負荷をもつ項目を集めて「因子」の数だけ尺度をつくれれば、それらは比較的高い内的整合性（internal consistency）をもち、相互には比較的独立した尺度得点となることが期待できる。

ここで筆者が述べようとしているのは、心理学研究のすべての局面ではなく、このような尺度構成の段階における項目水準のデータの「因子分析」としては、狭義の因子分析よりも主成分分析を用いた方がよいという主張である。研究が進行し、少数の絞り込まれた尺度間の関係が精査される段階では、共分散構造分析等を含む因子分析が効果をあげる可能性を否定するつもりはない。

4 「因子分析」の効用

単一の項目に対する反応（の分布）を手掛かりにものを言っていこうとする人たちは、1つの項目に対する反応の意味は単純明快であり、それらを合成した尺度得点は、複数の意味が混合されたあいまいなものになると考える傾向があるようだ。しかし、実際のところ、単一の項目に対する反応には実に多くの要因が影響しており、項目反応の意味はむしろ不明確である場合が多い。それに対して、複数の項目得点を合成した尺度得点においては、それぞれの項目に影響する雑多な要因は相殺され、多数の項目に共通する成分だけが強調される結果、かえって単純な意味をもつようになると考えられる。ともかく、それが尺度構成を行なうための、あるいは「因子分析」が意味をもつための前提である。

項目得点を「真の値 + 誤差」とモデル化する古典的テスト理論や、項目分散を「共通性 + 独自性」に分割する因子分析のモデルはこの状況を近似的に（あくまでも、近似的に）定式化しており、前述のように、心理学者の期待する結果をだいたい生み出してきた。特に、1次元の尺度しか扱えないテスト理論と比べて、項目間の相関構造に関する仮説があいまいなままでも、「自動的に」項目の分類を達成してくれる「因子分析」は、強力なツールとして重宝されてきた。

項目反応の共通性（communality）は一般に小さく、尺度得点の内的整合性（Cronbach の係数によって評価されることが多い）の値は、多数の項目を加算しなければ十分な大きさ（0.7 前後）にならないことが多い。したがって、項目数が多い割に「因子」の数が少ないというのが、こうした分析においては普通である（あった）。こういう

ケースでは、因子分析と主成分分析の結果は殆ど変わらない。前述のように、主因子法と主成分分析は、固有値分解の対象となる相関行列の対角要素が共通性（の推定値）であるか、1であるかの違いだけであるが、対角要素は相関行列の全要素のうちの $1/p$ （ p は項目数）に過ぎず、その影響は p が増加するにつれて小さくなるからである。このあたりが、因子分析と主成分分析の区別があいまいなまま今日にいたった理由の一つであろうと思われる。

確かに「因子分析」によって尺度の意味だけでなく、個々の項目反応の意味が明らかになる場合も多く、「因子分析」が心理学の研究の発展に対してした貢献は大きい。ただ、このことが心理学者を「因子分析」を始めとする統計手法に（悪く言えば）盲目的に依存させる傾向を招いたのも事実であるように思われる。

5 理論の進歩が却って不自由を招く

最新のソフトウェアを手にして、分析手続き自身は簡単になったのに、昔に比べて何か分析が不自由になったと感じている心理学者は多い。いつまでたっても分析が終わらず、一向に論文が書けない大学院生も散見される。これは一体どうしたわけなのだろうか。

前述のように、ここで取り上げているような項目水準での分析の場合、項目数は因子数に比してはるかに多い場合が大部分である。こうした場合における因子分析の当てはまりは、一般に良くない。因子分析のユーザーの中で、主因子法を用いても残差行列を検討する人や、最尤法による当てはまりの検定結果に神経を使う人にとっては、ここで述べているような分析の結果は許容しがたいものとなる。また、たとえ研究者本人は納得している結果であっても、「こんなデータは、因子分析には適さない」とか、「もっと共通性を高める工夫をせよ」といった無理難題をもちかける人もいる（狭義の因子分析にこだわる限り、むしろ正当な批判と言うべきなのであろう）。そこで、因子数を増やすという対応をすれば、1つの因子に負荷する項目の数が少なくなって内的整合性の低い尺度しか得られないし、しばしば不適解を生ずることになる。結局、項目水準の因子分析は、「許容しがたい当てはまりの悪さ」と「不適解」の2つの困難の間で立ち往生することになることがしばしばである。

6 なぜ多くの項目が必要なのか？

そこで浮上するアイデアは、項目数を減らすことであり（更に必要なら）確認的な方法に訴えることである。これによって許容し得る当てはまりと、合理的な（美しい）解が得られる場合もある。しかし、筆者はそれをよい解決法とは思わない。

なぜ、少数の尺度を構成するために多数の項目が必要なのであるのか？ その理由は2つ考えられる。

1) 多数の項目を加算することにより、内的整合性を向上させるため。

2) 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。

心理測定の用語で言えば、1) が信頼性、2) が妥当性に関わる部分である。1) は、個々の項目反応の不安定性を相殺して、共通部分を際立たせるにはある程度の項目数が必要であるということである。2) は、性格や態度に関する質問ではやや説明しにくいので、アチーブメントテストを例にとつて言うと、たとえば半期の授業の評価をするためには、全体を満遍なくカバーする項目（数）だということである。そして、このところで、因子分析のモデルと心理測定との間のある種の間隙が見えてくることになる。

1つの項目には1つの因子だけ関わっているという場合があるでしょう（筆者には現実にもそのような事が起こるとは考えられないが、数理的にそういうモデルが成立している状況を想像することはできる）。もし、通常の因子分析の基本モデルが正確に成り立つとすれば、すべての項目反応はランダム誤差を除けば、すべて単一の因子だけを反映しているのであり、意味的には相互に redundant なものである。そこには、2) で言うような広がりや考慮する必要はないし、誤差には項目間で重複する要素は全くない。そこでもし、個人の尺度得点を正確に求める必要がなく、因子間の相関や、因子得点の群間の平均差を求めればよいという条件なら、項目数は2つ以上あれば幾つでも同じことである。実際、尺度得点間の相関係数を直接計算するとすれば、いわゆる希薄化（attenuation）により相関係数は低下するから、相関係数を用いたそれ以上の分析には注意が必要になる。しかし、因子モデルによって、因子得点間の相関係数の希薄化は修正されているはずだから、項目の数を増やす必要はない。項目が多くなければならない理由が1) だけならそれでもよいはずである。

だが、2) を考えるとそうはいかなくなる。たと

えば、15コマの授業において、10コマ目の講義で教えられた内容の習得に関しては、他のコマとは異なる達成度の個人差があると考えるのが普通であろう。それは、最初の数コマの達成度とは余り相関がないが、9回目や11回目の達成度とはかなり相関があると考えるのが自然であろう。もしそうなら、この達成度に関する 15×15 の行列はいわゆる simplex の形態をもつことになり、 Ψ のような対角行列ではなく、厳密に言えば少数の因子で説明することもできないことになる。

いずれにしても、項目を減らして得られた因子間相関は、数値上は合理的な範囲にあるとしても、意図した構成概念の間の関係の指標としては、著しく偏ったものである可能性がある。

7 因子モデルと尺度構成は別のこと

恐らく、どのような構成概念もある程度の幅を想定せざるを得ない。したがって、その測定のための項目間には、単なる誤差とは言えず、かつ簡単にはモデル化できない共変関係が存在していると考えられる。このあたりの事情を筆者は、Wittgenstein の用語を借りて、「項目間には家族的類似（family resemblances）の関係が存在する」と表現したことがあるが、最近、全く独立に同じ主張をする人たちを発見して意を強くした（Saucier & Goldberg, 1996）。

単純構造に近づけられた項目の分類は（現実世界に関するわれわれの認識がそうであるように）基本的には事態を「考えやすくする」ために導入されるものであって、動かしがたい客観的事実であるわけではない。また、項目の変動は、アプリアリに共通成分と誤差に分かれて存在しているわけではない。それらは、一緒に分析される他の項目次第で変わり得る。もちろん、実際の現象が、因子モデルに従っておこる保証はないし、変数をモデルに合うように選択したり、特定の誤差の間に相関を仮定してモデルを修正したるするのは、現実を捻じ曲げるものである。つまり、ランダム誤差というものの性質から考えれば極めて自然に見える因子分析のモデルの誤差の構造に関する仮定は、心理学的データにおいては成り立ちがたいと思われる。

したがって（狭義の）因子分析の m 次元モデルがあてはまるということと、項目から m 次元の尺度が得られる（測定可能性）ということには、微妙だが決定的な違いがあると考えられる。モデルのあてはまりは悪くても、一定の内的整合性をもて

ば、測定尺度としては成立しうることを認めるべきである。

いずれにせよ、心理学的データ、特に質問紙データの性質を考えて、少なくとも研究の初期の段階においては、因子分析より主成分分析が用いられるべきであると考え。これは単に不適解を避けるといった消極的な理由からではなく、通常、研究者が自信をもって述べることのできる仮説のない場合について、仮説を生成するための機能、あるいは、構成概念を創出していくための機能を（回転を伴う）主成分分析はもっていると考えるからである。

ただ、通常、質問項目はデタラメに集められているのではなく、何か1つの上位概念にカバーされると想定されるような範囲（自己意識、親和動機、自己不全感）で収集されるから「因子分析」の結果として、それらがいくつかの下位概念（公的自己意識、私的自己意識）等からなるという、一種の階層構造をなすことが発見されることはよくある。この点で、回転は直交回転より斜交回転が望ましいであろう。

もう1つ、実際に実施された項目に対して最適なモデルを追求するという考え方自体も、余り適当ではない。2～3項目の、ある程度の相互相関をもつ項目群が発見されたときには、それと内容的に近い項目を追加することにより、あらかじめ予期していなかった下位尺度を（次の調査において）得ることができる可能性もあるからである。

このような、広い意味における発見的機能を「因子分析」には認めるべきであるし、研究の結果も常にある程度は暫定的なものと考えられるべきである。ともかく、最近奨励されているような不自由きわまるやり方をしていたのでは、Big-Five Theoryのような壮大な理論は決して発見されなかったであろう。もっとも、「先験的に」パーソナリティは5次元であると決め付けて、それにあわせて項目を刈り込んでいくことはできたであろうが、それは全く別問題である。

8 教科書の記述を鵝呑みにしないで

因子分析の機能（の一部）を主成分分析で肩代わりさせようとすることに反対する人たちは、しばしば主成分分析を狭く捉えすぎているように思われる。たとえば、主成分は回転してはいけない、特に、斜交回転をしてはいけないという考え方がある。本当だろうか。

主成分分析の一つの定義の仕方として、 Z を列

ごとに中心化、標準化した（つまり、変数を標準得点に変換した） $n \times p$ のデータ行列、 V を $p \times r$ の重み行列とし、 $V'V = I$ とする制約条件の下で、合成変量 $Y = ZV$ の分散の和、 $\text{tr } n^{-1}Y'Y = \text{tr } V'RV$ を最大化するというものである。ここで、 $R = n^{-1}Z'Z$ は変数間の相関行列、 n は被験者の数、 p は変数の数、 r は主成分の数で、常識的に $n > p > r$ である。

このとき、 V には直交回転に関する不定性がある。すなわち、 T を任意の $r \times r$ の正規直交行列とし、 $V^* = VT$ を考えると $V^*V^* = I$ 、 $\text{tr } V'RV = \text{tr } V^*RV^*$ である。この場合、回転された $YT = ZV^*$ の列間には相関が生ずるので斜交回転になる。このように、主成分分析の本来の定式化の中に、斜交回転の可能性は含まれている。

一方で、しばしば誇らしく語られる因子分析の優れた特徴と言われるものも、多くの場合、非現実的な誤差の構造に関する仮定から導かれたものであり、現実のデータについて見るとあまり価値があると思えないものが多い。たとえば、因子不変性に関する性質というのがある。主成分は項目を入れ替えると変化するが、因子は項目の入れ替えに関して不変であるという（無条件で成り立つなら）因子分析の絶対的優位を保障するような性質であるが、よく聞いてみれば、「因子モデルがあてはまる項目の範囲内」という条件の中での話に過ぎない。

ともかく、調査票を片手に負荷行列と主成分間相関行列を見ながら、1つ1つの項目と尺度の意味が明らかになっていく「ワクワク感」は、自分のモデルが検証されているかどうかを祈るような気持ちで見つめる「ドキドキ感」に、研究の促進という観点から見ると勝っていると筆者は信じている。

文 献

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かわるのか？ 行動計量学（掲載予定）

Saucier, G. & Goldberg, L.R. (1996) The language of personality: Lexical perspectives on the five-factor model. In J. S. Wiggins, (Eds.) *The five factor model of personality: Theoretical perspective*. Guilford.