教育統計における諸問題 III

——因子分析法の応用——

青 山 博 次 郎

(1961年2月受付)

On Some Statistical Problems in Education III

——Applications of Factor Analysis——

Hirojiro AOYAMA

In this paper we disccussed certain problems introduced in the course of application of factor analysis. Whenever we are intend to study the effect of certain teaching methods, we used to apply these methods in some classrooms separately and to analyse the test scores obtained after some training period.

First we discussed the determination of the number of common factors, comparing Thurstone's multiple factor analysis with the principal component analysis.

If we would like to decide what ability is concerned with the transfer of training, we must make use of factor analysis. However, we are in this case confronted with the difficulty of small size of sample. Usually we handle this problem by comparing the variation of factor loadings (c. f. Fleishman), but we treated it by comparison of the variation of mean factor scores because of the inadequacy of small size of sample. This procedure is carried out by the least square method.

In a similar consideration we obtained the factor loadings of certain tests which were ommitted previously from the initial factor analysis because of the lower correlation with the other tests included in previous analysis or of the trouble in computation.

Institute of Statistical Mathematics

1. はしがき

本稿においては因子分析法の実際問題への適用に際して生じる二，三の問題について考察する．

実際問題に対してどのような分析法を用いるかは，それぞれの学者の立場によって異る．心理学者は Thurstone の多因子分析を好み，あるいは Holzinger や Harman の二因子分析，Lawley の統計的方法，Whittle の最小自乗法によるものなどを推し．統計学者は Hotelling による主成分解析法をとるという具合である．これらは何れもそれぞれ長短があり，実際問題への適用においても，ややともすれば単なるモデルに終って了い，実際抽出された因子は何が何だか分らぬということも多い．ここではそれらの比較検討が目的ではないので，Thurstone の多因子法を用いてゆくと

きに生じる問題についてのみ考察することにする.

2. 問題の設定

教育効果の測定の問題に関して，幾つかの学級をえらび，それぞれ別箇の教授法を行い，若干期間を経てテストを行って何れの教授法が最も効果があったかを研究することがある．この際等質と考えられる学級をえらぶことは勿論であるが，実験開始前と，終了後のデータには相関があることに注意しなければならない．しかしこのことは当面の問題ではないのでこれ以上深入りはしない.

また教師の質的の差を考察する必要のないとき，例えば同じ教師が教えるとして教授法による効果の差を比較し，更にどのような能力が訓練によって伸びたかを研究したいとしよう．例えばある学科の教授法として A 法，B 法，C 法 (control) の3方法を考え，これらをそれぞれ A 組，B 組，C 組に適用し，それぞれがどのような効果をもったか，効果があったとすればどのような能力を伸ばすことができたためかを研究しようとする．この際各学級の生徒数は約 50 人位であるのが通常であろう.

問題の核心は特別の教授法によってどのような心的能力を伸張させたかが分れば，その学科の狙っている教育効果を最もあげるようにするには，どの心的能力を伸すのがよいか，それにはどの方法がよいかというような端緒がつかまれるのである.

このように問題を設定すれは，当然因子分析法を適用しようということになってくる．しかしながら因子分析法では相関係数を利用するために，サンプル数が大きくなくては誤差の影響が大きい．このため共通因子の抽出がいつ終了したかを判定することが困難になってくる．実験開始前では A, B, C 3 学級の生徒数は合わせて 150 人位であるから，まず何とか分析にかかる．しかし実験終了後は各学級毎に因子分析法を行い，共通因子の因子負荷量の変化をみて心的能力の変化をしらべるわけには行かない．サンプル数が不足するのである．ここに問題が生じる.

3. 共通因子の抽出

ここでとり上げたのは20問から成るテストで，これについて Thurstone の重心法と，Hotelling の主成分法* の両方を適用してみた．そうして両者の比較をしながら因子負荷量を評価しようとした.

文献 [1] によれば米国では IBM-704 電子計算機を用いて，24 問の心理テストの固有根及び固有ベクトルを求めるのに約8分を要したという．われわれの用いたのは FACOM-128 継電器式計算機で，電子計算機に比べると性能が劣り記憶装置が少いので固有根及び固有ベクトルを求めるのに多大の時間を要した．最大根は Power method で求めるが，大抵の場合に収斂が速いのでよいが，その他の根は逐次求めていくので途中で acceleration 法 (Aitken の δ^2 法) を用いても相当の時間を要する．因みにその様子をかかげると次のようになっている.

根の値	逐次近似の回数	所要時間（分）	備考
$\lambda_1=7.757473$	7	43	
$\lambda_2=1.642871$	38	212	
$\lambda_3=1.404958$	46	282	
$\lambda_4=1.072928$	150	675	
$\lambda_5=1.0495$*	35*	140*	* 印は途中までのものである

各固有値の値を相当多くの有効数字まで計算しておかないと，次の根を精確に求められないので

* 主因子法はこれを若干改変したものだが，モデルとしては特殊因子を考察するか否かという点で根本的な相違がある.

このように長時間を要したのである.

以上の主成分法の代りに重心法を用いるときは毎回ほぼ1時間を要した. 従って多くの有効数字を要する時には, われわれの現在もっている FACOM-128 を用いる限りでは重心法を用いる方が適当ということになる. 勿論記憶装置が大きくなれば, 同時に固有根を求め得るので主成分法が優位に立つであろう. なお重心法は主因子解, すなわち共通性を推定し, 相関行列の対角線にこれを入れた相関行列について主成分法を適用した時の解の近似計算に当っている訳だから, 計算機の速さが増すと共にこのような解法も容易に行えるようになるであろう.

以上のような理由から, ここでは重心法によって第6因子まで抽出操作を行った. 主成分法と, 重心法による解の一部を比較してみると次表のようになっている.

テスト	主成分法				重心法			
	a_1	a_2	a_3	a_4	a_1	a_2	a_3	a_4
1	0.84	0.24	−0.04	0.06	0.81	0.27	−0.07	−0.08
2	0.86	0.19	0.11	−0.05	0.84	0.26	0.09	−0.06
3	0.74	0.40	−0.08	−0.00	0.71	0.37	−0.18	−0.12
4	0.77	0.29	0.16	−0.02	0.74	0.28	0.11	−0.20
5	0.82	−0.00	−0.03	−0.07	0.80	0.09	−0.04	0.16
6	0.81	−0.14	0.15	−0.02	0.81	−0.12	0.17	−0.10
7	0.68	0.20	−0.02	0.16	0.67	0.27	0.05	0.34
8	0.78	−0.07	−0.30	−0.03	0.77	−0.08	−0.28	0.13
9	0.66	−0.40	0.21	−0.08	0.66	−0.36	0.28	0.10
10	0.59	0.32	−0.05	0.01	0.55	0.31	−0.13	−0.22
11	0.53	−0.20	0.17	−0.34	0.51	−0.10	0.12	0.20
12	0.58	−0.21	0.19	−0.15	0.56	−0.16	0.07	−0.15
13	0.49	−0.11	−0.20	−0.49	0.47	−0.13	−0.20	0.07
14	0.52	−0.09	−0.10	0.12	0.50	−0.07	−0.12	−0.23
15	0.44	−0.26	−0.69	0.13	0.43	−0.23	−0.48	0.23
16	0.34	0.51	0.17	0.16	0.31	0.35	0.18	0.18
17	0.43	−0.45	−0.43	0.02	0.42	−0.33	−0.25	0.06
18	0.35	−0.36	0.12	0.64	0.35	−0.19	0.13	−0.10
19	0.26	−0.43	0.58	−0.13	0.26	−0.25	0.38	−0.03
20	0.43	−0.18	0.22	0.41	0.41	−0.13	0.18	−0.17
$\sum_i a_{ij}^2$	7.76	1.64	1.41	1.07	7.36	1.14	0.85	0.55

この中でどの因子までを共通因子と考えればよいかについては, 判定法が種々あって, どれも満足すべきものではない. Thurstone によれば k 個の共通因子を抽出するに必要なテストの数 n は

$$n \geqq \{(2k+1)+\sqrt{8k+1}\}/2$$

を満足すればよいが, 少くともこの 2, 3 倍のテストは要するといわれている. 従って4つの共通因子を抽出しようとすると, n は8以上となるが, 16〜24 のテストは実際上は必要になる.

Saunders の判定法では第 k 因子残差を $r_{k.ij}$, 第 i テストの共通性を h_i^2, 被験者数を N とするとき, 始めて

$$\sum_{i,j} r_{k\cdot ij}^2 < \left(\frac{n-k}{n}\right)^2 (n-\sum_i h_i^2)^2 \Big/ N$$

が成立つ k の値が共通因子の個数であるといわれている [2]. われわれの場合 $|r_{6\cdot ij}|<0.1$ であるが, $k=6$ でもまだ上の不等式は成立しない.

また Kaiser によると,「共通因子の数は相関行列の1より大きい固有値の個数に等しい」とい

う経験法則が示されている．これは統計的な有意性，代数的な必要条件，心理測定の信頼度，心理学的な意味などを考慮して作られたといわれている [1]．われわれの場合固有根は λ_5 までしか求めてないが，少くとも5つは共通因子があるということになる．

統計的な有意性の検定は，Lawley や，Bartlett によって近似的な方法が示されているが，これを適用すると殆んど Thurstone のあげた必要条件の式の場合と同じく，あまりにも共通因子の数が多くなって了う．

例えば Bartlett の検定 [3] では

$$\chi^2=-\left\{N-1-\frac{1}{6}(2n+5)-\frac{2}{3}k\right\}\log\left\{\prod_{j=k+1}^{n}\lambda_j\Big/\left(\frac{1}{n-k}\sum_{j=k+1}^{n}\lambda_j\right)^{n-k}\right\}$$

N は生徒数，n はテスト数，k は共通因子の数，λ_j は固有根

は近似的に自由度 $\frac{1}{2}(n-k)(n-k-1)$ のカイ自乗分布に従うことを用いる．われわれの例では $k=5$ としても $\chi^2=263.4$ となって著しく有意である．これは $\sum_{j=1}^{5}\lambda_j=12.93$ で $n=20$ の約 65% にしか達しないことからも当然であって，相当多数の共通因子が抽出されない限り，このような検定を用いること自体が無理なのである．Kendall はこの点について，この検定は共通因子の数を検定しているものではないとのべている [4]．

そこで共通因子の数 k を5として，Thurstone の Simple structure にもって行くと，因子負荷量が小さく，因子の解釈が困難となる．このようなわけで $k=4$ として解釈することにした．

以上のように共通因子の数についての決定は，統計学的にのみ行うことができ難いところに種々の曖昧さが入りこんでくる．そうして主成分法が唯一つの解を与えるという意味で Kendall などの統計学者は心理学者に反対するのである [3]．

4.　能力の伸びの推定

因子分析法の応用として，訓練によってどのように能力が変動していくかについては，Fleishman の精神運動性 (psychomotor) の例がある [5]．この場合は被験者は測定の前後で同一であるから，訓練の後で再び因子分析を行い，因子負荷量の変化によって訓練効果を見ようとした．われわれの学級における教授法の例では，実験終了後の被験者の数は，特定の教授法を用いた学級ばかりとなるから減少して了う．従って相関係数のもつ誤差が大きくなることと，更にこのように別々に因子分析を行うからには，実験開始前のテストについての因子分析と対応させなければならないから，両回とも 50 人たらずの人数で分析しなければならなくなる．

そこで考え方を変えて，因子負荷量はテストのもつ各能力に対する重みづけであるから，実験開始前でも，終了後でも同一テストを用いる限り不変であると考える．そうして変化したのは因子得点であるとする．このように考えれば，実験開始前の因子得点の平均点が，特別の教授法によってどのような平均点に変ったかを統計的に検定していけばよいことになる．

いま第 i テストの標準得点を x_i, k 個の共通因子の因子負荷量を $a_{ij}(j=1,2,\cdots,k)$，第 j 因子の因子得点を z_j とおき，第 i テストの独自因子負荷量* を b_i，独自因子得点を u_i とする．このとき

$$x_i=a_{i1}z_1+a_{i2}z_2+\cdots+a_{ik}z_k+b_iu_i \tag{1}$$

というモデルにおいて．x_i, z_j, u_i がすべて平均 0，分散1のガウス分布に従うとし，パラメーター a_{ij} を求めようというのが Thurstone の多因子法の考え方である．

実験開始前のものに添字 (1)，終了後のものに添字 (2) を付すと，テストの実際の得点を y_i，平均を $\bar{y}_i$，標準偏差を s_i として

* 特殊因子負荷量 c_i，誤差因子負荷量 d_i なるとき $b_i^2=c_i^2+d_i^2$ を満足する b_i をいう．

$$x_i^{(1)}=\frac{y_i^{(1)}-\bar{y}_i^{(1)}}{s_i^{(1)}}=\sum_{j=1}^{k}a_{ij}z_j^{(1)}+b_iu_i^{(1)} \tag{2}$$

が成立つ．$\bar{y}_i^{(1)}$ に対応する終了後の値を $\bar{y}_i^{(2)}$ とおくと，$\bar{x}_i^{(1)}$（これは 0）からの伸び t_i として

$$t_i=\frac{\bar{y}_i^{(2)}-\bar{y}_i^{(1)}}{s_i^{(1)}}=\sum_{j=1}^{k}a_{ij}\,\bar{z}_j^{(2)}+b_i\bar{u}_i^{(2)} \tag{3}$$

が得られる．

われわれが実際に適用した場合では 20 のテストの中で，共通性が大きい値をもつ 9 つのテストについてだけ (3) 式を考慮した．その理由は，共通性 h_i^2 が大きくないと，a_{ij} も大きい値をもつものが少く，$\bar{u}_i^{(2)}$ の影響が利いて $\bar{z}_j^{(2)}$ の推定の精度が悪くなる心配があるからである．更にもう一つの理由は $\bar{z}_j^{(2)}$ 及び $\bar{u}_i^{(2)}$ を求める計算が厄介になるということもある．すなわち全テストを用いるときは，20+4=24 の未知数を求めなければならぬが，上のようにすると 9+4=13 の未知数を求めればよいことになる．

さてこのようにして l 個（ここでは 9 個）のテストについて，各学級毎に実験開始前と終了後との平均の差から t_i を求め，先に求めた a_{ij} と，$b_i^2=1-h_i^2$ を用いて $\bar{z}_j^{(2)},\bar{u}_i^{(2)}$ を (3) から求めればよい．このためには最小自乗法を用いて

$$\left.\begin{aligned}&\sum_{j=1}^{k}[\boldsymbol{a}_i\boldsymbol{a}_j]\bar{z}_j^{(2)}+\sum_{j=1}^{l}[\boldsymbol{a}_i\boldsymbol{b}_j]\bar{u}_j^{(2)}=[\boldsymbol{a}_i\boldsymbol{t}] \qquad i=1,2,\cdots,k\\ &\sum_{1=j}^{k}[\boldsymbol{b}_i\boldsymbol{a}_j]\bar{z}_j^{(2)}+b_i^2\bar{u}_i^{(2)}=[\boldsymbol{b}_i\boldsymbol{t}] \qquad i=1,2,\cdots,l\end{aligned}\right\} \tag{4}$$

を解けばよい．ここで

$$[\boldsymbol{a}_i\boldsymbol{a}_j]=\sum_{m=1}^{l}a_{mi}a_{mj},\qquad [\boldsymbol{a}_i\boldsymbol{b}_j]=a_{ji}b_j,$$

$$[\boldsymbol{a}_i\boldsymbol{t}]=\sum_{m=1}^{l}a_{mi}t_m,\qquad [\boldsymbol{b}_i\boldsymbol{t}]=b_it_i$$

である．

(4) から $\bar{z}_j^{(2)}$ が求められると，因子得点の伸びがあったかどうかは，平均の差の検定をすればよいから，生徒数 N を用いて

$$\zeta=\bar{z}_j^{(2)}\sqrt{N}$$

が近似的に標準ガウス分布に従うとして検定することができる．

われわれの行った例では 20 コのテストの中で 9 つだけ用い，4 つの因子得点の平均 $\bar{z}_j^{(2)}$ 及び 9 つの独自因子得点の平均 $\bar{u}_i^{(2)}$ を求めると次表のようになった．ここでテスト 5, 6, 7, 9 はB法で特に考慮して教育した項目に関係したテストで，第 3 因子を最も重視したのであったが，A組でも，B組でも共に第 3 因子は有意に伸び，C組では逆に有意に減少している．このようなことから，A, B 法の何れを用いても共通の心的能力（第 3 因子）が伸びるので，A, B 法の何れによっても訓練の転移が行われたとみることができるのである．ここで独自因子負荷量は，テストの数を増せば更に他の共通因子負荷量を含むかも知れないし，また特殊因子，誤差因子の因子負荷量だけで

教授法 / 平均因子得点	A 法	B 法	C 法
$\bar{z}_1$	−0.08	1.08	−0.70
$\bar{z}_2$	0.39	0.45	2.78
$\bar{z}_3$	0.92	0.44	−2.14
$\bar{z}_4$	0.03	0.67	−0.05
$\bar{u}_1$	−0.82	−1.49	−2.22
$\bar{u}_2$	1.27	−0.24	−1.72
$\bar{u}_3$	0.40	0.54	−0.65
$\bar{u}_4$	1.45	−0.13	1.45
$\bar{u}_5$ *	−1.04	−0.97	−0.17
$\bar{u}_6$ *	1.06	−0.17	0.24
$\bar{u}_7$ *	1.37	−0.29	−1.54
$\bar{u}_8$	0.62	0.13	1.72
$\bar{u}_9$ *	0.30	−0.48	1.90

注　添字（2）はすべて省略した．
* 印はB法にて訓練した項目に関するテスト

あるかも知れない．従ってこれらの独自因子得点の平均点の解釈は明瞭につけることはできない．

5. 除外したテストの因子負荷量の推定

テストの数が増す程因子分析の手数は急激に増大する．前述の例では最初 27 コあったテストの中から 20 コをとって因子分析を行った．この基準として相関係数の低いものや，明らかに他の心的能力を測定しているものを除外した．しかながら，もしこれらを加えて因子負荷量を求めたいときは，第 i, j テストの得点の相関係数が

$$r_{ij}=\sum_{l=1}^{k} a_{il}a_{jl} \tag{5}$$

なる関係式を満足することから，i を因子分析に採用したテスト，j を採用しなかったテストとして a_{jl} を求めればよい．このためには再び最小自乗法を用いて，各テスト j ごとに

$$\sum_{l=1}^{k}[\boldsymbol{a}_i\boldsymbol{a}_l]a_{jl}=[\boldsymbol{a}_i\boldsymbol{r}_j], \qquad i=1,2,\cdots k \tag{6}$$

を解いて a_{jl} ($l=1,2,\cdots k$) 求めればよい．ここで

$$[\boldsymbol{a}_i\boldsymbol{a}_l]=\sum_{m=1}^{n'} a_{mi}a_{ml}, \qquad (n' \text{ は因子分析に採用したテストの数})$$

$$[\boldsymbol{a}_i\boldsymbol{r}_j]=\sum_{m=1}^{n'} a_{mi}r_{mj}$$

とおいてある．

われわれの例では $n'=9$ として，前述の心的能力の伸びを検定したものと同じテストを採用すると，次表のような結果が得られた．具体的に因子の性質に当ってみても，ほぼ妥当な解釈を与えることができた．

因子負荷量 / テスト	a_1	a_2	a_3	a_4
21	0.28	0.46	0.59	0.12
22	0.06	0.27	0.06	0.22
23	0.17	0.23	0.21	0.14
24	0.08	0.45	0.30	−0.06
25	0.12	0.19	0.14	0.04
26	0.12	0.30	0.32	−0.17

注 テスト 27 はテスト 1～9 と負の相関係数をもつもので両極因子が入るので省略

6. あとがき

本稿で取扱った具体的な問題は，お茶の水女子大付属高校の長命俊子氏の「外国語としての英語学習能力の心理学的研究」(未発表) によるものであり，因子分析に関する種々の計算は能城昌子さんが行ったものである．FACOM-128 のプログラミング，計算には駒沢 勉，三枝八重子，勝山ヨシ子，丸山愛子の諸氏の援助も受けた．ここに謝意を表するものである．

なお固有値の計算は未完であるので，暇をみて継続して行きたいと考えている．

統計数理研究所

参考文献

[1] Harman, H. H.: Modern Factor Analysis, Univ. of Chicags, 1960.

[2] Cattell, R. B: Factor Analysis, New York, 1952.

[3] Kendall, M. G: A Course in Multivariate Analysis, Griffin, 1957.

[4] ―――――: Review of Uppsala Symposium on Psychological Factor Analysis, J. Roy. Stat. Soc., Ser. A. Vol. 107, 1954.

[5] Fleishman, E. H. and Hempel, W. E.: Changes in factor structure of a complex psychomotor test as a function of practice, Psychometrika, vol. 19, 1954.