

因素分析简介(续)

陈 龙

韩 淼

中国科学院心理研究所

北京市计算中心

三、因素分析的基本步骤

因素分析的基本步骤包括如下三个组成部分:

- (1) 收集资料, 并按因素分析目的, 将原始数据形成相应的关联性测度矩阵的形式, 预备进一步分析。
- (2) 抽取初始因素: 即将相差矩阵或协方差矩阵进行处置, 求得初始因素解。
- (3) 因素轴的旋转与解释: 将初始因素解进行因素轴旋转求得最终因素解, 并予解释。

考虑到因素分析的目的, 其中每一基本步骤都可分作两种类型进行分析。本文将这些类型的差异与主要解法分述之。

1. 形成关联性测度矩阵

因素分析可分为Q型和R型分析两种。前者是对样方进行因素分析, 后者是对变量进行因素分析。一般因素分析都是对变量进行分析, 对样方进行Q型因素分析的方法和原理与R型分析类同, 但在文献中并不多见。

因素分析的第一步是首先确定要进行分析的对象, 收集表征样方特征的诸变量观测值, 并将这些原始数据按要求排列成规定的形式, 进而计算变量间或样方间的关联程度。一般在R型分析中用变量间相关系数, Q型分析中用其它关联性测度, 最终将表征分析对象特征的关联性测度形成矩阵, 以便进一步分析。

在R型分析中, 主要是形成变量间相关矩阵或协方差矩阵。

2. 抽取初始因素

抽取初始因素的目的主要是探索用一组与原观测变量有关联的新变量集合进行数据简缩的可能性。有两种可能性, 一是对原则测变量的数学转换方法, 另一种是建构一种对原观测变量及其变异来源结构的推论性假设。这两种分析的结果都设初始因素间是彼此独立的, 则在可能的数据缩减过程中并不特别强调初始因素的各维度之意义。

根据抽取初始因素的不同准则和标准, 有如下几种主要方法可供选择。

(1) 主成分分析法 (principal component analysis)

把一个相关矩阵 R 分解为一组互相正交的主轴或主成分的方法称为主成分分析。由主成分分析法得到的主轴或主成分应对应于相关矩阵 R 的特征值和相应的特征向量。特征值按其数目由大到小依次提取, 而特征向量相互正交。第一个主成分说明变量总变异量的最

大部分,第二主成分说明剩余变异量中的最大部分,如此继续下去,每一个新的主成分说明变量原始总变异量中的愈来愈小的部分,直至最后一个主成分基本说明全部总变异量为止。用这种方法可以使所有主成分能再现变量之间的相关关系。在因素分析中,若用前 m ($m > n$, $m = 1, 2, \dots, m$)个主成分可以说明原始变量 n 的总变异量中的绝大部分时,则可将 m 个主成分作为初始因素。

在主成分分析中,相关系数矩阵的主对角线元素皆为1。若其它条件不变,而以对“共通性”的某种估计量代替1,置于矩阵中主对角线位置,其用主成分分析的方法所得结果则称之为主因素解(principal factors solution)。

在现成的计算机程序中,对主成分分析可分迭代和非迭代法两种。应用前者,可以由计算机自动加以迭代计算,直至求出最佳解。所以,在近来的因素分析应用中受到重视。

可以看出主因素法已不是一种单纯的数学转换方法,它已涉及到如何选择公因素“共通性”使得约相关矩阵 $\hat{R}$ 的秩恰好等于公因子空间的维数的问题了。这在数学上即是成为解约相关矩阵的特征方程问题,有关这方面的方法还有:

(2) 最大似然法(maximum-likelihood method)

D.N.Lawley于四十年代把A.Fisher的“最大似然法”应用于因素分析之中,目的在于用统计的方法在公因素“共通性”和因素个数之间找出一种最佳组合,使因素矩阵在合理的显著性水平下再现原观测变量的相关矩阵(令其剩余为最小)。但这种方法实用于因素分析中是在五十年代由电子计算机广泛应用之后的事情,由于最大似然法在求解时是对公因素空间单独处理的,因此,经旋转后解法不尽相同,所以要求在计算时应首先给出一些限定条件,结果才更有意义。

最大似然法的一个变式是典范因素分析(canonical factor analysis)

此法是C.R.Rao于1955年用H.Hotelling 1936年提出的典范相关分析方法来做因素分析时提出的。主要是要求出假设的因素变量集合与观测变量间的典范相关系数,这和主成分分析的目的正好相反,主成分分析主要是要求每一因素应当解释观测变量的尽可能多的变异,而本方法则是把因素作为一个集合来考虑。

(3) 最小二乘法(least squares method)

这种方法试图用序贯的方法确定一个最小的 K 值,使因素模型能更好地拟合原始数据,通过逐步增加 K 值,直至达到某个 K 值能使模型充分拟合数据为止。

通常做法为($K = 1, 2, \dots, K$):

- ①假设存在的 K 因素可以拟合观测变量之相关;
- ②得出某种对“共通性”的初始评估;
- ③用最小二乘法抽取能最好拟合观测变量相关矩阵的 K 值。得出因素矩阵;
- ④为更好由因素矩阵再现相关矩阵,对“共通性”进行进一步调整;
- ⑤重复①——④,直至 K 为最满意解。

这种方法的一个变式是由哈曼(H.H.Harman)提出的最小残差法(minimum residuals method)。

最小残差法的原则是瑟斯顿于1947年提出的:“因素分析的目的是利用少量的导出变量,构造一些统计检验或它们的交叉相关,这里所谓少量,是指与可接受剩余误差相一

致。”即使残差相关极小化的求解方法。但是这种算法直到1965年才由哈曼等人在计算机上得以实现。这里的残差是指在求得因素解之后，——将两变量在各因素上之因素负荷量相乘再求和，以推算这两个变量之间原来的相关系数，而把这种方法和原来相关系数矩阵之差算作残差(residual)，从残差的大小，来估计因素分析的结果是否良好，这种分析的结果也是寻求“共通性”的大小，从而寻求最佳结果。但这种方法是一种反复迭代的方法，造成了计算上的困难。

(4) “心理测量”法 (Psychometric method)

“心理测量”法有两种，一种是 α 法 (Alpha factor analysis)，另一种是“映像”法 (image factor analysis)。这两种方法出于对心理测量普遍含义的考虑而命名。在心理测量中，对某一群体进行研究，通常采用统计推论的方法，即从某群体中抽取若干样方，并对这些样方的属性进行观测，通过对变量的统计得出对这一群体的属性的统计性推理，在这里，取样的单位是个人，当把某观测变量作为对变量母体中进行研究时取样的单位时，对这些观测变量的因素分析即是对变量母体的心理统计意义上的推论。

① α 法：

凯泽 (H.F.Kaiser) 和卡弗里 (J.Caffrey) 于1965年提出了 α 因素分析法。这里， α 是借用库达—理查森Kuder-Richardson的信度系数或称克龙巴奇赫(L.J.Cronbach) α 系数的名称和含义。

这种方法主要寻求与所对应的全体公因素具有最大相关值的那些作为观测样本中的公因素的集合，一般是以多元回归系数的平方作为对“共通性”的估计，并依 n 个观测变量只是作全体变量的一个样本的假设，重新调整矩阵，这个迭代过程直至“共通性”逐渐收敛为止。

其公式为

$$\alpha_p = \frac{n}{n-1} \left(1 - \frac{1}{\lambda_p} \right)$$

式中 λ_p 为特征值。

②映像分析：

映像因素分析是格特曼 (L.Guttman) 1953年提出的。格特曼用复相关系数平方法 (SMC法) 来估计“共通性”。他认为用 SMC 来作为变量的共通性 h^2 并置于相关系数矩阵的主对角线上、以此来分析一个变量的变异量中所能由其它变量所预测的部分，并将这部分称为该变量的映像 (image)，即是该变量的共通性，而变量的独特性部分称为反映像 (anti-image)，即不能由其余变量之线性组合所预测的部分。自变量样本中得出的“映像”和“反映像”称之为变量总体的“偏映像”和“偏反映像”。

但是由这两种因素分析方法所得的结果不尽相同。

(5) 早期的几种方法：

本节介绍几种早期的因素分析方法。这些方法由于种种原因，现在已不多用，但考虑到这几种方法在历史上地位，简加介绍对我们更深刻地认识因素分析的原理和演变是不无裨益的。

①单因素法 (uni-factor method)

最早用因素分析方法研究智能活动的是斯皮尔曼，他所用的是单因素法。虽然他所建构的智能理论称为“双因素理论”，但在实际分析中他仅用了单因素，某变量在因素分析结果

中所涉及的公因素个数称为该变量的“复杂性”(complexity)。在斯皮尔曼的分析结果中,他只给出了符合线性拟合的标准,而这公因素的维度是一,即变量的复杂性最少,这样,代表变量的矢量完全符合于参照轴,而各个变量即可从这单一维度测出。斯皮尔曼的“双因素理论”的“双”只代表有两种能力,一种是普通的共同的,另一种是特殊的。近来在谈到因素分析结果时,只提到公因素的数目,这方面的例子有Holzinger,他所提出的两因素理论(Bi-Factor Theory)即是除独特因素外,尚有两个公因素。

②多组法(Multiple-group Method)

此法是霍尔津格和瑟斯顿在四十年代提出来的。用这种方法可以从一个相关矩阵中同时提取多组因素,这些因素互相斜交。因此,这种分析方法的结果要给出不同的因素模型和因素结构两个矩阵,由此可得出一个约相关矩阵,并决定残差矩阵,通过对残差矩阵的考验和调整可以找出最佳解。用多组法在做因素分析时,着重考虑的不是分析方法的本身,而是形成科学假设的一种途径,并可以检验已形成的理论的妥当性。

在每次对残差矩阵进一步分析时,得出的组因素之间是直交的,而在每次抽取出的组因素阶段中的因素是彼此斜交的,这个特点是与其它方法所不同的。

③重心法(Centroid Method)

重心法是伯特于1917年提出的,后来经瑟斯顿在三十年代发展而成。这是主成分分析方法的一种近似方法。尤其在用手工计算时,比主成分分析要省去许多繁琐的计算,所以,在30—50年代中,这种方法对因素分析的发展起了很大作用,但自从电子计算机广泛应用以来,主成分分析的方法逐渐取代了重心法。

3. 因素的轴旋转与解释

(1) 因素轴旋转

前文提到抽取因素的多种方法,用这些方法都可以抽取出因素来,但是得到的初始因素解在空间上的位置是任务设定的,而因素分析的最终解要求明了各观测变量在因素空间中的组型,当我们适当将因素轴旋转后,则各观测变量在n维因素空间中的关系就会清楚地显现出来。

瑟斯顿等为此曾做过大量工作,并于1937年提出了“简单结构原理”。其主要要求是因素轴经旋转后应能:①每一因素对很多变量来说没有负荷,②每一变量只与少数因素发生关系,即因素“复杂性”越低越好。这样,将初始因素矩阵做一恰当线性变换后,减低了因素的复杂性,增加了因素的简单性,从而使因素的解释工作便于进行。瑟斯顿的简单结构原理提出后,在实际应用和理论研究中,大约有95%的因素分析是用来找出简单结构的,但所用的方法都是正交旋转(orthogonal rotation),即旋转过程中因素轴互相正交,始终保持初始解中因素间不相关的特点。但是,这种正交简化的解法,曾一度妨碍了因素分析的发展。

1952年卡特尔(R.B.Cattell)提出了“决定性因素原理”作为对简单结构原理的补充和修正。他提出:“因为使用因素分析是为了科学研究。所以,按照某一科学的实际模型进行旋转更为重要”。所以在卡特尔的解法中所得到的结果是:因素矩阵应表示为与决定性因素相一致。卡特尔的决定性因素理论提出后二十年来,受到了大多数从事因素分析研究工作者的重视。他的理论也反映了客观事物的普遍规律。人们从实验数据中发现,在绝大多数科学领域中相互影响的各种因素不大可能是彼此无关的,斜交因素是普遍存在

的，而正交因素是较少的特例。在自然界中引起事物变化的各种内在因素之间始终存在着某种错综复杂的相互联系，当对因素再进行因素分析时，则会得出二阶或高阶因素的体系。这些结果用正交旋转是得不到的，即使在总体中各因素是不相关的，那么在子样中则有可能相关。而且在实际分析工作中，运用斜交旋转（oblique rotation）方法取得了较好和较满意的结果。正是从这几方面，卡特尔的决定性因素原理解决了简单结构所不能很好解释的问题。

一般说来，用斜交因素旋转所得到的因素模型与自然模型更为相近，近来用得也很广泛的也是斜交旋转法。

(2) 正交旋转法与斜交旋转

正交旋转法的特点已如前文所述，通常有图解法和解析法，在大多专用软件中都备有这两类程序。一般还可在终端显示出图形，由研究者选择合理的旋转角，当根据分析图形，并依简单结构准则进行判断后，则可固定该旋转角，但在很大程度上依赖于研究者的主观判断。因此，用解析法较为客观，本文不一一讨论正交旋转的各种方法，仅给出正交旋转的普遍公式，依标准不同，可分作四种方法。

$$\sum_{j=1}^r \left(\left(\sum_{i=1}^n b_{ij}^2 \right)^2 - \frac{r}{n} \left(\sum_{i=1}^n b_{ij}^2 \right)^2 \right) = \text{极大} \quad (16)$$

$$r = \beta(\alpha + \beta)$$

其中 b_{ij} 为 i 变量在 j 因素上之负荷。

当：

$r = 0$ 四次幂极大法 Quartimax method

$r = 1$ 方差极大法 Varimax method

$r = \frac{r}{2}$ 等公因素极大法 equimax method

$r = .5$ 双四次幂极大法 biquartimax method

表 4 不同旋转方法之因素分析解

变 量	初 始 因 子 解 (最大似然法)			正 交 旋 转 (方差极大法)		斜 交 旋 转			
	F ₁	F ₂	共通性	F ₁	F ₂	模 型 矩 阵		结 构 矩 阵	
X ₁	.766	-.232	.640	.783	.163	.800	.000	.800	.320
X ₂	.670	-.203	.490	.685	.143	.700	.000	.700	.280
X ₃	.574	-.174	.360	.587	.123	.600	.000	.600	.240
X ₄	.454	.533	.490	.143	.685	.000	.700	.280	.700
X ₅	.389	.457	.360	.123	.587	.000	.600	.240	.600
X ₆	.324	.381	.250	.102	.489	.000	.500	.200	.500

斜交旋转实际上是去掉正交旋转的某些限定和假设而将因素轴进行旋转的方法。限于篇幅这里不再介绍。

表4、给出了未经旋转和用不同旋转方法所得的因素分析解。

读者可以根据表中所列数据,看出经斜交旋转后因素的模型和因素结构都变得简洁,这对理解和解释因素的意义带来很大便利。

通常,对因素分析的结果中还应包括因子计量(measurement of factors),即给出用变量的线性组合表示因子的解析式或矩阵来。在计算机结果中这个矩阵称为因素分系数(Factor-score coefficient(or factor-estimate matrix),以此来衡量对某一因子进行评估的准确度。

几点小结

1. 因素分析是在对客观事物运动规律进行深入研究过程中发展起来的一种多元统计分析方法,应用因素分析有助于发现那些事物本身固有的、决定事物性质的本质特征,它通过对可观测的事物在发展中所表现的外部形态和联系的特征得出对客观事物普遍本质的概括和反映,用因素分析所提出的假设,可以用表征产生某种自然现象的内在因素数目和性质来表示和刻划。其基本逻辑模型为:“因素是作用于变量的原因”并最终因素模型、因素结构和因素计量来解释变量与因素之间的交互关系。

2. 因素分析根据对多变量的观测分析,较全面地反映了所研究对象的各个侧面。这种方法克服了传统单变量和双变量实验的片面性和不足,当对实验条件不可能进行控制或不可能对现象进行时序观测时仍可得出变量间的关系以及造成变量变异原因关系的研究,这在科研活动中无疑是一大进步。因此,用因素分析的方法建立假设的考虑正是基于客观实际中存在着的多种决定性主要因素这一前提出发,并在建立假设后,可以进行反复试验,以验证这些假设。所以,因素分析是一种既提出假设又检验假设的统计推断理论与方法,正是这一优点推动了因素分析的不断发

3. 进行因素分析的三个基本阶段中(收集资料、抽取初始因素、因素轴的旋转与解释),每一阶段都有多种选择,这些方法是多年来研究者探索的成果,每种方法的考虑都有其特定的渊源和目的,这就给研究者带来了某种程度的抉择困难,在应用因素分析进行研究时,应当遵循如下两种准则:①进行分析的统计学考虑,②特定研究领域和目的的有意义解释的考虑。在广泛应用计算机的今天,研究者不妨用种种方法进行尝试,以选定本学科最适宜的方法。

4. 因素分析对心理学的发展起了巨大推动作用。斯皮尔曼的多因素理论,霍尔津格的两因素理论,瑟斯顿的多因素理论,以及吉尔福德的智能结构等一系列研究和理论都是用因素分析的方法对人们的智能进行了逐步深入的研究,R.B.卡特尔等人对人格的研究和理论把因素分析方法的重要性提高到了一个新的高度。然而,这毕竟只是某研究者毕生努力建立理论体系的范例。在大量实践研究中,现代人可以用较为成熟的因素分析和理论建构量表,验证理论的研究,这些报告在近二十年中是非常大量的,因此,本文简要介绍了对因素分析结果的解释和基本求解过程,以期引起我国心理学工作者的重视,并在工作中能逐步得以应用。近年来,我国对学习理论、人格理论研究逐渐增加,但对国外

基于因素分析理论建构起的理论并未加以因素分析方法上的“假设验证”工作，这不能不说是一种遗憾，对国外的理论和量表的考察，除了作理论上探讨和验证外，还应加入我国自己多元统计分析结果的考察，否则，便是不完备的。

5. 国外通行的含因素分析的程序软件有 SPSS、OSIRIS、SAS、BMD 程序。我国地质、生物等学科也有自己的程序，目前，科学院心理研究所和北京市计算中心正在调制用于心理学的专用程序，待研究成功，希望能得到国内专家的评审，并期望能在心理学研究中得到推广。

参 考 文 献

- [1] Cattell, R.B. *Factor Analysis*, New York: Harper Bros. (1952).
- [2] Fruchter, B. *Introducton to Factor Analysis* (1954). New York: Van Nostrand. (1954)
- [3] Schuessler, K *Anal yzing Social Data*. (1971). Boston: Hougheon Mifflin.
- [4] Comrey, A.L. *A First Course in Factor Analysis*, (1973) New York: Acaclemic press.
- [5] Lawley. D.N. and Maxwell, A.E., *Factor Analysis as a Statistical Method*, (1971). London: Butterworth andco.
- [6] Milaik, S.A., *The Foundations of Factor Analysis* (1972) New York: McGraw-Hill.
- [7] Harman, H.H., *Modern Factor Analysis* (1976)., Chicago and London:
- [8] The University of Chicago press.
- [9] 于崇文: 《数学地质的方法与应用》, 北京, 冶金工业出版社, 1980
- [10] 阳会熙、卢泽愚: 植物生态学的数量分类方法, 1981. 北京, 科学出版社,
- [11] 谢小庆 因素分析: 研究生物科学和社会科学的重要工具。北京《百科知识》1983年第4期
- [12] 中国科学院计算中心概率统计组《数字计算机上用的数学方法》(第三卷) 1981. 上海科学技术出版社
- [13] 张尧庭、方开泰: 《多元统计分析引论》1982. 北京, 科学出版社,