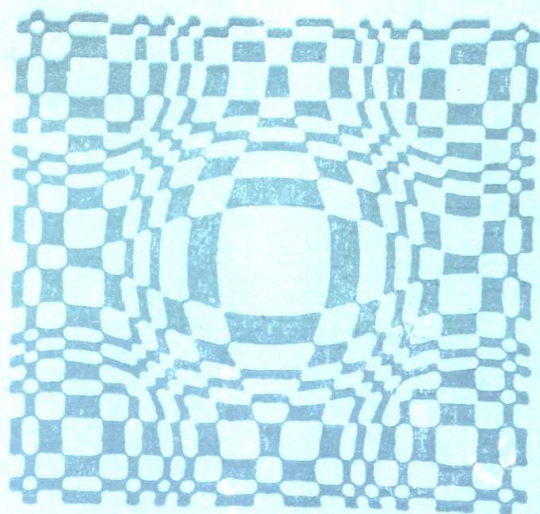


谢小庆
王丽
编著

YINSUFENXI



YINSUFENXI

因素分析

中国社会科学出版社

因 素 分 析

一种科学研究的工具

谢小庆 王丽 编著

中国社会科学出版社

责任编辑 杨 青 李树琦
责任校对 杨惠然
封面设计 谭国民
版式设计 张汉林

因素分析

Yinsù fēnxi

中国社会科学出版社 出版
发 行

新华书店 经销
北京印刷二厂印刷

787×1092毫米 32开本 6.25印张 2插页 131千字

1989年12月第1版 1989年12月第1次印刷

印数 1—3 500 册

ISBN 7-5004-0440-9/D·77 定价: 2.40元

• 前 言 •

因素分析是处理多变量数据的一种数学方法，它可以揭示多变量之间的关系，其主要目的是从为数众多的可观测的变量中概括和推论出少数的“因素”，用最少的“因素”来概括和解释大量的观测事实，从而建立起最简洁、最基本的概念系统，揭示出事物之间最本质的联系。因素分析是适用于广泛的科学研究领域的一种重要的数学工具。

因素分析方法是由心理学家在关于人的智力的研究中发展起来的，今天，它已经发展成为数学的一个分支。因素分析虽然已有半个多世纪的发展历史，但由于它那庞大繁复的计算非人力所及，在长时期内未能得到广泛的应用。只是在最近的十几年中，借助于现代计算技术才得到了迅速发展。今天，它已经被广泛地应用于心理学、教育学、社会学、经济学、考古学、文献学、以及地质学、气象学、医学等领域的研究之中。今天，许多计算机中心都备有现成的因素分析程序。

在科学技术迅速发展的今天，在自然科学和社会科学的许多领域中，追求数量化已成为潮流。因素分析方法的确是使许多研究工作达到科学化和定

量化的一种良好途径。考虑到目前我国自然科学和社会科学工作者的一般数学水平，本书着重于介绍因素分析的基本原理和应用范围，使具有一般统计学基础知识的人即可阅读，帮助他们借助计算机来解决自己研究领域中所面临的问题。

因素分析本身是现代科学理论化和计算机化的产物，通过阅读本书，将有助于读者提高自己在一般科学研究方法论方面的素养。

本书主要面对一般科学领域中从事实际科学研究的工作者，力图在他们与计算机之间建立起一座桥梁。首先，本书将帮助读者决定在自己的研究工作中是否可以采用因素分析方法。通过阅读第一、三、九章，这一问题即可得到回答。如果在你的研究工作中可以采用因素分析方法，本书将告诉你怎样在计算机站工作人员的帮助下，结合你对自己的研究领域的知识素养，适当地运用因素分析方法。

本书亦可作为大学本科生和研究生学习数理统计课程的参考书。

内 容 提 要

因素分析（即因子分析）是处理多变量数据的一种数学方法，它可以从为数众多的可观测变量中概括和推论出少数的重点“因素”，从而建立起最基本、最简洁的概念系统，揭示出事物之间最本质的联系。因素分析方法在自然科学和社会科学领域有着广泛的应用。

目 录

前 言	(1)
-----	-------

第一章 导言	(1)
--------	-------

一、因素分析的基本过程	(2)
-------------	-------

二、运用因素分析发现规律	(4)
--------------	-------

三、运用因素分析发展理论	(7)
--------------	-------

四、简单性标准与因素分析	(9)
--------------	-------

五、因素分析的应用举例	(10)
-------------	--------

第二章 线性代数知识	(14)
------------	--------

一、矩阵的定义和表示	(14)
------------	--------

二、矩阵运算	(17)
--------	--------

(一) 相等	(17)
--------	--------

(二) 加减法	(17)
---------	--------

(三) 乘法	(19)
--------	--------

(四) 逆矩阵	(22)
---------	--------

(五) 数乘矩阵	(24)
----------	--------

三、常用统计量的矩阵表示	(25)
--------------	--------

(一) 平均数	(25)
---------	--------

(二) 方差	(26)
--------	--------

(三) 标准分数	(26)
----------	--------

(四) 相关系数	(27)
----------	--------

第三章 因素分析模型	(30)
------------	--------

一、因素分析的起源	(30)
-----------	--------

二、因素分析的代数模型	(32)
三、因素分析的几何模型	(42)
(一) 将变量看作坐标轴	(42)
(二) 变量作为因素的函数	(45)
(三) 从变量出发的几何模型	(47)
第四章 因素抽取	(49)
一、特征值与特征向量	(50)
二、因素抽取的特征值方法	(53)
三、公共因素方差的估计	(64)
(一) 什么是公共因素方差	(66)
(二) 公共因素方差的上限和下限	(67)
(三) 用多重回归系数估计公共因素方差	(68)
(四) 用信度估计公共因素方差	(70)
(五) 以最大相关系数估计公共因素方差	(71)
(六) 用迭代法估计公共因素方差	(71)
(七) 关于必要性的讨论	(73)
四、公共因素数目的确定	(75)
(一) 因素模型与数据的充分协调性	(77)
(二) 公共因素数目确定的统计方法	(81)
(三) 确定公共因素数目的代数方法	(87)
(四) 计算方差的百分比以确定公共因素 数目	(94)
(五) 图解法确定公共因素数目	(95)
(六) 旋转过程中对因素数目的调整	(98)
五、其它因素抽取方法	(98)
(一) 对角线法	(99)

(二) 分组法·····	(104)
(三) 最小残差分析·····	(106)
(四) 影像分析·····	(107)
第五章 旋转变换与因素解释·····	(110)
一、因素旋转变换的原则·····	(113)
(一) 简单结构原则·····	(113)
(二) 成比例轮廓·····	(115)
二、视觉直觉旋转·····	(115)
三、旋转的代数基础·····	(116)
四、分析旋转·····	(118)
(一) 最大方差旋转·····	(119)
(二) 最大四次方旋转·····	(127)
五、斜交旋转·····	(133)
(一) 直接视觉斜交旋转·····	(134)
(二) 分析斜交旋转·····	(136)
六、各种旋转方法的比较·····	(145)
(一) 内部标准·····	(145)
(二) 外部标准·····	(147)
(三) 一些比较研究的结果·····	(149)
第六章 高阶因素·····	(150)
一、对高阶因素的解释·····	(150)
二、高阶因素的抽取·····	(152)
(一) 不适用性·····	(152)
(二) 更高要求·····	(152)
(三) 显著性考验·····	(153)
三、原始变量与高阶因素的关系·····	(158)

四、两个例子.....	(155)
第七章 因素分数.....	(157)
一、因素分数的计算.....	(158)
二、对因素分数的近似计算.....	(160)
三、对因素分数的检验.....	(163)
第八章 不同因素分析研究的比较.....	(167)
一、不同因素分析之间进行比较的依据.....	(167)
(一) 因素分数之间的相关.....	(168)
(二) 标记变量.....	(168)
(三) 与已明确因素的相关系数.....	(168)
(四) 因素在总变异中所占比重.....	(169)
(五) 在进一步研究中的作用.....	(170)
二、对相同变量、相同样本的不同 分析过程的比较.....	(170)
三、变量相同、样本不同情况下的比较.....	(171)
(一) R 、 R_{c} 和 W 已知的情况.....	(171)
(二) R 、 R_{c} 和 W 未知的情况.....	(172)
四、因素的搭配.....	(175)
第九章 因素分析的应用.....	(176)
一、因素分析的一般程序.....	(176)
二、提出理论概念并使之数量化.....	(180)
(一) 建立概念.....	(180)
(二) 概念的数量化.....	(184)
三、对因素分析当前应用的批评.....	(185)
四、因素分析的未来.....	(187)
参考文献.....	(189)

第一章

导 言

对于平面上的一个圆，你可以测量的数据有半径、直径、周长、面积、外接正方形面积等等，但所有这些数据之中，最基本的只有一个：半径。对于一架飞行中的飞机，你可以测量的数据也有很多，离北京多远、离上海多远、出发了多长时间，等等，但最基本的数据只有四种：经度、纬度、高度和时间。当我们描述这一类事件时，我们往往可以直观地看出这些基本的数据。在科学研究的领域中，有许多事件是难以直观地看出那些最基本的数据的。例如：对于一个地区的经济发展，可以测量的指标很多，如总产值、利润、投资、效益、劳动生产率、万元产值能耗、固定资产、流动资金周转率、物价、信贷、税收，等等，其中最基本的指标是哪儿个呢？对于一个人，从身高、体重到兴趣、爱好、情操，可以用一定方式测查的方面也是很多的，在了解或介绍一个人的时候，应该采用哪些数据呢？对于一种岩石，可以测查或化验的指标有许多，如颜色、硬度、含碳量、含硫量，等等，应该用哪些基本指标来描记这种岩石呢？对于一种鱼，可以测量的特征也有许多，哪几种指标是最基本的呢？在这些情况下，仅凭直观就难以找到那些最基本的数据。这时，就需要借助数学工具。因素分析正是解决这些问题的一种工具。

• 1 •

在系统地介绍因素分析的过程、方法和应用之前，我们有必要将因素分析方法放到整个科学研究背景中，指出它在科学研究中的意义，尤其是指出它作为一种达到简单性的数学工具在现代科学中的意义。我们也有必要通过一些实际的例子，指出它广泛的应用范围。

一、因素分析的基本过程

因素分析是一种统计技术，它的目的是从为数众多的可观测的“变量”中概括和推论出少数的“因素”，用最少的“因素”来概括和解释大量的观测事实，从而建立起最简洁、最基本的概念系统，揭示出事物之间最本质的联系。

对于一种研究对象，从不同方面收集资料，可以得到一组变量。这组变量反映了事物的什么特性？变量间的关系如何？每个变量在多大程度上反映了事物的特点？因素分析能帮助我们回答这些问题。

例如，对一班学生进行了词汇（1）、倒背数字（2）、算术推理（3）、语言流畅（4）、故事推理（5）和四则运算（6）这样六个测验，得到一组原始分数（表1.1）。那么这些测验分数反映了学生们的什么特点呢？我们先计算出这六个变量之间的相关系数，列出相关矩阵（表1.2）。对这个相关矩阵进行因素分析，经过因素抽取、因素数目确定、转轴等步骤，得到包含 F_1 、 F_2 两个因素的因素矩阵（表1.3）。从表1.3可以看出，测验1、4、5与因素 F_1 有关，测验2、3、6与因素 F_2 有关。根据我们关于这一领域的知识，可以将 F_1 命名为“语文能力”，将 F_2 命名为“数字能力”。表

表1.1 一班学生在六项测验上的得分

学生/测验	1	2	3	4	5	6
1. 李 明	23	19	40	27	13	12
2. 陈 伟	27	19	41	25	17	19
3. 王晓萍	31	25	40	26	17	17
4. 刘一凡	33	17	43	29	16	15
⋮						
45. 胡 胜	30	25	40	19	15	17

表1.2 六项测验得分的相关矩阵

测验	1	2	3	4	5	6
1	1	0	.16	.56	.72	0
2	0	1	.72	0	.27	.54
3	.16	.72	1	.14	.42	.48
4	.56	0	.14	1	.63	0
5	.72	.27	.42	.63	1	.18
6	0	.54	.48	0	.18	1

表1.3 因素矩阵

测验/因素	F_1	F_2	h^2
1	.999843	-.017736	.77574
2	.018288	.999833	.8093
3	.249369	.968411	.77622
4	.999332	-.036536	.70486
5	.948084	.31803	.85487
6	-.037386	.999301	.60799
总方差中比例	.44393	.31089	
累积方差	.443934	.754829	

1.3中的 h^2 反映了 F_1 、 F_2 这两个因素在多大程度上决定了各个变量。例如，这两个因素决定了词汇测验(1)分数变异的77.57%，决定了倒背数字测验(2)分数变异的80.93%。表1.3下面的两行中给出了每个因素所决定的变异(方差)在全部六项测验的总变异(总方差)中所占的比例和累积比例。因素 F_1 和 F_2 所决定的变异在全部变异中所占的比例分别是44.39%和31.09%，两个因素的累积比例是75.48%，也就是说，这两个因素决定了全部六项测验总变异的75.48%。

至此，我们不仅回答了“这些测验测的是什么”和“各测验间关系如何”的问题，而且通过 h^2 和累积百分比回答了“语文和数字能力在多大程度上决定了各个测验的成绩”的问题。

二、运用因素分析发现规律

科学研究的主要目的是揭示事物运动发展的规律。规律是一个事物随其它事物的变化而变化的关系，是两个或多个变量共同变化的关系。物理学中的运动定律，是指位置变量随时间变量而变化的关系；生物学中的遗传规律，是指子代的性状变量与亲代的性状变量之间的关系；心理学中的感觉定律，是指感觉变量与刺激、需要、准备状态等变量之间的共同变化关系；经济学中的价格规律是指价格变量与成本、税收、利润、供求等变量之间的共同变化关系。

一个事物的变化往往受到众多因素的影响。在传统的物理学、生理学和实验心理学研究中，总是控制其它变量，考察两个变量之间的关系，改变一个变量看另一个变量随之变

化的情况。例如，控制压力、气体成份等变量，改变温度而考察气体体积随温度变化的情况。在许多情况下，尤其是在生物科学和社会科学领域中，由于影响因素为数众多，这种方法受到局限。

一方面，变量的控制有时是不可能的。要想比较两种药物的疗效，就要给两组病人分别服用，这就要求两组病人在体质、患病程度、休养条件、精神状态等方面基本相同；要想比较两种教学方法，就要对两个班的同学分别用两种方法施教，这就要求两班学生在智力、知识经验、动机等方面基本相同。这些都是很难做到的。另一方面，变量的控制有时是无意义的。在事物的运动发展中，各种因素是互相影响的，往往只有将各种因素同时放到整个系统中加以综合考察，才能揭示出它们之间的关系。控制了其它变量，就揭示不出所研究变量之间真正的关系。由于不同因素的组合，可能会使某一单一因素产生不同的作用，所以，单独考察某一因素是没有意义的。例如，在一定条件下，锄草既可能因保证了作物养分而增产，也可能因使害虫的天敌无处栖居而造成害虫横行，产量下降；鞣酸蛋白具有强收敛止泻作用，在腹泻初期服用可能由于使病毒不能排出而引起人畜中毒以至死亡，在腹泻后期服用则可能由于防止虚脱而达到好的疗效；草乌类有毒药物对体质强的患者有很好疗效，对体质弱的患者则可能造成危害；降低价格可能增加销售量，也可能因使商品失去信誉而降低销售量；艰苦的生活环境可能毁灭一个有才能的人，也可能造就一个天才。

因此，在许多情况下，尤其是在生物科学和社会科学的研究领域中，就不能采用传统的单因素实验方法，而需采用

多因素实验、观察或调查的方法，收集各种变量的交叉资料，从对这些资料的分析中揭示出事物之间的联系。

当影响因素很少时，我们可以直观地看出或“猜出”各种变量之间的关系，但当影响因素众多时，我们就面临一个庞大的数据群，从这些数据中，首先，我们很难直观地发现各种变量之间的关系。在实际的研究中，我们经常面临几十个以至上百个变量（即表1.1有几十上百列），每个变量可能包含成百上千的观测数据（即表1.1有成百上千行），这时，发现规律的任务就可以归结为一个化简问题，即将众多的变量归结为一些主要的、正交的、互不相关的“因素”，用尽可能少的因素来概括尽可能多的数据资料。

其次，我们很难凭借直观的方法来了解各个变量之间、变量与“因素”之间的联系上的不同程度。例如，如果一个变量与表1.2中的变量1、4、5的相关都是0.35，与变量2、3、6的相关均为0，它将被归于因素 F_1 ，但是，与变量1、4、5相比，它与 F_1 的联系是较弱的。因素分析的方法可以为变量之间的联系程度提供一个数量指标。

既然直观的方法无法从这个庞大的数据群中发现规律，我们就需要借助于数学工具。多重回归、因素分析就是这样的一些数学工具。正如多重回归是简单相关和直线回归的深化产物一样，因素分析是多重回归的深化产物。多重回归可以考察一个变量与一组变量之间的相互关系，而因素分析可以同时考察一组变量之间的相互关系。多重回归的输出数据是关于单一变量的多重回归方程式，而因素分析的输出是一个“因素矩阵”，即一个反映了各个变量之间相互关系的多重回归方程组。更重要的是，因素分析可以得到更高级的

“因素”，这在提出假设和验证假设方面，在推动理论的发展方面，有着其特殊的优点和重要的意义。

三、运用因素分析发展理论

任何科学研究的直接目的都是将观察得到的经验事实概括成抽象的概念并将这些概念组成理论的公式，或将众多的较低级的概念和理论公式概括成为更高级、更抽象的概念和理论。科学研究的过程就是产生概念和理论的过程，科学发展的历史就是概念产生和发展的历史。在物理学中，人们首先将张三的锤、李四的锤、大锤、小锤、木锤、铁锤这些各式各样的锤子概括成脱离开具体事物的抽象概念“锤子”，又将锤子与法码、子弹、星球等较低级概念概括成较高级概念“质点”，将质点概念与“力”、“加速度”等概念联结在一起，就形成了物理学的运动定律；在生物学中，高矮、粗细、黑白这些较低级概念被概括成较高级的概念“性状”，性状概念与“遗传”、“基因”等概念联结在一起，就形成了生物学中的遗传定律；在心理学中，重量、波长、频率、分子数这些较低级概念被概括成较高级的“刺激量”概念，刺激量概念与“感觉量”等概念联结在一起，就形成了心理学中的感觉定律。

任何科学理论的建立都是从创造概念开始的，新概念的引入总是大大地推动了理论的发展。“位移电流”概念的引入，导致了麦克斯韦电磁学说的建立；“规范场”概念的引入，导致了弱电统一理论的建立；“基因”概念的引入，导致了现代基因遗传学说的建立；“条件反射”概念的引入，导致