DOI:10.13700/j.bh.1001-5965.1998.02.031
1998年 4月 北京航空航天大学学报 April 1998
第24卷 第2期 Journal of Beijing University of Aeronautics and Astronautics Vol.24 No.2

多元分析在大学研究生院规模研究中的应用

徐惠娟 (北京航空航天大学 校长办公室)
王惠文 (北京航空航天大学 管理学院)

摘 要 运用因子分析和聚类分析等多元统计分析方法,对中国设立研究生院的18所理工科大学研究生院1994年的部分数据进行分析、处理,将表征规模的各统计变量综合为博士规模因子和硕士规模因子两个综合变量,并根据每一样本点的因子得分对18所理工科大学研究生院办学规模进行排序和分类,为评估理工科大学研究生院办学规模提供参考依据.

关键词 因子分析; 排序; 聚类分析
分类号 O 212.4

我国自1984年试办研究生院,至今已有33所试办研究生院的高校正式建立研究生院.这些研究生院通过十多年的实践,在扩大研究生办学规模,保证和提高研究生培养质量和学位授予质量,开展学科建设等方面进行了积极的探索,对促进我国研究生教育事业的发展,起了重要的作用.为总结经验,也为研究生教育的总体布局和新建研究生院提供决策依据,我国于1995年首次开展了对全国33所研究生院的评估工作.这次评估的数据处理涉及了多种数学方法,如加权和法、百分线性规划法、区域线性规划法、平均值规划法等.国外也经常评估研究生院,各国评估研究生院的数据处理方法各异,有用简单数据处理方法,也有用多元统计分析方法,如用对应分析法等.本文运用因子分析和聚类分析等多元统计方法,对我国18所设立研究生院的理工科大学表征规模指标的数据进行分析、处理,试图揭示影响理工科大学研究生院办学规模的综合因素,并对评估研究生院办学规模的数据处理方法进行了初步探讨.

1 研究方法

本文以全国18所理工科大学的研究生院为研究对象,选取象征研究生院办学规模的1994年部分统计数据,进行因子分析与聚类分析.原始数据[1]见表1.

1.1 因子分析

因子分析的基本目的是用少数几个变量去描述多个变量间的协方差关系.其思路是将观测变量分类,将相关性较高即联系比较紧密的变量分在同一类中,每一类的变量实际上就代表了1个本质因子,从而可将原观测变量表示为新因子的线性组合.本研究运用因子分析的主要目的是简化观测系统,将已知的8个变量减少为几个新因子,以再现它们之间的内在联系.

因子分析的结果不仅给出因子模型,还得出变量和因子间的相关系数,这些相关系数构成因子结构.1个完全的因子解包括因子模型和因子结构两个方面,因子结构反映变量与因子间的相关关系,而因子模型则是以回归方程的形式将变量表为因子的线性组合.

本研究主要经过以下几个步骤:

1) 将原始数据进行标准化处理.设原始数据矩阵为

$$\boldsymbol{X}=(x_{ij})_{n\times p}$$

则

$$x'_{ij}=\frac{x_{ij}-\bar{x}_j}{s_j}$$

$$i=1,2,\cdots,n \qquad j=1,2,\cdots,p$$

其中

$$\bar{x}_j=\frac{1}{n}\sum_{i=1}^{n}x_{ij}$$

$$s_j^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2$$

2) 计算所有变量的相关矩阵 R.

3) 因子提取.本研究采用主成分分析法,利用相关系数矩阵 $\boldsymbol{R}$ 进行因子提取.可通过研究公共因子在变量总方差中所占的累计百分数 Q_m(一

收稿日期:1997-02-13 第一作者:女,31岁,助研,100083,北京

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

表1 我国18所理工科大学研究生院部分数据

单位	变量名							
	学位授权点		在校研究生		已授学位		重点学科	博士后流动站
	博士 x_1/个	硕士 x_2/个	博士 x_3/名	硕士 x_4/名	博士 x_5/名	硕士 x_6/名	x_7/个	x_8/个
清华大学	64	107	1 202	2 492	941	7 233	29	14
北航空航天大学	20	60	284	1 215	231	3 456	5	4
北京理工大学	20	57	278	1 140	156	2 721	4	3
北京科技大学	15	38	268	1 005	235	2 831	6	3
天津大学	24	75	399	1 784	230	4 336	6	4
大连理工大学	21	65	276	1 415	292	3 657	4	5
东北大学	23	50	330	1 185	139	3 266	2	3
哈尔滨工业大学	29	67	501	1 586	342	4 952	7	5
同济大学	20	62	271	1 177	167	2 632	4	4
上海交通大学	34	75	267	1 755	278	4 086	8	5
东南大学	21	62	302	1 181	206	3 004	4	1
浙江大学	35	84	658	1 882	367	4 598	9	10
中国科技大学	24	46	246	778	225	2 433	4	5
华中理工大学	29	76	495	1 737	299	4 087	4	4
中国地质大学	13	20	215	472	201	1 942	5	2
国防科技大学	14	40	212	776	85	2 182	2	1
西安交通大学	33	65	690	1 694	403	4 812	11	8
西北工业大学	19	54	277	1 102	207	3 078	3	3

般为85%以上)确定所需要的公共因子数.

4) 因子旋转.因子分析的目的不仅是要找出主因子,更重要的是知道每个主因子的意义,为便于对主因子进行解释,一般须对因子载荷矩阵进行旋转,以达到结构简化的目的.由于代表研究生院规模的各变量之间不可能是彼此无关的,故本研究作了斜交旋转.

因子模型的表达式为

$$x_i = a_{i1}F_1 + a_{i2}F_2 + \cdots + a_{im}F_m + a_iu_i$$

其中 $F_1, F_2, \cdots, F_m$ 为公共因子; a_{ij} 为因子载荷,它是第 i 个变量在第 j 个主因子上的负荷; u_i 为特殊因子, a_i 为特殊因子载荷.

5) 计算每一样本点的因子得分,用于诊断模型,并进行因子排序.在本文中,因子得分还作为进一步聚类分析的原始数据.

1.2 聚类分析

聚类分析是将一批样品或变量,按其性质的亲疏程度进行分类.本研究采用了系统聚类法.

定义各样本点之间距离采用的是平方欧氏距离,其公式为

$$d_{ij} = \left[\sum_{k=1}^{p}(x_{ik} - x_{jk})^2\right]^{\frac{1}{2}}$$

其中 d_{ij} 表示样本点 i 与样本点 j 之间的距离.

定义各类之间距离用最短距离法,其公式为

$$D_s(p,q) = \min\{d_{jk} \mid j \in G_p, k \in G_q\}$$

式中 $D_s(p,q)$ 为 G_p 类和 G_q 类最近的两个样本点间的距离.

通过系统聚类,可以得到树状谱系图,谱系图能明确清晰地描述各个样本点在不同层次上聚合分类的情况.

1.3 绘制散点图

根据因子得分,绘制散点图,在散点图上能清楚地观察不同的类别及其性质.

2 计算结果与分析

从相关矩阵 $\boldsymbol{R}$ 得知,各变量之间相关性很强,因此,有可能存在一些变量共享因子.同时,计算表明,几乎总方差的95.3%可由第1和第2因子解释,由此可选择两个主因子.表2为经斜交旋转后的因子模型矩阵和因子结构矩阵.

表2 斜主因子模型矩阵和结构矩阵

变量	模型矩阵		结构矩阵	
	F_1	F_2	F_1	F_2
x_7	1.089 98	−0.146 90	0.976 63	0.694 07
x_5	1.006 22	−0.026 17	0.986 03	0.750 18
x_3	0.850 73	0.147 81	0.964 77	0.804 19
x_8	0.829 76	0.143 05	0.940 12	0.783 24
x_1	0.703 31	0.325 48	0.954 43	0.868 12
x_2	−0.029 06	1.006 64	0.747 61	0.984 21
x_4	0.074 26	0.934 17	0.795 02	0.991 47
x_6	0.455 50	0.578 78	0.902 05	0.930 22

注: F_1 为第1斜因子, F_2 为第2斜因子.

it

分析表2看出,经斜交旋转后,第1主因子主要由与博士规模有关的变量即博士学位授权点数、在校博士生数、已授博士学位数、博士后流动站数及重点学科表征,而重点学科是从博士学位授权点遴选产生,因此,因子1可解释为博士规模因子.第2主因子主要由与硕士规模有关的变量

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

即硕士学位授权点数、在校硕士生数、已授硕士学位数表征,因此,因子 2 可解释为硕士规模因子.

表 3 为斜交因子得分和按因子 1 和因子 2 对 18 所院校的博士规模和硕士规模的排序表.

表 3　斜交因子得分与因子排序表

序号	单　位	F_1	F_2	P	M
1	清华大学	3.521 38	2.280 12	1	1
2	北航空航天大学	−0.323 22	−0.179 17	10	9
3	北京理工大学	−0.557 53	−0.391 70	16	13
4	北京科技大学	−0.354 86	−1.027 46	11	16
5	天津大学	−0.181 63	0.800 76	7	4
6	大连理工大学	−0.242 02	0.142 84	8	8
7	东北大学	−0.536 41	−0.340 08	15	12
8	哈尔滨工业大学	0.293 18	0.553 66	4	6
9	同济大学	−0.514 33	−0.264 73	13	11
10	上海交通大学	0.040 71	0.754 49	5	5
11	东南大学	−0.582 54	−0.213 07	17	10
12	浙江大学	0.811 28	1.146 50	3	2
13	中国科技大学	−0.273 16	−0.974 00	9	15
14	华中理工大学	−0.053 77	0.804 48	6	3
15	中国地质大学	−0.520 32	−2.017 11	14	18
16	国防科技大学	−0.932 88	−1.162 12	18	17
17	西安交通大学	0.920 23	0.521 81	2	7
18	西北工业大学	−0.514 11	−0.435 22	12	14

注:F_1 为斜因子 1 得分,F_2 为斜因子 2 得分,P 为博士规模大学排序,M 为硕士规模大学排序.

此外,本文通过对样本点进行聚类分析,得到谱系图 1,从中可看出,我国理工科大学 18 所研究生院的规模情况大致分为 4 类(见表 4).4 类各项指标的平均值及总平均值见表 5.

图 2 为取第 1 斜因子和第 2 斜因子作因子得分散点图(序号意义见表 3).

分析表 5 及图 2 的 4 大类情况,可以看出:第 1 类为清华大学.在图 2 中,清华大学的博士规模因子与硕士规模因子的得分值远大于其它院校的平均水平.实际上,该校各项研究生规模指标均大于其它院校的同类指标,其学位授权点数、在校研究生数及已授学位人数等均列全国高校榜首.

第 2 类包含了浙江大学等 6 所大学,此类大学的各项研究生规模指标均在 18 所大学研究生院总平均规模之上.这类大学大都建校历史悠久,学科门类较为齐全.

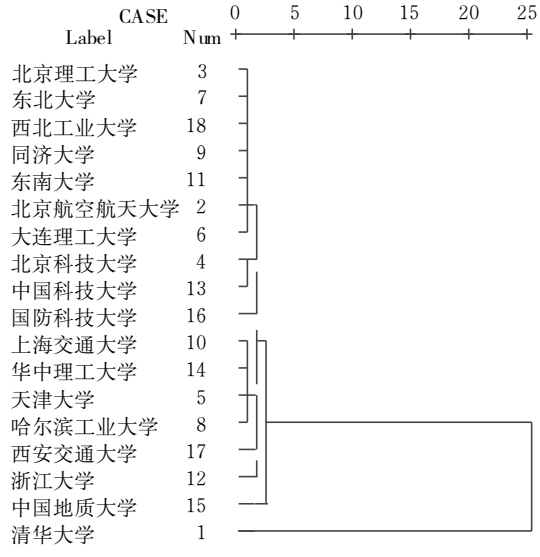


图 1　18 所大学研究生院的系统聚类谱系图

表 4　18 所理工科大学研究生院办学规模分类

类别	该类所包含的大学
1 类	清华大学
2 类	浙江大学、天津大学、哈尔滨工业大学、上海交通大学、华中理工大学、西安交通大学
3 类	北京航空航天大学、北京理工大学、北京科技大学、大连理工大学、东北大学、同济大学、东南大学、中国科技大学、国防科技大学、西北工业大学
4 类	中国地质大学

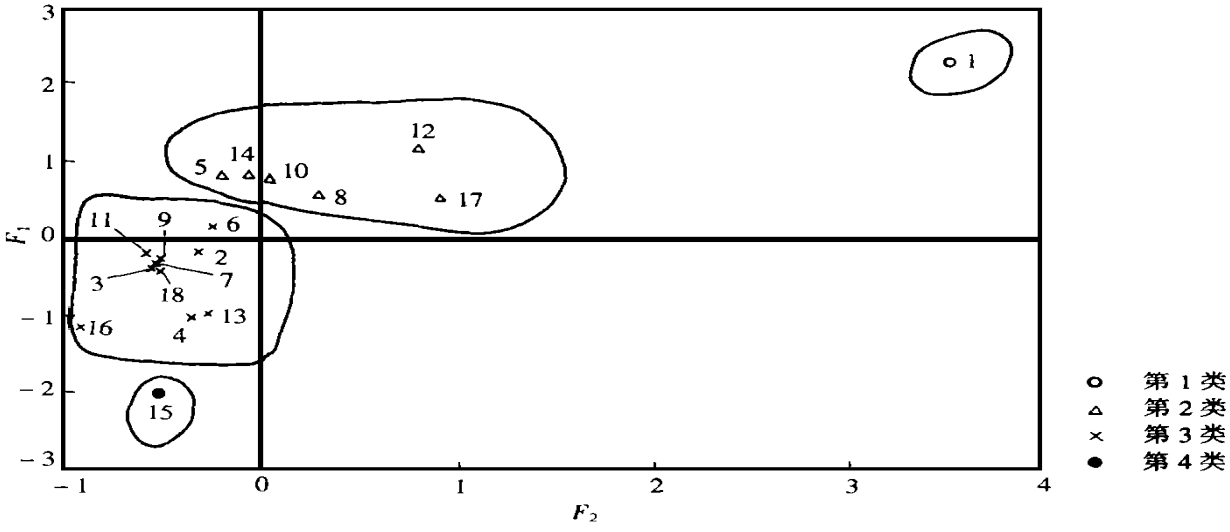


图 2　用博士规模因子得分 F_1 和硕士规模因子得分 F_2 绘制散点图

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

表5 1994年各类大学研究生院平均规模

类别	项目							
	学位授权点		在校研究生		已授学位		重点学科	博士后流动站
	博士	硕士	博士	硕士	博士	硕士		
1	64.0	107.0	1 202.0	2 492.0	941.0	7 233.0	29.0	14.0
2	30.7	73.7	501.7	1 739.7	319.8	4 478.5	7.5	6.0
3	19.7	53.4	274.4	1 097.4	194.3	2 926.0	3.8	3.2
4	13.0	20.0	215.0	472.0	201.0	1 942.0	5.0	2.0
总平均值	25.4	61.3	398.4	1 354.2	278.0	3 628.1	6.5	4.7

第3类包含北京航空航天大学等10所大学，此类大学研究生院各项平均规模指标略低于第2类大学，其中大部分大学有较强的行业特色，并有较强的与行业有关的特色学科，属多科性理工科大学.由图2可见，这一类中，东北大学等5所大学规模非常接近，几乎为重叠点.从表3可见，这5所大学博士和硕士规模排序也很接近.

第4类为地质大学.由图2可看出，该校的硕士规模因子得分值较小，实际上，其各项硕士规模指标均小于第3类同类指标的最低值.这与该校学科门类较为单一，特别是与该校工学学位授权点数相对较少有关.然而，该校有较多的重点学科，其已授博士学位的人数高于第3类的平均值.

3 结束语

利用因子分析将众多的变量综合为几个本质因子，并对样本进行聚类分析，这项工作在高等教育研究中很有意义.本文通过对我国18所理工科大学研究生院1994年的部分统计数据进行因子分析和聚类分析，将表征规模的各统计变量综合为博士规模因子和硕士规模因子两个综合变量；同时，通过对两个综合变量的排序，可大致看出1994年18所理工科大学研究生院培养博士生规模和培养硕士生规模的实力；通过聚类分析和绘制散点图，清楚地看出哪些大学研究生院的规模是属于同类大学.本研究得到了一些符合高等教育规律的结论，从而为科学评估我国理工科大学研究生院人才培养的规模以及评估研究生院整体条件和水平提供了一种新的途径.

致谢 本文研究过程中得到徐枞巍老师、杜智敏老师的帮助，特此致谢.

参考文献

1 陆叔云，范文曜.中国普通高等学校研究生院.北京：北京理工大学出版社，1995
2 方开泰.实用多元统计分析.上海：华东师范大学出版社，1986
3 王亚杰，刘 桔.研究生院评估的收获与不足.学位与研究生教育，1996，(3)：65～68

Application of Multivariate Analysis in Studying the Graduate Schools' Scales of Universities in China

Xu Huijuan
(Beijing University of Aeronautics and Astronautics, President's Office)
Wang Huiwen
(Beijing University of Aeronautics and Astronautics, School of Management)

Abstract Using part of the graduate schools' data in 1994 from 18 Chinese polytechnic universities with the State Council approved graduate schools, and by means of factor analysis and cluster analysis, two factors are extracted from the eight variables standing for the graduate schools' scales of the 18 Chinese polytechnic universities. One is related to the doctoral scales, the other is related to the master's scales. The graduate schools' scales of the 18 Chinese polytechnic universities in 1994 are rated and classified according to the two comprehensive factors' scores. The study method can be a reference for educational assessment on the graduate schools' scales of polytechnic universities.

Key words factor analysis; sequencing; cluster analysis

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net