

doi:10.16055/j.issn.1672-058X.2015.0011.018

客观聚类分析对区域基础教育发展水平评价

皮进修¹, 刘晓静²

(1.重庆师范大学 数学学院,重庆 401331;2.重庆师范大学 图书馆,重庆 401331)

摘要:首先,以国内外建立的教育指标体系为指导,结合我国年度统计用指标,选取 13 个指标并建立综合评价指标体系;其次,借助因子分析法将评价指标综合为具有代表性的影响因子,算得全国区域综合得分;最后,通过客观聚类分析对区域基础教育发展水平进行综合评价。

关键词:基础教育;因子分析;客观聚类分析

中图分类号:O212

文献标志码:A

文章编号:1672-058X(2015)11-0074-06

基础教育的发展一直受到社会各界的关注,如何衡量全国各区域基础教育的发展水平就显得尤为重要。对全国各个区域的基础教育发展水平作出评价,是一个系统性的工作,影响因素来自社会各个方面,如果全部考虑进去,显然不符合实际。因此,首要任务就是考虑主要影响因素,确定评价我国基础教育发展水平的指标体系。已发现经济合作与发展组织(简称经合组织)、欧盟以及世界银行建立了系统性的教育评价指标体系^[1]。国内也有大量学者对教育指标体系进行理论研究,陈学军^[2]介绍了经合组织教育指标体系概念框架及其内容的演变与发展;王燕^[3]揭示了经合组织教育指标体系的理论框架指标研制的过程与方法。但很少有学者对某一阶段性的教育发展状况构建评价指标体系并且进行实证研究。此处以国内外建立的教育指标体系为指导,并结合我国年度统计用指标体系,确定从区域的人口因素、经济因素、教育经费投入、师资力量、政治因素等中选取 13 个指标,建立综合评价指标体系。通过因子分析将各个指标综合,得到了影响区域基础教育发展主因子,结合我国统计年鉴数据,算得各省市的综合得分,得到初步评价结果;基于因子得分将全国各区域进行客观聚类分析,对各区域基础教育发展现状行深入的评价和分析。

1 区域基础教育发展水平评价模型的构建

1.1 评价指标的选取

根据国际教育经济学的研究以及 OECD, UNESCO, 世界银行和欧盟教育指标体系研究,区域基础教育发展水平受到人口、经济、教育经费投入、师资力量、政治等众多因素的影响。根据影响因素作用的特点,可将其分为内生因素与外生因素^[4]。在分析参考国内外评价指标体系,结合影响因素的作用机理,外生因素选取了人口因素、宏观经济因素;内生因素选取了总教育经费、生均教育经费、发展规模、师资力量、基础能力设施构建评价指标体系,如表 1 所示。

收稿日期:2015-03-31;修回日期:2015-05-11.

作者简介:皮进修(1989-),男,湖南常德人,硕士研究生,从事经济系统分析研究.

表 1 评价指标体系及其作用机理

	影响因素	符号	代表性因素	影响方式	评价目标
基础教育 发展 评价指 标体系	人口因素	X_1	文盲比重(X_{11})	间接影响	基础 教育 发展 水平
	宏观经济因素	X_2	人均 GDP(X_{21})		
			人均消费支出(X_{22})		
			人均可支配收入(X_{23})		
	总教育经费	X_3	教育经费占 GDP 的比重(X_{31}) 教育经费占政府财政支出的比重(X_{32})		
	生均教育经费	X_4	小学生均教育经费(X_{41})		
			初中生均教育经费(X_{42})		
发展规模	X_5	每万人中的小学生人数(X_{51})	直接影响		
		每万人中的中学生人数(X_{52})			
师资力量	X_6	小学生师比(X_{61})			
		中学生师比(X_{62})			
基础能力设施	X_7	教育固定资产投资(X_{71})			

1.2 数据来源

文中所涉及各项指标的具体数据均来源于《中国统计年鉴 2013》《中国教育经费统计年鉴 2013》中经网数据库以及 2013 年中国 31 省市的相关统计数据。

1.3 因子分析及主成分选取

1.3.1 因子相关性检验

因子分析的前提是变量间有较强的相关性.经计算结果 KMO 值为 0.706>0.5,并且巴特利特球形检验,近似卡方值为 552.023,显著水平值为 0.000<0.005,达到显著性水平要求,表明相关系数矩阵与单位矩阵有显著差异,即原始变量间可能存在有意义的关系,因此可进一步做因子分析.

1.3.2 因子分析过程和结果

运用统计软件 SPSS19.0,对标准化的数据进行处理得到指标相关系数矩阵,并且求得相关系数矩阵的特征值、方差贡献率以及累积方差贡献率.利用主成分分析方法^[5],可以发现前 3 个因子的方差贡献率较大,累积方差贡献率达到 85.574%>85%.因此根据主成分分析方法选取主成分的原则,选取前 3 个因子代表评价区域基础教育发展水平的 13 个指标的公共因子,是合适的.如表 2 是用 SPSS19.0 进行主成分分析,得到的主成分分析方差分解图.

表 2 主成分分析方差分解

因子	初始特征值			提取主成分		
	特征值	方差贡献率/%	累积贡献率/%	特征值	方差贡献率/%	累积贡献率/%
1	6.682	51.401	51.401	6.682	51.401	51.401
2	2.813	21.637	73.037	2.813	21.637	73.037
3	1.63	12.537	85.574	1.63	12.537	85.574

为确定每个公共因子的意义以便作出科学合理的分析,可以对因子载荷阵实施一次旋转变换^[6].旋转前,教育经费占 GDP 的比重(X_{31})、小学生师比(X_{61})同时在两个因子(因子 1 和因子 2)上的因子负荷超过了 0.5,可以考虑进行因子旋转.利用 Kaiser 标准化正交旋转法,对因子载荷矩阵进行旋转得到旋转后的因子载荷矩阵如表 3 所示。

表 3 旋转后因子载荷矩阵表

变 量	因子 1	因子 2	因子 3	变 量	因子 1	因子 2	因子 3
文盲率 X_{11}	-0.143	0.341	-0.758	中学生均教育经费 X_{42}	0.877	-0.313	-0.160
人均 GDP X_{21}	0.94	-0.177	0.215	每万人中的小学生人数 X_{51}	-0.543	0.771	-0.181
人均消费支出 X_{22}	0.907	-0.101	0.361	每万人中的中学生人数 X_{52}	-0.618	0.672	-0.188
人均可支配收入 X_{23}	-0.239	0.355	-0.858	小学生师比 X_{61}	-0.044	0.858	0.252
教育经费占 GDP 的比重 X_{31}	0.926	-0.109	0.260	中学生师比 X_{63}	-0.298	0.863	-0.120
财政教育经费比重 X_{32}	-0.009	0.333	0.749	教育固定资产投资 X_{71}	0.027	0.159	0.800
小学生均教育经费 X_{41}	0.883	-0.331	-0.27				

从表 2、表 3 知,第一、二、三主成分对区域基础教育发展状况的解释贡献度分别为 51.401%, 21.637%, 12.537%, 是所有主成分中最为重要的 3 个成分. 第一主成分主要与人均 GDP、人均消费支出、教育经费占 GDP 的比重、小学生均教育经费、中学生均教育经费等指标有较强的关系, 不难发现, 这些指标能从各个方面反映出区域经济实力状况, 可以将它称为经济实力因子; 第二主成分主要与每万人中的小学生人数、每万人中的中学生人数、小学生师比、中学生师比等指标有较强的关系, 而这些因素能反映出各区域基础教育普及情况以及学校师资力量配备情况, 因此可以间接得到基础教育在该地区的相对发展规模状况, 称为生师比因子; 第三主成分主要与人均可支配收入、财政教育经费比重、教育固定资产投资等指标有较强的关系, 从这些指标来看, 能够体现出各个区域对基础教育事业的投入状况, 因此可称其为社会投入因子.

根据 3 个主因子的方差贡献度(作为权重), 旋转后因子载荷矩阵表以及各项指标极差标准化后的数据, 利用主成分的计算公式可得各个区域基础教育发展状况的综合得分, 其计算公式如下:

$$P_i = 0.514 \times x_{i1} + 0.216 \times x_{i2} + 0.125 \times x_{i3} (i = 1, 2, \dots, 31) \quad (1)$$

为了更加直观地了解各区域基础教育发展状况的综合得分, 将式(1)的计算过程通过各区域基础教育主成分得分及综合得分排名表来反映(表 4).

表 4 区域基础教育主成分得分及综合得分

地 区	经济实力因子		生师比因子		社会投入因子		综合得分	排 名
	得分	排名	得分	排名	得分	排名		
上 海	2.490 3	2	-0.862 3	29	1.979 2	1	1.408 4	1
北 京	2.627 7	1	-0.827 1	28	1.168 2	6	1.388 9	2
浙 江	0.995 0	4	0.868 6	6	1.441 2	2	0.906 0	3
天 津	1.631 1	3	-0.511 4	21	-0.191 8	17	0.748 0	4
广 东	0.236 1	10	1.806 6	1	1.272 7	5	0.677 0	5
江 苏	0.933 0	5	0.678 0	9	-0.162 1	16	0.630 9	6
福 建	0.503 7	8	0.696 1	8	-0.224 1	18	0.394 9	7
山 东	0.213 7	11	1.467 1	3	-0.368 3	20	0.386 5	8
辽 宁	0.637 3	6	-0.016 5	17	-1.131 4	28	0.199 8	9
内 蒙 古	0.608 3	7	-0.616 8	24	-1.013 0	25	0.069 2	10
河 北	-0.194 0	17	0.727 2	7	-0.669 6	23	-0.031 6	11
湖 南	-0.142 1	16	0.479 3	10	-0.466 0	22	-0.031 6	12
湖 北	0.086 4	13	0.192 2	14	-1.153 8	29	-0.056 0	13
河 南	-1.078 7	29	1.773 3	2	0.968 5	7	-0.079 5	14
陕 西	-0.006 3	14	0.145 6	15	-1.030 4	26	-0.100 8	15

续表4

地 区	经济实力因子		生师比因子		社会投入因子		综合 得分	排 名
	得分	排名	得分	排名	得分	排名		
广 西	-0.751 9	23	0.988 0	4	0.605 5	10	-0.117 7	16
四 川	-0.391 3	21	0.407 8	11	0.027 5	13	-0.120 2	17
吉 林	0.453 0	9	-0.743 2	27	-1.882 3	31	-0.150 8	18
重 庆	-0.140 8	15	-0.395 6	19	-0.026 4	14	-0.164 9	19
安 徽	-0.333 7	19	0.240 0	12	-0.292 7	19	-0.165 3	20
江 西	-0.906 4	28	0.924 3	5	0.696 7	9	-0.203 6	21
山 西	-0.306 6	18	0.228 0	13	-1.032 4	27	-0.245 7	22
黑 龙 江	0.112 8	12	-0.702 4	25	-1.788 3	30	-0.314 3	23
云 南	-0.828 1	26	0.023 8	16	0.495 0	12	-0.381 0	24
海 南	-0.504 7	22	-0.592 7	23	-0.070 6	15	-0.409 9	25
新 疆	-0.348 9	20	-0.544 6	22	-0.856 9	24	-0.413 5	26
宁 夏	-0.787 4	24	-0.487 1	20	0.603 4	11	-0.455 8	27
青 海	-0.802 5	25	-1.274 0	30	0.777 2	8	-0.612 2	28
甘 肃	-0.873 3	27	-0.735 3	26	-0.370 6	21	-0.677 6	29
贵 州	-1.703 7	31	-0.098 9	18	1.324 3	4	-0.777 5	30
西 藏	-1.427 8	30	-3.238 3	31	1.371 1	3	-1.300 5	31

2 客观聚类分析过程和结果

2.1 客观聚类分析概念

Ivakhnenko 院士将数据分组处理方法理论的原理与核心概念应用于聚类分析研究,从而得到一种新的聚类分析方法——客观聚类分析方法^[7],贺昌政、张九龙、林嫔^[8]改进客观聚类分析法,使得新的客观聚类分析法能通过事先给定的一致性准则自动、客观地确定聚类个数并找到最优的聚类方案.客观聚类分析的基本思路:计算样本之间的距离,构建距离矩阵;利用偶极子将样本分组;在一致性准则下进行聚类分析.

2.2 产生偶极子及子集 A 与 B

将各区域的 3 个主成分得分作为客观聚类分析的 3 个指标集,并按综合得分排名对全国 31 个省市进行编号,比如编号 1 代表上海市,编号 21 代表江西.因为在计算综合得分时,各个因子得分的比重为其方差贡献度,因此,在计算样本间的距离时要对距离公式进行调整,并且因考虑了权重的影响,还需要对各个因子得分进行标准化再参与计算,于是得到调整后的距离如下:

$$d_{ij} = \sqrt{0.541 \times (x_{i1}^* - x_{j1}^*)^2 + 0.216 \times (x_{i2}^* - x_{j2}^*)^2 + 0.125 \times (x_{i3}^* - x_{j3}^*)^2} \quad (2)$$

其中, $i=1,2,\dots,31;j=i,i+1,\dots,31;x^*,y^*$ 表示标准化后的得分数据.

利用调整后的欧式空间距离公式以及表 4 中的数据计算得到各区域之间的距离并得到 496 个偶极子,将偶极子按其值从小到大排序,取前 15 个不重复的偶极子对,于是得到集合 A, B 如下所示:

偶极子:(29,28),(22,11),(23,16),(24,13),(10,9),(25,12),(18,15),(2,1),(19,17),(27,26),(4,3),(20,14),(7,5),(8,6),(31,30).

样本子集 A = {29,22,23,24,10,25,18,2,19,27,4,20,7,8,31}.

样本子集 B = {28,11,16,13,9,12,15,1,17,26,3,14,5,6,30}.

2.3 按子集 A, B 利用一致性准则 α_c 聚类

通过计算,得到各种聚类方案的一致性准则 α_c 的值,除了 $k=15, k=1$ 的情况, $k=4$ 时,即为如下聚类方案时, α_c 的值为 0,且在检验集 C 与 D 上聚类(除去 $k=15, k=1$), α_c 的值也为 0,因此可以判断聚为 4 类时为最佳聚类方案,其聚类方案如表 5 所示.

表 5 当 $k=4$ 时的聚类结果

子集合	样 本 编 号															
A	2	7	10	8	4	22	23	24	18	25	20	27	19	27	29	31
B	1	5	9	6	3	11	16	13	15	12	14	26	17	26	28	30

3 结论及建议

通过客观聚类分析,将全国各区域基础教育发展状况划分为 4 类,其结果及分析^[9]如下:第一类: {上海,北京};第二类: {福建,山东,天津,广东,辽宁,江苏,浙江,内蒙古};第三类: {河北,湖南,河南,湖北,广西,陕西,四川,安徽,重庆,吉林,江西,山西,黑龙江,云南,海南};第四类: {新疆,西藏,宁夏,青海,甘肃,贵州}.利用判别分析的方法,显然可以将江西省判到第三类中.第一类地区,基础教育发展水平在国内最好,但从生师比来看,存在教师资源利用不够合理现象.第二类地区,主要出于沿海地区,地理优势突出,经济发展水平均衡,对基础教育的发展起到了支撑作用,生均教育经费较高,教育资源利用合理;第三类地区,主要分布在中国的中部以及东北部,因为地理条件的影响,经济发展相对落后于第一、二类地区,从社会投入因子得分来看,这一类地区的排名总体比较靠后,这说明这些区域对基础教育的投入力度还有待提高.第四类地区的经济实力因子得分排在末尾,是导致这 6 个区域基础教育发展水平排在最后一类的最主要的原因;因子排名基本上是最后几名,这体现出我国中西部省份的教育资源、师资力量是比较欠缺的,有待进一步的改善和提升.综合因子分析及客观聚类分析的结果,可见影响区域基础教育发展水平的重要因子是区域的经济实力、教育资源以及社会投入,尤其是政府主导的社会投入.因此促进区域基础教育发展,可以从 3 方面着手:

(1) 加大对基础教育资源的投入,尤其是教育经费占政府财政支出低于 15% 的地区,比如黑龙江、吉林、青海等省市的均未达到要求,鼓励社会捐赠以及适当放宽对教育事业筹措资金的要求.

(2) 优化对基础教育资源的配置效率,建立完善的基础教育绩效评价制度以及各项评价指标,并将绩效目标的各指标具体化,政府要及时收集绩效信息,调整资源导向以及预算安排.

(3) 提高师资力量水平,鼓励优秀教师到基层执教,缩小城乡师资水平的差距,完善教师资格定期审查制度,加大对在职教师定期培训力度.

参考文献:

- [1] 张良才,孙继红.国内外教育指标体系分析与比较[J].教育学报,2009,12(6):60-68
- [2] 陈学军.经合组织教育指标体系概念框架及其内容的演变与发展[J].比较教育研究,2006(8):25-29
- [3] 王燕.经合组织教育指标研制:理论过程与方法[J].比较教育研究,2012(2):80-84
- [4] 杨大楷,王鹏.我国教育环境指标评价体系构建与测评[J].统计与决策,2014(12):27-30
- [5] 岳斯玮.基于主成分聚类分析对区域教育综合发展水平评价[J].西南民族大学学报:自然科学版,2012(7):37-43
- [6] 于秀林,任雪松.多元统计分析[M].北京:中国统计出版社,2011
- [7] IVAKHNENKO A G, MUELLER J A. Parametric and Nonparametric Selection Procedures in Experimental Systems Analysis[J]. Systems Analysis Modelling Simulation, 1992, 9(5): 157-175
- [8] 贺昌政,张九龙,林斌.基于数据分组处理方法的聚类分析模型[J].系统工程学报,2008,23(2):222-237

Research on the Development Level Evaluation of Regional Basic Education Based on Objective Cluster Analysis

PI Jin-xiu¹, LIU Xiao-jing²

(1.School of Mathematical Sciences, Chongqing Normal University, Chongqing 401331, China;

2.Library, Chongqing Normal University, Chongqing 401331, China)

Abstract: The development of basic education has attracted the attention of social circles, and how to measure the development level of the regional basic education is particularly important. Firstly, on the guidance of education index system at home and abroad, this paper selects 13 indexes and establishes the comprehensive evaluation index system combining China's annual statistical index; secondly, the evaluation indexes are synthesized into representative impact factors by factor analysis, and the comprehensive scores of national region are calculated; finally, the development level of the regional basic education is evaluated comprehensively by cluster analysis, Through the objective cluster analysis for the comprehensive evaluation of the level of development of regional basic education.

Key words: basic education; factor analysis; objective cluster analysis

(上接第 73 页)

- [10] 闻道君,万波.一类新的广义非凸变分不等式问题的近似解[J].云南大学学报:自然科学版,2014,36(1):1-5
- [11] WEN D J, CHEN Y A. Strong Convergence of Modified General Iterative Method for Generalized Equilibrium Problems and Fixed Point Problems of k -strict Pseudo-contractions[J]. Fixed Point Theory and Appl, 2012(2012): 125
- [12] KUROKAWA Y, TAKAHASHI W. Weak and Strong Convergence Theorems for Nonspreading Mappings in Hilbert Spaces[J]. Nonlinear Anal, 2010(73): 1562-1568
- [13] KANGTUNYAKARN A. The Methods for Variational Inequality Problems and Fixed Point of k -strictly Pseudononspreading Mapping[J]. Fixed Point Theory and Appl, 2013(2013): 171-175

On Fixed Point Problem of k -strictly Pseudononspreading Mapping

LUO Guang-yao, GONG Qian-fen

(1.School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing 400067, China; 2. School of Computer Science and Information Engineering, Chongqing Technology and Business University, Chongqing 400067, China)

Abstract: This paper introduces a class of new k -strictly pseudononspreading mappings with some examples in Hilbert space. Equivalence between the fixed point problem of strictly pseudononspreading mapping and variational inequality is established. This alternative equivalent formulation can be used to analyze some numerical methods for finding a fixed point of pseudononspreading mapping based on the current technique such as projection, resolvent operator and relaxed iteration.

Key words: nonspreading mapping; k -strictly pseudononspreading mapping; fixed point; variational inequality; Hilbert space