

基于因子和聚类分析的全国科普 示范县(市、区) 科普综合实力评价研究

张 艳 石顺科

(中国科普研究所, 北京 100081)

[摘 要] 如何对创建单位进行有效的管理是全国科普示范县(市、区)活动目前面临的问题之一。以 2011—2015 年全国科普示范县(市、区)为样本,设计评价指标,应用因子分析和聚类分析方法对我国 32 个省级行政单位的全国科普示范县(市、区)的科普综合实力进行分析,得出基于因子分析的全国科普示范县(市、区)科普综合实力评价体系和基于全国科普示范县(市、区)科普综合实力的分类标准。应用该结论能够对全国科普示范县(市、区)进行评价和分类,为全国科普示范县(市、区)管理提供一种可以量化的参考。

[关键词] 全国科普示范县(市、区) 因子分析 聚类分析 评价体系

[中图分类号] N4 **[文献标识码]** A **[文章编号]** 1673-8357(2012)03-0030-07

Assessing the Capability of National SP Exemplary Counties (Cities/Districts) by Using Clustering and Factor Analysis Methods

Zhang Yan Shi Shunke

(China Research Institute for Science Popularization, Beijing 100081)

Abstract: Based on the sampling data of those exemplary counties (cities/districts) analysis in 2011-2015, the paper starts from setting up a speculative evaluation framework. By using factor analysis and cluster analysis, this essay explores further on the general capability of those counties scattered across the country. The exploration yields as a result a possible practical evaluation framework for the assessment of those exemplary counties (cities/districts) and thereupon gives rise to a scaling assortment of their constructing condition. The conclusion may contribute to the assessment and classification of the National Science Popularization Exemplary Counties (Cities/Districts), which offers a quantifiable way for the management of the runners.

Keywords: National Science Popularization Exemplary Counties (Cities/Districts); factor analysis; clustering analysis; evaluation framework

CLC Numbers: N4 **Document Code:** A **Article ID:** 1673-8357(2012)03-0030-07

收稿日期: 2012-02-15

作者简介: 张 艳, 中国科普研究所博士后, 研究方向为基层科普, Email: geneclone@126.com;

石顺科, 译审, 中国科普研究所科普理论研究室主任, 研究方向为国外科普、基层科普、科普创作等,

Email: shishunke@yahoo.com.cn。

全国科普示范县(市、区)是中国科协在全国县域范围内发起的一项科普示范创建活动。活动以提高辖区内公民科学素质水平、服务当地经济社会发展为目的。自全国科普示范县(市、区)创建活动开展以来取得了良好的社会效益,创建规模也逐步扩大。目前实施方式是以5年为周期,创建和示范分阶段进行,创建成功后进入示范阶段。在2011—2015年新一轮创建中,创建单位达到902个县(市、区),几乎占到了我国县级行政区划单位的1/3。基数如此庞大,示范期内如何进行管理,如何使其充分发挥示范带动作用,成为全国科普示范县(市、区)发展中亟待解决的问题。

全国科普示范县(市、区)具有统分结合的特点。由于中国科协制定的创建标准^[4]只规定了最低的创建“门槛”,各省在此基础上又有各自的要求,这就使科普示范县(市、区)的科普综合实力受本地情况制约,表现出了层次性的特点,在管理上不能采取同等要求。因此,有重点、有针对性地在示范期促进全国科普示范县(市、区)发挥示范作用,需要客观、准确地分析各创建单位的科普状况。建立全国科普示范县(市、区)科普能力的评价体系,能够衡量各单位的发展水平,找出各单位科普实力档次,为分类管理、制定有效的激励手段提供数量依据。

本文以2011—2015年全国科普示范县(市、区)为样本,运用因子分析和聚类分析方法,建立全国科普示范县(市、区)科普综合实力评价体系^[1-3],并利用该体系分析全国科普示范县(市、区)的总体状况,划分全国的类别标准,为全国科普示范县(市、区)的管理提供参考。

1 全国科普示范县(市、区)科普综合实力评价体系的构建方法

评价体系根据指标权重确定方法的不同,可分为主观赋值法和客观赋值法。主观赋值法因由主观判断确定权重,更适用于无法采

集指标数据或有人为导向性时评价使用。客观赋值法是根据各指标间相关关系或各项指标值变异程度来确定权数,能够比较客观地反映各指标间的现实关系。

《全国科普示范县(市、区)创建工作手册》^[4]是全国科普示范县(市、区)活动的操作规范,其中“创建工作基本概况表”以数据填报的方式采集了县域科普创建工作的主要内容,可作为构建评价体系的数据来源。为了客观反映各指标之间的关系,本文采用客观赋值法构建全国科普示范县(市、区)评价体系。

评价体系的构建流程如下:基于2011—2015年全国科普示范县(市、区)申报材料^[5]中“创建工作基本概况表”的各项统计指标,对众多采集到的数据进行归类,从中提取体现全国科普示范县(市、区)科普综合实力的基本要素,理清各要素间的层次关系,筛选出适用的评价指标;运用因子分析法,对评价指标进行因子分析,最终得到反映包含全国科普示范县(市、区)科普综合实力主成分因子,进而形成综合评价体系。

全国科普示范县(市、区)数量较多,且各省之间的自设标准存在一定差异,但一省之内创建标准是统一的。因此利用各省创建指标的平均值来进行因子分析,能够全面地反映全国整体情况,同时也能避免抽样造成的误差和总体分析带来的繁重工作量。本文使用各省平均值来进行分析,建立评价体系,然后用聚类分析法对评价结果进行归类,从而得到全国科普示范县(市、区)总体综合实力类别分布状况。这就是因子分析和聚类分析的过程。基于此,建立起的评价体系可以对每个创建单位进行评价,并可根据评价结果对照评价标准进行分类,以方便管理。

1.1 因子分析法

因子分析^[6](Factor Analysis)是通过对变量的相关系数矩阵的内部结构进行分析,从中找出少数几个能够控制原始变量的因子,利用公共因子再现原始变量之间的相关关系,达到简

化变量,发现变量内在联系,对变量再解释及命名的目的。根据因子分析,可以得到指标间的内在联系;跟踪因子分析的计算过程,通过各指标的因子载荷,可以比较得到指标间的权重,从而建立起评价体系。

1.2 聚类分析法

聚类就是按照一定的类定义对研究对象分类^[2]。本文使用系统聚类方法中的最小距离法进行分析。假定原始数据中包括 m 个样本, p 个变量。先将每个样本看做一类,那么 m 个样本就分成 m 类;然后按照最小距离法度量样本之间的距离,并将距离最近的样本合并为一类,从而形成了 $m-1$ 类;最后计算出新产生的类别与其他各类别之间的距离,并将距离最近的两个类别合并为一类;继续重复这一步,直至所有的类别都合并为一类。

2 评价模型构建

2.1 评价指标的选取

根据 2011—2015 年全国科普示范县(市、区)申报材料,将各创建单位填报的“创建工作基本概况表”中与科普综合实力相关的统计指标归类合成 7 个评价指标:科普人员、科普组织、科普资源建设、科普经费、科普基础设施、科普活动、科普示范体系(如表 1)。使用评价指标进行因子分析。

2.2 数据来源和预处理

依据表 1 中的全国科普示范县(市、区)科普综合实力评价指标,计算 2011—2015 年各省全国科普示范县(市、区)创建单位统计指标平均数值。使用软件 SPSS16.0 将平均数值进行 Z-Score^[6]法标准化处理,然后等权重计算出 7 个评价指标数值(如表 2)。

2.3 使用因子分析构建评价体系

2.3.1 KMO 数据检验

Kaiser Meyer Olkin(抽样适度测定值,简称 KMO)统计量是用来判别是否适宜作因子分析的一个检验指标,其体现的是变量间简单相关系数和偏相关系数的相关性,其取值在 0~1 之间,KMO 值越接近于 1,表明原始变量相关性越强,越适宜作因子分析。运用 SPSS16.0 统计软件,对 32 个地区的数据进行分析可得 KMO 值为 0.617,表明各变量间有较强的相关性,可以作因子分析。

2.3.2 因子提取

主成分分析是因子分析中的一种提取方法。该方法不仅能够尽可能全面地反映出原有指标的信息量,又能解决指标之间信息重叠的问题,简化原指标体系。采用主成分分析法使用 SPSS16.0 软件对表 2 中各省评价指标的数值进行分析。分析得到因子的特征值、方差贡献率、累积贡献率(如表 3)。

表 1 全国科普示范县(市、区)科普综合实力评价指标

评价指标	统计指标	评价指标	统计指标
科普人员	科普专职人员	科普基础设施	县级综合性科普活动场所(个)
	科普兼职人员		乡镇(街道)级科普活动站(室)
	注册科普志愿者		村(社区)级科普活动站(室)
科普组织	乡镇科协数量		科普宣传栏(展示长度 5 延米以上)
	街道科协数量		省、市、县级青少年科普教育基地
	企业科协数量		年度工作主题科普活动次数
科普资源建设	科普广播栏目	科普活动	全国科普日重点活动项目数
	科普电视栏目		各类科普培训人数占县域常住人口比例
	科普报刊栏目	科普示范体系	科普示范基地个数(含省市县级)
	发放科普资料份数		科普示范乡镇个数(含省市县级)
科普经费	科普专项经费(万元)		科普示范街道(含省市县级)
	社会资金来源的科普经费(万元)		科普示范村(含省市县级)
			科普示范社区(含省市县级)

表 2 全国科普示范县(市、区)科普实力评价指标省份标准化处理值

序号	名称	科普人员	科普组织	科普资源建设	科普经费	科普基础设施	科普活动	科普示范体系
1	北京市	1.79	0.73	-0.78	1.43	2.02	0.66	0.83
2	天津市	-0.48	-0.61	0.84	-0.06	-0.22	0.15	0.47
3	河北省	-0.27	-0.21	-0.17	-0.28	-0.03	-0.42	-0.59
4	山西省	0.17	-0.15	-0.05	-0.35	-0.12	-0.23	0.26
5	内蒙古自治区	-0.53	-0.56	-0.50	-0.30	-0.72	-0.29	-0.65
6	辽宁省	-0.12	0.44	-0.40	-0.29	-0.22	-0.34	0.16
7	吉林省	0.70	0.16	-0.32	-0.33	0.07	-0.26	-0.18
8	黑龙江省	0.22	0.76	-0.58	-0.31	-0.38	-0.31	-0.23
9	上海市	1.00	-0.39	-0.62	2.09	1.39	0.01	0.57
10	江苏省	-0.42	0.69	0.30	2.93	0.55	-0.20	0.71
11	浙江省	-0.10	0.53	-0.01	0.07	-0.02	-0.22	0.07
12	安徽省	-0.40	-0.04	-0.42	-0.25	-0.34	-0.34	-0.30
13	福建省	0.12	0.45	0.85	-0.12	-0.14	-0.21	-0.22
14	江西省	1.03	0.03	-0.53	-0.34	-0.33	-0.41	-0.39
15	山东省	0.24	0.75	0.27	-0.08	1.52	-0.36	0.85
16	河南省	0.16	0.35	0.14	-0.01	0.40	-0.37	1.57
17	湖北省	-0.20	-0.04	0.34	-0.16	0.22	-0.28	0.00
18	湖南省	0.30	0.43	-0.17	-0.17	1.02	0.04	0.88
19	广东省	0.83	-0.36	-0.44	-0.31	-0.90	0.06	-0.41
20	广西壮族自治区	-0.62	0.21	0.19	-0.23	-0.02	-0.25	-0.03
21	海南省	-0.55	-0.81	-0.74	-0.28	-0.73	-0.08	-0.75
22	重庆市	-0.11	0.73	0.29	-0.12	0.42	-0.31	0.39
23	四川省	-0.41	0.98	0.40	-0.19	0.47	3.26	1.11
24	贵州省	-0.28	-0.01	-0.36	-0.25	-0.19	-0.35	0.01
25	云南省	0.09	0.41	-0.55	-0.25	-0.79	-0.31	-0.52
26	西藏自治区	-0.84	-1.64	-0.98	-0.39	-1.40	0.86	-1.27
27	陕西省	-0.32	-0.17	0.12	-0.24	-0.22	-0.28	-0.15
28	甘肃省	-0.17	0.19	0.63	-0.34	-0.36	0.17	-0.09
29	青海省	-0.58	-0.93	1.32	-0.15	0.06	-0.17	-0.64
30	宁夏回族自治区	-0.69	-0.57	0.55	-0.35	-0.31	-0.04	0.32
31	新疆维吾尔自治区	-0.42	-0.55	0.29	-0.19	-0.52	-0.22	-0.45
32	新疆生产建设兵团	0.82	-0.83	1.07	-0.18	-0.18	1.07	-1.33

表 3 各因子解释原始变量方差

因子	初始特征值			提取的因子特征值			旋转后的因子特征值		
	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(%)	特征值	方差贡献率(%)	累积贡献率(%)
1	2.825	40.364	40.364	2.825	40.364	40.364	2.066	29.518	29.518
2	1.288	18.397	58.761	1.288	18.397	58.761	1.523	21.753	51.271
3	0.942	13.455	72.216	0.942	13.455	72.216	1.244	17.773	69.044
4	0.804	11.482	83.699	0.804	11.482	83.699	1.026	14.655	83.699
5	0.656	9.376	93.075						
6	0.326	4.657	97.732						
7	0.159	2.268	100.000						

从表中看出,前4个因子的累积贡献率已达到83.699%,这4个因子包含的信息量已达到全部信息量的绝大部分,可提取前4个因子作为主成分,来反映全国科普示范县(市、区)的综合科普实力。

2.3.3 主成分因子的解释和命名

提取的主成分因子中所包含的信息需要解释和命名,才能表明其含义。使用SPSS16.0软件进行因子载荷旋转,使主成分因子的释义更加清晰。

根据表4,可将指标分成4类。因子1在指标“科普组织”和“科普示范体系”上有较大的载荷,这两个指标可以认为是反映科普组织系统的因子;因子2在指标“科普经费”、“科普基础设施”和“科普人员”上有较大的载荷,因子2可以认为是反映科普投入的因子;因子3在指标“科普资源建设”上有较大的载荷,可以认为是反映科普资源的因子;因子4在指标“科普活动”上有较大的载荷,可以认为是反映科普产出的因子。这4个因子概括地描述了全国科普示范县(市、区)的主要特征,可作为评价体系的一级指标。

表4 评价指标与主成分因子的相关系数
(旋转后的主成分因子载荷矩阵)

评价指标	主成分			
	因子1	因子2	因子3	因子4
科普人员	0.245	0.409	-0.663	0.126
科普组织	0.918	0.014	-0.131	-0.014
科普资源建设	0.089	0.086	0.876	0.138
科普经费	0.126	0.913	-0.060	-0.029
科普基础设施	0.649	0.649	-0.056	0.103
科普活动	0.040	0.008	0.057	0.987
科普示范体系	0.847	0.305	0.102	0.069

2.3.4 权重的获取及评价体系的建立

因子分析的指标权数是根据因子的方差贡献率计算而得到的,即各指标的权重由它对综合评价的贡献率确定。对表3中4个主成分因子方差贡献率,进行归一化处理,得到评价体系一级指标的权重。科普组织系统指标权重:科普投入指标权重:科普资源建设指标权重:科普活动指标权重=40.364%:18.397%:

13.455%:11.482%,通村标准化处理,得到各指标的权重分别为0.48,0.22,0.16,0.13。

每个主成分因子所解释的评价指标视为评价体系中一级指标下的二级指标。根据各主成分因子的方差贡献率(见表3)和二级指标与主成分因子的相关系数(见表4),计算二级指标(各因子包含指标)权重。

$$\text{二级指标权重} = \frac{\sum \text{指标在主成分因子上的载荷} \times \text{主成分因子方差贡献率}}{\sum \text{一级指标下各二级指标在主成分因子上的载荷} \times \text{主成分因子方差贡献率}}$$

以科普组织指标权重为例说明计算过程。

科普组织指标权重 = $\sum$ 科普组织在因子上的载荷 $\times$ 因子方差贡献率 / ($\sum$ 科普示范体系在因子上的载荷 $\times$ 因子方差贡献率 + $\sum$ 科普组织在因子上的载荷 $\times$ 因子方差贡献率) = $(40.364\% \times 0.918 + 18.397\% \times 0.014 + 13.455\% \times 0.131 + 11.482\% \times 0.014) \div (40.364\% \times 0.918 + 18.397\% \times 0.014 + 13.455\% \times 0.131 + 11.482\% \times 0.014 + 40.364\% \times 0.847 + 18.397\% \times 0.305 + 13.455\% \times 0.102 + 11.482\% \times 0.069) = 0.46$

基于此,得到全国科普示范县(市、区)科普实力的综合评价体系,见表5。

2.4 全国科普示范县(市、区)评价体系的应用及评价标准的划分

应用评价体系对全国32个省份全国科普示范县(市、区)科普综合实力进行评价,得到反映全国总体状况的评价结果。为了便于直观比较,把综合测评结果加上90进行坐标轴平移,使综合得分集中在80~100之间。各省的评价结果和排名见表6。

利用表6中的综合评价结果,通过SPSS16.0软件进行聚类分析,可得聚类图,见图1。

从图中看出,全国科普示范县(市、区)综合实力平均得分情况分成3个类别。

第一类:得分在91.554~90.450之间,包含四川省、江苏省、北京市、河南省、山东省、湖南省、上海市、重庆市8个地区,这些地区创建单位的平均科普综合实力状况处于同一水平,占全国科普示范县(市、区)总数的32.15%。

表 5 全国科普示范县（市、区）科普实力综合评价体系

一级指标	二级指标	三级指标
科普组织系统 (权重 0.48)	科协组织 (权重 0.46)	乡镇科协数量
		街道科协数量
		企业科协数量
	科普示范体系 (权重 0.54)	科普示范基地个数 (含省市县级)
		科普示范乡镇个数 (含省市县级)
		科普示范街道 (含省市县级)
		科普示范村 (含省市县级)
		科普示范社区 (含省市县级)
科普投入 (权重 0.22)	科普人员 (权重 0.14)	科普专职人员
		科普兼职人员
		注册科普志愿者
	科普基础设施 (权重 0.30)	县级综合性科普活动场所 (个)
		乡镇 (街道) 级科普活动站 (室)
		村 (社区) 级科普活动站 (室)
		科普宣传栏 (展示长度 5 延米以上)
		省、市、县级青少年科普教育基地
	科普经费 (权重 0.58)	科普专项经费 (万元)
		社会资金来源的科普经费 (万元)
科普活动 (权重 0.16)	科普活动 (权重 1)	年度工作主题科普活动次数
		全国科普日重点活动项目数
		各类科普培训人数占县域常住人口比例
科普资源 (权重 0.14)	科普资源 (权重 1)	科普广播栏目
		科普电视栏目
		科普报刊栏目
		发放科普资料份数

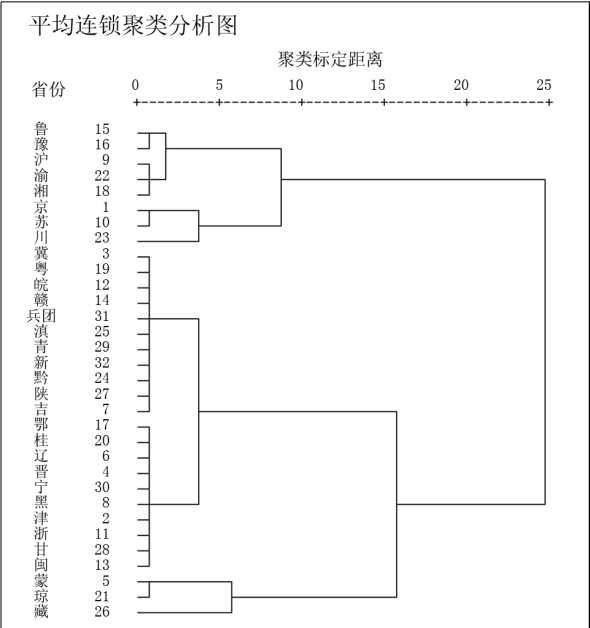


图 1 各省全国科普示范县（市、区）科普综合实力聚类结果

表 6 各省全国科普示范县（市、区）科普综合实力评价结果

省份	得分	省份	得分
北京市	91.037	上海市	90.455
天津市	90.173	江苏省	91.111
河北省	89.493	浙江省	90.179
山西省	89.926	安徽省	89.583
内蒙古自治区	89.199	福建省	90.243
辽宁省	89.968	江西省	89.590
吉林省	89.830	山东省	90.749
黑龙江省	89.884	河南省	90.749
湖北省	90.007	云南省	89.616
湖南省	90.542	西藏自治区	88.565
广东省	89.506	陕西省	89.779
广西壮族自治区	89.994	甘肃省	90.143
海南省	89.048	青海省	89.694
重庆市	90.450	宁夏回族自治区	89.938
四川省	91.554	新疆维吾尔自治区	89.558
贵州省	89.754	新疆生产建设兵团	89.685

第二类：得分在 90.243~89.493 之间，包含 21 个地区，占全国科普示范县（市、区）总数的 62.09%。由于这一类数量较多，为了更好地区别这一类创建单位的特征，根据聚类结果，这一类又可以分为两个层次。

第一层次得分在 90.243~89.884 之间，包含福建省、安徽省、天津市、甘肃省、湖北省、广西壮族自治区、辽宁省、宁夏回族自治区、山西省、黑龙江省 10 个地区，这些地区创建单位的平均科普综合实力状况处于第二类第一层次，占全国科普示范县（市、区）总数的 30.16%。

第二层次得分在 89.830~89.493 之间，包含吉林省、陕西省、贵州省、青海省、新疆生产建设兵团、云南省、江西省、安徽省、新疆维吾尔自治区、广东省、河北省 11 个地区，这些地区创建单位的平均科普综合实力状况处于第二类第二层次，占全国科普示范县（市、区）总数的 31.93%。

第三类：得分在 89.199~88.565 之间，包含内蒙古自治区、海南省、西藏自治区 3 个地区，这些地区创建单位的平均科普综合实力状况处于同一水平，占全国科普示范县（市、区）总数的 5.77%。

由此，通过省际间的全国科普示范县（市、区）科普综合实力评价结果分类，建立起 3 类得分的评价标准。当需要了解某全国科普示范县（市、区）的综合实力时，可以按照构建的评价体系，得出一个具体分值，与上述 3 类的分值区间对比，便可判断该全国科普示范县（市、区）综合实力状况处在的类别，为全国科普示范县（市、区）的分类管理提供了依据。

3 研究结论

第一，通过因子分析的方法建立了一套全国科普示范县（市、区）科普能力评价的指标体系，该体系严格按照统计口径建立，能够进

行定量分析，为全国科普示范县（市、区）的研究提供数量依据。

第二，通过因子分析理清了指标间的层次关系，并通过因子的贡献率区分出全国科普示范县（市、区）中科普工作的影响因素。其中科普组织体系是最重要的因素，这表明科普示范体系和科协组织系统对科普示范县（市、区）创建有较大的影响，应成为全国科普示范县（市、区）的工作重点；科普投入是次重要的因素，说明在全国科普示范县（市、区）中科普经费等投入有了硬性保证，不再是科普示范县（市、区）科普工作中的主要影响因素。但由于发展的不平衡，科普投入在有科普示范县（市、区）中仍需要加强。

第三，利用评价体系对全国科普示范县（市、区）进行评估，反映出全国科普示范县（市、区）科普实力的整体状况，在此基础上对各省样本进行聚类分析，划分出了在该体系下全国科普示范县（市、区）的 3 个类别，可作为衡量各创建单位的评价标准。

参考文献

- [1] 贾俊平. 统计学基础[M]. 北京：中国人民大学出版社，2010.
- [2] 张田，魏浩. 基于主成分分析法和聚类分析法的江苏省十三市科技服务业发展竞争力综合评价模型研究[EB/OL]. (2009-08-16) [2012-01-12]. <http://www.paper.edu.cn>.
- [3] 刘艳莉. 基于主成分分析和聚类分析的地区专利发展评价研究[J]. 科技管理研究，2008 (11)：80-85.
- [4] 中国科学技术协会. 全国科普示范县（市、区）创建工作手册（2011—2015 年）[R]. 北京：2009.
- [5] 档案编写组. 2011—2015 年度全国科普示范县（市、区）档案材料汇编[R]. 中国科协内部资料，2011.
- [6] 贾俊平. 统计学[M]. 第三版. 北京：中国人民大学出版社，2008.

(责任编辑 颜燕)