

# 基于改进主成分分析法的全国城市化水平研究\*

彭书凤, 田 盈

(重庆师范大学 数学与计算机科学学院, 重庆 400047)

**摘 要**: 从区域经济、人口、基础设施和生活方式等 4 个方面入手, 构建城市化水平评价指标体系。由于传统的主成分分析法未考虑原始评价指标对评价问题的重要性上的差异, 因此运用改进的主成分分析法对全国 31 个地区及城市城市化水平分析并排名。由改进的主成分分析法计算得到的结果看我国 31 个地区城市化水平有较大的差异, 沿海地区、珠江三角洲地区和长江流域地区发展水平较好, 西北部地区城市化进程则相对较慢, 阐述了具体地区的客观经济及人文环境造成此种差异的原因, 并对未来城市化发展方向讨论了一些改进思路。

**关键词** 城市化; 改进主成分分析法; 综合评价

中图分类号: F293

文献标识码: A

文章编号: 1672-6693(2007)03-0069-05

## Research into the Urbanization Levers of 31 Regions and Cities Based on the Developed Methods of Principal Component Analysis

*PENG Shu-feng, TIAN Ying*

(College of Mathematics and Computer Science, Chongqing Normal University, Chongqing 400047, China)

**Abstract**: This paper is to build the level of urbanization evaluation index system from the regional economy, population, infrastructure and lifestyle approach. Since traditional principal component analysis method does not consider the differences between original evaluation and the importance of the evaluation, it uses from the improved principal component methods to analysis the urbanization Levers of 31 Regions and Cities and ranks them. The concrete steps: standardized indicators of the original data; weight standardized data; the use of AHP to calculate the index weight; to get implant and eigenvector for the feature matrix; calculation of the composite ranking score; Finally, it can be seen that the level of urbanization in China is a greater difference. Coastal areas, Zhujiang River areas and Changjiang River areas develop better, northwest urbanization is slower. It presents the concrete reasons in objective economic and cultural environment for such differences and discusses the future development direction in those areas and gives some ideas for improvement.

**Key words**: urbanization; improved methods of principal component analysis; comprehensive evaluation

城市是人口和社会经济活动高度集中的场所, 是产业布局的特殊形态。它是工业、农业、交通运输业及商业布局等在空间上的聚合, 是地区经济的中心。近代以来, 工业化带动城市化, 城市发展带动区域发展, 构成了各国经济和社会发展的一大特征。城市化战略是我国城市发展的基本战略之一。因此, 分析和评价全国 31 个地区的城市化发展水平, 具有十分重要的意义。由于资源、环境、区位、基础、历史等多方面的原因, 这些地区的发展水平

差异不断扩大, 因此有必要对此进行分析比较, 从

\* 收稿日期 2006-12-31 修回日期 2007-04-27

资助项目: 中国博士后科学基金( No. 20060390570 )

作者简介: 彭书凤( 1982- ) 女, 重庆人, 硕士研究生, 研究方向为经济系统分析。

而为不同城市的区域发展战略和宏观调控区域经济提供依据和参考。从这一角度,本文构建了城市化水平指标体系,力求全面准确的衡量城市化水平。近年来,数理统计、运筹学、模糊数学、灰色系统、物元分析等理论和方法在系统综合评价中广泛应用,本文综合运用多元统计和运筹学理论较为成熟的主成分分析,并综合层次分析法对其改进,建立系统综合评价的改进的主成分分析法,给出了全国 31 地区城市化水平的评价实例。

### 1 城市化水平指标体系的构建

全面系统性原则:城市化过程是人口、经济、社会文化等多要素的系统转化过程,乡村型地域向城市型地域的转化,也有城市型地域的城市功能的进一步强化。层次性原则:众多城市化指标的性质和特点各异,反映的城市化内涵也不尽相同,应合理确定其在指标体系中的层次与位置。可比性原则:主要选取近些年我国统计部门稳定使用的指标,该指标能在资料中查到或通过有关资料可计算获得。针对性原则:城市化水平评价指标体系必须从城市化的本质出发,兼顾城市化的功能和影响,以期所筛选出的指标具有可行性和代表性<sup>[1]</sup>。

### 2 区域城市化指标体系

根据以上原则,本文采用复合指标法建立一套指标,从不同层次、方位来描述城市化的程度速度和分布<sup>[2]</sup>。总体可分为 4 大类,10 个具体指标(图 1)。筛选过程中,某些意义相仿、量化性差或不易获得的指标已被舍弃,所选指标均来自《中国统计年鉴(2005)》。

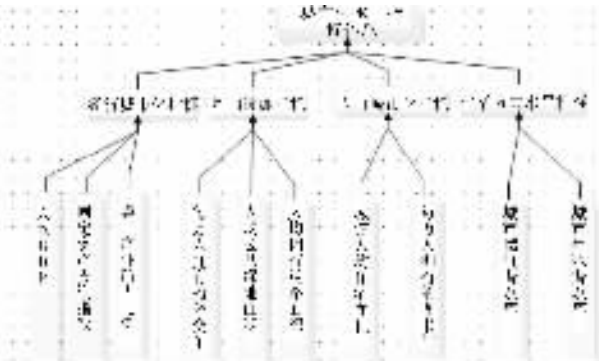


图 1 区域城市化指标体系

### 3 改进的主成分分析法

本文城市化水平的评价指标涉及区域经济、人口、设施和文化等各个方面,指标较多,但各指标之间并非独立,而是统一整体的不同部分,彼此之间具有一定程度的相关性,传统主成分分析法是利用原始指标数据所包含的方差信息量大小来对主成分分析施行客观赋权,没有考虑原始评价指标对评价问题的重要性上的差异,不能体现各指标之间在社会环境上的差异。因此,本文采用改进的主成分分析法,力求系统模型与实际环境相适应,更科学化并容易接受。改进的主成分分析方法如下。

#### 3.1 标准化原始数据

设要做主成分分析的目标为  $X_1, X_2, X_3, \dots, X_n$  作主成分分析,首先要对原变量做标准化处理,  $X_i = (X_{ij} - \bar{X}_i) / S_i, i = 1, 2, 3, \dots, n$ , 其中  $\bar{X}_i, S_i$  分别为第  $i$  个指标所有样本的均值和均方差,本文直接利用 Matlab 软件对全国 31 个地区各指标的原始数据(均来自《中国统计年鉴 2005》)进行标准化处理<sup>[3]</sup>。

#### 3.2 运用层次分析法求出各指标的权重

1)建立梯阶层次结构体系。采用层次分析法,根据影响城市化水平评价指标体系的主要因素建立梯阶层次分析结构体系(图 1)<sup>[4]</sup>。

2)构造判断矩阵。在建立梯阶层次结构体系后,根据上下层次间的隶属关系构造判断矩阵<sup>[4]</sup>。

3)计算各层次中因素的权重。根据判断矩阵提供的信息,用幂法求解得到任意精度的最大特征根和特征向量,特征向量代表该层次各因素对上层某因素影响大小的权重<sup>[4]</sup>。

4)检验。在两两比较判断矩阵中,为计算权重系数,必须保持判断思维的一致性,避免思维逻辑颠倒,以免各判断值出现  $a > b, b > c, c > a$  等混乱现象,这就必须对两两比较判断矩阵进行检验<sup>[4]</sup>。

#### 3.3 对标准化数据加权处理

由上两步已经得到的标准化数据和各指标的实际归一化权重  $w_1, w_2, w_3, \dots, w_n$ , 将各标准化数据进行加权处理,常用方法为各标准化数据乘以此数据对应的权重  $w_1, w_2, w_3, \dots, w_n$  或乘以改进的权重  $1 + w_1, 1 + w_2, \dots, 1 + w_n$ <sup>[5]</sup>。本文采用第二种方法对标准化数据进行加权。设对原始数据标准化处理的数据阵是  $X^* = (X_{ij}^*) = (X_1^*, X_2^*, \dots, X_n^*)$ , 由此构造新数据  $X' = (X_{ij}') = [(1 + w_1)X_1^* (1 + w_2) \cdot X_2^* \dots (1 + w_n)X_n^*]$  则  $\text{var}[(1 + w_j)X_j^*] = (1 +$

$w_j) \text{var}(X_j^*)^{[6]}$ 。使用此方法被赋予更大权数, 在评价系统中较为重要的变量得到更多重视, 将主客观赋权有机结合起来, 使评价结果更符合综合评价问题的目标和实际。

### 3.4 求目标矩阵的特征值与特征向量

运用数学统计软件 Matlab, SPSS 等将经过第三步处理的矩阵输入, 可直接求出目标矩阵的特征值与特征向量, 这里就不再详细说明。

### 3.5 计算综合排名得分

求出各特征值的贡献率, 提出包含原始信息 85% 以上的主成分因子, 由主成分因子计算综合排名得分。

## 4 全国 31 个地区城市化水平综合评价

1) 由目标体系从《中国统计年鉴 2005》上得到对应各城市的原始数据, 利用 Matlab 软件对其进行标准化处理, 得到标准化数据。

2) 城市化水平是一个比较复杂的系统, 对它的评价需要考虑许多因素, 对城市化水平目标体系运用层次分析法对人们的主观判断作客观描述, 它把人们对复杂问题的决策思维过程层次化和数学化, 通过各个因素的比较计算, 得出各个指标的权重。本文运用 Matlab 软件设计建立判断矩阵, 并对层次总排序进行一致性检验, 得出目标体系权重, 见表 1。

表 1 城市化水平指标权重

指标	人均 GDP	固定资 产投资	第三产业 所占比例	万人拥有公 共交通车	人均公共 绿地面积	人均拥有 道路面积	人均住宅 建筑面积	每万人拥 有图书馆	城市用水 普及率	城市燃气 普及率
	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$
$w_i$	0.1711	0.1401	0.1106	0.0590	0.0716	0.05811	0.1102	0.0850	0.1038	0.0904

3) 当标准化完成后, 将原变量数据作加权处理, 即  $X = (1 + w_i)X_i$ , 见表 2, 使运算数据准确并符合实际。

4) 借助 Matlab 软件对加权后的标准化数据进行主成分分析运算, 得到该指标数据的特征值与特征向量(表 2、3)。计算各特征值的贡献率, 按总贡献率大于 85% 提取主成分。

5) 每个主成分  $F_j$  是  $n$  个  $X'_i$  的线性组合  $F_j = \sum_{i=1}^n C_{ij}X'_i$   $j = 1, 2, 3, \dots, n$ ,  $F_j$  称为第  $j$  个主成分,

$C_{ij}$  为第  $j$  个主成分的  $X'_i$  的系数, 所以求解主成分表达式就是要求出  $C_{ij}$  个元素。按特征值大于 1 的原则提取 4 个主成分:  $F_7, F_8, F_9, F_{10}$ (表 2)。提取的 4 个特征根的值分别为 2.9980, 3.8148, 2.5990, 4.7189, 各成分的贡献率分别为 26.9017, 21.7475, 17.0911, 14.8165, 累积贡献率为 87.1466, 即此 4 个主成分包含了原始信息的 87.15%; 根据各主成分因子的贡献率及在各城市上的得分可算出各城市综合得分及排名状况(表 4)。

表 2 特征值及贡献率

指标	人均住宅 建筑面积	城市用水 普及率	城市燃气 普及率	每万人拥 有公交车	人均拥有 道路面积	人均公共 绿地面积	每万人拥 有图书馆	固定资产 投资	人均 GDP	第三产业 比例
	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$
特征值	0.3342	0.4851	0.4017	1.3020	0.3921	0.4955	2.9980	3.8148	2.5990	4.7189
贡献率	0.0191	0.0277	0.0229	0.0742	0.0224	0.0282	0.1709	0.2175	0.1482	0.26897

表 3 相关矩阵特征值

$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$
0.3957	-0.2801	0.0405	-0.2239	0.5690	0.3477	-0.1193	-0.4493	0.1174	0.0707
0.3139	-0.1131	0.0468	0.1465	-0.0997	0.5133	0.0280	0.6510	0.4432	-0.0840
0.3923	-0.2204	0.0109	0.1034	0.1950	-0.7723	-0.1575	0.2207	0.2869	-0.0806
0.0880	-0.1979	0.5897	-0.2218	-0.0149	-0.0515	0.0603	0.2284	-0.3302	0.5653
0.1771	-0.0472	0.0534	-0.4853	0.0822	-0.0092	0.2002	0.2479	-0.4206	-0.6715
0.3175	-0.0489	-0.1372	0.3580	0.0574	-0.0336	0.8327	-0.1882	-0.1236	0.1027
0.0236	-0.1822	0.3260	0.6934	0.0451	0.1077	-0.2782	-0.0678	-0.3743	-0.3306

0.4165	-0.3561	-0.2199	-0.1002	-0.7348	0.0257	-0.2125	-0.2547	-0.1570	0.0619
0.5226	0.7996	0.1794	0.0162	-0.0385	0.0060	-0.1664	-0.0727	-0.0726	0.0438
0.0626	0.1304	-0.6632	0.1168	0.2732	0.0549	-0.2700	0.3192	-0.4860	0.2922

$$F_7 = -0.4493X'_1 + 0.6510X'_2 + 0.2207X'_3 + 0.2284X'_4 + 0.2479X'_5 - 0.1882X'_6 - 0.0678X'_7 - 0.2547X'_8 - 0.0727X'_9 + 0.3192X'_{10}$$

$$F_8 = -0.4493X'_1 + 0.6510X'_2 + 0.2207X'_3 + 0.2284X'_4 + 0.2479X'_5 + 0.1882X'_6 - 0.0678X'_7 - 0.2547X'_8 - 0.0727X'_9 - 0.4860X'_{10}$$

$$F_9 = 0.1174X'_1 + 0.4432X'_2 + 0.2869X'_3 - 0.3302X'_4 - 0.4206X'_5 - 0.1236X'_6 - 0.3743X'_7 - 0.1570X'_8 - 0.0726X'_9 - 0.4860X'_{10}$$

$$F_{10} = 0.0707X'_1 + 0.0840X'_2 + 0.0806X'_3 + 0.5653X'_4 + 0.6745X'_5 + 0.1027X'_6 + 0.3306X'_7 + 0.0619X'_8 + 0.0438X'_9 + 0.2922X'_{10}$$

表 4 各地区城市化得分及排名

城市	北京	天津	河北	山西	内蒙古	辽宁	吉林	黑龙江	上海	江苏	浙江
得分	0.2455	0.2073	0.1138	-0.0629	-0.3529	0.0686	-0.0536	-0.1131	0.3529	0.1063	0.3173
排名	3	5	9	22	30	15	21	23	1	11	2
城市	安徽	福建	江西	山东	河南	湖北	湖南	广东	广西	海南	重庆
得分	0.0388	0.0691	0.1064	0.1447	0.1011	0.0803	0.1924	0.2222	-0.2547	0.1386	0.0687
排名	19	14	10	7	12	13	6	4	27	8	16
城市	四川	贵州	云南	西藏	陕西	甘肃	青海	宁夏	新疆		
得分	-0.0173	-0.1775	0.0642	-0.3357	0.0630	-0.3861	-0.1377	-0.2628	-0.2449		

注:排名顺序按《中国统计年鉴 2005》

## 5 各城市地区城市化水平分析

由主成分分析的结果可以看出这 31 个地区综合城市化水平的高低。由表 5 可以看出我国 31 个地区城市化水平有较大的差异,沿海地区、珠江三角洲和长江流域发展水平较好,它们在政治、经济、文化、历史、地理和人文素质上都是具有较大的优势,客观上比中小城市具有更强的集聚力和扩散力<sup>[3]</sup>。这些地区在我国经济发展过程中一直是周边地区经济和社会发展的“引路者”和重要动力,其发展水平的高低、经济实力的大小、辐射和扩散能力的强弱对所在省及其周边地区各个方面都有重大影响。当今时代,城市化水平的高低已经成为衡量一个国家地区现代化程度的重要标志之一。从总体上看,我国城市化仍严重滞后于工业化和经济发展水平。因此在未来一个时期内城市化还将有一个大的发展,如何加速城市化的进程应成为区域经济发展的战略重点。从表 4 可以看出,综合得分较高的有上海、浙江、北京、广东、天津,其余地区的综合得分都较低,因此,中国城市化发展战略首先要加大大城市建设,充分发挥大城市的优势,促进区域经济的协调发展。大城市在政治、经济、文化、历史、地理和人文素质上都具有更强的集聚力和扩散力。大城市往往是所在区域的政治、经济、文化、商贸和信息中心,具有较完备的基础设施和公共服务设施,对区域经济的发展起着重要的作用。发挥大城市已有和潜在的综合优势,促进区域经济的协调发展,符合资源配置的经济规律<sup>[8]</sup>。盲目严格控制大城市的发展,既不利于其综合优势的利用,也不利于区域内中小城市的发展。庞大的经济区域缺少大城市或特大城

市的带动,这是我国城市经济发展中的一大缺陷。例如福建省和广东省相比,虽然同是最早改革开放的省份,但福建发展水平和速度明显不如广东省,这一方面除了区位条件、交通条件、政策等原因外,与福建缺乏大城市和特大城市的带动明显有关。作为福建省的龙头城市福州和厦门的城市规模和经济总量都明显偏小,对全省经济的带动能力偏弱。当然在发展大城市的同时,还应强化中小城市的功能和积极发展小城镇。

西北部地区城市化进程则相对较慢,但在西部各省、市、自治区中,西部省区位次只有陕西(19位)、重庆(16位)、云南(17位)、陕西(18位)、四川(20位)位次较前,而新疆(26位)、广西(27位)、宁夏(28位)、西藏(29位)、甘肃(31位)却排在最后6位且综合得分很低。这与各地区经济、历史原因、社会、人口及其它因素密切相关,是整体一致发展的过程。因此,加大西部各省投资力度,带动西部发展西部大开发战略是在近几年国内外经济形势和环境出现了新特点、新变化的背景下提出的。加快西部地区发展,不能齐头并进,同时起步,而是要鼓励有条件的地方率先突破,选择一部分地区重点开发。要充分发挥西部中心城市经济实力较强、投资环境较好、人力资源比较集中的优势,优先开发,培育其辐射和扩散功能,使这些城市成为西部大开发的依托和支撑点,成为对外开放的窗口和示范区,成为发展高技术和技术创新的基地和策源地。因此,今后 10 年开发的重点,应是加强一些中心地区城市城市化建设,使之成为经济增长的极核,从而带动整个区域的发展。例如,通过西安(陕西)的建设可以带动关中地区,重庆的建设可以带动四川盆地,乌鲁木齐(新疆)的建设可以带动新疆天山北

坡地带的城市化建设。

中国在客观上正逐渐形成长江三角洲地区、珠江三角洲地区、武汉经济区、成渝经济区、关中经济区、辽中南经济区、山东半岛经济区、北部湾经济区和天山北部经济区<sup>[5]</sup>,而各个经济区都必须以一定的中心地区(城市)为依托。加大全国中心地区(城市)的城市化开发开放力度,使之成为全国或地区有引导作用的经济增长极,对带动全国或区域经济的发展具有重要意义。

#### 参考文献:

- [1] 陈彭述. 城市化与城市地理信息系统[M]. 北京: 科学出版社, 2001.
- [2] 许学强,周一星,宁越敏. 城市地理学[M]. 北京: 高等

教育出版社, 2001.

- [3] 陈胜昌. 层次分析法与城市化进程[J]. 城市发展研究, 2000(3): 90-94.
- [4] 刘利华. 主成分分析法在粮食产量影响因素分析中的应用[J]. 重庆师范大学学报, 2000(5): 36-39.
- [5] 王圆圆. 安徽城市效率与对策[J]. 地域研究与开发, 2004, 23(2): 34-37.
- [6] 任俞. 城市化水平综合测度模型研究[J]. 四川师范大学学报, 2001(2): 78-81.
- [7] 房镜. 主成份分析在城市综合实力评价中的应用[J]. 西安工程科技学院学报, 2006(4): 243-246.
- [8] 蔡俊豪, 陈兴渝. “城市化”本质含义的再认识[J]. 城市发展研究, 1999(5): 22-23.

(责任编辑 欧红叶)

## 研究快讯

# 大学生经济自立词汇内隐观的初步探讨\*

李祚山<sup>1,2</sup>, 王丽平<sup>1</sup>, 刘晶<sup>1</sup>

(1. 重庆师范大学应用心理学重点实验室, 重庆 400047; 2. 西南大学心理学院, 重庆 400715)

关键词: 大学生, 经济自立, 内隐观

中图分类号: B848.2; B844.2

文献标识码: B

文章编号: 1672-6693(2007)03-0073-01

本研究以人格词汇学研究和内隐观人格假设为方法学基础,旨在研究大学生经济自立词汇内隐观的结构和特征,为建立经济自立的人格理论提供相应的前期研究依据。研究分两阶段进行,第一阶段要求大学生根据自己的理解,尽可能多地使用形容词来描述经济自立的大学生具有的心理和行为特征,通过对127名大学生进行预调查,按照形容词频数出现的高低编制出形容词调查表;第二阶段使用编制出的形容词调查表,要求被试按照7级评分制

对符合自己的程度进行评价,对414名大学生进行正式调查。结果表明(1)大学生认为经济自立重要性的前10项主要以积极意志品质的词为主,符合程度前10项则以描述伦理和目标的词为主(2)大学生经济自立的结构可以包含5个因子:成就能动性、社交性、独立性、个人中心性、俭朴性(3)性别、年级、自评家庭经济条件三者之间的主效应和交互作用效应在各因子上均不显著。

## A Preliminary Study of the Implicit Verbal Theory of Undergraduates' Economic Self-Supporting

LI Zuo-shan<sup>1,2</sup>, WANG Li-ping<sup>1</sup>, LIU Jing<sup>1</sup>

(1. Key Applied Psychology Lab, Chongqing Normal University, Chongqing 400047;

2. School of Psychology, Southwest University, Chongqing 400715, China)

**Abstract:** Based on the verbal theory of personality and the implicit personality theory, in order to study the structures and features on implicit verbal theory of undergraduates, economic self-supporting, 127 undergraduates were pretested, and 414 undergraduates were

\* 收稿日期: 2007-05-16

资助项目: 教育部社科规划课题(No. 01JAXLX002)

作者简介: 李祚山(1969-)男,四川大竹人,教授,博士研究生,研究方向为人格过程与心理健康。

tested by formal verbal questionnaires. The results show that (1) the former 10 adjectives which are of the most importance of self-supporting that undergraduates' thought are all positive, self-belief, and full of willpower; and the former 10 adjectives, according to the conformity, are affiliated with ethic and aims; (2) undergraduates' economic self-supporting structures consist of 5 factors, i. e. achievement-activity, sociability, independence, individualism, thrift; (3) there are not significant differences in main effect and interactive effect between sex and grades, sex and self-judged family's economic condition, or grades and self-judged family's economic condition.

**Key words** : undergraduates ; economic self-supporting ; implicit theory

(责任编辑 李若溪)