

# 中国各省市投资环境的统计分析\*

张雅清

(重庆师范大学 数学与计算机科学学院, 重庆 400047)

**摘要** 综合利用主成分分析法和因子分析法两种评价方法,对我国31个省市、自治区的投资环境进行排序和分类,为规划各地区的投资环境提供了一定的科学依据。两种评价方法的综合运用克服了单一评价方法的片面性,评价结果较为全面客观。

**关键词** 投资环境;主成分分析;因子分析;Wilcoxon符号秩检验

中图分类号:O212.4

文献标识码:A

文章编号:1672-6693(2006)01-0067-04

## Statistical Analysis on the Investment Environment in Various Regions of China

ZHANG Ya-qing

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

**Abstract** This paper gives an order and a classification for the investment environments in various regions of China, comprehensively utilizing the principle component analysis and factor analysis. It provides some scientific basis for programming the investment environment of each region. The comprehensive application of two assessing methods eliminates the one-sidedness of single assessing method, and the results are more comprehensive and objective.

**Key words** investment environment; principle components analysis; factor analysis; Wilcoxon signed ranks test

中国是一个地域辽阔、人口众多的国家,各省市、自治区的投资环境存在较大差异。本文运用比较科学的统计方法,对我国31个省市、自治区的投资环境进行排序和分类,从分类结果中不仅可以了解我国各地区投资环境的状况,而且为优化各地区的投资环境提供了一定的科学依据。

### 1 指标的选择

投资环境,一般是指投资所处的自然、技术、经济和社会的各种条件。它包括地理位置、资源条件、基础设施、工资水平、技术水平、原料供应、市场情况、信息渠道、纳税负担、资金融通、投资政策、法律保障、政治局势、社会秩序等<sup>[1]</sup>。

本文遵循系统性、科学性、可操作性的原则,根据中国统计年鉴——2004的统计数据<sup>[2]</sup>(原始数据略)选取了具有代表性的5类23个指标<sup>[1,3-4]</sup>。

(1)经济水平和市场潜力:国民生产总值(亿元),人均国民生产总值(元),第二产业比重(%),第三产业比重(%),固定资产投资额(亿元),城镇居民人均全年消费性支出(元),农村居民人均全年

消费性支出(元),城镇居民人均可支配收入(元),农村居民平均纯收入(元),社会消费品零售总额(亿元)。

(2)交通通讯条件:货运周转量(亿吨公里),客运周转量(亿人公里),人均邮电业务总量(元),本地电话年末用户(万户),长途电话数(万次)。

(3)经济外向度:外商直接投资额(万美元),实际利用外资金额(亿元)。

(4)人口素质:每万人中高校在校人数(人),每万人中专业技术人员数(人)。

(5)能源供应和环境质量:城市自来水普及率(%),城市燃气普及率(%),人均公共绿地面积(平方米),工业废水排放达标率(%)。

### 2 分析方法

#### 2.1 主成分分析法

主成分分析法就是设法将原来的多个指标重新组合成一种新的互相无关的几个综合指标,同时根据实际需要从中选取几个较少的综合指标,尽可能多地反映原来指标的信息,它是一种降维的统计分

\* 收稿日期 2005-06-13 修回日期 2005-09-08

作者简介:张雅清(1982-)女,山西太谷人,硕士研究生,研究方向为随机经济系统分析。

析方法<sup>[5]</sup>。

本文采用主成分分析法对投资环境做了评价(利用 SPSS11.5<sup>[6]</sup>在计算机上完成其计算过程)具体步骤如下。

(1)原始数据标准化,消除量纲的影响。

(2)建立指标之间的相关系数矩阵,并求其特征值和特征向量。按照特征根大于 1.00 的原则,选取前 4 个主成分(见表 1),其方差累计贡献率为 85.969%,即反映了原始指标中 85.969% 的信息,将这 4 个主成分作为评价投资环境的综合评价指标。由于每个评价指标在各个主成分上的载荷不是很明确,所以用因子分析法来确定其经济意义。这 4 个特征值相应的特征向量如表 2 所示。

表 1 特征值

主成分	特征值		
	特征根	方差贡献率	累计贡献率
1	11.664	50.713	50.713
2	5.268	22.904	73.617
3	1.752	7.617	81.234
4	1.089	4.735	85.969

表 2 相应的特征向量

指标	成分			
	1	2	3	4
国民生产总值	0.787	-0.583	-0.047	-0.017
人均国民生产总值	0.816	0.510	0.001	-0.212
第二产业比重	0.412	-0.399	0.510	-0.484
第三产业比重	0.309	0.745	-0.442	0.178
固定资产投资额	0.747	-0.577	-0.079	-0.060
城镇居民人均全年消费性支出	0.763	0.444	-0.416	0.000
农村居民人均全年消费性支出	0.853	0.350	-0.097	-0.077
城镇居民人均可支配收入	0.848	0.371	-0.257	-0.022
农村居民平均纯收入	0.910	0.341	-0.018	-0.084
社会消费品零售总额	0.817	-0.516	-0.099	0.039
货运周转量	0.670	0.129	0.126	-0.527
客运周转量	0.482	-0.780	-0.022	0.208
人均邮电业务总量	0.806	0.518	-0.073	0.035
本地电话年末用户	0.716	-0.662	-0.008	0.068
长途电话数	0.793	-0.484	-0.160	0.149
外商直接投资额	0.835	-0.321	-0.194	-0.083
实际利用外资金额	0.797	-0.369	-0.291	-0.051
每万人中高校在校人数	0.616	0.631	0.230	-0.061
每万人中专业技术人员数	0.263	0.679	0.331	-0.042
城市自来水普及率	0.547	0.229	0.376	0.374
城市燃气普及率	0.830	0.181	0.280	0.197
人均公共绿地面积	0.677	0.131	0.364	0.460
工业废水排放达标率	0.573	-0.127	0.587	0.070

(3)得出主成分表达式,并利用这一线性组合计算出各主成分值。由表 2 可得

$$F_1 = 0.787X_1 + 0.816X_2 + 0.412X_3 + 0.309X_4 + \dots + 0.573X_{23}$$

其中  $F_1$  表示第 1 主成分,  $X_i$  表示第  $i$  个指标 ( $i = 1, 2, \dots, 23$ )。同理可得其它 3 个主成分  $F_2, F_3, F_4$  的表达式。用  $F_1, F_2, F_3, F_4$  来替代原来的 23 个指标,对各省、自治区的投资环境进行分析。为此,可求出各省、自治区在各个主成分上的得分。

(4)求综合得分。由于在第一步已经对数据进行了标准化处理,消除了量纲的影响,所以能够以各主成分的方差贡献率占 4 个主成分总方差贡献率的比重作为权重进行加权汇总,得出各省、自治区的主成分综合得分  $F$ (见表 3),即

$$F = \frac{50.713F_1 + 22.904F_2 + 7.617F_3 + 4.735F_4}{85.969}$$

其中  $F_1, F_2, F_3, F_4$  分别为主成分 1、主成分 2、主成分 3、主成分 4 的得分。

(5)按主成分综合得分排序(见表 3)。

表 3 得分及排名

地区	主成分得分	主成分排名	因子得分	因子排名	综合得分	综合排名
北京	1.91	1	0.26	6	1.65	3
天津	0.98	3	-0.17	15	0.61	7
河北	0.15	9	0.2	7	0.26	10
山西	-0.36	23	-0.45	24	-0.61	25
内蒙古	-0.34	21	-0.44	23	-0.59	24
辽宁	0.23	8	0.16	9	0.29	9
吉林	-0.12	12	-0.39	21	-0.39	18
黑龙江	-0.14	13	-0.32	19	-0.35	16
上海	1.67	2	0.9	4	1.96	2
江苏	0.49	6	1.34	2	1.39	5
浙江	0.9	4	1.16	3	1.57	4
安徽	-0.44	25	-0.15	14	-0.45	19
福建	0.37	7	0.19	8	0.42	8
江西	-0.18	17	-0.22	16	-0.31	15
山东	0.01	10	0.88	5	0.67	6
河南	-0.28	20	0.15	10	-0.1	12
湖北	-0.26	19	0.06	12	-0.15	13
湖南	-0.24	18	0.04	13	-0.15	14
广东	0.76	5	2.09	1	2.18	1
广西	-0.56	28	-0.26	17	-0.63	26
海南	-0.17	15	-0.5	25	-0.52	21
重庆	-0.46	26	-0.29	18	-0.57	22
四川	-0.14	14	0.07	11	-0.05	11
贵州	-0.69	30	-0.53	27	-0.93	29

续表 3

地区	主成分得分	主成分排名	因子得分	因子排名	综合得分	综合排名
云南	-0.37	24	-0.39	22	-0.58	23
西藏	-0.97	31	-0.5	26	-1.12	31
陕西	-0.17	16	-0.33	20	-0.38	17
甘肃	-0.55	27	-0.56	29	-0.84	28
青海	-0.34	22	-0.72	30	-0.81	27
宁夏	-0.61	29	-0.73	31	-1.02	30
新疆	-0.06	11	-0.54	28	-0.45	20

## 2.2 因子分析法

因子分析法也是一种降维的统计分析方法。其基本思想是通过变量的相关系数矩阵内部结构的研究,找出能控制所有变量的少数几个因子去描述多个变量之间的相关关系,再根据相关性的大小把变量分组,使得同组内变量之间的相关性较高,但不同组的相关性较低。

本文利用因子分析法<sup>[7]</sup>对投资环境进行分析,其步骤如下。

首先,与主成分分析法相同,对原始数据进行标准化,并建立指标间的相关系数矩阵,求其特征值和特征向量。由于前 4 个公因子的方差累计贡献率已达 85.969%,因此可以选用这 4 个公因子作为综合评价指标(特征值与表 1 相同),建立因子载荷矩阵。

其次,对因子载荷矩阵进行方差最大化旋转,得出旋转后的因子载荷矩阵(表 4),进而建立因子模型  $X = AF$ ,其中  $X$  是由 23 个指标构成的随机向量, $F$  为公共因子, $A$  为旋转后的因子载荷矩阵(表 4)。

从表 4 看出,公因子 1 在国民生产总值、固定资产投资额、社会消费品零售总额、客运周转量、本地电话年末用户、长途电话数、外商直接投资额、实际利用外资金额这几个因子上的载荷比较高,所以公因子 1 主要反映了经济总量和交通通讯情况。公因子 2 在人均国民生产总值、第三产业比重、城镇居民人均全年消费性支出、农村居民人均全年消费性支出、城镇居民平均可支配收入、农村居民平均纯收入、货运周转量、人均邮电业务总量、每万人中高校在校人数、每万人中专业技术人员数的比重比较高,主要反映了居民生活情况和文化水平。公因子 3 在城市自来水普及率、城市燃气普及率、人均公共绿地面积、工业废水排放达标率上的载荷较高,主要反映了环境状况。而公因子 4 在第二产业比重上的载荷较高,反映了这些地区的第二产业的基本情况。

表 4 旋转后的因子载荷矩阵

指标	成分			
	1	2	3	4
国民生产总值	0.930	0.150	0.160	0.221
人均国民生产总值	0.126	0.912	0.252	0.244
第二产业比重	0.370	-0.053	0.160	0.811
第三产业比重	-0.212	0.793	0.023	-0.449
固定资产投资额	0.905	0.147	0.099	0.227
城镇居民人均全年消费性支出	0.269	0.921	0.058	-0.167
农村居民人均全年消费性支出	0.303	0.835	0.254	0.112
城镇居民平均可支配收入	0.332	0.884	0.176	-0.029
农村居民平均纯收入	0.325	0.843	0.324	0.176
社会消费品零售总额	0.919	0.220	0.178	0.143
货运周转量	0.247	0.592	0.041	0.588
客运周转量	0.895	-0.232	0.170	0.034
人均邮电业务总量	0.157	0.884	0.345	0.006
本地电话年末用户	0.940	0.022	0.200	0.177
长途电话数	0.907	0.221	0.198	0.013
外商直接投资额	0.802	0.414	0.076	0.155
实际利用外资金额	0.841	0.381	0.006	0.074
每万人中高校在校人数	-0.134	0.758	0.446	0.206
每万人中专业技术人员数	-0.416	0.523	0.404	0.177
城市自来水普及率	0.117	0.315	0.722	0.003
城市燃气普及率	0.339	0.534	0.646	0.144
人均公共绿地面积	0.282	0.321	0.798	-0.031
工业废水排放达标率	0.308	0.094	0.644	0.420

最后利用回归法得出因子得分系数,进而得出因子得分函数,得出各省市、自治区在这 4 个公因子上的得分。类似于主成分分析法,由于开始已经对原始数据进行了标准化处理,消除了量纲的影响,所以以各因子的方差贡献率占 4 个公因子总方差贡献率的比重作为权重对各因子得分进行加权汇总,得出因子综合得分,并以此进行排序(见表 3)。

## 2.3 Wilcoxon 符号秩检验和综合评价

采用不同的评价方法往往会导致评价结果有差异,如果评价结果明显不一致,则评价无实际意义。为了考察两种评价结果是否一致,进行了 Wilcoxon 符号秩检验。 $H_0$ :两种评价结果一致, $H_1$ :两种评价结果不一致。表 5 给出了 Wilcoxon 符号秩检验的检验结果,由表 5 可知,负号秩总和为 233.5,正号秩总和为 262.5。 $Z$  检验统计量为 -0.285,对应的概率  $p$  值为 0.776。如果显著性水平  $\alpha$  为 0.05,由于概率  $p$  值大于显著性水平  $\alpha$ ,因此不能拒绝零假设,即认为两种评价结果是一致的。

表 5 Wilcoxon 符号秩检验结果 1  
秩

		<i>N</i>	秩均值	秩总和
主成分排名 - 因子排名	负号秩	15 <sup>a</sup>	15.57	233.50
	正号秩	16 <sup>b</sup>	16.41	262.50
	总和	31		

注 : a 主成分排名 < 因子排名

b 主成分排名 > 因子排名

表 6 Wilcoxon 符号秩检验结果 2  
检验统计<sup>b</sup>

	主成分排名 - 因子排名
<i>Z</i> 检验统计量	-0.285 <sup>a</sup>
双尾二项分布累积概率	0.776

注 : a 基于负号秩

b Wilcoxon 符号秩检验

综合两种评价方法,将因子分析得分<sup>[7]</sup>与主成分分析得分标准化,取其算术平均值作为综合得分,并给出相应的排序(见表 3)。可以看出,综合方法的评价比单一用因子分析或主成分分析的结果要全面、客观,避免了单一评价方法的片面性。

### 3 结果分析

由表 3 中给出的投资环境最后的综合得分排序,把 31 个省市、自治区分类如下。

第 1 类 : 广东、上海、北京、浙江、江苏、山东、天津、福建、辽宁、河北。

第 2 类 : 四川、河南、湖北、湖南、江西、黑龙江、陕西、吉林、安徽、新疆、海南、重庆。

第 3 类 : 云南、内蒙古、山西、广西、青海、甘肃、贵州、宁夏、西藏。

从以上统计可以看出,投资环境受地域的影响很大,排名第 1 类的全部是沿海或东部的发达省市,而排名第 3 类的多为边远省市。广东凭借优越的地理位置和优惠政策以及国家的大量投入,经济迅速增长,投资环境明显高于其它省,而北京、上海基础好,条件优,在投资环境上要明显高一些。西藏、宁夏、贵州、甘肃各项因子条件都很差,软件硬件都存在一定的缺陷。云南、内蒙古、山西、广西、青海的投资环境相对较差,急需加强改善和优化建设。这些分析统计结果与实际情况基本上是相吻合的。

### 参考文献 :

- [1] 李丽辉. 中国东西部地区投资环境的差异问题及对策分析[J]. 统计与信息论坛, 2002, 17(6): 73-76.
- [2] 中国统计局. 中国统计年鉴[M]. 北京: 中国统计出版社, 2004.
- [3] 高会丽, 胡燕京. 西部地区投资环境的因子分析及政策建议[J]. 西安石油学院学报, 2002, 11(2): 7-11.
- [4] 天津市科学技术委员会. 投资环境的评价与改善[M]. 天津: 科学技术出版社, 1993.
- [5] 何晓群. 现代统计分析方法与应用[M]. 北京: 中国人民大学出版社, 2003.
- [6] 薛薇. SPSS 统计分析方法及应用[M]. 北京: 电子工业出版社, 2004.
- [7] 叶宗文. 重庆城镇居民消费结构变动的因子分析[J]. 重庆师范大学学报(自然科学版), 2004, 21(4): 71-73.

(责任编辑 游中胜)