

# 主成分分析在企业发展评估的应用

易采桥<sup>a</sup>, 卢天秀<sup>b\*</sup>, 岳健民<sup>b</sup>

(四川理工学院 a.管理学院; b.数学与统计学院, 四川 自贡 643000)

**摘要:**在某国有银行财务报表中选取10个主要指标进行主成分分析,利用SPSS20软件求解出第一主成分和第二主成分的主成分得分和综合得分,并根据综合得分进行了排名。最后,总结出平均总资产回报率、风险加权资产收益率、不良贷款率、拨备覆盖率、一级资本充足率和总权益对总资产比率这6个指标的变化会对该企业的发展产生显著的影响。

**关键词:**企业发展;主成分分析;SPSS软件

**中图分类号:**F272.5:F224 **文献标志码:**A **文章编号:**1673-1891(2018)03-0060-04

## Discussion on the Application of Principal Component Analysis in Enterprise Development Evaluation

YI Cai-qiao<sup>a</sup>, LU Tian-xiu<sup>b\*</sup>, YUE Jian-min<sup>b</sup>

(a. School of Management; b. School of Mathematics and Statistics, Sichuan University of Science and Engineering, Zigong, Sichuan 643000, China)

**Abstract:** Ten key indicators are chosen from the financial statements of a state-owned bank. Using SPSS software, based on principal component analysis, the scores of the first principal component and the second principal component are obtained. The overall score ranking is given. Finally, it is point out that average return on total assets, risk weighted return on assets, NPL(non-performing loans) ratio, provision coverage, ratio of tier one capital and ratio of total equity to total assets have significant impact to enterprise development.

**Keywords:** development of enterprise; principal component analysis; SPSS software

### 0 引言

在西方经济学中,企业发展评估即是企业在国家产业政策和区域经济规划中的地位,其主要目的是分析和评判企业的公平市场价值。可通过企业盈利能力、资产质量及资产充足率等指标来反映和考核企业未来的发展能力和前景。

在多元统计分析中,主成分分析(Principal Component Analysis, 简记为PCA)是以数据信息丢失最少为原则,为突出数据间的相似和差异,通过线性变换将多指标转化为少数几个有具体性的综合指标的多元统计分析方法<sup>[1-2]</sup>。在企业发展评估中,为了全面反映企业发展状况和避免遗漏重要信息,应尽可能考虑多指标,但随着指标数目的增多,问题亦越复杂,而通过主成分分析法筛选出的几个综合指标可反映原来指标至少85%的信息<sup>[3]</sup>,且相

互独立,无重叠信息,因此,从数学的角度,PCA实际上是一种降维处理技术,将众多相关变量进行综合压缩,从而简化数据结构。主成分分析是因子分析的特殊形式,分析计算软件为SPSS20。

### 1 主成分分析建模原理及步骤

#### 1.1 原理

将原变量指标记为 $x_1, x_2, \dots, x_p$  ( $p \in \mathbb{N}$ ,  $\mathbb{N}$ 为自然数),综合指标记为 $z_1, z_2, \dots, z_m$  ( $m \in \mathbb{N}$ ,  $m \leq p$ )为,即有

$$\begin{cases} z_1 = l_{11}x_1 + l_{12}x_2 + \dots + l_{1p}x_p \\ z_2 = l_{21}x_1 + l_{22}x_2 + \dots + l_{2p}x_p \\ \dots \\ z_m = l_{m1}x_1 + l_{m2}x_2 + \dots + l_{mp}x_p \end{cases} \quad (1)$$

式(1)中, $z_1, z_2, \dots, z_m$ 分别为原指标 $x_1, x_2, \dots, x_p$ 的第1、第2、...、第 $m$ 个主成分; $l_{11}, l_{12}, \dots, l_{ip}$ 称为 $x_1, x_2, \dots, x_p$

收稿日期:2017-12-11

基金项目:国家自然科学基金:几类空间上的线性混沌研究(11501391);人工智能四川省重点实验室开放基金项目(2015RZJ01);四川理工学院教学改革研究项目(JG-1708、JG-1803)。

作者简介:易采桥(1996—),女,四川成都人,本科生,研究方向:多元统计分析。\*为通信作者。

在主成分  $z_i$  上的荷载 ( $1 \leq i \leq m$ )。

### 1.2 步骤

事实上,各指标包含变异程度上的差异信息和相互影响程度上的信息,主成分分析结果会受量纲的影响,若各变量单位不同,每次改变量纲输出结果亦不同,故一般会对数据进行标准化处理<sup>[4]</sup>,然后用相关系数矩阵进行分析。但实际上,数据标准化过程中,在消除量纲影响的同时,可以使得各指标的方差均为1,从而去除了各指标变异程度的差异信息,因此只需对数据进行均值化处理即可。

主成分个数的选取主要由累计贡献率来决定,贡献率越大,所包含的原始数据的信息也越全面,当累计贡献率达到85%及以上时,即:

$$\sum_{i=1}^m \lambda_i / \sum_{i=1}^p \lambda_i \geq 85\%$$

满足要求<sup>[4]</sup>。

因此,用主成分分析法建模的步骤为:

step1:对原始数据进行标准化;

step2:计算样本的相关系数矩阵;

step3:计算相关系数矩阵的特征值和对应的特征向量;

step4:选取重要的主成分,并给出主成分表达式;

step5:计算主成分得分。

## 2 应用实例

表1是2013—2016年某国有银行的财务状况,根据所采集的各指标数据分析影响该银行发展的主要指标。

表1 财务状况原始数据

一级指标	二级指标	各年度数据/%			
		2013年	2014年	2015年	2016年
盈利能力指标	平均总资产回报率	1.44	1.40	1.30	1.20
	净利息收益率	2.57	2.66	2.47	2.16
	风险加权资产收益率	2.45	2.26	2.16	2.01
	成本收入比	28.03	26.75	25.49	25.91
资产质量指标	不良贷款率	0.94	1.13	1.50	1.62
	拨备覆盖率	257.19	206.90	156.34	136.69
	贷款拨备率	2.43	2.34	2.35	2.22
资本充足率指标	一级资本充足率	10.57	12.19	13.48	13.42
	资本充足率	13.12	14.53	15.22	14.61
	总权益对总资产比率	6.76	7.46	8.11	8.21

下面应用软件SPSS20(该软件内部自带默认数据标准化处理)对采集的10项财务指标进行主成分分析。

首先,导入数据后按照步骤分析→降维→因子分析执行。

接下来再执行下列步骤:

描述→系数→抽取→相关性矩阵→碎石图→旋转→最大方差法→载荷图→确定。

执行完毕后得到各财务指标的相关系数。

用  $x_1, x_2, \dots, x_{10}$  依次代表平均总资产回报率、净利息收益率、风险加权资产收益率、成本收入比、不良贷款率、拨备覆盖率、贷款拨备率、一级资本充足率、资本充足率和总权益对总资产比率,相关系数矩阵如下:

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$
$x_1$	+1.000	+0.921	+0.961	+0.808	-0.969	+0.947	+0.905	-0.865	-0.638	-0.912
$x_2$	+0.921	+1.000	+0.794	+0.526	-0.810	+0.749	+0.815	-0.604	-0.291	-0.682
$x_3$	+0.961	+0.794	+1.000	+0.881	-0.963	+0.980	+0.945	-0.936	-0.777	-0.962
$x_4$	+0.808	+0.526	+0.881	+1.000	-0.924	+0.950	+0.685	-0.991	-0.962	-0.976
$x_5$	-0.969	-0.810	-0.963	-0.924	+1.000	-0.990	+0.837	-0.953	+0.793	+0.978
$x_6$	+0.947	+0.749	+0.980	+0.950	-0.990	+1.000	+0.861	-0.980	-0.851	-0.995
$x_7$	+0.905	+0.815	+0.945	+0.685	-0.837	+0.861	+1.000	-0.778	-0.575	-0.820
$x_8$	-0.865	-0.604	-0.936	-0.991	+0.953	-0.980	-0.778	+1.000	+0.938	+0.995
$x_9$	-0.638	-0.291	-0.777	+0.962	+0.973	-0.851	-0.575	+0.938	+1.000	+0.898
$x_{10}$	-0.912	-0.682	-0.962	-0.976	+0.978	-0.995	-0.820	+0.995	+0.898	+1.000

此矩阵显示的是各指标间的相关性,若两指标间的相关系数的绝对值越大,说明这两指标的联系也越大。从上表中可得  $x_{10}$  与  $x_6$  和  $x_{10}$  与  $x_8$  的相关系数的绝对值均为0.995,相对其余指标间的相关系数绝对值是最大的,从而,总权益对总资产比率与拨备覆盖率、总权益对总资产比率与一级资本充足率的联系最大;而  $x_2$  与  $x_9$  的相关系数的绝对值只有0.291,说明资本充足率与净利息收益率的相关性最差,指标间的相关性都可按此分析。

根据相关系数矩阵得出的特征值和特征值的累计贡献率见表2。

表2 指标方差分解主成分提取分析表

主成分	初始特征值/%			提取平方和载入/%		
	合计	方差	累计	合计	方差	累计
$f_1$	8.730	87.298	87.298	8.730	87.298	87.298
$f_2$	1.066	10.657	97.954	1.066	10.657	97.954
$f_3$	0.205	2.046	100.000			
$f_4$	$1.314 \times 10^{-15}$	$1.314 \times 10^{-15}$	100.000			
$f_5$	$2.983 \times 10^{-16}$	$2.983 \times 10^{-16}$	100.000			
$f_6$	$9.934 \times 10^{-17}$	$9.934 \times 10^{-17}$	100.000			
$f_7$	$-6.018 \times 10^{-17}$	$-6.018 \times 10^{-17}$	100.000			
$f_8$	$-1.171 \times 10^{-16}$	$-1.171 \times 10^{-16}$	100.000			
$f_9$	$-2.800 \times 10^{-16}$	$-2.800 \times 10^{-16}$	100.000			
$f_{10}$	$-3.921 \times 10^{-16}$	$-3.921 \times 10^{-16}$	100.000			

现在来确定主成分。主成分的个数选取原则为累计贡献率至少为85%的前  $m$  个主成分,且累计

贡献率越大时选取的主成分个数越精准。由上表可知,前两个主成分累计贡献率达到 97.954%,即这两个主成分可包含原始数据的大部分信息,但仅仅根据累计贡献率来选取主成分个数有一定误差,还可结合主成分对应的特征值大于 1 这一原则进行主成分选取<sup>[3]</sup>,SPSS20 软件中的碎石图可反映各主成分所对应的特征值大小,如图 1 所示。

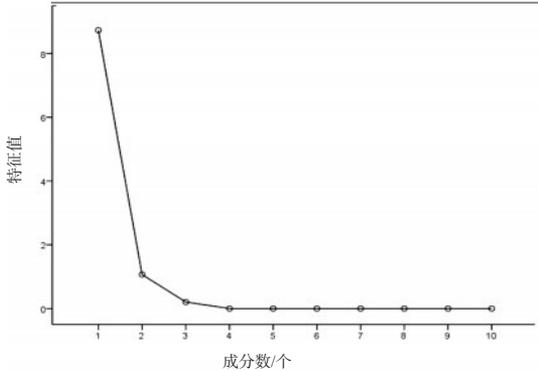


图 1 碎石图

综合主成分提取分析表和碎石图,这里可选取前两个主成分继续分析。这两个主成分分别为:

$$f_1 = 0.32x_1 + 0.26x_2 + 0.33x_3 + 0.32x_4 - 0.33x_5 + 0.34x_6 + 0.30x_7 - 0.33x_8 - 0.28x_9 - 0.34x_{10}$$

$$f_2 = 0.27x_1 + 0.60x_2 + 0.09x_3 - 0.33x_4 - 0.04x_5 + 0.03x_6 + 0.31x_7 + 0.23x_8 + 0.53x_9 + 0.13x_{10}$$

将旋转中的载荷图勾选上,确定后则可生成图 2。此载荷图比较直观地表明了这些指标之间的亲疏关系。观察图 2 可知,  $x_5$ 、 $x_8$  和  $x_{10}$  三个指标聚集在一起,表明这三个指标联系紧密;而  $x_1$ 、 $x_3$ 、 $x_6$  和  $x_7$  四个指标聚集在一起,表明这四个指标联系紧密。此外,指标位置与中心位置距离越大,说明指标与主成分间的相关系数越大。

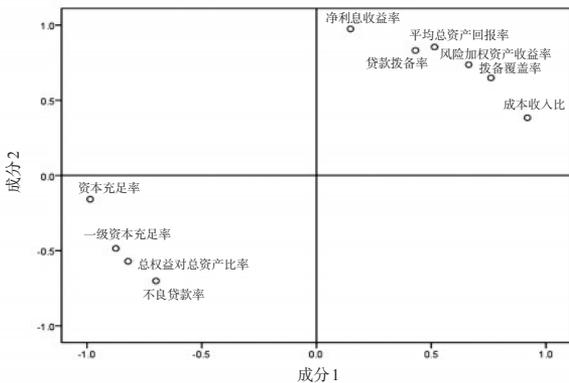


图 2 各财务指标在主成分空间上的投影

各指标主成分载荷矩阵为:主成分载荷矩阵显示了各指标与主成分之间的关系,若指标与某主成分的关联系数的绝对值越大,则二者联系越紧密。易得指标  $x_1$ 、 $x_3$ 、 $x_5$ 、 $x_6$ 、 $x_8$  和  $x_{10}$  在第一主成分上的载荷比较高,说明第一主成分较好的反映了这 6 个指标

	$f_1$	$f_2$
$x_1$	+0.956	+0.282
$x_2$	+0.767	+0.619
$x_3$	+0.987	+0.095
$x_4$	+0.937	-0.337
$x_5$	-0.989	-0.045
$x_6$	+0.999	-0.034
$x_7$	+0.880	+0.322
$x_8$	-0.973	+0.232
$x_9$	-0.834	+0.549
$x_{10}$	-0.991	+0.133

的信息;同理可知,  $x_2$ 、 $x_4$ 、 $x_7$  和  $x_9$  在第二主成分上的载荷比较高,说明第二主成分较好的反映了这 4 个指标的信息。所以用两个主成分已可以较全面的解释原来的指标。

为了纵向分析企业的发展,再次利用软件 SPSS20 计算出各主成分得分为:

	$f_1$	$f_2$
2013 年	88.777 6	-1.330 4
2014 年	70.055 3	-1.330 4
2015 年	51.484 0	1.337 3
2016 年	44.888 6	1.258 0

最后,计算得出综合得分进行排名,得到该国有银行综合得分和年度效益排名为:

$$Score(2013)=77.007 2; Score(2014)=61.013 8;$$

$$Score(2015)=45.086 8; Score(2016)=39.320 7;$$

从综合得分可知 2013 年的发展情况最好,而 2016 年的企业发展情况最差,还可以清晰地看出该国有银行的发展日渐衰败,有走下坡路的趋势。

综上所述,通过对企业的各指标进行主成分分析之后,可从大量的指标中选出主要指标,用少数几个指标代替,同时也较大程度上保持了原来信息的完整性。通过本文的分析,可以确定影响该企业的主要指标为拨备覆盖率、总权益对总资产比率、不良贷款率、风险加权资产收益率、一级资本充足率和平均总资产回报率,其次为净利息收益率、资本充足率、成本收入比和贷款拨备率。

### 3 结语

国有企业始终是中国追求将创新作为基础的发展道路的基本力量,对国有企业的发展评估显得尤其重要。此类评估涉及多个变量,变量个数太多就会增加课题的复杂性,人们自然希望变量个数较少而得到的信息较多。但在很多情形当中,变量之间是有一定的相关关系的,当两个变量之间有一定相关关系时,可以解释为这两个变量反映此课题的

信息有一定的重叠。

主成分分析利用降低指标维数、去除指标间的相关性等方法,对于原先提出的所有变量,将重复的变量(关系紧密的变量)删去多余,建立尽可能少的新变量,使这些新变量是两两不相关的,而且这些新变量在反映课题的信息方面尽可能保

持原有的信息,这就大大减少了工作量<sup>[6-10]</sup>。

主成分分析的应用涉及许多领域,尤其在评估问题中,伴随数学变换过程生成了信息量权数和系统效应权数,比认为确定权数更客观,因此,主成分分析在企业发展评估中将会得到更加广泛的应用。

#### 参考文献:

- [1] 张文霖.主成分分析在SPSS中的操作应用[J].市场研究,2005(12):31-34.
- [2] 白慧强.主成分分析法在SPSS中的应用——以文峪河河岸带林下草本群落为例[J].科技情报开发与经济,2009,19(9):173-176.
- [3] 吴亚非,李科.基于SPSS的主成分分析法在评价体系中的应用[J].当代经济,2009(3):166-168.
- [4] 林海明,杜子芳.主成分分析综合评价应该注意的问题[J].统计研究,2013,30(8):25-31.
- [5] 高王翠,单飞,杨宇翔.基于SPSS主成分分析法在公路客运量预测中的应用[J].山东交通科技,2011(2):7-10.
- [6] 狄升.因子分析在证券市场股票评价中的应用[J].大学数学,2005,21(4):13-19.
- [7] 吴海英,张杰.学生成绩排名的综合评价模型[J].大学数学,2006,22(4):142-145.
- [8] 朱宁,徐标.学生成绩的学分绩与主成分排序模型[J].大学数学,2007,23(1):32-37.
- [9] 陈江丽.基于均值化主成分分析的城镇居民消费性支出应用研究[J].西昌学院学报(自然科学版),2016,30(1):74-76.
- [10] 王莺,王静,姚玉璧,等.基于主成分分析的中国南方干旱脆弱性评价[J].生态环境学报,2014,23(12):1897-1904.

(责任编辑:蒋召雪)

(上接第50页)

- [14] 金成俊.2014年政府工作报告[EB/OL]. [2018-03-01]. <http://www.ahfeixi.gov.cn/content/detail/54b5c351af88bc6867c74594.html>.
- [15] 周先惠.2015年三河镇人民政府工作报告[EB/OL]. [2018-03-01]. <http://zwgk.hefei.gov.cn/zwgk/public/spage.jsp?doAction=view&indexno=003008585/201505-00006>.
- [16] 许灵然.浙江省特色小镇品牌影响力评价及其传播优化策略[D].杭州:浙江传媒学院,2017.
- [17] 孙博洋.专家热议特色小镇:应成为人们对美好生活向往的载体[EB/OL]. (2017-12-20). [2018-03-01]. <http://finance.people.com.cn/n1/2017/1220/c1004-29718982.html>.

(责任编辑:蒋召雪)