

(13) 63-68

主成分分析法在农村电气化发展中的应用

Application of Primary Component Analysis Method
in Development of Rural Electrification

冯析善 陈真 孙辉
Feng Qishan Chen Zhen Sun Hui

S24

(重庆大学工商管理学院)

摘要 在阐明主成分分析法原理的基础上,以浙江省农村电气化的近20种技术经济指标,进行主成分计算和分析,获得影响农村电气化发展的关键因素。并以此预测我国农村电气化发展方向。

关键词 主成分分析; 农村电气化; 关键变量

中国图书资料分类法分类号 S24

ABSTRACT After expounding the principles of primary component analysis, the authors have calculated and analyzed on computer, about twenty technical and economic data of rural electrification in zhejiang province, and found out the key factors which influence the rural electrification in china. According to the research results, the authors forecast the developing trend of rural electrification in our country.

KEYWORDS primary component analysis; rural electrification; key variables

在研究多个指标(自变量)对总体目标的影响时,往往由于自变量个数太多,且彼此之间存在着一定的相关性,使得观测到的数据反映的信息在一定程度上有所交迭。另外,当自变量数较多时,在高维空间中研究样本的分布规律也会遇到较大的困难。主成分分析利用各主分量来代替原来的变量,主分量之间相互独立,主分量数与原自变量相等。为了便于分析、控制及预测,常有必要减少变量数。而根据各主分量对目标变量的贡献大小,可消去对总体影响极小的变量,简化指标体系,并让保留下来的变量负载着原有的绝大部分信息。

1 原理与方法

1.1 主分量及意义

设 X 和 Y 均为 P 维列向量,且满足

$$Y = CX \quad (1)$$

* 收文日期 1992-12-05

国家自然科学基金资助项目

其中 \underline{C} 为正交矩阵, $\underline{C} \cdot \underline{C}' = \underline{I}$, \underline{I} 为单位矩阵。显然, \underline{Y} 的协方差为

$$\underline{Y} \underline{Y}' = (\underline{C} \underline{X}) \cdot (\underline{C} \underline{X})' = \underline{C} \underline{X} \underline{X}' \underline{C}' = \underline{\Lambda} \quad (2)$$

这里 $\underline{\Lambda}$ 为对角阵, 对角元为 $\lambda_i (i = 1, \dots, p)$ 。

很明显, 特征值 λ 依赖于初始变量的计量单位, 当单位变化时, 会得到不同的特征值。为此, 需对初始变量进行标准化处理, 将 x_{ik} 转换为 x'_{ik} :

$$\begin{aligned} \bar{x}_i &= \frac{1}{n} \sum_{k=1}^n x_{ik} \\ s_i^2 &= \frac{1}{n-1} \sum_{k=1}^n (x_{ik} - \bar{x}_i)^2 & i &= 1, \dots, n \\ & & k &= 1, \dots, p \\ x'_{ik} &= \frac{x_{ik} - \bar{x}_i}{s_i} \end{aligned}$$

为方便计, 仍设 \underline{X} 为已标准化处理后的数据矩阵, 则其相关矩阵为

$$\underline{R} = \underline{X} \underline{X}' \quad (3)$$

相关矩阵各元素为

$$r_{ij} = \left(\sum_{k=1}^n x_{ik} \cdot x_{jk} \right) / (n-1) \quad i, j = 1, \dots, p$$

从(2)式及(3)式, 有 $\underline{R} \underline{C}' - \underline{C}' \underline{\Lambda} = 0$ (4)

为得到齐次方程组(4)的非零解, 要求关于 $c_{ij} (j = 1, \dots, p)$ 的系数行列式为零, 即

$$|\underline{R} - \lambda_i \underline{I}| = 0 \text{ 进一步可写成 } |\underline{R} - \underline{\lambda} \underline{I}| = 0$$

利用雅可比法可求出方程(6)的 P 个非负特征值, \underline{R} 的 P 个特征值可按大小排列为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, 设相应于 λ_i 的特征向量为 \underline{C}_i , 构造矩阵 \underline{C} 为

$$\underline{C} = \begin{bmatrix} C_{11} & C_{21} & \dots & C_{p1} \\ C_{12} & C_{22} & \dots & C_{p2} \\ \dots & \dots & \dots & \dots \\ C_{1p} & C_{2p} & \dots & C_{pp} \end{bmatrix} = [\underline{C}_1 \quad \underline{C}_2 \quad \dots \quad \underline{C}_p]$$

显然应有

$$\underline{C}_i \underline{C}_j = \sum_{k=1}^p C_{ik} C_{jk} = \begin{cases} 1 & (i=j) \\ 0 & (i \neq j) \end{cases}$$

而相应于 y_i 的方差为

$$\text{Var}(y_i) = \text{Var}(\underline{C}_i \underline{X}) = \underline{C}_i \underline{X} \underline{X}' \underline{C}_i' = \underline{C}_i \underline{R} \underline{C}_i' = \lambda_i$$

即 y_1 有最大的方差 λ_1 , y_2 有次大的方差 λ_2 , \dots , y_p 有最小的方差 λ_p , 而协方差为

$$\text{Cov}(\underline{C}_i \underline{X}', \underline{C}_j \underline{X}) = \underline{C}_i' \underline{R} \underline{C}_j = 0$$

故 P 维列向量 \underline{X} 经过正交变换后, 可用新的 P 维列向量 \underline{Y} 来表示, 且满足 $\underline{Y} = \underline{C} \underline{X}$, 新变量的坐标相互正交。根据它们的方差从 λ_1 到 λ_p 依次递减, 相互独立的新变量 y_1, y_2, \dots, y_p , 分别称之为第一, 第二, \dots 第 p 个主分量。

1.2 主要变量的筛选

从以上分析中可以发现, 方差的大小体现式(3)主分量 y_i 负载的信息的多少。当前面的 m 个主分量 $y_1, \dots, y_m (m < p)$ 的方差占全部方差的比例接近于 100% 时(例如大于 95%), 可以认为前面的 m 个主分量基本上保留了原来变量的信息, 这样就可以用 m 个综合变量

——主分量来代替原来的 P 个变量。但由于主分量是原来 P 个变量的线性组合,因而难以直接说明原变量对总体目标的作用和影响。为了便于更直观地进行分析、控制和预报等工作,有必要从原始变量集合中选择出关键变量,构成最佳变量子集合。一般说来,首次主成分分析中最后一个主分量的特征值近似为零($\lambda_p \approx 0$),这表明该主分量 g_p 对总体目标的贡献最小(几乎没有贡献),根据与此特征值相对应的特征向量 C_p ,可以找出对该主分量贡献最大的原变量,即具有最大系数 C_{pi} 的变量,该变量对总体目标贡献极微,可以删除。对剩余的原变量再进行主成分分析,可删去另一个原变量。如此反复,直到最后进行主成分分析时,所有特征值都不是很小为止。此时的剩余原变量即为对总体目标有较高相关关系的自变量。如有必要可在此基础上建立起回归方程,通过回归方程来分析、预测总体目标今后的变化趋势及各主要变量对总体目标的影响。

2 实例计算与分析

以下将对浙江省体现农村电气化程度的18个指标的历史纪录值进行分析计算,数据的起迄时间为1976年到1988年,每年一组。希望能找出影响农村电气化的关键因素,并根据这些关键因素来分析预测我国农村电气化的未来发展方向及道路。

选用的18个指标分别属于以下6大类:小水电装机、小水电年发电量、电力设备及变电所(包括高压线路长度、低压线路长度、变电所容量、配电站容量)、产值(总产值、工业产值、农业产值)、用电设备(总用电设备、农村用电设备、县城用电设备)、用电构成(总用电量、农业生产用电、排灌用电、乡镇企业用电、照明用电、农村加工用电)等。对以上指标进行标准化处理并求出相关矩阵后,即可求出特征值 λ 及相应的特征向量,分别如表1、表2所示。

表1 计算所得特征值表

序号	特征值	每个特征值占总体百分比(%)	累计百分比(%)
1	14.9534	83.07	83.07
2	1.4997	8.32	91.40
3	0.9265	5.15	96.54
4	0.3452	1.92	98.46
5	0.1647	0.92	99.38
6	0.0582	0.32	99.70
7	0.0300	0.17	99.87
8	0.0157	0.09	99.95
9	0.0060	0.03	99.99
10	0.0014	0.01	99.99
11	0.0007	0.00	100.00
12	0.0001	0.00	100.00
13	0.0001	0.00	100.00
14	0.0001	0.00	100.00
15	0.0001	0.00	100.00
16	0.0000	0.00	100.00
17	0.0000	0.00	100.00
18	0.0000	0.00	100.00

如按保留 95% 的信息来选取主分量, 只须保留前三个分量即可。这三个主分量包含了全部信息的 96.54%, 它们代表了原 18 个指标对农村电气化的影响, 主分量的采用虽使原结构得到简化, 却也使对原变量影响的分析复杂化。通过观察可以发现, 表 1 中后几个特征值为零或近似为零, 表现它们对总体目标没有大的影响, 可将这些特征值所对应的特征向量中权数最大的变量剔除。例如对最后一个主分量而言, 工业产值所占权数最大, 为 0.6304 见表 2, 可将工业产值指标剔除。

对剩下的变量再反复进行主成分分析, 删去其他一些指标。经过四次主成分分析后, 剩下六个影响较大的指标: 小水电装机容量、低压线路长度、变电所容量、农业产值、排灌用电、农村产品加工用电。

如果进行主成分分析时, 不考虑三个产值指标, 仅对技术指标进行分析, 经过四次主成分分析后, 剩下八个影响较大的指标: 小水电装机容量、小水电年发电量、低压线路长度、变电所容量、排灌用电、农村加工用电、照明用电、县城用电设备容量。两次主成分分析结果基本一致。

计算流程见图 1。

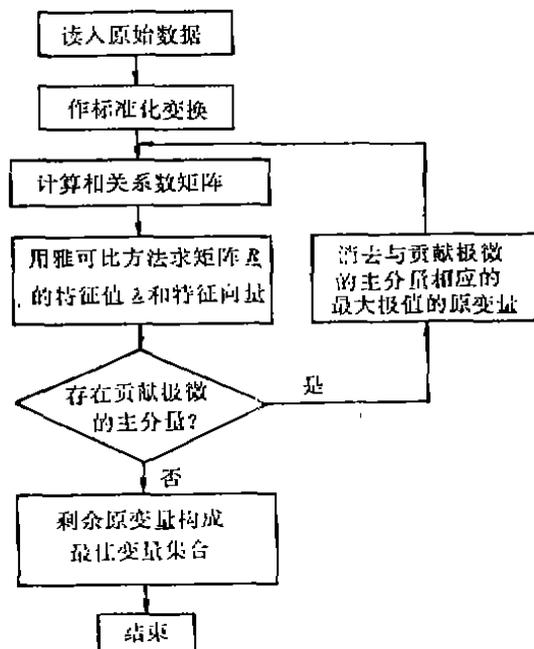


图 1 主成分分析及主要变量筛选流程图

3 讨 论

从以上分析计算中, 可以得到以下论点:

表2 特征向量表

C_1	0.24	-0.29	0.03	0.01	0.01	0.26	-0.00	-0.10	0.23	-0.31	-0.35	-0.25
	0.06	-0.15	-0.38	-0.41	0.30	0.04						
C_2	0.23	-0.19	-0.15	-0.22	-0.26	-0.70	-0.36	0.22	0.13	0.02	-0.19	0.07
	0.04	0.02	-0.03	-0.09	-0.09	0.02						
C_3	0.25	-0.01	-0.02	0.12	0.07	0.20	-0.37	-0.32	0.08	0.43	0.09	0.32
	0.45	0.27	-0.02	-0.15	0.05	0.00						
C_4	0.23	-0.29	0.12	0.09	-0.06	0.43	-0.33	0.54	-0.18	0.15	0.09	-0.05
	0.26	-0.04	0.11	0.25	-0.05	-0.02						
C_5	0.23	0.30	-0.02	0.02	0.32	-0.16	-0.05	0.24	-0.11	0.23	0.27	-0.46
	-0.14	-0.27	0.19	-0.29	0.26	-0.01						
C_6	0.25	-0.08	-0.05	-0.03	0.01	-0.10	-0.16	-0.40	0.26	-0.47	0.53	-0.09
	0.08	0.01	0.19	0.24	0.12	-0.04						
C_7	0.25	0.16	-0.03	-0.08	-0.16	0.14	0.13	-0.01	0.07	-0.09	-0.10	0.01
	-0.14	-0.14	0.18	-0.14	-0.39	-0.74						
C_8	0.24	0.22	0.07	-0.35	-0.12	0.21	0.08	-0.02	0.09	-0.07	-0.01	-0.01
	-0.12	-0.14	0.20	-0.10	-0.42	0.63						
C_9	0.21	-0.07	-0.33	-0.78	-0.22	-0.08	0.24	0.03	-0.01	-0.06	-0.04	-0.01
	-0.08	-0.05	0.17	-0.02	-0.12	0.21						
C_{10}	0.25	-0.02	-0.09	-0.05	0.18	-0.05	0.16	-0.18	0.04	0.24	-0.08	0.50
	0.39	-0.56	0.01	0.03	0.16	0.00						
C_{11}	0.25	-0.05	-0.14	-0.06	-0.02	-0.00	-0.00	-0.41	-0.17	0.37	-0.22	-0.52
	0.32	0.17	-0.07	0.23	-0.17	-0.01						
C_{12}	0.25	0.01	0.00	-0.03	0.54	-0.14	0.42	0.20	0.41	0.01	-0.08	0.04
	-0.06	0.39	-0.11	0.17	-0.08	0.00						
C_{13}	0.25	0.13	-0.04	-0.05	-0.16	0.06	0.11	0.10	-0.23	-0.07	0.22	0.18
	0.48	0.48	-0.03	-0.48	0.08	-0.01						
C_{14}	0.22	0.32	0.20	0.23	0.35	-0.03	-0.41	0.00	-0.35	-0.38	-0.31	0.12
	0.08	0.01	-0.06	0.14	-0.13	0.02						
C_{15}	0.13	0.02	0.86	0.20	-0.25	-0.23	0.17	-0.11	0.09	0.14	0.02	-0.03
	0.01	-0.01	0.00	-0.00	0.04	-0.00						
C_{16}	0.14	-0.66	0.12	-0.12	0.23	-0.03	0.17	0.01	-0.54	-0.07	0.14	0.04
	-0.20	-0.02	0.12	-0.05	-0.16	-0.00						
C_{17}	0.25	0.16	-0.08	-0.04	-0.22	0.02	0.12	0.14	-0.20	-0.00	0.32	0.03
	-0.20	-0.15	-0.71	0.29	-0.03	-0.01						
C_{18}	0.25	0.10	-0.04	-0.20	-0.25	0.09	0.17	0.09	-0.19	-0.06	-0.30	0.06
	-0.22	0.15	0.33	0.34	0.57	-0.00						

3 讨 论

1) 小水电仍是水利资源丰富的地区农村电气化的重要组成部分。两组分析中,小水电装机容量都为主要影响因素,这说明小水电装机容量是农村电气化的重要标志之一。1988年浙江省小水电装机容量为67.9万千瓦,年发电量15.64亿度,分别占全省总装机容量和发电量的51%和61%,它们对浙江的经济发展起了积极作用,同时也说明浙江省的农村电气化水平是比较高的。

2) 从主成分分析中可以看到,低压线路长度和变电所容量在两次分析中均为重要变量,由于高压线路长度与低压线路长度有高度相关性,二者的相关系数为0.9176,可以认为

这两个变量即代表着电力网。因此,发展电力网,是实现农村电气化的重要环节。目前,一些地区存在只建电站不建电网的问题。在小水电地区,只有既强调电站建设又强调电网建设,才能作到发、供、用相结合,更好地以电养电,促进农村电气化水平的提高。

3) 农村加工用电随着农村经济的发展将不断增长,但随着农村加工业的扩大,部分将与乡镇企业用电交融,两者将成为影响农村用电的重要因素,也是农村电气化水平的重要因素之一:

4) 在两次分析结果中,都显示了排灌用电的重要地位。很明显,排灌用电是农村电气化的主要体现者之一,要实现农村电气化,应该首先实现排灌电气化。

5) 照明用电和县城用电设备容量也在较大程度上影响着农村电气化的实现。只有保证了人民生活用电的需要,才能说明农村电气化给人民群众带来的实际效益。从原始数据中可以看到农村照明用电速度有较快的增长趋势,这一点应该在农村电气化的规划中得到足够重视。

4 结 语

抓好小水电站及地方电网的建设工作,分析用电发展趋势及用电构成,是农村电气化工作中应切实作好的几个方面,也是农村电气化的发展方向。此外,积极发展联网,与大电网相互调剂余缺,则是提高农村电气化水平的重要手段。限于资料,以上的分析未考虑联网因素的影响,这一点有待于在今后的工作中加以完善。另外,本文的分析虽然是在浙江省历史数据基础上得出的,但对于其它省份或地区也不无参考价值。当然,由于各地情况的差异,在研究时必须具体情况具体分析。

参 考 文 献

- 1 罗积玉等. 经济统计方法与预测. 北京:清华大学出版社,1987
- 2 B. C. Jian. Rural Energy Centres Based on Renewables—Case Study on an Effective and Viable Alternative. IEEE Tran on Energy Conversion, 1987(3)