

文章编号:1672-058X(2012)11-0022-05

# 基于主成分分析的农村居民消费性支出研究

## ——以中国 31 个省市为例

罗时超<sup>1</sup>, 李曦染<sup>2</sup>

(1. 四川农业大学 生命科学与理学院, 四川 雅安 625014, 2. 四川农业大学 食品学院, 四川 雅安 625014)

**摘要:**基于主成分分析的评价模型,在一定程度上避免了评价的主观性、克服了指标之间的相关性,且信息损失率较少;运用主成分分析法筛选出影响我国各省市农村居民家庭消费性支出的主导成分,再通过系统聚类将消费支出水平相近的省市归类,构成谱系聚类图用以描述我国各省市农村居民消费性支出水平,取得了较为合理的结果;力求对中国目前的农村居民消费性支出提出了相应策略。

**关键词:**农村居民消费性支出;主成分分析;系统聚类

**中图分类号:** O305

**文献标志码:** A

消费是指人们为了满足生活需要,对物质产品的消耗和享受服务的行为过程。随着近几年我国社会和经济发展,农村居民消费水平不断提高,消费结构发生了显著变化。而且作为社会总需求最重要的组成部分,居民的消费增长直接影响到整个国民经济增长的速度和质量。国内外对居民消费增长问题的研究由来已久,其中对消费结构变化的研究已成为近年来学者们研究的热点<sup>[1]</sup>。从目前的研究思路看,主要是运用传统消费函数理论对其进行解释,如生命周期理论、持久收入假说等,存在着一定的缺陷。经过 20 多年的改革,我国的经济运行机制有了明显的变化,市场导向作用大大增强,消费对经济增长的作用日趋明显,消费理论研究在经济研究中也具有越来越重要的地位。但目前我国消费经济学的研究思路侧重于定性分析,定量研究还较欠缺,不能形成一个完整的理论体系<sup>[2]</sup>。而对我国来说各地区的经济发展不平衡,居民的消费水平和消费性支出也存在着很大的差异。因此,对我国各地区农村居民人均消费性支出的各项指标进行统计分析,明确我国居民的消费性支出差异的数量特质,就显得尤为必要。

## 1 主成分分析法原理及步骤

主成分分析是一种常用的多元统计分析(即多指标的统计分析)方法,是一种化繁为简,将指标数尽可能压缩的降维(即空间压缩)技术,也是一种综合评价方法<sup>[4]</sup>。

主成分分析的基本步骤如下:

(1) 利用公式  $x_{ij}^* = \frac{x_{ij} - \bar{X}_j}{S_j}$  ( $\bar{X}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$ ;  $S_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{X}_j)^2}$ ;  $i = 1, 2, \dots, n$ ;  $j = 1, 2, \dots, p$ ) 将

原始数据进行标准化,这样可以消除量纲的影响以及各指标在数量级上的差别。其中,  $x_{ij}^*$  是  $x_{ij}$  的标准化数据,  $\bar{X}_j$  和  $S_j$  分别是第  $j$  个指标的样本均值和样本标准差。

(2) 建立标准化数据的相关系数矩阵  $R = [r_{ij}]_{p \times p}$ ,  $r_{ij}$  是  $x_i^*$  与  $x_j^*$  的相关系数,  $x_i^*$  和  $x_j^*$  分别是  $x_i$  和  $x_j$  的标准化指标。

(3) 求出相关系数矩阵  $R$  的特征值  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$  及对应的特征向量  $e_1, e_2, \dots, e_p$ , 其中  $\lambda_m$  是第  $m$  个成分  $f_m$  的方差, 方差越大, 对总方差的贡献越大。计算出  $f_m$  的贡献率  $\alpha_m = \frac{\lambda_m}{\sum_{j=1}^p \lambda_j}$  和累计贡献率  $\sum_{j=1}^m \alpha_j$ ,

找到使累计贡献率大于 85% 的最小整数  $m$ , 这样就确定了前  $m$  个成分为主成分。

$$(4) \text{ 线性表出每个主成分: } \begin{cases} f_1 = l_{11}x_1 + l_{12}x_2 + \dots + l_{1p}x_p \\ f_2 = l_{21}x_1 + l_{22}x_2 + \dots + l_{2p}x_p \\ \vdots \\ f_m = l_{m1}x_1 + l_{m2}x_2 + \dots + l_{mp}x_p \end{cases}$$

其中,  $l_{ij} = \sqrt{\lambda_i} e_{ij}$ ,  $e_{ij}$  为特征向量  $e_i$  的第  $j$  个分量,  $i, j = 1, 2, \dots, p$ ; 进而可以计算出每个样本点在各个主成分上的得分。

## 2 农村居民消费性支出水平的实例分析

本研究选取《中国统计年鉴》(2010) 提供的包括全国 31 个省、直辖市和自治区(港、澳、台除外)的各地区农村居民家庭人均年消费性支出数据作为观测样本<sup>[3]</sup>, 所选指标包括食品  $x_1$ , 衣着  $x_2$ , 居住  $x_3$ , 家庭设备及服务  $x_4$ , 交通和通讯  $x_5$ , 文教娱乐用品及服务  $x_6$ , 医疗保健  $x_7$ , 其他商品及服务  $x_8$ 。

### 2.1 数据分析

由指标的原始数据经标准化后建立相关系数矩阵, 求出相关系数矩阵的特征值、贡献率及累计贡献率如表 1, 主成分载荷如表 2 所示。

表 1 相关系数矩阵的特征值、贡献率及累积贡献率

特征值	差值	贡献率/%	累积贡献率/%
6.730 4	6.175 8	84.130 6	84.130 6
0.554 6	0.279 7	6.933 0	91.063 6
0.274 9	0.124 0	3.436 2	94.499 9
0.150 9	0.033 1	1.886 4	96.386 3
0.117 8	0.027 3	1.472 6	97.858 9
0.090 5	0.031 9	1.130 8	98.989 8
0.058 6	0.036 3	0.731 9	99.721 7
0.022 3	—	0.278 3	100.000 0

表 2 主成分载荷

标准化变量	主成分 Print1	主成分 Print2
食 品	0.346 1	0.531 1
衣 着	0.342 0	-0.494 7
居 住	0.355 1	-0.012 0
家庭设备及服务	0.361 9	0.053 6
交通和通讯	0.378 7	0.072 0
文教娱乐用品及服务	0.363 1	-0.068 1
医疗保健	0.340 4	-0.524 2
其他商品及服务	0.339 1	0.430 8

由表 1 可知相关矩阵的特征值为  $\lambda_1 = 6.730 4$ ,  $\lambda_2 = 0.554 6$ 。前两个特征值的累计贡献率为 91.063 6%, 说明前 2 个成分构成了原指标数据的主成分。由表 2 主成分载荷矩阵得到各主成分得分的表达式, 以各主成分的贡献率为权重得到综合得分表达式如下:

$$F_1 = 0.346 1x_1 + 0.342 0x_2 + 0.355 1x_3 + 0.361 9x_4 + 0.378 7x_5 + 0.363 1x_6 + 0.340 4x_7 + 0.339 1x_8$$

$$F_2 = 0.531 1x_1 - 0.494 7x_2 - 0.012 0x_3 + 0.053 6x_4 + 0.072 0x_5 - 0.068 1x_6 - 0.524 2x_7 + 0.430 8x_8$$

$$F_{\text{总}} = 0.360 2x_1 + 0.278 3x_2 + 0.327 2x_3 + 0.338 4x_4 + 0.355 4x_5 + 0.330 3x_6 + 0.274 6x_7 + 0.346 1x_8$$

### 2.2 成分命名及解释

从第一主成分  $F_1$  的表达式来看, 它在每个标准化变量上有相近的正负荷, 说明每个标准化变量对  $F_1$  的

重要性都差不多,且按第一主成分得分从小到大进行排序后的结果可以看出,标准化后,每个地区的消费性支出的总和与第一主成分得分基本成正比,也就是  $F_1$  反映的是消费性支出的综合水平,可认为第一主成分  $F_1$  是综合消费性支出成分。

从第二主成分  $F_2$  的表达式来看,它在每个标准化变量  $x_1$ (食品)和  $x_8$ (其他商品及服务)有中等程度的正载荷,在  $x_2$ (衣着)和  $x_7$ (医疗保健)有中等程度的负载荷。说明  $F_2$  反应的是两个方面的对比,一个方面是衣着和医疗保健的消费性总支出,另一个方面是食品和其他商品及服务的消费总支出,且按第二主成分得分从小到大进行排序后的结果标准化后,每个地区两个方面消费总支出的差与第二主成分得分基本成反比,并且南方地区在食品和其他商品及服务上的消费性支出比较大,北方地区在衣着和医疗保健上的消费支出比较大;这大概跟南北方气候差异有关,南方气候温暖,人们的消费倾向于食品和其他商品及服务,北方气候寒冷,人们的消费倾向于衣着和医疗保健。所以,可认为第二主成分  $F_2$  是消费倾向成分。

### 2.3 各地区综合得分

由上述表达式计算出各个样本点在主成分上的得分、各样本点的综合得分如表 3 所示。经过主成分分析法处理后的信息数据,主成分之间的相关性减小,通过降维将分散的变量指标上的信息集中到了 2 个综合指标上,且损失的信息量较少,大大简化了计算。为下一步的聚类分析提供了较为客观的信息。

表 3 各省市在各主成分上的得分及综合得分

地区	第一主成分 得分 $F_1$	第二主成分 得分 $F_2$	综合总 得分 $F_{\text{总}}$	排名	地区	第一主成分 得分 $F_1$	第二主成分 得分 $F_2$	综合总 得分 $F_{\text{总}}$	排名
北京	7.535 1	-1.527 3	6.845 5	2	湖北	-0.555 9	0.674 2	-0.462 3	13
天津	0.107 3	-0.114 0	0.090 5	10	湖南	-0.581 1	0.713 7	-0.482 6	14
河北	-1.095 3	-0.451 4	-1.046 3	21	广东	0.912 0	2.057 8	0.999 3	6
山西	-0.959 9	-0.514 7	-0.926 1	19	广西	-1.862 9	0.579 9	-1.677 0	26
内蒙古	-0.159 8	-0.505 3	-0.186 1	12	海南	-1.806 5	1.112 4	-1.584 3	25
辽宁	0.530 6	-0.503 9	0.451 9	7	重庆	-1.610 8	-0.124 1	-1.497 7	24
吉林	0.031 0	-0.970 8	-0.045 2	11	四川	-0.605 3	0.146 2	-0.548 1	15
黑龙江	0.519 0	-1.003 7	0.403 2	9	贵州	-2.807 9	-0.009 9	-2.595 0	31
上海	7.886 8	0.759 5	7.344 7	1	云南	-2.123 8	0.142 3	-1.951 4	28
江苏	2.439 8	0.781 9	2.313 7	4	西藏	-2.471 1	-0.103 9	-2.291 0	30
浙江	5.329 1	0.111 3	4.932 2	3	陕西	-0.969 6	-0.615 0	-0.942 6	20
安徽	-0.77 80	0.157 1	-0.706 9	16	甘肃	-2.223 9	-0.217 8	-2.071 3	29
福建	1.106 0	1.191 3	1.112 5	5	青海	-1.405 7	-0.563 5	-1.341 6	23
江西	-1.203 8	0.374 9	-1.083 7	22	宁夏	-0.882 5	-0.383 9	-0.844 6	17
山东	0.445 4	-0.104 8	0.403 5	8	新疆	-1.768 3	-0.895 4	-1.701 9	27
河南	-0.970 1	-0.193 1	-0.911 0	18					

### 2.4 聚类分析

下面对各省市的第一主成分,第二主成分得分进行聚类分析。在此采用切比雪夫距离  $d_{ij} = \max_{1 \leq k \leq p} |x_{ik} - x_{jk}|$ ,选取 WARD 法,对我国 31 个地区进行聚类分析,得农村居民消费性水平谱系聚类图如图 1。结合实际情况将我国 31 个省市分为 4 类,见表 4。

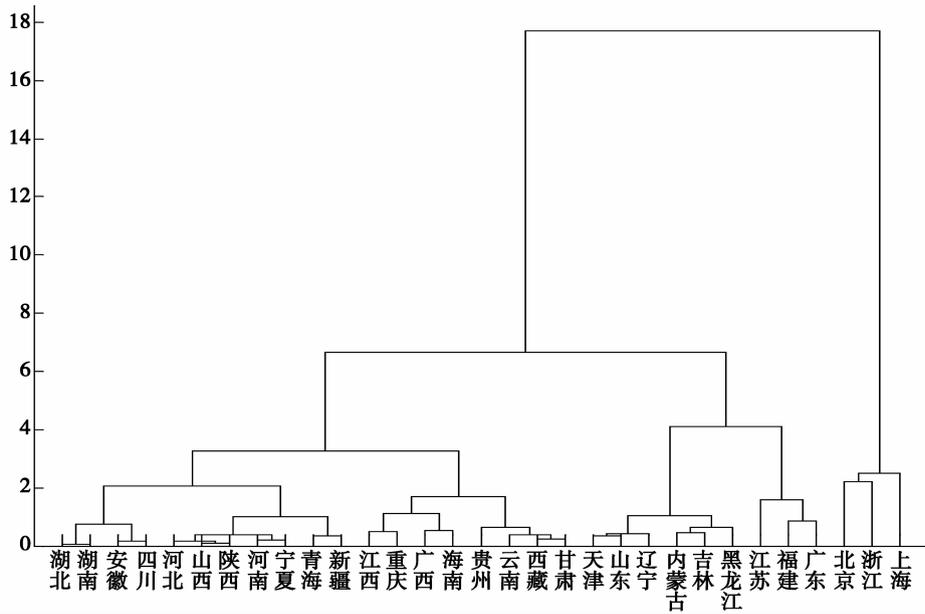


图 1 各省市农村居民消费性水平谱系聚类图

表 4 分 4 类时的组合结果

类 别	地 区
第一类	北京 上海 浙江
第二类	江苏 福建 广东
第三类	辽宁 山东 黑龙江 天津 吉林 内蒙古
第四类	河北 山西 安徽 江西 河南 湖北 湖南 海南 四川 陕西 青海 宁夏 新疆 广西 重庆 贵州 云南 西藏 甘肃

### 3 结果分析与建议

(1) 随着我国经济的迅速发展,各地区在消费性支出方面都有了较大幅度的提高。但是就反映综合消费性支出成分  $F_1$  而言,我国 31 个地区综合消费水平发展极不平衡,存在较大的差异。得分最高的上海和得分最低的贵州相差 10.4967 分之多,并且只有上海、北京、浙江、江苏、福建、广东、辽宁、黑龙江、山东、天津、吉林 11 个地区得分在 0 分以上,其余 2/3 的地区得分都小于 0,这与我国各地区整体经济发展水平的实际情况吻合,说明大部分地区还处于全国消费平均水平之下,有待进一步提高和引导消费。为此,对于那些  $F_1$  得分小于 0 的地区,只有大力发展经济,使得大家的可支配收入水平提高,才能有效地扩大内需,拉动消费。

(2) 就  $F_2$  是消费倾向成分而言,以北京、黑龙江、吉林、新疆等北方地区的第二主成分比较小,中部地区次之,而以广东、福建、海南等南方地区较大,说明第二主成分是因地域差异所造成的消费倾向成分。另外北方地区在衣着和医疗保健上的消费支出比较大,这大概跟南北方气候差异有关。南方气候温暖,人们的消费倾向于食品和其他商品及服务,北方气候寒冷,人们的消费倾向于衣着和医疗保健。

(3) 从总的消费性支出水平表达式,可以看出我国目前农民主要消费集中在食品、交通和通讯,而在衣着和医疗保健相对较少。

(4) 根据聚类分析的结果可知,我国31个省市目前的农村居民消费性服务水平呈现“两头小,中间大”的态势。北京、上海、浙江农村居民消费性支出水平相对较高,而广西、贵州、云南、西藏、甘肃的农村居民消费性支出水平相对薄弱,其他省市都处于中间态势。然而,比较理想的消费水平结构应该是呈现“倒金字塔”型的。

综上,我国在农村居民消费性支出过程中应当继续发挥北京、上海、浙江的优势,着力帮扶广西、重庆、贵州、云南、西藏、甘肃建立更加完善的居民消费性支出制度。提高农村居民可支配收入,进一步减轻农村居民负担。在政策上可以直接给予其生产减免税收,增加补贴,这对增加农村居民收入能起到立竿见影的效果,对刺激消费有不可或缺的作用,争取农民消费支出的全面胜利。

#### 参考文献:

- [1] 冯建中,何先平. 基于因子和聚类分析的城镇居民家庭消费性支出应用研究[J]. 长江大学学报:自然科学版,农学卷, 2010,7(1):97-100
- [2] 李喆. 因子分析在我国农村居民消费性支出中的应用[J]. 科学经济市场,2006
- [3] 中华人民共和国国家统计局. 中国统计年鉴2010[M]. 北京:中国统计出版,2010
- [4] 余锦华,杨维权. 多元统计分析与应用[M]. 广州:中山大学出版社,2005
- [5] 朱建平. 应用多元统计分析[M]. 北京:科学出版社,2006
- [6] 余明江,季丽,胡云霞. 我国城镇居民消费性支出的因子分析[J]. 安徽工业大学学报,2008,25(2):222-225
- [7] 胡丽平,何春花. 基于因子分析的河南省各地区城镇居民消费评价研究[J]. 河南科学,2011,29(6):751-753
- [8] 张志红,任杰. B样条小波边缘检测改进算法[J]. 四川兵工学报,2010(2):136

## Research on Consumption Expenditure of Rural Inhabitants Based on Principal Component Analysis ——Taking 31 Provinces and Municipalities of China as an Example

**LUO Shi-chao<sup>1</sup>, LI Xi-ran<sup>2</sup>**

(1. School of Life Science and Science, Sichuan Agriculture University, Sichuan Yaan 625014, China;  
2. School of Food Science, Sichuan Agriculture University, Sichuan Yaan 625014, China)

**Abstract:** The evaluation model based on principal component analysis avoids the subjectivity of evaluation to some extent, overcomes the correlation between indicators and has smaller information loss rate. This paper uses principal component analysis to select the leading factors affecting family consumption expenditure of rural inhabitants of each province and municipality of China, then classifies the provinces and municipalities which have similar consumption expenditure level through system clustering analysis to construct hierarchical clustering diagram to describe consumption expenditure level of rural residents of all provinces and municipalities of China and obtains relatively reasonable result in order to provide corresponding strategies for current rural inhabitants consumption expenditure in China.

**Key words:** rural inhabitants consumption expenditure; principal component analysis; system clustering analysis

责任编辑:代小红