

对应分析在农村家庭人均消费支出结构中的应用

张永利

(黑龙江科技学院 理学院, 哈尔滨 150027)

摘要:消费水平是衡量一个地区经济发展水平的重要指标之一。我国幅员辽阔,地区间经济发展水平不同,同时由于人们对消费的偏好程度不同,所以支出也不尽相同。本文通过对我国 2009 年的 31 个省份地区农村居民人均消费支出及消费结构的数据进行研究分析,得出一定的消费规律性,并针对我国目前农村的问题提出了一些科学可行的建议,对政策部门具有一定的指导意义。

关键词:对应分析;因子分析;因子载荷;概率矩阵

中图分类号:F224 **文献标志码:**A **文章编号:**1671-1807(2012)05-0131-04

消费水平是衡量一个地区经济发展水平的重要指标之一。考虑一个地区的消费水平不能只考察某个测量指标,也不能将众多测量指标同等对待。如果仅仅抓住某个测量指标考察,就有可能陷入考察结果有失偏颇、以偏概全的危险境地。如果对各测量指标都进行考察,就有可能陷入考察结果主次不分、特征不显的尴尬境地。具体到我国农村家庭人均消费支出情况,由于我国幅员辽阔,地区间经济发展水平不同,同时由于人们对消费的偏好程度不同,所以支出也不尽相同。本文通过对我国 2009 年的 31 个省份地区农村居民人均消费支出及消费结构的数据进行研究分析,得出一定的消费规律性,并针对我国目前农村的问题提出了一些科学可行的建议,对政策部门具有一定的指导意义。

1 研究方法

因子分析方法是少数几个公共因子去提取研究对象的绝大部分信息,既减少了因子的数目,又把握住了研究对象的相互关系。在因子分析中根据研究对象的不同,分为 R 型和 Q 型;由于 R 型因子分析和 Q 型因子分析都是反映一个整体的不同侧面,因此它们之间一定存在内在的联系^[1]。对应分析就是通过一个过渡矩阵 Z 将二者有机地结合起来,具体地说,给出变量点的协差阵 $A = Z'Z$ 和样品点的协差阵 $B = ZZ'$,由于 $Z'Z$ 和 ZZ' 有相同的非零特征根 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m, 0 < m \leq \min(p, n)$,如果特征根 λ_i 对应的特征向量为 U_i ,则 B 的特征根 λ_i 对应的特征向量就是 $ZU_i \Delta V_i$,根据这个结论,就可以很方

便的借助 R 型因子分析而得到 Q 型因子分析的结果。因为求出 A 的特征根和特征向量后很容易地写出变量点协差阵对应的因子载荷阵,记为 F 。则

$$F = \begin{bmatrix} u_{11} \sqrt{\lambda_1} & u_{12} \sqrt{\lambda_2} & \cdots & u_{1m} \sqrt{\lambda_m} \\ u_{21} \sqrt{\lambda_1} & u_{22} \sqrt{\lambda_2} & \cdots & u_{2m} \sqrt{\lambda_m} \\ \vdots & \vdots & & \vdots \\ u_{p1} \sqrt{\lambda_1} & u_{p2} \sqrt{\lambda_2} & \cdots & u_{pm} \sqrt{\lambda_m} \end{bmatrix}$$

这样一来样品点协差阵 B 对应的因子载荷阵记为 G ,则

$$G = \begin{bmatrix} v_{11} \sqrt{\lambda_1} & v_{12} \sqrt{\lambda_2} & \cdots & v_{1m} \sqrt{\lambda_m} \\ v_{21} \sqrt{\lambda_1} & v_{22} \sqrt{\lambda_2} & \cdots & v_{2m} \sqrt{\lambda_m} \\ \vdots & \vdots & & \vdots \\ v_{n1} \sqrt{\lambda_1} & v_{n2} \sqrt{\lambda_2} & \cdots & v_{nm} \sqrt{\lambda_m} \end{bmatrix}$$

由于 A 和 B 具有相同的非零特征根,而这些特征根又正是各个公共因子的方差,因此可以用相同的因子轴同时表示变量点和样品点,即把变量点和样品点同时反映在具有相同坐标轴的因子平面上,以便对变量点和样品点一起考虑进行分类。

在处理实际问题时,如果指标(变量)的量纲不同以及数量级相差很大时,通常先将指标作标准化处理,然而这种标准化处理对样品就不好进行了。换句话说,标准化处理对于指标和样品是非对等的,为了使之有对等关系性,以便将 R 型与 Q 型建立起联系,就需设法将原始数据阵 $X = (x_{ij})$ 变换成矩阵 $Z = (z_{ij})$,即将 x_{ij} 变换成 z_{ij} 之后,使 z_{ij} 对指标和样品具有对等性。

收稿日期:2012-03-16

作者简介:张永利(1976—),男,黑龙江宁安人,黑龙江科技学院讲师,硕士,研究方向:概率论和统计。

$A = Z'Z$ 和 $B = ZZ'$ 。A 与 B 两矩阵明显的存在着简单的对应关系,而且将原始数据 x_{ij} 转换成 z_{ij} 后, z_{ij} 对于 i, j 是对等的,即 z_{ij} 对变量和样品具有对等性^[1]。

第一步 由原始资料阵 X 出发,计算规格化的概率矩阵

$$P = (p_{ij}) = \left(\frac{x_{ij}}{T}\right) \quad T = \sum_{i=1}^n \sum_{j=1}^p x_{ij}$$

第二步 计算过渡矩阵 $Z = (z_{ij})$

$$z_{ij} = \frac{x_{ij} - \frac{x_{.j}x_{i.}}{x_{..}}}{\sqrt{x_{.j}x_{i.}}} \quad i = 1, \cdots, n$$

$$j = 1, \cdots, p$$

第三步 进行因子分析

R 型因子分析

(1)计算协差阵 $A = Z'Z$ 的特征根 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$, 按其累计百分比

$$\sum_{a=1}^m \lambda_a / \sum_{a=1}^p \lambda_a \geq 85\%$$

取前 m 个特征根 $\lambda_1, \cdots, \lambda_m$, 并计算相应的单位特征向量记为 $u_1, \cdots, u_m$, 从而得到因子载荷阵:

$$F = \begin{bmatrix} u_{11} \sqrt{\lambda_1} & u_{12} \sqrt{\lambda_2} & \cdots & u_{1m} \sqrt{\lambda_m} \\ u_{21} \sqrt{\lambda_1} & u_{22} \sqrt{\lambda_2} & \cdots & u_{2m} \sqrt{\lambda_m} \\ \vdots & \vdots & & \vdots \\ u_{p1} \sqrt{\lambda_1} & u_{p2} \sqrt{\lambda_2} & \cdots & u_{pm} \sqrt{\lambda_m} \end{bmatrix}$$

表 1 各地区农村居民家庭平均每人全年消费性支出(2009 年)

指标	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
北京市	5 936.11	1 795.68	1 290.22	1 225.68	2 767.85	2 654.98	1 389.45
天津市	5 404.53	1 362.56	1 505.70	911.92	1 968.37	1 740.85	1 273.38
河北省	3 250.77	1 190.19	1 142.83	628.49	1 151.15	982.21	971.29
山西省	3 071.93	1 162.00	1319.45	563.82	1 095.77	1 070.60	789.92
内蒙古	3 772.63	1 857.19	1 246.21	797.77	1 557.03	1 504.36	992.73
辽宁省	4 680.85	1 338.84	1 293.00	607.51	1 493.17	1 283.68	1 018.44
吉林省	3 637.32	1 419.12	1 394.94	543.69	1 305.45	1 028.06	1 120.44
黑龙江	3 397.41	1 403.72	1 026.77	547.87	922.77	956.85	978.79
上海市	7 344.83	1 593.08	1 913.22	1 365.39	3 498.65	3 138.98	1 002.14
江苏省	4 773.67	1 297.95	1 148.85	923.32	1 721.87	1 968.03	808.37
浙江省	5 604.72	1 614.66	1 485.9	828.96	3 290.63	2 295.32	984.62
安徽省	4 051.40	1 080.06	1 219.83	589.73	1 013.38	1 225.36	716.87
福建省	5 336.36	1 171.88	1 394.91	859.06	1 993.77	1 504.96	591.50
江西省	3 881.56	1 053.01	935.44	761.85	1 145.16	1 066.94	550.25
山东省	3 954.34	1548.75	1 280.04	885.04	1 719.68	1 332.97	885.16
河南省	3 272.75	1 270.74	1 004.37	684.79	1 033.99	1 048.14	875.52
湖北省	4 160.51	1 210.32	999.49	759.24	953.69	1 208.46	694.61
湖南省	4 174.55	1 146.25	1 074.69	798.40	1 233.82	1 207.72	784.66
广东省	6 225.22	1 064.33	1 814.00	1 052.57	2 979.88	2 168.88	925.62

(2)在两两因子轴平面上作变量点图。

Q 型因子分析

对上述所求的 m 个特征根 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ 计算其对应于矩阵 $B = ZZ'$ 的单位特征向量 $Zu_1 \bar{\Delta}V_1, Zu_2 \bar{\Delta}V_2, \cdots, Zu_m \bar{\Delta}V_m$ 从而得到 Q 型因子载荷阵:

$$G = \begin{bmatrix} v_{11} \sqrt{\lambda_1} & v_{12} \sqrt{\lambda_2} & \cdots & v_{1m} \sqrt{\lambda_m} \\ v_{21} \sqrt{\lambda_1} & v_{22} \sqrt{\lambda_2} & \cdots & v_{2m} \sqrt{\lambda_m} \\ \vdots & \vdots & & \vdots \\ v_{n1} \sqrt{\lambda_1} & v_{n2} \sqrt{\lambda_2} & \cdots & v_{nm} \sqrt{\lambda_m} \end{bmatrix}$$

在与 R 型相应的因子平面上作样品点图。

2 数据和结果分析

2.1 数据来源

本文以 2009 年我国 31 个省、市、自治区农村居民家庭人均消费支出数据为研究对象,这里选取了 7 个指标作为反映农村居民家庭人均消费支出结构的因素,这 7 个指标分别是:

X_1 :食品支出比重; X_2 :衣着支出比重; X_3 :居住支出比重; X_4 :家庭设备及服务支出比重; X_5 :交通和通讯支出比重; X_6 :文教娱乐、用品及服务支出比重; X_7 :医疗保健支出比重。

续表 1

指标	X_1	X_2	X_3	X_4	X_5	X_6	X_7
广西	4 129.55	855.60	1 021.11	754.79	1 598.68	1 111.13	538.17
海南省	4 507.81	581.66	1 000.32	585.72	1 548.76	961.95	604.15
重庆市	4 576.23	1 503.49	1 120.60	1 043.06	1 189.03	1 351.90	982.73
四川省	4 391.73	1 178.38	973.02	679.16	1 416.49	1 150.73	648.31
贵州省	3 755.61	1 012.14	747.57	589.35	983.13	11 46.35	535.43
云南省	4 460.58	1 102.14	943.67	393.22	1 587.19	798.69	708.78
西藏自	4 581.60	1 086.42	689.76	356.86	1 062.83	465.84	352.31
陕西省	3 988.57	1 209.96	1 018.23	683.51	1 071.48	1 430.22	863.36
甘肃省	3 359.30	1 169.70	801.21	559.06	894.35	1 025.47	746.77
青海省	3 548.85	1 043.40	790.50	505.32	975.91	889.32	701.37
宁夏	3 432.23	1 260.58	1 128.12	636.88	1 363.63	1 075.88	921.86
新疆	3 386.33	1 357.05	856.78	552.50	1 198.65	855.53	684.01

数据来源《中国统计年鉴 2010 年》

2.2 基本过程

接下来就是通过 SPSS 软件生成对应分析图再进行分类：

实验步骤^[2]：

1) 数据录入。打开 SPSS 文件，按顺序:File-New-Data 打开一个空白数据的文件，首先进行变量的编辑，点击 Variable View 选项，录入三个变量，并对变量 “province” 和 “consume” 进行赋值。

2) 按顺序:Data-Weight Cases，打开 “Weight Cases” 对话框。把 “消费” 选入到 “Frequency Variable” 框中，单击 “OK”，返回到 “Data View” 窗口。

这一步骤是为了激活 “消费”，至关重要。如果 “消

费” 没有被激活，之后要进行的对应分析的命令仍会执行，但结果是错误的

3) 按顺序:AnalyData-Reduction Correspondence Analysis 选择对应分析的菜单，进入到对应分析的主对话框，将 “地区” 选入到 Row 框中，并点击 “Define Range” 打开其对话框，在 “Minimum Value” 中输入 “1”，在 Maximum Value 中输入 “31”，点击 “Update”，点击 “OK”，返回到主对话框；将 “消费类别” 选入到 “Column” 框中，按同样的方法，定义其范围为 1 到 7，返回到主对话框，单击 “OK”，即可输出结果。

2.3 分类和结果分析

图 1 所示为分类结果，由图可知将变量点与样品点进行分类，分类如下：

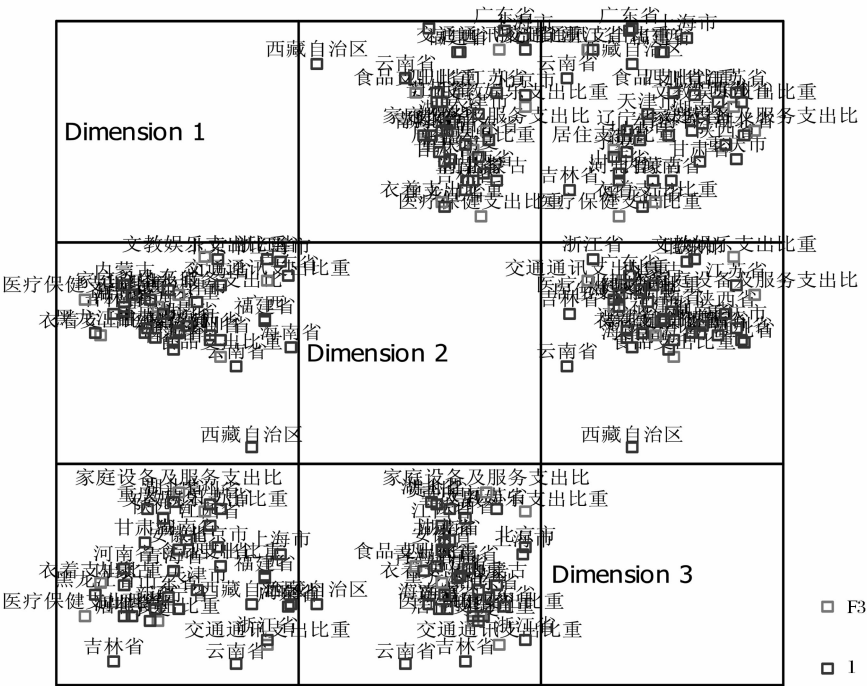


图 1 分类结果

第一类:变量为 X_1, X_2
样品点为黑龙江,重庆,湖南,安徽,甘肃,新疆,湖北,江西,贵州,四川,青海,云南,辽宁,山西,西藏

第二类:变量为 X_3, X_4, X_7
样品点为内蒙古,天津,河南,宁夏,陕西,吉林,河北,山东

第三类:变量为 X_5, X_6
样品点为北京,上海,广东,浙江,江苏,福建,广西,海南

在第一类中,变量为衣着,食品支出占总支出的比重,省份有:黑龙江,重庆,湖南,安徽,甘肃,新疆,湖北,江西,贵州,四川,青海,云南,辽宁,山西,它们位于我国北部与西南部,说明这些省份的消费支出结构相似。且由恩格尔系数的解释知,食品支出比重较大的地区,经济发展较落后,因此国家应对这些省份做出相应的扶持,并提升区域政策的灵活性。

在第二类中,变量为居住,医疗保健,家庭设备及服务支出占总支出的比重,省份有:内蒙古,天津,河

南,宁夏,陕西,吉林,河北,山东,它们在我国的中东部,说明这些省份的消费支出结构相似。

在第三类中,变量为交通和通讯,文教娱乐用品及服务支出占总支出的比重,省份有:北京,上海,广东,浙江,江苏,福建,广西,海南,相对于其它省份这几个省份经济要高很多,因此这几个省份的消费结构相似。

通过上面的分类,可以发现类别的区分,具有明显的地域性,说明农村消费结构与地域有一定的关联性。东部沿海地区的消费较高,而东北及西部地区主要以满足生活需求消费指标为主,这也与我国实际地区的发展情况相符。

参考文献

[1] 于秀林,任雪松.多元统计分析[M]. 北京:中国统计出版社,1999:199—210.

[2] 杨纶. SPSS 统计分析从基础到实践[M]. 广州:华南理工大学出版社,2005:88—107.

The Application of Correspondence Analysis in the Structure of the Rural Family Consumption

ZHANG Yong-li

(College of Science, Heilongjiang Institute of Science and Technology, Haerbin 150027, China)

Abstract: The level of consumption is an important index for a national economic strength or region. Due to the different level of economic development, the different culture, the different customs, the different consumer preferences the spending is not the same. The data of per capita consumption and consumption structure of rural residents of 31 provinces in China on the 2009 is analyzed using statistical methods and SPSS software in this paper. The classification result of 31 areas is given and the regulativity of consumption and very significance of conclusions are obtained. So we can give some science feasible of recommendations for the problem of currently rural areas in China.

Key words: correspondence analysis; factor analysis; factor loadings; probability matrix

(上接第 45 页)

Analysis on the Competition Ability of Hi-tech Companies in Nanjing

XU Ai-hua, LI Hong

(Institute of Scientific and Technical Information of Nanjing, Nanjing 210018, China)

Abstract: In this paper the datas of Hi-tech companies of twelve deputy provincials are selected. And the competition ability of the Hi-tech companies in nanjing is discussed based on the comparative analysis on the total, per capita, economic and scientific indexes. All of the results indicate that the Hi-tech companies in nanjing are more competitive in some indexes, such as per capita output value, labour productivity and so on. But the the indexes of export scale and R&D financial input have their drawbacks.

Key words: Nanjing; hi-tech companies; competition ability