

基于主成分分析法的中国城镇居民消费结构研究*

Principal Component Analysis on China's Consumption Structure of Urban Residents

赵丽棉, 黄基廷

ZHAO Li-mian, HUANG Ji-ting

(河池学院数学系, 广西宜州 546300)

(Department of Mathematics of Hechi University, Yizhou, Guangxi 546300, China)

摘要: 以 2009 年全国 31 个省、市、自治区的城镇居民家庭平均每人全年消费性支出的食品、衣着、居住、家庭设备用品及服务、医疗保健、交通与通讯、娱乐教育文化服务、其它商品和服务等 8 个指标数据为依据, 利用 R 统计软件, 采用主成分分析法对当前城镇居民消费结构进行分析。结果显示: 娱乐教育文化服务、交通通讯、家庭设备用品、居住、食品是影响消费大小变动的主要因素, 而衣着、医疗保健、居住、食品是影响消费结构变动的主要因素; 各省市城镇居民消费大小与其经济发达程度密切相关; 相邻省市消费结构比较相似; 沿海地区与内地消费结构有较大的差别。

关键词: 城镇居民 消费结构 相关性 主成分分析法

中图分类号: O212.4 文献标识码: A 文章编号: 1005-9164(2012)02-0121-04

Abstract: Based on the eight indicator data of the Per Capita Annual Living Expenditure of Urban Residents from 31 provinces, cities and autonomous regions of China in 2009, the current consumption of urban residents structure was analyzed using the R statistical software to adopt the principal component analysis. The results showed that recreation, education and cultural services, transportation and communications, household appliances, housing, and food are the main factors affecting the size of changes in consumption, while clothing, medical care, housing, and food are the main factors that affect the changes in consumption structure; the size of consumption of urban residents in provinces and cities is closely related to the level of their economic development; it is quite similar to the adjacent provinces and cities in the consumption structure; and there are large differences in consumption structure between the inland and coastal regions.

Key words: urban residents, consumption structure, correlation, principal component analysis

国民经济是一个有机整体, 是生产、交换、分配、消费四大环节不断运动的统一。消费具体表现为消费数量、消费增长速度、消费结构等。消费结构是消费的核心内容, 它是社会经济发展的结果, 又是制约社会经济发展的重要因素。因此分析当前居民消费结构对于消费结构的优化、生产力的合理配置、各产业的协调发展以及国家经济政策的制定都有非常重要的意义。卢方元等^[1]利用 Panel Data 模型, 分析中国农村

居民不同收入层次之间的消费差异, 以及这种差异对农村居民消费结构的影响; 祝伟等^[2]运用平行数据的基本模型、变截距模型和变系数模型, 分析中国城镇居民收入差异及其变动对消费结构变化的影响; 张雅清^[3]、王巧英^[4]应用扩展线性支出系统对中国城镇居民消费结构进行了分析。本文选取消 2009 年中国 31 个省、市、自治区城镇居民消费指标, 利用 R 统计软件, 采用主成分分析法^[5], 研究中国城镇居民的消费结构。

1 中国城镇居民消费指标选取

根据 2009 年全国 31 个省、市、自治区的城镇居民家庭平均每人全年消费性支出的 8 个指标数据 (数据来源: 2010 年中国统计年鉴): 食品 X_1 、衣着

收稿日期: 2011-09-07

修回日期: 2011-12-29

作者简介: 赵丽棉 (1964-), 女, 副教授, 主要从事数理统计教学与应用研究工作。

* 新世纪广西高等教育教改工程项目 (广西教育厅 2011JGA098), 广西自然科学基金项目 (2010GXNSFB013051) 资助。

X_2 、居住 X_3 、家庭设备用品及服务 X_4 、医疗保健 X_5 、交通与通讯 X_6 、娱乐教育文化服务 X_7 、其它商品和服务 X_8 ,算出各项指标支出占总消费支出的比重(表1)。

表1 各项指标消费支出比重(%)

Table 1 Ratio of different consumptions

范围 Range	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
全国	36.52	10.47	10.02	6.42	6.98	13.72	12.01	3.87
北京	33.18	10.04	7.21	6.85	7.77	15.47	14.84	4.66
天津	36.51	9.21	10.17	6.16	8.60	13.30	11.76	4.28
河北	33.59	12.30	11.81	6.49	10.04	11.89	10.15	3.74
山西	32.84	12.42	14.10	6.03	8.44	11.71	11.44	3.01
内蒙古	30.50	15.01	10.07	6.45	8.03	12.59	12.16	5.19
辽宁	37.98	10.86	10.49	4.93	8.26	12.12	10.42	4.94
吉林	33.33	13.00	12.78	4.98	10.27	11.96	9.42	4.26
黑龙江	35.28	14.58	10.66	5.69	10.16	9.58	9.94	4.11
上海	34.99	7.59	9.11	6.50	4.77	16.67	14.95	5.41
江苏	36.29	9.87	8.73	7.02	6.15	13.09	14.96	3.88
浙江	33.59	9.68	8.91	4.97	5.90	19.72	13.76	3.47
安徽	39.59	10.55	11.92	5.76	7.00	9.90	11.97	3.30
福建	39.67	8.71	10.37	6.39	4.40	14.82	11.19	4.45
江西	39.85	10.81	9.60	7.82	5.65	11.76	10.95	3.55
山东	32.92	12.89	10.66	7.37	7.37	14.32	11.10	3.39
河南	34.21	13.28	10.50	7.16	9.15	10.81	10.96	3.94
湖北	40.42	11.76	9.71	7.38	6.75	9.26	11.74	2.99
湖南	38.55	10.59	9.92	7.37	7.25	11.39	11.15	3.77
广东	36.93	6.31	10.76	6.24	5.49	17.68	12.87	3.72
广西	39.89	8.26	9.86	7.29	5.20	15.44	10.73	3.32
海南	44.69	5.77	9.92	5.81	5.99	15.35	9.54	2.94
重庆	37.68	12.38	9.23	8.59	8.09	9.79	11.13	3.10
四川	40.44	10.85	8.96	6.25	5.97	13.04	10.60	3.89
贵州	41.51	11.19	8.26	6.51	5.92	10.87	12.67	3.08
云南	43.72	10.80	9.25	3.85	6.95	15.56	7.83	2.03
西藏	50.71	12.03	7.63	3.95	3.90	11.76	5.16	4.86
陕西	37.26	11.30	9.51	6.38	8.06	10.01	13.36	4.11
甘肃	37.78	13.16	9.01	6.29	8.40	10.06	11.53	3.77
青海	40.39	11.88	9.00	5.75	7.98	11.11	10.12	3.78
宁夏	33.39	12.26	10.97	6.20	8.97	13.26	10.47	4.48
新疆	36.30	14.55	9.19	5.92	7.33	12.85	9.17	4.68

基于联合国粮农组织提出的用恩格尔系数(即食品支出占消费支出的比重)判定生活发展阶段的一般标准:60%以上为贫困;50%~60%为温饱;40%~50%为小康;40%以下为富裕.从表1结果可以看出:2009年中国城镇居民生活整体上都达到小康水平.除了西藏、海南、云南、贵州、四川、湖北、青海外其它24个省、市、自治区都达到富裕水平.从全国平均

数看,基本需求比重居第一位的是食品(36.52%),其次是交通通讯13.72%、娱乐教育文化服务(12.01%);而衣着、居住类的基本需求比重居中;居最后三位的是医疗保健(6.98%)、家用设备用品及服务(6.42%)、其它商品和服务(3.87%),说明目前城镇居民对食品的需求主要是基本需求,而对医疗保健、家用设备用品及服务等的的需求主要属于发展和享受型的需求.

2 中国城镇居民消费结构

2.1 消费指标的相关性分析

从表2可以看到,各项指标都显正相关关系,且系数都比较大.除了衣着与食品、居住、家庭设备用品及服务、交通与通讯;医疗保健与食品、交通及通讯以外,其它各项指标之间的相关系数都达0.5以上,尤其是娱乐教育文化服务与其它指标的相关系数比较大,达0.8以上的就有4个,说明了中国城镇居民消费性支出的8个指标是相互促进的.

表2 各项指标的相关系数

Table 2 Correlation coefficient between different consumptions

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	1.0000	0.2525	0.6899	0.7421	0.2970	0.8827	0.8281	0.7826
X_2	0.2525	1.0000	0.4257	0.4978	0.7133	0.3973	0.5342	0.6102
X_3	0.6899	0.4257	1.0000	0.6979	0.6058	0.7911	0.7764	0.7335
X_4	0.7421	0.4978	0.6979	1.0000	0.5142	0.7332	0.8920	0.7561
X_5	0.2970	0.7133	0.6058	0.5142	1.0000	0.4331	0.5692	0.5571
X_6	0.8827	0.3973	0.7911	0.7332	0.4331	1.0000	0.8781	0.7897
X_7	0.8281	0.5342	0.7764	0.8920	0.5692	0.8781	1.0000	0.8494
X_8	0.7826	0.6102	0.7335	0.7561	0.5571	0.7897	0.8494	1.0000

2.2 消费指标的主成分分析

8项指标值做主成分分析(Components,简写成Comp.)得表3结果和载荷因子矩阵.

表3 主成分分析结果

Table 3 Results of principal component analysis

	主成分标准差 (特征值的开方) Standard deviation	贡献率 Contribution rate	累积贡献率 Cumulative contribution rate
Comp. 1	2.3748	0.7049	0.7049
Comp. 2	1.0748	0.1444	0.8493
Comp. 3	0.6475	0.0524	0.9018
Comp. 4	0.5455	0.0372	0.9389
Comp. 5	0.4296	0.0231	0.9620
Comp. 6	0.4073	0.0207	0.9828
Comp. 7	0.2981	0.0111	0.9939
Comp. 8	0.2216	0.0061	1.0000

载荷因子矩阵:

	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6	Comp. 7	Comp. 8
X_1	-0.355	-0.423	0.106	0.169	0.388		0.662	0.253
X_2	-0.265	0.629	0.460	0.260	-0.265	-0.325	0.261	0.103
X_3	-0.363		-0.653		-0.618		0.164	0.143
X_4	-0.371		0.237	-0.771	-0.185		0.163	-0.378
X_5	-0.283	0.582	-0.455	-0.142	0.576	0.114		
X_6	-0.379	-0.267		0.311	0.136	-0.566	-0.278	-0.515
X_7	-0.403		0.117	-0.231		-0.157	-0.530	0.670
X_8	-0.385		0.260	0.361		0.727	-0.285	-0.200

从载荷因子矩阵可以看出,前3个特征值的累积贡献率已达90%,从第4个特征值开始,其数值有显著减少(都小于 $0.5455^2 = 0.2976$),所以只取前3个主成分,它们分别是

$$z_1 = -0.355X_1 - 0.265X_2 - 0.363X_3 - 0.371X_4 - 0.283X_5 - 0.379X_6 - 0.403X_7 - 0.385X_8,$$

$$z_2 = -0.423X_1 + 0.629X_2 + 0.582X_5 - 0.267X_6,$$

$$z_3 = 0.106X_1 + 0.460X_2 - 0.653X_3 + 0.237X_4 - 0.455X_5 + 0.117X_7 + 0.260X_8.$$

第一主成分对应系数全为负,其值在-0.265到-0.403之间,它反映了一个省市人均消费大小程度:消费较大的省市,它的8项指标数值都比较大,因此,第一主成分的值就较小;相反,消费较小的省市,第一主成分的值就较大,称第一主成分为大小因子。第二主成分是衣着、医疗保健与食品、交通与通讯的差,衣着、医疗保健支出较大而食品、交通与通讯支出较小的省市第二主成分的值就较大,表明该省市倾向于发展和享受型消费支出。第三主成分是各项指标(除交通通讯外)与居住、医疗保健的差,居住与医疗保健支出较大的省、市第三主成分的值就较小,表明该省市也倾向于发展和享受型消费支出。所以也可称第二、三主成分为类型或结构因子。从31个省、市、自治区的主成分值排序结果(表4)可以看出,经济发达地区的消费能力明显高于欠发达地区;相邻省市消费结构比较类似但是南方地区与北方地区的消费结构差别较大;内地与沿海的消费结构差别较大,符合中国城镇居民消费的实际情况。

3 讨论

(1) 娱乐教育文化服务、交通通讯、家庭设备用品、居住、食品是影响城镇居民消费大小变动的主要因素,而居住、衣着、医疗保健是影响城镇居民消费结构变动的主要因素。在主成分表达式中系数绝对值越大表明该项对主成分的影响就越大,在第一主成分表达式中,系数绝对值从大到小排列依次为:其它商品和服务、娱乐教育文化服务、交通通讯、家庭设备用

表4 各地区的主成分值

Table 4 Principal component value in different regions

序号 No.	地名 Place name	Comp. 1	地名 Place name	Comp. 2	地名 Place name	Comp. 3
1	上海	-7.5160	海南	-2.1851	广东	-1.3006
2	北京	-5.2798	广东	-1.9400	吉林	-1.1929
3	广东	-3.7598	广西	-1.6277	山西	-1.0655
4	浙江	-3.5402	福建	-1.5496	天津	-0.9689
5	天津	-2.7302	上海	-1.4151	海南	-0.8511
6	内蒙古	-1.3794	西藏	-1.2533	河北	-0.7996
7	福建	-1.1725	江西	-0.8113	安徽	-0.5740
8	江苏	-1.1620	贵州	-0.7614	辽宁	-0.4707
9	辽宁	-0.6904	四川	-0.5667	云南	-0.3780
10	重庆	-0.6697	云南	-0.5329	宁夏	-0.3726
11	山东	-0.6496	江苏	-0.3815	黑龙江	-0.1639
12	吉林	0.0410	安徽	-0.3429	浙江	-0.0566
13	湖南	0.5788	浙江	-0.1891	广西	-0.0426
14	陕西	0.6063	湖南	-0.1843	河南	-0.0252
15	宁夏	0.6676	湖北	-0.1262	湖南	0.0502
16	四川	0.8812	青海	-0.0908	陕西	0.1018
17	河北	1.1311	陕西	0.3309	山东	0.1222
18	湖北	1.1613	甘肃	0.4095	青海	0.1738
19	安徽	1.1963	山西	0.4633	福建	0.1854
20	黑龙江	1.2053	新疆	0.5945	甘肃	0.3599
21	河南	1.2342	辽宁	0.5975	湖北	0.3715
22	广西	1.3576	天津	0.6198	重庆	0.4697
23	山西	1.4730	宁夏	0.7957	四川	0.5098
24	江西	1.6048	山东	0.8909	江西	0.5116
25	新疆	1.6811	河南	0.9036	江苏	0.5480
26	海南	1.9594	河北	0.9316	上海	0.5795
27	云南	1.9821	重庆	0.9427	贵州	0.6489
28	甘肃	2.1252	北京	1.2190	新疆	0.7655
29	青海	2.4069	黑龙江	1.5463	内蒙古	0.8074
30	贵州	2.4157	吉林	1.6377	北京	0.8442
31	西藏	2.8410	内蒙古	2.0748	西藏	1.2128

品、居住、食品、医疗保健、衣着,而前六项的系数绝对值都大于0.35,所以它们是影响消费大小变动的主要因素。在第二主成分表达式中,绝对值大于0.45的项有(从大到小)衣着、医疗保健;生第三主成分表达式中,绝对值大于0.45的项有(从大到小):居住、衣着、医疗保健。说明居住、衣着、医疗保健是影响消费结构变动的主要因素。虽然食品消费支出占总消费支出的比重最大,但已不是影响消费大小变动的第一主要因素,说明居民的消费观念在改变,消费重点已经逐渐由物质产品消费转向精神产品和服务消费,由基本生活型逐步转向享受型发展。

(2) 各省市城镇居民消费大小与其经济发达程度密切相关,经济越发达消费能力越大。从第一主成分得分情况(表4)看,从小到大排列前10名依次为上海28837.78(1),北京26738.48(2),广东21574.72(4),浙江24610.81(3),天津21402.01(5),内蒙古15849.19(9),福建19576.83(7),江苏20551.72

(6) 辽宁 15761.38(10)、重庆 15748.67(11),其中省市后面数值及括号里的数字分别表示该省市 2009 年人均可支配收入(数据来源:2010 年中国统计年鉴)和收入排名,如上海 28837.78(1),表示上海 2009 年人均可支配收入 28837.78 元,排在全国 31 个省、市、自治区的第 1 位(按高到低排)。这些省市除了内蒙古和重庆外都是东部地区,经济都比较发达。而内蒙古是国家重点扶植的少数民族地区;重庆作为全国直辖时间最短的城市,有很大的发展潜力,正在施行的“宜居重庆”举措,为重庆带来了经济发展机遇。这 10 个省市的人均可支配收入都排在全国的前 11 位。而第一主成分得分排在后 10 名的依次为广西 15451.48(12)、山西 13996.55(23)、江西 14021.54(21)、新疆 12257.52(30)、海南 13750.85(25)、云南 14423.93(15)、甘肃 11929.78(31)、青海 12691.85(28)、贵州 12862.53(27)、西藏 13544.41(26),这些省市除了海南外都是中、西部地区,经济都比较落后,人均可支配收入都排在全国的后 10 位(除广西、云南外)。这充分说明了各省市城镇居民消费大小与其经济发达程度密切相关,经济越发达消费越大,也说明了我们的分析结果是符合实际情况的。

(3) 相邻省市消费结构比较相似。从第二主成分看(表 4)得分越高,说明该省市在衣着、医疗保健上的支出比在食品、交通通讯上的支出越多。从表 4 看到,北方地区得分较高而南方地区得分较低,且相邻省市的第二主成分得分比较接近,比如黑龙江、吉林、内蒙古得分都比较高,在 1.5463 到 2.0748 之间;海南、广东、广西、福建得分都比较低,在 -2.1851 到 -

1.5496 之间;此外,四川与贵州、云南;江苏与安徽;湖南与湖北;陕西与甘肃、山西;辽宁与天津;山东与河南、河北的得分都比较接近,说明了相邻省市消费结构比较相似,这种相似性的产生有着多方面的原因,诸如:经济发展水平相近、居民消费观念相似、生活方式类似、生产方式近似、产业结构雷同等等。

(4) 沿海地区或中东部地区与内地消费结构有较大的差别。从第三主成分看(表 4)得分越低,说明该省市在居住、医疗保健上的支出比在其它指标上的支出越多,越倾向于发展和享受型消费支出。从表 4 看到,沿海地区或中东部地区得分较低,而内地得分较高,排在前八名的都是沿海地区或中东部地区。说明中国的改革开放政策,拉动了沿海地区的经济发展,从而促进了居民的居住和医疗保健的消费,逐步拉开了沿海与内地的生活水平差距。

参考文献:

- [1] 卢方元,鲁敏.中国农村居民消费结构的 Panel Data 模型分析[J].数理统计与管理,2009,28(1):122-127.
- [2] 祝伟,李元生.中国城镇居民消费结构的平行数据分析[J].数理统计与管理,2006,25(6):645-648.
- [3] 张雅清.中国城镇居民消费结构的实证分析[J].长春师范学院学报:自然科学版,2005,24(5):75-76.
- [4] 王巧英.近年来中国城镇居民消费结构变化的定量分析[J].内蒙古财经学院学报,2005(5):67-71.
- [5] 薛毅,陈立春.统计建模与 R 软件[M].北京:清华大学出版社,2007.

(责任编辑:尹 闯)

研究查明植物耐高温逆向调控机制

植物在高温胁迫下会产生应激防御反应,启动体内大量热激转录因子(HSF)和热激蛋白(HSP)基因的转录,从而维护细胞和叶绿体的稳定性,但对于启动这些热激响应基因表达的信号来源并不清楚。最近,中科院上海生命科学研究院植物生理生态所的科研人员开展了植物高温胁迫响应蛋白的鉴定工作。他们的研究表明,叶绿体核糖体蛋白 RPS1 参与类囊体膜蛋白的翻译,并且 RPS1 的表达水平以剂量依赖的方式调控类囊体膜结构的稳定性。尤为重要的是,RPS1 表达水平下调导致拟南芥突变体对高温胁迫极度敏感,其原因是 RPS1 突变体在高温胁迫条件下热激转录因子 HsfA2 及其下游靶基因的表达受到严重抑制。同时,组成型表达 HsfA2 可将 RPS1 类囊体膜稳定性和耐热性恢复至野生型水平。就此,科研人员提出了植物细胞热激反应逆向调控机制模型:RPS1 作为叶绿体蛋白翻译调控的关键因子,其蛋白表达水平受高温胁迫的诱导;RPS1 表达增强可提高类囊体膜蛋白的翻译效率,对于维持高温胁迫下叶绿体的功能状态和产生质体逆向信号是必要的。产生的质体信号通过相关的热激信号转导组分传递到细胞核,从而启动 HsfA2 和其下游靶基因的热激响应表达。而 HsfA2 下游靶基因编码的叶绿体定位的热激蛋白如 HSP21 等进入叶绿体,对高温胁迫下的叶绿体类囊体膜系统进行保护。专家认为,这项研究为细胞核-质体信号互作参与植物逆境胁迫适应机制提供了新证据,为进一步研究植物的耐热性状形成机理开启了新视角。同时,该发现为通过调控质体翻译效率增强农作物的耐高温胁迫能力提供了全新的遗传改良操作路径。

(据科学网)