

我国各地区城镇居民消费性支出的分析研究

王学民

(上海财经大学 统计系, 上海 200433)

摘要:我国各地区的经济发展水平存在着较大的差异。本文运用主成分分析方法对我国各地区城镇居民消费性支出进行分析研究,研究表明:各地区城镇居民消费性支出的差异主要是由两方面引起的,首先是地区的经济发展水平,我国东部、中部和西部地区的消费水平存在着较大差异;其次是由地区气候因素引起的消费倾向,我国南北地区明显有别。

关键词:消费性支出;主成分分析;消费水平;消费倾向

中图分类号:F063 **文献标识码:**A **文章编号:**1001-9952(2002)01-0064-06

一、引言

我国地域辽阔,各地区的经济发展很不平衡,各地区城镇间的消费性支出(由表 1 中的 X_1, \dots, X_8 这八个指标来描述)存在着较大的差异,而且由于多种因素的影响,这种差异呈现加速扩大的态势。如何客观、准确、有效地分析这些差异,具有重要的理论和实践意义。消费性支出的指标有许多,如果直接从诸多指标来分析各地区的差异,那末分析的结果很可能将是繁杂和不得要领的,很难给出直观有效的结论。如果仅用消费性总支出这个指标,则显得太粗糙,丢失的有用信息太多,不能较充分地反映各地区的消费差异。那么,如何能使得所作的分析研究即不繁杂又不损失太多的信息呢?这正是本文所要解决的问题。

本文运用多元统计分析中的主成分分析方法,将描述各地区城镇居民消费性支出的八个指标压缩成两个综合指标(称为主成分),这两个主成分保留了原始八个指标的绝大部分信息,在指标压缩的同时能够最大限度地反映出各地区消费性支出的差异。本文给出了这两个主成分的符合实际意义的解释,并用它们来描述分析各地区城镇居民消费性支出的主要差异。

二、方法研究与实证分析

(一)样本数据选择

本文选择的样本数据来自于《中国统计年鉴》(2000)提供的全国 31 个省、直辖市和自治区(港、澳、台不在其中)的城镇居民家庭平均每人全年消费性支出的 9 项主要指标(单位:元),指标代码和指标名称列于表 1,原始数据列于表 2。

表 1 指标代码和指标名称

指标代码	指标名称	指标代码	指标名称
X_1	食品	X_6	娱乐教育文化服务
X_2	衣着	X_7	居住
X_3	家庭设备用品及服务	X_8	杂项商品和服务
X_4	医疗保健	$S = (\sum_{i=1}^8 X_i)$	消费性总支出
X_5	交通和通讯		

收稿日期:2001-09-28

作者简介:王学民(1964-),男,上海人,上海财经大学统计系讲师。

表 2 原始数据表

(单位:元)

指标 地区	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	S
北京	2959.19	730.79	749.41	513.34	467.87	1141.82	478.42	457.64	7498.48
天津	2459.77	495.47	697.33	302.87	284.19	735.97	570.84	305.08	5851.53
河北	1495.63	515.90	362.37	285.32	272.95	540.58	364.91	188.63	4026.30
山西	1406.33	477.77	290.15	208.57	201.50	414.72	281.84	212.10	3492.98
内蒙古	1303.97	524.29	254.83	192.17	249.81	463.09	287.87	192.96	3468.99
辽宁	1730.84	553.90	246.91	279.81	239.18	445.20	330.24	163.86	3989.93
吉林	1561.86	492.42	200.49	218.36	220.69	459.62	360.48	147.76	3661.68
黑龙江	1410.11	510.71	211.88	277.11	224.65	376.82	317.61	152.85	3481.74
上海	3712.31	550.74	893.37	346.93	527.09	1034.98	720.33	462.03	8247.69
江苏	2207.58	449.37	572.40	211.92	302.09	585.23	429.77	252.54	5010.91
浙江	2629.16	557.32	689.73	435.69	514.66	795.87	575.76	323.36	6521.54
安徽	1844.78	430.29	271.28	126.33	250.56	513.18	314.00	151.39	3901.81
福建	2709.46	428.11	334.12	160.77	405.14	461.67	535.13	232.29	5266.69
江西	1563.78	303.65	233.81	107.90	209.70	393.99	509.39	160.12	3482.33
山东	1675.75	613.32	550.71	219.79	272.59	599.43	371.62	211.84	4515.05
河南	1427.65	431.79	288.55	208.147	217.00	337.76	421.31	165.32	3497.53
湖北	1783.43	511.88	282.84	201.01	237.60	617.74	523.52	182.52	4340.55
湖南	1942.23	512.27	401.39	206.06	321.29	697.22	492.60	226.45	4799.51
广东	3055.17	353.23	564.56	356.27	811.88	873.06	1082.82	420.81	7517.81
广西	2033.87	300.82	338.65	157.78	329.06	621.74	587.02	218.27	4587.22
海南	2057.86	186.44	202.72	171.79	329.65	477.17	312.93	279.19	4017.75
重庆	2303.29	589.99	516.21	236.55	403.92	730.05	438.41	225.80	5444.23
四川	1974.28	507.76	344.79	203.21	240.24	575.10	430.36	223.46	4499.19
贵州	1673.82	437.75	461.61	153.32	254.66	445.59	346.11	191.48	3964.35
云南	2194.25	537.01	369.07	249.54	290.84	561.91	407.70	330.95	4941.26
西藏	2646.61	839.70	204.44	209.11	379.30	371.04	269.59	389.33	5309.12
陕西	1472.95	390.89	447.95	259.51	230.61	490.90	469.10	191.34	3953.25
甘肃	1525.57	472.98	328.90	219.89	206.65	449.69	249.66	228.19	3681.50
青海	1654.69	437.77	258.78	303.00	244.93	479.53	288.56	236.51	3903.76
宁夏	1375.46	480.89	273.84	317.32	251.08	424.75	228.73	195.93	3547.99
新疆	1608.82	536.05	432.46	235.82	250.28	541.30	344.85	214.40	4163.98

(二)方法研究

为了达到对 X_1, \dots, X_8 这八个指标的降维目的,我们可以使用多元统计分析中的主成分分析法。主成分分析是一种通过降维技术把多个指标约化为少数几个综合指标的统计分析方法,这些称为主成分的综合指标能够反映出原始指标的绝大部分信息。利用它们可进行简洁、有效的分析研究,甚至还可借助于图形作直观的分析。主成分分析的基本步骤如下:

设有 n 个地区, p 个消费性支出指标,则数据矩阵为 $X=(X_{ij}):n \times p$,其中 X_{ij} 为第 i 个地区的第 j 个消费性支出指标。

(1)各指标数值的大小相差较大(X_1 值较大,而 X_4, X_5, X_7 和 X_8 值较小),如果直接使用主成分分析方法,则主成分会过于照顾数值较大的指标(如 X_1),而对数值较小的指标却照顾不够。为使主成分分析能够均等地对待每一个指标,消除由于数值大小的差异而可能带来的一些不利影响,需对各指标作标准化处理。即令:

$$X_{ij}^* = \frac{X_{ij} - \bar{X}_j}{S_j}$$

其中 \bar{X}_j 为第 j 个消费性支出指标的样本均值, S_j 为第 j 个消费性支出指标的样本标准差, $j=1, \dots, p; i=1, \dots, n$ 。指标 X_1, \dots, X_p 经标准化后,记为 X_1^*, \dots, X_p^* 。

(2)求出 X_1^*, \dots, X_p^* 的协方差矩阵,也就是 X_1, \dots, X_p^* 的相关矩阵 R 。

(3)求 R 的特征值及相应的一组正交单位特征向量。

(4)计算累计贡献率,确定主成分个数,使累计贡献率达到一个较高的百分比。给出所使用主成分的符合实际意义的解释。

(5)计算 n 个地区的主成分得分,然后根据得分对各地区的消费性支出情况作比较分析。

(三)实证分析

本文下面的所有计算和作图都是通过 SAS6.12 软件做成的。首先求得相关矩阵 R(从略),再算得 R 的前三个特征值、特征向量以及贡献率,如表 3 所示。

表 3 R 的前三个特征值、特征向量以及贡献率

特征向量	t ₁	t ₂	t ₃
X ₁ *	0.401	-0.077	0.415
X ₂ *	0.132	0.749	0.332
X ₃ *	0.375	0.065	-0.442
X ₄ *	0.320	0.345	-0.478
X ₅ *	0.388	-0.232	0.279
X ₆ *	0.406	0.027	-0.310
X ₇ *	0.326	-0.496	-0.034
X ₈ *	0.396	0.096	0.345
特征值	5.098	1.352	0.574
贡献率	0.637	0.169	0.072
累计贡献率	0.637	0.806	0.878

从表 3 可以看出,前两个主成分的累计贡献已达 80.6%,因此可以考虑只取前面两个主成分,它们能很好概括 X₁…X₈ 这八个指标。前两个主成分可分别表述为:

$$Y_1 = 0.401X_1^* + 0.132X_2^* + 0.375X_3^* + 0.320X_4^* + 0.388X_5^* + 0.406X_6^* + 0.326X_7^* + 0.396X_8^*$$

$$Y_2 = -0.077X_1^* + 0.749X_2^* + 0.065X_3^* + 0.345X_4^* - 0.232X_5^* + 0.027X_6^* - 0.496X_7^* + 0.096X_8^*$$

第一主成分 Y₁ 对所有变量(除对 X₂* 的载荷稍偏小外)都有近似相等的载荷,因此可认为是对所有消费性支出指标的度量,反映了综合消费性支出的水平,它与消费性总支出 S 有很强的正相关性(ρ=0.989)。第二主成分 Y₂ 在变量 X₂* 上有很高的正载荷;而在其余变量上有负载荷或很小的正载荷。可以认为这个主成分是由于度量受地区气候影响的消费性支出(主要是衣着 X₂,其次是医疗保健 X₄^①)在所有消费性支出中占的比重(也可理解为一种消费的倾向)。前两个主成分与原始变量的相关矩阵列于表 4。

表 4 主成分与原始变量的相关矩阵

原始变量 \ 主成分	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
Y ₁	0.906	0.298	0.847	0.723	0.876	0.916	0.737	0.895
Y ₂	-0.090	0.871	0.076	0.401	-0.270	0.032	-0.577	0.112

表 5 和表 6 是把 31 个地区分别按第一主成分和第二主成分得分从小到大重新排序后的结果,在地区一栏中地名后面跟着的是汉语拼音的头一个字母。从表 5 可以看出,东部地区的第一主成分得分普遍较高,中部地区一般,而西部地区则普遍较低。从表 6 可见,北方地区的第二主成分得分普遍较高,而南方地区则普遍较低,这是由于北方地区气候寒冷,用于衣着、医疗保健等消费的比重相对较高,而南方地区则相反。图 1 是关于第一和第二主成分的散点图,它对各地区的综合消费性支出和受地区气候影响的消费性支出占的比重有较直观的描述。图中各散点是用相应地区地名汉语拼音的头一个字母标出的,对重复的字符,可对照表 5 和表 6 加以辨别。

表 5 按第一主成分排序的 31 个地区

地区	S	Y ₁	Y ₂	地区	S	Y ₁	Y ₂
江西 j	3482.33	-2.234	-1.867	新疆 x	4163.98	-0.697	0.647
河南 h	3497.53	-1.947	-0.388	四川 s	4499.19	-0.533	0.041
黑龙江 h	3481.74	-1.927	0.636	广西 g	4587.22	-0.251	-2.052
吉林 j	3661.68	-1.859	0.151	山东 s	4515.05	-0.147	0.983
山西 s	3492.98	-1.848	0.404	福建 f	5266.69	0.201	-1.337
内蒙古 n	3468.99	-1.826	0.509	湖南 h	4799.51	0.219	-0.203
安徽 a	3901.81	-1.796	-0.519	江苏 j	5010.91	0.407	-0.311
甘肃 g	3681.50	-1.549	0.526	云南 y	4941.26	0.435	0.479
宁夏 n	3547.99	-1.501	0.906	西藏 x	5309.12	0.437	2.365
辽宁 l	3989.93	-1.313	0.844	重庆 c	5444.23	1.115	0.409
贵州 g	3964.35	-1.298	-0.341	天津 t	5851.53	2.006	0.044
海南 h	4017.75	-1.157	-1.913	浙江 z	6521.54	3.583	0.531
青海 q	3903.76	-1.045	0.426	北京 b	7498.48	5.426	2.466
陕西 s	3953.25	-0.859	-0.501	广东 g	7517.81	5.583	-3.072
河北 h	4026.30	-0.769	0.580	上海 s	8247.69	5.866	-0.195
湖北 h	4340.55	-0.717	-0.247				

表 6 按第二主成分排序的 31 个地区

地区	Y ₁	Y ₂	地区	Y ₁	Y ₂	地区	Y ₁	Y ₂
广东 g	5.583	-3.072	湖南 h	0.219	-0.203	浙江 z	3.583	0.531
广西 g	-0.251	-2.058	上海 s	5.866	-0.195	河北 h	-0.769	0.580
海南 h	-1.157	-1.913	四川 s	-0.533	0.041	黑龙江 h	-1.927	0.636
江西 j	-2.234	-1.867	天津 t	2.006	0.044	新疆 x	-0.697	0.647
福建 f	0.201	-1.337	吉林 j	-1.859	0.151	辽宁 l	-1.313	0.844
安徽 a	-1.796	-0.519	山西 s	-1.848	0.404	宁夏 n	1.501	0.906
陕西 s	-0.859	-0.501	重庆 c	1.115	0.409	山东 s	-0.147	0.983
河南 h	-1.947	0.388	青海 q	-1.045	0.426	西藏 x	0.437	2.365
贵州 g	-1.298	-0.341	云南 y	0.435	0.479	北京 b	5.426	2.466
江苏 j	0.407	-0.311	内蒙古 n	-1.826	0.509			
湖北 h	-0.717	-0.247	甘肃 g	-1.549	0.526			

从散点图可以看出,上海、广东和北京在最右边,城镇居民综合消费性支出是最高的,其次是浙江和天津;江西在散点图的最左边,表明综合消费性支出是最低的;北京和西藏在散点图的最上边,说明受地区气候影响的消费性支出占的比重最高;广东在最底部,表明受地区气候影响的消费性支出占的比重最低。

三、分析讨论

本文通过主成分分析将消费性支出的八个原始指标 X_1, \dots, X_8 约化为两个综合指标 Y_1 和 Y_2 , Y_1 和 Y_2 能解释各地区城镇居民消费性支出差异的 80.6%。主成分分析表明,各地区的消费差异首先表现在综合消费性支出的水平上,其次表现在受地区气候影响的消费倾向上。从这两个综合指标出发,画一个如图 1 所示的散点图,可以从直观的图形上来比较各地区消费性支出的差异。并可分别按 Y_1 和 Y_2 数值的大小将 31 个地区分成若干集团。

根据综合指标 Y_1 的数值,我们可以比较出各地区综合消费水平的高低。按在 Y_1 上的得分从高到低大致可分成这样三个集团(每个集团内都按从高到低的得分依次排序):(1)上海、广东、北京、浙江和天津为第一集团。这些地区都为我国经济较发达的东部地区,城镇居民的收入较高,因此有较强的消费能力。(2)重庆、西藏、云南、江苏、湖南、福建、山东、广西、四川、新疆和湖

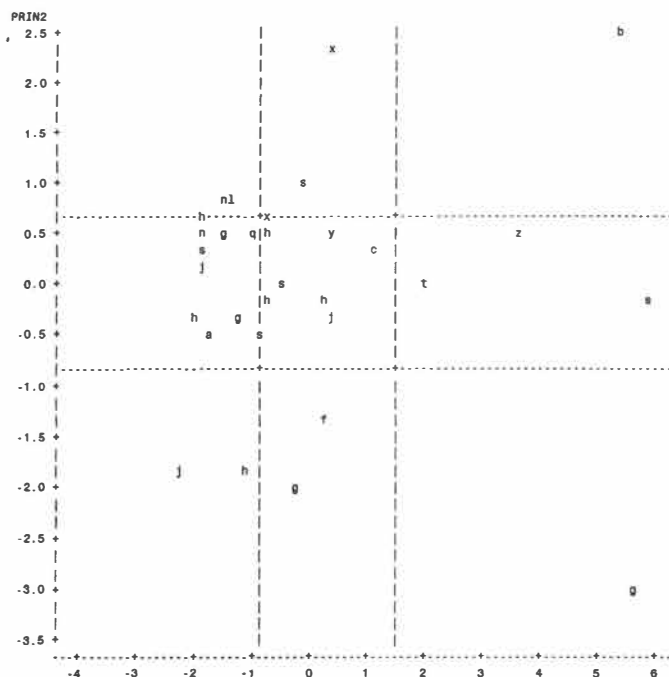


图1 第一和第二主成分散点图

北为第二集团。这些地区主要为我国中部和沿海相对不发达的地区,经济发展水平在我国属于中等,城镇居民的收入水平一般,因而只有中等的消费能力。(3)河北、陕西、青海、海南、贵州、辽宁、宁夏、甘肃、安徽、内蒙古、山西、吉林、黑龙江、河南和江西为第三集团。这些地区大多为我国经济发展水平较落后的中西部地区,消费能力较弱。这三个集团的划分如图1中横轴上的参考线所示。

当前我国国民经济运行的主要问题之一是需求不足,消费需求是整个总需求的一个重要方面。上述分析表明,我国的消费需求呈现出极大的地区不平衡性。这说明,我国当前消费需求不足,就区域而言,是一种地区性需求不足,处于第一集团的沪、粤、京、浙和津已经具有较高的消费需求。而处于第三集团的赣、豫、黑、吉、晋等十五个省、区,则是我国消费需求严重不足的地区,这些地区消费还有很大的提升空间。如何扩大这些地区的消费需求,刺激这些地区的消费增长,应是当前宏观调控的重要内容。

从表5中可以看出,如果按消费性总支出S来划分消费水平高低的集团,所得的结果与上述按反映综合消费性支出的 Y_1 划分的出集团基本相同,而且各集团内的地区排序也大致一样,这是由于 Y_1 与S之间存在着高达 $\rho=0.989$ 的正相关。虽然 Y_1 和S这两个指标的含义很相近,但还是有所不同的,S是原始指标 X_1, \dots, X_8 的总和, X_1, \dots, X_8 中的各指标对S的作用可以有很大的不同,如 X_1 的作用就特别大;而 Y_1 是对 X_1, \dots, X_8 作标准化变换后得到的,依据 Y_1 的表达式可知, X_1, \dots, X_8 中的每个指标对 Y_1 的作用是大致相同的。某地区的 Y_1 值取决于该地区 X_1, \dots, X_8 中的每个指标在所有31个地区中的相对大小,它是这八个指标相对大小的综合值。

根据综合指标 Y_2 的数值,可以比较出各地区受地区气候影响的消费性支出(主要是衣着 X_2 ,其次是医疗保健 X_4)在所有消费性支出中占的比重(或者说是受地区气候影响的消费倾向)。按这种消费倾向从高到低的得分也可大致分为以下三个集团(每个集团都按消费倾向由高到低排序):(1)北京、西藏、山东、宁夏、辽宁、新疆和黑龙江为第一集团。这些地区在地理上大多属于我国的北方地区,气温较为寒冷,衣着支出 X_2 和医疗保健支出 X_4 在所有消费性支出中占的比重都较高。(2)河北、浙江、甘肃、内蒙古、云南、青海、重庆、山西、吉林、天津、四川、上海、湖南、湖北、江苏、贵州、河南、陕西和安徽为第二集团。这些地区多为我国南北向的中部地区。(3)福建、

江西、海南、广西和广东为第三集团。海南、广东和广西位于我国大陆地区的最南端,气温较高,衣着支出 X_2 和医疗保健支出 X_4 在所有消费性支出中占的比重都较低。这三个集团的划分如图 1 中的纵轴参考线所示。

每个地区在 Y_1 和 Y_2 上属于哪个集团,可以从图 1 中一目了然。如果将 Y_1 和 Y_2 结合起来看图 1,则有四个地区(北京 b、上海 s、广东 g 和西藏 x)是明显的“离群点”。表明这四个地区的城镇居民消费性支出状况与绝大多数地区有显著的不同,有一定的特殊性。上海的特殊性完全是由综合消费性水平引起的,西藏的特殊性完全是由受地区气候影响的消费倾向引起的,而北京和广东的特殊性既由综合消费性水平引起又由受地区气候影响的消费倾向引起,因此最具有特殊性。

分析研究各地区城镇居民消费性支出的不同特点,对国家的宏观经济调控和更好地落实扩大内需政策是非常有益的,它有助于政府有的放矢地制定出更合理的政策和减少政策在制定和执行中的盲目性,对我国国民经济的健康发展具有重要意义。

注释:

①可从表 2 计算出医疗保健在消费性总支出中占的比率 X_4/S , 然后进行由大到小的排序,各地区的顺序依次为:宁夏、黑龙江、青海、河北、辽宁、北京、浙江、陕西、甘肃、山西、吉林、河南、新疆、内蒙古、天津、云南、山东、广东、湖北、四川、重庆、湖南、海南、江苏、上海、西藏、贵州、广西、安徽、江西和福建,大致由寒冷的北方地区排到温暖的南方地区。这是由于气候的寒冷易导致医疗保健费用的增加,因此,可以认为除衣着 X_2 外医疗保健 X_4 也是受地区气候影响的指标。

参考文献:

- [1]国家统计局. 中国统计年鉴(2000)[M]. 北京:中国统计出版社,2000.
- [2]张尧庭,方开泰. 多元统计分析引论[M]. 北京:科学出版社,1982.
- [3]王学民. 应用多元化分析[M]. 上海:上海财经大学出版社,1999.

Comparative Analysis on Urban Residents' Living Expenditure by Regions

WANG Xue-min

(Department of Statistics, Shanghai University of Finance and Economics, Shanghai 200433, China)

Abstract: There are many differences with reference to the level of economic development by region. This article applies the principal component analysis to the comparative research of the urban residents' living expenditure by region and comes to a conclusion that there are two important factors leading to the differences. First is the regional level of economic development. In China, the consumption level in the east, middle and west varies a lot. Second is the consumption tendency caused by the climate factor. There are obvious differences between the north and the south.

Key words: living expenditure; principal component analysis; consumption level; consumption tendency