

# 我国城镇居民收入与支出的多元统计分析

## 摘要

本文搜集并整理中国统计年鉴上面历年来我国各省市的城镇居民收入与消费支出数据，主要利用多元统计学中的系统聚类法和主成分分析法作为分析工具，对我国各地区城镇居民年均家庭收入与支出进行统计分析，分析多个变量之间的观测的数据，推断与分析出我国城镇居民收入和支出之间存在的关系，找出影响存在的问题的主要因素。然后对我国居民收入与消费支出现状提出了一些的政策性建议。

**关键词：**收入 支出 系统聚类 主成分分析

# Multivariate statistical analysis of urban residents' income and expenditure of the China

## Abstract

This article has collected and analyzed over the years, China Statistical Yearbook on top of income of urban residents and consumer spending data Province in China, the main use of multivariate statistics in hierarchical clustering and principal component analysis as an analytical tool for the region China's average annual urban residents household income and expenditure for statistical analysis, analysis of data between multiple variables observation, inference and analysis of the relationship between China's urban residents' income and expenditure, to identify the main factors affecting the existing problems. Then the status of the income and consumption expenditure made some policy recommendations.

**Keywords:** Income ;Expenses;System cluster; Principal component analysis

# 目录

1、引言 .....	4
1.1 研究问题的背景及意义 .....	4
2、多元统计分析方法的简单介绍 .....	5
2.1 聚类分析---系统聚类法 .....	5
2.1.1 聚类分析的概念 .....	5
2.1.2 聚类分析的基本思想 .....	5
2.1.3 系统聚类法的基本思路 .....	5
2.2 主成分分析 .....	5
2.2.1 主成分分析的概念 .....	5
2.2.2 主成分分析的数学模型 .....	5
2.2.3 主成分的计算过程 .....	6
3、我国城镇居民人均年收入的多元统计分析 .....	7
3.1 系统聚类分析 .....	7
3.2 用主成分分析对居民收入进行分析 .....	10
3.3 收入的综合分析结果 .....	14
4、我国城镇居民人均年消费性支出的多元统计分析 .....	15
4.1 系统聚类分析 .....	15
4.2 主成分分析 .....	18
4.3 支出的综合分析结果 .....	22
5、研究结论及建议 .....	23
参考文献 .....	25
致谢 .....	26

# 1、引言

## 1.1 研究问题的背景及意义

对于和谐社会一直以来都有诸多定义，然而他们都有一个共同基本的含义，即和谐社会应该是一个稳定有序、充满活力、人和自然和谐相处的社会。实现和谐社会和人们的消费行为密不可分。要想满足人们需要，那就必须消费，只有人们的需求得到满足，人们才会感到幸福，社会才会保持安定和谐，理性、健康、合理的消费模式可以使人们物质和精神方面都得到满足，对提升整个社会活力与创造力也至关重要。

中国经济发展速度迅速，城镇居民的收入与消费水平不断提高。然而收入与消费的相关性问题也逐渐凸现出来。许多经济学家指出：市场经济条件下，收入主要因素是居民消费，城镇居民的消费性支出与收入有着内在的关系，收入分配差距过大，会直接影响居民消费需求。平衡好城镇居民收入与支出的关系对全面建设和谐社会起着至关重要的作用。

其次，消费结构可以反映经济发展水平和居民生活水平的质量。一般来说，经济越发达的地区，高消费支出所占的比例就会较大。相反，在经济落后的地区中，日常生活必需品消费支出会占大多数比例。相对过去而言，居民在日常生活必需品支出比重在逐渐下降，而在交通通讯、文化娱乐等方面的消费支出比重越来越大。消费结构的变化，反映居民需求的变化。研究我国城镇居民消费性支出情况及其影响因素，对建立合理的消费结构，扩大内需，搞活社会主义市场经济，制定产业政策，促进国民经济的发展都有着极其重要的意义。

本文就是在此基础之上，利用统计学的数学工具对历年我国各省市的城镇居民收入与支出数据进行多元统计分析，找出问题的主要影响因素，以及各省市的发展状况。本文统计分析参考的数据主要取自《中国统计年鉴(2003-2013)》中的《我国各地区城镇居民平均每人每年家庭收入》和《我国各地区城镇居民平均每人每年家庭消费支出》两部分数据。城镇居民和收入的统计数据主要采用了全国 31 个省、市、自治区（不包括港澳台地区）的数据。

## 2、多元统计分析方法的简单介绍

### 2.1 聚类分析---系统聚类法

#### 2.1.1 聚类分析的概念

聚类分析又称群分析、点群分析,是定量研究样品或指标分类问题的一种多元统计方法。其中类指相似元素的集合。[3]

#### 2.1.2 聚类分析的基本思想

认为所研究的样品或指标之间存在着程度不同的相似性,根据一批样品的多个观测指标,找出能够度量样品或变量之间相似程度的统计量,并以此为依据,采用某种聚类法,将所有的样品或变量分别聚合到不同的类中,使同一类中的个体有较大的相似性,不同类中的个体差异较大。[2]

#### 2.1.3 系统聚类法的基本思路

首先,将  $n$  个样品看成  $n$  类,定义各样品之间的距离;其次,按照某种原则将最近的两类合并为一个新类,于是得到  $n-1$  类,定义并计算各类间距离,然后再从中找出最近的两类合并成一个新类,重复以上步骤,直到全部样品合并成一类为止,将上述合并过程画成聚类谱系图,据此图可将全部样品分类。[6]

### 2.2 主成分分析

#### 2.2.1 主成分分析的概念

主成分分析是考察多个定量(数值)变量间相关性的一种多元统计方法。它是研究如何通过少数几个主成分来解释多变量的方差-协方差结构。

#### 2.2.2 主成分分析的数学模型

设原始变量为  $x_1, x_2, \dots, x_p$  考虑它们的线性变换

$$y_i = a_{1i}x_1 + a_{2i}x_2 + \dots + a_{pi}x_p \quad (i=1, 2, \dots, p)$$

其中  $y_1, y_2, \dots, y_p$  满足以下条件

第一  $\text{cov}(y_i, y_j) = 0, (i \neq j);$

第二  $D(y_1) \geq D(y_2) \geq \dots, \geq D(y_p);$

即  $y_1$  是  $x_1, x_2, \dots, x_p$  的一切线性组合中方差最大者,  $y_2$  是方差次大者, 依此类推, 称  $y_1$  为  $x_1, x_2, \dots, x_p$  的第一主成分,  $y_2$  为第二主成分,  $y_p$  为第  $p$  个主成分。

### 2.2.3 主成分的计算过程

第一步: 计算样本点数据矩阵  $X_{m \times n}$  的协方差矩阵  $\text{Cov}(X)$

第二步: 计算  $m$  阶的实对称矩阵  $\text{Cov}(X)$  的  $m$  个实特征根, 并按照大小排列,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$

第三步: 计算相应特征根的长度为 1 的、相互正交的特征向量  $B_1, B_2, \dots, B_m$

第四步: 按以下方法求主成分  $Y_i$

$$Y_1 = B_1^T X, Y_2 = B_2^T X, \dots, Y_m = B_m^T X$$

$$Y = B^T X$$

进一步展开得到

$$Y_1 = b_{11} X_1 + b_{21} X_2 + \dots + b_{m1} X_m$$

.....

$$Y_m = b_{1m} X_1 + b_{2m} X_2 + \dots + b_{mm} X_m$$

这就是主成分的计算公式, 这样计算出的主成分满足

第一、 $\text{Cov}(Y)$  为对角矩阵, 即  $Y_1, Y_2, \dots, Y_m$  之间互不相关, 而且  $Y_i$  的方差为

$$\lambda_i。$$

第二、 $Y$  与  $X$  的协方差矩阵的对角线上的元素之和相等, 即  $\text{trcov}(Y) = \text{trcov}(X)$ , 这表明上述变换没有方差上的损失。

第三、对于  $Y$ , 我们可以得到  $\text{trcov}(Y) = \sum_{j=1}^m \lambda_j = m$ , 若取  $\lambda_1, \lambda_2, \dots, \lambda_p$ , 使得

$$(\lambda_1 + \lambda_2 + \dots + \lambda_p) / (\lambda_1 + \lambda_2 + \dots + \lambda_m) \geq Q\%$$

则  $Y_1 = B_1^T X, Y_2 = B_2^T X, \dots, Y_p = B_p^T X$  等式右端的  $Q$  叫累计贡献率, 表示计算后得到的  $p$  维主超平面近似代替原始群点的精度, 一般可取在 75%-90% 之间。

### 3、我国城镇居民人均年收入的多元统计分析

在本章我们主要利用前面叙述的多元统计方法对我国各地区城镇居民家庭人均年收入进行探索性统计分析，目的在于通过统计数据了解我国城镇居民家庭收入的总体现状与发展趋势。

#### 3.1 系统聚类分析

首先，我们对《中国统计年鉴(2003-2013)》中的《我国各地区城镇居民平均每人每年家庭收入》数据取均值，得到我国城镇居民人均年收入的均值数据(如表 1 所示)。

表 1 我国城镇居民人均年收入来源 (2003-2013) 单位：元

地区	可支配收入	总收入	工资收入	经营净收	财产性收	转移性收
北京	22889.7	25509.2	17261.2	678.4	530.9	6458.5
天津	17917.4	19526.9	11970.6	742.3	321.1	5900.5
河北	12429.5	13149.9	8188.7	959.2	189.2	3497.0
山西	12079.5	12915.0	8770.7	622.4	209.2	3041.2
内蒙古	13233.8	13988.1	9482.2	1385.8	279.2	2541.0
辽宁	13279.9	14583.1	8545.1	1221.2	242.0	4213.2
吉林	11926.7	12650.8	7975.9	1081.9	143.0	3152.2
黑龙江	9516.5	10171.6	6142.4	896.5	91.4	2722.4
上海	24810.9	27425.2	18788.1	1165.1	567.1	6036.0
江苏	17237.1	18645.6	11264.8	1761.6	377.3	4819.6
浙江	21314.4	23248.0	14137.9	2628.2	1105.4	4791.1
安徽	12168.5	13272.0	8562.1	1008.2	321.6	3067.2
福建	16812.9	18317.7	11828.5	1592.5	920.7	3558.4
江西	11856.9	12482.3	8241.6	950.0	257.5	2797.5
山东	15126.7	16315.9	12019.9	1164.9	363.1	2596.5
河南	12071.7	12776.6	8094.7	1107.9	176.5	3120.6
湖北	12324.0	13320.0	8834.5	949.0	240.1	3058.4
湖南	12955.0	13701.6	8588.1	1365.5	411.7	3096.4
广东	19005.5	20953.9	14679.5	1856.0	838.6	3130.6
广西	12937.6	13987.6	9273.0	1028.1	410.3	3007.4
海南	12037.7	12941.5	8471.5	1099.3	378.1	2695.5
重庆	13716.7	14667.7	9865.1	907.9	266.1	3341.5
四川	12009.0	12985.4	8353.4	940.0	331.6	3005.7
贵州	11154.0	11672.9	7553.1	1028.0	170.8	2658.2
云南	12679.5	13605.5	8538.5	1013.8	658.3	3114.7
西藏	12020.9	13212.9	11803.8	337.2	174.7	830.5
陕西	11951.3	12829.1	8901.2	429.4	157.9	3041.3
甘肃	10521.1	11273.6	7948.0	572.1	95.9	2437.5
青海	10881.4	12003.0	7748.0	670.1	73.0	3192.9
宁夏	10889.5	11909.4	7556.3	1268.9	141.2	2673.6
新疆	10975.4	12111.0	8949.7	785.9	96.1	2015.8

利用 SPSS 的系统聚类法得到分析结构如下

表 2 收入聚类表

阶	群集组合		系数	首次出现阶群集		下一阶
	群集 1	群集 2		群集 1	群集 2	
1	7	16	0.001	0	0	6
2	12	23	0.002	0	0	4
3	14	17	0.008	0	0	4
4	12	14	0.009	2	3	9
5	4	27	0.010	0	0	13
6	3	7	0.010	0	1	11
7	24	30	0.012	0	0	11
8	20	21	0.016	0	0	9
9	12	20	0.022	4	8	14
10	28	29	0.023	0	0	13
11	3	24	0.030	6	7	17
12	5	18	0.032	0	0	16
13	4	28	0.039	5	10	15
14	12	22	0.043	9	0	16
15	4	31	0.051	13	0	19
16	5	12	0.052	12	14	17
17	3	5	0.061	11	16	18
18	3	6	0.086	17	0	21
19	4	8	0.093	15	0	21
20	1	9	0.095	0	0	29
21	3	4	0.119	18	19	24
22	13	19	0.120	0	0	27
23	15	25	0.221	0	0	24
24	3	15	0.229	21	23	26
25	2	10	0.246	0	0	28
26	3	26	0.343	24	0	28
27	11	13	0.398	0	22	29
28	2	3	0.650	25	26	30
29	1	11	1.028	20	27	30
30	1	2	1.696	29	28	0



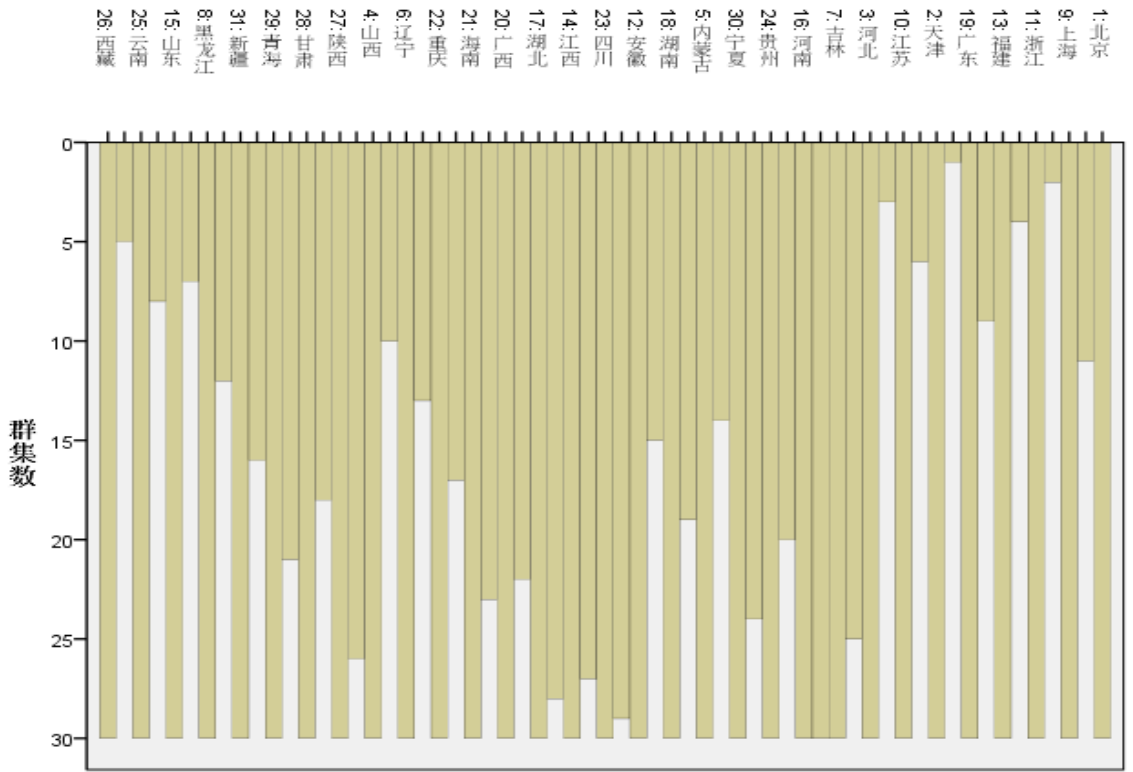


图1 收入的群集数

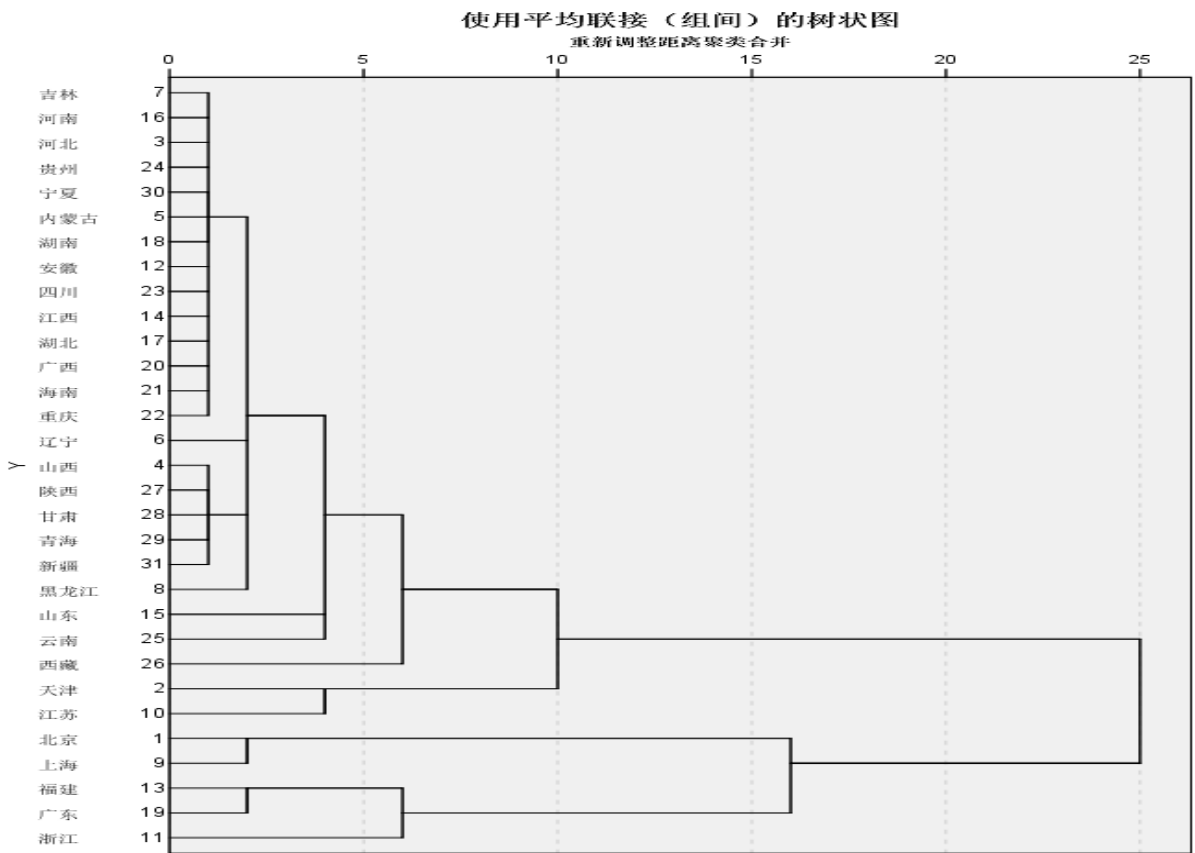


图2 收入的树状图

对以上均值数据进行聚类分析，利用 spss 统计软件进行统计分析。按照收入水平高、较高、中、低的原则，初步把各个省份分为四类：

高收入地区：北京和上海

较高收入地区：广东和浙江

中等收入区：天津、江苏和福建

低收入区：其余的省市地区

以上数据可以看出，除上海、北京、广东、浙江、天津、江苏、福建属中高收入地区，其他省市都属于低收入地区，低收入地区所占比重过大。中高收入地区大部分处于东南沿海地区，西北地区居民的收入水平还不容乐观。

### 3.2 用主成分分析对居民收入进行分析

表 3 解释的总方差

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	4.392	73.208	73.208	4.392	73.208	73.208
2	1.031	17.181	90.389	1.031	17.181	90.389
3	0.410	6.833	97.222			
4	0.165	2.754	99.977			
5	0.001	0.021	99.998			
6	0.000	0.002	100.000			

表 4 成份矩阵

	成份	
	1	2
可支配收入	0.985	-0.153
总收入	0.983	-0.168
工资收入	0.911	-0.246
经营净收入	0.584	0.757
财产性收入	0.812	0.467
转移性收入	0.790	-0.357

在第一主成分中，可支配收入、总收入、工薪收入、财产性收入和转移性收入相对应的载荷系数较高，说明第一主成分主要反应了上面五项指标的信息，我们称之为综合性因子；而在第二主成分中，仅经营性收入的载荷较高，因此第二主成分主要体现了它的信息，我们称之为经营性因子。从贡献率看，第一主成分的贡献率为 73.208%，而第一、二主成分的累积贡献率为 90.389%，由此可见，我们利用主成分分析的结果，用两个主成分变量即可较准确的反应全部指标的信息。

从分析结果可以看出，主成分对应的特征值中，大于 1 的有两个， $\lambda=4.392$  和 1.031，因此确定主成分向量有两个。用主成分载荷矩阵中的数据除以主成分相对应的特征值开平方根便得到两个主成分中每个指标所对应的系数，形成特征向量，将得到的特征向量与标准化后的数据( $ZX_i$ )相乘,就可以得出主成分表达式。

特征向量矩阵如表 5

表 5 收入的特征向量矩阵

	Z1	Z2
可支配收入X1	0.470	-0.151
总收入X2	0.469	-0.165
工资收入X3	0.435	-0.242
经营净收入X4	0.279	0.746
财产性收入X5	0.387	0.460
转移性收入X6	0.377	-0.352

主成分表达式  $F_1=0.470X_1+0.469X_2+0.435X_3+0.279X_4+0.387X_5+0.377X_6$

$F_2=-0.151X_1-0.165X_2-0.242X_3+0.746X_4+0.460X_5-0.352X_6$

表 6 收入的主成分

地区	因子1	因子2	成分1	成份2
1:北京	2.10912	-2.56884	3.46916516	-2.6755539
2:天津	0.90021	-1.81977	1.11220538	-1.2288111
3:河北	-0.43232	-0.23	-1.0163676	0.16153846
4:山西	-0.57853	-0.63	-1.244114	-0.3589415
5:内蒙古	-0.20201	0.74	-0.2444389	0.41002587
6:辽宁	-0.05514	-0.03	-0.3778763	0.40714446
7:吉林	-0.55416	0.04824	-1.173501	0.35795993
8:黑龙江	-1.11738	0.12	-2.2502347	0.65947376
9:上海	2.53912	-1.86	4.51705864	-2.5428612
10:江苏	0.94019	0.38	1.59748854	0.30977809
11:浙江	2.41725	2.53983	6.61478755	3.43279511
12:安徽	-0.36872	0.19379	-0.7528603	0.26246173
13:福建	1.09469	1.43746	2.29246195	0.43001632
14:江西	-0.55723	0.13155	-1.0905959	0.22927801
15:山东	0.21907	0.14567	0.58628569	-0.497967
16:河南	-0.50294	0.14004	-1.0521493	0.38239388
17:湖北	-0.41628	-0.07736	-0.869853	0.05172819
18:湖南	-0.12109	0.88032	-0.1929357	0.7317729
19:广东	1.51935	1.42343	3.28062986	-0.1164702
20:广西	-0.17259	0.28773	-0.3203819	0.14726861
21:海南	-0.38872	0.57873	-0.6817309	0.44849746
22:重庆	-0.13863	-0.37178	-0.3733726	-0.3111713
23:四川	-0.43037	0.15317	-0.8706381	0.2383972
24:贵州	-0.75279	0.25848	-1.4606631	0.45723669
25:云南	-0.06587	0.7651	-0.0864792	0.521356
26:西藏	-0.79777	-0.76423	-1.1990977	-1.4536456
27:陕西	-0.67612	-1.03923	-1.4771697	-0.6682341
28:甘肃	-1.0069	-0.5467	-1.9975114	-0.2257844
29:青海	-0.83373	-0.6763	-1.8188925	-0.0938622
30:宁夏	-0.70417	0.59422	-1.3394146	0.75781115
31:新疆	-0.86554	-0.20602	-1.579763	-0.2236418

$$\text{综合主成分表达式 } F = \frac{4.392}{4.392 + 1.031} * F1 + \frac{1.031}{4.392 + 1.031} * F2$$

表 7 各地区收入的得分及排名

地区	F1得分	F1排名	F2得分	F2排名	总得分	总排名
1:北京	4.4201037	3	-2.60835	31	3.08388	3
2:天津	1.886579	7	-1.84776	29	1.176621	7
3:河北	-0.906017	19	-0.22971	22	-0.77744	19
4:山西	-1.212431	23	-0.63952	25	-1.10351	24
5:内蒙古	-0.423354	14	0.754022	6	-0.19952	12
6:辽宁	-0.115557	9	-0.03297	19	-0.09986	11
7:吉林	-1.161359	21	0.048982	18	-0.93125	22
8:黑龙江	-2.341704	31	0.121724	17	-1.87337	31
9:上海	5.3212589	1	-1.88912	30	3.95045	2
10:江苏	1.9703655	6	0.38487	9	1.668937	6
11:浙江	5.0658548	2	2.578897	1	4.593044	1
12:安徽	-0.77273	15	0.196771	12	-0.58841	16
13:福建	2.2941527	5	1.459571	2	2.135485	5
14:江西	-1.167792	22	0.133573	16	-0.92038	21
15:山东	0.4591072	8	0.147911	14	0.399944	8
16:河南	-1.054016	20	0.142194	15	-0.8266	20
17:湖北	-0.872402	17	-0.07855	20	-0.72148	18
18:湖南	-0.25377	11	0.893861	4	-0.03559	10
19:广东	3.1841168	4	1.445325	3	2.853544	4
20:广西	-0.361699	13	0.292156	10	-0.23739	13
21:海南	-0.814644	16	0.587632	8	-0.54805	15
22:重庆	-0.290528	12	-0.3775	23	-0.30706	14
23:四川	-0.901931	18	0.155526	13	-0.70089	17
24:贵州	-1.577629	26	0.262456	11	-1.2278	25
25:云南	-0.138044	10	0.776869	5	0.035895	9
26:西藏	-1.671894	27	-0.77599	27	-1.50157	27
27:陕西	-1.416951	24	-1.05522	28	-1.34818	26
28:甘肃	-2.11017	30	-0.55511	24	-1.81453	30
29:青海	-1.747256	28	-0.6867	26	-1.54563	29
30:宁夏	-1.475736	25	0.60336	7	-1.08047	23
31:新疆	-1.813921	29	-0.20919	21	-1.50884	28

### 3.3 收入的综合分析结果

第一主成分主要反映的是各地区的综合性收入水平，高收入省市主要有上海、浙江、北京、广东，并且这四个省市的收入水平明显高于全国其它地区，而低收入的省份则主要集中在东北和西部地区：第二主成分反映了各地区经营性收入水平的差异，经营性收入居前列的省份是浙江、福建、广东和湖南。由于经营性收入在总收入中所占的比重较小，所以第二主成分反映的情况对整体综合性收入的影响并不大。

从各地区的城镇居民的收入统计分析结果我们不难看出，我国城镇居民收入水平与分配状况中存在着不少问题。居民收入的实际情况和和谐社会的目标还存在着一定的差距。总结起来主要有以下几个特点：

第一、收入水平地区差异明显，传统格局没有明显的改变，东部和东南沿海地区的收入水平明显高于东北和西北地区。

第二、高、中、低收入地区间的差异较大，两极分化的状况依然严峻。

第三、低收入地区间的差异有缩小的趋势，即低收入地区的收入水平有所提高，尤其是西部地区的收入水平显著提高。

第四、从各省市本身的收入结构特性来看，不管是高收入地区还是低收入地区，大部分地区收入分配结构失衡。

## 4、我国城镇居民人均年消费性支出的多元统计分析

### 4.1 系统聚类分析

首先，对《中国统计年鉴(2003-2013)》中《我国各地区城镇居民平均每人每年家庭消费支出》数据取均值，得到我国城镇居民家庭人均年消费性支出数据(如表 8 所示)

表 8 我国城镇居民家庭人均年消费性支出 (2003-2013 的均值) 单位: 元

地区	现金消费	食品	衣着	居住	家庭设备	医疗保健	交通和通	文教娱乐	其它消费
北京	16041.2	5179.2	1582.4	1271.0	1211.8	1474.6	2567.4	2024.2	730.6
天津	12688.1	4640.6	1102.5	1171.2	882.4	1077.3	1763.8	1520.9	529.6
河北	8312.5	2820.6	982.1	858.8	639.8	782.0	1062.7	882.2	284.2
山西	7997.5	2554.4	1038.8	855.8	548.2	675.6	1068.1	974.4	282.3
内蒙古	10125.7	3175.1	1526.7	880.2	694.0	869.1	1332.7	1150.7	497.3
辽宁	10070.9	3711.1	1148.8	865.9	669.7	854.9	1269.0	1095.8	455.9
吉林	9000.0	2990.0	1164.5	881.2	593.3	888.2	1080.3	1021.6	380.9
黑龙江	8134.1	2894.6	1128.9	714.9	550.8	763.0	921.9	837.2	322.8
上海	17563.8	6362.9	1298.9	1414.8	1118.4	1230.0	2933.9	2317.7	887.4
江苏	11333.1	4167.0	1100.9	912.0	765.5	827.9	1554.9	1570.1	434.8
浙江	14459.0	5051.5	1408.2	1101.7	847.6	1236.6	2501.7	1779.4	532.4
安徽	8912.0	3623.1	953.6	801.9	505.9	665.6	1025.3	1028.5	308.2
福建	11616.4	4632.8	985.5	1047.2	712.4	766.8	1650.0	1354.7	467.0
江西	8037.6	3226.6	867.7	715.5	547.3	506.6	964.5	926.4	282.1
山东	10048.1	3331.7	1321.8	975.6	718.8	812.3	1466.1	1054.7	367.3
河南	8261.3	2807.6	1087.7	726.1	650.6	712.2	990.4	959.8	326.9
湖北	9061.8	3597.3	1084.8	770.7	605.0	669.0	1031.5	1037.7	265.7
湖南	9499.6	3481.9	1003.3	833.9	669.3	766.4	1304.4	1099.2	341.1
广东	14662.4	5400.9	906.6	1291.6	900.3	1271.8	2482.8	1851.3	557.1
广西	8986.8	3587.3	673.7	804.9	601.2	625.7	1322.2	1082.4	289.5
海南	8669.4	3911.2	484.5	776.5	516.2	659.5	1188.6	867.1	265.6
重庆	10681.6	4144.3	1275.0	865.4	777.7	874.6	1298.6	1103.8	342.1
四川	9342.6	3765.3	966.0	760.9	631.7	672.8	1220.0	976.3	349.5
贵州	7970.1	3187.6	872.7	676.1	518.5	540.2	1017.4	895.0	262.7
云南	8916.6	3704.2	952.0	626.8	518.7	763.4	1249.6	848.4	253.7
西藏	8490.4	4092.7	1083.0	563.3	338.2	639.9	799.2	548.0	504.8
陕西	9222.9	3337.8	1016.9	805.3	640.4	775.0	1136.0	1155.2	356.6
甘肃	8081.4	2971.2	1013.0	719.9	546.5	677.3	974.4	862.7	316.5
青海	7889.8	2992.8	933.4	710.5	562.7	641.4	941.4	782.6	325.3
宁夏	8720.5	2984.8	1075.6	813.6	607.7	762.6	1144.7	949.5	382.1
新疆	8398.5	3052.2	1190.7	659.3	553.4	703.0	1065.1	814.7	360.2

利用 SPSS 的系统聚类法得到分析结构如表 9

表 9 支出聚类表

阶	群集组合		系数	首次出现阶群集		下一阶
	群集 1	群集 2		群集 1	群集 2	
1	14	24	0.006	0	0	10
2	28	29	0.010	0	0	10
3	18	27	0.013	0	0	8
4	8	31	0.022	0	0	9
5	16	30	0.033	0	0	9
6	12	17	0.033	0	0	11
7	3	4	0.034	0	0	13
8	18	23	0.036	3	0	11
9	8	16	0.044	4	5	13
10	14	28	0.046	1	2	19
11	12	18	0.055	6	8	18
12	20	21	0.070	0	0	23
13	3	8	0.072	7	9	17
14	6	22	0.078	0	0	16
15	10	13	0.079	0	0	22
16	6	15	0.085	14	0	21
17	3	7	0.090	13	0	18
18	3	12	0.104	17	11	20
19	14	25	0.110	10	0	20
20	3	14	0.122	18	19	23
21	5	6	0.146	0	16	25
22	2	10	0.226	0	15	25
23	3	20	0.266	20	12	27
24	11	19	0.276	0	0	29
25	2	5	0.355	22	21	28
26	1	9	0.410	0	0	29
27	3	26	0.419	23	0	28
28	2	3	0.563	25	27	30
29	1	11	0.692	26	24	30
30	1	2	3.037	29	28	0



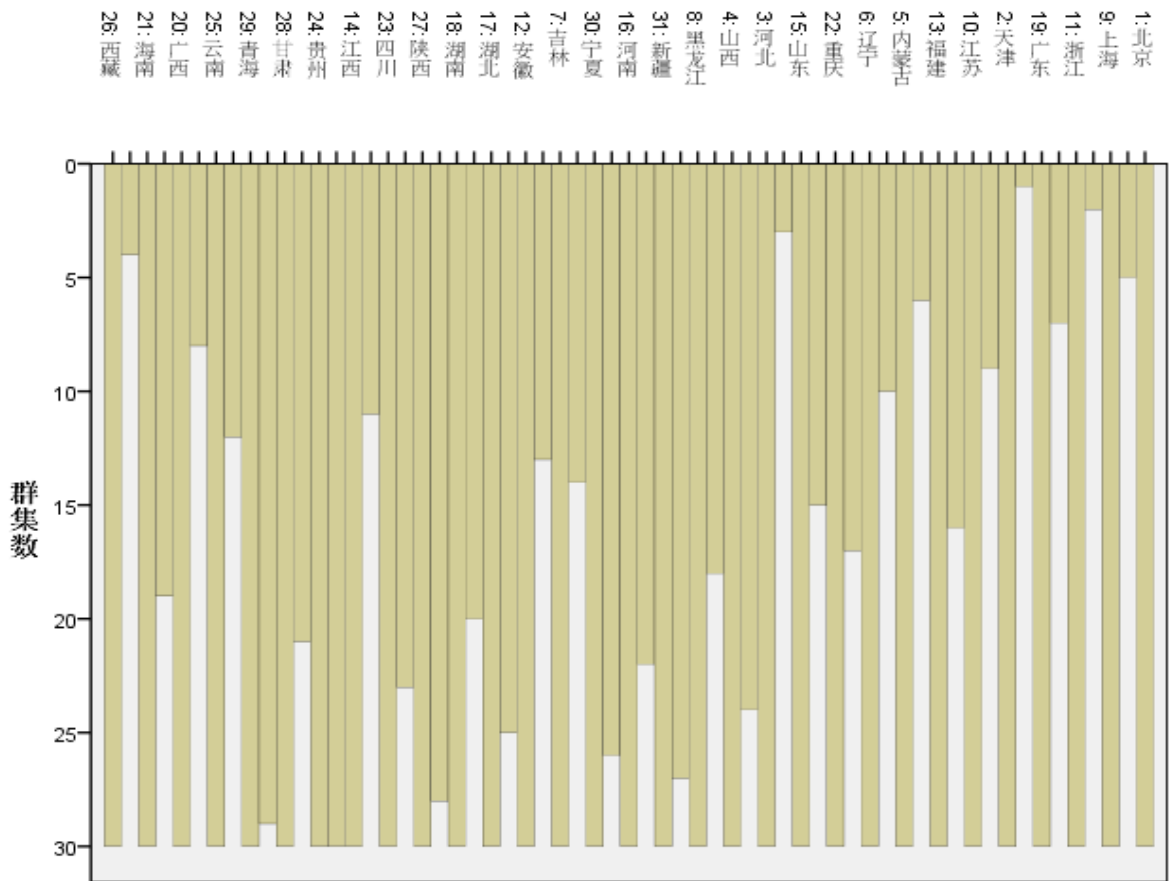


图 3 支出的群集数

通过聚类分析表，可以把各地区城镇居民的支出聚成四类，分别为

高消费地区：上海

较高消费地区：北京、广东和浙江

中等消费区：天津、江苏、福建和重庆

低消费区：其余的省市地区

以上数据可以看出，我国消费支出水平高的地区和收入水平高的地区基本吻合。而且除上海、北京、广东、浙江、天津、江苏、福建、重庆属中高消费地区，其他省市都属于低消费地区，低消费地区所占比重过大。中高消费地区大部分处于东南沿海地区，西北地区居民的消费支出状况还比较保守。

## 4.2 主成分分析

表 10 解释的总方差

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	7.444	82.708	82.708	7.444	82.708	82.708
2	0.869	9.660	92.368	0.869	9.660	92.368
3	0.312	3.462	95.831			
4	0.129	1.433	97.264			
5	0.108	1.202	98.466			
6	0.063	0.697	99.163			
7	0.049	0.544	99.706			
8	0.026	0.294	100.000			
9	6.470E-006	7.189E-005	100.000			

表 11 支出的成分矩阵

	成份	
	1	2
现金消费支出	0.993	-0.081
食品	0.878	-0.333
衣着	0.555	0.820
居住	0.945	-0.110
家庭设备及用品	0.947	0.063
医疗保健	0.940	0.132
交通和通信	0.971	-0.145
文教娱乐	0.967	-0.119
其它消费	0.911	0.102

在第一主成分中，除衣着消费支出外，其它变量所对应的载荷系数都很高，说明第一主成分主要反应了除衣着外的八项指标的信息，我们称之为综合性因子；而在第二主成分中，仅衣着消费支出的载荷较高，因此第二主成分主要体现了它的信息，我们称之为衣着消费因子。从贡献率看，第一主成分的贡献率为 83.265%，而第一、二主成分的累积贡献率为 92.681%，由此可见，我们利用主成分分析的结果，用两个主成分变量即可较准确的反应全部信息。

从分析结果可以看出，主成分对应的特征值中，可以确定主成分向量有两个的有两个， $\lambda = 7.444$  和  $0.869$ ，用主成分载荷矩阵中的数据除以主成分相对应的特征值开平方根便得到两个主成分中每个指标所对应的系数，形成特征向量，将得到的特征向量与标准化后的数据( $ZX_i$ )相乘,就可以得出主成分表达式。

表 12 支出的特征向量矩阵

	Z1	Z2
现金消费支出	0.364	-0.087
食品	0.322	-0.357
衣着	0.203	0.880
居住	0.346	-0.118
家庭设备及用品	0.347	0.068
医疗保健	0.345	0.142
交通和通信	0.356	-0.156
文教娱乐	0.354	-0.128
其它消费	0.334	0.109

主成分表达式

$$F1=0.364X1+0.322X2+0.203X3+0.346X4+0.347X5+0.345X6+0.356X7+0.354X8+0.334X9$$

$$F2=-0.087X1-0.357X2-0.880X3-0.118X4+0.068X5+0.142X6-0.156X7-0.128X8+0.109X9$$

表 13 支出的主成分

地区	因子1	因子2	成分1	成份2
1:北京	2.56387	1.34337	6.995185	1.252292
2:天津	1.10213	-0.4312	3.007018	-0.40197
3:河北	-0.54	0.14	-1.47332	0.130508
4:山西	-0.67	0.36	-1.82801	0.335593
5:内蒙古	0.23	2.30	0.627525	2.144064
6:辽宁	0.06576	0.45728	0.179418	0.426277
7:吉林	-0.24166	0.87976	-0.65934	0.820114
8:黑龙江	-0.67	0.87	-1.82801	0.811015
9:上海	2.91188	-0.813	7.944685	-0.75788
10:江苏	0.46505	-0.2685	1.268828	-0.2503
11:浙江	1.63976	0.45201	4.473871	0.421364
12:安徽	-0.54344	-0.47258	-1.48271	-0.44054
13:福建	0.49613	-1.04852	1.353626	-0.97743
14:江西	-0.85627	-0.63953	-2.33622	-0.59617
15:山东	0.11019	1.16283	0.300639	1.083992
16:河南	-0.57489	0.6639	-1.56851	0.618889
17:湖北	-0.47403	0.11449	-1.29333	0.106728
18:湖南	-0.20291	-0.18979	-0.55361	-0.17692
19:广东	1.74491	-1.91761	4.760759	-1.7876
20:广西	-0.51813	-1.76913	-1.41365	-1.64919
21:海南	-0.73779	-2.60228	-2.01296	-2.42585
22:重庆	0.19795	0.73732	0.540081	0.687331
23:四川	-0.36463	-0.42341	-0.99485	-0.3947
24:贵州	-0.90378	-0.58456	-2.46585	-0.54493
25:云南	-0.63032	-0.38542	-1.71975	-0.35929
26:西藏	-0.78634	0.35294	-2.14543	0.329011
27:陕西	-0.27483	-0.00192	-0.74984	-0.00179
28:甘肃	-0.72976	0.24901	-1.99105	0.232128
29:青海	-0.80506	-0.05947	-2.1965	-0.05544
30:宁夏	-0.39703	0.4822	-1.08324	0.449508
31:新疆	-0.61746	1.04275	-1.68466	0.972053

$$\text{综合主成分表达式 } F = \frac{7.444}{7.444 + 0.869} * F1 + \frac{0.869}{7.444 + 0.869} * F2$$

最后算得 F1、F2 和综合主成分的得分及排名如表 14

表 14 各地区支出的得分及排名

地区	F1得分	F1排名	F2得分	F2排名	总得分	总排名
1:北京	6.995185	2	1.252292	2	6.394851	2
2:天津	3.007018	5	-0.40197	23	2.650659	5
3:河北	-1.47332	19	0.130508	15	-1.30566	18
4:山西	-1.82801	24	0.335593	12	-1.60184	25
5:内蒙古	0.627525	8	2.144064	1	0.786057	8
6:辽宁	0.179418	11	0.426277	10	0.205223	11
7:吉林	-0.65934	13	0.820114	5	-0.50469	12
8:黑龙江	-1.82801	24	0.811015	6	-1.55214	23
9:上海	7.944685	1	-0.75788	27	7.034962	1
10:江苏	1.268828	7	-0.2503	20	1.110026	6
11:浙江	4.473871	4	0.421364	11	4.050242	4
12:安徽	-1.48271	20	-0.44054	24	-1.37377	20
13:福建	1.353626	6	-0.97743	28	1.109949	7
14:江西	-2.33622	30	-0.59617	26	-2.15432	30
15:山东	0.300639	10	1.083992	3	0.382527	10
16:河南	-1.56851	21	0.618889	8	-1.33985	19
17:湖北	-1.29333	17	0.106728	16	-1.14697	17
18:湖南	-0.55361	12	-0.17692	19	-0.51423	13
19:广东	4.760759	3	-1.7876	30	4.076226	3
20:广西	-1.41365	18	-1.64919	29	-1.43827	22
21:海南	-2.01296	27	-2.42585	31	-2.05612	29
22:重庆	0.540081	9	0.687331	7	0.555474	9
23:四川	-0.99485	15	-0.3947	22	-0.93211	16
24:贵州	-2.46585	31	-0.54493	25	-2.26505	31
25:云南	-1.71975	23	-0.35929	21	-1.57753	24
26:西藏	-2.14543	28	0.329011	13	-1.88676	27
27:陕西	-0.74984	14	-0.00179	17	-0.67164	14
28:甘肃	-1.99105	26	0.232128	14	-1.75865	26
29:青海	-2.1965	29	-0.05544	18	-1.97268	28
30:宁夏	-1.08324	16	0.449508	9	-0.92301	15
31:新疆	-1.68466	22	0.972053	4	-1.40694	21

### 4.3 支出的综合分析结果

第一主成分主要反映的是各地区消费支出的整体水平，高消费支出省市主要有上海、北京、广东、浙江和天津，这五个省市的消费支出水平明显高于全国其它地区，而低收入的省份则主要集中在东北、西南和西北部地区；第二主成分反映了各地区以衣着消费支出的差异，其中位居前列的省份是内蒙古、北京、山东、新疆、吉林。由于衣着消费在总消费支出中所占的比例较小，所以第二主成分反映的情况对整体综合性收入的影响并不大。

从各地区的城镇居民的支出统计分析结果我们可以看出以下几点：

第一 、消费水平地区差异明显，传统格局没有明显的改变，东部和东南沿海地区的消费水平明显高于东北和西北地区。

第二 、高、中、低消费地区间的差异较大，两极分化的状况依然严峻。

第三 、中高消费地区在日常生活用品方面支出较少，而低消费地区在日常生活用品方面支出过多。

## 5、研究结论及建议

本文通过对我国各地区城镇居民家庭人均收入及支出进行研究分析，我们不难发现，我国实际的居民收入与消费结构仍然存在一些不合理。

通过对我国居民收入聚类分析以及主成分分析发现，我国的收入差距并有明显改善。虽然各个省之间发展差距的主要原因是经济发展之间的差距，但是差距必须保持在一个合适的度。贫富之间的差距特别是不合理的因素，不仅使社会心理扭曲，诱发或加重失业率上升，引发社会治安恶化等一系列社会风险，而且也阻碍经济的发展。

本文通过统计分析，我们了解到，我国大部分省市还处于低收入水平阶段，激发中低收入群体被压抑的消费趋向转变为消费行动是增加市场需求的关键所在。这一措施的主要推动力在于资金，只有增加低收入群体的收入水平，释放他们的消费能力，并大大增加消费，全社会总消费需求量才能稳步提高，消费结构才会更加和谐稳定，潜在的消费市场才会扩大。并且由消费市场的启动，带动企业投资的增加，保证社会需求的正常增长，国民经济平稳较快发展。在经济稳定增长的条件下，加大对低收入群体的扶持力度，对于缓解需求不足对经济的制约作用至关重要。

中等收入者群体虽然有较高的消费欲望和消费能力，但是面临着未来收入和支出的不确定性，缺乏良好的预期和消费信心。要提高中等收入阶层的消费需求，首先要深化与居民收入、消费关系密切的体制改革，稳定中等收入者的收入预期，增强其消费信心和物质基础。

此外，必须适当调节部分居民的高收入。收入分配差距的迅速拉大，不仅降低了总体的消费需求，也影响了我国的社会稳定与和谐发展。即使是合理合法的收入，如果收入水平过高，超过了大部分居民的收入水平，也必须通过适当的手段进行调节。

总之，从促进消费需求增长、提高边际消费倾向的角度来看，我国收入分配政策调控的目标应是形成“橄榄”型的收入分配结构，即中等收入者占据居民收入的较大比重，低收入和高收入者都只占社会居民的一部分，使目前我国社会“金字塔”型的收入分配结构向和谐社会的“橄榄”型的收入分配结构转变。

以下是我提出的一些相关建议：

第一，大力发展低收入地区的地方经济，其中大力发展生产力是前提。居民收入水平的提高依赖经济的增长，如果某一地区经济落后，居民的生活水平不高，在这种情况下调节收入，对收入差距的影响微乎其微，并不能从根本上解决问题。只有经济发展了，缩小收入差距才能有现实的基础。

第二，政府应当加大对落后地区的财政投入，注重落后地区教育事业的发展。科技是第一生产力，人才是技术创新、技术进步的前提。特别是对低收入水平地区而言，只有技术进步，经济才能较快发展，居民的生活水平才能提高，收入才能增加，与高收入地区的差距才能缩小。

第三，完善完善社会保障体系，为城镇居民提供基本的社会保障，加大对失业人口和贫困人口的扶持力度，对于贫困群体和低收入群体给予适当的补贴，保证贫困群体和低收入群体的最低生活保障，提高社会福利水平。

第四，深化分配制度改革，建立合理的分配制度。让多种要素参与到社会主义市场经济体制分配制度中，调动要素所有者的积极性，使更多生产要素参与到财富生产。此外，还要整顿不合理收入，加强垄断行业的监督和管理，引入良性竞争机制，取缔非法收入。

第五，出台关于政策、税收等方面的优惠措施来激发人们的投资热情。降低高收入者的投资成本，让投资者获取合理合法收入，另一方面也要扩大经济不发达地区的就业机会。

第六，建立良好的消费环境，促进高收入者的消费，在规范和引导不良消费前提下，给予高消费者应有的尊重，提供正常的消费环境。

本文在多元统计分析的基础上，主要结合我国政府创建和谐社会的政策号召，对我国城镇居民家庭收入与支出的现状进行研究分析。虽然大部分内容是通过统计数据进行研究分析，但有一点可以肯定的是，我国政府十分重视建设和谐的收入与消费的分配结构。然而在实施发展的道路上，有各种各样的问题和困难阻挠。在此我仅希望通过这篇学术分析作为解决这一问题的其中一道工具。



## 参考文献

- [1]许国琼. 基于聚类方法的我国各省市城镇居民消费结构分析[J]. 科技风, 2009, (6).
- [2]吴旭芳, 郭洪娟. 咸安区各乡镇经济发展实力的统计分析[N]. 咸宁学院学报, 2010-08-15.
- [3]陈龙. 多元统计分析在中国城镇居民收入中的应用[N]. 兰州文理学院学报, 2014-09-10.
- [4]郭显光. 如何利用 SPSS 软件进行主成分分析[J]. 统计与信息论坛, 1998, (2).
- [5]马庆国. 管理统计—数据获取、统计原理 SPSS 工具与应用研究[M]. 北京科学出版社, 2002, 308-372.
- [6]苏丹. 面向产品的参数化模块设计技术及其应用研究[D]. 西安理工大学, 2007-03-01.
- [7]张文霖. 主成分分析在 SPSS 中的操作应用[J]. 市场研究, 2005, (12).
- [8]张国华. 中国城镇居民消费的聚类分析[N]. 重庆工商大学学报(社会科学版), 2008-7-25.
- [9]刘建勋. 中国城镇居民消费结构及趋势分析[J]. 现代商贸工业, 2009, (2).
- [10]武佳. 和谐生活背景下中国城镇居民收入与消费支出结构的多元统计分析[D]. 南开大学, 2007.
- [11]谢江红, 李雪梅, 王生原. 多元统计分析在宏观经济分析中的应用[J]. 微型机与应用, 2009, (8).
- [12]李志辉. SPSS for windows 统计分析教程(第二版)[M]. 电子工业出版社, 2005.
- [13]丁咏梅, 周晓阳. 我国城镇居民消费结构的统计分析[J]. 市场研究, 2004, (1).
- [14]何晓群. 多元统计分析[M]. 中国人民大学出版社, 2004.
- [15]余建莱, 何旭宏. 数据统计分析与 SPSS 应用[M]. 人民邮电出版社, 2003.
- [16]李泓欣, 丁孟春. 我国居民消费结构存在的问题及对策[J]. 工业技术经济, 2003, (10).
- [17]何天立. 从城镇居民消费结构看改革 30 年的成就[N]. 山东工商学院学报, 2008-6-12.
- [18]张宇辉, 曹颖琦. 城镇居民消费支出与收入的典型相关分析[J]. 经济论坛, 2005, (4).

## 致谢

本论文是在L老师的亲切关怀和细心指导下完成的。他渊博的专业知识，严肃的科学态度，严谨的治学精神，精益求精的工作作风，深深地感染和激励着我。另外还要特别感谢系主任Z老师，Z老师多次为我指点迷津，帮我拓宽思路。

从课题的选择到项目的最终完成，两位老师都始终给予我细心的指导和不懈的支持。在此感谢两位老师的谆谆教导和细心栽培，感谢校领导、各位老师为我提供了良好的研究条件和学习环境，最后向我的家人、朋友、同学致谢，感谢他们对我的支持和帮助。