

基于聚类分析的我国城镇居民消费支出结构的研究

邱晨晨

上海工程技术大学管理学院, 上海

收稿日期: 2024年1月10日; 录用日期: 2024年1月31日; 发布日期: 2024年4月11日

摘要

居民消费结构不仅反映地区居民消费偏好程度以及对居民收入、消费品价格等因素的敏感程度, 而且能反映地区居民的生活质量与区域经济发展水平。由于31省市所处的地理位置不同, 其消费存在较大的差异性, 消费结构层次也明显具有差异。本文以2022年31省、市、自治区城镇居民的人均消费支出的八个项目支出数据为依据, 首先运用层次聚类法对各地区人均消费结构进行聚类。然后运用K-Means聚类法, 对各地区人均消费结构进行K-Means聚类, 不仅从总体上掌握了消费结构类型的地区分布, 而且系统阐述了各地区消费结构的特点, 最后在此基础上提出了促进城镇居民消费结构合理化的政策建议, 为我国各地区因地制宜地制定消费政策提供了科学合理的决策借鉴。

关键词

城镇居民, 消费支出, 层次聚类, K-Means聚类

The Study of Consumption Expenditure Structure of Chinese Urban Residents Based on Cluster Analysis

Chenchen Qiu

School of Management, Shanghai University of Engineering Science, Shanghai

Received: Jan. 10th, 2024; accepted: Jan. 31st, 2024; published: Apr. 11th, 2024

Abstract

The consumption structure of residents not only reflects the degree of consumption preference of

文章引用: 邱晨晨. 基于聚类分析的我国城镇居民消费支出结构的研究[J]. 运筹与模糊学, 2024, 14(2): 181-190.

DOI: 10.12677/orf.2024.142124

regional residents and the sensitivity to factors such as resident income and consumer prices, but also reflects the quality of life of regional residents and the level of regional economic development. Because the 31 provinces and municipalities are in different geographical locations, their consumption is different, and the level of consumption structure is also obviously different. Based on the eight project expenditure data of urban residents in 31 provinces, municipalities and autonomous regions in 2022, this paper firstly uses hierarchical clustering method to cluster the per capita consumption structure of each region. Then, it uses K-Means clustering method to cluster the per capita consumption structure of each region. It not only grasps the regional distribution of consumption structure types on the whole, but also systematically expounds the characteristics of consumption structure in each region. Finally, on this basis, it puts forward policy suggestions to promote the rationalization of urban residents' consumption structure, which provides a scientific and reasonable reference for each region to formulate consumption policies in accordance with local conditions.

Keywords

Urban Dwellers, Consumer Spending, Hierarchical Clustering, K-Means Clustering

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

消费是拉动经济增长的“三架马车”之一，消费是社会再生产的重要环节，是社会经济活动的出发点和归宿，消费结构的状况反映社会经济发展的水平。近年来，随着我国经济快速发展，居民消费结构也发生很大变化。由于各地区的经济发展不均衡，使得各地区的消费结构仍然存在着明显差别。为了进一步改善消费结构，正确引导消费，提高我国城镇居民的消费水平，有必要考察我国各地区城镇居民的消费结构之间的差异并进行比较研究，从宏观上观测各地区城镇居民的消费现状和不同地区消费结构及消费水平的差异，为提高我国各地区消费水平提供决策依据。本文选取 2022 年我国城镇居民人均消费支出数据，首先运用层次聚类法对各地区人均消费结构进行聚类。然后运用 K-Means 聚类法，对各地区人均消费结构进行 K-Means 聚类，不仅从总体上掌握了消费结构类型的地区分布，而且系统阐述了各地区消费结构的特点，最后在此基础上提出了促进城镇居民消费结构合理化的政策建议，为我国各地区因地制宜地制定消费政策提供了科学合理的决策借鉴。

2. 国内外研究现状

许多学者运用消费结构模型分析体系中的因子分析法、聚类法、VAR 模型、面板模型以及灰色关联分析等对消费结构进行了深入研究。Matyas *et al.* [1] (1992)采用面板数据(Panel Data)模型来研究居民消费结构中关于消费的相关变量。卢方圆等[2] (2009)是将中国农村居民的消费结构作为研究对象，并且通过 Panel data 模型，得出层次不同的农村居民的消费支出差异较大。查道中等[3] (2011)通过 var 模型重点研究了经济增长和产业结构二者与城乡的居民消费结构之间的关系，研究表明经济增长和产业结构两者中，经济增长对消费结构的作用更明显，而产业结构的作用相对较弱。马银晓等[4] (2012)通过因子分析和聚类分析二者相结合研究了我国 31 个省的消费结构，并将 31 个省市的消费结构可以划分为五大类别。朱玉春等[5] (2013)运用灰色关联分析法研究了农村的收入变动对消费结构的变动影响，研究表明随着收入的增加，居民消费结构有所改善，但仍然存在着一些制约其改善的因素。缪迎迎[6] (2017)通过聚类分析和

灰色关联分析发现合肥的居民消费结构正在优化,但与安徽的消费倾向仍有一定差距,并通过灰色预测模型预测出未来几年消费支出进一步提高,但仍存在一些医疗保健发展较慢等问题。柯善淦等[7](2017)研究了武汉城市群的消费结构时空演变特征,通过运用基尼系数、偏离一份额分析法对武汉城市群居民消费支出进行了分析,研究表明消费结构变化的特点是由中心向外围辐射。戴林送[8]通过2000年和2015年两个年份我国31个地区评价消费质量相关指标数据,运用因子分析法得出近十几年来我国城镇居民消费质量发展变化的特点。唐琦等[9](2018)运用QUAIDS模型分析了1995~2013年城镇居民消费结构的变动,发现居住消费的增长较快并在一定程度上影响了其他类消费的支出。侯淑静[10]首先用因子分析的方法将8个消费指标降为3个公共因子,然后用综合得分法分析了排名前五名和排名后五名省市的城镇居民消费水平。陆丹妮[11]基于2002、2007和2012年数据,采用因子分析方法提取隐性抽象因子,分析15年来居民消费结构变动。张芬等[12]以2019年31省市数据为依据,运用因子分析建立了消费水平评价函数模型,分析了影响31省市的居民消费结构差异的因素。

聚类分析法可以将数据按照相似性进行分组,有助于帮助用户发现数据中的隐藏模式和规律,综上所述,本文在以上学者采用的方法基础上,运用K-Means聚类方法对我国城镇居民2022年消费数据进行分析。

3. 数据来源及可视化分析

3.1. 数据来源

样本数据来自中国统计年鉴2022年我国城镇居民人均消费支出。数据包含8个变量分别为:食品烟酒(X_1)、衣着(X_2)、居住(X_3)、生活用品及服务(X_4)、交通通信(X_5)、教育文化娱乐(X_6)、医疗保健(X_7)、其他用品及服务(X_8)。数据包含31个“观测值”,分别对应中国31个省、市、自治区城镇居民的人均消费支出。为了消除原来各指标量纲,使各指标之间具有可比性,需要对原始数据进行标准化处理。 X_{ij} 标准化的计算公式为

$$X_{ij} = \frac{X_{ij} - X_j}{\sqrt{\text{Var}(x_j)}} (i = 1, 2, \dots, n; j = 1, 2, \dots, p),$$

经过处理后的数据的相关系数矩阵由 R 表示,其中 $r_{jk} = \frac{S_{jk}}{\sqrt{S_{jj}S_{kk}}} = \frac{S_{jk}}{S_j S'_k}$,表示第 j 和第 k 个变量之间的样本相关系数。

3.2. 数据可视化分析

3.2.1. 城镇居民消费结构分析

表1列出了2022年全国城镇居民在食品烟酒、衣着、居住、生活用品及服务、交通通信、教育文化娱乐、医疗保健、其他用品及服务八项支出结构。

Table 1. Proportion of consumption structure of urban residents in 2022 (Unit: Yuan)

表1. 2022年城镇居民消费结构占比(单位:元)

指标	食品烟酒	衣着	居住	生活用品及服务	交通通信	教育文化娱乐	医疗保健	其他用品及服务
费用	8958	1735	7644	1800	3909	3050	2481	814
占比	29.48%	5.71%	25.15%	5.92%	12.86%	10.04%	8.16%	2.68%

从表 1 可以看出 2022 年，全国城镇居民人均食品烟酒消费支出为 8958 元，食品烟酒在各项消费支出中的比重最大为 29.48%，接下来依次是居住、交通通信、教育文化娱乐、医疗保健、生活用品及服务、衣着、其他用品和服务。整体来看，我国居民的消费支出结构中满足日常生活需求的比例较高，消费结构较为合理。

3.2.2. 星图分析

以星图为例对样本数据进行可视化分析。利用 Rstudio 所做星图各半径与原变量的对应关系为：从右起水平半径对应第一个变量，逆时针旋转，星图的各半径分别对应第二、第三到第八个变量。根据星图各半径的长短，可以清晰地判断出各地区对应变量的相对水平，以此来分析各地区的消费水平及消费结构。同时，也可以利用星图对各地区进行分类。2022 年各地区消费支出星图如图 1 所示。

根据星图分析结果为：北京和上海的八项消费支出指标均超出其他地区，位于全国第一梯队；天津、江苏、浙江、广东八项消费支出指标较高，大致位于第二梯队；福建、山东、重庆、湖南等大致位于第三梯队；山西、内蒙古、吉林、黑龙江、河南的星图形状相似，其消费支出结构偏向第二变量(X_2)即衣着消费，其余省份的八项消费支出指标较为低，广西的八项消费支出指标最少。



Figure 1. Star chart of consumer spending by region, 2022

图 1. 2022 年各地区消费支出星图

4. 城镇居民人均消费支出的聚类分析

聚类分析是对多种属性统计样本进行分类的一种多元统计分析方法。其基本思想是：研究样本或变

量之间存在着不同程度的相似性。于是根据一批样本的多个观测变量，具体找出一些能够度量样本或变量之间相似程度的统计量，以这些统计量为划分类别的依据，把一些相似程度较大的样品聚为一类。关系密切的聚为一个小的分类单位，关系疏远的聚为一个大的分类单位，直到把所有样本或变量都聚类完毕，这样就可以形成一个由小到大的分类系统。

4.1. 层次聚类

聚类是研究不同组样本间的相关关系，它以各个类的相关度来衡量，将相似的聚到同一类别中，而使得同一类别中的差别尽量小，不同类的差距尽量大。层次聚类有三种分别为：简单连接法、完全连接法、平均连接法。

4.1.1. 简单连接法

简单连接法定义两族群间相隔最近的两个个体间的距离为两族群的距离。31省市城镇居民人均支出简单连接聚类如图2所示。

简单连接法公式： $D(A, B) = \min \{d(y_i, y_j), \text{for } y_i \text{ in } A \text{ and } y_j \text{ in } B\}$

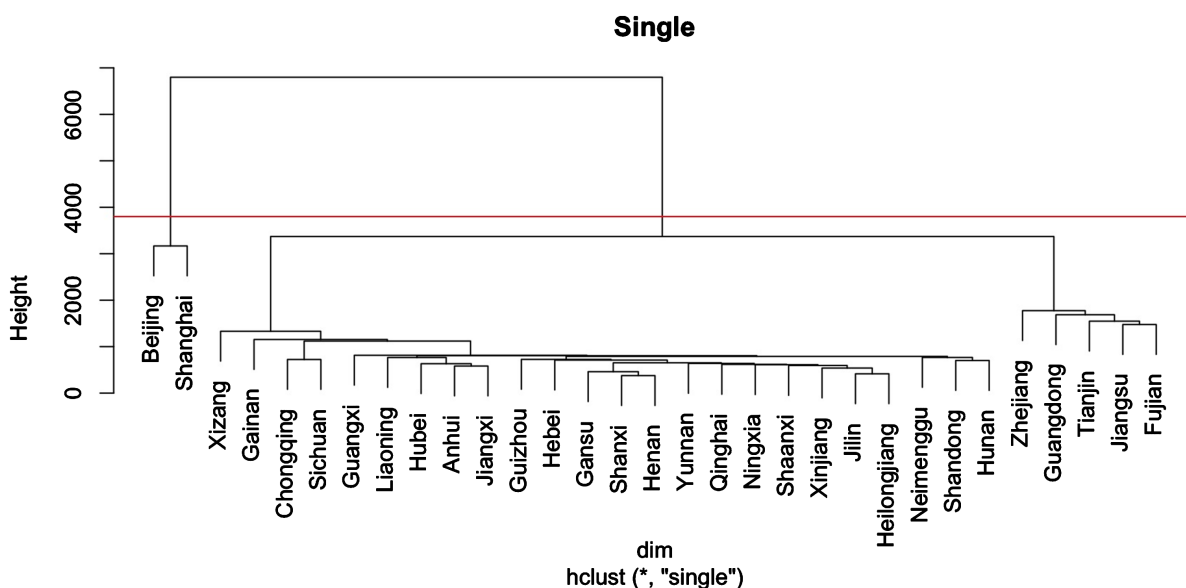


Figure 2. Simple connection method hierarchical clustering

图2. 简单连接法层次聚类

从图2可以看出，简单连接法可以将31个省市城镇居民人均消费情况大致聚为两个群体，具体为北京、上海聚为一个群体，该群体地区物价较高、房价较高、教育发达、娱乐设施齐全且先进。因此，该群体各方面的消费支出都远超全国平均水平；其余省市为一个群体，除北京、上海外的地区均聚为一个群体使得群体样本过多，样本间差异性过大，聚类过于笼统，不能更好地反映实际情况。因此，简单连接聚类效果并不好。

4.1.2. 完全连接法

完全连接法定义两族群间相隔最远的两个个体间的距离，为两族群的距离。31省市城镇居民人均支出完全连接聚类如图3所示。

完全连接法公式： $D(A, B) = \max \{d(y_i, y_j), \text{for } y_i \text{ in } A \text{ and } y_j \text{ in } B\}$

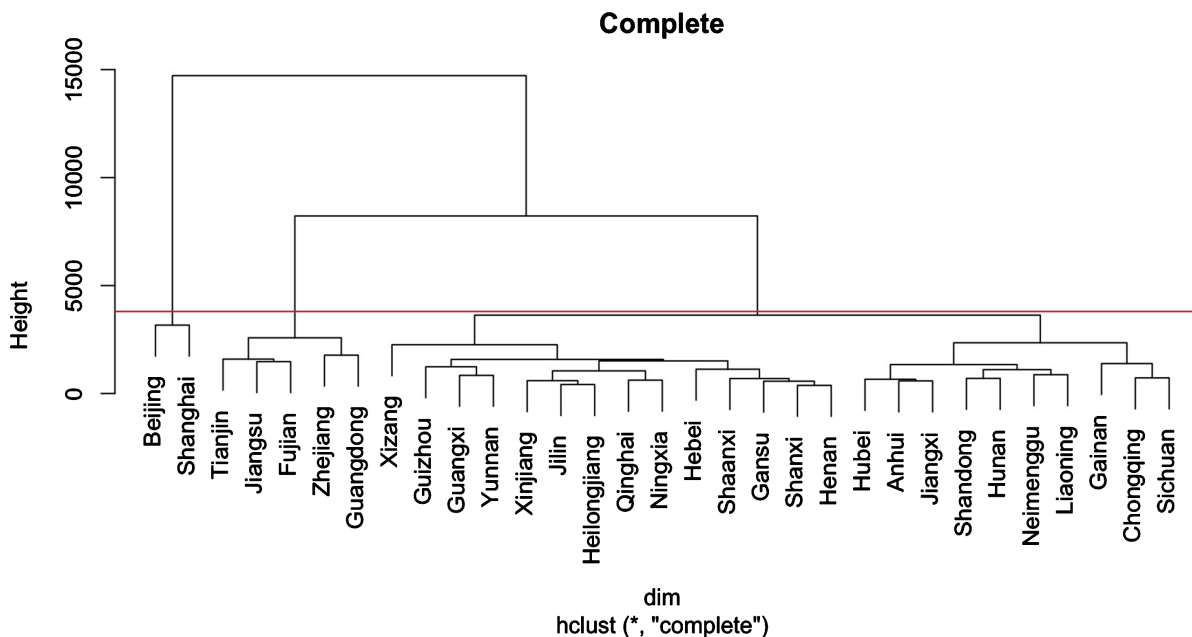


Figure 3. Complete connection method hierarchical clustering
图 3. 完全连接法层次聚类

从图 3 可以看出，完全连接法将 31 个省市城镇居民人均消费支出情况大致聚为四类，具体为：北京、上海聚为一类；浙江、广东、天津、江苏、福建聚为一类；陕西、河南、山西、吉林、黑龙江、广西、新疆、贵州、宁夏聚为一类；河北、湖北、安徽、江西、辽宁、甘肃、山东、湖南、内蒙古、云南、青海、西藏、海南、重庆、四川聚为一类。

前三类聚类效果较好符合实际情况，但在第四类组别中，样本数量较多，情况较为复杂，此组别聚类效果也不完全准确。其中，山东地区经济一直较为发达，湖北、湖南、安徽、重庆等地区近几年来经济发展有显著进步，更多的年轻人青睐于武汉、长沙、重庆等新一线城市，且这些新一线城市房价并不算高。所以，在饮食、教育文化娱乐、医疗保健等方面的支出比例变大，居民消费的结构呈合理化趋势。然而，甘肃、内蒙古、青海、西藏等地区城镇居民收入较低，这些地区的经济发展仍需引起国家措施，加大扶持力度。

4.1.3. 平均连接法

平均连接法为：两族群之间的距离定义为 n_A 个 A 集合点和 n_B 个 B 集合点产生的所有 $n_A n_B$ 个距离数值的平均。31 省市城镇居民人均支出平均连接聚类如图 4 所示。

$$\text{平均连接法公式为： } D(A, B) = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} d(y_i, y_j)$$

从图 4 中可以看出，平均连接法将 31 省市城镇居民人均消费支出大致聚为三类，具体为：北京、上海聚为一类；浙江、广东、天津、江苏、福建聚为一类；其余 24 个省份聚为一类，可以大致分为一线发达地区、较发达地区、欠发达地区。其中，第三类组别的样本含量较大，组内样本差异偏大，如山东与甘肃虽聚为一个组别，但两地区经济发展有很大差异。

4.2. K-Means 聚类分析模型构建

在层次聚类中，一旦个体被分入一个族群，它将不可再被归入另一个族群，而 K-Means 聚类能弥补

层次聚类的不足，它的建模思想为：在所有要聚类的对象中，任意选择 k 个对象作为初始聚类中心，计算每个对象与聚类中心的马氏距离，计算公式为：

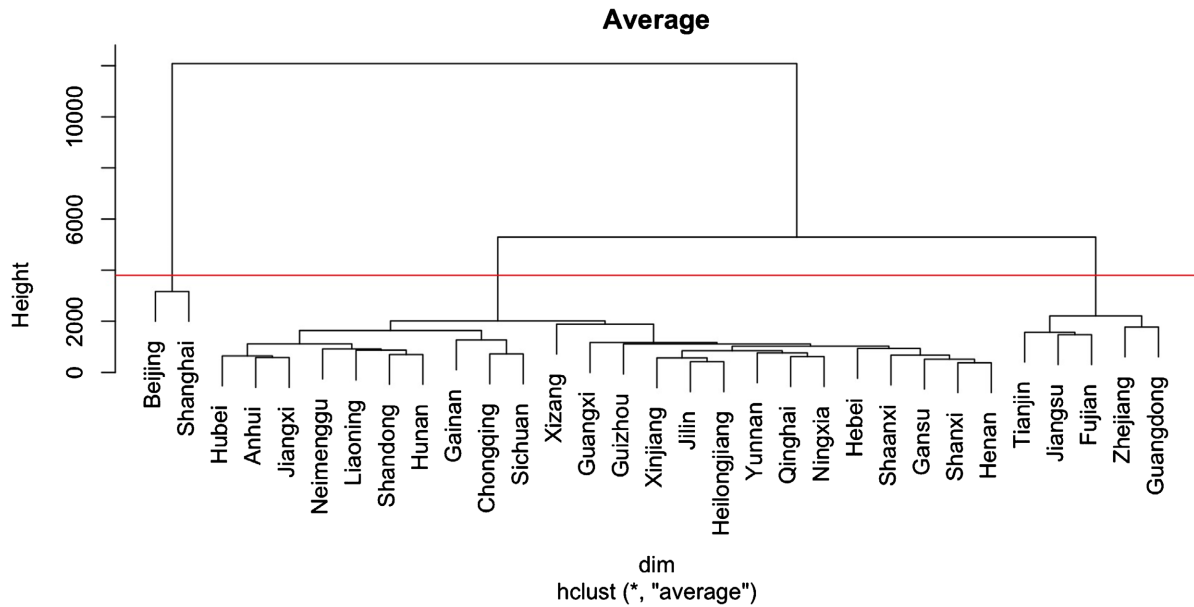


Figure 4. Average join method clustering
图 4. 平均连接法聚类

$$D_s(x, y) = \sqrt{(x - y)^t S^{-1}(x - y)}$$

并根据最小距离重新划分，循环计算聚类中心至聚类中心不再变，划分使得下面式子最小：

$$E = \sum_{j=1}^k \sum_{x_i \in w_j} \|x_i - m_j\|^2$$

根据图 5 碎石图显示，应当聚为 3 类，2022 年我国城镇居民人均消费支出聚类结果如表 2 所示，聚

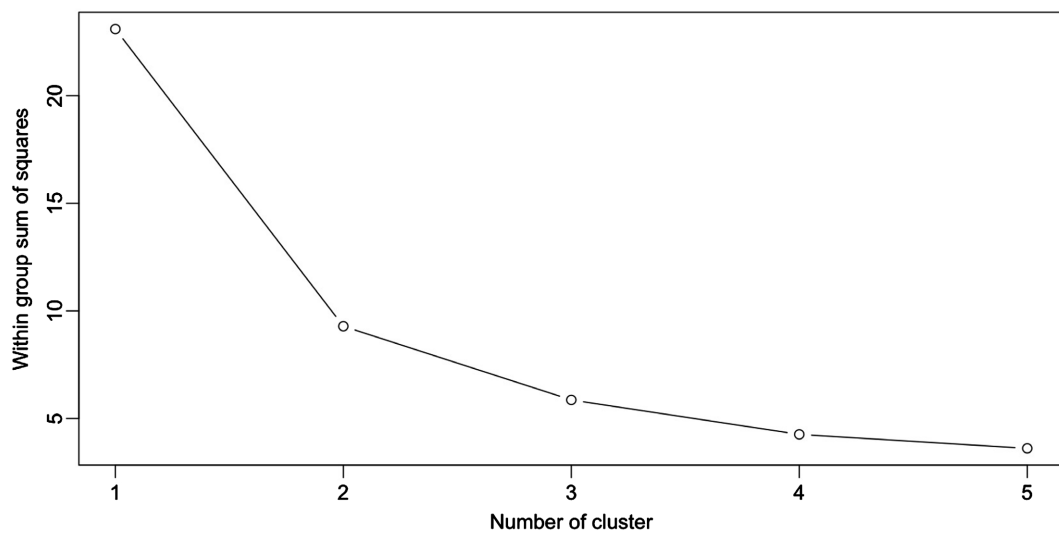


Figure 5. Lithotriptic plan
图 5. 碎石图

类散点图如图 6 所示。具体聚类结果为：北京、上海聚为一类；天津、辽宁、江苏、浙江、福建、山东、湖南、广东、重庆聚为第二类；其余省份聚为第三类。

2022 我国城镇居民人均消费支出聚类分析结果表明：北京、上海聚为第一类地区即发达地区，这类地区经济发达，城镇居民收入偏高，人均消费支出也偏高，聚类结果符合实际情况；天津、江苏、浙江、福建、山东、广东、重庆、辽宁、湖南九个地区聚为第二类地区即较发达地区，但从图 6 聚类结果散点图看出，这类组别内的样本差异较大，较为分散，如辽宁虽被聚到了较发达地区，但与同组的浙江、天津、江苏等地区的差距较大。其余山西、黑龙江、吉林等 20 个省份聚为第三类地区即欠发达地区，同组内的差距也比较大。K-Means 聚类虽然存在一定缺陷但总体聚类效果对比于三种层次聚类较好。

Table 2. Clustering result

表 2. 聚类结果

样本编号	地区	聚类结果
1	北京	1
2	天津	2
3	河北	3
4	山西	3
5	内蒙古	3
6	辽宁	2
7	吉林	3
8	黑龙江	3
9	上海	1
10	江苏	2
11	浙江	2
12	安徽	3
13	福建	2
14	江西	3
15	山东	2
16	河南	3
17	湖北	3
18	湖南	2
19	广东	2
20	广西	3
21	海南	3
22	重庆	2
23	四川	3
24	贵州	3
25	云南	3
26	西藏	3

续表

27	陕西	3
28	甘肃	3
29	青海	3
30	宁夏	3
31	新疆	3

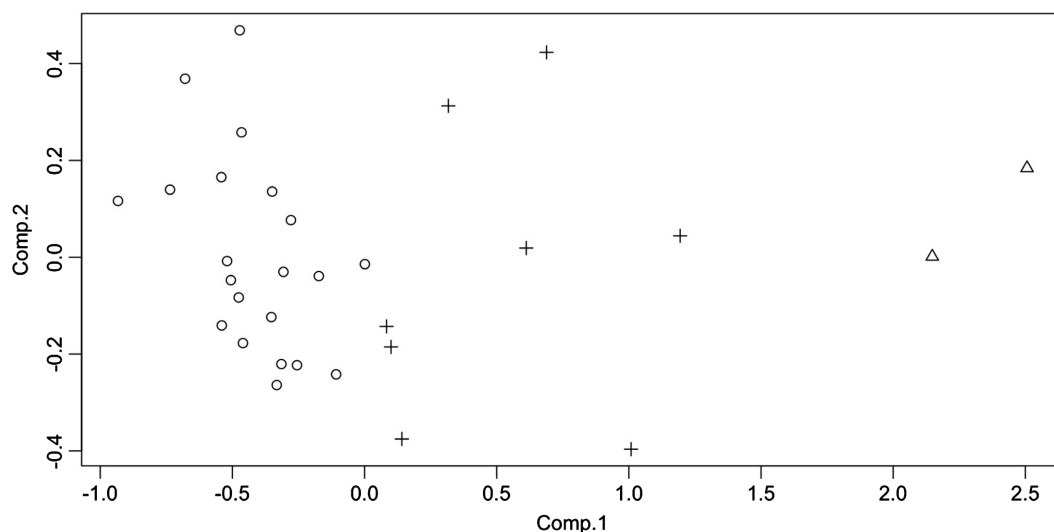


Figure 6. Scatter plot of clustering results
图 6. 聚类结果散点图

5. 促进我国消费结构合理化的发展建议

根据上面两种统计分析表明：经济发展较好的地区，居民的消费水平较高，消费结构也越合理。而且，居民收入水平是影响居民消费需求最直接、最根本的因素，并最终决定着居民的消费层次和消费结构。但是，除收入水平外，还有很多因素影响居民消费结构，如：价格水平、消费习惯、替代品及相关商品的价格、消费环境、消费者预期等。因而，政府也要继续出台切实有效的政策措施，增加居民消费积极性，并引导居民消费结构向更健康、合理的方向发展，以此全面提高我国各地区城镇居民的生活质量。鉴于此，提出以下几点建议。

5.1. 提高居民收入水平、扩大消费基础

收入是影响居民消费需求最直接、最根本的因素，并最终决定着居民的消费层次和消费结构。因此，在国家经济增长的同时，进一步提高城镇居民的收入水平。同时，国家还需采取措施改善经济发展不平衡的现象，尽可能地缩小不同地区间的收入差距，真正发挥大国经济的内部拉动作用，使经济发展走上需求投资拉动的良性发展的道路。

5.2. 改善收入分配结构

改善收入分配结构，增强中低收入人群的消费能力，提高劳动报酬在国民收入分配中的比重，建立健全工资正常增长机制，完善和落实促进农民增收减负的政策措施，将所得税征收重点放在高收入人群，减少中低收入人群的税负，增加城镇居民人均可支配收入。

5.3. 刺激消费，积极培育新的消费热点

当前，我国经济发展收到疫情的影响，由于居民收入会较疫情前收入有所降低，因此消费需求相对不足，旅游方面的消费支出的影响尤其严重。一方面，我们应该进一步巩固已有成果，另一方面还须培育新的消费热点，如医疗保健、保险、健身等。

5.4. 加快中西、东北部发展，缩小地区差异

中西部及东北部地区由于所处地理位置和基础设施建设较差，大多数地区经济基础差，这些地区发展缓慢，发展相对滞后，因此必须加快这些地区的转型发展。政府应在资金、政策上给以重点扶持并加大扶持的力度，为这些地区提供一个良好的外部投资环境，吸引更多的投资资源。此外，各地区也应积极探索发展新思路，借鉴东部沿海经济发达地区的成功经验，同时利用自身的资源优势和后发优势实现跨越式发展，因地制宜探索出一条适合本地区的发展之路，尽可能缩小地区发展的差距。

6. 结语

本文根据 2022 年我国城镇居民人均消费支出数据，利用聚类分析法，对各地区 31 个省、市、自治区城镇居民人均消费进行分类，先运用层次聚类法对各地区人均消费结构进行聚类。然后运用 K-Means 聚类法，对各地区人均消费结构进行 K-Means 聚类，不仅从总体上掌握了消费结构类型的地区分布，而且系统阐述了各地区消费结构的特点，最后在此基础上提出了促进城镇居民消费结构合理化的政策建议，为我国各地区因地制宜地制定消费政策提供了科学合理的决策借鉴。

参考文献

- [1] Matyas, L. and Sevestre, P. (1992) *The Econometrics of Panel Data: Handbook of Theory and Applications*. Kluwer Academic, Dordrecht, 49-87.
- [2] 卢方圆, 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.
- [3] 查道中, 吉文惠. 城乡居民消费结构与产业结构、经济增长关联研究[J]. 经济问题, 2011(7): 19-22.
- [4] 马银晓, 柳虹. 因子分析法在城镇居民消费结构分析中的研究[J]. 浙江工业大学学报, 2012(3): 337-339.
- [5] 李翔, 朱玉春. 农村居民收入与消费结构的灰色关联分析[J]. 统计研究, 2013, 30(1): 76-78.
- [6] 缪迎迎. 合肥市城镇居民消费结构的分析与预测研究[D]: [硕士学位论文]. 合肥: 安徽大学, 2017.
- [7] 柯善淦, 卢新海. 武汉城市群消费结构时空演变特征[J]. 城市问题, 2017(12): 95-103.
- [8] 戴林送. 基于因子分析的城镇居民消费质量发展状况研究[J]. 安庆师范大学学报(自然科学版), 2017, 23(4): 16-21.
- [9] 唐琦, 夏庆杰, 李实. 中国城市居民家庭的消费结构分析: 1995-2013[J]. 经济研究, 2018, 53(2): 35-49.
- [10] 侯淑静. 基于因子分析的全国各省消费水平评价[J]. 中国商论, 2019(6): 67-68.
- [11] 陆丹妮. 因子分析在区域消费结构分析中的应用——基于全国各地城镇居民消费情况[J]. 商讯, 2019(5): 9-10.
- [12] 张芬, 张艳邦, 李晓霞. 我国城镇居民消费评价指标分析——基于因子分析方式[J]. 现代商贸工业, 2022, 43(7): 23-24.