**聚类分析：城镇居民月平均消费**

张世翔

(信计212，210118)

**摘要：**本研究使用“城镇居民月平均消费”文件中的数据，通过系统聚类和动态聚类两种方法，对我国各省市的消费情况进行了聚类分析。分析的目的是为了了解不同地区的消费特征。

**关键词：**聚类分析；消费特征；系统聚类；动态聚类；Z-Score标准化；类别数；广东省；数据分析

# 1 绪论

数据详细反映了我国各省市居民在多个方面的消费情况。具体来说，数据包括八个消费指标：人均粮食支出、人均副食支出、人均日用品支出、人均两时支出等。这些指标综合反映了我国各地区居民的消费习惯、消费能力以及消费结构。本文基于该文件中的数据，对我国不同地区的消费状况进行聚类分析，旨在揭示各地区的消费特征和差异，从而为相关政策的制定和研究提供科学依据。

两种聚类方法都突出了广东省的消费特征，其多项消费指标在全国范围内均排在前列，因此被单独分为一类。这表明广东省的消费水平和结构与其他省市有显著差异。聚类分析揭示了我国不同地区的消费特征，为进一步研究和政策制定提供了依据。

# 2 系统聚类分析

## 2.1 分析操作及其结果

采用系统聚类中的 Ward 法，使用欧氏距离的平方作为样品间距离指标进行聚类分析，给出聚类树型图和凝聚状态表，绘制类间最小距离系数随类数变化的折线图， 确定最后划分归类。

在SPSS中采用【分析-聚类-系统聚类分析】进行系统聚类，选择变量为8个消费指标，个案标注依据为城市名称，选择Ward法，使用欧氏距离的平方作为样品间距离指标，由于数据量级有所差异，进行Z-Score标准化。

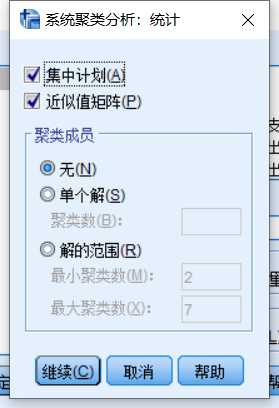


图1 系统聚类分析

结果1：聚类树型图

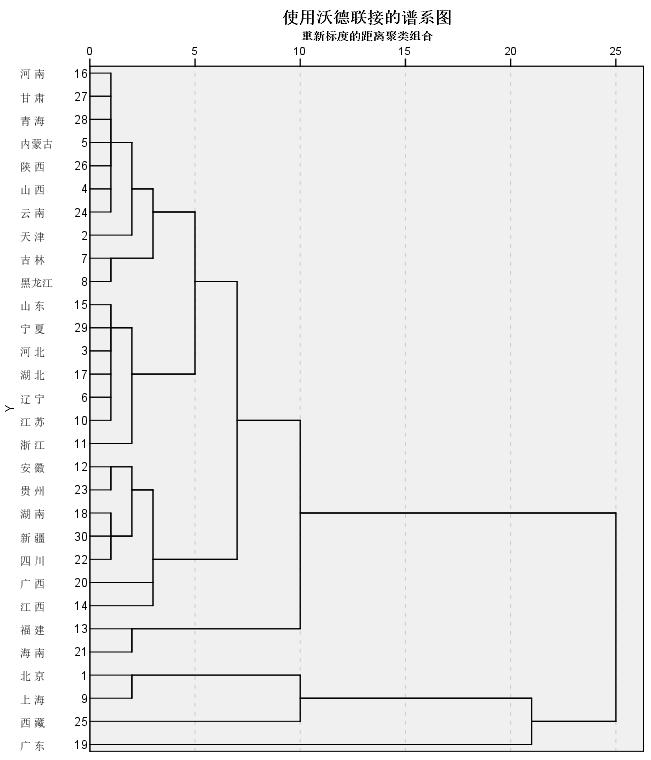


图2 聚类树型图

结果2：凝聚状态表

表1 凝聚状态表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **集中计划** | | | | | | |
| 阶段 | 组合聚类 | | 系数 | 首次出现聚类的阶段 | | 下一个阶段 |
| 聚类 1 | 聚类 2 | 聚类 1 | 聚类 2 |
| 1 | 16 | 27 | .247 | 0 | 0 | 5 |
| 2 | 15 | 29 | .635 | 0 | 0 | 15 |
| 3 | 3 | 17 | 1.034 | 0 | 0 | 13 |
| 4 | 5 | 26 | 1.527 | 0 | 0 | 9 |
| 5 | 16 | 28 | 2.146 | 1 | 0 | 14 |
| 6 | 18 | 30 | 2.771 | 0 | 0 | 11 |
| 7 | 7 | 8 | 3.397 | 0 | 0 | 23 |
| 8 | 12 | 23 | 4.173 | 0 | 0 | 18 |
| 9 | 4 | 5 | 5.051 | 0 | 4 | 12 |
| 10 | 6 | 10 | 6.224 | 0 | 0 | 13 |
| 11 | 18 | 22 | 7.718 | 6 | 0 | 18 |
| 12 | 4 | 24 | 9.265 | 9 | 0 | 14 |
| 13 | 3 | 6 | 11.567 | 3 | 10 | 15 |
| 14 | 4 | 16 | 14.174 | 12 | 5 | 19 |
| 15 | 3 | 15 | 16.785 | 13 | 2 | 20 |
| 16 | 1 | 9 | 19.595 | 0 | 0 | 26 |
| 17 | 13 | 21 | 22.511 | 0 | 0 | 27 |
| 18 | 12 | 18 | 26.035 | 8 | 11 | 21 |
| 19 | 2 | 4 | 30.393 | 0 | 14 | 23 |
| 20 | 3 | 11 | 35.157 | 15 | 0 | 24 |
| 21 | 12 | 20 | 40.462 | 18 | 0 | 22 |
| 22 | 12 | 14 | 46.340 | 21 | 0 | 25 |
| 23 | 2 | 7 | 52.792 | 19 | 7 | 24 |
| 24 | 2 | 3 | 63.015 | 23 | 20 | 25 |
| 25 | 2 | 12 | 77.785 | 24 | 22 | 27 |
| 26 | 1 | 25 | 99.571 | 16 | 0 | 28 |
| 27 | 2 | 13 | 122.133 | 25 | 17 | 29 |
| 28 | 1 | 19 | 172.223 | 26 | 0 | 29 |
| 29 | 1 | 2 | 232.000 | 28 | 27 | 0 |

结果3：折线图

在Excel中绘制类间最小距离系数随类数变化的折线图。可以发现，当类别数为4时，折线的下降趋势趋缓，故可将类别数设定为4。

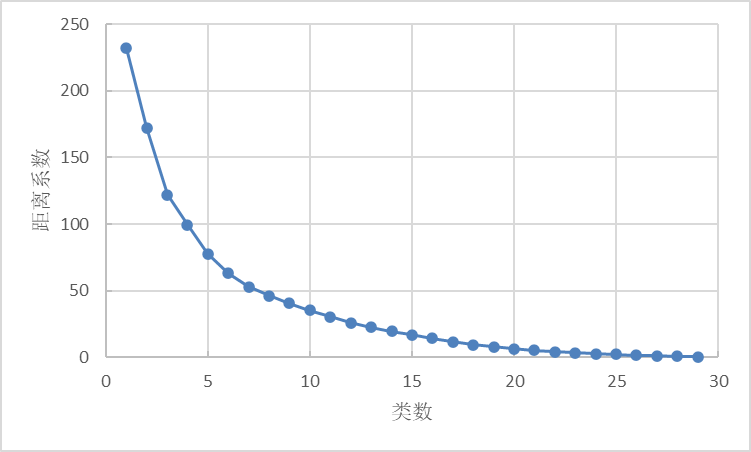


图3 折线图

结果4：每个样品所属类别

表2 分类结果

|  |  |
| --- | --- |
| **类别** | **城市** |
| 1 | 北京、上海、西藏 |
| 2 | 天津、河北、山西、内蒙古、辽宁、吉林、黑龙江、江苏、浙江、安徽、江西、山东、河南、湖北、湖南、广西、四川、贵州、云南、陕西、甘肃、青海、宁夏、新疆 |
| 3 | 福建、海南 |
| 4 | 广东 |

## 2.2聚类结果解释

在系统聚类分析中，采用了Ward法和欧氏距离的平方作为样品间距离指标，并对数据进行了Z-Score标准化。聚类结果显示，最佳的类别数为4，这是通过观察类间最小距离系数随类数变化的折线图确定的。在这四个类别中，北京、上海、西藏被归为一类；天津、河北等多个城市被归为第二类；福建和海南被归为第三类；广东单独成为第四类。

这一发现是基于类间最小距离系数随类数变化的折线图分析得出的。在确定的四个类别中，北京、上海、西藏归为一类，反映了这些地区可能有相似的消费模式或水平；天津、河北等多个城市归为第二类，显示出另一种消费特征；福建和海南归为第三类；而广东因其独特的消费指标，在全国范围内显著不同，因此单独成为第四类。

结果揭示了中国不同地区之间在消费行为上的差异，为进一步的经济分析和政策制定提供了有价值的洞见。特别是广东省的独特性可能指向该地区的经济发展水平和居民消费能力的高度发展。系统聚类分析方法为理解和解释区域经济差异提供了一个强有力的工具。

# 3 动态聚类分析

## 3.1分析操作及其结果

采用动态聚类法进行聚类（分类数与系统聚类法一致），比较分类结果与系统聚类法的一致性**第一步，**利用【分析-描述性分析-描述】对数据进行Z-Score标准化。

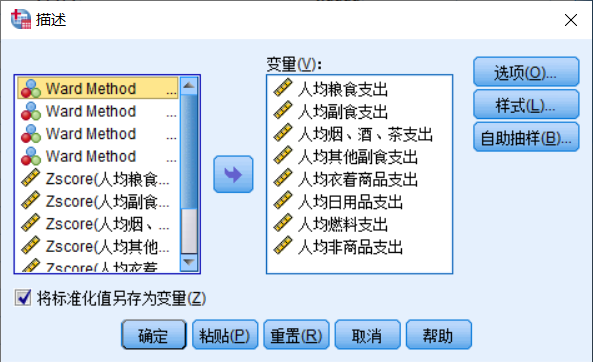


图4 数据标准化

**第二步，**采用【分析-聚类-K-均值聚类分析】进行动态聚类，将聚类数选为4。

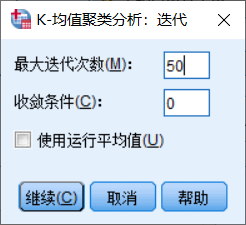
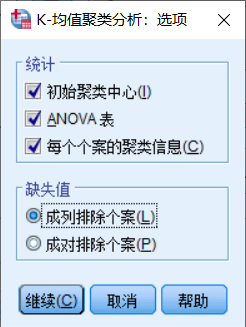
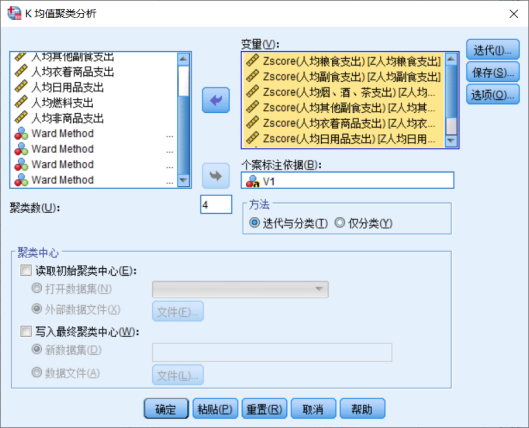


图5 动态聚类

结果1：最终聚类中心及其之间的距离

图6 最终聚类中心及其之间的距离

结果2：聚类结果

表3 分类结果

|  |  |
| --- | --- |
| **类别** | **城市** |
| 1 | 北京、上海、浙江 |
| 2 | 天津、河北、山西、内蒙古、辽宁、吉林、黑江苏、安龙江、徽、江西、山东、河南、湖北、湖南、广西、四川、贵州、云南、陕西、甘肃、青海、宁夏、新疆、福建、海南 |
| 3 | 西藏 |
| 4 | 广东 |

## 3.2聚类结果解释

动态聚类分析的结果与系统聚类法大致相同，但有细微差别。动态聚类将西藏单独归为一类，并将浙江与北京、上海归为同一类，同时将福建和海南并入到第二类中。这些差异反映了不同聚类方法在处理数据时的特点。动态聚类法对于中国不同地区的消费特征也得出了类似的结论，但在一些细节上可能存在微小的差异。这一结果为我们理解区域经济差异提供了更全面的视角。

# 4 研究的具体意义及解释

## 4.1聚类结果现实意义

了解各地区的消费结构有助于政府制定更有针对性的经济政策。例如，对于像广东这样消费水平高的地区，可以重点发展高端服务业和消费品市场。而对于西藏这种消费结构独特的地区，政府可以考虑提供更多的基础消费品和服务，以满足当地居民的特殊需求。企业可以根据不同地区的消费结构制定差异化的市场策略。例如，在广东省，企业可以推出更多高档消费品和个性化服务，而在西藏，企业可能需要更加关注基础生活用品和传统文化相关的商品。通过分析各地区的消费特点，可以更好地理解区域经济的发展水平和潜力。例如，浙江、北京和上海的消费结构相似，表明这些地区具有强大的消费市场，适合进行高投入高回报的经济活动。而像西藏这样的地区，可能需要更多的基础设施建设和政策支持，才能推动经济发展。政府和非营利组织可以根据各地区的消费结构，合理分配社会资源。例如，在消费能力较弱的地区，可以加大教育、医疗和社会保障的投入，提高居民的生活水平和消费能力。

## 4.2总体聚类结果解释

通过对城镇居民月平均消费数据的聚类分析，我们可以发现，两种聚类方法的结果大致相同，都将广东省单独分为一类。这表明广东省在多个消费指标（如人均粮食支出、人均副食支出、人均日用品支出和人均两时支出）上均显著高于全国其他地区，显示出其独特的消费特征。广东省作为经济发达地区，其居民的消费能力和消费结构明显不同于其他省市。

两种聚类方法也存在一定的差异。动态聚类方法将西藏单独归为一类，而静态聚类方法并未将其单独分出。这一差异可能反映了西藏在特定时期内消费结构的独特性。西藏地处偏远，其经济发展水平、文化习俗和地理环境都与内地有显著不同，这些因素共同作用，使得西藏在动态聚类分析中呈现出独特的消费模式。

动态聚类方法还将浙江归入北京、上海一类中。这表明浙江的消费结构与北京、上海这两个直辖市有较高的相似性。作为经济发达地区，北京、上海和浙江的居民在消费习惯、消费能力和消费偏好上具有一定的一致性，这也反映了这些地区在经济发展和城市化进程中的共同特征。

另一个显著的差异在于，动态聚类方法去除了福建和海南的单独一类，将其归入到大部分城市所在的类别中。这一结果表明，福建和海南在消费结构上与全国大部分地区更为相似。尽管福建和海南在地理位置和经济发展水平上有一定的独特性，但其消费结构却与全国大部分地区趋同，可能是因为这些地区的居民消费习惯和消费能力与全国平均水平较为接近。

两种聚类方法都揭示了我国不同地区的消费特征及其差异性。静态聚类方法主要基于各地区在某一时点上的消费数据进行分析，而动态聚类方法则考虑了时间维度，分析各地区消费结构的变化和趋势。通过比较这两种方法的结果，我们不仅可以了解各地区在某一时点上的消费特征，还可以观察到这些特征在时间上的动态变化。

动态聚类方法在处理西藏、浙江、福建和海南这些地区时显示出其优势。通过将西藏单独归为一类，动态聚类方法更准确地反映了该地区的独特消费结构。将浙江与北京、上海归为一类，则突显了这些经济发达地区在消费结构上的一致性。去除福建和海南的单独一类，动态聚类方法则揭示了这些地区与全国大部分城市在消费结构上的相似性。

这种差异也为我们提供了进一步研究的线索。为什么西藏的消费结构如此独特？浙江与北京、上海的消费模式为何如此相似？福建和海南在消费结构上的趋同性背后有哪些经济和社会因素？这些问题都值得我们在后续研究中深入探讨。

总体而言，本文通过两种聚类方法对我国各地区的消费结构进行了系统分析，揭示了各地区在消费习惯和消费能力上的显著差异。这些发现不仅有助于我们更好地理解我国各地区的经济和社会发展状况，还为政府和企业制定针对性的政策和策略提供了宝贵的参考。通过进一步深入分析这些聚类结果，我们可以获得更多关于我国不同地区消费特征和趋势的洞见，从而为推动我国经济的均衡发展和居民生活水平的提升贡献力量。

**参考文献**

[1]冯梅.聚类分析在公务员招聘中的应用及SPSS实现[J].数学的实践与认识,2006,(10):46-52.

[2]宫长亮,李磊.我国区域可持续发展的聚类分析[J].哈尔滨理工大学学报,2004,(04):59-61+65.

[3]韩胜娟.SPSS聚类分析中数据无量纲化方法比较[J].科技广场,2008,(03):229-231.

[4]胡纯严,胡良平.合理进行多元分析——变量聚类分析[J].四川精神卫生,2023,36(S1):61-66.

[5]轩书婷.基于离散哈希的聚类分析[D].烟台大学,2023.

[6]黄淑娉.基于聚类分析和集成学习的多组学联合分析研究[D].华中农业大学,2023.

[7]王骏,王士同,邓赵红.聚类分析研究中的若干问题[J].控制与决策,2012,27(03):321-328.

[8]许丽利.聚类分析的算法及应用[D].吉林大学,2010.

[9]郑兵云.多指标面板数据的聚类分析及其应用[J].数理统计与管理,2008,(02):265-270.

