Advances in Applied Mathematics 应用数学进展, 2024, 13(1), 360-375
Published Online January 2024 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2024.131038

Hans汉斯

基于多元统计方法的全国消费支出分类和差异研究

翁鹏飞[1]，夏　莉[1,2*]

[1]广东财经大学统计与数学学院，广东 广州
[2]广东财经大学大数据与教育应用统计实验室，广东 广州

收稿日期：2023年12月25日；录用日期：2024年1月19日；发布日期：2024年1月25日

摘　要

本文主要以我国31个省、自治区和直辖市的经济发展水平为研究对象，选取能反映经济发展水平的8项人均消费支出为变量，通过平行坐标图、星图、热图和聚类图等可视化得到分类差异，接着使用聚类分析和主成分分析对各地区的消费水平进行分类，最后分析差异并给出部分结论及建议。各地区消费支出存在明显的差异与规律，其中食品烟酒、居住和交通通信是绝大部分地区消费支出的主要构成部分且表现出一定的地区特征，具有较强的异质性。基于此，建议政府加强重视消费支出的数据分析，及时了解各地区消费趋势及其对于经济社会发展的影响；鼓励消费者合理分配消费支出，使其多元化、合理化，推动各地区消费结构的升级；希望进一步激发中西部地区消费潜力，推动经济发展的全面均衡等。

关键词

数据可视化，分类差异，聚类分析，主成分分析

Research on the Classification and Difference of Consumption Expenditure in China Based on Multivariate Statistics

Pengfei Weng[1], Li Xia[1,2*]

[1]School of Statistics and Mathematics, Guangdong University of Finance and Economics, Guangzhou Guangdong
[2]Laboratory of Big Data and Educational Application Statistics, Guangdong University of Finance and Economics, Guangzhou Guangdong

Received: Dec. 25th, 2023; accepted: Jan. 19th, 2024; published: Jan. 25th, 2024

*通讯作者。

文章引用: 翁鹏飞, 夏莉. 基于多元统计方法的全国消费支出分类和差异研究[J]. 应用数学进展, 2024, 13(1): 360-375.
DOI: 10.12677/aam.2024.131038

Abstract

This paper mainly takes the economic development level of 31 provinces, autonomous regions and municipalities directly under the central government as the research object, and selects 8 per capita consumption expenditures that can reflect the economic development level as variables. The classification differences were obtained through the visualization of parallel coordinate maps, star maps, heat maps and cluster maps, and then the consumption levels of each region were classified by cluster analysis and principal component analysis; finally, the differences were analyzed and some conclusions and suggestions were given. There are obvious differences and patterns in consumer expenditure in different regions, among which food, tobacco and alcohol, housing, transportation and communication are the main components of consumer expenditure in most regions, and they show certain regional characteristics and have strong heterogeneity. Based on this, it is suggested that the government should pay more attention to the data analysis of consumer expenditure, and keep abreast of the consumption trends in various regions and their impact on economic and social development. Encourage consumers to rationally allocate consumption spending, diversify and rationalize it, and promote the upgrading of consumption structure in various regions; it is hoped that the consumption potential of the central and western regions will be further stimulated and the economic development will be comprehensively balanced.

Keywords

Data Visualization, Classification Difference, Cluster Analysis, Principal Component Analysis

1. 引言

2020 年新冠疫情暴发以来对全球经济造成了巨大冲击，引发了人们对疫情影响下和疫情常态化时期经济政策的深入思考和讨论。与此同时，中国对“一带一路”建设的持续关注和经济高质量发展的探索成为热门话题。在新的经济发展背景下，高质量发展逐渐取代传统的单纯经济增长，成为关注焦点。在进入新一轮结构改革的背景下，中国经济结构预计将持续优化，改革的深度和广度也将进一步加强，新的经济增长动力也将得到更大的开发和挖掘[1]。

随着中国经济的持续发展以及人民生活水平的提高，消费支出已经成为评价国民经济发展和生活质量的重要指标之一。人均消费支出水平反映了一个地区居民的生活水平和消费习惯，对该地区的经济社会发展研究具有非常重要的意义。为此本文旨在利用多元统计方法，对中国 31 个地区的消费支出进行分类和差异研究[2] [3]。首先，通过对消费支出数据的可视化分析，将不同地区划分为几种类型，并分析各类型地区之间的消费支出差异，并对这些差异进行深入分析[4]。其次，运用主成分分析和聚类分析等多元统计方法，探索不同地区消费支出的主要特征及地区之间的差异性。因为主成分分析可以帮助我们找出影响消费支出的主要因素，而聚类分析能够揭示地区之间在消费结构上的相似性与差异性，有助于深入理解不同地区的消费行为[5] [6]。最后，本文根据研究结果提出相应的政策建议，旨在促进地区经济发展和消费结构的优化升级，为政府部门、学术界和企业提供参考，从而推动区域经济的协调发展，并提高人民的生活水平。

2. 数学模型的建立与求解

2.1. 数据来源和指标选取

本文数据源于《2022 中国统计年鉴》[7]，具体数据见附表 1。为更加全面地评价 2021 年我国 31 个省、自治区和直辖市(以下简称 31 个地区)的经济发展状况，本文结合各省市经济发展实际情况和数据的科学性、可得性及可操作性等原则，选取能够反映我国 31 个省市经济发展水平的 8 个指标：食品烟酒支出(x_1)、衣着支出(x_2)、居住支出(x_3)、生活用品及服务支出(x_4)、交通通信支出(x_5)、教育文化娱乐支出(x_6)、医疗保健支出(x_7)和其他用品及服务支出(x_8)。

2.2. 研究方法

首先使用 R 语言完成数据可视化及分析，然后结合主成分分析和聚类分析进一步研究得出部分结论并给出相关建议。

3. 数据可视化

通过 RStudio 绘制平行坐标图、星图、热图、聚类图等，把各地区人均消费支出数据可视化，从而得到直接美观更能显示差异的结果比较图。然后再运用主成分分析、聚类分析等方法对 2021 年全国居民消费水平和消费结构进行分类比较研究，给出相关的优劣情况。

3.1. 平行坐标图可视化[8]

3.1.1. 均值条图

均值条图通常用来比较各变量在不同观察单位上的均值变化大小，笔者对附表 1 的 31 个地区八项指标作均值比较条图，如下图 1。可以看出，不同地区的人均消费差异比较明显。

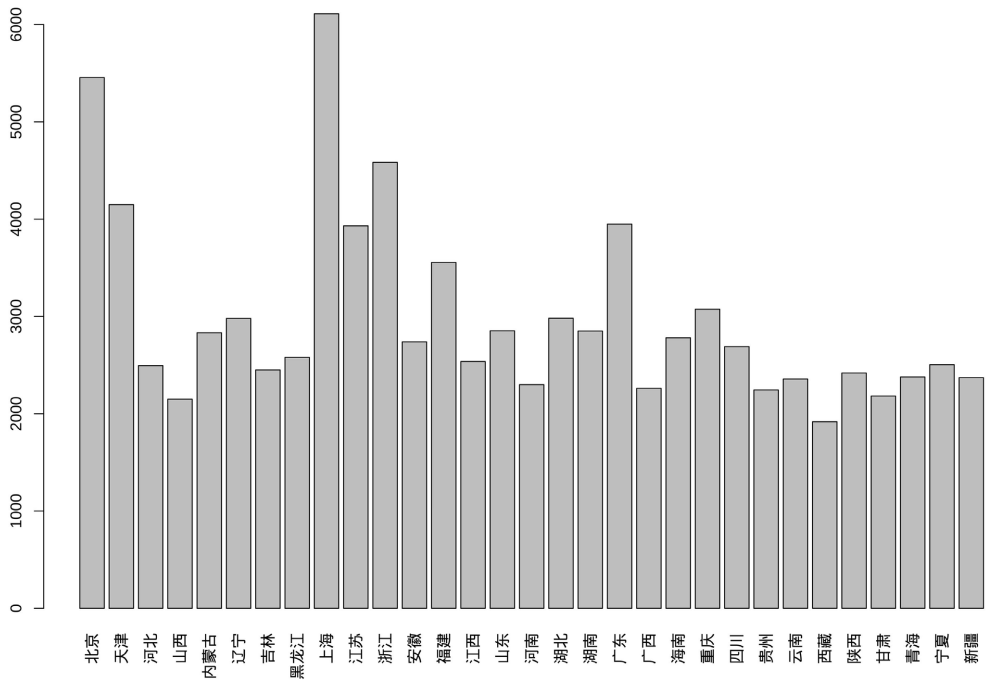

Figure 1. Mean bar chart
图 1. 均值条图

3.1.2. 饼图和箱尾图

通过附表 1 的数据，使用 RStudio 绘制出下面的饼图(图 2)和箱尾图(图 3)。从中可以看出，食品烟酒和居住消费支出远高于其他项目，这也和现实中房价居高不下、烟草的税费快持平甚至反超我国国防军费等情况相互印证。

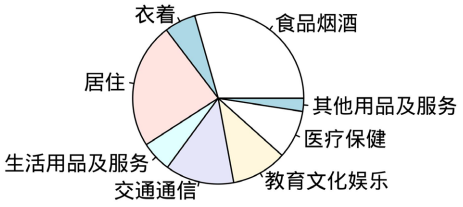

Figure 2. Pie charts
图 2. 饼图

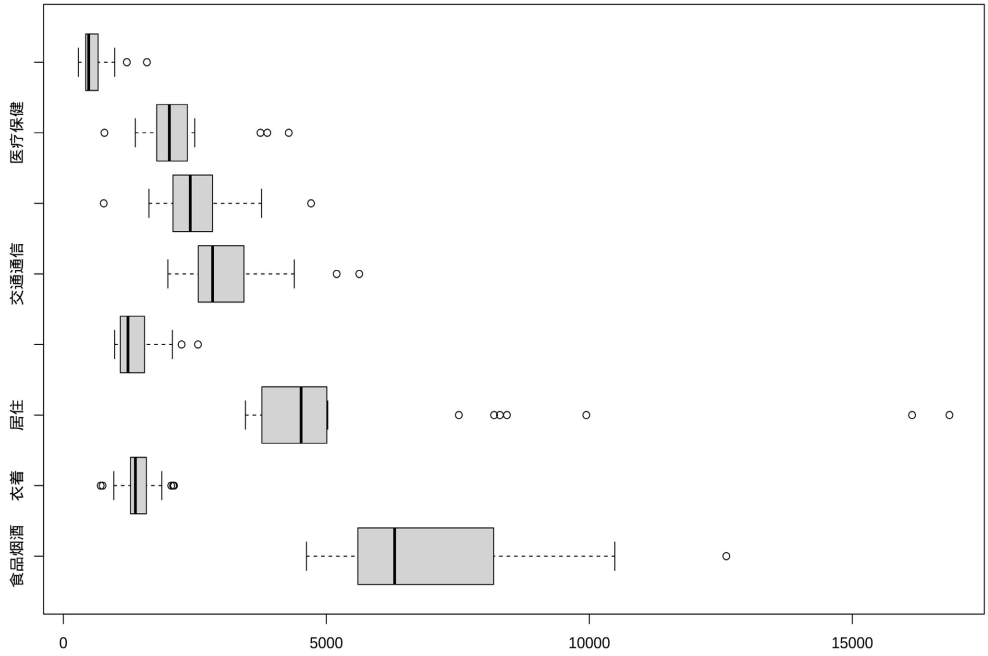

Figure 3. Box plots
图 3. 箱尾图

3.2. 星图[8]

每一个星图都是由多个角构成，用线段离中心的长度来表示变量值的大小，用于展示多个变量的个体，每个变量的图形相互独立，即每个角都有一条轴线与中心点连接起来这多条轴线，它们分别对应了数据的维度，数值越大，轴线越长，画出来的星图也就越大。笔者通过附表 1 数据使用 RStudio 绘制出了彩色的 360 度星图，如下图 4。

图 4 包括 31 个地区的星相图，其中星相图每种颜色对应不同变量，每个角的大小对应表示该变量值的大小。从图中可以看出，上海、北京、浙江三个地区的消费情况属第一梯队尤为突出，天津、广东、江苏三个地区的消费情况属第二梯队较为突出，令人意外的是广东省的人均消费不能排进前五，进一步查阅相关资料可以印证广东富的只是珠三角地区，还真应了“最富的在广东，最穷的也在广东”。

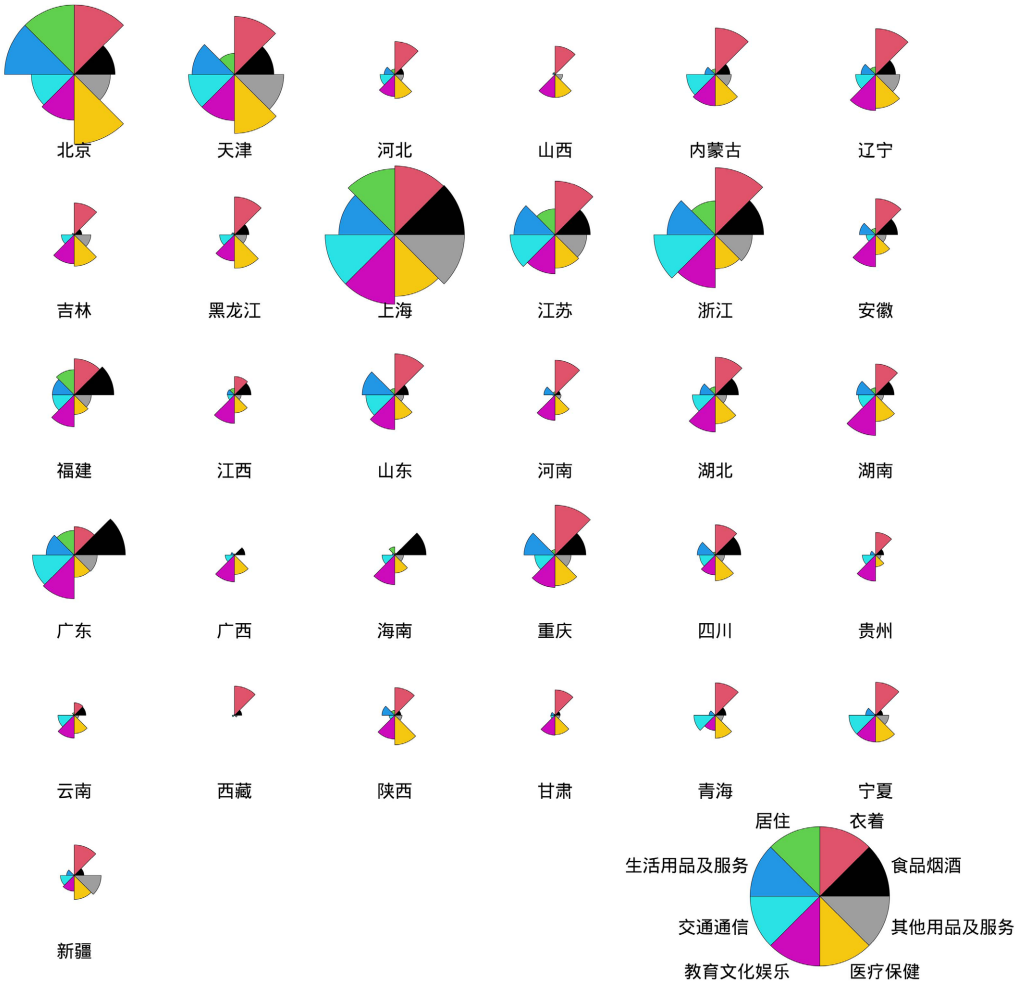

Figure 4. Star chart
图 4. 星图

除此之外从图中还可以看出，各地区在不同消费支出领域上的分布情况存在明显差异。食品烟酒、居住和交通通信是各地区消费支出的主要构成部分。而教育文化娱乐、医疗保健和其他用品及服务等方面的消费支出则表现出较大的地区差异。

3.3. 热图[8]

热图是一种可以直观地展示矩阵数据的可视化工具。它将每个数据点的值映射为颜色，并在颜色映射条旁边显示对应的数据值，形成一个热度分布图。热图通常用于显示多维数据的相关性、分类等信息，以及帮助用户在大量数据中找出规律和异常值等。热图可以非常有效地说明数据间的相似性和差异性。在矩阵数据中，每个行列值对应一个数据值，在热图中，相似的数据会聚集在一起，使相似的颜色单元格集中在热图的中心区域，从而更容易辨认和快速分析数据间的联系。此外，通过调整颜色映射条的取值范围和调色板的选取等参数，还可以突显出矩阵数据中的重点区域、异常值等信息。总之，热图是一种可以直观展示矩阵数据关系的有用工具，可以通过热图帮助用户更好地理解和有效地分析数据。

如此绘制下图 5 和图 6 热力图，发现前面提到的大部分分析情况也可以在热图中明显观测到且更具一番美感。

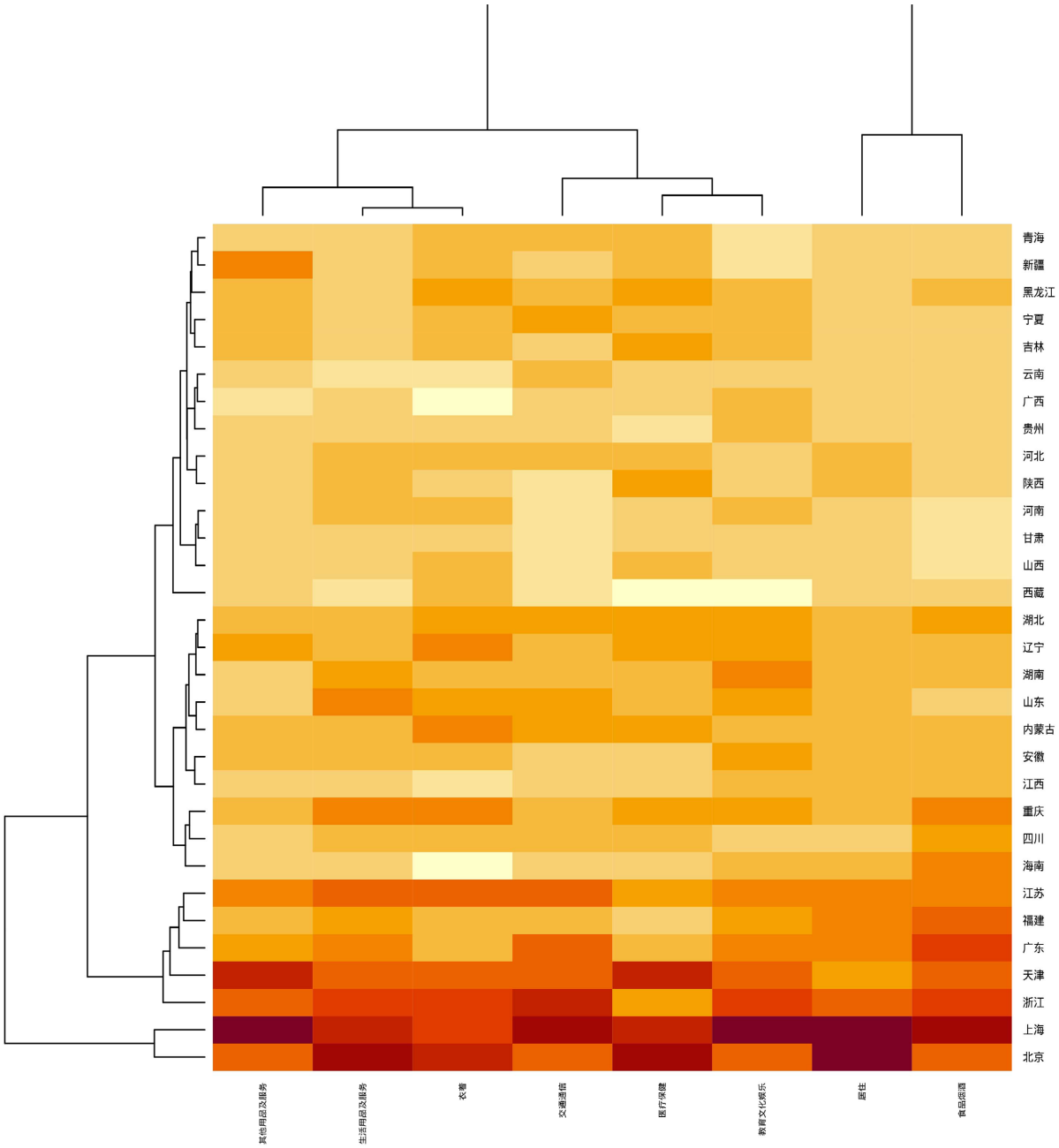

Figure 5. Heat map
图 5. 热图

4. 基于主成分分析和聚类分析的研究评价比较

4.1. 主成分分析

下面应用主成分分析方法，以附表 1 的八个指标作为原始变量，通过 R 对我国 31 个省的人均消费水平作分析评价，并根据因子得分和综合得分对各省的人均消费水平进行比较综合分析[9] [10]。

1) 计算相关矩阵，结果如下表 1：

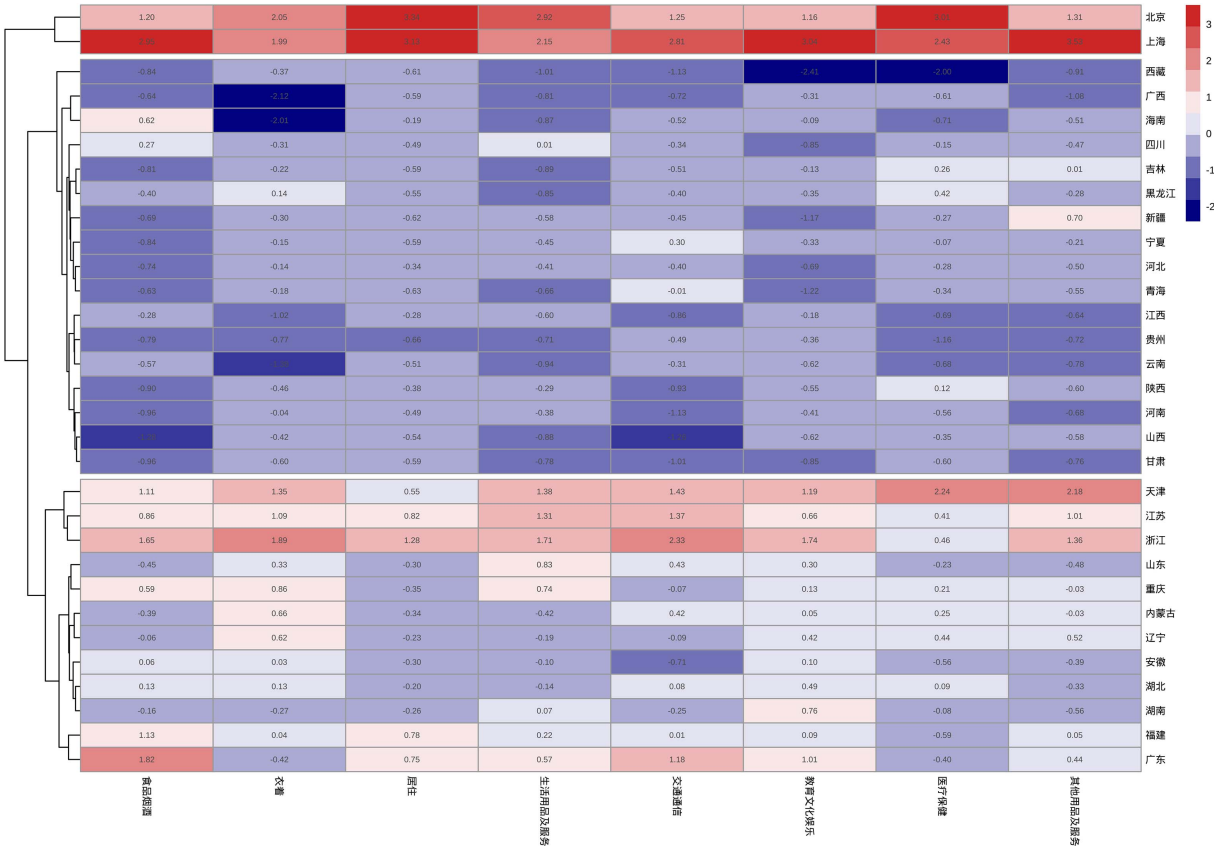

Figure 6. Numerical heat map
图 6. 数值热图

Table 1. Numerical heat map
表 1. 相关矩阵结果

成分	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
食品烟酒	1.0000000	0.5851628	0.8285290	0.7994380	0.8465594	0.8215304	0.5954472	0.7983909
衣着	0.5851628	1.0000000	0.6964916	0.8268234	0.7479897	0.6357014	0.7583285	0.7810476
居住	0.8285290	0.6964916	1.0000000	0.8866806	0.7944438	0.7738795	0.7674076	0.8063342
生活用品及服务	0.7994380	0.8268234	0.8866806	1.0000000	0.8435361	0.7934366	0.7919232	0.7920094
交通通信	0.8465594	0.7479897	0.7944438	0.8435361	1.0000000	0.8363051	0.7083823	0.8665745
教育文化娱乐	0.8215304	0.6357014	0.7738795	0.7934366	0.8363051	1.0000000	0.7345718	0.7866788
医疗保健	0.5954472	0.7583285	0.7674076	0.7919232	0.7083823	0.7345718	1.0000000	0.8169766
其他用品及服务	0.7983909	0.7810476	0.8063342	0.7920094	0.8665745	0.7866788	0.8169766	1.0000000

2) 求得相关矩阵的特征值和主成分负荷：

Table 2. The standard deviation, difference proportion, and cumulative proportion of correlation matrices
表 2. 相关矩阵的标准差、差异比例及累计比例情况

成分	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
标准差	2.5375479	0.7477244	0.5426164	0.5206558	0.4475964	0.3253921	0.2862566	0.2192676
差异比例	0.8048937	0.0698864	0.0368040	0.0338853	0.0250428	0.0132350	0.0102428	0.0060097
累计比例	0.8048937	0.8747801	0.9115842	0.9454695	0.9705123	0.9837473	0.9939902	1.0000000

Table 3. The load matrix of the correlation matrix
表 3. 相关矩阵的载荷矩阵

成分	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
食品烟酒	0.345	0.559	0.162	0	0.198	0.340	0.538	0.306
衣着	0.330	−0.575	0.529	−0.118	−0.169	0.377	−0.102	0.292
居住	0.360	0.107	0	0.646	0.248	0	−0.580	0.190
生活用品及服务	0 .370	0	0.217	0.427	−0.309	−0.198	0.329	−0.619
交通通信	0.365	0.181	0.267	−0.373	0	−0.749	−0.106	0.224
教育文化娱乐	0.351	0.297	−0.374	−0.278	−0.634	0.283	−0.287	0
医疗保健	0.338	−0.461	−0.650	0	0	−0.163	0.376	0.278
其他用品及服务	0.365	0	0	−0.407	0.607	0.187	−0.144	−0.517

3) 确定主成分

从表2中的累计比例可以看出，前面2个主成分的方差和占全部方差的比例为87.478%，已超过85%，即基本上保留了原始指标的大部分信息，大致能够解释各省份之间的人均消费水平差异，因此取前 2 个成为作为主成分。

从表 3 中的主成分载荷矩阵可以看出，变量 x_1，x_2，x_3，x_4，x_5，x_6，x_7，x_8 在成分 1 上的因子载荷系数为 0.345、0.330、0.360、0.370、0.365、0.351、0.338、0.365，成分 2 上的因子载荷系数 0.559、−0.575、0.107、0、0.181、0.297、−0.461、0，把这两个主成分的得分记为 F_1，F_2，主成分系数可以根据因子荷载矩阵(表 3)和累计贡献率(表 3)得出，其方法为：分别将因子载荷矩阵的第 m 列的每个元素除以其对应特征根的平方根，即可得到主成分分析的第 m 个主成分的系数，可以得到主成分表达式：

$$F_1 = 0.2165768x_1 + 0.3816305x_2 + 0.4887154x_3 + 0.5127744x_4 + 0.5455688x_5 + 0.6153236x_6 + 0.6317408x_7 + 0.7794811x_8$$

$$F_2 = 0.3509172x_1 - 0.6649623x_2 + 0.1452570x_3 + 0.2705423x_5 + 0.5206584x_6 - 0.8616346x_7$$

以所选取的第 1 和第 2 主成分的方差贡献率 α_1 和 α_2 作为权数，构建综合评价模型：$F = \alpha_1 F_1 + \alpha_1 F_2$。其中 F 为综合评价指标，分别带入综合评价模型，可得出综合评价模型：

$$F = 0.8048937F_1 + 0.0698864F_2$$

此外，主成分 Comp.1 在居住、生活用品及服务、教育文化娱乐、交通通信和其他用品及服务上的载荷值都很大，说明第一主成分基本可以反映出这五个指标的信息；Comp.2 在食品烟酒、衣着和医疗保健上有较大的载荷，表明第二主成分基本可以反映出这三个指标的信息。

4) 各地区的主成分得分及排名情况

有了各个主成分的解释，结合各个省、市、自治区在两个主成分上的得分和综合得分，就可以对各省、市、自治区的综合人均消费水平进行评价了。

最后，由加权法估计出综合得分，以各主成分的方差贡献率占两个主成分贡献的比重作为权重进行加权汇总，得出各个地区的综合得分，即(具体情况如下表 4 所示)

$$PC = \frac{2.537^2 \times \text{Comp.1} + 0.747^2 \times \text{Comp.2}}{2.537^2 + 0.747^2}$$

Table 4. Principal component scores and rankings of 31 places in China
表 4. 全国 31 地的主成分得分及排名情况

地区	Comp.1	Comp.2	PC (得分)	rank (排名)
北京	5.82981759	−1.32574323	5.258157461	2
天津	4.10350824	−0.79196247	3.712407456	4
河北	−1.26530818	−0.45651935	−1.200693776	20
山西	−2.14613893	−0.67840494	−2.028881177	28
内蒙古	0.05160672	−0.63367733	−0.003140851	12
辽宁	0.48931861	−0.53934191	0.407138575	9
吉林	−1.05172673	−0.57739529	−1.013832230	18
黑龙江	−0.84383425	−0.65246171	−0.828545430	17
上海	7.93334756	0.72490377	7.357462589	1
江苏	2.72113807	0.01760585	2.505151987	5
浙江	4.47595286	0.47311017	4.156164428	3
安徽	−0.68289067	0.17873024	−0.614055491	14
福建	0.62916384	0.98388479	0.657502619	7
江西	−1.62813723	0.61452995	−1.448969808	24
山东	0.16447060	−0.24098957	0.132078256	10
河南	−1.68925507	−0.56395352	−1.599354351	25
湖北	0.07269817	0.12806940	0.077121800	11
湖南	−0.26987700	0.29220476	−0.224972100	13
广东	1.81368583	1.99424423	1.828110704	6
广西	−2.44621736	1.01240163	−2.169907144	30
海南	−1.51277443	1.82766056	−1.245905971	21
重庆	0.72918996	−0.34485389	0.643384239	8
四川	−0.84403936	0.06280355	−0.771591378	15
贵州	−2.02022329	0.38505309	−1.828064964	26
云南	−2.06527761	0.63533903	−1.849524448	27
西藏	−3.32646410	−0.17359152	−3.074580066	31
陕西	−1.44339613	−0.61332472	−1.377081445	22
甘肃	−2.21690636	−0.30125381	−2.063864244	29
青海	−1.51612874	−0.43322610	−1.429615292	23
宁夏	−0.84108612	−0.41239158	−0.806837570	16
新疆	−1.20421649	−0.59145007	−1.155262377	19

5) 主成分作图

以第一主成分为横轴，第二主成分为纵轴，绘制各地区的成分图，见下图 7。

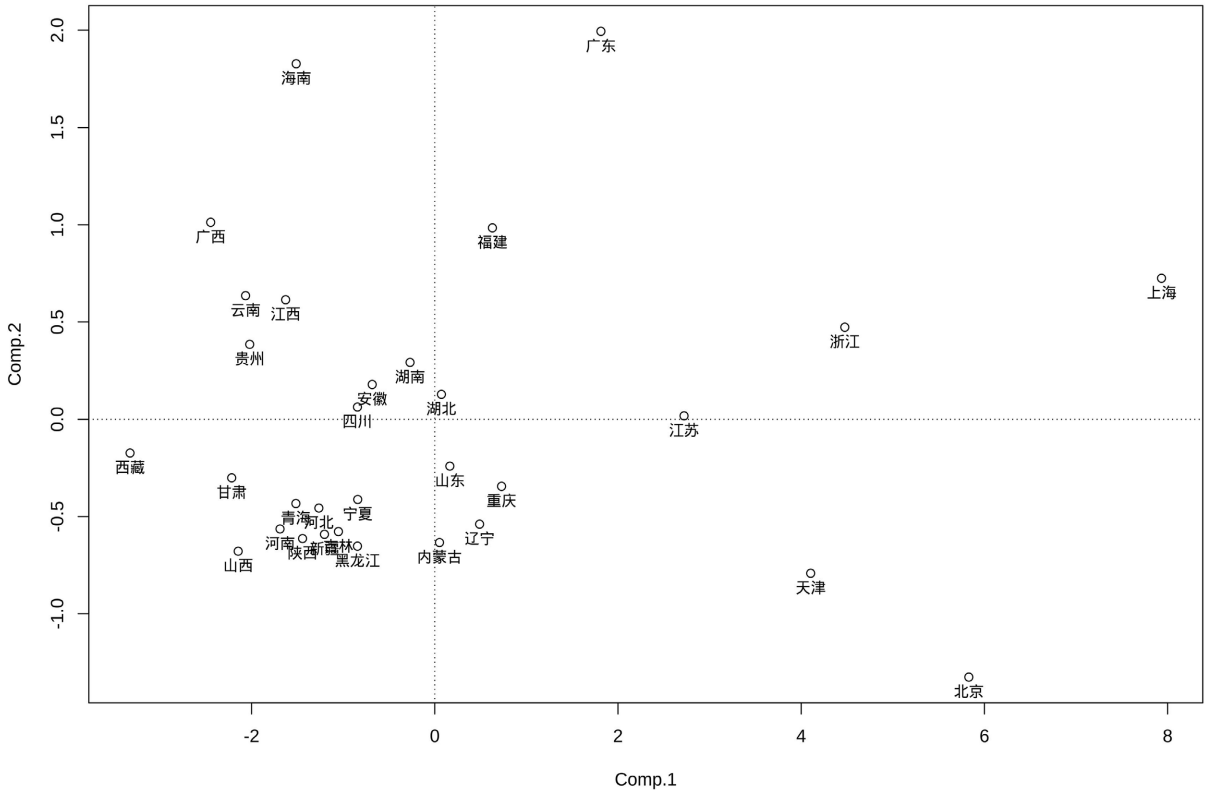

Figure 7. Principal component plot
图 7. 主成分图

在主成分 Comp.1 上得分最高的前五个地区依次是上海、北京、浙江、天津和江苏，广东排在第六，且上海、北京绝对值明显高于其他省、市、自治区，这就是说以居住、生活用品及服务、教育文化娱乐、交通通信和其他用品及服务等为主的消费主成分而言，上海、北京的消费水平远远高于其他省、市、自治区；而广西和西藏在这方面的消费相对较低些。北京在主成分 Comp.2 上的得分负值最高，反映的是北京在食品烟酒、衣着和医疗保健方面的人均消费水平远超全国，而广东在主成分 Comp.2 上的得分正值最高反映的是广东在这些领域的公共服务投入更充分，或者说广东的消费者更愿意在这些领域进行消费。这些数据和往年的情况有较大的不同，所谓的“北上广深”，在人均消费上广东虽说还是不错，但并没有比较大的优势。这些情况可能与此次疫情有较大关系。当然更可能的原因是在经济落后的地区，如西部和部分中部地区等地，居民消费水平较低，主要支出在衣食医疗保健等方面，与当地的经济水平有关。当地经济发展相对比较落后，消费支出水平和消费倾向自然有所不同。

4.2. 聚类分析

以上的几种图，只能大致的对样本间的相似性进行比较，没有办法做出明确直观的分类结果。接下来，笔者将通过系统聚类分析和 K-means 聚类图将这 31 个样本进行分类[11]。笔者将聚类分类的相关代码放在附录，并将所得分类结果图转化成了表 5。

Table 5. Organize the clustering chart results by class
表 5. 按类整理聚类图结果

分两类	第一类		第二类		
	北京；上海		天津；河北；山西；内蒙古；辽宁；吉林；黑龙江；江苏；浙江；安徽；福建；江西；山东；河南；湖北；湖南；广东；广西；海南；重庆；四川；贵州；云南；西藏；陕西；甘肃；青海；宁夏；新疆		
分三类	第一类		第二类	第三类	
	北京；上海		天津；内蒙古；辽宁；江苏；浙江；安徽；福建；山东；湖北；湖南；广东；重庆	河北；山西；吉林；黑龙江；江西；河南；广西；海南；四川；贵州；云南；西藏；陕西；甘肃；青海；宁夏；新疆	
分四类	第一类	第二类	第三类	第四类	
	北京；上海	天津；江苏；浙江	河北；山西；吉林；黑龙江；江西；河南；广西；海南；四川；贵州；云南；西藏；陕西；甘肃；青海；宁夏；新疆	内蒙古；辽宁；安徽；福建；山东；湖北；湖南；广东；重庆	
分五类	第一类	第二类	第三类	第四类	第五类
	北京	天津；江苏；浙江	河北；山西；吉林；黑龙江；江西；河南；广西；海南；四川；贵州；云南；西藏；陕西；甘肃；青海；宁夏；新疆	内蒙古；辽宁；安徽；福建；山东；湖北；湖南；广东；重庆	上海

从表 5 可以看出类别之间存在着明显的差异。综合考虑以上的分类结果，认为从我国各地区的消费情况来看，分为四类较为合适。下面是对分成四类后每一种归类的分析解释：

第一种归类是北京和上海。作为中国的一线城市，北京和上海拥有丰富的经济资源和高端产业，尽管受到疫情的影响，这两个城市的人均消费水平在全国仍保持最高级别。其可能原因包括拥有较为完善的经济体系和大量高收入人群，同时这两个城市的产业结构也比较多元化。除此之外，上海和北京拥有优质的教育和文化资源，包括高水平的大学和研究机构、博物馆、艺术馆等。这些资源吸引了众多学生和文化爱好者前往这些城市学习和体验，因此在教育和文化消费方面投入较多。上海和北京得到了国家和地方政府的政策支持和特殊投资吸引力。这些政策和投资促进了当地经济的增长和发展，进而带动了消费水平的提升。

第二种归类是天津、江苏和浙江。这些地区为中国东部发达地区，拥有较高的经济水平和人均消费水平。这些地区的经济发展水平较高，拥有较多的制造业和服务行业，同时人口密集，形成较好的消费市场。所以尽管受到疫情的冲击，人均消费水平仍相对较高，其经济体系比较完善、外向型产业比较发达。

第三种归类包括河北、山西、吉林等 17 个地区。这些地区大多为中国中西部地区，经济发展相对滞后，人均消费水平较为普通。这些地区的经济主要依靠农业、资源等传统产业，相对发展缓慢；同时，一些地区还面临自然资源和环境等方面的局限，所以总体水平低于其他地区。

第四种归类包括内蒙古、辽宁、广东、重庆等 9 个地区。这些地区的人均消费水平较高，尽管受到疫情的一定冲击，但由于有较为发达的制造业和服务业，以及较为丰富的经济发展资源，这些地区的人均消费水平仍然比较乐观。如广东的制造业和福建的对外贸易。一些地区则受益于重要的经济政策和利于商贸发展的地理位置，如重庆和湖北，但与第一、二类地区相比人均消费稍低一些。

从分类结果可以看出，四分类大体符合我国经济发展水平的客观情况，反映了我国地区经济发展不

平衡的现状，人均消费水平受到地区发展、城市吸引力等因素的影响。这个结果可以为政府和企业提供疫情后的经济发展改变，以便他们在政策、市场研究和业务决策等方面做出更为精准的判断和决策。

取 4 类的 K-means 聚类图，如下图 8 所示。

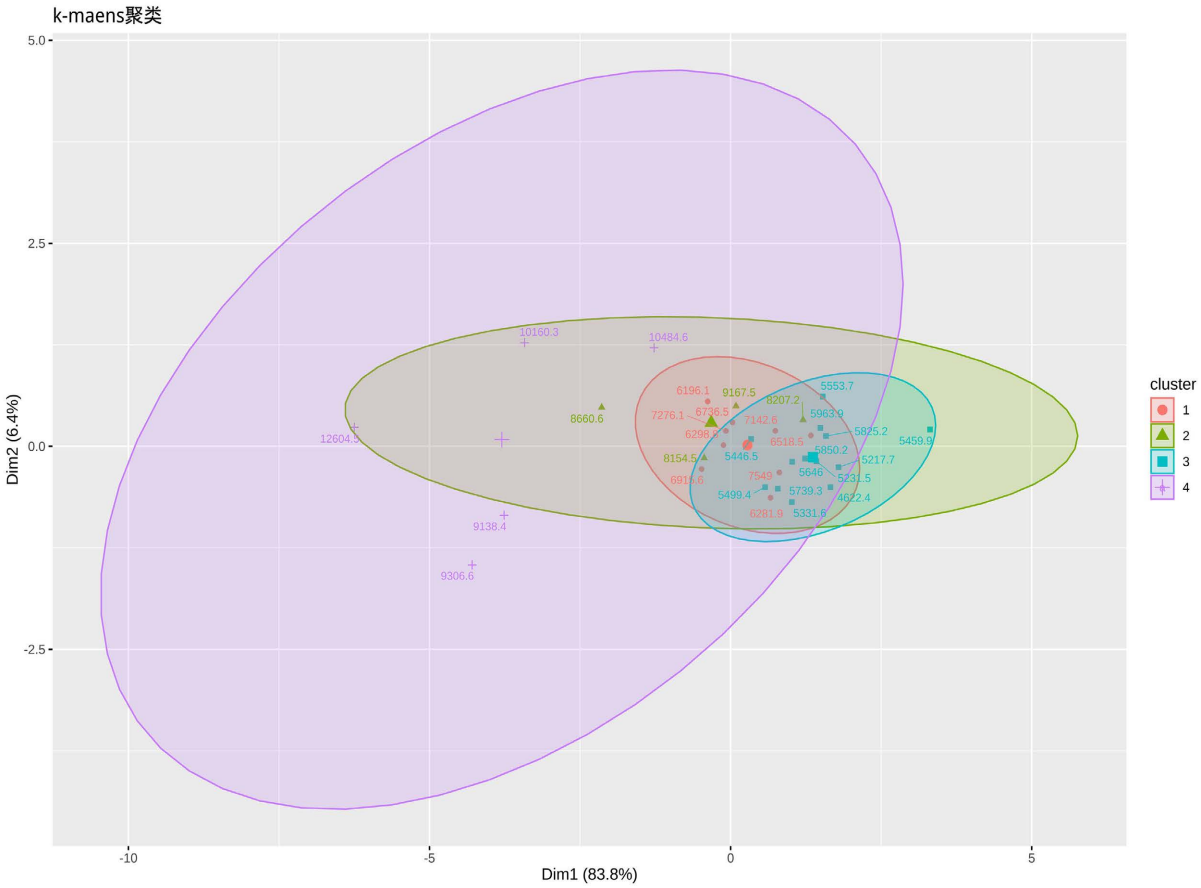

Figure 8. K-means clustering plot
图 8. K-means 聚类图

5. 比较主成分分析和聚类分析的优劣情况

通过上述分析结果，可以发现两种分析方法各有优劣，并且主成分分析和聚类分析在比较全国 31 个地区的人均消费水平时可以提供不同的信息和视角：

主成分分析可以确定主要影响人均消费水平差异的变量或因素，从而能够解释地区之间人均消费水平的不同。这有助于揭示不同地区消费模式和特征的差异性。通过主成分分析，可以发现哪些消费项目对聚合在一起解释了大部分的方差。但主成分分析可能忽略了地区之间真实的分类信息，因为它主要关注总体差异的捕捉，而不是区分不同的地区群组。

而聚类分析则可以将这 31 个地区根据人均消费水平划分为不同的群组或类别。这有助于识别哪些地区在人均消费水平上相对接近，以及哪些地区有显著的差异。这种方法能够从整体上比较这些地区的消费水平，并发现是否有类似地区聚集成为一组。但聚类分析可能会忽略不同地区之间的细微差异，因为它主要关注相似性。此外，聚类分析提供的结果通常是分类的，它对于确定哪些因素造成了分类结果的差异性并不够明确。

因此，笔者认为综合使用主成分分析和聚类分析可以提供一个更全面的观察，既可以通过聚类分析找到相似的地区群组，又可以通过主成分分析进一步揭示这些群组之间的差异性，进而帮助全面理解全国 31 个地区的人均消费水平的优劣情况。

6. 结论与建议

通过对 2021 年全国 31 个地区的人均消费支出数据进行分析，可以得知在不同地区的居民消费水平和消费结构是不同的，且能够有效地反映出地区居民的经济水平和消费倾向。得出结论如下：

1) 各地区消费支出存在明显的差异与规律，其中食品烟酒、居住和交通通信是绝大部分地区消费支出的主要构成部分；

2) 各地区在不同消费支出领域上的分布情况存在明显差异，表现出一定的地区特征；

3) 各地区在不同消费支出领域上的人均消费支出变化趋势存在一定程度的差异，不同地区的消费潜力具有较强的异质性。

基于此，提出以下建议：

1) 政府应重视对消费支出的数据分析，及时了解各地区消费趋势及其对于经济社会发展的影响；

2) 鼓励消费者加大对环保、健康、文化等方面的消费支出，推动各地区消费结构的升级；

3) 希望进一步激发中西部地区消费潜力，推动经济发展的全面均衡；

4) 综上所述，消费支出在中国不同地区的侧重点各不相同，这也是与中国的实际生活和地区情况相吻合的。不同地区的居民要学会树立理性的消费观，合理分配消费支出，使其多元化、合理化。

基金项目

国家自然科学基金资助项目(11971200)；广东省教育厅委托项目(0835-210Z33606691)。

参考文献

[1] 罗润东, 谢香杰, 杨鸣. 2021 年中国经济学研究热点分析[J]. 经济学动态, 2022(2): 105-123.

[2] 何远霞, 王兰, 焦登丹. 基于主成分-聚类分析法的 31 个省市经济发展水平的综合评价[J]. 中国管理信息化, 2023, 26(14): 177-179.

[3] 李晓红. 我国各地区居民消费支出情况及分析[J]. 现代营销(经营版), 2019(9): 36-37.

[4] 郑波. 中国旅游经济时空分异及影响因素研究[D]: [硕士学位论文]. 太原: 山西师范大学, 2019.

[5] 李煜涵. 基于主成分分析和聚类分析的河南省各城市经济综合实力评价[J]. 中小企业管理与科技, 2023(10): 149-151.

[6] 赵静. 基于主成分分析的不同地区居民消费水平与消费结构[J]. 科技经济市场, 2022(5): 68-70.

[7] 中华人民共和国统计局. 中国统计年鉴[M]. 北京: 中国统计出版社, 2022.

[8] Chang, W. (2018) R Graphics Cookbook: Practical Recipes for Visualizing Data. 2nd Edition. O'Reilly Media, Sebastopol, California.

[9] 李佳霖. 基于多元线性回归分析及主成分分析的我国居民消费水平建模[J]. 产业与科技论坛, 2017, 16(12): 79-81.

[10] 林碧茹. 聚类分析和主成分分析在消费数据中的应用[D]: [硕士学位论文]. 曲阜: 曲阜师范大学, 2016.

[11] 秦杨杨. 基于主成分分析和 K-means 聚类分析的地区经济发展研究[J]. 中国商论, 2020(4): 214-215. https://doi.org/10.19699/j.cnki.issn2096-0298.2020.04.214

附　录

Table S1. Eight per capita consumption expenditure data in all regions of the country in 2021
附表 1. 2021 年全国各地区的八项人均消费支出数据

地区	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
北京	9306.6	2104.4	16846.7	2559.7	4226.8	3348.0	4285.7	962.5
天津	9138.4	1872.0	7519.5	1940.6	4390.4	3372.5	3747.6	1207.5
河北	5646.0	1371.8	4520.9	1216.9	2755.1	2007.3	1983.9	451.8
山西	4622.4	1277.4	3850.8	1029.0	1988.0	2059.1	1935.2	429.3
内蒙古	6298.8	1641.0	4532.6	1214.7	3488.4	2543.7	2354.7	584.5
辽宁	6915.6	1627.9	4913.6	1307.5	3033.7	2809.4	2485.1	738.0
吉林	5499.4	1346.3	3707.0	1025.8	2655.8	2413.1	2360.7	596.4
黑龙江	6281.9	1466.3	3842.6	1040.7	2761.2	2254.1	2475.2	513.9
上海	12604.5	2086.9	16136.8	2248.1	5626.2	4709.9	3877.9	1589.1
江苏	8660.6	1783.9	8433.6	1911.7	4335.7	2984.7	2463.4	877.9
浙江	10160.3	2051.3	9943.0	2072.9	5196.5	3768.7	2498.9	976.6
安徽	7142.6	1430.8	4664.7	1343.0	2479.5	2584.8	1783.6	482.0
福建	9167.5	1431.9	8300.8	1472.5	3121.1	2572.2	1768.5	605.5
江西	6518.5	1079.7	4721.2	1141.9	2342.5	2381.8	1693.8	410.6
山东	6196.1	1530.3	4682.7	1716.4	3495.6	2728.6	2015.5	455.6
河南	5231.5	1405.2	4027.0	1228.9	2103.6	2209.2	1786.8	399.0
湖北	7276.1	1464.5	4991.8	1327.2	3186.4	2863.3	2238.7	498.0
湖南	6736.5	1329.3	4811.5	1411.2	2891.0	3061.3	2122.2	435.0
广东	10484.6	1278.0	8189.6	1614.1	4164.6	3241.6	1900.9	715.9
广西	5825.2	710.2	3697.6	1058.7	2473.1	2283.9	1752.8	286.3
海南	8207.2	745.8	5028.0	1033.9	2650.7	2444.5	1682.9	448.9
重庆	8154.5	1708.3	4490.3	1682.5	3049.8	2601.4	2325.8	585.2
四川	7549.0	1315.4	4035.5	1387.6	2807.4	1891.9	2071.9	459.3
贵州	5553.7	1162.2	3461.8	1097.7	2678.0	2247.7	1368.2	387.9
云南	5963.9	959.3	3954.0	1005.6	2837.7	2059.0	1700.1	371.3
西藏	5459.9	1294.4	3622.5	975.8	2104.9	768.0	781.4	335.4
陕西	5331.6	1264.6	4401.5	1267.1	2284.2	2110.7	2264.6	422.2
甘肃	5217.7	1217.3	3706.0	1068.0	2215.4	1893.8	1761.4	376.6
青海	5850.2	1358.7	3580.2	1119.0	3108.7	1627.5	1938.1	437.8
宁夏	5446.5	1370.1	3693.1	1203.1	3378.5	2273.2	2126.6	532.5
新疆	5739.3	1320.6	3598.3	1149.6	2707.7	1664.4	1990.7	789.9

R 程序代码

```
X=read.table('clipboard',header=T)#复制的就是附表 1 的数据
barplot(apply(X,1,mean),las=3)#按行作均值条形图

pie(apply(X,2,mean))#按列作均值饼图

boxplot(X,horizontal=T)#箱尾图中图形按水平放置

stars(X,full=T,draw.segments=T,key.loc=c(13,2))#具有图例的 360 度彩色星相图

mat=as.matrix(X)
heatmap(mat,scale="column",margins = c(4,3),cexRow=0.6,cexCol = 0.45)#热图

mat3=scale(mat)
#install.packages("pheatmap")
library(pheatmap)
pheatmap(mat3,color=colorRampPalette(c("navy","white","firebrick3"))(10),
         display_numbers=TRUE,
         cellheight_row=6,
         fontsize=7,
         treeheight_row=50,treeheight_col=35,
         cutree_col=2,
         cutree_row=3,
         cluster_col =FALSE)#数值热图

case=read.table("clipboard",header = T)#数据来源附表 1
z=scale(case)
hc=hclust(dist(z))#最长距离法（complete）：计算每个簇中亮点之间的最长距离
plot(hc);rect.hclust(hc,2);cutree(hc,2)#分 2 类
plot(hc);rect.hclust(hc,3);cutree(hc,3)#分 3 类
plot(hc);rect.hclust(hc,4);cutree(hc,4)#分 4 类
plot(hc);rect.hclust(hc,5);cutree(hc,5)#分 5 类

library(factoextra)
library(ggplot2)
df=X[,-c(2,3)]
mat2=as.matrix(df)
rownames(mat2)=df[,1]
km=kmeans(mat2,centers=4)
fviz_cluster(km,mat2[,-1],
```

```
                repel=TRUE,
                ellipse.type="norm",
                labelsize=8,
                pointsize=1.5,
                main="k-maens 聚类")

X=read.table('clipboard',header=T)
cor(X)#计算相关矩阵
PCA=princomp(X,cor=T)#主成分分析
PCA#特征根开根号结果
summary(PCA)
PCA$loadings#主成分载荷
PCA$scores[,1:2]#主成分得分
princomp.rank(PCA,m=2)#主成分排名
princomp.rank(PCA,m=2,plot=T)#主成分作图
```