Research on Per Capita Consumption Expenditure Based on Fuzzy C-Means Clustering and Factor Analysis

Xiaohua Xu[1], Xiaofei Hu[2]

[1]Network Information Center, Zhaotong University, Zhaotong, Yunnan, 657000, China

[2]School of Mathematics and Statistics, Zhaotong University, Zhaotong, Yunnan, 657000, China

Keywords: Fuzzy C-Means, Clustering, Factors, Per Capita Consumption Level

Abstract: With the continuous improvement of people's living standards, there are obvious differences in the per capita consumption level of each region. For this reason, 8 main factors affecting the per capita consumption expenditure of 31 provinces, cities, autonomous regions and municipalities in china are studied by clustering and factor analysis. The fuzzy c-means clustering algorithm is used to cluster the regions into four categories. Factor analysis yields two public factors, namely the long-term consumption factor and the short-term consumption factor. The cumulative contribution rate of these two factors is 94.2851%, and the sum of the scores of the two public factors of per capita consumption of water expenditure is ranked. The results show that there are obvious differences in per capita consumption expenditure in each region, which can directly reflect the region's comprehensive economic capacity and development level.

1. Introduction

People's Consumption is Constrained by Factors Such as Geography, Environment, and Climate, Which Has Led to a Certain Degree of Difference in Consumption Levels in Different Regions of China [1-2]. to Study the Current Situation of Per Capita Consumption Expenditure in Each Region and Evaluate the Consumption Level in This Region Has Important Reference Value for the Rapid Economic Development and Balanced Development of Each Region [3-4].

2. Fuzzy C-Means Clustering and Factor Analysis

In the Mid-1960s, l.a. Zadeh Put Forward the Fuzzy Set Theory to Deal with the Clustering Problem, Which is Called Fuzzy Clustering Analysis [5-8]. Fuzzy Clustering is to Divide n Given Sample (p Factors of Each Sample) into c Class $(2 \le c \le n)$, and Record $V = (v_1, v_2, \cdots, v_c)$ as c Clustering Centers, Where $v_i = (\upsilon_{i1}, \upsilon_{i2}, \cdots, \upsilon_{ip})$, $(i = 1,2,\cdots c)$. the Samples Are Classified into a Certain Category with a Certain Degree of Membership, U Represents the Degree of Membership Matrix $\left(U = (u_{ik})_{c\times n}, u_{ik}\right.$Represents the membership degree of class i of the k sample$\left.\right)$, , and the Sum of the Degree of Membership of Each Sample is 1. the Criterion of the Fuzzy C-Means Clustering Method is to Find U、V So That the Objective Function $J(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^{m} d_{ik}^{2}$ Obtains the Minimum Value $(dik = \|x - v_i\|)$.

Factor analysis is to use the idea of reduction to decompose each original variable into two parts, one is a small number of public factors, the other is a unique factor of this variable, which needs to be explained reasonably [8-10]. The key of factor model is factor load matrix and contribution rate.

3. Fuzzy C-Means Clustering Analysis

This paper studies the per capita consumption of food, tobacco and alcohol, clothing, housing, daily necessities and services, transportation and communication, education, culture and

entertainment, medical care, other supplies and services. The statistics of 31 provinces, cities, autonomous regions and municipalities directly under the central government (excluding Hong Kong, Macao and Taiwan) in 2017 published on the website of the National Bureau of statistics are selected as the basic data. Based on these data, cluster analysis was performed on each region using fuzzy C-means clustering method. The C-means fuzzy clustering algorithm is implemented in the MATLAB environment, and its main function is fcm, and its format is as follows;

$$[\text{center}, u, \text{obj\_fcn}] = \text{fcm}(\text{data}, \text{cluster\_n}, \text{options}) \quad (1)$$

Its data represents input parameters and is a clustered data set. Each row represents a sample and each column corresponds to a variable. The article's data is a 32×9 matrix. cluster_n represents the number of classifications. Article data will be classified into 4 categories, so n is equal to 4. options is a vector of 4 elements. The first element is the power exponent of membership in the objective function shown, and the power exponent of the article is set to 4. The maximum number of iterations of the second element is 200. The third element is the termination tolerance of the objective function, which is set e^{-6} in this paper. The fourth element is the iterative process, which is not shown in this paper and is set to 0. Its options function setting expression is $\text{options} = [4,200,1e-6,0]$. center is a matrix of N rows and P columns. The paper is a matrix of 4×8. U is the membership matrix shown, indicating that the sample belongs to the class I membership. After data reading, standardization, clustering and other steps for analysis, the iterative process has gone through 65 steps. The membership matrix U of 31 regions is shown in Table 1.

Table 1 Membership Matrix Of 31 Regions (Part)

Region	Category 1 membership	Category 2 membership	Category 3 membership	Category 4 membership
Beijing	0.6772	0.1002	0.1053	0.1173
Tianjin	0.2656	0.2144	0.2326	0.2874
Hebei	0.0780	0.3413	0.3780	0.2027
…	…	…	…	…
Qinghai	0.0681	0.2693	0.4689	0.1936
Ningxia	0.0762	0.2912	0.4294	0.2032
Xinjiang	0.1952	0.2669	0.2695	0.2684

It can be seen from the table that the four categories of affiliations in the first row of Beijing are 0.6772, 0.1002, 0.1053 and 0.1173 respectively. As the first category is larger than other categories, Beijing is classified as the first category, and the classification of other cities is similar. The classification results are shown in Table 2.

Table 2 Classification Results Of 31 Regions

Category	Region
First kind	Beijing, Shanghai
Second category	Shanxi, Anhui, Jiangxi, Henan, Guangxi, Hainan, Guizhou, Yunnan, Tibet, Gansu
Third category	Hebei, Jilin, Heilongjiang, Shandong, Hubei, Hunan, Sichuan, Shaanxi, Qinghai, Ningxia, Xinjiang
Fourth category	Tianjin, Inner Mongolia, Liaoning, Jiangsu, Zhejiang, Fujian, Guangdong, Chongqing

Shanghai and Beijing belong to the first category. They have obvious regional advantages. They have developed infrastructure, medical conditions, education level, urban intelligence and so on. The per capita consumption expenditure far exceeds the national per capita consumption expenditure. Most areas in the second category belong to tourism areas. In recent years, the GDP growth rate is higher than the national average, and the economic growth is faster. They are also the main areas for the country to tackle poverty, so the per capita consumption expenditure is slightly lower than the national average. In the third category, the GDP and economic growth rate of most regions have steadily increased in recent years. In a small number of regions, the total population is small and the territory is vast. Therefore, the per capita consumption level is basically the same as

that of the country, or slightly higher than the national average. Most of the fourth area is located near the coast, with major industrial belts in the country, convenient transportation, and many other factors. Per capita consumption is also higher than the country.

4. Factor Analysis

Factor analysis is implemented in the Matlab environment. According to the factor loading matrix, special variance, and contribution rate factors, in the eight common factor models, it is more appropriate to choose two common factors. The two factor loading matrices are shown in Table 3 below.

Table 3 Factor Load Matrix

Food and tobacco	Clothing	Live	Daily necessities and services	Traffic communication	Education, culture and entertainment	Medical care	Other supplies and services
0.9054	0.3383	0.8685	0.7780	0.0418	0.7689	0.4435	0.8380
0.1304	0.1816	0.3617	0.4775	0.1580	0.5250	0.8935	0.4968

From the perspective of load matrix, the load of the first public factor in food, tobacco, alcohol and housing is relatively large, which shows that this factor is the main component of per capita consumption and can be interpreted as a long-term consumption factor. The second factor, which can be explained as a short-term consumption factor, has a large load on health care and other goods and services. The contribution rate of the two factors to the data is 47.3237% and 46.9314% respectively. The scores and rankings of 2 public factors of 31 regions are shown in Table 4 below.

Table 4 Factor Score Ranking

Region	First factor	Second factor	Two factors and	Comprehensive ranking
Shanghai	3.2778	0.7825	4.0603	1
Beijing	1.9364	2.1009	4.0373	2
Tianjin	0.9799	1.4454	2.4253	3
Liaoning	-0.0451	1.1268	1.0817	4
Zhejiang	1.1151	-0.1249	0.9902	5
Jiangsu	1.1973	-0.5446	0.6527	6
Guangdong	1.912	-1.3249	0.5871	7
Hubei	-0.7279	1.0994	0.3715	8
Inner Mongolia	0.0172	0.3479	0.3651	9
Jilin	-1.0043	1.202	0.1977	10
Heilongjiang	-0.93	1.1079	0.1779	11
Shaanxi	-0.9786	0.9424	-0.0362	12
Chongqing	-0.0890	-0.0090	-0.098	13
Qinghai	-0.7534	0.6002	-0.1532	14
Shandong	-0.3577	0.1592	-0.1985	15
Hunan	-0.2487	-0.0224	-0.2711	16
Ningxia	-0.8285	0.5473	-0.2812	17
Fujian	1.1175	-1.4057	-0.2882	18
Xinjiang	-0.7678	0.3263	-0.4415	19
Hebei	-0.5993	0.0889	-0.5104	20
Sichuan	-0.2892	-0.2394	-0.5286	21
Shanxi	-0.8568	0.1428	-0.714	22
Anhui	-0.1072	-0.7261	-0.8333	23
Henan	-0.4981	-0.338	-0.8361	24
Hainan	-0.175	-0.7780	-0.953	25
Gansu	-0.7862	-0.1733	-0.9595	26
Guangxi	-0.5853	-0.6210	-1.2063	27
Yunnan	-0.8428	-0.3814	-1.2242	28
Jiangxi	0.1045	-1.3834	-1.2789	29
Guizhou	-0.2347	-1.2674	-1.5021	30
Tibet	0.0478	-2.6805	-2.6327	31

It can be seen from table 4 that the top two cities of per capita consumption are Shanghai and Beijing, followed by Tianjin, and the last three are Jiangxi, Guizhou and Tibet. This is the same as the result of fuzzy c-means clustering algorithm. But factor analysis is more intuitive and obvious.

5. Conclusion

According to the fuzzy C-means clustering and factor analysis method, the per capita consumption expenditure in each region is very different, and the composition of per capita consumption in each region is also different, which is related to many factors such as economic development and cultural level. Therefore, the state should create favorable conditions, narrow the gap in consumption levels, and promote the positive and healthy development of the economy.

References

[1] Yang Yi. Empirical Study on regional consumption differences in China [J]. Management Science (5): 62-70

[2] Zhang Tao. Study on classification of regional consumption level based on cluster analysis [J]. Economic Research Guide, 2017 (11): 100-102

[3] Ma Jianyue. Cluster analysis and factor analysis of consumption level of urban residents in China [J]. China business theory, 2018 (2): 74-75

[4] Ling biaocan, Wei Hongxia. Hierarchical clustering and factor analysis of consumption level differences among cities [J]. Journal of North China University of science and technology, 2017 (01): 116-122

[5] Zeng huanglin. Rough set theory and its application (4): rough set theory and fuzzy set [J]. Journal of Sichuan Institute of Technology (self SCIENCE EDITION), 9 (4): 20-26

[6] Wang Xiaochuan, Shi Feng, Yu Lei. Analysis of 43 cases of MATLAB neural network [M]. Beijing University of Aeronautics and Astronautics Press, 2019 (3): 324-377

[7] Xie Zhonghua. Matlab statistical case analysis of 43 cases [M]. Beijing University of Aeronautics and Astronautics Press, 2012 (7): 358-360

[8] Chen Xing, Lu Da Dao, Zhang Hua. Comprehensive measurement and dynamic factor analysis of China's urbanization level [J]. Journal of geography, 2009, 64 (4): 387-398

[9] Jiao Yuanmei, Xiao duning, Ma Mingguo. Analysis of spatial distribution characteristics and influencing factors of residential land in Oasis Landscape [J]. Journal of Ecology (10): 146-154

[10] Su Weihua. Research on the theory and method of multi index comprehensive evaluation [D]. Xiamen University, 2000