Application of fuzzy clustering analysis on the consumption level of residents by region in Hainan

Chen Yijuan

School of Mathematics and Statistics

Hainan Normal University

Haikou,China

E-mail:867146636@qq.com

Zhang Chengyi*

School of Mathematics and Statistics

Hainan Normal University

Haikou,China

E-mail:chengyizh@sina.com

He Lifang

School of Mathematics and Statistics

Hainan Normal University

Haikou,China

E-mail:634224845@qq.com

Abstract—The previous literatures generally study consumption level of the single urban or rural residents in Hainan.But there are few documents about the residents' consumption level in Hainan province. According to the the data in range of year 2003-2012,this paper analyzes and evaluates on consumption level of urban and rural residents in Hainan from the whole perspective. We compare and analyse the data of cities. The main conclusions are as follows: It unveils that the urban and rural resident consumption trend shows strong positive correlation to each respective local economic status. The education expense accounts for gradual increasing part of resident consumption cost. In addition, the population growth, the urbanization rate, and the income differentiation between urban and rural resident play a significant role in consumption behavior.The last, countermeasures and suggestions are offered for these phenomena and their reasons. And the result from this study provides references to the future urbanization planning.

Keywords-Hainan;Consumption Level;FCM Algorithm;

Fuzzy Cluster Analysis;Residents

I. INTRODUCTION

Study of the residents' consumption in terms of expense level and portion is expected to not only unveil the actual conditions of local people's livelihood, but also in turn to yield vital significance inputs to the strategic decisions in improving government investment, optimizing investment structure, and increasing domestic demand.The previous related literatures[1-7]

generally concluded the studies with respect to data from individual urban or rural in Hainan.But there are few documents about the residents' consumption level in Hainan.These researches select food, clothing, housing, supplies, medical, transportation, entertainment,

miscellaneous goods,and services as indicators in modeling the consumption structure, but few takes the social development degree into account.

We think that statistics analysis of overall spending behavior in both urban and rural eras would give better understanding of the consumption structure. We therefore employed weighted FCM algorithm to carry out the analysis in this perspective.

We categorized the factors that drive consumption behavior as essential factors, major factors and other factors. We choose per capita GDP as the essential factor. The major factors included per capita disposable income of urban, per capita net income of rural, and the income difference between urban and rural eras. Other factor covers the number of university students the education funds ,and urbanization rate and the natural population growth rate.

The weighted FCM algorithm was carried out to classify Hainan residents' consumption expenses. The classification results were subsequently discussed towards each driver. Recommendations were thereafter proposed to policy decision in future urbanization strategies.

II. THE IMPROVED ALGORITHM

A. brief introduction fuzzy c-means clustering analysis method

Bezdek has proved fuzzy c-means algorithm will be convergent from iterative calculation:

Step1: select class number $c(2 \leq c \leq n)$ and take an initial fuzzy classification matrix $R^{(0)} \leq M_{fc}$. Then the iteration step by step $l = 0,1,2,...$

Step2:To $R^{(l)}$ and calculate matrix clustering center $V^{(l)}=(V_1^{(l)},V_2^{(l)},...,V_c^{(l)})^T$ .In the formula

$$V_i^{(l)}=\sum_{k=1}^{n}(r_{ik}^{(l)})^q w_k u_k \Big/ \sum_{k=1}^{n}(r_{ik}^{(l)})^q .$$

Step3:Modify fuzzy classification matrix $R^{(l)}$,and take

$$r_{ik}^{(l)}=[\sum_{j=1}^{c}(\frac{\left\|w(u_k-V_i^{(l)})\right\|}{\left\|w(u_k-V_j^{(l)})\right\|})^{\frac{2}{q-1}}]^{-1}$$

$(k=1,2,...n;j=1,2,...c),$ (1)

Step4:Compare $R^{(l)}$ with $R^{(l+1)}$ .If the accuracy of fixed $\varepsilon > 0$ has $\max\{\left|r_{ik}^{(l+1)}-r_{ik}^{(l)}\right|\}\leq\varepsilon$,so $R^{(l)}$ and $V^{(l)}$ are requested, and stop. Otherwise, $l=l+1$,back to the step2and repeated.

B. *improvement of algorithm*

We consider the role of index weight to the level of consumption in order to more appropriate to the needs of the actual situation in application. Assume that index weight vector is $w=(w_1,w_2,...,w_m)$ .So general European power distance $\left\|w(u_k-V_i)\right\|=\{\sum_{j=1}^{m}[w_j(u_{kj}-v_{ij})]^2\}^{1/2}$ is available to say the difference between the sample clustering u_k and category V_i .So in (1) the distance will be the improved generalized. European power distance.

III. DATA ACQUISITION AND PROCESSING

A. *data acquisition and determination of the weights*

In order to ensure the scientific nature of the study, data of this article from statistical yearbook of Hainan province.According to the information entropy method to determine the factors impact on the consumption level of importance weight

$$E_j=-k\sum_{i=1}^{m}p_{ij}\ln p_{ij},\forall j \qquad (2)$$

$$d_j=1-E_j \qquad (3)$$

$$w_j=\frac{d_j}{\sum_{j=1}^{n}d_j},\forall j \qquad (4)$$

By type(2),(3)and(4)to determine the entropy of each attribute E_j ,the degree of difference d_j and standardized weights w_j .Calculated separately and be shown in the table below:(down to one over ten thousand and we denote by x_1 Per Capita GDP(yuan), x_2 Per Capita Disposable Income of Urban(yuan), x_3 Per Capita Net Income of Rural(yuan), x_4 Income Difference Between Urban And Rural (yuan), x_5 Number of University Students(person), x_6 Education(10000yuan), x_7 Urbaniz ation Rate(%),and x_8 Natural Growth Rate (%))

TABLEI. SPECIFIC DATA OF ENTROPY, DIFFERENCE DEGREE AND WEIGHT

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
E_j	0.9825	0.9887	0.9912	0.9896	0.9776	0.9787	0.9925	0.9938
d_j	0.0175	0.0113	0.0088	0.0104	0.0224	0.0213	0.0075	0.0062
w_j	0.1657	0.1073	0.0838	0.0987	0.2130	0.2018	0.0708	0.0589

Therefore, the weight of each index for: W=(0.1657, 0.1073, 0.0838, 0.0987, 0.2130, 0.2018, 0.0708, 0.0589)

B. *data normalization and determine the initial fuzzy classification matrix*

Different impact on the result of the classification because the characteristic parameters of dimension and order of magnitude is not necessarily the same.In order to eliminate the influence of different characteristic index order of magnitude.So use type(5) to normalize the data processing.

$$u'_{ij}=0.1+\frac{u_{ij}-m_j}{M_j-m_j}\times(0.9-0.1) \qquad (5)$$

$i=1,2,...,144; j=1,2,...,10$

At the same time the corresponding initial fuzzy classification matrix $R^{(0)}$ .And determine the classification number $c=4$,take $q=2$,precision $\varepsilon=0.001$.

C. *clustering results*

The results from iterative calculation are summarized as in the below TABLEII.

TABLEII. CLUSTER CENTERS OF VARIOUS TYPE

Cluster centers	Indicators							
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
V_1	0.3985	0.6116	0.6323	0.5488	0.3320	0.2974	0.6006	0.5488
V_2	0.6586	0.7911	0.8066	0.3530	0.4385	0.4258	0.6739	0.5811
V	0.2930	0.4412	0.4389	0.7170	0.2071	0.1973	0.7100	0.5998
V_4	0.2447	0.3361	0.4054	0.8748	0.1805	0.1681	0.4526	0.6836

We defined the classification rule as that an object u_k would be allocated to the closest vector clustering center. Therefore, the classification results are shown as in TABLEIII.

TABLEIII. OUTPUT OF CLUSTERING BY REGION IN HAINAN

Class	Region
1	Wenchang, Qionghai, Chengmai, Lingao, Danzhou, Dongfang
2	Haikou, Sanya
3	Wuzhishan, Tunchang, Qiongzhong, Baoting, Whitesands, Changjiang
4	Wanning, Ding'an, Ledong, Lingshui

The data and information from Hainan statistical yearbook indicate that the eight index of Haikou and Sanya are attributed with the values greater than the ones from Wenchang, Qionghai, Chengmai, Lingao, Danzhou and Dongfang, which are in turn greater than the rest eras. Wuzhishan, Tunchang, Qiongzhong, Baoting, White sands and Changjiang are ranked at the bottom. We brief it in the classification language that the class 2 is ranked as the highest level of consumption, the class 1 as higher, class 4 as lower, and class 3 seating itself at the lowest.

IV. ANALYSIS AND EVALUATION THE CLUSTERING RESULTS AND GIVING THE CORRESPONDING SUGGESTIONS

The above argument would be submitted to following phenomena.

First: The clusters are classified consumption level in each era., Data in TABLEIV. reveal that the economic development aspects in Hainan province four aspects.For instance,economic development in Haikou and Sanya scores the highest. Qionghai, Chengmai, Wenchang, Danzhou and Wanning makes the second group, Baisha, Qiongzhong, Ledong ,Lingao and Tunchang dominates the lowest. In addition, from the classification results and the data, we can know the level of consumption of class 2 is the highest , class1 is the second and class3 is the lowest.So the results of the classification and the economic development of cities and counties show strong relationship.

Second:Sanya, as the most developed city in term of tourism industry, exhibits higher economic level than other eras. As shown in TABLEV, Sanya's essential factors and main factor are comparable better, while the number of university student, the education investment, and urbanization rate in Sanya are comparable lower than in Haikou. The classification result categorizes the two cities in the same class that induces the education expenses contributing to a larger proportion of residents' consumption. This statement as well can be proved by the weight.

Third:A Data from Haikou and Danzhou (see TABLEV.) are similar, but the two cities are not in the same class. The reason is that the urbanization rate in Danzhou is low and the population growth rate, in the contory, is higher.Thus, the population growth and urbanization rate are also the influencing factors that determine residents' consumption level.

TABLEIV. COMPOSITE SCORES OF REGIONAL ECONOMIC DEVELOPMENT IN HAINAN

Region	Synthesis Scores	Ranking	Economic Benefits	Economic Scale	Economic Structure
Average of Eastern group	55.85		60.63	51.96	59.43
Qionghai	58.17	3	59.58	57.10	58.99
Wanning	53.52	7	61.68	46.82	59.88
Average of South group	54.75		61.53	48.87	61.01
Sanya	95.92	1	100.00	94.69	92.14
Lingshui	49.10	9	52.66	46.99	49.18
Baoting	42.29	11	47.97	34.08	58.25
Ledong	31.70	16	45.50	19.73	44.47
Average of Western group	44.85		50.30	42.95	40.63
Danzhou	54.04	6	54.32	56.15	46.63
Lingao	35.26	15	46.16	28.21	37.50
Changjiang	49.98	8	50.63	53.37	37.67
Dongfang	40.10	12	50.08	34.09	40.73
Average of Northern group	59.06		59.97	57.10	63.69
Haikou	85.85	2	67.99	91.95	100.00
Wenchang	55.00	5	59.80	52.49	54.04
Ding'an	40.00	13	51.59	27.79	57.77
Chengmai	55.37	4	60.51	56.19	42.93

Average of central group	34.56		42.16	23.74	55.40
Wuzhishan	43.63	10	43.24	33.43	77.65
Qiongzhong	29.85	17	37.82	18.61	51.37
Tunchang	35.86	14	47.49	23.42	54.33
Baisha	28.91	18	40.10	19.51	38.24

Source：Thematic analysis of Hainan province and Hainan statistical yearbook --2011.

TABLEV.   CONTRAST OF DATA(HAIKOU, SANYA, DANZHOU)

Region	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Haikou	35338	19730	7191	12539	29003	95291	60.08	1.22
Sanva	42236	20472	7582	12890	10385	46150	50.34	1.98
Danzhou	39716	16789	6781	10008	19501	40388	37.49	2.18

Source：Hainan statistical yearbook --2012,Hainan bureau of statistics,NBS survey office in Hainan

Fourth: The income difference between urban and rural eras takes the role as the contribution factor that influences classification result. The data from Hainan statistical yearbook shows that both Lingao and Changjiang are underdeveloped eras, while Changjiang plays higher among the two (see TABLEVI.). However, Lingao, is assigned to class 1 with wenchang, qionghai, ChengMai, danzhou and Orient together, and ChangJiang is in class 3. It is noticeable that urban-rural income difference in ChangJiang, seating the third among 18 eras, is wider than in Lingao. This scenario differentiates Changjiang from Lingao and its counter partner eras in the class 1, though the latter economic level is low. In summary to this argument, the level of consumption shows a strong positive correlation with the economic development; the mean time,the income gap between urban and rural areas is a significant influencing factor that corrects residents' consumption level.

TABLEVI.   CONTRAST OF DATA BETWEEN LINGAO DAN CHANGJIANG

Region	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Lingao	19612	15035	5542	9493	8924	29642	29.34	1.72
Changjiang	32576	18033	5684	12349	4769	18347	34.84	0.76

Source：Hainan statistical yearbook --2012,Hainan bureau of statistics,NBS survey office in Hainan

Fifth:The data in TABLEVII.show consumption level in Hainan central region, including Wuzhishan, Qiongzhong, Tunchang and Baisha, is the lowest.The values are small, and all of the four eras are assigned to class 3. It is justifiable that the central eras, though rich in ecological resources, is with closed economies dominated in long term that caused considerably underdeveloped, and the residents' education expenses are less than in other eras.

TABLEVII.   STATISTICS OF EIGHT INDICATORS BY REGION IN HAINAN

Region	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
North								
Haikou	35338	19730	7191	12539	29003	95291	60.08	1.22
Wenchang	26676	17985	7248	10737	10607	32974	20.48	1.59
Ding'an	16511	16430	5954	10476	6097	25758	28.88	0.79
Chengmai	31389	17829	7212	10617	10091	30470	33.91	1.13
South								
Sanya	42236	20472	7582	12890	10385	46150	50.34	1.98
Lingshui	22113	15003	5415	9588	6740	21279	26.22	1.95
Baoting	16240	15300	4482	10818	3102	13210	46.56	1.93
Ledong	15174	14257	5925	8332	9704	36134	27.83	1.03
East								
Qionghai	27615	17477	7220	10257	8961	26167	30.16	1.07
Wanning	21857	17470	6933	10537	10970	31888	30.32	0.22
West								
Ligao	19612	15035	5542	9493	8924	29642	29.34	1.72
Danzhou	39716	16789	6781	10008	19501	40388	37.49	2.18
Dongfang	23722	16987	6372	10615	8020	30391	25.1	1.39
Changjiang	32576	18033	5684	12349	4769	18347	34.84	0.76
Middle								

Wuzhishan	15257	13584	4780	8804	2017	6527	46.31	-1.37
Qiongzhong	14621	14593	4383	10210	4135	14147	47.21	1.12
Tunchang	15411	15123	5908	9215	5484	18413	34.58	1.54
Baisha	17716	15006	4738	10268	3623	13637	40.84	0.04

Source：Hainan statistical yearbook --2012,Hainan bureau of statistics,NBS survey office in Hainan

Based on the above arguments, we derive the following suggestions.

The first: More high quality education is required to satisfy the people; the mean time,promote education investment.These project is established in terms the three strategic sides.First,to consolidate the fundamental educations.Second,to expedite the construction of vocational training institutes.Finally to establish faculties and specialties per the local economic and industrial needs in the local colleges and universities.

The second:Narrow the income gap between urban and rural areas.First,constantly improve farmers rich peasants policy,and implement various subsidies.Second, we must vigorously develop modern agriculture.Third, we need to make full use of local advantages of resources, develop characteristic economy, especially support the agricultural product processing industry to expand farmers' employment,such as developing chicken of Wenchang, cattle of Hainan and sericulture of Qiongzhong.

The third:Control the growth of the population.In particular,although the economy of Danzhou,Sanya,Baoting and Lingshui is more developed, the residents' consumption level is not high because the population growth rate is high.Economy,culture,medical and health care are the main factors influencing population growth rate.Therefore,we should control the growth of population from the following three aspects.First,strengthen the propaganda of family planning.Second,to the personnel of super living, take economic punishment .Third, improve the medical and health level.

The fourth: narrow the gap between other regions and the central of Hainan.The central of Hainan is not only the national ecological protection area and ethnic minority areas,but also the important barrier of ecological environment in Hainan.It has rich natural tourism resources and cultural tourism resources.So if we want to raise the consumption level of residents in central,we can develop rich variety of natural ecological tourism resources and unique cultural ecological tourism resources from the view of sustainable development. Strengthen ecological function areas in central, give full play to the regional resource advantages,develop ecological tourism ,focus on construction tourism projects of Li nationality and Miao nationality cultural,and develop mountainous planting and breeding. At the same time, we will continue to improve the ecological compensation and to support the construction of ecological functional areas in central.Finally increase the investment in the rural vocational education and skills training of migrant workers to train more skilled migrant workers.

V. CONCLUSION

In this paper,we select eight indicators from factors influencing the level of consumption ,and use the weighted fuzzy c-means clustering algorithm in analyzing the residents' consumption in 18 eras in Hainan. By comparing the similarities, partial similarities, and differences in the statistic results from varies of eras, we identify each characteristics, differences and influencing factors. Consequently, we discuss these driving factors and propose corresponding suggestions to each phenomenon.

REFERENCES

[1] Gao Jing-chun. Discuss consumer demand and development trend of residents in Hainan[J].Journal of Hainan Open University,2005(2):36-38.(in Chinese)

[2] Zhang Yu-mei,Lu Jiang-yong,Guo Jian Jian-chun.An Empirical Analysis of Urban Residents'Consumption Structure in Hainan Province[J].Journal of Anhui BusinessCollege(socialsciences),2005,4(4):24-27,67.(in Chinese)

[3] Guo Jian-chun,Zhang Yu-mei,Lu Jiang-yong.The determination of demand elasticity and its use in the marketing strategy--analysis on consumption demand of urban residents in Hainan[A].Academy of management annual conference 2006, the Chinese market for the fourth time proceedings of the national member of congress[C],2006,6(4):1842-1850.(in Chinese)

[4] Zou Wen-tao,Sun Rong-jun,Cai Lei etc.Empirical analysis on consumption structure of rural residents in Hainan province[J].Inner Mongolia Agricultural Science and Technology,2008(5):19-20.(in Chinese)

[5] Wang Yan-ying,Miao Jun-hong.Comparative study on consumption structure of the rural residents in Hainan province[J].Journal of Hainan Normal University(Natural Science), 2011,29(4):319-323.(in Chinese)

[6] Miao Jun-hong.Study on rural residents' consumption structure based on factor analysis and cluster analysis in Hainan[J].Journal of Hainan Open University,2012(1):68-71.(in Chinese)

[7] Zhou Jun-mei.Analysis on development trend of urban residents'consumption structure--Hainan Province[J].

   a) Science,2013(3):178,208.(in Chinese)

[8] Liu De-fen,Ji Li-wen,An Wen-juan.Evaluation on comprehensive consumption level of the region using principal component analysis and cluster analysis--based on studies of the urban household consumer expenditures from China's 31 provinces and cities.Popular Business,2010(116):14.(in Chinese)

[9] Mei Cui,Shan Pei.The cluster analysis of regional income gap and its consumption impact[J].Journal of Nanjing Audit University,2005,2(3):7-10.(in Chinese)

[10] Chen Shui-li,Li Jing-gong,Wang Xiang-gong.Fuzzy set theory and its application(the first Edition)[M].Science Press, 2005:94-118.(in Chinese)