Financial Economics Research; Vol.3, No.2; 2026
E-ISSN: 3006-3663 P-ISSN: 3006-3655
DOI: https://doi.org/10.70267/fer.250302.137148
Published by: Zeus Press

An Analysis of Regional Differences in Consumption Expenditure Levels in China Based on Hierarchical Clustering

Yawen Shi*

School of Mathematics and Statistics, Beijing Technology and Business University, Beijing 100071, China

**Corresponding author: Yawen Shi.*

Abstract

There are significant regional disparities in the consumption expenditure levels of urban residents in China. Scientifically identifying the regional differentiation characteristics of consumption structure is of great significance for formulating differentiated regional development policies. Based on the data of per capita consumption expenditure of urban residents across eight major categories in 31 provinces, autonomous regions, and municipalities from the *China Statistical Yearbook 2024*, this study conducts an empirical analysis of regional differences in consumption expenditure levels in China using the systematic clustering method, specifically the within-cluster sum of squares method and squared Euclidean distance. First, descriptive statistics and boxplots are employed to identify the distribution characteristics of variables and detect outliers. Second, Pearson correlation coefficients are used to examine the relationships among variables. Subsequently, systematic clustering based on the within-cluster sum of squares method is performed, and the optimal number of clusters is determined with the aid of scree plots and dendrograms. Finally, one-way analysis of variance (ANOVA) is applied to test the statistical significance of the clustering results. The empirical results indicate that the 31 provinces in China can be classified into three consumption tiers. The first category is characterized as "high consumption–comprehensive development," including Beijing, Shanghai, and Zhejiang, where the consumption structure has shifted toward development- and enjoyment-oriented consumption. The second category is defined as "moderate consumption–structural optimization," covering Tianjin, Inner Mongolia, Jiangsu, Fujian, Guangdong, and Tibet, where the consumption structure is in a transitional stage of upgrading. The third category is described as "relatively low consumption–survival-oriented," comprising the remaining 22 provinces, where consumption remains concentrated on essential goods. Significant inter-group differences are observed across all eight categories of consumption expenditure among the three groups ($p < 0.01$), and the clustering results are highly consistent with China's regional economic development patterns. This study provides empirical evidence for understanding regional differences in consumption structure in China and for formulating differentiated consumption promotion policies.

Keywords

systematic clustering method, regional consumption disparity, consumption structure, urban residents, within-cluster sum of squares method

1. Introduction

1.1 Research Background

Consumption is the primary driver of economic growth and an important indicator for measuring high-quality regional economic development [1]. In recent years, the consumption levels of urban residents in China have significantly increased, with the consumption structure gradually shifting from survival-oriented to development- and enjoyment-oriented. However, influenced by multiple factors such as regional resource endowments, economic development levels, and institutional environment, there exist significant imbalances in the consumption expenditure levels of urban residents across different provinces in China [2]. Existing studies indicate a pronounced "polarization" in consumption tiers between the eastern and central-western regions, and the rapid increase in rigid expenditures such as housing and education has exerted a significant crowding-out effect on consumption in other regions, limiting the full potential of domestic demand [3, 4]. How to scientifically identify such regional differentiation characteristics has become a key issue to be addressed in formulating differentiated regional coordinated development strategies.

1.2 Research Questions

Although existing literature has paid attention to regional differences in consumption structure, most studies focus on single-variable analysis or a simple division into the eastern, central, and western regions, lacking fine-grained classification based on the latest multidimensional data. Specifically, current studies have not fully addressed the following questions: At the provincial level in China (31 provinces), based on per capita consumption expenditure data across eight major categories, what clustering patterns emerge among different regions? What are the essential statistical differences in consumption structure among these categories? These questions require rigorous empirical analysis to clarify and fill the gap in the precision of existing classification studies.

1.3 Research Objectives and Significance

This study aims to use the latest data from the *China Statistical Yearbook 2024* and apply the systematic clustering method to scientifically classify the consumption expenditure levels of urban residents in 31 provinces, thereby revealing the spatial differentiation patterns of regional consumption structure.

Theoretically, this research enriches the interdisciplinary perspective of regional economics and consumption economics and provides new empirical evidence for understanding the characteristic pattern of regional consumption gradients in China. Practically, the empirical results assist government authorities in accurately identifying the consumption development stages of different regions, enabling the formulation of more targeted consumption promotion policies. In particular, proposing differentiated strategies for "high consumption–comprehensive development" and "relatively low consumption–survival-oriented" regions has important practical significance for promoting a healthy development pattern of "high consumption leading, moderate consumption optimizing, and low consumption potential releasing" [5].

1.4 Research Methods

This study adopts quantitative empirical research methods. First, the per capita consumption expenditure of urban residents in eight major categories across 31 provinces, autonomous regions, and municipalities is selected as the research object. Second, descriptive statistics and boxplots are used to identify data distribution characteristics and detect outliers, and Pearson correlation coefficients are applied to examine the relationships among variables. In the core analysis stage, clustering is conducted using the systematic clustering method, specifically the within-cluster sum of squares method combined with squared Euclidean distance, and the optimal number of clusters is determined with the aid of scree plots and dendrograms. Finally, one-way analysis of variance (ANOVA) is employed to test the statistical significance of the clustering results, ensuring the scientific rigor and robustness of the classification.

1.5 Structure of the Paper

The paper is organized into five chapters. Chapter 1 is the introduction, outlining the research background, questions, objectives, methods, and framework. Chapter 2 describes the variable selection and data sources,

and conducts descriptive statistics and preliminary data processing. Chapter 3 presents the core empirical analysis, detailing the systematic clustering process, the optimal classification results, and the consumption structure characteristics of each category. Chapter 4 tests the statistical significance of the clustering results using one-way ANOVA and provides discussion. Chapter 5 concludes the study and offers policy recommendations, summarizing the main findings and proposing corresponding measures.

2. Variables and Data Sources

The data used in this study are sourced from the "Living Standards" section of the *China Statistical Yearbook 2024*, specifically the "Per Capita Consumption Expenditure of Urban Residents by Region." The dataset covers per capita expenditures of urban residents in eight major consumption categories across 31 provinces, autonomous regions, and municipalities (excluding Hong Kong, Macao, and Taiwan). The variables include: food, tobacco, and alcohol; clothing; housing; household goods and services; transportation and communication; education, culture, and recreation; healthcare; and other goods and services.

All data correspond to actual expenditures in 2023 (unit: RMB). The data have been standardized to eliminate dimensional effects and facilitate subsequent clustering analysis.

3. Descriptive Statistical Analysis of Variables

3.1 Means and Standard Deviations of Variables

Before performing clustering analysis, descriptive statistics were conducted on the eight consumption expenditure categories, including sample size, minimum, maximum, mean, and standard deviation, to preliminarily understand the central tendency, dispersion, and distribution characteristics of regional consumption expenditures. The results are shown in Table 1.

Table 1: Means and Standard Deviations of Variables (Unit: RMB)

	N	Min	Max	Mean	SD
Food, tobacco, alcohol	31	6216	13568	9129.52	1802.764
Clothing	31	1029	2893	1880.61	368.897
Housing	31	4361	20261	7364.42	3790.403
Household goods & services	31	1286	2848	1818	383.375
Transportation & communication	31	3099	7558	4288.48	1011.633
Education, culture & recreation	31	1628	5339	3359.74	805.001
Healthcare	31	1551	4733	2873.32	694.477
Other goods & services	31	473	1849	929.32	338.834

From the perspective of mean values, expenditure on food, tobacco, and alcohol ranks first at 9,129.52 RMB, consistent with Engel's law, indicating that food expenditure remains a core component of urban residents' consumption structure, reflecting that China's urban consumption is still dominated by basic living needs. Housing expenditure ranks second at 7,364.42 RMB, indicating that housing-related costs are another major pillar in urban residents' consumption structure, highlighting the significant role of housing costs in daily spending. Expenditures on transportation and communication, education, culture and recreation, and healthcare are 4,288.48 RMB, 3,359.74 RMB, and 2,873.32 RMB, respectively, demonstrating that residents' spending on travel, communication, human capital investment, and health protection has reached a certain scale, with development- and enjoyment-oriented consumption gradually emerging. In contrast, clothing, household goods and services, and other goods and services have relatively lower means of 1,880.61 RMB, 1,818.00 RMB, and 929.32 RMB, respectively, reflecting that these expenditures account for a relatively stable or still low proportion of total spending, with other goods and services exhibiting the highest elasticity, typically increasing significantly only after basic needs are met.

Regarding standard deviations reflecting dispersion, consumption variables exhibit significant regional differences. Housing expenditure has a standard deviation of 3,790.40 RMB, far exceeding other variables, with a maximum-to-minimum ratio of approximately 4.6, indicating substantial gaps in housing costs across regions. This phenomenon is closely related to differences in regional housing prices, rents, and property management fees and is a key factor causing imbalances in regional consumption expenditure levels.

Expenditure on food, tobacco, and alcohol also shows considerable variation (SD = 1,802.76 RMB), indicating that even basic living consumption differs significantly across regions, likely due to variations in prices, income levels, and dietary culture. Transportation and communication and education, culture, and recreation have moderate standard deviations (1,011.63 RMB and 805.00 RMB, respectively), indicating regional differences, with economically developed or geographically large regions investing more in transportation and communication and spending more on human capital and cultural consumption. Healthcare shows a relatively lower standard deviation (694.48 RMB), reflecting its status as a rigid expenditure, though it may still vary due to population structure and health insurance policies. Clothing, household goods and services, and other goods and services exhibit small standard deviations (368.90 RMB, 383.38 RMB, and 338.83 RMB, respectively), indicating strong convergence nationwide, high market liquidity, and limited influence from regional differences.

Overall, consumption variables show varying degrees of regional dispersion, with housing and food, tobacco, and alcohol exhibiting the most pronounced differences, fully reflecting the significant regional imbalance in urban residents' consumption structure. To further eliminate the effects of differences in units and magnitudes among variables and ensure the scientific and comparable nature of clustering results, the eight variables are standardized in subsequent analyses to construct a reasonable basis for regional classification.

3.2 Outlier Analysis of Variables

To examine the presence of outliers in the data and avoid their potential interference with subsequent clustering analysis, boxplots were used to identify outliers for the eight categories of consumption expenditure. Boxplots visually display the median, quartiles, extreme values, and possible outlier observations for each variable, aiding in assessing the distribution shape and the degree of deviation in the sample data.

Figures 1 to 8 present the boxplots for the eight categories of consumption expenditure. Overall, there are noticeable differences in the occurrence of outliers across variables among regions.

Figure 1: Boxplot of Food, Tobacco, and Alcohol

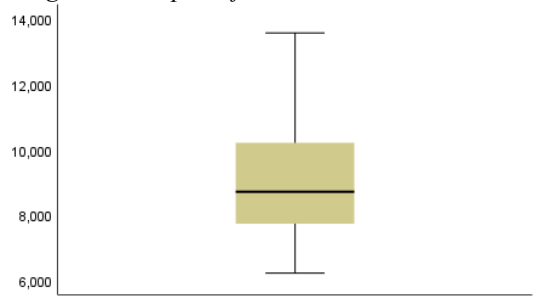


Figure 2: Boxplot of Clothing

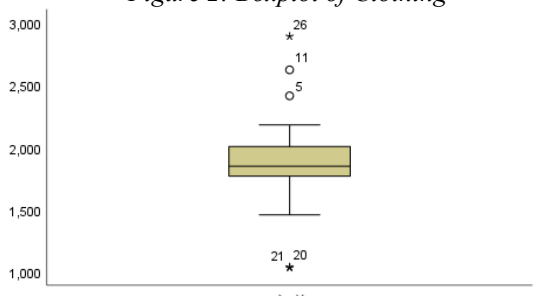


Figure 3: Boxplot of Housing

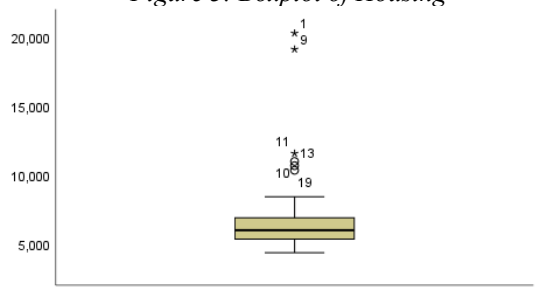


Figure 4: Boxplot of Household Goods and Services

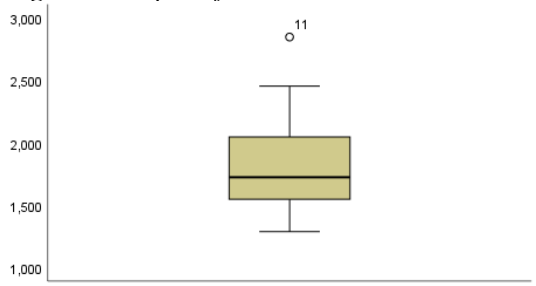


Figure 5: Boxplot of Transportation and Communication

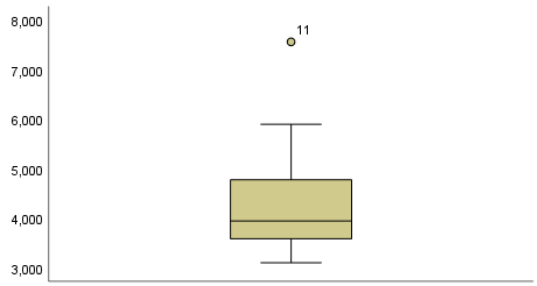


Figure 6: Boxplot of Education, Culture, and Recreation

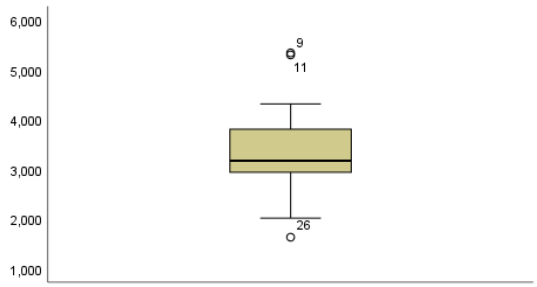


Figure 7: Boxplot of Healthcare

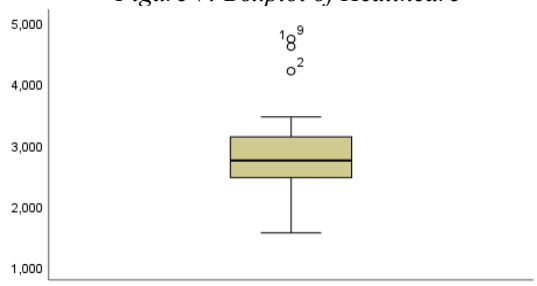
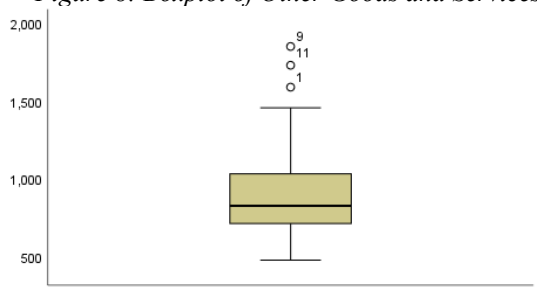


Figure 8: Boxplot of Other Goods and Services



Based on the boxplot identification results, high-value outliers are observed in variables such as housing, food, tobacco and alcohol, transportation and communication, and education, culture, and recreation, particularly concentrated in economically developed eastern regions such as Beijing, Shanghai, Zhejiang, and Guangdong. The objective presence of outliers indicates significant regional imbalance in consumption structures, which may affect the measurement of inter-sample distances in clustering analysis. Since the complete linkage method is sensitive to outliers, the presence of extreme values may cause clustering results to be overly influenced by these points, potentially distorting the formation and classification of clusters. Therefore, in the choice of clustering methods and interpretation of results, caution should be exercised when using complete linkage, and results should be compared with other clustering approaches to enhance robustness and reliability.

3.3 Correlation Analysis of Variables

To examine the relationships among the eight consumption expenditure variables and prevent potential bias in subsequent clustering analysis due to multicollinearity, Pearson correlation coefficients were calculated to assess pairwise linear correlations among variables. The results are presented in Table 2.

As shown in Table 2, the consumption expenditure variables exhibit varying degrees of positive correlation, with most correlations being statistically significant. Specifically, strong positive correlations exist between housing and other goods & services ($r = 0.795$), transportation & communication and other goods & services ($r = 0.789$), and healthcare and other goods & services ($r = 0.745$), indicating that regions with higher housing costs and stronger transportation and communication demands also tend to spend more on other high-end and quality-oriented services. Moreover, significant positive correlations are observed between food, tobacco, and alcohol and housing ($r = 0.734$), household goods & services and transportation & communication ($r = 0.754$), and education, culture & recreation and transportation & communication ($r = 0.722$), reflecting interlinkages among different expenditure categories, where economically developed regions generally show higher spending across multiple consumption dimensions.

Notably, some variable pairs show weak or insignificant correlations, such as clothing and food, tobacco, and alcohol ($r = 0.285$, $P > 0.05$), clothing and education, culture & recreation ($r = 0.140$, $P > 0.05$), and healthcare and food, tobacco, and alcohol ($r = 0.300$, $P > 0.05$), indicating that these categories vary independently across regions and are influenced by different economic and social factors.

Table 2: Correlation Matrix of Variables

		Food, Tobacco, Alcohol	Clothing	Housing	Household Goods & Services	Transportation & Communication	Education, Culture & Recreation	Healthcare	Other Goods & Services
Food, Tobacco, Alcohol	Pearson correlation coefficient	1	0.285	.734**	.700**	.675**	.653**	0.3	.684**
	Significance (two-tailed)		0.121	0	0	0	0	0.101	0
Clothing	Pearson correlation coefficient	0.285	1	0.275	.706**	.488**	0.14	0.211	.551**
	Significance (two-tailed)	0.121		0.134	0	0.005	0.454	0.254	0.001
Housing	Pearson correlation coefficient	.734**	0.275	1	.657**	.592**	.677**	.655**	.795**
	Significance (two-tailed)	0	0.134		0	0	0	0	0
Household Goods & Services	Pearson correlation coefficient	.700**	.706**	.657**	1	.754**	.636**	.368*	.704**
	Significance (two-tailed)	0	0	0		0	0	0.042	0
Transportation & Communication	Pearson correlation coefficient	.675**	.488**	.592**	.754**	1	.722**	.437*	.789**
	Significance (two-tailed)	0	0.005	0	0		0	0.014	0
Education, Culture & Recreation	Pearson correlation coefficient	.653**	0.14	.677**	.636**	.722**	1	.593**	.653**
	Significance (two-tailed)	0	0.454	0	0	0		0	0
Healthcare	Pearson correlation coefficient	0.3	0.211	.655**	.368*	.437*	.593**	1	.745**
	Significance (two-tailed)	0.101	0.254	0	0.042	0.014	0		0
		Food, Tobacco, Alcohol	Clothing	Housing	Household Goods & Services	Transportation & Communication	Education, Culture & Recreation	Healthcare	Other Goods & Services
Other Goods & Services	Pearson correlation coefficient	.684**	.551**	.795**	.704**	.789**	.653**	.745**	1
	Significance (two-tailed)	0	0.001	0	0	0	0	0	

***p < 0.01 (two-tailed), significant; *p < 0.05 (two-tailed), significant*

Overall, although certain correlations exist, multicollinearity is not severe, and the variable structure retains sufficient discriminative power, making it suitable for clustering analysis to explore regional consumption patterns. Since systematic clustering is based on inter-sample distances, high correlations among variables may imply overlapping information, but the method's validity does not depend on variable independence. To fully capture comprehensive differences across multiple consumption dimensions, squared Euclidean distance is adopted as the distance measure in subsequent clustering analysis. This distance metric amplifies differences, enhances cluster discrimination, and more clearly reveals the spatial differentiation of consumption expenditure levels across regions in China.

4. Analysis of Regional Differences in Consumption Expenditure Levels in China

4.1 Introduction to the Systematic Clustering Method

The systematic clustering method is one of the most widely used clustering techniques. Its fundamental concept is an "agglomerative" bottom-up clustering strategy. Specifically, the clustering process proceeds as follows: First, each sample is considered as an independent cluster, forming the initial classification. Second, the distances between clusters are calculated, and the two closest clusters are merged into a new cluster, with distances recalculated between the new cluster and the remaining clusters. This merging and distance updating process is repeated iteratively, reducing the number of clusters by one at each step until all samples are ultimately combined into a single overarching cluster. The clustering process can be intuitively visualized using a dendrogram, and researchers can determine an appropriate number of clusters based on the hierarchical structure of the dendrogram and complementary methods such as scree plots.

Considering the characteristics of single linkage, complete linkage, average linkage, centroid method, and the within-cluster sum of squares (Ward) method, as well as the distributional features of the data, the Ward method was selected to avoid the influence of outliers while balancing intra-cluster homogeneity and inter-cluster heterogeneity. Squared Euclidean distance was adopted as the distance measure, as it amplifies differences between samples and improves cluster discriminability. To determine the optimal number of clusters, a comprehensive assessment combining the hierarchical structure of the dendrogram and the elbow point in the scree plot was employed, allowing scientific identification of regional patterns in consumption expenditure levels and providing a reliable basis for subsequent regional classification and policy analysis.

4.2 Scree Plot for Clustering Analysis

Clustering analysis was conducted using the Ward method with squared Euclidean distance. The dendrogram and scree plot were jointly used to determine the optimal number of clusters, enabling scientific identification of differences in consumption expenditure patterns across regions in China.

The elbow point of the agglomeration coefficient indicates a relatively optimal clustering. Observation of the scree plot shows that when the number of clusters decreases from four to three, the agglomeration coefficient exhibits a significant increase, suggesting that dividing the 31 provinces into three clusters is reasonable. This classification scheme effectively distinguishes different consumption patterns while maintaining simplicity and interpretability.

4.3 Clustering Analysis Results

Based on the Ward method and squared Euclidean distance, a systematic clustering analysis was performed on urban residents' consumption expenditure data across 31 provinces. The samples were ultimately divided into three clusters, as shown in Table 3.

Figure 9: Dendrogram of Clustering

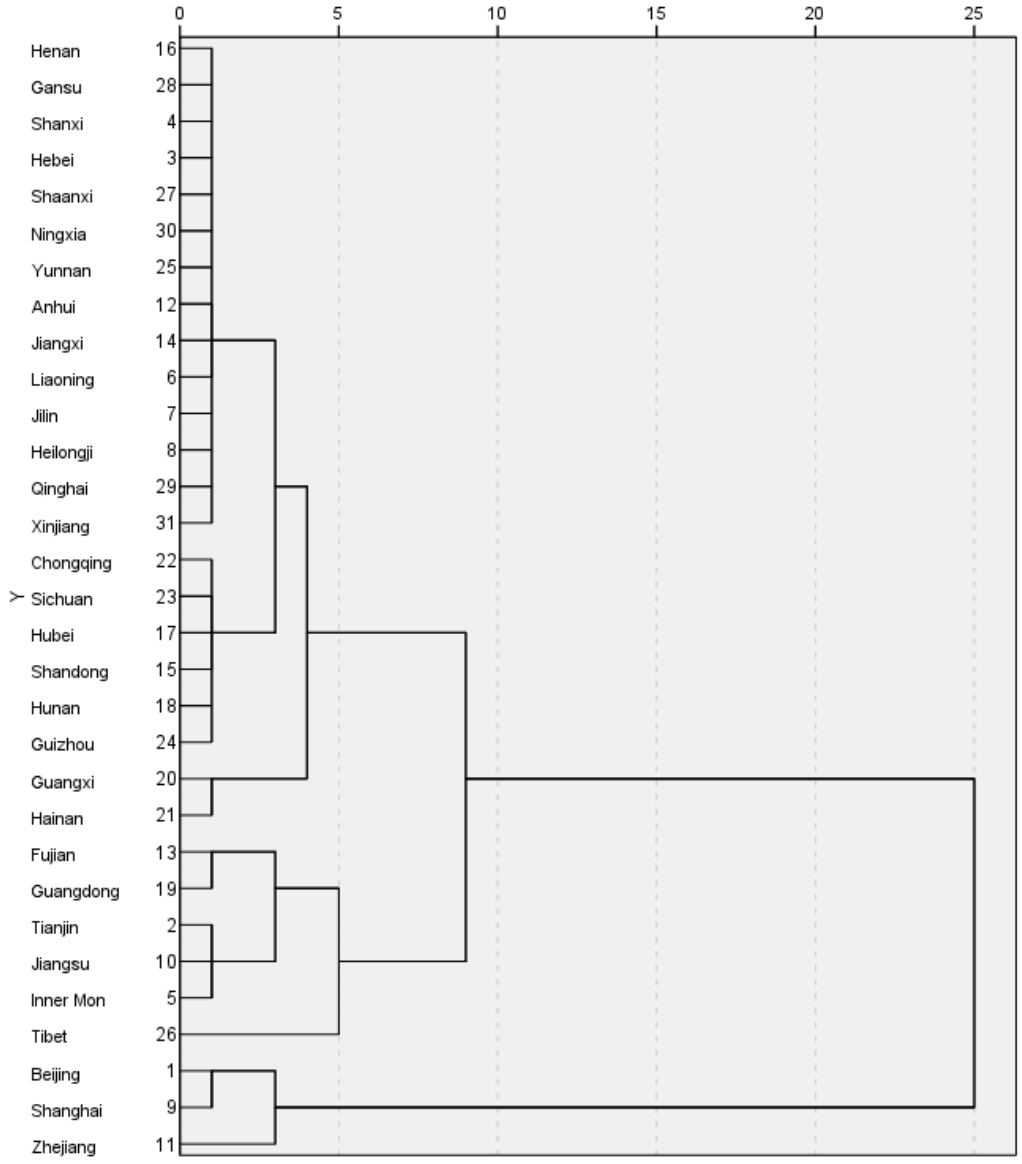


Figure 10: Scree Plot of Clustering Analysis

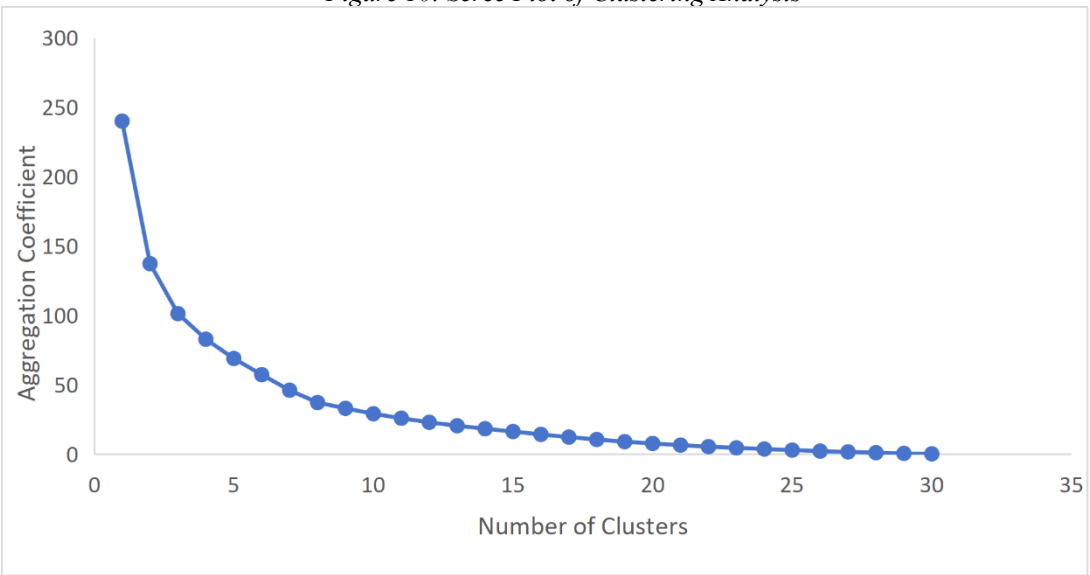


Table 3: Clustering Results

Cluster	Number of Provinces	Provinces
Cluster 1	3	Beijing, Shanghai, Zhejiang
Cluster 2	6	Tianjin, Inner Mongolia, Jiangsu, Fujian, Guangdong, Tibet
Cluster 3	22	Hebei, Shanxi, Liaoning, Jilin, Heilongjiang, Anhui, Jiangxi, Shandong, Henan, Hubei, Hunan, Guangxi, Hainan, Chongqing, Sichuan, Guizhou, Yunnan, Shaanxi, Gansu, Qinghai, Ningxia, Xinjiang

To further reveal the heterogeneity of consumption structures among clusters, Table 4 presents the within-cluster means and standard deviations for the eight consumption categories.

Table 4: Within-Cluster Means and Standard Deviations

Ward Method		Food, Tobacco, Alcohol	Clothing	Housing	Household Goods & Services	Transportation & Communication	Education, Culture & Recreation	Education, Culture & Recreation	Other Goods & Services
1	Mean	12353.67	2293.33	16970.33	2553	6175	4929.33	4196.67	1721.33
	Average	1567.597	290.107	4729.904	260.04	1266.432	674.368	823.906	131.127
2	Mean	10643.67	2136.67	8730.17	2086.83	5052.67	3469	2773.83	1133.5
	Average	1274.912	471.048	2298.228	152.089	780.176	953.398	918.727	242.05
3	Mean	8276.91	1754.05	5682.05	1644.45	3822.82	3115.91	2719.09	765.64
	Average	1058.368	271.542	675.442	261.997	511.753	498.376	381.887	137.236

From the group mean values reported in Table 4, it can be observed that the three clusters exhibit clear gradient differences in both consumption expenditure levels and consumption structure. A detailed analysis is as follows.

Cluster 1 represents high-consumption, comprehensive development regions. This cluster includes Beijing, Shanghai, and Zhejiang, which are among the most economically developed core regions in China. In terms of consumption levels, this group ranks at the top across all eight categories of consumption expenditure. In particular, expenditures on housing (16,970.33 yuan), transportation and communication (6,175.00 yuan), education, culture and entertainment (4,929.33 yuan), and healthcare (4,196.67 yuan) are especially prominent. These development- and enjoyment-oriented expenditures are significantly higher than those of the other two clusters. The consumption structure of urban residents in this cluster has largely moved beyond a survival-oriented pattern and has fully shifted toward development- and enjoyment-oriented consumption, representing the forefront of consumption upgrading in China. These regions typically feature higher income levels, well-developed infrastructure, and comprehensive public service systems, with consumption behavior reflecting typical characteristics of a post-industrial society.

Cluster 2 corresponds to moderate-consumption, structurally optimized regions. This cluster includes Tianjin, Inner Mongolia, Jiangsu, Fujian, Guangdong, and Tibet. In terms of composition, it encompasses both economically developed coastal provinces (Jiangsu, Fujian, Guangdong) and regions with distinctive policy or geographic conditions (Tianjin, Inner Mongolia, Tibet). Although there are differences in economic development levels, the consumption patterns are relatively similar. In terms of consumption levels, expenditures in this cluster fall within the national mid-range. Basic consumption categories such as food, tobacco and alcohol (10,643.67 yuan) and housing (8,730.17 yuan) are notably higher than those in Cluster 3, but there remains a substantial gap compared with Cluster 1 in development- and enjoyment-oriented consumption. It is noteworthy that the inclusion of Tibet suggests that consumption structure may be influenced by multiple factors, including policy support, demographic structure, and regional characteristics. Overall, Cluster 2 is at a critical stage of transformation and upgrading in consumption structure, exhibiting both features of advanced consumption patterns and constraints associated with regional development disparities.

Cluster 3 represents lower-consumption, survival-dominated regions. This cluster includes 22 provinces, mainly located in central, western, and northeastern China, where the level of economic development is relatively lagging and constitutes a major component of regional imbalance in China. In terms of consumption structure, this group remains at a generally low level across all eight categories of consumption expenditure. Consumption is still concentrated in essential categories such as food, tobacco and alcohol (8,276.91 yuan),

while the proportion of development- and enjoyment-oriented consumption is relatively limited. The consumption pattern in this cluster exhibits typical survival-oriented characteristics, with disposable income constraints limiting consumption upgrading. At the same time, this cluster covers a wide range of regions with considerable internal heterogeneity, and the consumption potential of some provinces has not yet been fully realized. From the perspective of expanding domestic demand and promoting coordinated regional development, this cluster represents an important potential space for future consumption growth and a key target for policy intervention and structural adjustment.

Overall, the classification of the three clusters reveals a clear gradient progression in both consumption levels and consumption structure, which fully validates the effectiveness of the clustering results. Cluster 1 shows a highly diversified consumption structure, Cluster 2 exhibits transitional characteristics of structural optimization, and Cluster 3 remains dominated by survival-oriented consumption. The differences among the three clusters not only reflect the imbalance in economic development levels but also reveal the spatial differentiation of consumption structure among urban residents in China, providing empirical evidence for the formulation of differentiated regional consumption policies.

4.4 One-way Analysis of Variance

To further verify the rationality and validity of the clustering results, a one-way analysis of variance (ANOVA) is employed to test the significance of differences among the three clusters across the eight categories of consumption expenditure. ANOVA determines whether statistically significant differences exist among groups by comparing the relative magnitude of between-group mean squares and within-group mean squares. If the between-group differences are significantly larger than the within-group differences, the clustering results can be considered to have strong discriminative validity. The results are presented in Table 5.

Table 5: One-way ANOVA Results

		Sum of Squares	df	Mean Square	F	Significance
Food, Tobacco, Alcohol	Between groups	60934003.92	2	30467001.96	23.331	0
	Within groups	36564741.82	28	1305883.636		
	Total	97498745.74	30			
Clothing	Between groups	1256817.433	2	628408.716	6.226	0.006
	Within groups	2826196.955	28	100935.606		
	Total	4083014.387	30			
Housing	Between groups	350280747.1	2	175140373.5	60.742	0
	Within groups	80733908.46	28	2883353.873		
	Total	431014655.5	30			
Household Goods & Services	Between groups	2716899.712	2	1358449.856	22.475	0
	Within groups	1692384.288	28	60442.296		
	Total	4409284	30			
Transportation & Communication	Between groups	18951261.14	2	9475630.568	22.579	0
	Within groups	11750776.61	28	419670.593		
	Total	30702037.74	30			
Education, Culture & Recreation	Between groups	8770473.451	2	4385236.725	11.507	0
	Within groups	10670340.49	28	381083.589		
	Total	19440813.94	30			
Education, Culture & Recreation	Between groups	5836417.456	2	2918208.728	9.457	0.001
	Within groups	8640529.318	28	308590.333		
	Total	14476946.77	30			
Other Goods & Services	Between groups	2721423.517	2	1360711.758	52.709	0
	Within groups	722839.258	28	25815.688		
	Total	3444262.774	30			

From the results in Table 5, it can be seen that at the significance level of $\alpha = 0.01$, the F-statistics for all eight categories of consumption expenditure pass the significance test, with corresponding p-values less than 0.01. Among them, variables such as housing and other goods and services exhibit relatively high F-values, indicating particularly strong between-group differences. Although the F-value for clothing is relatively smaller, its P-value is 0.006, which still indicates a statistically significant difference at the 0.01 level.

These results indicate that the three clusters derived from hierarchical clustering exhibit significant differences in group means across all eight consumption expenditure categories. In other words, the consumption structures of different clusters are not only characterized by overall gradient differences but are also statistically distinguishable across specific consumption categories. This provides statistical evidence supporting the rationality and validity of the clustering results and demonstrates that the classification effectively captures the true structure of regional differences in consumption expenditure levels in China, thereby providing a reliable basis for subsequent regional comparisons and policy analysis.

5. Conclusion

Based on data from the *China Statistical Yearbook 2024*, covering per capita consumption expenditures of urban residents across eight categories in 31 provinces, autonomous regions, and municipalities, this study applies hierarchical clustering using Ward's method and squared Euclidean distance to classify regional differences in consumption expenditure levels in China. Through descriptive statistical analysis, boxplot-based outlier detection, correlation analysis, hierarchical clustering, and one-way ANOVA, the following main conclusions are obtained.

First, there are significant regional disparities in the consumption structure of urban residents in China, which can be clearly classified into three categories with different consumption levels. The clustering analysis divides the 31 provinces into three groups: Cluster 1 includes Beijing, Shanghai, and Zhejiang, representing high-consumption, comprehensive development regions; Cluster 2 includes Tianjin, Inner Mongolia, Jiangsu, Fujian, Guangdong, and Tibet, representing moderate-consumption, structurally optimized regions; and Cluster 3 includes the remaining 22 provinces, representing lower-consumption, survival-dominated regions. This classification is validated by both the hierarchical structure of the dendrogram and the elbow point observed in the scree plot, indicating strong internal stability of the clustering scheme.

Second, the differences among the three clusters are reflected not only in the gradient distribution of total consumption expenditure but also more fundamentally in the internal composition of consumption structure. Cluster 1 exhibits significantly higher expenditures in development- and enjoyment-oriented categories such as housing, transportation and communication, education, culture and entertainment, and healthcare, with a consumption structure characteristic of a post-industrial society, where survival-oriented consumption declines and service- and development-oriented consumption becomes dominant. Cluster 2 is in a transitional stage of structural upgrading, with relatively balanced expenditures, reflecting both a trend toward consumption upgrading and a noticeable gap compared with Cluster 1 in higher-level consumption. Cluster 3 remains concentrated on essential consumption such as food, tobacco and alcohol, with a consumption pattern focused on meeting basic needs, a relatively low share of development- and enjoyment-oriented consumption, and underdeveloped consumption potential. The "high–medium–low" gradient pattern across the three clusters systematically reflects the regional differentiation of consumption structure in China.

Third, the clustering results demonstrate strong statistical significance and practical interpretability and are highly consistent with China's regional economic development pattern. The results of one-way ANOVA show that the mean differences in all eight consumption expenditure categories among the three clusters are statistically significant ($p < 0.01$), indicating strong discriminative validity of the classification. In terms of geographical distribution, Clusters 1 and 2 are mainly concentrated in the eastern coastal regions and economically developed or policy-advantaged areas, while Cluster 3 is widely distributed in central, western, and northeastern regions. Consumption levels exhibit a gradient decline from the eastern coastal areas to the inland regions. This distribution is highly consistent with the overall pattern of regional economic development in China and reveals a significant positive relationship between urban residents' consumption levels and regional economic development.

In summary, the empirical analysis of regional differences in consumption expenditure levels based on hierarchical clustering is methodologically rigorous and provides strong practical interpretability. The findings offer a reference for identifying regional consumption structure characteristics, formulating differentiated consumption promotion policies, and promoting coordinated regional development. Future research may further incorporate panel data to examine the dynamic evolution of consumption structure and introduce additional socioeconomic variables to explore the mechanisms underlying regional consumption disparities and potential policy interventions.

References

[1] Shi, M., Jiang, Z., & Zhou, X. (2019). Consumption upgrading or consumption downgrading? *China Industrial Economics*, (7), 42–60. https://doi.org/10.19581/j.cnki.ciejournal.2019.07.003

[2] Fu, H. (2020). Regional differences in residents' consumption structure in China: An empirical analysis based on the ELES model. *Consumer Economics*, 36(4), 77–84.

[3] Tian, Q., Ma, J., & Gao, T. (2008). Regional differences in factors affecting urban residents' consumption in China. *Management World*, (7), 27–33. https://doi.org/10.19744/j.cnki.11-1235/f.2008.07.004

[4] Tang, Q., Xia, Q., & Li, S. (2018). Consumption structure of urban households in China: 1995–2013. *Economic Research Journal*, 53(2), 35–49.

[5] Zang, X., & Sun, W. (2003). Consumption structure of urban and rural residents: A comparative analysis based on the ELES and AIDS models. *Journal of Shandong University (Philosophy and Social Sciences)*, (6), 122–126.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Copyrights