The spatial differences of consumer expenditure patterns in London

Student Number: 20022308

Word count:1737

Data and Code used in this paper can be found here:
https://github.com/akiakutaji/QM/tree/main/Assessment%204/data%20and%20code

1 Introduction

The study of resident behavior is crucial to social welfare policy (Gangopadhyay and Wadhwa, 2004). At the same time, resident expenditure is the fundamental driving force for the economic growth of a country or region (Zhen, 2008). Therefore, the study of resident expenditure behavior patterns also plays a vital role in regional development. Due to differences in natural endowment, geographic location, industrial layout, and population distribution in different places, different areas have different spatial expenditure structure. Therefore, studying the spatial differences in consumer expenditure patterns between different boroughs in London can help the local government to make a scientific and accurate evaluation of the expenditure pattern in London, so as to make different policy formulations according to the different expenditure behavior characteristics.

2 Literature Review

In the past few decades, the convergence or differentiation of consumer behavior has aroused widespread concern among researchers. Some scholars believe that although the development of technology and income status have become similar, the differences in consumer behavior have become stronger (De Mooij and Hofstede, 2002). There are many cases of consumer behavior analysis from a macro scale, including the analysis of different consumer expenditure patterns in various EU countries (Nowak and Kochkova, 2011) and a comparison of expenditure patterns in 13 countries(Goldberger and Gamaletsos, 1970). However, there are few studies on the micro-scale. Based on this, this paper chooses factor analysis and cluster analysis to study the consumer expenditure patterns among different boroughs in London, thus to provide a scientific basis for the management of local governments.

3 Data

The differentiation of consumer behavior is reflected in the consumption and use of products and services (Mooij, 2003). Therefore, this paper uses consumer expenditure data in London from the London Data Store [1] for analyzing. The "Aggregated Borough Base" file was chosen and downloaded, which provides consumer expenditure data from 2000 to 2010 and that from 2011 to 2036 (estimated), which are broken down by London borough. The Aggregated category contains spending data on 12 sectors. This paper chooses the data in 2010 to analysis (seen in Table 1).

According to Table 1, in 2010, the total expenditure in Wandsworth was the largest among all boroughs, reaching 8331.94£mn, which is more than thirty times

[1] Available at:
https://data.london.gov.uk/dataset/london-consumer-expenditure-estimates-2011-2036

Expenditure types	City of London	Camden	Westminster	Wandsworth
Accomodation Services	4.67	128.64	138.96	145.26
Convenience	22.36	672.10	673.81	793.84
DIY	1.61	53.80	53.10	62.47
Gardening	0.48	13.33	14.71	15.11
Leisure	6.87	195.55	202.57	223.13
On Licence (i.e. Pubs and Wine Bars)	6.05	166.47	179.83	187.98
Other Goods and Services	19.86	563.10	611.33	648.43
Other Spending (Mostly Household related, Health and Education)	139.36	3794.25	3926.44	4364.45
Restaurants and Caffees	9.62	264.84	286.09	299.06
Takeaway / Snack Spending	4.40	121.07	130.79	136.71
Comparison - Bulky	16.14	500.80	588.73	533.41
Comparison - Not Bulky	28.52	783.68	836.09	922.11
Total expenditure	259.92	7257.64	7642.44	8331.94

Table 1: Sample data of consumer expenditure data in London in 2010

more than the minimum 259.92£mn. In addition, in most boroughs, spending on health and education accounts for almost half of the total expenditure.

4 Methodology

4.1 Factor analysis

To avoid repetitive information provided by these 12 indicators, it is necessary to reduce the dimensionality of data, and the factor analysis method is a practical multivariate statistical method for this. This method is to recombine the original numerous indicators into a new set of uncorrelated comprehensive indicators (Rummel, 1988). Compared with the principal component analysis method which is just to simplify the data, factor analysis is more practical for the factor statement. There are basically two types of factor analysis: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). The former mainly attempts to discover the properties of structures that affect a series of reactions, while the latter is used to verify whether a set of specific structures affect the reactions in a predictive manner (DeCoster, 1998). Since this study is to avoid the correlation between these variables to obtain new indicators, the EFA method is used. The main steps of EFA are: collecting the measured values, obtaining the correlation coefficient matrix, deciding the number of factors, obtaining the initial factor set, rotating factor set, and explaining factor structure.

4.2 Cluster analysis

In order to identify the spatial differences in the expenditure structure of London residents, the paper will classify and analyze the boroughs of different expenditure patterns. Considering cluster analysis is the most suitable statistical method to analyze data grouping in heterogeneous samples (Duda et al., 2006), this paper uses cluster analysis to group boroughs based on similar consumption behaviors. The specific steps of the clustering algorithm are: (1) Measure the distance between all objects; (2) Classify objects according to these distances. The cluster analysis method can be divided into hierarchical clustering and non-hierarchical clustering according to whether the clustering process produces a clustered subset (Pieri et al., 2015). The former can be divided into the agglomerative cluster analysis (a "bottom-up" approach) and the divisive cluster analysis (a "top-down" approach) (Rokach and Maimon, 2005). The dendrogram generated by the agglomerative hierarchical cluster analysis method can show many neighbor relationships and classification relationships in the data, which is a convenient representation (Murtagh and Contreras, 2012). Therefore, this paper chooses to use the agglomerative cluster analysis method.

5 Results and Discussions

5.1 The result of factor analysis

5.1.1 The correlation matrix

Considering that the total expenditure of each district varies greatly, before analyzing, the data is standardized in to the same scale. Then the correlation coefficient is calculated (Figure 1). According to the figure, Accommodation services have a very strong positive correlation with three variables (Licence, Restaurants Cafes and Takeaway Snack), and have a strong negative correlation with other three variables (Convenience, Leisure and DIY). But it seems to have no relationship to Gardening. The reason for this may be that the higher the cost the residents spend on housing means the basic living price is high, which leads to the more money they have to spend on food. The fact that the cost of living and money spending on gardening are not related means that the two are independent of each other. Whether it is high or low expenditure in housing, it does not affect its expenditure in Gardening.

5.1.2 The number of factors

Before determining the factors, first it is necessary to determine the number of factors. It is best to plot the eigenvalues of the correlation matrix in descending order, and then use the number of eigenvalues that appeared before the last significant drop (Cattell, 1966). After the eigenvalues of the correlation matrix are arranged in

Figure 1: The correlation matrix

descending order, the first four eigenvalues are 0.853, 0.130, 0.012, 0.004, which drop by 84.8%, 90.8%, and 66.7% respectively. Therefore, it is decided that the number of final factors is 2.

5.1.3 Rotated factor matrix

After determining the number of factors, the factors can be extracted. After extracting and rotating the factors, the final rotated matrix as shown in Table 2. The factor F_1 has the greatest positive correlation with Restaurants and Caffees, and it also has a certain positive correlation with Takeaway/Snack and Accommodation Services. Therefore, this factor mainly reflects the expenditure level of London residents on basic living cost such as food, accommodation, health and education. The factor F_2 is the opposite of F_1, it is negatively correlated with the above variables, and mainly reflects other aspects of expenditure such as leisure and entertainment. Therefore, F_2 can be considered as an entertainment expenditure factor.

5.1.4 Factor score

Factors F_1 and F_2 have different explanatory for the expenditure status of London residents. Table 3 shows the sample of factor scores and score ranking[2]. The higher the score, the higher the expenditure level of this borough's resident in the aspect reflected by the factor. From the perspective of basic living expenditure factor F_1, Kensington and Chelsea have the highest scores, indicating that residents in theses

[2]The full table can be accessed here

Table 2: The rotated factor matrix

	F_1	F_2
Other Spending (Mostly Household related, Health and Education)	0.000491	-0.02446
Restaurants and Caffees	0.001638	-0.00097
On Licence (i.e. Pubs and Wine Bars)	0.00103	-0.00061
Accomodation Services	0.000796	-0.00047
Takeaway / Snack Spending	0.000749	-0.00044
Gardening	0.000107	0.000172
DIY	-0.00065	0.000583
Leisure	-0.00079	0.00144
Comparison - Bulky	0.004548	0.00347
Comparison - Not Bulky	-0.00014	0.004513
Convenience	-0.00947	0.00601
Other Goods and Services	0.001693	0.010773

area have the highest expenditure levels in food and accommodation [3], followed by Richmond upon Thames and Westminster; while Newham, Hackney and Lewisham are the lowest. According to the entertainment expenditure factor F_2, Bexley, Hillingdon and Havering's entertainment and leisure expenditure level accounted for most of the local expenditure, while it in the City of London, Hammersmith and Fulham and Islington are lowest.

Table 3: The sample of factor scores and ranks

Boroughs	**Comp1**	**Rank1**	**Comp2**	**Rank2**
City of London	1.077	4	-1.387	33
Hackney	-1.507	32	-0.914	24
Hammersmith and Fulham	0.772	9	-1.284	32
Haringey	-0.641	25	-0.623	22
Hillingdon	0.394	13	1.411	3
Islington	-0.137	21	-1.284	31
Kensington and Chelsea	2.239	1	-0.866	23
Lewisham	-1.486	31	-0.283	21
Newham	-2.296	33	-0.281	19
Richmond upon Thames	1.220	3	-0.282	20
Westminster	1.532	2	-1.064	27

[3] Because the data is standardized into the same scale, the highest here does not refer to the total expenditure, but the expenditure ratio

5.2 The result of cluster analysis

Based on the results of factor analysis, this paper uses the agglomerative hierarchical clustering method to do the cluster analysis of London boroughs. The result is shown in a thematic map (Figure 2). There are 11 boroughs are assigned to Cluster 1; 6 boroughs are assigned to Cluster 2; and 16 boroughs are assigned to Cluster 3 (see Table 4 for the specific boroughs of each cluster). Based on the factor score results in Table 3, the main expenditure characteristics of these three clusters can be summarized. Boroughs in Cluster 1 have low scores on both factors, indicating that the expenditure structure of these boroughs is relatively balanced compared to other boroughs. Boroughs in Cluster2 get higher score on factor F_1 and lower score on factor F_2, indicating that these areas have a higher proportion of expenditures on accommodation, food and health. The three boroughs with the highest total expenditure (Wandsworth, Westminster and Camden) belong to this cluster. Boroughs in Cluster 3 have higher score on factor F_2, indicating that these boroughs spend a higher proportion of leisure and entertainment.

Table 4: The result of clustering

Cluster	**Boroughs**
Cluster1: Boroughs with a balanced consumer expenditure structure	Barking and Dagenham, Greenwich, Hackney, Haringey, Islington, Lambeth, Lewisham, Newham, Southwark, Tower Hamlets, Waltham Forest
Cluster2: Boroughs with high living costs	Camden, City of London, Hammersmith and Fulham, Kensington and Chelsea, Wandsworth, Westminster
Cluster3: Boughs with high entertainment expenditure	Barnet, Bexley, Brent, Bromley, Croydon, Ealing, Enfield, Harrow, Havering, Hillingdon, Hounslow, Kingston upon Thames, Merton, Redbridge, Richmond upon Thames, Sutton

According to Figure 2, boroughs in Cluster 1 are mainly located in Inner London, indicating that most of the expenditure structure of Inner London is relatively balanced. However, in the west of Inner London, there are several boroughs where the cost of accommodation and food are high. This may be due to the higher cost of living in these areas. While the majority of boroughs in Outer London are spent more on leisure and entertainment, which shows that the basic cost of living in these places is relatively low.

6 Conclusions

Based on factor analysis and cluster analysis, this paper analyzes London's expenditure structure and its spatial differences, and has some key findings as follows: (1) Wandsworth, Westminster and Camden have the highest total expenditure. (2) More than half expenditure is spent on education and health in most boroughs. (3)

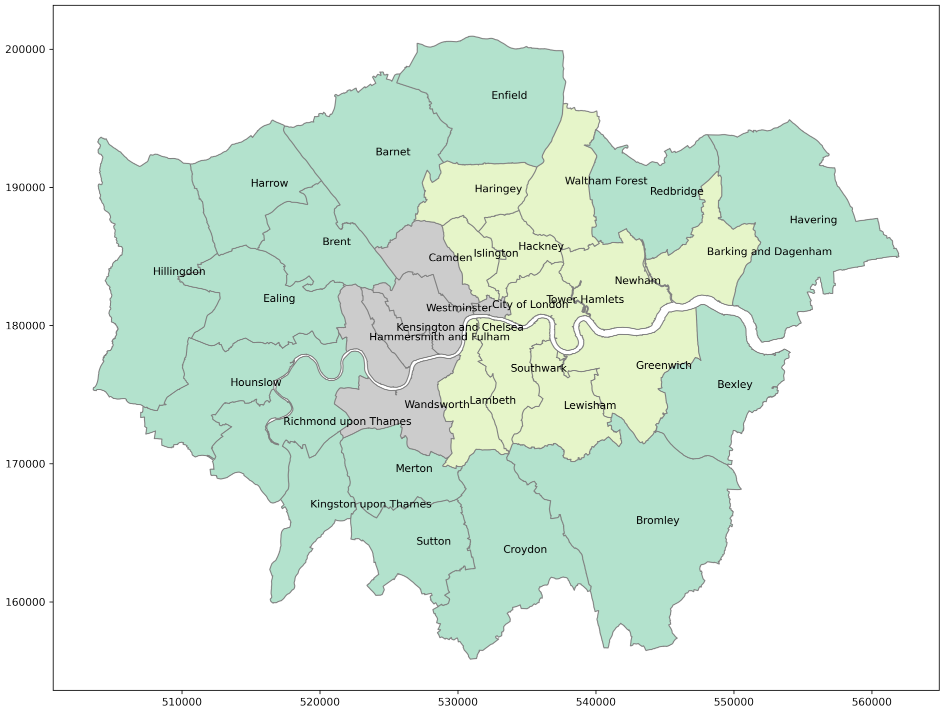

Figure 2: The the thematic map of the clustering result

The expenditure on accommodation is significantly positively correlated with that on license, while is significantly negatively correlated with transportation cost. (4) London boroughs can be divided into three clusters: boroughs with a balanced consumer expenditure structure, boroughs with high living costs, and boughs with high entertainment expenditure. (5) Most boroughs in Inner London have a balanced consumer expenditure structure and others have high cost of living, while boroughs in Outer London are basically have more entertainment expenditure. Therefore, in order to reduce the spatial differences in consumer expenditures in London, the government can formulate different management policies based on different expenditure patterns. For boroughs with higher living costs, price limits and subsidies can be used to control high costs; for Outer London areas, expenditure can be stimulated by increasing residents' incomes.

References

Cattell, R. B. (1966), 'The scree test for the number of factors', *Multivariate behavioral research* 1(2), 245–276.

De Mooij, M. and Hofstede, G. (2002), 'Convergence and divergence in consumer behavior: implications for international retailing', *Journal of retailing* 78(1), 61–69.

DeCoster, J. (1998), 'Overview of factor analysis'.

Duda, R. O., Hart, P. E. et al. (2006), *Pattern classification*, John Wiley & Sons.

Gangopadhyay, S. and Wadhwa, W. (2004), 'Changing pattern of household consumption expenditure', *Society for Economic Research & Financial Analysis, New Delhi, The Planning Commission, Government of India* .

Goldberger, A. S. and Gamaletsos, T. (1970), 'A cross-country comparison of consumer expenditure patterns', *European Economic Review* 1(3), 357–400.

Mooij, M. D. (2003), 'Convergence and divergence in consumer behaviour: implications for global advertising', *International Journal of advertising* 22(2), 183–202.

Murtagh, F. and Contreras, P. (2012), 'Algorithms for hierarchical clustering: an overview', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2(1), 86–97.

Nowak, J. and Kochkova, O. (2011), 'Income, culture, and household consumption expenditure patterns in the european union: convergence or divergence?', *Journal of International Consumer Marketing* 23(3-4), 260–275.

Pieri, S. P., Santamouris, M. et al. (2015), 'Identifying energy consumption patterns in the attica hotel sector using cluster analysis techniques with the aim of reducing hotels' co2 footprint', *Energy and Buildings* 94, 252–262.

Rokach, L. and Maimon, O. (2005), Clustering methods, *in* 'Data mining and knowledge discovery handbook', Springer, pp. 321–352.

Rummel, R. J. (1988), *Applied factor analysis*, Northwestern University Press.

Zhen, Q. (2008), The long term interactive relationship of economic growth, residents consumption and insurance development——based on var model and impulse responses function [j], *in* 'Statistics & Information Forum', Vol. 7.