

Mathematical modeling - food price analysis

Bingxian Leng, Yunfei Fu, Siyuan Li

Mathematics and Physics College, Dongying Teachers College, Shandong, China

Abstract: This paper mainly uses the idea of pedigree clustering analysis, gray prediction and principal component analysis. The clustering analysis model, GM (1,1) model and principal component analysis model were established by using SPSS software to analyze the correlation matrices and principal component analysis. MATLAB software was used to calculate the correlation matrices. In January, The difference in price changes of major food prices in cities is calculated, and had forecasted the various food prices in June 2016. For the first issue, the main food is classified and the data are processed. After that, the SPSS software is used to classify the 27 kinds of food into four categories by using the pedigree cluster analysis model and the system clustering. The four categories are made by EXCEL. The price of food changes over time with a line chart that analyzes the characteristics of food price volatility. For the second issue, the gray prediction model is established based on the food classification of each kind of food price. First, the original data is cumulated, test and processed, so that the data have a strong regularity, and then establish a gray differential equation, and then use MATLAB software to solve the model. And then the residual test and post-check test, have $C < 0.35$, the prediction accuracy is better. Finally, predict the price trend in June 2016 through the function. For the third issue, we analyzed the main components of 27 kinds of food types by celery, octopus, chicken (white striped chicken), duck and Chinese cabbage by using the data of principal given and analyzed by principal component analysis. It can be detected by measuring a small amount of food, this predict CPI value relatively accurate. Through the study of the characteristics of the region, select Shanghai and Shenyang, by looking for the relevant CPI and food price data, using spss software, principal component analysis, the impact of the CPI on several types of food, and then calculated by matlab algorithm weight, and then the data obtained by the analysis and comparison, different regions should be selected for different types of food for testing.

Key words: pedigree clustering, gray prediction, principal component analysis, SPSS software, MATLAB software

Chapter 1: Problem Statement

Food price is an important part of the consumer price index. It directly affects the cost of living and farmers' income. It is an important strategic issue related to the people's livelihood. Since 2000, China's urban household food consumption expenditure accounted for the proportion of total expenditure has been maintained at 36% or more. In the case of slow income growth, rising food prices will make the cost of living increased, especially food prices will reduce the life quality of low-income groups.

The Consumer Price Index (CPI) is an indicator of price changes based on statistics on the prices of products and services related to the life of the residents. It is usually used as an important indicator of the level of inflation.

According to the information on the annex, Shanghai and Shenyang find the relevant information and establish mathematical models to solve the following problems:

- (1) According to the annex and the relevant statistical website data, analysis of China's food price fluctuations.
- (2) To predict the trend of food prices in June 2016.
- (3) The current statistical departments need to monitor a large number of food price changes to calculate the residents of the consumer price index changes. Can accurate calculation be found out by monitoring little data of the food price changes Are the types of food and the number of species selected in the two different regions consistent with the same accuracy requirements

Chapter 2: Problem Analysis

The consumer price index is a fluctuating amount, which objectively expresses the level of inflation. Food prices are an important part of the price level of consumers. To solve the above three problems, it is necessary to clarify the change of food prices.

In response to the problem, it is required to analyze the characteristics of food price fluctuation in China according to the known statistical data. It can be seen that the price of food is the main

object of this study. However, due to the 27 kinds of known food types, the amount of data is relatively large, if analyses each food price fluctuations one by one, will inevitably lead to cumbersome process. So you can first classify 27 kinds of food based on the degree of correlation between the food prices. As each type of food prices have the same trend, so the analysis can be derived from China's food price fluctuations.

In response to question 2, it is necessary to predict the trend of food prices in June 2016, which can predict the price trend of food in each category on the basis of question one, because the price trend of each kind of food is roughly the same. Then the GM (1,1) gray prediction model is used to solve the forecast and analyze the price.

For the third problem, we use the data and find the data, using SPSS software principal component analysis to analyses 27 kinds of food ingredients in the main component. It can be detected by a small amount of food species, can be relatively accurate predict the CPI value. And select more characteristic of the two cities, by looking for the relevant CPI and food price data, using the same operation, and compare the type of food, the number is consistent.

Chapter 3: Model Assumptions

- (1) To collect the relevant data are accurate and reliable, high credibility.
- (2) The average price and the average daily price of food is very small, thus can be ignored.
- (3) The classification of food is divided according to the price trend, the same type of food price changes may vary, assuming that as long as the same price trend can be satisfied.
- (4) Assuming that there is no economic development in the forecast period (such as natural disasters), which can be a significant factor in food price fluctuations.

Chapter 4: Symbol Description

: The price of the i th food at the j th time point

: Distance between the two categories of p and q

- : The raw data of the time series
- : The original data after a cumulative data
- : The variance of the residual estimate of the data after one accumulation and the residual of the primary data of the original data
- : The variance of the original data

Chapter 5: Model Establishment and Solution

5.1 Establish a pedigree cluster analysis model

The problem requires the analysis of 50 urban food price fluctuations based on known statistical data. Therefore, we can see from the title of the requirements, the food price fluctuations is the object we want to analyze and study. But according to the data in the annex we can see that there are 27 types of food and the amount of data is relatively large. If each food price fluctuations are analyses one by one. The process is very cumbersome, and has no general concise. So you can be first classifying 27 kinds of food based on the relative extent of food prices. As each type of food prices has the same trend, you can get 50 urban food prices fluctuations and can find the difference between.

5.1.1 System clustering model

With n samples, that is, the title of 27 food categories, according to the chronological order of each sample, measured m values, set the time as an indicator. ($i = 1, 2, \dots, n; j = 1, 2, \dots, m$). The data in the table is called the observation data array, and its mathematical expression is

Where the column vector = (x_{1j}, \dots, x_{nj}) , representing the j th variable ($j = 1, 2, \dots, m$); the row vector = (x_{i1}, \dots, x_{im}) , representing the i th variable.

(1) The basic idea of system clustering

Samples (or variables) that are similar to each other are clustered into clusters, and the process is carried out continuously. Each sample (or variable) can always be clustered into the appropriate class. The system clustering process is: Assuming a total of n samples (or variables), the first step will be each sample (or variable) is alone in its own class, means that is a total of n class; second step is, according to the determined sample (or variable) 'distance' formula, the two samples (or variables), which are closer to each other, are aggregated into one class, and the other samples (or variables) are each clustered into one class, together into $n-1$ class; the third step will be the 'distance' of the nearest of the two samples (or variables). Two classes are further aggregated into a class of $n-2$; the above steps have been carried out, and finally all the samples (or variables) into a class. In order to visually reflect the above system clustering process, the entire classification system can be painted into a pedigree. So sometimes system clustering is also known as pedigree analysis.

(2) The basic steps of system clustering

1, Select the clustering method

The mean value of the distance between the classes is the average of the squares of the two types of elements.

=

In a certain step of the cluster, merged and into , and the distance between any class is:

=

=+)

=+

2, The system cluster process

1) Assuming a total of n samples (or variables), the first step will be each sample (or variable) is alone in its own class, means that is a total of n class.

2) According to the determined sample (or variable) 'distance' formula, the two samples (or variables), which are closer to each other, are aggregated into one class, and the other samples (or variables) are each clustered into one class, Class 1.

3) The 'distance' of the two closest class into a class, copolymerization into $n-2$ class.

The above steps have been going on and finally all the samples (or variables) into a class.

4) Spectral chart.

5) Determine the number of categories and various types of members.

(3) Determine the number of clusters

By the appropriate threshold, we determine the selected clustering method, according to the method of system clustering and class, get a line of clustering map, the cluster diagram only reflects the relationship between the samples, it does not give classification. Given a critical similarity scale, the classification of the sample is obtained by dividing the cluster map, such as a given critical value d . When the distance between the samples or between classes is less than d , it is considered that these samples are closely related to the class.

5.1.2 Food classification in 50 cities

From the pedigree clustering method above we can see: first determine the 50 urban food classification which have same indicators for rate of decline, and then system cluster analysis model was used to classify the average method, and then use the results to determine the number of categories, draw the price of various types of food price chart, and the figure analysis of food price growth.

According to Annex 1, this information is the food price list and table of food price rise of 27 kinds in 50 cities in the January 1, 2016 to May 10, 2016. SPSS software is used to increase the table data clustering system classification results.

(1) From the pedigree (shown in Figure 1), that is four types of results (as shown in Table 1):

Fig.1 Cluster diagram of the average spine system

(2) The final result of food classification:

First Type Cabbage

Second Type Cucumber, tomatoes

Third Type Beans

Forth Type Rice, flour (flour), flour, soy products, peanut oil, soybean oil, rapeseed oil, pork (hind legs), pork (pork), beef, lamb, chicken (white striped chicken), chicken (chicken breast), duck, egg, live carp, live grass carp, octopus, rape, celery, potato, apple, banana

Table 1 Food Category Table

5.1.3 Analysis of the fluctuation characteristics of each kind of food price

(1) Price chart

In order to further explain the rationality of various food classification and the characteristics of the average price of various types of food, the combination of various types of food price trend chart with more intuitive description is used, since the fourth category contains more food types , so we select the average price trend of all food plot. For first category and third category which contains only a food, we just make the average price of each food to one chart. Charts and each type of food features are shown in Figure 2 below.

Figure 2 Food Category Chart

(2) Price fluctuation characteristics analysis

During this period, the average price of food in first category non-stop changes around 3.5, the volatility is roughly 1.5, the overall price of such food is in a slow upward trend.

During this period, the average price of food in second category non-stop changes around 8, the volatility is roughly 2, the overall price of such food is in a downward trend.

During this period, the average price of food in third category non-stop changes around 14, the volatility is roughly 4, there is a substantial price increases in the middle of the period, such food prices

are not stable.

During this period, the average price of food in fourth category non-stop changes around 18.9, the volatility is roughly 0.5, the overall price is relatively stable, such food prices will not change much.

5.2 Prediction of food price movements in June 2016 using the gray forecasting model GM (1,1)

Asked to predict the trend of food prices in June 2016. Since each kind of food in the price trend is roughly the same, if each of the 27 kinds of food is predicted, obviously the process is cumbersome, there is no representation and unity. Method above classified 27 kinds of food into four categories, and uses the trend of each category to replace the various categories of food prices in the trend. The gray forecasting model is used to test and deal with the data, and the original data is accumulated once, so that the data has a strong regularity, and then the gray differential equation is established, and then the MATLAB model is used to solve the model. After that the residual test and post-check test is carried out, have $C < 0.35$, the prediction accuracy is better. Finally, we predict the price trend in June 2016 through the function.

5.2.1 Modeling steps are as follows:

(1) GM (1, 1) represents a differential equation of albino form:
 $a+u(1-1)$

In the formula, u, a, is the parameter that needs to be obtained by modeling; it is the cumulative generation (AGO) value of the original data.

(2) The first data column of the first k elements of the cumulative data generated after the first k elements of the k, this is the data processing. Expressed as:

$(k)=(n) (1-2)$

We not using the original data modeling directly, but the original, irregular data processing, making it more regular, and then use the generated data to analyze the data column, which is the characteristics of the gray system theory one.

(3) for GM (1,1), the data matrix is:

$B= (1-3)$

Vector =

(4) For least squares estimation, with matlab software parameters a, u:

$=(BTB)-1BTYN (1-4)$

(5) Set the time response function, find the solution of the differential equation (1-1):

$(1)(t+1)=(X(0)(1)-) + (1-5)$

The above is the gray prediction model to be built.

5.2.2 Solving the model

Taking the first category of cabbage as an example, the trend of food prices in June 2016 is forecasted. The average price changes of 50 urban cabbages are shown in Table 2 below. The data are from Annex 1.

TIME	1.1-1.10	1.11-1.20	1.21-1.30	2.1-2.10	2.11-2.20
	2.21-2.29				
PRICE	2.31	2.43	3.2	2.36	3.89
	3.74				
TIME	3.1-3.10	3.11-3.20	3.21-3.30	4.1-4.10	4.11-4.20
	4.21-4.30	5.1-5.10			
PRICE	3.98	4.69	5.13	4.8	3.88
	3.43	3.19			

Table 2 Average price changes of cabbage

(1) Cumulative generation

Logarithm={2.31,2.43,3.20,2.36

3.89,3.74,3.98,4.59,5.13,4.80,3.88,3.43,3.19} undergo accumulation and become:

={2.31,4.74,7.94,10.30,14.19,17.93,21.91,26.60,31.73,36.53,40.41,43.84,47.03}

(2) Construct the data matrix B and the data vector

According to 3 the construction matrix $B= , =$

(3) Calculate $=(BTB)-1BTYN$

Substituting matrix B and vector into (4) yields =

(4) To arrive at the prediction model

According to (1-5) can be obtained model (1), from (1)(t+1)=129.41-127.1, take t for the stress order k, by

(1) (1)(k+1)=129.41-127.1 (X(0)(1)=2.31,-127.1)

We can get the cumulative array (1)(k+1)(k=1,2,.....)

(5) Residual test

1, According to the forecast formula, calculates (1) (k), get:

(1) (1)(k)={ 5.55 , 8.86 , 12.26 , 15.75 , 19.32 , 22.98 , 26.74 , 30.58 , 34.53 , 38.57 , 42.71 , 46.96}

And the actual ones are:

={4.74,7.94,10.30,14.19,17.93,21.91,26.60,31.73,36.53,40.41,43.84,47.03}

Calculate the average relative error of 0.1, the residual variance = 1.741, the variance = 17.515,

2, Forecast residuals

$C=0.0994 < 0.35$

Accuracy Test Level Reference Table:

Indicator	Evaluation	c	p
Level 1	Good	0.35	0.95
Level 2	Qualified	0.50	0.80
Level 3	Barely qualified	0.65	0.70
Level 4	Unqualified	0.80	0.60

Table 3 Accuracy Test Level Reference Table

From Table 3, we can see that the prediction accuracy is good.

(6) Forecast data

$K=12 , =-4.35$

$K=13 , =-4.46$

$K=14 , =-4.57$

$K=15 , =-4.69$

$K=16 , =-4.80$

$K=17 , =-4.92$

Therefore, the price of cabbage 6.1-6.10 is 4.68 yuan, the price of 6.10-6.20 is 4.80 yuan, the price of 6.20-6.30 is 4.92 yuan.

By repeat the above operation, you can get the price of food in June, as shown in Table 4 below

	6.1-6.10	6.10-6.20	6.20-6.30		
FIRST CATEGORY	Cabbage	4.69	4.80	4.92	
SECOND CATEGORY	Cucumber		4.65	4.39	
	4.15				
	Tomato	4.68	4.42	4.17	
THIRD CATEGORY	Beans	12.4	12.25	12.10	
FORTH CATEGORY	Rice	6.26	6.26	6.26	
	Flour(Fu Qiang)	6.06	6.06	6.07	
	Flour(Standard)	4.89	4.90	4.92	
	Soy products	4.69	4.69	4.71	
	Peanut oil	27.52	27.56	27.58	
	Soybean oil	9.93	9.95	9.99	
	Rapeseed oil	13.76	13.78	13.80	
	Pork (hind legs)	31.49	31.51	31.68	
	Pork(belly)	32.54	32.63	32.70	

Beef	66.87	66.89	66.90
Mutton	59.98	60.22	61.01
Chicken (white striped chicken)		21.89	21.91
21.95			
Chicken (Chicken Breast)	20.87	20.89	20.91
Duck	18.89	18.90	18.93
Egg	9.08	9.09	9.10
Live carp	15.36	15.36	15.37
Live carp	15.91	15.93	16.01
Octopus	34.16	34.18	34.20
Rape	5.47	5.89	6.51
Celery	7.66	7.92	8.44
Potato	5.89	5.94	6.02
Apple	12.56	12.78	12.80
banana	6.39	6.41	6.43

Table 4 June food price forecast table

5.3 Principal component analysis model

The consumer price index, referred to as CPI, is a measure of the relative prices of residents' food prices over time, reflecting the changes in the purchase of food by residents. In this paper, the average price of 27 kinds of major food and CPI were analyzed in 50 cities from May 2015 to April 2016, screening out the development trend of the larger CPI varieties of food. It is necessary to use the principal component analysis method to find the minimum price of food from a large quantity of food prices to calculate accurately and to forecast the consumer price index.

And then select the two characteristics of the city of Shanghai, Shenyang. According to the regional differences, select the appropriate types of goods and the number of the use of the same mathematical model to operate.

5.3.1 Changes in consumer price index of national residents

To determine whether we can predict CPI by using small amount of food prices, we need to determine the impact of food prices on the impact of CPI. When the weight ratio is high, you can think that food prices can significantly affect the CPI, thus we can get relatively accurate calculations to predict the consumer price index. According to the size of the weight ratio, choose one or several food, with how far the linear regression method, to find this one or several food prices and the relationship between CPI, and then through a small amount of food prices to predict the purpose of CPI.

(1) The method of principal component analysis

1, calculate the correlation coefficient matrix

$R =$

In the above equation, $(i, j = 1, 2, 3, 4, \dots, 27)$ is the correlation coefficient between the original variable and the original variable,

$=$

Since R is a real symmetric matrix ($=$), it is only necessary to calculate the upper triangular element.

2. Calculate eigenvalues and eigenvectors

First, solve the characteristic equation $|I-R|=0$, eigenvalues $i (1, 2, 3, \dots, 27)$ are usually obtained by Jacobi (Jacobi), and they are arranged in order of size, that is, $1 > 2 > \dots > 27 > 0$, Then find the eigenvector $(1, 2, 3, \dots, 27)$ corresponding to the eigenvalue. Here, we need $|=1$, which $=1$, where represent the j th component of the vector .

3. Calculate the contribution rate of the main component and the cumulative contribution rate

The contribution rate of the principal component is:

The cumulative contribution rate is: $(i=1, 2, 3, \dots, 27)$

According to Annex I and the data (see Table 1 for the data, see

Table 1), this data includes the data on the average food price change of 50 cities in May 2015 and April 2016 and the corresponding national consumer price index for each month. By using 27 kinds of food as a factor, we get the data table through the CPI data and the spss software which shown in Table 5 below.

Description of the total number of variants

Element	Start Feature Value	Fetches Square and Load	Total variation (%)	Total variance (%)	Total
1	15.387	56.991	56.991	15.387	56.991
2	5.620	20.815	77.805	5.620	20.815
3	2.846	10.542	88.347	2.846	10.542
4	1.506	5.576	93.923	1.506	5.576

Table 5 eigenvalue and contribution rate of R

The first three standardized samples of the main classification and contribution rate has reached 88.35%, so we only need to take the first three main components. The coefficients of the normalized variables $= (i=1, 2, 3)$ in the first three normalized sample principal components are the corresponding eigenvectors.

Calculated by excel, the first, second and third main components of the load as shown in Table 6 below:

The first principal component	The second principal component	The third principal component
0.2386	0.0645	-0.0528
0.2371	-0.0806	0.0937
0.2450	0.0713	-0.1008
0.2493	-0.0460	0.0723
0.2330	-0.1000	-0.0439
-0.2292	-0.0882	0.0593
0.2381	0.0413	-0.1227
0.1841	-0.0949	0.1328
0.1930	-0.1050	0.1197
0.1629	-0.0979	0.3675
-0.1797	0.2430	0.2235
0.2429	0.1000	-0.0202
0.1066	0.3328	-0.2253
0.0994	0.2379	0.3687
0.0250	-0.1890	0.4256
-0.0928	0.3214	0.2970
-0.0897	0.3290	0.2650
0.2333	0.1287	0.0225
0.1242	0.3311	-0.0670
0.2080	-0.0759	0.2282
0.2098	0.2050	0.1381
0.2009	-0.1198	0.1263
0.2381	-0.0996	-0.0089
0.2338	0.0456	0.0089
0.1591	0.2215	-0.2768
-0.1968	0.2012	0.1843
0.0311	0.3990	-0.0806

Table 6 Main component load table

The weight (P) of each food is calculated according to the contribution rate and the load of the first principal component, the second principal component and the third principal component:

$= (J \text{ means } 27 \text{ kinds of food, } j = 1, 2, \dots, 27)$

Calculated by MATLAB software, the impact of the food price on the CPI weight as shown in Table 7 below:

Food Name	Weight
Rice	0.1628
Floor(Fu Qiang)	0.1451
Floor (Standard)	0.1628
Peanut oil	0.1586

Peanut oil	0.1215
Soybean Oil	-0.1616
Rapeseed oil	0.1487
Pork (hind legs)	0.1122
Pork(belly)	0.114
Beef	0.1258
Mutton	-0.032
Chicken (white striped chicken)	0.1778
Chicken (Chicken Breast)	0.1203
Duck	0.1642
Egg	0.0224
Live carp	0.0513
Live carp	0.0513
Octopus	0.1835
Cabbage	0.1501
Rape	0.1435
Celery	0.2001
Cucumber	0.1164
Tomato	0.1291
Beans	0.1626
Potato	0.1218
Apple	-0.0575
Banana	0.1045

Table 7 Weight ratio

Through the table analysis, the large food consumption index which affects residents can be obtained. They were celery, octopus, chicken (white striped chicken) and duck.

Little data of celery, octopus, chicken (white striped chicken), and ducks such food is enough for the relatively accurate calculation of food price and the forecast of consumer price index.

5.3.2 Selected food types and numbers of Shanghai and Shenyang
Similarly, the principal component analysis method can be used to calculate the food and quantity of food selected in Shanghai and Shenyang, as shown in Fig.3.

Figure 3 Shanghai, Shenyang weight ratio table

From the chart, we can clearly see the Shanghai and Shenyang, the weight of various types of food is not the same, so in the same precision requirements, the selected number of food types are not the same.

Chapter 6: The evaluation of the model

For problem 1, a reasonable classification of class is perform by the use of pedigree clustering method of the food involved in the retail price movements and other fuzzy nature of the quantitative determination.

For the GM (1,1) gray prediction model based on the least squares method in question 2, the exponential function fitting method is adopted. Compared with the polynomial fitting model, linear fitting, quadratic polynomial fitting, and as well as prediction accuracy are greatly improved.

For the analysis model of problem 3, taking into account the impact of multiple variables on a single variable, the accuracy of the forecast greatly improved.

This article also has some shortcomings, such as the requirements of a kind of requirements on the classification of food. This article only classified the similarity of the average price of food, and thus the use of food retail price is not many. This making the classification of indicators than a single, not on the internal relationship between the deep mining of food, which may lead to the final classification results are too uneven (such as the number of different types of food is different).

Chapter 7: The promotion of the model

The correlation analysis model and the gray forecasting model

involved in this paper have good promotion value. Gray system theory proposed a new method of analysis - correlation degree analysis method, that is, according to the development trend of factors similarity or degree of dissimilarity to measure the degree of correlation between the factors, it reveals the dynamic relationship between things features and extent. Due to the development trend as a foothold, the amount of the sample size is not an excessive requirements, it do not need a typical distribution of the law, the calculation is small or even hand count, and does not appear correlation degree of quantitative results and qualitative analysis of the situation is inconsistent . This method has been applied to the agricultural economy, water conservancy, macroeconomic and other aspects, which achieved good results.

The main task of gray system theory modeling is to develop, and according to the characteristic data, use the explicit information and hidden information in the data of the specific gray system to find the mathematical relationship among the factors or the factors themselves. The usual approach is to use a discrete model to create a time-by-segment model. However, the discrete model can only do short-term analysis of the development of objective systems, and cannot meet the requirements of long-term analysis. Although the discrete approximation model of continuous systems is useful for many engineering applications, it is often desirable to use differential equations in some areas of research. In fact, the system of differential equations describes the nature of the physical or chemical processes within the system we want to identify.

At present, the gray system theory has been successfully applied to engineering control, economic management, futuristic research, ecosystems and complex and agile agricultural systems, and has made gratifying achievements. Gray system theory is likely to analyze, model, predict, and control the abstract system of society and economy. It may become a new theoretical tool for people to understand the objective system of objective system transformation.

REFERENCES

- [1] System clustering analysis theory.
- [2] Gray forecasting model of network public opinion events and case analysis 'Information Science' 2013/12
- [3] Gray theoretical basis [M]. Deng Ju Long. Wuhan: Huazhong University of Science and Technology Press, 2002
- [4] gray prediction and decision model research [M] Beijing: Science Press, 2009
- [5] National Statistical Office of the People's Republic of China <http://www.stats.gov.cn/>
- [6] CPI and food prices and the relationship between the trend Xiong Jingwen
- [7] Shanghai Municipal Government Data Service Network <Http://www.datashanghai.gov.cn/query!queryDataByField.actiondataField=9>
- [8] Shenyang Price Bureau Network <http://wjxx.price.ln.cn/web/main.aspx>