Research on the Sales Distribution Law of Vegetable Products Based on Linear Fitting

Meng Wang[1, *, #], Aixin Leng[2, #], Chuhan Yu[1, #]

[1]School of computer science and technology, Shandong University of Technology, Shandong, China, 255000

[2]School of Mathematics and Statistics, Shandong University of Technology, Shandong, China, 255000

*Corresponding author: 19553305960@163.com

#These authors contributed equally.

Abstract. Firstly, the hourly sales volume changes of vegetable categories and individual products are analyzed from the time dimension, and frequency distribution histograms are used to describe the distribution patterns of vegetable categories and individual product sales volume. Then, Spearman correlation analysis is used to obtain the distribution patterns of individual product sales volume, and linear fitting is used to compare and analyze the continuous sales volume of the six major categories for three years using a bar chart, From the demand side perspective, there is often a certain correlation between the sales volume of vegetable products and time. This article analyzes the relationship between vegetable categories and individual product sales volume from the time dimension. At the same time, based on the distribution patterns of the six major categories, Spearman correlation analysis was used to obtain the distribution patterns of individual product sales volume. Linear fitting was performed, and the continuous sales volume of the six categories for three years was compared and analyzed using a bar chart to obtain the general monthly and seasonal patterns of vegetable sales. Then, descriptive statistical analysis is used to analyze the distribution patterns of vegetable categories and single item sales volume, and the correlation relationships between different categories or single items of vegetable products are obtained. Finally, a grey correlation analysis model is established to predict the distribution patterns and interrelationships of vegetable categories and single item sales volume based on this model.

Keywords: Spearman correlation analysis, grey correlation analysis, linear fitting.

1. Introduction

The shelf life of vegetable products in supermarkets is relatively short and is limited by the purchase transaction time. Supermarkets usually replenish daily based on the historical sales and demand of each product without knowing the specific individual product and purchase price[1]. The pricing of vegetables generally adopts the "cost plus pricing" method [2], and supermarkets usually offer discounts on goods with transportation losses and poor quality. The sales volume and time of vegetable products often have a certain correlation, and the supply variety is relatively rich from April to October. Analyzing market demand and supermarket sales space is based on the basic data of vegetables, establishing a mathematical model, and formulating appropriate replenishment and pricing decisions.

2. Analysis of distribution patterns of various categories based on linear fitting

BP This article first summarizes the sales volume of six categories. From July 2020 to June 2023, for a total of 36 months, linear fitting was performed on the corresponding sales volume of each category on a monthly basis, and the trend of change was obtained. Before summarizing, perform dummy variable processing on the relevant data,Then use Matlab to obtain the changes in sales of each category over the past 36 months, as shown in Figure 1, and analyze the overall trend of changes in various categories of vegetables [3].

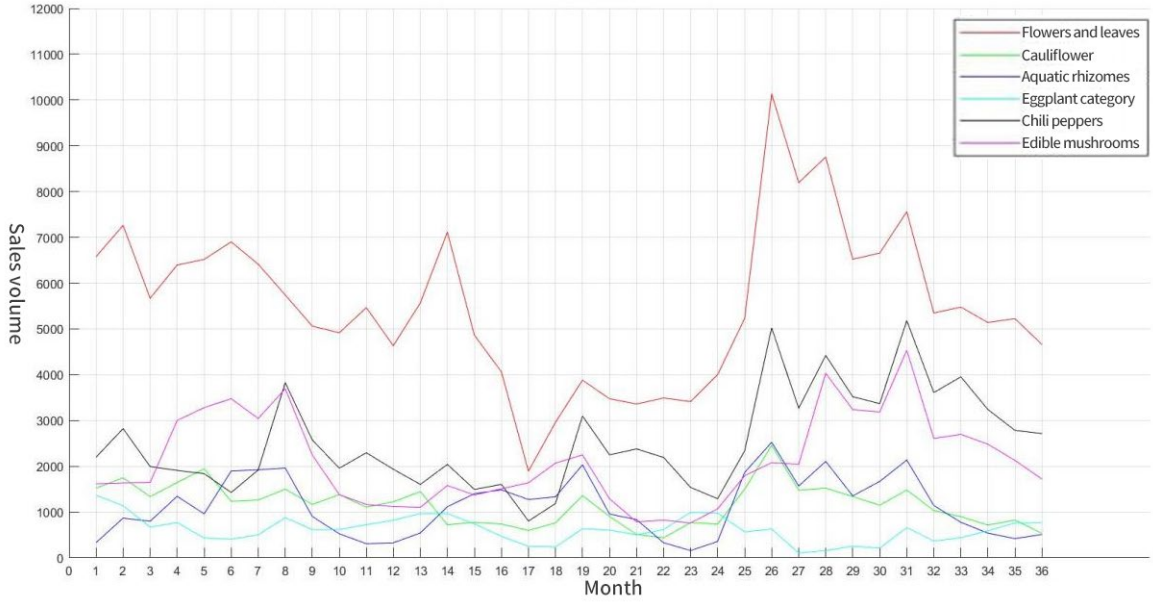


Fig 1. Line chart of monthly sales

From the above Fig 1, it can be seen that the overall sales volume shows a bimodal distribution, with a long "sales trough" period from the 16th to the 24th week (i.e. from October 2021 to June 2022), and then rapidly increasing after the 24th week. The overall data of each category fluctuates greatly, but it is not difficult to see that the sales of leafy vegetables are always the highest, while eggplant vegetables are always very few.

Secondly, due to the possibility of seasonal changes in vegetable sales, this data was divided into three years and analyzed separately for six categories by month. Draw a bar chart using Excel as shown in Fig 2 and observe its variation pattern.

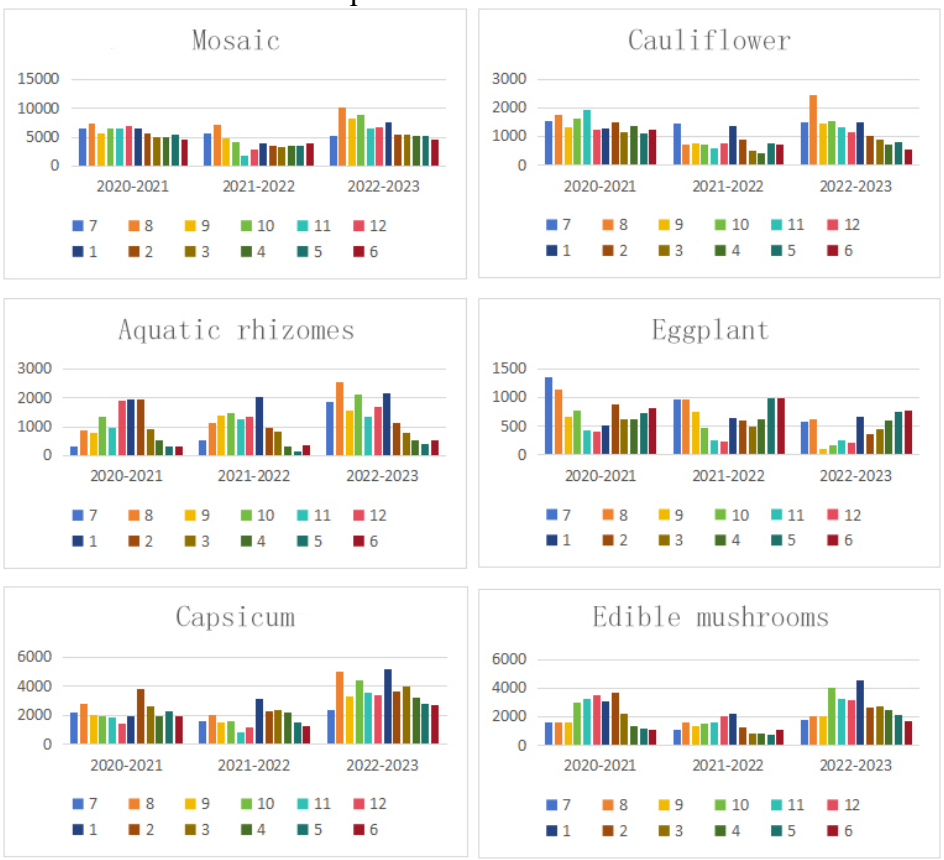


Fig 2 Comparison of sales volume of six categories over the past three years

The comparison of the total sales volume of each category over the past three years shows that [4]: the flower leaf category has the highest sales volume among the six categories, while the eggplant category has the lowest sales volume; The peak sales season is from August to January of the following year, and the sales volume is relatively low from March to July each year. The sales of flowers and leaves are relatively stable, with the peak season from July to October every year; The peak season for cauliflower is from July to January of the following year; Aquatic rhizomes have a peak season from October to February each year, but over time, the peak season tends to move forward; Solanaceae are very different from other categories, with September to December being the off-season each year; The sales of chili peppers are relatively stable, with January to April as the peak season every year; The peak season for edible mushrooms is from October to February of the following year.

Based on the above results, linear fitting was performed on the changes in sales volume, and a chart was drawn using Matlab. The total sales volume of the six categories within three years was calculated by month, and the continuous trend of each category over the past three years was observed, as shown in Fig 3.

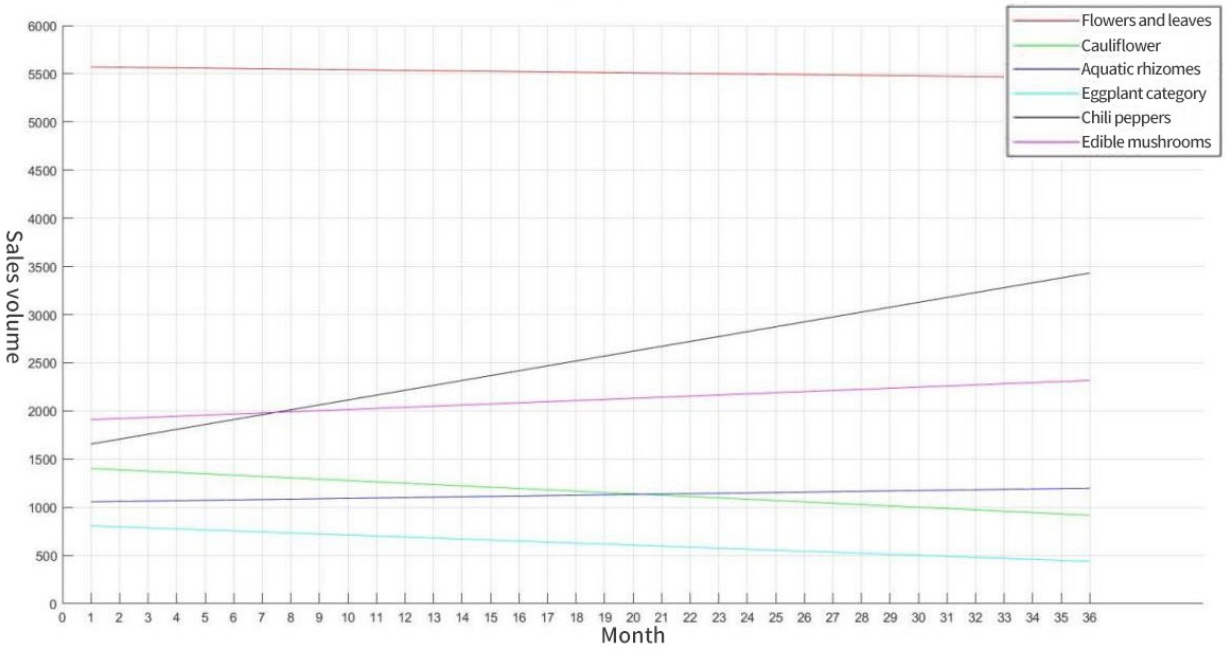


Fig 3 Linear fitting diagram of monthly sales volume

Analyzing Fig 3, it can be seen that the sales of the three categories of flowers and leaves, cauliflower, and eggplants have shown a slight downward trend overall over the past three years; The sales of aquatic rhizomes have slowly increased slightly for three consecutive years, showing a flat growth trend; The sales of chili peppers and edible mushrooms have shown an overall growth trend, with chili peppers showing a significant annual growth trend. It is expected that their sales will continue to increase.

Overall, there has been no particularly significant change in the overall trend of each product (except for chili peppers) over the past three years. The sales of flowers and leaves are the highest, while the sales of eggplants are the least, and most of the peak seasons are autumn and winter.

2.1. Analysis of the Distribution Pattern of Single Product Sales Based on Spearman Correlation Coefficient

This article first summarizes the sales volume of 246 individual products, as shown in Fig 4, and observes the distribution of sales volume for different individual products [5].

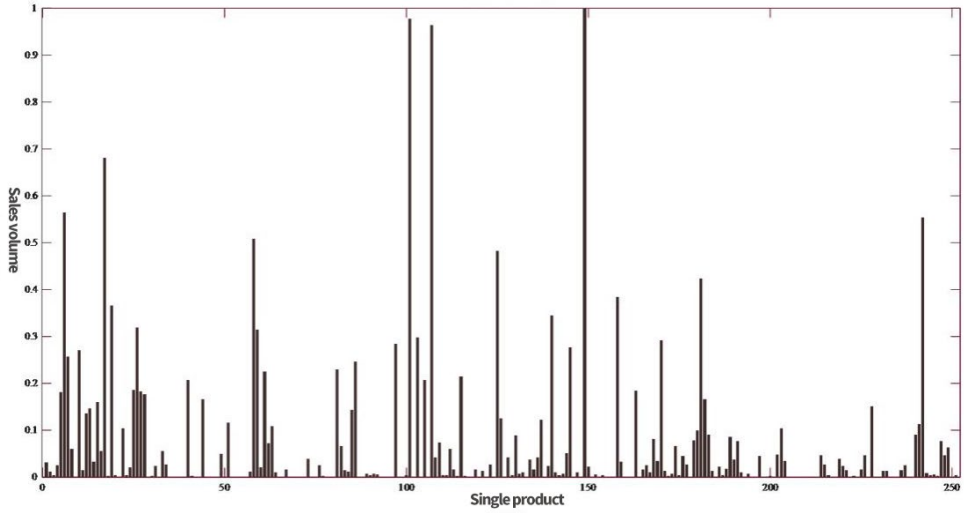


Fig 4 Summary of Single Product Sales

Next, process the data, sequentially code the individual items 1-246, and sequentially code the days 1-1095 to obtain a summary table of daily sales for each item. Use this data [6] for analysis. There are currently 251 units corresponding to 1095 day sales data. This article uses Matlab to analyze the Spearman [7] correlation coefficient between each two units and observe its distribution pattern, as shown in Table 1.

Table 1 Spearman correlation coefficient between individual products (partial)

1	-0.031836959	-0.067742798	0.021204337	-0.009768456	-0.045381433
-0.031836959	1	-0.084360832	0.230726865	0.2803918	-0.019116413
-0.067742798	-0.084360832	1	-0.064034716	-0.220991683	-0.218490096
0.021204337	0.230726865	-0.064034716	1	-0.111917161	0.073733456
-0.009768456	0.2803918	-0.220991683	-0.111917161	1	0.313785458
-0.045381433	-0.019116413	-0.218490096	0.073733456	0.313785458	1
-0.006608302	-0.004197189	-0.14821307	-0.21152335	0.719593929	0.384951819
-0.021096152	-0.171026363	-0.134039331	-0.182161464	0.352564585	0.601645579

Based on the above analysis, it can be concluded that:

1. Flower and leaf species: Red coral (coarse leaves), red oak leaves, and green avocado have the strongest correlation coefficient, all of which are r=1.

2. Cauliflower: Green stem scattered flowers have the strongest correlation with Flammulina velutipes (1) (edible mushroom), r=0.8732. However, within the cauliflower category, the correlation between green stem scattered flowers and branch river green stem scattered flowers is strongest, with r=-0.4802.

3. Aquatic rhizomes: Honghu lotus root (Crisp lotus root) has the strongest correlation with local yellow heart rapeseed (Flower and Leaf type), r=0.7149. Within aquatic rhizomes, the correlation between water chestnut and Honghu lotus root is strongest, with r=0.5798.

4. Eggplant: Green eggplant (2) and purple eggplant (1) have the strongest correlation, r=0.6924.

5. Chili peppers: Xiaomi pepper has the strongest correlation with Xiaowrinkled skin, r=0.8799.

6. Edible mushroom: Golden needle mushroom (1) and apricot mushroom (1) have the strongest correlation, r=0.8959.

Analysis shows that most vegetables have a stronger correlation with similar vegetables; But even though some vegetables have high single item sales, their correlation with other vegetables is not significant. For items with high cross category correlation, it is speculated that they are either hidden bundled sales or bundled sales models conducted by supermarkets due to the combination of dishes;

There are also negatively correlated vegetable items, mostly due to similar vegetable varieties, customers will not purchase them together.

3. Analysis of the relationship between sales volume of various categories and individual products based on grey correlation analysis

Based on the above data, the grey correlation analysis method [8] is used to analyze the interrelationships between various categories and individual product sales. Firstly, this article divides the three-year period into 36 months by month, and then obtains the corresponding monthly sales of categories and their individual products. Grey correlation analysis is conducted on 6 categories to obtain the correlation between individual product sales and categories [9].

This article takes the 36 month sales volume of the category as the parent sequence, which includes the corresponding 36 month sales volume of each individual product as the sub sequence, and establishes a grey correlation analysis model[10].

Parent sequence:

$$\mathrm{y}_1 = (y_1(1), y_1(2), \ldots\ldots, y_1(36))^T \tag{1}$$

Subsequence:

$$x_1 = (x_1(1), x_1(2), \ldots\ldots, x_1(36))^T \tag{2}$$

$$x_n = (x_n(1), x_n(2), \ldots\ldots, x_n(36))^T \tag{3}$$

Set the maximum value of the two poles as a and the minimum value as b:

$$\mathrm{a} = \max(\mathrm{i})\max(\mathrm{k})|\mathrm{y}_1(\mathrm{k}) - \mathrm{x}_\mathrm{i}(\mathrm{k})| \tag{4}$$

$$\mathrm{b} = \min(\mathrm{i})\min(\mathrm{k})|\mathrm{y}_1(\mathrm{k}) - \mathrm{x}_\mathrm{i}(\mathrm{k})| \tag{5}$$

The grey correlation degree is:

$$\mathrm{Y}(\mathrm{y}_1, \mathrm{x}_\mathrm{i}) = \frac{1}{\mathrm{n}}\sum_{\mathrm{k}=1}^{\mathrm{n}} \mathrm{y}(\mathrm{y}_1(\mathrm{k}), \mathrm{x}_\mathrm{i}(\mathrm{k})) \tag{6}$$

Gama is:

$$\mathrm{y}(\mathrm{y}_1(\mathrm{k}), \mathrm{x}_\mathrm{i}(\mathrm{k})) = \frac{\mathrm{b}+\rho\mathrm{a}}{|\mathrm{y}_1(\mathrm{k})-\mathrm{x}_\mathrm{i}(\mathrm{k})|+\rho\mathrm{a}} \tag{7}$$

In Gama, ρ For resolution, this article sets it to 0.5.

There are relatively few single items of cauliflower vegetables, and according to the analysis above, the sales volume of each item is quite considerable. So this article takes cauliflower and its individual products as examples, and uses SPSS software to obtain the grey correlation between their individual product sales and categories, as shown in Table 2 and Fig 7.

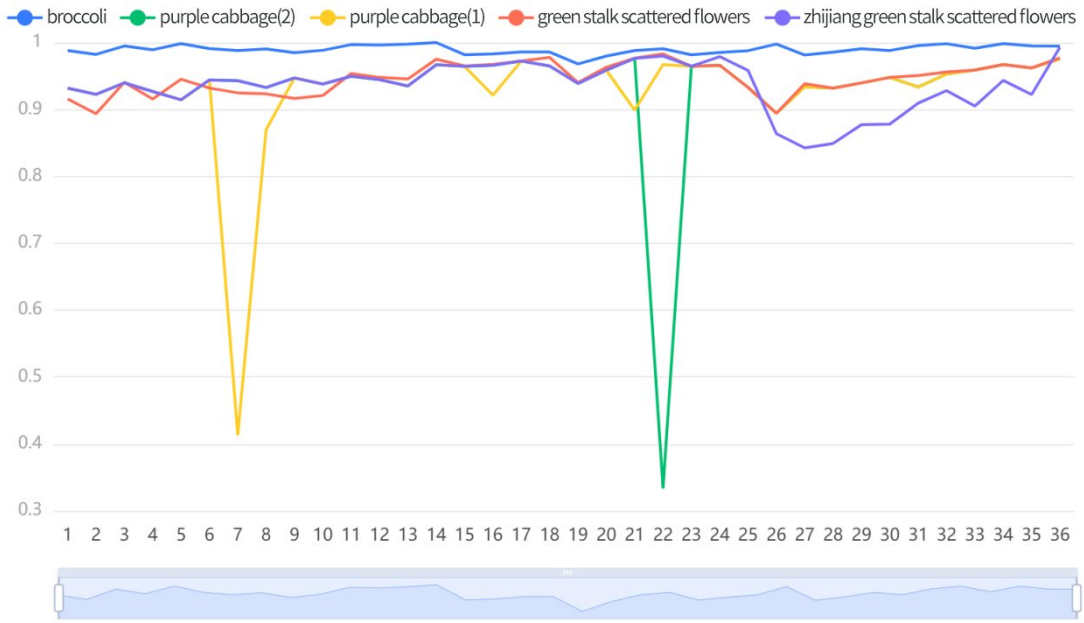


Fig 7 Grey correlation between individual product sales and category

Table 2 Correlation degree result

Evaluation item	Degree of association	Ranking
broccoli	0.989	1
Green stalk scattered flowers	0.947	2
Zhijiang green stalk scattered flowers	0.935	3
purple.cabbage(2)	0.93	4
purple cabbage(l)	0.928	5

According to the results, it can be seen that the top five items have a high to low correlation with the category:

1. Cauliflower: broccoli, green stem scattered flowers, Zhijiang green stem scattered flowers, purple cabbage (2), purple cabbage (1).

2. Flowers and leaves: Shanghai green, yellow cabbage (2), milk cabbage, baby cabbage, spinach.

3. Aquatic rhizomes: Clean lotus root, Honghu lotus root (powder lotus root), water chestnut, high melon (2), high melon (1).

4. Eggplant category: purple eggplant (2), green eggplant (1), round eggplant (2), long line eggplant, and big dragon eggplant.

5. Chili peppers: Wuhu green pepper, screw pepper, red pepper (1), millet pepper (portion), and green pepper.

6. Edible mushrooms: White Jade Mushroom (bag), Xixia Mushroom (1), Golden Needle Mushroom (box), White Fungus (bud), Apricot Mushroom (bag)

Overall, it can be found that the correlation between individual product sales and varieties is not solely determined by the total sales volume. Even if a high selling item happens to be in a high selling category, it cannot be concluded that the high sales of the item are caused by the category. To obtain more reliable relationship results, it is necessary to connect with the month and analyze the correlation between the two parties.

4. Conclusions

The above analysis indicates that there is a correlation between different categories or individual items of vegetable products. The distribution pattern of sales volume of various categories and individual products is closely related to the season, with most peak seasons occurring in autumn and

winter. Some individual products have strong correlation, most of which are similar vegetables, while some highly correlated exotic vegetables are speculated to be invisible bundled sales due to the menu. There are also significant differences in the correlation between individual products and categories, and the sales of products with strong basic correlation are also considerable.

References

[1] Gu Sihong. Research on Dynamic Pricing of Fresh Products from H Retailers Considering Freshness Changes [D]. Donghua University, 2023

[2] Mao Lisha. Research on Pricing Strategy and Production and Sales Model of Vegetable Wholesale Market from the Perspective of Supply Chain [D]. Central South University of Forestry and Technology, 2023

[3] Yang Haoxu. Analysis and Research on the Correlation between Vegetable Price and Sales Volume: Taking Yamai Cai as an Example [J]. Science and Technology Literature Park

[4] Lu Yajie. Research on Dynamic Pricing of High Quality Fresh Vegetables in Chinese Supermarkets [D]. Beijing Jiaotong University, 2010

[5] Wang Xiaoyan, Li Meimei. Discussion on the correlation coefficient between grades and Spearman grades [J]. Guangdong Light Industry Vocational and Technical College, 2006, 04 (05): 26-27

[6] Tian Yumei, Zhou Dan, Zhang Wei, et al. Comparison of Five Inorganic Elements in Curcuma and Correlation Analysis of Sample Elements [J]. Zhejiang Key Laboratory of Solid Surface Reaction Chemistry, 2008,09 (28): 2193-2195

[7] Zhang Peng, Zhou Hui, Liang Yongqing, et al. Spearman correlation analysis of FFDM imaging signs and ER, PR, C-erbB-2 expression in breast cancer [J]. Qingyuan, Guangdong, 2022

[8] Chi Fuling, Li Feng, Ruan Huiming, et al. Comprehensive evaluation of sightseeing fruits and vegetables based on grey correlation analysis [J]. Hubei Agricultural Science, 2021, 60 (14): 85-89

[9] Chen Wenwen. Research on the Factors Influencing the Development of China's Sports Industry Based on Grey Correlation Analysis [J]. Shanghai Institute of Physical Education

[10] Yuan Conghao, Li Ning, Niu Ke. Analysis of Traffic Congestion Based on Cluster and Correlation Analysis [J]. Journal of Beijing University of Information Technology, 2018,33 (03): 37-41