Research on Pricing and Replenishment of Vegetables Based on Grey Relational Model and Seasonal Index Prediction Model

Keyu Li*

School Of Biomedical Engineering,Southern Medical University, Foshan, China, 528305

*Corresponding author: keyuli2024@163.com

Abstract. This article mainly studies the problem of automatic pricing and replenishment decisions for vegetable products in supermarkets. A grey correlation model and ARIMA time series model are comprehensively established, and based on clustering analysis and regression analysis methods, the distribution patterns of sales volume for each vegetable category and item over time, as well as the relationship with cost plus pricing and mutual correlation, are analyzed to make reasonable predictions for future sales volume. In order to make decisions on product pricing, sales combination, and replenishment, and maximize supermarket profits, the data is preprocessed through normal test and quartile method. Through screening and sorting analysis, the distribution patterns of sales volume for each vegetable category and item over time are obtained; In order to analyze the relationship between different categories of vegetables and between different single products, a grey correlation model is established to determine the analysis sequence, and the correlation coefficients between each category and each single product are calculated to analyze the mutual relationship. And combined with the hierarchical clustering analysis method, the vegetable categories were divided into three sales types: high, medium, and low, expanding the relationship between categories. At the same time, a periodic chart was drawn to determine the periodicity of sales volume distribution in each category. Then, statistical analysis was conducted on the data to determine the linear correlation between sales volume and cost plus pricing. Based on this, a linear regression model was established to analyze the linear relationship between sales volume and cost plus pricing for the six categories; In order to provide the total daily replenishment amount and pricing strategy for the next week, this article establishes a seasonal index prediction model for sales forecasting. Combining the seasonal distribution of each category, the seasonal index is used to predict the sales volume for the next week. Considering the goal of maximizing profits, the target function is listed to comprehensively analyze the replenishment and pricing strategy for the next week.

Keywords: Grey Correlation Analysis, Cluster Analysis, Linear Regression, Prediction Model.

1. Introduction

The In fresh food supermarkets, due to the relatively short shelf life of general vegetable products and the deterioration of product quality with the increase of sales time, supermarkets need to use the "cost plus pricing" method to discount and sell products with transportation damage and poor product quality, and replenish daily based on the historical sales and demand situation of each product without knowing the specific individual product and purchase price[1]. From the demand side perspective, there is often a certain correlation between the sales volume and time of vegetable products; From the supply side perspective, the variety of vegetables available is relatively abundant from April to October, and there are certain limitations to the sales space of supermarkets[2]. Therefore, reliable market demand analysis and reasonable sales mix analysis are needed to make replenishment and pricing decisions for supermarkets, Firstly, preprocess the data by listing a matrix to eliminate outliers and extreme sales values[3]; Then, by screening, the sales volume of each individual product is obtained, and the monthly sales distribution map for each product and category is analyzed and drawn to obtain the distribution pattern; Using the grey relational analysis method again, the degree of correlation between each individual product and each category is analyzed[4,5]. Based on different levels of sales volume, vegetable categories are divided into three categories: low sales, medium sales, and high sales through cluster analysis. The distribution pattern of periodic changes is determined based on the sales time. Then, conduct statistical analysis on the average, maximum, and minimum

values of the total sales volume and cost markup pricing to determine their changes; Further fitting is performed using regression analysis method, with cost plus pricing as the independent variable and total sales of vegetable categories as the dependent variable. A regression equation is established with total sales=Kx cost plus pricing+b. The relationship between total sales of various vegetable categories and cost plus pricing is analyzed[6].

2. Research on Vegetable Categories Based on Grey Relational Analysis

2.1. Grey correlation analysis process

The analysis sequence of the model in this article can be divided into reference sequence and comparative sequence. Reference sequence is a data sequence that reflects the main behavioral characteristics of the studied system, while comparative sequence is a data sequence composed of factors that will affect the behavioral characteristics of the system. Set the sales volume of various vegetable categories at different times as the reference sequence (X0), and set the sales volume of single vegetable products as the comparison sequence (Xi) according to the quarterly distribution.

$$X_0 = \{X_0(1), X_0(2), X_0(3), X_0(4), X_0(5), \cdots\cdots, X_0(n)\} \tag{1}$$

$$X_i = \{X_i(1), X_i(2), X_i(3), \ldots\ldots X_i(n), i = 1,2,3, \ldots\ldots, m\} \tag{2}$$

Due to the different units of influencing factors and significant differences, it is not convenient for subsequent comparative analysis. Therefore, the collected data is subjected to dimensionless processing[7]. The common methods for dimensionless normalization include initialization and averaging, which are commonly used in economic sequences. This article uses the method of initialization to perform dimensionless processing on the data of vegetable single products and vegetable categories.

$$X_i' = \frac{x_i}{x_i(1)} = \left(x_i'(1), x_i'^{(2)}, x_i'^{(3)}, \ldots\ldots, x_i'^{(n)}\right), i = 0,1,2, \ldots\ldots, m \tag{3}$$

Among them, the reference sequence is X'0 (k), the comparison sequence is Xi '(k), k=1,2, N. Differencing sequenceΔ0i (k). Calculate the absolute value of the difference between annual housing prices and other corresponding influencing factors for each year.

$$\Delta_{0i}(k) = |x_0'(k) - x_i'(k)|, i = 0,1,2 \ldots\ldots, n \tag{4}$$

Find the extremum.

$$\Delta_{\max} = \max_i\{\max_k[\Delta_{0i}(k)]\} \tag{5}$$

$$\Delta_{\min} = \min_i\{\min_k[\Delta_{0i}(k)]\} \tag{6}$$

Among them, Δ Max, Δ Min is the maximum and minimum value in each time period, usually Δ Min=0.

Calculate the correlation coefficient

$$\gamma_{0i}(k) = \frac{\Delta_{min} + \rho_{\Delta max}}{\Delta_{0i}(k) + \rho_{\Delta max}} \tag{7}$$

Among them, γ 0i (k) is the correlation coefficient at time t=k, ρ For the resolution coefficient, ρ ∈ (0,1), this article takes ρ= 0.5. Firstly, filter and process the data for each category of vegetables based on their sales volume in different quarters. The filtered data is shown in Table 1.

Table 1 Sales of Vegetables by Different Quarters (Unit: kg)

	florescent vegetables	Flower leaf class	Chili peppers	Solanaceae	edible fungi	Aquatic rhizomes
Q3 2022	5434.072	23568.74	10630.103	1301.403	5918.81	5958.415
Q4 2022	4009.779	21955.56	11315.49	623.969	10461.574	5131.952
Q1 2023	3417.192	18399.38	12752.495	1472.014	9844.323	4060.141
Q2 2023	2074.555	15031.12	8741.345	2120.821	6329.136	1472.82

This article regards cauliflower as a reference sequence and calculates the correlation coefficients between flowering, chili, eggplant, edible mushrooms, and aquatic rhizomes, respectively.

Similarly, analyzing the correlation between other vegetable categories using a grey correlation model can yield correlation coefficients of -0.1, 0.37, 0.23, and 0.30 for the categories of flowers and leaves, chili peppers, eggplants, edible fungi, and aquatic rhizomes, respectively; The correlation coefficients between chili peppers, eggplants, edible fungi, and aquatic rhizomes are 0.59, -0.11, and -0.03, respectively; The correlation coefficients between eggplants, edible fungi, and aquatic rhizomes are -0.53 and -0.80, respectively; The correlation coefficient between edible fungi and aquatic rhizomes is -0.47; There is a strong positive correlation between chili peppers and eggplants, while eggplants have a strong negative correlation with edible fungi and aquatic rhizomes[8].

According to the hierarchical clustering analysis algorithm, the analysis of vegetable categories and sales volume, as well as the existence of temporal periodicity, can be concluded that flower and leaf vegetables belong to the high sales category, chili peppers and edible fungi belong to the medium sales category, and cauliflower, eggplant, and aquatic rhizomes belong to the low sales category. Use drawing software to plot its distribution pattern over time as shown in Fig 1.

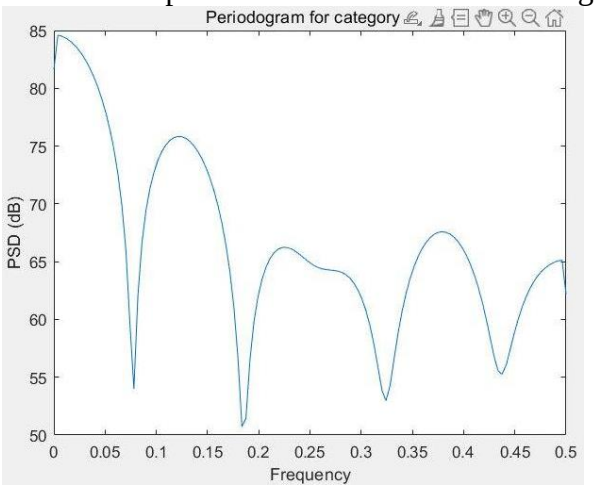

Fig 1 Periodic chart

Periodic chart is a method of representing time series data in the frequency domain, which can be used to detect the periodic or cyclic patterns of data. According to the peak phenomenon in the periodic chart shown in Fig 1, it can be inferred that sales undergo periodic changes over time.

3. The relationship between the total sales volume of vegetable categories and cost pricing

Firstly, in order to analyze the relationship between the total sales volume and cost plus pricing of various vegetable categories, statistical analysis was conducted on the total sales volume and cost plus pricing[9]. Then, this article takes cost plus pricing as the independent variable and the total sales volume of vegetable categories as the dependent variable, establishes a regression model, and fits the regression equation.

Secondly, in order to provide the replenishment volume and pricing strategy for each vegetable category in the next week, it is necessary to predict the sales volume of each vegetable category in the next week. This article uses the ARIMA time series model for prediction analysis. To maximize the profits of supermarkets, list the objective function:

$$W_i = e_j \times D_j \times (1 - \alpha) - c_j \times b_j \tag{8}$$

The considered loss rate is based on the data of each individual product, and the weighted loss rate of each vegetable category is calculated based on the sales volume and loss rate of each product to represent the loss rate of a vegetable category. Combine constraint conditions for target prediction and planning.

3.1. Establishment and Solution of Linear Regression Model

By analyzing the data on the total sales volume and cost markup pricing of various vegetable categories, a scatter plot is obtained as shown in Fig 2, and a linear regression model is established to obtain the regression equation.

Cauliflower:

$$y = -0.1333x + 15.645 \tag{9}$$

Flowers and leaves:

$$y = -32.827x + 311.13 \tag{10}$$

Chili peppers:

$$y = -28335x + 99.915 \tag{11}$$

Solanaceae:

$$y = -3.2644x + 45.739 \tag{12}$$

Edible mushrooms:

$$y = -2.5245x + 74.221 \tag{13}$$

Aquatic rhizomes:

$$y = -1.2682x + 40.334 \tag{14}$$

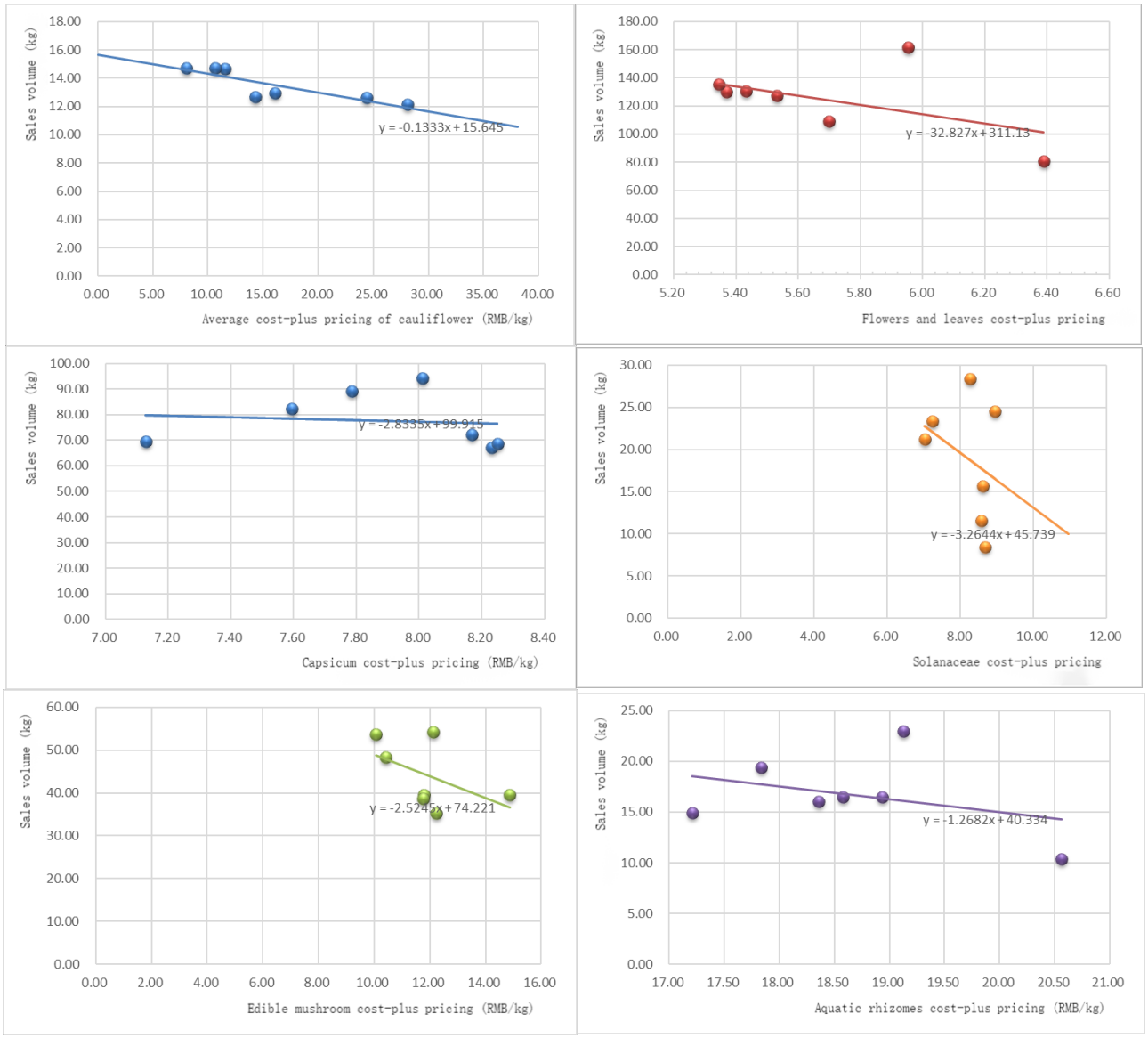

Fig 2 Scatter plot of total sales and cost markup pricing for vegetable categories

3.2. Establishment and Solution of Seasonal Index Prediction Model

In order to predict the sales and pricing of various vegetable categories in the next week more reasonably, a seasonal index prediction model is established. The seasonal index prediction method, also known as the Winston method, is based on the actual values of each period in history[10]. Firstly, a prediction model is established and the trend values of each period in history are obtained. Then, the seasonal index is calculated by dividing the actual values by the trend values. If there are two or more fluctuation cycles, it is necessary to calculate the average seasonal index for the same period and use it to correct the predicted value.

A matrix of 12 rows and 6 columns was created based on the quarterly sales distribution data of various vegetable categories in the past. Each row represents a quarter, and each column represents a vegetable category. The element value represents the sales volume of that category for that quarter. Calculate the proportion of sales volume for each category in each quarter to the total sales volume for that quarter, and predict the sales volume for the next 7 days using the seasonal index of the most recent quarter, as shown in Table 2.

Table 2 Sales Forecast for Each Vegetable Category in the Next Week

Vegetable categories	July 1st	July 2st	July 3st	July 4st	July 5st	July 6st	July 7st
florescent vegetables	256.4	613.5	582.2	540.7	869.9	264.8	318.1
Flower leaf class	119.2	939.8	645.6	479.5	639.3	544.7	647.3
Chili peppers	543.9	721	522.5	993.7	218.7	105.8	1090.7
Solanaceae	63.6	404.6	448.4	365..8	763.5	627.9	772
Edible fungi	932.9	972.7	192.0	138.9	696.3	93.8	525.4
Aquatic rhizomes	530.3	861.1	484.9	393.5	671.4	741.3	520.1

4. Conclusions

The Firstly, this article preprocesses and removes return data, as well as outliers and extreme sales values of each individual product. Then, by filtering, the sales volume of each individual product is obtained, and a monthly sales distribution map for each product and category is analyzed and drawn to obtain the distribution pattern.

Secondly, in order to analyze the degree of correlation between individual products and categories, it is considered that in the field of applied mathematics, grey systems represent systems with incomplete information cognition. Among them, grey correlation analysis method is a commonly used research method. In the process of system development, the degree of correlation between elements is measured based on the similarity of development trends between elements in the system. For this purpose, this article establishes a grey correlation analysis model to analyze the degree of correlation between each single product and each category. Based on different levels of sales volume, vegetable categories and single products are divided into three categories: low sales volume, medium sales volume, and high sales volume through cluster analysis. The distribution pattern of periodic changes is judged based on the sales time.

References

[1] Xu Tian. Research on Product Demand Prediction and Supplier Evaluation of Fresh Dealers [D]. Xi'an University of Technology: 2023

[2] Yang Liqiang, Weng Xiangying, Zhang Jiahui, Fan Chengyi, Zhuang Qianwei. Comprehensive Evaluation and Optimization of Online Sales Platforms for Agricultural Products: Grey Relational Analysis Based on AHP Entropy Weight Method [J]. Strait Science, 2022 (03): 67-72+108

[3] Zhang Lianzhen. Research on Supplier Selection and Evaluation in Equipment Manufacturing Industry Based on Grey Relational Degree [J]. Motor Technology, 2023 (04): 61-64

[4] Mao Hui. Seasonal prediction model for short time series and its application research [D]. Supervisor: Zhang Kui. Wuhan Light Industry University, 2015

[5] Xing Mian. Study on the Combination Grey Neural Network Model for Seasonal Prediction [J]. Systems Engineering Theory and Practice, 2001, (01): 31-35

[6] Bai Yifan. Sales prediction of new energy vehicles based on SARIMA and BP neural network [D]. Xiangtan University, 2019.

[7] Zhang Junkai, Sun Zhifeng. Sales prediction based on optimized grey Markov chain [J]. Modern Manufacturing Engineering, 2019, (04): 7-13

[8] Zhang Xing, He Lili, Zheng Junhong. Research on Sales Prediction Based on Grey Relational Analysis and XGBoost Model [J]. Software Guide, 2020,19 (09): 6-10

[9] Yasaman E, Hassanzadeh S A, Guoqing Z, et al. Time-series forecasting of seasonal items sales using machine learning–A comparative analysis [J]. International Journal of Information Management Data Insights, 2022,2(1):

[10] Zhang Yu, Xu Yiming, Zhang Yan et al. A new K-means clustering algorithm based multiple linear regression prediction model for substation line loss rate [J]. Journal of Electric Power Science and Technology, 2021, 36 (05): 179-186