Research on pricing and replenishment based on optimal programming model

Chaolan Chen *, Xi Wu

Department of Civil Engineering, Sichuan University of Science & Engineering, Zigong, China, 643002

* Corresponding Author Email: chacha0031203@163.com

Abstract. This paper studies the law and correlatoin of different categories of vegetable commodities, and makes the total amount of replenishment and pricing strategy of each vegetable category in the next week, so as to maximize the profit of supermarket. Firstly, the correlatoin between different categories of vegetable products was analyzed based on Pearson correlatoin analysis. The results showed that flower-leaf, cauliflower, pepper had strong correlatoin with edible fungi, while aquatic rhizome and solanaceae had relatively low correlatoin with other categories of vegetables. Then, the monthly and daily sales volume of each category and single vegetable were counted, and the monthly, weekly and daily distribution function of each category and single vegetable was obtained by fitting with Python software using the least square method. The vegetable category cluster analysis based on systematic clustering was used to cluster 6 categories of vegetables, and the 6 categories of vegetables were divided into 3 categories. Then the least square method was used to fit the processed data, and the fitting function expression between the sales volume and the cost plus pricing of each category of vegetables was obtained. The replenishment quantity decision model based on time series was established, and the daily replenishment quantity decision of each vegetable category in the next week was obtained. Finally, according to the correlatoin between sales volume and price, the optimal pricing model based on dynamic market demand is established, and the most reasonable pricing is obtained and the optimal replenishment decision is made to maximize the profit of supermarket.

Keywords: Cluster Analysis, Least Square Method, Time Series, Correlation Analysis.

1. Introduction

This paper as an indispensable food in residents' life, fresh vegetables can provide rich nutrients for the human body. With the promotion of agricultural product sales in the 14th Five-Year Plan, vegetables are playing an increasingly important role in China's agricultural development. [1-3] The vegetable industry is one of the important livelihood issues, which is closely related to the “vegetable basket" of citizens and the "money bag" of farmers. [4-7] In the field of fresh vegetables, accurate prediction of market demand can effectively avoid the imbalance between supply and demand, determine the optimal replenishment and pricing strategy, and greatly improve market sales profits. As far as the main body of enterprise management is concerned, the reasonable prediction of the unknown market demand can achieve scientific production according to demand and quantity in the production process, making the market supply and demand correlatoin relatively stable, and reducing the circulation cost of fresh vegetables. [8] Domestic scholars put forward a forecasting model of agricultural product price in view of the fact that the price of agricultural products is greatly affected by the short-term market, and carried out an example calculation to verify the accuracy of the model. [9-10] At present, the work of change point detection and time series segmentation of linear system based on Bayesian estimation has been done abroad, but the mechanism of simultaneous change of time series parameters has not been revealed. The coupled pricing and parameter estimation methods based on Bayes method are used by foreign scholars to model the events in which time series parameters change simultaneously. [11] This paper studies the correlatoin between various vegetable categories, and uses the replenishment decision model based on time series and the optimal pricing model based on dynamic market demand to determine the maximum profit of the supermarket.

2. Explore the correlatoin between vegetable categories and their sales distribution law

2.1. Preliminary treatment

First of all, through the analysis of question C datas of the 2023 CUMCM, it is found that five kinds of vegetables, including "local spinach", "lotus root", "Wuhu green pepper (2)", "Wuhu green pepper (part)" and "Cordyceps flower (box) (1)", do not have any sales volume. In the data observation, it is found that there is "return" in the sales type of the data. After analyzing the law of commodity trading, this paper excludes this part of "return" sales data, which cannot reflect the real sales situation of the supermarket, so it is considered to retain this part of sales data. Then, EXCEL function VLOOK is used to merge the data for the next step.

2.2. The correlatoin between each category

Correlatoin analysis refers to the analysis of two or more variable elements with correlatoin, so as to measure the degree of correlatoin between two variable factors.

2.2.1. Explore linear correlatoin

Analyze the data and draw a matrix scatter plot, as shown in Fig 1.

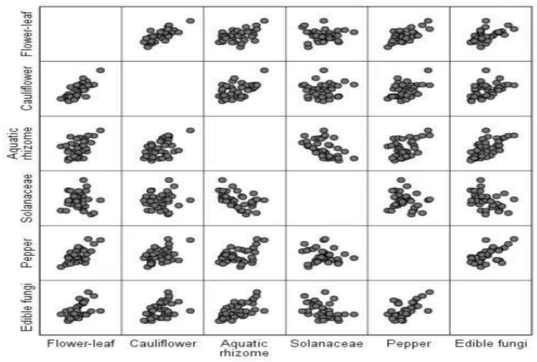

Fig 1. Scatter plot of various vegetable matrices.

It can be seen from Fig.1 that although the data distribution of various vegetables is not absolutely linear, it can basically be regarded as linear correlatoin.

2.2.2. Normality test

First of all, MIN-MAX standardization and sample equalization of the table data were processed, followed by normality test.

(1)S-W test results

Using the S-W test, the significance P-value of all kinds of vegetables is greater than 0.05, the level does not show significance, and the null hypothesis cannot be rejected, so the data meet the normal distribution.

(2) Normality test histogram and P-P plot

The Mosaic class normal test histograms and P-P plots are shown here.

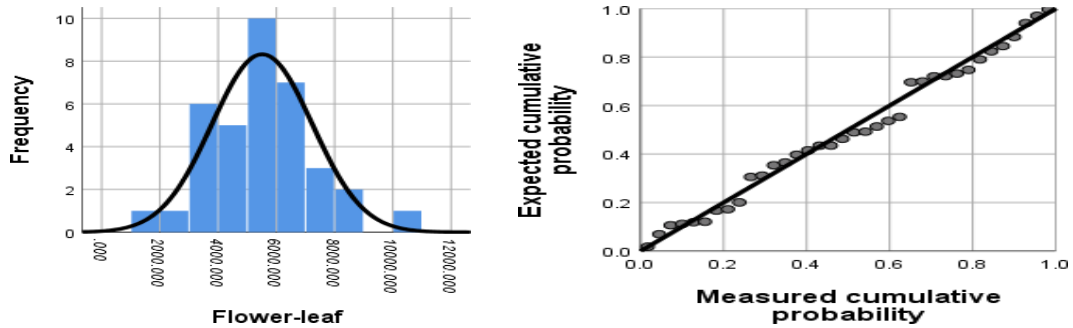

Fig 2. Normality test diagram.

As can be seen from Fig. 2. If the normal graph basically presents a bell shape (high in the middle, low at both ends), it indicates that the data is not absolutely normal, but basically acceptable as a normal distribution. The P-P graph has a high degree of fitting and follows the normal distribution.

2.2.3. Correlatoin analys

The above data analysis results show that the data meets the applicable conditions of Pearson correlatoin analysis. Therefore, Pearson correlatoin analysis was chosen to explore the correlatoin among various vegetable categories.

(1) Correlatoin coefficient heat map

Visualize the correlatoin coefficient and draw the correlatoin coefficient heat map.

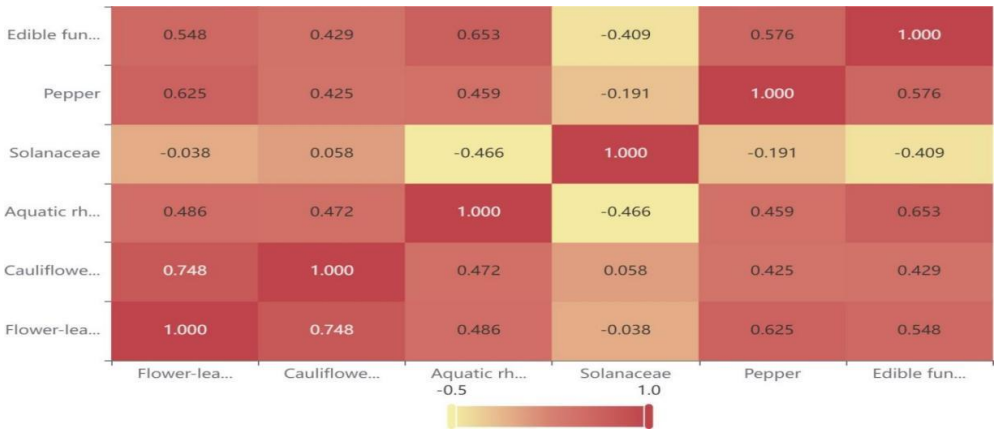

Fig 3. Correlatoin coefficient heat map.

It can be seen from Fig. 3 that there is a strong correlatoin between the four categories of vegetables, namely, flower-leaf, cauliflower, pepper and edible fungi. Because of the general habit of cooking flower-leaf, cauliflower, edible fungi and pepper together in our daily diet, it is easy to buy and sell together; and the correlatoin between aquatic rhizome and solanaceae and other types of vegetables is relatively low.

2.3. Cluster analysis of vegetable category based on systematic clustering

The so-called birds of a feather flock together and people are grouped together. Cluster analysis, which is, based on the characteristics of the research objects, classifies them into categories, so that the differences between individuals of the same category are relatively small and the similarities are relatively large, and the differences between individuals of different categories are large and the similarities are small. Considering the above descriptive statistical analysis of the sales data of 6 categories of vegetables, it is easy to know that these 6 categories of vegetables have a certain correlatoin, so this paper uses the systematic clu**st**er analysis method to analyze the data.

2.3.1. Confirm the optimal distance formula

When analyzing the sales volume of various vegetables by cluster analysis, the optimal distance formula should be determined first. Common distance formulas are absolute distance, Euclidean distance, Minkowski distance, Chebyshev distance, Mahalanobis distance and so on. Among them, Euclidean distance is the most common distance calculation method, which has the advantages of simple and intuitive, so this paper uses the Euclidean distance formula. In mathematics, the Euclidean distance or Euclidean metric is the "ordinary" (i.e., straight line) distance between two points in Euclidean space. Using this distance, the Euclidean space becomes the metric space, and the associated norm is called the Euclidean norm. Its expression is as follows.

$$d_{i,j} = \sqrt{\sum_{k=1}^{m} (x_{i,k} - x_{j,k})^2}, \quad (1)$$

Where, $d_{i,j}$ represents Euclidean distance; x stands for sample and i, j, k stands for subscript.

2.3.2. Perform a systematic cluster analysis

First of all, the original data is normalized to avoid large errors caused by dimension and other influences. Use a z-score

The normalized method processes the data.

$$X = \frac{x - \mu}{\delta}, \tag{2}$$

Where, X represents the sample data after normalization; x represents the observed sample; μ represents the population mean; δ represents the population standard deviation.

SPSS software was used for systematic cluster analysis of the normalized monthly sales data of vegetables. According to the results of systematic cluster analysis, as shown in Table.1, the6 categories of vegetables can be roughly divided into three categories, and the abnormal points of all kinds of data are roughly the same.

Table 1. Statistical description results.

Category	Including vedetable categories
Category1	Cauliflower, Aquaticrhizome, Solanaceae
Category2	Pepper, Edible fungi
Category3	Flower-leaf

2.3.3. Visualize cluster analysis

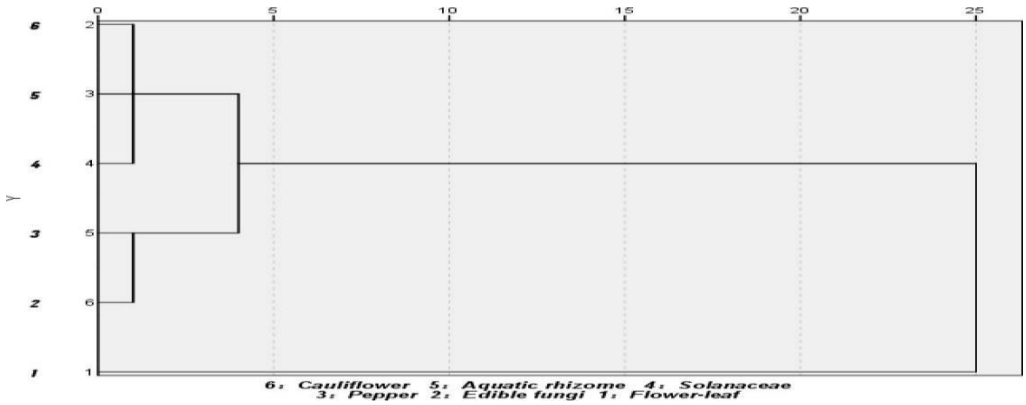

Fig 4. Pedigree diagram using Ward joins.

From Fig. 4 we can see that it is most reasonable to divide vegetables into three categories. Among them, cauliflower, aquatic rhizome and solanaceae are divided into one category, which shows that these three categories of vegetables have a strong mutual correlatoin, while the correlatoin between pepper and edible fungi is more intense, compared with other categories of vegetables, flower-leaf are very unique, systematic clustering method will be it as a single category.

3. Exploring the correlation between sales and cost and constructing a replenishment model

3.1. Correlatoin between sales volume and pricing based on least square method

The least squares method finds the best function match of the data by minimizing the sum of squares of the errors. Using least squares method, the papper can quickly find unknown data and fit the data even if the sum of squares of the errors between the obtained data and the actual data is

minimized. In this paper, the least square method is used to fit the monthly sales volume and the monthly average unit price data, and the correlatoin between the sales volume and the cost-plus pricing of each category of vegetables is obtained.

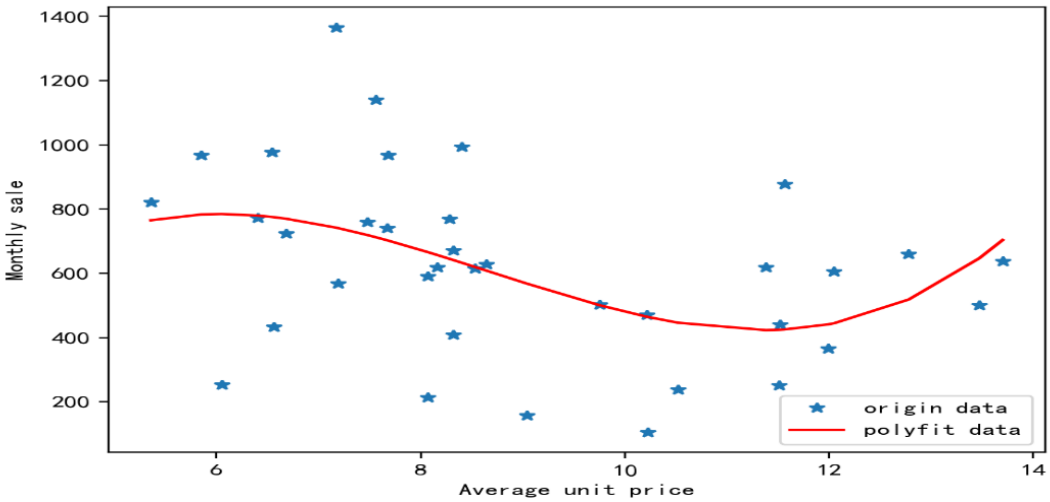

Fig 5. Solanaceae curve fitting.

Solanaceae fitting function expression.

$$Z = 4.871x^3 - 127x^2 + 1001x - 1700, \quad (3)$$

It can be seen from Fig. 5 that some vegetables fit well, such as edible fungi and solanaceae. It can be seen from the side that there is a good correlatoin between the sales volume and pricing of these two types of vegetables. However, the correlatoin between sales volume and pricing of Mosaic, cauliflower and pepper is not strong, and their curve fitting is slightly inferior. As for the reasons, in combination with their vegetable characteristics and sales volume, this paper shows that compared with other categories, Mosaic, cauliflower and pepper vegetables account for a larger proportion in the transaction circulation of supermarket, and their transaction frequency is higher. Therefore, the correlatoin between their sales volume and corresponding unit price is difficult to fit.

3.2. Replenishment volume decision model based on time series

Observe the data, vegetable sales show time series distribution. In this paper, the time series model can make full use of time series data and grasp the dynamic determination of model parameters. The sales data of cauliflower vegetables have an upward trend, and there are obvious periodic fluctuations.

As an important part of modern econometrics, time series analysis is widely used in the research of economic, commercial and social issues. It plays an important role in index prediction and is an important method to study the dynamic and periodic characteristics of statistical indicators and their correlatoin.

a. ADF test results, as shown in Table.2.

Table 2. ADF test.

Variable	Differential order	t	P	Critical value AIC		
				1%	5%	10%
Cauliflower	0	-3.053	0.030**	121.654	-3.689	-2.972	-2.625
	1	-4.079	0.001***	124.291	-3.711	-2.981	-2.63
	2	-4.242	0.001***	123.67	-3.753	-2.998	-2.639

b. ACF and PACF

Order the time series model through ACF and PACF, build the time series model ARIMA (0, 0, and 1)

c. White noise test model parameter table, as shown in Table.3.

Table 3. Model parameter table.

ARIMAModel (0,0,1) Test		
Term	**Character**	**Value**
Sample size	N	30
QStatistic	Q6(PValue)	0.161(0.688)
	Q12(PValue)	4.765(0.574)
	Q18(PValue)	8.793(0.721)
	Q24(PValue)	14.416(0.702)
	Q30(PValue)	25.099(0.400)

For the test results of Q statistics, the P values are all greater than 0.1, indicating that the model residuals conform to white noise

d. Combined with the predicted value of the loss rate correction

According to the time series analysis of cauliflower vegetable data, we have preliminarily obtained the predicted value from July 1 to 7, 2023. Considering that the impact of unknown factors such as climate and means of transportation during vegetable transportation, the forecast of sales volume needs to fully consider the loss rate. Therefore, we define the final forecast replenishment volume as.

$$b_i = \frac{b}{1-r_i'}, \tag{4}$$

Where, b_i represents the final planned replenishment quantity; b stands for preliminary forecast replenishment volume; r' stands for attrition rate.

3.3. The optimal pricing model based on dynamic market demand

3.3.1. Quantify the index and determine the objective function

Taking into account the market demand law of supermarket trading, the price elasticity of demand in economics is introduced, which means that the relative change of commodity demand in a certain period of time responds to the relative change of commodity price. When calculating income, the price elasticity of demand is the key point that must be taken into account. The formula is as follows.

$$E_d = -\frac{dQ}{dP} * \frac{P}{Q} \approx -\frac{\Delta Q / Q}{\Delta P / P}, \tag{5}$$

Where, E_d stands for price elasticity coefficient; P tands for commodity price; Q stands for quantity demanded; dQ stands for change in demand; dP represents the value of the price change.

According to the statistical data, the elasticity coefficient of residents' emand for vegetables, grain and other commodities should be between [0, 1]. In order to ensure the authenticity of the calculation results and reflect the dynamic demand of the market, this paper takes 0.5 as the elasticity coefficient of market demand to make the results more real and reliable.

(1) Cost plus pricing model

The basic idea of cost-plus pricing is that the target price of the product = cost base + percentage. Its basic formula is.

$$\begin{cases} p = c'(1+r) \\ c' = \dfrac{c}{j} \\ r = \dfrac{p-p'}{p'} \times 100\% \end{cases}, \tag{6}$$

Where, p stands for pricing; c' stands for unit cost; c stands for total cost; j stands for quantity purchased; p' stands for purchase price; r is for profit margin.

(2) Explore the correlatoin between daily average pricing of a category and daily sales volume

Establish reasonable pricing for each category from July 1 to 7, 2023 to maximize revenue for the supermarket. Due to the time particularity from July 1 to 7, 2023, we select the historical sales volume, wholesale price, sales pricing and other data from July 1 to 7, 2022 to explore the correlatoin between pricing and sales volume of 6 categories.

$$S_i = f_i(p_j), \tag{7}$$

Where, S_i represents the relationship coefficient between average pricing and sales volume; f_i stands for relational expression; p_j stands for average pricing for different categories.

3.3.2. Determine the constraints

(1) Attrition rate

According to the information consulted, for the commodities that suffer from the loss, due to its poor appearance, poor taste and other reasons, the supermarket usually discounts the part of vegetables to reduce the loss. However, the loss rate of different types of vegetables is different, and a reasonable loss rate can reduce the loss of revenue of the supermarket.

$$r' \in [a, b], \tag{8}$$

Where, a represents the minimum loss rate of the category of vegetables; b represents the maximum rate of consumption of vegetables of this category.

(2) Profit rate

The profit margin of vegetables according to the season and production, vegetables are relatively short of winter, some varieties can reach more than 60% of the profit, summer is the peak season of vegetable production, the profit is about 30%, generally need to ensure the supply of fresh and clean, fresh and clean dishes can make consumers leave a good impression, so as to attract consumers to bu y. Therefore, the profit margin of vegetables should be maintained in the range of [30%, 60%], the expression is.

$$r = \frac{p - p''}{p''} \times 100\% \in [0.3, 0.6], \tag{9}$$

Where, r stands for profit rate; p Stands for pricing; p'' Represents the purchase price.

3.3.3. Model summary

$$\begin{cases} \max M = \dfrac{\sum_{i=1}^{n} S_i \times \sum_{j=1}^{m} (p_j s_j - \dfrac{p_j}{r_j'} c_j') Ed}{\sum_{i=1}^{n} s_i'} \\ s.t. \begin{cases} r = \dfrac{p - p'}{p'} \times 100\% \in [0.3, 0.6] \\ r' \in [a, b] \end{cases} \end{cases}, \tag{10}$$

Where, S_i represents the relationship coefficient between average pricing and sales volume; p_j represents the average price of a product; s_j represents the sales volume of a single product; r' stands for item loss rate; c' stands for unit cost of a product; s' stands for unit sales; Ed stands for elasticity of demand.

4. Conclusions

This paper analyzes the sales of different categories of vegetable products, and explores the correlatoin between vegetable categories based on Pearson correlatoin analysis. It can be seen that the correlatoin between flower-leaf, cauliflower, pepper and edible fungi is strong, while the correlatoin between aquatic rhizome and solanaceae and other categories of vegetables is relatively low. Through the vegetable category cluster analysis based on systematic clustering, the cluster classification of 6 categories of vegetables makes the exploration more scientific and valuable. Cauliflower, aquatic rhizoid and solanaceae are divided into one category, pepper and edible fungi fungi are one category, and flower-leaf is one category. Using time series analysis to predict the quantity of replenishment, in line with the distribution law of data changes with time, using the time series-based replenishment quantity decision model and the optimal pricing model based on dynamic market demand, to determine the daily total quantity of replenishment and pricing strategy of each vegetable category in the next week, so as to maximize the profit of supermarket.

References

[1] Cui Yunhao. Research on forecast of fresh vegetable sales based on CNN-PSO- LSTM combination model [D]. Anhui University of Agriculture Xue, 2023.

[2] Wang Zhangzhang. Research and Implementation of Price prediction Model based on Machine Learning [D]. Chang 'a university, 2019.

[3] Li Lian. Research on Cross-variety price transfer in bulk vegetable Market [D]. Huazhong Agricultural University, 2023.

[4] Zhang Jixiao. Study on pro-environment behavior of leaf vegetable growers [D]. Henan Agricultural University, 2023.

[5] Yang Haoxu. Correlatoin analysis between Vegetable Price and sales volume -- A case study of rosella [J]. Food Safety Guide, 2018(21).179- 181.

[6] Mao Lisha. Research on Pricing Strategy and Production and Marketing Model of vegetable Wholesale Market from the perspective of supply chain [D]. Central South University of Forestry and Technology, 2023.

[7] Joint Research Group of Chengdu Development and Reform Commission and School of Economics, Sichuan University, Guo Xiaoman, Zhang Rui et al. Research on Optimization of Quantitative Price forecasting and Application of Agricultural products -- A case study of Major agricultural products in Chengdu [J]. Macroeconomic Research, 2021(06).107-116.

[8] Zhang Yingqi, Tang Heng, Xu Guangcai. Analysis on vegetable consumption preference and its influencing factors of urban and rural residents in Beijing [J]. Agricultural Outlook, 2023, 19(6).127 -134.

[9] Lv Wang. Research on the trend forecasting of fresh vegetable sales based on improved SVM [D]. Anhui Agricultural University, 2021.

[10] Zhang Qianqian. Research on vegetable demand forecasting system based on SVM[D]. Beijing Jiaotong University, 2015.

[11] TOKINAGA S, IKEDA Y. Modeling for Events characterizing Simultaneous Change in Time Series Parameters by Using Coupled Pricing and Parameter Estimation[J]. A - Abstracts of IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences (Japanese Edition), 2014, J97-A (7).