Comprehensive Research on Vegetable Market Data Analysis and Strategy Formulation

Wu Qiao [1], Kaijie Yang [2, *], Ziyu Liu [1], Yuxiang Zhao [1], Lei Chen [1], Lianxiang Yang [3]

[1] School of Information Engineering, Xi'an Mingde Institute of Technology, Xi'an, China, 710124

[2] School of Economics and Management, Xi' an Mingde Institute of Technology, Xi' an, China, 710124

[3] School of Management, Xiamen University Tan Kah Kee College, Zhangzhou, China, 363105

* Corresponding Author Email: browmjason00@gmail.com

Abstract. In the current context of heightened public awareness of healthy eating, there is an increasing demand for high-quality fresh food products. This trend has brought fresh food supermarkets into the public spotlight. This study will primarily focus on vegetable products and aim to explore replenishment and pricing strategies for vegetable products under different conditions, to maximize the supermarket's revenue. The study will begin by checking whether the sales volume data for various vegetable categories conform to a normal distribution. After confirming normality, the Pearson correlation coefficient method will be used to analyze the relationships between sales volumes of different vegetable categories. Additionally, an ARIMA time series model will be employed to forecast vegetable sales data for the next seven days, and single-objective optimization will be utilized to obtain the best outcome. Finally, by constructing an optimal pricing and replenishment strategy hierarchy analysis system, decisions related to the sale of various vegetable products will be assessed and formulated.

Keywords: Pearson Correlation Coefficient Method, ARIMA Time Series Model, Single Objective Optimization, Hierarchical Structure Analysis.

1. Introduction

With the development of the times and the improvement of people's living standards, residents' consumption habits and dietary preferences have undergone significant changes[1]. Therefore, this study aims to provide fresh food to residents to the maximum extent and maximize the profits of supermarkets[2][3].

After conducting a K-S single-sample test on the sales data of various vegetable categories, a further S-W test suitable for small samples was used to confirm that all six vegetable categories follow a normal distribution[4]. Subsequently, Pearson correlation coefficient analysis was performed to analyze the relationship between the sales volumes of various vegetable categories[5].

The ARIMA time series model was employed to forecast the sales data of vegetables for the next seven days[6][7]. Then, based on the forecasted sales volume with a variation of ±5% and the relationships between sales volume, replenishment quantity, selling price, purchase price, and the maximum profit for each vegetable category, an optimization model was established to obtain the optimal values[8].

Finally, a comprehensive and efficient hierarchy analysis system for optimal pricing and replenishment strategies was constructed[10]. The construction of this system is based on a comprehensive consideration of key factors in supply chain management, including supplier information, seasonal factors, and inventory quantities. The analysis revealed the correlation between these factors and the pricing and replenishment strategies for vegetable products. Ultimately, decisions related to the sale of relevant vegetables were made based on this correlation.

2. Statistical Testing and Analysis of Vegetable Sales Data

2.1. Distribution Patterns of Sales Volumes for Different Vegetable Categories

The data attachment obtained from the website http://www.mcm.edu.cn/index_cn.html provides detailed transaction records and wholesale prices for three years at a fresh supermarket. Due to the large volume of data, it is necessary to preprocess the data, checking for outliers, missing values, and duplicate data. Based on the results of data preprocessing, 36 sets of sample data were obtained, with sales dates aggregated by month as the unit. Initially, the K-S single-sample test was applied to test the hypothesis of whether sales volumes for different vegetable categories follow a normal distribution pattern. The test results yielded the normality parameters for each vegetable category, as presented in Table 1:

Table 1: Two-Tailed Significance for Various Vegetable Categories under Normal Parameters

Name	Leafy Vegetables	Cauliflower	Aquatic Root and Stem Vegetables	Eggplant /Tomato	Chili Peppers	Edible Mushrooms
Normal parameters	0.200	0.200	0.191	0.200	0.052	0.124

Observing the asymptotic significance of various vegetable categories under normality parameters in Table 1, it is evident that all six categories of vegetables follow a normal distribution pattern under normality parameters. The K-S single-sample test is suitable for cases with a large amount of data. To ensure accuracy, a further S-W test, suitable for small samples, was conducted, and the test results are presented in Table 2:

Table 2: Shapiro-Wilk Test for Various Vegetable Categories (Partial Data)

Name	Leafy Vegetables	Cauliflower	Aquatic rhizomes	Eggplant /Tomato	Chili Peppers	Edible Mushrooms
Distinctiveness	0.854	0.140	0.090	0.724	0.063	0.068

Based on the test results in Table 2, it can be concluded that all six categories of vegetables follow a normal distribution pattern.

2.2. Testing the Distribution Patterns of Sales Volumes for Various Vegetable Categories

To further confirm that all six categories of vegetables adhere to a normal distribution pattern, normal Q-Q plots were generated for the six vegetable categories (see Figure 1).

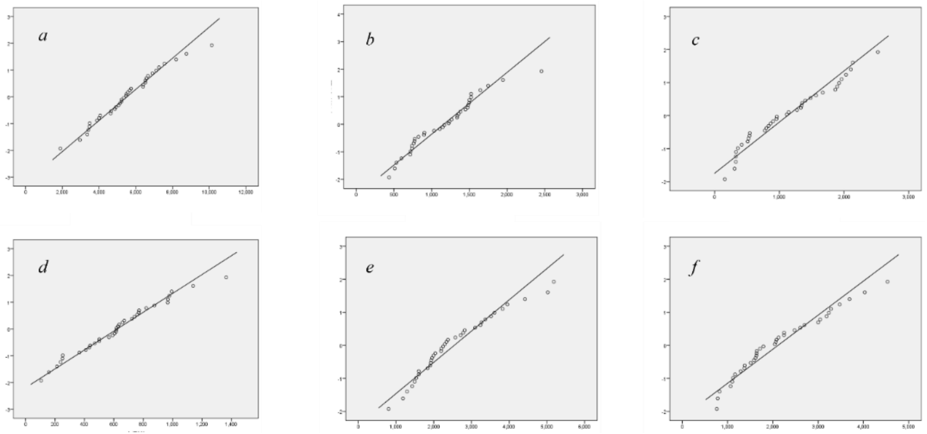

Figure 1: Normal Q-Q Plots for Six Vegetable Categories. The paragraph describes various normal Q-Q plots for different plant categories, including flower leaves, flower heads, aquatic stems, eggplants, chili peppers, and edible fungi. These plots are labeled as a through f, representing the normal distribution of quantile-quantile plots for each plant type.

Based on Figure 1, a scatter plot was created with the quantiles of the actual data for the six vegetable categories on the X-axis and the corresponding quantiles for the expected normal distribution on the Y-axis. It was observed that the measured quantiles closely align with the expected normal quantiles, and the scatter points generally fall along the diagonal line, confirming that the six vegetable categories follow a normal distribution pattern.

2.3. Interrelationships of Sales Volumes for Various Vegetable Categories

The distribution patterns for sales volumes of various vegetable categories all follow a normal distribution, and all six vegetable categories are considered interval-level variables, making them suitable for Pearson correlation coefficient analysis. Some of the Pearson correlation coefficients for the vegetable categories are presented in Table 3:

Table 3: Pearson Correlation Coefficients for Vegetable Categories (Partial Data)

Vegetable categories	Pearson correlation	Vegetable categories	Pearson correlation
Mosaic and cauliflower	0.747	Nightshades and aquatic rhizomes	0.653
Cauliflower and aquatic rhizomes	0.422	Chili peppers and mosaic leaves	0.624
Aquatic rhizomes and edible fungi	-0.466	Edible mushrooms and chili peppers	0.576

Analyzing Table 3, it can be observed that there is a strong positive correlation between Leafy Vegetables and Cauliflower. Cauliflower exhibits a relatively strong positive correlation with Aquatic Root and Stem Vegetables, Chili Peppers with Leafy Vegetables, Eggplant/Tomato with Aquatic Root and Stem Vegetables, and Edible Mushrooms with Chili Peppers. Additionally, there is a relatively strong negative correlation between Aquatic roots and Stem Vegetables and Edible Mushrooms.

3. Vegetable Sales Forecasting and Strategies

3.1. Using the ARIMA model to forecast vegetable sales

The Cost-Plus Pricing method refers to setting the product price by adding a profit margin to the product cost, expressed in the formula:

$$P_i = C_i\left(1+\alpha_i\right) \tag{1}$$

Where P_i represents both the cost-plus pricing and the selling price of various vegetable categories, C_i is the wholesale price for each vegetable category, a_i denotes the profit margin for each vegetable category, S_i represents the total sales volume for various vegetable categories, and i=1,2,⋯,6 represents the six vegetable categories.

The ARIMA model combines the characteristics of autoregression and moving averages, allowing for more accurate predictions of time series data. The ARIMA model results for various vegetable categories are presented in Table 4:

Table 4: ARIMA Models for Various Vegetable Categories

Vegetable categories	ARIMA(d,p,q)	R^2	Distinctiveness
Mosaic and leafy	ARIMA(3,1,2)	0.967	0.064
Cauliflower	ARIMA(2,1,2)	0.996	0.920
Aquatic rhizomes	ARIMA(3,1,3)	0.996	0.056
Eggplant/Tomato	ARIMA(0,1,0)	0.997	0.151
Chili Peppers	ARIMA(1,1,1)	0.981	0.717
Edible Mushrooms	ARIMA(3,1,1)	0.970	0.022

In the case of Leafy Vegetables, as shown in Table 4, we observe an R-squared value of 0.967 and a significance level of P = 0.064, which is greater than 0.05. Therefore, it reflects that the fit of the data for the Chili Peppers category is good, which is the expected outcome. Calculating the ARIMA model statistics for the remaining vegetable categories, it can be concluded that the fit of the models for the other categories is also acceptable. However, further residual analysis is required. Taking the Chili Peppers category as an example, an examination was conducted using ACF and PACF plots, as shown in Figure 2 and Figure 3:

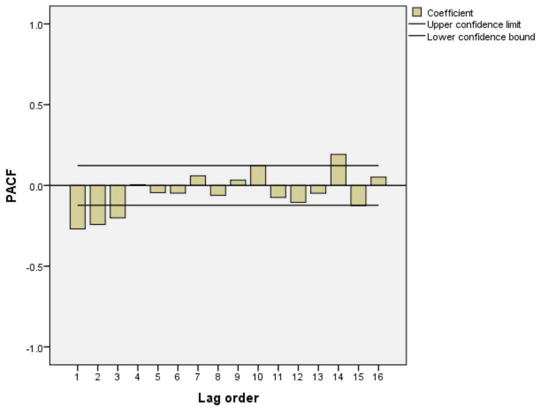

Figure 2: ACF Plot for Chili Peppers Category Time Series

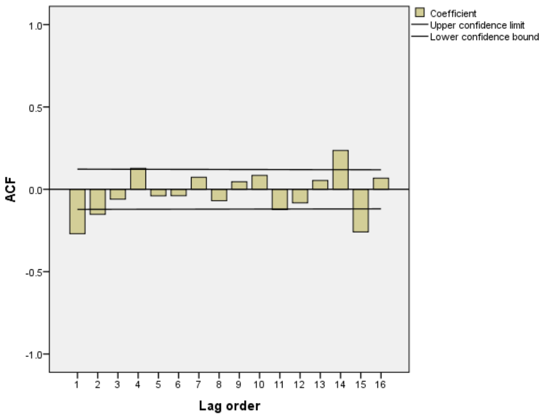

Figure 3: PACF Plot for Chili Peppers Category Time Series

After performing one differencing operation, we obtain Figure 2 and Figure 3. From the information in the above figures, it is observed that the coefficients of autocorrelation and partial autocorrelation are mostly significant and fall within the upper and lower confidence intervals. Based on Figure 2 and Figure 3, we determine the lag orders for autocorrelation and partial autocorrelation to be 3 and 2, respectively, for the Leafy Vegetables category. Therefore, the model for Leafy Vegetables is ARIMA(3,1,2). Similarly, we can derive the ARIMA models for other vegetable categories. Finally, based on the ARIMA models for various vegetable categories, we obtain the forecasted values for replenishment quantity, selling price, and wholesale price for the next 7 days (as shown in Table 6, taking Leafy Vegetables as an example). These values are used as constraints for better establishing a single-objective optimization model.

3.2. Establishing a Single-Objective Optimization Model

The objective function of the optimization model is to maximize the supermarket's revenue. One of the constraints is to use the sales volume forecast values with a margin of plus or minus 5%. Additionally, based on the correlation analysis between vegetable categories and using the curve estimation method, we calculate the functional relationships between various vegetable categories. The best-fitting result is a cubic function. Furthermore, another constraint for the optimization model is based on the relationship between the total replenishment quantity of vegetable products and the selling price, as well as the relationship between the selling price and the purchase price. The respective fitting goodness of mosaic-cauliflower, cauliflower-aquatic rhizomes, aquatic rhizomes-edible fungi, nightshades-aquatic rhizomes, chili peppers-mosaic leaves, edible mushrooms-chili peppers are 0.566, 0.374, 0.453, 0.337, 0.437, 0.389.

$$\max Z = \sum_{i=1}^{7}\sum_{j=1}^{6}\left(1-h_j\right)\cdot\left(b_j-k_j\right)\cdot a_{(i,j)} \tag{2}$$

Table 5: Sales Volume Forecast Values for Leafy Vegetables for the Next 7 Days

Vegetable categories	**Day 1**	**Day 2**	**Day 3**	**Day 4**	**Day 5**	**Day 6**	**Day 7**
Leafy Vegetables	149.083	193.42	180.44	115.218	138.974	139.73	139.48

Obtained the daily replenishment quantities and pricing schemes for various vegetable categories from July 1st to July 7th, 2023, as shown in Table 5, Table 6 and Table 7:

Table 6: Daily Replenishment Quantity Strategy for Various Leafy Vegetable Categories for the Upcoming Week

	July 1	**July 2**	**July 3**	**July 4**	**July 5**	**July 6**	**July 7**
Leafy Vegetables	177.56	196.829	189.726	176.596	178.555	171.1	153.048

Table 7: Pricing Strategy for Various Vegetable Categories for the Upcoming Week

	Leafy Vegetables	**Cauli-flower**	**Aquatic rhizomes**	**Eggplant /Tomato**	**Chili Peppers**	**Edible Mushrooms**
price	5.02	11.66	14.04	8.50	7.22	10.38

4. Decision Analysis in Vegetable Sales

Step1: Constructing the Hierarchical Structure System for Optimal Pricing and Replenishment Strategy, as shown in Figure 4:

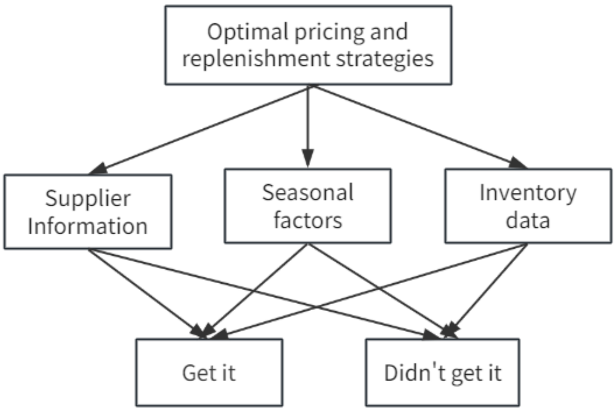

Figure 4: Construction of the Hierarchical Structure for Optimal Pricing and Replenishment Strategy

Step 2: Constructing the Comparison Matrix

The hierarchical structure system can reflect the hierarchical relationships between supplier information, seasonal factors, inventory data, and the target layer. The comparison matrix P is as follows:

$$p=\begin{bmatrix} 1 & 2 & 0.25 \\ 0.5 & 1 & 0.062 \\ 4 & 16 & 1 \end{bmatrix} \tag{3}$$

Step 3: Calculate the m-dimensional vectors for supplier information, seasonal factors, and inventory quantity.

Step 4: Normalize the m-dimensional vectors to obtain the weight vectors.

Step 5: Calculate the maximum eigenvalue.

$$\lambda_{\max}=\frac{1}{n}\sum_{j=1}^{n}\frac{(\phi c)_i}{c_i} \tag{4}$$

Calculate and obtain $\lambda_{\max} = 3.054$.

Step 6: Calculate the final value of the Consistency Index (C.I.).

$$C.I. = \frac{\lambda_{\max} - n}{n-1} = 0.027 \tag{5}$$

According to the R.I. table, find the corresponding R.I. value, which is R.I. = 0.525.

Step 7: Calculate the Consistency Ratio (C.R.).

$$C.R. = \frac{C.I.}{R.I.} = 0.051 < 0.1 \tag{6}$$

Analyze the consistency test. When C.R. < 0.1, it is considered that the judgment matrix P complies with the consistency test. When C.R. > 0.1, the judgment matrix P does not comply with the consistency test. Calculations yield C.R. < 0.1, demonstrating that supplier information, seasonal factors, and inventory quantity are consistent with the replenishment and pricing strategy for vegetable products in practice. The weightings of the influencing factors for the criteria layer are shown in Table 8:

Table 8: Weightings of Influencing Factors in the Criteria Layer

index	Weights
Supplier Information	$\bar{\omega}_1 = 15.919$
Seasonal factors	$\bar{\omega}_2 = 6.372$
The quantity in stock	$\bar{\omega}_3 = 77.709$

According to Table 9, the influencing factors in the criteria layer are ranked in ascending order of importance as follows: Seasonal Factors, Supplier Information, and Inventory Quantity.

Based on the above model, it is evident that supplier information, seasonal factors, and inventory quantity are closely related to the pricing and replenishment strategy for vegetable products. This can help address replenishment and pricing issues effectively.

(1) Businesses can utilize inventory quantity data to effectively monitor stock levels and adjust the prices of vegetable products to stimulate sales and increase revenue.

(2) Seasonal factors influence the demand and timing of fresh products, so businesses need to proactively replenish and adjust pricing strategies.

(3) Consideration of different suppliers' pricing strategies, discount policies, and supply capabilities allows businesses to implement flexible replenishment and pricing strategies.

5. Conclusions

Using the K-S test and S-W test distribution patterns, it was found that all six categories of vegetables follow a normal distribution. Then, through Pearson correlation analysis, it was determined that there is a strong correlation among the sales volumes of various vegetable categories. After using the ARIMA time series and optimizing with single-objective programming, the projected revenue for the next seven days was calculated as 5512 yuan. The daily replenishment quantities for categories such as Leafy Vegetables, Cauliflower, Aquatic Root and Stem Vegetables, Eggplant/Tomato, Chili Peppers, and Edible Mushrooms are approximately 180, 40, 20, 35, 50, and 40 kg, with average prices of 5.016, 11.662, 14.042, 8.498, 7.225, and 10.377 yuan, respectively. Lastly, using the Analytic Hierarchy Process (AHP), it was determined that the influencing factors in the criteria layer, in ascending order of importance, are: Seasonal Factors, Supplier Information, and Inventory Quantity. This confirms the close relationship between these three categories and the pricing and replenishment strategy for vegetable products, ultimately leading to the selection of the optimal replenishment and pricing strategies.

This study provides a comprehensive forecasting method for replenishment and pricing strategies for vegetable products under different conditions, based on ARIMA time series forecasting and single-objective programming optimization. It demonstrates the feasibility of this approach.

References

[1] Yu Han, Wang Jiaqing, Liu Xiaoyu. Analysis of the Development Path of Fresh Agricultural Products E-commerce under the Background of Rural Revitalization[J]. Commercial Modernization, 2023, (24): 26-29.

[2] Shen Qianming. Research Progress Analysis of Organic Vegetable Consumption Based on Literature Metrology[J/OL]. China Vegetables, 1-8 [2023-12-18] https://doi.org/10.16861/j.cnki.zggc.20231215.006.

[3] Continuous Improvement in Consumer Market Recovery[N]. China Information Daily, 2023-12-18 (002).

[4] Xie Zhuang. Optimization Research on Overall Drive Generator Maintenance Scheme Based on Normal Distribution[J]. Transportation Science and Management, 2023, 4(23): 19-21+25.

[5] Guo Wei, Wu Yanxing, Guo Shujin, et al. Correlation Analysis of Quality Traits and Yield of Jin Da Zhi Lu 1 Soybean[J]. Shanxi Agricultural Science, 2023, 51(12): 1388-1394.

[6] Zhao Keyi. Forecast of the Development of the Tertiary Industry in Yunnan Province Based on ARIMA Model[J]. Value Engineering, 2023, 42(33): 44-48.

[7] Yuan Bingxin, Zhang Shuang, Shan Miaomiao, et al. Prediction Analysis of Henan Province's Total Health Expenditure and Its Composition Based on the ARIMA-GM (1,1) Combination Model[J]. Health Soft Science, 2023, 37(12): 64-68.

[8] Liu Longzhe, Liu Gang, Xu Hongpeng, et al. Loss Function Optimization Design for Single-Stage Target Detection[J/OL]. Electric Light and Control, 1-11 [2023-12-18] http://kns.cnki.net/kcms/detail/41.1227.TN.20231115.1522.018.html.

[9] Zhu Xiping. Multi-Objective Optimization Prediction Model for Gas Production in Gas Storage[J/OL]. Oil & Gas Storage and Transportation, 1-10 [2023-12-18] http://kns.cnki.net/kcms/detail/13.1093.TE.20231115.1650.004.html.

[10] Wang Shaosen. Fusion and Progression: A Cultural Hierarchical Analysis of Architecture[J]. Architecture and Culture, 2023, (12): 1.