Agricultural & Forestry Economics and Management (2023)
Clausius Scientific Press, Canada
DOI: 10.23977/agrfem.2023.060310
ISSN 2616-2202 Vol. 6 Num. 3

Automatic Pricing and Replenishment Decision Analysis of Vegetable Products Based on ARIMA Optimization Model

Junyan Li[1,*]

[1]*School of Shipping and Naval Architecture, Chongqing Jiaotong University, Chongqing, China*
[*]*Corresponding author: 17876004089@163.com*

Keywords: Vegetable Commodities, Correlation Analysis, ARIMA Model, Linear Programming Analysis

Abstract: Due to its short shelf life, the pricing and replenishment decision of vegetables in fresh food supermarkets is often the focus of supermarkets. In this paper, we first establish and solve the Pearson correlation model between vegetable categories and sales distribution, and obtain the correlation relationship between each category and each item. Then, we establish the ARIMA prediction model to predict the replenishment quantity and pricing strategy of each vegetable category in the coming week in the fresh food superstore, and finally, we use linear programming to optimize the ARIMA prediction model to solve the automatic pricing and replenishment decision of vegetable items when the maximum revenue is obtained under the demand of the market. Finally, the ARIMA prediction model was optimized using a linear programming method to solve the automatic pricing and replenishment decision of vegetable items when meeting the market demand and maximizing the revenue. The ARIMA optimization model used in this paper can effectively predict and optimize the automatic pricing and replenishment decision of vegetable products, which is of great significance to ensure the normal operation and profit of fresh food superstores.

1. Introduction

Fresh food supermarkets can provide residents with easily accessible fresh fruits, vegetables, meat, seafood and other food products, enriching their shopping choices and providing useful assistance to residents in obtaining a balanced diet and maintaining a healthy lifestyle. The shelf life of various fresh food products in fresh food superstores varies, while the pricing and replenishment decisions of vegetable products are often the focus of superstores due to their shorter shelf life. Fresh food superstores usually replenish their products every day according to the historical sales and demand of each product, so how to establish the correlation between vegetable categories and sales distribution, and predict and optimize the automatic pricing and replenishment decision of vegetable products based on this correlation is of great significance to ensure the normal operation and profit of fresh food superstores [1, 2].

This paper takes the commodity information of each vegetable category of a fresh food superstore this year, the related data of sales and wholesale of each vegetable category, and the loss rate of each single product of each category as the object of study, establishes the Pearson correlation model of vegetable category and sales distribution, and solves the Pearson correlation coefficient of the sales distribution law and the interrelationships of all the single products of cauliflower category with the use of MATLAB and SPSS software, and obtains the Pearson correlation coefficient of each category and each single product. The correlation coefficients between each category and each individual product were obtained. At the same time, the ARIMA prediction optimization model of commodity pricing replenishment was established with the profit of the superstore as the objective function, and the number of displays, the total number of items available for sale, the market demand and the cost balance as the constraints, and the ARIMA prediction model was solved using the linear programming method, and the total number of items replenished at a specific date of the fresh food superstore, and the pricing strategy with the maximum revenue under the satisfaction of the market demand were obtained [3].

2. Model establishment

2.1 Pearson's correlation modeling of vegetable category and sales distribution

According to a fresh food superstore this year, the commodity information of each vegetable category, the sales of each vegetable category and wholesale data can be summarized to obtain the sales of each category in different time periods, as shown in Table 1.

Table 1: Distribution of sales data by vegetable category (unit: kg)

Date	Cauliflower	Philodendron	Capsicum	Eggplant	Edible Mushroom	Aquatic Rhizomes
2020/7/1-2020/12/1	71239.23	32421.37	10759.45	4380.99	11178.80	4302.98
2020/12/1-2021/5/1	71349.00	29030.01	11702.88	3023.19	13837.88	7209.41
2021/5/1-2021/10/1	5278.18	27640.81	9368.01	4221.38	6342.30	3699.28
2021/10/1-2022/3/1	4362.02	16266.75	8941.74	2203.70	8755.83	7077.89
2022/3/1-2022/8/1	3956.04	19490.93	9735.18	3657.46	5231.13	3544.92
2022/8/1-2023/1/1	7940.05	40259.37	19597.28	1357.58	14575.32	9221.34
2023/1/1-2023/6/30	5488.89	33411.69	21484.06	3587.45	16165.44	5525.50
Total	41766.45	198520.97	91588.62	22431.78	76086.72	40581.35

In order to make the data more intuitive to show the distribution of sales data of vegetable categories, you can visualize the above table, which can be obtained as shown in Figure 1 of the total sales volume of different vegetable categories accounted for the share of the map, and from the map it can be seen that the share of each vegetable category in the order of share of the following: foliage > chili peppers > mushrooms > aquatic roots and tubers > foliage class > eggplant class.

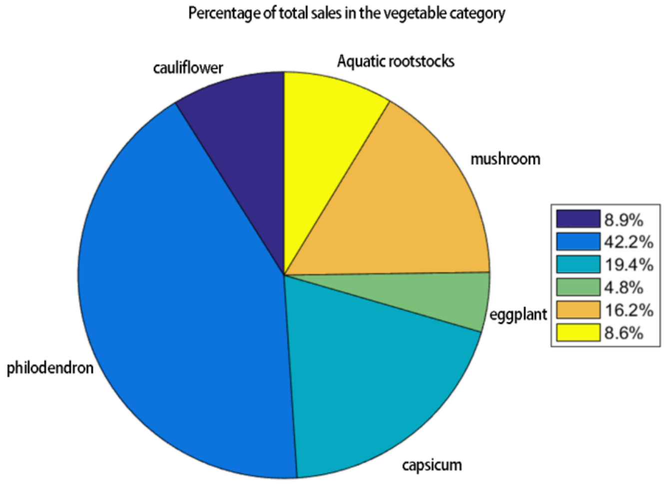

Figure 1: Histogram of total sales volume of different vegetable categories

Due to the large amount of statistical data on single product sales under different categories of vegetables, only the statistical results of the cauliflower category are analyzed in the analysis process of this paper, and the single product sales summary table of the cauliflower category as shown in Table 2 as well as the single product sales chart of the cauliflower category as shown in Figure. 2 are obtained, and the analysis of the charts can be understood that in the single product of the vegetable category, the percentage of the sales share is as follows: broccoli > green peduncle loose flower > zhijiang green peduncle loose flower > purple cabbage.

Table 2: Summary of individual item sales in the cauliflower category

Vegetable Singles	Summary of Sales/kg
Broccoli	27537.23
Purple Cabbage (1)	13.215
Green Peduncle Loose Flower	8393.786
Zhijiang Green Peduncle Loose Flower	5821.571
Purple Cabbage (2)	0.615

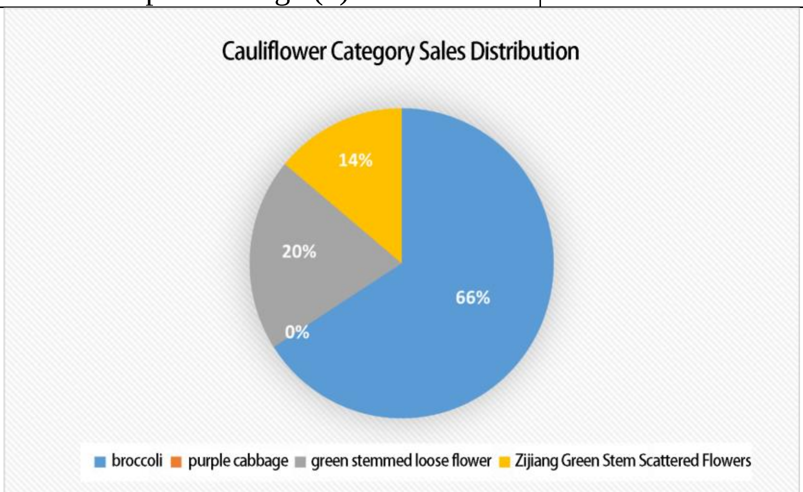

Figure 2: Sales of cauliflower vegetables

In statistics, the Pearson correlation coefficient, also known as PPMCC or PCCS, and often represented by r or Pearson's in articles, is used to measure the correlation (linear correlation) between two variables X and Y, with a value between -1 and 1. In the natural sciences, this coefficient is widely used to analyze the degree of correlation between variables. Considering that the sales volume of each category is quantitative data, Pearson correlation coefficient can be used to analyze the correlation between each category [4].

The Pearson correlation coefficient is calculated as follows:

$$\alpha=\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2}} \tag{1}$$

2.2 ARIMA forecasting modeling for merchandise pricing replenishment

Since the total sales and cost-plus pricing of each vegetable category is a "forecasting + optimization" problem, it is necessary to establish a forecasting model to give the replenishment quantity and pricing strategy for the coming week, and then establish an optimization model to maximize the revenue of the superstore on this basis. Firstly, the time series model is analyzed to determine the smoothness of the series, if it fails to pass, the smoothness treatment is carried out, and if it passes, the autocorrelation coefficient and the average shift coefficient are determined, then, the ARIMA model is established and model optimization is carried out [5]. Finally, the model is evaluated and the data are predicted. The flow chart is shown in Figure 3:

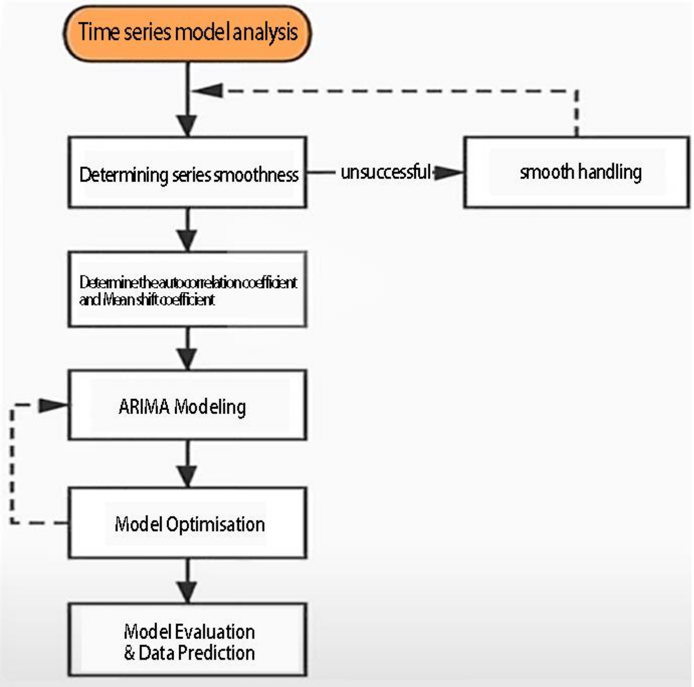

Figure 3: Flowchart of ARIMA prediction optimization model building

Before building the ARIMA model in the time series model, it must satisfy the smoothness requirement and must have autocorrelation, if the autocorrelation coefficient (φ_i) is less than 0.5, it is not suitable.

The formula for the ARIMA model is as follows:

$$Y(t)=c+\phi_1 Y(t-1)+\phi_2 Y(t-2)+\ldots+\phi_p Y(t-p)+\varepsilon t+\theta_1\varepsilon(t-1)+\theta_2\varepsilon(t-2)+\ldots+\theta_q\varepsilon(t-q) \tag{2}$$

where $Y(t)$ is the value of the time series, t is the time point, c is the constant term, $\phi_1,\phi_2,\ldots\phi_p$

is the natural regression coefficient, εt is the disturbance term, $\theta_1, \theta_2, \ldots, \theta_q$ and is the moving average coefficient.

3. Model solving

3.1 Solving Pearson's correlation coefficient based on MATLAB and SPSS

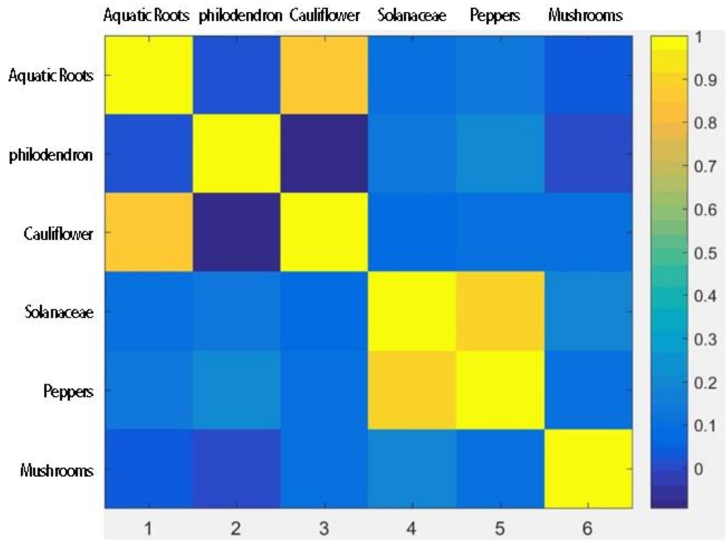

Figure 4: Pearson's correlation coefficient analysis graph

From Figure 4, we can analyze that the correlation between "Foliage" and "Peppers", "Eggplant" and "Edible Mushrooms" is low, which indicates that the sales pattern of "Foliage" may be independent from the sales pattern of other categories.

For the single product statistics in the "cauliflower category", because the data value of purple kale (2) is very small, so it needs to be eliminated, and put the remaining data into the SPSS software to analyze, which can get the correlation between the single product of cauliflower vegetables. And the correlation analysis of the distribution data of sales data of each vegetable category in Table 1, the correlation between vegetable categories can be solved by SPSS software.

SPSS software can get the Pearson correlation coefficient between different vegetables, and the larger value represents the stronger correlation. Firstly, the group analyzed the correlation between cauliflower commodities, and found that except for the highest correlation between broccoli and Zhijiang green stemmed loose flowers, the correlation between the rest of the commodities is very low, which belongs to the weak correlation, and it can be inferred that the degree of the sales of broccoli and Zhijiang green stemmed loose flowers in the superstore is basically the same among cauliflower vegetable commodities. Secondly, taking the most common cauliflower vegetables as an example, it can be seen that the correlation between cauliflower vegetables and leafy vegetables, chili peppers, eggplant vegetables, edible mushrooms, and root vegetables is weak; the correlation with leafy vegetables, chili peppers, and edible mushrooms is stronger; and the correlation between aquatic root vegetables and eggplant vegetables is stronger, and combining with the dietary characteristics of the two regions of Sichuan and Chongqing, it can be seen that leafy vegetables leafy vegetables, chili vegetables, and edible fungus vegetables are sold to the same extent in supermarkets. The fact that leafy vegetables, chili vegetables and edible fungus vegetables are often consumed together proves the accuracy of the analysis results.

3.2 ARIMA predictive model solving

The ARIMA prediction model was solved using MATLAB to give the replenishment quantities and pricing strategies for the fresh produce superstore vegetable items for the coming week as shown in Table 3 and Table 4.

Table 3: Total optimal replenishment for each category (unit: kg)

Date	Cauliflower	Philodendron	Capsicum	Eggplant	Edible Mushroom	Aquatic Rhizomes
July 1st	134.009	21.8901	34.647	57.0863	85.368	66.3805
July 2nd	144.52796	44.36496	24.532	65.445	67.77	73.67
July 3rd	99.2551	47.4775	36.58788	23.456	57.4325	67.88
July 4th	155.5774	37.58543	43.356	64.32	47.77	90.0001
July 5th	110.007	36.5	23.5632	39.455	86.237	65.788
July 6th	68.998	36.47	35.35	34.661	67.001	59.382
July 7th	121.67101	47.3648	26.2326	46.2279	68.2	61.751
Total	834.04547	271.65279	224.26868	330.6512	479.7785	484.8516

Table 4: Optimal Pricing Strategies by Category (unit: yuan)

Date	Cauliflower	Philodendron	Capsicum	Eggplant	Edible Mushroom	Aquatic Rhizomes
July 1st	0.5834	0.5834	0.6441	0.5001	0.5185	0.6534
July 2nd	0.6441	0.5334	0.5941	0.4509	0.5294	0.5701
July 3rd	0.6001	0.5434	0.6041	0.4693	0.4894	0.6305
July 4th	0.5605	0.6034	0.6641	0.5573	0.4685	0.5885
July 5th	0.5294	0.6634	0.7241	0.5273	0.5873	0.5201
July 6th	0.6185	0.5734	0.6341	0.5873	0.5873	0.4569
July 7th	0.6373	0.6334	0.6941	0.4873	0.5685	0.5509
Total	4.1733	4.1338	4.5587	3.5795	3.7489	3.9704

Visual characterization of the above replenishment quantities and pricing strategies leads to the graphs of optimal replenishment totals for each category and optimal pricing strategies for each category as shown in Figures 5 and 6.

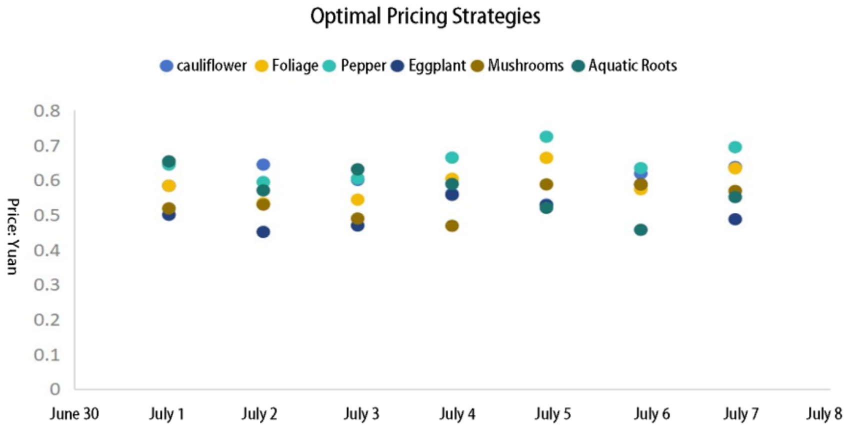

Figure 5: Graph of total optimal replenishment by category

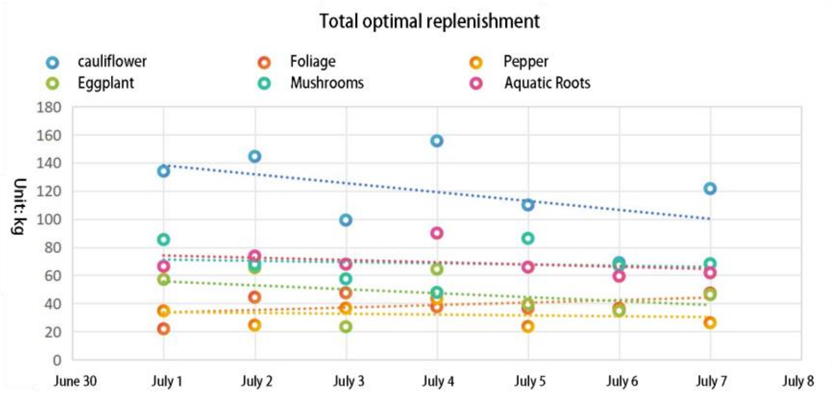

Figure 6: Optimal pricing strategies by category

3.3 Optimization of ARIMA model based on linear programming approach

The ARIMA prediction model in the previous section solves for the replenishment quantity and pricing strategy of vegetable items in the fresh produce superstore for the coming week, but the prediction model needs to be optimized in order to maximize the superstore's revenue under the given conditions.

Based on the field survey of the fresh produce superstore, it is required that under the requirement that the total number of available items is more than 27 and less than 33, and the order quantity of each individual item meets the minimum display quantity of 2.5 kg, the total quantity of replenishment of individual items and the pricing strategy are given by the model solution on July 1 based on the available varieties from June 24, 2023, to June 30, 2023, so that the superstore can maximize the revenue under the satisfaction of the market demand. Therefore, the ARIMA model can be optimized using linear programming methods.

Objective function:

The total revenue of the superstore is the objective of maximization, which can be expressed as:

$$Max = \sum_{i=1}^{n} (p_i \cdot x_i) \tag{3}$$

Constraints:

1) The total number of sellable items is constrained:

The superstore wants to keep the total number of saleable items between 27-33. This can be expressed as the following inequality:

$$27 \leqslant \sum_{i=1}^{n} x_i \leqslant 33 \tag{4}$$

2) Minimum display constraints:

The replenishment of each vegetable category is not less than the minimum display quantity of 2.5 kg, which can be expressed as:

$$x_i \geqslant 2.5, \quad i = 1, 2, ..., n \tag{5}$$

3) Market demand constraints:

The total amount of replenishment in the superstore needs to meet the market demand. This can be expressed as the following equation:

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} D_i \tag{6}$$

4) Cost-balancing constraints:

Replenishment levels and pricing strategies need to satisfy the balance between market demand and cost. For each vegetable category , the cost balance can be expressed as the following inequality:

$$x_i \cdot C_i \leqslant x_i \cdot p_i \tag{7}$$

Thus this optimization model can be expressed as:

$$Max = \sum_{i=1}^{n} (p_i \cdot x_i)$$

$$s.t. \begin{cases} 27 \leqslant \sum_{i=1}^{n} x_i \leqslant 33 \\ x_i \geqslant 2.5, \quad i = 1, 2, ..., n \\ \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} D_i \\ x_i \cdot C_i \leqslant x_i \cdot p_i \end{cases} \tag{8}$$

A sample of 27 groups is taken from the individual items, and the above model can be solved by MATLAB to obtain the replenishment quantity and pricing strategy of each individual item, so as to maximize the revenue of the superstore.

4. Conclusion

This paper takes the commodity information of each vegetable category in a fresh food superstore this year, the related data of sales and wholesale of each vegetable category, and the loss rate of each single product of each category as the object of study, establishes the Pearson correlation model of vegetable category and sales distribution, and solves the Pearson correlation coefficient of the sales distribution law and interrelationships between all the single products of cauliflower category by using the software of MATLAB and SPSS. The correlation coefficients between each category and each individual product were obtained. Meanwhile, the ARIMA prediction optimization model of commodity pricing and replenishment was established with the profit of the superstore as the objective function, and the number of displays, the total number of saleable items, the market demand and the cost balance as the constraints, and the ARIMA prediction optimization model was solved using the linear programming method, which obtained the total number of replenishment of individual items on a specific date of the fresh food superstore, and the pricing strategy with the maximum benefit under the market demand. In order to make better replenishment and pricing decisions for vegetable commodities, the superstore can more accurately predict and plan the replenishment quantity and pricing strategy of commodities by comprehensively analyzing the sales trend, customer demand and supply chain situation, so as to improve the goods turnover rate, reduce the backlog of inventory and satisfy the customers' demand, and to achieve the balance of supply and demand and maximize the operating efficiency.

References

[1] Gu Sihong. A study on dynamic pricing of fresh products in H-retailers considering freshness variation [D]. Donghua University, 2023. DOI:10. 27012/d. cnki. gdhuu. 2023. 001318.

[2] Lifang Wu. A multi-stage pricing model for fresh produce considering freshness and consumer utility [J]. Comprehensive Transportation, 2022, 44(10):124-129.
[3] Fan Bo, Song Wenbin. Research on product sales volume prediction based on ARIMA model [J]. Industrial Control Computer, 2021, 34(05):128-129+125.
[4] Cheng Juanjuan. Empirical Study on the Relationship between Research and Teaching in Colleges and Universities--Analysis Based on Pearson's Correlation Coefficient [J]. Science and Technology in Chinese Colleges and Universities, 2022(10):46-52. DOI:10. 16209/j. cnki. cust. 2022. 10. 016.
[5] Junjun Gui, Ting Kang. A joint decision model of inventory control and promotion optimization based on demand forecasting[C]// IEEE International Conference on Automation & Logistics. IEEE, 2009. DOI:10.1109/ICAL.2009.5262965.