Optimal Decision on Pricing and Replenishment of Vegetable Products Based on Goal Planning

Ge Chen *, #, Yewei Chen #

School of Mathematics, Hangzhou Normal University, Hangzhou, China, 311121

* Corresponding author: 19858155354@163.com

#These authors contributed equally.

Abstract. Developing replenishment and pricing plans has always been the key for fresh food supermarkets to gain more profits. This paper analyzes various situations and provides the optimal solution through Spearman correlation analysis, Lasso regression, goal programming and other methods and models. We use normality testing and box plots to present the distribution patterns of major categories and individual products from both vertical and horizontal perspectives. Establish a single objective programming model and solve it with Lingo to obtain a maximum profit of 98513 yuan for 7 days. The largest total replenishment amount is for the flower and leaf category, the smallest is for edible mushrooms, the highest average cost markup rate is for aquatic roots and stems, and the lowest is for cauliflower. This has certain reference value and theoretical significance for fresh food supermarkets to formulate replenishment and pricing plans.

Keywords: Normality test, Spearman correlation analysis, Lasso regression, ARIMA model, Single objective planning.

1. Introduction

Among the products sold by fresh produce supermarkets, vegetable products often have a short shelf life, and as sales time increases, the phase change of vegetable products is poor, resulting in most categories of vegetables that cannot be sold the next day if they are not sold on the same day [1]. How to use appropriate models to assist fresh food supermarkets in formulating the best replenishment and pricing plans has always been a topic worth exploring.

Supermarkets usually develop replenishment plans based on product categories. So the supermarket is looking forward to further optimizing its replenishment plan, limiting the number of available items to 27 to 33, and allowing the order quantity of each item to reach 2.5 kilograms, which is the minimum display quantity for the supermarket [2].

Hence, we use methods such as normality testing and box plots, present and analyze the distribution patterns of major categories and individual products from two perspectives: vertical (in time) and horizontal (in space). After that, we obtain the relationship between the total sales volume of vegetable categories and the cost markup rate of each vegetable category through function fitting. Therefore, first determine the linear relationship and multicollinearity between each independent variable. In the case of strong multicollinearity between each independent variable, use regression to finally obtain the expression of the total sales volume of each category and the cost markup rate of each vegetable. By collecting the sales price, sales volume, replenishment volume, transportation and storage fees for each sales cycle, and considering the discounted sales of some overnight products, a target function is established with the goal of maximizing the average daily profit of supermarkets within a production cycle. The target function is restricted by sales price, sales volume of various vegetables, cost markup rate, and inventory level. At the same time, the impact of weather conditions, consumer feedback, and unexpected events on supermarket revenue was also considered.

2. Preliminary

2.1. Assumptions

1. The loss rate is the proportion of fruits that cannot be sold due to transportation losses and other reasons [3].
2. Products that can be sold overnight are not discounted on the first day.
3. The profit from selling discounted vegetables is much smaller than the total profit [4].

2.2. Notations

The primary notations used in this paper are listed in Table 1.

Table 1. Notations

Symbols	Notations
w	the total vegetable inventory of the supermarket
p_i	Purchase price of the *i* th vegetable in a marketing cycle
c_i	Cost markup rate of the vegetable *i* over a marketing cycle
Q_i	Sales of the *i* th vegetable per marketing cycle
$\min c(i, j)$	The minimum value of the cost markup rate for vegetable *i* on day *j*
$\max c(i, j)$	The maximum value of the cost markup rate for vegetable *i* on day *j*
x_i	Availability of next-day resale of vegetable *i*
T_i	Marketing cycle for vegetable *i*
q_j	Sales volume on day *j*
$Trans_i$	transportation cost of each replenishment of the vegetable *i*
t_i	Vegetable unit transportation cost
S_i	Total storage cost for a marketing cycle of the vegetable *i*

3. The distribution pattern of sales volume among vegetables

3.1. The distribution pattern of major categories

Firstly, it is possible to explore whether each sales volume conforms to a normal distribution pattern. As the number of sales samples by day for each major category is, it can be considered as a small sample. Therefore, a test can be used, and the final results are shown in Table 2.

Table 2. S-W test results

Variable Name	Average	Standard deviation	Skewness	Kurtosis	S-W test
Flower leaf class	180.597	85.229	3.074	28.684	0.843(0.000***)
florescent vegetables	38.256	22.775	1.565	4.338	0.899(0.000***)
Aquatic rhizomes	36.889	31.195	2.597	12.967	0.805(0.000***)
Solanaceae	20.39	12.913	1.888	6.72	0.871(0.000***)

The sales distribution of eggplants, cauliflower, and aquatic rhizome vegetables is relatively stable, while the sales of flowers, leaves, chili peppers, and edible mushrooms are easily affected by seasonal factors[5]. At the same time, it can be seen from the median and the spacing between the top and bottom four digits of each box chart that the sales distribution of leafy vegetables, chili vegetables, edible mushrooms, and cauliflower vegetables is relatively symmetrical.

3.2. Distribution pattern of individual products

The distribution of Yunnan lettuce, Xixia shiitake (1), and Xiaomi spicy (portion) is relatively stable, while most vegetables experience significant fluctuations in sales due to seasonal and holiday influences. At the same time, it can also be seen that the minimum sales of Wuhu green pepper (1), broccoli, clean lotus root (1), and purple eggplant (2) are all greater than 0, indicating that the supply is relatively stable and there is no inability to supply due to various reasons. It is possible to choose to establish long-term cooperation with these suppliers.

Around February, there may be a sharp increase in sales due to the Chinese New Year holiday. By consulting information, it can be found that lotus roots are generally harvested from July to December, which explains the clear decreasing trend in February to July. There is a clear increasing trend from July to February of the following year.

3.3. Spearman correlation analysis

By introducing the rank, which is the average descending position of the original data in the overall data, the Pearson correlation coefficient formula can be transformed into the following one [6].

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{1}$$

Among them, the difference in the position (level) of the paired variables after sorting two variables separately, and is the number of observation objects. The Spearman correlation coefficient can be obtained.

The value of the obtained correlation coefficient varies between+1 and -1. A value of 1 indicates a perfect correlation between variables. As the correlation coefficient value approaches 0, it means that the relationship between variables will become weaker. When the correlation coefficient is 0, it is completely uncorrelated. The direction of the relationship is represented by the sign of the coefficient; A positive sign indicates a positive relationship, while a negative sign indicates a negative relationship.

As the prerequisite for finding the relationship between sales volume is that each product is being sold simultaneously, the daily sales volume of the six categories is first removed from rows with 0 to avoid the impact of the period when sales cannot be sold due to special reasons on the correlation coefficient. Further Spearman correlation analysis can be used to obtain the correlation coefficients between the two categories of each product.

3.4. The interrelationships between vegetable categories

The correlation coefficient between almost all vegetable categories shows a positive correlation, as there are often significant differences in nutrition, taste, and other aspects between different categories of vegetables. They are usually purchased together rather than just one type. Only two sets of relationships show a negative correlation, and the absolute value of the negative correlation coefficient is very small, indicating that the degree of negative correlation is not significant. Among them, eggplants and fungi may be caused by similar taste.

4. The relationship between vegetable categories and cost plus pricing

4.1. Establishment of the Model

It is easy to know that the cost plus method divides the sales price into two parts: cost and profit, and the profit is obtained by multiplying the ratio of cost and cost plus. Defined as the cost markup rate of the vegetable, upon observing the data provided in the question, it was found that the pricing of the vegetable in the data did not vary significantly over time. Therefore, the range of values for the independent variable of the function fitted using sales price (i.e. pricing) was too small, making this function unable to be widely used in real-life situations. Therefore, here we consider using the cost markup rate of vegetables to replace the selling price of vegetables for fitting. Next, we will use flowers and trees as an example for fitting, and the other fitting methods are the same [7].

$$w = (1 + c)p, \tag{2}$$

Due to the large number of data samples, seasonality in the types and quantities of vegetable sales, and considering the maximum profit of supermarkets in the next question, we will select data from late June to early July each year for fitting.

Divide the data by days and filter the sales volume, sales price, and cost price of each category of vegetables during the above time period. Use formula 2 to calculate the cost markup rate for each vegetable per day. Remove outliers in the cost markup rate during data preprocessing.

4.2. Multi-collinearity test

Firstly, calculate the contribution rates of each variable to the dependent variable, and sort them in descending order according to their contribution rates. Select the independent variables in sequence until the cumulative contribution rate of the independent variables exceeds 90%. For vegetables with a remaining cumulative contribution rate of 10%, it can be considered that the change in their cost markup rate has a relatively small impact on the total sales volume and can be ignored. Therefore, the cost markup rate of this part can be treated as a constant. Next, perform a linear relationship judgment on the selected variables. Determine whether the null hypothesis is valid through significance testing. If the original hypothesis is true, multiple linear regression can be used for fitting [8].

$$\mathrm{VIF} = \frac{1}{1 - R^2} \tag{3}$$

$$R^2 = 1 - \frac{\sum(\hat{y}_i - y_i)^2}{\sum(y_i - \hat{y})^2} \tag{4}$$

If VIF >10 is obtained after testing, it is considered that there is strong multi-collinearity between the independent variables, and multiple linear regression cannot be used for subsequent calculations.

Consider linear models:

$$\mathrm{y} = X\beta + \varepsilon, E(\varepsilon) = 0, \mathrm{cov}(\varepsilon) = \sigma^2 I. \tag{5}$$

Estimating unknown parameter vectors using the least squares method aims to minimize the square of the error term, i. e:

$$\varepsilon^2 = \| y - X\beta \|^2 = (y - X'\beta)(y - X\beta). \tag{6}$$

$$\hat{\beta} = (X'X)^{-1} X'y. \tag{7}$$

Set a minimum objective function now

$$Q(\beta) = \varepsilon^2 + \lambda \| \beta \|_1 , s.t. \sum_1^j \beta_j < s. \tag{8}$$

In fact, in a Cartesian coordinate system, the constraint condition can be drawn as a rectangular region, gradually approaching the estimated value through the boundaries of the rectangular region, and ultimately obtaining better results. The specific steps are shown in Figure 1.

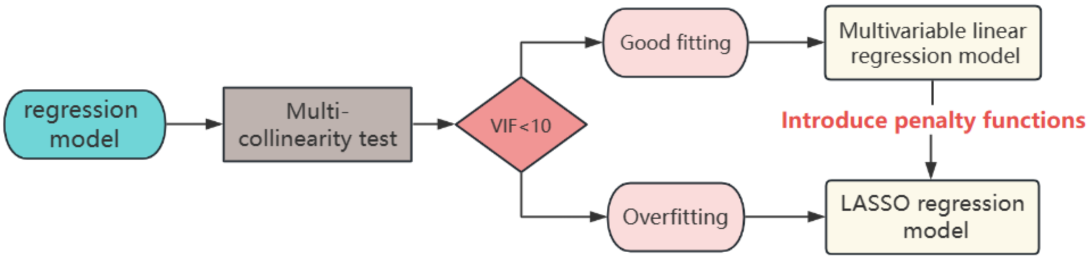

Figure 1. Specific steps of the LASSO regression model

Taking category one as an example, we will solve it. Firstly, verify that there is a linear relationship between the independent variables of category one. The significance tests and tests were conducted on each independent variable, and the results of the sum of the values obtained are shown in Table 3.

Table 3. Value and value test results for some vegetables

Vegetable	*VIF*
Yunnan lettuce	2.235
Yunnan Youmai Cai	1.552
Sweet potato tip	2.623
Bamboo leafed vegetable	10.822
Yellow Cabbage (2)	1.925
Vegetable heart	4.61
Chinese cabbage	13.921
Shanghai Qing	3.561

Based on the above data, it can be concluded that there is a linear relationship between the independent variables, as they have passed the significance test. Furthermore, from the values of each independent variable, it can be seen that the cost markup rate data of some vegetables have strong multi-collinearity, so multiple linear regression cannot be directly used to solve the problem, and regression needs to be used to solve the multi-collinearity problem.

After the above calculation, the relationship between the total sales volume of category one vegetables and the cost plus pricing of each vegetable is obtained as follows

$$Q = 156.5 - 37.8c_1 - 81.8c_2 - 48.4c_3 - 35.4c_4 - 78.7c_5 + 82.6c_6 - 72.1c_7 \tag{9}$$

4.3. Solutions of the model

Due to the inability of supermarkets to change vegetable purchase prices during restocking, and the fact that vegetable purchase prices have a certain temporal pattern, it is considered to use the ARIMA model to predict the purchase prices of various vegetables from July 1st to July 7th, 2023.

The following selection of vegetable sales, vegetable transportation costs and storage costs indicators to build a model to help the program better developed, the following explanation of the reasons for the selection of these three indicators.

Let the i th type of vegetables sold every day to gain Mi, the title mentioned that most of the vegetables overnight can not be re-sold, so the vegetables will be divided into two categories for consideration, where the two categories are divided into specific see Figure 2.

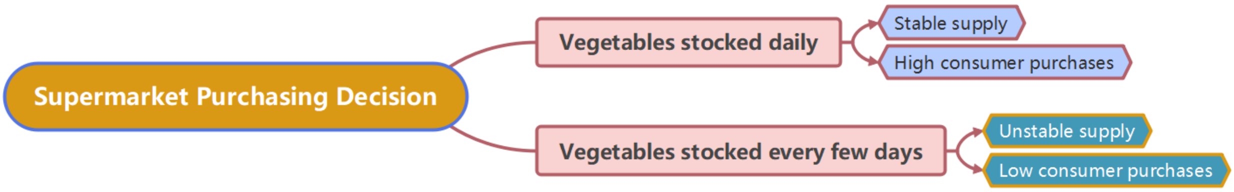

Figure 2. Supermarket purchase decision chart

Considering that some vegetables can be sold overnight and some vegetables can be sold only on the same day, the components of profit are analyzed with the objective of having the highest average profit per day in a sales cycle, where the source of profit is the selling price of vegetables and the expense components are the transportation costs of vegetables and the storage costs of overnight vegetables. Therefore, the effect of vegetable sales, vegetable transportation costs and storage costs on supermarket profits will be considered quantitatively below.

(1) Income from sales

Let the i th kind of vegetable sales per sales cycle for Qi, the vegetable in a sales cycle of the purchase price for pi, the corresponding cost plus rate for ci, let the i th kind of vegetable sales cycle for Ti, assuming that the first day of the vegetables are not discounted, the second day and every day thereafter the vegetable discounts for discount (j), j=1, 2,…, Ti. the j th day of the sales of is qj, , then there are

$$Q_i = \begin{cases} q_i c_i p_i, x_i = 0 \\ c_i p_i \left[q_i + \sum_{j=2}^{T_i} discount(j) q_j \right], x_i = 1 \end{cases} \tag{10}$$

Therefore there is a need to collect data on sales, pricing, selling price and volume of each vegetable at each sales cycle.

(2) Transportation costs

Transportation costs may vary among suppliers based on distance and other reasons, and suppliers with low wholesale prices may have high transportation costs, so they can be optimized in conjunction with transportation costs.

Let the transportation cost of each replenishment of the i th vegetable be Transi, the unit transportation cost of the vegetable be ti, and the volume of each replenishment be ri, then there are

$$Trans_i = t_i r_i \tag{11}$$

Therefore, data on the cost of each transportation for each vegetable is needed.

(3) Storage costs

It is known that most vegetables cannot be sold on the next day if they are not sold on that day, so for such vegetables, no storage fee is required, while for other vegetables that can be sold on the next few days, they can be sold on at a reduced price, etc., in order to make a larger profit, and for this group of vegetables, a storage fee is required. Use x_i to distinguish whether the i th vegetable can continue to be sold every other day, let

$$x_i = \begin{cases} 1, T_i > 1, \\ 0, T_i = 1. \end{cases} \tag{12}$$

Let the total storage cost per sales cycle be Si, the unit storage cost per day for vegetable i be ki, thus it can be obtained:

$$s_i = \begin{cases} 0, x_i = 0 \\ k_i r_i + \sum_{j=2}^{T_j} k_i \left[r_i - \sum_{t=1}^{j-1} q(j) \right], x_i = 1 \end{cases} \tag{13}$$

According to the data measured above, with the premise that the superstore's profit in a sales cycle is maximized, i.e., the average total profit per day is required to be maximized, let the number of all kinds of vegetables be N, and you can construct the superstore's profit function per day in a sales cycle as M.

$$M = \sum_{i=1}^{N} \frac{Q_i - Trans_i - s_i}{T_i} \tag{14}$$

The relevant constraints are given below.

(1) Supermarket Inventory Capacity Limitations

Let the i th type of vegetable replenishment of the maximum share of δ_i, the total inventory of the superstore is w, need to meet the amount of vegetables purchased in each sales cycle does not exceed the maximum share of the vegetable in the total inventory and require:

$$r_i \le w \cdot \delta_i, i = 1,2,\cdots,N \tag{15}$$

(2) Cost-plus Rate Limitation

Let the minimum and maximum values of the cost-plus rate of the i th vegetable on the j th day be minc (i, j) and maxc (i, j), respectively. In order to ensure a reasonable market structure, and taking into account the impact of emergencies on the relationship between supply and demand, it is considered to add a relaxation factor u, which needs to be satisfied:

$$(1\text{-}u)\min c(i,j) \le c(i,j) \le (1+u)\max c(i,j), \forall i, j. \tag{16}$$

4.4. Results

By LINGO [9], the final total profit of the 7-day supermarket was 98513 yuan, with an average of 14073.28 yuan per day. At the same time, the total replenishment volume of each category from July 1st to July 7th and the weighted average cost markup rate of each corresponding category per day were obtained, as shown in Tables 4 and 5, respectively.

Table 4. Value and value test results for some vegetables

Vegetable	Flowers leaves	Cauliflower	Aquatic roots	Stems and eggplants
2023/7/1	213.9	91.8	82.2	86.1
2023/7/2	213.5	91.8	82.2	86.1
2023/7/3	94.3	91.8	82.2	86.1
2023/7/4	213.5	91.8	82.2	100.9
2023/7/5	137.6	91.8	82.2	57.2
2023/7/6	248.7	91.8	82.2	57.2
2023/7/7	243.8	91.8	82.2	57.2

Table 5. Weighted average cost markup rate for various vegetable from July 1st to 7th

Vegetable	Flowers leaves	Cauliflower	Aquatic roots	Stems and eggplants
2023/7/1	3.9	2.8	9.1	6.9
2023/7/2	4.1	2.8	9.1	6.9
2023/7/3	6.0	2.8	9.8	6.9
2023/7/4	4.1	2.8	9.8	6.1
2023/7/5	3.2	2.8	9.8	7.1
2023/7/6	3.5	2.8	9.1	7.1
2023/7/7	3.4	2.8	9.8	7.1

4.5. Analysis

Through the above planning model, in practical life, only the parameters to be collected need to be determined to obtain the optimal solution for a sales cycle, and to formulate a replenishment plan and pricing discount decision for a sales cycle, so as to maximize the profit of supermarkets. In addition to considering the data to be collected mentioned above, the following factors can also be appropriately considered according to the actual situation to make the model more comprehensive and closer to real-life situations.

(1) Government policies. In addition to the aforementioned factors such as weather and suppliers that can affect the prices of vegetable products, government policies can also affect the prices of vegetable products. Understanding relevant government policies can help supermarkets better grasp the market situation and make replenishment and pricing decisions [10].

(2) Competition from other supermarkets. In general, there may be other supermarkets near the supermarket. It is necessary to understand the customer flow, vegetable product categories, product pricing, and promotion situation of the nearby supermarkets, and flexibly adjust replenishment and pricing decisions based on this data.

(3) Weather in vegetable producing areas. Low temperatures, cold waves, and rainy weather have a direct impact on vegetable prices. Excessive rainfall in vegetable producing areas can lead to reduced vegetable production, increased vegetable prices, and a decrease in sales volume. By paying attention to weather data in vegetable producing areas, it is possible to roughly infer whether vegetables have decreased production and make better replenishment and pricing decisions.

5. Conclusion

We analyzing the distribution patterns of categories and individual products from multiple perspectives, exploring various hidden patterns, the model is comprehensive, and most of the methods are universal, with a wide range of applications. By dividing supermarket revenue into sales revenue, transportation costs, and storage costs, and categorizing all vegetables into two categories based on whether they are available for sale overnight, parameters such as sales cycles and discount rates can be changed to make them suitable for complex real-life situations. However, we do not discussing the discounted sales of vegetables may result in a certain difference between the obtained results and the optimal solution.

References

[1] Liu, Jiahe. "Vegetable sales data mining: association rule analysis." *Journal of Education, Humanities and Social Sciences* 25 (2024): 97 - 104.

[2] Chen, Ting, et al. "Using fresh vegetable waste from Chinese traditional wet markets as animal feed: Material feasibility and utilization potential." Science of The Total Environment 902 (2023): 166105.

[3] Chen, Tingting, Lihong Zhou, and Senxin Guo. "Research on Pricing and Replenishment Strategy for Vegetable Commodities based on Non-linear Programming." Transactions on Economics, Business and Management Research 1 (2023): 19 - 24.

[4] Zhang, Huifeng, and Ziheng Yan. "Automatic Pricing and Replenishment Decisions of Vegetable Products Based on Machine Learning." Highlights in Business, Economics and Management 25 (2024): 228 - 236.

[5] Parsafar, Bohlul, et al. "The impact of fruit and vegetable waste on economic loss estimation." (2023): 871 - 884.

[6] Hamilton M A, Russo R C, Thurston R V. Trimmed Spearman-Karber method for estimating median lethal concentrations in toxicity bioassays [J]. Environmental Science & Technology, 1977, 11 (7): 714 - 719.

[7] Ozerova, M. G., A. V. Sharopatova, and J. A. Olentsova. "The development level and economic efficiency of vegetable production in the Krasnoyarsk region." IOP Conference Series: Earth and Environmental Science. Vol. 421. No. 3. IOP Publishing, 2020.

[8] Zhang, Biao, et al. "Effect of householder characteristics, production, sales and safety awareness on farmers' choice of vegetable marketing channels in Bei**g, China." British Food Journal 119.6 (2017): 1216 - 1231.

[9] Beetsma R M W J, Schotman P C. Measuring risk attitudes in a natural experiment: Data from the television game show Lingo [J]. Papers, 1998.

[10] Thompson, Jennifer Jo, and Julia Gaskin. "An extension specialist's reflections from the field: discovering ag of the middle in the shift from direct sale to wholesale vegetable production." Culture, agriculture, food and environment 40. 2 (2018): 124 - 129.