Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 5, No. 3, 2023

Research on Vegetable Commodity Pricing and Replenishment based on Planning Models and Genetic Algorithm

Tao Chen †, Zikun Luo †, Jinhui Liu †

School of Mechanical Engineering & Automation, Wuhan Textile University, Wuhan, China
†These authors also contributed equally to this work

Abstract: This paper focuses on an in-depth study of supermarket vegetable pricing and replenishment problems, utilizing a variety of methods such as statistics, prediction models, planning models and other methods of analysis. First, the data were preprocessed, and frequency distribution histograms were drawn, revealing the distribution pattern and correlation between each category and each single product of vegetables through descriptive statistical analysis. Secondly, for the relationship between sales and pricing, the sales unit price was averaged through the cost-plus pricing formula, and MATLAB was used to nonlinearly fit the relationship between the total sales volume of the categories and the sales price, and the fitting result was further optimized through neural network, and a nonlinear planning model was established, and a genetic algorithm was used to solve the daily replenishment volume of supermarkets and the pricing strategy in order to achieve the maximization of revenue. Finally, the top-rated individual products were screened out by entropy weighting method, and a linear programming model was established to predict the replenishment quantity and pricing strategy for the coming day, which further provided effective decision support for the sales management of the supermarket.

Keywords: Vegetable Pricing; Replenishment Strategy; Nonlinear Programming; Genetic Algorithm.

1. Introduction

In the fresh food supermarket sector, vegetable commodities have a relatively short shelf life and their quality and appearance deteriorate over time. To meet this challenge, supermarkets make daily replenishment decisions based on historical sales data and market demand. Typically, supermarkets use a "cost-plus pricing" strategy to price different vegetable items, while offering discounts for items that have deteriorated in quality or appearance. Due to the wide variety of vegetables and their diverse sources, supermarkets need to rationalize their replenishment and pricing strategies without knowing the exact cost and purchase price of each item. There is a correlation between sales volume and time of year, especially during the period from April to October when the variety of vegetables is more abundant. Therefore, developing a reasonable sales mix to maximize revenue within a limited sales space has become an urgent problem. The purpose of this study is to investigate the correlation between different vegetable categories and individual products, and to analyze the relationship between the total sales volume and the "cost-plus pricing" strategy. At the same time, this paper will propose a set of daily replenishment and pricing plans for the coming week to maximize the supermarket's revenue [1-3].

2. Vegetable Correlation Analysis

2.1. Distribution Pattern by Category

To reduce the effect of chance, this paper further uses descriptive statistical analysis monthly. The monthly sales volume of each category was summarized through Excel and a line graph was plotted to observe the change in sales volume between July 2020 and June 2023, as shown in Figure 1.

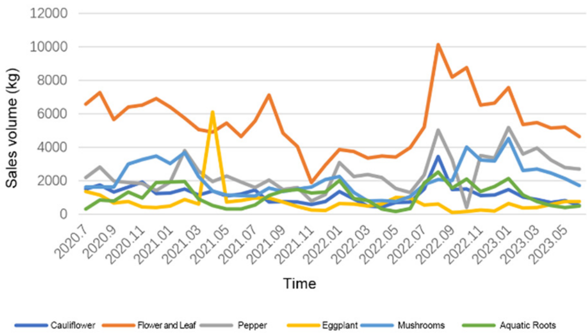

Figure 1. Category distribution pattern in terms of months

Figure 4 demonstrates that the flower and leafy category had the highest total sales volume and was popular with consumers during the study cycle, while eggplant vegetables had the lowest total sales volume. It is worth noting that after June 2022, the sales volume of all categories increased compared to the previous period, which may have been affected by the epidemic. Overall, the volatility of sales volume is high.

In this paper, we first divided the sales volume by multiple years and integrated and presented the sales data for each year through frequency distribution histograms. The division of sales data into multiple intervals aims to analyze and reveal the general trend in sales volume to better understand the changes and distribution characteristics of sales data. As shown in Figure 2, the graphs demonstrate the distribution status and change the trend of sales volume in different years. Through comparative analysis, the changes in sales volume in different years can be more clearly grasped, providing a basis for further analysis.

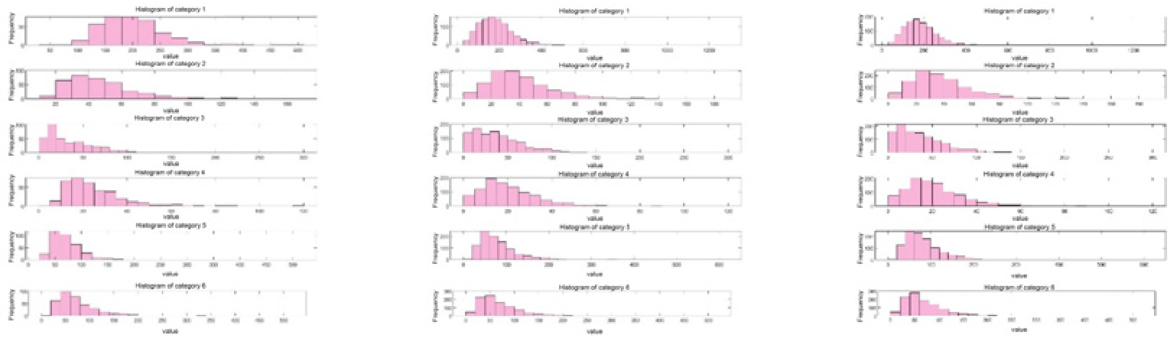

Figure 2. Histogram of frequency distribution

Figure 2 demonstrates the broad similarity and apparent cyclicality of sales volume by category from year to year. Among them, the sales volume of the chili pepper category increases year by year, showing that this category is becoming more and more popular among consumers. Observing the distribution of sales volume in each year, it is not difficult to find that the peaks of the data are generally skewed to the left, indicating that the sales data do not conform to a normal distribution. Therefore, this paper chooses to use Spearman correlation test to analyze the correlation between the sales volume of each category. Spearman correlation test does not require the data to be normally distributed, which is more suitable for the data characteristics and needs of this study, and can effectively assess the correlation of the sales volume between each category.

2.2. Correlation Analysis

Figure 3 illustrates the correlation coefficients of sales volume among different vegetable categories, thus reflecting the correlation of sales volume among these categories. Significantly, the correlation coefficient between the foliage and cauliflower categories reaches 0.9, which is almost close to 1, indicating a strong positive correlation between these two. In contrast, the correlation coefficient between eggplant and aquatic rootstocks was -0.72, and although the absolute value of this value is also closer to 1, the negative sign indicates a strong negative correlation between them. While the correlation coefficients of eggplant with cauliflower and pepper with eggplant were -0.18 and -0.17, respectively, and these values were closer to 0, indicating a relatively weak correlation between them [4].

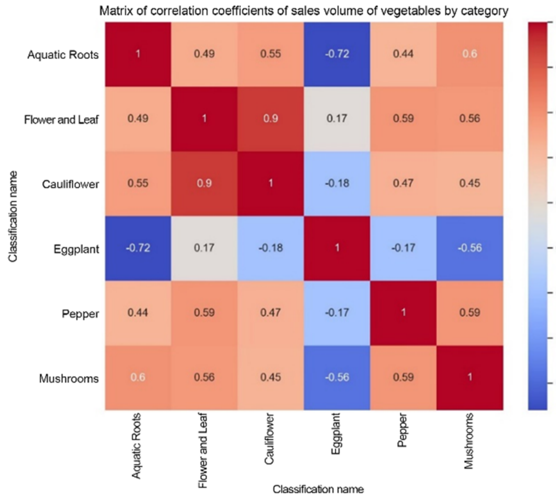

Figure 3. Spearman correlation coefficient plot

Because there are a total of 252 individual items, the Spearman correlation coefficient can be used to draw conclusions, but it is not easy to observe. Therefore, this paper divides the individual products into 6 categories and then studies the correlation between each individual product. As shown in Figure 4.

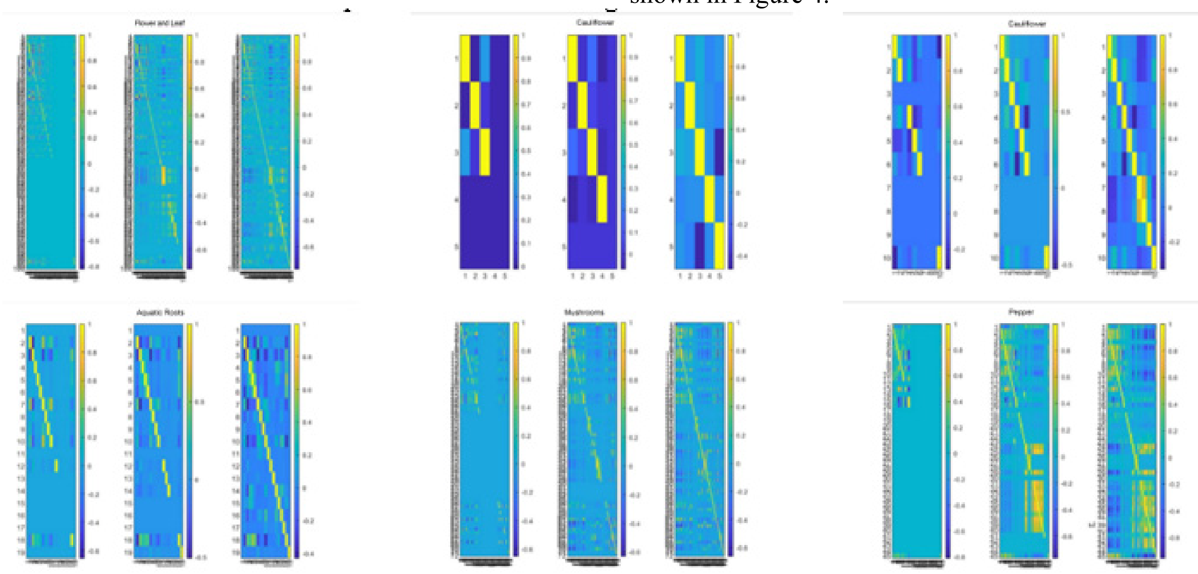

Figure 4. Correlation between individual items in the category

Figure 7 shows the correlation of sales volume between different vegetable items through color shades, with darker colors indicating stronger correlation and lighter colors indicating weaker correlation. Based on the pattern shown in the figure, this paper draws the following conclusions:

a) In the flower and leaf category, the strongest correlation was found between red coral (coarse leaf), red oak leaf and green butter with a correlation coefficient of almost 1, showing a high positive correlation between their sales volumes.

b) In the cauliflower category, the strongest correlation was found between green peduncle loose flower and zhi jiang green peduncle loose flower, indicating a high degree of synchronization between the changes in the sales volume of these two types of cauliflower.

c) In the aquatic roots and tubers category, the strongest correlation was demonstrated between rhododendron and Honghu lotus root strips, pointing to a high degree of synchronization between changes in the sales volume of these two types of cauliflower.

d) In the eggplant category, the strongest correlation was found between green eggplants (2) and purple eggplants (1), implying that they may be affected by similar market factors and that their sales volume movements show a high degree of correlation.

e) Among the chili peppers category, the most significant correlation was found between millet peppers (1) and small wrinkled peppers (1), showing a high degree of consistency in their sales volume movements.

f) In the edible mushroom category, the strongest correlation was demonstrated between seared mushrooms (2) and shimeji mushrooms (2), indicating a high degree of synchronization in the characteristics of their sales volume movements.

3. Pricing and Replenishment Strategies for the Category

3.1. Modeling

This paper addresses the analysis of data from July 1 to July 7, 2023, which covers exactly one week's time, therefore, all the dates were divided into seven days, from Monday to Sunday. The sales price of vegetables and the total sales volume of the six categories were extracted, calculated, and summarized through Excel. Considering that April to October is a period of abundant vegetable supply, this paper chose to consider only the sales price and total sales volume during this time, and discarded data that did not fall within this time frame. After collation, a dataset that meets the needs of the analysis is obtained. To explore the relationship between the total sales volume of each category of vegetables and cost-plus pricing, this paper refers to the formula of cost-plus pricing and finds that cost-plus pricing is very close to the sales price of vegetables. Therefore, it chooses to use the sales price of vegetables directly instead of cost-plus pricing, i.e., focusing the analysis on the relationship between the total sales volume and sales price of each category of vegetables. After obtaining the required data, this paper utilized MATLAB to fit the data non-linearly, taking the data of July 2 (Sunday) as an example, and the fitting results are shown in Figure 5.

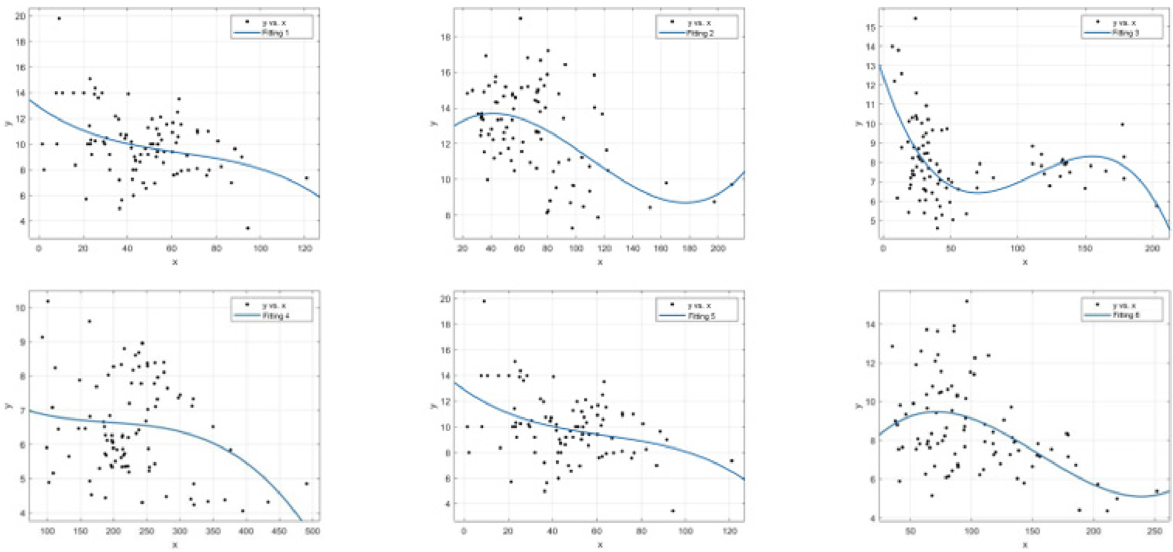

Figure 5. Fitting results

In this paper, the functional relationship between total sales volume and sales price for each of the six categories of vegetables is obtained [5]:

$$\begin{aligned} y_{12} &= 1.4245{x_{12}}^2 - 41.9281x_{12} + 365.9303 \\ y_{22} &= -6.1719x_{22} + 150.2939 \\ y_{32} &= 1.9656{x_{32}}^2 - 45.3236x_{32} + 334.9964 \\ y_{42} &= 0.4828{x_{42}}^3 - 15.169{x_{42}}^2 + 144.365x_{42} - 365.712 \\ y_{52} &= 1.4245{x_{52}}^2 - 41.9281x_{52} + 365.9303 \\ y_{62} &= 0.0001{x_{62}}^3 - 0.045{x_{62}}^2 - 2.29x_{62} + 71.415 \end{aligned} \tag{1}$$

Considering the attrition rate of 6 categories, here in this paper, we directly use the average of the attrition rate of individual products in each category to represent the attrition rate of this category:

$$P_i = \frac{\sum_{k=1}^{N_i} P_{ik}}{N_i} \tag{2}$$

At the same time the question is to consider the wholesale price of the category, so this paper uses the average of the wholesale price of a single product in a category to represent the wholesale price of this category:

$$Z_i = \frac{\sum_{k=1}^{N_i} Z_{ik}}{N_i} \tag{3}$$

To avoid situations such as infinity, this paper adds a constraint term L to constrain the objective function of the problem. If the sales price is high, the constraint term can be used to make its sales volume lower; if the sales volume is too high, its vegetable sales price can be reduced. The constraint term L is formulated as follows:

$$L = \frac{y_{ij}}{1+\left(\frac{y_{ij}-\mu(y_{ij})}{\sigma(y_{ij})}\right)^2} \tag{4}$$

Thus, the objective function can be listed in this paper:

$$\max w = x_{ij}y_{ij} - \frac{x_{ij}}{1-P_i}Z_i + \frac{y_{ij}}{1+\left(\frac{y_{ij}-\mu(y_{ij})}{\sigma(y_{ij})}\right)^2} \tag{5}$$

3.2. Genetic Algorithm Optimization

Genetic Algorithm (GA) is a stochastic search algorithm inspired by the mechanism of natural selection, aiming at finding the optimal solution of a problem. By simulating the selection, crossover and mutation processes in nature, the optimal solution is finally found among the candidate solutions through multiple combinations and iterations, thus realizing the principle of "survival of the fittest" in biology [6].

Based on the above model, the objective function is solved in this paper. The specific solution process is shown in Figure 6. The process shows how to apply genetic algorithms to solve the problem, including the generation of the initial solution, selection operation, crossover operation, mutation operation and the judgment of the termination conditions and other key steps. Through the solution process of genetic algorithm, this paper is able to find the optimal solution in the solution space of the problem, which provides an effective method and support for solving practical problems.

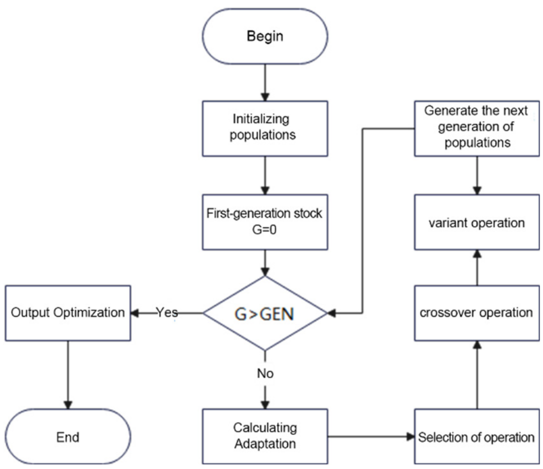

Figure 6. Flowchart of genetic algorithm

The specific algorithm is as follows:

- Step1 Chromosome coding: the more common methods are floating point coding, permutation coding and binary coding. In this paper, binary coding is used to form chromosomes using the number 0 and the number 1.
- Step2 population initialization: establish a suitable model to constitute the initialization population;
- Step3 fitness function: to ensure that good genes will be inherited to the next generation, using the "survival of the fittest, the survival of the fittest" principle, the establishment of fitness function. Its selection directly affects the speed of solving and the optimal solution.

$$\text{Fit}\ (f(x)) = \begin{cases} C_{\max} - g(x), C_{\max} > g(x) \\ 0,\ \text{other} \end{cases} \tag{6}$$

- Step4 Selection operator: select the individuals with higher adaptation in the population to be retained as parents, this paper adopts the roulette wheel betting method to establish the probability that has a mapping relationship with the adaptation function:

$$P_i = \frac{f(x_i)}{\sum_{j=1}^{n} c}, i=1,2,...,n; j=1,2,...l \tag{7}$$

- Step 5 crossover operator: select appropriate chromosomes and use crossover operation to produce offspring.
- Step6 mutation operator: perform mutation operation on certain individuals, often denoted by M.
- Step7 judgment of fitness function value: jump to Step3 to recalculate and judge after the completion of the previous step and continue the cycle.
- Step8 termination condition: terminate the operation when the number of iterations reaches a predetermined number of iterations.

Using Matlab, the solution can be used to derive the stocking data and pricing data for the next seven days (2023.7.12-023.7.7).

4. Pricing and Replenishment Strategies for Individual Products

4.1. Data Screening based on Topsis and Entropy Weighting Methods.

In this study, the total daily sales volume of each individual product obtained was first utilized for regression analysis with time, aiming to reveal the pattern of sales volume changes over time. However, due to the large number of vegetable individual items, direct regression analysis may lead to excessive complexity of the analysis. To simplify the analysis process and improve the efficiency and accuracy of the analysis, this paper chooses to first apply the entropy weight method to screen all the vegetable single products, to select the 33 single products with the highest comprehensive score as the object of analysis [7].

Entropy weight method is a weight determination method based on information entropy, which can objectively reflect the importance of each index. Through the entropy weight method, this paper successfully screened out the 33 single products with the highest comprehensive score, providing a clear research object for the subsequent regression analysis.

Based on the 33 screened products, this paper carried out linear regression analysis to explore the law of sales volume change over time. Through the establishment and optimization of the regression model, this paper can predict the sales volume of each single product on July 1, 2023 (the eighth day) more accurately. This prediction not only provides a useful reference for the sales management of the supermarket, but also provides data support for further optimizing the supply chain and sales strategy of the supermarket.

Step1 Calculation of weights of each indicator for vegetable singles

1) Extremely Large Indicators

$$v'_{ij} = \frac{v_j - v_{\min}}{v_{\max} - v_{\min}} \tag{8}$$

2) Very small indicators

$$v'_{ij} = \frac{v_{\max} - v_j}{v_{\max} - v_{\min}} \tag{9}$$

Step2 Normalization of data

The formula for data normalization is:

$$t_j = \frac{x_j}{\sqrt{\sum_{i=1}^{n} x_j}} \tag{10}$$

Step3 Positive Idealization and Negative Idealization Solving

$$\begin{aligned} T^+ &= \max_{1 \leqslant i \leqslant n} \{T_{ij}\} = (T_1^+, T_2^+, \ldots T_n^+) \\ T^- &= \min_{1 \leqslant i \leqslant n} \{T_{ij}\} = (T_1^-, T_2^-, \ldots T_n^-) \end{aligned} \tag{11}$$

Step4 Solving the weights of the indicators.

The weights occupied by each indicator were calculated using the entropy weighting method, as shown in Table 1 below.

Table 1. Entropy weighting method weights

	Information entropy	Weights
v_1	0.730	0.424
v_2	0.976	0.119
v_3	0.751	0.446
v_4	0.924	0.010

Step5 Euclidean distance calculation

$$\begin{aligned} D_i^+ &= \sqrt{\sum_{j=1}^{m} [\omega_i (T_J^+ - T_{ij})^2]} \\ D_i^- &= \sqrt{\sum_{j=1}^{m} [\omega_i (T_J^- - T_{ij})^2]} \end{aligned} \tag{12}$$

Step6 Optimal Value Calculation and Ranking

Calculate the relative closeness of the ith individual product to the ideal solution: the

$$W_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{13}$$

The composite indicator values for each individual item of all these vegetables were then ranked and the final ranking results are shown in Table 2 below.

Table 2. Individual product ranking results (top 10)

Vegetables	Rankings	Vegetable Singles	Rankings
Steak Mushroom	1	Purple Round Eggplant	6
Wuhu Green Pepper (1)	2	Chili Pepper	7
Ginger, Garlic & Millet Pepper Combo	3	Crabmeat Mushroom (2)	8
Iceweed (box)	4	Wood Ear Vegetable	9
Amaranth	5	Suizhou Bubble Green	10

4.2. Pricing and Replenishment Strategies

First of all, this paper wants to predict the data for the next 1 day, this paper uses a linear regression analysis model, which is implemented through the fitting toolbox in MATLAB, and it can be found that the final daily sales volume with respect to the time of the day is:

Ginger, Garlic, and Millet Pepper Combo Pack (small portion):

$$n_1 = -1.2x + 36.3 \tag{14}$$

Sweet Cabbage:

$$n_2 = -8.39x + 38.8 \tag{15}$$

By analogy one can find the sales volume as a function of time for these 33 individual items.

This question adds more constraints to the previous one, with a total of more than 50 items appearing between June 24, 2023, and June 30, 2023, and the merchant keeps the number of available items between 27 and 33 each to save selling space:

$$27 \leqslant N \leqslant 33 \tag{16}$$

Requiring a minimum display quantity of 2.5kg for a single item order, this paper can hereby convert the single item order quantity into the total number of single items sold on the date of July 1, 2023:

$$y_{k8} \geqslant 2.5 \tag{17}$$

To maximize the profit, the profit is taken as the objective function, which can be written according to the given data:

$$w = x'_{kj} y'_{kj} - \frac{x'_{kj}}{1 - P'_k} Z'_k \tag{18}$$

So, the final planning model can be written as:

$$\max\ w = x'_{kj} y'_{kj} - \frac{x'_{kj}}{1 - P'_k} Z'_k$$

$$s.t. \begin{cases} 27 \leqslant N \leqslant 33 \\ y_{k8} \geqslant 25 \\ w = x'_{kj} y'_{kj} - \frac{x'_{kj}}{1 - P'_k} Z'_k \\ n_1 = -1.2x + 36.3 \\ . \\ . \\ . \\ n_{33} = -8.39x + 38.8 \end{cases} \tag{19}$$

The pricing and replenishment of individual items can be found through Matlab.

5. Conclusion

This study simplifies and visualizes the pattern of change in vegetable sales data using descriptive statistical analysis models and histograms and line graphs. The application of the Spearman correlation test demonstrates a high degree of reliability in data sets where outliers exist. The subsequent analysis of fitness reveals a high degree of fit relationship, while the application of the planning model makes the problem-solving process concise and clear. The implementation of the inventory decision model for just-in-time replenishment based on actual demand helps to manage inventory levels more accurately, which is particularly suitable for scenarios with short delivery times and stable demand. However, there are limitations to this study, such as fitting and machine learning can only produce approximate solutions, and the references to data averages on certain categories do not accurately reflect the overall relationship. In summary, this paper provides useful insights into the regularity of vegetable sales and inventory management, and future research can further explore the improvement of model accuracy and the effective application of data in decision-making for superstore operations.

References

[1] LI Wanyin, YANG Yanli. Research on employee training system of H fresh food supermarket under new retail mode [C]// Chengdu Information Engineering University, School of Management. Proceedings of the Symposium on Labor Security Research 2023. [Publisher unknown], 2023:4. DOI: 10.26914/c.cnkihy.2023.024363.

[2] Yu-Ying Song. Research on optimization of marketing strategy of MH fresh food supermarket [D]. Henan University of Finance and Politics, 2023. DOI:10. 27113/ d.cnki. ghncc. 2023. 000443.

[3] Yang Shuai, Huang Xiangmeng,Wang Junbin. Research on joint optimization strategy of shelf allocation and pricing for fresh food[J]. Supply Chain Management,2022,3(08):49-59. DOI:10.19868/j.cnki.gylgl.2022.08.005.

[4] Yuan Li, Yutian Liu,Li-Wei Feng. Nonlinear dynamic process feature extraction and fault detection based on Spearman correlation analysis[J]. Journal of Shandong University of Science and Technology(Natural Science Edition), 2023, 42 (02): 98-107.DOI: 10.16452/j.cnki.sdkjzk.2023.02.011.

[5] Li XF, Chang ZP,Gao YZ et al. Prediction of headless rivet installation interference using multivariate nonlinear regression model [J/OL]. Journal of Xi'an Jiaotong University, 2024(01):1-10[2023-10-27].http:// kns.cnki.net/ kcms/detail/ 61. 1069.t.20230904.1848.004.html.

[6] XU Xinhua, Gao Gang,XU Yongjia. An integrated energy optimization scheduling method based on improved genetic algorithm[J]. Electrical Switch,2023,61(05):35-38+42.

[7] LIU Huiheng, ZHENG Peiping,YANG Hui et al. Multi-objective optimization study on thermal insulation scheme of winter Olympic temporary office building envelope under severe cold mountain conditions based on entropy weight-TOPSIS method[J/OL]. HVAC:1-7 [2023-10-27]. https:// doi.org/ 10.19991/j.hvac1971.2023.11.25.