

Research on Establishing Inbound Strategies for Supermarkets Based on Time Series and Multi-Objective Planning Models

Jialin Qian*

School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai, China

*Corresponding Author: yayuqwq@gmail.com

ABSTRACT

This paper focuses on the pricing and replenishment decision problems for vegetable products in supermarkets. Pricing and replenishment decisions need to study the interconnections between categories and individual products, so as to make reasonable sales combinations; it also needs to study the correlation between sales volume and time, and predict the future purchase price and sales volume, through visual analysis and correlation analysis, multiple regression model, reveals the interconnections between categories and individual products, and combines the correlation between sales volume and time to predict the future price and sales volume, and developed a profit-maximizing replenishment pricing strategy. For each vegetable category, a multi-objective planning mathematical model was developed using multiple regression and time series models to determine the optimal total daily replenishment and time-of-day pricing strategies. At the individual item level, the entropy weight TOPSIS model was used to comprehensively evaluate and rank the individual items, and the top 33 individual items were selected as the target for stocking, and a multi-objective model was further established to develop the replenishment quantity and pricing strategy for July 1, 2023 for individual items. By focusing on verifying the impact of market price trends, this paper conducts a time series forecasting analysis by analyzing the relevant sales data of a large superstore in order to comprehensively optimize the superstore's vegetable selling strategy.

KEYWORDS

Multi-objective planning, Multiple regression, Monte Carlo simulation, ARMA time series forecasting

1. INTRODUCTION

As the quality of life improves, dietary needs, especially vegetable consumption, have become a focus of attention in Chinese society. China is the world's largest producer of vegetables, and the vegetable industry contributes significantly to the country's economy. However, vegetable commodities pose inventory management challenges for merchants due to their short freshness period and time-sensitive sales. Fresh produce merchants need to decide daily on replenishment strategies based on historical sales, customer demand, and the characteristics of the commodity (many varieties, many origins, and early arrival), and use "cost-plus pricing" to set prices without determining costs. Sales and supply volumes vary over time, and shipping losses and product quality affect pricing, so market analysis is critical to effective replenishment and pricing [1].

In this paper, we collect the commodity information of 6 vegetable categories and 252 individual vegetable items of a supermarket for three years, as well as its sales flow details and wholesale data from July 1, 2020 to June 30, 2023, and the loss rate of merchandise purchase, to explore the distribution pattern and interconnections of each category of vegetables, as well as to analyze the

relationship between the total amount of sales of each vegetable and the cost plus pricing on the basis of the category, and to design a pricing plan under the circumstance of maximizing the profitability of the commodity supermarket. In order to maximize the profitability of the supermarket, design and propose the total replenishment and pricing strategy for each vegetable in the coming week.

2. RELEVANCE ANALYSIS

2.1. Statistical descriptions

After pre-processing the data, we first visualize the data by category. The use of Excel pivot table to derive the total sales volume of each category during the three-year period and the total sales volume of the top ten single product histograms in Figure 1 and 2. It can be clearly seen that, in each category: the total sales volume of foliage category is the highest, followed by pepper, eggplant total sales volume is the lowest; sales volume of the top ten single product in the total sales volume of the Wuhu green pepper is the highest, followed by broccoli.

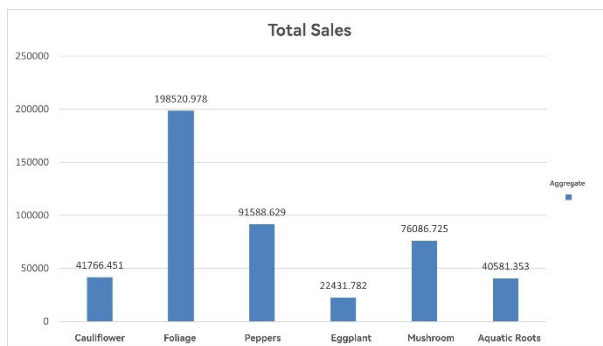


Figure 1. Histogram of total sales volume of vegetable category

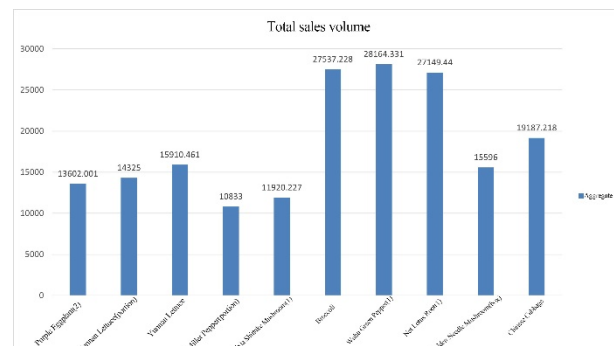


Figure 2. Histogram of Total Sales of Vegetable Singles (Top 10)

In this paper, we found that there is often a certain correlation between the sales volume and time of the vegetable category. In order to ensure the comprehensiveness of the model study, we respectively take the hour, month and quarter as the unit, and make the line graph of the sales volume of each category according to the pivot table as in Figure 3, to analyze the distribution law of the sales volume in different time dimensions.

From the hourly dimension, it is obvious that the peak sales volume of each category is reached around 10:00 a.m. and 5:00 p.m., which indicates that these two time points are the peak periods for trading vegetable products, i.e., the morning and evening peaks. The trough value of sales volume for each category (except early morning and night when the superstores open and close) is reached at around 13:00 pm, indicating that this is the trough period for trading of vegetable commodities.

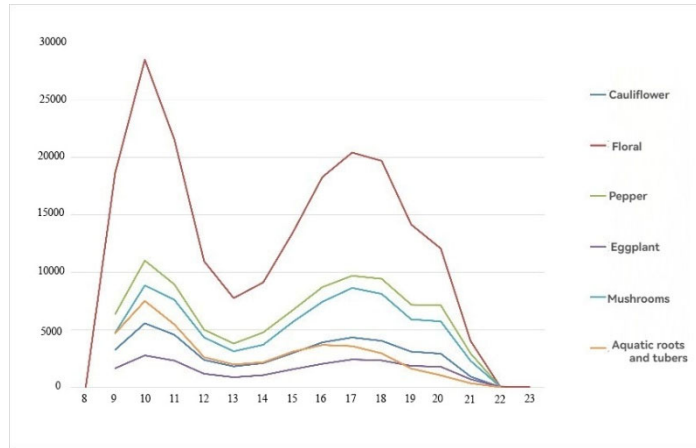


Figure 3. Sales volume of each category at different points in time

The month dimension can be seen in Figure 4, where different types of vegetables are subject to different seasonal factors and have different annual sales trends. However, there is a general increase of different magnitudes in June-August and November-December, and a decrease of different magnitudes in January-May. On the whole, the sales volume of vegetable commodities show a "down-up-down" trend during the year.

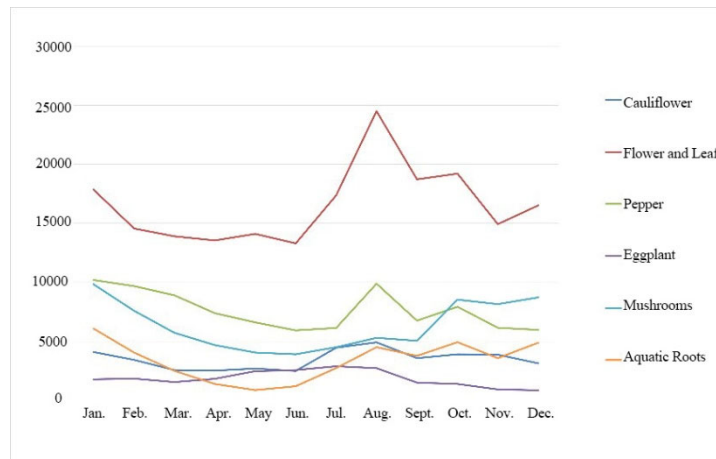


Figure 4. Sales volume of each category in different months

From Figure 5, it can be seen that the quarterly dimension, different categories of vegetables in each quarter will have different ranges of increase, based on the trend of change in sales volume, it can be roughly concluded that foliage, cauliflower and peppers in the "decline - rise - decline", edible mushrooms and aquatic roots and tubers in the "decline - rise - rise," and eggplant in the "rise - rise - decline".

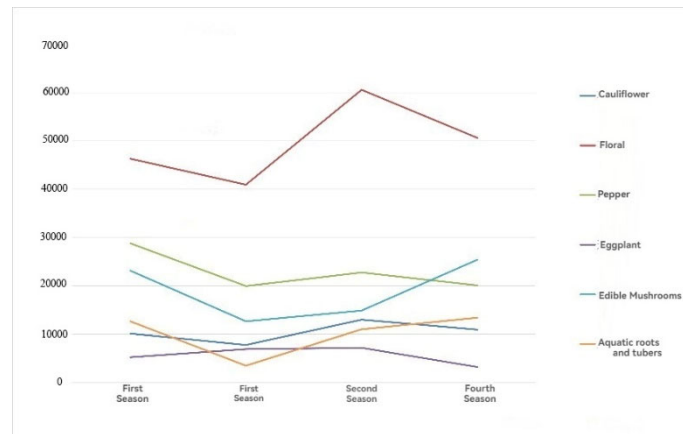


Figure 5. Sales volume of each category in different quarters

3. CORRELATION ANALYSIS

After data visualization and analysis, the correlations among different categories of vegetables were further analyzed. Correlation analysis of six different categories of vegetables was carried out using SPSS software.

1) Make a matrix scatterplot between the categories as in Figure 6, from the scatterplot can be found that there is a certain linear relationship between certain categories.

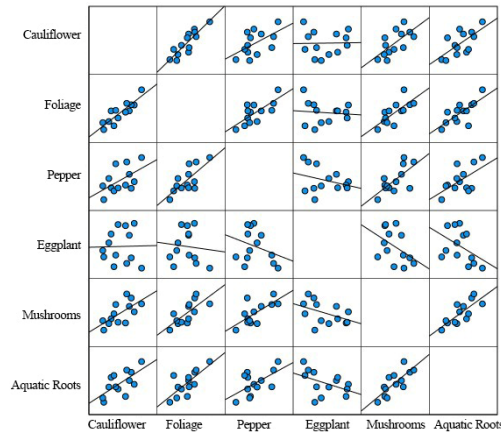


Figure 6. Scatterplot of the matrix for each category

2) The corresponding Pearson correlation coefficients between each vegetable category were calculated using MATLAB software programming as in Table 1.

Table 1. Pearson's correlation coefficient for each vegetable category

R	Cauliflower	Foliage	Chili	Eggplant	Mushroom	Aquatic rootstocks
Cauliflower	1.00000	0.87413***	0.54591**	0.01884	0.67072**	0.67072**
Foliage	0.87413***	1.00000	0.73526	-0.09468	0.73040***	0.73420***
Chili	0.54591**	0.73526***	1.00000	-0.29728	0.68436	0.57918**
Eggplant	0.01884	-0.09468	-0.29728	1.00000	-0.43833	-0.43890
Mushroom	0.67072**	0.73040***	0.68436***	-0.43833	1.00000	0.7786***
Aquatic rootstocks	0.67072**	0.73420***	0.57918**	-0.43890	0.7786***	1.00000

3) Analyze the Pearson correlation coefficient table.

At 99% confidence intervals, foliar species are correlated with cauliflower, edible mushrooms with foliar species, edible mushrooms with chili peppers, aquatic rhizomes with foliar species, and aquatic rhizomes with edible mushrooms.

At 95% confidence interval, chili peppers are correlated with cauliflower, edibles with cauliflower, aquatic rhizomes with cauliflower, aquatic rhizomes with chili peppers, at 95% confidence interval [2].

And eggplant category was negatively correlated with cauliflower, chili pepper, aquatic rhizomes and edible mushrooms, while all other variables were positively correlated.

4. CATEGORY PRICING AND REPLENISHMENT STRATEGIES

4.1. Modeling

On the basis of the above, we have established that time has an effect on the total sales volume of each vegetable category. Next, in order to explore the relationship between the total sales volume of each vegetable category and cost-plus pricing, the effect of time on these two figures is also considered [3].

After reviewing the relevant literature, cost-plus pricing refers to the cost of goods based on the demand of the retail body plus an appropriate proportion of profit, which ultimately forms the retail price of goods.

Utilizing the EXCEL calculation function, integrate the relevant data to get the single product information (including the name of the single product, the category of the single product, the loss rate of the single product, the cost-plus pricing of the single product, etc.) for each account flow of the superstore from 2020.7.1-2030. 6. 30.

Cost-plus pricing is related to the selling price of the goods, and the selling price is related to whether or not the discount promotion; therefore, we make a cross-plot of the sales volume at each point of time in a day and the frequency of discount promotion at each point of time as in Figure 7.

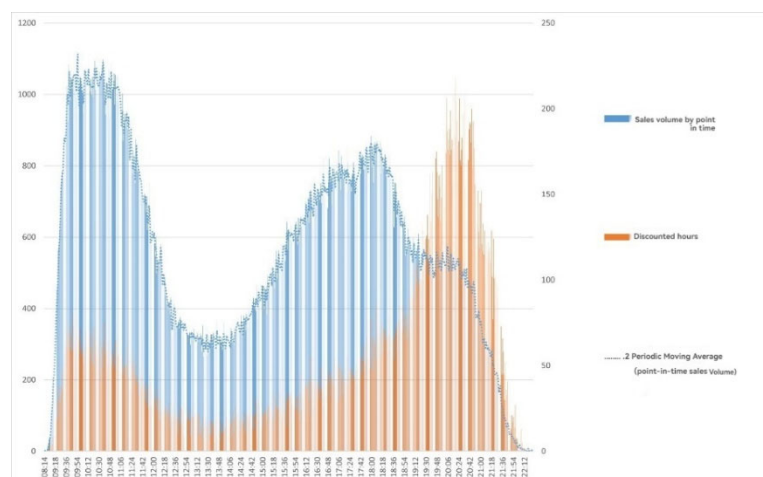


Figure 7. Cross-plot of sales volume at each point in time and frequency of discount promotions at each point in time

Analysis of the cross-tree plot: in addition to the morning and evening peaks in sales volume found just above, there are some small fluctuations in sales volume at the 18:30 pm discount peak at that time. And the increase before each sales peak is different, which may have some time error effects on the relationship between total sales and cost-plus pricing for each vegetable category.

Therefore, in order to ensure the robustness of our modeling, we divide the data into three special fluctuations: 8:00-13:30, 13:31-18:30, and 18:31-23:00. named, respectively, the sales volume of vegetables in the morning, the sales volume of vegetables in the afternoon, and the sales volume of vegetables in the evening. The sales volume of six varieties of vegetables and different data relationship models are established from the perspective of three different time periods.

(1) Using EXCEL's pivot data table to organize the data, the sales and cost pricing of the six types of vegetable varieties and other variables were extracted for the three time periods from 2020.7.1-2023.6.30, every day in the morning, afternoon and evening, respectively. A total of 18 sets of data were integrated, each containing 1084 time samples as shown in Table 2 below.

(2) Correlation tests were carried out on five variables for the 18 sets of data. It is not difficult to find that the p-value of sales volume and cost-plus pricing is greater than 0.05, the original hypothesis is

rejected, the correlation test is not passed, and it is not possible to establish a simple linear model of sales volume and cost-plus pricing, and it is necessary to add the interaction variables to facilitate the establishment of the model.

Table 2. Table of correlation coefficients between sales and cost pricing and other variables

	Cost-plus pricing	Sales volume (kg)	Sales unit price (yuan/kg)	Wholesale price (yuan/kg)	Loss rate (%)
Cost-plus pricing	1(0.000***)	-0.009(0.768)	1(0.000***)	0.232(0.000***)	-0.235(0.000***)
sales volume	-0.009(0.768)	1(0.000***)	-0.009(0.765)	-0.009(0.762)	0.008(0.796)
Sales unit price (yuan/kg)	1(0.000***)	-0.009(0.765)	1(0.000***)	0.241(0.000***)	-0.244(0.000***)
Wholesale price (yuan/kg)	0.232(0.000***)	-0.009(0.762)	0.241(0.000***)	1(0.000***)	-0.372(0.000***)
Loss rate (%)	-0.235(0.000***)	0.008(0.796)	-0.244(0.000***)	-0.372(0.000***)	1(0.000***)

(3) Organize the data again, add the variable of the sum of daily cost-plus pricing, and carry out the correlation test again to get the correlation coefficients shown in Table 3, the *P*-values of the three variables are less than 0.05, and pass the correlation test to establish the multiple regression model [4].

Table 3. Table of correlation coefficients for cost-plus pricing, cost-plus pricing composite, and sales volume

	Cost-plus pricing	Sales volume (kg)	Total cost-plus pricing
Cost-plus pricing	1(0.000***)	-0.236(0.000***)	0.158(0.000***)
Sales volume (kg)	-0.236(0.000***)	1(0.000***)	0.858(0.000***)
Total cost-plus pricing	0.158(0.000***)	0.858(0.000***)	1(0.000***)

(4) Based on the above, it is necessary to establish the relationship between total sales and cost-plus pricing for each vegetable category, so we make two regression equations, the first: the independent variable is cost-plus pricing and total cost-plus pricing; the dependent variable is sales. The second: the independent variable is cost-plus pricing; the dependent variable is the sum of cost-plus pricing. Join the two equations to get the model we need. An example of the morning cauliflower category data is shown in Table 4:

Table 4. Analysis of regression results

Linear regression analysis results n=1084									
	Non-standardized coefficient		Standardized coefficient	t	P	VIF	R ²	Adjustment of R ²	F
	B	standard error	Beta						
constant	16.524	0.465	-	35.552	0.000***	-	0.877	0.877	F=386 5.823 P=0.000***
Cauliflower: cost-plus pricing average price	-1.629	0.046	-0.381	-35.344	0.000***	1.026			
Cauliflower: cost-plus pricing aggregate	0.048	0.001	0.918	85.084	0.000***	1.026			

Dependent variable: cauliflower: sales (kg)

The analysis of the results of the *F*-test can be obtained that the significance *P*-value is 0.000***, which presents significance at the level and rejects the original hypothesis that the regression

coefficient is 0. Therefore, the model basically meets the requirements. For the covariate covariance performance, VIF is all less than 10, so the model has no multicollinearity problem and the model is well constructed.

The model equation is as follows: $y = 16.524 - 1.629 * \text{Cauliflower: Cost Plus Pricing Average Price} + 0.048 * \text{Cauliflower: Cost Plus Pricing Total}$.

Similarly, another model equation is obtained as: $y = 178.639 + 12.83 * \text{Cauliflower Category: Cost Plus Pricing Average Price}$.

Collate to get:

$y = 16.524 - 1.629 * \text{Cauliflower: Cost-plus Pricing Average} + 0.048 * (178.639 + 12.83 * \text{Cauliflower: Cost-plus Pricing Average})$.

Following the above steps, three separate models are built for each vegetable category, which are added up to obtain a model of one day's sales.

4.2. Solving the model

In the above section, the daily sales volume, purchase price and wastage rate of each category of vegetables have been organized. For the time series data, we have used the traditional time series forecasting model to predict the sales volume as well as the incoming price of each category of vegetables for the next seven days (2023.7.1-2023.7.7) in order to facilitate the subsequent development of a mathematical model for multi-objective planning. Since the rate of wastage does not change with time, it is not analyzed. Time series modeling is a theory and method of building mathematical models through curve fitting and parameter estimation based on time series data obtained from systematic observation. Common models include the ARMA model [5].

SPSS software was used to perform time series analysis of two sets of time series data of sales volume and purchase price. After importing the sales volume time series data, the model fit table was obtained by SPSS analysis, and the *R*-squared value was 0.782, the smooth *R*-squared value was 0.597, and the goodness-of-fit was qualified; and the prediction model of the six categories basically passed the residual white noise test with a confidence level of 0.95. Therefore, the time series forecasting model is considered reliable. The results of the predicted sales of each category of vegetables for the next seven days are shown in Table 5:

Table 5. Forecasted sales results for each category of vegetables for the next seven days

Sales volume (kg)	cauliflower	foliage	chili	eggplant	mushroom	Aquatic rootstocks
2023-7-1	22.91	141.04	77.26	22.23	49.35	20.06
2023-7-2	20.02	136.38	77.52	21.68	49.54	19.41
2023-7-3	19.60	132.26	80.30	18.71	52.16	19.39
2023-7-4	20.51	137.33	80.64	17.33	52.12	19.93
2023-7-5	21.63	134.91	78.96	17.66	53.60	19.73
2023-7-6	23.70	132.18	80.55	19.91	51.24	20.18
2023-7-7	24.63	129.20	82.11	22.44	49.11	19.95

The *R*-squared value of the import price time series data obtained after the SPSS analysis of the prediction is 0.853, and the smooth *R*-squared value is 0.607, which is a good fit; and the prediction model of the six categories basically passes the residual white noise test with confidence level of 0.95. Therefore, the time series forecasting model is considered reliable [6]. The results of the predicted incoming prices for the next seven days for each category of vegetables are shown in Table 6:

Table 6. Forecasted results of incoming prices for each category of vegetables for the next seven days

Purchase price (yuan/kg)	cauliflower	foliage	chili	eggplant	mushroom	Aquatic rootstocks
2023-7-1	6.75	3.73	5.43	7.32	7.59	14.90
2023-7-2	6.15	3.73	5.43	7.38	7.34	14.89
2023-7-3	5.57	3.73	5.43	7.36	7.34	14.88
2023-7-4	4.99	3.73	5.43	7.33	7.34	14.87
2023-7-5	4.40	3.73	5.43	7.36	7.34	14.86
2023-7-6	3.82	3.73	5.43	7.39	7.34	14.85
2023-7-7	3.24	3.73	5.43	7.42	7.34	14.84

5. INDIVIDUAL PRODUCT PRICING AND REPLENISHMENT STRATEGIES

We summarized the daily sales volume of individual vegetables in each category to derive the daily sales volume of vegetables in each category; averaged the daily sales price of individual vegetables in the same category as a substitute for the daily sales price of each category; averaged the daily wholesale price of individual vegetables in the same category as a substitute for the daily wholesale price of each category; and averaged the attrition rate of individual vegetables in the same category as a substitute for the daily attrition rate of each category. The loss rate of individual vegetables of the same category is averaged to replace the loss rate of each category. The shopping mall needs to replenish the goods without knowing the specific items and the purchase price. Therefore, by substituting the above parameters for each vegetable category into Equation 5, the daily cost-plus pricing for each vegetable category can be obtained, and it is also easy to establish the relationship between the total sales volume of each vegetable category and the cost-plus pricing.

The profit margin P satisfies the relation:

$$P = \frac{N}{M} \quad (1)$$

The relationship equation is satisfied between the selling price of vegetables N and the cost-plus pricing X :

$$N = \frac{M}{M + K} \cdot X \quad (2)$$

The relationship between sales volume and cost-plus pricing of vegetables:

$$Y = F(X) \quad (3)$$

Also the value of profit satisfies the relation:

$$W = P \cdot Y \cdot N \quad (4)$$

Substituting the above equations gives the final objective function:

$$W = \frac{M}{(M + K)^2} \cdot X \cdot F(X) \quad (5)$$

Observing the above equation, the purchase price M and the wastage rate K can be substituted into the relevant values according to the specific category, so the total profit W is only related to the cost-

plus pricing. The objective function for specific vegetable categories needs to be substituted into the relevant values to obtain a specific objective function.

The following constraints are given based on the interrelationships between the data:

Constraint 1: The sales volume should not be greater than the stocking volume.

The incoming quantity of each vegetable category is based on the predicted sales volume obtained using time series analysis as described above, so the regression equation of sales volume on selling price should be satisfied:

$$Y \leq Y_{ij} \quad (6)$$

Constraint 2: Pricing should be greater than the purchase price and less than two times the purchase price.

Obviously, the pricing should be greater than the purchase price. It was found through the data that the pricing of vegetable commodities is usually not higher than two times of the purchase price. Since the exact inlet price cannot be known with certainty, the predicted inlet price obtained based on the above using time series analysis is used as a reference standard. The mathematical expression for the constraint should be:

$$M < X < 2M \quad (7)$$

Constraint 3: The selling price of vegetables in the morning should be greater than that in the afternoon and lowest in the evening.

According to the vegetable market pattern shown in the data and relevant information, the selling price of vegetables in a day should usually meet the highest in the morning, followed by the afternoon and the lowest in the evening:

$$X_1 > X_2 > X_3 \quad (8)$$

After the objective function and constraints are established, the idea of Monte Carlo simulation is utilized and traversal is implemented with the help of matlab programming to obtain the maximum value of the objective function and the selling prices of the three time periods corresponding to the maximum value. Similarly, the selling price corresponding to the time when the profit of the remaining categories of vegetables is maximized can be found. Based on the results obtained, the total amount of replenishment and the pricing strategy for the coming week can be roughly determined [7].

6. CONCLUSIONS

For the interrelationships between vegetable categories and categories, we used correlation analysis. The matrix scatter plot between the categories was first made to determine the existence of a certain linear relationship between the categories. Then MATLAB was used to solve the Pearson's correlation coefficient R and the P value of the hypothesis test between the categories, in order to determine that there is a certain degree of correlation between different categories of vegetables of different individual categories, and it is closely related to time. Next, in order to study the relationship between the total sales volume of each vegetable category and cost-plus pricing, we continue to consider the effect of the time variable. By comparing the sales of vegetables at each time point of the day with the frequency of discount promotions at each time point of the day, the daily data were categorized into three categories, representing morning, afternoon and evening. Organize and extract the required data, and use the multiple regression model to calculate each category into three time periods: morning, afternoon, and evening to get three different regression equations, and then summarize them

to get the regression equations of the sales volume of a single category in a day with the cost-plus pricing of the morning, midday, and evening time periods.

After obtaining the regression equations, we used a time series forecasting model to predict the sales volume as well as the purchase price of each vegetable category for the coming week. Then, the mathematical model of multi-objective planning was established, and the objective function and three constraints of total profit were established based on the regression equations and the predicted results. Finally, the idea of Monte Carlo simulation was used to substitute the model into a matlab program to obtain the replenishment and pricing of each category of vegetables for the next seven days, thus maximizing the daily profit.

REFERENCES

- [1] ZHAO An-Ping, ZHAO You-Sen, WANG Xiao-Dong et al. Study on the balance of supply and demand of vegetables in Beijing[J]. China Agricultural Information,2023,35(02):53-62.
- [2] GUO Yangnan, LEI Shaogang, REN Lixin et al. Detection of spatial heterogeneity of vegetation change based on Pearson's time-series algorithm[J]. Journal of Hunan University of Science and Technology(Natural Science Edition),2023,38(04):119-124.DOI:10.13582/j.cnki.1672-9102. 2023. 04.015.
- [3] An Qi. Research on the pricing of scientific and technical research services based on cost-plus pricing method[J]. Library Research and Work, 2021, (10):25-31+24.
- [4] SUN Jiya, LIU Weining, WU Qiang et al. Research on the cost-effectiveness of different scales of hog farming--an analysis based on TOPSIS-multiple regression model[J/OL]. Price Theory and Practice,1-5[2024-02-20]. <https://doi.org/10.19851/j.cnki.CN11-1010/F.2023.11.439>.
- [5] Nie Shuyuan. Comparison of GDP forecasting effects based on four types of time series models[J]. Statistics and Decision Making,2024,40(02):63-67. DOI:10.13546/j.cnki.tjyjc.2024.02.011.
- [6] MA Yaling, YANG Xiangfei, XIN Congcong. Construction of Evaluation Index System for Comprehensive Measurement of Financial Risk of Railway Logistics Based on SPSS Analysis[J]. China Logistics and Purchasing, 2023, (15):107-108.DOI:10.16079/j.cnki.issn1671-6663. 2023. 15. 023.
- [7] Guo YG, Wang XN. Research on the calculation of the risk rate of diffuse dam based on the improved Monte Carlo method[J]. People's Yellow River,2023,45(S2):102-104.