Advances in Computer, Signals and Systems (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2023.070909
ISSN 2371-8838 Vol. 7 Num. 9

Automated Pricing and Replenishment Decisions for Supermarket Fresh Vegetables

Zhichao Zhang[1,†], Rinong Wu[2,†], Haolin Cui[3,†]

[1]*College of Information Engineering, Inner Mongolia University of Technology, Hohhot, China*
[2]*College of Science, Inner Mongolia University of Technology, Hohhot, China*
[3]*College of Aeronautics, Inner Mongolia University of Technology, Hohhot, China*
[†]*These authors also contributed equally to this work.*

Keywords: Correlation Analysis; Linear Programming; Gray Forecasting; Automatic Pricing; Replenishment Decisions

Abstract: In today's vegetable superstore market, vegetable items have a short shelf life due to their short shelf life. Supermarkets usually replenish the goods on a daily basis based on the historical sales and demand of each item. Therefore, this paper conducts a relevant research on automatic pricing and replenishment decisions for vegetable items based on the measured data of a superstore. First, the trends of different categories under different seasons are plotted. Then, Python linear regression is used to fit the functional relationship equation between sales volume and cost-plus pricing, and an optimization model is constructed with the total daily replenishment as the decision variable and the superstore's revenue as the objective function, so as to derive the predicted sales volume table and pricing strategy table for each category. Finally, the gray prediction model is used to predict and analyze the sales volume of individual items, so as to maximize the superstore's revenue under the premise of trying to meet the market demand for each category of vegetable goods. The model developed in the paper can help superstores predict demand more accurately, make replenishment plans, adjust pricing strategies, and improve market competitiveness.

1. Introduction

In today's fresh food superstores, the freshness period of all vegetable commodities is relatively short, and the quality deteriorates with the increase of selling time, such as of course not sold, the next day can not be sold. Therefore, superstores usually make pricing and replenishment decisions based on the sales and demand of each commodity [1].

Considering the realities of the situation, such as the variety of vegetables, origin, trading hours, etc., merchants have to make pricing and replenishment decisions accordingly in the absence of certainty. Vegetables are priced using the "cost-plus pricing" method, and superstores usually sell at a discount for goods with shipping losses and poor quality. Reliable market demand analysis is important for replenishment and pricing decisions. From the demand side, there is often a correlation between the sales volume of vegetables and the time of day; from the supply side, the supply of vegetables is more plentiful in certain months, and the limitations of the superstore's sales space make a reasonable sales mix extremely important.

In order to solve the above problems, this paper takes the measured data of a superstore as an example, firstly, to find out the distribution law of the sales volume of each vegetable category and single product and their interrelationship. Then, analyze the relationship between the total sales volume of each vegetable category and the cost-plus pricing, and give the total daily replenishment volume and pricing strategy of each vegetable category in the coming week to maximize the revenue of the superstore. Finally, due to the limited sales space of vegetable items in the superstore, a replenishment plan for new individual items is formulated to maximize the revenue of the superstore under the premise of trying to satisfy the market's demand for vegetable items in each category [2].

2. Model formulation and solving

2.1 Sales volume distribution pattern and correlation analysis

Descriptive statistics of sales by category were analyzed as shown in Table 1. The foliage category is the largest category in terms of sales, while the aquatic rhizome category is the smallest category in terms of sales, so this data will serve as the key measurement basis for our individual product indicators below.

Starting from spring, in terms of the overall trend, the sales of aquatic roots and tubers showed a continuous increase, while other categories had a tendency to decline; in summer, due to the warmer temperatures, the sales of vegetables increased and fruits decreased; in winter, due to the cold weather, the sales of aquatic roots and tubers continued to increase, while other parts of the category were affected by the weather and experienced a decline. By analyzing these data, it can be seen that different seasonal factors have a significant impact on the sales of different categories, thus contributing to the development of the market as a whole.

Pearson is used to calculate the correlation coefficient when both variables are normal continuous variables and they show a linear relationship [3]. In statistics, known as the Pearson correlation coefficient, r or Pearson is commonly used in articles to de-measure the correlation between two variables, which has a value given between -1 and 1. In the natural sciences, it is used to measure the degree of association between two variables and to measure the linear relationship of the variables. Pearson's correlation coefficient is calculated as:

$$r = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{N\sum y_i^2 - (\sum y_i)^2}\sqrt{N\sum x^2 - (\sum x_i)^2}} \tag{1}$$

The closer the correlation coefficient is to 1 or -1, the larger the absolute value of the correlation coefficient is, and the stronger the correlation is the closer the correlation coefficient is to 0, the weaker the correlation is. Through the above correlation analysis, we arrive at the correlation results as shown in Table 1.

Table 1: Correlation coefficients for various types of vegetables

	philodendron	cauliflower	Aquatic rhizomes	eggplant	capsicum	edible mushroom
philodendron	1	-0.38	-0.23	-0.22	-0.13	-0.18
cauliflower	-0.38	1	0.20	0.51	0.44	0.51
meristem	-0.23	0.20	1	0.51	0.16	0.24
eggplant	-0.22	0.51	0.32	1	0.66	0.86
capsicum	-0.13	0.44	0.16	0.66	1	0.95
edible mushroom	-0.19	0.51	0.24	0.86	0.95	1

An analysis of the correlations corresponding to the categories in the above table shows that. Among them, there is a negative correlation between foliage and cauliflower, aquatic rootstocks, eggplant, chili and edible mushrooms, i.e., the abundance of foliage may reduce the abundance of

other categories. This may be due to reasons such as competition between plants or limited resources in the same season or under the same growing environmental conditions. There was a positive correlation between cauliflower and aquatic rootstocks, eggplant, chili and edible mushrooms, while there was a negative correlation with flowering and foliage categories. This may be caused by reasons such as similarity in growing environmental conditions among them or competition for resources among them. While eggplant, pepper and edible fungi have a strong correlation between the three of them, in conjunction with the actual situation this may be the influence of artificial selection between them, thus showing the corresponding correlation pattern.

The correlation heat map of various vegetables is shown in Figure 1. It is important to note that in Heat Map 1 x_1 for the phanerogamous species, x_2 is the cauliflower class, x_3 for aquatic rhizomes, x_4 is the eggplant class, x_5 is the pepper group, x_6 for edible mushrooms.

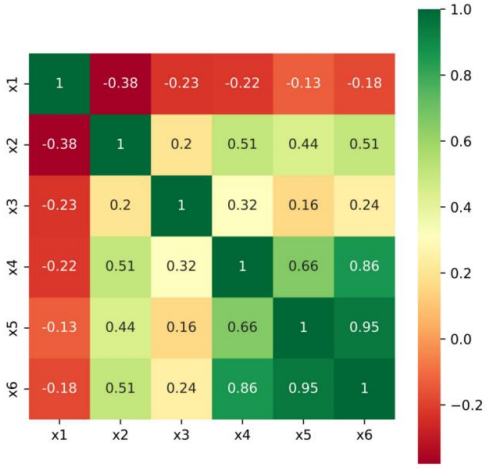

Figure 1: Heat map of correlation of various types of vegetables

It is known that the category of flowers and leaves has the highest percentage of sales, here we will filter the sample of single products in this category. Through Appendix I and Appendix II, we counted the top ten items in the flower and leaf category with the highest data sample size. They are, Yunnan oleander, Chinese cabbage, yellow cabbage, Yunnan lettuce, spinach, milky cabbage, sweet potato tips, choy sum, corns, and baby vegetables (in descending order). Taking them as the measurable indicators, they go to roughly predict the distribution pattern and characteristics of measuring the sales of single items in other categories, as shown in Figure 2:

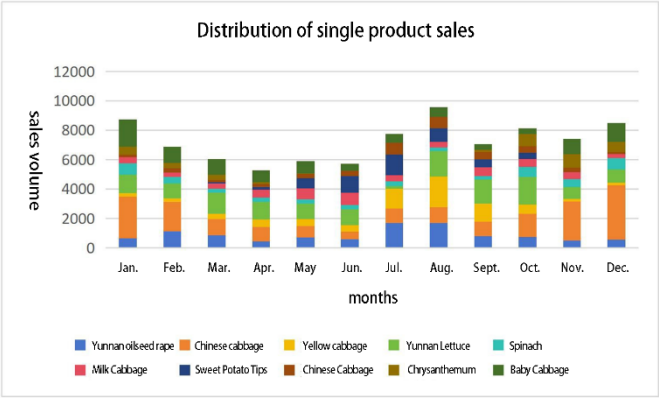

Figure 2: Distribution of sales by individual product

Through the analysis of the above chart can be seen, Yunnan oil wheat cabbage in addition to higher sales in July and August, its sales in the year are in a relatively stable situation; we will be cabbage, yellow cabbage, milk cabbage as a class for analysis can be seen, cabbage in the autumn

and winter seasons sales are higher and show a trend of increasing and then decreasing in the spring and summer seasons, sales are lower to show a relatively stable state. Yellow cabbage and it is just the opposite, indicating that the two have a certain pattern of seasonal change and substitution. Milk cabbage sales tend to be a relatively stable trend throughout the year. Overall, the demand for vegetable single product shows a seasonal trend, in summer and winter relative high demand, in spring and summer relative low demand.

For the correlation analysis of this trivia question, we still use the formula for correlation analysis in the above trivia question and its correlation intensity table, which leads to the correlation results shown in Figure 3.

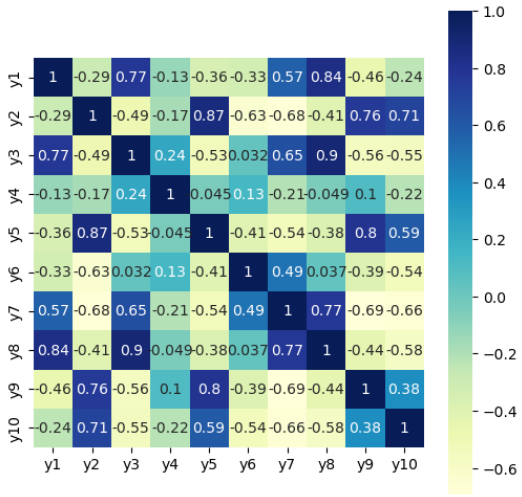

Figure 3: Relevance heat map of each individual product

The heat map above and its coefficients can be analyzed for correlation as follows. Exploring the correlation of individual products in the same category, we here analyze our this chart from an overall perspective. On the whole, negative correlation is dominant, which means that an increase in one of the two variables may lead to a decrease in the other, which is analyzed from a realistic point of view, i.e., most of them are used as substitutes for each other. The smaller number of variables with strong correlations indicates that they are interdependent, i.e., they act as substitutes for each other.

2.2 Individual product replenishment program development

The relationship between sales volume and cost-plus pricing is first fitted using the available data, and in this paper, regression analysis is chosen for fitting and prediction. Regression analysis is a statistical method to study the correlation between random variables. It establishes a quantitative relationship between one variable and another, i.e., a regression equation, by analyzing and calculating the actual observations of the variables. This subsection uses the pricing of each category (y) as the dependent variable and sales volume (x) as the independent variable to perform a one-way regression analysis and build a one-way regression model:

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{2}$$

whereε is the random error, obeying a normal distribution$N(0, \sigma^2)$, , *and* β_0 , *and* β_1 is the regression coefficient. The specific steps of regression analysis are:

Step 1: Determine the regression coefficients from the observations using the least squares method β_0 , *the* β_1 The estimated value of $\widehat{\beta}_0, \hat{\beta}_1$.

Step 2: The regression equation was tested for significance, F-test for linear relationship with t-test for regression coefficients.

Step 3: Prediction using regression equations

A one-way linear regression was performed using Python programming, and the resulting regression equation was calculated using aquatic rhizomes as an example:

$$y = 0.6149 - 0.0061x \tag{3}$$

The regression equation is tested for significance. The first test is the test of linear relationship, which is to test whether the linear relationship between the independent variable and the dependent variable is significant or not, and its test statistic is:

$$F = \frac{MSR}{MSE} \tag{4}$$

whereMSR is called the mean square regression andMSE is called the mean square residual. It is obtained that$F = 2703.557$ the significance of the relationship between the independent variable and the dependent variable$p = 0.000$, it is considered that there is a significant linear relationship between the independent variable and the dependent variable, so it passes the linear relationship$F$ test.

Next, the significance of the regression coefficients is tested, the significance test of the regression coefficients is to test the significance of the effect of the independent variable on the dependent variable, and its test statistic is:

$$t = \frac{\hat{\beta}_i - \beta_i}{s_{\hat{\beta}_i}} \tag{5}$$

where$s_{\hat{\beta}_i} = \frac{Se}{\sqrt{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}}$, is called the$\hat{\beta}_i$ the estimated standard deviation and the results are obtained as shown in Table 2.

Table 2: Table of regression results

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.6149	0.002	311.226	0	0.61	0.62
x	-0.0061	0	-51.996	0	-0.006	-0.006

According to the above graph then the effect of the independent variable on the dependent variable is considered to be significant and the regression coefficient of the$t$ test is passed.

Finally, the coefficient of determination is calculatedR^2 to determine the goodness of fit of the regression line, which depends on the magnitude of SSR and SSE and is calculated as:

$$R^2 = \frac{SSR}{SST} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \tag{6}$$

R^2The closer to 1, the better the fit, and $R^2 = 0.998$, indicating that the fit is excellent. In summary, the regression equation equation is significant.

Through the above series of tests, it shows that the one-way linear regression is reasonable. From the regression results, it can be seen that the regression coefficients are all negative, which is consistent with our prior expectations and in line with the actual situation, i.e., an increase in sales volume will lead to a decrease in pricing, i.e., the sales volume of the goods and pricing are inversely proportional to each other. Through the one-way regression analysis, it is found that the functions are all decreasing functions, but the degree of their reduction is small, so it can be concluded that the total sales of each vegetable category is negatively correlated with the cost-plus pricing, but it is not obvious, so although it has a linear correlation in nature but the impact is not significant.

For solving the total daily replenishment and pricing strategy for each vegetable category for the coming week, we use the cauliflower category as an example for solving the problem, and the rest of

the dishes eventually show the calculation results.

Set the decision variable to$x_{icauliflower}, i = 1，2，3 \ldots \ldots 7$，denoting the total number of sales of cauliflower category on day$i$ day total sales of cauliflower category.

Since pricing is done using cost-plus pricing, cost-plus pricing is a pricing strategy that is commonly used in manufacturing and service industries to determine the selling price of a product or service. The core idea of this pricing method is to combine costs with the required profit to determine the final selling price. Thus, the daily rate is set to be$W_{icauliflower}$，the final pricing$y_{icauliflower}$ can be expressed as $y_{icauliflower} = c_{icauliflower}(1 + W_{icauliflower})$.

Where.$c_{icauliflower}$ denotes the cost of entering the cauliflower category on that day, so for the cauliflower category on day i The profit for the day can be expressed as $W_{icauliflower} = w_{icauliflower} \times x_{icauliflower}$ 。Thus the total profit of the superstore in a week can be expressed as:

$$W_{\text{total}} = \sum(W_{ifoliage} + W_{icauliflower} + W_{iaquatic\,rootstock} + W_{ieggplant} + W_{icapsicum} + W_{imushroom} \quad (7)$$

The above analysis yields the predicted sales volume as shown in Table 3 and the pricing strategy as shown in Table 4.

Table 3: Forecasted sales volume

	philodendron	cauliflower	Aquatic rhizomes	eggplant	capsicum	edible mushroom
July 1	152.94	19.79	20.42	23.28	95.04	54.14
July 2	110.99	14.36	14.82	16.89	68.97	39.29
July 3	96.13	12.44	12.84	14.63	59.74	34.03
July 4	113.61	14.70	15.17	17.29	70.60	40.22
July 5	124.97	16.17	16.69	19.02	77.66	44.24
July 6	138.95	17.98	18.56	21.15	86.35	49.19
July 7	135.46	17.53	18.09	20.62	84.18	47.95

Table 4: Pricing strategy

	philodendron	cauliflower	Aquatic rhizomes	eggplant	capsicum	edible mushroom
July 1	0.64	0.49	0.49	0.57	0.64	0.58
July 2	0.65	0.51	0.52	0.57	0.60	0.60
July 3	0.66	0.52	0.54	0.58	0.60	0.60
July 4	0.65	0.51	0.52	0.57	0.60	0.60
July 5	0.64	0.50	0.51	0.57	0.61	0.60
July 6	0.64	0.49	0.50	0.57	0.62	0.59
July 7	0.64	0.50	0.50	0.57	0.62	0.59

2.3 Individual product replenishment strategy development

The gray prediction model GM(1,1) is a prediction method for predicting gray systems, which is used to predict systems that contain both known and uncertain information [5-6]. This method is mainly through the identification of the degree of dissimilarity between the development trend of the system factors, that is, correlation analysis, and the generation of raw data processing to find the law of the system changes, to generate a strong regularity of the data sequences, and then establish the corresponding differential equation model, so as to predict the status of the development trend of things in the future, and finally get the model of its development [5].

Establishment of a gray prediction model: according to the sequence shows a monotonically

increasing law of exponential form after accumulation, associating the differential equation$y' = ay$ has an exponential form of the solution$y = e^{ax}$ Thus, the first-order gray equation model is proposed, i.e., GM(1,1) model, in which the first 1 represents the first-order differential equation, and the second 1 represents the gray model containing only one variable.

It is known that the reference data column$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), .., x^{(0)}(n))$ that the sequence (1-AGO) is generated by 1-accumulation.

$x^{(1)}$The mean generating sequence for the:

$$z^{(1)} = (z^{(1)}(2), z^{(1)}(3), .., z^{(1)}(n)) \tag{8}$$

of which $z^{(1)}(k) = 0.5x^{(1)}(k)$ +0.5$x^{(1)}(k-1)$, k=2,3,. ...,n.

The GM(1,1) model prediction steps are as follows:

(1) Testing and processing of data

In order to ensure the feasibility of the modeling method, it is necessary to make the necessary tests on the known data columns. Calculate the rank ratio of the reference series

$$\lambda(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)}, k = 2,3, \dots, n \tag{9}$$

If all the grade ratios λ_k all fall within the tolerable coverage $\Theta = (e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}})$ within the tolerable coverage, then the sequence$x^{(0)}$ can be modeled as$GM(1,1)$ of the data for gray prediction, otherwise, the sequence needs to be$x^{(0)}$ to be exchanged so that they fall within the tolerable coverage, i.e., take the appropriate number of normals$c$ and make a translation transformation

$$y^{(0)}(k) = x^{(0)}(k) + c, k = 1,2, \cdots, n \tag{10}$$

such that the sequence$y^{(0)} = (y^{(0)}(1), y^{(0)}(2), \cdots, y^{(0)}(n))$ rank ratio of a sequence

$$\lambda_y(k) = \frac{y^{(0)}(k-1)}{y^{(0)}(k)} \epsilon \Theta, k = 2,3, \cdots n \tag{11}$$

Meet the requirements.

(2) Modeling

Modeling Differential Equations

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b \tag{12}$$

The model is a differential equation of order 1 in one variable, denoted as$GM(1,1)$.

In order to identify the model parametersa, b , in the interval$k-1 < t \leq k$ on the interval, let

$$x^{(1)}(t) = z^{(1)}(k) = \frac{1}{2}\left[x^{(1)}(k-1) + x^{(1)}(k)\right] \tag{13}$$

$$\frac{dx^{(1)}(t)}{dt} = x^{(1)}(k) - x^{(1)}(k-1) = x^{(0)}(k) \tag{14}$$

Then Eq. (14) reduces to a discrete model

$$x^{(0)}(t) + az^{(1)}(k) = b, k = 2,3, \dots, n \tag{15}$$

Eq. (15) is called the gray differential equation and Eq. (14) is called the corresponding whitening equation.

Noting that $u = [a, b]^T, Y = [x^{(0)}(2), x^{(0)}(3), \cdots, x^{(0)}(n)]^T, B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}$, then by the

method of least squares, find the value that minimizes the value of$J(u) = (Y - B_u)^T(Y - B_u)$ reach the minimum value of$u$ The estimate of the value of$\hat{u} = [\hat{a}, \hat{b}]^T = (B^T B)^{-1} B^T Y$,and so solving equation (15.3) yields

$$\hat{x}^{(1)}(t) = \left(x^{(0)}(1) - \frac{\hat{b}}{\hat{a}}\right) e^{-\hat{a}t} + \frac{\hat{b}}{\hat{a}} \tag{16}$$

That is, we get the predicted value.

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{\hat{b}}{\hat{a}}\right) e^{-\hat{a}k} + \frac{\hat{b}}{\hat{a}} \qquad k = 0,1,2 \ldots \tag{17}$$

(not only ...) but also$\hat{x}^{(0)}(1) = \hat{x}^{(1)}(1), \hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \quad k = 0,1,2\ldots$

(3) Error checking

The following two tests can be used:

a. Relative error test

Calculate the relative error:

$$\delta(\mathrm{k}) = \frac{\left|\mathrm{x}^{(0)}(\mathrm{k}) - \hat{x}^{(0)}(k)\right|}{\mathrm{x}^{(0)}(\mathrm{k})} \qquad k = 0,1,2 \ldots \tag{18}$$

Here.$\hat{x}^{(0)}(1) = \mathrm{x}^{(0)}(1)$, ifδ (k) < 0.2, the general requirements are considered to be met; ifδ (k) < 0.1, the higher requirement is considered to be met.

b. Grade ratio deviation value test.

The class ratio is first calculated from the reference series$\lambda(\mathrm{k})$ and then using the development coefficients$\hat{a}$ to find the corresponding grade ratio deviation:

$$\rho(\mathrm{k}) = \left|1 - \left(\frac{1-0.5\hat{a}}{1+0.5\hat{a}}\right)\right| \lambda(\mathrm{k}) \qquad \mathrm{k} = 2,3,\ldots,\mathrm{n} \tag{19}$$

If$\rho(k) < 0.2$, then the general requirements are considered to be met; ifρ(k) <01, the higher requirement is considered to be met.

(4) Predictive forecasting

The predicted values of the specified points are obtained by the$GM(1,1)$ The model obtains the predicted value of the specified point, and gives the corresponding prediction forecast according to the needs of the actual problem.

Linear programming (linear programming, LP) is an important branch of operations research, which originated in the decision-making problem of industrial production organization and management, with the optimal value of ensuring that a number of linear variables satisfy linear constraints. Linear programming three elements of the approximate conditions constitute. Generally speaking, the decision variables are those quantities that the decision maker wants to control in order to achieve the predetermined goal, and the solution of the problem is to find out the final values of the decision variables; the objective function is the index that the decision maker wants to optimize, which is a linear function of the decision variables, describing the relationship between the decision variables and the predetermined goal; these constraints to be meaningful.

The general form of linear programming is:

$$\max(min) \quad z = \sum_{j-1}^{n} c_j x_j \tag{20}$$

$$\text{s.t.} \begin{cases} a_{ij} x_j \leq (\geq, =) b_i \\ x_j \geq 0 \end{cases}, \tag{21}$$

In the above expression, (1) is called the objective function, (2) is called the constraints, and$c$ is called the value vector;x is the decision vector. By using the above principle, the following equation is listed:

$$\max z = \sum_i P_i[Q_i(1-\beta)] - C_i\,Q_i \tag{22}$$
$$\begin{cases} Q_i \geq 2.5 \\ 27 \leq i \leq 33 \\ P_i \geq 0 \end{cases}$$

The solution is solved using the scipy.optimize module of the Python program, first reduced to the standard form in SciPy:

$$\min z = \sum_i -P_i[Q_i(1-\beta)] + C_iQ_i \tag{23}$$
$$\begin{cases} Q_i \geq 2.5 \\ 27 \leq i \leq 33 \\ P_i \geq 0 \end{cases}$$

The final total profit obtained was $670.64.

3. Conclusion

In this paper, based on the measured data of a superstore, we conduct a relevant research on automatic pricing and replenishment decision of vegetable items. The trends of different categories under different seasons are plotted. The correlation relationship existing between the sales volume of different categories and individual items is analyzed, and line graphs, histograms, and Pearson correlation coefficients are considered to be used to solve the problem. The model established in the paper can help superstores predict demand more accurately, make replenishment plans, adjust pricing strategies, and improve market competitiveness.

References

[1] Wu Lifang. A multi-stage pricing model for fresh produce considering freshness and consumer utility[J]. Comprehensive Transportation, 2022, 44(10):124-129.

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

[3] Cheng Juanjuan. Empirical research 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.

[4] Zhang Liang, Yan Yonghong, Zhao Rong et al. Linear regression analysis and prediction of college grades based on SPSS software[J]. Computer and Digital Engineering, 2023, 51(05):1086-1090.

[5] D. C. Li, J. F. Dong, J. W. Su et al. Prediction of hot keywords and topics development in intelligence based on gray prediction GM(1, 1) model[J]. Technology and Market, 2023, 30(08):137-142.