ISSN 2616-5902 Vol. 6, Issue 5: 147-153, DOI: 10.25236/AJBM.2024.060520

Research on automatic pricing of vegetable commodities based on optimization model

Yashi Guo*, Xiaozhen Chen

School of Mathematics and Finance, Putian University, Putian, 351100, China
**Corresponding author: 18759507571@163.com*

Abstract: *This paper studies the relationship between vegetable sales volume and the effect of season on pricing. MATLAB was used to normalize the sales volume of category and single product, and descriptive statistical analysis was used to calculate the mean, variance and skewness, and the distribution of edible fungi was concentrated, while the distribution of cauliflower was close to normal, and the sales volume distribution of single product was also obtained. SPSS was used to analyze the relationship between category and single product sales, and Pearson correlation analysis was used to find that caullower and nightshade had a significant correlation. Among single products, beef head lettuce had the greatest correlation with Xixia mushroom. Multiple linear regression model was used to analyze the model equation of the total sales volume of different categories in different quarters, and 2022.6.30-2023.6.30 was divided into four quarters, and the sales volume of the next week was predicted by ARIMA model.*

Keywords: *Pearson Correlation, Multiple Linear Regression, Time Series Model, Linear Programming Model*

1. Introduction

By introducing association rules and mining the relevant data generated in the process of product production and marketing, the optimization of tea combination sales and economic benefit scheme was realized by calculating the effect coefficient of combination sales or joint cross-sales[1]. A dynamic irrigation optimal decision-making model with the aim of minimizing inter-plant evaporation was established[2]. Aiming at maximization of crop system income, using interval nonlinear programming model to optimize the allocation of water resources not only avoids the waste of water resources, but also increases the system income[3]. This paper considers the effect of vegetable sales volume and season on vegetable pricing based on the optimization model.

2. The distribution and relationship of the sales volume of each category and single item of vegetables

2.1 Data preprocessing for categories and units

In this paper, for the distribution law and mutual relationship of each category, the item code of each category is screened respectively, and the Excel table of the sales volume of each item of Mosaic, cauliflower, aquatic root, nightshade, chili and edible fungi is made. The box diagram is drawn, outliers are obtained and deleted. And calculate the total sales volume of each category. As for the distribution rule and mutual relationship between single products, the missing value is deleted according to the total sales volume of each single product in each category. Because the data is too large, the data with a small number of single product sales samples does not have reference value, so it is deleted and only 100 single products are collected for analysis.

2.2 Distribution law of categories

SPSS was used to calculate the mean value, variance and skewness of each category, and the following results were obtained, as shown in Table 1.

The mean value can reflect the central tendency of this set of data, the variance can reflect the dispersion degree of this set of data, and the skewness can reflect the distribution form of this set of data. It can be seen from Table 1 that the distribution law of each category is as follows: (The higher the

ISSN 2616-5902 Vol. 6, Issue 5: 147-153, DOI: 10.25236/AJBM.2024.060520

ranking of central trend, the more concentrated the ranking of dispersion degree, the lower the ranking of distribution form, the closer to normal)

Table 1: Descriptive statistics of categories

	N	Mean value		Variance	Kurtosis	
		N	Standard error		N	Standard error
Cauliflower	5	8353.2902	5068.32393	128439537.1	1.683	0.913
Mosaic	100	1985.2098	360.03232	12962327.29	2.551	0.241
Capsocum	45	2035.3029	713.88211	22933244.77	4.166	0.354
Aquatic rhizomes	19	2135.8607	1427.31631	38707405.15	4.022	0.524
Edible fungi	72	1056.7601	285.88628	5884629.307	4.491	0.283
Solanum	10	2243.1782	1316.09971	17321184.39	2.729	0.687

Table 2: Distribution law

	Central tendency	Degree of dispersion	Distribution pattern
Cauliflower	6	6	1
Mosaic	2	2	2
Capsocum	3	4	5
Aquatic rhizomes	4	5	4
Edible fungi	1	1	6
Solanum	5	3	3

As can be seen from Table 2, the distribution of cauliflower is partial normal, the distribution of flowers and leaves is concentrated and tends to normal, the distribution of capsicum is concentrated, the distribution of aquatic rhizomes is discrete, the distribution of edible fungi is concentrated, and the distribution of nightshade is discrete.

2.3 Interrelation of categories

In this paper, Pearson correlation coefficient is widely used to measure the degree of correlation between two variables by using Pearson correlation analysis. Pearson correlation coefficient:

$$r = \frac{cov(x_i, y_i)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

(x_iRefers to the category variable,$y_i$refers to the category variable.)

Using spss to calculate can be obtained as shown in Fig 1:

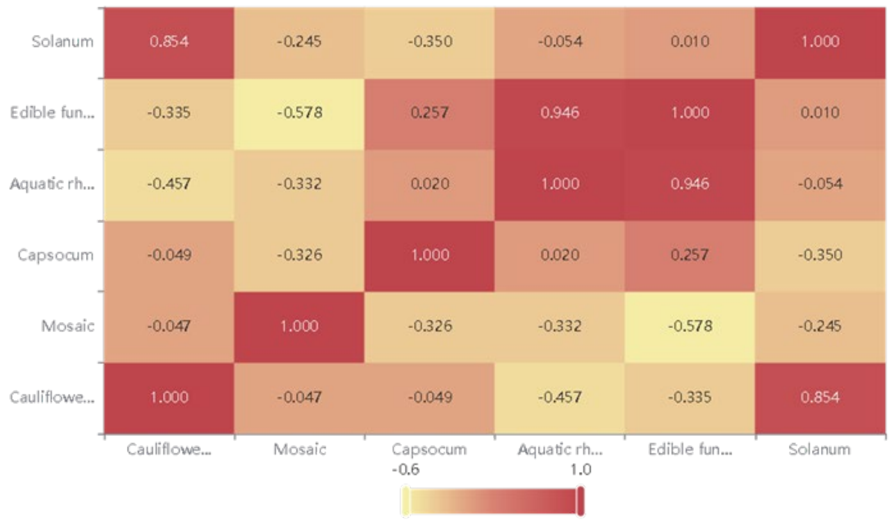

Figure 1: Category correlation heat map

From the heat map can be obtained (ranking order):

ISSN 2616-5902 Vol. 6, Issue 5: 147-153, DOI: 10.25236/AJBM.2024.060520

Table 3: Category ranking order table

Correlation	Cauliflower	Mosaic	Capsocum	Aquatic rhizomes	Edible fungi	Solanum
Cauliflower	/	1	3	5	4	1
Mosaic	3	/	4	3	5	4
Capsocum	2	3	/	2	2	5
Aquatic rhizomes	5	4	2	/	1	3
Edible fungi	4	5	1	1	/	2
Solanum	1	2	5	4	3	/

As can be seen from Table 3, cauliflower has a greater correlation with nightshade, aquatic rhizome has a greater correlation with edible fungi, capsicum has a greater correlation with edible fungi, and flowers and leaves have a greater correlation with cauliflower.

2.4 Distribution rule of single product

In this paper, the sales volume of each single product is summarized, statistical description is carried out by spss, and the following results are obtained:

By analyzing the mean value of aubergine, we can find that the sales volume of aubergine is concentrated; by analyzing the variance, we can find that the sales volume of aubergine is not discrete; by analyzing the skewness, we can find that the sales volume of aubergine is close to the normal distribution. The distribution of other items follows the same pattern.

2.5 Relationship of single products

This paper analyzes the relationship between the total sales volume of a single product. Because of the huge data, 100 single products are used for analysis. By analogy with the correlation analysis of various categories, Pearson correlation was used to analyze the correlation of 100 items, and SPSS was used to analyze the correlation of each item in the supporting materials.

2.6 Verification of the model

Spearman correlation was used to test the correlation of various categories, and Spsspro was used to calculate the correlation, as shown in the following Fig 2:

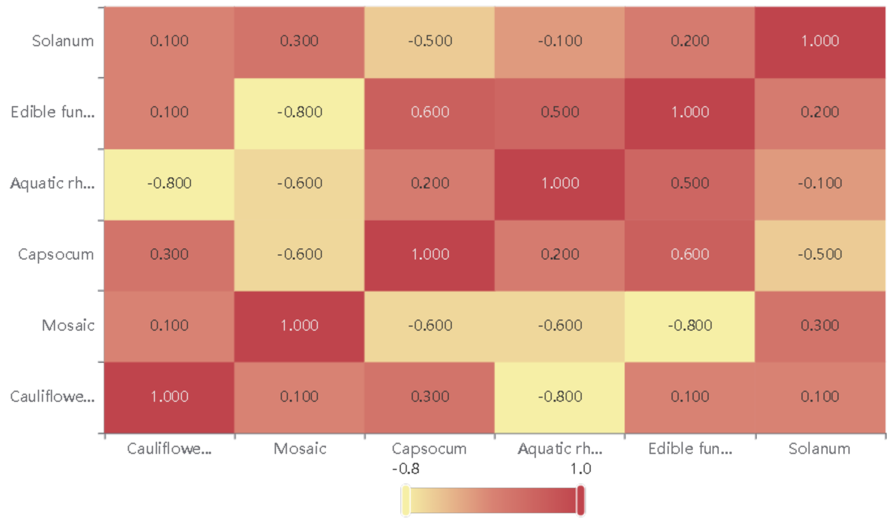

Figure 2: Spearman tests the heat map

According to the heat map, it can be found that cauliflower has a greater correlation with nightshade, aquatic rhizome has a greater correlation with edible fungi, capsicum has a greater correlation with edible fungi, and flower and leaf have a greater correlation with cauliflower.

ISSN 2616-5902 Vol. 6, Issue 5: 147-153, DOI: 10.25236/AJBM.2024.060520

3. Daily replenishment volume and pricing strategy

3.1 Quarterly effect [4]

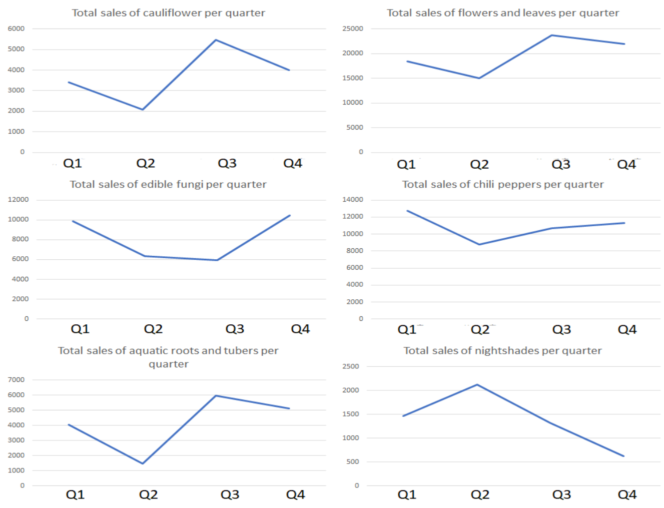

Figure 3: Quarterly-Total Sales Volume of Each Category

As can be seen from Fig 3, the total sales volume of each category is affected by quarter. Moreover, the sales volume of cauliflower, mosaics and aquatic roots were the lowest in the second quarter and the highest in the third quarter; the sales volume of edible fungi was the lowest in the third quarter, the peppers were the lowest in the second quarter, and the sales of nightshades were the highest in the second quarter.

3.2 Fitting

This article uses matlab for polynomial fitting, using the average pricing per quarter as the independent variable and total sales volume as the dependent variable. It uses a multiple linear regression model and uses matlab for fitting. Get the following results:

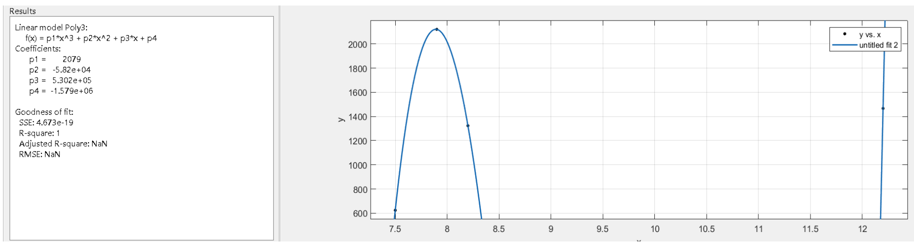

Figure 4: Solanum fitting

It can be seen from the figure 4 that the closer R^2 is to 1 and the SSE is closer to 0, the better the fitting effect is, and the fitting equation of nightshade is better. The same applies to other categories of fitting equations.

Table 4: Fitting equation expression

Category	fitting equation
Cauliflower	$Y = -1023828.96 + 331497.459 * x^1 - 35046.752 * x^2 + 1215.967 * x^3$
Mosaics	$Y = -2461648.356 + 1471512.668 * x^1 - 284600.941 * x^2 + 17986.677 * x^3$
Chili peppers	$Y = 44710.68 - 9889.257 * x^1 + 693.773 * x^2$
Aquatic rhizomes	$Y = -59764.184 + 14029.114 * x^1 - 825.103 * x^2 + 9.675 * x^3$
Edible fungi	$Y = 4387379.638 - 1240186.063 * x^1 + 116473.62 * x^2 - 3626.499 * x^3$
Nightshades	Y = −1579106.456 + 530205.571 ∗ x1-58200.059*x2 + 2078.677 ∗ x3

ISSN 2616-5902 Vol. 6, Issue 5: 147-153, DOI: 10.25236/AJBM.2024.060520

It can be seen from Table 4 that in the eggplant category, cost-plus pricing and sales volume showed a negative correlation in the first quarter. In the pepper category, the relationship between the two showed a positive correlation in the first, second and third quarters, and a negative correlation in the fourth quarter. In edible fungi, the relationship between the two showed a positive correlation in the first and second quarters, and a negative correlation between the third and fourth quarters. In cauliflower, the relationship between the two showed a negative correlation in the first, second and fourth quarters, and a positive correlation in the third quarter. In the mosaic category, the relationship between the two showed a negative correlation in the first, second and third quarters, and a positive correlation in the fourth quarter.

3.3 Establishment of Arima Model

This article predicts the changes in total sales volume and wholesale price of each category in the next week, that is, based on existing historical time data, predicting future trends, and the relationship between sales volume and quarter has been proven previously.

$$AR: y_t = c + \sum_{i=0}^{p} \Gamma_i * y_{t-i} + \varepsilon_t \tag{2}$$

$$MA: y_t = c + \sum_{i=1}^{q} \Theta_i * \varepsilon_{t-i} + \varepsilon_t \tag{3}$$

Where X_t is the current value, c is the constant term, epsilont is the error (white noise).

Sequence stationarity Stationarity means that the distribution of data does not change when it is shifted in time[5-7].Therefore, non-stationary data shows fluctuations due to trends and must be transformed for analysis.

Table 5: ADF inspection form for peppers

ADF inspection form							
variable	Differential order	t	P	AIC	critical value		
					1%	5%	10%
Total sales	0	-3.663	0.005***	1420.821	-3.477	-2.882	-2.578
	1	-6.237	0.000***	1421.109	-3.478	-2.882	-2.578
	2	-8.144	0.000***	1425.269	-3.479	-2.882	-2.578

It can be seen from the table 5 that the ADF significance p value of the original sequence is <0.05, so it is stable and significant at the level. The sequence is a stationary time series. As can be seen from the Fig 5 below, the sequence is basically a relatively even random oscillation near the mean.

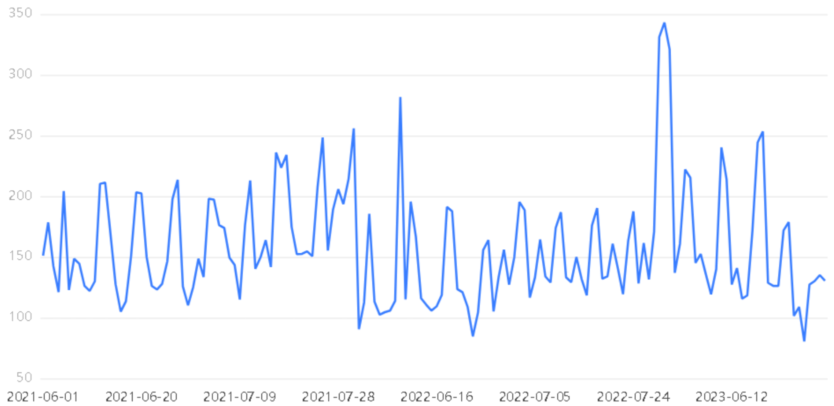

Figure 5: Best Differential Sequence Diagram

After the selection of the model in this article has completed the sequence stabilization, SPSSPro is used to analyze ACF and PACF[8-9], the p value and q value are determined, and SPSSPro is used for analysis. The results are as shown in the Fig. It can be seen from the Fig that p=0 and q=1. So the fitting model is the ARIMA model (0,0,1), using multiple linear regression, we get y(t) = 157.386 + 0.469 ∗ (t − 1),as shown in Fig6 and Table 6.

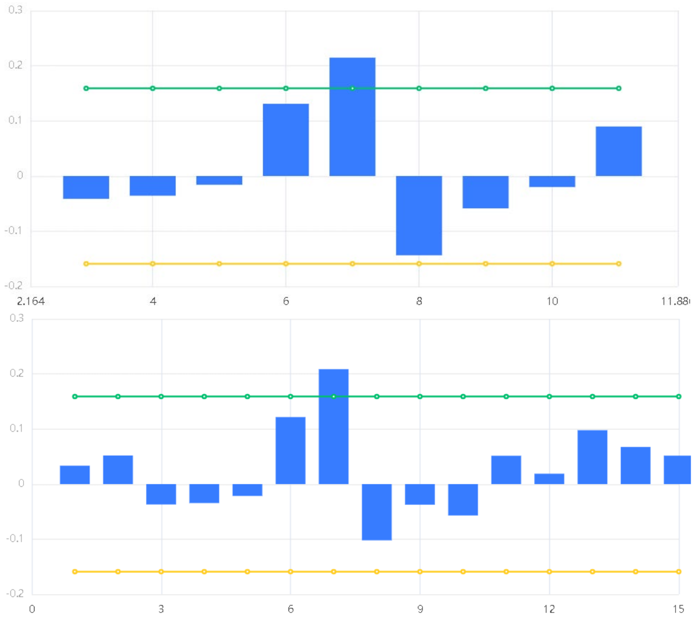

Figure 6: ACF and PACF

Table 6: Model parameter table

Model parameters						
	coefficient	standard deviation	t	P>t	0.025	0.975
constant	157.386	4.947	31.816	0	147.69	167.081
ma.L1 Total sales	0.469	0.068	6.891	0	0.336	0.602

3.4 Optimization model

The loss amount is understood as the loss amount during transportation and the loss amount of the goods that are not sold, so the replenishment amount = the purchase amount - the loss amount. To sum up, the following linear programming model is established:

$$\begin{cases} a_i < A_i, i = 1,2,3,4,5,6 \\ q_i < p_i \leq P_i, i = 1,2,3,4,5,6 \\ 0 < j_i - j_i * a_i \leq j_i, i = 1,2,3,4,5,6 \end{cases} \tag{4}$$

(j_i refers to the replenishment amount of each category,p_i refers to the cost plus pricing of each category,q_i refers to the wholesale of various categories[10],P_i refers to the maximum unit price of each category market,a_i refers to the rate of attrition,A_i refers to the maximum loss rate)

The maximum profit can be obtained by bringing in the predicted total sales volume, wholesale price (cost price), and damage rate of each category from July 1 to July 7.

4. Conclusions

This article mainly studies the relationship between product pricing and category sales volume, single product sales volume and season. Through descriptive statistical analysis, the distribution trend and abnormal values of the data can be visually displayed, and the patterns and anomalies in the data can be derived. Through Pearson correlation analysis, the linear relationship between categories and single products can be calculated and obtained. Under ideal conditions, we believe that the results obtained in this article are feasible. As we all know, weather is an important factor that cannot be ignored. Drastic changes in weather may cause a sudden decrease or increase in the sales of a certain variety. We have not discussed these in the article. Of course we hope that we can improve this shortcoming in the near future.

ISSN 2616-5902 Vol. 6, Issue 5: 147-153, DOI: 10.25236/AJBM.2024.060520

References

[1] GAO H F. Economic benefit optimization method of tea sales based on correlation analysis model [J]. Fujian Tea, 2019, 46(03):53-55.
[2] Shi Rongchao, Guo Wenzhong. Research progress of agricultural irrigation water resources optimal allocation [J/OL]. Journal of agricultural engineering: 1-13 [2024-03-17].
[3] Harigu Abduresiti. Research on optimal allocation of water resources in reservoir irrigation area [J]. Haihe Water Resources, 2024(02):12-16.
[4] Yu Weige, Wu Huarui, Peng Cheng. Research on short-term vegetable price forecast combination model based on Lasso regression and BP neural network [J]. Smart Agriculture (Chinese and English), 2020, 2(03): 108-117.
[5] Tu Xinqi, Teng Jiamin, Fu Lianlian. The impact of agricultural product circulation efficiency on agricultural product prices [J]. Business Economics Research, 2024(05):108-111.
[6] Li Weiwei. Analysis on the correlation and fluctuation characteristics of different types of vegetable prices - taking Beijing as an example [J]. Price Monthly, 2015(12):36-41.
[7] Wang Fang. Research on dynamic pricing strategies of fresh food in supermarkets [D]. Southeast University, 2006
[8] Mao Lisa. Research on pricing strategy and production and marketing model of vegetable wholesale market from the perspective of supply chain [D]. Central South University of Forestry and Technology, 2022.
[9] Zhou Yubing. Research on fresh product inventory control and dynamic pricing based on deep reinforcement learning [D]. Central South University, 2022.
[10] Han Junhua, Qian Shengdao. Review of cost-plus pricing method [J]. Accounting Monthly, 2012(22):74-75