

Approximate Pricing Strategy Prediction Model Based on Pearson Correlation Coefficient and Golden Section Method

Mingwei Han^{1*}

¹ North China University of Technology, Beijing, China

Email Address

1443931306@qq.com (Mingwei Han)

*Correspondence: 1443931306@qq.com

Received: 26 February 2024; Accepted: 15 March 2024; Published: 11 April 2024

Abstract:

Fresh supermarket for the general public to provide fresh quality vegetables, due to the short shelf life of vegetables, the supermarket usually according to the historical sales and demand of each product for replenishment and pricing. Based on the given data and conditions, the statistical model based on Pearson correlation coefficient and the approximate pricing strategy model based on the golden section method are established in order to maximize the profit of the supermarket, and the prediction of the replenishment volume and pricing strategy of the category and item is realized. Problem 1: First of all, the commodity information of 6 vegetable categories sold by a supermarket is combined with the data of the detailed data of vegetable sales, and the data of each category is extracted separately according to the category of vegetables for data processing. Then, the processed data is visualized to obtain the relevant statistical results such as line chart, pie chart, box chart and scatter matrix. Finally, Pearson correlation coefficient theory was used to establish the correlation statistical model, and heat map was used to show the results: the correlation coefficients of flowers and leaves and cauliflower, flowers and leaves and edible fungi, aquatic rhizoids and capsicum, aquatic rhizoids and edible fungi were all greater than 0.6, and the heat map was darker, showing a strong correlation; The correlation coefficients between nightshade and the other five categories were all less than 0.35, and the color of the heat map was light, showing a weak correlation. Problem 2, it is necessary to give the daily replenishment total amount and pricing strategy of each vegetable category in the next week by category. Based on the given data, the total sales volume and average sales volume of supermarket vegetable items for a total of 36 months from July 2020 to June 2023 were first calculated, and the scatter plot and fitting curve were obtained, and then the fitting equation of sales volume was obtained. Then, the approximate pricing strategy prediction model based on the Golden Section method was established according to the increasing profit. The results are shown in Table 2.

Keywords:

Pearson Correlation Coefficient, Data Visualization, Scatter Plot Fitting, Golden Section

1. Introduction

1.1. Problem background

In the daily life of fresh supermarkets, usually the freshness period of vegetable products is relatively short, and the moisture, nutrition, color and color will deteriorate with the increase of sales time. Most vegetables are suitable for the same day of purchase and sold out on the same day. In this special case, supermarket staff will generally according to the historical sales situation of each product. Daily replenishment based on sales experience and public demand.

As a daily supply of nutrition for the general public, supermarkets sell a wide variety of vegetables, with different tastes and producing areas all over the country. In order to ensure fresh and on-time sales, the purchase and trading time of vegetables is usually at 3:00-4:00 a.m., so businesses must make replenishment decisions for each vegetable category in advance without clearly knowing the specific item and purchase price. The method of vegetable pricing is “cost plus pricing”, the supermarket causes appearance damage during transportation, and the goods that are wilted and poor are usually sold at a discount. In the process of purchase, convincing market demand analysis is extremely important for merchants’ replenishment and pricing strategy.

1.2. Question raising

By building relevant mathematical models, we solve the following problems:

Question 1: There may be a variety of correlation between different categories or different items of vegetable products. Analyze the distribution law and the relationship between the sales volume of each category and item of vegetable.

Question 2: Fresh supermarket makes replenishment plan daily by category. Firstly, the relationship between the total sales volume of each vegetable category and the cost plus pricing is analyzed, and then the daily replenishment volume and pricing strategy of each vegetable category are given in the next week (July 1-7, 2023), so as to maximize the profit of the supermarket.

2. Problem Analysis

For the first question, it is required to analyze the distribution law and mutual relationship of the sales volume of each category and single product of vegetables. It can be subdivided into two questions: the distribution law and mutual relationship of various categories of vegetables, and the distribution law and correlation relationship of specific items within the vegetable category.

For the second question, when analyzing the total volume and cost of vegetable sales, first of all, we need to obtain the vegetable sales data from July 2020 to June 2023 from the sales records of the supermarket, sort and number these data by month, and calculate the total sales volume of vegetable products in the supermarket each month. The sales quantity of each vegetable variety in each month can be obtained from the table, and the total sales volume of the month can be added together. Divide the total number of sales by 30 to get the average number of sales per day. In order to more intuitively show the changing trend of sales data, a scatter plot is established to show the relationship between sales volume and month number. In order to describe the sales trend more accurately, it is necessary to build the fitting curve and find the fitting equation that best fits the data. By fitting the equation, correlation analysis can

be performed on the data to further predict the average sales volume of the next month's numbering.

3. Model Assumptions

(1) It is assumed that each vegetable category is abundant, and the merchant's purchase quantity of each category in a certain period will not appear the problem of category shortage.

(2) It is assumed that the total amount of vegetables purchased by consumers every day fluctuates within a certain range, and there will be no unexpected circumstances, such as group buying and natural disaster factors.

(3) Assuming that the daily wholesale price of vegetables fluctuates within a certain range, the purchase quantity of merchants can basically meet the total daily demand of consumers.

4. Symbol Description

Table 1 shows the variables needed in the following experiment process and their symbolic representation.

Table 1. Variable name and symbol.

Symbol	Instructions	Unit
V	Total sales volume	kg
D	Display quantity	kg
R	Replenishment volume	kg
S	Selling price	yuan
C	Cost price	yuan
L	Defective amount	\
P	profit	yuan
X	A collection of items within a class	\
F	Golden section	\

5. The Establishment and Solution of the Model

5.1. Model building and solving of problem 1

5.1.1. Model building of problem 1

In order to facilitate data calculation and processing, the vlookup function is used to merge the data of Attachment 1 and Attachment 2, and then the data of each category is extracted separately into a new worksheet, and six categories are obtained: Cauliflower, Mosaic, nightshade, chili, edible fungi and aquatic rhizomes, using excel function formula to calculate the total monthly sales and average unit price of each category.

This paper introduces Pearson correlation coefficient [1,2], also known as Pearson product moment correlation coefficient (PPMCC or PCCs), which is used to measure the degree of correlation between two variables or vectors X and Y, and its value is between -1 and 1. The Pearson correlation coefficient is calculated as

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X - \mu_x)(Y - \mu_y))}{\sigma_X \sigma_Y}$$

$$\rho_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)} \cdot \sqrt{E(Y^2) - E^2(Y)}} \quad \# \quad (1)$$

Where $E(X)$ and $E(Y)$ represent the mean of the X and Y vectors, respectively. According to Pearson correlation coefficient theory, 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. The following will be the corresponding data visualization according to the theory.

5.1.2. Data visualization

By calculating the proportion of total sales of each category, a pie chart is drawn, as shown in Figure 1. Flower-leaf accounted for the majority of sales, while nightshade, cauliflower and aquatic rhizomes accounted for less.

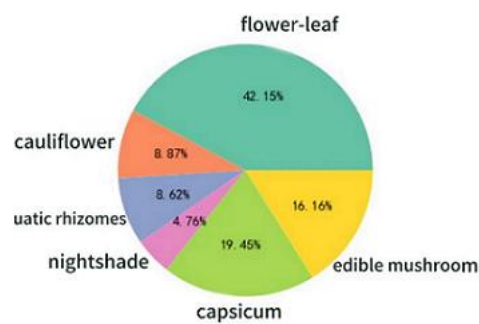


Figure 1. Total sales by category.

The distribution matrix of the sales volume of six vegetable categories is shown in Figure 2. The sub-graphs show the correlation between different feature columns respectively, such as the close linear relationship between capsicum and edible fungi, edible fungi and aquatic rhizoids, and flowers and leaves and peppers, indicating a strong correlation between these features.

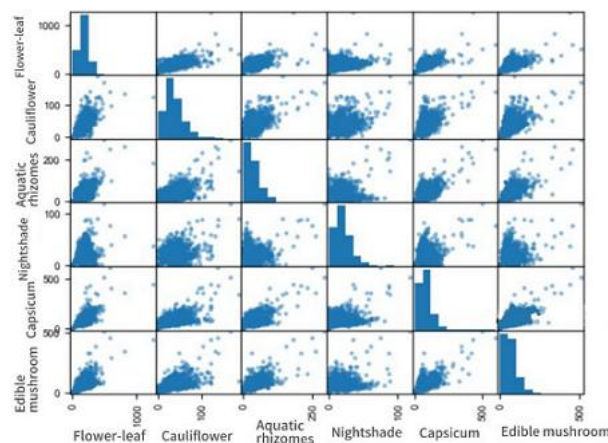


Figure 2. Scatter matrix graph.

The scatter matrix “qualitatively” gives the relationship between the features in the form of a graph, and the correlation coefficient is needed for further “quantitative” analysis. Therefore, the Pearson correlation coefficient between the six categories is programmed to calculate, and the heat map (Figure 3) is generated to visually and obviously indicate the correlation of various dishes with data and colors. The correlation coefficients of flowers and leaves and cauliflower, flowers and leaves and

edible fungi, aquatic rhizoids and capsicum, aquatic rhizoids and edible fungi were all greater than 0.6, showing a strong correlation. The correlation coefficients between nightshade and the other five categories were all less than 0.35, and the color of the heat map was light, showing a weak correlation.

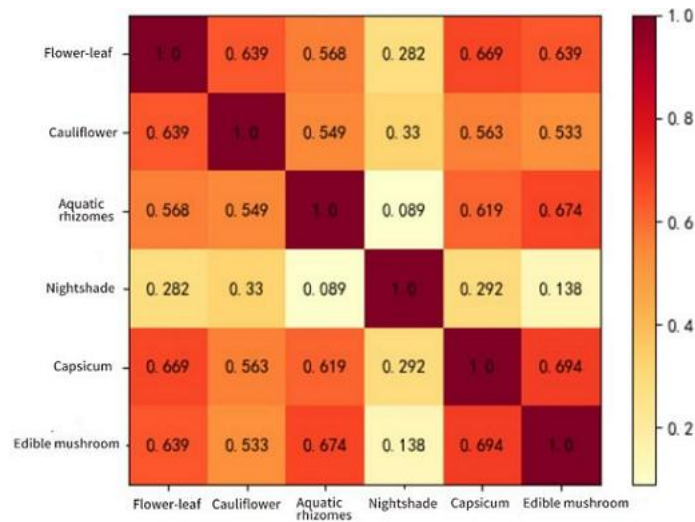


Figure 3. Pearson correlation coefficient heat map.

5.2. Model establishment and solution of problem 2

5.2.1. Model establishment of problem 2

Create a new workbook to store Date of Sale and Sales Volume and export the relevant data from the original data. Next, a preliminary table and chart can be created by analyzing the data to calculate the change in sales from July 2020 to June 2023 and calculating the total sales for each month, including the average sales volume for each day of the month, serial number values, scatter plots and fitting curves.

In the new worksheet, create a table that includes the date of sale, serial number, average daily sales, total sales, and days of the month; In the Average Daily Sales column, divide the total number of sales per month by the corresponding number of days using the formula; In the “Serial Number Value” column, enter serial number values from 1 to 36, corresponding to July 2020 to June 2023; You can then create a scatter plot based on the Serial Number Values column and the Average Sales per Day column; Then the fit curve and fit function image are generated by scatter plot. As shown in Figure 4 below:

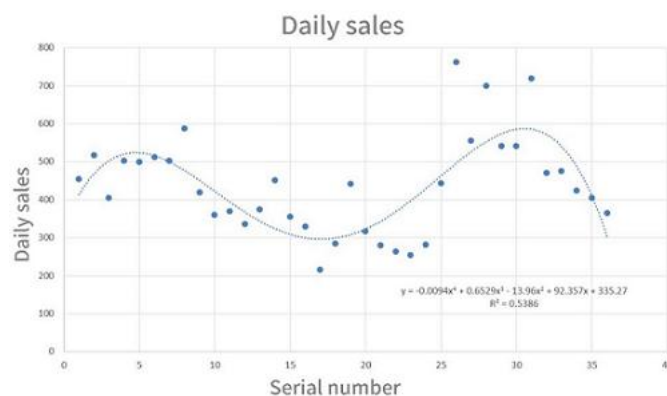


Figure 4. Fitting curve.

The function expression in the figure is:

$$y = -0.0049x^4 + 0.6529x^3 - 13.96x^2 + 92.357x + 335.27 \quad (2)$$

When $x=37$, the value of y represents that the average daily sales volume in July 2023 is 95.4693, and the total sales volume from July 1 to 7 is 668.2851. From this, we can get the total replenishment of July 1-7, 2023.

Suppose V is the sales volume of the supermarket, C is the cost price, S is the pricing, and P is the profit

$$P = V(S - C) \quad (3)$$

Thus, an approximate model of pricing strategy [3,4] can be obtained:

$$S = C + \frac{\max P}{V} \quad (4)$$

Where V can be approximated by y in formula (2) above, only an approximation of $\max P$ is required.

After analyzing the data in the table, it is necessary to determine a reasonable pricing strategy to maximize the profit of the supermarket. Based on the analysis, we calculate the profit of different types and different items. By arranging the profit of different items of each category in descending order, using the Golden Section method (0.618 method) [5] F, we take about 38% of the items with the largest profit of each category as the main factor affecting the maximum profit, and calculate the average profit of these items as the approximate value of the maximum profit S , that is

$$\sum \max P = \dots F(\text{Rankp}(X)) \quad (5)$$

Where, represents the collection of items of each class, represents the ordering of items in X by profit from largest to smallest, $F(*)$ represents the implementation of the golden section method for *. Thus, the average profit and average unit price of each category are obtained, which is the pricing strategy, as shown in Table 2.

Table 2. Average profit and average unit price of each category.

Column 1	Cauliflower	Flower-leaf	Capsicum	Nightshade	Edible mushroom	Aquatic rhizomes
Average profit	4.76	1.69	2.18	2.84	3.06	3.54
Average unit price	11.06	6.54	10.46	8.52	15.4	9.05

6. Conclusions

We established a statistical model by using Pearson correlation coefficient theory, and the results were shown by heat map: there was a strong correlation between flowers and cauliflower, flowers and leaves and edible fungi, aquatic rhizoids and capsicum, aquatic rhizoids and edible fungi; The correlation coefficients between nightshade and the other five categories were all less than 0.35, showing a weak correlation. In addition, we set up the prediction model of approximate pricing strategy based on golden section method, and gave the pricing strategy of various types of vegetables.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

- [1] Zhang, Y.L.; Dang, Y.; He, P.A. Quantitative analysis of biological relationship by Pearson correlation coefficient. *Computer Engineering and Applications*, 2005, 41(33), 5.
- [2] Li, S.H.; Wang, X.D.; Song, Y.F. Conflicting evidence combination method based on Pearson Coefficient and uncertainty measure. *Journal of Electronic Measurement and Instrumentation*, 2021, 8, 38-41.
- [3] Han Junhua. Review of cost-plus pricing. *Finance and Accounting Monthly*, 2012, 08, 74-75.
- [4] Pan, X.F.; Xie, Z.H.; Wang, S.Y. Optimal decision of fresh products' super preservation effort and pricing considering loss avoidance. *Journal of Highway and Transportation Science and Technology*, 2022, 06(6), 177-186.
- [5] Yang, X.D.; Li, Q.; Shen, C.Y. Data fitting of polymer melt viscosity model based on Golden Section method. *Chinese Journal of Mechanical Engineering*, 2014, 50(2), 7.



© 2024 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)