A study on commodity pricing and sales optimization based on linear programming model: a case of vegetable category

Yutong Chen*
Harbin Normal University, Harbin, China, 150500
* Corresponding Author Email: 18645626279@163.com

Abstract. In the fresh food retail industry, vegetable commodities require daily replenishment decisions based on historical sales and demand due to their short freshness period and susceptibility to quality. In this paper, the sales volume, wholesale price and unit price of each category are first extracted, and the cost-plus ratio and related data are calculated. The relationship between the sales volume of each category and the cost markup rate is analyzed through linear regression, and it is concluded that the sales volume of most categories decreases with the increase of the markup rate, and through the distribution of the scatter plot of the total sales volume of the vegetable category and the markup rate, it can be seen that there is a negative correlation between the total sales volume of the vegetable category and the cost markup pricing, and then through the establishment of the optimization model and the prediction of the sales volume of each category for the coming week, the wholesale price and the related data, and the profits are used to Pricing to ensure revenue at the maximum markup rate to develop the total daily capture and pricing strategy for the coming week.

Keywords: Linear Programming Models, Merchandise Pricing and Sales Optimization, Vegetable Category.

1. Introduction

In the fresh food retail industry, vegetable commodities require daily replenishment decisions based on historical sales and demand due to their short freshness period and susceptibility to quality. Trading hours are usually in the early morning hours, and merchants need to make replenishment decisions without knowing the specific commodity and purchase price. Vegetable pricing is generally based on a cost-plus pricing method, and items of poor quality and condition are usually sold at a discount [1-2]. Based on a linear programming model, this study focuses on solving linear programming in optimization class problems, especially those involving key factors such as sales volume, wholesale price, unit price, and cost-plus rate [3]. By carefully analyzing the sales data of each category and the linear regression method to reveal the relationship between sales volume and cost markup rate, we found a negative correlation trend between total sales volume and cost markup rate in the vegetable category. Subsequently, by building an optimization model to predict key data such as sales volume and wholesale price for the coming week and formulating a corresponding pricing strategy, we aimed to ensure reasonable pricing while maximizing profits, thus providing a scientific basis for enterprises to formulate their daily capture totals and pricing strategies for the coming week [4-5]. This study provides strong support for companies to develop effective business strategies in a competitive market.

2. Modeling and solving for replenishment and pricing

2.1. Data preprocessing

Historical sales data is first collected, including sales volume, cost-plus pricing, and time stamps. Data cleaning is then performed to deal with missing values and outliers to ensure data integrity and reliability. And then feature engineering is performed to convert timestamps to numerical format, while sales volume and cost-plus pricing are normalized to eliminate the effect of different scales. The data was partitioned into training and test sets in preparation for building and evaluating linear
regression models. Target variables were also processed to meet model assumptions. Data visualization and analysis provided insight into the relationship between sales volume and cost-plus pricing. A reliable basis for model building and solving was provided [6-7].

2.2. Linear regression modeling and prediction

This question superstores do replenishment planning on a category-by-category basis, converting the analysis of the relationship between total sales and cost-plus pricing for each vegetable category into an analysis of the relationship between total sales and cost-plus rate for each vegetable category, introducing the concept of the markup rate, which is the ratio of cost to sales price because the cost-plus rate is more reflective of the profitability of the item than just the cost-plus pricing itself. By analyzing the relationship between total sales and cost-plus ratio, you can better understand the impact of sales strategy on profitability and develop a more targeted pricing strategy. The cost-plus ratio is calculated as:

\[
\text{Cost-plus ratio} = \frac{\text{Selling price} - \text{cost}}{\text{cost}} \times 100\%
\]  

(1)

This to develop a linear regression model of total sales and cost-plus rate for each vegetable category in the form of:

\[
Y = \beta_0 + \beta_1 X + \epsilon
\]  

(2)

Where \( Y \) denotes total sales, \( X \) denotes cost-plus pricing, \( \beta_0 \) and \( \beta_1 \) are parameters of the model, and \( \epsilon \) denotes the error term.

Observations and error degrees of freedom are used to train the data points and degrees of freedom of the model, and root mean square error is used as a measure of model fit. It represents the average deviation between the model's predicted and actual sales totals, and R-squared measures the degree of model fit, which represents the proportion of the variability in the dependent variable (sales totals) that is explained by the model. r-squared equals to 0 means that the model is unable to explain the variability in the dependent variable, i.e., the model is a very poor fit to the data. r-squared equals to 1 means that the model explains the variability in the dependent variable perfectly, i.e., the model fits the data very well. f-statistic is used to test the overall fit of the model. The F statistic is used to test the overall fit of the model. Here, the p-value of the F-statistic is very close to 0, indicating that the overall fit of the model is very significant [8-10].

The linear regression equation for the rate of addition and sales of foliage category is:

\[
Y = 0.5743 + 0.0062 * X
\]  

(3)

The scatterplot of the linear regression analysis of the additive rate of foliage category and sales is shown in Figure 1:

![Figure 1](image-url)
The linear regression equation of cauliflower additive rate and sales volume is:

\[ Y = 0.4847 - 0.0032X \]  (4)

The scatterplot of the linear regression analysis of cauliflower category additive rate versus sales is shown in Figure 2:

![Figure 2](image-url)

**Figure 2** Linear regression analysis of cauliflower additive rate and sales volume.

The linear regression equation for the rate of aquatic rhizome addition and sales volume was:

\[ Y = 0.6946 - 0.0110X \]  (5)

The scatter plot of the linear regression analysis of aquatic rhizome addition rate and sales volume is shown in Figure 3:

![Figure 3](image-url)

**Figure 3** Linear regression analysis of aquatic rhizome addition rate and sales volume.

The linear regression equation of additive rate and sales volume for chili category is:

\[ Y = 0.4646 - 0.0041X \]  (6)

The scatterplot of the linear regression analysis of the additive rate and sales of chili peppers is shown in Figure 4:
The linear regression equation of eggplant markup rate and sales volume is:

\[ Y = 0.4925 + 0.0012 \times X \quad (7) \]

The scatterplot of the linear regression analysis of eggplant markup rate versus sales is shown in Figure 5:

![Figure 5](image)

**Figure 5.** Linear regression analysis of additive rate and sales volume of eggplant.

The linear regression equation between the rate of addition of edible mushrooms and sales volume is:

\[ Y = 0.5026 + 0.0020 \times X \quad (8) \]

The scatter plot of the linear regression analysis of the rate of addition of edible mushrooms and sales volume is shown in Figure 6:
2.3. Building optimization models

Decision variables: the decisions to be made are the total daily replenishment of each vegetable category in the coming week (let the daily replenishment of each vegetable, in the order of foliage, cauliflower, aquatic rootstocks, eggplant, chili peppers, and edibles, be \( A_1, A_2, A_3, A_4, A_5, A_6 \)) and the pricing strategy (let the pricing strategy of each vegetable category, in the order of foliage, cauliflower, aquatic rootstocks, eggplant, chili peppers, and edibles, be \( B_1, B_2, B_3, B_4, B_5, B_6 \)).

Objective function: total revenue = total sales revenue - total costs:

\[
\max z = \sum_{i=1}^{n} (R_i - C_i)
\]  

Let the total revenue be \( z \), \( R_i \) be the total sales revenue of the total sales revenue of the category \( i \) and \( C_i \) be the total cost.

Solve for total sales revenue:

\[
R_i = \sum_{j=1}^{7} (P_i \cdot S_j)
\]

Let \( R_i \) be the total sales revenue of the category \( i \), \( x_i \) denote the daily replenishment of the category \( i \), and \( P_i \) denote the pricing strategy (price per unit) of the category \( i \) \((i=1,2,3,4,5,6)\) in the order of Foliage, Cauliflower, Aquatic Roots and Tubers, Eggplant, Chili Peppers, and Edible Mushrooms.

Solve for the total cost

\[
C_i = c_i \cdot x_i
\]

Let \( c_i \) be the cost of the category.

Constraints:
1. The minimum range of the pricing strategy should be higher than the cost of the vegetable category, i.e. \( P_i > C_i \).
2. The upper and lower limits of total replenishment are constrained.
3. There should be a linear relationship between cost and replenishment volume.

\[
C=ax + b
\]

Let \( C \) be the total cost, \( x \) be the total replenishment, and \( a, b \) be an unknown parameter.
2.4 Give the total daily replenishment and pricing strategy for the coming week

The total daily capture and pricing strategy for cauliflower category is analyzed after the above modeling process as shown in Table 1:

<table>
<thead>
<tr>
<th>Date of sale</th>
<th>Cauliflower stock (kg)</th>
<th>Additive rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2023/7/1</td>
<td>55.851</td>
<td>0.387</td>
</tr>
<tr>
<td>2023/7/2</td>
<td>47.944</td>
<td>0.921</td>
</tr>
<tr>
<td>2023/7/3</td>
<td>27.955</td>
<td>0.084</td>
</tr>
<tr>
<td>2023/7/4</td>
<td>42.64</td>
<td>1.94</td>
</tr>
<tr>
<td>2023/7/5</td>
<td>31.186</td>
<td>1.949</td>
</tr>
<tr>
<td>2023/7/6</td>
<td>50.962</td>
<td>0.091</td>
</tr>
<tr>
<td>2023/7/7</td>
<td>48.356</td>
<td>0.171</td>
</tr>
</tbody>
</table>

The total daily catch and pricing strategy for the foliar category is shown in Table 2:

<table>
<thead>
<tr>
<th>Date of sale</th>
<th>Foliar stock (kg)</th>
<th>Additive rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2023/7/1</td>
<td>195.295</td>
<td>0.522</td>
</tr>
<tr>
<td>2023/7/2</td>
<td>188.672</td>
<td>0.35</td>
</tr>
<tr>
<td>2023/7/3</td>
<td>116.605</td>
<td>0.088</td>
</tr>
<tr>
<td>2023/7/4</td>
<td>133.21</td>
<td>0.126</td>
</tr>
<tr>
<td>2023/7/5</td>
<td>164.461</td>
<td>0.629</td>
</tr>
<tr>
<td>2023/7/6</td>
<td>134.044</td>
<td>1.085</td>
</tr>
<tr>
<td>2023/7/7</td>
<td>129.05</td>
<td>1.234</td>
</tr>
</tbody>
</table>

The total daily catch and pricing strategy for the chili category is shown in Table 3:

<table>
<thead>
<tr>
<th>Date of sale</th>
<th>Peppers stocked (kg)</th>
<th>Additive rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2023/7/1</td>
<td>55.999</td>
<td>7.222</td>
</tr>
<tr>
<td>2023/7/2</td>
<td>58.67</td>
<td>6.137</td>
</tr>
<tr>
<td>2023/7/3</td>
<td>47.458</td>
<td>6.571</td>
</tr>
<tr>
<td>2023/7/4</td>
<td>49.728</td>
<td>4.902</td>
</tr>
<tr>
<td>2023/7/5</td>
<td>51.119</td>
<td>6.198</td>
</tr>
<tr>
<td>2023/7/6</td>
<td>39.874</td>
<td>6.655</td>
</tr>
<tr>
<td>2023/7/7</td>
<td>42.7</td>
<td>5.715</td>
</tr>
</tbody>
</table>

3. Conclusions

This paper establishes a framework for merchandise pricing and sales optimization based on linear programming models. Through meticulous data preprocessing and linear regression analysis, the relationship between the total sales volume of vegetable categories and the cost-plus rate was deeply understood. Taking leafy flowers, cauliflower, aquatic roots and tubers, chili peppers, eggplants, and edible mushrooms as examples, the sales characteristics of each category were revealed through the establishment of a linear regression model, which provided data support for subsequent decision-making. By introducing the concept of cost plus rate, an optimization model was constructed to effectively formulate the total daily replenishment and pricing strategy for the coming week to maximize the total revenue. Ultimately, a scientific and feasible replenishment plan was provided for the superstore, providing strategic decision support to meet the market challenges.
References


