Study on Vegetable Pricing Based on Time Series and Exponential Smoothing Model

Yunhang Lv#, Xiulan Han#, Mengyu Xu#, Yunxiang Tan*,#

School of Ecology and Environment, Hainan University, Hainan, Haikou, 507228, China

*Corresponding author: bioecotan@163.com
#These authors contributed equally.

Abstract: Vegetable products have the characteristics of short shelf life. If they can't be sold out in time, it will cause waste, which will lead to the loss of merchants. Therefore, it is of great significance to formulate reasonable purchase quantity and pricing strategy. In view of this, this paper studies the purchase quantity and pricing strategy of vegetable commodities by establishing multiple linear regression model, ARIMA time series model and exponential smoothing time model. The results show that: (1) the relationship between consumers' purchase of various vegetables can be understood through correlation analysis; (2) By establishing time series, exponential smoothing model and combining multiple linear regression, the daily sales volume and daily replenishment volume can be better predicted, thus guiding pricing decision. (3) In the process of vegetable pricing, combining the historical pricing method can make the pricing decision more reliable, and adopting different pricing strategies for different vegetables according to consumers' purchasing psychology is conducive to increasing vegetable sales.

Keywords: Time series model, Multiple linear regression, Exponentially smoothed time series, ABC classification

1. Introduction

The shelf life of vegetables is usually very short, and with the increase of sales time, the quality of vegetables will become worse. If most varieties of vegetables are not sold on the same day, they can't be sold on the next day, which will lead to losses for businesses. Analyzing the sales volume and pricing strategy of vegetables through mathematical model can effectively guide the purchase quantity and pricing decision, reduce the loss of merchants and maximize their benefits.

There have been many studies on vegetable pricing. Lu [1] has established a pricing model of supermarket fresh vegetables based on value loss, and proposed to dynamically adjust the price of vegetables sold on the same day according to the value loss of vegetables, so as to promote the increase of vegetable sales. Zhao [2] used the semi-parametric Copula method to fit the joint distribution of vegetable yield fluctuation and price fluctuation, and studied the insurance pricing of vegetable income. Ning [3] put forward four principles and pricing strategies of vegetable pricing. Zhang [4] extended the pricing strategy in the opaque field to the agricultural products field, and discussed the optimal pricing level of merchants and the revenue improvement effect of the process. Zeng [5] found through the research that the dynamic pricing strategy adopted by a fresh community supermarket can reduce the losses and bring higher profits for the supermarket. By issuing questionnaires, Zhou [6] analyzed the characteristics of customers' perceived value of various types of products, and put forward that different pricing strategies should be formulated for rice, fruits, vegetables and eggs. Gu [7] constructed a dynamic pricing model of fresh products considering the change of freshness, and obtained the optimal pricing of fresh products in the fresh period and the decline period of freshness with the goal of maximizing profits.

The above-mentioned vegetable pricing strategies seldom involve time series model and exponential smoothing time model. In view of this, this paper establishes vegetable pricing strategies with multiple linear regression model, ARIMA time series model and exponential smoothing time model, which provides guiding significance for vegetable commodity sales.
2. Model building

Model assumptions:

(1) Assume that the goods are returned on the same day, and the second sale will not be affected after the return.

(2) It is assumed that the recent loss rate of vegetables will be consistent in the next week.

(3) It is assumed that the change of the wholesale price of each category of vegetables on the second day can be ignored.

(4) It is assumed that other competitive enterprises will not interfere with vegetable sales.

2.1 Pearson correlation analysis

Pearson correlation analysis is carried out between each category and the sales volume of each single product, and the calculation formula is as follows:

\[
\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{n}(x_i - E(X))(y_i - E(Y))}{\sqrt{\sum_{i=1}^{n}(x_i - E(X))^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - E(Y))^2}}
\]

Where, \(E(X)\) and \(E(Y)\) represent the expectation, that is, the average value; \(\sigma\) represents the standard deviation.

When the detected samples basically meet the Pearson correlation detection conditions, it is necessary to test the significance of the Pearson new correlation coefficient, and a statistic \(T\) can be constructed as follows:

\[
t = r \cdot \sqrt{\frac{n-2}{1-r^2}}
\]

Where is the number of samples and Pearson correlation coefficient. This statistic is proved to be consistent with the T-distribution with degree of freedom of \(t-2\), which can be used to test the correlation.

2.2 ARIMA time series model

Taking Flowers and leaves vegetables as an example, a time series model was established by SPSS with pricing and wholesale price as independent variables and total sales of cauliflower as dependent variables.

Time series is a group of random variables sorted by time, and time series data essentially reflect the trend changing with time, which may represent the evolution of one or more random variables with time. The core of time series prediction method is to mine the rules from the data and use the rules to predict the future data.

(1) sequence stabilization test, to determine the d value

Plot the sales series in the sorted data, conduct ADF test and observe whether the series is stable; Non-stationary time series should be converted into stationary time series by D-order difference.

(2) determine the p value and q value.

Autocorrelation coefficient (ACF) is a statistical index used to measure the correlation between different lags (delays) in time series data, indicating the degree of correlation of the same event in two different periods, and the Q value can be roughly judged by the maximum lag point of the ACF chart. The calculation formula of autocorrelation coefficient is as follows:

\[
ACF(k) = \frac{\sum_{t=k+1}^{n}(Z_t - \bar{Z})(Z_{t-k} - \bar{Z})}{\sum_{t=1}^{n}(Z_t - \bar{Z})^2}
\]

Where, \(\bar{Z}\) is the mean of the sequence, \(t\) is the time, \(k\) is the interval, and both \(Z_t\) and \(Z_{t-k}\) represent the time series.
Partial autocorrelation coefficient (PACF) is a statistical index used to measure the correlation between a certain lag period and the current observed value in time series data, but when calculating the partial autocorrelation coefficient, the influence of other lag periods will be excluded. The value of $p$ can be roughly judged by the maximum lag point of the partial autocorrelation coefficient (PACF) graph. The calculation formula of partial autocorrelation coefficient is as follows:

$$\text{PACF}(k) = \frac{E(Z_k - EZ) (Z_{t-k} - EZ_{t-k})}{\sqrt{E(Z_k - EZ)^2} \cdot \sqrt{E(Z_{t-k} - EZ_{t-k})^2}}$$

(4)

2.3 Multiple Linear Regression Model

The main formula of multiple linear regression is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon$$

(5)

Where, $Y$ is the dependent variable, $X_1, X_2, \ldots, X_n$ is the independent variable, $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ is the regression coefficient, and $\varepsilon$ is the error term.

After obtaining the linear regression formula, you can calculate the $F$ statistic to test whether the linear relationship between independent variables and dependent variables is significant. The formula for calculating the $F$ statistic is as follows:

$$F = \frac{\text{SSR}/k}{\text{SSE}/(n-k-1)} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 / k}{\sum_{i=1}^{n} (y_i - \hat{y})^2 / (n - k - 1)} \sim F(k, n - k - 1)$$

(6)

After determining the significance level $\alpha$ and degree of freedom $(k, n - k - 1)$, then the critical value $F_\alpha$ was found.

If $F > F_\alpha$, or the $p$-value is less than $\alpha$, reject $H_0$, the linear relationship is not significant.

2.4 Exponential Smoothing Time Model

The follow are that steps to construct an exponential smooth sequence of text:

Let the time series be $y_1, y_2, \ldots, y_t$, $\alpha$ be the weighting coefficient, $0 < \alpha < 1$, and the formula of the first smoothing index is:

$$S_t^{(1)} = \alpha y_t + (1 - \alpha) S_{t-1}^{(1)} = S_{t-1}^{(1)} + \alpha (y_t - S_{t-1}^{(1)})$$

(7)

Among them, $y_t$ is the time series of observation and $S_t^{(1)}$ the first smoothing index of $t$ period.

The recursive formula of moving average is:

$$M_t^{(1)} = M_{t-1}^{(1)} + \frac{y_t - y_{t-N}}{N}$$

(8)

Among them, $M_t^{(1)}$ is a smooth exponential moving term in $t$ period.

$$M_t^{(1)} = M_{t-1}^{(1)} + \frac{y_t - y_{t-N}}{N} = \frac{y_t}{N} + \left(1 - \frac{1}{N}\right) M_{t-1}^{(1)}$$

(9)

With $M_t^{(1)}$ as the best estimate of $y_{t-N}$, there is:

Order $\alpha = \frac{1}{N}$ take $S_t$ instead $M_t^{(1)}$ to obtain:

$$S_t^{(1)} = \alpha y_t + (1 - \alpha) S_{t-1}^{(1)}$$

(10)

In order to further understand the significance of data smoothing, Equation (8) is expanded in turn, and we can get:

Among them, $S_t^{(1)}$ is the weighted average of all historical data, and the weighted coefficients are
\[ \alpha, \alpha(1 - \alpha), \ldots, \alpha(1 - \alpha)^j \] and you can get:

\[ \delta_t^{(i)} = \alpha y_t + (1 - \alpha) \left[ \alpha y_{t-1} + (1 - \alpha) \delta_{t-1}^{(i)} \right] = \alpha \sum_{i=0}^{\infty} (1 - \alpha)^j y_{t-j} \tag{11} \]

\[ \sum_{i=0}^{\infty} (1 - \alpha)^j = \frac{\alpha}{1 - (1 - \alpha)} = 1 \tag{12} \]

Thus, a smooth exponential model can be derived:

\[ \hat{y}_{t+1} = \alpha y_t + (1 - \alpha) \hat{y}_t \tag{13} \]

3. Result analysis

3.1 Analysis of the relationship between sales volume of each category and sales volume of each single product

3.1.1 Analysis of the relationship between sales volume of each category

The correlations among the six categories are shown in Table 1. In order to make the results clearer, the correlation coefficients are plotted into a heat map as shown in Figure 1. From Table 1 and Figure 1, we can see the reciprocal substitution and antagonism between these six kinds of vegetables. There is a high correlation between cauliflower vegetables and Flowers and leaves vegetables, and the correlation between cauliflower vegetables and other vegetables is low. Pepper vegetables are negatively correlated with Solanula vegetables, and there is a high and significant correlation with aquatic rhizomes, edible fungi vegetables, and Flowers and leaves vegetables. Aquatic root vegetables and edible fungi vegetables are negatively correlated with Solanula vegetables, and there is a high positive correlation with edible fungi vegetables. Besides the negative correlation with solanaceous vegetables, Flowers and leaves vegetables have a high correlation with other vegetables.

<table>
<thead>
<tr>
<th></th>
<th>Cauliflower</th>
<th>Peppers</th>
<th>Solanula</th>
<th>Aquatic rhizomes</th>
<th>Edible fungi</th>
<th>Flowers and leaves</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cauliflower</td>
<td>1***</td>
<td>0.042</td>
<td>0.192</td>
<td>0.177</td>
<td>0.08</td>
<td>0.332**</td>
</tr>
<tr>
<td>Peppers</td>
<td>0.042</td>
<td>1***</td>
<td>-0.388**</td>
<td>0.529***</td>
<td>-0.61***</td>
<td>0.564***</td>
</tr>
<tr>
<td>Solanula</td>
<td>0.192</td>
<td>-0.388**</td>
<td>1***</td>
<td>-0.487***</td>
<td>-0.501***</td>
<td>-0.242</td>
</tr>
<tr>
<td>Aquatic rhizomes</td>
<td>0.177</td>
<td>0.529***</td>
<td>-0.487***</td>
<td>1***</td>
<td>0.667***</td>
<td>0.582***</td>
</tr>
<tr>
<td>Edible fungi</td>
<td>0.08</td>
<td>0.61***</td>
<td>-0.501***</td>
<td>0.667***</td>
<td>1***</td>
<td>0.586***</td>
</tr>
<tr>
<td>Flowers and leaves</td>
<td>0.332**</td>
<td>0.564***</td>
<td>-0.242</td>
<td>0.582***</td>
<td>0.586***</td>
<td>1***</td>
</tr>
</tbody>
</table>

Note: ***, ** and * represent the significance levels of 1%, 5% and 10% respectively.

Figure 1: Correlation heat map between six categories
3.1.2 Analysis of the relationship between the sales volume of each single product

Drawing the heat map, the results show that the correlation between each item is not high, which is close to no linear relationship.

3.2 Analysis of daily replenishment amount and pricing of each category in the coming week

3.2.1 Analysis of the relationship between the total sales volume of each category and cost-plus pricing

The results of time series model are shown in Table 2. The F-test of the time series model is significant horizontally, so the original hypothesis that the regression coefficient is 0 is rejected, and the model basically meets the requirements. For the collinearity of variables, VIF is all less than 10, so the model has no multicollinearity problem and is well constructed.

Table 2: Analysis of Flowers and leaves Linear Regression Results

<table>
<thead>
<tr>
<th>B</th>
<th>Standard error</th>
<th>Standardization coefficient</th>
<th>t</th>
<th>P</th>
<th>VIF</th>
<th>R²</th>
<th>Adjust r</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>-1255.285</td>
<td>743.344</td>
<td>-</td>
<td>1.689</td>
<td>0.103</td>
<td>-</td>
<td>0.232</td>
<td>0.176</td>
</tr>
<tr>
<td>make a price</td>
<td>430.751</td>
<td>153.716</td>
<td>0.502</td>
<td>2.802</td>
<td>0.009***</td>
<td>1.131</td>
<td>0.406</td>
<td>0.028**</td>
</tr>
<tr>
<td>wholesale price</td>
<td>-41.773</td>
<td>99.311</td>
<td>-0.075</td>
<td>-0.421</td>
<td>0.677</td>
<td>1.131</td>
<td>1.131</td>
<td>1.131</td>
</tr>
</tbody>
</table>

Note: ***, ** and * represent the significance levels of 1%, 5% and 10% respectively.

From this, we can get the linear relationship between the pricing and wholesale price of Flowers and leaves vegetables:

\[ Y_j = 430.751P_j - 41.773W - 1255.285 \]  \hspace{1cm} (14)

Among them, \( Y_j \) is the total daily sales, \( P_j \) is the single-day Pricing, and \( W \) is the Wholesale price.

Other kinds of vegetables are the same, and finally get:

Cauliflower vegetables:

\[ Y_j = 3.178P_j - 39.831W + 398.1 \]  \hspace{1cm} (15)

Pepper vegetables:

\[ Y_j = -59.64P_j - 87.118W - 1255.285 \]  \hspace{1cm} (16)

Solanula vegetables:

\[ Y_j = 6.052P_j - 15.438W + 224.86 \]  \hspace{1cm} (17)

Edible fungi vegetables:

\[ Y_j = -4.381P_j + 1.941W + 320.943 \]  \hspace{1cm} (18)

Aquatic rhizomes vegetables:

\[ Y_j = 9.532P_j - 32.054W + 500.988 \]  \hspace{1cm} (19)

3.2.2 Determine the daily replenishment amount of each category in the coming week

Based on the ARIMA time series forecasting model, the daily sales volume from July 1 to 7, 2023 is predicted, as shown in Table 3.

R-square test results show that the fitting effect of the model is relatively good, that is, the model can better explain the variability of actual data. In addition, in order to evaluate the stability and
reliability of the model, the stability of the model is tested. Stability test is to judge whether there is systematic deviation or instability in the model by analyzing the residual of the model. If the result of stability test can't reject the hypothesis that the residual of the model is a white noise sequence, that is, there is no obvious pattern or trend of the residual, it shows that the performance of the model is relatively consistent in different time periods or different samples, which meets the requirements of stability. Based on the above analysis, the model performs well in R-square test and stability test, which shows that the model has good performance in fitting data and forecasting ability.

Table 3: Daily Total Sales Forecast Table

<table>
<thead>
<tr>
<th>Date</th>
<th>Cauliflower</th>
<th>Flowers and leaves</th>
<th>Aquatic rhizomes</th>
<th>Peppers</th>
<th>Solanula</th>
<th>Edible fungi</th>
</tr>
</thead>
<tbody>
<tr>
<td>2023-07-01</td>
<td>264.04</td>
<td>636.29</td>
<td>299.21</td>
<td>515.39</td>
<td>178.16</td>
<td>241.59</td>
</tr>
<tr>
<td>2023-07-02</td>
<td>235.65</td>
<td>578.03</td>
<td>283.88</td>
<td>551.00</td>
<td>158.82</td>
<td>232.72</td>
</tr>
<tr>
<td>2023-07-03</td>
<td>228.50</td>
<td>588.89</td>
<td>294.59</td>
<td>565.58</td>
<td>153.64</td>
<td>223.81</td>
</tr>
<tr>
<td>2023-07-04</td>
<td>227.72</td>
<td>601.45</td>
<td>296.21</td>
<td>578.68</td>
<td>159.18</td>
<td>229.39</td>
</tr>
<tr>
<td>2023-07-05</td>
<td>227.82</td>
<td>615.73</td>
<td>297.08</td>
<td>590.45</td>
<td>168.87</td>
<td>230.83</td>
</tr>
<tr>
<td>2023-07-06</td>
<td>227.72</td>
<td>628.41</td>
<td>297.93</td>
<td>601.04</td>
<td>170.75</td>
<td>236.19</td>
</tr>
<tr>
<td>2023-07-07</td>
<td>227.46</td>
<td>628.00</td>
<td>298.74</td>
<td>610.55</td>
<td>169.66</td>
<td>236.31</td>
</tr>
</tbody>
</table>

As the daily sales volume and daily replenishment volume satisfy the following formula:

\[ R_j = Y_j \times (1 + L) \]  \(\text{(20)}\)

Where, \(R\) is the Daily replenishment and \(L\) is the Loss rate.

Table 4 and Figure 2 can be obtained by calculating the daily replenishment amount from the daily sales amount and loss rate.

Table 4: Total daily replenishment of each category

<table>
<thead>
<tr>
<th>Date</th>
<th>Cauliflower</th>
<th>Flowers and leaves</th>
<th>Aquatic rhizomes</th>
<th>Peppers</th>
<th>Solanula</th>
<th>Edible fungi</th>
</tr>
</thead>
<tbody>
<tr>
<td>2023-07-01</td>
<td>276.38</td>
<td>657.31</td>
<td>333.96</td>
<td>550.79</td>
<td>185.38</td>
<td>254.75</td>
</tr>
<tr>
<td>2023-07-02</td>
<td>246.66</td>
<td>597.12</td>
<td>316.84</td>
<td>588.85</td>
<td>165.26</td>
<td>245.40</td>
</tr>
<tr>
<td>2023-07-03</td>
<td>259.17</td>
<td>608.35</td>
<td>328.80</td>
<td>604.43</td>
<td>159.86</td>
<td>236.00</td>
</tr>
<tr>
<td>2023-07-04</td>
<td>238.37</td>
<td>621.32</td>
<td>330.61</td>
<td>618.43</td>
<td>165.64</td>
<td>241.89</td>
</tr>
<tr>
<td>2023-07-05</td>
<td>238.46</td>
<td>636.08</td>
<td>331.58</td>
<td>631.01</td>
<td>175.72</td>
<td>243.40</td>
</tr>
<tr>
<td>2023-07-06</td>
<td>238.36</td>
<td>649.18</td>
<td>332.53</td>
<td>642.33</td>
<td>177.67</td>
<td>249.05</td>
</tr>
<tr>
<td>2023-07-07</td>
<td>238.09</td>
<td>648.75</td>
<td>333.44</td>
<td>652.49</td>
<td>176.54</td>
<td>249.18</td>
</tr>
</tbody>
</table>

Figure 2: Line chart of replenishment quantity in the next week and Sunday

3.2.3 Formulate pricing strategies for each category in the coming week

Assuming that the average price of vegetables in one week remains unchanged, the average daily sales of vegetables in the next week will be calculated to get the average daily sales of vegetables in the next week. Now the average total sales volume and average wholesale price are known. Substituting these two items into the multiple linear regression equation, we can get the preliminary pricing as shown in Table 5 and Figure 3.
Table 5: Preliminary pricing strategies for each category

<table>
<thead>
<tr>
<th>Category</th>
<th>Cauliflower</th>
<th>Flowers and leaves</th>
<th>Aquatic rhizomes</th>
<th>Peppers</th>
<th>Solanula</th>
<th>Edible fungi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pricing (yuan/kg)</td>
<td>6.935</td>
<td>4.65</td>
<td>17.47</td>
<td>13.2</td>
<td>0.53</td>
<td>22.49</td>
</tr>
</tbody>
</table>

By, it can be seen that the pricing of pepper, Solanula and edible fungi vegetables deviates from the historical pricing seriously, which shows that these pricing does not meet the actual market demand and consumer buying habits. This is because the demand elasticity coefficient of these three kinds of vegetables is small, and the price fluctuation has little influence on the consumer demand, which shows that the fitting degree of the multiple linear regression model of the corresponding vegetables is relatively low. Therefore, in order to formulate a more reasonable pricing strategy, it is necessary to re-price these three categories of vegetables by using the historical average price pricing method. By investigating the average price in the past period and combining the market demand and supply, a more reasonable and stable pricing level can be obtained. The final pricing strategy is shown in Table 6.

Table 6: Final pricing strategies of each category

<table>
<thead>
<tr>
<th>Category</th>
<th>Cauliflower</th>
<th>Flowers and leaves</th>
<th>Aquatic rhizomes</th>
<th>Peppers</th>
<th>Solanula</th>
<th>Edible fungi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pricing (yuan/kg)</td>
<td>6.935</td>
<td>4.65</td>
<td>17.47</td>
<td>6.33</td>
<td>7.88</td>
<td>5.60</td>
</tr>
</tbody>
</table>

3.3 Pricing Analysis of Individual Items

3.3.1 Determine the daily sales volume and daily replenishment quantity of each single product

Based on the details of sales flow from June 24 to 30, 2023, each single product was screened, so that the demand of each single product met the requirement of minimum display volume of 2.5kg, and finally 29 vegetable single products were screened out.

Use MATLAB to establish an exponential smoothing time series for the daily sales total data of single products from June 1 to 30, 2023, and predict the daily sales total of each single product on July 1, 2023. Then through the relationship between daily sales volume and daily replenishment volume, the daily replenishment volume of each item on July 1 can be predicted, such as Figure 4 shown.

3.3.2 Formulating pricing strategies for individual items

Based on the multiple regression model, firstly, with pricing and cost price as independent variables
and daily sales volume as dependent variables, multiple linear regression models were established for 29 items, and linear relationships were obtained. Taking Flammulina velutipes (box) as an example, a multiple linear regression model was established, and the test results showed that the model was well constructed.

Because the sales volume of each category of vegetables is different, we can adopt different pricing strategies for different vegetables on the basis of regression model to improve consumers' interest in buying.

Firstly, these 29 vegetable items were classified by ABC classification. The essence of ABC classification is to queue all materials according to their "importance", so that management can grasp the key points and grasp the key points.[8]. In this paper, Class A commodities are mainly commodities with high importance and accumulated sales of more than 60%-80%. These commodities are usually key products that customers often buy and contribute greatly to the sales and profits of merchants. Class B commodities are mainly commodities with medium importance and accumulated sales of more than 20%-30%, which play a medium role in business sales and have certain market demand and profit potential. Class C commodities are mainly commodities with low importance and low sales frequency. The market demand of these commodities is relatively small, and their contribution to the sales and profits of merchants is low. The classification is shown in Table 7.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Single Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class A single product</td>
<td>Yunnan lettuce (portions); Net lotus root (1); baby Chinese cabbage, millet pepper (part); Honghu lotus root belt; purple eggplant (2); Wuhu green pepper (1); edible amaranth; Screw pepper; screw pepper (part); broccoli; Xixia Flower Mushroom (1)</td>
</tr>
<tr>
<td>Class B single product</td>
<td>Shanghai green; Bosporus mushroom (box); Milk cabbage; Small wrinkled skin (part); Small vegetables (1); Malabar spinach; Zhijiang green stem scattered flowers; bengal dayflower herb; Spinach (part); Flammulina mushroom (box); High melons (1)</td>
</tr>
<tr>
<td>Class C single product</td>
<td>Yunnan oilseed lettuce (portion); Ginger, garlic and pepper combination (small part); Fresh mushroom (bag); Sweet potato tip; Green eggplant (1); fresh fungus</td>
</tr>
</tbody>
</table>

Then, the odd mantissa pricing strategy and integer pricing strategy are used to price different kinds of vegetable items respectively, thus affecting consumers' purchasing strategies.[9].

Odd mantissa pricing strategy is a marketing pricing strategy, and its basic idea is to set the price of products as numbers ending in odd numbers. Research by Jungkeun Kim and Peter Beomcheol Kim proves that prices with non-zero endings will lead to increased sales and more positive customer attitudes.[10]. For example, the price can be set at 19.99 yuan, 29.95 yuan, etc. When facing the commodity price, consumers usually pay more attention to the rightmost and leftmost figures. Compared with the middle figures, consumers are 19-29% more likely to notice the changes in the left and right figures, but if the price after the price increase ends with the number 9, the probability of consumers noticing the price increase will be reduced by 11%.[11]. Therefore, by setting the price to a number ending in an odd number, it can attract consumers' attention and make the price more attractive and close to the people, which is mainly used for goods with large sales volume.

Figure 5: Bar chart of pricing of individual items on July 1
Integer pricing strategy is a marketing pricing strategy that sets the price of products or services as an integer. This means that the price does not contain decimal points or decimal parts, but uses integers for pricing. The basic principle is to set the price as an integer to simplify consumers' understanding and decision-making process of the price. Integer prices are easier to remember and compare, and may appear more credible and reasonable in consumers' minds, mainly used for goods with small sales volume.

Based on the above pricing strategy, it is decided to price Class A and Class B with reference to odd mantissa pricing strategy and Class C with reference to integer pricing strategy on the basis of multiple linear regression. The result is shown in Figure 5.

4. Conclusion

In order to guide vegetable sales, based on Pearson correlation analysis, multiple linear regression model, ARIMA time series model and exponential smoothing time model, this paper obtains the following research results:

1) There is a certain correlation between some kinds of vegetables. Through Pearson correlation analysis, we can understand the correlation of consumers buying vegetables, that is, we can know which vegetables are easy to be bought together and which vegetables are unlikely to be bought together. Correlation analysis can be used to guide the arrangement of the relative positions of various vegetables.

2) By establishing multiple linear regression model and ARIMA time series model, we can get the relationship between total daily sales of vegetable commodities and single-day pricing and wholesale price, so as to predict the future total daily sales and future daily replenishment. Finally, we can get the final pricing strategy of each category of vegetable commodities by combining historical pricing method.

3) The daily sales volume and daily replenishment amount are determined by establishing exponential smoothing time series, and a multiple linear regression model is established to formulate the initial pricing strategy of each item. Then, the vegetable items are classified to obtain the final pricing strategy by adopting different pricing strategies for different types of vegetables.

4) The fitting degree of some regression models in this paper is not high enough, and this direction is worth further study.

References


