Merchandise replenishment and pricing decisions based on time series model forecasting and log-linear analysis with split logarithms

Haiyang Qiu*

Shanghai University of Finance and Economics, Shanghai, China

*Corresponding author: 1473559172@qq.com

Abstract. In this paper, in the process of researching the replenishment and pricing decisions of vegetable products, we use the Spearman model and log-linear analysis with logarithms to determine the correlation between individual vegetable products and categories, and analyze the relationship between the total amount of sales of each vegetable category and the cost-plus pricing, and then build a time-series model to fit the prediction to get the daily replenishment and pricing strategy for the coming week. Then a time series model is fitted to predict the daily replenishment and pricing strategy for the coming week. In this paper, based on the model of Spearman's correlation coefficient calculation, we set up a log-linear analysis model with logarithmic scores, and use MATLAB to calculate the correlation coefficients of each type of vegetables, and find that the total sales volume of each type of vegetables has a strong correlation with the cost-plus pricing and a high degree of fit. Then we establish ARIMA model in time series model to fit the time with the sales volume, cost and pricing of vegetables respectively, and predict the sales volume, cost- and cost-plus pricing in the coming week after passing the white noise test to get the total daily replenishment and pricing strategy.

Keywords: ARIMA, Spearman, Log-linear Analysis of Fractional Logarithms.

1. Introduction

With increasing competition in the fresh produce market, sound inventory and pricing strategies have become critical to superstore operations. In this analysis, we explore how to predict future sales of vegetable categories based on past sales data, wholesale prices, and item attrition rates, and provide replenishment and pricing recommendations accordingly [1].

We first analyzed historical sales data for each vegetable category and used the moving average method to predict their sales in the coming week. We adjusted the predicted sales volume to calculate the replenishment volume, taking into account the wear and tear rate of the items. For the pricing strategy, we used the "cost-plus pricing" method to set retail prices for each category and item based on the latest wholesale prices [2-3].

For the item replenishment analysis, we focused on the 22 most popular items. These items were selected based on the most recent week's sales data. We provided replenishment volume forecasts for these individual items and pricing strategies for them based on wholesale prices and a 20% markup percentage [4].

We obtained replenishment and pricing recommendations for the vegetable category and 22 individual items for the coming week. These recommendations can help superstores optimize their inventory to avoid overstocking and out-of-stocks, and ensure that items are priced to match market demand [5-6].

By deeply analyzing historical sales data, we can provide superstores with valuable insights about future sales, replenishment and pricing. However, these recommendations are based on historical data and actual market feedback and sales may vary. Therefore, superstores should remain flexible in implementing these recommendations and make adjustments accordingly. In addition, to more accurately predict future sales and demand, superstores may consider collecting more market and consumer data and applying more sophisticated forecasting models [7].

2. The fundamental of data analysis

The Spearman model is a nonparametric statistical method for evaluating the correlation between two variables in which the measure of the variable is expressed as a rank order.

The Spearman model analyzes correlations based on the rank of the variables rather than the actual observations. It measures the degree of association between variables by converting raw data to ordinal data and then calculating the correlation between the ordinal scores.

Correlation coefficients range from -1 to 1, where -1 indicates a perfect inverse relationship, 1 indicates a perfect positive relationship, and 0 indicates no relationship [8].

The following is the procedure for analyzing correlations using the Spearman model:

1. import data related to daily sales volume of various types of vegetables.
2. Sort the data to obtain ordinal data and assign ranks.
3. Calculate the ordinal mean between each ordinal data and deal with possible duplicate values.
4. use the ordinal data to calculate the Spearman correlation coefficient:

Based on the Spearman model, we first normalize the sales volume of each vegetable, remove the effect of the scale, and then use the Spss tool to calculate the Spearman correlation coefficient, and get the correlation coefficient table, and in order to make the results more significant, we use MATLAB to draw a heat map of the correlation coefficient as shown in Fig. 1, and the correlation coefficient of the vegetable sales volume is shown in Fig. 1, and the correlation coefficient of the vegetable sales volume is shown in Fig. 1.

Table 1: Correlation coefficients for various types of vegetables

R	Cauliflower	Foliage	Eggplant	Peppers	Mushrooms	Aquatic Roots
Cauliflower	1	0.712	-0.160	0.667	0.715	0.739
Foliage	0.712	1	-0.182	0.735	0.730	0.696
Eggplant	-0.160	-0.182	1	-0.230	-0.329	-0.247
Peppers	0.667	0.735	-0.230	1	0.729	0.723
Mushrooms	0.715	0.730	-0.329	0.729	1	0.732
Aquatic Roots	0.739	0.696	-0.247	0.723	0.732	1

From this we can conclude:

1. cauliflower is strongly correlated with chili peppers, edible mushrooms, and aquatic rhizomes, and weakly correlated with Lycopersicon. 2. cauliflower and cauliflower, pepper, edible fungi, aquatic rhizomes have a strong correlation, and eggplant has a very weak correlation. 3. eggplant and the other five categories.

3. tomato and the other five categories are only very weak correlation. 4.

4. the hot pepper class in addition to the eggplant class has a very weak correlation, and the other four categories have a very strong correlation.

5. edible mushrooms have a strong correlation with the other four categories except for a weak correlation with eggplant.

6. Aquatic roots and tubers are strongly correlated with the other four categories, except for a weak correlation with Solanum.

We also used random sampling to analyze the correlation of 18 vegetables: Shanghai bok choy, baby cucumber, Yunnan oilseed rape, broccoli, green stemmed cauliflower, zhijiang green stemmed cauliflower, net lotus root (1), lobelia, gourami (2), purple eggplant (2), flower eggplant, round eggplant (2), red bell pepper, red bell pepper (1), lantern bell pepper (2), white fungus (dosage), almond mushrooms (pouch), and seaweed mushrooms (pouch)(3) (see Annex), and obtained the correlation results. The correlation analysis of these 18 vegetables (please refer to the Annex for details) shows the correlation between these 18 vegetables [9-10].

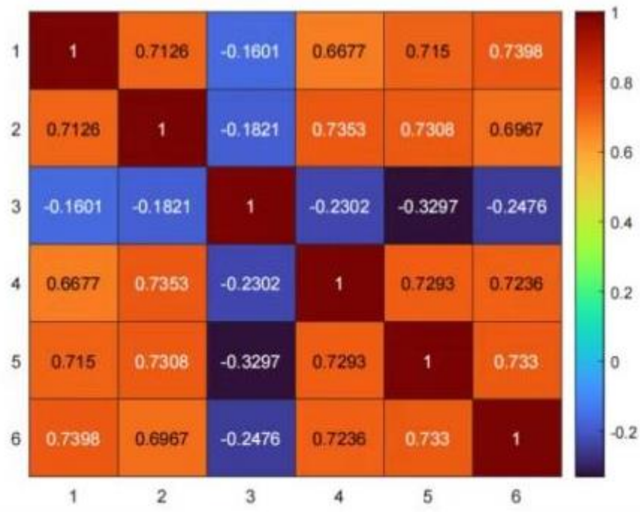

Figure 1: Heat map of correlation coefficients

For a Spearman correlation coefficient rs, we set the original hypothesis r = 0. In the large sample case, the statistic is approximated as a standard normal distribution N(0,1). The probability density function (PDF) of the standard normal distribution is shown in Figure 2:

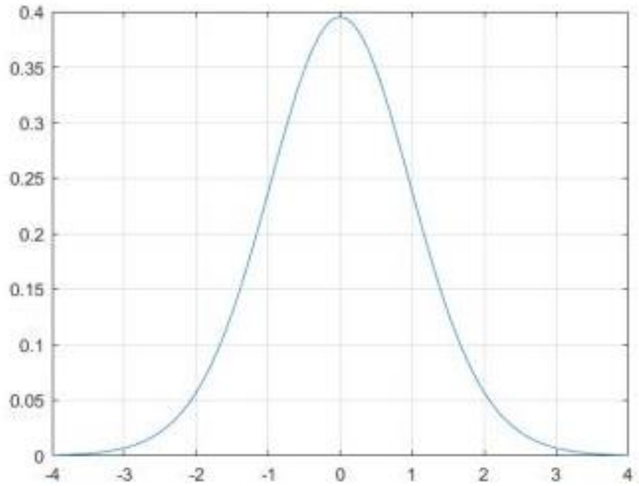

Figure 2: Probability density function

The p-values of the probability density functions are shown in Table 2: Since the p-values of each vegetable variety are less than 0.05, the original hypothesis is rejected at the 95% confidence level. Therefore, we can conclude that there is a correlation between the vegetable categories.

Table 2: Various vegetables p-value

R	Cauliflower	Foliage	Eggplant	Peppers	Mushrooms	Aquatic Roots
Cauliflower	1.0000	0.0000	0.0024	0.0000	0.0000	0.0000
Foliage	0.0000	1.0000	0.0005	0.0000	0.0000	0.0000
Eggplant	0.0024	0.0005	1.0000	0.0000	0.0000	0.0000
Peppers	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
Mushrooms	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
Aquatic Roots	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

3. Log-linear analysis of fractional logarithms

3.1. Steps in log-linear analysis of fractional logarithms

We use the following steps for log-linear analysis using split logarithms:

1. Data preparation: Filter the data for total sales and cost-plus pricing for the vegetable category.

2. Logarithmic transformation: Logarithmically transform the two variables to reduce the skewness and variance of the data. The natural logarithm is used for the transformation, noting that the transformed variable is ln(Sales) for total sales and ln(Price) for cost-plus pricing. 3.

3. Fitting a log-linear model: Construct a log-linear model to describe the relationship between the variables. The general form of the log-linear model is.

$$\log(Y_{ij}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \gamma_1 Z_1 + \gamma_2 Z_2$$

Where Y_{ij} is the logarithmic value of the frequency or proportion of categorical data in the subgroups after grouping, X_1 and X_2 are independent variables, Z_1 and Z_2 are indicator variables for subgroups or hierarchies, β_0, β_1 and β_2 are regression coefficients for the independent variables, and $\gamma 1$ and $\gamma 2$ are regression coefficients for the indicator variables, which represent the overall effect of the independent variables on the categorical data and the differences between subgroups, respectively.

4. Analysis: Log-linear analysis with logarithmic scores using the Spss tool. 5.

5. Correlation Analysis: The correlation between the variables is determined by testing the coefficient $\beta 1$ of the model. If the coefficient $\beta 1$ is less than the pre-determined significance level of 0.01, then it is considered that there is a correlation between total sales and cost-plus pricing.

By comparing the parameters of the three models, we choose R^2 as the three main parameters to compare the accuracy of the models. Finally, we found that Holt model is more accurate among these three models.

We studied the total sales volume of each vegetable category in relation to time and found that: the total sales volume of cauliflower, tomato, chili, and aquatic root vegetables varied approximately the same over time. Total sales of foliage and flowers fluctuated steadily over time, except for a sudden increase in sales at certain times. 3. 3 Edible mushroom sales are relatively stable in the middle of the time period, with a surge at the beginning and end of the time period. Based on this, we build a time series model to forecast in daily units.

The ARIMA(p1,d,q) model is built first, we studied the total sales volume of each vegetable category in relation to time and found that: the total sales volume of cauliflower, tomato, chili, and aquatic root vegetables varied approximately the same over time.

Based on this model, the forecasted sales volume and cost-plus pricing for each vegetable category are shown in Table 3.

Conclusion: Based on the above predictions, we follow the principle of market supply equilibrium, so we take the predicted daily sales volume of each type of vegetable as the daily capture volume, and the predicted average cost-plus pricing as the pricing, as shown in Table 3.

Table 3: Forecast of sales volume and cost plus pricing of various types of vegetables

Model(July 2023）	**1** day	**2** days	**3** days	**4** days	**5** days	**6** days	**7** days
Kilograms sold: Cauliflower	22.57	20.67	19.95	21.22	22.13	23.36	25.13
Average cost plus pricing. Cauliflower	11.58	11.58	11.58	11.58	11.58	11.58	11.58
Volume in kg: Cauliflower	138.61	145.81	145.32	148.59	147.98	149.78	152.12
Average cost plus pricing. Foliar	5.03	5.03	5.05	5.02	4.99	5.00	5.00
Volume in kilograms: Solanum	21.53	20.00	18.16	18.30	17.16	20.41	21.14
Average cost plus pricing. Tomato	7.92	7.92	7.92	7.92	7.92	7.92	7.92
Volume in kg: Peppers	83.41	83.67	81.28	80.91	80.31	80.66	81.02
Average cost plus pricing. Peppers	6.19	6.23	6.21	6.19	6.28	6.22	6.25
Volume kg: Edible Mushrooms	48.54	49.02	52.04	52.48	53.82	51.85	49.99
Average cost plus pricing. Edible Mushrooms	5.44	5.49	5.58	5.627	5.60	5.54	5.62
Volume in kg. Aquatic Roots	20.36	19.41	19.70	19.90	20.04	20.16	20.27
Average cost plus pricing. Aquatic Roots & Tubers	15.61	15.63	15.65	15.67	15.69	15.70	15.72

4. Conclusions

In this paper, in the process of researching the replenishment and pricing decision-making process of vegetable commodities, this paper analyzes the traditional total sales volume and cost-plus pricing model of vegetable categories by constructing the correlation coefficient model and linear analysis model, determines the correlation between individual vegetable products and categories and analyzes the relationship between the total sales volume and cost-plus pricing of each vegetable category by using Spearman's model and log-linear analysis of logarithmic numbers. Then we establish a time series model to fit the prediction and get the daily replenishment quantity and pricing strategy for the coming week and conduct a prediction study on the total sales quantity, cost and pricing, and get the ideal result through reasonable means, which provides technical means for the supervision and management of the market operation.

References

[1] Mao Lisha. Research on Pricing Strategy and Production and Marketing Mode of Vegetable Wholesale Market under the Perspective of Supply Chain [D]. Central South Forestry University of Science and Technology,2022.

[2] Fang Li. Research on the pricing method of vegetable raw materials in college logistics [J]. College Logistics Research, 2021, (10): 22-24.

[3] Wei Danlin. Research on the risk and pricing of vegetable income insurance in Shandong [J]. Rural Economy and Technology, 2021, 32(02): 69-70+160.

[4] Ding Wenbo. Analysis on the Development of Facility Vegetable Industry Based on Economics Perspective [J]. Industry and Technology Forum, 2020, 19(09): 13-14.

[5] Ma'mun Sarma, Marthin Nanere, Philip Trebilcock. Pricing strategies for organic vegetables based on Indonesian consumer willingness to pay [J]. Management Science Letters, 2020.

[6] Zhao Yu. Pricing research on vegetable revenue insurance based on semiparametric Copula method [J]. Agricultural Modernization Research, 2019, 40(02): 308-315.

[7] Sun Xiaolu. Vegetable price index insurance pricing based on policyholders' expectations [D]. Northeast University of Finance and Economics, 2018.

[8] Sun Xiaolu. Vegetable price index insurance pricing based on policyholder expectations [J]. Journal of Insurance Vocational College, 2018, 32(04): 22-28.

[9] QIAN Hua, YU Jusheng, LUO Qiang, ZHANG Chen. Research on the integration of Shanghai vegetable supply chain for collaborative pricing [J]. China Agronomy Bulletin, 2013, 29(35): 386-390.

[10] ZHENG Hao-hao, ZHANG Wen, DING Xue-feng, FU Ya-san. Research on joint pricing model of stakeholders in China's vegetable market [J]. Logistics Technology, 2013, 32(11): 121-123.