Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 3
CECS 2023

Automated pricing and replenishment decisions for vegetable products based on evaluation optimization models

Zhichun Wei

College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou, China

1325926749@qq.com

Abstract. Based on the commodity information of the supermarket in the annex, the detailed data of historical sales flow, the wholesale price of vegetable commodities and the recent loss rate of vegetable commodities, and through the data analysis of each category and each single product, the automatic pricing and replenishment decision-making model of commodities is established. Use the optimization evaluation algorithm to formulate the total daily replenishment and pricing strategy of each category and each single product. In order to solve the first problem, firstly, the outliers in the original data of Annexes 2 and 3 are cleaned, normalized, feature selected and dimensionally reduced. Secondly, a quarter is taken as a sales cycle of supermarkets, so as to find the proportion of sales volume of a certain category in the same quarter of three years to the total sales volume, and give the distribution law of sales volume of different categories, the results are shown. Considering different periods again, the daily sales volume distribution law is calculated by taking one day as a sales cycle, and the results are shown. Finally, the Pearson grade correlation coefficient is used to judge the relationship between the processing indicators, and the matrix heat map is obtained. According to the two results, it was concluded that there was a significant positive correlation between the sales volume of mosaic and cauliflower vegetables, and a significant negative correlation between the sales volume of nightshade and aquatic root vegetables. In view of the second problem, firstly, considering the functional relationship between the total sales volume and the cost pricing, the correlation analysis and linear fitting were carried out to obtain the linear relationship between the sales price of each category and the maximum value of the sales volume of each category in July of the previous year can be described as $p = -0.1300m_t + 14.4080$ Through further nonlinear fitting and optimization problem solving, the total daily replenishment volume and pricing strategy of each vegetable category in the coming week (July 1-7, 2023) are shown in Table 1 and Table 2, which makes the supermarket have the largest revenue In response to the third question, based on the known data, we can analyze the data requirements for each data: we need to know the sales volume of various vegetables during this period, we need to determine the purchase cost of each vegetable, we need to understand the past pricing strategy and response, and we need to know the inventory of various vegetables on June 30. On this basis, a multi-objective dynamic programming model is established, and the total number of saleable items is 30 by using the greedy algorithm to obtain the replenishment quantity of single items on July 1, and the pricing strategy is further solved by using the linear equation fitted in problem 2. In response to the fourth problem , on the basis of the existing sales, wholesale price and loss rate data, in order to better formulate the replenishment and pricing decisions of vegetable products, supermarkets also need to consider and collect the following 12 aspects of relevant data to assist in planning the pricing and replenishment decisions of vegetable products, such as: customer preference and satisfaction survey, seasonality and availability of vegetables, competitor information, inventory costs and storage conditions, historical sales data and trend analysis, customer flow and purchase period, nutritional value and health benefits of vegetables, Socio-economic factors, external environmental factors, policy and regulatory factors, technological and innovation factors, and supply chain and logistics information to ensure more comprehensive and accurate decision-making. Among them, the analysis of historical sales data and trends is mainly carried out.

Keywords: Pauta criterion, Pearson correlation coefficient, multi-objective dynamic programming, greedy algorithm.

1. Restatement of the problems

Vegetable products in fresh food superstores have a short shelf life and deteriorate in quality over time, so superstores are replenished daily based on historical sales and demand.

Supermarkets sell a wide variety of vegetables from different origins, which are brought in between 3:00 a.m. and 4:00 a.m. Merchants need to make decisions about replenishment of the day's vegetable categories without knowing exactly what the products are and the prices at which they will be purchased. Vegetables are often priced on a "cost-plus" basis, and supermarkets often offer discounts for shipping losses and poor quality items. Reliable market demand analyses are critical to replenishment and pricing decisions. Sales volumes are time-dependent, and on the supply side, with a wide variety of vegetables available from April to October, the supermarket's sales mix strategy is critical.

Annex 1 gives the commodity information of six vegetable categories distributed by a superstore; Annex 2 and Annex 3 give the sales flow details and wholesale price data of each commodity from 1 July 2020 to 30 June 2023 respectively; Annex 4 gives the recent wastage rate data of each commodity. A mathematical model is developed to solve the following problems based on the annexes and the actual situation:

In question 1, there may be a certain correlation between different categories of vegetables or different single product, please analyze the distribution pattern of the sales volume of each category and single product of vegetables and the interrelationship.

Question 2: Assuming that the supermarket has a replenishment plan for each category, analyze the relationship between the total sales volume and cost pricing for each vegetable category, as well as the total daily replenishment volume and pricing strategy for the coming week (1-7 July 2023) to maximize the superstore's revenue.

Question 3: Due to the limited sales space for vegetable items, the superstore wishes to further develop a replenishment plan for individual items, requiring that the total number of saleable items be limited to 27-33 and that the replenishment of each item satisfy the minimum display quantity of 2.5 kg. Based on the varieties available for sale from 24-30 June 2023, give the replenishment quantity of individual items and pricing strategy for 1 July that will maximize the superstore's revenue while trying to meet the market's demand for each category of vegetable merchandise.

2. Question analysis

2.1. Judgement of the need for data pre-processing

Outliers in the raw data in Annexes 2 and 3 can undermine the accuracy of statistical analyses, and pre-processing is needed to identify and deal with outliers, which can be detected and corrected using either statistical-based or model-based methods. Data quality is the basis for reliable analysis, and pre-processing includes data cleansing, such as correcting data entry errors, removing duplicate data and standardizing data formats to ensure data accuracy and consistency.

2.2. Analysis of question 1

Question 1 requires a quantitative analysis of the distribution patterns and interrelationships of the sales volume of various categories and individual products of vegetables, and the main task is to integrate and analyze the data related to the sales volume of categories and individual products. For category sales analysis process, the single product data into category data, and the title of the data given to the super sales cycle for two definitions: first of all, a quarter as a super sales cycle, so as to find out three years in the same quarter of a category sales accounted for the proportion of the total sales of sales of different categories of sales distribution patterns, and then different times of the day as a Sampling data with different time of the day as a sales cycle, repeat the above analysis, using

Pearson's rank correlation coefficient, to deal with the relationship between the indicators to judge. Based on the two results combined to draw conclusions.

2.3. Analysis of question 2

Problem 2 is an optimization type of problem that takes into account the functional relationship between total sales and cost pricing. For the solution of the relationship, correlation analysis and non-linear fitting were performed to obtain a model for cost pricing and a mathematical basis. For the second part of the prediction, the historical data is processed and the mathematical model of non-linear programming and the cost-pricing relationship presented in the previous question are used to solve the prediction for the maximum value and to arrive at the most profitable solution.

2.4. Analysis of question 3

Problem 3 requires that the revenue of the superstore be maximized while satisfying the prerequisite market demand for each category of vegetables. Since there are two objectives, multi-objective planning is used. The analysis yields the constraints: the total number of single items available for sale, and the order quantity of each individual item to satisfy the minimum display quantity. According to the first constraint, the individual items were analyzed, and the relationship between the order quantity and the actual selling quantity in the second question was also considered. In this way, suitable items were selected and the wastage rate was optimized in order to select 36 items for further selection and the greedy algorithm was used to obtain the optimal solution to this problem by finding the local optimal solution.

3. Description of the symbol

symbol	illustrate	unit
i	**The number of sales per item**	**kg**
j	**The wholesale price of each item**	**Yuan/kg**
ρ	**Sample correlation coefficient**	\
d_i	**The difference in the rank value of the first data pair$\boldsymbol{i}$**	\
m_t	**Category sales were the highest in July of the previous year $\boldsymbol{t}$**	**kg**
η	**The attrition rate of the product category**	\
η_i	**The attrition rate of the $\boldsymbol{i}$ single item**	\
λ	**profit**	**Yuan**
p	**The selling price of each category**	**Yuan**
a	**Wholesale prices by category**	**Yuan/kg**
x_i	**Sales volume of single items $\boldsymbol{i}$**	**kg**
$\boldsymbol{y_i}$	**The estimated sales volume of the category $\boldsymbol{i}$**	**kg**
$\boldsymbol{z_j}$	**The known number of sales of a single item $\boldsymbol{j}$**	**kg**
$\boldsymbol{l}$	**Constraint maximum**	\
$\boldsymbol{\delta_i}$	**Forecast sales per item $\boldsymbol{i}$**	**kg**
$\boldsymbol{n_i}$	**The amount of replenishment of the item $\boldsymbol{i}$**	**kg**

4. Data preprocessing

In response to the data in Annex 2, some unreasonable data were found, which required data preprocessing. Data preprocessing includes removing useless data, dealing with missing data, dealing with outliers, and resolving data duplication to ensure data accuracy and consistency. Visualizing the raw data, as the sample data are approximately normally distributed, using the 3σ principle, the sample data with negative sales volume fitting in the questions are first removed, and then the unreasonable wholesale prices are first zeroed, and then the unreasonable wholesale prices are filled in by processing the missing values. From the practical point of view, but also will be processed to remove the sales of large data back to the processing results table, because vegetable sales are large in real life is objective and credible.

In a normal distribution, σ represents the standard deviation, μ represents the mean, and $x = \mu$ is the axis of symmetry of the image. The 3σ principle is

(1) The probability that the values are distributed in $(\mu-\sigma,\mu+\sigma)$ is 0. 6826;

(2) The probability that the values are distributed in $(\mu-2\sigma,\mu+2\sigma)$ is 0. 9544;

(3) The probability that the values are distributed in $(\mu-3\sigma,\mu+3\sigma)$ is 0. 9974.

It can be assumed that the values of Y are almost entirely concentrated in the interval $(\mu-3\sigma,\ \mu+3\sigma)$, with less than 0.3 per cent probability of being outside this range. The 3σ criterion is based on equal-precision repeated measurements that are normally distributed and cause interference or noise in the singular data that makes it difficult to satisfy the normal distribution. If the absolute value of the residual error $v_i > 3\sigma$ for a measurement in a set of measurement data, that measurement is bad and should be rejected. An error equal to $\pm 3\sigma$ is usually taken as the limiting error. For a normally distributed random error, the probability of falling outside $\pm 3\sigma$ is only 0.27%, and it is very unlikely to occur in a finite number of measurements, so the 3σ criterion exists. The 3σ criterion is the most commonly used and the simplest criterion to discriminate the gross errors, which is generally applied when the number of measurements is sufficiently large $(n \geq 30)$ or when $n > 10$ to make a rough discrimination. The formula is expressed as follows

$$\mathbf{878504\mu_1 = \sum_{t=1}^{878504} i_t,\ 878504\mu_2 = \sum_{t=1}^{878504} i_t.}$$

$$\boldsymbol{\sigma_1 = \sqrt{\frac{((i_t-\mu_1)\wedge 2 + (i_t-\mu_1)\wedge 2 + \cdots + (i_t-\mu_1)\wedge 2)}{878504}}, t = 1,2,\cdots\cdots 878504.}$$

$$\boldsymbol{\sigma_2 = \sqrt{\frac{((j_t-\mu_2)\wedge 2 + (j_t-\mu_2)\wedge 2 + \cdots + (j_t-\mu_2)\wedge 2)}{878504}}, t = 1,2,\cdots\cdots 878504.}$$

$$\boldsymbol{\begin{cases} i \sim N(\mu_1, \sigma_1^2) \\ P\{|i-\mu_1| < k\sigma_1\} = 2\phi(k) - 1 \\ P\{|i-\mu_1| < \sigma_1\} = 2\phi(1) - 1 = 0.6826 \\ \{P\{|i-\mu_1| < 2\sigma_1\} = 2\phi(2) - 1 = 0.9544 \\ P\{|i-\mu_1| < 3\sigma_1\} = 2\phi(3) - 1 = 0.9974 \end{cases}}$$

$$\boldsymbol{\begin{cases} j \sim N(u,\ \sigma_2^2) \\ P\{|j-\mu| < k\sigma_2\} = 2\phi(k) - 1 \\ P\{|j-\mu| < \sigma_2\} = 2\phi(1) - 1 = 0.6826 \\ \{P\{|j-\mu| < 2\sigma_2\} = 2\phi(2) - 1 = 0.9544 \\ P\{|j-\mu| < 3\sigma_2\} = 2\phi(3) - 1 = 0.9974 \end{cases}.}$$

Where i_t is the sales volume of each individual item in Annex 2, j_t is the wholesale price of each individual item in Annex 3, μ_1 represents the mean value of the sales volume of each individual item, σ_1 represents the standard deviation of the sales volume of each individual item, μ_2 represents the mean value of the wholesale price of an individual item, and σ_2 represents the standard deviation of the wholesale price of each individual item.

Data that are clearly outliers in the scatterplot should be removed based on the 3σ principle .

5. Modelling and solution of problem 1

5.1. Problem 1: Model building

After preprocessing the data, problem 1 can be divided into four parts, namely the distribution pattern of each category, the distribution pattern of each individual product, the interrelationship of each category and the interrelationship of each individual product. The sales data are time series in nature and can be analyzed using time series analysis to study the distribution pattern and trend of sales volume.

5.1.1. The distribution of sales volume of each category on a quarterly basis

For the first part, that is, the distribution of the sales volume of each category of vegetables, the data in Annex 2 after processing are divided into 12 cycles with a quarterly cycle, and the distribution of each category of vegetables in each cycle is calculated respectively, and a histogram reflecting the distribution law is made. The formula for calculating the sales volume of each vegetable category on a quarterly basis is as follows

$$y_{k,\ i} = \sum_{k=1}^{12} x_{k,\ h} \quad k = 1, 2, \cdots 6, l = 1, 2, \cdots 12. \tag{1}$$
$$h = f(t).$$

Where $y_{k,\ i}$ is the total product sales of category k in the i-th cycle, $x_{k,\ h}$ is the total product sales of category k on day h, and h is a function of time denoting day h of the i-th cycle.

5.1.2. The distribution of sales volume of each category on a one-day cycle

For the second part, i.e. the distribution pattern of the sales volume of each vegetable item, we take different hours of the day as a sales cycle, and take uniform samples from the data processed in Appendix 2 to calculate the distribution of each vegetable category in each cycle, and then make the quarterly sales volume three-dimensional bar charts and single-item scatter plots reflecting the distribution pattern. The formulae for calculating the sales volume of each vegetable category in different time periods are as follows

$$y_{k,\ i} = \sum_{k=1}^{12} x_{k,\ h} \quad k = 1, 2, \cdots 6, l = 1, 2, \cdots 12. \tag{2}$$
$$h = f(t).$$

Where $y_{k,\ i}$ is the total product sales of category k in cycle i, $x_{k,\ h}$ is the total product sales of category k at time h, and h is a function of time denoting time h in cycle i.

5.2. Review of correlation analysis

5.2.1. Correlation test based on SPSS tool

Since the data after the annex preprocessing satisfy the conditions of continuous data, normal distribution, and linear relationship, this paper firstly adopts the Pearson rank correlation coefficient, i.e. Pearson correlation coefficient, to judge the relationship between the treatment indicators. The Pearson correlation coefficient method uses monotonic equation to evaluate the correlation of two statistical variables if there is no repeated value in the data, and when two variables are completely monotonically correlated, the Pearson correlation coefficient is +1 or -1. The formula is as follows

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 (y_i - \bar{y})^2}}.$$

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}.$$

where ρ is the correlation coefficient and takes the value of [-1, 1]. d_i denotes the equal variance, x_i and y_i are the independent and dependent variables, respectively, and n is the total number of observed samples. Confidence intervals and null tests for the correlation coefficient ρ were obtained using the Fisher transform as follows

$$F(r) = \frac{1}{2}\ln\frac{1+r}{1-r} = arctanh(r).$$

$$z = \sqrt{\frac{n-3}{1.06}}F(r).$$

where $F(r)$ is the Fisher transform of r and z is the z -value of r . r approximately obeys a standard normal distribution when the statistical dependence is zero, i.e., $\rho = 0$. The significance of the Fisher transform is

$$\varphi = r\sqrt{\frac{n-2}{1-r^2}}.$$

It approximately obeys the standard normal distribution at the time, the degree of freedom of the distribution $\rho = 0$ is $n-2$, and the significance is φ.

The matrix heat map drawn by Python from the above formula is shown below

As can be seen, in terms of quarterly sales, cauliflower and foliage have a significant positive correlation, while eggplant and the other five categories show a negative correlation, especially a strong negative correlation with aquatic root vegetables. The remaining categories of vegetables showed medium-strength positive correlations. It can be seen that from all time periods, aquatic root vegetables showed weak positive correlations with edible mushrooms, eggplant, chilli, moderate positive correlations with cauliflower, and strong positive correlations with foliar vegetables. The remaining categories showed very strong positive correlations with each other.

6. Modelling and solution of problem 2

6.1. Establishment of evaluation models

6.1.1. Model building

Based on the results of problem 1, the data of problem 1 is further analyzed, firstly, the functional relationship between total sales volume and cost pricing needs to be considered to solve the optimization problem, and we obtain the model and mathematical basis of cost pricing by correlation analysis and nonlinear fitting. For the prediction of the second part, we use the processed historical data, the mathematical model of nonlinear programming and the cost-pricing relationship proposed in the previous question to solve the maximum value prediction, and obtain the most profitable scheme.

$$\eta = \sum\frac{x_i}{\sum x_i}\eta_i.$$

$$p = \frac{m_t}{1-\eta}\eta.$$

where η is the average attrition rate, η_i is the attrition rate of item i, and x_i is the sales volume of item i. p denotes the selling price of each category, and m_t denotes the maximum value of category sales in the tth July of the previous year. Firstly, x_i is analysed by using 0. 9 of the historical data of category sales in July of the previous years as the data for prediction, then using the ratio of the sales of each category to the total sales to determine the weight of η_i, and then using η_i to find out the average attrition rate η, and after that, using the attrition rate to find out the amount of incoming goods. The values of x_i are then used to fit a nonlinear fit to the required p.

6.1.2. Model solution

Based on the summary analysis of the previous model, the data in Annex 2 and Annex 3 were imported into Python and nonlinearly fitted according to the formulae established in the above model, and the results are as follows.

As can be seen, the actual fit basically meets the expectation of the preset fitting formula, and the final relationship is obtained as follows

$$p = -0.1300m_t + 14.4080.$$

6.2. Optimization model building

In 6. 1, in order to further narrow down the range of raw material suppliers, supermarkets are considered to do replenishment planning on a category-by-category basis, analyzing the total sales volume of each vegetable category in relation to the cost mark-ups.

6.2.1. Determination of the objective function

This model is a nonlinear programming model, the objective function is profit, and its mathematical expression is λ

$$p = \frac{m_t}{1-\eta}\eta.$$

$$\lambda = 0.9max\sum_{t=1}^{6}(p-a)m_t\,, t = 1,2,\cdots 6.$$

Where m_t is a logistic variable denoting the maximum value of category sales in the t-th July of the previous year, p denotes the selling price of each category, and a denotes the wholesale price of each category.

6.2.2. Determine the constraints

The decision variables are first set, where the decision variable p denotes the selling price of each category, the decision variable a denotes the wholesale price of each category, and the decision variable m_t denotes the maximum value of the sales volume of the category in the t-th July of the previous year.

Through the historical data analysis of the relationship between the sales volume and price of each category of vegetables, combined with the real-life pricing standards for vegetables is the minimum market price shall not be lower than the standard price of 30% of the maximum shall not exceed the standard price of 40%, three constraints are identified, the model is solved respectively.

$$\begin{cases} 0.9m_t \geq 0 \\ a \geq 0 \\ 1.3a \leq p \leq 1.4a \end{cases}.$$

6.2.3. Solving the model

The most economical replenishment program planning model obtained is as follows

$$\lambda = 0.9max\sum_{t=1}^{6}(p-a)m_t\,, t = 1,2,\cdots 6.$$

$$s.t\begin{cases} 0.9m_t \geq 0 \\ a \geq 0 \\ 1.3a \leq p \leq 1.4a \end{cases}. \tag{3}$$

At the same time, according to the non-linear relationship equation evaluated in problem two above, the total daily replenishment and pricing strategy of each vegetable category in the coming week which makes the supermarket gain the most can be obtained. The final results are as follows.

Table 1. The total daily replenishment plan for each vegetable category in the coming week (unit: kg).

date	Cauliflower	Mosaic and leafy	Chili peppers	Nightshades	edible fungi	Aquatic rhizomes
July 1st	30. 521	36. 636	25. 224	34. 568	27. 180	13. 513
July 2nd 2nd	34. 506	36. 952	34. 110	39. 474	25. 880	16. 908
July 3rd	31. 549	37. 575	26. 437	37. 096	29. 362	16. 474
July 4th	31. 673	37. 179	26. 632	34. 968	28. 578	20. 677
July 5th	31. 367	36. 775	26. 171	35. 798	27. 976	16. 434
July 6th	31. 668	36. 791	23. 848	36. 712	28. 284	20. 970
July 7th	31. 260	36. 842	27. 308	37. 382	28. 956	18. 144

Table 2. Daily pricing strategy planning table for each vegetable category in the coming week (unit: yuan/kg).

date	Cauliflower	Mosaic and leafy	Chili peppers	Nightshades	Edible fungus	Aquatic rhizomes
July 1st	18. 62	19. 51	18. 05	19. 59	18. 32	16. 25
July 2nd	19. 17	19. 55	19. 33	20. 32	18. 14	16. 71
July 3rd	18. 77	19. 64	18. 22	19. 96	18. 64	16. 65
July 4th	18. 78	19. 58	18. 25	19. 65	18. 52	17. 22
July 5th	18. 74	19. 53	18. 19	19. 77	18. 44	16. 65
July 6th	18. 78	19. 53	17. 85	19. 91	18. 48	17. 26
July 7th	18. 73	19. 53	18. 35	20. 01	18. 58	16. 88

7. Modelling and solution of problem 3

7.1. Replenishment volume strategy

For this question, according to the data processing of questions 1 and 2 and the attachments given in the title, a total of four data can be extracted, namely, sales data from June 24 to 30, 2023, supply prices of vegetables, historical pricing data of vegetables and inventory data. Based on these four data, it is possible to analyze the data requirements for each data: the need to know the sales volume of various vegetables during this period, the need to determine the cost of purchase of each vegetable, the need to understand the past pricing strategy and response, and the need to know the inventory of various vegetables as of June 30. On this basis, a multi-objective dynamic programming model is established, and the local optimal solution of the subproblem is obtained firstly by using the greedy algorithm to solve each subproblem, and then the local optimal solution of the subproblem is synthesized into a solution of the original solution of the problem

7.1.1. Dynamic modelling

According to the requirements of the topic, the overall analysis of the single product data, for the preprocessed attachment data, first of all, through the 3 σ principle to remove the data that do not meet the expectations, and then based on the available varieties of 24-30 June 2023 filtered out the sales volume of 49 types of single product, in order to calculate the filtered single product type of the average daily income, in order to get the generality of the regulation.

Table 3. Statistics of abnormal profitable categories obtained by data screening

date	Single product category	Take profit on the day
2023-06-25	Cordyceps flowers (servings)	-0. 36
2023-06-27	Cordyceps flowers (servings.)	-0. 37
2023-06-28	Hong hu lotus root	-1. 15
2023-06-30	Tall melon (1)	-4. 19
2023-06-30	Crab mushroom and white mushroom double	-0. 44

From the analysis of Table 3, it can be seen that "Cordyceps flowers (portions)" for two consecutive days of single product profits are negative, the other single product losses in the table may be fortuitous, therefore, consider first of all, the category will be excluded, to obtain the 48 types of single product in the 24-30 June in the year 2023.

Based on the analysis in Table 3, the combined analysis allows for further screening of the 48 single product types within June 24-30 within 2023. Based on the scatter distribution, the 15 single product types with low scatter average daily returns are excluded, and further discussion is carried out among the remaining 33 single product types.

After comprehensive analysis, the above problem can be solved using multi-objective dynamic programming. Dynamic programming is a mathematical optimization method used to solve multi-stage decision problems. It is a method of solving complex problems by decomposing the original problem into simplified subproblems. After consideration, it is decided to solve the problem by the knapsack problem, which introduces the unit sales, unit price and total replenishment as planning factors. The knapsack problem is a classical problem in computer science and optimization. In this problem, given a set of items, each of which has its own weight and value, the objective is to select a number of items from the set such that their total weight does not exceed a given limit and their total value is maximized.

Table 4. Statistical table of loss rate by category

category	Cauliflower	Mosaic and leafy	Chili peppers	Nightshades	Edible mushrooms	Aquatic rhizomes
Attrition rate	**16. 70%**	**14. 99%**	**9. 95%**	**7. 53%**	**10. 14%**	**21. 15%**

7.1.2. Identification of decision variables

In the knapsack problem, the decision variable can be the choice or non-choice of the item, setting the choice of the item to 1 and the non-choice to 0.

7.1.3. Determining the state and state transfer equations

The state defines the various possible cases of the problem, while the state transition equation describes the relationship between the states. In the backpack problem, the state can be the weight of the current item and the backpack, and the state transition equation describes how the value of the backpack changes with or without the current item. The loss rate in Table 4 is used as the evaluation standard.

7.1.4. Build a model to solve

Using the greedy algorithm, each vegetable category is viewed as a space, and by sampling the replenishment of stacked individual items in a given space, the following formula can be obtained.

$$y_i - \sum_{j=1}^{l} z_j < 0. \tag{4}$$

where y_i is the predicted sales volume of vegetable category i, z_j is the predicted sales volume of item j under the category corresponding to the descending order by how much profit is made, and l is the constrained maximum value that makes equation (4) hold for the first time.

$$\delta_j = \begin{cases} z_j, j < l \\ y_i - \sum_{1}^{l-1} z_j, j = l \end{cases} .$$

$$n_i = \frac{1}{1-\eta_i}\delta_j$$

Where δ_j is the predicted sales volume of item i, n_i is the replenishment volume of item i, and η_i is the wastage rate of item i . Using Python solution can be obtained, there are 30 kinds of sellable single product vegetable number, the table is as follows.

Table 5. Replenishment statistics of single items on July 1

Singles	category	Replenishment amount
broccoli	**Cauliflower**	**15. 076**
Zhijiang green stalk scattered flowers	**Cauliflower**	**4. 322**
Yunnan lettuce (serving)	**Mosaic and leafy**	**37. 977**
Yunnan oily wheat vegetable (serving)	**Mosaic and leafy**	**25. 038**
Bamboo leafy vegetables	**Mosaic and leafy**	**13. 296**
Baby cabbage	**Mosaic and leafy**	**10. 428**
Shanghai green	**Mosaic and leafy**	**3. 861**
Milk cabbage	**Mosaic and leafy**	**6. 422**
amaranth	**Mosaic and leafy**	**8. 923**
Fungus vegetables	**Mosaic and leafy**	**5. 933**
Green Vegetables (1)	**Mosaic and leafy**	**4. 900**
Sweet potato tip	**Mosaic and leafy**	**4. 506**
Millet pepper (serving)	**Chili peppers**	**23. 795**
Screw pepper	**Chili peppers**	**7. 598**
Ginger, garlic and millet pepper combination (small portion)	**Chili peppers**	**7.000**
Small wrinkles (servings)	**Chili peppers**	**11. 285**
Red Pepper (2)	**Chili peppers**	**2. 035**
Purple Eggplant (2)	**Nightshades**	**11. 790**
Long-line eggplant	**Nightshades**	**4. 557**
Green Eggplant (1)	**Nightshades**	**2. 675**
West Gorge Mushroom(1)	**edible fungi**	**5. 145**
Bisporus mushroom (box)	**edible fungi**	**11. 128**
Seafood mushrooms (pack)	**edible fungi**	**9. 856**
Enoki mushroom (box)	**edible fungi**	**17. 964**
Pure Lotus Root(1)	**Aquatic rhizomes**	**7. 642**
Honghu lotus root	**Aquatic rhizomes**	**5. 114**
water caltrop	**Aquatic rhizomes**	**1. 746**
Tall melon (1)	**Aquatic rhizomes**	**3. 004**
Wild powdered lotus root	**Aquatic rhizomes**	**0. 348**

7.2. Pricing strategy

The replenishment quantity strategy has been modelled and successfully solved, and the replenishment quantity of a single product on 1 July is shown in Table 5. Since the correlation analysis and linear fitting of the functional relationship between total sales and cost pricing in Problem 2 are basically consistent with the actual situation, the pricing strategy can be solved directly using the data in Table 5, and the formula is as follows

$$p = -0.1300m_t + 14.408.$$

Where p denotes the selling price of each category and m_t denotes the maximum value of category sales in the tth July of the previous year.

The result of the solution is shown in the following table

Table 6. Pricing strategy per unit as of July 1

Singles	category	Pricing
broccoli	Cauliflower	5. 50
Zhi jiang green stalk scattered flowers	Cauliflower	4. 66
Yunnan lettuce (serving)	Mosaic and leafy	1. 06
Yunnan oily wheat vegetable (serving)	Mosaic and leafy	1. 54
Bamboo leafy vegetables	Mosaic and leafy	1. 43
Baby cabbage	Mosaic and leafy	1. 78
Shanghai green	Mosaic and leafy	3. 86
Milk cabbage	Mosaic and leafy	2. 25
amaranth	Mosaic and leafy	1. 49
Fungus vegetables	Mosaic and leafy	2. 19
Green Vegetables (1)	Mosaic and leafy	2. 37
Sweet potato tip	Mosaic and leafy	2. 14
Millet pepper (serving)	Chili peppers	4. 34
Screw pepper	Chili peppers	4. 60
Ginger, garlic and millet pepper combination (small portion)	Chili peppers	2. 26
Small wrinkles (servings)	Chili peppers	1. 06
Red Pepper (2)	Chili peppers	5. 82
Purple Eggplant (2)	Nightshades	2. 72
Long-line eggplant	Nightshades	6. 02
Green Eggplant (1)	Nightshades	1. 91
West Gorge Mushrooms(1)	edible fungi	10. 08
Bisporus mushroom (box)	edible fungi	1. 95
Seafood mushrooms (pack)	edible fungi	0. 95
Enoki mushroom (box)	edible fungi	0. 51
Pure Lotus Root(1)	Aquatic rhizomes	4. 30
Hong hu lotus root	Aquatic rhizomes	3. 62
water caltrop	Aquatic rhizomes	4. 83
Tall melon (1)	Aquatic rhizomes	1. 70
Wild powdered lotus root	Aquatic rhizomes	9. 93

8. Strengths, weaknesses and improvements of the model

8.1. Strengths of the model

(1) In question 1, Pearson's correlation is used to analyze the correlation between each vegetable category on a quarterly basis and at each time of the day, and under appropriate conditions, statistical tests can be used to verify and interpret the significance of the correlation, and to analyze the linear relationship between the variables with a higher degree of precision.

(2) The relationship between pricing and sales volume in question two, based on the basis of correlation analysis, we used linear fitting to obtain the quantitative analysis of pricing and sales volume of the relationship between the formula, and as a constraint to answer the second question of question two, to obtain the final answer to question two.

(3) Solving problem 3 uses a greedy algorithm in the optimal establishment of the replenishment scheme model, which can locally optimize multiple objective functions, is usually faster in execution, performs better in terms of spatial efficiency, finds the optimal solution in each local space and obtains the optimal scheme in the overall space, and the scheme performs well in satisfying certain ratio conditions.

8.2. Weaknesses of the model

(1) Grey correlation analysis between the indicators of the independent variables was not performed in the first question, and the independent variables may have uncertainty or ambiguity, which may

have the undesirable effects of increased complexity of the model, difficulty in solving the problem, decreased accuracy, difficulty in interpreting and applying the model, data requirements, reliability problems, increased computational costs, and decreased certainty and usefulness of the objective function.

(2) A greedy algorithm was used to solve the objective function in Problem 3. The greedy algorithm considers only the current optimal solution at a time, the greedy algorithm is more problem dependent and in the case of Problem 3, the greedy algorithm may lead to sub-optimal solutions and incorrect results.

8.3. Improvements to the model

The quantitative supply characteristics in the data processing are crude. It is possible to broaden the idea by introducing more decision variables, such as assessing the service level of the supplier or the quality of the product supplied. As a result, the method of determining the weights can be improved.

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] Li Yuan. Research on Joint Optimization of Pricing and Inventory for Dual-Channel Retailers Considering Reference Price Effect [D]. Yanshan University, 2022.

[3] Ting Wang, Lei Zhu, Yuelei Zhang et al. Econometric analysis of photo pricing problem based on greedy algorithm [J]. Journal of Chifeng College (Natural Science Edition), 2018, 34(07): 16-19.

[4] Yang Haoxu. Research on the correlation analysis of vegetable price and sales volume--Taking the example of oleaginous vegetables[J]. Food Safety Guide, 2018(21): 179-181.

[5] Li Xuan. Application of data processing methods in mathematical modelling competitions [J]. Digital Technology and Application, 2022, 40(11): 61-63.

[6] Liu JC, Li F, Wang HH et al. A research review on evolutionary high-dimensional multi-objective optimization algorithms [J]. Control and Decision Making, 2018, 33(05):879-887...JOURNAL OF RETAILING AND CONSUMER SERVICES [J]. Journal of retailing and consumer services, 2008, 15(1).

[7] Marta C, Cristina M, Valerio A, et al. Efficient and accurate inference for mixtures of Mallows models with Pearson distance [J]. Statistics and Computing, 2023, 33(5).

[8] Angel. Research on Pricing of Science and Technology Novelty Search Service Based on Cost-Plus Pricing Method [J]. Library Research and Work, 2021(10): 25-31.

[9] An Qi. Research on the pricing of scientific and technical research services based on cost-plus pricing method [J]. Library Research and Work, 2021(10): 25-31.

[10] Hao Yannan. Research on supply chain coordination of fresh food e-commerce considering preservation effort and promotion effort [D]. Yunnan University of Finance and Economics, 2022.

[11] Yang Shuai, Huang Xiangmeng, Wang Junbin. Research on joint optimization strategy of shelf allocation and pricing for fresh food [J]. Supply Chain Management, 2022, 3(08):49-59.