Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 10, No. 2, 2024

Research on Crop Planting based on Linear Programming and Multiple Regression Functions

Bo Zhai *, Hanxue Zhu, Hongzhuang Wan

School of Computer Engineering, Tongda College, Nanjing University of Posts and Telecommunications, Yangzhou, Jiangsu, China
*** Corresponding author:** Bo Zhai (Email: 2292476389@qq.com)

Abstract: The purpose of this paper is to explore how to optimize crop planting strategies through linear programming and multiple regression functions to help rural revitalization and agricultural modernization. In the study, we consider a variety of factors such as crop planting seasons, plot types, and potential planting risks. By analyzing the crop suitability of different plots and combining constraints such as arable land area limitations and crop rotation requirements, we established a linear programming model with the objective of maximizing economic benefits. At the same time, for the two cases of stagnant and wasteful marketing and selling at a reduced price when the yield exceeds the expectation, binary decision variables are used to find the optimal solution. In addition, this paper uses multiple regression functions to predict the growth rate of sales of major crops such as corn and wheat and predicts the future price growth of edible mushrooms through a gray prediction model (GM (1,1)). In the course of the study, we also discuss the impact of extreme weather due to monsoon climate in the northern region and give the optimal planting scheme based on these perturbation factors. Through gray correlation analysis, this paper further explores the correlation coefficients between planting cost and selling price relative to expected sales volume and combines the principles of complementarity and substitutability to derive the highest yielding planting strategy. Finally, the stability of the model is verified by sensitivity analysis, and its adaptability in other agricultural scenarios is explored.

Keywords: Optimal Planting Strategy; Linear Programming; Gray Correlation Analysis; GM (1,1) Prediction; Multiple Linear Regression.

1. Introduction

In the context of global climate change and market volatility, the rural revitalization strategy [1] has become an important initiative to promote agricultural modernization and enhance the rural economy. The aim of this paper is to study the optimal planting strategy [2] for a rural village in North China to improve the utilization rate of arable land and production efficiency, and to ensure the stability of food production. To this end, we comprehensively use a variety of computer-related algorithms and models, including linear programming [3], gray prediction model [4] and multivariate function fitting [5], etc., in an effort to find a scientific and reasonable solution in a complex agricultural environment.

This paper analyzes crop cultivation options for 34 arable plots in the countryside through a linear programming model to maximize economic benefits, setting constraints such as arable land area, heavy cropping restrictions and crop rotation. Multivariate function fitting and gray prediction model are used to predict the sales volume and price in the next seven years and adjust the planting strategy to cope with the market risk. At the same time, the substitutability and complementarity between crops are explored [6], and the relationship between sales volume, price and cost is studied through gray correlation analysis to provide theoretical support and data basis for rural agricultural development and policy formulation.

2. Optimal Cropping Programs for Crops in the Event of Slow Sales and Price Reductions

The optimal cropping program for crops in the village from 2024 to 2030 is given for two scenarios, respectively, when the total production of crops per season exceeds the corresponding expected sales volume of stagnant sales and sales at a reduced price. The solution is analyzed according to the following steps.

2.1. Situation Analysis of Available Cropland in Villages

2.1.1. Matching Crops to Plots and Seasons

There are mainly six different types of cultivated plots in this countryside, namely, flat dry land, terraced land, hillside land, watered land, 16 ordinary greenhouses and 4 intelligent greenhouses, of which flat dry land, terraced land and hillside land can only grow one season of grain crops per year, watered land can grow one season of rice or two seasons of vegetables per year, ordinary greenhouses can grow one season of vegetables and one season of edible fungi per year, and intelligent greenhouses can grow two seasons of vegetables per year. Each cultivated land type corresponds to the plot crop and season matching specific information.

2.1.2. Outlier Test

With regard to the statistical data related to crop cultivation and sales indicators, taking into account the fact that there are several kinds of crop indicator data that seem to be “abnormal”, but in real life, different types of crops cultivation and sales indicators may indeed have greater differences.

Box plot can visualize the distribution of data, through the median, quartiles and other statistics to show the distribution of data, including data concentration trends, dispersion and outliers. The box plot sorts the data set from smallest to largest, and selects the data values in the 25th and 75th percentile as the first and third quartiles of the box, and the number in the 50th percentile as the median, and if the median is located in the center of the box, the median is the first and third quartiles.

The length of the box, the maximum value outside the box; the minimum value is, and the values located outside the maximum and minimum values are called outliers. Accordingly, we plot the box-and-line diagrams of the three types of crops under three indicators of planting cost, acreage yield, and selling unit price as shown in Figure 1.

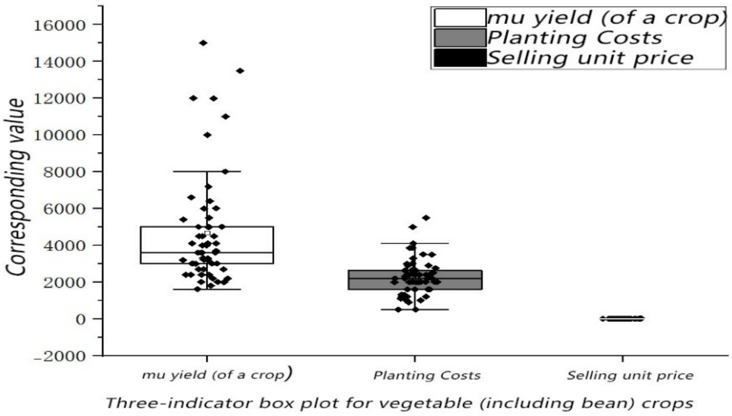

Figure 1. Vegetable Crops Triple Indicator Box Char

And the sales unit price is not a component data, so it is not possible to use the CLR logarithmic transformation. The median values corresponding to grain, vegetable and edible mushroom are 6.5, 7.2 and 42.5, respectively, 8.375, 8 and 78.75, respectively, and 4, 6 and 19.5, respectively, by sorting the unit sales price of each crop under the categories of grain, vegetable and edible mushroom by quartile order in Annex 2. Therefore, the boxes corresponding to the three categories are 4.375, 2 and 69.25, respectively; and the corresponding great values are 14.938, 11, and 182.625, respectively; and the corresponding very small values are 0, 3, and 0, respectively. Therefore, for the unit price of sales, there is no crop that produces outliers, and there is no need to analyze it separately.

2.2. Linear Programming based on Optimal Crop Planting Schemes

In this section, we need to study the optimal crop planting program during the period from 2024 to 2030, and consider that the planting strategy with the highest annual economic benefit is the optimal planting strategy for the year under the conditions of complying with the area of arable land, the requirement of crop rotation, and the convenience of cultivation operation and field management.

2.2.1. Uncertain Parameters

(1) Expected sales volume

Expected sales volume refers to the market demand for the crop, as the crops planted in each season are only sold in the current season, and the production of crops planted in each season cannot all be guaranteed to be marketable, if the production in the current season exceeds the demand in the current season, i.e., there is oversupply, which can lead to stagnation (not sold at all) or the need to sell at a reduced price (sold at 50% of the sales price).

Let the expected sales volume be $S_{jk},(j=1,2,...,41;k=0,1,2)$, that is, the total production of the crop numbered j in k seasons in 2023, where $k=0,1,2$ represents a single season, the first season and the second season, respectively, is a categorical variable, the matching relationship between the crop seasons is shown in Fig. 1, $x_{ijk2023}$ is the actual acreage of the crop j in the k seasons in 2023 in the plot i, $Y_{ijk2023}$ is the acreage of the crop j in the k seasons in 2023 in the plot i, so the total production of all seasons in 2023 is Therefore, the total production of each season in 2023 is:

$$S_{jk2023}=\sum_{i=1}^{54}x_{ijk2023}\times Y_{ijk2023} \tag{1}$$

Solve for it to obtain the total production of the crop in 2023 for each season (i.e., the expected sales volume of the crop for each season)

(2) Unit price

The economic efficiency of the crop is directly affected by the sales unit price, the range of sales unit price of different seasons of each crop is obviously only related to the crop and sales season, let the sales unit price in 2023 is, then

$$P_{jk2023}=P_{jk2023\min}+\rho\frac{P_{jk2023\max}-P_{jk2023\min}}{10},\rho\in[0,10] \tag{2}$$

Where ρ is the perturbation factor, P_{jk2023} represents the unit sales price of crop j in 2023 k season, Tables $\max$ and $\min$ below show the maximum and minimum values of the unit sales price, respectively. respectively, the formula can more accurately represent the range of fluctuations in the unit sales price. In this paper, for the sake of simplifying the model, we take $\rho=5$, which is used for the calculation of the objective function in the following. This method can better balance the possibility of high price and low price without more information.

2.2.2. Construction of the Objective Function

Since the crops grown in each season are sold only in the current season, the annual economic benefit should be the sum of the net returns from a single-season crop and a two-season crop. Let the countryside in t year economic return is Q_t, where t indicates the year; x_{ijkt} is the decision variable, indicating t year when the k seasonal crop j in i plots of planting area; $Y_{ijk2023}$, $C_{ijk2023}$ are known parameters, respectively, indicate the seasonal crop j in i plots of mu yield and planting costs.

(1) The portion of the crop that exceeds the expected sales volume is not sold, resulting in wastage. At this time, if the total production of each season of the crop is greater than the expected sales volume, the excess part cannot get revenue, sales volume to take the expected sales volume; if the total production of each season of the crop is less than the expected sales volume, the production of each season can be sold, sales volume to take the actual production, as shown in Equation (3):

$$\min\left(\sum_{i=1}^{54} x_{ijkt}\times Y_{ijk2023}, S_{jk}\right) \tag{3}$$

Therefore, the objective function is:

$$Q_t=\sum_{k=0}^{2}\left[\sum_{j=1}^{41}\min\left(\sum_{i=1}^{54} x_{ijkt}\times Y_{ijk2023}, S_{jk}\right)\times P_{jk2023}-\sum_{i=1}^{54} x_{ijkt}\times C_{ijk2023}\right] \tag{4}$$

(2) The excess is sold at a reduced price of 50 percent of the 2023 sales price. Based on the first case, the cost of planting remains the same and the sales proceeds will increase by the portion of the proceeds that exceeds the expected sales volume. The difference between the actual crop yield and the expected sales volume is made, and then compared with 0 to determine: whether the expected sales volume is exceeded or the yield in excess of the expected sales volume. This can be expressed as:

$$\max\left(\sum_{i=1}^{54} x_{ijkt}\times Y_{ijk2023}-S_{jk}, 0\right) \tag{5}$$

Therefore, the objective function is:

$$\begin{aligned}Q_t=&\sum_{k=0}^{2}\sum_{j=1}^{41}\min\left(\sum_{i=1}^{54} x_{ijkt}\times Y_{ijk2023}, S_{jk}\right)\times P_{jk2023}\\&+\max\left(\sum_{i=1}^{54} x_{ijkt}\times Y_{ijk2023}-S_{jk}, 0\right)\times\frac{1}{2}P_{jk2023}-\sum_{i=1}^{54} x_{ijkt}\times C_{ijk2023}\end{aligned} \tag{6}$$

2.2.3. Construction of Constraints

The constraints are as follows

(1) Area control of individual cropland plots

(2) Crop and plot and season matching

(3) Crop heavy cropping restrictions

(4) Legume rotation requirement

(5) Linkage of area to binary variables

For case I, i.e., the crop production exceeds the expected production sales lag, there is an objective function with constraints as:

$$\max\ Q_t=\sum_{j=1}^{41}\sum_{k=0}^{2}\min\left(\sum_{i=1}^{54} x_{ijkt}\times Y_{ijkt}, S_{jk}\right)\times P_{ijkt}-\sum_{i=1}^{54} x_{ijkt}\times C_{ijkt} \tag{7}$$

$$s.t.\begin{cases}\lambda D_i\leqslant\sum_{j=1}^{41}\sum_{k=0}^{2} x_{ijkt}\leqslant D_i\\ z_{ij0t}+z_{ij0(t+1)}\leqslant 1\\ z_{ij1t}+z_{ij2t}\leqslant 1\\ \sum_{t'=t}^{t+2}\sum_{j} z_{ijkt'}\geqslant 1, j\in\{1,2,3,4,5,17,18,19\}\\ x_{ijkt}\leqslant M\times z_{ijkt}\end{cases} \tag{8}$$

For case II, i.e., the case where the crop production exceeds the expected production volume sold at a reduced price, there is an objective function and constraints as:

$$\begin{aligned}Q_t=&\sum_{k=0}^{2}\sum_{j=1}^{41}\min\left(\sum_{i=1}^{54} x_{ijkt}\times Y_{ijk2023}, S_{jk}\right)\times P_{jk2023}\\&+\max\left(\sum_{i=1}^{54} x_{ijkt}\times Y_{ijk2023}-S_{jk}, 0\right)\times\frac{1}{2}P_{jk2023}-\sum_{i=1}^{54} x_{ijkt}\times C_{ijk2023}\end{aligned} \tag{9}$$

$$s.t.\begin{cases}\lambda D_i \leqslant \sum_{j=1}^{41}\sum_{k=0}^{2} x_{ijkt} \leqslant D_i \\ z_{ij0t} + z_{ij0(t+1)} \leqslant 1 \\ z_{ij1t} + z_{ij2t} \leqslant 1 \\ \sum_{t'=t}^{t+2}\sum_{j} z_{ijkt'} \geqslant 1, j \in \{1,2,3,4,5,17,18,19\} \\ x_{ijkt} \leqslant M \times z_{ijkt}\end{cases} \tag{10}$$

For the stagnant waste, selling at a reduced price is more economically efficient, but the limitation of taking the production volume in 2023 as the expected sales volume in the next few years is that the economic efficiency under this condition is even higher than that in 2023, which is due to the fact that in real life, due to the existence of weather and climatic factors, natural disasters, market and economic regulation, and governmental policies, etc., so that the crop planting and sales There is a great deal of uncertainty in the indicators, and there are a variety of potential planting risks, which may have an extremely important impact on the economic benefits.

3. Crop Cultivation Programs under Variations and Risks of Various Factors

The various crop planting and marketing indicators have great uncertainty and other potential planting risks are added into consideration to re-give the optimal planting program for crops in the coming years.

3.1. Consideration and Analysis of Uncertainties

3.1.1. Simplify the Uncertainty

(1) Wheat and corn expected sales volume growth trend prediction

Check the relevant information, the average grain production in each region relative to the national share as shown in Figure 2, of which North China accounted for about 13.7%.

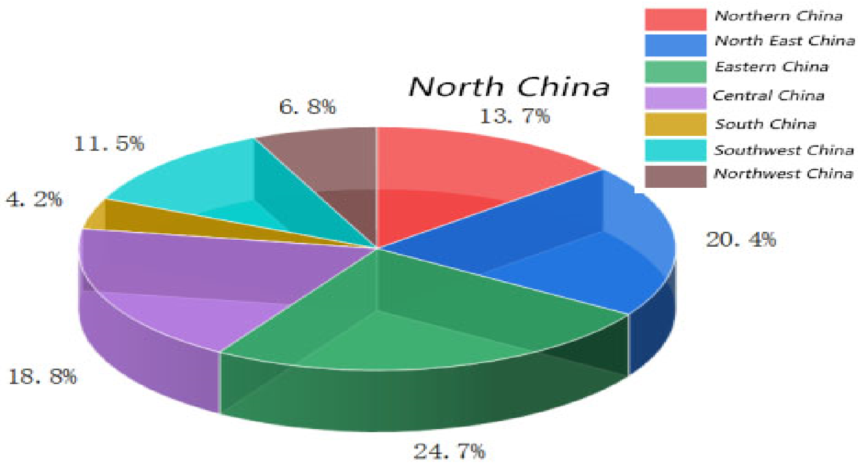

Figure 2. Grain output share of each region

This paper assumes that all the output of wheat and corn can be normally sold, then the total output of wheat and corn in North China in the past ten years is its sales volume.

Goodness of fit Used to measure how well the model fits the observed data, in short, it indicates the extent to which the model effectively describes or predicts the actual data and indicates the proportion of variation in the dependent variable that is explained by the independent variables, which is calculated by the formula:

$$R^2 = 1 - \frac{TSS - RSS}{TSS} \tag{11}$$

where TSS and RSS denote the total sum of squares and regression sum of squares, respectively, from which the various curve fits were calculated R^2.

The third fit has nonlinear existence, which is rounded off, and secondly when the fitted curve is a power function, there is a maximum of R^2, indicating that the power function fit is the best, so the power function curve fit was carried out for wheat and maize, and the results of the fit were, respectively:

$$y = 1668.429x^{0.049} \tag{12}$$

$$y = 3210.516x^{0.061} \tag{13}$$

According to this fitting result, the expected sales volume of wheat and corn in 2023-2030 are calculated respectively, and at the same time, the average annual growth rate of the expected sales volume of wheat and corn in the next seven years are calculated according to this table respectively.

(2) Gray prediction GM (1, 1) model

For the sales unit price of edible mushrooms, it will decrease by 1%-5% every year. By reviewing the relevant data, the sales unit price (RMB/kg) of shiitake and white mushroom from 2015 to 2021 in North China, represented by Beijing, was obtained, as detailed in the supporting documents.

Considering the small amount of data, the gray prediction GM (1, 1) model was used to predict the sales unit price from 2023 to 2030.

The gray prediction method does not require much data, as long as the amount of data is greater than or equal to 4, it can be predicted. It is of great importance in many practical problems, especially in those areas where it is difficult to obtain sufficient data.

3.2. Linear Programming based on Planting Scheme

3.2.1. Introduction of Unexpected Factors

Consider the extreme weather in North China, North China is a warm temperate semi-humid continental monsoon climate with four distinct seasons, low precipitation, hot and rainy summers, cold and dry winters, and short springs and autumns, and the unique geographic location and climatic characteristics are the cause of the frequent occurrence of cold waves, high temperatures, and other extreme weather events in the North China region.

3.2.2. Comprehensive Linear Planning

Introducing several impact parameters, and individual impact parameters have been specific values through specific prediction methods, so this question here only give specific constraints on the impact parameters, to the mushroom as an example shown below:

$$P_{j^{0},k2023} \equiv 0(j=1,5,10,12,16,18,19,25,27,28,30,32,33,38,40) \tag{14}$$

where u indicates the year, 2023 in $u=0$ cases and 2030 in $u=7$ cases. For all crops the cost of cultivation increases by 5% per year, so there is the recursive equation.

For shiitake mushroom No. 39, the growth rate of sales price was 2.21%, 2.22%, -2.11%, -4.39%, 7.10%, 4.52%, 4.39% and -13.61%.

$$P_{39,k,u+1} = P_{39,k,u} \times R_u (u=0,1,2,...7) \tag{15}$$

$$C_(i,j,k,u+1) = C(i,j,k,u) \times (1+5\%),(u=0,1,2,...7) \tag{16}$$

4. Substitutability and Complementarity among Crops

4.1. Correlation Analysis based on Gray Correlation Analysis

Gray correlation analysis is a multi-factor statistical analysis method, based on gray coefficient theory, mainly used to analyze the degree of correlation between the factors in the system, applicable to the case of small sample data, incomplete information or system uncertainty.

This paper analyzes the correlation between the expected crop sales volume and sales price and planting cost based on gray correlation analysis.

Step1 Determine the sequence to be compared with the test sequence

In the selection of the mother sequence and the two sequences to be compared, in order to quickly explore the correlation between each agricultural product, it is considered that the numbering sequence of the three sequences is controlled in the same order, and the incremental change of the numbering of crops is taken as the potential independent variable, and at the same time the numbering of the two sequences is also subjected to correlation due to the type of crops, so that the gray correlation coefficient of the depicted labels indicate the correlation between the expected selling quantity and the selling price and the planting cost of each type of agricultural product. The correlation between from the meaning of the question, the parent series is selected as the expected sales volume of each agricultural product, and the two-comparison series are the corresponding sales price and planting cost of each crop.

Step2 Preprocessing of variable data

Z-score standardization can transform the data into a standard normal distribution with a mean of 0 and a standard deviation of 1, and the process is not affected by whether the data set contains negative numbers or not, and it is also applicable to the case where the indicators do not have theoretical extreme values, and the specific formula is:

$$\hat{Z} = \frac{X-\mu}{\sigma} \tag{17}$$

where $\hat{Z}$ is the normalized dataset, $\hat{Z}_{ij}$ denotes the value of each element in the normalized dataset, X is the value of the data point, μ is the mean of the dataset, and σ is the standard deviation of the dataset. The dataset of the parent series and the two-comparison series normalized by *Z-score* is derived from Eq.

Step3 Substitution of gray correlation algorithm

In this problem, the planting cost and selling price are selected to have equal weights, i.e., they have equal importance, and the gray correlation coefficient is:

$$\xi_i(k) = \frac{\min_s \min_t |x_0(t) - x_s(t)| + \rho \max_s \max_t |x_0(t) - x_s(t)|}{|x_0(k) - x_i(k)| + \rho \max\max_s |x_0(t) - x_s(t)|} \tag{18}$$

Where, $\rho[0,1]$ is the discrimination coefficient, which is taken as 0.5 in this question, the correlation coefficient of each indicator is expressed as:

$$y(x_0(k),x_i(k)) = \frac{a+\rho b}{|x_0(k)-x_i(k)|+\rho b}, i=1,2,...,m; k=1,2,...,n \quad (19)$$

where $a=\min_i \min_k |x_0(k)-x_i(k)|$ is the two very small differences and $b=\max_i \max_k |x_0(k)-x_i(k)|$ is the two very large differences.

Then the gray weighted correlation is calculated:

$$r_i = \sum_{k=1}^{n} \omega_i \xi_i(k) \quad (20)$$

Where r_i represents the gray-weighted correlation of agricultural products corresponding to different numbers on the ideal object.

Step4 Solve the gray correlation coefficient

In this paper, the gray correlation coefficient table is obtained through the SPSSPRO calculation (the length is too long to show, see the appendix for details), and the correlation coefficient graph is shown in Figure 3.

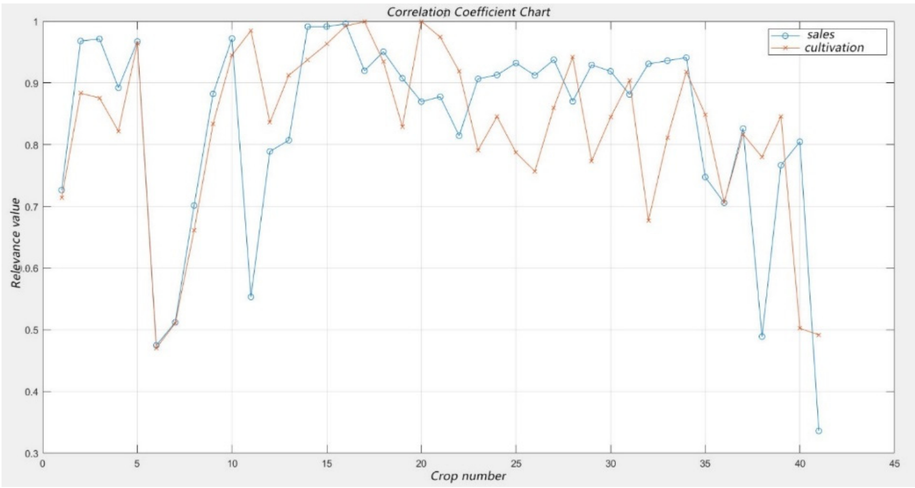

Figure 3. Correlation coefficient graph

The two curves in this graph are the gray correlation between the expected sales volume of each agricultural product and its corresponding sales price and planting cost. Using the expected sales volume as the “reference value” (parent series), the correlation between the two evaluation items (sales price and planting cost) and the expected sales volume is examined.

The correlation value of the vertical coordinate is between 0 and 1, and the larger the value is, the stronger the correlation between it and the “reference value” (parent series), which means that it is more highly evaluated. As a result, the correlation between the unit sales price and the cost of cultivation and its sales volume were 0.835 and 0.826, respectively.

4.1.1. Complementarity and Substitutability Analysis of Crops

Consider crop type, planting season, plot type, cost of planting, and acreage between each crop, where acreage and cost of planting are the difference conditions, and the threshold is set at 20%. The plot type and planting season are the matching conditions, and crops that belong to the same plot type and have the same planting season must fulfill the prerequisites before they can be a replacement group.

4.1.2. Optimal Planting Program of Crops and the Relationship between Crop Indicators

Through the gray correlation analysis, it is concluded that the correlation ratio of sales price and planting cost for the expected sales volume is 0.502:0. 498.It can be approximated that it has the same correlation, i.e., it is considered to have equal weights.

Secondly, through the comprehensive examination of substitutability and complementarity, the following planting strategy is given: After considering the impact due to climatic factors based on the second question, in the screening of plan table agricultural products, the main focus is on the pairing of agricultural products that can complement each other.

5. Conclusion

This paper discusses the optimization of rural crop planting strategies through regression functions and other factors, aiming to improve yields and economic benefits and provide solutions for rural revitalization. The study used linear programming, multiple regression function and gray prediction model (GM (1,1)), combined with plot type, climate and other factors, to construct a planting model for maximizing economic benefits. Linear programming was used to analyze the planting optimization scheme under a stable market, and set restrictions on cultivated land area and crop rotation. The multiple regression function predicts crop sales growth, and the gray prediction model predicts changes in the price of edible mushrooms and adjusts planting strategies to cope with market risks. The study verifies its feasibility and potential for replication in real agricultural scenarios, which helps to achieve the strategic goal of rural revitalization.

References

[1] Jin Guanping. Doing enough special articles to promote rural revitalization [N]. Economic Daily, 2024-10-21(001).

[2] Song Xuefen. Research on high-yield and high-efficiency planting strategy of crops in black soil test area of Mengcheng sand ginger[J]. Anhui Agricultural Science,1995, (03):229-231.

[3] GUO Jidong, XU Cuiling, LI Kaijie,et al. Optimization and simulation of customized furniture oil milling production line efficiency based on linear programming[J]. Electromechanical Engineering Technology,2024,53(09):81-89.

[4] Zhang Xiaochun. Research on the forecasting of gross regional product of Guangdong Province based on gray GM (1,1) model[J]. Value Engineering,2024,43(29):1-3.

[5] DONG Hui. Segmentation of defects in X-ray weld images based on binary function fitting[J]. Welding Technology, 2023,52(07): 18-22.DOI: 10.13846/j.cnki.cn12-1070/tg. 2023. 07.007.

[6] Cui Guohu, Xu Yujun. Analysis of the Irreplaceability and Complementarity of Accounting Supervision and Audit Supervision[J]. Taxation,2019,13(15):136.