

Research on the Influence Mechanism and Optimization Path of Market Dietary Structure on the Health of Rural Residents

Rongjie Fan¹, Han Zhou^{2,*}, Ziqian Lou³

¹ School of Artificial intelligence and Big Data, Henan University of Technology, Zhengzhou, China

² School of Mathematics and Statistics, Henan University of Technology, Zhengzhou, China

³ School of Physics, Henan University of Technology, Zhengzhou, China

* Corresponding author: Han Zhou (Email: 18159778166@163.com)

Abstract: With the development of global economic integration and the diversification of the food market, the dietary structure of rural residents is undergoing profound changes, and their health status has also received widespread attention. This study aims to explore the influence mechanism of market dietary structure on the health of rural residents and propose optimization paths. Based on the national rural tracking survey data of the China Rural Revitalization Association from 2022 to 2024, this study uses multiple linear regression models and Logistic regression models to analyze the influence mechanism of market dietary structure on the health of rural residents. The results show that meat consumption is significantly positively correlated with the BMI of rural residents, vegetable consumption is significantly negatively correlated with BMI and can reduce the probability of chronic diseases, age growth will cause a slight increase in BMI, and an increase in BMI will increase the risk of chronic diseases. In addition, rural residents have a high acceptance of animal-derived certified foods, and staple foods and protein expenditures account for a large proportion in the dietary structure. This study provides a theoretical basis and practical guidance for improving the nutritional health status of rural residents and promoting the healthy development of rural areas.

Keywords: Market Dietary Structure; Health of Rural Residents; Multiple Linear Regression; Logistic Regression.

1. Introduction

As a key factor measuring the nutritional and health level of residents, the dietary structure profoundly affects an individual's physiological functions, disease susceptibility, and overall quality of life. A reasonable dietary structure not only provides the human body with sufficient energy and various nutrients to maintain normal metabolism and functions of the body but also plays an irreplaceable role in preventing chronic diseases and promoting healthy longevity [1]. With the development of global economic integration and the diversification of the food market, the dietary structure of residents is undergoing profound changes, and this change is particularly significant in rural areas [2].

Rural residents are an important part of China's population, and their health is related to the national overall health level and the implementation effect of the rural revitalization strategy [3]. In recent years, with the rapid growth of the rural economy, the gradual improvement of the market circulation system, and the change of lifestyle, the dietary structure of rural residents has changed greatly. The traditional dietary pattern dominated by plant-based foods such as grains and vegetables are gradually shifting to one rich in more animal foods, processed foods, and excessive intake of oil, salt, and sugar. At the same time, the health problems of rural residents have become increasingly prominent [4]. The prevalence rates of chronic diseases such as overweight, obesity, cardiovascular diseases, and diabetes are increasing, and the phenomenon of nutritional imbalance such as micronutrient deficiency is still common. These health problems have seriously restricted the improvement of the quality of life of rural residents and the sustainable development of rural

society.

A large number of studies have shown that there is a close and complex relationship between dietary structure and residents' health. Unreasonable dietary structure has been confirmed to be an important risk factor leading to the occurrence and development of various chronic diseases [5]. For rural residents with relatively scarce medical and health resources and relatively low health awareness and knowledge level, this impact may be more far-reaching [6]. In-depth exploration of the influence mechanism of market dietary structure on the health of rural residents is of great theoretical and practical significance for accurately formulating strategies to improve the nutritional health status of rural residents and improving the level of rural medical and health services. However, at present, for the specific group of rural residents, the achievements of systematic research on the relationship between market dietary structure and health are relatively limited, and there are still many gaps to be filled especially in the influence mechanism and optimization path.

In view of this, this study aims to analyze the influence mechanism of market dietary structure on the health of rural residents and put forward feasible optimization paths, in order to provide a theoretical basis and practical guidance for improving the nutritional health status of rural residents and promoting the healthy development of rural areas.

2. Research Methods and Model Construction

2.1. Data Acquisition

The data of this study are from the national rural tracking survey database of the China Rural Revitalization Association

from 2022 to 2024. The data cover 3,500 rural families in 50 counties (cities) of 12 provinces including Henan, Shandong, and Guangdong, and use a multi-stage stratified sampling method to take into account rural areas of different geographical types such as plains, mountains, and suburban areas. The data collection methods include household questionnaire surveys (accounting for 70%), market observation records (accounting for 15%), and focus group interviews (accounting for 15%), and quote the national rural tracking survey data of the association in the same period, including rural residents' food consumption logs, family income and expenditure details, and health examination files. All data have undergone three logical verifications and reliability and validity tests, and the effective recovery rate is 98.2%.

2.2. Multiple Linear Regression Model

Multiple linear regression aims to reveal the linear relationship between multiple independent variables and a continuous dependent variable. In this study, it is assumed that the health status (BMI) can be approximately represented by the linear combination of five variables: meat, vegetable, and dairy consumption, age, and exercise.

2.2.1. Model Setting

Let the dependent variable Y_{BMI} be the health status (BMI) of rural residents, the independent variable be meat, vegetable, and dairy consumption, age, and exercise. The general form of the model is:

$$Y_{\text{BMI}} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon \quad (1)$$

Among them, β_0 is the intercept term, representing the expected value of Y_{BMI} when all independent variables take a value of 0; $\beta_i (i = 1, 2, \dots, 5)$ are the regression coefficients, measuring the impact degree of each independent variable on the dependent variable; ε is the random error term, which follows a normal distribution with a mean of 0 and a variance of σ^2 , that is, $\varepsilon \sim N(0, \sigma^2)$.

2.2.2. Matrix Representation

Represent the above - mentioned model in matrix form. Let Y be an $n \times 1$ dependent - variable vector, X be an $n \times k$ independent - variable matrix, β be a $k \times 1$ regression - coefficient vector, and ε be an $n \times 1$ error vector. Then the model can be written as:

$$Y_{\text{BMI}} = X\beta + \varepsilon \quad (2)$$

Among them,

$$Y_{\text{BMI}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_{11} & X_{21} & X_{31} & X_{41} & X_{51} \\ 1 & X_{12} & X_{22} & X_{32} & X_{42} & X_{52} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1n} & X_{2n} & X_{3n} & X_{4n} & X_{5n} \end{bmatrix},$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

2.2.3. Parameter Estimation

The goal of the ordinary least - squares method is to find a set of regression coefficients β such that the sum of squared residuals

$$S(\beta) = \sum_{i=1}^n \varepsilon_i^2 = (Y_{\text{BMI}} - X\beta)^T (Y_{\text{BMI}} - X\beta) \quad (3)$$

Take the partial derivative of $S(\beta)$ with respect to β and set it equal to 0, we can get:

$$\frac{\partial S(\beta)}{\partial \beta} = -2X^T (Y_{\text{BMI}} - X\beta) = 0 \quad (4)$$

After rearrangement, we obtain the normal equation system:

$$X^T X \beta = X^T Y_{\text{BMI}} \quad (5)$$

If $X^T X$ is invertible, then the OLS estimated value of the regression coefficient is:

$$\beta = (X^T X)^{-1} X^T Y_{\text{BMI}} \quad (6)$$

2.3. Logistic Regression Model

Logistic regression is suitable for the case where the dependent variable is a binary variable, and maps the independent variables of the linear combination to the probability space through the logistic function. In this study, the dependent variable is whether rural residents suffer from chronic diseases (sick = 1, not sick = 0), where X_1 represents meat food consumption, X_2 represents vegetable food consumption, X_3 represents dairy product consumption, X_4 represents age, X_5 represents exercise, and Y_{BMI} is the health status of rural residents. In order to explore the factors affecting these health indicators, we establish a Logistic regression model.

2.3.1. Model setting

Let $P(Y = 1)$ represent the probability that rural residents suffer from chronic diseases, and $P(Y = 0) = 1 - P(Y = 1)$ represent the probability that they do not suffer from chronic diseases. The logistic regression model first constructs a linear combination:

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 Y_{\text{BMI}} \quad (7)$$

Then convert Z into a probability through the logistic function (also known as the sigmoid function):

$$P(Y = 1) = \frac{1}{1 + e^{-Z}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 Y_{\text{BMI}})}} \quad (8)$$

The value range of the logistic function is between (0, 1), which conforms to the definition of probability.

2.3.2. Log-odds ratio form

The odds ratio (Odds) is defined as the ratio of the probability of an event occurring to the probability of it not occurring, that is, $\text{Odds} = \frac{P(Y=1)}{1-P(Y=1)}$. Taking the natural logarithm of the odds ratio gives the log-odds ratio (Log Odds):

$$\text{logit}(P) = \ln \left(\frac{P(Y=1)}{1-P(Y=1)} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 Y_{\text{BMI}} \quad (9)$$

This form is more convenient to understand the meaning of the regression coefficient. The regression coefficient β_i represents the change in the log-odds ratio when the independent variable X_i increases by one unit while other independent variables remain unchanged.

2.3.3. Parameter estimation

The likelihood function of the entire sample is the product of the likelihood functions of each observation:

$$L(\beta) = \prod_{i=1}^n L_i(\beta) = \prod_{i=1}^n P(Y = y_i)^{y_i} (1 - P(Y = y_i))^{1-y_i} \quad (10)$$

For the convenience of calculation, the logarithm of the likelihood function is usually taken to obtain the loglikelihood function:

$$\ln L(\beta) = \sum_{i=1}^n [y_i \ln P(Y = y_i) + (1 - y_i) \ln(1 - P(Y = y_i))] \quad (11)$$

Using the gradient descent algorithm to maximize the log-likelihood function $\ln L(\beta)$ and determine a set of optimal regression coefficient β estimates.

3. Results

As shown in Table.1, 68.96% of rural residents produce "safe" vegetables or pork for self-consumption, which reflects the residents' attention to the safety of food sources. They hope to control the quality of ingredients and ensure food safety through independent production. The 31.04% of residents who do not engage in self-production may be restricted by factors such as scarce land resources, tight daily working hours, or lack of relevant planting and breeding techniques. In terms of the purchase of certified foods, 67.15% of residents have purchased certified vegetables, and 74.45% and 74.43% of residents have purchased certified pork and certified dairy products, respectively, indicating that residents have a higher acceptance of safety certification for animal-derived foods. This may be because meat and dairy products are more prone to safety risks such as antibiotic residues and additive use in the production and processing process. In contrast, the purchase rate of certified vegetables is slightly lower, which may be due to the wide variety of vegetables, residents' insufficient awareness and attention to certification marks, and the higher price of certified vegetables may also affect the 32.85% of residents who did not purchase certified vegetables.

Table 1. Statistics on consumption trends of rural residents

| Behavior type | Yes | No |
|---|--------|--------|
| Produce "safe" vegetables/pork for self-consumption | 68.96% | 31.04% |
| Purchase certified vegetables | 67.15% | 32.85% |
| Purchase certified pork | 74.45% | 25.55% |
| Purchase certified dairy products | 74.43% | 25.57% |

Table 2. Linear regression results

| Variable | Coefficient (β) | Standard error | t-value | p-value | 95% confidence interval |
|-----------------------|-------------------------|----------------|---------|---------|-------------------------|
| Meat consumption | 0.35 | 0.12 | 2.92 | 0.004 | [0.11, 0.59] |
| Vegetable consumption | 0.5 | 0.1 | -5 | <0.001 | [-0.70, -0.30] |
| Age | 0.02 | 0.01 | 2 | 0.046 | [0.001, 0.04] |

From the linear regression results in Table.2, meat consumption is significantly positively correlated with the health status (BMI) of rural residents, with a coefficient of 0.35 and a p-value of 0.004, indicating that when other variables are constant, every one unit increase in meat consumption is expected to increase BMI by 0.35 units, and the 95% confidence interval is [0.11, 0.59], indicating that this positive correlation is relatively stable. Vegetable consumption is significantly negatively correlated with BMI, with a coefficient of -0.50 and a p-value less than 0.001, meaning that every one unit increase in vegetable consumption will decrease BMI by 0.50 units, and the 95% confidence interval is [-0.70, -0.30], showing that vegetable consumption has a strong inhibitory effect on BMI. The

Figure 1 is a radar chart of dietary structure, which shows the expenditure distribution of rural residents in different food categories. The four dimensions of the radar chart represent staple food expenditure, protein expenditure, dairy product expenditure, and non-dietary expenditure.

It can be seen from the figure that rural residents have the highest value in staple food expenditure, reaching 47.50, indicating that staple food occupies an important position in their dietary structure. Followed by protein expenditure, with a value of 77.60, which may reflect the attention of rural residents to protein source foods such as meat, eggs, and soy products, which are important sources of protein needed by the body.

The value of non-dietary expenditure is 103.39, and this high value may indicate that residents' expenditure on non-food items also accounts for a considerable proportion, which may include daily necessities, clothing, etc. The value of dairy product expenditure is 60.42. Although it is not the highest, it also shows that residents have a certain concern for dairy products as a source of nutrition.

Combining these data, we can see that the dietary structure of rural residents is relatively balanced, and the expenditure on staple foods and proteins accounts for a large proportion, which helps to meet the body's basic needs for energy and nutrition.

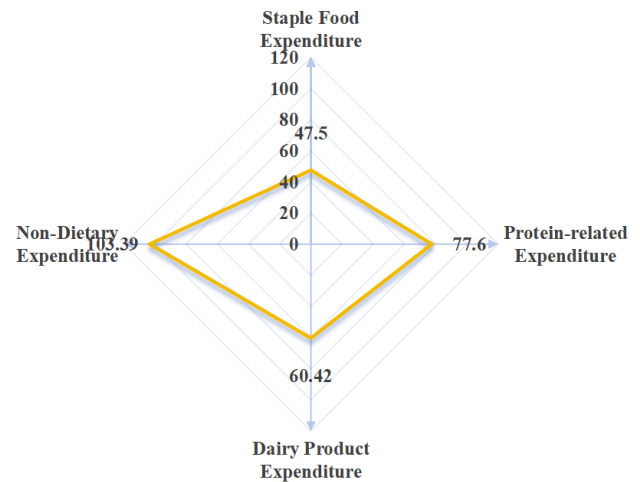


Figure 1. Radar chart of dietary structure

coefficient of age and BMI is 0.02, the p-value is 0.046, reaching a significant level, and the 95% confidence interval is [0.001, 0.04], indicating that with the increase of age, BMI has a slight upward trend.

Table 3. Logistic regression results

| Variable | OR value | p-value | 95% confidence interval |
|-----------------------|----------|---------|-------------------------|
| Vegetable consumption | 0.8 | 0.02 | [0.65, 0.98] |
| BMI | 1.15 | 0.01 | [1.03, 1.28] |

In the Logistic regression results in Table.3, the OR value of vegetable consumption is 0.80, the p-value is 0.02, and the

95%. The OR value of vegetable consumption is 0.80, the p-value is 0.02, and the 95% confidence interval is [0.65, 0.98]. This indicates that vegetable consumption can reduce the probability of chronic diseases in rural residents. An OR value less than 1 means that for each unit increase in vegetable consumption, the odds of suffering from chronic diseases are 0.8 times the original, and the result is statistically significant. The OR value of BMI is 1.15, the p-value is 0.01, and the 95% confidence interval is [1.03, 1.28], meaning that for each unit increase in BMI, the risk of chronic diseases is 1.15 times the original, and this association is significant, reflecting that an increase in BMI will increase the possibility of suffering from chronic diseases.

4. Conclusions

This study deeply explores the influence mechanism of market dietary structure on the health of rural residents and proposes optimization paths through the analysis of the national rural tracking survey database of the China Rural Revitalization Association from 2022 to 2024. The study finds that the dietary structure of rural residents is changing from the traditional model dominated by grains and vegetables to one rich in more animal foods, processed foods, and excessive oil, salt, and sugar, leading to the rise in the prevalence of chronic diseases and the prominent problem of nutritional imbalance. Regression analysis shows that meat consumption is significantly positively correlated with BMI, while vegetable consumption is significantly negatively correlated with BMI and can significantly reduce the risk of chronic diseases. In addition, age growth is slightly positively correlated with BMI, and an increase in BMI is an important risk factor for chronic diseases in rural residents. In terms of food consumption behavior, rural residents show attention to the safety of food sources. Most residents produce "safe" vegetables or pork for self-consumption, and have a high acceptance of animal-derived certified foods. In terms of dietary expenditure, staple foods and protein foods account for a relatively high proportion, and the dietary structure is relatively balanced, but there is still room for improvement in vegetable and dairy consumption.

Aiming at the above problems, the following optimization paths and suggestions are put forward:

(1) Guide reasonable dietary consumption: Strengthen the publicity of nutritional health knowledge, improve the awareness of rural residents about reasonable diet, guide them

to increase vegetable consumption, control meat intake, and form scientific dietary habits.

(2) Ensure food supply and safety: Pay attention to the production and supply of certified vegetables, reduce the price of certified vegetables, and improve the purchasing power and willingness of rural residents for certified vegetables; strengthen food market supervision to ensure the safety and quality of various foods.

(3) Pay attention to the health of key populations: Formulate personalized health management plans for rural residents of different age groups, especially the elderly, carry out regular health check-ups, and intervene in health problems in a timely manner.

(4) Promote the construction of rural nutritional health service system: Improve rural medical and health service resources, strengthen the construction of professional talent teams for nutritional health in rural areas, and provide professional nutritional consultation and health guidance services for rural residents.

References

- [1] Odoms-Young A, Brown A G M, Agurs-Collins T, et al. Food insecurity, neighborhood food environment, and health disparities: state of the science, research gaps and opportunities [J]. *The American journal of clinical nutrition*, 2024, 119(3): 850-861.
- [2] World Health Organization. *The State of Food Security and Nutrition in the World 2023: Urbanization, agrifood systems transformation and healthy diets across the rural-urban continuum* [M]. Food & Agriculture Org., 2023.
- [3] Yue W A N, Hao H U, Wuyang H U. Agricultural production structure, market conditions and farmers' nutritional intake in rural China [J]. *Journal of Integrative Agriculture*, 2022, 21(6): 1812-1824.
- [4] Hawkes C. Uneven dietary development: linking the policies and processes of globalization with the nutrition transition, obesity and diet-related chronic diseases [J]. *Globalization and health*, 2006, 2: 1-18.
- [5] Ma R, Yin S. Rural Economic Transformation and Household Consumption Structure: An Empirical Study in the Context of Urbanization [J]. *Journal of the Knowledge Economy*, 2024: 1-17.
- [6] Zeyu W, Weijing Z, Yanxi W, et al. Spatiotemporal Differences of Dietary Structure Index of Chinese Rural Residents and Its Influencing Factors [J]. *Economic geography*, 2023, 43(11): 165-172.