A study of factors affecting the physical health of urban residents based on Pearson's correlation analysis

Yuqi Cai*

School of Intelligent Science and Engineering, Harbin Engineering University, Harbin, China, 150006

* Corresponding Author Email: caiyuqi@hrbeu.edu.cn

Abstract. Over the years, the prevalence of chronic diseases has continued to rise in China, posing a great threat to people's health. In order to better understand the prevalence of chronic diseases and the factors affecting them, according to the Dietary Guidelines for Chinese Residents issued by the Chinese Nutrition Society, combined with the data from the Epidemiological Questionnaire on Chronic Non-Communicable Diseases and their Associated Factors, we used stratified sampling to select samples from people in different regions, ages and genders, and applied person correlation analysis and neural network analysis to construct a standardized healthy life evaluation model and a chronic disease-related degree model. methods, and constructed a standardized healthy life evaluation model and a chronic disease-related degree model. The degree of influence of living habits and dietary habits on physical health was analyzed, and the correlation between diet, living habits, age, gender, marital status, education level, occupation and other factors and the prevalence of chronic diseases was studied. Through modeling analysis, the risk of chronic diseases can be more accurately assessed, providing a scientific basis for prevention and treatment. Through the implementation of these recommendations, people can effectively reduce the risk of chronic diseases and improve the quality of life.

Keywords: Chronic Diseases, Living Habits, Dietary Standards, Health Standards, Correlation Analysis.

1. Introduction

With the accelerated pace of modern life, chronic diseases have become a major disease burden worldwide. How many factors such as lifestyle, environmental factors, genetics and other factors affect the occurrence of these diseases has become an urgent problem to be solved. Exploring the principles of the occurrence of chronic diseases and the prevention of chronic diseases, how to prevent chronic diseases and treat chronic diseases has attracted more and more people's attention.

Tao Zhang, Xiaodong Wang, Jiahui Ma [1] proposed a model for constructing a balanced diet to study the qualitative nature of the experience of healthy eating and physical activity social support in elderly patients with chronic diseases. Wang Jincheng [2] investigated the epidemiology of chronic obstructive pulmonary disease (COPD) among permanent residents aged ≥40 years in Mianyang city and analyzed its related influencing factors. The prevalence of COPD was compared among different investigative factors, and the independent influencing factors of COPD were analyzed by multifactorial logistic regression analysis for the meaningful factors. Wen Guoli [3] analyzed the influencing factors on the health of middle-aged and elderly people, from the self-assessment of health, self-care ability in daily life, the number of chronic diseases, the degree of depression and other four aspects, respectively, on the overall, physiological and psychological aspects of the health of the middle-aged and elderly people, and specifically analyzed the impact of individual factors on the health of the middle-aged and elderly people.

At present, scholars have less research on chronic diseases associated with residents' living habits and dietary habits, so this paper uses PERSON correlation to study the relationship between living habits and dietary habits on the physical health of urban residents' influencing factors, and analyzes the different populations in terms of dietary habits, living habits, and age and gender, to find the prevention of chronic diseases sub-factors, and to provide reasonable suggestions[4].

2. Data sources

2023 "Shenzhen Cup" Mathematical Modeling Challenge A - Analysis of Factors Affecting the Health of Urban Residents

http://www.m2ct.org(Obtained by removing outliers from the data and filling in missing values with the median)[5]

2.1. Healthy Living Evaluation Model

Dietary quality assessment criteria are an important part of the evaluation of nutritional status, and one of the commonly used methods is the Dietary Assessment Index (DAI). The Dietary Evaluation Index (DEI) is based on specific dietary guidelines and is used to assess and calculate dietary habits that are often difficult to quantify accurately through data.

For dietary habits, as the data mentioned in the data is too redundant, after screening by correlation analysis, some representative data were selected for detailed analysis. For example, the consumption of staple foods, including rice and wheat[8]; meat, including pork, beef, lamb and poultry; fresh vegetables and fresh fruits; and salt and oil were analyzed in detail[6]. For lifestyle habits, we decided to select smoking status, alcohol consumption, and physical activity for Pearson correlation tests[7].

2.2. Indicators for the person correlation model

Person correlation coefficient:

$$r = \frac{\sum_{i-1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{1}$$

For space reasons, the more relevant parts of the paper are selected for presentation.

(1) Eating habits:

Condiments: (retain two decimal places), The correlations between flavoring and factors is shown in Table 1.

Table 1. Correlations between flavoring and factors

	Birth year	Distinguishing between the sexes	Marital status	Educational attainment	Nation	Careers
Animal fat	0.04 (0.00)	-0.02 (0.24)	0.00 (0.99)	-0.05 (0.00)	0.07 (0.00)	-0.01 (0.25)
Vegetable oil	-0.03 (0.02)	0.01 (0.64)	-0.01 (0.54)	-0.13 (0.00)	-0.03 (0.03)	0.06 (0.00)
Ssalt	-0.02 (0.22)	0.012 (0.33)	-0.00 (0.78)	-0.07 (0.00)	-0.01 (0.41)	-0.00 (0.73)
Jealousy (in love rivalry)	-0.08 (0.00)	0.03 (0.02)	0.05 (0.00)	0.11 (0.00)	0.04 (0.00)	0.00 (0.76)
Soya sauce	-0.046 (0.00)	0.04 (0.00)	0.01 (0.50)	-0.13 (0.00)	-0.03 (0.03)	0.05 (0.00)
Mayonnaise	-0.01 (0.68)	-0.03 (0.01)	0.01 (0.71)	0.01 (0.71)	0.05 (0.00)	-0.02 (0.22)
Monosodium glutamate (MSG)	0.00 (0.98)	-0.03 (0.01)	-0.01 (0.28)	0.02 (0.14)	0.08 (0.00)	0.00 (0.85)

From the above table, it can be seen that there is a significant positive correlation between age and consumption of animal oil, the older the age the higher the consumption of animal oil. There is a significant negative correlation between literacy level and the various condiments, as shown by the fact that the higher the literacy level, the lower the consumption of animal oil, vegetable oil, salt and soy sauce[8].From the perspective of ethnicity, when the ethnicity is Han Chinese, it is manifested in higher consumption of animal oil, vegetable oil, salt, soy sauce, vinegar and salt. Similarly, it is

known from the correlation coefficients that there is a significant positive correlation between occupation and the consumption of each type of condiment[9].

Habits.

Smoking and drinking: the correlations between smoking and drinking and factors is shown in Table 2.

Table 2. Correlations between smoking and drinking and factors

	Birth year	Careers	Marital status	Educational attainment	Distinguishing between the sexes
Are you a smoker?	0.02 (0.03)	0.16 (0.00)	0.02 (0.14)	-0.02 (0.14)	0.57 (0.00)
Whether or not alcohol is consumed	-0.08 (0.00)	0.12 (0.00)	0.03 (0.01)	-0.12 (0.00)	0.33 (0.00)

From the correlation in the above table, it can be seen that the probability of young people smoking increases slightly compared to the middle-aged and the elderly, probably due to the higher pressure of life, and there is also a significant correlation between occupation and whether or not to smoke, as can be seen from the data, there is a positive correlation between occupation and whether or not to smoke, that is, it is easier to smoke in certain occupations. There is a significant correlation between gender and whether or not one smokes, as can be seen from the data, men are more likely to smoke. There was also a significant correlation between whether one smoked or not and whether one drank alcohol or not, specifically, the data showed that whether one drank alcohol or not was positively correlated with the probability of whether one smoked or not, i.e., those who drank alcohol were more likely to smoke.

(2) Physical exercise

The correlations between physical activity and factors is shown in Table 3.

Table 3. Correlations between physical activity and factors

	Birth year	Distinguishing between the sexes	Nation	Marital status	Careers	Educational attainment
Participation in physical exercise	-0.19 (0.00)	-0.04 (0.00)	-0.01 (0.50)	-0.01 (0.65)	0.12 (0.00)	0.05 (0.00)

As can be seen from the above data, there is a significant negative correlation between year of birth, gender and participation in physical activity, with young people being less likely to participate in physical activity than middle-aged and older people, and women being less likely to participate in physical activity than men. There is a significant positive correlation between occupation and participation in physical activity, specifically, the data show that certain occupations are more likely to participate in physical activity. There is a significant positive correlation between literacy and participation in physical activity, with those with higher literacy being more likely to participate in physical activity.

After synthesizing the data from the PERSON correlation test and the dietary and lifestyle habits, it can be seen:

a. Age:It was found that age has a greater impact on eating habits, the older the person, the greater the impact of age on eating habits, but due to the age of the body's functions gradually decline, the metabolism slows down, so the diet, life need to pay more attention.

b. Literacy level: Literacy level has a significant impact on lifestyle and dietary habits. People with higher levels of education are usually more concerned with healthy lifestyles and eating a balanced diet. On the contrary, people with lower levels of education may have more poor living and eating habits.

c. Marital status: Married or not in the diet and living habits of the larger impression, may be after marriage, the new family may be formed with the partner to change their own dietary habits as well

as habits, at the same time, their own consumerism will also be influenced by their partners, in the dietary habits of the pre-marriage and after marriage there may be a more obvious differences.

d. Occupation: different occupations in the dietary habits there will be a big difference, for example, urban white-collar workers in the choice of diet may favor vegetables, fruits and other fat-reducing meals, and go to reduce the proportion of staple foods, and physical workers because of the need to be engaged in the relevant physical labor, will be the proportion of staple foods to increase the physical workers in the movement of the time more, some white-collar workers due to overtime that time is long and sports Some white-collar workers have more time for exercise because they work overtime for a long time, so there is a strong correlation between occupation and dietary and living habits.

e. Gender: There are also big differences between men and women in terms of diet and lifestyle habits, with men smoking and drinking more than women.

f. Ethnicity: Due to the different religious beliefs and living cultures of different ethnic groups, there are more obvious differences and promises in terms of diet.

3. Factors affecting chronic diseases

Based on the person correlation analysis of the degree of influence of the above mentioned occupation, gender, age, education, etc. on living habits and dietary habits, in-depth analysis of the relationship and degree of correlation between chronic common diseases and smoking, drinking, living habits, dietary habits, nature of the work, exercise and other factors, due to the amount of data is too large and redundant, the first need to be for the data for the downgrading of the data processing, according to the data of the actual situation. Principal Component Analysis (PCA) was used to reduce the dimensionality of the data.

For principal component analysis, before analyzing, we need to perform KMO test and Bartlett's test on the data to determine whether it is possible to do principal component analysis, for the KMO value: 0.8 is very suitable for principal component analysis, 0.7-0.8 is generally suitable for between 0.6-0.7 is not very suitable for between 0.5 and 0.6 indicates poor, 0.5 means extremely unsuitable for under For Bartlett's test, if P is less than 0.05 and the original hypothesis is rejected, it means that principal component analysis can be done, if the original hypothesis is not rejected, it means that these variables may provide some information independently and are not suitable for principal component analysis.

3.1. Modeling the extent of chronic disease correlation

Using SPSSPRO software, principal component analysis downscaling was performed for diet-related factors. KMO test as well as Bartlett's test were performed and the results are as follows: the KMO test and Bartlett's test is shown in Table 4.

Table 4. KMO test and Bartlett's test

	KMO value	0.724
Bartlett's test of sphericity	approximate chi-square (math.)	14673.673
	df	561
	P	0.000

All of the above data meet the needs of principal component analysis, so principal component analysis can be used to reduce the dimensionality of the above redundant data. According to the gravel plot statistics, when using 12 principal components, the original data can maintain a relatively good degree of interpretation, therefore, the subsequent data dimensionality reduction processing are based on 12 principal components as the standard. The Diagram of debris is shown in Figure 1. The results of principal component selection and weights is shown in Table 5.

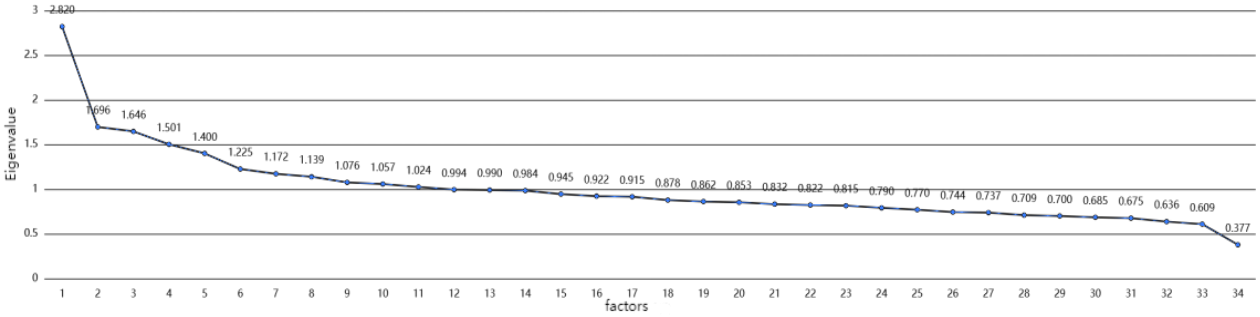

Figure 1. Diagram of debris

Table 5. Results of principal component selection and weights

Name (of a thing)	Explanation of variance (%)	Cumulative variance explained (%)	Weight (%)
Principal Component 1	0.139	13.86	28.945
Principal Component 2	0.058	19.663	12.118
Principal Component 3	0.054	25.059	11.27
Principal Component 4	0.047	29.804	9.908
Principal Component 5	0.041	33.878	8.51
Principal Component 6	0.037	37.556	7.68
Principal Component 7	0.036	41.159	7.524
Principal Component 8	0.035	44.622	7.232
Principal Component 9	0.033	47.885	6.815

The principal component analysis described above leads to the following formula for calculating a composite score based on dietary habits, which can then be used as the basis for subsequent calculations of the correlation between chronic diseases and their other factors.

F=(0.083/0.49)×F1+(0.05/0.49)×F2+(0.048/0.49)×F3+(0.044/0.49)×F4+(0.041/0.49)×F5+(0.036/0.49)×F6+(0.034/0.49)×F7+(0.033/0.49)×F8+(0.031/0.49)×F9+(0.031/0.49)×F10+(0.029/0.49)×F11+(0.029/0.49)×F12

3.2. Model Data Fitting

After the data dimensionality reduction process, the comprehensive score about dietary habits was obtained, which was used as a variable for evaluating dietary habits; at the same time, the data related to diabetes mellitus and hypertension were dimensionality reduction processed by the same method to obtain the corresponding variables, and then the neural network model was used to regressively analyze the data on living habits, dietary habits, whether to smoke or not, whether to drink or not, and so on[10].

The diabetes related data, for example, is brought into the established neural network model and the results are obtained after several training sessions: as is shown in Figure 2.

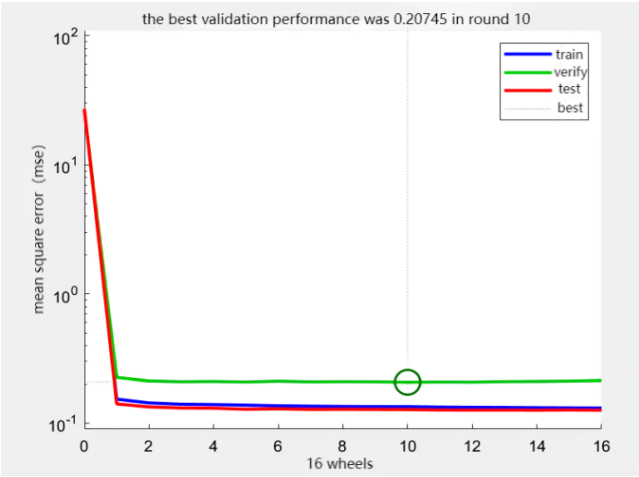

Figure 2. 16 rounds of mean square error

In round 10, the best mean square error of 0.20745 was obtained, while the vast majority of the test points fell within the error allowance after testing as is shown in Figure 3.

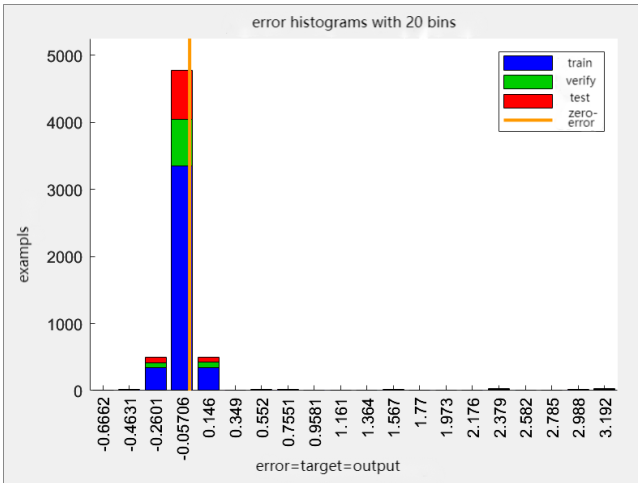

Figure 3. Error histogram for round 10 with 20 bins

The regression equation is as follows:the regression equation is shown in Figure 4.

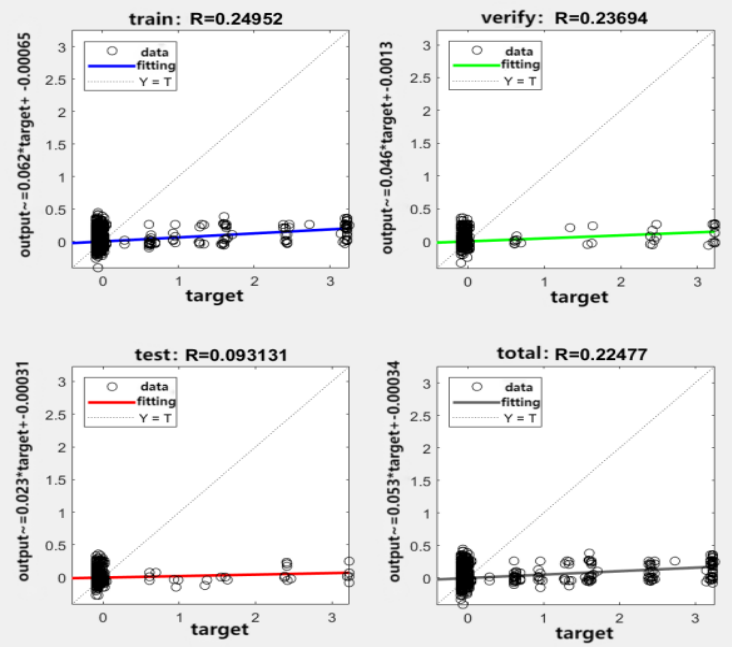

Figure 4. Regression equation

After the data were processed and verified by Pearson's correlation coefficient, we obtained its significance p-value of 0.000, meanwhile, according to the regression equation, it can be seen that the model basically meets our requirements, and the point offset is within our error margin.

Similarly, the same model was used to verify the correlation between other diseases and the above mentioned factors[11], and the fit obtained was good, indicating that there is a strong correlation between chronic diseases and factors such as smoking, alcohol consumption, lifestyle habits, dietary habits, nature of work, and exercise.

4. Conclusion

Based on analyzing the factors of citizens' physical health, this paper establishes a model for evaluating living habits and dietary habits, and uses PERSON correlation analysis to analyze the factors influencing the physical health of living habits and dietary habits.The results show that the

probability of residents suffering from chronic diseases has a close relationship with dietary habits and living habits, and that there is a dietary evaluation and living habits evaluation model that has full application value.

Of course, the model is analyzed based on the given questionnaire data, so the results are limited by the data. The results obtained do not fit well with the dietary and living habits of the nation's population. If the data are not representative or complete, the evaluation results of the model may be biased, and in the actual physical health factors of the population, multiple factors are received. Therefore, each city is analyzed by using the evaluation model according to local habits, fully introducing local climate, humanities and other influencing factors. This is conducive to localities to get in line with the local health factors affecting urban residents, to get more accurate prediction data, to project the probability of suffering from chronic diseases, and to make scientific and rational decision-making.

References

[1] ZHANG Tao,WANG Xiaodong,MA Jiahui et al. A balanced diet mathematical modeling method[P].Geriatrics Research,2023,4(03):37-42.

[2] Wang Jincheng, Yi Ke, Zeng Qiang, etc. Epidemiological investigation of chronic obstructive pulmonary disease and analysis of influencing factors [J]. Shandong Pharma, 2021,61 (05): 47-50.

[3] Wen Guoli. Analysis of the health influencing factors of middle-aged and elderly people [D]. Guizhou University of Finance and Economics, 2020.DOI:10.27731/d.cnki.ggzcj. 2019.000244.

[4] Xie Congshang. A review of studies on the correlation between dietary habits and chronic diseases[J]. China Primary Health Care,2022,36(11):66-67+72.

[5] Chinese Nutrition Society. Dietary Guidelines for Chinese Residents 2022 [M]. Beijing: People's Health Publishing House, 2022.

[6] WANG Siqi, XU Jianwei, BAI Yamin, et al. Progress of research on the effects of salt reduction interventions in populations [J]. China Chronic Disease Prevention and Control, 2020, 28(6):463-466.

[7] ZHU Na,LIU Xia-Yang,YUAN Yuan et al. Exploring the construction of data model for standardized meal evaluation service[J]. China Standardization,2023(08):98-104.

[8] STOOKEY J D, WANG Y, GE K, et al. Measuring diet quality in China: the INFH-UNC-CH diet quality index[J]. European Journal of Clinical Nutrition,2000,54(11):811-821.

[9] U.S. Department of Agriculture. Healthy Eating Index (HEI) [EB/ OL]. [2020-12-25]. https://www.fns.usda.gov/resource/healthyeating-index-hei.

[10] Chinese Medical Association Diabetes Branch. Guidelines for the prevention and treatment of type 2 diabetes mellitus in China (2020 edition) [J]. Chinese Journal of Diabetes Mellitus, 2021, 13(4):315-409.

[11] China Hypertension Prevention and Control Guidelines Revision Committee, Hypertension Alliance China, Chinese hypertension prevention and treatment guidelines (2018 revised edition) [J]. Chinese Cardiovascular Journal, 2019, 24(1):24-56.