



Prevention of Chronic Diseases and Promotion of Health in the Context of Social Practice at Applied Universities

Shengjie Cui^{a,1}, Xin Yu^{b,2}, Chunyan Wang^{a,3*}, Weiliang Li^{c,4}, Sheng Wang^{d,5},
Hui Wang^{e,6}, Jiali Wu^{f,7}, Sihang Wang^{g,8}

^aORDOSINSTITUTE OF TECHNOLOGY, College of Medical, Ordos, China

^bHENG XING University, The School of Mechanical and Automotive Engineering, Qingdao, China

^cInner Mongolia University of Finance and Economics, School of Law, Hohhot, China

^dORDOSINSTITUTE OF TECHNOLOGY, College of Large Aircraft, Ordos, China

^eORDOSINSTITUTE OF TECHNOLOGY, Department of Mathematics and Computer Engineering, Ordos, China

^fInner Mongolia University of Finance and Economics, School of Finance, Hohhot, China

^gORDOSINSTITUTE OF TECHNOLOGY, Mechanical and Traffic Engineering Department, Ordos City, China.

¹3408843435@qq.com, ²2466547751@qq.com, ^{3*}289214927@qq.com,
⁴1w118879189253@126.com, ⁵wangsheng2_0218@qq.com,
⁶huiwang@oit.edu.cn, ⁷1851940832@qq.com,
⁸1911646235@qq.com

Abstract. With changes in lifestyle, the incidence of chronic non-communicable diseases (such as cardiovascular and cerebrovascular diseases, diabetes, malignant tumors, and chronic obstructive pulmonary disease) continues to rise, exerting tremendous pressure on the social healthcare system. This study aims to analyze the dietary habits and their rationality of the population, explore the relationship between living habits and demographic characteristics, and study the association between chronic diseases and various lifestyle factors to formulate targeted health promotion strategies, effectively reduce the incidence of chronic diseases, and improve the overall health level of the population.

Keywords: data preprocessing, dimensionality reduction, quality check, principal component analysis, correlation analysis

1 Intrudction

Chronic non-communicable diseases (such as cardiovascular and cerebrovascular diseases, diabetes mellitus, malignant tumors and chronic obstructive pulmonary disease) have become an important issue affecting the health of our population^[1]. As lifestyles change, the incidence of chronic diseases continues to rise, putting enormous pressure on the social healthcare system. Studies have shown that the health status of the population is closely related to a number of factors, including age, dietary habits, physical

activity and occupation^[2]. Therefore, how to promote health through rational diet, moderate exercise and healthy lifestyle has become a topic of general concern to the whole society. To this end, we need to analyze in depth the dietary habits of the population and their rationality, explore the relationship between living habits and demographic characteristics, and study the association between chronic diseases and various lifestyle factors in order to formulate targeted health promotion strategies, so as to effectively reduce the incidence of chronic diseases and improve the overall health of the population^[3].

2 Problem Analysis

2.1 Questionnaire Validity and Reliability Tests

Reliability test is a test of the reliability of the questionnaire, which refers to the degree of consistency of the results obtained by using the same method to measure the same object, i.e., the degree of response to the actual situation. It mainly shows the consistency, consistency, reproducibility and stability of the results. In this paper, the reliability coefficient method *Cronbach α (Alpha)* was used to test the reliability of the questionnaire.

$$\alpha = \frac{N}{N-1} \left[1 - \frac{\sum S_i^2}{S^2} \right] = \frac{N}{N-1} \left[1 - \frac{N}{N-2r} \right] \quad (1)$$

Where N is the number of questionnaire questions, S_i^2 is the in vivo variance of the first question, S^2 is the total score variance, and r is the sum of the correlation coefficients between the questions^[4]. First we need to perform KMO test and Bartlett's spherical test on the questionnaire to determine the covariance or correlation between the questions.

$$KMO = \frac{\sum_{i \neq j} r_{ij}}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} \alpha_{ij}^2} \quad (2)$$

Where r_{ij} denotes simple correlation coefficient and $\alpha_{j \cdot 1, 2, 3, \dots, k}$ denotes partial correlation coefficient.

Table 1. *Cronbach α* table of coefficients

Cronbach α coefficient	Standardized Cronbach α coefficient
0.769	0.792

With the results in Table 1 we can see that the *Cronbach α* coefficient is 0.729. Usually, if the *Cronbach α* coefficient is above 0.9, the reliability of the test or scale is very good, between 0.8 and 0.9 indicates good reliability, between 0.7 and 0.8 indicates acceptable reliability, between 0.6 and 0.7 indicates fair reliability, between 0.5 and 0.6 indicates less than optimal reliability, and if it is below 0.5, the questionnaire has to be considered for restructuring^[5]. Therefore, through Table 2 we can see that the reliability of our questionnaire is acceptable.

2.2 Indicator Downscaling

In order to analyze the correlation of lifestyle and dietary habits with factors such as age, gender, marital status, literacy, and occupation, we need to downscale the indicators to reduce redundancy. Four key indicators were selected: the number of cigarettes smoked per week and the number of days of passive smoking to represent smoking; the amount of alcohol consumed per week to indicate alcohol consumption; and the average time spent exercising per day to reflect physical activity. These indicators effectively replace the original multiple indicators, thus simplifying the analysis^[6].

In terms of dietary habits, for the huge data of 109 indicators, we performed a dimensionality reduction process and streamlined it to 39 items. We focused on indicators with overlap, such as dining locations^[7]. The analysis led to the selection of six indicators to characterize meal locations, covering interrelated factors such as the number of days without breakfast and breakfast at home. This dimensionality reduction treatment not only improves the efficiency of data analysis, but also ensures the representativeness of the indicators and provides a clearer basis for subsequent research. With these simplified indicators, we are able to explore more effectively the relationship between lifestyle habits and demographic characteristics and their impact on chronic diseases^[8].

At the same time, we imported the data into SPSS for correlation tests. Here, we introduced the PERSON correlation coefficient for determining the relationship between the indicators^[9].

Sample covariance:

$$Cov(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1} \quad (3)$$

Sample standard deviation:

$$s_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (4)$$

Sample Pearson correlation coefficient:

$$r_{xy} = \frac{Cov(X, Y)}{s_x s_y} \quad (5)$$

Using the correlation coefficients to first study the habits of life in relation to age, gender, marital status, education, occupation and other factors Smoking is very clearly related to the sex of the person and is not very much linked to other indicators. For physical activity there was a strong correlation with year of birth and level of education.

For eating habits, SPSS was used to calculate the effect with age, gender, and literacy, and it was concluded that eating place is closely related to the year of birth, i.e., the age of the person; there is also a more pronounced relationship with marital status. The literacy correlation is not as pronounced as for several other indicators.

2.3 The Degree of Correlation Between Common Chronic Diseases and Each Indicator

In studying the relationship between chronic diseases and factors such as smoking, alcohol consumption, dietary habits, lifestyle, nature of work and exercise, multiple linear regression models need to be constructed to obtain exact functional expressions. Since dietary habits involve 29 initially downscaled indicators, a linear regression may not be appropriate, so data downscaling techniques were introduced. First, the correlation of these secondary indicators was assessed by KMO test and Bartlett's spherical test to select the appropriate downscaling method. Component analysis is commonly used for dimensionality reduction as it is effective in extracting metrics with strong correlations. However, if the test results were unsatisfactory, the t-SNE method was used to downscale the multidimensional nonlinear indicator to a two-dimensional series to ensure the validity of data processing. KMO test and Bartlett's spherical test help to determine the covariance or correlation between the indicators, so as to optimize the subsequent analysis process and enhance the accuracy and reliability of the model^[10].

$$KMO = \frac{\sum \sum_{i \neq j} r_{ij}}{\sum \sum_{i \neq j} r_{ij}^2 + \sum \sum_{i \neq j} a_{ij}^2} \quad (6)$$

Where, r_{ij} denotes simple correlation coefficient and $\alpha_{ij}, 1, 2, 3, \dots, k$ denotes partial correlation coefficient. The results of the numerical processing, imported into SPSS for multidimensional indicator KMO test and Bartlett's spherical test were obtained as shown below.

Table 2. Results of KMO test and Bartlett's test of sphericity

Indicator name		Dietary situation
KMO value		0.839
Bartlett	approximate chi-square (math.)	57700.152
Sphericity test	df	741
	P	0.000**

Note: ***, **, * represent 1%, 5%, and 10% significance levels, respectively.

Therefore, for the results of this test, we used principal component analysis to downscale the indicators for the level of dietary status.

Standardized processing: The units of the collected indicators are different, for example, the unit of weight data is kilogram, but the unit of rice intake is two, which is a different unit; there is a big difference in the magnitude of the values between the collected data. Therefore, in order to eliminate the differences in scale and magnitude between indicators, the data collected need to be standardized.

Numerical values collected for each indicator a_{ij} were standardized and transformed into standardized indicators \tilde{a}_{ij} .

$$\tilde{a}_{ij} = \frac{a_{ij} - \mu_j}{s_j} \quad (7)$$

where denotes the first indicator and the first sample for which data were collected. μ_j is the mean and standard deviation of the first indicator.

Correlation coefficient matrix: Calculate the correlation coefficient between the collected indicators and construct a correlation coefficient matrix $\mathbf{R} = (r_{ij})_{5 \times 5}$. Convenient for subsequent calculations.

$$r_{ij} = \frac{\sum_{k=1}^n \tilde{a}_{ki} \cdot \tilde{a}_{kj}}{n-1} \quad (8)$$

Where n is the total number of samples for which data r_{ij} were collected and is the correlation coefficient between the first i indicator and the first j indicator.

Calculate the eigenvalue eigenvectors: Using the relational data matrix $\mathbf{R} = (r_{ij})_{5 \times 5}$ derived in the second step, calculate the eigenvalues of the matrix $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$, m is the total number of indicators related to meteorological conditions collected as well as the corresponding eigenvectors, from which the eigenvectors $u_1, u_2, u_3, \dots, u_m$ form a new indicator.

Selection of principal components: Calculate the contribution rate and cumulative contribution rate according to the eigenvalues obtained in the third step, select the principal components, and discard the principal components with lower contribution rates. Select the former p most important eigenvectors as the base vectors of the new coordinate system to form the new principal components. Where the contribution rate, cumulative contribution rate is calculated as:

$$g_j = \frac{\lambda_j}{\sum_{k=1}^m \lambda_k} \quad (9)$$

$$1 - j = \frac{\sum_{k=-1}^p \lambda_k}{\sum_{k=-1}^m \lambda_k} \quad (10)$$

where g_j is the contribution of the first j principal component and $1 - j$ is the cumulative contribution.

Calculation of the composite score: each sample of the original data set of meteorological conditions will be collected and projected onto the new coordinate system, and the contribution as a principal component g_j will be taken as a weight to be calculated, so as to obtain the principal component score of the sample. The formula for the principal component score is shown below:

$$Z = \sum_{j=1}^p g_j y_j \quad (11)$$

The selection of principal components through the gravel plot and the contribution rate, it is concluded that the selection of the first 11 principal components for dimensionality reduction can get a better degree of explanation of the original variables.

Therefore, the first 11 principal components are selected for the subsequent dimensionality reduction process. In order to visualize the dimensionality reduction process, the factor loadings are plotted here, and the spatial distribution of the principal components is presented in the form of quadrant plots by dimensionality reduction of multiple factors into two or three principal components.

2.4 Establishment and Solution of Multiple Linear Regression Models

For the analysis of common chronic diseases (e.g., hypertension, diabetes, etc.) selected hypertension, diabetes, respectively, only one indicator to indicate the relevant chronic disease, such as y_1 for hypertension and y_2 for diabetes. Multiple linear regression equations of y_1 with factors ($x_1, x_2, x_3, x_4, x_5, x_6$) such as smoking, alcohol consumption, dietary habits, lifestyle, nature of work, and exercise were established. Y_2 is the same. Here we take hypertension as an example, for the indicators related to hypertension, which are too redundant, and choose to deal with downscaling. KMO and Bartlett's tests were performed and the results are shown below.

The results were found to be better and could be directly dimensionalized using principal component analysis. Therefore, the hypertension dataset was downscaled to one piece of data. In order to explain the process of implementing principal components, a gravel plot for principal component selection and a matrix heat map for describing the relationship between principal components and the original variables are plotted here using SPSSPRO.

The fragmentation plot with the eigenvalues of the hypertension indicator in Table 1 shows that for the downscaling of the hypertension indicator two principal components can be chosen to achieve 74% of the explained variance. Therefore, there is a better interpretation of the results, and in order to more adequately explain the meaning of its newly generated principal components, the molecular heat matrix was plotted. Eventually, a column of indicators used to express the data related to hypertension was obtained, and a multivariate linear regression model was built with this indicator as y . By double analyzing the results of our image and Person test during data processing, we can see the relationship between these variables. Therefore, we develop a multiple linear regression model as follows

$$\begin{cases} y_{\text{fingh blood pressure}} = \alpha_0 + \alpha_1 x_{\text{Cigarette}} + \alpha_2 x_{\text{Paxsive}} + \alpha_3 x_{\text{Drinking}} + \alpha_4 x_{\text{Exercise}} + \alpha_5 x_{\text{Catering}} \\ \beta_{\text{Draberes}} = \beta_0 + \beta_1 x_{\text{Treat}} + \beta_2 x_{\text{Help}} + \beta_3 x_{\text{Constant}} + \beta_4 x_{\text{Exercise}} + \beta_5 x_{\text{Catering}} \end{cases} \quad (12)$$

$$\xi \sim N(0, \sigma^2)$$

Where, ξ obeys a normal distribution and, $\beta_1, \beta_2, \dots, \beta_4$ is the coefficient of the respective variable. The significance P-value is 0.000***, which presents significance at the level and rejects the original hypothesis that the regression coefficient is 0. Therefore, the model basically meets the requirements. For the covariate covariance performance, VIF is all less than 10, so the model has no multicollinearity problem and the model is well constructed. The same process can be done for diabetes, and the results are shown directly here. The significance P-value is 0.000***, which presents significance at the level and rejects the original hypothesis that the regression coeffi-

cient is 0. Therefore, the model basically meets the requirements. For the covariate covariance performance, VIF is all less than 10, so the model has no multicollinearity problem and the model is well constructed.

3 Conclusion

The comprehensive evaluation model demonstrates several advantages in terms of data analysis and processing, indicator evaluation system, comprehensive evaluation method and clarity of conclusions. First, the model ensures the accuracy and credibility of the data through comprehensive data analysis, including anomalous data exclusion, numerical processing and validity confidence test. Second, the model constructs a scientific and comprehensive indicator evaluation system and analyzes the correlation between primary and secondary indicators. In addition, various methods such as KMO test, Bartlett's sphericity test, principal component analysis and t-SNE were applied and combined with the TOPSIS method for comprehensive evaluation so as to provide objective evaluation results. However, there are some drawbacks to the model, mainly including assumption dependency, data limitations and subjectivity issues. The reliability of the model depends on the reasonableness of the assumptions, and the representativeness of the data and the selection of evaluation indicators may be influenced by the subjective judgment of the researcher. In addition, the models are based on data from specific times, places or populations, and the generalizability of their conclusions needs to be further verified. To increase the potential for generalization, the model can be extended not only to other domains such as workplace learning and social media, but also to different types of data analysis such as observational and experimental data. The methods and techniques in the model can be applied to the study of a wider range of issues, while the indicator evaluation system and the comprehensive evaluation method can be adapted to specific needs. The rollout needs to focus on the adaptability of the model to ensure adequate validation in the new environment to improve its accuracy and reliability. In conclusion, the model's potential for application in multiple dimensions and room for improvement make it of strong generalization value. The comparison results of this model with the general model are shown in Table 3.

Table 3. Model comparison

performance index	Comprehensive evaluation model	
Data analysis and processing	Including exception data exclusion, numerical processing, and validity confidence testing	May not include all of these steps
Index evaluation system	Build a scientific and comprehensive index evaluation system	May not be comprehensive or scientific
Comprehensive evaluation method	Combining KMO test, Bartlett spherical test, principal component analysis and t-SNE method	Single or less common assessment methods may be used
Conclusion Definition	The conclusion is clear and easy to understand	The conclusions may not be clear enough

References

1. Lin Tian, Weijian Han, Weilun Tian. Digital financial inclusion and residents' health: evidence from rural China [J]. *Discover Social Science and Health*, 2024, 4 (1):
2. Debra R Keast, Patricia M Guenther. Sodium Content and Sodium Intake Contributions of Store-Bought and Restaurant-Prepared Foods in Their As-Eaten Form: National Health and Nutrition Examination Survey, 2009–2018 [J]. *Current Developments in Nutrition*, 2024, 8 (10):
3. Yiyang Yuan, Adrita Barooah, Kate L Lapane, Deborah Mack, Anthony J Rothschild, Christine M Ulbricht. Health Profile Transitions and the Association With Cognitive Impairment in Older Nursing Home Residents With Suicidal Ideation. [J]. *International journal of geriatric psychiatry*, 2024, 39 (10):
4. E I Aksenova, N N Kamynina, D P Derbenev, P S Turzin. [Priority areas for the development of healthy cities that contribute to the preservation of the health of their inhabitants]. [J]. *Problemy sotsial'noi gigieny, zdravookhraneniia i istorii meditsiny*, 2024, 32 (Special Issue 2):
5. V. Haley Zitlin, O. Trapp, S. Alasvand. Assessment of Healthy Aging Indicators of Residents Residing in a Neighborhood Environment That Promotes Physical Activity [J]. *Journal of the Academy of Nutrition and Dietetics*, 2024, 124 (10S):
6. Xiufen Gu, Dayong Wang. Resolving sustainable development conflicts caused by vinyl leakage accidents in the USA: From the perspective of graph model under fuzzy preferences with self-confidence [J]. *Journal of Cleaner Production*, 2024, 473
7. Jinhee Kim, Jennifer Green, Erica McIntyre, Christopher Standen. Considering Residents' Health and Well-Being in the Process of Social Housing Redevelopment: A Rapid Scoping Literature Review [J]. *Journal of Urban Health*, 2024, (prepublish):
8. Charles Brown, Brenna Dotson, Juandria Montgomery, Chastity Sutterfield, Geeta Maharaj. Evaluating the Effectiveness of Using the Teach-Back Method to Improve the Health Literacy of Individuals in the Community. [J]. *Journal of community health nursing*, 2024,
9. Dian Ding, Yueqi Jiang, Shuxiao Wang, Jia Xing, Zhaoxin Dong, Jiming Hao, Pauli Paasonen. Unveiling the health impacts of air pollution transport in China [J]. *Environment International*, 2024, 191
10. Yingsen Zhang, Xinwei Lu, Xiufeng Han, Tong Zhu, Bo Yu, Zhenze Wang, Kai Lei, Yufan Yang, Sijia Deng. Determination of contamination, source, and risk of potentially toxic metals in fine road dust in a karst region of Southwest China. [J]. *Environmental geochemistry and health*, 2024, 46 (10):

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

