

Study on the Influencing Factors of Digital Health Ability of the Elderly Based on Stepwise Regression

Siyu Luo[#], Hanbo Min[#], Yi Ren[#], Muyang Liu[#], Yihan Zhao[#]

College of Science, Minzu University of China, 100081 Beijing, China.

[#] These authors contributed equally.

Abstract: This study focuses on the ability improvement of the elderly in the field of digital health in post-epidemic era noting the digital transformation's significant impact on elderly medical service access. Despite government and corporate efforts to promote digital adoption and "Internet + medical health," elderly individuals still struggle with technology proficiency, health constraints, and privacy issues. A questionnaire survey was conducted to assess the digital health capabilities of the elderly in a specific region, revealing a mixed proficiency level with some individuals lagging. Using projection pursuit and stepwise regression analysis, the study identified key factors influencing digital health proficiency, including psychosocial elements, support systems, and privacy concerns. Findings indicate that interpersonal and community support positively affect digital health ability, while technology anxiety negatively impacts it significantly. The model's high fit suggests reliable results, emphasizing the importance of social and community support in enhancing elderly digital health proficiency and the need to address technology anxiety. The study aims to facilitate smoother elderly integration into digital society and enjoyment of digital conveniences

Keywords: Elderly, Digital Health, Stepwise Regression

1. Introduction

Post-epidemic era refers to the profound social, economic and cultural changes experienced after the COVID-19 epidemic was brought under control. During this period, digital transformation has changed the way of life and health service mode, and telemedicine and health management services have sprung up.

However, older groups are at risk of marginalization in the wave of digitization, which not only exacerbates health inequalities. It may also lead to a "digital divide". In order to meet this challenge, it is particularly urgent and critical to assess and improve the digital health ability of the elderly.

In 2022, Jia et al. proposed suggestions to improve the user demand, resource supply, organizational coordination and digital construction of the digital health service platform of HJ company in smart elderly care[1]. In 2023, Xia et al. studied the industrial ecology of digital exercise therapy in China, analyzed international experience, and discussed its application and development strategies in China. They proposed to strengthen data security, improve clinical trial standards, optimize the regulatory system and payment mechanism, and promote the high-quality development of digital exercise therapy[2]. In 2023, Hu et al. discussed how digital technology affects the satisfaction of medical services. Through questionnaire surveys and data analysis, they verified the impact of digital empowerment on the satisfaction of medical services. They found that digital empowerment, perceived usefulness, and perceived ease of use have a positive impact on the satisfaction of medical services, and perceived usefulness and perceived ease of use play a mediating role[3]. In 2023, Lv et al. studied the barriers and influencing factors faced by the elderly in using digital health technologies at home. The study found that low technology acceptance, education level, and digital literacy are the main barriers faced by the elderly population[4]. In 2024, Xie et al. analyzed the causes of the digital divide among the elderly

from the perspective of healthy aging, and proposed that the governance of the digital divide among the elderly needs to empower the elderly, empower the digital environment, implement classified governance, promote coupling, and form a synergistic effect[5]. In 2024, Liu et al. studied the digital divide in online health services for the elderly in the digital healthcare era in China, analyzed the dilemmas faced by the elderly in utilizing online health services, and proposed key solutions[6]. In 2024, Chen et al. used the logistic regression analysis method to study the influencing factors of anxiety in the use of digital health technologies by elderly patients with chronic diseases. The conclusion shows that elderly patients with chronic diseases have relatively low anxiety about digital health technologies, and the key factors are education, income, and learning ability. It is suggested to improve acceptance through age-friendly design and targeted training[7]. In 2024, Wan et al. surveyed and analyzed the current status of digital health literacy of the elderly, and used logistic regression analysis to explore the influencing factors. It was found that low education, smoking, and health anxiety are risk factors, and regular physical examinations are protective factors. It is suggested to pay attention to the elderly with low education and strengthen health education and intervention[8].

At present, there is still room for further research on the influencing factors of digital health ability of the elderly. Based on data collection and statistical analysis, this paper puts forward a new idea to study the digital health ability of the elderly.

Stepwise regression analysis was used in this study. It systematically explores many potential influencing factors, such as social psychological factors, interpersonal support, privacy issues and so on. In order to accurately identify the key variables that have a significant impact on the digital health ability of the elderly in a certain place. This quantitative analysis method from multiple perspectives not only pays attention to the skill level of the elderly themselves.

It also takes into account the social and psychological background factors that support its digital health management. It is helpful to diagnose and solve the difficulties and obstacles faced by the elderly in this field more comprehensively and deeply. This study has innovative analysis ideas and targeted research conclusions. It provides a reference for formulating policies and measures to improve the digital health ability of the elderly, and has high academic value and practical significance.

The structure of this paper is arranged as follows: The first chapter is the introduction, which expounds the research background, current situation and research content; The second chapter introduces the projection pursuit method and the stepwise regression method; The third chapter is the experiment, including data analysis and model construction; The fourth chapter is the conclusion, which analyzes the influencing factors of the digital health ability of the elderly in a certain place; The fifth chapter summarizes the research findings and envisions future research directions.

2. Related Theories

Projection Pursuit (PP) is a method for dimensionality reduction and weighting of high-dimensional data, particularly suited for handling data with complex distributions or significant non-linear structures. The core idea is to use optimization algorithms like simulated annealing to search for the "optimal projection direction" in the high-dimensional space, such that the data features along that direction are most prominent, thereby revealing the inherent relationships between the indicators. For the weighting problem, PP can automatically assign weights to each secondary indicator by optimizing the projection direction, thereby determining the importance of each indicator. First, an projection index function is established. Since this model projects complex high-dimensional data to low-dimensional space, the solving process is to find the

optimal projection direction $\vec{c} = (c(1), c(2), \dots, c(p))$, where c is a unit vector, p is the number of indicators, which represents the weight of each indicator, to obtain the projection value, expressed as:

$$z(i) = \sum_{j=1}^p c(j)x(i, j) \quad (1)$$

The distribution characteristics of the projection value Z are the basis for the construction of the projection index function, which should be scattered as a whole and dense locally, so it can be constructed as:

$$Q(c) = S_z D_z \quad (2)$$

$$\text{s.t.} \begin{cases} S_z = \sqrt{\frac{\sum_{i=1}^n (z(i) - E(z))^2}{n-1}} \\ D_z = \sum \sum_i^n (R - r(i, j)) \times u(R - r(i, j)) \end{cases} \quad (3)$$

$E(z)$ is the mean of the projected values, S_z is the sample standard deviation, R is the window radius of the local density, D_z is the local density of the projection, $r(i, j) = |z(i) - z(j)|$, $u(t)$ is the unit step function, $t \geq 0, u(t) = 1; t < 0, u(t) = 0$.

Then, the projection index function Q is optimized. The change of the projection direction leads to the change of $Q(c)$, so the characteristics of the high-dimensional data are more effectively captured in the low-dimensional space, that the index function $Q(c)$ reaches a maximum.

$$\begin{aligned} \max Q(c) &= S_z * D_z \\ \text{s.t.} \sum_{j=1}^p c^2(j) &= 1, c(j) \geq 0 \end{aligned} \quad (4)$$

The projection value of each sample is obtained by finding the best projection direction \vec{c} and bringing it into the formula of $z(i)$.

Stepwise regression is a variable selection method that builds a regression model by adding variables step by step. In stepwise regression with forward selection, the process begins with a model that contains only the intercept term. A t-test was then performed on all candidate independent variables, and their p-values were calculated to assess their effect on the dependent variable. The independent variable with the smallest p-value (i.e., the highest significance) was selected to be added to the model, and the model parameters were re-estimated. After that, the F test was used to evaluate whether the newly added independent variables significantly improved the explanatory power of the model. This process is repeated until no candidate argument has a p-value less than a set threshold (usually 0.05). The selection process is stopped at this point and the final regression model is obtained and evaluated[9]. The multiple regression model is as follows:

$$Y = \beta_0 + \beta_i x + \varepsilon, i = 1, 2, \dots, p \quad (5)$$

The essence of the step-by-step method is that the selection is made during the process of introducing variables into the model, and once the variables are included in the model, even if its importance decreases in the subsequent analysis, it will not be excluded. The advantage of this approach is the ability to evaluate the impact of each variable on the model in an orderly manner. However, the disadvantage is that it may contain some unimportant variables, which will affect the accuracy and prediction performance of the model. Therefore, in the actual operation, the evaluation and verification of the final model is a crucial link.

3. Experiments

The study collected data by administering questionnaires to the elderly population in a specific region. Perform statistical analysis on the e-HEALS health literacy scale scores collected from the survey data[10].

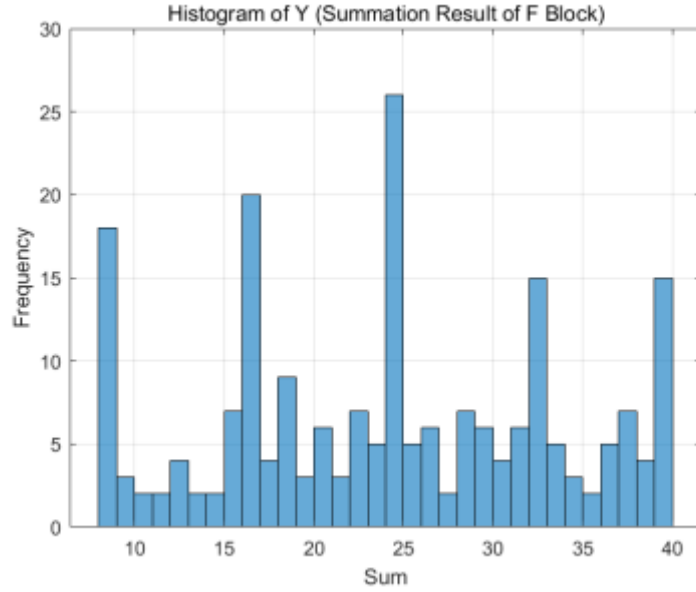


Figure 1: Histogram of the total score distribution of the electronic information literacy scale

Based on the data in Figure 1, Table 1 was obtained:

Table 1: Digital Health Ability Scores

Average value	Median	Mode	25% quantile	75% quantile
23.81	24.00	24.00	16.00	32.00

It can be seen from Table 1 that in the digital health ability test with a total score of 40 points (5 points for each question, 8 questions in total). The average score of the elderly was only 23.81, indicating that the overall performance was close to the passing line ($40 \times 6 = 24$ points). It shows that the average level of digital health ability of the elderly is not optimistic. The distribution of scores showed that more than half of the individuals scored at the passing level, but 25% of the individuals scored below 40% and needed additional support.

The overall score distribution shows a right deviation, indicating that a small number of individuals with high scores have raised the average value, which can be seen from the collected academic data. This phenomenon may imply that a small number of highly educated elderly people in a certain place have a higher level of overall digital health ability. Most of the individual scores are concentrated in the lower range, indicating that although a certain place is a cultural center, science and technology are highly developed. However, there are still some people with low digital ability, which suggests that we need to pay attention to and improve the digital health ability of low-score groups. In order to narrow the ability gap

and seek the health and well-being of all the elderly.

Then the projection pursuit algorithm based on simulated annealing optimization is used to explore the importance of the factors affecting the digital health ability of the elderly. The target variable was the numerical health ability score. The expression is as follows:

$$Y = \sum_{i=1}^8 Fi \quad (6)$$

Where Y represents the numeric e-HEALS ability score and Fi represents the score for each entry on the E-HEALS scale. So the prior matrix is: $Y = (Y_1, Y_2, \dots, Y_{215})$.

After sorting, the design matrix is defined as $X_i = (A_{11}, A_{12}, \dots, A_{1,29}, A_2, A_3, A_4, A_5, A_6, A_7)$, where, A_{1j} represents the variables coded with labels, $j=1,2,\dots,29$. $A_2, A_3, A_4, A_5, A_6, A_7$ represented gender, occupation, marital status, living conditions, diagnosed chronic diseases and physical discomfort symptoms.

Therefore, the overall design matrix of the sample is: $X = (X_1, X_2, \dots, X_{215})$

The projection results are as follows:

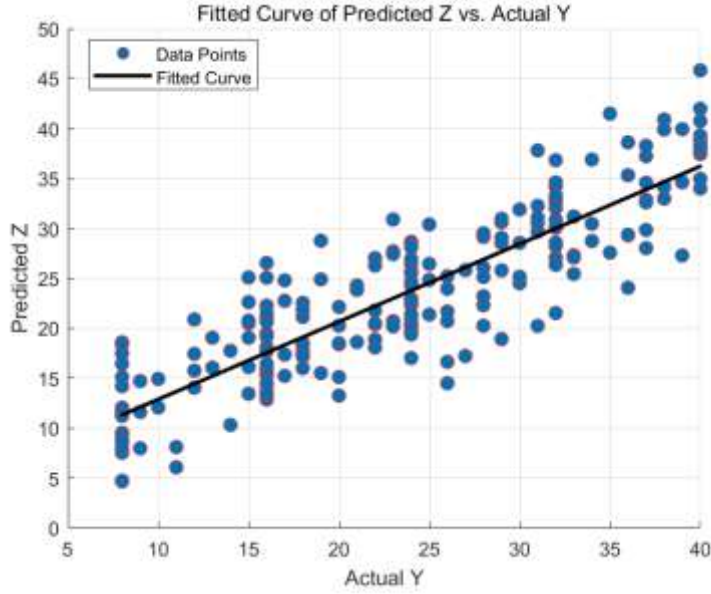


Figure 2: projection result and true value

Figure 2 shows that most of the data points are close to the 45-degree slope, indicating that the model predicts well in general. However, there are still some data points deviated, suggesting that the model has limitations under specific circumstances and needs further optimization. The scatter of the data points indicates the possible presence of unaccounted for factors or noise and the lack of a clear linear or nonlinear pattern.

Due to the exploratory study on the influencing factors of the digital health ability of the elderly based on the projection pursuit algorithm, the results are not satisfactory. Next, the

$$\sum_{i=1}^8 Fi = Y = \beta_0 + \beta_1 X_C + \beta_2 X_D + \beta_3 X_E + \beta_4 X_G + \beta_5 X_H + \varepsilon \quad (7)$$

Among them, $\sum_{i=1}^8 Fi$ represents the total score of 8 items of e-HEALS scale, X_C, X_D, X_E, X_G, X_H represent the scores on the scales corresponding to the secondary indicators of psychological health level, family and friends support, community assistance, technological anxiety, and privacy issues, respectively.

β_0 is the intercept term, which is the only parameter of the initialization model; $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are the regression coefficients of X_C, X_D, X_E, X_G, X_H , respectively; ε is the random error.

Through the above research and analysis, a preliminary conclusion about the influencing factors of digital health ability of the elderly group in a certain place is drawn. As shown below.

stepwise regression model was used for the study.

Stepwise regression analysis was used to screen out the factors that had significant effects on the digital health ability of the elderly from multiple potential variables. To help improve the digital health capacity of the elderly to meet the digital challenges.

Firstly, all the variables were included in the model, and then through the stepwise regression method, the significance level was set at 0.05. Variables with $p > 0.05$ will be eliminated, and the stepwise regression formula will be constructed as follows:

4. Results

The results of Table 2 are obtained by stepwise regression fitting of the data of the digital health ability of the elderly in a certain place and its influencing factors.

Table 2: Stepwise Linear Regression Results

Variables	Regression Coefficient	T-value	P-value
Constant term	31.7018	11.453	0.000
X_D	1.3308	2.589	0.010
X_E	1.6609	4.688	0.000
X_G	-5.6995	-13.994	0.000

Therefore, the regression equation obtained by the final model is:

$$y = 31.7018 + 1.3308X_D + 1.6609X_E - 5.6995X_G + \varepsilon \quad (8)$$

Table 3: Model Evaluation Results

F value	P-value	AIC	R ²	Adjusted R ²	MSE
173.0	0.000	1303	0.712	0.708	24.8701

R² can be seen from Table 3 that it reaches 0.712, so it is considered that the model has a high degree of fitting and good explanatory power.

Through the above research, the following conclusions can be drawn:

The regression results of the model remove the two independent variables of social psychological factors and privacy issues. These two factors have no significant effect on the dependent variable. However, although psychosocial factors and privacy issues are not the main factors in the

current sample, Doesn't mean mental health and privacy issues aren't important. On the contrary, it suggests that more attention should be paid to the psychological feelings of the elderly, and privacy education for the elderly should be strengthened. To increase their awareness and ability to deal with privacy issues.

According to the sign and magnitude of regression coefficient of each variable, it can be concluded that the higher the score of the influencing factors of interpersonal support and community support. The higher the score of digital health ability, that is, they have a positive impact on the score of digital health ability. The positive influence degree of the two factors is close; the higher the score of technology anxiety factors, the lower the score of digital health ability. It shows that it has a negative impact on the score of digital health ability. Overall, among all the independent variables, technology anxiety has the greatest impact on the dependent variable. And has a negative effect on that digital health ability score.

Through trust and dependence, emotional support and experience sharing, interpersonal support plays an important role in improving personal digital health ability. A positive effect, the sense of community support provides rich digital health information, enhances the sense of belonging and identity, and promotes mutual assistance and feedback mechanism. It effectively improves the personal digital health ability. However, technology anxiety has a significant negative impact on the score of digital health ability. It is the most influential factor of all independent variables. Technological anxiety stems from the unfamiliarity and uncertainty of new technologies. It leads to fear and anxiety, cognitive impairment and behavioral inhibition, which limits the improvement of their digital health ability.

5. Conclusions

In this study, the digital health ability of the elderly in a certain place and its influencing factors were analyzed. It was found that interpersonal support and community support had a positive impact on the digital health ability of the elderly. However, technology anxiety has a negative impact, and the degree of impact is the largest. Studies have shown that positive interpersonal relationships and community support can enhance the confidence of older people in using digital technology. Technological anxiety, on the other hand, hinders their ability to improve. In the future, we can improve the

digital health skills of the elderly through education and training, and pay more attention to their mental health. And popularize knowledge of privacy protection. In this way, the elderly will be able to use digital technology more confidently and safely, and better integrate into the digital society. Enjoy the convenience and opportunities brought by the digital age.

References

- [1] Jia Mengxia. Research on Improvement Countermeasures of Digital Health Service Platform of HJ Company under the Background of Intelligent Old-age Care [D]. Zhejiang University of Science and Technology, 2022.
- [2] Xia Xue. Research on Industrial Ecology of Sports Digital Therapy in China under Digital Enabling [D]. Wuhan Institute of Physical Education, 2023.
- [3] Hu Rong. Research on the Impact of Digital Empowerment on Medical Service Satisfaction [D]. Guizhou University, 2023.
- [4] Lv Hekai. Study on the application barriers and influencing factors of home digital health technology for the elderly [D]. Bengbu Medical College, 2023.
- [5] Xie Yongfei, Liu Yifeng. The Digital Divide of the Elderly from the Perspective of Healthy Aging: Causes and Governance [J]. Journal of Minzu University of China (Philosophy and Social Sciences Edition), 2024, 51 (05): 125-134.
- [6] Liu Jingliang, Wang Wenkai, Han Jieru, et al. Research on Digital Divide and Countermeasures of Online Health Service for the Elderly in the Digital Medical Age [J]. Journal of Library and Information Science, 2024, 9 (09): 41-45.
- [7] Chen Yunhao, Wang Hansong, yuan Jiajun, et al. Study on the status and influencing factors of digital health technology anxiety in elderly patients with chronic diseases [J]. Journal of Medical Informatics, 2024, 45 (11): 9-14 + 21.
- [8] Wan Hui, Peng Jun, Wen Hongxia, et al. Current situation and influencing factors of digital health literacy of the elderly in Shanghai [J]. Journal of Medical Informatics, 2024, 45 (10): 53-58.
- [9] Analysis of influencing factors of carbon emission in construction industry based on stepwise regression [J]. Journal of Engineering Management, 2021, 35 (02): 16-21.
- [10] Fu Yi. Research on electronic health literacy of the elderly in Guangdong Province based on eHEALS [J]. Library Research, 2024, 54 (01): 93-101.