



# Analysis of Influencing Factors and Suggestions for Improvement of Women's Training Willingness Under the Background of Rural Revitalization Strategy

Ke Zhao<sup>1,a</sup>, Yanan Miao<sup>1,b</sup>, Ying Tu<sup>2,c</sup>, Liyuan Cai<sup>2,d\*</sup>

<sup>1</sup>Chengdu Technological University, Department of Network & Communication Engineering, Chengdu, China

<sup>2</sup>Chengdu Technological University, Department of Big Data and Artificial Intelligence, Chengdu, China

<sup>a</sup>38024084@qq.com, <sup>b</sup>miaoyanan33@163.com, <sup>c</sup>2040775112@qq.com, <sup>d\*</sup>8415708@qq.com

**Abstract.** Starting from the implementation of rural revitalization strategy, the main factors affecting the training willingness of rural women are verified by proposing research hypotheses, investigating, and collecting on relevant data, formulating scales, and adopting Lasso regression principal component analysis to analyze the correlation between major factors. The results show that the factors affecting rural women's willingness to receive training are the main types of household income sources, balance of household expenses, leisure time management, participation in training and types of training information. Some suggestions to improve the training willingness of rural women are proposed in this paper.

**Keywords:** Rural women; Training willingness; Influencing factors; Path for improvement

## 1 Introduction

It is necessary to prioritize the rural human capital and strengthen the supporting role of talents in rural revitalization to implement the rural revitalization strategy<sup>[1]</sup>. Rural women have always been an important subject and power in promoting rural revitalization<sup>[2]</sup>. Therefore, it is of great significance to develop the human resources of rural local women, strengthen the training work, take the demand as guidance, explore the influencing factors of rural women's training willingness, and improve the training effect to promote the overall implementation of the rural revitalization strategy<sup>[3]</sup>.

At present, there are several related research on the issue of farmers' training at home and abroad. For example, increasing the investment in human capital of farmers can improve agricultural production efficiency, achieve an increase in per capita income of farmers<sup>[4]</sup> and promote the prosperity of rural industries through farmer training<sup>[5]</sup>. There are supply and demand differences in training investment, training

methods, training content, and training teachers for farmer training<sup>[6, 7]</sup>. The rural practical talent demonstration training base has better cultivation experience and mode<sup>[8]</sup>. Training tracking service, government and family support for training will affect the training effect<sup>[9, 10]</sup>. However, these discussions and research are mostly analyzed by SPSS statistical software, and its practicality needs to be enhanced. The above research subjects are mostly rural farmers, including quantitative analysis with a sufficient number of samples. However, there few empirical research methods are used to quantitatively explore and verify the impact of various factors on the training willingness of rural local women from a group perspective. This paper analyzes the influencing factors of rural women's training willingness from an empirical perspective by establishing a Lasso regression model, clarifies the composition and interrelationships of influencing factors, and explores effective paths and methods to enhance rural women's training willingness and improve training effectiveness.

## 2 Research Hypothesis

This paper proposes a hypothesis on the impact of local women's training willingness in rural areas through collecting the relevant literature at home and abroad, and combining with the actual situation of rural training.

(1) The main types of household income sources, such as aquaculture or other industrial production, are considered to be the direct influencing factors of rural women's training needs and the type of demand information is determined which greatly affects the supply and demand of training content. Proposing hypothesis 1: Main household income source has a positive and significant impact on the willingness of local rural women to receive training.

(2) The balance of household expenses, such as other balances in addition to input of the means of production and daily living expenses, are considered to have higher annual household income. Women tend to be more motivated and actively involved in training, which is considered to have a positive impact on rural women's training motivation. Proposing hypothesis 2: Balance of household expense has a positive and significant impact on the willingness of local rural women to receive training.

(3) The management of leisure time, such as watching TV, playing mahjong, etc. are considered to be the objective condition related to the training willingness. Only when rural women have leisure time after work can they have the need to pursue higher level training. Proposing hypothesis 3: Leisure time management has a positive and significant impact on the willingness of local rural women to receive training.

(4) The participation in training, if there is no organized training or training organization in the location, it will directly affect the decision of rural women to participate in training. The lack of demand and supply of training will stimulate the willingness to participate in training. Proposing hypothesis 4: Participation in training has a positive and significant impact on the willingness of local rural women to receive training.

(5) The types of training information, such as the agricultural product sales, migrant workers' rights and interests protection, children's education and other aspects of training information, determine whether to meet the real needs of rural women. The

more accurate the information is, the more effective the training is. Proposing hypothesis 5: Training information types have a positive and significant impact on the willingness of local rural women to receive training.

### **3 Empirical Research and Analysis**

#### **3.1 Questionnaire Distribution**

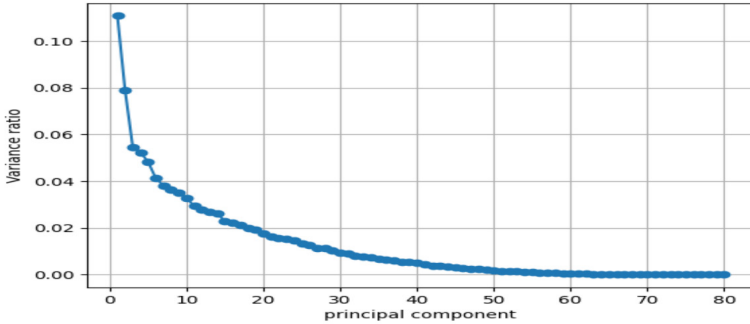
The project team conducted an empirical research on the age, educational level, marital status, family economic status, production labor, and social environment of rural women from a demographic perspective. "Questionnaire of Rural Women in Sichuan Province" was compiled and filled out by rural women themselves. 130 questionnaires were distributed and 117 were recovered, with a recovery rate of 90 %. Among them, 26 questionnaires were invalid and 91 questionnaires were valid, with an effective rate of 77.8 %. In this analysis, only the question group information related to rural women's willingness to participate in training is adopted. The age of subjects are as follows: under 25 years old, 26-35 years old, 36-45 years old, over 46 years old. The survey shows that women over 25 years old are the main force of rural revitalization.

#### **3.2 Data Pre-processing**

The question group related to the research was retrieved, with a total of 66 independent variables and 1 dependent variable. We have removed irrelevant and duplicate data from the original data set, smoothed noisy data, filtered out data unrelated to mining topic, and processed missing values and outliers, etc. And a multicollinearity test was also conducted, and it was found that there was a certain degree of multicollinearity between the data.

#### **3.3 Principal Component Analysis**

According to the above assumptions and related questionnaire data, the SPSS software was used to draw the principal component analysis model of rural women's training willingness, the contribution of eigenvalues was obtained and the principal component variance ratio graph was drawn. As shown in Figure 1.



**Fig. 1.** Principal component variance ratio graph (Import questionnaire data into SPSS software to obtain the graphic)

It is found that the interpretation of the data can be increased to 85% by extracting 26 principal components including age, household income source, training status, physical and mental health. However, the proportion of the first three components in the results is rather high, and the proportion of the latter components is extremely flat. Therefore, the principal component model analysis is not used in this study.

### 3.4 Lasso Regression Model

#### *Choose Lasso Regression Algorithm*

Since the poor performance of principal component analysis and considering the possibility of multicollinearity, autocorrelation, and heteroscedasticity in the data, Lasso regression is adopted in this paper.

The Lasso regression formula can be obtained as follow

$$\min \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2 + \lambda \sum_{j=1}^p |\beta_j| \tag{1}$$

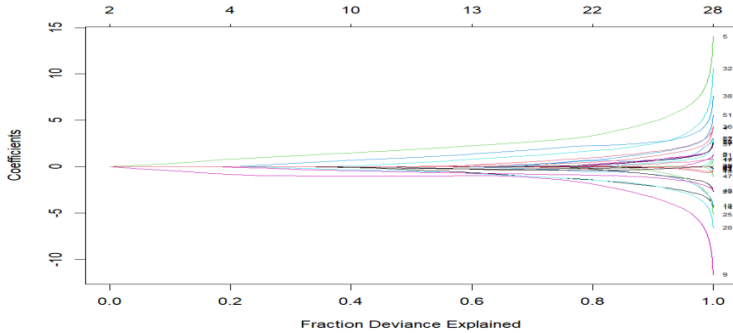
Where,  $x_{ip}$  is the  $p$ -th characteristic variable of  $i$ -th problem;  $y_i$  is the target variable of  $i$ -th problem;  $\beta$  is the regression coefficients for each characteristic variable;  $\lambda$  is a parameter of  $L_1$  regularization term.

The determination of  $\lambda$  in the whole model is the most critical, and the final effect of the whole model will be directly affected by the value of  $\lambda$ .

#### *Determination of Regularization Term Parameters*

Firstly, Python's lasso function is used to draw the path of coefficient variation with the regularization term. It can be seen from Figure 3. Figure 3 shows the regression coefficient ( $y$  axis) of each variable and the path of coefficient variation with the  $L_1$  regularization term ( $x$  axis). The larger the  $L_1$  regularization term  $\sum_{j=1}^p |\beta_j|$  is, the smaller the  $\lambda$  is to minimize the formula as a whole. In fact, with the gradual decrease

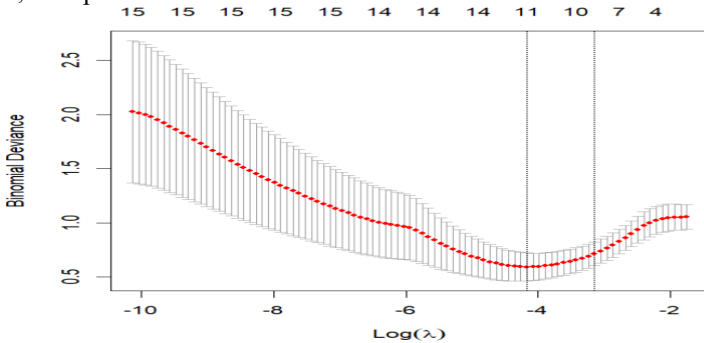
of  $\lambda$ , more and more regression coefficients of characteristic variables are gradually zeroed (more and more characteristic variables enter the model). As shown in Figure 2, variables 5, 32, 38, 9, 20 are introduced in the early stages of regularization term variations. The contained information is strongly correlated with the target variable and plays a crucial role in explaining the variations in the target variables.



**Fig. 2.** Path of coefficient variation with regularization term (Regularize the questionnaire data using Lasso function in Python)

**Cross-validation Estimation Parameter  $\lambda$**

This paper adopts the cross-validation to estimate the adjustment parameters. The above figure shows a 10-fold cross-validation curve changing with  $\log(\lambda)$ . Each red dot represents the mean squared error (MSE) of a 10-fold cross-validation, and the length of its upper and lower error bars represents the standard deviation of 10 MSEs in each cross-validation. The figure helps us to understand the average level of model performance and the range of performance variations. The first of the two vertical dashed lines in the figure is the value that minimizes the cross-validation error. The second is to give the  $\log(\lambda)$  value of the most regularized ( $\lambda$  as large as possible) model when the cross-validation error is kept within a standard deviation range of the minimum error. This paper adopts  $\lambda$  to minimize the cross-validation error. After calculation, the optimal value can be determined to be 0.030. It is shown in Figure 3.



**Fig. 3.** Cross-validation estimation parameter  $\lambda$  (Estimate the questionnaire data using Lasso function in Python)

**Lasso Regression Results and Tests**

Each variable coefficient of Lasso regression can be obtained by using the optimal  $\lambda$ . The regression results are shown in Tables 1 and 2.

**Table 1.** Lasso regression results (Data were obtained through regression using questionnaire data)

(Intercept)	Q6_2	Q15_4	Q13_6	Q6_1	Q15_2
3.372704	2.264617	1.3952	0.821233	0.198423	0.130605

The results in Table 1 show a positive correlation between the corresponding features and the target variables. Among the variables selected by Lasso, Q6\_2, Q15\_4, and Q13\_6 have a particularly significant impact on Y.

**Table 2.** Lasso regression results (Data were obtained through regression using questionnaire data)

Q6_6	Q12_3	Q15_6	Q7_5	Q13_8	Q16	Q12_1	Q1
0.980	0.770	0.706	0.698	0.289	0.258	0.058	0.028

The results in Table 2 show a negative correlation between the corresponding features and the target variables. Among the variables selected by Lasso, Q6\_6, Q12\_3, Q15\_6 and Q7\_5 have a particularly significant impact on Y. The main statistical variables obtained are shown in Table 3.

**Table 3.** Description of statistical variables (Custom questionnaire variables)

Symbols	Statistical variables
Q1	Age of respondents
Q6_1	Planting as the main source of household income
Q6_2	Aquaculture as the main source of household income
Q6_6	Other industrial production as the main source of household income
Q7_5	Balance in the main household expense
Q12_1	Leisure time of the respondents spent is watching TV
Q12_3	Leisure time of the respondents spent is playing mahjong
Q13_6	No such organization in the location of respondents
Q13_8	No participation in organized training for respondents
Q15_2	Hope to receive training on agricultural product sales for respondents
Q15_4	Hope to receive training on migrant workers' rights and interests protection for respondents
Q15_6	Hope to receive training on children's education for respondents
Q16	Labor hours of personal agricultural and household for respondents

## 4 Results and Analysis

The results show that the assumed key variables are all in the optimal combination of predictors. The rural women's willingness to receive training has the most significant positive correlation with the main household income sources. Especially in the aquaculture as the main source of income, the Lasso regression coefficient is 2.265. There is a positive correlation between the rural women's willingness to receive training and the type of information provided by training, and the Lasso regression coefficient is 1.395. There is a positive correlation between the rural women's willingness to receive training and the training status that the respondents have participated in. The willingness to participate in training is stronger for those who have not participated in similar training, and the Lasso regression coefficient is 0.821. There is a positive correlation between the rural women's willingness to receive training and the leisure time management status. The willingness to participate in training is stronger for those respondents spent their leisure time playing mahjong, and the Lasso regression coefficient is 0.770. There is a positive correlation between the rural women's willingness to receive training and the balance in the main household expense, and the Lasso regression coefficient is 0.698. And there is only a weak correlation between the other variables. At the same time, there is a certain correlation between the characteristics of the influencing factors of rural women's training willingness, that is, there is a multicollinearity among the variables. Therefore, it is necessary to screen the variables and select the features with high scoring proportions.

The top 5 factors strongly correlated with the willingness of rural women to receive training are shown in Table 4.

**Table 4.** Top 5 strongly correlated factors affecting the willingness of rural women to receive training (Data were obtained through regression using questionnaire data)

Influencing factors of rural women's training willingness	Lasso regression coefficient
Types of household income sources	2.265
Types of training information	1.395
Status of attended training	0.821
Status of leisure time management	0.770
Balance of household expense	0.698

It can be seen that the willingness of rural women to participate in training is strongly correlated with the main types of household income sources through Pearson correlation coefficient analysis. Especially in households dominated by aquaculture, women have a strong willingness to receive training. It can bring benefits to household production if women attach great importance to training and better stimulate training willingness and motivation. There is a strong correlation between the rural women's willingness to receive training and the type of information provided by training. Rural women need more training information closely related to rights and interests protection, agricultural product sales, children's education, etc. They have a strong willingness to participate in such training. There is a strong correlation be-

tween the rural women's willingness to receive training and the training status. Those who have never participated in the training have a strong willingness to participate in learning and training and put forward demands for participation. There is a strong correlation between the rural women's willingness to receive training and the leisure time management status. The willingness to participate in training is stronger for those women spent their leisure time watching TV or playing mahjong. There is a positive correlation between the rural women's willingness to receive training and the balance in the main household expense. Those who have more annual household income balances are more likely to participate in training. And there is only a weak correlation between the other variables. In addition, it is known that women's willingness to participate in training is also related to the organizers of training, training teachers, teaching methods, etc. from the interviews during the research period. There is also a certain correlation between individual characteristics of women, such as the age and education level.

All the hypotheses of this study have been supported after investigation, test and analysis, which better clarifies the influencing factors and related functions of rural women's training willingness.

## 5 Conclusions

There are 66 statistical variables including the types of household income sources, balances of household expense, status of leisure time management, etc. are designed in this paper based on the previous research. Through the investigation and collection of rural women's training willingness data, Lasso regression principal component analysis is used to empirically verify the correlation between the main factors. Finally, the main factors affecting the willingness of rural women to receive training were verified, and 5 main factors were obtained. This paper explored the effective ways to improve the training willingness of rural women. According to the conclusion, it is suggested to improve the training willingness of rural women from the following aspects.

1. Construct a modular curriculum system. According to the development of the local rural revitalization industry, the course content is set up, embedded on-site inspections and simulation training, etc., and allow the students to choose their own learning modules independently.

2. Increase legal information such as children's education, rights and interests protection, and physical and mental health training content to meet the actual needs of rural women.

3. Combine all social forces to form a joint entity. The government, training institutions, and social forces jointly participate in the organization. The universities, training institutions, and farmer associations are encouraged to undertake training, integrate and share training resources.

4. Optimize the social environment through policy formulation. Women's agriculture and side business related to production and living tasks need to be reduced, and



their innovation and entrepreneurship ability to increase income also needs to be improved to promote comprehensive development.

## Acknowledgment

Soft Science Project of Science & Technology Department of Sichuan Province: Research on Comprehensive Development Strategies for Rural Left-Behind Women in Poverty Alleviation and Development in Sichuan Province. (Project No.2016ZR0179)

## References

1. Zhang, J.Y. , Chen, J. (2021) Strengthen Effective Supply of Rural Talent Support and Realize Effective Connection of Poverty Alleviation and Rural Revitalization. *Macroeconomic Management*, 8: 54-60.
2. Liu, T. (2022) Value Certification, Realization Difficulties and Countermeasures of Local Female Talents under Rural Revitalization. *Rural Economy*, 9: 72-74.
3. Mo, C.W., Ma, J.F. (2021) Dilemma and Countermeasures of Vocational Education and Training for Rural Left-Behind Women Under the Background of Rural Revitalization. *Jiangsu Education*, 64: 40-44.
4. Wang, W.B., Zhang, Y.Y. (2020) Agricultural Productivity, Factor, and Rural Household Income--Theoretical and Empirical Research based on CHIP Data. *Finance and Trade Research*, 31: 61-74.
5. Li, W., Yang, S.Y., Lan, D.X. (2020) Difference Between Supply and Demand of New Professional Farmers' Training and Influence on Training Effect. *Rural Economy*, 2: 138-144.
6. Zhai, S.X., Peng, C. (2024) Can Training Increase Farmers' Income: An Empirical Study Based on National Rural Fixed Observation Point Data. *Journal of Huazhong Agricultural University (Social Sciences Edition)*, 2: 108-121.
7. Wang, Y.L., Xiang J., Liu, S.L. (2020) Research on Supply and Demand Influencing Factors of Farmers Participating in Government Organized Agricultural Technology Training. *Chinese Agricultural Bulletin*, 36: 144-151.
8. Jiang, T.J., Jiang, Y.X. (2023) Explore a New Model for Cultivating Rural Practical Talents-- Based on the Practice of Demonstration Training for Rural Practical Talents in Hunan Province. *Farmer Technology Training*, 6: 43-46.
9. Liu, D. (2022) Problems and Countermeasures of High-Quality Farmers Training Needs Analysis. *Farmers' Technology Training*, 19: 55-56.
10. Chen, X.Y., Yang, J.X., Zhu, Y.R. (2017) Research on Effectiveness and Influencing Factors of Practical Talent Training in Rural Areas. *Rural Economy*, 11: 108-113.

**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.

