Proceedings of the 2nd International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/39/20240544

The research on influence factors related to depression in rural zones

Jianing Sun

School of Foreign Language, Peking University, Beijing, 100000, China

2300018106@pku.edu.cn

Abstract. Depression is famous for its capacity to make devastating impact on people's physical and mental well-being and sense of happiness. Many researches have tried to identify different factors that function on different groups of people. In this research, the method Binary Logistic Model is used to deal with the data from a study about the life condition of people in rural zones, and was first published online in 2019 and compiled in 2020, with 1429 individuals as its samples. It is concluded that although depression has no connection with sex, marriage, number of children, members in family, gained asset, durable asset, saved asset, living expenses, other expenses, income from salary, income from farm, income from business, income from non-business, income from agriculture, farm expenses, labor primary, lasting investment, non-lasting investment, it has a relatively strong connection with age and education level. This paper focuses on people in rural zones, and concerns 20 factors that cover many aspects of their lives simultaneously, which hopes to help further study of depression.

Keywords: Depression, rural zones binary logistic model.

1. Introduction

As a sort of disease that has aroused attention all around the world in recent years, depression can make worrisome and even horrifying influence on the recognition of people to their own lives. Generally speaking, adolescent depression patients with non-suicidal self-injury are experiencing negative meaning of life, unable to find a worthy reason for living [1]. Besides making negative impact singularly, depression also acts as a mediation that can intensify the trauma or harm caused by other factors. Li et al. concluded that for nurses with work stress, depressive symptoms would play the role of mediation and increase the secondary trauma [2]. Moreover, combined with other diseases, depression can remarkably reduce the quality of life. Parkinson's disease patients accompanied by depression are more affected in aspects including physical activity, mental health, social support and so on, compared with those who have no depressive symptoms [3]. Therefore, to prevent and cure depression is the key to decrease trauma triggered by it, to avoid deteriorating the consequence of other diseases, and to protect the happiness and health of human beings.

In order to better recognize depression, many researches have been dedicated to this area in the past several decades. It has been found that dysfunctional cognitions, stress in life, maternal depression, interpersonal dysfunction, and being female are all robust depression risk factors [4]. Other scholars studied factors associated with depression among a certain group of people. The depression and anxiety symptoms of medical students are peculiarly linked with school location and tuition scholarship [5].

While for breast cancer patients, depression is associated with being rural resident, non-Orthodox Christian and suffering for extend symptom [6]. However, for the former sort of research, many factors are so hard to define and judge, that in daily life they cannot serve as useful predictors of depression. To solve such problem, this essay would focus on the clear and universal indexes, and explore whether there exists a relationship between those indexes and depression. And many researches like the latter one focus on a specific group of people, but seldom does them consider people in rural areas, and this paper aims to fill the blank.

To search the answer to the general question—the factors related with depression, many methods have been used by precedent scholars in specific fields. Generally speaking, logistic regression is most commonly adopted. Mao et al. used logistic regression to analyze influencing factors, and receiver operating characteristic curve (ROC) to evaluate the its efficiency [7]. Hu also adopted the model, though he preprocessed the data, and further used restricted cubic spline models to assess more accurate relationship between factors and depression [8]. Liu et al. used unifactorial and multifactorial logistic regression and the chi-square automatic interaction detection (CHAID) to construct a correlation decision tree [9]. But other models are also used, in rarer studies. Huang categorized depressive symptoms into two dimensions, then applied univariable and multivariable zero-inflate negative Poisson regression (ZINB), with those symptoms used as a dependent variable [10]. It should be noticed that depression is an area in which many studies as mentioned above have been conducted in the past; and therefore, in recent years, systematically reviewing those studies and summarizing them has become a widely used way of research. For example, after thoroughly screening relevant articles, Razzak et al. selected 14 articles, refined and concluded their results, to offer a more comprehensive sight toward this topic [11].

In summary, depression is a disease that should, and has been concerned. Though many studies have contributed to this direction, this article would differ from them by exploring the relationship between clear factors and depression for people in rural zones.

2. Methodology

2.1. Data source

The data used in this article comes from a dataset downloaded from Kaggle, an online data science community. It was originally published in 2019 by a user named Frank, and a year later compiled by Diego Babativa, a data scientist from Colombia. The data was the results of a study about the life conditions of people who live in rural zones.

2.2. Variable selection

The data counts a total of 1429 people in 292 different villages. It is basically an all-round investigation about the lives of people that covers many fundamental information and important aspects. Variables include sex, age, marriage, number of children, members in family, educational level. Other variables concerning their economic condition are divided especially specifically. Asset (gained, durable and saved), expenses (for living, farm and others), investment (lasting and non-lasting) are all recorded respectively. In terms of their income, researchers surveyed whether they are from salary, one's own farm, agriculture, business or non-business. They also concerned whether labor is primary in the farms of rural people. The whole structure of variables is shown in Figure 1 below. In a word, the data are gathered under the principle of knowing more about both the family and the career condition of rural people, with an emphasis on economy.

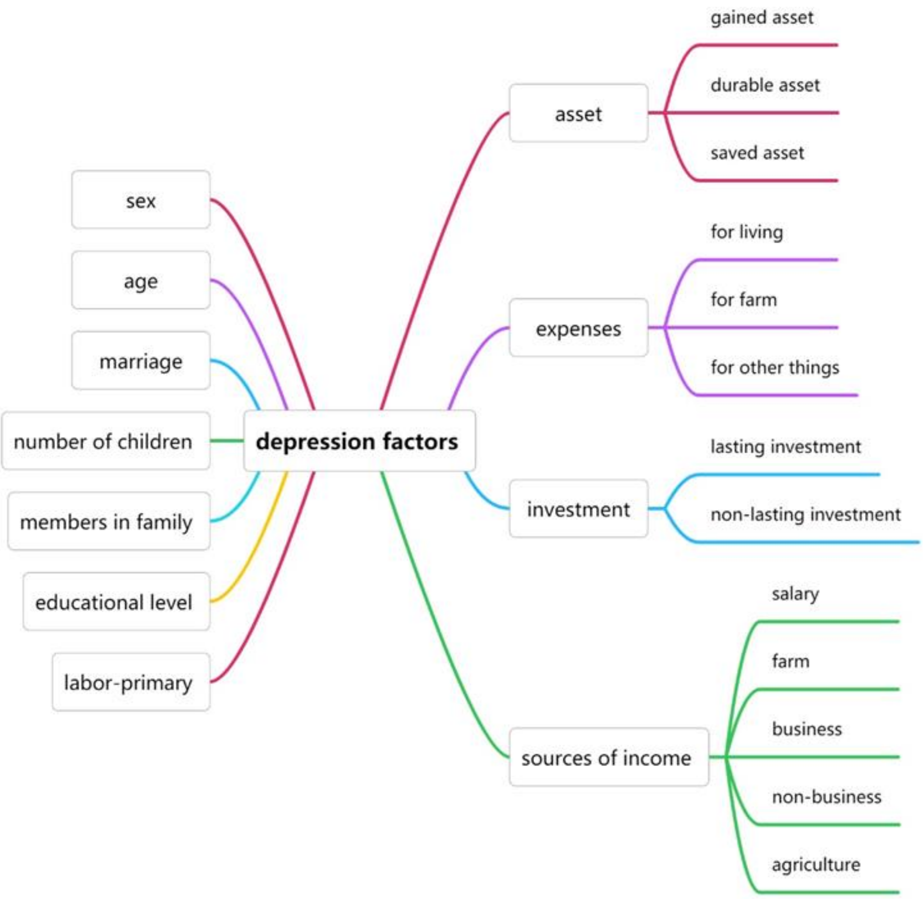

Figure 1. Schematic of the structure of variables.

2.3. *Research protocol*

This paper uses the dependent variable (Y) as whether or not the person suffers from depression, and the 20 factors are the independent variables (X). For the dependent variable and several independent variables, among which marriage, sources of income, whether or not labor-primary, 0 represents no and 1 represents yes. For sex, 0 represents male and 1 represents female. This paper analyzes the relationship between the effect of X on Y, i.e., the relationship between the 20 factors and depression. As there are 20 variables in total, analysis of variance is firstly used to make initial exploration of the relationship between X and Y, so as to reject irrelevant X and simplify the model. Next, this paper uses the Binary Logistic Regression model to further analyze the data.

2.4. *Model principle*

Binary Logistic Model is used to analyze the influence of quantitative or categorical data X to categorical data Y. The model is built in this way:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m \tag{1}$$

Where p represents the probability of Y happens (the value is 1), and (1-p) represents the probability of Y does not happen (the value is 0). When using this model, if the original data is 1, then the predicative value of model fitting would try its best to approximate 1; if the original data is 0, then the predicative value of model fitting would try its best to approximate 0.

3. Results and discussion

3.1. Data presentation

Figure 2 and 3 serve as two examples to display how both quantitative and categorical data are collected. From Figure 2, it can be seen that most samples age about 25 to 30; from Figure 3, it can be seen that female (represented by 1) consists of the majority of these samples.

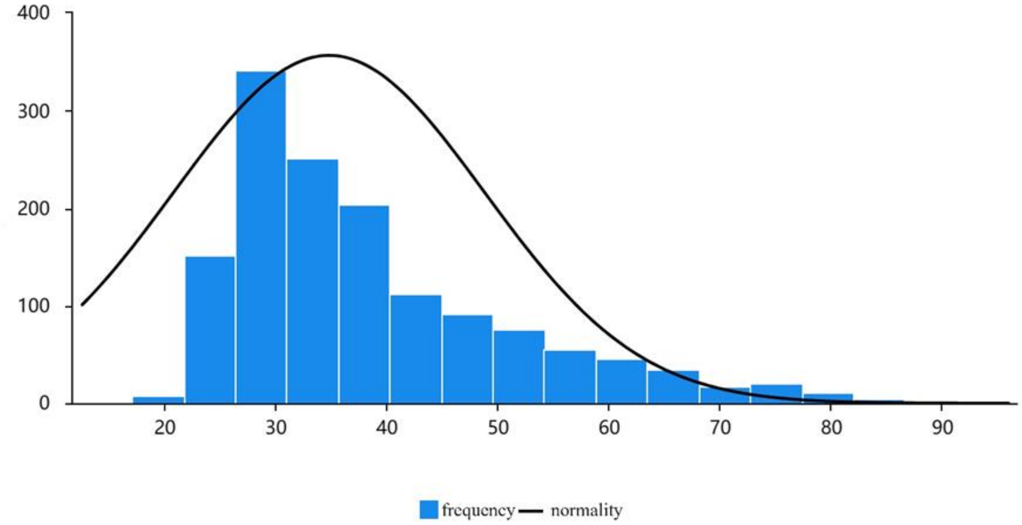

Figure 2. Histogram of age data.

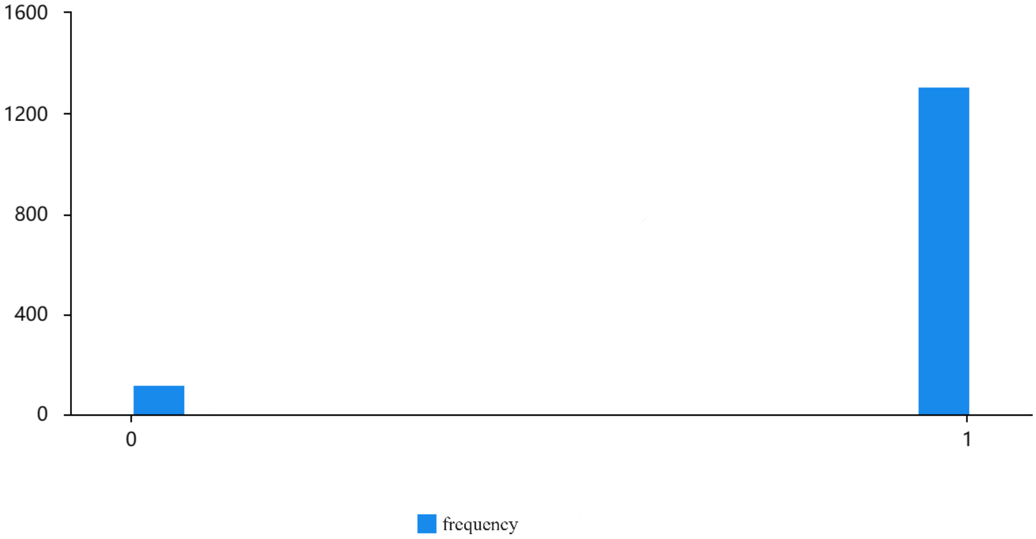

Figure 3. Histogram of sex data.

3.2. Factor analysis of variance

Since there are 20 factors concerned in the study, it is easy to think about that to directly explore their relationship with depression would be really tricky, as they might influence on one another, and make the model complicated, if not ineffective. To deal with the problem, factor analysis of variance is used, and the result is shown in Table 1 below.

Table 1. The result of analysis of variance.

Element	p	F
Sex	0.894	0.018
Marriage	0.019*	5.534
Number of children	0.885	0.021
Education level	0.000**	13.850
Members in family	0.185	1.756
Age	0.000**	16.130
Gained asset	0.868	0.028
Durable asset	0.126	2.345
Saved asset	0.732	0.017
Living expenses	0.287	1.137
Other expenses	0.518	0.418
Income from salary	0.882	0.022
Income from farm	0.619	0.247
Income from business	0.287	1.132
Income from non-business	0.335	0.928
Income from agriculture	0.470	0.523
Farm expenses	0.853	0.034
Labor-primary	0.628	0.235
Lasting investment	0.876	0.024
Non-lasting investment	0.051	3.811

* p<0.05 ** p<0.01

It can be seen that different samples in depression show no significance (p>0.05) to 17 factors, viz., sex, number of children, members in family, gained asset, durable asset, saved asset, living expenses, other expenses, income from salary, income from farm, income from business, income from non-business, income from agriculture, farm expenses, labor-primary, lasting investment, non-lasting investment. This means that for different samples in depression, these 17 factors all show consistency, without any differences. On the other hand, different samples show significance (p<0.05) to marriage, age and education level, which means that they would cause discrepancy among different samples.

3.3. Model testing

It can be seen from Table 2 that as model fitting quality is judged by model prediction accuracy. The overall prediction accuracy of the research model is 83.34%, and the model fitting is acceptable.

Table 2. Binary probit regression prediction accuracy summary.

		Predicted value		Forecast accuracy	Prediction error rate
		0	1		
True value	0	1191	0	100.00%	0.00%
	1	238	0	0.00%	100.00%
	summary			83.34%	16.66%

3.4. Binary logistic regression

For the reason mentioned above, next, with marriage, age and educational level as independent variables and depression as the dependent variable, the Binary Logistic Regression model is used to analyze them.

Table 3. Likelihood ratio test results of binary Logit regression model.

model	-2x log likelihood value	Chi-square	df	p	AIC	BIC
Intercept	1287.167					
Final model	1266.888	20.279	3	0.000	1274.888	1295.947

The null hypothesis for model testing here is that the quality of the model remains the same regardless of whether the independent variables (marriage, age, educational level) are included (Table 3). The p-value in Table 3 is less than 0.05, indicating a rejection of the null hypothesis, i.e., the independent variables included in this model construction are effective and meaningful. The model uses this equation:

$$\ln\frac{p}{1-p} = -1.509 + 0.012 * \text{Age} - 0.109 * \text{Marriage} - 0.055 * \text{educationlevel} \tag{2}$$

Where p represents the probability of depression being 1, while 1-p represents the probability of depression being 0.

Table 4. Summary of Binary Logit Regression Analysis Results.

elements	Regression coefficient	Standard error	z value	Wald χ2	p value	OR value	95% CI
Age	0.012	0.006	2.228	4.964	0.026	1.012	1.001 ~ 1.024
Marriage	-0.109	0.179	-0.608	0.370	0.543	0.897	0.632 ~ 1.273
Education level	-0.055	0.025	-2.161	4.670	0.031	0.947	0.901 ~ 0.995
Intercept	-1.509	0.394	-3.829	14.661	0.000	0.221	0.102 ~ 0.479

*Dependent variable: depression

The result is manifested in Table 4. By again judging whether p-value is less than 0.05, it is inferred whether these elements can influence depression. For age and education level, the p-value is less than 0.05. The regression coefficient value of age is greater than 0, indicating that it will have a significant positive impact on depression. And the odds ratio (OR value) is 1.012, which means that when age increases by one unit, the magnitude of depression would increase to 1.012 times. Whereas the regression coefficient value of educational level is less than 0, indicating that it will have a significant negative impact on depression. And the odds ratio (OR value) is 0.947, which means that when educational level increases by one unit, the magnitude of depression would decrease to 0.947 times. In terms of marriage, the p-value is larger than 0.05.

3.5. Further discussion

From the analysis above, it can be concluded that age and education level would influence the probability of depression, while marriage would make no difference to it. To be more specific, the younger and more cultivated people are, the less possible they would get depressed; while the older and less cultivated people are, the more possible they would suffer from depression. These results are rather interesting, and shall be looked forward to gain further explanation. For example, young people may tend to be more optimistic and have more chance to change their destiny, but as they get old, they gradually realize the roughness and frustration of life. Also, people with higher educational level may be taught to view life in a wider vision, so that tend to understand, accept and cope with incidents in life, especially those miserable ones, in a more rational and serene way, making them more immune to depression.

4. Conclusion

This article uses data gathered from people in rural area, to study what factors are related to depression. The factors in the data include sex, marriage, dumber of children, education level, members in family, age, gained asset, durable asset, saved asset, living expenses, other expenses, income from salary, income from farm, income from business, income from non-business, income from agriculture, farm expenses, labor primary, lasting investment, non-lasting investment. Through calculation and analysis, the result is that only age and education level is related with the possibility of suffering from depression. Age gives depression a positive impact and education level gives depression a negative impact.

It should be noted that, although the data used in this study is collected from 291 villages, how extensively do those villages distributed in the world is unknown. As it's known, different rural areas can offer totally different political, economic, cultural and social background or context, and thus influencing the possibility of depression implicitly and distinctly. Therefore, the result might be limited into certain region, country or ethnic, and is not able to be generalized to the rural area in every place.

The result shows age and education level as two influential factors, but to find out the reason of their power is not the main purpose in this paper, so further studies can continue to research on their causal factors. Also, the data in this paper innovatively selected as many as 20 potential factors, with an emphasis on the economic condition of rural people, but it turns out that all the factors about property and money are irrelevant to depression. Taking this result into consideration, further research can convert to another aspect on people's life condition, and find other breaches. Moreover, this paper only focuses on rural people, so other social groups can be taken care of in future studies.

All in all, this paper hopes that these results can help the prevention, diagnosis and treatment of depression, and help people in rural zones.

References

[1] Zhu X 2022 Qualitative research on the life meaning experience of adolescent depression patients with non-suicidal self-injury. Zhejiang University of Traditional Chinese Medicine.

[2] Li Y, et al. 2024 The mediating effect of depression symptoms and empathy satisfaction among nurses on the impact of work stress and secondary trauma. Industrial Health and Occupational Disease, 41-45.

[3] Zheng C 2023 The impact of depression on quality of life and related factors in Parkinson's disease patients, Master's thesis, Nanchang University.

[4] Hammen C 2018 Risk factors for depression: an autobiographical review. Annual review of clinical psychology, 14, 1-28.

[5] Brenneisen M F, et al. 2016 Factors associated to depression and anxiety in medical students: a multicenter study. BMC medical education, 16, 1-9.

[6] Tsaras K, et al. 2018 Assessment of depression and anxiety in breast cancer patients: prevalence and associated factors. Asian Pacific journal of cancer prevention: APJCP, 19(6), 1661.

[7] Mao Y, Du L, Zhang X, Wang Z and Zhou L 2023 Epidemic status and influencing factor model construction of postpartum depression among lying-in women in Zhengzhou City. Shenzhen Journal of Integrated Traditional Chinese and Western Medicine,23, 15-18+137.

[8] Hu D 2019 The relationship between adult carbohydrate intake and depressive symptoms, Master's thesis, Qingdao University.

[9] Liu H, et al. 2019 Construction of a risk prediction model and exploration of the pathogenesis of non-suicidal self-injury behavior in middle-aged and young patients with severe depression based on decision tree algorithm. Chongqing Medical, 1-13.

[10] Huang J 2018 The influencing factors of depression symptoms among residents in Guangdong Province based on the zero-inflation negative binomial model, Master's thesis, Guangdong Pharmaceutical University.

[11] Razzak H A, Harbi A and Ahli S 2019 Depression: prevalence and associated risk factors in the United Arab Emirates. Oman medical journal, 34(4), 274.