

# Investigation on Depression of College Students Based on Bayesian Model

Deqiang Yuan, Lili Yang, Yuyang Wu, Chao Wei

**Abstract**—This paper studies the influencing factors of college students' depression and carries out population identification. Firstly, the possible influencing factors are analyzed by using multivariate logistic regression. Secondly, the equations of tolerance, acceptance test and five suspicious factors are established respectively. Finally, Bayesian model is used to identify the depression group of college students and provide accurate identification technology for schools.

**Index Terms**—college students' depression, multivariate logistic regression, Bayesian model, identification technology

## I. INTRODUCTION

### 1.1 Background

Nowadays, with the rapid development of society, the proportion of mental diseases among college students in China is increasing year by year, especially the psychological crisis caused by depression. In recent years, the incidence of depression has been increasing among college students. Some researches reported that in Asia, 11 percent of college students were suffered from depression, while the aggregate prevalence of anxiety disorders was 7 percent ([14-16]). Some literature indicated that the incidence of depression among college students is higher than that in general population. In China, more than 20 percent of Chinese college students were suffered from depression and this ratio has kept growing in the past few decade ([5-6]). At 2017 "International Neuropsychiatric Disease Summit Forum", Professor Fang Yiru, chairman of the Psychiatry Basic and Clinical Branch of the Chinese Society of Neuroscience, said: " in 2015, there were more than 320 million people suffered with depression in the world, and there were two peak ages for depression: around 20s and around 40s. Therefore, the group of college students is a high incidence group of depression. In China, suicide ranks first among the cause of death observed in those aged 15-34 years, accounting for 19% of all deaths in this age group [17-19]. Therefore, it is of great practical significance to screen out the depression groups of college students in time and effectively curb them.

Manuscript received October 12, 2021; revised September 13, 2022.

Deqiang Yuan is a Professor of School of Mathematics and Statistics, Anyang Normal University, Anyang 455000, China (corresponding author to provide phone: 13783851179; Email: aytongji@126.com).

Lili Yang is an undergraduate student of School of Mathematics and Statistics, Anyang Normal University, Anyang 455000, China (Email: 277565527@qq.com)

Yuyang Wu is an undergraduate student of School of Mathematics and Statistics, Anyang Normal University, Anyang 455000, China (Email: 279438812@qq.com)

Chao Wei is a Professor of School of Mathematics and Statistics, Anyang Normal University, Anyang 455000, China (Email: chaowei0806@aliyun.com).

### 1.2 Research status

In recent years, the mental health of university students has been studied by many authors ([1]). College students' psychological problems has been highly valued by country and society. A lot of researches have studied the problem of college students' depression at home and abroad and showed that college students are the high incidence group of depression ([20-21]). Cooper ([2]) indicated that the overall prevalence of depression among Chinese college students is 23.8% and has a rising trend. Some authors focused on gender, grade, major, nationality and family economic status ([7-9]), left behind experience ([10-13]), parents' marriage status, interpersonal communication and employment prospects ([4]). Li et al. ([3]) investigated the relationship between family economic situation, interpersonal relationship, employment prospects, love status and college students' depression.

### 1.3 The research content of this paper

This paper is organized as follows. In Section 2, the design and implementation of research scheme are given. In Section 3, the research process and main results are provided. We derive that family economic status, academic performance, future planning, interpersonal communication and parental education are highly correlated with depression. The conclusion is given in Section 4.

## II. DESIGN AND IMPLEMENTATION OF RESEARCH SCHEME

### 2.1 Research ideas

This study includes threestages as follows:

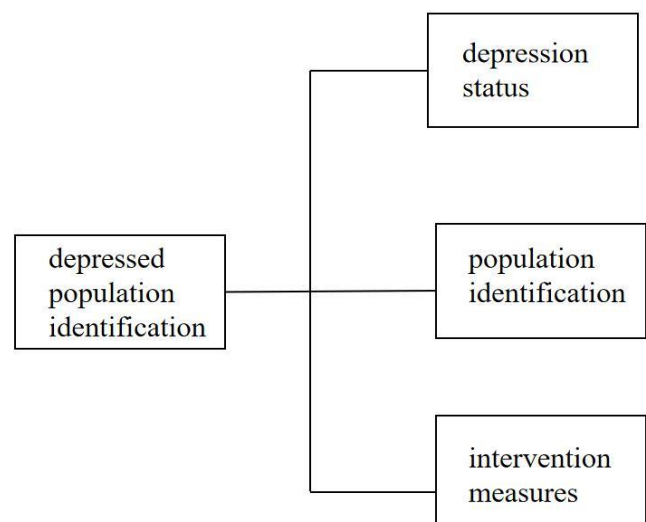


Fig 1. Research road map

In Figure 1, the three stages are depression status, identification of college students and intervention measures. Depression status includes depression mood, depression cognition and depression effects. Identification of college students includes model and applications. Intervention measures includes innovative education and hierarchical management.

2.2 Research project design

(1) Respondents

We select undergraduates in Henan as the research object. We issues 23670 questionnaires and 23635 are recovered. The effective questionnaire is 23412, with an effective efficiency of 98.91%. In the questionnaire statistical analysis, a few data can not track the survey object to complete the information because of only 1-2 empty items. Based on the large missing of 1-2 blank values, the statistical analysis of the whole data will not be affected.

(2) Investigation methods

The project team, together with the mental health education and consultation center, distributes questionnaires to all departments for online survey.

III. RESEARCH PROCESS AND ORGANIZATION

3.1 Research organization and research process

3.1.1 The Research process

(1) The questionnaire design stage

In order to test the validity of the questionnaire, we communicates with teachers of the school mental health studio. According to the data collection in previous years and the feedback of the results, we modify the deficiencies of the questionnaire as follows:

1) For the investigation of recent psychological state, there are four levels to facilitate their choice.

2) The interviewees think that the difference between "most of the time" and "most of the time" in the questionnaire is not big, so it is difficult to distinguish them. Therefore, "most of the time" is changed to "generally more time".

3) Considering the concerns of the respondents about the length of the questionnaire, the form of the scale was changed to the matrix form. By combining some questions into one, the respondents felt that there were not many questions visually.

(2) The questionnaire structure

The questionnaire consists of general information questionnaire and SDS scale. The specific contents are shown in Appendix 2. The basic structure of the questionnaire is as follows:

1) Questionnaire questions: multiple choice questions, scale questions, fill in the blank questions, combined with open questions and closed questions.

2) Questionnaire framework: basic information, the degree of understanding of depression among the subjects, mental health status of the subjects.

(3) The formal investigation stage

We use social media to survey the whole school by using online questionnaire. Questionnaire survey refers to the survey method that the investigators use the unified design questionnaire to understand the situation or seek opinions

from the selected respondents, and understand the respondents' views and opinions on depression.

3.1.2 Analysis of influencing factors of depression

We mainly analyze gender, grade, interpersonal communication, family economic situation, whether or not the only child, student origin, staying experience, academic achievement, future planning and parents' education level. Through the questionnaire survey, the following data are obtained, which is the proportion of people with different degrees of depression in the total number of this factor.

3.1.3 The quality control of survey data

(1) The quality control in questionnaire design

Before designing the questionnaire, the team members consult the relevant information and get familiar with the questionnaire making process and how to deal with various situations in the survey process; In the design of the questionnaire, the structure and content of the questionnaire were further improved by communicating with several teachers in the mental health education and consultation center of our school.

TABLE 1 Number and proportion of depression

Variable		Proportion	Depression detection rate
Gender	Male	30.01%	12.9%
	Female	69.99%	16.7%
grade	Freshman	29.80%	12.4%
	Sophomore	25.28%	12.5%
	Junior	24.83%	20.7%
	Senior	20.09%	17.7%
only-child or not	Yes	21.08%	26.2%
	No	78.92%	12.7%
Place of residence	Countryside	75.01%	15.9%
	City	24.99%	14.5%
Academic achievement ranking	Top 30%	41.01%	17.0%
	30%-60%	48.00%	13.2%
	After 60%	10.99%	20.7%
Is there a clear and practical plan for the future	Planned	31.99%	14.2%
	Yes, but not clear	44.02%	18.3%
	No planning	23.99%	12.3%
Family economic status	Good	6.99%	14.5%
	Commonly	79.93%	14.7%
	Bad	13.08%	21.4%
Father's education	Primary school and below	23.87%	24.2%
	Middle school	63.61%	13.8%

	University and above	12.52%	7.83%
Mother's education	Primary school and below	31.82%	18.6%
	Middle school	60.01%	14.6%
	University and above	8.17%	10.2%
Have you ever been left behind	Never	68.72%	19.6%
	Less than 1 year	8.97%	12.0%
	1-3 years	15.93%	19.4%
	More than 3 years	6.38%	12.5%
Views on gain and loss	If you don't get it, you can't let go	59.69%	36.2%
	Can't be relieved	8.55%	13.1%
	There is no substitute	20.93%	15.8%
	It doesn't matter if you lose it	10.83%	12.3%

(2)The quality control of questionnaire collect- ion

After the questionnaire, the quality of the questionnaire was checked. If the answer time is too short or too long, the answers are inconsistent, and the same option is repeatedly selected, the invalid questionnaire will be eliminated.

### 3.1.4 The data sources

(1) In cooperation with the mental health education and consultation center, the depression data of undergraduates were obtained through the general survey of the whole university.

(2) From the mental health education and consultation center of the University, we get the data about the mental health education of the past college students.

## 3.2 The data analysis method

### 3.2.1 The data visualization

Pie chart, bar chart, histogram, scatter chart and other data visualization methods are used to clearly show the basic situation, centralized trend and distribution of data.

### 3.2.2 The multivariate statistical analysis

Through the use of SPSS software, the multivariate statistical method of regression equation model is adopted to explore the statistical laws among the variables.

## 3.3 Analysis Of Data

### 3.3.1 Distribution of the number of depressed people

With the help of descriptive statistical analysis, this survey obtained the basic information and depression distribution of

college students in our school. The specific data are shown in the table below:

TABLE 2 Number and proportion of depression

Degree of depression	Number of depression	Proportion
No depression	19767	84.42%
Mild to mild depression	2576	10.98%
Moderate to severe depression	1007	4.29%
Severe depression	71	0.31%
Total	23421	1

### 3.3.2 College Students' awareness of depression

Students are prone to psychological problems in college, and a large part of them have depression tendency. According to the survey of the World Health Organization in 2019, most domestic students' cognitive level of depression is at the level of having heard of the psychological disease of depression, but they rarely understand the depression and how long the bad mood will last to reach depression. The statistical results obtained from the questionnaire are 72% college students know a little about depression. Only 17% college students know a lot about depression.

### 3.3.3 Self awareness of depression in college students

In the contemporary university campus, many college students will be depressed. If they stay in this state for a long time, they may be depressed. However, many students can not realize whether they have entered the state of depression at the first time, which will lead to the initial mild depression to moderate or even severe depression. In the later stage of data analysis, through on-the-spot interviews and social media, we communicated with people with different degrees of depression, and found that nearly half of the students did not know that they were in depression, and the proportion of those who knew their illness well was 17.85%, which showed that college students could not be effectively aware of their own illness, so it was necessary for schools to find an efficient method to identify depressed people. The specific findings are 17.85% of college students are very aware of the disease, 36.24% are vague, and 45.91% are not aware of the disease.

## 3.4 The reliability and validity of the questionnaire

The reliability and validity analysis of the questionnaire is one of the criteria to test whether the questionnaire is qualified. Therefore, when analyzing the questionnaire, we should test the reliability and validity of the questionnaire to ensure that the next analysis is meaningful.

### (1) The reliability test

This paper uses Cronbach  $\alpha$  The reliability measures the internal consistency coefficient of the questionnaire items,

and the value range of the reliability coefficient is [0,1]. The calculation formula of Cronbach coefficient is as follows:

$$\alpha = \frac{k}{k-1} \left( 1 - \frac{\sum S_i^2}{S_T^2} \right)$$

Among them,  $k$  is the total number of questions on the scale,  $S_i^2$  is the in-question variance of the  $i$  question score, and  $S_T^2$  is the variance of the total score of all the question items. It can be seen from the formula that Cronbach  $\alpha$  The coefficient evaluates the internal consistency of the scores of each item in the scale, and there is a positive correlation between the reliability of the questionnaire items and the reliability coefficient. Using SPSS software to test the reliability of the questionnaire, Cronbach was obtained  $\alpha$  The coefficients are shown in Table 3.

TABLE 3 Cronbach's Alpha coefficient

Cronbach a	Based on standardized items Cronbach a	The number of projects
0.847	0.864	15

From Table 3, it can be seen that the Cronbach  $\alpha$  coefficient of the questionnaire on the mental health of college students is 0.847, greater than 0.7, so the reliability of the questionnaire is good, that is, the questionnaire has stability and reliability.

(2) The validity test

KMO test is used to check the partial correlation between variables, and the value is between 0-1. In practical analysis, KMO statistic is better when it is above 0.7; When KMO statistic is less than 0.5, it is not suitable to use factor analysis. The KMO and Bartlett coefficients of the questionnaire were calculated by SPSS software:

TABLE 4 KMO and Bartlett verification

Validity of Kaiser Meyer Olkin measurement sampling		0.842
Bartlett's spherical test	The Chi square value	132.511
	df	28
	The Significance	0.000

It can be seen from Table 2 that the kmo value of the questionnaire is 0.842, greater than 0.7, and Bartlett's sig is 0.000, which means that the zero hypothesis of spherical test is rejected. The correlation coefficient matrix has significant difference, indicating that the data is suitable for factor analysis.

3.5 Recognition of Depressed College Students

In the process of exploring the influencing factors of depression, establishing multiple regression models and multiple linear regression models to construct associated factors with college students' depression, using stepwise regression, as variable screening, the risk factors that finally

enter the equation are related factors that affect depression and anxiety symptoms. The original model incorporates all the ten variables involved in the questionnaire, including gender, grade, interpersonal communication, family economic situation, whether it is an only child, birthplace, left-behind experience, academic performance, future planning, and parents' education level. Secondly, the multiple regression model is used and the related variables deleted before are adjusted to construct the correlation model of family economic situation, interpersonal relationship, academic performance ranking, future planning and parents' education level.

3.6 Multiple logistic regression to screen suspicious factors

Different degrees of depression of college students (0=no depression, 1=slight depression, 2=moderate depression, 3=severe depression) as dependent variables, the ten influencing factors mentioned above are used as independent variables. After analysis of multiple logistic regression in SPSS, it was found that 5 suspicious influencing factors ( $p < 0.1$ ) were screened out, namely, family financial situation, interpersonal communication, academic performance, future planning, parents' education level, and the occurrence of depression have obvious correlation ( $p < 0.05 \sim p < 0.01$ ).

3.7 Establishment of multiple logistic regression equation model

This article mainly analyzes the degree of depression through endurance and the test accepted. First, establish a binary logistic regression equation between the factors selected above and endurance, with endurance as the dependent variable. Secondly, establish the regression equation between accepting the test, taking the accepting test (acceptable test=1, unacceptable test=0) as the dependent variable. Through logistic regression analysis in SPSS, the correlation coefficient before each variable is obtained, and its regression is established equation.

3.7.1 The relationship between tolerance and depression

This paper divides the endurance as a dependent variable into four levels, with five suspicious factors as independent variables, among which  $p_2, x_1, x_2, x_3, x_4, x_5$  represents the probability of endurance being 1, the value of endurance, family economic situation, interpersonal relationship, and academic performance.

TABLE 5 Logistic regression analysis of tolerance and various influencing factors

Variables	B	Standard Error	Sig.	Exp(B)
Family financial	0.127	0.256	0.071	1.039
Interpersonal relationship	-0.483	0.673	0.072	0.714
Grades ranking	0.519	0.461	0.064	1.481
Future plan	-0.832	0.341	0.131	0.719
Educational level of parents	0.135	0.364	0.062	0.836
Constant	3.25	0.746	0.051	0.582

From Table 5, the relationship between endurance and five factors can be obtained as:

$$\ln\left(\frac{p_1}{1-p_1}\right) = m(x_1, x_2, x_3, x_4, x_5) = 3.25 + 0.127x_1 - 0.483x_2 + 0.519x_3 - 0.832x_4 + 0.135x_5$$

Then, we can get that

$$y_a = \frac{\exp(m(x_1, x_2, x_3, x_4, x_5))}{1 + \exp(m(x_1, x_2, x_3, x_4, x_5))}$$

### 3.7.2 The relationship between being tested and depression

The acceptance test to be studied in this article is measured by the difference between ideal and reality. First, establish the equations of ideal, reality and the five influencing factors selected by the above logistic analysis. Secondly, the difference between the two types is tested and the relationship between the various factors. Finally, through SPSS analysis, the relationship between the acceptance test value and the depression value is obtained.

We can get the correlation between the ideal value, the actual value and the five factors. First, make the relationship matrix between the ideal value and the five factors, and we can obtain the test conditions of the variables in the equation, and the fitting conditions of the above steps are respectively given. The P value of the family economic situation is slightly greater than 0.05, but it is still acceptable because the exclusion criterion (0.1 by default) is used here, and this variable can be left in the equation. Among them,  $p_2, x_1, x_2, x_3, x_4, x_5$  are the probability of ideal value 1, family economic situation, interpersonal relationship, academic performance, future planning, parents' education level, and the regression equation can be obtained.

TABLE 6 Logistic regression analysis of ideal value and various influencing factors

Variables	B	Standard Error	Sig.	Exp(B)
Family financial	0.090	0.181	0.062	1.094
Interpersonal relationship	-0.328	0.596	0.026	0.265
Grades ranking	0.609	0.240	0.011	1.838
Future plan	-0.614	0.294	0.037	0.541
Educational level of parents	0.031	0.215	0.041	0.746
Constant	1.072	0.685	0.038	0.342

From Table 6, we can obtain:

$$\ln\left(\frac{p_2}{1-p_2}\right) = n(x_1, x_2, x_3, x_4, x_5) = 1.072 + 0.090x_1 - 0.328x_2 + 0.609x_3 - 0.614x_4 + 0.031x_5$$

Then, we can obtain the equation:

$$p_2 = \frac{\exp(n(x_1, x_2, x_3, x_4, x_5))}{1 + \exp(n(x_1, x_2, x_3, x_4, x_5))}$$

The second is the regression equation of the actual value and each factor, among which  $p_3, y_b$ .

TABLE 7 Logistic regression analysis of actual value and various influencing factors

Variables	B	Standard Error	Sig.	Exp(B)
Family financial	0.083	0.142	0.012	1.087
Interpersonal relationship	-2.134	0.425	0.032	0.315
Grades ranking	0.537	0.140	0.046	1.724
Future plan	-0.592	0.324	0.027	0.481
Educational level of parents	0.072	0.274	0.016	0.713
Constant	-1.728	0.617	0.036	0.361

From Table 7, we obtain

$$\ln\left(\frac{p_3}{1-p_3}\right) = r(x_1, x_2, x_3, x_4, x_5) = 1.728 + 0.083x_1 - 1.143x_2 + 0.537x_3 - 0.592x_4 + 0.072x_5$$

It is easy to see that

$$p_3 = \frac{\exp(r(x_1, x_2, x_3, x_4, x_5))}{1 + \exp(r(x_1, x_2, x_3, x_4, x_5))}$$

Through the above analysis, we can get the relationship between the test and the five suspicious factors as follows:

$$y_b = p_2 - p_3$$

### 3.7.3 The relationship between depression, endurance, and acceptance

In summary, the final equation model can be obtained as:

$$\begin{cases} y_a = \frac{\exp(m(x_1, x_2, x_3, x_4, x_5))}{1 + \exp(m(x_1, x_2, x_3, x_4, x_5))} \\ y_b = p_2 - p_3 \\ y = 5.327y_a - 5.825y_b \end{cases}$$

Among them, (-2.391, -1.731) is a severely depressed group, (-1.731, -0.644) is a moderately depressed group, (-0.644, 0) is a mild to mildly depressed group, and (0, 0.808) is a non-depressed group.

### 3.7.4 Model checking

(1) SPSS test results

TABLE 8 Fitting information of rrgression equation model

Model	Model fitting conditions		Likelihood ratio test	
	-2Log likelihood	Bangla	Degree of freedom	Significance
Intercept only	425.871			
Finally	331.282	94.589	66	0.012

According to the above table, it can be seen that the significance is  $0.012 < 0.05$ , indicating that the model fits well.

(2) Substitution inspection

Among people with different depression levels, 500 students were randomly selected, including 300 without depression, 100 with mild depression, 75 with moderate to severe depression, and 25 with severe depression. The data of five factors collect-ed in the previous period were substituted into the regression In the equation model, it can be obtained that the distribution of depression degree is concentrated on the following data:

$(2.748, 6.963)$ ,  $(-0.518, -0.132)$ ,  $(-1.413, -0.762)$ ,  $(-2.139, -1.876)$ ,

and the equation fits well within the interval range of the above regression model.

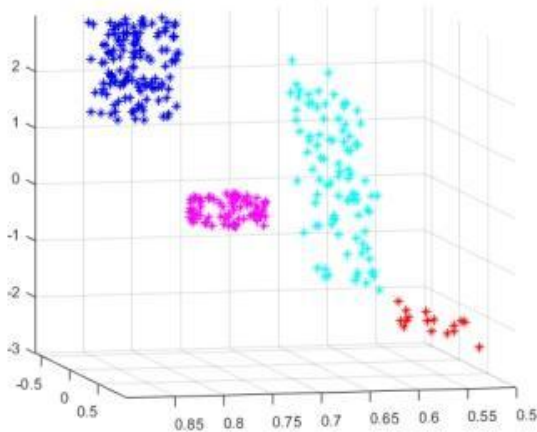


Fig 2. Three-dimensional scatter plots of different depression levels

In Figure 2,  $(2.748, 6.963)$  indicates 300 without depression,  $(-0.518, -0.132)$  indicates 100 with mild depression,  $(-1.413, -0.762)$  indicates 75 with moderate to severe depression,  $(-2.139, -1.876)$  indicates 25 with severe depression.

3.8 Dynamic recognition of depressed people

This project team uses logistic regression equation model to identify the depression groups of college students with different degrees. However, based on the actual situation,

some students may hide their true condition even though they have shown a tendency to be depressed. In this case, the regression result will show a certain deviation, so it is necessary to realize the dynamic recognition of the depression state of college students.

3.8.1 Principles of Bayesian Statistics

We use Bayesian methods to further explore the probability for depression of college students who deliberately hide the real situation, and make corresponding countermeasures based on different results.

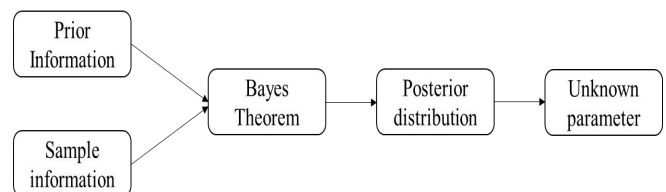


Figure 3. The basic form of Bayesian statistical inference

Figure 3 describes the core idea of Bayesian methods.

3.8.2 Bayes statistical model

Bayes theorem makes the connection between conditional probability and its inverse. Let  $B_1, B_2, \dots, B_n$  events be mutually exclusive and constitute A complete event and  $P(B_i) > 0$  if there is event A, and event A occurs at the same time as events in  $B_1, B_2, \dots, B_n$ , then:

$$P(B_i | A) = \frac{P(A | B_i) P(B_i)}{\sum_{j=1}^n P(B_j) P(A | B_j)}$$

In the formula,  $P(B_i)$  represents the prior probability, which refers to the possibility of the occurrence of the event judged by the observer according to the prior knowledge or existing data before the event occurs.  $P(B_i | A)$  represents the posterior probability, which refers to the more realistic probability after the knowledge of the prior of the random event and the integration of the Bayesian formula.  $P(A | B_i)$  is the condition, which is the probability of the occurrence of event B, given that random event A occurs.

The behavior habits of students change dynamically with time. For example, some moderately depressed students have significantly reduced their depression level through school education management. At this time, there is no need for the school to focus on them and they need to be removed from the moderately depressed group, otherwise it will cause a waste of educational resources. Therefore, the data of college students are collected regularly, and the above formula is used to update the data of depression factors in time, so as to identify people with different degrees of depression.

3.8.3 Bayes estimation

Define the probability function of sample population with parameter  $\theta$  as  $p(x | \theta)$ .

According to the prior information, we obtain the prior distribution  $\pi(\theta)$ .

Then, the joint conditional probability function of sample  $X = (x_1, x_2, \dots, x_n)$  is

$$p(X|\theta_0) = p(x_1, x_2, \dots, x_n | \theta_0) = \prod_{i=1}^n p(x_i | \theta_0)$$

Thus, we get the joint distribution of sample  $X$  and the parameter  $\theta$

$$h(X, \theta) = p(X|\theta)\pi(\theta).$$

Let

$$h(X, \theta) = \pi(\theta|X)m(X),$$

where

$$m(X) = \int h(X, \theta)d\theta = \int p(X|\theta)\pi(\theta)d\theta.$$

It is obviously that  $m(X)$  does not depend on the parameter  $\theta$ .

Therefore, we can obtain the posterior distribution

$$\pi(\theta|X) = \frac{h(X, \theta)}{m(X)} = \frac{p(X|\theta)\pi(\theta)}{\int p(X|\theta)\pi(\theta)d\theta}.$$

Consider the stochastic differential equation

$$dX_t = f(\theta, X_t)dt + dW_t, X_0 = x_0,$$

where  $W_t$  is the standard Wiener process and  $\theta$  is the parameter.

Then, we can obtain the likelihood function

$$L_T(\theta) = \int_0^T [f(\theta, X_t) - f(\theta_0, X_t)]dX_t - \frac{1}{2} \int_0^T [f(\theta, X_t) - f(\theta_0, X_t)]^2 dt.$$

### 3.8.4 Application of Bayesian model

By using regression analysis, we derive that five main factors for affecting the depression of college students are family economic condition, interpersonal relationship, academic performance ranking, future planning and the level of education of parents. Because the family economic situation and the level of education of parents does not have a large change in a short time, so in dynamic recognition mainly consider the remaining three factors. The first batch of data was collected in January 2020. Two months later, some samples were selected from people with different degrees of depression for information collection again, so as to obtain the probability information of each influencing factor on depression.

We obtain the prior probability of interpersonal communication situations, academic performance and future planning as follows:

(1) The conditional probability of two data

We get the conditional probability information of two time periods as follows:

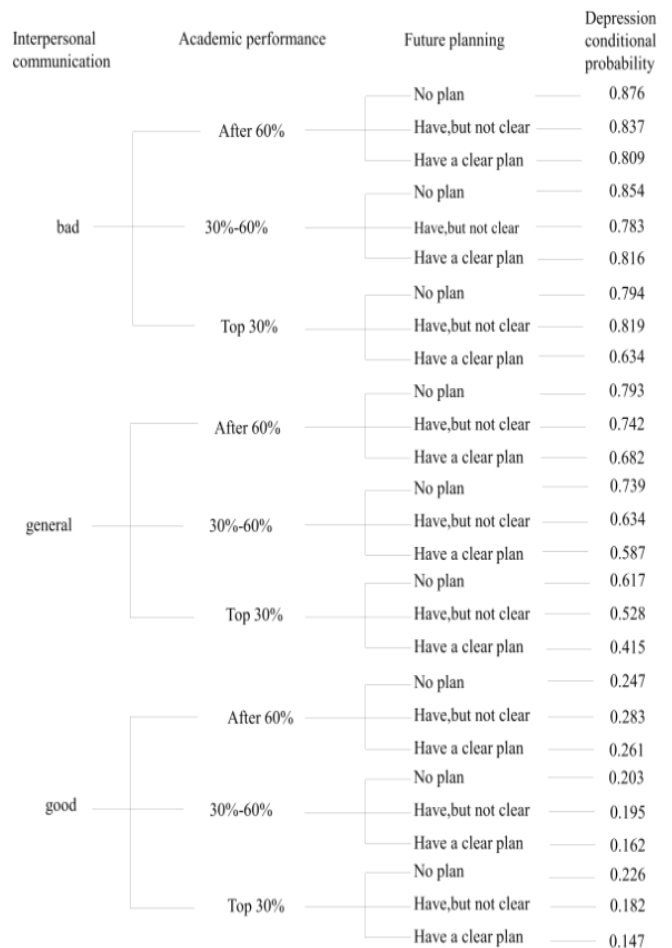


Fig 4. Conditional probability tree diagram of initial data

### (2) Conditional probability tree of the second data

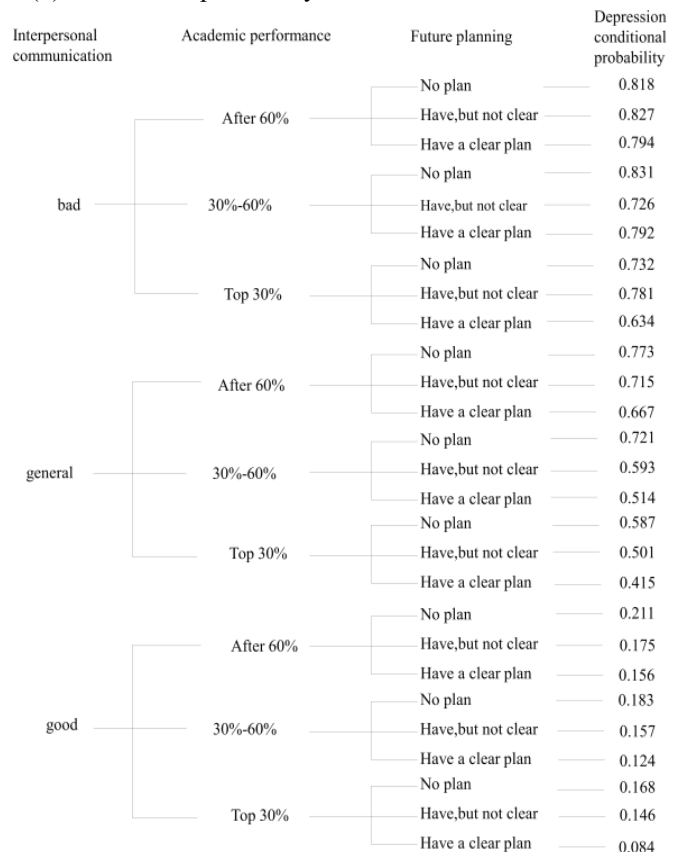


Fig 5. Conditional probability tree of the second data

Figures 4 and 5 give the probability of depression at two different time periods, namely the initial stage and two months later. Can be seen from the diagram depression significantly lower probability, this means that after a period of time of the education intervention, the situation of some people in depression can improve, so for this part of the groups ,the data of depression should be replaced in time. Substitute the new into the regression equation of the model mentioned above, to correctly identify the degree of depression, and thus for the corresponding education management.

Remark 1: There is another method called maximum likelihood estimation could also estimate the parameters. We introduce it as follows:

Let the probability function of whole be  $p(x; \theta)$ ,  $\theta \in \Theta$  is an unknown parameter,  $\Theta$  is parameter space of  $x_1, x_2, \dots, x_n$ . Treat the joint probability function of the sample as the function of  $\theta$  and use  $L(\theta; x_1, x_2, \dots, x_n)$  to describe it. Then, we obtain

$$\begin{aligned} L(\theta; x_1, x_2, \dots, x_n) \\ = p(x_1; \theta)p(x_2; \theta) \cdots p(x_n; \theta). \end{aligned}$$

where  $L(\theta; x_1, x_2, \dots, x_n)$  is called the likelihood function.

Then, the maximum likelihood estimation is

$$\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_n)$$

and satisfies

$$L(\hat{\theta}) = \max L(\theta).$$

#### IV. CONCLUSION

The aim of this paper is to study the influencing factors of college students' depression and carry out population identification. The possible influencing factors have been analyzed by using multivariate logistic regression. The equations of tolerance, acceptance test and five suspicious factors have been established respectively. Bayesian model has been used to identify the depression group of college students and provide accurate identification technology for schools.

#### REFERENCE

- [1] Z.Xueting, et al, "Perceived social support as moderator of perfectionism, Depression, and Anxiety in College Students." *Social Behavior & Personality An International Journal*, vol. 41, no. 7, pp. 1141-1152, 2013.
- [2] K.M.Cooper, et al, "An exploratory study of students with depression in undergraduate research experiences." *Plos One*, vol. 19, no. 2, pp. 447-458, 2020.
- [3] W. Li, et al, "Prevalence, correlates of major depression: A mental health survey among undergraduates at a mainland Chinese university." *Asia-pacific Psychiatry*, vol. 8, no. 3, pp. 206-214, 2016.
- [4] Y.Huang, W. Lv, J. Wu, "Relationship between intrinsic motivation and undergraduate students depression and stress: the moderating effect of interpersonal conflict." *Psychological Reports*, vol. 119, no. 2, pp. 527-538, 2016.
- [5] A.Qaisy, L.Ming, "The relation of depression and anxiety in academic achievement among group of university students." *International Journal of Psychology & Counselling*, vol. 3, no. 5, pp. 96-199, 2011.
- [6] W.Gao, S.Ping, X.Liu, "Gender differences in depression, anxiety, and stress among college students: a longitudinal study from China." *Journal of Affective Disorders*, vol. 263, no. 5, pp. 292-300, 2019.
- [7] J.G.Zong, et al, "Coping flexibility in college students with depressive symptoms." *Health & Quality of Life Outcomes*, vol. 8, no. 66, pp. 103-105, 2010.
- [8] D.Eisenberg, et al, "Prevalence and correlates of depression, anxiety, and suicidality among university students." *American Journal of Orthopsychiatry*, vol. 77, no. 4, pp. 534-542, 2010.
- [9] S.Andrew, et al., "Effects of parental warmth and academic pressure on anxiety and depression symptoms in Chinese adolescents." *Journal of Child&Family Studies*, vol. 15, no. 24, pp. 106-116, 2015.
- [10] S.R.Jenkin, et al., " First-generation undergraduate students' social support, depression and life satisfaction." *Journal of College Counseling*, vol. 16, no. 2, pp. 129-142, 2013.
- [11] E.Hards, et al. "Negative view of the self and symptoms of depression in adolescents." *Journal of Affective Disorders*, vol. 262, no. 1, pp. 143-148, 2019.
- [12] S.D.Amarasuriya, A.F.Jorm , N.J.Reavley , "Prevalence of depression and its correlates among undergraduates in Sri Lanka." *Asian Journal of Psychiatry*, vol. 15, no. 10, pp. 32-37, 2015.
- [13] A.A.Ceyhan, E.Ceyhan, Y.Kurty, "Investigation of university students' depression." *Eurasian Journal of Educational Research (EJER)*, vol. 36, no. 36, pp. 75-90, 2009.
- [14] R.Dyson, K.Renk, "Freshmen adaptation to university life: depressive symptoms, stress, and coping. " *Journal of Clinical Psychology*, vol. 62, no. 10, pp. 1231-1244, 2010.
- [15] N.Bayram, N.Bilgel, "The prevalence and socio-demographic correlations of depression, anxiety and stress among a group of university students." *Social Psychiatry and Psychiatric Epidemiology*, vol. 43, no. 8, pp. 667-672, 2008.
- [16] X.Liu, S.Ping, W.Gao, "Changes in Undergraduate Students' Psychological Well-Being as They Experience University Life." *International Journal of Environmental Research and Public Health*, vol. 16, no. 16, pp. 28-34, 2019.
- [17] S.Yao, et al, "The depressive experiences questionnaire: construct validity and prediction of depressive symptoms in a sample of Chinese undergraduates." *Depression and Anxiety*, vol. 26, no. 10, pp. 930-937, 2008.
- [18] H. Meng, et al, "Rural/urban background, depression and suicidal ideation in Chinese college students: a cross-sectiona study." *Plos One*, vol. 26, no. 10, pp. 13-21, 2013.
- [19] C.Lu, et al, "Depression among Chinese university students: prevalence and socio-demographic correlates." *Plos One*, vol. 8, no. 11, pp. 451-453, 2013.
- [20] J. Cheng, Y. H. Sun, "Depression and anxiety among left-behind children in China: a systematic review." *Child Care Health Dev*, vol. 41, no. 4, pp. 515-523, 2014.
- [21] C.P.Li, H.B.Hao, "Likelihood and Bayesian estimation in stress strength model from generalized exponential distribution containing outliers," *IAENG International Journal of Applied Mathematics*, vol.46, no.2, pp.155-159, 2016.