Dean&Francis

ISSN 2959-6157

Influencing Factors of Mental Health Issues among Adolescents

Xiaoyan Song[1, *]

[1]Department of Software, Northwest University, Xi'an, 710000, China
*Corresponding author: song080862@outlook.com

Abstract:

This paper focuses on exploring the influencing factors of mental health issues. Utilizing a binary Logit regression model and a dataset comprising 1,000 samples, it examines the impact of four key variables-stress level, sleep duration, working hours, and physical activity duration-on mental health status. The analytical results reveal that the p-values of all four variables exceed 0.05, indicating that none of them exert a statistically significant effect on mental health. Additionally, the overall prediction accuracy of the model stands at 54.30%, reflecting poor predictive performance. The study suggests that the underperformance of the model may be attributed to weak correlations between the selected variables, the omission of complex factors such as family environment, and limitations in measurement methods. To enhance the explanatory power and prediction accuracy for population mental health issues and provide a solid foundation for formulating intervention strategies, future research should expand the range of variables, optimize measurement methods, and experiment with alternative models.

Keywords: Mental health; influencing factors; binary logit regression model; lifestyle.

1. Introduction

The population is the fundamental force for national development, and their mental health status is an important cornerstone of social harmony and stability, as well as the focus of attention from all sectors of society. Individuals' physiological functions and psychological cognition continue to develop and change at different stages of life. When facing multiple challenges such as life pressure and work competition, they may experience psychological problems to varying degrees [1]. In recent years, surveys have shown that approximately 10% of the population has mental health levels below the average [2]. Mental health is a necessary condition for individual survival and development, and its importance is no less than that of physical health. Once the defense line of mental health is breached, it will not only affect the current quality of life, but also may leave long-term hidden dangers in the life trajectory.

Regarding the college student population, compared with 2018, the scores of all symptom factors in the SCL-90 have significantly increased. This indicates that the mental health status of college students has

ISSN 2959-6157

continued to decline over the past decade [3]. For the middle-aged group, a certain medical college recently compared the results of over 6,000 physical examinations conducted at the affiliated hospital's outpatient department. They found that more than half of the middle-aged people suffered from various diseases to varying degrees, and the main reason was various mental health problems [4]. Regarding the elderly population, 26.4% of them exhibit depressive symptoms, among which 6.2% are of moderate to severe severity. Depression has become a significant risk factor threatening the physical and mental health as well as the quality of life of the elderly [5]. Therefore, it can be seen that people of all ages have varying degrees of mental health problems.

The formation of mental health problems is not the result of a single factor, but a complex outcome of the interweaving and interaction of environmental factors and individual factors. Family environment, social environment, work environment, as well as individual physiological factors, personality traits and coping abilities, all have an impact on the mental health of the population [6]. In terms of family, family noise level (i.e., family chaos) can cause emotional and sleep problems in individuals, which in turn lead to mental health problems [7]. Regarding the school aspect, current research indicates that the number of teenagers who have experienced school bullying is significantly higher than those who have not, and they report more mental health problems [8]. In terms of society, studies have confirmed that lower social class is a risk factor for mental health problems, and people from lower social classes have lower mental health levels [9]. In terms of individual reasons, factors such as genes and heredity may cause some mental health problems, and a person's personality traits can affect emotions, thereby affecting mental health [10]. In addition, individual differences in the ability to cope with stress also have an impact - when facing the same life setbacks or interpersonal conflicts, people with problem-solving skills are more likely to regain psychological balance, while those who habitually adopt negative coping strategies such as avoidance and self-blame are more likely to accumulate psychological problems.

In related studies, scholars have used various methods to conduct research, providing references for this field. For example, some studies use a multiple mediation model from a social-ecological perspective to analyze the influencing mechanisms of mental health problems in low-income populations; some studies explore the correlation between family noise levels and population mental health through correlation analysis and regression analysis. These methods highlight the necessity of research from a multi-dimensional and multi-model integrated perspective.

However, some studies only analyze the three aspects of family, work and society separately, with insufficient examination of the interaction of these factors. This makes it difficult to reasonably explain some psychological problems.

Therefore, this study intends to adopt the framework of ecosystem theory, integrate data on demographic characteristics (age, gender, occupation, country), mental health status (whether there are mental health problems, severity of problems, consultation history) and lifestyle factors (pressure level, sleep duration, working hours, physical activity duration), and use the binary Logit regression model to analyze the impact of pressure level, sleep duration, working hours and physical activity duration on population mental health. The study aims to provide empirical basis for building a hierarchical intervention system (such as family communication guidance programs, workplace psychological support courses, etc.) and help improve the accuracy and effectiveness of population mental health prevention and treatment.

2. Methods

2.1 Data Description

The database is available on the Kaggle platform website. The dataset contains 1,000 observations and involves 12 variables, with no missing values. The data cover demographic characteristics (age, gender, occupation, country), mental health status (whether there are mental health problems, severity of the problems, history of consultation), and lifestyle factors (pressure level, sleep duration, working hours, physical activity duration).

2.2 Sample Selection

The study selected samples from all age groups for analysis, and finally included 1000 valid samples. According to the core research objective, 4 key independent variables and 1 dependent variable were selected:

Dependent variable: Mental health status. "Having mental health problems" is coded as 1, "Having no mental health problems" is coded as 0, serving as the prediction target of the binary Logit model.

Independent variables: Stress level, Sleep hours, working hours, physical activity time. Among them, the stress level is an ordered categorical variable (low = 1, medium = 2, high = 3), while the rest are continuous variables (table 1).

Table 1. Variable Explanation

Name of Variables	Definition of the Variables
Mental_Health_Condition	Mental health status, 1 = Problem, 0 = No problem
Stress_Level	Stress level: 1 = Low, 2 = Medium, 3 = High
Sleep_Hours	Average daily sleep duration (hours)
Work_Hours	Average daily working / study duration (hours)
Physical_Activity_Hours	Average daily duration of physical activity (hours)

There was a total of 515 cases (accounting for 51.50%) in the sample who had mental health problems, while 485 cases (accounting for 48.5%) had no mental health issues. The sample distribution was basically balanced, meeting the research requirements (Figure 1).

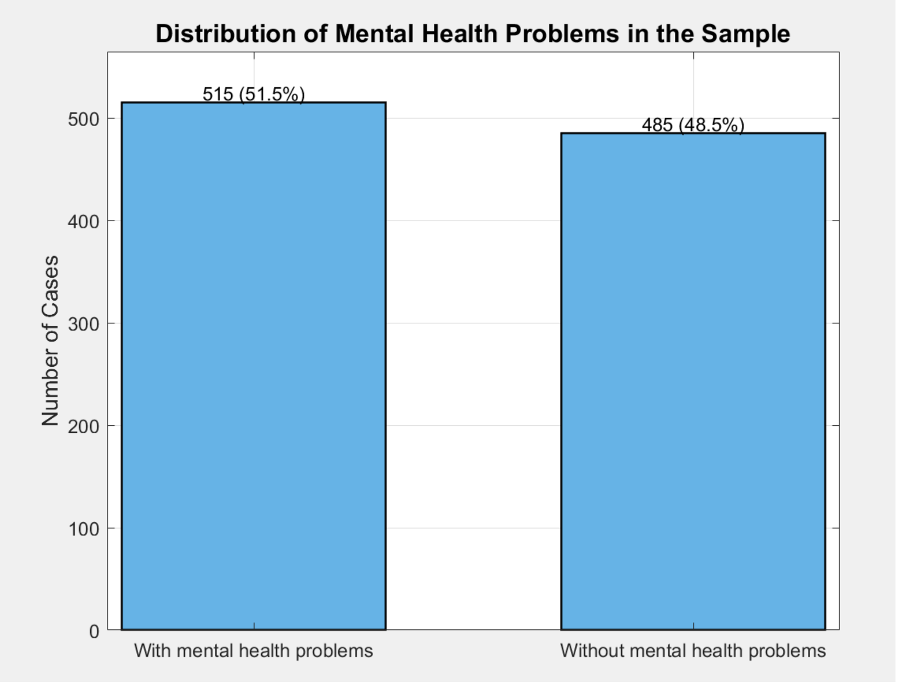

Fig. 1 Distribution of Mental Health Problems in the Sample (Picture credit: Original)

2.3 Sample Selection

This study adopts the binary Logit regression model, with Mental_Health_Condition as the dependent variable and Stress_Level, Sleep_Hours, Work_Hours, Physical_Activity_Hours as independent variables. The model is constructed to explore the impact of independent variables on the dependent variable. The model formula is:

$$ln(p/1-p) = -0.131 + 0.050 Stress\_Level - 0.029 Sleep\_Hours + 0.005 Work\_Hours + 0.004 Physical\_Activity_{H}ours \quad (1)$$

Where p represents the probability that Mental_Health_Condition is 1, and 1-p represents the probability that it is 0. The validity and goodness-of-fit of the model are evaluated through likelihood ratio test, Hosmer-Lemeshow test, model prediction accuracy, etc. The influence of each variable is analyzed in combination with regression coefficients, p-value and OR value.

3. Result and Discussion

3.1 Model Results

Based on the binary Logit regression coefficient estimation results, the following conclusions are drawn: for each additional unit of “Stress_Level”, the log-odds of having mental health problems increases by 0.050 with a p-value of 0.523; for each additional unit of “Sleep_Hours”, the log-odds decreases by 0.029 with a p-value of 0.427; for each additional unit of “Work_Hours”, the log-odds increases by 0.005 with a p-value of 0.235; and for each ad-

ISSN 2959-6157

ditional unit of "Physical_Activity_Hours", the log-odds increases by 0.004 with a p-value of 0.843 (Table 2). The p-values of all variables are greater than the 0.05 significance level, indicating that none of these four factors have a statistically significant impact on mental health status in this model. In terms of coefficient magnitude, "Stress_Level" has the largest coefficient (0.050) and "Physical_Activity_Hours" has the smallest (0.004). However, since none of them passed the significance test, they cannot be used as effective indicators for predicting mental health status (table 2).

Table 2. Binary Logit Regression Information

Term	Coefficient	P-value
Stress_Level	0.050	0.523
Sleep_Hours	-0.029	0.427
Work_Hours	0.005	0.235
Physical_Activity_Hours	0.004	0.843
Intercept	-0.131	0.751

The confusion matrix was used to evaluate the model's predictive performance. The overall accuracy of the binary Logit model in predicting mental health conditions is 54.30%, which is relatively low and close to random guesses. The results show that the model performs poorly in predicting cases without mental health issues (true value = 0), with a prediction accuracy of only 32.99% and a high false positive rate of 67.01%. In contrast, it performs slightly better for cases with mental health issues (true value = 1), with a prediction accuracy of 74.37% and a false negative rate of 25.63%. In total, 543 individuals were correctly classified (either true positive or true negative), while 457 were misclassified (325 false positives and 132 false negatives).

3.2 Model Discussion

The low accuracy of the model may be attributed to several factors. First, the selected variables (stress level, sleep hours, work hours, sleep hours, work hours, and physical activity hours) may have weak correlations with mental health conditions, as indicated by the near-zero R-squared values (McFadden $R^2 = 0.002$, Nagelkerke $R^2 = 0.003$). Mental health is a complex issue influenced by multiple factors such as social support, family environment, and psychological resilience, which were not included in this model. Second, the sample characteristics or variable measurement methods may limit the model's performance. For example, stress level was measured as an ordinal variable without detailed grading, which might fail to capture its true relationship with mental health. Additionally, the model's failure to pass the likelihood ratio test ($p = 0.637$) further confirms that the included variables do not effectively improve the model's explanatory power.

In comparison with related studies on health prediction models, the accuracy of this model is significantly lower. For instance, similar logistic regression models in medical research often achieve accuracies above 80% when predicting disease outcomes, which highlights the inadequacy of the current model. This suggests that relying solely on the four selected variables to predict mental health conditions is incomplete.

4. Conclusion

This study explored the impact of stress level, sleep duration, working hours, and physical activity duration on adolescent mental health through a binary Logit regression model. The results showed that the p-values of these four variables were all greater than the 0.05 significance level, indicating that none of them had a statistically significant impact on adolescent mental health status in this model. In terms of the model's predictive effect, the overall accuracy was only 54.30%, which was at a low level. Among them, the prediction accuracy for samples without mental health problems was extremely low, and the false positive rate was high, indicating that the model had limited ability to predict adolescent mental health problems.

The poor performance of the model may be due to various reasons. On the one hand, the four selected variables had weak correlations with adolescent mental health. As can be seen from the near-zero R-squared values, their explanatory power for mental health status was negligible. Adolescent mental health is a complex issue, which is also affected by many factors not included in this study, such as family environment, campus atmosphere, social support, and personality traits. On the other hand, there may be limitations in the variable measurement methods. For example, stress level, as an ordinal variable, had insufficiently detailed grading, which failed to fully reflect the true relationship with mental health. At the same time, sample characteristics may also have a certain impact on the model results.

Based on the above research results, future studies can be improved in the following aspects: first, expand the scope of variables, include more potential influencing factors such as family functions, campus environment, and social support into the analysis to more comprehensively explore the causes of adolescent mental health problems; second, optimize variable measurement methods, conduct more detailed grading of variables such as stress level or use more accurate measurement tools to improve data quality; third, try other more appropriate statistical models or machine learning algorithms, such as structural equation models and random forests, to improve the prediction accuracy and explanatory power of adolescent mental health problems. In addition, longitudinal studies can be carried out to track the dynamic changes of adolescent mental health status and the long-term effects of influencing factors, so as to provide a more solid theoretical basis and practical guidance for formulating more effective intervention strategies for adolescent mental health.

References

[1] Chen Zihua, Liu Guo, Fan Yuanyuan, et al. Analysis of the Mental Health Status and Influencing Factors of Adolescents Aged 10-19 in Jingzhou City. Modern Hospital, 2023, 23(09): 1430-1434.

[2] Chen Qing. Research on the Existing Problems and Intervention Strategies of Teenagers' Mental Health. Zhengzhou Normal Education, 2023, 12(04): 43-47.

[3] Li Dan. Analysis of Factors Affecting College Students' Mental Health Based on Logistic Regression. Science and Education Journal, 2025, (13): 91-93.

[4] Lu Haidong, Jiang Chuan. Analysis of Factors Affecting the Mental Health of Middle-aged People and Sports Fitness Strategies. Science and Information (Science Teaching Research), 2008, (22): 533+603.

[5] Yang Xinyi, Li Xinyu, Zhu Lihua, et al. Analysis of Factors Affecting the Depression Status of the Elderly -An Empirical Study Based on the China Health and Retirement Longitudinal Study (CHARLS). Chinese Rural Health Management, 2025, 45(06): 444-450.

[6] Chen Xuefeng. Influencing Factors and Responses of Adolescent Mental Health Problems. People's Forum, 2024, (16): 90-93.

[7] Yu Guoliang, Zhang Zhe. Review of Research on the Impact of Family Noise Level on the Mental Health of Children and Adolescents. Shanxi Normal University Journal (Social Sciences Edition), 2024, 51(06): 80-87.

[8] Hakan K, Mats H. Bullying at School and Mental Health Problems among Adolescents: A Repeated Cross-sectional Study. Child and Adolescent Psychiatry and Mental Health, 2021, 15(1): 74-74.

[9] Chen Bing, Hu Chuchen, Zhao Yufang. Why Do Low-Income Adolescents More Easily Experience Mental Health Problems? - A Multiple Mediation Analysis Based on the Social Ecological Perspective. Liaoning Normal University Journal (Social Sciences Edition), 2024, 47(02): 105-113.

[10] Dritsas E, Trigka M. Stroke risk prediction with machine learning techniques. Sensors, 2022, 22(13): 4670.