Team Number :	0671

Problem Chosen :	A

2017 APMCM summary sheet

Keep Healthy to Keep Healthy

summary

Basing on the statistics, Chinese people's insomnia rate is on the rise. If people don't pay attention to rest, long-term insomnia will lead to the occurrence of various diseases. So, in order to maintain physical and mental health, a reasonable sleep plan is needed.

To solve the question 1, firstly, this paper analyzes the data in Annex I and excludes abnormal data. Then, the correlation between indicators and sleep quality is determined according to the new data, analyzing the Pearson correlation coefficient of each attribute using SPSS, which eliminates an unrelated attribute named Reliability. At the same time, the results of diagnoses in Annex I is divided into six groups relying on sex and age, and the unrelated indexes to sleep quality are excluded from these six data. Finally, a statistical classification model is used to obtain the multivariate regression relation between sleep quality and given index.

For question 2, at the beginning, this paper analyzes the data in Annex II table and excludes abnormal value. Then, according to the International Classification Standard, this paper tries to use qualitative analysis to classify the types of diagnoses by reducing dimension. Next, making quantitative analysis of data to find the ratio of male and female was closer to 1:2. Establishing iteration model as a basis, depending on the principle of depth of machine learning, sets up evaluation matrix of diagnoses. Iterating down the lower dimensional data, continually, until the iteration accuracy to the optimal solution of the model, then stop iterating immediately, to obtain the optimal solution. Finally, the correctness of the final iteration judgment reached 72.1%, which means that the disease variables and the sleep indexes have a higher degree of matching.

Directing attention to question 3, bringing the data in Annex III to the model of question 2, employing MATLAB algorithm carries out the depth machine learning.

Second, using clean data in Annex II as a sample set, and sending Annex III set into the sample concentration for training, the evaluation index will be obtained. After sorting and filtering the final evaluation criteria to get the optimal solution, in the light of the optimal solution, the corresponding disease is found. Ultimately, it is confirmed that half of the people have depression.

Focusing on question 4, this paper processes random simulations by MATLAB. Planning a healthy sleep schedule in terms of the results of the Hadoop, which is more effective as the number of iterations increases, and finds that the indexes what affected people's sleep are mostly restored to normal values.

Keywords:dimension reduction;pearson correlation coefficient; multiple regression analysis; sleep regression model

Contest

1 Introduction

1.1 Background

Since the 21st century, people's awareness of health has been unprecedentedly increased, and the new concept of "having health to have everything" has gained people's hearts. Therefore, the issue of sleep has aroused the concern of the international community. According to the World Health Organization survey of 25,916 primary care patients in 14 countries and 15 regions, found that 27% of people have sleep problems. It is reported that the incidence of insomnia in the United States is as high as 32~50%, 10~14% in Britain, 20% in Japan, 30% in France, and more than 30% in China. And fifty percent of students have insufficient sleep. Sleep disorders have a negative impact on quality of life, but a considerable number of patients have not been properly diagnosed and treated. Sleep disorders have now become a prominent issue that threatens the public worldwide.

Many factors that affect insomnia can generally be divided into two major categories of objective factors and subjective factors. Objective factors are environmental changes, before going to bed to drink tea or coffee, etc .; and subjective factors are generally the stress of life, emotional loss, mental excitement and other mental factors. However, in their growth and development, young people are easily tired due to the pressure of study and work. Therefore, they must pay special attention to bed rest to ensure good health.

1.2 Restatement of the Problem

In order to study the effects of sleep on the human body, the following issues were solved based on the data in Annexes I and II:

- Based on the data in Annex I, analyze the relationship between the given indicators and quality of sleep. If there is no correlation between one or more of the metrics and sleep quality, identify and exclude them.
- Analyze the relationship between diagnosis and sleep.
- According to the data in Annex III, the patient is diagnosed and the result is given.
- How to arrange a break to maintain good health? Develop sleep plans and assess their effectiveness.

1.3 Literature Review

In order to measure the quality of sleep,The University of Pittsburgh psychiatrist Dr. Buysse and others prepared Pittsburgh Sleep Quality Index (PSQI) in 1989. The scale is suitable for patients with sleep disorders, mental disorders, but also for the general assessment of the quality of sleep.

Based on the Pittsburgh Sleep Quality Scale, Yan Youwei and Lin Rongmao further studied the relationship between anxiety and sleep quality and their underlying mechanisms by collecting questionnaires and constructing structural equation models.[1] The results showed that high PSQI scores of anxiety-sensitive students were significantly higher than those of low anxiety-sensitive students. The higher the level of anxiety sensitivity, the worse the quality of sleep.

It is precisely because sleep problems occur in all age groups.Therefore, the study of sleep in the elderly is also necessary.In this regard, Zhao Peiqiu passed the clinical observation of 65 elderly patients nursing,analysis of the causes of insomnia in the elderly,Including physiological factors, psychological factors, environmental factors, drug factors, food factors in five areas.[2]

Further research is based on a comparison between the model and the reality.Liu Xianchen, Tang Maoqin proposed neurotic patients as the test object,test Reliability and Validity of Pittsburgh Sleep Quality Index Using Statistical Analysis Techniques.[3]

2 Assumptions and Justifications

To simplify the problem and make it convenient for us to simulate real-life conditions, we make the following basic assumptions, each of which is properly justified.

- Suppose the data in the title is suitable for everyone.
- Suppose the doctor is professional, but there are also judgments error.、
- Assume the data source is true and reliable.

3 Notation

Serial number	Symbol	Description
1	*p*	The patient to be diagnosed
2	*n*	Patient ID
3	*a*	Patient's age
4	*sex*	Patient's gender
5	*sou*	Patient source
6	*sq*	Sleep quality
7	*sl*	Sleep latency
8	*st*	Sleep time
9	*se*	Sleep efficiency
10	*sd*	Sleep disorder
11	*hg*	Hypnagogue
12	*dd*	Daytime dyfunction

4 Establishing The Question I Model And Solving

Question 1 asked for finding the correlation between the index and the quality of sleep, at the same time, one or more indicators that have no correlation should be eliminated. This is a high latitude data function fitting problem.

4.1 Data analysis

For a high dimension function fitting, it is necessary to low down the dimension of problem analysis.

First step, We need to analyze the data types of six indicators and dependent variables in the data source. The age indicator is a non-negative positive integer set, which is a set of integers from 16 to 87 in the sample set provided by the original data. Gender indicators are made up by male and female, and we abstracts it into a Boolean set consisting by 0 and 1: 0 for women and 1 for men; The reliability indicators are described by non-negative floating-point sets, which are two floating sets from 6.79 to 71.70 in the sample set provided by the original data; Psychoticism index is a set of two floating Numbers varying from 0 to 97.8; Nervousness index is a pair of floating Numbers varying from 0 to 84.12. Character index is a set of two floating Numbers varying from 21.71 to 79.13. The dependent variable is sleep quality, which is a non-zero integer set and a finite set of integers consisting of 0, 1, 2, and 3.

Tabel 1 The changing range in the original data set

	Number	Age	Sex	Source	Sleep quality	Reliability	Psychoticism	Nervousness	Character
Max	-	16	1	-	0	6.79	0	0	21.71
Min	-	87	0	-	3	71.7	97.8	84.12	79.13

4.1.1 Outliers

The second step is that removing the necessary outliers and improving the quality of the data before lowing down the dimension. figure 1 to figure 4 are the original indicators of Reliability, Psychoticism, and Character in 6439 lines. We can conclude that from figure 5 to figure 8, some data indicators fluctuates greatly and needs to be eliminated. If we use these data directly, we will have an error in the analysis results.

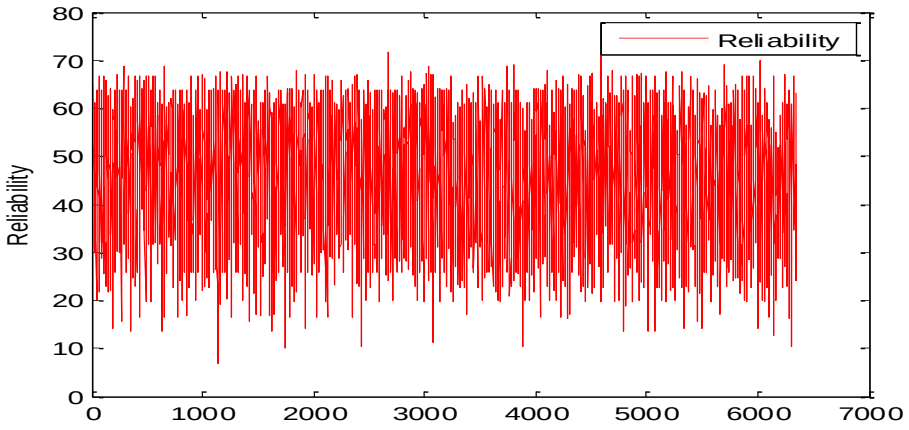


Figure 1 The original data of Reliability

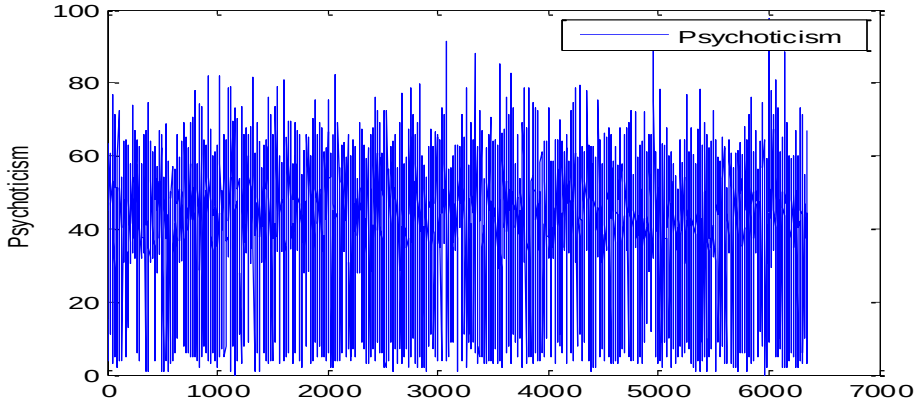


Figure 2 The original data of Reliability

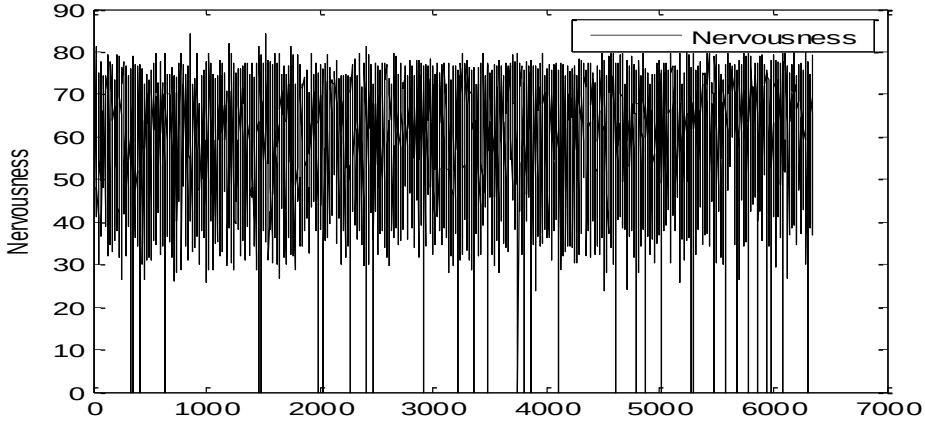


Figure 3 The original data of Psychoticism

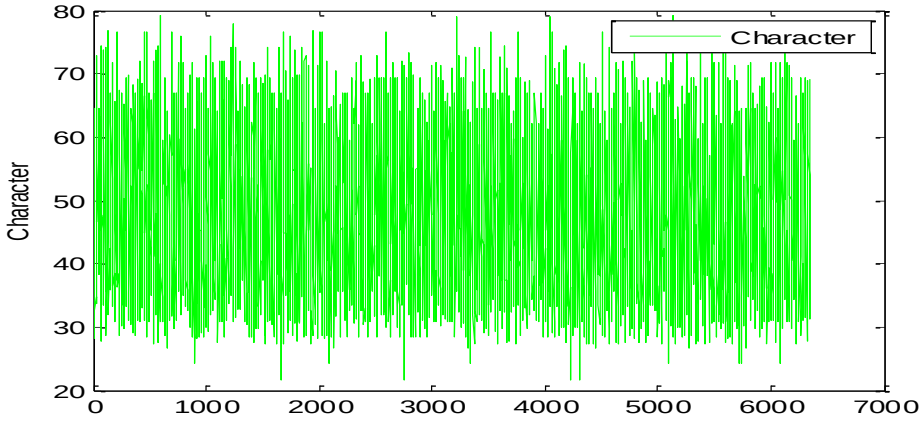


Figure 4 The original data of Nervousness

The method of abnormal value elimination adopts the Pauta criterion, 3σ criterion. When the deviation is greater than 3σ, for the original index data of the line 6349, the data is abnormal and needs to be eliminated in a timely way. The σ calculation formula is:

$$\sigma=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} \tag{1}$$

Specially, $x_i\,(i=1\sim 6349)$ means the original data including Reliability, Psychoticism, Nervousness and Character in the 6439 lines, $\bar{x}$ is the average of the original index data including Reliability, Psychoticism, Nervousness and Characte, and sample size is n=6439。

When the deviation is greater than 3σ, the data is outliers, and the criterion of the outliers is as follows:

$$|x_i-\bar{x}|>3\sigma \tag{2}$$

Therefore, The normal data range is $x_i>3\sigma+\bar{x}$ and $x_i<\bar{x}-3\sigma$. After stripping out the abnormal data, the rest is normal value. the original index data including Reliability, Psychoticism, Nervousness and Character in 6439 lines have 676 outliers. After stripping out all the outliers, the total rows of normal data sequences is 5763, conclusion shown from figure A to figure B.

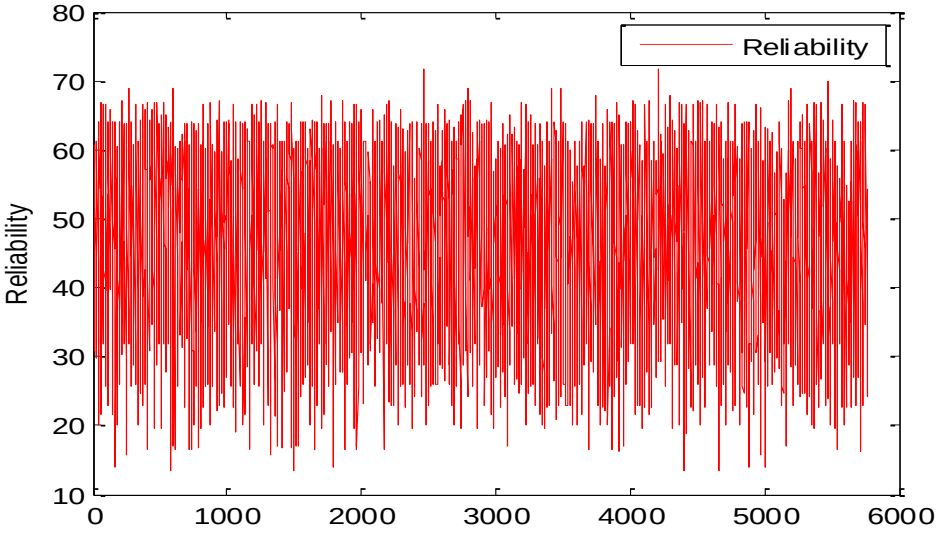


Figure 5 The post-processing data of Reliability

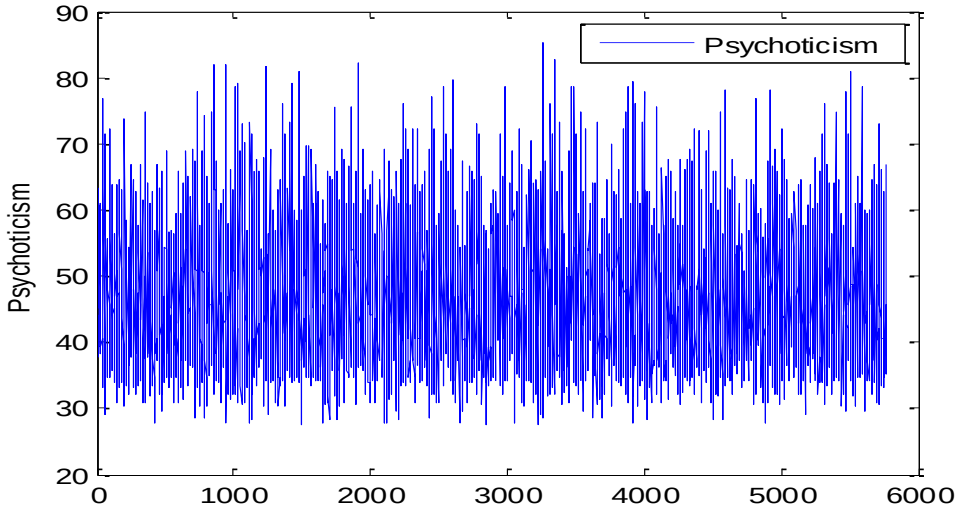


Figure 6 The post-processing data of Psychoticism

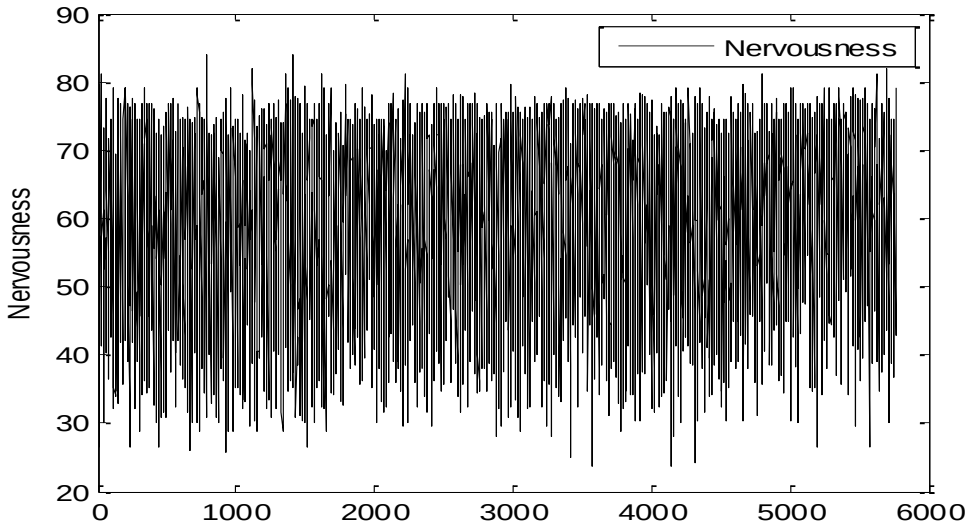


Figure 7 The post-processing data of Nervousness

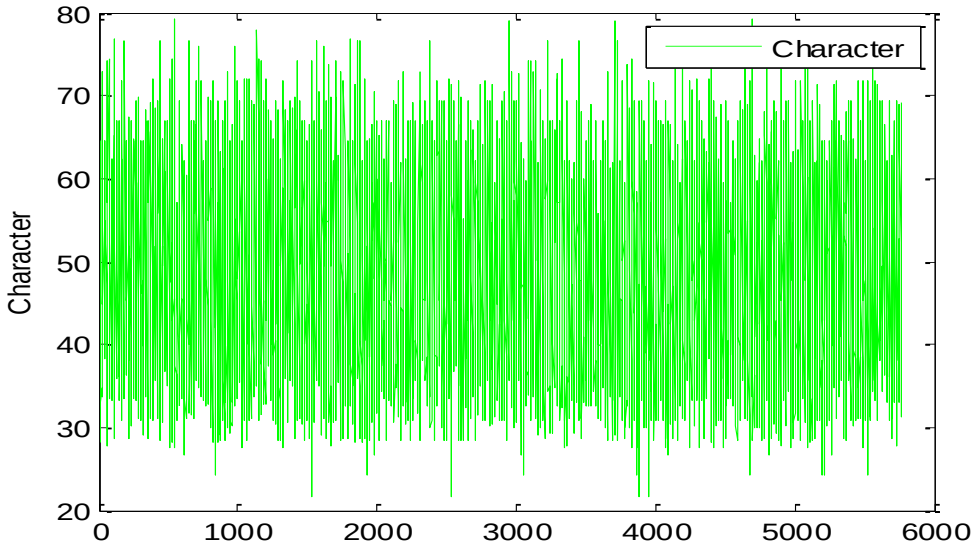


Figure 8 The post-processing data of Character

The data set Changed after data cleaning is:

Table 2 The range of data set changes after the table is cleaned

	Number	Age	Sex	Source	Sleep quality	Reliability	Psychoticism	Nervousness	Character
MAX	-	16	1	-	0	6.79	0	0	21.71
MIN	-	87	0	-	3	71.7	97.8	84.12	79.13

4.1.2 Dimension reduction

The third step, we need to find the correlation between the various indexes in the data and the quality of sleep after obtaining a better data source..

First, each index is used as a dimension to analyze the correlation between each dimension and the dependent variable. the raw data and the cleaned data were imported into SPSS to analyze, and the scatter plot analysis and correlation analysis were carried out, according to the Pearson's correlation analysis. The result is shown in figure 9 to 10:

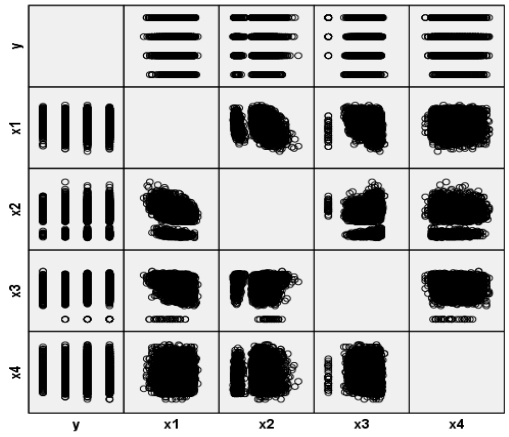


Figure9 Data pre-pre-processing scatter diagram

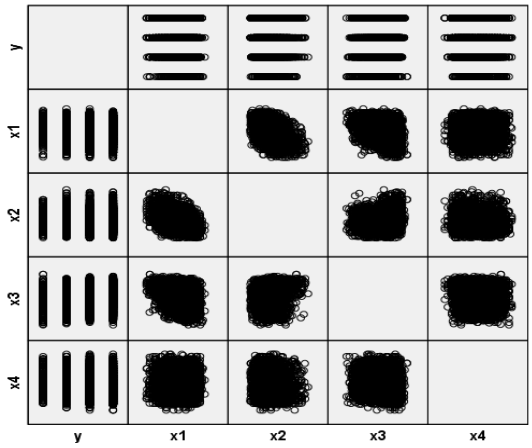


Figure 10 Scatter diagram after data preprocessing

Tabel 3 Pearson correlation between data preprocessing indicators:

Pearson correlation significance (bilateral) N	y	x1	x2	x3	x4
y	1 5763	.025 .058 5763	.042** .001 5763	.090** .000 5763	-.031* .019 5763
x1	-	-.419** .000 5763	-.399** .000 5763	-.399** .000 5763	-.008 .551 5763
x2	-	-	.245** .000 5763	.245** .000 5763	-.064** .000 5763
x3				1	-.195**

	- -	-	-	5763	.000 5763 1
x4		-	-	-	5763

Analyzing the two figures, results show that the linear correlation of pretreatment is significantly higher than the original data, which also verifies the necessity of preprocessing the data from the other side. Through the table, we can see that there is little correlation between Reliability and Sleep Quality. Therefore, this index can be excluded.

4.1.3 Rising Dimension

By establishing accurate mathematical model, in order to confirm a strong correlation between the individual data index, at the same time, through irrelevant indexes analyzing to prove the correctness of the above analysis.

We divided people into six categories and used SPSS to carry out Pearson correlation to analyze, as shown in the figure:

Tabel 4 Pearson correlation between six populations and sleep quality

Age and gender		Reliability	Psychoticism	Nervousness	Character
Male underage	Sleep quality	-0.028	0.147	0.336*	0.038
Young men		-0.082**	0.077**	0.145**	-0.005
Male elderly		-0.057*	0.065*	0.095**	-0.022
Female minor		-0.019	0.044	0.089**	-0.038
Young girl		0.051**	0.026	0.087**	-0.038*
Female elderly		0.039*	0.020	0.078**	-0.050**

Tabel 5 Pearson correlation between six groups of people and their sleep quality

	Male underage	Young men	Male elderly	Female minor	Young girl	Female elderly	overall
Related to sleep quality	N	R、P、N	R、P、N	N	R、N、C	R、N、C	P、N、C
Not related to sleep quality	R、P、C	C	C	R、P、C	P	P	R

We can conclude that the overall quality of sleep is independent of Reliability according to the figure. After dividing the population into six categories, sleep quality was not only about sex, but about age. Among them, the sleep quality of underage male and underage female is only related to that just only one, and the quality of sleep quality and character of the young and old men are irrelevant, and whether the sleep

quality of female youth or old age has nothing to do with psychosocial. Now, We set up the regression equation according to the correlation index.

4.2 Multiple Linear Regression Equations

- Multiple linear regression equations：

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \varepsilon \\ \varepsilon \sim N\left(0, \sigma^2\right) \end{cases} \tag{3}$$

In the formula, $\beta_0, \beta_1, \cdots, \beta_m, \sigma^2$ are unknown parameter having no correlation with $x_1, x_2, \cdots, x_m$. $\beta_0, \beta_1, \cdots, \beta_m$ is called as the regression coefficient.

Now, we get n independent observation data $\left(y_i, x_{i1}, \cdots, x_m\right), i = 1, \cdots, n, n > m,$ according to （3）, we calculate:

$$\begin{cases} y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} + \varepsilon_i \\ \varepsilon_i \sim N\left(0, \sigma^2\right), i = 1, \cdots, n \end{cases} \tag{4}$$

Marked as:

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1m} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nm} \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 & \cdots & \varepsilon_n \end{bmatrix}^T, \beta = \begin{bmatrix} \beta_0 & \beta_1 & \cdots & \beta_m \end{bmatrix}^T \tag{5}$$

Formula (7.1) is expressed as:

$$\begin{cases} Y = X\beta + \varepsilon \\ \varepsilon \sim N\left(0, \sigma^2 E_n\right) \end{cases} \tag{6}$$

E_n is the unit n matrix in there.

- according to theory of multiple regression using MATLAB to solve the regression equation, it is concluded that the quantitative relationship between sleep quality and index:

Tabel 6 Multiple regression simulation equations

Age section	Multiple linear regression equations
Underage male	$y_1 = 0.1346 + 0.0249x_3$
Middle-aged male	$y_2 = 1.5710 - 0.0014x_1 + 0.0025x_2 + 0.0086x_3$
Old male	$y_3 = 1.1947 + 0.0083x_1 + 0.0097x_2 + 0.0012x_3$
Underage female	$y_4 = -0.7255 + 0.0373x_3$
Middle-aged female	$y_5 = 1.3473 + 0.0073x_1 + 0.0088x_3 - 0.0008x_4$
Old female	$y_6 = 2.5066 - 0.0024x_1 + 0.0006x_3 - 0.0036x_4$
Total	$y = 1.7873 + 0.0019x_2 + 0.0056x_3 - 0.0010x_4$

5 Establishing The Question II Model And Solving

5.1 Data Analysis And Preprocessing

- **Elimination Of Outliers**

This paper based on the Diagnosis data given in Annex II, after researching and analyzing, found the table outliers exist in 7. The test Number respectively is: Number 161124064, Number 161005060, Number 160924001, Number 160110017, Number 160504048, Number 160505089, Number 160505089. The abnormal value should be given out processing. Because this data brought into this problem to analyze and solve, which will result in the deviation of the result of the data analysis, which can not draw the correct conclusion .After the elimination of the abnormal value, the data of 6346 normal sequence were obtained, and the specific data condition was seen (Supporting Material Annex II).

5.2 Qualitative Analysis

In this paper, the mental illness in Annex II translation table was divided into 11 categories according to the Classification Criteria Of International Diseases And the Expert Opinions, and the classification basis of the Annex II (Supporting Materials). The specific mental illness categories are listed as follows:

Table 7 Category Of Mental Illness

Number	Category Of Mental Illness	The number of sick people
1	Organic mental disorders	6
2	Mental illness caused by psychoactive substances	4
3	Schizophrenia and other psychiatric disorders	118
4	Mood disorders	1595
5	Neuropathic disorder	1769
6	Organ/somatic form disorder and dissociative disorder	60
7	Psychological factors correlate with physical obstacles	1797
8	Stress related disorder	360
9	Personality disorders and psychosexual disorders	5
10	Composite condition	586
11	Other conditions	59

Through the above table, it can be found that there are more than 1500 patients with mental disorders, neuropathic disorders and psychological factors. Other mental illness diseases are relatively few; However, mental disorders and psychoactive substances have the least number of diseases.

5.3 Quantitative analysis

- **Pittsburgh Sleep Quality**

Pittsburgh sleep quality (Pittsburgh Sleep Quality Index, PSQI) evaluates subjective sleep quality. In this paper, from the sleep quality, sleep latency, sleep time, sleep efficiency, sleep disorder, hypnagogue, daytime dysfunction, where seven aspects to evaluate the sleep quality problem. Each of aspects, according to the 0 to 3 points, cumulates score to the total score. The higher the score is worse sleep quality. Divided into demarcation is equal to the total score 7 points and the sensitivity is 98.3%, the specific degrees is 90.2%.

Table 8 The correlation coefficient of sleep quality in patients with depression anxiety scores(r)

Project	PSQI	Sleep quality	Sleep latency	Sleep time	Sleep efficiency	Sleep disorder	Hypnagogue	Daytime dyfunction
Depression score	0.25*	0.24*	0.13	0.07	0.12	0.10	0.23	0.23
Anxiety score	0.54**	0.36**	0.40**	0.19	0.23	0.24*	0.39**	0.57**

Note：*P<0.05; **P<0.01

- Sleep Quality Differences In Between Male and female

The different average effect between male and female in sleep was shown in table 4. According to the standard of Cohen_l, it can be seen that PSQI total score and the effect amount of each factor are small. (The effect quantity is < 0.5). Furthermore, the confidence interval was used to investigate whether the PSQI total score and the factors were different in sex. The results showed that PSQI was statistically different from sleep disorders and daytime dysfunction factors.

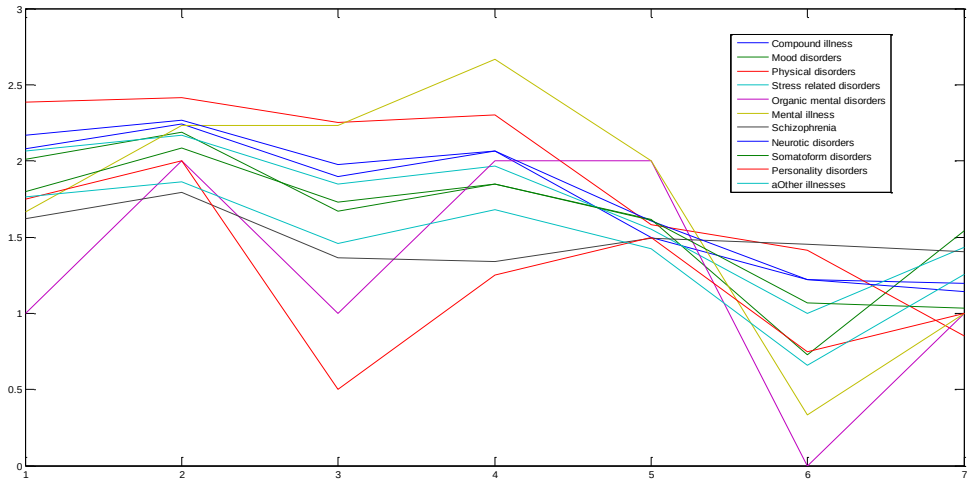


Figure 11 The Mean Of Each Factor With The Classification Of Mental Disease

A linear regression analysis of the dependent variables was based on PSQI score and factor scores before table 2 controls sample size,

Table 9 Linear Regression Analysis Of Dependent Variables

Variables	Partial Regression Coefficient				Partial Regression Coefficient	R^2
	Coefficient（β）	Standard Error（SE）	t	P		
Sleep quality	-0.002	0.008	-0.10	0.919	-0.011	<0.001
Sleep latency	0.001	0.009	0.12	0.904	0.013	<0.001
Sleep time	-0.009	0.012	-0.71	0.478	-0.079	0.006
Sleep efficiency	-0.008	0.010	-0.86	0.392	-0.096	0.009
Sleep disorder	0.013	0.008	1.55	0.125	0.179	0.029
Hypnagogue	<0.001	0.009	0.01	0.998	<0.001	<0.001
Daytime dyfunction	-0.005	0.010	-0.51	0.612	-0.057	0.003
PSQI	-0.018	0.030	-0.58	0.604	-0.062	0.004

A linear regression analysis of the dependent variables with $PSQI$ total score and factor scores after Table 3 controls sample size.

Table 10 Linear Regression Analysis Of Dependent Variables

variables	Partial Regression Coefficient				Partial Regression Coefficient	R^2
	Coefficient（β）	Coefficient（β）	R^2	P		
Sleep quality	-0.004	0.007	-0.85	0.319	-0.098	0.004
Sleep latency	0.007	0.009	-0.59	0.554	-0.065	0.007
Sleep time	-0.004	0.014	0.74	0.464	0.079	0.001
Sleep efficiency	-0.008	0.010	-0.30	0.764	-0.035	0.009
Sleep disorder	-0.002	0.009	-0.80	0.453	0.089	0.001
Hypnagogue	-0.001	0.009	-0.27	0.795	0.029	0.003
Daytime dyfunction	-0.002	0.006	-0.67	0.805	-0.027	0.006
PSQI	-0.026	0.033	-0.85	0.506	-0.098	0.02

- **The sleep quality of male and female**

A comprehensive study of the sleep quality of male and female found that no significant change in the quality of sleep. The possible reason is that as society becomes more diversified, male and female face various pressures and challenges, and they also have more and more ways to relieve stress and meet challenges. Many studies have found that close links between male and female's sleep quality and mental health problems such as anxiety and depression. From the existing research on the quality of sleep between male and female, most studies have focused on the

phenomenon, and few studies focused on how to improve the quality of sleep. Future research can focus on how to improve the sleep quality of male and female, which is important for improving the overall psychological state of male and female.

Table 11 Different Average Effect Of Sleep Quality And 95% Confidence Interval

Variables	$\bar{d}$	95% CI	
		Low	High
Sleep quality	0.02	-0.002	0.12
Sleep latency	0.02	-0.06	0.08
Sleep time	0.01	-0.05	0.23
Sleep efficiency	-0.03	-0.11	0.12
Sleep disorder	0.13	0.09	0.25
Hypnagogue	-0.01	-0.18	0.06
Daytime dyfunction	0.07	0.05	0.16
PSQI	0.08	0.02	0.33

- **Gender Differences In Sleep Quality Between Male And Female**

In terms of gender differences in sleep quality between male and female, the findings are inconsistent with previous studies. The inconsistencies in the results are likely due to random factors in different studies, the timing of the survey and the different locations of the site. In this study, the effects of these random factors were effectively balanced by meta-analysis. The results of the study showed that girls poor than boys in sleep disorders and daytime dysfunction. that is to say, the girl sleep in the night and the daytime of mental state are worse than male, it could be the girl's physiological cycles. Menstrual abnormalities (such as dysmenorrhea) are likely to have a significant impact on sleep quality. The women had higher scores on sleep disorder and daytime dysfunction, which resulted in a woman's discomfort during the night and a lack of energy during the day. In addition, this study also found that girls of PSQI scores higher than boys, that is means, the girl sleep quality worse than boys, which supports that the male sleep quality better than female.

5.4 Sleep Iterative Model

Next, we set up the Bayesian Discriminant function （F_i） model and the Fisher Discriminant function model （y_j）, using the seven indicators given in Appendix II, to estimate the influence of each factor on sleep.

$$F_i = \beta_{i1}x_1 + \beta_{i2}x_2 + \beta_{i3}x_3 + \beta_{i4}x_4 + \beta_{i5}x_5 + \beta_{i6}x_6 + \beta_{i7}x_7 + \beta_{i0} \quad (7)$$

$$y_j = \beta_{j1}x_1 + \beta_{j2}x_2 + \beta_{j3}x_3 + \beta_{j4}x_4 + \beta_{j5}x_5 + \beta_{j6}x_6 + \beta_{j7}x_7 + \beta_{j0} \quad (8)$$

Specially: β_{ij} stands for the coefficient of classification; β_{ji} means that the canonical discriminant function coefficient; x_1 for Sleep quality; x_2 for Sleep latency; x_3 for Sleep time； x_4 for Sleep efficiency; x_5 for Sleep disorder; x_6 for Hypnagogue; x_7 for Daytime dyfunction.

On the basis of the mentioned above, this paper is divided into 11 categories according to the International Classification Standard Of Diseases.

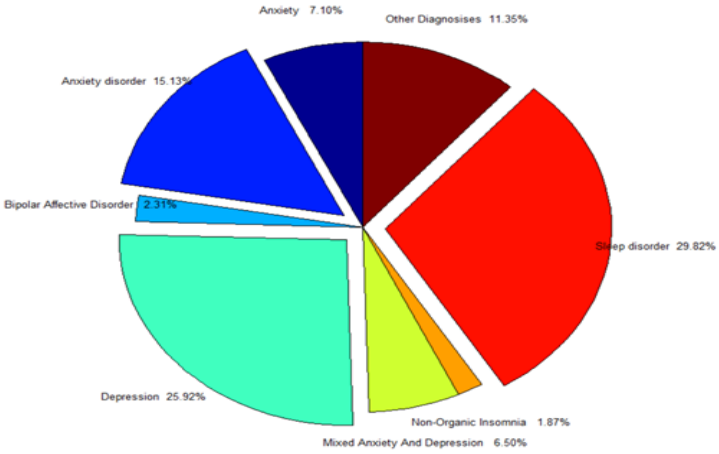
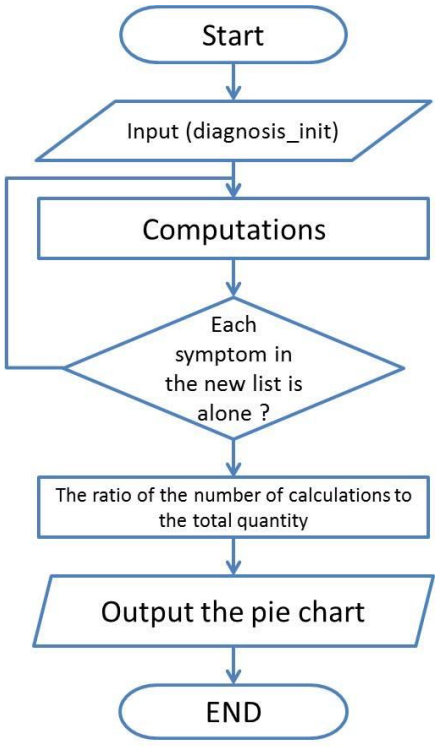


Figure 12 The Percentage Of Different Categories Of Diseases

In accordance with the figure above, there are a number of different types of disorders, such as mood disorders, neuropathic disorder, and psychological factors.

Quantitative analysis flow chart:



- Step 1: Clean the data to ensure the reliability of the source.
- Step 2: By processing the data provided by age, sex and evaluation matrix, the age distribution of the disease was found to be distribution of this diagnosis, and the sex ratio was close to 1:2.
- Step 3: Through MATLAB program to analyze diagnosis conditions, drying out of meaningless data, and the statistics condition of type is 100 kinds. Classified on the basis of this, for the various disease conditions, the average

amount is less than to be classified as single category, integrating separated after the final eight kinds of type, and counted the number of each condition and the percentage of total samples. Finally, making the pie charts shows the characteristics of the data clearly.

- Step 4: Using the evaluation matrix and combining the result of quantization classification, the data import SPSS is used to make the discriminant analysis. The final discriminant equation is obtained.

1. First Optimization

Through SPSS software, On the basis of the analysis of normal raw data after the abnormal value having been eliminated, the accuracy of the judgment is only 30.7%. Due to the low accuracy of regression, this paper will continue to reduce the mental illness in Annex II table. Finally, considering mood disorders, neuropathic disorder, and psychological factors, there are three kinds of diseases, and the data is more persuasive.

Tabel 12 The Canonical Discriminant Function Coefficient

	Function				
	1	2	3	4	5
V2	.500	-.781	-.466	1.229	-.041
V3	.011	-.373	.155	-.088	-.313
V4	.230	.301	.523	.017	.585
V5	.099	-.139	.592	-.567	-.218
V6	-.228	.349	-.575	-.305	1.300
V7	.269	.657	-.070	.138	-.275
V8	-.798	.248	.554	.451	-.147
(Constants)	-.819	.668	-1.251	-1.550	-1.504

Tabel 13 Coefficient Of Classification Function

	V1					
	1	2	3	4	5	6
V2	2.026	1.414	1.514	1.861	2.054	1.670
V3	1.562	1.451	1.506	1.451	1.521	1.498
V4	.430	.352	.157	.367	.422	.237
V5	.410	.309	.216	.240	.342	.269
V6	1.490	1.686	1.737	1.630	1.470	1.699
V7	-.144	-.122	-.342	-.019	-.039	-.189
V8	-.076	.646	.840	.177	-.050	.345
(Constant)	-8.147	-7.279	-7.228	-7.608	-8.072	-7.186

Tabel 14 The Classification Resultsa,c

		V1	Predict Group Membership Information						Total
			1	2	3	4	5	6	
Primordial	Counting	1	34	10	18	9	27	8	106
		2	47	63	131	40	47	40	368
		3	145	205	753	99	95	171	1468
		4	162	110	188	119	200	81	860
		5	456	170	237	203	478	148	1692
		6	74	49	119	54	51	56	403
		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	%	1	26.4	9.4	17.0	8.5	29.2	9.4	100.0
		2	12.8	16.3	36.1	10.9	12.8	11.1	100.0
		3	9.9	14.4	50.7	6.7	6.5	11.8	100.0
		4	18.8	12.9	22.1	13.0	23.7	9.4	100.0
		5	27.3	10.0	14.0	12.1	27.8	8.7	100.0
		6	19.4	12.4	31.0	13.4	12.7	11.2	100.0

a. The 30.7 per cent of the original groups were classified correctly.

In this paper, 29.8 per cent of the cases of cross-verification have been classified.

2. The Fifth Data Optimization Processing

Based on this data optimization, the people with Depression and Sleep disorder were selected as the sample sets, and the Depression was expressed as 1, 2 for Sleep disorder.

Table 15 The Canonical Discriminant Function Coefficient

	Function
	1
Sleep quality	0.479
Sleep latency	0.016
Sleep time	0.254
Sleep efficiency	0.130
Sleep disorder	-0.240
Hypnagogue	0.276
Daytime dyfunction	-0.796
(Constants)	-0.828

Fisher's discriminant function is:

$$y_1 = -0.828 + 0.479x_1 + 0.016x_2 + 0.254x_3 + 0.130x_4 - 0.240x_5 + 0.276x_6 - 0.796x_7 \quad (9)$$

Table 16 Coefficient Of Classification Function

	Diagnosis	
	1	2
Sleep quality	1.538	2.083
Sleep latency	1.459	1.477
Sleep time	0.163	0.452
Sleep efficiency	0.282	0.430
Sleep disorder	1.609	1.337
Hypnagogue	-0.265	0.048
Daytime dyfunction	0.883	-0.022
(Constants)	-6.115	-7.046

The Bayesian discriminant function is:

$$F_1 = -6.115 + 1.538x_1 + 1.459x_2 + 0.163x_3 + 0.282x_4 + 1.609x_5 - 0.265x_6 + 0.883x_7 \quad (10)$$

$$F_2 = -7.046 + 2.083x_1 + 1.477x_2 + 0.452x_3 + 0.430x_4 + 1.337x_5 + 0.048x_6 - 0.022x_7 \quad (11)$$

Tabel 17 The Classification Results

		Bianhao	Predict Group Membership Information		Total
			1	2	
Primordial	Counting	1	1310	458	1768
		2	547	1282	1829
		Ungrouped Case	869	892	1761
	%	1	74.1	25.9	100.0
		2	29.9	70.1	100.0
		Ungrouped Case	49.3	50.7	100.0
Cross Validation B	Counting	1	1306	462	1768
		2	548	1281	1829
	%	1	73.9	26.1	100.0
		2	30.0	70.0	100.0

b. The grouping cases of 71.9% cross-validation were classified correctly.

Conclusion: through the data further optimized for many times found that in the final selection two sick people for the sample set, and get back to determine type generation accuracy reached 72.1%, in order to meet the necessary requirements. In this paper, we can conclude that the accuracy of the fourth optimization is highly reliable, and the diagnosis results have a high degree of matching with each index.

5.5 Problem Three Model Is Established And Solved

5.5.1 Analysis And Positioning Of The Problem

For problem 3, we need to look for the relationship between various indicators and symptoms, and make a diagnosis based on the known indicators.

5.5.2 Establishment Of Model

The method solving the problem is: through the MATLAB software to make depth of machine learning, after cleaning, the Annex II data as sample set, through indicators into the sample should be focused on training, evaluation indexes for the final ranking selection, thus it is concluded that the optimal solution and found out the corresponding condition to determine the type of illness.

The problem is a data analysis problem. Solving the problem of the key is to find the right way of evaluation, using appropriate sample set as the training set, to index into the training of patients, to find the optimal solution.

For patientp_i, the indicators are $n_i, a_i, sex_i, sou_i, sq_i, sl_i, st_i, se_i, sd_i, hg_i, dd_i$; sex_i is the distribution set of Boolean distribution after binary processing of sex:

$$sex_i = \begin{cases} 1, & female; \\ 0, & male. \end{cases} \tag{12}$$

sou_i is p_i patient source and training sample source for comparison of the Boolean judging set:

$$sou_i = \begin{cases} 1, & p_i = T_i; \\ 0, & p_i \neq T_i. \end{cases} \tag{13}$$

Because the patient number n_i and the training sample number Tn_i are not affected by the result, it belongs to the irrelevant index, so it is no longer considered.

Therefore, the evaluation matrix of patients in group is:

$$A_{ki} = (\, a_i \quad sex_i \quad sou_i \quad sq_i \quad sl_i \quad st_i \quad se_i \quad sd_i \quad hg_i \quad dd_i \,) \tag{14}$$

For m group training sample evaluation matrix:

$$TA_{mi} = (\, Ta_i \quad Tsex_i \quad Tsou_i \quad Tsq_i \quad Tsl_i \quad Tst_i \quad Tse_i \quad Tsd_i \quad Thg_i \quad Tdd_i \,)$$

The variance operation of the same column in the two matrices:

$$S_{mi} = \sqrt{A_{ki}(k)^2 - TA_{mi}(k)^2} \tag{15}$$

Finally, the optimal judgment matrix is selected as:

$$C_{mi} = \min\{Smi\}, Smi > 0; \tag{16}$$

The sample set of the training set for C_{mi} is T_c, which contains samples of group c.

In particular, for C_{mi} number greater than 1 in the collection of items, if the n of the same item, v in different diagnosis, criterion according to the first v disease symptoms D_v in C_{mi} accounted for the symptoms score again in total weights of the training sample set, total D_v number of training samples is w, and the weight of the D_v is U_v as follows:

$$U_v = \frac{v}{w} \tag{17}$$

The optimized judgment matrix is:

$$C_{mi} = \min\{Smi\} \times U_{\mathrm{v}}, Smi > 0; \quad (18)$$

5.5.3 Solution Of Model

According to the training of the sample set, the results of 10 people in table Annex III are as follows:

Tabel 18 Diagnosis table

Number	Diagnosis
1	Depression
2	Depression
3	Anxiety,Sleep disorder
4	Postpartum Depression
5	Depression
6	Anxiety disorder
7	Depression
8	Depression
9	Sleep disorder
10	Sleep disorder,Anxiety disorder

5.6 problem 4 model establishment and solution

5.6.1 research and discussion on the importance of sleep

The human brain just changes the way it works during sleep， carrying out a series of active adjustment and reorganization, there are obvious changes in the system of the human body. Sleep has the ability to restore fatigue, maintain physical strength, and play a positive role in the normal operation of the nervous system.

In tension after a day's work and study, both mental and physical strength are at the height of fatigue state, only the reasonable and scientific sleep, talent is the body's cells in a state of completely relax and rest, especially the brain nerve cells. So sleep becomes the best way to restore normal energy and physical fatigue.

In general, a reasonable human sleep time is about 10 hours per day for preschoolers. School-age children should sleep between 9 and 10 hours per night; Young people under the age of 20 can sleep about nine hours a day. Adults sleep about eight hours a day. Sleep not only time should be sufficient, should notice the height of sleep quality more. Here are some details on how to improve the quality of human sleep:

Factors that affect sleep:
Personal factors: emotional excitement, sadness, anger, depression, etc., it is difficult to fall asleep.

Age: newborns and children slept longer, while middle-aged and older adults slept less.
Habit factor: people of the same age range have different amounts of sleep.
Physical activity factor: after physical activity, general sleep is deeper.
Drug factor: should take certain medicine, strong tea, coffee and other excited central nervous system, and affect sleep.

- **A comfortable sleep environment**

It is very important to have a quiet, gentle and temperate environment.

- **Get regular hours**

We need to live a regular life. Modern scientific research has also shown that there is a sophisticated "biological clock" in the human body, and it is constantly in the process of determining the course of human life. Washing your feet with hot water is more conducive to sleep. Reading books, drinking coffee, drinking tea, smoking and so on before going to bed can cause excitement and insomnia.

- **Pay attention to the moderation**

  When a person lacks exercise and labor, it is difficult to fall asleep. On the other hand, when a person is too tired, it is difficult to fall asleep at once, so pay attention to the appropriate amount of work to ensure good sleep. Diet.

- **Take the correct position and position:**

Lying on the right side is the best posture for sleep, which is good for the blood flow of the brain and liver, and is an effective guarantee to maintain normal brain cells and liver cells.

5.6.2 Sleep regression model

The following regression model is used to establish the following regression model to estimate the influence of each factor on sleep.

$$\begin{aligned} S_i = {} & \beta_1\left(\text{Daytime dysfunction}_i\right)+\beta_2\left(\text{Sleep latency}_i\right)+ \\ & \beta_3\left(\text{Sleep time}_i\right)+\beta_4\left(\text{Hypnotic}_i\right)+\beta_5\left(\text{Sleep disorder}_i\right)+ \\ & \beta_6\left(\text{Sleep efficiency}_i\right)+\beta_7\left(\text{Sleep quality}_i\right)+\beta_0 \end{aligned} \tag{19}$$

- **Based on the above model analysis, the following assumptions can be made.**

(1) if the long day function obstacle effect of sleep is higher than the substitution effect, then; If the long-term replacement effect is higher than the daytime dysfunction, then $\beta_1 < 0$。

(2) if the sleep time is normal, then $\beta_2 > 0$。

(3) if the amount of sleep is less than the amount of sleep that should be given in all ages, $\beta_3 < 0$。

(4) if a hypnotic drug is needed to get to sleep normally, no sleep duration of hypnotic drugs can be shortened $\beta_4 < 0$。

(5) if there are sleep disturbances, and the length of sleep, $\beta_5 < 0$。

(6) if sleep efficiency is not normal, the sleep duration is shorter， $\beta_6 < 0$。

(7) if the human body causes a poor quality of sleep due to certain factors, then $\beta_7 > 0$， ，$\beta_8 > 0$， and $\beta_7 < \beta_8$。

- **Age group, the shorter sleep time.**

Biomedical research has shown that nerve cells in the brain decrease with age and neurological activity begins to show abnormalities that affect the quality of sleep. In other words, the older you get, the worse your sleep quality, the less you get. According to the type (9)、(18) known，$\frac{dt_s}{da_1} = \frac{a_0}{2a_1^2} > 0$ 。As a result, older groups tend to sleep less.

- **Women sleep longer than men.**

Because women's brains tend to work more, they can do a lot of things at the same time and are more flexible, and use more complexity in brain activity than men. Therefore, the energy dissipation is faster than the male. According to equation (9), we can see:

$$\frac{dt_s}{df} = \frac{a_1 T + a_0}{\left(2a_1 + f\right)^2} > 0 \qquad (20)$$

The more you get, the longer you sleep. From this perspective, women need more sleep than men.

5.6.3 MATLAB Random Number Simulation

Through the simulation of MATLAB random number, a large number of data analysis is carried out to the attachment, so as to find out the optimal evaluation index which is not easy to get sick, and to determine the reasonable sleep plan according to the index.

- **MATLAB random number simulation procedure:**

First of all, in the second question, m set TA_{mi} of training sample with $\mathrm{r} = 11$, calculating each index range of domain, scope of use of the domain as a random number generated by the domain constraints, we create$\mathrm{h}_r = 100$ group, $\mathrm{r} = 11$ random matrix.

The second step, taking the data of the question 2 as the sample set and using the random number matrix as the patient source p_r, to carry out the calculation in the mathematical model of question 3, we get the evaluation matrix C_{mh_r}.

The third step, to establish a new evaluation model to screen out excellent patient source indicators, which are almost free of disease and far from the disease, finally,

we establish $j = 3$ evaluation indexes J_j

$$J_1 = \frac{sum\left(S_{mh_r} \mid S_{mh_r} > \left\{\max\left(S_{mh_r}\right) - \min(S_{mh_r}) \times 0.8\right\} + \min(S_{mh_r}),\ S_{mh_r} > 0\right.}{m}$$

$$J_2 = \frac{sum\left(S_{mh_r} \mid S_{mh_r} < \left\{\max\left(S_{mh_r}\right) - \min(S_{mh_r}) \times 0.1\right\} + \min(S_{mh_r}),\ S_{mh_r} > 0\right.}{m}$$

$$J_1 = \frac{sum\left(S_{mh_r} \mid S_{mh_r} > \left\{\max\left(S_{mh_r}\right) - \min(S_{mh_r}) \times 0.5\right\} + \min(S_{mh_r}),\ S_{mh_r} > 0\right.}{m}$$

The fourth step, through using the new set of evaluation indexes to the first three questions to influence people's sleep data to re-screen, came up with an ideal solution.

- **MATLAB Software Stochastic Simulation Of The Implementation Of The Operation**

After the random simulation of MATLAB software, the indexes of the reasonable sleep average are as follows:

Table 19 A Reasonable Average Of Sleep

Sleep quality	Sleep latency	Sleep time	Sleep efficiency	Sleep disorder	Hypnagogue	Daytime dyfunction
1.79	1.16	1.38	1.71	1.59	1.67	1.71

From the above table: people who achieve the above values are the least likely to develop diagnosis.

- Simulation Results Of MATLAB Software

For the first three questions, the results of the re-screening of the indicators that affect people's normal sleep are shown in the following table:

Table 20 The Numerical Value Of The Influence Index After Re-Screening

Number	J1	J2	J3	hr
1	42.53%	0.00%	65.95%	3
2	47.86%	0.03%	66.62%	13
3	39.16%	0.00%	65.62%	23
4	46.91%	0.00%	66.50%	32
5	43.65%	0.03%	66.34%	36
6	44.19%	0.05%	66.42%	46
7	40.37%	0.03%	65.82%	54
8	45.92%	0.03%	66.33%	70
9	50.93%	0.00%	66.70%	85
10	36.79%	0.02%	65.11%	89
11	48.24%	0.00%	66.50%	94
12	49.31%	0.02%	66.73%	99

Based on the analysis and solution of the previous three questions, by using the MATLAB software, we found that most of the indexes that affect people's sleep were restored to normal values. Therefore, this paper concludes that the effect of the new sleep program is more effective.

According to the analysis and solution of the previous three questions, by using the MATLAB software, we found that most of the indexes that affect people's sleep were restored to normal values. Therefore, this paper concludes that the effect of the new sleep program is more effective.6. Sensitivity Analysis and Model Test

6 Sensitivity analysis

In order to further verify the iteration accuracy, prove the existence of its contingency factors, therefore, this article will be on this basis will rise dimension processing. In the optimization of the data at the same time, we will also consider the effect of gender and age of iteration accuracy.

1） χ^2 distribution $\chi^2(n)$

If the random variables $X_1, X_2, \cdots, X_n$ are independent and obey the normal distribution $N(0,1)$, then the random variable:

$$Y = X^2_1 + X^2_2 + \cdots + X^2_n \tag{21}$$

The χ^2 distribution of the obedience degrees of freedom is n, marked χ^2.

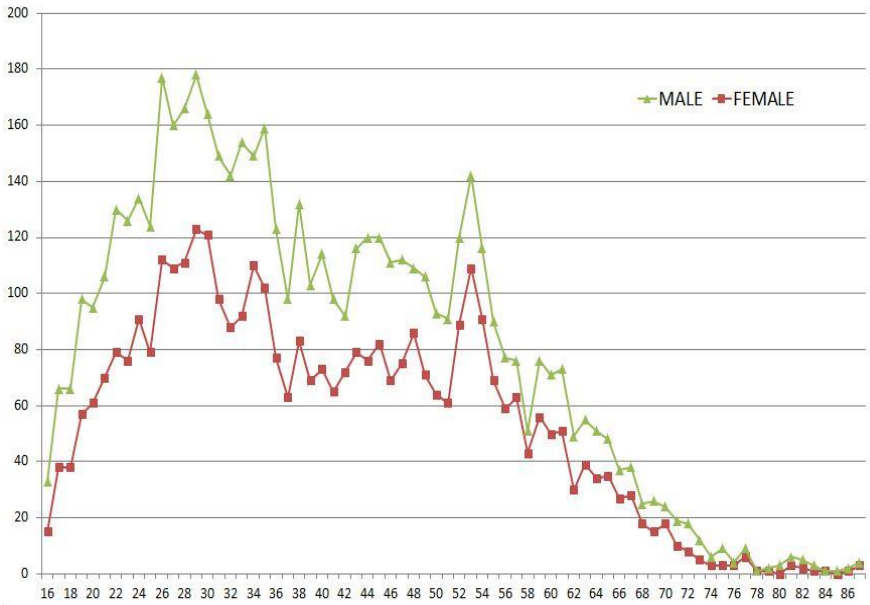


Figure 13 The Age Distribution Of Male And Female

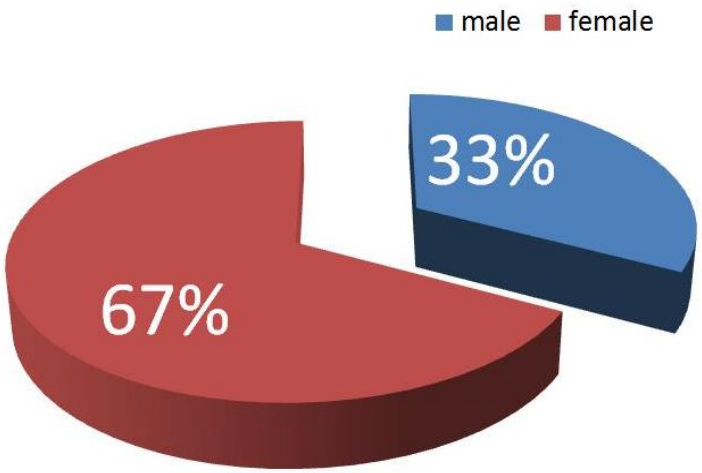


Figure 14 Sex distribution of male and female

It can be seen from FIG. X1 and X2: The age distribution of male and female was tested by sample and found that the distribution of male and female in figure X1 was mainly between the ages of 18 and 60. The gender distribution of male and female showed that the number of people with mental illness was greater than that of men with mental illness. Therefore, after the fourth optimization based on the higher dimension, the number of patients with the male and female patients was returned to the regeneration judgment test.

Tabel 21 The Male Canon Discriminating Function Coefficients

	Function
	1
Sleep quality	.475
Sleep latency	-.108
Sleep time	.246
Sleep efficiency	.175
Sleep disorder	-.219
Hypnagogue	.336
Daytime dyfunction	-.725
(Constants)	-.595

Tabel 22 The Male Classification Function Coefficient

	Diagnosis	
	1	2
Sleep quality	1.751	2.186
Sleep latency	1.360	1.261
Sleep time	0.212	0.438
Sleep efficiency	0.144	0.304
Sleep disorder	1.796	1.596
Hypnagogue	-0.338	-0.030
Daytime dyfunction	0.927	0.262
(Constants)	-6.203	-6.718

Note: fisher linear discriminant functionThe Bayesian discriminant function is:

F_1= -6.203+1.751x_1+1.360x_2+0.212x_3+0.144x_4+1.796x_5-0.338x_6+0.927x_7

F_2= -6.718+2.186x_1+1.261x_2+0.438x_3+0.304x_4+1.596x_5-0.030x_6+0.262x_7

Table 23 Male Classification Results

		Diagnosis	Predict Group Membership Information		Total
			1	2	
Primordial	Counting	1	293	111	404
		2	168	299	467
	%	1	72.5	27.5	100.0
		2	36.0	64.0	100.0
Cross Validation B	Counting	1	288	116	404
		2	168	299	467
	%	1	71.3	28.7	100.0
		2	36.0	64.0	100.0

a. The classification of 68.0% original groups was correctly classified.
b. In this paper, 67.4 per cent of the cases of cross-verification have been classified.

Tabel 24 The Female Canon Discriminating Function Coefficients

	Function
	1
Sleep quality	0.479
Sleep latency	0.114
Sleep time	0.216
Sleep efficiency	0.142
Sleep disorder	-0.281
Hypnagogue	0.258
Daytime dyfunction	-0.856
(Constants)	-0.890

Fisher's discriminant function is:

y_1=-0.890+0.479x_1+0.114x_2+0.216x_3+0.142x_4-0.281x_5+0.258x_6-0.856x_7

Tabel 25 The Female Classification Function Coefficient

	Diagnosis	
	1	2
Sleep quality	1.578	2.167
Sleep latency	1.568	1.709
Sleep time	0.143	0.409
Sleep efficiency	0.340	0.515
Sleep disorder	1.638	1.292
Hypnagogue	-0.362	-0.045
Daytime dyfunction	0.834	-0.219
(Constants)	-6.361	-7.435

Fisher linear discriminant function is:

The Bayesian discriminant function is:

F_1= -6.361+1.578x_1+1.568x_2+0.143x_3+0.340x_4+1.638x_5-0.362x_6+0.834x_7

F_2= -7.435+2.167x_1+1.709x_2+0.409x_3+0.515x_4+1.292x_5-0.045x_6-0.219x_7

Tabel 26 Female Classification Results

		Diagnosis	Predict Group Membership Information		Total
			1	2	
Primordial	Counting	1	703	233	936
		2	279	707	986
	%	1	75.1	24.9	100.0
		2	28.3	71.7	100.0
Cross validation b	Counting	1	701	235	936
		2	280	706	986
	%	1	74.9	25.1	100.0
		2	28.4	71.6	100.0

a. The classification of 73.4% of the original groups was correctly classified.

b: The classification of the 73.2 cases of cross-verification has been carried out correctly.

Conclusion: getting through rising dimension of sex and age, the results of the fourth optimization test were obtained from the perspective of gender and age, and the results showed that the correct accuracy of the male group was about 68%. And women back to the generation of determination of the type test accuracy is approximately 73.4%, further optimization and get back to the fourth generation of accuracy is approximately 72.1%, the accuracy of results of men and women back to the generation of decision for the fourth time optimization numerical swinging, this article can draw a conclusion that for the fourth time back to optimize income generation having high credibility, at the same time, the correct diagnosis and every index has high compatibility.

7 Strengths And Weaknesses

- **Strengths**

Our models can be well applied in other places and we just need to change a little specific conditions.

We come up with various criteria to compare different situations. Therefore, overall comparison can be made based on these criteria.

The data were quantitative analysis and qualitative analysis, which make the result more clear and organized.

The formation of our models is simple, where the relationship between the argument and dependent variable is clear and straightforward.

- **Weaknesses**

Although we have try our best. Time is finite and some data are missed. As a result, the missing data can still bring the errors in evaluation.

In our model, large quantities of statistics required to ensure the prediction is accurate and reliable. Thus, the model is dependent on the database to guarantee the accuracy.

Only using the title of the data for research in this paper, the amount of data is not enough and the calculation of the results has error.

8 Conclusion

As our team setting out finding a strategy that what would be the most efficient way to describe the relationship between sleep quality and various indicators, we took into major consideration was the relevance between at the first aspect. We use the SPSS to analyze the data and excluded irrelevant indicators in there.

Then we divided the data into six categories by sex and age .In the last, We use multivariate regression to calculate the relationship between argument and independent variables.

Similar models and methods appear in the second question. Only the amount of computation has increased. In addition, we analyzed the data quantitatively and qualitatively, gradually, improving the correct rate to 73.4%.

Combined with the analysis of the first three questions, we have formulated a scientific schedule to stay healthy, and using random number simulation method gives a reasonable sleep plan to us.

9 References

[1] Yan Youwei.Relationship between anxiety and sleep quality of adolescent students and its intrinsic mechanism[J].Journal of Psychological Science.2011,34(4):987-992.

[2] Zhao Peiqiu.Analysis of the Causes of Insomnia in Elderly Patients and the Countermeasures[J].National Medical Frontiers of China.2010,5(9):77-78.

[3] Liu Xianchen.Reliability and Validity of Pittsburgh Sleep Quality Index[J].Chinese Journal of Psychiatry.1996,29(2):103-107.

[4] International Classification of Diseases，https://wapbaike.baidu.com/item/%E5%9B%BD%E9%99%85%E7%96%BE%E7%97%85%E5%88%86%E7%B1%BB/11071527

[5] Chung Ka-Fai£¬Cheung Miao-miao. Sleep-wake patterns and sleep disturbance amongHong Kong Chinese adolescents[J], Sleep, 2008,31(2), 185-194.

[6] Gongziping;Xiao Jieping, sleep intervention combined with psychological behavior of refractory hypertension [J], Contemporary Medicine, 2017 (12), 13-58.

Appendix

```
// Operating environment: EXCEL;
=EXACT($C$2,C16)*EXACT($D$2,D16)*(SQRT(POWER(($B$2-B16),2))+SQRT(POWE
R(($E$2-F16),2))+SQRT(POWER(($F$2-G16),2))+SQRT(POWER(($G$2-H16),2))+SQRT(
POWER(($H$2-I16),2))+SQRT(POWER(($I$2-J16),2))+SQRT(POWER(($J$2-K16),2))+SQ
RT(POWER(($K$2-L16),2)))
%%  AUTHOR：        Contestants of APMCM;
%%  FUNCTION：      A program used for data processing on MATLAB;
%%  TIME：          25th,Nov ember,2017.
%%  Initialization;
clear
clc
%%  Basic parameters configuration;
upperlimit=5665;                    %    The number of rows to process.
diagnosis_less_limit=100;           %    Lowest limit of classification.
filename='data_clear.xlsx';         %    The original data stored in this file.
sheet='原始数据';                    %    The name of the sheet in the EXCEL.
%%  Loading the original data;
% number_init=xlsread(filename,sheet,'A2:A6350');
% age_init=xlsread(filename,sheet,'B2:B6350');
% [temp1,temp2,sex_init]=xlsread(filename,sheet,'C2:C6350');
% [temp3,source_init,temp4]=xlsread(filename,sheet,'D2:D6350');
[temp5,diagnosis_init,temp6]=xlsread(filename,sheet,'E2:E6350');
% sleep_quality_init=xlsread(filename,sheet,'F2:F6350');
% sleep_latency_init=xlsread(filename,sheet,'G2:G6350');
% sleep_time_init=xlsread(filename,sheet,'H2:H6350');
% sleep_efficiency_init=xlsread(filename,sheet,'I2:I6350');
% sleep_disorder_init=xlsread(filename,sheet,'J2:J6350');
% hypnagogue_init=xlsread(filename,sheet,'diagnosis_species 2:diagnosis_species 6350');
% daytime_dyfunction_init=xlsread(filename,sheet,'L2:L6350');
clear temp* sheet filename    %    Free up memory space.
%%  Convert sexual attributes to Boolean distribution;
% for i=1:upperlimit-1
%     if(length(sex_init{i}) == 4)
%         sex(i,1)=1;
%     else
%         sex(i,1)=0;
%     end
% end
%% Number the illness types
diagnosis_names=cell(upperlimit-1,1);
 i=1;diagnosis_species =1;flag=0;
 for m=1:upperlimit-1
     for i=1:diagnosis_species
         temp=strcmp(diagnosis_names{i},diagnosis_init{m});    %  The same is returned
1, defference is returned 0;
         if(temp)
         else
             flag=flag+1;
         end
     end
     if(flag==diagnosis_species )
                 diagnosis_names{diagnosis_species } = diagnosis_init{m};
```

```
                        diagnosis_species =diagnosis_species +1;
        end
        flag=0;
  end
diagnosis_species=diagnosis_species-1;
clear i m temp flag
%%    Calculate the number of each condition;
temp=0;flag=0;m=0;
for i=1:diagnosis_species
      for m=1:upperlimit-1
        temp=strcmp(diagnosis_names{i},diagnosis_init{m});
        if(temp == 1)
            flag=flag+1;
        end
      end
      diagnosis_number(i,1)=flag;
      flag=0;
end
clear temp i flag m
%%    The explanation format of the conversion disorder is a character;
%%    Find out the rare ones and find the quantity;
for i=1:diagnosis_species
      if(diagnosis_number(i,1) <= diagnosis_less_limit)
            boolean_evaluation(i,1)=1;
      else
            boolean_evaluation(i,1)=0;
      end
end
clear i
other_diagnosis_species=sum(boolean_evaluation);
other_diagnosis_number=sum(boolean_evaluation.*diagnosis_number);
%%      call any required information;
% for i=1:diagnosis_species
%         if(temp(i,1)==1)
%               number_less(i,1)=number_init(i,1);
%               age_less(i,1)=age_init(i,1);
%               sex_less(i,1)=sex_(i,1);
%               source_less{i}=source_init{i};
%               diagnosis_less{i}=diagnosis_init{i};
%               sleep_quality_less(i,1)=sleep_quality_init(i,1);
%               sleep_latency_less(i,1)=sleep_latency_init(i,1);
%               sleep_time_less(i,1)=sleep_time_init(i,1);
%               sleep_efficiency_less(i,1)=sleep_efficiency_init(i,1);
%               sleep_disorder_less(i,1)=sleep_disorder_init(i,1);
%               hypnagogue_less(i,1)=hypnagogue_init(i,1);
%               daytime_dyfunction_less(i,1)=daytime_dyfunction_init(i,1);
%         end
% end
% clear temp
%%    Classify and make pie charts by number;
%     Processing data;
k=1;
data=(~boolean_evaluation).*diagnosis_number;      %      Input data
final_classification_species=sum(~boolean_evaluation)+1;
for i=1:diagnosis_species
```

```
    if(data(i) == 0)
    else
        data_useful(k,1)=diagnosis_number(i,1);
        k=k+1;
    end
end
final_classification_numbers=[data_useful;other_diagnosis_number];
%%  Make sure the names of classification;
k=1;
for i=1:diagnosis_species
    if(boolean_evaluation(i) == 0)
        final_classification_names{k}=diagnosis_names{i};
        k=k+1;
    end
end
final_classification_names{final_classification_species}='Other Diagnosises';
clear i k data data_useful
label=final_classification_names;    %    Enter the label;
data=final_classification_numbers'; %    Define the salient parts;
for i=1:final_classification_species

if(final_classification_numbers(i,1)>(sum(final_classification_numbers)/final_classification_s
pecies))
        explode(i,1)=1;
    else
        explode(i,1)=0;
    end
end
explode=explode';
bili=data/sum(data);                    %    Calculate the ratio;
baifenbi=num2str(bili'*100,'%1.2f');%    Calculate the percentage;
baifenbi=[repmat(blanks(2),length(data),1),baifenbi,repmat('%',length(data),1)];
baifenbi=cellstr(baifenbi);
Label=strcat(label,baifenbi');
%%  Drawing;
pie(data,explode,Label)
clear i label Label explode data bili baifenbi data_useful
%% Output analysis result.
disp('According to the data you provided, the final result of the output analyzed by the
mathematical modelis：')
disp('Each categorie of disorders is：')
final_classification_species'
disp('Each names of the categories is：')
final_classification_names
disp('Each number of patients with disease is')
final_classification_numbers'
disp('The resulting of pie chart is shown in the window.')
%%  AUTHOR：        Contestants of APMCM;
%%  FUNCTION：      A program used for data processing on MATLAB;
%%  TIME：          25th,Nov ember,2017.
clc,clear, close all;
data=load('shuju.txt'); %Enter the text document shuju.txt in the workspace named data form
plot(data(:,1),'r')
ylabel('Reliability');
```

```
legend('Reliability');
figure;
plot(data(:,2))
ylabel('Psychoticism');
legend('Psychoticism');
figure;
plot(data(:,3),'k')
ylabel('Nervousness');
legend('Nervousness');
figure;
plot(data(:,4),'g')
ylabel('Character');
legend('Character');
%%  AUTHOR：        Contestants of APMCM;
%%  FUNCTION：      A program used for data processing on MATLAB;
%%  TIME：          25th,Nov ember,2017.
x1=data(:,1);
x2=data(:,2);
x3=data(:,3);
x4=data(:,4);
x5=data(:,5);
x6=data(:,6);
x7=data(:,7);
y=data1(:,1);
x=[x1 x2 x3 x4 x5 x6 x7];

stepwise(x,y)
%%  AUTHOR：        Contestants of APMCM;
%%  FUNCTION：      A program used for data processing on MATLAB;
%%  TIME：          25th,Nov ember,2017.
x1=data(:,1);
x2=data(:,2);
x3=data(:,3);
x4=data(:,4);
x5=data(:,5);
x6=data(:,6);
x7=data(:,7);
y=data1(:,1);
x=[x1 x2 x3 x4 x5 x6 x7];
X=[ones(1623,1) x1 x2 x3 x4 x5 x6 x7];
    b=regress(y,X)
%%  AUTHOR：        Contestants of APMCM;
%%  FUNCTION：      A program used for data processing on MATLAB;
%%  TIME：          25th,Nov ember,2017.
clc,clear, close all;
data=load('yuchuli.txt'); %Enter the text document shuju.txt in the workspace named data
form
plot(data(:,1),'r')
ylabel('Reliability');
legend('Reliability');
figure;
plot(data(:,2))
ylabel('Psychoticism');
legend('Psychoticism');
```

```
figure;
plot(data(:,3),'k')
ylabel('Nervousness');
legend('Nervousness');
figure;
plot(data(:,4),'g')
ylabel('Character');
legend('Character');
```