# 2022 Mathematical Modeling Competition for College Students Essay

Thesis title: Optimization Model for Alzheimer's Disease Identification

Team number: 2022092412600

# Optimized model for Alzheimer's disease identification


## Abstract:

In this paper, we discuss the problem of age, gender, marriage and other data characteristics and Alzheimer's disease diagnosis, using descriptive statistics for overall analysis and Pearson's algorithm for correlation analysis between data characteristics and Alzheimer's disease, using the accompanying structural brain characteristics and cognitive-behavioral characteristics data to build an XGBoost model with training set data to design an intelligent diagnosis of Alzheimer's disease. Then, for the three subclasses included in MCI (SMC, EMCI and LMCI ), the clustering was continued to refine into three subclasses using the K-means algorithm with dosage outlining, and their relationship with time points was analyzed using time series based on the annexed features included in the collection at different time points to reveal the patterns of different classes of diseases evolving over time. Finally five types of early intervention and diagnostic criteria for CN, SMC, EMCI, LMCI and AD are described.

In response to question 1, this paper used the preprocessing of the data in the Appendix first, the KMO and Bartlett's test to determine whether the principal component analysis, the analysis variance explanation table and the gravel plot were obtained, and then the statistical indicators of the five data AGE, APOE4, CDRSB_bl, ADAS11_bl, ADASQ4_bl, and ADAS13_bl were The overall descriptive analysis was performed and obtained by normality test. Since the normal distribution was satisfied so correlation analysis was performed on these data using Pearson's algorithm and the results were reached indicating a strong correlation.

For question 2, from the attached data of structural brain features and cognitive-behavioral features, the features obtained by correlation analysis in the first question were used for classification, and the XGBoost regression model was built through the training set data, and then the feature importance was calculated through the established XGBoost, and the model evaluation result was obtained: R2, indicating that the model is highly accurate, and then the prediction results were obtained from the data (see the figure of question 2 for details)

In the clustering algorithm, the K-MEANS clustering algorithm is used to calculate the Euclidean distance, and the three subclasses (SMC, EMCI, and LMCI) contained in MCI are refined into three subclasses according to the minimum distance, and the central object of each cluster is recalculated until each cluster no longer changes, and finally the number of SMC, EMCI, and LMCI is obtained.

For problem four, first data preprocessing, as spss cannot handle variables of character type, so dummy variables were created for gender and person type. The two variables female-1; female-2 were created, while later time variables were created and time series graphs were drawn (as shown), and the model results were obtained using additive time series analysis: the seasonal cycle length of MOD_3, the calculation of the moving average, and the seasonal factor.

For question five, five types of early interventions and diagnostic criteria for CN, SMC, EMCI, LMCI, and AD were found by referring to the relevant literature. Early intervention is mainly psychiatric and psychological intervention, and if the symptoms are severe, medication is needed to intervene, while the diagnostic criteria are behavioral, memory, and emotional changes in the elderly, as well as international diagnostic criteria such as NIA-AA for diagnosis and clinical observation.

**Keywords: Alzheimer's disease identification, correlation analysis, XGBoost model, K-means algorithm**

# 一、Restatement of the problem

## 1.1 Research Background and Significance

Alzheimer's disease (AD), commonly known as dementia, is a chronic neurodegenerative disease with an insidious onset, and most patients are over 60 years old. According to statistics, there are more than 7 million people suffering from AD in China, with the prevalence rate of 5.6% in people over 65 years old and up to 20% in people over 85 years old. It is the world's largest and fastest growing population with AD, which brings a heavy burden to patients, families, society and medical care in China. According to data in the journal Neurology, more than 500,000 patients a year die from Alzheimer's disease (AD), and the massive brain cell death caused by Alzheimer's disease is irreversible and therefore needs to be closely prevented.

Therefore, it is important to assess the structural and cognitive-behavioral characteristics of the brain for the accurate diagnosis of Alzheimer's disease, and to provide adjuvant therapy on the side.

Since the elderly present a complex situation in terms of gender (male and female), age (50-90), and marital status (divorced or not). This thesis is based on the data given in the competition and other relevant data to diagnose the type of Alzheimer's disease in the annex and reveal the evolution of different categories of disease over time, aiming at early intervention and diagnostic criteria for patients

## 1.2 Problem formulation

This paper will address the following questions.

(1) Preprocessing the feature indicators of the attached data and investigating the correlation between the data features and the diagnosis of Alzheimer's disease.

(2) To design an intelligent diagnosis of Alzheimer's disease using the attached structural brain features and cognitive-behavioral features .

(3) First, CN, MCI and AD are clustered into three major categories. Then, for the three subclasses contained in MCI (SMC, EMCI and LMCI), the clustering continues to be refined into three subclasses.

(4) The same samples in the Appendix contain characteristics collected at different time points; please analyze them in relation to the time points to reveal the evolutionary patterns of different categories of diseases over time.

(5) Please review the relevant literature to describe the early intervention and diagnostic criteria for the five categories of patients with CN, SMC, EMCI, LMCI, and AD

# 二、Analysis of the problem

For each of the five problems presented in this paper, we do the following analysis.

Analysis of Problem 1: Problem 1 requires pre-processing of the feature indicators of the data. In addition, the features given in the table should be filtered by using the explanatory information of the given documents to eliminate redundant features, and then the correlation should be calculated by using Pearson's correlation coefficient.

Analysis of Problem 2: Using the features obtained from the correlation analysis in the first problem, the XGBoost regression model is built and the feature importance is calculated from the training set data. The XGBoost regression model was applied to the training and testing data to obtain the model evaluation results. Since XGBoost has randomness, the result of each operation is not the same, if this training model is saved, the subsequent data can be directly uploaded to this training model for calculation of prediction.

Analysis of Problem 3: Firstly, all the data are quantified and unified, and the quantified data are clustered and analyzed. In the clustering algorithm is to use the K-MEANS clustering algorithm to calculate the Euclidean distance, and to recalculate the three subclasses (SMC, EMCI and LMCI) contained in MCI according to the minimum distance, and the corresponding objects are refined into three subclasses, and the center of each cluster is recalculated objects until no more changes occur in each cluster.

Analysis of problem four: Since spss cannot handle variables of character type, dummy variables were created for gender and person type, time variables were created, and time series graphs were drawn for time series analysis to obtain seasonal factors for each quarter

Analysis of problem five: Relevant literature was reviewed to describe the early intervention and diagnostic criteria for the five types of CN, SMC, EMCI, LMCI, and AD.

# 三、Modeling and solving

## 4.1 Modeling and Solution of Problem 1

Question 1 asked to preprocess the characteristic indicators of the attached data and to investigate the correlation between the data characteristics and the diagnosis of Alzheimer's disease. In this paper, we first performed an overall descriptive analysis of each statistical indicator for the five data overall, AGE, APOE4, CDRSB_bl, ADAS11_bl,

ADASQ4_bl, and ADAS13_bl. The results obtained were checked for normal distribution, and the Shapiro-Wilk test was performed on the data to check their significance. Then Pearson correlation coefficient hypothesis test was performed, and finally correlation analysis was performed with spss.

### 4.1.1Pre-processing of data

(1) Data processing

A. Analyze the data in the annex for outliers, missing values and other parts that affect the modeling results, and find that the data do not have such problems.

B. Since the missing data of "age" in Annex 2 is not easy to fill, and the amount of data is huge, the missing data of age is deleted.

The missing data of "age" in Annex 2 is not easy to fill and the data volume is large, so the missing data of age is deleted.

C. The indicators with missing values greater than 50% were deleted, and the indicators with missing values less than 50% were filled by EM estimation. (See Appendix for details)

D. First, KMO and Bartlett's test were performed to determine whether principal component analysis could be performed. The two most dominant data sets (d1,d2) were selected by principal component analysis.

(2) Principal component analysis

First, KMO and Bartlett's test were performed to determine whether principal component analysis could be performed.

A. For the KMO value: between 0.7-0.8 is generally suitable, for Bartlett's test, because P is less than 0.05, rejecting the original hypothesis, it means that the principal component analysis can be done .The result of KMO test shows that the value of KMO is 0.771, meanwhile, the result of Bartlett's spherical test shows that the significance P value is 0.000***, the level presents significance, the original hypothesis is rejected, there is correlation between the variables, and the principal component analysis is valid to an average degree.

**KMO test and Bartlett's test**

**KMO 和巴特利特检验**

| | | |
|---|---|---|
| KMO 取样适切性量数。 | | .835 |
| 巴特利特球形度检验 | 近似卡方 | 84465.213 |
| | 自由度 | 325 |

| | |
|---|---|
| 显著性 | .000 |

4.1.4 Figure 1 KMO test and Bartlett's test plot

KMO = 0.835 Good for principal component analysis

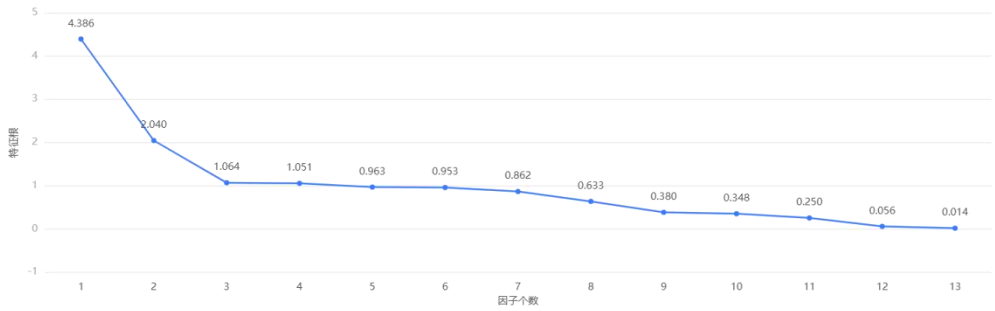
B. By analyzing the variance interpretation table and the gravel plot, the variance interpretation table of the number of principal components mainly looks at the contribution of the principal components to the explanation of variables. The number of principal components to be selected is confirmed by the slope of the decline in eigenvalues in the gravel plot, and the combination of these two confirms or adjusts the number of principal components. According to the analysis it can be obtained that at principal component 5, the eigenroot of the total variance explained is below 1.0 and the contribution of the variables explained reaches 73.109 .

**Table of Variance Explanation**

| Total variance explained | | | |
|---|---|---|---|
| Ingredients | 特征根 Characteristic roots | Explana of variance(%) | Cumulative variance explain(%) |
| 1 | 4.386 | 33.74 | 33.74 |
| 2 | 2.04 | 15.691 | 49.431 |
| 3 | 1.064 | 8.188 | 57.619 |
| 4 | 1.051 | 8.082 | 65.701 |
| 5 | 0.963 | 7.408 | 73.109 |
| 6 | 0.953 | 7.333 | 80.442 |
| 7 | 0.862 | 6.628 | 87.07 |
| 8 | 0.633 | 4.87 | 91.94 |
| 9 | 0.38 | 2.925 | 94.865 |
| 10 | 0.348 | 2.678 | 97.543 |
| 11 | 0.25 | 1.924 | 99.467 |
| 12 | 0.056 | 0.427 | 99.894 |
| 13 | 0.014 | 0.106 | 100 |

4.4.1 Figure2 explanation of variance table chart

**Gravel map**




4.1.1 Gravel map

C. The importance of the hidden variables in each principal component can be analyzed by analyzing the principal component loading coefficients and heat maps. The hidden variable analysis of each

principal component can be combined with specific business. Based on the principal component loading diagram by reducing the dimensionality of multiple principal components into two-principal components or three-principal components.

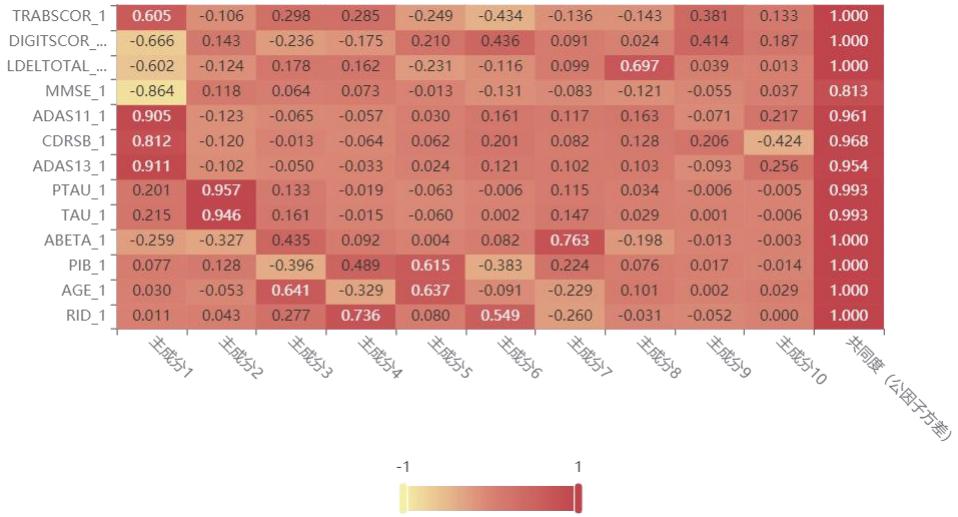
**Table of factor loading coefficients**

Table of factor loading factors

| | Factor loading factor | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Principal Components 1 | Principal Components 2 | Principal Components 3 | Principal Components 4 | Principal Components 5 | Principal Components 6 | Principal Components 7 | Principal Components 8 | Principal Components 9 | Principal Components 10 | Commonality (common factor variance) |
| RID_1 | 0.011 | 0.043 | 0.277 | 0.736 | 0.08 | 0.549 | -0.26 | -0.031 | -0.052 | 0.000 | 1 |
| AGE_1 | 0.03 | -0.053 | 0.641 | -0.329 | 0.637 | -0.091 | -0.229 | 0.101 | 0.002 | 0.029 | 1 |
| PIB_1 | 0.077 | 0.128 | -0.396 | 0.489 | 0.615 | -0.383 | 0.224 | 0.076 | 0.017 | -0.014 | 1 |
| ABETA_1 | -0.259 | -0.327 | 0.435 | 0.092 | 0.004 | 0.082 | 0.763 | -0.198 | -0.013 | -0.003 | 1 |
| TAU_1 | 0.215 | 0.946 | 0.161 | -0.015 | -0.06 | 0.002 | 0.147 | 0.029 | 0.001 | -0.006 | 0.993 |
| PTAU_1 | 0.201 | 0.957 | 0.133 | -0.019 | -0.063 | -0.006 | 0.115 | 0.034 | -0.006 | -0.005 | 0.993 |
| ADAS13_1 | 0.911 | -0.102 | -0.05 | -0.033 | 0.024 | 0.121 | 0.102 | 0.103 | -0.093 | 0.256 | 0.954 |
| CDRSB_1 | 0.812 | -0.12 | -0.013 | -0.064 | 0.062 | 0.201 | 0.082 | 0.128 | 0.206 | -0.424 | 0.968 |
| ADAS11_1 | 0.905 | -0.123 | -0.065 | -0.057 | 0.03 | 0.161 | 0.117 | 0.163 | -0.071 | 0.217 | 0.961 |
| MMSE_1 | -0.864 | 0.118 | 0.064 | 0.073 | -0.013 | -0.131 | -0.083 | -0.121 | -0.055 | 0.037 | 0.813 |
| LDELTOTAL_1 | -0.602 | -0.124 | 0.178 | 0.162 | -0.231 | -0.116 | 0.099 | 0.697 | 0.039 | 0.013 | 1 |
| DIGITSCOR_1 | -0.666 | 0.143 | -0.236 | -0.175 | 0.21 | 0.436 | 0.091 | 0.024 | 0.414 | 0.187 | 1 |
| TRABSCOR_1 | 0.605 | -0.106 | 0.298 | 0.285 | -0.249 | -0.434 | -0.136 | -0.143 | 0.381 | 0.133 | 1 |

4.1.1 Figure 3 Table of factor loading factors

**Factor load matrix heat map**




4.1.1 Fig. 4 Thermal diagram of the factor load matrix

D. The spatial distribution of principal components is presented by means of a quadrant diagram. The principal component composition formula and weights are derived by analyzing the component matrix.

**Table of component matrix**

Component Matrix

| Name | Ingredients | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Ingredients 1 | Ingredients 2 | Ingredients 3 | Ingredients 4 | Ingredients 5 | Ingredients 6 | Ingredients 7 | Ingredients 8 | Ingredients 9 | Ingredients 10 |
| RID_1 | 0.002 | 0.021 | 0.26 | 0.701 | 0.083 | 0.576 | -0.302 | -0.049 | -0.138 | 0 |
| AGE_1 | 0.007 | -0.026 | 0.602 | -0.313 | 0.661 | -0.095 | -0.266 | 0.16 | 0.006 | 0.084 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| PIB_1 | 0.018 | 0.063 | -0.372 | 0.466 | 0.638 | -0.402 | 0.26 | 0.12 | 0.044 | -0.039 |
| ABETA_1 | -0.059 | -0.16 | 0.408 | 0.088 | 0.004 | 0.086 | 0.886 | -0.312 | -0.033 | -0.01 |
| TAU_1 | 0.049 | 0.464 | 0.151 | -0.014 | -0.063 | 0.002 | 0.17 | 0.046 | 0.002 | -0.018 |
| PTAU_1 | 0.046 | 0.469 | 0.125 | -0.019 | -0.065 | -0.006 | 0.134 | 0.054 | -0.016 | -0.015 |
| ADAS13_1 | 0.208 | -0.05 | -0.047 | -0.031 | 0.025 | 0.126 | 0.119 | 0.162 | -0.245 | 0.736 |
| CDRSB_1 | 0.185 | -0.059 | -0.012 | -0.061 | 0.064 | 0.211 | 0.095 | 0.202 | 0.542 | -1.218 |
| ADAS11_1 | 0.206 | -0.06 | -0.061 | -0.054 | 0.031 | 0.168 | 0.136 | 0.258 | -0.186 | 0.624 |
| MMSE_1 | -0.197 | 0.058 | 0.06 | 0.069 | -0.014 | -0.138 | -0.096 | -0.191 | -0.144 | 0.108 |
| LDELTOTAL_1 | -0.137 | -0.061 | 0.167 | 0.154 | -0.24 | -0.122 | 0.115 | 1.101 | 0.102 | 0.038 |
| DIGITSCOR_1 | -0.152 | 0.07 | -0.222 | -0.167 | 0.218 | 0.457 | 0.105 | 0.038 | 1.089 | 0.537 |
| TRABSCOR_1 | 0.138 | -0.052 | 0.28 | 0.271 | -0.259 | -0.456 | -0.157 | -0.225 | 1.002 | 0.383 |

Figure 5 Component matrix

**Factor weighting analysis**

| Name | Explanation of variance(%) | Cumulative variance explained(%) | Weighting(%) |
|---|---|---|---|
| 主成分1 | 33.74 | 33.74 | 34.59 |
| 主成分2 | 15.691 | 49.431 | 16.086 |
| 主成分3 | 8.188 | 57.619 | 8.394 |
| 主成分4 | 8.082 | 65.701 | 8.286 |
| 主成分5 | 7.408 | 73.109 | 7.595 |
| 主成分6 | 7.333 | 80.442 | 7.518 |
| 主成分7 | 6.628 | 87.07 | 6.795 |
| 主成分8 | 4.87 | 91.94 | 4.993 |
| 主成分9 | 2.925 | 94.865 | 2.999 |
| 主成分10 | 2.678 | 97.543 | 2.746 |

4.1.1 Figure 6 Factor weighting analysis diagram

## 4.1.2Descriptive statistics

First, an overall descriptive analysis of each statistical indicator was performed first for the five data, AGE, APOE4, CDRSB_bl, ADAS11_bl, ADASQ4_bl, and ADAS13_bl. Secondly, we analyze the indicators that are abnormal or show more prominent performance, such as high variance, high mean, etc.

**Description of the algorithm:**

Descriptive statistics are activities that describe the characteristics of data using tabulations and classifications, graphs, and the calculation of generalized data. Descriptive statistical analysis involves the statistical description of data related to all variables in the survey population, including frequency analysis, concentration trend analysis, dispersion analysis, distribution, and some basic statistical graphics.

**Data output results:**

| Variable name | Sample size | Maximum value | Minimum value | Average value | Standard deviation | Median | Variance | Kurtosis | Skewness | Coefficient of variation of variance (CV) |
|---|---|---|---|---|---|---|---|---|---|---|
| AGE | 15898 | 91.4 | 54.4 | 73.313 | 6.98 | 73.4 | 48.717 | -0.326 | -0.137 | 0.09520500540450336 |
| APOE4 | 15898 | 2 | 0 | 0.517 | 0.648 | 0 | 0.419 | -0.329 | 0.874 | 1.2522732623650317 |
| CDRSB_bl | 15898 | 10 | 0 | 1.219 | 1.516 | 0.58 | 2.299 | 3.719 | 1.78 | 1.2440662869477734 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ADAS11_bl | 15898 | 42.67 | 0 | 9.269 | 5.782 | 8 | 33.435 | 2.615 | 3<br>1.36<br>3 | 0.6238008568094767 |
| ADAS13_bl | 15898 | 54.67 | 0 | 14.67 | 8.627 | 13 | 74.426 | 0.988 | 0.99<br>6 | 0.5880912940177317 |
| ADASQ4_bl | 15898 | 10 | 0 | 4.694 | 2.843 | 4 | 8.083 | -0.924 | 0.33<br>4 | 0.6056979238743139 |

4.1.2 Figure 6 d1 data output graph

**Data analysis:**

Based on AGE, the coefficient of variation (CV) is 0.095, which is less than 0.15, and there is a small probability of outliers in the current data, and the mean value is used for descriptive analysis. Based on APOE4, the coefficient of variation (CV) is 1.252, which is greater than 0.15. There may be outliers in the current data, and the indicators that are abnormal or have a more prominent performance are analyzed. Based on CDRSB_bl, the coefficient of variation (CV) is 1.244, which is greater than 0.15. There may be abnormal values in the current data, and the indicators that are abnormal or more prominent are analyzed. Based on ADAS11_bl, the coefficient of variation (CV) is 0.624, which is greater than 0.15. There may be abnormal values in the current data, and the indicators with abnormal or prominent performance will be analyzed. Based on ADAS13_bl, the coefficient of variation (CV) is 0.588, which is greater than 0.15. There may be abnormal values in the current data, and the indicators that are abnormal or have outstanding performance are analyzed. Based on ADASQ4_bl, the coefficient of variation (CV) is 0.606, which is greater than 0.15. There may be abnormal values in the current data, and the indicators with abnormal or outstanding performance will be analyzed.

| Variable name | Sample size | Maximum value | Minimum value | Average value | Standard deviation | Median | Variance | Kurtosis | Skewness | Coefficient of variation of variance (CV) |
|---|---|---|---|---|---|---|---|---|---|---|
| APOE4 | 15898 | 2 | 0 | 0.517 | 0.648 | 0 | 0.419 | -0.329 | 0.874 | 1.2522732623650317 |
| ADAS13_1 | 15898 | 85 | 0 | 16.751 | 9.556 | 16.735 | 91.309 | 5.132 | 1.643 | 0.5704426271081451 |
| ADASQ4_bl_1 | 15898 | 10 | 0 | 4.694 | 2.843 | 4 | 8.083 | -0.924 | 0.334 | 0.6057042836658916 |
| ADAS11_1 | 15898 | 70 | 0 | 10.889 | 7.019 | 10.875 | 49.266 | 9.922 | 2.368 | 0.6445885891629438 |

4.1.2 Figure 7 d2 data output graph

**数据分析：**

Based on APOE4, the coefficient of variation (CV) is 1.252, which is greater than 0.15. There may be outliers in the current data, and the indicators that are abnormal or have a more prominent performance are analyzed. Based on ADAS13_1, the coefficient of variation (CV) is 0.57, which is greater than 0.15. There may be abnormal values in the current data, and the indicators with abnormal or prominent performance will be analyzed. Based on ADASQ4_bl_1, the coefficient of variation (CV) is 0.606, which is greater than 0.15. There may be abnormal values in the current data, and the indicators that are abnormal or have outstanding

performance are analyzed. Based on ADAS11_1, the coefficient of variation (CV) is 0.645, which is greater than 0.15. There may be abnormal values in the current data, and the indicators with abnormal or prominent performance will be analyzed.

Chart description.

The above table shows the results of descriptive statistics, including sample size, maximum value, minimum value and other statistics, which are used to study the overall situation of quantitative data.

1. analyze each statistical indicator and perform an overall descriptive analysis of each statistical indicator.

2. Analyze the indicators that are abnormal or show more prominence, such as high variance, high mean, etc.

### 4.1.4Normal distribution calibration

Shapiro-Wilk (test was performed on the data to check its significance. If it does not show significance (P>0.05), it means that it meets the normal distribution, and vice versa means that it does not meet the normal distribution (PS: it is usually difficult to meet the test in real research situations, if the absolute value of its sample kurtosis is less than 10 and the absolute value of skewness is less than combined with the normal distribution histogram, PP plot or QQ plot can be described as basically meeting the normal distribution).

**Algorithm description.**

Kolmogorov-Smirnov is a test that compares a frequency distribution f(x) with a theoretical distribution g(x) or with the distribution of two observations. Its original hypothesis H0:the two data distributions agree or the data conform to the theoretical distribution. d=max| f(x)-g(x)|, when the actual observation D>D(n,α) then H0 is rejected, otherwise the H0 hypothesis is accepted.

The KS test differs from other methods like the t-test in that the KS test does not require knowledge of the distribution of the data and can be considered a nonparametric test. Of course, the cost of this convenience is that when the distribution of the data tested conforms to a specific distribution, the sensitivity of the KS test is not as high as the corresponding test. When the sample size is relatively small, the KS test is the most non-parametric test is quite commonly used to analyze whether two sets of data are different from each other.

**Overall description of the results.:**

| Variable name | Variable name | Median | Average | Standard deviation | Skewness | Kurtosis | S-W test |
|---|---|---|---|---|---|---|---|
| ADAS11 | 2421 | 8.67 | 10.92 | 8.125 | 1.872 | 5.513 | 0.85(0.000***) |
| ADAS13 | 2421 | 14 | 16.883 | 11.361 | 1.331 | 2.523 | 0.905(0.000***) |
| ADASQ4 | 2421 | 5 | 5.059 | 3.063 | 0.204 | -1.109 | 0.943(0.000***) |
| MMSE | 2421 | 28 | 26.835 | 3.478 | -1.86 | 5.528 | 0.82(0.000***) |

| | | | | | | |
|---|---|---|---|---|---|---|
| mPACCdigit | 2421 | -4.313 | -6.152 | 7.617 | -1.122 | 2.183 | 0.929(0.000***) |
| mPACCtrailsB | 2421 | -3.843 | -5.724 | 7.186 | -1.174 | 2.716 | 0.924(0.000***) |

4.1.4 Figure 8 d1 description results

**Chart description.:**

The above table shows the results of ADAS11, ADAS13, ADASQ4, MMSE, mPACCdigit, mPACCtrailsB descriptive statistics and normality tests, including median, mean, etc., for testing the normality of the data.

1. There are usually two tests for normal distribution, one is the Shapiro-Wilk test for small sample data (sample size ≤ 5000) and the other is the Kolmogorov-Smirnov test for large sample data (sample size > 5000).

2. If it presents significance (P<0.05), it means that the original hypothesis is rejected (the data meets the normal distribution) and the data does not satisfy the normal distribution, and vice versa.

**Chart Analysis:**

ADAS11 sample N < 5000, using S-W test, significance P-value is 0.000***, level presents significance, ADAS13 sample N < 5000, using S-W test, significance P-value is 0.000***, level presents significance, ADASQ4 sample N < 5000, using S-W test, significance P-value is 0.000***, level presents significance, MMSE sample N < 5000, using S-W test, significance P-value is 0.000***, level presents significance, mPACCdigit sample N < 5000, using S-W test, significance P-value is 0.000***, level presents significance. level presents significance, MMSE sample N < 5000, using S-W test, significance P-value is 0.000***, level presents significance, mPACCdigit sample N < 5000, using S-W test, significance P-value is 0.000***, level presents significance. mPACCtrailsB sample N < 5000, using S-W test , the significance P-value is 0.000*** and the level presents significance.

| Variable name | Variable name | Median | Average | Standard deviation | Skewness | Kurtosis | S-W test |
|---|---|---|---|---|---|---|---|
| MMSE_bl | 2421 | 28 | 27.375 | 2.653 | -1.119 | 0.624 | 0.862(0.000***) |
| ADAS13_bl | 2421 | 14 | 15.894 | 9.567 | 0.923 | 0.601 | 0.938(0.000***) |
| ADASQ4_bl | 2421 | 5 | 4.981 | 2.952 | 0.216 | -1.081 | 0.943(0.000***) |
| mPACCtrailsB_bl | 2421 | -3.668 | -5.065 | 5.916 | -0.581 | -0.512 | 0.955(0.000***) |
| mPACCdigit_bl | 2421 | -4.192 | -5.441 | 6.246 | -0.58 | -0.485 | 0.957(0.000***) |
| ADAS11_bl | 2421 | 8.67 | 10.106 | 6.538 | 1.294 | 1.991 | 0.905(0.000***) |

4.1.4 Fig. 9 General description diagram of d2

**Chart Description:**

The above table shows the results of MMSE_bl, ADAS13_bl, ADASQ4_bl, mPACCtrailsB_bl, mPACCdigit_bl, and ADAS11_bl descriptive statistics and normality tests, including median and mean, for testing the normality of the data.

1. There are usually two tests for normal distribution, one is the Shapiro-Wilk test for small sample information (sample size ≤ 5000) and the other is the Kolmogorov-Smirnov test for large sample information (sample size > 5000).

2. If it presents significance (P<0.05), it means that the original hypothesis is rejected (the data meets the normal distribution) and the data does not meet the normal distribution, and vice versa.

PS: It is usually difficult to meet the test in realistic research situations. If the absolute value of its sample kurtosis is less than 10 and the absolute value of skewness is less than 3, combined with the histogram of normal distribution, PP chart or QQ chart can be described as basically meeting the normal distribution.

Graphical analysis：

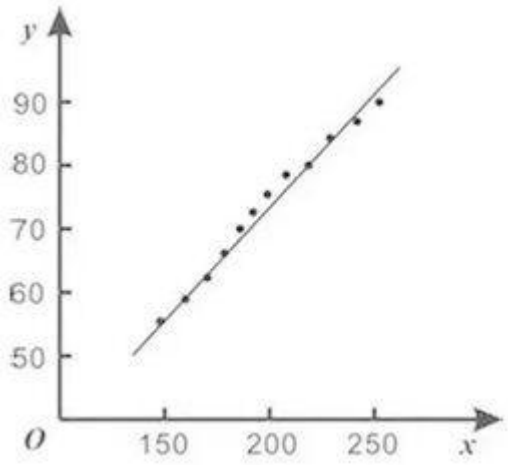
MMSE_bl sample N < 5000, using S-W test, significance P-value is 0.000***, level presents significance , ADAS13_bl sample N < 5000, using S-W test, significance P-value is 0.000***, level presents significance , ADASQ4_bl sample N < 5000, using S-W test, significance P-value is 0.000***, level presents significance , mPACCtrailsB_bl sample N < 5000, using S-W test, significance P-value is 0.000***, level presents significance 0.000***, the level presents significance, mPACCtrailsB_bl sample N < 5000, using S-W test, the significance P-value is 0.000***, the level presents significance , mPACCdigit_bl sample N < 5000, using S-W test, the significance P-value is 0.000***, the level presents significance ADAS11_bl sample N < 5000, using S-W test, significance P-value is 0.000***, level presents significance

## 4.1.5Pearson correlation coefficient correlation analysis

First, we test whether there is a statistically significant relationship (P<0.05) between XY, analyze the positive and negative direction of the correlation coefficient as well as the degree of Pearson correlation coefficient correlation, and then summarize the results of the analysis.

Algorithm introduction：

Pearson product moment correlation coefficients (also known as PPMCC or PCCs) are used to measure the correlation (linear correlation) between two variables X and Y, with a value between -1 and 1. This coefficient is widely used to measure the correlation between two variables




4.1.5 Pearson coefficient diagram

Table of relevant factors.

| | ADAS11 | ADAS13 | ADASQ4 | MMSE | mPACCdigit | mPACCtrailsB |
|---|---|---|---|---|---|---|
| ADAS11 | 1(0.000***) | 0.982(0.000***) | 0.766(0.000***) | -0.831(0.000***) | -0.882(0.000***) | -0.877(0.000***) |
| ADAS13 | 0.982(0.000***) | 1(0.000***) | 0.862(0.000***) | -0.833(0.000***) | -0.919(0.000***) | -0.915(0.000***) |
| ADASQ4 | 0.766(0.000***) | 0.862(0.000***) | 1(0.000***) | -0.675(0.000***) | -0.854(0.000***) | -0.843(0.000***) |
| MMSE | -0.831(0.000***) | -0.833(0.000***) | -0.675(0.000***) | 1(0.000***) | 0.928(0.000***) | 0.923(0.000***) |
| mPACCdigit | -0.882(0.000***) | -0.919(0.000***) | -0.854(0.000***) | 0.928(0.000***) | 1(0.000***) | 0.982(0.000***) |
| mPACCtrailsB | -0.877(0.000***) | -0.915(0.000***) | -0.843(0.000***) | 0.923(0.000***) | 0.982(0.000***) | 1(0.000***) |

Note: ***, **, * represent 1%, 5%, 10% level of significance respectively

4.1.5 Coefficient Table 1

| | CDRSB_bl | ADAS11_bl | ADAS13_bl | ADASQ4_bl | MMSE_bl |
|---|---|---|---|---|---|
| CDRSB_bl | 1(0.000***) | 0.731(0.000***) | 0.75(0.000***) | 0.649(0.000***) | -0.724(0.000***) |
| ADAS11_bl | 0.731(0.000***) | 1(0.000***) | 0.976(0.000***) | 0.772(0.000***) | -0.728(0.000***) |
| ADAS13_bl | 0.75(0.000***) | 0.976(0.000***) | 1(0.000***) | 0.878(0.000***) | -0.748(0.000***) |
| ADASQ4_bl | 0.649(0.000***) | 0.772(0.000***) | 0.878(0.000***) | 1(0.000***) | -0.655(0.000***) |
| MMSE_bl | -0.724(0.000***) | -0.728(0.000***) | -0.748(0.000***) | -0.655(0.000***) | 1(0.000***) |

Note: ***, **, * represent 1%, 5%, 10% level of significance respectively

4.1.5 Coefficient Table 2

**Chart description.:**

The above table shows the table of the results of the parameters of the model test, including the correlation coefficient, and the significant P-value.

1. The existence of a statistically significant relationship between XY is first tested to determine whether the P-value presents significance (P<0.05).

2. If it presents significance, it means that there is a correlation between the two variables, and vice versa, there is no correlation between the two variables.

3. analyze the positive and negative direction of the correlation coefficient and the degree of correlation.

**Correlation coefficient heat map:**


4.1.5 Fig. 10 Heat map of d1 correlation coefficient


4.1.5 Fig. 11 Heat map of d2 correlation coefficient

**Chart Description:**

The above figure shows the value of the correlation coefficient in the form of a heat map, mainly by color shades to indicate the magnitude of the value.

## 4.2Modeling and solving Problem 2

Intelligent diagnosis of Alzheimer's disease is designed from the attached data of structural brain features and cognitive-behavioral features . First, the features obtained by correlation analysis using the first question are classified. The training set data is used to build an XGBoost regression model to calculate the feature importance. The established XGBoost regression model is applied to the training and testing data to obtain the model evaluation results.

### 4.2.1Build XGBoost regression models from training set data

**Algorithm Introduction**

Training dataset and testing dataset are two concepts in the field of machine learning, which arise from different ways of data slicing. Common practice: when slicing the original data, 80% of the original data is used as training data to train the model, and the other 20% is used as test data to directly judge the effect of the model through test data, and continuously improve the model before it enters the real environment.

**Model Parameters**

| Parameter name | Parameter values |
|---|---|
| Training time | 1.62s |
| Data slicing | 0.7 |
| Data shuffling | 是 |
| Cross-validation | 5 |
| Base learners | gbtree |
| Number of base learners | 100 |
| Learning rate | 0.1 |

| | |
|---|---|
| L1 regular terms | 0 |
| L2 regular term | 1 |
| Sampling rate of sample signatures | 1 |
| Tree feature sampling rate | 1 |
| Node feature sampling rate | 1 |
| Minimum weight of samples in leaf nodes | 0 |
| Maximum depth of the tree | 10 |

4.2.1 d1 model parameters

| Parameter name | Parameter values |
|---|---|
| Training time | 1.62s |
| Data slicing | 0.7 |
| Data shuffling | 是 |
| Cross-validation | 5 |
| Base learners | gbtree |
| Number of base learners | 100 |
| Learning rate | 0.1 |
| L1 regular terms | 0 |
| L2 regular term | 1 |
| Sampling rate of sample signatures | 1 |
| Tree feature sampling rate | 1 |
| Node feature sampling rate | 1 |
| Minimum weight of samples in leaf nodes | 0 |
| Maximum depth of the tree | 10 |

4.2.1 D2 model parameters

Graph description:

The above table shows the configuration of each parameter of the model and the training time of the model.

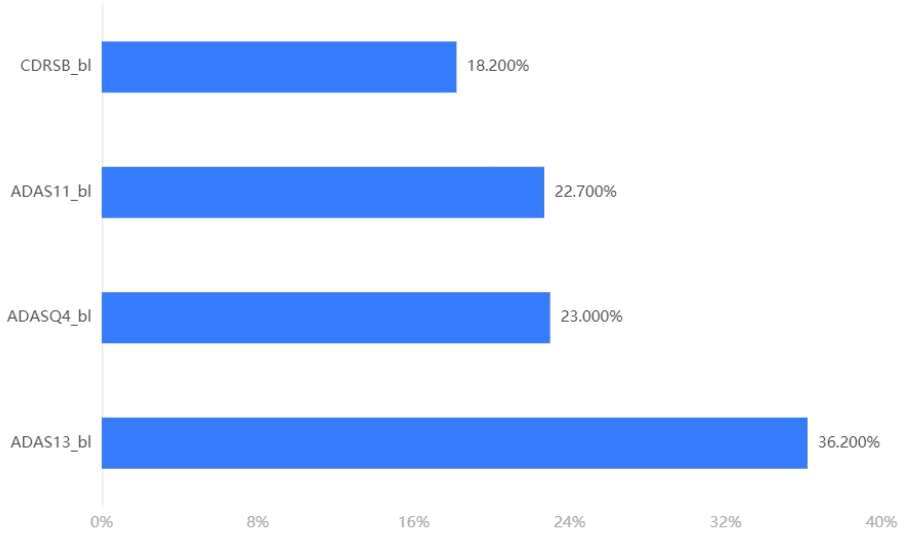
### 4.2.2 The feature importance is calculated by the established XGBoost.

Algorithm Introduction

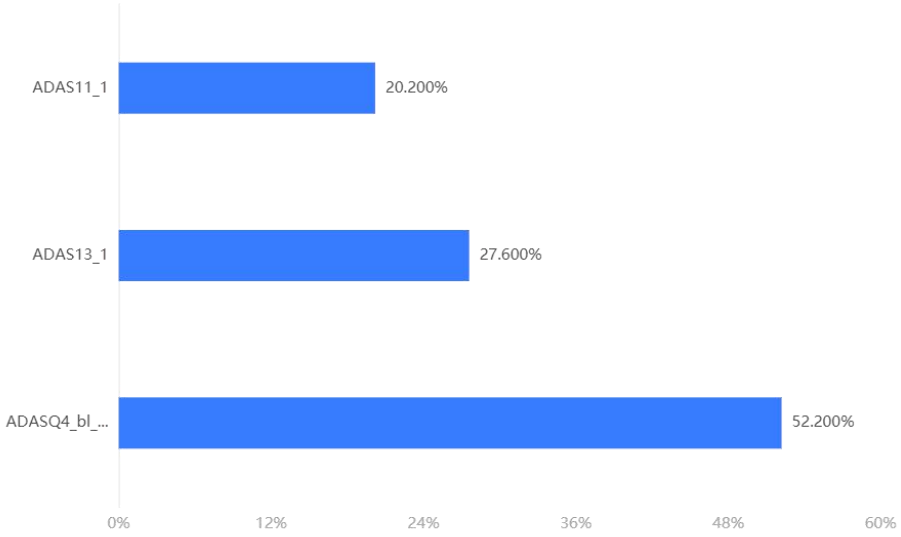
XGBoost is an optimized distributed gradient enhancement library designed to be efficient, flexible and portable. It implements machine learning algorithms in the

XGBoost provides parallel tree boosting (also known as GBDT, GBM). XGBoost is an improvement of the gradient boosting algorithm by solving the extreme value of the loss function using Newton's method, Taylor expansion of the loss function to the second order, and additionally adding a regularization term to the loss function. The objective function during training consists of two parts, the first part is the gradient boosting algorithm loss, and the second part is the regularization term.

Feature importance




4.2.2 d1Characteristic importance




4.2.2 D2Characteristic importance

**Graph description:**

The upper bar chart or table shows the proportion of importance of each characteristic (independent variable).

### 4.2.3 The XGBoost regression model developed to obtain the model evaluation results

Model evaluation results:

| | MSE | RMSE | MAE | MAPE | $R^2$ |
|---|---|---|---|---|---|
| Training set | 0.112 | 0.334 | 0.24 | 74.159 | 0.747 |
| Cross-validation set | 0.56 | 0.748 | 0.599 | 473.881 | -0.267 |
| Test set | 0.528 | 0.727 | 0.579 | 252.859 | -0.253 |

Graph of model evaluation results1

| | MSE | RMSE | MAE | MAPE | $R^2$ |
|---|---|---|---|---|---|
| Training set | 0.27 | 0.519 | 0.423 | 105.588 | 0.334 |
| Cross-validation set | 0.41 | 0.64 | 0.524 | 226.699 | -0.02 |
| Test set | 0.445 | 0.667 | 0.549 | 234.007 | 0.008 |

4.2.3 Graph of model evaluation results2

Graph description:

The above table shows the prediction evaluation metrics of the cross-validation set, training set and test set to measure the prediction effectiveness of XGBoost through quantitative metrics. Among them, the evaluation metrics of the cross-validation set can continuously adjust the hyperparameters to obtain a reliable and stable model.

● MSE (Mean Square Error): The expected value of the squared difference between the predicted and actual values. The smaller the value, the higher the accuracy of the model.

● RMSE (Root Mean Square Error): The square root of MSE, the smaller the value, the more accurate the model.

MAE (Mean Absolute Error): The average of the absolute errors, which reflects the actual situation of the prediction errors. The smaller the value, the higher the accuracy of the model.

MAPE (Mean Absolute Percentage Error): A variation of MAE, which is a percentage value. The smaller the value, the higher the accuracy of the model.

$R^2$: The closer the predicted value is to 1, the more accurate the model is compared to the case where only the mean value is used.

### 4.2.4 Due to the random nature of XGBoost, uploading data for computational prediction

Test data prediction results:

| Predicted outcome Y | APOE4 | CDRSB_bl | ADAS11_bl | ADASQ4_bl | ADAS13_bl |
|---|---|---|---|---|---|
| 0.865118145942688 | 1 | 6 | 33 | 10 | 45 |
| 0.29343461990356445 | 0 | 0 | 4 | 3 | 7 |

| | | | | | |
|---|---|---|---|---|---|
| 0.6995174288749695 | 0 | 0 | 5 | 5 | 11 |
| 0.08185886591672897 | 0 | 0.5 | 3.33 | 2 | 5.33 |
| 0.07391554117202759 | 1 | 1 | 4 | 5 | 9 |
| 0.985584020614624 | 1 | 0 | 10.67 | 3 | 14.67 |
| 0.24743741750717163 | 0 | 0 | 6.67 | 2 | 8.67 |
| 0.7881326675415039 | 1 | 3 | 20 | 10 | 31 |
| 1.104787826538086 | 0 | 2 | 13 | 5 | 18 |
| 0.7722784280776978 | 2 | 1.5 | 2.67 | 1 | 3.67 |
| 0.8947102427482605 | 0 | 0 | 5 | 5 | 10 |
| 0.866416335105896 | 0 | 1.5 | 12.33 | 5 | 20.33 |
| 0.12063263356685638 | 0 | 1 | 6 | 0 | 6 |
| 0.22346942126750946 | 0 | 0 | 6.33 | 5 | 11.33 |
| 0.8627852201461792 | 0 | 3 | 13 | 10 | 24 |

4.2.4 d1Test data prediction results

| Predicted outcome | YAPOE4 | ADAS11_1 | ADAS13_1 | ADASQ4_bl_1 |
|---|---|---|---|---|
| 0.30308660864830017 | 1 | 5 | 10 | 4 |
| 0.2373480200767517 | 1 | 5 | 9 | 4 |
| 0.36702221632003784 | 1 | 10.875483558141616.7348150124511 | | 4 |
| 0.26444828510284424 | 1 | 6 | 12 | 4 |
| 0.3735928535461426 | 1 | 4 | 7 | 4 |
| 0.36702221632003784 | 1 | 10.875483558141616.7348150124511 | | 4 |
| 0.31770047545433044 | 1 | 3 | 7 | 4 |
| 0.02031506411731243 | 1 | 1 | 1 | 2 |
| 0.2688102126121521 | 1 | 4 | 6 | 2 |
| 0.3052016794681549 | 1 | 10.875483558141616.7348150124511 | | 2 |
| 0.22048260271549225 | 1 | 2 | 2 | 2 |
| 0.3938906788825989 | 1 | 4 | 4 | 2 |
| 0.3052016794681549 | 1 | 10.875483558141616.7348150124511 | | 2 |
| 0.338725745677948 | 1 | 3 | 5 | 2 |
| 0.3052016794681549 | 1 | 10.875483558141616.7348150124511 | | 2 |

4.2.4 d2Test data prediction results

Chart Description.

The above table shows the preview results, only some data are shown, please click the download button to export the full data.

The above table shows the predictions of XGBoost on the test data

Test Data Prediction Chart:


4.2.4 d1Test data prediction


4.2.4 d2Test data prediction

Graph description.

The above graph shows the predictions of XGBoost on the test data.

## 4.3 Modeling and solving Problem 3

First of all, all the data are quantified and unified, and the quantified data are subjected to cluster analysis. In the clustering algorithm is to use the K-MEANS clustering algorithm, calculate the Euclidean distance, and according to the minimum distance to the three subclasses contained in MCI (SMC, EMCI and LMCI), the corresponding objects are refined into three subclasses, and the central object of

each cluster is recalculated until each clusters no longer change.

### 4.3.1Quantization of all data

**Introduction to the algorithm:**

The purpose of dimensionalization is to standardize the data in terms of units, some of which have practical significance, such as minimization, maximization, averaging, standardization, etc.; they represent data divided by the mean, data divided by the first number, data divided by the minimum, data divided by the maximum, data divided by the summation, data divided by the sum of squares, and standardized data with a mean of 0 and a standard deviation of 1.

The common method used in K-MEANS clustering to reveal the similarity between data is expressed by Euclidean distance.

It is defined as:

$$d_{ij}=\sqrt{\left|x_{1i}-x_{1j}\right|^2+\left|x_{2i}-x_{2j}\right|^2+\left|x_{3i}-x_{3j}\right|^2+\left|x_{4i}-x_{4j}\right|^2}$$

Descriptive statistics

| | N | Min | Max | Mean | Std Deviation |
|---|---|---|---|---|---|
| RID_1 | 1284 | 7 | 7092 | 3982.20 | 2368.936 |
| AGE_1 | 1284 | 54.4 | 90.1 | 73.928 | 7.3609 |
| ABETA_1 | 1284 | 210.90 | 1681.00 | 754.5026 | 161.23296 |
| ADAS13_1 | 1284 | .00 | 72.00 | 20.2839 | 9.03696 |
| CDRSB_1 | 1284 | .0 | 16.0 | 2.307 | 1.9910 |
| ADAS11_1 | 1284 | .00 | 57.00 | 12.7530 | 6.78729 |
| MMSE_1 | 1284 | 8.0 | 30.0 | 26.032 | 3.4100 |
| LDELTOTAL_1 | 1284 | .0 | 20.0 | 5.483 | 3.8563 |
| DIGITSCOR_1 | 1284 | .0 | 70.0 | 36.374 | 11.8738 |
| TRABSCOR_1 | 1284 | .0 | 300.0 | 135.034 | 78.3502 |
| 有效个案数（成列） | 1284 | | | | |

4.3.1 Descriptive statistical chart after pre-processing

### 4.3.2 Clustering analysis of K-means algorithm after quantization

**Algorithm introduction:**

1. Randomly select a sample as the first initial point

2. Calculate the shortest distance between each sample and the current existing clustering center, the larger the value, the greater

the probability of being selected as the probability center of the clustering center, and finally use the roulette wheel method to select the next clustering center

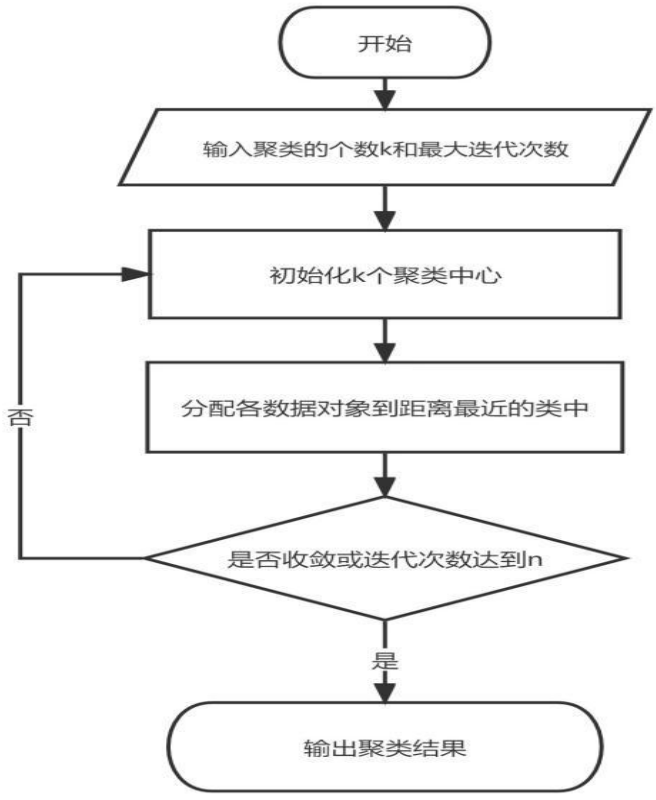
3. Repeat 2 until k clustering centers are selected, the initial point is selected, and the k-means algorithm is continued.

4. randomly select K data objects as the initial clustering centers (not necessarily our sample points) ;,

5. Calculate the distances of the remaining data objects to the K initial clustering centers, and assign the data objects to the cluster class of the center closest to it;

6. Adjust the new classes and recalculate the centers of the new classes;

7. Loop through steps 3 and 4 to see if the centers converge (no change), and stop the loop if they converge or if the number of iterations is reached.




K-Means Algorithm flow chart 4.3.2

K-Means clustering results:

Initial Clustering Centre

| | Clustering | | |
|---|---|---|---|
| | LMCI | SMC | EMCI |
| Zscore(RID_1) | .43513 | .48452 | .94296 |
| Zscore(AGE_1) | .77054 | .09128 | .39015 |
| Zscore(PIB_1) | 11.47127 | -.03702 | -4.42108 |
| Zscore(ABETA_1) | .07087 | -1.34776 | 2.75066 |
| Zscore(TAU_1) | -.05611 | 8.18241 | -1.75524 |
| Zscore(PTAU_1) | -.05809 | 9.68032 | -1.78830 |
| Zscore(ADAS13_1) | 1.29646 | -.17859 | .33707 |
| Zscore(CDRSB_1) | .09701 | -.65638 | .85039 |
| Zscore(ADAS11_1) | .77306 | -.15956 | .08501 |
| Zscore(MMSE_1) | -.30260 | -.30260 | .87042 |
| Zscore(LDELTOTAL_1) | -1.42170 | -1.42170 | -.90307 |
| Zscore(DIGITSCOR_1) | 1.65292 | .22119 | .47385 |
| Zscore(TRABSCOR_1) | -.57478 | -.19189 | -.10254 |

**Final Clustering Centre**

| | Clustering | | |
|---|---|---|---|
| | LMCI | SMC | EMCI |
| RID_1 | 1023 | 6553 | 4607 |
| AGE_1 | 74.4 | 73.0 | 74.4 |
| PIB_1 | 1.8423 | 1.8454 | 1.8526 |
| ABETA_1 | 746.75 | 752.77 | 764.88 |
| TAU_1 | 300.68 | 308.49 | 300.23 |
| PTAU_1 | 29.58 | 30.36 | 29.53 |
| ADAS13_1 | 19.85 | 21.20 | 19.82 |
| CDRSB_1 | 2.1 | 2.5 | 2.3 |
| ADAS11_1 | 12.52 | 13.28 | 12.47 |
| MMSE_1 | 26.0 | 25.8 | 26.3 |
| LDELTOTAL_1 | 5.4 | 5.5 | 5.6 |
| DIGITSCOR_1 | 36.8 | 35.9 | 36.4 |
| TRABSCOR_1 | 129.5 | 142.2 | 133.7 |

Using the weighted Euclidean distance of K . The mean clustering method was used to cluster the data in the Appendix into three categories, LMCI, SMC, and EMCI. The clustering results are given in Table 4.

**Number of cases in each cluster**

| | | |
|---|---|---|
| cluster | LMCI | 453.000 |
| | SMC | 422.000 |
| | EMCI | 409.000 |
| Valid | | 1284.000 |
| Missing | | .000 |

## 4.4 Modeling and solving Problem 4

A time series, also known as a dynamic series, is a sequence of values of indicators of a phenomenon in chronological order. Time series analysis can be broadly divided into three main parts, which are describing the past, analyzing the law and predicting the future. In this problem, we use the seasonal decomposition model to mathematically model the relevant variables after our pre-processing in the Appendix. Firstly, the time series is decomposed into long-term trend of change, seasonal change pattern, cyclical change pattern, and irregular change (random disturbance term) which are four kinds of changes with mutual influence relationship, then the product model should be used:

$$Y=T*S*C*I$$

Symbol Description:

| Symbols | Description of symbols | Remarks |
|---|---|---|
| $Y$ | Final change in indicator value | |
| $T$ | Long-term trend change | |
| S | Seasonal variation | |
| $C$ | Cyclical variation | |
| $I$ | Irregular variation | |

Illustrative diagram of symbols

### 4.4.1Data pre-processing

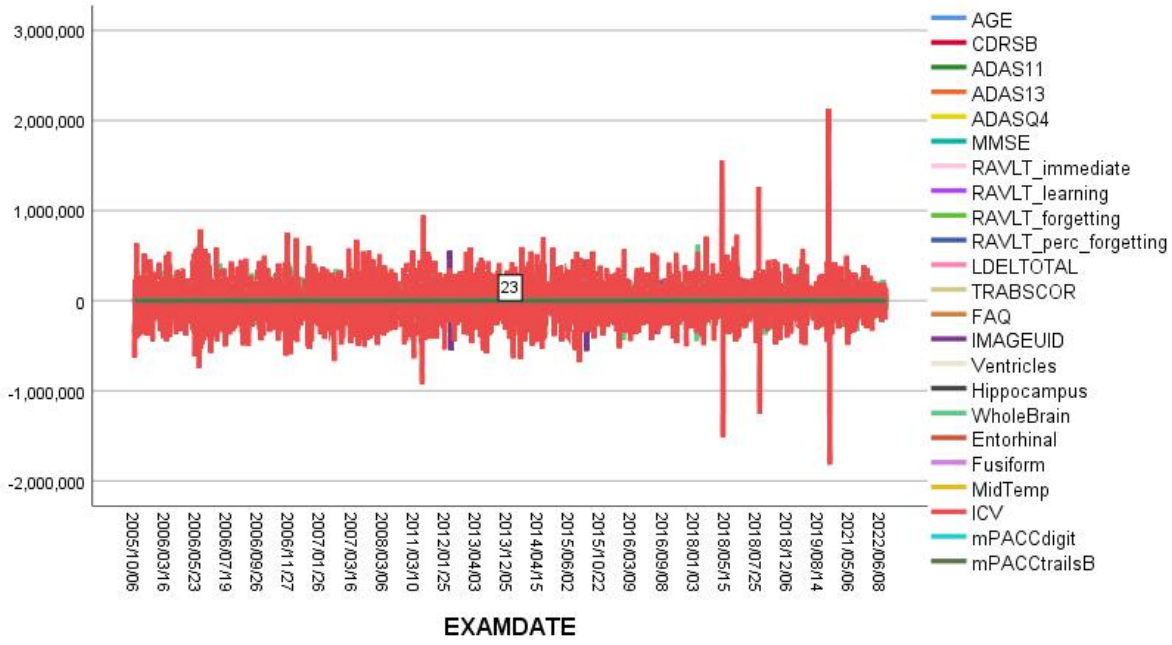
Since spss cannot handle variables of character types, dummy variables are created for gender and person types

**Variable creation**

| | Tags |
|---|---|
| Female_1 | PTGENDER=Female |
| Female_2 | PTGENDER=Male |

4.4.1 Variable creation diagram

### 4.4.2Create time variables and draw time series graphs




转换：季节性差异(1，周期 4)

## 4.4.3Perform time series analysis

**Model description**

| | | |
|---|---|---|
| Model name | | MOD_3 |
| Model type | | Additivity |
| Sequence | 1 | YEAR, not periodic |
| name | 2 | QUARTER, period 4 |
| Seasonal cycle length | | 4 |
| Calculation of the moving average | | Span equal to cycle length plus 1 and endpoints weighted by 0.5 |

Model specifiers from MOD_3 are being applied

4.4.3Model description diagram

**Seasonal factors**

| Series Name | Period | Seasonal Factor (%) |
|---|---|---|
| YEAR, not periodic | 1 | 100.0 |
| | 2 | 100.0 |
| | 3 | 100.0 |
| | 4 | 100.0 |

| | | |
|---|---|---|
| QUARTER, period 4 | 1 | 120.0 |
| | 2 | 160.0 |
| | 3 | 40.0 |
| | 4 | 80.0 |

The seasonal factor of 1.2 for the first quarter indicates that the probability of developing Alzheimer's disease in the first quarter is 1.2 times the average seasonal probability, the probability of developing Alzheimer's disease in the second quarter is 1.6 times the average seasonal probability, the probability of developing Alzheimer's disease in the third quarter is 0.4 times the average seasonal probability, and the probability of developing Alzheimer's disease in the fourth quarter is 0.8 times the average seasonal probability. After analysis, the first and second quarters were found to be more likely to develop Alzheimer's disease.

## 四、Problem Analysis

### 5.1.1Early intervention

**Cognitively normal older adults (CN):**

(1) Maintain normal weight: Obesity in midlife increases the risk of Alzheimer's disease and cognitive dysfunction, but appropriate obesity in older adults can protect cognitive function.

(2) Use your brain more: To maintain normal function, it is important to take precautions, in addition to replenishing the necessary nutrients, and to provide stimulation and training to stimulate brain cells.

(3) Avoid head injury: people with traumatic brain injury are prone to Alzheimer's disease, and having stroke and brain atrophy can also affect their intelligence and reduce cognitive function.

(4) Protect your hearing: Studies have shown that people with hearing loss have more than twice the risk of developing Alzheimer's disease than normal people.

(5) Control blood pressure and blood sugar: With good control of blood pressure, the risk of brain damage is reduced. In addition, higher fasting insulin levels are associated with decreased language and memory.

**Subjective memory complaints (SMC):**

Community health care providers need to focus on the emotional problems of older adults in the community in dementia prevention and actively guide them to participate in community activities to reduce depression. Go to the hospital for cognitive screening or go to a specialist clinic for further examination to avoid the possible influence of psychosocial stress and mental anxiety that may interfere

with the memory process as well as psychosocial factors, genetic susceptibility and their interaction.

**Early Mild Cognitive Impairment (EMCI):**

Non-pharmacological interventions: mainly include moderate physical exercise, life behavior interventions, cognitive training, socialization and some educational activities

Pharmacological interventions: Folic acid and vitamin B12 supplementation for MCI caused by folic acid and vitamin B12 deficiency; hormone replacement therapy for MCI caused by hypothyroidism; active treatment for MCI caused by stroke to minimize the sequelae of cognitive impairment; vitamin B1 supplementation for MCI caused by alcoholism; and cholinesterase inhibitors for patients with indicators of AD and DLB. Cholinesterase inhibitors and other drugs can be tried, but individualized regimens should be implemented and monitored for efficacy.

**Late Mild Cognitive Impairment (LMCI):**

Pharmacological treatment has limited intervention in patients with LMCI. A large randomized, double-blind, placebo-controlled study using ginkgo biloba preparations for MCI found that ginkgo biloba preparations had a mild effect on delaying memory decline in normal elderly people but did not inhibit the conversion of MCI to dementia; eight randomized, double-blind, placebo-controlled studies of cholinesterase inhibitors for MCI (carbapenems) over a period of six months to four years overwhelmingly showed that these drugs did not reduce the conversion of MCI to dementia. Only one trial suggested that donepezil had a lower conversion rate than the control group during the initial 12 months of intervention, but there was no difference in the conversion rate between the two groups at the end of 3 years.

**Alzheimer's disease (AD):**

ChEIs may be used for treatment. If treatment with a particular cholinesterase inhibitor is not effective or is not tolerated due to adverse effects, the patient may be switched to another ChEIs or to a patch for treatment, depending on the patient's condition and the degree of adverse effects, and the patient should be closely observed for possible adverse effects during treatment. After explaining the benefits and possible risks of treatment to the patients, Ginkgo biloba, cerebroprotein hydrolysate, olanzapine or piracetam can be used as synergistic adjuvant drugs for AD patients.

### 5.1.2 Diagnostic criteria

**Cognitively normal elderly (CN):**

Normal memory, quick thinking, competent, normal life skills, no personality change, no difficulty in reading, smooth.

**Subjective memory complaints (SMC):**

SMC older adults have lower overall levels of cognitive function, mainly in the cognitive domains of abstraction, delayed memory,

visuospatial and executive function, and language; SMC has a lower risk of developing dementia within three years, suggesting that SMC provides the best window of time for early treatment of dementia.

Early Mild Cognitive Impairment (EMCI):

Ancillary tests that enable cognitive impairment disorders include body fluid tests, imaging tests, electrophysiological tests and genetic tests. The selection of appropriate ancillary tests can effectively assist in the diagnosis and differential diagnosis of cognitive impairment disorders and monitor the disease process. Impairment of cognition reported by patients or informed persons, or detected by experienced clinicians; objective evidence of impairment in one or more domains of cognitive function exists (from cognitive tests); complex instrumental daily abilities can be slightly impaired but maintain independent daily living abilities; progressive decrements in memory or other cognitive functions, but do not affect daily living abilities and have not reached the diagnosis of dementia.

Late Mild Cognitive Impairment (LMCI).

The etiological diagnosis of MCI is made by combining the onset and progression of LMCI, features of cognitive impairment, history and signs of the presence or absence of neurological primary disease, psychiatric disease (or stressful events) or systemic disease, and necessary ancillary tests For patients with a current diagnosis of MCI, at least 1 year of follow-up is recommended to further clarify the diagnosis.

Alzheimer's disease (AD).

AD is divided into 3 stages, namely preclinical stage of AD, mild cognitive impairment of AD origin and dementia stage of AD, and the clinical diagnosis of AD can be made according to the NINCDS-ADRDA of 1984 or the AD diagnostic criteria proposed by the NIA-AA of 2011. When molecular imaging and cerebrospinal fluid testing of AD are available, AD diagnosis can be made according to the 2011 NIA-AA or the 2014 IWG-2 diagnostic criteria.